Identity and Identification in an Information Society

Augmenting Formal Systems of Identification with Technological Artefacts

by

WILL ABRAMSON

Thesis submitted in partial fulfilment of the requirements

of Edinburgh Napier University, for the award of

Doctor of Philosophy

School of Computing

Edinburgh Napier University

AUGUST 2022

Dedication

This thesis is dedicated to past, present and future healthcare professionals around the world in recognition of their selfless commitment to the care of others and the dedication, integrity and compassion with which they perform this trusted role in society. I hope that the future of healthcare integrates digital information and communication technologies that empower these trustworthy individuals, rather than burdening or displacing them.

Author's declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where states otherwise by reference or acknowledgment, the work presented is entirely my own.

SIGNED: ..... DATE: .....

Acknowledgements

Undertaking this PhD has been daunting, exhilarating, stressful, inspiring, humbling, exhausting and joyful at different moments throughout the journey. The challenge of persisting with something over years. The sense of awe and fascination that comes when diving deep into a branch of human knowledge. The thrill of experiencing your mind bursting with ideas and possibilities. The struggle of converting these into a coherent, consistent whole, whilst fighting the fear, uncertainty and doubt that bubbles up and threatens to overwhelm on a regular basis. Building resilience. Learning to trust myself.

A PhD is often framed as an individual pursuit of, and contribution to, knowledge. Although, as with most things in life, to consider the individual individually overlooks much of the complexity that enables them to thrive. The environment, interactions, relationships, circumstances and serendipity. Certainly, my personal experience has been full of these elements and without them, it is doubtful that you would be reading these words today. This is my attempt to acknowledge, praise and recognise all those who made this work possible.

First I must thank Blockpass, who had the foresight to fund the research lab at Edinburgh Napier, and my PhD position within it. In doing so they made a contribution to science that I have been the grateful beneficiary of. I have huge respect for the way in which they did this, freely and without an expectation of anything in return. They placed trust in my Director of Studies, and through him, in me. I hope this thesis is worthy of that trust.

The Blockpass Identity Lab at Merchiston Campus, Edinburgh Napier University provided me with a rich, stimulating environment populated with a diverse range of people who I could learn from and collaborate with. Despite not setting foot in this lab much since the start of the pandemic, I felt part of a research culture that helped to nurture my skills and refine my thoughts. Thanks to everyone who contributed to this culture. In particular thanks to my fellow research students with whom I have shared parts of this journey; Nilu, Marwan, Pavlos and Adam. I am confident that this research lab and its alumni will play an important role in shaping the potential futures of our information society.

To my Director of Studies, Professor William J Buchanan, thank you for giving me the freedom to direct my own research, believing in my ability and always trusting that I had a contribution to make. Even when I did not always trust myself. Thanks for all the nudges and support that you provided. Dr Owen Lo, as my second supervisor, I always valued your input and am grateful for the diligent and timely review of any of the work I

showed you. Thank you both for guiding me through this process.

Dr Manreet Nijjar and Dr Henry Goodier, my time working with you at truu provided me with a concrete system to study and the first-hand experience through which to study it. I am grateful that you shared with me your the wealth of domain-specific knowledge gained through your lived experience interacting with the NHS as a health-care professional. It is because of the tireless work of Manny and Henry that the case study presented in this thesis has a very real possibility of being integrated into U.K. Healthcare systems. I applaud and respect your tireless dedication to the healthcare profession. Thank you.

I must also thank the Royal College of Physicians of Edinburgh and in particular Professor Derek Bell and Pernille Marqvardsen. Being able to run a workshop at this prestigious institution with a diverse range of healthcare professionals and key stake-holders helped to contextualise the existing identification processes and pain points that are regularly encountered. Thanks to everyone who gave up their valuable time to share your wisdom with me. Pernille, without your coordination efforts this workshop would not have been possible. Bill, thank you for putting me in touch with the RCPE and with Dr Nicole van Deursen who planned and co-facilitated the day with me. Nicole, it was an absolute pleasure to collaborate with you on this workshop. Your input and experience were immensely valuable.

While naming all the places, people and spaces that were influential throughout the course of my studies, I must mention a few more. Dr Nick Spencer and Nicky Hickman, I thoroughly enjoyed collaborating with you both during our tentative exploration of SSI metrics - our differences in perspective and styles of thought were hugely beneficial. Ian Barclay, thanks for introducing me to iStar. Andrew Trask, I thank you for creating such a vibrant and welcoming community at Open Mined. Adam Hall, thank you for inviting me to participate in Open Mined with you, we produced some great work together. Lohan Spies, I always enjoyed our late night hacking, even if they were sometimes a distraction. Philip Sheldrake, I am incredibly grateful for your effort in prodding, poking and challenging those working on decentralised identity to consider the wider, messier and infinitely more complex human systems of interaction that these technologies are developed within and applied to. It is thanks to you that I got lost for months down social rabbit holes, immersed in the likes of Bateson, Luhmann and others. I think this thesis is better for it. Or at least it is different, and for me that is to be celebrated. Finally, Edinburgh, Scotland thank you for being a beautiful part of the world that has contributed greatly to both my creativity and mental well-being.

Of course I must also thank all my family and friends. All those who showed interest in my work, but also those who provided me with a much needed distraction from it. A special shout out goes to Moritz, who provided a friendly review for much of my early iterations and ideas.

This thesis would not have been possible without my parents who read through the entirety of the thesis between them and have provided insightful feedback and much needed encouragement throughout the course of my journey. George, your input towards the end of this thesis kept me going and undoubtedly helped me to realise a better version of this work.

Finally, my heartfelt gratitude goes to my darling Kathy, the bringer of teas, who has had to put up with my rollercoaster of emotions that has defined the many bumps along this journey. Kathy, without your love, care and support I am not sure I would have made it to the finish line. I feel blessed to have you by my side.

Abstract

Information and Communication Technology (ICT) are transforming society's information flows. These new interactive environments decouple agents, information and actions from their original contexts and this introduces challenges when evaluating trustworthiness and intelligently placing trust.

This thesis develops methods that can extend institutional trust into digitally enhanced interactive settings. By applying privacy-preserving cryptographic protocols within a technical architecture, this thesis demonstrates how existing human systems of identification that support institutional trust can be augmented with ICT in ways that distribute trust, respect privacy and limit the potential for abuse. Importantly, identification systems are located within a sociologically informed framework of interaction where identity is more than a collection of static attributes.

A synthesis of the evolution and systematisation of cryptographic knowledge is presented and this is juxtaposed against the ideas developed within the digital identity community. The credential mechanism, first conceptualised by David Chaum, has matured into a number of well specified mathematical protocols. This thesis focuses on CL-RSA and BBS+, which are both signature schemes with efficient protocols that can instantiate a credential mechanism with strong privacy-preserving properties.

The processes of managing the identification of healthcare professionals as they navigate their careers within the Scottish Healthcare Ecosystem provide a concrete case study for this work. The proposed architecture mediates the exchange of verifiable, integrity-assured evidence that has been cryptographically signed by relevant healthcare institutions, but is stored, managed and presented by the healthcare professionals to whom the evidence pertains.

An evaluation of the integrity-assured transaction data produced by this architecture demonstrates how it could be integrated into digitally augmented identification processes, increasing the assurance that can be placed in these processes. The technical architecture is shown to be practical through a series of experiments run under realistic production-like settings.

This work demonstrates that designing decentralised, standards-based, privacy-preserving identification systems for trusted professionals within highly assured social contexts can distribute institutionalised trust to trustworthy individuals and empower these individuals to interface with society's increasingly socio-technical systems.

TABLE OF CONTENTS

DEDICATION i

AUTHOR'S DECLARATION ii

ACKNOWLEDGEMENTS iii

ABSTRACT v

TABLE OF CONTENTS vi

LIST OF TABLES xvii

LIST OF FIGURES xix

1 INTRODUCTION 1

1.1 Public Key Cryptography 5

1.2 Artefacts 6

1.3 Thought Collectives 9

1.4 Research Questions 12

1.5 Thesis Contributions 12

1.5.1 Publications 13

1.6 Thesis Structure 14

2 IDENTITY, TRUST AND PRIVACY IN AN INFORMATION SOCIETY 16

2.1 Human Identity 17

2.1.1 Individuals . . . 18
2.1.2 Interaction . . . 19
2.1.3 Groups . . . 21
2.1.3.1 Teams . . . 21
2.1.3.2 Categories . . . 22
2.1.4 Identity . . . 22
2.1.5 Social Systems and Structure . . . 24
2.1.6 Identity and Power . . . 25
2.1.7 Identity and Change . . . 26
2.1.8 Identity in an Information Society . . . 27
2.2 Trust . . . 29
2.2.1 Interpersonal Trust . . . 30
2.2.2 Systems Trust . . . 33
2.2.3 Trust, Risk and Assurance . . . 34
2.2.4 Trust in an Information Society . . . 35
2.2.4.1 The impact of Information Technologies on Trust . . . 36
2.2.4.2 Designing Technologies for Trust . . . 37
2.3 Privacy . . . 41
2.3.1 Privacy Throughout History . . . 42
2.3.2 Defining Privacy . . . 43
2.3.2.1 Privacy in Private and in Public . . . 44
2.3.2.2 Privacy as Contextual Integrity . . . 45
2.3.3 The Value of Privacy . . . 46
2.3.3.1 Privacy and Individual Autonomy . . . 46
2.3.3.2 Privacy and Social Relationships . . . 47
2.3.3.3 Privacy and Society . . . 47
2.3.4 Privacy in an Information Society . . . 48
2.3.4.1 Surveillance Capitalism . . . 49
2.3.4.2 Privacy-enhancing Technologies . . . 52
2.3.4.3 Changing Norms . . . 53

2.4 Critical Discussion 54
2.5 Conclusion 55

3 IDENTIFICATION SYSTEMS 57

3.1 Terminology 60
3.1.1 Legal and Virtual Persons and Identities 62
3.1.2 Privacy and Data Minimisation 64
3.2 Digital Identification Systems 65
3.3 Decentralised Identification Systems 73
3.3.1 Roles, Interactions and Infrastructure 73
3.3.2 Open Standards 75
3.3.3 Digital Agents and Wallets 77
3.3.4 Machine-Readable Governance 78
3.4 Lessons from Identification Systems for Development 81
3.4.1 Privacy and Surveillance 82
3.4.2 Exclusion of Individuals and Groups 84
3.4.3 Accuracy, Misrepresentation and Falsification 85
3.4.4 Complexity of Identification Systems 85
3.4.5 Power and Abuse 87
3.4.6 Governance of the Identification System 87
3.5 Critical Discussion 88
3.6 Conclusion 89

4 SECURITY WITHOUT IDENTIFICATION 90

4.1 Laying the Foundations 96
4.1.1 Early Ideas and Concepts 96
4.1.2 Key Cryptographic Primitives 100
4.1.2.1 Digital Signatures 101
4.1.2.2 Hash Functions 102
4.1.2.3 Cryptographic Commitments 103
4.1.2.4 Zero-Knowledge Proofs of Knowledge 103

4.1.2.5 Cryptographic Accumulators 105
4.1.3 Understanding the Mathematical Setting 105
4.1.4 The Security of Cryptographic Primitives and Protocols 109
4.1.5 Credential Mechanisms 1976-2001 110
4.2 Protocols for a Credential Mechanism from a Signature Scheme with Efficient Protocols 113
4.2.1 Definitions 114
4.2.2 Pseudonym Generation 115
4.2.3 Credential Issuance 117
4.2.4 Credential Presentation 120
4.2.5 Supporting Protocols 123
4.2.5.1 Revocation 123
4.2.5.2 Delegation 124
4.2.5.3 Accountability 125
4.2.5.4 Linkability 127
4.3 From Theory to Practice 128
4.3.1 Implementing Cryptographic Algorithms 129
4.3.2 Key Management 131
4.3.3 System Architecture 133
4.4 State-of-the-Art 136
4.4.1 Pairing Cryptography 137
4.4.2 Universally Composable Security 138
4.4.3 Credential Mechanism Constructions 139
4.4.3.1 A Signature Scheme with Efficient Protocols 140
4.4.3.2 Message Authentication Codes 141
4.5 Critical Discussion 141
4.6 Conclusions 143

5 DEFINING A TECHNICAL ARCHITECTURE 145

5.1 Design principles for digital identification systems 145

5.2 Requirements for Credential-based Identification Architectures 148
5.2.1 The Credential Data Model 151
5.2.2 Identifiers 151
5.2.3 Protocols 152
5.2.3.1 Transport and Messaging 152
5.2.3.2 Credential Issuance 153
5.2.3.3 Credential Verification 153
5.2.3.4 Credential Revocation 153
5.2.4 Key Management 154
5.2.5 Cryptographic Signature Suites 154
5.2.6 Data Storage 155
5.2.7 Recovery and Backup 155
5.2.8 Schema Management 155
5.2.9 Complexity / Ease of Use 156
5.2.10 Adoption 156
5.2.11 Transparency and Governance 157
5.3 Evaluation of Existing Implementations 157
5.4 The Hyperledger Verifiable Information Exchange Platform (HVIEP) . . 159
5.4.1 Indy-based Distributed Ledgers 161
5.4.2 Aries Agent Architecture 164
5.4.2.1 Aries Agents Facilitate Peer to Peer Interaction 164
5.4.2.2 Aries Agents Understand A Specific Set of Protocols . . 165
5.4.2.3 Aries Agents are Event-Driven 166
5.4.2.4 Aries Agents Interface with Ursa and Indy 166
5.5 Conclusion and Critical Discussion 167

6 AN IDENTIFICATION SYSTEM FOR PROFESSIONALS WITHIN THE SCOTTISH HEALTHCARE ECOSYSTEM 170

6.1 The Healthcare Context 171
6.1.1 The Scottish Healthcare Ecosystem 174

6.2 Engaging With Healthcare Professionals through an Interactive Workshop 175
6.2.1 Workshop Methodology 176
6.2.1.1 Healthcare Ecosystem Process Mapping 176
6.2.1.2 Re-imagining Identification Processes using Decentralised Identification Technologies 178
6.2.1.3 Evaluating the Design Principles 180
6.2.2 Results 181
6.3 Modelling the Scottish Healthcare Ecosystem 185
6.3.1 High-level Healthcare System 189
6.3.2 Healthcare Professionals ID Artefacts and Dependencies 190
6.3.3 A Medical Student becoming Junior Doctor 193
6.3.4 Doctor Onboarding 196
6.4 Implementation 200
6.4.1 Scottish Healthcare Ecosystem Proof of Concept 201
6.4.1.1 Medical Student Becoming a Junior Doctor 201
6.4.1.2 Doctor Onboarding 204
6.5 The Aries Jupyter Playground 206
6.6 Conclusion 208

7 EVALUATION 212

7.1 Ledger Footprint 213
7.2 Aries Agent Performance 221
7.2.1 Issue Credential 223
7.2.2 Present Proof 229
7.2.3 Discussion 236
7.3 Cryptographic Protocols 240
7.3.1 Key Generation 241
7.3.2 Credential Issuance 243
7.3.3 Credential Presentation 244
7.3.4 Discussion 245

7.4 Evaluation Limitations 246
7.5 Future Work: A Participatory Evaluation Engaging with Healthcare Professionals 247
7.6 Conclusion 250

8 CONCLUSION 252

8.1 Technologically Augmenting an Existing Identification System 253
8.1.1 Modelling the System 253
8.1.2 The Hyperledger Verifiable Information Exchange Platform . . . 255
8.1.3 Schema Design and its Impact on Performance 256
8.1.4 Analysing the Ledger Footprint 256
8.1.5 Cryptographic Constraints 257
8.1.6 Aries Juypter Playground 259
8.2 The Genesis, Development and Standardisation of Digital Identification 260
8.2.1 Cryptographic Thought 260
8.2.2 Digital Identity Practitioners 261
8.2.3 The Intellectual Interaction between Cryptography and Digital Identity 262
8.2.4 A Pivotal Moment in History 264
8.3 Identification Systems in an Information Society 264
8.3.1 Identity 265
8.3.2 Trust 265
8.3.3 Institutions and Identification 266
8.3.4 Advanced Digital ICTs 267
8.3.5 Asymmetries of Knowledge and Power 268
8.3.6 Structured Transparency 269
8.4 Limitations and Future Work 271
8.4.1 Identification Systems within Professional Contexts 271
8.4.2 The Structure of Cryptographic Thought 273
8.4.3 Social Sciences 274

8.5 Closing Remarks 275

REFERENCES **278**

APPENDIX A NUMBER THEORY **323**

A.1 Sets 323

A.2 Groups 324

A.3 Fields 326

A.4 Elliptic Curve Groups 327

A.5 Bilinear Maps 330

A.6 Integer Factorization and the RSA Group 331

APPENDIX B CRYPTOGRAPHIC PRIMITIVES AND PROTOCOLS **333**

B.1 Hash Functions 333

B.2 Public Key Cryptosystems 334

B.3 Pedersen Commitments 334

B.4 Zero Knowledge Proofs of knowledge 335

B.4.1 The Sigma Protocol 336

B.4.2 An Example 336

APPENDIX C BBS+: A SIGNATURE SCHEME WITH EFFICIENT PROTOCOLS **338**

C.1 Key Generation 338

C.2 Signing 339

C.3 Verification 339

C.4 Signing Committed Messages 340

C.5 Proof of Knowledge of A Signature 341

APPENDIX D EVALUATION OF SOFTWARE FRAMEWORKS FOR CREDENTIAL-BASED IDENTIFICATION SYSTEMS **343**

D.1 Hyperledger Verifiable Information Exchange Platform (HVIEP) 343

D.1.1 Data Model 344

D.1.2 Identifiers . . . 344
D.1.3 Protocol support . . . 345
D.1.3.1 Transport & Messaging . . . 345
D.1.3.2 Credential Issuance . . . 345
D.1.3.3 Credential Presentation . . . 345
D.1.3.4 Credential Revocation . . . 346
D.1.4 Key Management . . . 346
D.1.5 Cryptographic signature suites . . . 346
D.1.6 Data Storage . . . 347
D.1.7 Recovery and Backup . . . 347
D.1.8 Schema Management . . . 347
D.1.9 Complexity / Ease of use . . . 348
D.1.10 Adoption . . . 348
D.1.11 Transparency / Governance . . . 349
D.2 I Reveal My Attributes (IRMA) . . . 349
D.2.1 Data Model . . . 350
D.2.2 Identifiers . . . 350
D.2.3 Protocol Support . . . 351
D.2.3.1 Transport and Messaging . . . 351
D.2.3.2 Credential Issuance . . . 351
D.2.3.3 Credential Presentation . . . 351
D.2.3.4 Credential Revocation . . . 352
D.2.4 Key Management . . . 352
D.2.5 Cryptographic Signature Suites . . . 352
D.2.6 Data Storage . . . 352
D.2.7 Recovery and Backup . . . 353
D.2.8 Schema Management . . . 353
D.2.9 Complexity / Ease of Use . . . 353
D.2.10 Adoption . . . 353
D.2.11 Transparency / Governance . . . 354

D.3 Serto / Veramo 354
D.3.1 Data Model 354
D.3.2 Identifiers 355
D.3.3 Protocol Support 355
D.3.3.1 Transport and Messaging 355
D.3.3.2 Credential Issuance 355
D.3.3.3 Credential Presentation 355
D.3.3.4 Credential Revocation 356
D.3.4 Key Management 356
D.3.5 Cryptographic Signature Suites 356
D.3.6 Data Storage 356
D.3.7 Recovery and Backup 357
D.3.8 Schema Management 357
D.3.9 Complexity / Ease of Use 357
D.3.10 Adoption 358
D.3.11 Transparency / Governance 358

APPENDIX E THE TECHNOLOGICAL AUGMENTATION OF COMPLEX ADAPTIVE HUMAN SYSTEMS 359

E.1 Introduction 359
E.2 Human Systems 360
E.3 Individuals 361
E.4 Interaction 362
E.5 Environment 363
E.6 Technological Impact on the System 363
E.7 Digital Identity Systems 365
E.8 Looking to the Future 366

APPENDIX F IDENTITY AND INTERACTION IN COMPLEX HUMAN SYSTEMS 368

APPENDIX G PUBLICATIONS **381**

LIST OF TABLES

TABLES **Page**

1.1 Academic publications produced as part of this thesis and their associated sections in the thesis 14

4.1 Table of Credential Mechanisms 1976-2001 112

4.2 Comparison of digital signature sizes that achieve 128-bit security level (adapted from [241]) 138

5.1 List of design principles 149

5.2 The list of design principles distilled from the literature 150

7.1 Trimmed Mean Issuance Times for Credentials with Varying Attribute Size 226

7.2 Trimmed Mean Issuance Times for Credentials with Varying Attribute Number 228

7.3 Trimmed Mean Presentation Times for Credentials with Varying Attribute Size 231

7.4 Trimmed Mean Presentation Times for a Single Attribute Presented from a Credential with Varying Attribute Number in Schema 233

7.5 Trimmed Mean Presentation Times for Varying Number of Attributes Disclosed 235

7.6 Trimmed Mean Presentation Times for 5 Attributes from a Varying Number of Credentials 237

7.7 Comparison of CL-RSA and BBS+ Key Generation Times 242

7.8 Comparison of CL-RSA and BBS+ Credential Issuance Times 244

7.9 Comparison of CL-RSA and BBS+ Credential Presentation Times 245

7.10 Stakeholders of the Scottish Healthcare Ecosystem 249

A.1 Generators for Additive Group Modulo 5 325

A.2 Multiplicative Group Modulo 13 with g=2 325

List of Figures

Figures **Page**

1.1 Three orders of Technological Artefacts (adapted from Floridi [9, Chapter 2]) 8

1.2 ICTs Layer of Indirection 9

2.1 Extended Blockchain Engineering Framework [100] 39

2.2 Google Street View Timeline (adapted from the EPIC [136] 50

3.1 Relationships between Real and Virtual Entities (adapted from [173]) 64

3.2 Timeline of Evolution of Digital Identification Systems 69

3.3 Roles, Interactions and Infrastructure for Decentralised Identity (adapted from [34]) 75

3.4 Trust over IP Stack (Taken from [222]) 81

4.1 A timeline of the emergence of and relationships between cryptographic knowledge 94

4.2 Structure and hierarchy of knowledge in cryptography 95

4.3 The relation between abstractions, concrete instantiations and practical implementations within cryptography 96

4.4 Alice registering a pseudonym with Bob and later authenticating against it 117

4.5 Overview of Credential Issuance protocol 119

4.6 Overview of the Credential Presentation protocol 122

4.7 Actors and interactions in Attribute-Based Credential system [30] 134

4.8 Credential Issuance Architecture Diagram (Attribute Based Credentials for Trust (ABC4Trust)) [240] 134

5.1 Hyperledger Verifiable Information Exchange Platform (adapted from hyperledger.org) 161
5.2 Decentralised Identifier (DID) Communication Between Alice and Bob . . 165
5.3 Decentralised Identifier (DID)Comm dependencies [388] 165
5.4 Application Architecture Hyperledger Aries Based Decentralised Identifier (DID)Comm Interaction 167

6.1 Time Estimates for a Healthcare Professional's Identification Interactions . 177
6.2 Royal College of Physicians of Edinburgh (RCPE) Workshop Representation of Healthcare Professional's Digital Wallet after completing a number of career interactions [41] 179
6.3 Principles ranked by importance 181
6.4 Principles individually rated 181
6.5 iStar 2.0 Modelling Language Components [411] 187
6.6 Overview of the Scottish Healthcare Ecosystem (SHE) 188
6.7 Healthcare Professional Credential Dependencies 192
6.8 Healthcare Professional Credential Dependencies Focused on Medical Licensing and Employee Onboarding Evidence 194
6.9 SD View of Medical Student Becoming a Junior Doctor 196
6.10 SR View of Medical Student Becoming a Junior Doctor 197
6.11 SD View of Junior Doctor Onboarding at a Hospital 199
6.12 SR View of Junior Doctor Onboarding at a Hospital 200
6.13 General Medical Council (GMC) Notebook to Establish Connection with Healthcare Professional 209
6.14 Healthcare Professional Accepting Connection with General Medical Council (GMC) 210

7.1 Indy Transaction Relationships Between NYM, SCHEMA and CLAIM_DEF . 216

7.2 Transaction Author Ledger Footprint for an National Health Service (NHS) Staff Passport Pilot Project (Sept 2020) [42] 219
7.3 Transaction Endorser Ledger Footprint for an National Health Service (NHS) Staff Passport Pilot Project (Sept 2020) [42] 220
7.4 Interactions benchmarked between Issuer and Holder while engaging in the issue-credential protocol (Aries Request for Comment (RFC) 0036) [376] . . 223
7.5 Box Plots of Times to Issue Non-Revocable (left) and Revocable (right) Credentials with Varying Attribute Size 225
7.6 Trimmed Mean Issuance Times for Credentials with Varying Attribute Size 226
7.7 Box Plots of Timed to Issue Non-Revocable (left) and Revocable (right) Credentials with Varying Attribute Number 227
7.8 Trimmed Mean Issuance Times for Credentials with Varying Attribute Number228
7.9 Interactions benchmarked between a Verifier and Holder while engaging in the present-proof protocol (Aries RFC 0037) [427] 229
7.10 Box Plots of Times to Present Attribute from Non-Revocable (left) and Revocable (right) Credentials with Varying Attribute Size 231
7.11 Trimmed Mean Presentation Times for Credentials with Varying Attribute Size232
7.12 Box Plots of Times to Present a Single Attribute from Non-Revocable (left) and Revocable (right) Credentials with Varying Attribute Number in Schema 233
7.13 Trimmed Mean Presentation Times for a Single Attribute Presented from a Credential with Varying Attribute Number in Schema 234
7.14 Box Plots of Times to Disclose a Varying Number of Attributes from a Non-Revocable (left) and Revocable (right) Credential 235
7.15 Trimmed Mean Presentation Times for Varying Number of Attributes Disclosed236
7.16 Box Plots of Times to Present 5 Attributes from Varying Number of Non-Revocable (left) and Revocable (right) Credentials 237
7.17 Trimmed Mean Presentation Times for 5 Attributes from a Varying Number of Credentials 238

A.1 Elliptic Curve Graphs - taken from [442] 328

A.2 Elliptic Curve over finite field 191 - taken from [443] 328
A.3 Elliptic Curve point addition - taken from [444] 329

F.1 Process of Experience 369
F.2 Trust Placed in The Present (Based on Luhmann's Constancies and Events [2, pp. 12-20]) 370
F.3 Human Interaction 371
F.4 Identities and Interaction 372
F.5 Collective 373
F.6 Trust and Distrust 374
F.7 The Social and Historical Context that Influences Interaction 375
F.8 Digital Mediation 376
F.9 Interaction Design Space 377
F.10 Technology Design Space 378
F.11 Formal System of Identification 379
F.12 Synthesis of Technologies Built on Scientific Fundamentals, Produced By and Embedded Within Human Systems of Interaction 380

Acronyms

ABC Attribute Based Credential

ABC4Trust Attribute Based Credentials for Trust

ACA-Py Aries Cloud Agent Python

API Application Programming Interface

AWS Amazon Web Services

BBS Boneh-Boyen-Shacham

BLS Boneh-Lynn–Shacham

CA Certificate Authority

CCG Credentials Community Group

CCPA California Consumer Privacy Act

CHAPI Credential Handler Application Programming Interface

CL Camenisch and Lysyanskaya

DBS Disclosure and Barring Service

DID Decentralised Identifier

DIF Decentralized Identity Foundation

DLT Distributed Ledger Technology

ECDSA Elliptic Curve Digital Signature Algorithm

EU European Union

GDPR General Data Protection Regulation

GMC General Medical Council

HTML Hypertext Markup Language

HTTP Hypertext Transfer Protocol

HVIEP Hyperledger Verifiable Information Exchange Platform

ICT Information and Communication Technology

IETF Internet Engineering Task Force

IIW Internet Identity Workshop

IoT Internet of Things

IPFS Interplanetary File System

IRMA I Reveal My Attributes

JSON JavasScript Object Notation

JWS JSON Web Signatures

JWT JSON Web Token

KMS Key Management Service

KVAC Keyed-Verification Anonymous Credentials

MAC Message Authentication Code

MD5 Message Digest Algorithm 5

NES National Health Service Education for Scotland

NHS National Health Service

NIST National Institute of Standards and Technology

NIZK Non-Interactive Zero Knowledge Proof

PKI Public Key Infrastructure

PMQ Primary Medical Qualification

PoC Proof of Concept

RAM Random Access Memory

RCPE Royal College of Physicians of Edinburgh

REST Representational State Transfer

RFC Request for Comment

RFID Radio Frequency Identification

ROM Random Oracle Model

RSA Rivest, Shamir and Adelman

SHA Secure Hash Algorithm

SHE Scottish Healthcare Ecosystem

SSI Self-Sovereign Identity

SSL Secure Sockets Layer

SSO Single Sign On

TCP/IP Transmission Control Protocol/Internet Protocol

TLS Transport Layer Security

UK United Kingdom

URL Uniform Resource Locator

VC Verifiable Credential

VDR Verifiable Data Registry

VP Verifiable Presentation

W3C World Wide Web Consortium

WACI Wallet and Credential Interactions

Web KMS Web Key Management System

XML Extensible Markup Language

ZKP Zero Knowledge Proof

First Chapter

Introduction

Identities and trust are emergent properties of human systems of interaction that exist within and are shaped by their environment [1]. Both are learnt, constructed, projected, applied and continuously evolved as individuals and groups attempt to structure possibilities from experience, select actions and influence the action selection of others through interaction [2]. Identities assign meaning to informational inputs and trust extrapolates positive future expectations from this meaningful information, structuring the perceived possibilities from which actions are selected [2, 3, 4]. Individuals, through the actions they select and their expressions both *given and given off* when observed act as the primary source of information and are the irreducible and indeterminable agents of change within any human social system [5, 6]. Privacy as contextual integrity, articulated by Nissenbaum, fits into this framework of interaction - the perception that information, especially information identifying subjects, has flowed appropriately throughout these human systems [7]. All are highly contextual, complex phenomena. This has been true throughout human history, although it is only during the last century that we have become self-aware of ourselves as goal-seeking adaptive informational agents embedded in complexity, seeking to understand our environment, ourselves, and the other agents we encounter within this environment [5, 8, 9, 10].

Our environment and the potential for complex social interaction within it have been continuously transformed by the introduction, proliferation and integration of ICTs within society. The written word, printing press, postal service, telegram and the

telephone are all early historical examples of these technologies that transformed our ability to produce, disseminate, collect, process, manage, experience and make sense of information at an increasing scale and over an increased distance at ever greater speeds [11, 12]. Informational inputs were no longer dependent on being co-present in the same spatial-temporal coordinates. This reduced our ability to rely on personal relationships and the expressive, redundant signals given off by entities encountered in the environment to provide comment on their trustworthiness during an interaction [13, 5].

ICTs were instrumental in the establishment of institutions and the consolidation of their power within social systems, the emergence of the nation state in the western world throughout the 18th century provides a primary example [14]. These institutions sought to regulate behaviour within specific social contexts by formalising and enforcing rules, (dis)incentivising behaviour and defining roles, responsibilities and accountabilities for individual actors. These formal social pressures added a layer of institutional trust on top of existing interpersonal trust relationships, *inducing* certain behaviour within socio-economic interactions outwith a dependency on personal motivations that must be judged from existing relationships [2, 15]. Institutions took on the risk of misplaced trust in untrustworthy actors, thus increasing the capacity and willingness for social systems and the actors within it to engage in risky interactions. As Botsman articulates, the adoption of new technologies is inherently risky, requiring a *trust leap* to bridge the uncertainty associated with the unknown [16]. Institutionalised trust can be seen as a key driver for the adoption of new technologies and the evolution of society into increasingly complex configurations [2, 15].

Indeed it was institutions, primarily state actors, that funded and supported the development of computing technologies. First with Babbage and Lovelace's pioneering work on the analytical engine in the 19th century, then the work of Turing that theoretically conceptualised a computer [17] and Shannon's theory of information [18], next the actual development of the first digital computers and then the ARPANET, the precursor to the Internet [11]. Computing technology represented a step change in ICTs, these technological artefacts were capable of performing calculations on information

at a speed and accuracy previously unattainable by human actors. Furthermore, the capabilities of these devices have grown exponentially and the cost of computing power has decreased, leading to their proliferation throughout society and into the hands of individuals. In the words of information philosopher Floridi *ICTs are enveloping the world* [9].

Digital ICTs have unlocked new possibilities for structuring social interactions and are now integrated into many key subsystems of society including commerce, healthcare and government. However, they also introduce new risks around security and privacy as actions encoded, mediated and executed by ICTs on behalf of human actors must be authorised and may be recorded and correlated over time. ICTs add a layer of indirection between the human actor and their representation within an information system. In order for trust to be placed intelligently in this new medium, information about who the actor is and what authority they have must be accessible and assured to the appropriate level for the context of the interaction [19]. Assurance must also be gained in the binding between the human actor and the technical artefact they are interfacing with and interacting through. Within computer science the technical architecture to facilitate these interactions became known as identity management systems, which identified actors and virtually represented their *digital identity* within an information system. In the future, the same actor could authenticate against this representation, which the system could then use to determine access and authorisation to resources [20].

This thesis views this language around digital identity as legacy, originating from a time when digital ICTs and the systems they supported were simpler, often siloed and could, perhaps, legitimately be seen as distinct from the real-world [20]. Instead, this thesis prefers to use identification systems to encapsulate the complex set of actors, processes, rules and institutions that provision ID artefacts in order to facilitate identification. With identification defined as a formal process between an entity being identified and an entity performing the identification, often on behalf of an institution. This language is taken from researchers working on systems to provision legal identity for sustainable development [21, 22]. By adopting this language, it is possible to

re-characterise identity management systems as identification systems, in which an information system provisions account credentials (ID) and performs identifications of entities, who must identify themselves if they wish to access the service/resource protected by the information system. Using this language draws attention to the other, non-technical, aspects of an identification system, including the power conveyed to those who determine the rules and requirements for identification. This perspective is especially important as existing, non-digital identification systems are being augmented with technological artefacts that codify rules and sharpen the edges of these systems in ways that have been shown to disadvantage and exclude those most vulnerable in society [23].

This is primarily a technical thesis about the synthesis of digital identification technologies, however it is recognised that these technologies are part of socio-technical systems embedded in specific contexts against which their legitimacy can be judged [7]. With that in mind, the identification of healthcare professionals within the Scottish Healthcare Ecosystem has been selected to provide a concrete case study for how existing identification processes might be augmented with digital identification technologies. The requirement for an identification system to identify healthcare professionals to a high level of assurance is clearly necessary and legitimate. Furthermore, existing processes are widely stated as being inefficient, repetitive and burdensome for those being identified, as well as costly for the institutions performing the identifications [24]. Healthcare within Scotland is a highly structured subsystem of society, regulated by a multiplicity of institutions with well-defined identification processes and ID artefacts that healthcare professionals must collect, manage and present as they progress through their career. This ecosystem provides an excellent opportunity to demonstrate how digital identification technologies might augment these existing identification processes, reducing the burden of identification whilst maintaining the necessary levels of assurance.

1.1 Public Key Cryptography

This thesis is also about cryptography. Specifically how the introduction of public key cryptography in the 70s transformed the possibility space for structuring digital information flows [25, 26]. These new possibilities are now being realised to design and implement secure, privacy-preserving and decentralised identification systems. Public key cryptography removed the dependency for shared secrets from protocols to ensure message integrity, authenticity and confidentiality. At a high level a public key cryptosystem involves three algorithms:

- $KeyGen(.) \rightarrow (sk, pk)$
- $Encrypt(pk, m) \rightarrow e$
- $Decrypt(sk, e) \rightarrow m$

An interacting entity, Bob, can run, independently, the KeyGen(.) algorithm to generate a unique asymmetric keypair, (sk, pk), for themselves. By sharing his public key, pk, and keeping his secret key, sk, private allows Bob to engage in two fundamental cryptographic protocols with another entity - asymmetric encryption and digital signing. For example, Alice, with knowledge of Bob's public key, pk, can encrypt a message, $Encrypt(pk, m_1) \rightarrow e$, such that only an entity in control of the secret key, sk, (Bob) will be able to decrypt the information e, ($Decrypt(sk, e) \rightarrow m_1$) and read the message. Diffie and Hellman further proposed that the reverse of this protocol could be used to sign messages allowing their integrity and authenticity to be verified [26]. The signer in control of sk wishing to sign message, m_2 can run $Decrypt(sk, m_2) \rightarrow \sigma$. By sending σ along with m_2, anyone with knowledge of the corresponding pk can run $Encrypt(pk, \sigma) \rightarrow e_2$ an verify that $e_2 == m_2$. Thus giving them confidence that the message they received, was indeed the message sent. And assuming they know which entity controls sk, they can judge in the authenticity of the message.

The introduction of public key cryptography instigated the beginning of cryptography's modern scientific expression. Cryptography transitioned from a reliance on

shared secrets, code books and protocols that must be kept hidden from adversaries, towards a world of private, individually held secrets that could be used in public, published protocols whose security was based on the emerging science of computational complexity [11].

This thesis sets out to demonstrate that cryptography has become a highly systematised scientific discipline that has been shaped by early conceptual ideas. These ideas for cryptographic protocols were proposed, largely throughout the 80s, by researchers in response to risks they anticipated a society whose information flows became mediated by computing technologies would face. Chaum's proposal for achieving security without pervasive over identification is used to illustrate this point [27]. Furthermore, tracing the evolution of cryptography through the lens of Chaum's conceptual idea emphasises the strong foundations that have emerged, which have transformed abstract concepts into concrete, mature cryptographic protocols with well understood mathematical properties [28, 29, 30, 31]. This thesis demonstrates that it is now possible to use these protocols as blueprints to synthesise software artefacts that realise these properties within real-world systems. As technology is increasingly being used to augment human processes of identification and pervasive identification and correlation proliferates our digital interactions, Chaum's proposal appears more relevant than ever.

1.2 Artefacts

Artefacts, as discussed in Herbert Simon's publication The Sciences of the Artificial [10], are synthesised objects brought into existence by human beings to perform a function. They are designed or engineered constructs adapted towards a purpose which the artefact fulfils to varying degrees of success. How well adapted the artefact is depends on its goals, the inner environment that characterises the artefact and the outer environment that the artefact interfaces with. Although they may imitate aspects of naturally occurring phenomenon, they are distinct in that they are artificial.

Floridi, in work focusing on how ICTs are causing a revolution in our self-understanding as we become self-aware information processing organisms, focuses on the in-betweenness

of technological artefacts [9, Chapter 2]. These artefacts sit inbetween their users and what Floridi defines as a *prompter*, the circumstances that caused the artefact to be synthesised. Once constructed the artefact is intended to augment the user's ability to interface with these circumstances to achieve some purpose. Floridi further identifies three orders of technological artefacts. First order artefacts are prompted by and act upon natural objects. An axe for chopping wood or a boat to sail on water. Second order technologies are artefacts whose design was prompted by other technologies. The flint to sharpen the axe, or the hammer to knock nails into planks to build the boat. Finally, there are third order technological artefacts designed to be used by other technologies. Our world is now populated with these third order technologies, which form the backbone of advanced digital ICTs and the protocols that support their complex networked interaction.

Floridi also revisits the interfaces to these artefacts, identifying not one but two interfaces that should be considered. The first is the user facing interface, the means by which a user of a technology can interact with that technology and use it to achieve some purpose. For example, the handle on your axe, or the keys on your keyboard. The second interface defines how the artefact interfaces with the prompter, in second and third order technologies this is often described as the protocol. For example, how does the keyboard (the artefact), encode and communicate key presses exposed by the user interface to the computer. The schema for this model adapted from Floridi are shown in Figure 1.1.

When considering an identification system that provisions ID, the need for identifiers to identify entities is the prompter and the entity being identified becomes the user. In the simple case, when ID was paper-based these artefacts were first order technologies sitting inbetween the entity being identified and the entity performing the identification. However, with the introduction digital ICTs there was a need for information systems themselves to be able to perform and evaluate identifications of entities represented within these new environments. There is also an increasing move to digitise physical ID artefacts, which in doing so requires supporting third order digital technologies and the associated protocols able to issue, receive, store, present and verify

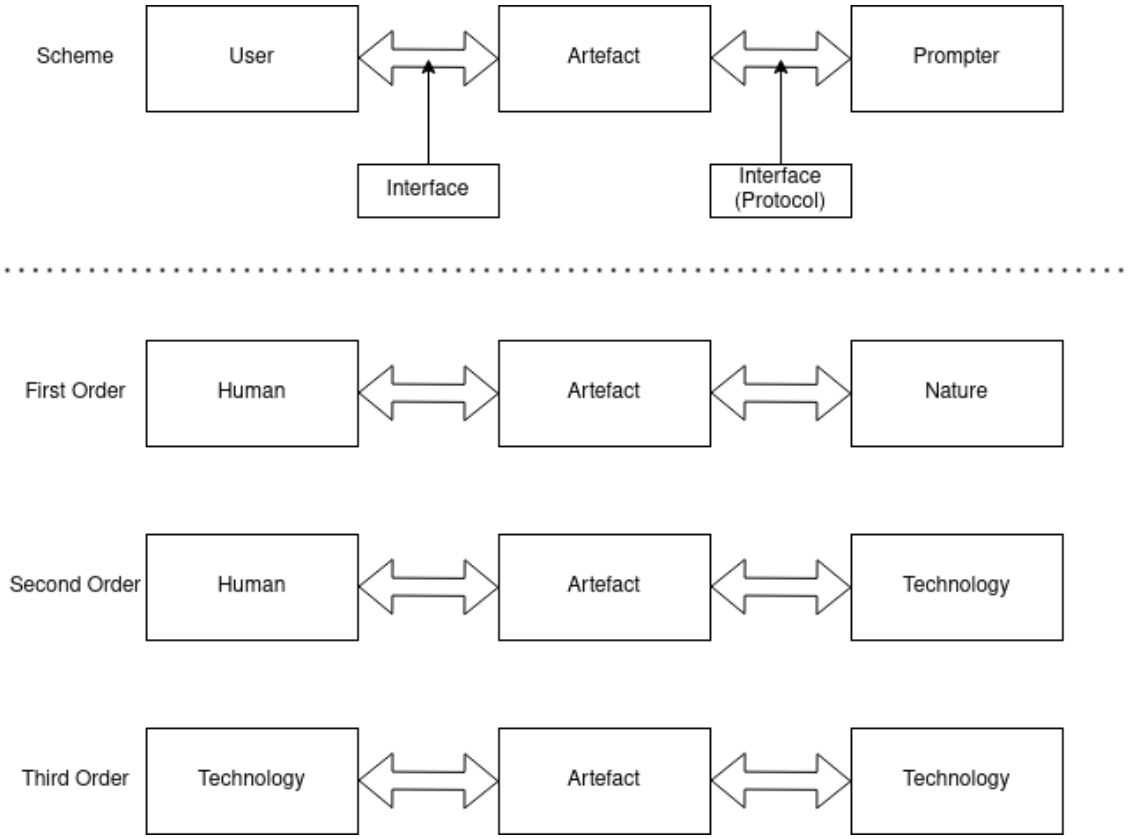

Figure 1.1: Three orders of Technological Artefacts (adapted from Floridi [9, Chapter 2])

these digital ID artefacts.

ICTs produce virtual, informational environments that human actors only can act within by interfacing with a technological artefact. This boundary between the human and the artefact introduces uncertainty through a layer of indirection. The human actor and the artefact they interface with are not the same. Trust must not only be placed in the actor and their actions, but the means by which the actor has been identified and represented within this informational environment. As well as the process by which their actions have been encoded and transmitted through a network of ICTs. The recipient of this information may be a software system that must make decisions around access and authorization based on a preprogrammed set of rules, an intelligent agent, or it may mediated through an artefact and displayed to another human through an interface (See Figure 1.2).

This thesis is focused on the ways in which these technological artefacts to facilitate and support identification systems have been conceptualised and synthesised. In particular two fields, or thought collectives, in computing have actively been working

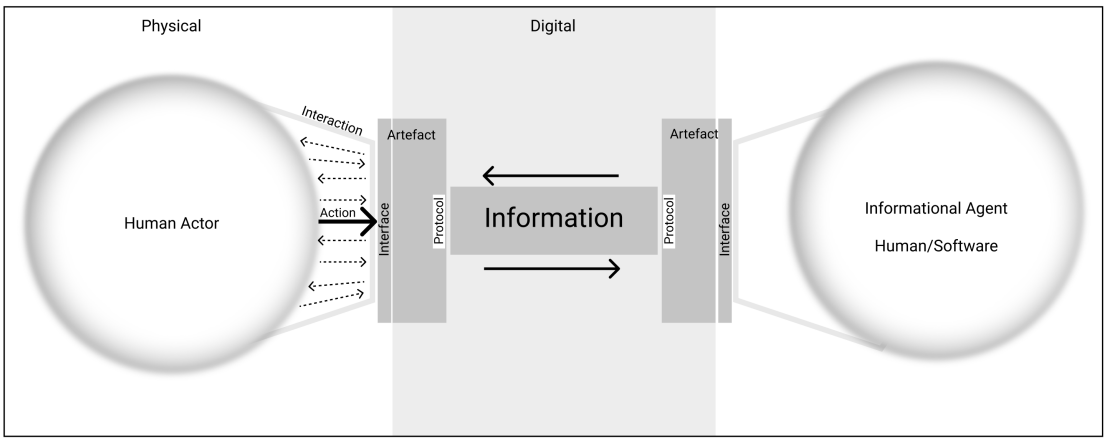

Figure 1.2: ICTs Layer of Indirection

on the design and implementation of these artefacts ever since the importance of identification within digital environments was recognised. These are the digital identity community and the cryptographic community.

1.3 Thought Collectives

A thought collective is a term defined by Ludwik Fleck in his work the Genesis and Development of a Scientific Fact [32]. In this work, Fleck attempts to demonstrate that the *conceptual creations of science, like other works of the mind, become accepted as fact through a complex process of social consolidation.* It is the thought collective, *a community of persons mutually exchanging ideas* that produce and consolidate these *thought products* into socially accepted facts [32, pp. 38-51]. With each thought product originating as a hazy proto idea, which exerts a kind of gravitational pull on the work of the collective ultimately leading to modern scientific expressions of the idea [32, pp. 78-79]. Each thought collective develops and acts as a carrier for a specific thought style, through which ideas are judged to be factual. No thought product is final, but undergoes transformation through continuous intra- and inter-collective interaction.

While the relevance of Fleck's work to this thesis is subtle, it underpins the general approach taken in the presentation of this work. In particular, Chapter 4 presents the systematisation of cryptographic knowledge since Diffie and Hellman outlined

public key cryptosystems [26] through the lens of Chaum's proto idea for achieving *security without identification* developing into a theoretically practical, provably secure cryptographic protocol [27, 28]. It is important to realise that theoretically practical and provably secure were not well-defined concepts when Chaum first proposed his idea but were developed within the thought collective in part to demonstrate that ideas such as Chaum's were *factually* possible when considered under the directed perception of the cryptographic thought style. Cryptographic knowledge is now highly systematised, with passive relations derived from the laws of mathematics and active constraints applied around what is considered secure and what purposes this knowledge has been developed for.

The other thought collective presented is characterised in this thesis as the digital identity community, which is reviewed in Section 3.2. The historical development of this thought style can be traced to the functional requirement to distinguish between actors and their actions performed within mainframe computers [33]. Digital identity has developed into a practical discipline focused on achieving functional, interoperable and implementable solutions to identification, authentication and authorisation within digital systems. The thought products of the digital identity community can be seen as the standards and specifications they produce, which define the way technical artefacts and their interfaces should be built, are clearly achieved through a process of social consolidation. Members of the thought collective negotiate within standards organisations, in an attempt to influence the imperatives that are standardised and effectively considered factual within the community. As the digital environment has changed, especially with the introduction of the Internet and the realisation of highly distributed systems the thought style of this community has evolved. With the latest set of standards, Verifiable Credentials and Decentralised Identifiers and the systems of identification they support having many similarities to the credential mechanism originally proposed by Chaum [34, 35, 27].

These two collectives are both focused on similar goals, the synthesis of technological artefacts to support identification. Although, apart from a brief intellectual interaction leading to the development of Public Key Infrastructures (PKIs) and Certi-

ficate Authorities, these two thought collectives have largely developed independently. Additionally, Brands argues that by failing to embed privacy, PKI overlooks much of the potential of cryptography [36]. The cryptographic community focuses on defining distributed protocols using mathematical equations and then analysing their theoretical security and privacy. However, they have always struggled to achieve adoption of these protocols within real-world systems [37]. Whereas the digital identity community has had success in the implementation and adoption of the standards it has produced, the privacy and security of these systems has been problematic [37]. This thesis seeks to demonstrate that these two thought styles need not be *alien* to each other and that the interweaving of ideas from both communities through *intercollective interaction* can produce the kinds of *special phenomena* that Fleck refers to [32, p. 10]. Indeed, this thesis argues strongly that it is possible to have cryptographically secure and privacy-preserving technological artefacts whilst also achieving the functional interoperability and implementablity that characterises the thought products developed within digital identity community.

Finally, both these communities of thought characterise identity primarily through the way entities are represented within, and by, a technological artefact. Under the National Institute of Standards and Technology (NIST) Digital Identity Guidelines for example, identity is defined as *an attribute or set of attributes that uniquely describe a subject within a given context* [38]. This thesis prefers to reserve identity for its more complex, multi-dimensional, relational and fluid meaning found within the social sciences [4]. Recognising that individuals have always constructed, negotiated and perceived identities as a powerful mechanism to influence and structure meaning during an interaction. Furthermore, identities are not just self-constructed but might be associated with a role, social group, category or label which we apply to ourselves and others [39, 4]. Identity under this understanding is the essence of what makes us human and is present in all aspects of our lives. As technological artefacts for identification are now being applied to mediate real-world interactions, this thesis is mindful of the impact these technologically augmented systems of identification can have on an identified individuals abilities to actively participate in the construction and

self-presentation of their identities [5, 40].

1.4 Research Questions

This thesis demonstrates how technical artefacts constructed from a strong foundation in cryptography can augment existing formal systems of identification found in highly assured social contexts in ways that distribute trust to those identified, balance power and preserve privacy. Whilst this is a technical thesis, it does not shy away from the fact that any system of identification is ultimately a socio-technical system which can only be trusted, effective and legitimate if it takes into account the experiences of those being identified and establishes clear constraints and accountabilities on those performing the identifications. The following questions have directed this thesis:

- **RQ1**: By analysing an existing healthcare identification infrastructure, can we integrate more trustworthy cryptographic primitives to bridge the gap between human and digital trust, and thus understand the key performance impact of these primitives?
- **RQ2**: What are the existing methods and software frameworks that can be integrated within a social structure which has high levels of human trustworthiness, in order to improve levels of digital trust?

1.5 Thesis Contributions

This thesis is primarily a technical thesis, and as such the concrete contributions of this work listed below are technical. However, the analysis of identification systems has been situated within a sociological framework for human interaction which underpins the presentation of this thesis. This interdisciplinary approach locates the technical work within a more meaningful understanding of our social reality and is itself a contribution to knowledge. It is only by developing detailed understanding of a specific social context, its use cases and requirements that a system of identification can be effectively and

justifiably augmented with digital identification technologies. This thesis has achieved this abstractly by presenting identity, trust and privacy from a sociological perspective (Chapter 2) and concretely by selecting a specific, highly structured sub system of society - healthcare - and engaging directly with key stakeholders to understand the requirements for identification and the existing processes in place that meet these (Chapter 6).

The technical contributions to knowledge of this thesis are as follows:

- The identification of a set of technical requirements that must be met in order to realise a credential-based identification architecture and evaluated the emerging architectures against these requirements.
- The construction of a model, developed with input from healthcare professionals, of the existing actors, institutions and identification processes within the Scottish Healthcare Ecosystem (Chapters 6).
- The production of a Proof of Concept (PoC) developed within an interactive web-based computing platform that augments the health care identification system to achieve high levels of trustworthiness within digital interactions.(Chapters 6).
- The implementation of set of experiments which evaluates both the individual operation of relevant cryptographic primitives and their performance within the developed software framework. (Chapter 7)

1.5.1 Publications

Table 1.1 lists a set of first author academic publications produced as a result of the research undertaken during the course of this thesis. Content from each of these publications has been included in different sections of this thesis and the relevant sections are identified in the table.

Publication Title	Citation	Relevant Thesis Sections
Trust-by-Design: Evaluating Issues and Perceptions within Clinical Passporting.	[41]	Section 5.1 and Section 6.2
Evaluating Trust Assurance in Indy-based Identity Networks Using Public Ledger Data	[42]	Section 7.1
PyDentity: A Playground for Education and Experimentation with the Hyperledger Verifiable Information Exchange Platform.	[43]	Section 6.5

Table 1.1: Academic publications produced as part of this thesis and their associated sections in the thesis

1.6 Thesis Structure

The structure for this thesis is as follows:

- Chapter 2 contextualises the thesis by presenting a sociological foundation for understanding identity, trust and privacy, emphasising how ICTs by changing the nature and environment of interaction has impacted how identities are formed, trust is placed and privacy can be realised. It underpins the technical work of the later chapters, highlighting that it is primarily humans participating within social contexts who: form identities; are identified; evaluate trustworthiness; and desire privacy.

- Chapter 3 introduces and defines identification systems. It presents the requirement for and evolution of digital identification systems and reviews the application and impact of augmenting existing identification systems with digital technologies.

- Chapter 4 traces the idea for achieving security without unnecessary identification within digital systems introduced by Chaum in 1985 through to its modern scientific expression. In doing so, a set of cryptographic building blocks that

provide useful properties when designing privacy-respecting digital identification systems are introduced. Along with this it presents a cryptographic protocol for realising *security without identification* constructed from these primitives.

- Chapter 5 defines the design principles and technical requirements for a digital identification system that leverages the selected cryptographic protocol with an existing software framework.

- Chapter 6 introduces the case study used in this thesis, with the identification of healthcare professionals throughout the Scottish Healthcare Ecosystem. This aimed engaged with key stakeholders through a facilitated workshop. Insights from this workshop are presented, including a set of figures depicting the existing identification system of healthcare professionals within Scotland. Chapter 6 concludes by outlining a PoC developed in a customised, educational environment that uses the Hyperledger Verifiable Information Exchange Platform (HVIEP) to augment key identification processes regularly encountered by healthcare professionals with digitally verifiable, cryptographically secure and privacy-respecting credential exchanges.

- Chapter 7 presents a set of experiments that evaluate the performance of the proposed PoC and the operation of the underlying cryptographic protocols in terms of time and space complexity.

- Chapter 8 presents the conclusions of this thesis and identifies areas that require further research.

Second Chapter

Identity, Trust and Privacy in an Information Society

> *"Are we to reserve the techniques and the right to manipulate people as the privilege of a few planning, goal-oriented, and power-hungry individuals, to whom the instrumentality of science makes natural appeal? Now that we have the techniques, are we in cold blood, going to treat people as things? Or what are we going to do with these techniques?"*
>
> Steps to an Ecology of Mind, Gregory Bateson [44]

Identity, trust and privacy are terms used widely throughout the computing disciplines, however their definitions within this domain often include implicit assumptions based on a technological worldview that some argue do not accurately reflect our real world experience [45, 46, 47]. Examples include the definition of identity as a collection of attributes related to an entity as enshrined in multiple ISO standards and the intentional exclusion of the human from the model [48]. As advanced ICTs become increasingly entangled with and embedded in human systems of interaction, a conceptualisation of these socio-technical systems purely from the perspective of the technical components and their interactions is no longer an adequate or appropriate response [45]. The accelerating transition from interactions managed by evolutionary social systems, to interactions that occur within intentionally designed socio-technical systems brings an urgency to this re-framing [15].

The chapter follows the advice of Connolly and situates this thesis in such a way as

to encourage the rich academic diversity required to ensure technologies protect our humanity and empower individuals as unique, complex human beings within the social systems they are embedded [45]. To achieve this an understanding of identity, trust and privacy has been synthesised from the sociological literature. Identities, trust and privacy are presented as emergent properties of human social systems that modulate the flow of information, the meanings interpreted from information flows and the scaled complexity that a social system can support [5, 4, 2, 7]. Each property is present, and uniquely perceived within a specific interactive setting and it is through the processes of interaction that the structure of society is continually created [6]. Furthermore, the impact of advanced ICTs, which have increased the quantity of information flowing around the system whilst reducing out ability to intelligently place trust, is reviewed.

Formal systems of identification, which are analysed in depth in the following chapter, are primarily social constructs applied as a mechanism to increase trust and mitigate the risk associated with individuals performing roles within a specific interactive setting. However, they can also influence and constrain the ability to participate in the construction of identities, amplify power imbalances and have serious implications for the privacy of those identified. This chapter provides the sociological foundations upon which the appropriate design, implementation and adoption of legitimate identification systems augmented with advanced ICTs can be justified.

2.1 Human Identity

Identity is a complex, sociological phenomenon studied across many distinct academic disciplines with roots that can be traced to influential researchers from across the last century such as Mead, Goffman, Bateson, Simon, Stryker and Margulis among others [49, 50, 5, 8, 51]. The individual is presented as a complex life form that is molded through their lived experience as they co-evolve in relation with their environment. They select actions from perceived possibilities as they navigate real-world complexity in a process of discovery while attempting to survive, reproduce and satisfy aspirations [10, 2, 52, pp. 28-20]. Though to consider individuals individually would be to overlook

the messy web of relationships and interactions that locate them within a social system. Their position in this network further shapes the possibilities they perceive and the actions they select. Identity, in this context, is fundamentally about meaning [4].

While each individual is unique, forming their own personal identities, they also participate in teams, fulfill roles idealised by society and belong to, or are categorised in, various social groupings [5, 39, 53]. This section presents identity theory, which views identity across three mutually interdependent, simultaneously applied dimensions: unique person identities, role identities and group identities. These identities are presented, applied and interpreted during an interaction to structure meaning and inform action selection (See Figure F.4 in Appendix F). Burke and Stets have developed identity theory over three decades of academic research which seeks to understand identities within an interactive setting. This sociological theory is summarised in this section [4].

This section provides the reader with the necessary knowledge of interaction and identity from which trust and privacy can be understood. It intentionally seeks to define human identity independent of technology. Humans have formed identities, fulfilled roles and participated in groups long before we had technologically mediated information exchanges. This analytical lens is then used to understand the impact of ICTs on human identity, looking at the spread of disinformation, the changing nature of interaction and the purposeful construction of virtual environments that we now inhabit.

2.1.1 Individuals

Individuals are intelligent, self-learning, self-organising, information processing complex adaptive systems [52]. They are the irreducible and indeterminable agents of change in any larger human system. They process information from memory and experience, interpret meaning, perceive possibilities and select actions to satisfy a multidimensional set of needs and aspirations [10, 2].

Sciences of the Artificial, a book widely recognised as foundational to the field of

artificial intelligence, analyses the individual as a goal-seeking information processing system molded by their environment [10, p. 22]. Simon introduced the concept of bounded rationality, whereby individuals make decisions within local constraints such as the time and information available as well as their *inner environments* ability to process this information [54]. Simon suggests that individuals are *satisficers*, seeking good enough alternatives over optimal solutions, pointing out that optimisation problems are notoriously computationally expensive [10, pp. 28-30]. Individuals have aspirations across many dimensions which they use as a differential mechanisms to select actions that will satisfy their aspirations, which gradually trend downwards over time [55].

Simon's notion of satisficing is echoed in biological studies of life. With both Wheatley and Margulis expressing views that life is in a constant process of discovery, as it creatively explores possibilities seeking good enough solutions, recognising diversity over efficiency as the key to resilience [51, 52]. Life adapts to its environment, and the environment adapts to the life existing within it. Both mutually interact and shape the other, as Bateson emphasises throughout his work, the constancy is in the relationship between the two, which to survive requires change in both related elements as they co-evolve together [44, pp. 153-156, pp. 338-339].

2.1.2 Interaction

The individual as a living, information processing, self-learning agent in their environment does most of its learning through and from their interactions with others. Interaction is the active process by which information is reciprocally exchanged between individuals as they attempt to shape meaning and coordinate action within their environment [5, 56]. When we interact, we encode information into our messaging and signals that provide comment on the relationship and the trustworthiness of the information encoded [57]. These signals are generally semi-voluntary and applied reflexively by the individual, implying an evolutionary emphasis placed on the importance of honesty within these interaction systems [13]. Within the systems of interaction that humans and other mammalian animals produce, it is recognised that signals can be

trusted, distrusted, falsified, denied, amplified and corrected, hence they are redundantly applied to messaging giving a greater contextual depth to the information [8]. A diagram representing interaction using the language in the literature can be found the in Appendix (see Figure F.3).

Goffman uses the conceptual lens of a dramaturgical performance to examine the art of constructing and maintaining the definition of a situation practiced by a *performer* during an interaction. The objective being to impress upon an audience, a definition that influences the direction of an interaction in a favourable manner for the performer. Goffman emphasises that this perspective is not perfect, life happens moment-to-moment with boundaries, roles and direction rarely defined or rehearsed. However, the expressive and relatable language of a performance usefully draws attention to the complexity of human interaction [5, pp. 10-27].

Individuals make use of the expressions *both given and given off* by other individuals in their presence. The observer interprets and processes this information in order to structure experience, influence behaviour and select actions from perceived possibilities [5, pp. 10-27]. To do this effectively, they must make a judgement about intention, truthfulness and commitment of the actors to their performance. This perception exists on a fluctuating continuum between convinced and cynical. With individuals often settling for *good enough* performances for the given interaction [5, p. 30]. There is value in maintaining a working consensus, or common ground, such that inter-subjective meaning can be established. This gives confidence in future actions of others, fostering cooperation and enabling the coordination of activity to achieve shared goals [56]. The tacit agreement not to challenge each others definitions is mutually beneficial, as all performers are likely to be concealing some aspect of their selves they believe incompatible with the definition of the situation they are trying to foster [5]. As O'Neill puts it, transparency is not necessarily conducive to strong relationships [19].

In an interaction, an individual's initial projection of themselves sets the tone for the encounter. The definition of this initial informational state is the point at which an individual has the greatest degree of freedom to present themselves. This in turn defines the way they would like to be treated and the expectations that others present in the

encounter can reasonably make of them [5]. It is also the time when those observing the individual are least able to make a judgement about them, as this must be done without any first-hand evidence. Instead, they rely on comparisons against similar experiences, cues drawn from the social context and bounded region in which the interaction occurs and generic categorisations learnt from the society they are embedded within [53]. Repeated interactions over time form a relationship, with memory and local meaning cultivated through shared experiences. This prior knowledge helps to regulate the interaction, with each party bound by the commitments they have previously made to the other.

2.1.3 Groups

2.1.3.1 Teams

A team, using Goffman's definition, is a group of individuals working to maintain a shared definition through a collection of interdependent individual performances [5, Chapter 2]. This creates a *bond of reciprocal dependence*, as any performer has the power to disrupt the teams projected definition [5, p. 108]. A team is a grouping in relation to a set of one or more interactions in which a consistent definition is maintained throughout [5, p. 108].

Team formation is fluid, and not always formally defined through roles and responsibilities. Appiah references a famous experiment in the 1950's that identified four steps to group formation, first individuals share a common purpose creating reasons for interaction, over time each individual's experiences and behaviour are shaped through the process of interaction within the group [53, pp. 29-31]. Gradually the structure of the group forms as individual status, roles and relationships stabilise. Individual members identify as part of the in-group in opposition to other group structures they encounter. Finally, the group evolves a set of norms to regulate behaviour of its members as they interact internally and externally with other individuals [58].

2.1.3.2 Categories

Social categorisations are a type of grouping we apply to ourselves and others to help us make sense of the world around us and the unpredictable entities we encounter within it. Grouping individuals into understood classifications allow us to apply generic traits and expectations to unknown others. It has been shown natural for children to group people into categories, apply generalisations to the groups enabling us to simplify the world around us [59]. We essentialise them [53, pp. 25-29].

Appiah describes in detail the mistakes we repeatedly make when we prescribe social identities based on commonalities perceived to bind a group of individuals into an arbitrary category. Many of which come with historical baggage, misconceptions and stereotypes. Appiah focuses of five broad complex classifications societies commonly apply: creed, country, colour, class and culture. He refutes the assumption that at the core of these identities, there is a common essence binding a group of people together throughout his work [53]. However, he recognises that a shared identity helps bind groups together around a common self, enabling them to motivate and coordinate action at scale [53, p. xvi]. Similar language is used by Wheatley as she describes how life organises around a self to give it form, meaning and purpose [52, pp. 46-54].

2.1.4 Identity

A theory of identity first introduced by Stryker [50, 60], although influenced by the earlier work of Mead [49], has formed the foundation of sociological research into identity. Identity theory situates identity, which Burke refers to as the *sets of meanings people hold*, within a social structure [39]. Through interactions, observation and education we learn to label and classify our experience and develop behavioural expectations in relation to these learnt structures of thought. They become symbols that carry meaning. These symbols assign us positions within a social structure, which come with associated roles. Roles develop and derive shared behavioural expectations from the social structure they are embedded within. By invoking the names of social positions in interactions, individuals can implicitly draw attention to commitments that they expect

others to uphold [60, 39, 4].

Stryker suggested that we additionally apply these labels to our self and by doing so reflexively incorporate the meanings these labels hold within our self-definitions [60]. This builds on the early work of Mead, who defined individuals as being in constant negotiation with society, a process in which they continuously redefine the roles they play, the values they maximise and the purposes they pursue [49]. Stets and Burke further emphasise this point, by perceiving ourselves as an object in relation to the roles we hold, the social categories we identify with and the others we interact with, our identities are continuously reformed [61].

Identity theory has been an active area of research since its introduction. Important contributions have recognised roles are not the only identity forming lens we apply to ourselves and others. Additional identities suggested include personal identities, category identities and group identities [61, 39, 4]. Category and group tend to be taken together, as social identities, although categories are broader labels categorising individuals against gender, creed and class with meanings determined by culture and history as discussed by Appiah [53]. Whereas group identities are derived when an individual self-selects to become a member of, or is placed in, a social group. Personal identities are those that arise from an individual's unique self expression and evolve through their personal interactions and the local meanings they develop as a result of these [62]. Personal identities are always active within an interaction, on top of which role and social identities are applied and aggregated. Through this application person identities are shaped, especially where there are conflicting meanings and expectations [39].

Other researchers have focused on the importance of resources to identity theory. These are defined as anything that sustains us and our identities across interactions, space and time [63, 62]. They do not need to be valuable or scarce, rather they can be anything that functions to sustain a system of interaction. The control and manipulation of both active and potential resources is seen as a key motivator for action in social systems [39]. The concept of control over potential resources, a resource that might become active at some point, recognises the future-oriented nature of individuals.

Finally, identities have been viewed by researchers as a perceptual control system comprising of four components [64, 62]: 1) The sets of meanings held for an identity, termed the *identity standard*; 2) The perceptions of meanings relevant to that identity derived from a specific interactive context often through observation of and feedback from others [5, 62]; 3) A comparator function contrasting meanings held within the identity standard and perceived meanings from a situation; and 4) The differentials output from this function between the idealised and perceived meanings. The view of identity as a cybernetic control system is that individuals will seek to counteract the perceived definition of a situation where there is a deviation from the identity standard they hold at that moment [39]. Within this context, identity verification is defined as the process by which people act to control the perception of identity-relevant meanings in a situation. Role verification is viewed as verification of the cognitive performance, or competence, of a role actor. Social verification increases feelings of self-worth as individuals verify themselves as being like others. Finally, personal verification leads to feelings of authenticity as individuals conform to the expectations they set themselves [39, 4]. Identity standards held by the individuals set their expectations for the way an interaction is supposed to go.

2.1.5 Social Systems and Structure

Social systems are defined by Luhmann as dynamic meaning-making systems of communication for structuring experience and selecting actions in a complex world [65, 2]. The social structures referred to by Burke and Stets in the identity theory literature appear to fit into this definition as an emergent property of these social systems [39]. Individuals perceive the structure of a social system and use this to locate themselves and others. These perceived positions assigns role identities and influences the social groupings an individual participates in, as well as the labels they apply to themselves and others [39, 4]. Furthermore, a social structure can be understood as the flows of resources and information around a human system. Nissenbaum describes a similar notion which she refers to as the social context, through which information flows between

senders and recipients can be meaningfully understood. A social context is defined by a set of values, roles and behavioural norms [7, pp. 129-145].

Norms are widely recognised as influential factors regulating behaviour in an interactive context by setting social expectations [5, 7]. Axelrod defines a norm tied to a social setting as *the extent that individuals usually act in a certain way and are often punished when see not acting in this way* [66]. While cooperating groups, or teams, have been shown to evolve norms [58], norms are drawn from and created in relation to an individual's or team's position within a social structure.

Social structures may be partially or completely organised and institutionalised, with roles, rules and responsibilities formally defined and explicitly assigned [2, pp. 53-67]. These are then enforced by organisations who are themselves subject to regulation and constrained by legal obligations (see Figure F.11). The context of healthcare, which this thesis is focused on, provides an example of a social system that has been structured and institutionalised over time [7, pp. 171-174].

2.1.6 Identity and Power

Power is a generalised symbolic communications media that provides a mechanism to exert influence over the selection of actions in others under unfavourable circumstances [2, pp. 119-131]. Through the construction of inter-subjective meaning, action chains can be established between subordinates and superoridates that open and constrain possibilities for actions. Actions taken are experienced through the established lens of meaning, attributing motives and intentions to the actor, thus facilitating a chain of interdependent actions over an extended period of time [2, pp. 132-142]. Luhmann emphasises the distinction between power and coercion, stating that power is open to possibility and increases when all participants in an action chain have greater optionality. Whereas he defines coercion as forcing person to take a single concrete action [2, pp. 122-123]. Both participants in a power relation perceive a course of action they wish to avoid, this *avoidance alternative* can referenced by the other participant as a threat of punishment used to motivate action selection. Although it is in the interest of

those who wield power to avoid exercising the sanction as it alters the relationship and reduces possibilities [2, pp. 122 123].

Depending on the identities you hold and have applied to you and their position within society as perceived by others, whether accurate or not, can either convey or reduce your privileges [53, p. 11]. Privileges in identity theory are understood as access to active or future resources, where resources are defined in the broad sense as anything that sustains the identities and interactions that maintain a social structure [62]. Thus, a position within the structure that conveys control over these resources, conveys power [67].

Burke et al define the dominance identity, as a person's *self-meanings of power and influence* in a specific interactive setting [67]. Furthermore, through an empirical study they identify key factors in determining an individual's punishment/reward strategy. These include the network structure, the reward available based on an individual structural location, the capability to punish and the dominance identity of a given individual. Although they also make clear that these factors are contingent on others including the history of the participants. This creates a unique interactive setting that evolves over time, influencing strategies, beliefs and actions [67].

2.1.7 Identity and Change

Identity control theory draws attention to the fact that the identities we hold are not static, despite the labels we apply. Individuals seeks to control the definition and perceived meaning in an interactive situation by bringing it into alignment with their *identity standards*. However, at the same time these standards undergo transformation as they slowly adapt to the projected definitions of others [62]. This change in the *identity standard* is understood to occur more quickly when an individual has limited ability to influence the meaning projected into a social context that they are embedded in [39, 68, 62]. This aligns with Goffman who suggested that controlling the definition of an interaction, influences the behaviour of those within the interaction [5].

It is not just identity control theory that emphasises that identities are in constant

flux. Researchers who study complex adaptive systems have been grappling with the desire to identify the object of their studies, which are by very definition constantly changing. Cummings and Collier identified four components required for an adequate definition of such a system: 1) its components, 2) the relationships between these components, 3) the location, spatial scale and appropriate constancy that the definition applies and 4) the temporal scale and definition of constancy through time [69]. The definition of a system is contingent on the boundaries drawn by the observer. Wheatley applies this to the self, describing a paradox whereby the self must draw boundaries to distinguish it from others and yet no individual self exists separately. Life is connected and entangled, it co-evolves in relation to others and through this process the self can redefine its boundaries changing the emergent identities it organises around [52, pp. 46-54].

2.1.8 Identity in an Information Society

ICTs have made society increasingly complex and created change at an unprecedented pace and scale. From the printing press and the penny post to the digital age we live in today, each new technology has introduced new possibilities for interaction that are no longer constrained to a single spatial temporal location. These benefits have come at a price. We are no longer able to rely on the redundantly encoded expressions given and given off during a physical interaction to provide comment on the information being exchanged [13, 5]. This reduction in contextual depth is further hindered by the removal of a shared environment, containing *sign equipment* that is often used to construct a common ground enabling understanding to flow between participants [5]. Finally, these technologies are developed and operated by organisations who take on a structural position of power as they mediate the information exchanges across these technologies. These organisations are often motivated by purposes that are not aligned with the individuals whose interactions they mediate [70].

In the information age, as the pace of innovation accelerates, these challenges have become more acute. Interactions now take place in designed, constructed, privatised

virtual environments that are no longer inanimate. Rather they are constantly reconfigured by their designers as they experiment with possibilities, personalise experiences and seek to optimise variables. Often this means maximising profit, whose proxy in the digital age has become attention [71]. Zuboff describes in detail how these technologies are being applied to engineer human behaviour at scale: *tuning* the *choice architecture* of these environments, *herding* behavior by controlling a person's immediate context to increase the probability of a specific action selection and *conditioning* a subject through scheduled reinforcement [72].

All of these processes are applied at scale and continuously improved through A/B testing to determine which informational inputs produce a desired set responses on a labeled and categorised candidate pool [72, 73]. Information is collected about us, often beyond our awareness, as we interact with and across these platforms. This information is then to construct rich dossiers on our lives from which we can be labeled, grouped, herded and tuned [7, Chapter 2]. We find ourselves with diminishing influence over how we choose to present ourselves, with our identities selected for us, often by algorithms we never see. These in turn are used to determine the information we receive and the make up of the virtual environments we interact within. Reflecting on identity control theory, it seems clear that this *shadow text* as Zuboff defines it, which is out of our control and exists largely beyond our awareness, will cause changes to our *identity standards* and hence the meanings we draw as we process experience from our environment and those we interact with within it [74, 62].

The ability to use ICTs to influence others is not new. In some ways it is simply another mechanism for impressing onto others a favourable definition of a situation which helps to define roles and sets expectations of others in the interaction [5]. However, a key difference is a change in scale and speed at which information can be disseminated and the unequal access to these technologies both in terms of availability and education. Mass information dissemination that once was the role of centralised state actors and travelled at the speed of a horse and later became the responsibility of newspaper publications. Now the dissemination of information is available to anyone with access to the Internet, it can be published anonymously and can travel round the world in

minutes. This has led to an increase in the intentional spread of disinformation, as unknown actors attempt to seed information that develops into narratives persisted by communities of real, authentic individuals [75]. There is an increasing realisation that this can have meaningful consequences in the real world. A shocking example of this being the storming of the US Capitol in January 2021 [76]. Cambridge Analytica and Russian election interference are other worrying examples how ICTs are being co-opted by powerful actors to shape meaning and influence democratic processes at an unprecedented scale [77].

In addition to the flood of disinformation, the information age has seen the introduction of digital (autonomous) agents. These are defined as software that have the ability to act in the environment, communicate with other agents, pursue objectives, process resources, possess and learn skills and offer services [78]. Individuals are no longer the only entities that structure experience and select actions within their environment. As the capabilities of these agents have increased, they are being given increasingly influential roles in decision making processes that have real world impact. For example, computer algorithms are used by law enforcement to filter and sort the population, direct attention and resources and identify suspects. These have repeatedly been shown to contain coded biases embedded into the design of these systems by those holding positions of power over this new medium [79, 80, 81].

2.2 Trust

If identities are formed primarily through the interactions of individuals as they exchange information in an environment as the evidence suggests, then understanding how these exchanges are mediated is crucial if we are to effectively develop humane relationships across technologically mediated information exchanges. This section argues that trust is foundational to understanding these human interactions. Trust has been defined as a multi-faceted, contextual and future-oriented feeling or belief that we place in others as we depend on them to fulfill commitments whilst navigating uncertainty with imperfect information [2, 82, 16].

Fortunately, trust has been the focus of research for over 50 years, and it is something we all experience in our lives. This section looks back over the research into trust in an attempt to understand and define this nebulous concept. It samples early ideas, drawing on the republished translation of Niklas Luhmann's work *Vertrauen* in 2018 as Trust and Power to celebrate 50 years since it was first published in German. This work framed trust within complexity, defining it as the way individuals and social systems use trust to structure possibility, process experience and select actions [83, 3, 2]. In addition to early theories, current characterisations of trust found in the literature are surveyed. Hawely frames trust around the concept of a commitment and like Luhmann emphasises the importance of understanding distrust within this framing [82]. Her later work contrasts with Luhmann on his notion of systems trust, questioning if this is in fact reliance [84]. O'Neill has made many thoughtful contributions considering the influences transparency and accountability have on trust first outlined in the 2002 BBC Reith Lecture series [85]. Since then, she has discussed trust and trustworthiness [86] and in 2020 presented her thoughts on trust and accountability in the digital age [70]. Finally, trust is explored within the context of an information society, looking at the impact of advanced ICTs on our ability to intelligently place trust and reviewing some of the ways in which technology is being used to understand trust and design trustworthy technical systems to mediate interactions in social systems.

2.2.1 Interpersonal Trust

Trust is widely acknowledged as fundamental to human life, social interactions and organisations [87]. It is an attitude from a trustor towards a trustee that emerges from social relationships within a framework of interaction [2]. DeSteno reviews evidence showing that from birth an individual makes judgements about what and who to trust in their environment, over time these experiences shape their dispositions to place trust [88]. They learn and adapt over time, using memory from lived experience to modulate attitudes of trust throughout any given interaction [2].

A review of the trust literature by Carter et al identifies that the majority of research-

ers view trust as a three-place relation; *A*, the trustor, trusts *B*, the trustee, to perform action *X* [87]. Hawely extends this definition to make a clear distinction between trust and reliance, stating that trust is only appropriate when *A* believes *B* has a commitment to do *X* and expects that they will meet this commitment [82]. She uses a broad notion of a commitment, as something either explicitly made or implicitly perceived from the social context an interaction is embedded within. She suggests that such a commitment does not need to be made directly to *A*, but rather *A* may believe *B* has a commitment to some other entity [82]. Furthermore, Hawely defines distrust as the expectation that a commitment made by *B* will not be met, highlighting the distinction between merely the absence of trust. This view is shared by Luhmann, who referred to distrust as the opposite and *functional equivalent* of trust [2, pp. 79-85]. See Figure F.8 in the Appendix for a representation of trust and distrust.

The act of placing trust is a mutual commitment initiated by the trustor and agreed to by the trustee [2, 82]. This commitment constrains actions of both parties as they present a self in this relationship that is bound by the trust commitment. To break this commitment, would mean presenting a different, untrustworthy self [5]. Over time the selves presented by trustor and trustee help to regulate the patterns of interaction and attitudes of trust within the relationship.

Interpersonal trust is based on familiarity, developed through social relationships in which each individual has a choice to place trust in the other [2]. It is widely recognised that placing trust is fundamental, we exist in a society, where we must constantly place trust in others as we depend on them for various aspects of our daily lives [85, 87]. Without trust, Luhmann writes, we would be left with chaos and fear, paralysed by our inability to navigate an uncertain environment [2]. Trust, while necessary, must be placed in the present with care, through active inquiry where the individual that bestows trust, the trustor, does so after a judgement based on available information and prior experiences [19] (See Figure F.2). Where claims are made, the trustor seeks out evidence to verify their accuracy and ascertain the authenticity of the source before considering the claim when weighing up a decision to place trust. Although whether this decision is best characterised as a calculation, or more of a belief or feeling is much

debated in the literature [87].

The decision to place trust is independent of whether a trustee is trustworthy within the context of a particular interaction. For trust to be effective it should be placed intelligently in the trustworthy and withheld from the untrustworthy [19]. Botsman identifies four traits of trustworthiness that a trustor is attempting to judge: competence, reliability, integrity and benevolence [16]. Competence and reliability relate to an individual's ability to complete a task that they have committed to. While integrity and benevolence are aspects of an individual's character, their honesty, empathy and goodwill.

Researchers generally categorise processes used by the trustor to evaluate trustworthiness of a trustee within interpersonal social relationships into three dimensions: behavioral, affective and cognitive [89, 90, 91]. With behavioral judgements, the trustor interprets the signals given off by the behavior of another within a specific interaction, through their own lens [5]. Initially this will be independent of the trustee, but over time, as the relationship develops, the behavioural nuances of the trustee's self-presentation can be learnt and used to modulate trust across repeat encounters [2]. Affective, or emotional evaluations, are based on feeling and influenced by the trustors unique standpoint developed through their lived experiences within society. Finally, the cognitive dimension is a judgement on ability of a trustee within the context of a specific action that the trustor is dependent on the trustee to fulfill. The trustee provides evidence of competence by disclosing claims and successfully completing tasks [89]. See Figure F.3 in the Appendix.

Interpersonal trust in human relationships must be paid continual attention to as it fluctuates dependent on factors such as risk, perception, roles and social context. It is formed gradually through low-risk interactions wherein the trustor presents the trustee with opportunities to demonstrate their trustworthiness [2]. Trust is fragile and inappropriate or overzealous examination can damage a trust relationship when if left alone it would have remained strong [92, 87, 93]. For example, through continuous monitoring of progress towards a commitment.

2.2.2 Systems Trust

Systems trust, or institutional trust, seeks to establish relationships of trust outwith a dependency on *familiarity* [16]. Luhmann suggests that social relationships built on personal trust developed across repeat interactions are no longer sufficient for achieving the levels of trust required to navigate an increasingly complex society [2]. Or perhaps, to have created such a complex society we needed to design different mechanisms for structuring context, processing experience and selecting action from an increasingly rich set of possibilities [2, pp. 53-67]. Logically this makes sense, we are no longer able to know with familiarity all those we interact with and yet we must still depend on them in our daily lives. The taxi driver, healthcare professional and schoolteacher are all examples of relationships maintained through this impersonal type of trust [94, 16].

A similar notion of trust was introduced by Walker which she calls default trust, the feeling of security within an environment established because an individual feels confident in what to expect, who to place trust in and how to judge trustworthiness. This definition is further extended to *diffuse, default trust*, whereby the normative expectation to uphold a commitment are directed towards an institution rather than a specific individual [95, 78]. Such attitudes have been demonstrated through qualitative research [96]. This points to the importance of the context, or zone of default trust, in which an individual performs roles. Roles give individuals an impersonal motive for action which can be judged in relation to trust in the system, independent from personal trust established with the individual actor [7, 5, 2].

Systems trust and interpersonal trust develop independently. Within any social organisation of individuals, interpersonal trust acts to balance the dependency on the system [2]. After all one should not place trust blindly. Instead, they should make a personal judgement about a specific interaction with an individual, while taking into account the social context surrounding this interaction which the system helps to define and regulate [19].

Luhmann identifies generalised communications media as a key mechanism used to develop and sustain impersonal systems trust. He goes on to detail three: money,

truth and legitimate political power [2, pp. 53-67]. Each of these objects of trust have different properties and dependencies that are largely known and understood by individuals, so they can be relied on when making decisions about the future from imperfect information. For example, in Luhmann's words, *truth is a medium which acts as a carrier for the reduction of inter-subjective complexity* [2, p. 57]. That is, interacting actors can reach consensus about a statement and have confidence this will be upheld by others, that is, everyone will agree the statement is true. The key difference for Luhmann from pure reliance as explored by Hawely [84], is that trust in a system has evolved through and is dependent on human processes, which must be performed by trustworthy actors [2, pp. 53-67].

The impersonal nature of trust and the important role it plays in maintaining complex social organisations presents many opportunities for abuse. Therefore, it demands processes of accountability, which themselves create new opportunities for abuse [97, 70]. Shapiro points out that this can create an *inflationary spiral* of processes for control over institutional trust decreasing the agency relationships of those participating within the system [97]. This is echoed by Schneier who points out that there can be too much security, which limits freedom, liberty and individualism within a social system and can lead to stagnation [15].

2.2.3 Trust, Risk and Assurance

Risk is the possibility for future harm perceived by individuals through emotions and cognitive predictions [93, 98]. While risk is often portrayed as a decision heuristic by way of assigning a probability to the likelihood of a specific outcome, Hansson points out that in reality we rarely make decisions against well-defined probabilities rather we are navigating uncertainty [98]. This aligns with Luhmann's perspective on trust a mechanism to reduce social complexity allowing us to take on risk in the face of imperfect information as we orient ourselves towards the future [2].

Trust is unpredictable and must be placed without guarantee, this creates a dependency on the trustee and leaves the trustor vulnerable [19]. This vulnerability cannot

easily be mitigated, for example, as previously mentioned, through continual monitoring of the trust relationship [87]. Some suggest that exposure to risk is a necessary component of trust; without risk trust would not be required [19]. Although Nickel points out that while the demand for trust increases with the risk associated to a particular action, it has also been shown that trust is not always based on risk [93, 99]. This suggests that trust and risk coexist and influence each other but are not tightly coupled.

As risk increases, the act of placing trust becomes harder for the trustor to justify [93]. If the trustor is to place trust intelligently, they must have a means to gain confidence that the trustee will meet the commitments specified by the trusting relationship [82]. In other words, they must demonstrate they are trustworthy within the context of the interaction [86]. The trustor must gather evidence, providing assurance in their judgement of the trustee's trustworthiness. Such evidence can be drawn from previous encounters with the trustee, emotional bonds and claims presented about the trustee. Claims must be judged for their accuracy, authenticity and relevance, which requires the ability to attribute claims to their sources whose authority and trustworthiness must also be assessed [85].

As society has organised into increasingly complex social groups, systems have developed to absorb the risk beyond the capacity an individual trustor is likely to be comfortable with. Individuals embedded within these systems, interact with others in roles governed by the norms and purpose of the social context they are participating in [7]. Luhmann and Shapiro both define this as impersonal trust, providing social security and assurance independent of the interpersonal trust that emerges from direct human relationships [2, 97]. The nature of these interactions allows distrust to be coordinated without damaging personal relationships, such as when the administrative staff in a hospital check employees qualifications.

2.2.4 Trust in an Information Society

When considering trust in the context of today's information society there are two angles we must take. First the impact of information technologies on our ability to

create and maintain trusting relationships and judge trustworthiness of actors and information presented to us in this new environment [94, 86]. Second, the development of technologies to understand, manage, promote and design trusted interactions and provide assurance within technologically augmented environments.

2.2.4.1 The impact of Information Technologies on Trust

Information technologies are making it ever easier to produce and disseminate information to wider audiences. In today's digital age, these mechanisms are no longer limited by geographical or temporal constraints [70]. As such information is regularly consumed from a vast array of sources that the average individual has likely never met, let alone had the chance to become familiar with. As such our ability to judge trustworthiness of a source, can no longer leverage the rich, semi-voluntary and redundantly reinforcing expressions both given and given off by an individual in a face to face interaction [5, 13].

An analysis of the literature on trust in sharing economy by Hawlitschek et al identified that digital interactions often include three parties, the consumer, the provider and the platform [100]. Keymolen refers to this as interpersonal systems trust, with the platform acting as the mediator and emphasising the active role technology plays in shaping trust attitudes between the consumer and provider [101]. Both the consumer and the provider must place a degree of trust in the platform they are using. This trust is based on beliefs about the intentions and capabilities of the organisations responsible for the platform [96].

O'Neill further addresses the impact of these platforms as new mediators of information and the resulting disintermediation of incumbent actors will have on relationships of trust. Outlining how digital platforms have been able to side step regulation and avoid accountability, which while never perfect, play an important role in society of holding those in positions of power responsible for their actions [86, 70, 74]. These new intermediaries hold ever stronger positions of influence over the information flows within society, while at the same time withdrawing from commitments to uphold the responsibilities and duties that actors in similar positions have historically been constrained by.

The fact that these information technologies are being used to spread disinformation and deception to specific target groups at scale is detrimental to our ability to form and maintain trust well [70]. Those who can afford it can now wage *information warfare* using these technologies. Leaving individuals to place trust without the ability determine the source of the information, its purpose, funding, aims or agenda [102, 103, 70]. We find ourselves drowning in information detached from context, leaving us without the effective means to sort, filter or judge its veracity or relevance [19]. Botsman identifies this loss of contextual space as a key driver for the loss of faith in evidence and the mechanisms by which truth can be determined from fiction. It is in these instances of *reality apathy*, that we revert back to *local trust* or interpersonal trust in direct relationships we participate in [16, 84].

2.2.4.2 Designing Technologies for Trust

Information technologies have an impact on our ability to place trust intelligently, both in the virtual environment, and increasingly the real world. They have enabled us to interact with peers throughout the world at a vastly larger scale than anything seen before in human history. Yet, the maintenance of trust, an active process that takes time and energy has been delegated to the platforms mediating our information exchanges [70]. They calibrate the rules, govern the algorithms and develop the user interface. Then they determine who can participate in these virtual environments through identification and authentication procedures. It is this which has conveyed these platforms so much power [74].

In response to the growing power of platforms, many have looked to designing trust-free or trust-less systems that leverage blockchain technology to disintermediate actors [84]. By relying on these public, distributed, immutable and consensus-driven state machines, individuals can exchange value without being dependent on an intermediary [104]. However, Glaser points out that while this is a valuable innovation, it is only one layer of the digital infrastructure required for complex interactions. He suggests trust-free systems are only possible in simple, closed ecosystems [105]. Furthermore, even in the case of the Bitcoin blockchain [106], a financial accounting ledger for the

bitcoin currency, trust still exists. Instead of being placed in people it has shifted to the algorithms that govern interactions within the system [107]. Although, as noted previously there is still debate about whether trust can be placed in such a thing as an algorithm [84, 87]. It is also worth highlighting that these algorithms are designed, implemented and maintained by human actors. So trust is implicitly or explicitly placed in the systems, processes and institutions governing these actors. Research has shown that claims of trustlessness in new technologies, have infact been used to obfuscate the new centers of power that these systems introduce [108]. Without intentional, designed governance processes, those who govern will be unaccountable.

Compelling evidence that the focus on designing trust-free systems is misplaced has been compiled by Hawlitschek et al [100, 109]. Their work extended a four layer blockchain engineering framework proposed by Notheisen et al after a thorough literature review. The original framework included four layers: 1) an environment layer capturing the ambient socio-economic context of actors constraining fields of action within the other layers, 2) the infrastructure layer providing the technological backbone for applications, 3) an application layer describing features and rules constraining types of interactions possible and 4) the agent layer capturing the actions of actors within an application [110]. Hawlitscheck et al extended this framework by including a behavioural layer to capture the human interactions of actors that take place outside of the technical system and must be translated into it, recognising it is in this layer that the majority of the literature on trust can be situated [100]. They identify separation between the behavioural layer and the agent layer that human actors interact with as the *trust frontier*. The framework is reproduced in Figure 2.1. While the focus of the literature search in this research was primarily on trust within the sharing economy, much of it appears to translate to any digital interactions that involve humans and require trust. A key point they make is that while blockchain is unlikely to lead to *trust-free* systems, it might allow us to redesign the distribution of boundaries and responsibilities involved when placing and maintaining trust over technologically mediated communication systems [100].

Another area of active research in the design of trust within technologically mediated

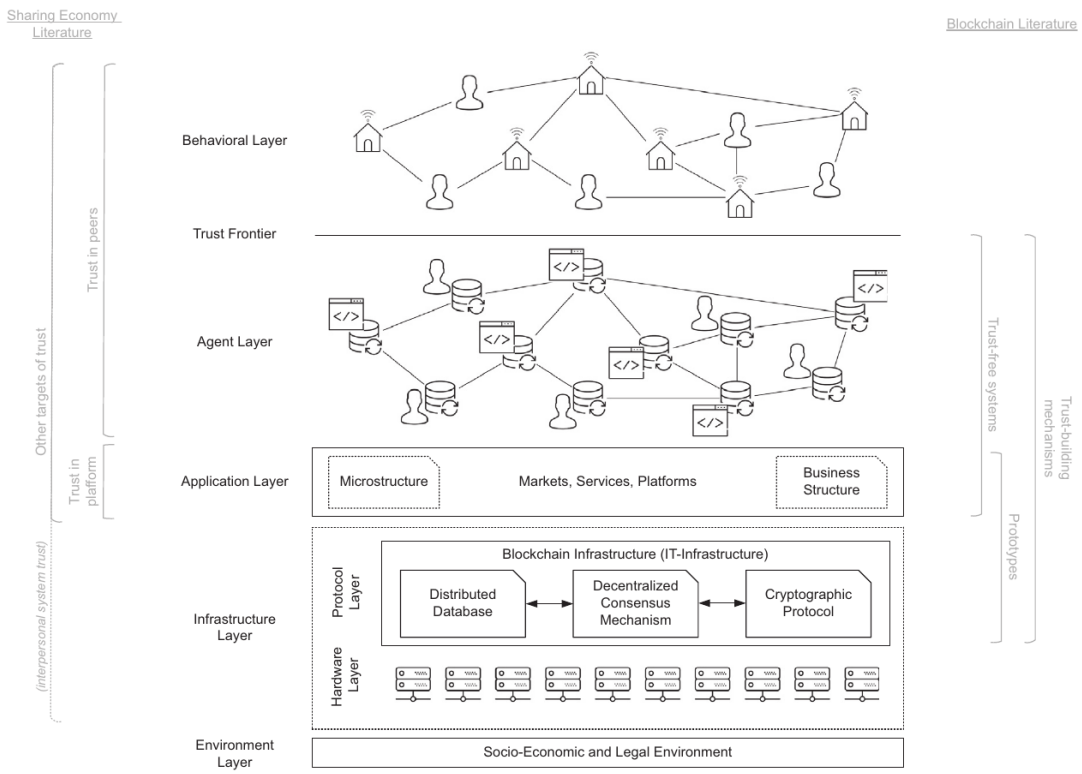

Figure 2.1: Extended Blockchain Engineering Framework [100]

interactions has involved framing it as a form of measurement or calculation [91]. This attempt to quantify trust appears logical. Although a large number of theorist would claim trust is a belief, feeling or attitude that is subjective and personal therefore cannot and should not be represented as a number [2, 82, 86]. Despite this, technological systems do have access to information that can be useful for actors making trust decisions. The TAPESTRY project presents an interesting example of how this information might be presented to individuals. They use timeline-activity proofs as evidence of trustworthiness which they attempted to convey to individuals through visualisations without providing a stamp of approval [111, 112]. The project intentionally aimed to represent timeline information through complex visualisations which individuals would have to interpret and learn. The ability to transform raw data into patterns presents a novel, decentralised model for how we might begin to design more useful mechanisms for evaluating trustworthiness and intelligently placing trust in an online environment. Mechanisms that remove the dependence on a central arbiter of trustworthiness. In addition to this academic research, many organisation have started to build in *trust*

pauses into their applications, adding friction and emphasising the responsibility of the individual to place trust [16].

Finally, understanding of trust in both virtual and potentially physical settings can be drawn from extensive research into multi-agent systems [113, 78]. Early experiments modeled game theoretical interactions between agents, such as the prisoner's dilemma. These were then simulated over multiple interactions, identifying different strategies and their effectiveness over different population sizes using an evolutionary model that replicated the most effective strategies [113]. Results indicated that the fewer repeat interactions between agents led to the replication of distrusting strategies. They also clearly illustrated the importance for recovery from mistakes, acknowledging that uncertainty can lead to misinterpretation of intent. For example, generous tit-for-tat strategy out-competing the classic tit-for-tat in simulations [114].

More recent research into multi-agent systems explored the notion of creating zones of *default, diffuse trust* in virtual environments. The research presented by Buechner and Tavani applied two policies to agent's decision making, with agents modelled under the SPIRE framework [78]. Agents collected into teams, with teams fulfilling a certain activity by delegating tasks to agents, each task had an associated incentive and time to completion. Agents could also receive external offers; outwith the teams they were participating in [115]. The intention with this framework it to model rational agents constrained by social norms and embedded within teams. The authors suggest this to be representative of default, defuse trust environments in real-life. These involve dynamic and distributed relationships, with human actors joining and leaving teams while fulfilling commitments over time. Within the environment simulated they applied two policies, one which used past behaviour to rank the commitment of an agent towards a task, the other a discount-based policy which determined the benefits an agent could get from breaking a commitment [78].

2.3 Privacy

The desire and expectation for privacy has been perceived by individuals and groups throughout history [116] and increasingly it is being acknowledged as foundational for individual freedom, social relationships and the functioning of society [7, Chapter 4]. Like identity and trust, privacy is intrinsically connected to the information flows produced through interaction. While identities structure meaning from informational inputs and trust extrapolates from this meaningful information to positive expectations about future actions in order to inform action selection, privacy can be understood within this interactive setting as the constraints entities attempt to place on information flows as they seek to influence the definition of a situation. This section presents privacy as it has evolved throughout history, demonstrating how the changing environment and introduction of successive ICTs have transformed the potential for information exchanges, which in turn has evolved the possibilities for control and influence over these flows. As the nature of interaction changed, new risks emerged, and new social norms and regulations were developed to manage them.

As with the previous sections on identity and trust, privacy is placed in its modern context - an information society. This has led to an explosion of new information flows, communication mediums and intermediaries. It has fundamentally reshaped our social fabric and the ability for some actors to wield power as they hold and abuse their positions at the centre of these digital information flows [70, 74]. While the information society has exacerbated the imbalances of knowledge and power, it has also led to new possibilities for structuring the transparency of societies information flows. Novel cryptographic techniques can be used to authenticate information and actors as well as define intelligent constraints on meaningful access to information. All with mathematically provable security guarantees [117]. This cryptography is now practically implementable, and rapidly being implemented and made widely available.

2.3.1 Privacy Throughout History

Privacy does not have a defined beginning, it is something that has been desired and managed by individuals, families and communities in some form for as long as humans have been communicating [116, p. 2]. It is the forms, features and expectations for privacy aspired to by individuals embedded in the social structure of their time that has evolved. The gradual development and inhabitation of the urban environment along with new technologies for information storage, publication and dissemination created new possibilities, constraints and threats for privacy.

Vincent uses the home to illustrate this point. Inside this domestic space the occupants enjoy relative privacy and entering this domain has long held significance. Walls, windows, shutters and doors shield actions and communication from observers. Privacy was expected even when walls were insufficient to contain the sounds from within. However, the home was often inhabited by many occupants within a familial unit, and within its walls privacy would have been limited, with only the very wealthy able to afford internal private spaces. These were often reserved for intimate communion with god [116, pp. 2-15].

Gradually urbanisation and the migration towards large cities changed the nature of privacy. Individuals might inhabit shared lodging with unfamiliar others and live in a city of strangers. The crowd created new possibilities for personal privacy and self-disclosure as those you interacted with on a daily basis were no longer intimately familiar with your life as was common in smaller rural villages. At the same time, it presented new risks for being observed by unknown others in the street. This led to the development of street etiquette, social norms based on seeing only what was necessary to navigate while avoiding windows and face to face contact [116, pp. 26-53].

Alongside the reconfiguration of the built environment, another transformation took place. The introduction of successive ICTs and the gradual education of the populace in their use [118]. Increasingly information exchanges were mediated by postal services, whose networks became ever more widely spread, cost effective and efficient since the passage of the Post Office act in 1660 [116, pp. 46-49]. This widening of the sphere of

privacy, created new challenges for existing authorities as Vincent illustrates with the case of teenage lovers arranging a private rendezvous beyond the awareness of their parents [116, p. 49]. This new dimension to information exchanges also created new risks, as it increased the likelihood that communications would be read by those other than the intended recipient. While the postal service was bound by to confidentiality by legislation, the British state reserved the right to open letters sent from those designated *enemies of the state* [116, pp. 50-51].

Two world wars, and the subsequent bolstering of the state had implications for privacy in a number of ways. Citizens had to be identified and registered with the state, passports identifying individuals as citizens of a certain state were required to travel abroad and vehicle licences were introduced [116, pp. 100-101]. Furthermore, the world wars emphasised the importance of the both the surveillance of and protection from surveillance of private communications. Accelerating the innovation in both offensive and defensive cryptographic methods [116, pp. 104-110]. These changes reflect the state exerting more influence over the private sphere of its citizens lives through identification, regulation and surveillance.

The history of privacy from the 1600's can be viewed in this manner. Successive information and communication technologies combined with changing environments have repeatedly transformed the potential for information exchanges and the reasonable expectation of privacy that individuals could place in these exchanges. These changing modes of communication have challenged existing power structures, while at the same time introducing new intermediaries and centers of power which must be trusted. As well as creating new opportunities for both realising privacy and surveillance.

2.3.2 Defining Privacy

One of the challenges for researchers, judges and legislators throughout history has been to parse the meaning of privacy and both determine and justify our rights in respect to this meaning [119]. Different environments, cultures and social settings give rise to

different expectations of privacy. They create different private spheres that might need to be protected or regulated in some way. In this section early conceptions of privacy are summarised before reviewing Nissenbaum's work defining privacy as contextual integrity that she presents as a justifactory framework adjudicating on privacy violations in novel future systems [7].

Privacy has been conceived of as control over the information others know about ourselves and the constraints around which this information can be exchanged with others. Westin's work Privacy and Freedom is frequently cited by researchers and uses the following definition *Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how and to what extent information about them is communicated to others* [120]. Gerety claims this is too vague a definition to be useful [121]. Other researchers have sought to define privacy in terms of the degree to which others have access to information about another, including an experiential aspect as opposed to solely informational [122, 123].

2.3.2.1 Privacy in Private and in Public

A region, defined by Goffman as a place bounded by barriers to perception [5, Chapter 3], is often used to delineate the boundaries of privacy. With regions that are highly bound, such as the home, a vehicle, are often referred to as the private sphere of an individual or families life in which their right to privacy should be protected from the arbitrary interference of others [124]. Although trusted actors can be granted special powers under specific conditions in which they are deemed morally and legally justified to intrude in this domain. The 1908 Children's Act in the UK provides an example of this [116, p. 80]. Nissenbaum notes that this public/private distinction is also applied to information, quoting the works of Parent, Wacks and Gerety all of whom describe some form of personal information - private facts about an individual which they often choose not to reveal themselves [7, pp. 96-98, 125, 121, 126]. Parent, in particular, believes that an individual should only have an expectation for privacy when personal information disclosed about themselves is not available in publicly accessible documents, for example, a newspaper or government records [125].

2.3.2.2 Privacy as Contextual Integrity

Privacy as contextual integrity is a framework introduced by Nissenbaum, that places information flows within the social context that they take place. As the earlier Section 2.1.5 on identity has already reviewed, the idea of a social context is taken from a well developed social theory with origins in the work of Goffman among others [5]. This in turn influenced Luhmann's theory of trust [2] and remains actively developed and applied by Burke and Stets in identity theory who refer to it as a social structure [39, 4]. Nissenbaum describes a social context as a structured social setting defined by a set of roles, activities, norms and values. Individual actors perform roles determined by their position within the social structure; these roles have normative expectations associated with them prescribing actions and behaviours of the actors. A social context can be implicit, with normative behaviours learnt and regulated through participating in interaction within the context, or they can be partially or fully formalised with roles and expectations institutionalised and regulated by organisations and legislation [7, pp. 96-98].

Contextual integrity identifies these social contexts as the system that regulates the appropriate flow of information between actors performing roles constrained by *context-relative informational norms*. These can be characterised by four parameters: contexts, actors, attributes and transmission principles. Information flows from a sender to a recipient and might identify a set of information subjects; all are instances of actors within this framework. Additionally, actors fulfil roles meaningful under a context. Identifying the roles is an important aspect of specifying the informational norms. Information flows are said to contain specific attributes describing the types of information being exchanged. Finally, transmission principles are the constraints placed on the flow of information between parties [7, pp. 140-147]. Nissenbaum identifies confidentiality, entitlement and consent as examples of transmission principles [7, pp. 145-156]. These norms *define and sustain essential activities and key relationships and interests, protect people and groups against harm, and balance the distribution of power* [7, p. 3]. The central thesis of privacy as contextual integrity, is that people feel

their privacy has been violated when these norms are breached.

Nissenbaum acknowledges the framework she proposes is conservative and could lead to beneficial information systems being rejected on the basis of privacy concerns. As such contextual integrity also provides a means by which breaches in informational norms may be evaluated. First against general moral considerations, such as potential harms or unfairness of the proposed system and then judged in terms of the new systems contribution to the purposes and goals of the social context [7, pp. 161-169]. Another criticism of the contextual integrity framework is that social contexts, actors, roles and transmission principles are rarely formalised and agreed upon. Rather they are subjectively perceived by unique individuals who will often come to different ideas about their expectations for appropriate flows of information within the context [127]. This criticism highlights the critical role of the design of these systems, ensuring that a diverse and representative population for whom the information system is designed to serve have a voice in shaping and evolving the informational norms governing the system.

2.3.3 The Value of Privacy

The right to privacy was first articulated by Brandeis and Warren in 1890 [119] and enshrined in the 1948 universal declaration of human rights [124]. Researchers have attempted to define the value inherent in the preservation of this right. What follows is a brief overview of these arguments, roughly following Nissenbaum's categorisations of its importance to individuals, social relationships and society [7, Chapter 4].

2.3.3.1 Privacy and Individual Autonomy

Privacy has been identified as a key property of an environment that promotes individual autonomy [7, pp. 82-84]. Gavison details its importance for enabling individual experimentation and creativity away from the judgement of others [123]. Even the perception of being viewed, in what Reiman describes as *an informational panopticon* can lead to self-censure as individuals view themselves from the perspective of others [128].

This has a chilling effect on both thought and behaviour as individuals begin to process experience and select actions based on that which is deemed socially acceptable.

Cohen makes a similar argument, stating that privacy allows for the *conscious construction of self* [129, 7, p. 76]. Emphasising privacy's importance in self-determination, identity formation and the capacity for independent moral judgement [7, p. 81, 123]. This aligns with Goffman's description of the backstage and its importance to the presentation of self [5].

2.3.3.2 Privacy and Social Relationships

Much has been written about the importance of privacy in fostering human relationships. Nissenbaum quotes philosopher Charles Fried who describes privacy as fundamental to relationships of respect, trust, love and friendship [7, pp. 84-85, 130]. Indeed, from the previous section on identity we know that an interactive context is co-defined by the self-presentation and selective disclosure of information about oneself to others over time (Section 2.1.2). The literature on trust indicates that making oneself vulnerable through the disclosure of sensitive information can provide opportunities to grow trust and demonstrate trustworthiness in a relationship (Section 2.2).

Rachels shares the view that privacy is foundational for social relationships. The ability to vary our behaviour and self-disclosure play a role in defining the relationship we participate in [131]. O'Neill points out that we are rarely transparent about all aspects of our lives in any of the relationships we have [85]. By disclosing the information deemed appropriate for the relationship, individuals are able to perform the diversity of roles we play in society without contradiction [131, 5].

2.3.3.3 Privacy and Society

While a large amount of attention has been placed on justifying the importance of privacy for protecting and empowering individuals within society, Nissenbaum reviews the work of Regan who argues for the importance of privacy as a common value, public value and collective value [7, pp. 85-88, 132]. As we saw in Section 2.1 on identity when reviewing the work of Burke and Stets, individuals are embedded within social

structures that they continuously recreate through their actions [39, 4]. Therefore, it appears logical that the role privacy plays in promoting individual autonomy and freedom of expression are vital aspects of a thriving free, democratic society, as Regan argues [132].

2.3.4 Privacy in an Information Society

While it might be argued that the information society began with the introduction of the printing press and the ubiquitous access to the postal service, it is the digital age that has accelerated this transformation and brought a difference of scale, speed and scope to information flows within society [116, Chapter 5]. Nissenbaum identifies three interdependent, mutually reinforcing classes of informational capabilities which have seen exponential improvements since the beginning of the digital age. These are: monitoring and tracking; dissemination and publication; and aggregation and analysis [7, pp. 21-64].

As with earlier ICTs, digital technologies have been leveraged by individuals, organisations and society to create socio-technical systems with new possibilities for both interaction and control. This has produced opportunities to challenge existing structures of power, whilst simultaneously creating new mechanisms for the powerful to exert influence over others by mediating societies information flows. In some cases, old intermediaries have been displaced by digital technologies and the organisations that produce and control them without the requisite mechanisms for accountability, regulation and recourse that societies have developed to constrain those in positions of power [133]. Zuboff coined this new frontier the age of surveillance capitalism, in which information about individuals both given and given off as they interface with digital systems is extracted as the raw material used to feed algorithms which infer and assign labels that are then used to predict effective informational nudges with the aim of influencing the future behaviour of an identified, classified population [74].

2.3.4.1 Surveillance Capitalism

The digital age unlocked new possibilities for communication, interaction and commerce. Aspects of life that were only possible through shared physical presence have been gradually translated and reproduced in virtual environments. The search engine unlocked unprecedented access to the world's information. However, in the early 2000s many of these companies were lacking a viable business model. Zuboff reviews a Google patent that highlights its transformation into a surveillance driven model for advertising, with information collected about an individual used to uniquely identify and target them for advertisements despite it not always being voluntarily given due to their privacy concerns [74, pp. 77-80, 134].

This business model developed an *extraction imperative* that spread to many other organisations giving rise to a race for control over the world's digital information and its mechanisms for production [74, pp. 128-131]. Spheres of life that only ever existed ephemerally, with information and meaning persisted in the human memories of those who were present, were subject to a *cycle of dispossession* in a process Zuboff compares to colonialism. In this cycle, companies first make *incursions* into unprotected digital territory, often meeting strong resistance from those being dispossessed. Next follows a *habituation* stage as this initial incursion becomes the new normal for society due to slow-moving legal challenges and intentional misdirection from companies involved. Once legal challenges and other avenues of complaints catch up, companies move on to the *adaption* phase, on which they apologise and promise to improve in the future. Finally, companies *redirect* their initial incursion efforts towards new *surplus* behavioral information usually increasing the scope of their data extraction [74, pp. 138-158]. Zuboff illustrates this cycle with the example of Google Street View which challenged our right to privacy on our streets [135] by turning a public good into a private asset without permission. This gave Google the power to decide how this information was used, presented and interpreted. See Figure 2.2 for an adapted timeline of this cycle of dispossession taken from the Electronic Privacy Information Center (EPIC) [136].

The expansion of social media platforms is another example of the increase in

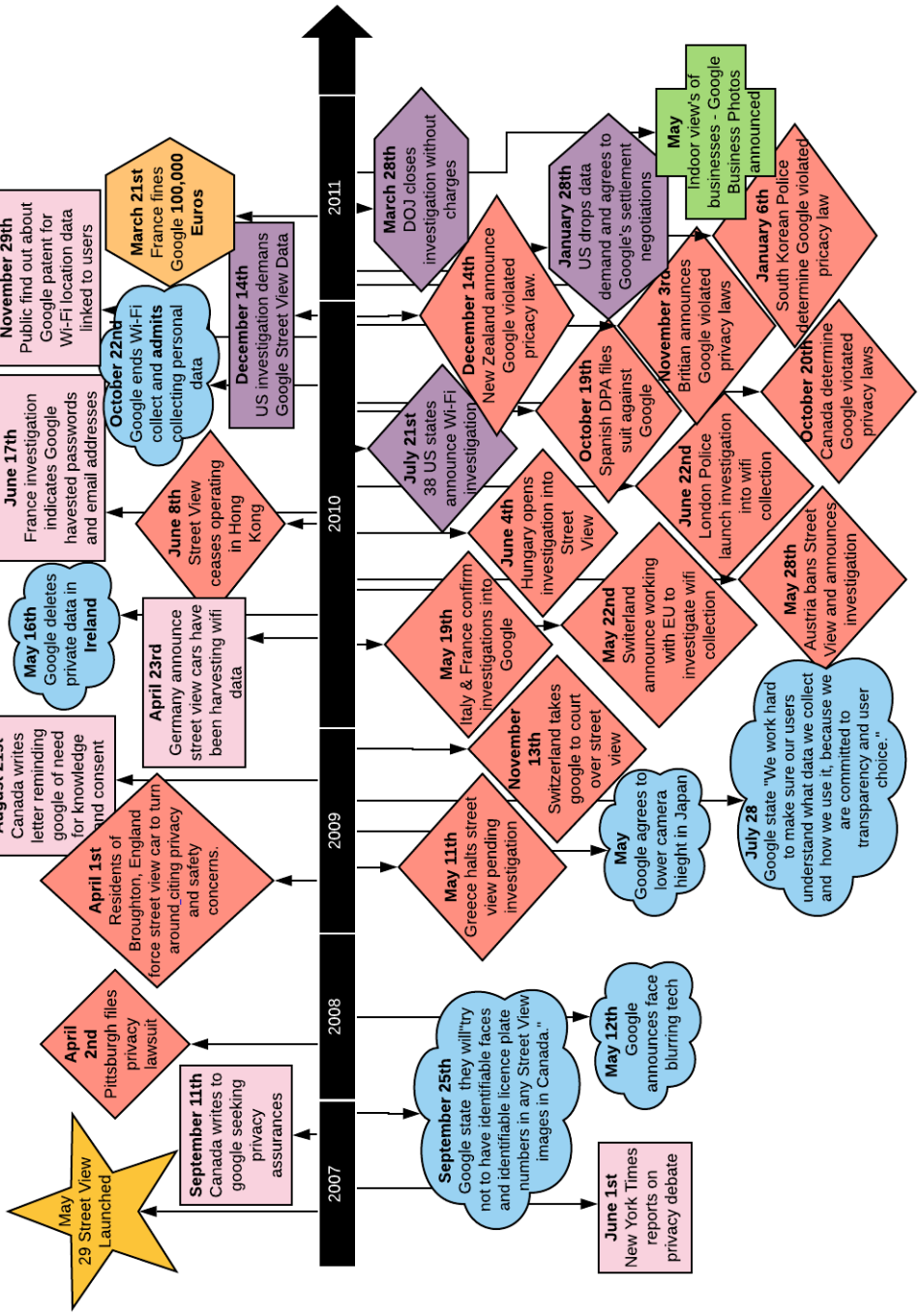

Figure 2.2: Google Street View Timeline (adapted from the EPIC [136])

the scale and scope of the information flows possible in society. With the platform providers becoming powerful intermediaries able to design virtual environments for their purposes. For example, the introduction of the Facebook like button designed to track, trace and monitor individuals across the internet, even those not registered as a Facebook user [137]. Furthermore, research has shown how information from social media interactions can be used to predict the personality of an individual [138, 139]. Rather than our social media profiles presenting ourselves as we would like to be presented, our idealised selves, they have repeatedly been shown to be accurate predictors of our actual personality. Kosinski and Stillwell proved how Facebook likes can be used to automatically predict private traits about an individual without their consent or knowledge [140], noting that these predictions are potentially facts the individual would not be willing to share. Further breakthroughs in this field of research in 2015, indicated that a computer can now more accurately predict your personality than a human [141].

These powerful prediction tools have begun to work themselves into hiring processes [142], credit scoring [143], policing [144] and even more insidiously, they are being used to manipulate democracy itself [145]. This is despite a plethora of evidence pointing to the biases encoded into these algorithms by their designers and the data used to train them [79]. Introducing, extending and amplifying discrimination against individuals from social categories in structurally less powerful positions in society. For example, racial and gender minorities [81, 146].

To detail all applications of digital technologies that threaten individual privacy, autonomy and freedom is not possible in this work. However, the trend is clear and can be seen in both the publications of Nissenbaum and Zuboff [7, 74]. Information is being captured through our interaction with and across digital systems. The primary purpose for this often appears innocuous and the data captured inconsequential. The concerns arise when this information is correlated with other information, aggregated and analysed for insights. Especially when this occurs beyond our awareness and is motivated by profit. The realisation that profit could be made through the surveillance capitalism model led to an increase in breadth, depth and scale of the data points col-

lected, aggregated and analysed. Our experiences, location, bodily health and emotions began to be viewed as business opportunities containing rich information sources that could enable our selves to be rendered in increasing detail [74, Chapters 8-9].

2.3.4.2 Privacy-enhancing Technologies

A set of countervailing technologies are being developed with the promise to preserve and enhance privacy in the information age. These ideas can be traced back to theoretical cryptographic protocols conceived in the decade following the publication of new directions in cryptography by Diffie and Hellman in 1976 [26]. Many of them proposed in direct response to the perceived impact the architecture of these systems could have *on the centralization of our economic system, on some of our basic liberties and even on our democracy* [147]. This resulted in an entire subculture loosely referred to as Cypherpunks, determined to make privacy a reality in an electronic society [148].

These early ideas include; untraceable electronic mail [149], blind signatures to allow the signer to create a signature on data they do not need to see [150], credential mechanisms to provide security to digital systems without the need for unnecessary identification of individuals [27], untraceable electronic cash [151], secure multi-party computation (SMPC) [152], homomorphic encryption to enable computation on encrypted data [153], group signatures allowing statements to be signed by individuals on behalf of a group such that the signer is indistinguishable from other group members [154], cryptographic commitment schemes supporting a binding private commitment to a value without revealing any information about the value committed to [155], zero-knowledge proof systems allowing a prover to demonstrate the completeness and soundness of a computation without revealing any information other than its result [156] and their special case zero-knowledge proofs of knowledge allowing a prover to demonstrate knowledge of some x which satisfies equation y while only revealing one bit of information to the verifier (True or False) [157, 158].

These ideas all apply mathematical knowledge to define and guarantee properties of digital information flows, determining which actors learn certain information under what set of constraints. This has been discussed by Trask et al, in what they describe

as the structuring of transparency within information flows building on the work of contextual integrity [117, 159]. Many of these theoretical ideas have now been shown to be practical, and are the focus of active, often open-source development highlighting the growing attention towards these technologies and their increasing availability and potential when designing real world socio-technical systems. Examples include; OpenMined[1] a open-source community developing tools for privacy-preserving machine library who's flagship library Syft contains implementations of Differential Privacy [160], Homomorphic Encryption [161] and Federated Machine Learning [162], ZCash[2] a privacy-preserving decentralised currency that leverages a Zk-SNARK based zero-knowledge proof system [163] including a protocol specification [164] and open-source implementations of the underlying primitives and Hyperledger Ursa[3] which has practical open-source implementations of a number of group signature schemes that support blind signatures and zero-knowledge proofs of knowledge on a signed, selectively revealable array of messages [29, 30, 31] as well as other cryptographic primitives.

2.3.4.3 Changing Norms

Context-relative informational norms are central to Nissenbaum's thesis outlining privacy as the appropriate flow of information [7]. The digital information age has created new contexts, for example social media, and disrupted many others. Where this has occurred the framework of contextual integrity asks us to examine these changes, the new actors, information flows and transmission principles introduced. Then evaluate these changes against the contextual values and normative expectations of those within the social context under study. The resulting evaluation should the be used to prescribe a course of action, be that resistance or acceptance towards the changes introduced by a socio-technical system [7, pp. 190-191].

Human society is at the beginning of the digitally accelerated information age, which has seen an amplification of our power to publish, disseminate, store, aggregate and analyse information. As discussed in Section 2.1.5, social norms take time to develop

[1] https://openmined.org/
[2] https://z.cash/
[3] https://github.com/hyperledger/ursa

this can be seen throughout the history of privacy as individuals figure out what is possible, socially acceptable and thus reasonable for them to expect of others [116]. Facebook appear to recognise the importance of social norms as evidenced by Mark Zuckerberg's statement in 2010 claiming that *users no longer have an expectation of privacy* [165]. Despite this claim, there is growing evidence that societies throughout the world are forming alternative perspectives. Legislation such as the European General Data Protection Regulation (GDPR) [166] and the California Consumer Privacy Act (CCPA) [167] seeks to protect data subjects and outline responsibilities of organisations holding personal information. Furthermore, widely disseminated documentaries released on Netflix have brought these issues further into the public consciousness. These include: The Great Hack (2019) outlining how big data was used to influence the US 2016 election and the United Kingdom (UK) Brexit referendum [168], The Social Dilemma (2020) a docu-drama providing insight into how social media platforms are designed to maximise user attention as a singular metric for success [169] and Coded Bias (2020) documenting Buolamwini's discovery of racial bias in facial recognition software and subsequent exploration of biases in algorithms used throughout society, often without any accountability [170].

2.4 Critical Discussion

While the construction of identities can be abused, there exists a very real need for inter-subjective meaning within social interactions. They define the behavioural expectations one can reasonably expect from another within a specific social context. Trust is the confidence that some unpredictable entity, understood in relation to a collection of identities, will uphold their commitments to the behavioural expectations defined by these identities. Trust, when placed in the trustworthy, allows us to select our actions from a richer set of possibilities and respond creatively to uncertainty. However, misplaced trust in illegitimate others leaves us vulnerable to manipulation. In order to place trust intelligently we require information about the entity, information that itself must be judged for trustworthiness which further requires attribution to its source and

evidence of its truthfulness [86, 70]. Where the interacting parties know each other with familiarity, trust can be based on information interpreted from first-hand experience or from sources whose authenticity can be judged personally. However, in order for society to scale mechanisms to allow trust to be placed impersonally in unknown others were required. Trust, in specific sub systems of society became institutionalised with roles and associated expectations explicitly defined, rules and consequences formalised and an entire apparatus of organisations, regulations and assurance processes with which to enforce these institutional pressures [15, 2, pp. 53-67].

Through systems trust, individuals are able to place trust in unknown others on the basis of the role they are assumed to be performing. Trust is placed in the shared meaning and behavioural expectations of this role within society and the system of institutional pressures designed to induce trustworthy behaviour, reduce defection and wield distrust impersonally [15, 2, pp. 53-67]. It is this aspect of trust that this thesis is primarily concerned with. Specifically, the systems of identification that have been developed to identify actors as holders of role identities within highly structured sub systems of society and how they might be augmented with advanced ICTs whilst protecting individual privacy. Which, as the privacy literature reviewed in this chapter emphasises, has been significantly impacted with the introduction of ICTs that have reconfigured societies information flows and led to the over identification of subjects and their correlation across contexts [7]. However, the literature also indicates that advanced ICTs have enabled the intentional structuring of these information flows through use of cryptographic protocols so that informational flows appropriately within a social context [117].

2.5 Conclusion

Identity, trust and privacy are all complex terms to define, while at the same time being integral to our lived experience. This thesis suggests that we can all recognise that we take on various identities at different moments in our life, while never being wholly defined by them. Whether it is that of a student, parent, Yorkshireman or something

personal that ties you in relationship to another unique entity. These are labels that we use to classify our experience, to infer meaningful information from it [4]. While the labels might remain static, the collection of meanings these labels identify are individually perceived and continuously shaped through a process of interaction. By applying these labels reflexively to ourselves and others within a specific interactive setting, individuals are implicitly bound to society through a web of commitments that set behavioural expectations and influence their selection of actions.

Identities are not solely self constructed, but have meanings that are intentionally curated and persisted within social systems to structure interactions within a social context. These meanings can intentionally, or unintentionally, characterise a certain group unfavourably in relation to another, which when learnt and applied by individuals introduces biases that influence the distribution of power and access to resources within a social system [53]. The intentionally constructed environments of our information society emphasise how the ability to understand and influence the identities individuals hold can be used to tune, herd and condition behaviour in both virtual and real environments [74]. Although Goffman's work shows the desire to project and control the definition of a situation has always been a part of human interaction [5], it is the change in scale and scope, as well as the displacement of human actors with algorithms coded by unknown others that should give us pause. Especially considering these have repeatedly been shown to encode the implicit biases of their designers [146].

Third Chapter

Identification Systems

> *"There are no greenfield identity systems - People have been managing identities for millennia, adapting their behaviour and practices in relation to each new identity technology that becomes commonplace. Even the most advanced digital identity platform will be perceived and engaged with in relation to all of the other identity systems, digital and analog, an individual has experience with."*
>
> Identities Report, Caribou Digital [40]

Identification systems encapsulate a set of processes, actors, infrastructures and interactions by which ID artefacts attesting to a set of attributes that characterises an entity, are issued, and validated, in order to facilitate the formal identification of individuals as they interface with institutions [21]. Using the language of identification systems, as opposed to identity systems, intentionally directs focus towards the administrative nature of these systems in which complex human beings with their relational, multi-dimensional and dynamic identities are represented within static records [23]. As Gelb and Metz point out, ICTs have primarily led to a revolution in systems of identification which are rapidly seeping into many aspects of our daily lives [22]. Although the distinction between identification and identity is useful, it is important to emphasise that they cannot be considered independently. Our identities influences how we are, and would like to be, identified and how we are identified and represented impacts how we are perceived, the resources we have access to and the field of action available to us [4]. Which in turns shapes the identities we form and apply. This complex, dynamic interplay between identification and identity has only accelerated and intensified with

the increased prevalence of ICTs [171, 21].

The chapter draws on literature from two distinct, but closely related fields of research in order to understand how identification systems have and are being created, why they are necessary and what risks they can introduce. These two fields are: 1) the provision of legal identity for development and 2) digital identification systems. Legal identity and the means to prove this status is seen as critical for sustainable development as acknowledged in the UN Sustainable Development Goal (SDG) 16.9 to provide legal identity for all by 2030 [14]. Setting this goal has accelerated the adoption of identification systems, many digitally augmented, throughout many different countries and contexts. The subsequent analysis of these systems and their impact has produced a rich and invaluable set of literature based on real-world experiences of the people who have had to navigate these systems [23, 172, 40]. In contrast, digital identification systems originated from a functional requirement for siloed information systems to perform identifications as a mechanism to authenticate actors against their virtualised representations simulated within digital environments in order to authorise their actions [173]. Both types of identification system have become inextricably intertwined. As the capabilities of ICTs increased, they began to facilitate increasingly complex, valuable socio-economic interactions. This required higher levels of assurance in the identifications associated with virtual entities in order to mitigate the institutional risk inherent in these interactions. To offset this risk, digital identification systems often require proof of legal identity, ID artefacts, which are used as *breeder documents* to establish adequate levels of assurance in virtual entities [174, 175]. At the same time, national, supra-national and non-governmental organisations (NGOs) are increasingly turning to technological identification solutions to bridge the legal identity gap and *make people visible as holders of rights* [23].

Advanced digital ICTs change the design space for realising identification systems. Personal information can be captured, stored and retrieved with relative ease; however it must be represented within a machine-readable format [23]. The information that is captured on registration, how it is verified, where it is stored, how it is used and by whom are important questions whose answers have implications for the privacy of

those identified and the agency they have over their representations within the system. It also has implications for security, accountability and the distribution of power within the system. ICTs must interface with both human users and other technological systems [9, Chapter 2]. The design of user interfaces impacts the accessibility and usefulness of these systems to end users, while the use of open, interoperable standards for protocols and data models prevents vendor lock-in and promotes technological neutrality [176]. Finally, the information captured and remembered through the process of identification as entities interface with ICTs, including metadata about when, where and whom were involved, can be beneficial for accountability but also can seriously compromise user privacy. This is especially true if it includes identifiers that can correlate individuals across contexts [22].

While ICTs have changed the design space, it is human actors, organisations and institutions within a social context that determine the requirements and purpose for identification systems. And primarily, it is human actors who are identified through the processes of identification. It is only through understanding a specific social context that the legitimacy of such a system can by judged [7, pp. 129-145]. Legal identity, especially in relation to sustainable development, provides a rich set of literature drawn from real-world experiences where identification systems intended as a public good designed to build capacity, recognise individual rights and enable opportunities such as civic participation have led to unintended consequences. These include exclusion from vital services, the appropriation and misuse of personal data and the facilitation of surveillance architectures [22, 23]. That evidence suggests these consequences disproportionately impact those already structurally disadvantaged in society; women, minorities and those from rural or lower income backgrounds indicates we must carefully consider the edge cases when designing systems of identification so as not to encode and amplify existing institutionalised discrimination [14, 23]. Furthermore, the literature surrounding systems of identification that facilitate the provision of legal ID documents emphasises that human processes for governance and accountability must be built-in and cannot be realised through technological solutions alone [22].

3.1 Terminology

The terminology presented in this section draws on language from both digital identification and identification in sustainable development literature. First identity, identification and ID are defined following the recommendation from Caribou Digital, who have expertise in researching and advising on real-world identification systems [21, 40]. Then the terminology for discussing entities, including, but not limited to, individuals, represented within identification systems is introduced. This has been adopted from Jaquet-Chiffelle et al who make a clear distinction between the real and virtual environments and the entities represented within these [173]. Finally, terms for discussing privacy following Pfitzmann and Hansen are introduced [48].

- **Entity**: Any object, person, group or thing that has a distinct existence. Can be both physical and abstract, either virtual or legal [173].

- **Identity**: Identity is reserved for that relational, intangible and fluid construct that we produce, hold, negotiate and perceive in ourselves and others as we engage in a process of interaction. The literature surrounding this understanding of identity has been reviewed in Section 2.1.

- **Identification**: Identification is intrinsic to interaction as individuals self-identify and perform identification of others they are interacting with, in order to apply the meanings they associate with these identifications [177]. However, in this thesis identification primarily refers to the formal, administrative kind. Identification in this context is the set of processes or interactions by which a subject becomes identified by an evaluator performing the identification, often on behalf of an institution [177, 21]. It generally involves the subject disclosing a set of claims about themselves, which the evaluator must assess for validity in order to make a decision about the access, rights and responsibilities the subject should be granted.

- **ID**: A tangible artefact, often issued by a third party, that can be used to support a set of claims. For example, a credential such as a driving licence.

- **Identity-related information**: Any information that characterises an entity, not necessarily uniquely, within a context [173]. Identity-related information is disclosed through a process of identification.

- **Identifier**: An identifier is anything that uniquely characterises a single entity within a specific context [173]. Identifiers might be assigned bit strings generated from a subset of identity-related information, or they might be provided by the subject being identified during registration (e.g. a username). Such identifiers are often referred to as pseudonyms [48]. However, this definition intentionally includes any set of identity-related information that can uniquely identify an entity within a context [173].

- **Authentication**: Authentication is the process by which assurance is increased in the association of entities with identifiers and identity-related information. This language comes primarily from digital identification systems, where entities present credentials alongside a pseudonym to authenticate against a previously established virtual identity [173]. However, we use authentication to include any verification of ID including both physical and digital artefacts. Authentication mechanisms used in these identification systems are categorised into three groups; something you know, something you have and something you are [178, 179].

- **Authorisation**: Authorisation is a decision about the privileges of an entity. What resources they can access, and which actions are they able to perform [33]. Authorisation decisions are made on the basis of identification of identity-related information and the authentication of these asserted claims [23].

From these definitions, an **identification system** can be defined as the assemblage of processes, institutions, technologies, laws, policies and other aspects that provision, issue, facilitate, authenticate, assure, audit and require ID to perform identification

in order to authorise access to services and resources [21]. Identification systems encapsulate the infrastructures and processes of identification that gives meaning and purpose to ID artefacts [171].

3.1.1 Legal and Virtual Persons and Identities

A legal person is an entity that can invoke the rights and responsibilities that an individual is normally able to when interacting with the law. It provides a layer of indirection distinguishing human beings from legal artefacts that represent them before the law. Furthermore, legal personhood can be applied to groups such as organisations or even natural objects like rivers [173]. The Universal Declaration of Human Rights states *everyone has the right to recognition everywhere as a person before the law* [124]. Legal identity is the means by which an entity can be recognised in the eyes of the law [171]. In the modern age, birth registration and the identification systems of nation states have become the defacto mechanism for individuals to exercise this right [14].

Jaquet-Chiffelle et al proposes a similar layer of indirection between the real and the virtual world [173]. The real world is where all actions originate, this includes human actors but also software agents. While the virtual world is an abstract product of human thought in which physical entities are represented and virtual entities with no link to the physical world are imagined. This model emphasises that the digital world is not the virtual world, but that the digital world can be thought of as creating representations of virtual entities [173]. The model then defines a virtual person as any virtual entity that has rights and responsibilities associated with them within a certain context. Under this definition, not all virtual entities are virtual persons, but all physical human persons are represented in digital environments by virtual persons [173]. Although, as with legal persons, they are not the only entities that can be, other examples are an intelligent software agent or a company.

This model explicitly moves away from a common misconception in digital identification systems that one virtual identity is (and should be) associated with one physical person. This comes from bureaucratic identification systems for managing legal iden-

tity which have the requirement to uniquely identify its citizens in order to provision services while preventing fraud. A virtual entity is modelled as having one tautological identity, the set of information that characterises an entity virtually represented within an information system [173]. This definition moves away from the sociological notion of identities as a collection of meanings perceived and applied during interaction presented within identity theory [61], to an understanding of identity as collection of information attributes correlated to an identifier [173]. Combining these two perspectives highlights that the collection of attributes correlated with an identifier characterising a virtual entity will be perceived through a lens of meaning by the informational agents that interact with it. Whether that be the human actor in control of, or represented by, the virtual entity, or the information system that must judge the affordances of the virtual entity within the system. Despite this dual meaning of identity, this thesis will use **virtual** identity in the manner proposed by Jaquet-Chiffelle et al [173].

The intentional distinction between the virtual and real worlds allows for actors, human or otherwise, to represent themselves as virtual persons in a way that supports the one-to-many and many-to-one relationships between the real and the virtual world. Human beings can represent themselves as a virtual person characterised by a being father, a citizen or an employee in the different contexts that they interact, without the requirement to collapse these virtualised representations to a single physical being. Furthermore, it supports the case where many physical actors wish to present their actions behind a single entity, for example an organisation or community. Finally, the concept of virtual persons includes the ability for a virtual person to mask their actions behind another virtual person. As such the model extends to the multiplicity of virtual entities we act through, with and behind, either as individuals or on behalf of a group (see Figure 3.1) [173].

Virtual entities when initially registered and represented within the system through a process of identification are assigned, or asked to provide, an identifier that uniquely characterises that entity within the domain of the information system they are represented within. This provides a means to correlate actions and information to a single virtual entity [48]. In contrast to Pfitzmann, the virtual persons model defines an iden-

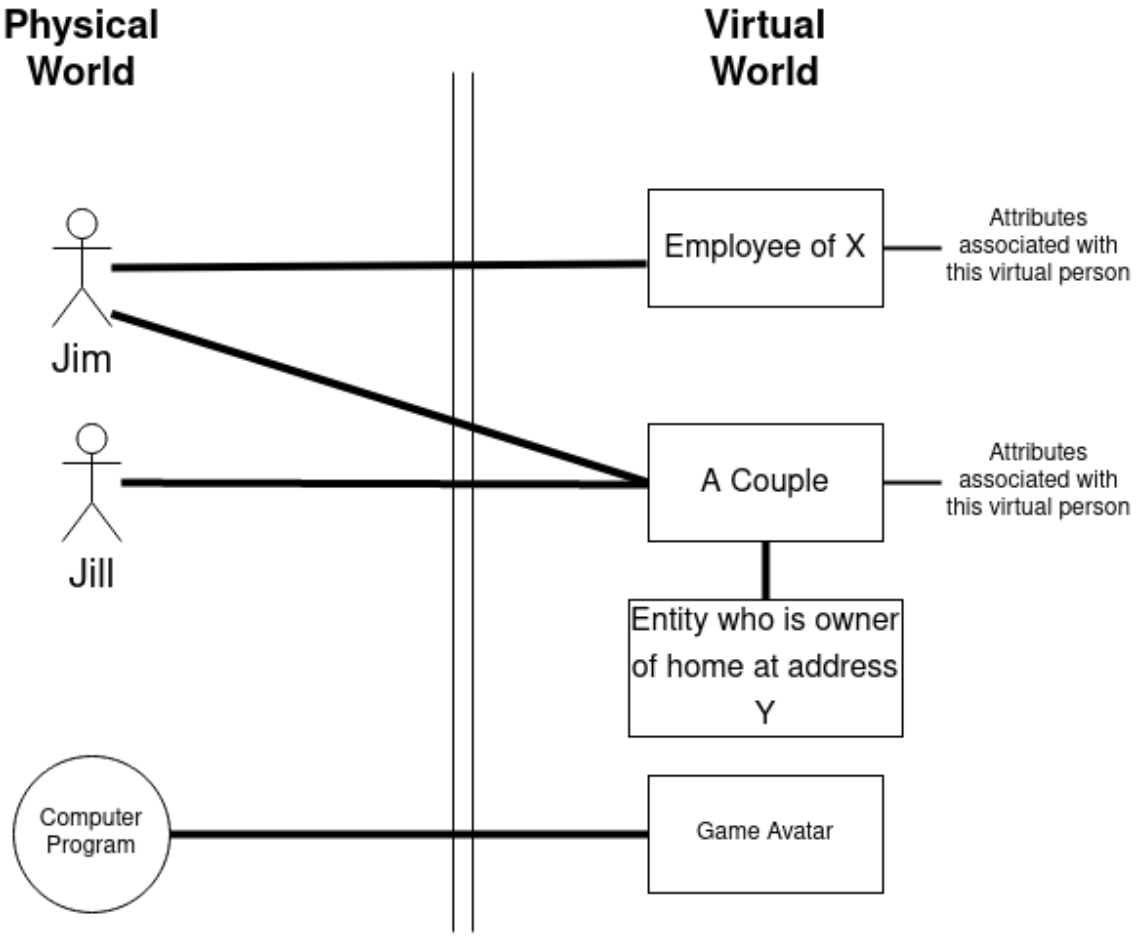

Figure 3.1: Relationships between Real and Virtual Entities (adapted from [173])

tifier as anything that uniquely characterises a single entity within a specific context. This acknowledges that attributes associated with a virtual person may also be used as identifiers [173, 48]. The use of identifiers, especially across multiple virtual entities, has implications for correlatability and privacy [48].

3.1.2 Privacy and Data Minimisation

Pfitzmann and Hansen developed terminology for discussing privacy and data minimisation within information systems modelled using entities (subjects and objects) and actions, with subjects executing actions on objects to change the state of the system. Generally, all information and communication systems can be represented as subjects (senders) sending objects (messages) to other subjects (recipients) [48]. They define Items of Interest (IOI) as the subjects, messages and actions represented within the system. While Pfitzmann and Hansen do not make the distinction between physical

and virtual entities for simplicity, modelling only virtual entities, this thesis prefers to use the model of Jaquet-Chiffelle et al to emphasise the distinction between physical entities and the virtual entities that represent them within informational spaces [173]. However, the terminology around privacy they introduce is useful and covered below:

- **Anonymity**: A subject is not identifiable within the set of subjects, known as the anonymity set.
- **Unlinkability**: An adversary is unable to determine if two or more IOI are related in some manner. Who the adversary is, depends on how the system has been characterised. **Linkability** is the negation of this.
- **Undetectablity**: An adversary cannot determine whether an IOI exists.
- **Unobservability**: An IOI is unobservable if it is undetectable by all subjects not involved with the IOI and all subjects involved remain anonymous even to each other.

All of these terms can be applied to senders and recipients independently as well as collectively to the relationship between a sender and recipient [48].

3.2 Digital Identification Systems

ICTs have, and continue to, transform how humans interact. Information exchanges can now be mediated across time and space by technological systems. This has required the design and implementation of identification systems functional within these new interactive settings. First communication technologies, from the postal service, to the telegraph and telephone created new mediums for information to flow between a sender and a recipient without requiring these entities to be co-present. The sender required a mechanism to identify and address the intended recipient, encode the information they wished to communicate within the medium of communication and identify themselves as the sender. The recipient had to decode the messages they received, identify the sender and interpret the meaning of the communication [11].

Each communications system came with its own set of properties, mechanisms for encoding and speed of information exchange. All early systems required human operators to mediate these information flows. The exchange of letters depended on the postal service for delivery, the telegraph required a telegraph operator to encode and decode telegrams before sending them over the wire and early telephone networks used switchboard operators to connect lines together [11]. Each new technology had implications for the privacy and security of the communications they facilitated. In early systems the recipient had no mechanism to judge the authenticity and integrity of a message. While for the sender it was hard to know the message and its meaning had been received solely by the intended recipient. Indeed, systems like the telegraph required the operator learn the message being exchanged. This led to an increase in defensive measures such as communicating in codes, which in many ways was a precursor to modern cryptography. For example, the Enigma machine used by the Germans in World War 2, gave them confidence that despite their messages being intercepted only those in possession of a corresponding machine and the codes required to work the machine could interpret the messages intended meaning.

The telephone is also illustrative of the evolution of communication systems and their mechanisms for identification of senders and recipients. A telephone is a technological artefact that encodes and transmits voice over a network. Initially, those with access to a telephone would call an operator who would connect them to the phone of the recipient they wished to communicate with. As telephone networks scaled, operators introduced a system for identifying phones to manage this increased complexity. Telephones registered with an operator were assigned a number, this number was correlated with an entity through an accompanying address book [11]. Eventually switchboard operators were replaced with a technological system removing human actors from the direct role of mediating information exchanges. Here we see an identifier assigned to a device so it can be identified within the communication system, this is combined with an information system (the address book) that associates identifiers to real-world entities. An early example of an identification system developed to meet the needs of a technologically mediated communication system.

The computer introduced a different set of problems. These machines, capable of storing and processing information at an unprecedented speed, scale and precision, were initially large, independent mainframe computers servicing a set of distinct human actors typically for research or administrative purposes. Access to the system was limited to those in direct physical presence of the machine, who would interface with it through a terminal. The number of potential bad actors and hence the risks of security breaches were minimal [180]. The information system itself became the recipient of informational inputs from senders and had to assign meaning to these inputs and execute actions on the basis of this meaning. Furthermore, these systems were first programmed by human actors who defined the set of actions, rules and capabilities of the system through software. It quickly became realised that the ability to identify and distinguish between the actions of unique actors represented virtually within an information system was a valuable feature. Thus, the account paradigm was introduced. Information systems were partitioned into a set of user accounts, with individuals identified within the system by a username. Passwords were introduced to authenticate individuals against usernames to limit and protect access within the system [33].

ICTs are continuing to transform society, Moore's law has increased the information these machines could store and the calculations they were able to compute. Computing devices have become smaller, more widely available and are gradually being integrated into everything [181, 9]. These information systems have been networked together, evolving into a communications system with access to ever expanding quantities of information. Actions can now be performed within networked information systems from virtually anywhere on the planet, increasing the threat surface and hence the requirement for security. Furthermore, information systems themselves have become intelligent agents capable of performing actions without input of human actors. As a result, ICTs have enabled increasingly rich and complex socio-economic interactions. Human societies have become dependent on these systems, nothing highlights this more strongly than the COVID-19 pandemic [182]. ICTs store increasing amounts of sensitive data and facilitate the transaction of increasing value. As a result they have become an attractive target for hackers and scammers who are in an evolutionary race

against the security engineers [15].

Digital identification systems have been conceptualised within this context. The need to identify and authenticate actors within information systems so that their actions can be authorised and correlated with an identifier. This was primarily a functional need, an information system can provide more complex, nuanced, personalised services if it has a mechanism to distinguish between the actions of unique actors [33]. As these systems became more complex, storing and facilitating increasingly valuable information exchanges, they became an attractive target for malicious actors seeking to abuse these systems to further their own interests. Achieving security within information systems by providing confidence in the authenticity of actions and actors became a strong focus in the design of these systems. The challenge is that the human actor is outside this system, represented within it as a virtual entity and given a means to authenticate against that representation. This introduces a *trust frontier*, because if the actions of the virtual entity are trusted because of the information known about the actor then the virtual entity representing that actor within the information system must be strongly bound to that actor only [100].

What follows is the evolution of the design and implementation of digital identification systems and the associated technical standards that have been developed to support these systems. For each iteration, the entity being identified as well as the entity performing the identification are considered as well as the types of interactions these identification systems facilitated. The digital identification systems are portrayed in chronological order; however, it is important to recognise that each new iteration did not replace the previous architectures. Rather they exist in parallel with different design choices made to fit the purposes of identification that the digital identification system intended to support (see Figure 3.2). These purposes expanded over time, which in part can be seen as a driver for new approaches to identification systems. The first three approaches siloed, centralised and federated all originated from an organisational need to provision access for its employees to its information systems [183]. These systems changed over time from single mainframe systems in the 60s and 70s, to multiple services relying on a single centralised identification system, to distributed services with

federated identification systems [33, 184]. Those identified were employees and the entities performing identification were information systems acting as agents on behalf of employing organisations. The focus primarily was on security and functionality from the perspective of the information system and its controlling entity [184, 20].

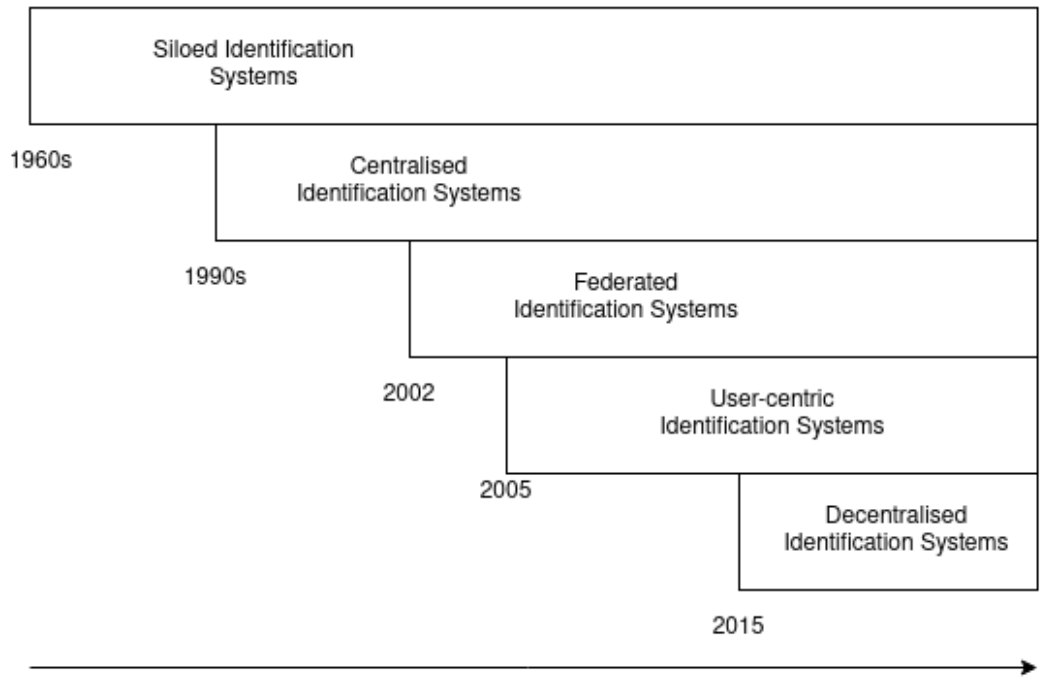

Figure 3.2: Timeline of Evolution of Digital Identification Systems

It wasn't until the mid 90s and into the 2000s that organisations began to interact with private individuals across the medium of ICTs, primarily to engage in commercial transactions. This was supported by the Internet, web browsers and public key infrastructures, in which Certificate Authoritys (CAs) identified organisations hosting websites and attested to their public keys and other attributes in the form of an integrity-assured certificate [185]. The Secure Sockets Layer (SSL), and subsequently Transport Layer Security (TLS) standards enabled individuals to browse websites, gain assurance in the authenticity of those web pages by tracing their certificates to a trusted root CA [186] and establish a secure communications channel using asymmetric key exchange protocols across which sensitive payment details could be shared [26]. Root CAs and intermediate CAs of which there might be hundreds, were and still are entities that must be trusted despite being intermediaries in an interaction that they have no contextual association to [187, 186]. The current model has been demonstrated to lack accountab-

ility and transparency, with instances of insecure or untrustworthy CAs compromising the security of electronic communication systems built on top of these infrastructures [187].

This transition also proliferated the account management strategies that organisations had adopted for employees onto the web and into the personal lives of individuals. With individuals able to, and often forced to, register themselves in the identification systems of the entities they transacted with online to support repeat transactions, personalised services and reduce fraud. However, every time an individual registers with a new information system, they are trusting that the authentication process is secure, that the service provider has implemented the appropriate protections over their authentication credentials and the personal information they store and that the service provider is not maliciously using this personal data [184]. There are countless examples where these practices have been inadequate, either through poorly managed account credentials like the case where Twitter revealed they had been logging user passwords in plaintext [188], or through compromised data servers as happened in the widely publicised Equifax hack [189].

Throughout the early 2000s the web increased in popularity and availability within the wider population, this trend was amplified by the invention of the interactive, collaborative co-constructed social web. It was recognised that the existing identification systems did not sufficiently take the needs of the individual being identified, the user, into account. Individuals were having to manage an increasing number of account credentials, widely acknowledged to be of dubious security susceptible to social engineering [190] or dictionary attacks [191] and commonly reused across services [192]. Furthermore, they had no mechanism to represent themselves as a single continuous entity across distinct information silos or transfer any reputation or information accrued within one silo to another. A new set of user-centric identity standards began to be incubated and developed within the World Wide Web Consortium (W3C) and a community of practitioners formed to spearhead this work in part thanks to the bi-annual Internet Identity Workshop (IIW) that has been running since 2005. Kim Cameron's widely cited Laws of Identity acted as a set of guiding principles for the design of these

user-centric systems [193]. OpenID, branded as the *internet identity layer*, was designed to allow individuals to bring their own identity provider [194]. Essentially letting them delegate the identification system they would like to be identified within across multiple internet services. OAuth further increased the utility for individuals when interacting with online services, giving them the ability to authorise specific services to access resources within other services on their behalf without having to compromise their account credentials [195]. Both standards became 2.0 versions as lessons were learnt and new functionalities were incorporated, eventually OpenID 2.0 was superseded by OpenID Connect which was designed to be more API friendly [196].

Whilst this was happening, the power dynamics within the internet between the individuals using services and the organisations provisioning those services gradually shifted. For example, Facebook grew from servicing 1 million users in 2004 to over a billion by 2012 [197]. By 2005, Google was valued at $52,000,000 with services including GMail and Google Maps. The web became dominated by a few giant organisations who, it has been well documented, pursued an economic model coined by Zuboff as surveillance capitalism to solidify and enhance their position of power in the emerging information society [74]. These organisations were well-positioned to take advantage of these new user-centric standards, quickly becoming the primary identification systems supported by other services on the web [198]. Effectively creating new forms of a centralised identification system. Individuals got an Single Sign On (SSO) solution usable across the web, but the price they paid was compromised privacy. A few large organisations became the intermediaries at the centre of the majority of their online interactions [199].

Another interesting aspect of the maturation of the social web in terms of identification, is that while entities were identified by services to provision access to these virtual spaces; inside these spaces they performed identifications of virtual identities represented to them by the information system [173]. Entities were also capable of constructing their own representation, to a certain degree, of the identity they virtually wished to present. These virtual spaces include social media, the gig economy, online marketplaces and many other environments [16]. While a detailed analysis of this is

outside the scope of this thesis, it illustrates how ICTs have amplified the entanglement between identification and identity. It also emphasises the sheer quantity of identity-related information that is produced, shared, recorded and available online to infer identity from in the information age [74, Chapters 8-9, 9].

The next major shift in identification systems has been driven by a number of mutually reinforcing technologies. The first is Bitcoin, and the subsequent innovation in Distributed Ledger Technologys (DLTs) which enables multiple untrusting entities to reach consensus about some state in a way that is independently verifiable [200, 201]. These provided an opportunity to revisit public key infrastructures without the dependency on centralised certificate authorities, effectively creating decentralised public key infrastructures [202]. In addition to this, DLT also exposed many more people to public key cryptography and have led to numerous innovations in novel, privacy-preserving cryptographic schemes. Next, the expansion of Internet of Things (IoT) in which any device can now become *smart* and networked by connecting to the Internet. The need to design identification systems for these devices has brought about challenges of scale with existing centralised approaches. Not to mention the security issues introduced by these devices due to their inability to effectively identify, authenticate and authorise the subjects they should accept messages from [181]. Finally, the rise of autonomous agents able to select and execute actions independent from any human input increases the urgency in which we need to design functional and secure identification systems that enable autonomous agents to intelligently place trust and demonstrate trustworthiness in this new technological landscape [78].

The identification systems that are emerging to meet the needs of this changed landscape are commonly referred to as Self-Sovereign Identity (SSI) or decentralised identity. These systems also have their own set of guiding principles initially proposed by Christopher Allen [203], but since iterated upon and challenged by numerous actors working on, with and adjacent to digital identification systems [204, 47, 205, 206, 207, 208]. While these discussions are important, they are largely outside the scope of this thesis. From a technical perspective, decentralised identity-based solutions aim to enable both entities interacting to identify and authenticate the other party in the

interaction using information systems and secrets under their control. Effectively provisioning both entities with their own digital identification system, giving them the ability to interact as peers, mutually able to present and verify integrity-assured digital ID artefacts in the form of Verifiable Credentials (VCs) [180]. This is reviewed in more detail in the following section.

3.3 Decentralised Identification Systems

The decentralised identity space contains numerous actors, implementers and application developers, each working towards their own version of a digital identification system and the software development stack to support it. These entities then collaborate on open standards in forums such as the W3C and the Decentralized Identity Foundation (DIF) as they attempt to achieve technical interoperability amongst their different implementations. This is still an emerging, and rapidly evolving space with a lot of work required to standardise the different layers of these technological systems. However, two key standards: Verifiable Presentation (VP) Data Model became version 1.0 recommendation from the W3C in 2019 and Decentralised Identifiers (DIDs) version 1.0 are currently a proposed recommendation awaiting final confirmation [35, 34].

3.3.1 Roles, Interactions and Infrastructure

Identification systems based on decentralised identity standards are credential-focused systems [209]. Entities in the role of the holder, receive an integrity-assured set of claims about a subject from an entity in the role of the issuer. The subject of the claims is often the holder, but does not have to be, such as in the important case of guardianship [210]. These credentials, or ID artefacts, are then stored and managed by the holder, who can choose when, and to whom, to disclose any subset of the claims within their credentials to when engaging in a process of identification with an entity in the role of a verifier. The verifier on receipt of a presentation is able to verify its integrity and authenticity using public key cryptography without needing to interact with the issuer of the claims. That is, they can check the claims have not been tampered with since

issuance and that they were signed by a specific public key pair. Additionally, they often include a mechanism to verify the claims presented were initially issued to the entity presenting them. They then must make a judgement, based on the context in which this identification occurs, whether the public key that signed the claims, and the entity this key is understood to represent, has the authority to make the claims presented. As well as determining, based on the information within the claims, what authorisations to grant the holder. Both credential issuance and credential presentation are a negotiation between two entities in which they exchange messages to determine which claims to issue or present in their respective situations.

A public key infrastructure accessible by all entities across both interactions is required to support these digital identification systems. An issuer must anchor their issuing public key to this infrastructure. Holders may also have to anchor a public key if this is the mechanism used to identify themselves as the authentic holder of a credential, they also require the ability to retrieve the issuers key in order to verify they have been issued a valid credential. Verifiers might not need to anchor any cryptographic material to the infrastructure, but will require access to it to retrieve both issuer and holder public keys from the claims presented in order to verify them. This infrastructure needs to be verifiable, integrity-assured, highly available and often, ideally, censorship resistant. These requirements made distributed ledgers an attractive solution, although other options exist [202]. The VC Data Model specification refers to this infrastructure as a Verifiable Data Registry (VDR) [34]. Furthermore, rather than storing plain public keys, DIDs were developed to add a layer of indirection between the identifier and the cryptographic material which brought some attractive benefits to these systems. These are discussed in more detail in the following section, for now it is only important to know that within a VP both issuers and potentially holders are identified by DIDs which can be resolved against a VDR, to retrieve the necessary public key material to verify claims. Finally, other data objects, in particular credential schema that enable the semantics of claims to be parsed need to be available to all the relevant entities. These objects may also be stored within a VDR, although not necessarily the same one.

The roles, interactions and infrastructure discussed in this section are shown in

Figure 3.3 which has been adapted from the W3C VC Data Model specification and is commonly used throughout the decentralised identity space [34].

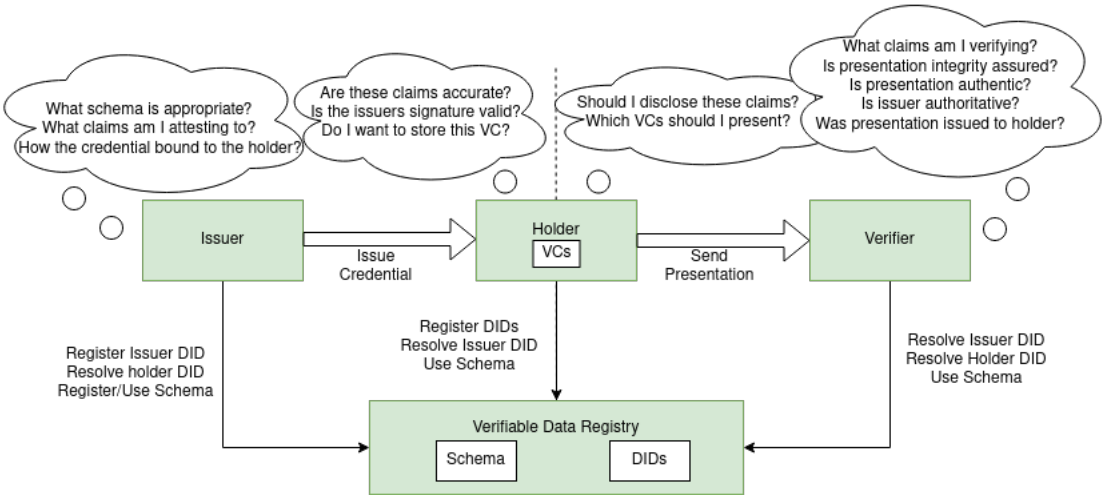

Figure 3.3: Roles, Interactions and Infrastructure for Decentralised Identity (adapted from [34])

3.3.2 Open Standards

Open standards are an important part of digital identification systems; they enable technical systems implemented by different actors to engage in mutually understood protocols and exchange messages in a consistent format. This is broadly termed interoperability and is designed to encourage competition, support permissionless innovation and prevent vendor lock-in allowing those who use these technologies to have the choice of solution providers and the flexibility to change as new implementations become available [176]. Achieving interoperability between actors and their chosen technologies, cryptographic protocols and data formats is challenging, especially within a distributed system such as the ones decentralised identity standards aim to support. Different actors advocate for different elements to be standardised and attempt to influence the language to be included in specifications being drafted based on their perspective of how these technologies should be developed. Ultimately standardisation involves compromise. However, the fact that you can open a browser of your choice and render virtually all Web pages accessible on the internet in a consistent manner emphasises how valuable open standards can be [176].

DIDs specify a set of common interfaces an implementer must define when creating a standards compatible DID method that supports Create, Read, Update and Deactivate (CRUD) functionality of DIDs anchored to a specific VDR. DIDs are represented in the format *did:method_name:identifier*, with the method name identifying a specification under which a DID method should be implemented in order to perform CRUD operations on the DID Document the identifier identifies. A DID Document contains public key material, the authentication mechanisms that this key material can be used with and service endpoints to communicate with the entity controlling the DID [211, 35]. Numerous DID methods have been registered anchoring DIDs and DID Documents to different VDRs including Bitcoin, Ethereum, Veres One, Interplanetary File System (IPFS) and Sovrin. There are also DIDs such as did:key and did:peer that are not anchored to VDRs at all, but instead used to manage ephemeral or private relationships. The full list of DID methods can be found in the W3C DID method registry [212]. It is up to implementers to determine which DID methods they wish to support.

The VC Data Model defines the structure that a credential issued by an issuer should follow. As well as a data structure for the composition of VCs into VPs, which can contain claims from multiple credentials. A VC is made up of credential metadata, such as an identifier for the issuer and subject, a set of tamper evident claims attesting to information about the subject and a set of proofs that can be used to verify the claims [34]. The VC Data Model defines a standard set of field names for expressing this information, enabling all entities to parse a VC for its information. However, the data model does not specify a format in which a VC should be represented currently different implementations use plain JavasScript Object Notation (JSON), JSON encoded as a JSON Web Token (JWT) and JSON Linked Data. Each format needs to be understood by both implementations, if a VC is to be able to meaningfully interoperable between them [213].

Another challenge to achieving interoperability across different VC implementations arises from the cryptographic suites they support. While the VC Data Model facilitates parsing of the information with a VC (assuming the data format is understood), unless the proofs contained within a presentation can be verified by the im-

plementation then the integrity and authenticity of the claims cannot be judged [213]. This is especially challenging when considering implementations that leverage Hyperledger Indy/Ursa/Aries open-source projects, which make use of more complex privacy-preserving cryptography to provide unlinkability across multiple presentations of the same VC [28]. This cryptography requires specific interactive flows when issuing and presenting credentials, that other implementations would have to understand and support. While progress has been made, most notably in the standardisation efforts around the BBS+ signature scheme and proof format [214, 30, 215], more work is needed before practical interoperability is achievable.

The other standards under development are supported unevenly throughout the players in the decentralised identity space. These include: DIDComm for establishing a private and secure message exchange channel using the information within DID Documents which originated from the Hyperledger community and has since moved to DIF for broader adoption [216]. The Credential Handler Application Programming Interface (CHAPI) fulfils a similar function and is being pursued primarily by Digital Bazaar [217]. A Verifiable Presentation Exchange Data Format has been ratified at DIF [218].

3.3.3 Digital Agents and Wallets

Digital agents and wallets are terms used to describe the actual software artefacts that entities acting in the roles of issuers, holders and verifiers use to engage in protocols with other entities. While the distinction between agents and wallets is not consistently made across implementations, this terminology will be used throughout this thesis with definitions following those in O'Donnell's report *The Current and Future State of Digital Wallets* [133]. A digital wallet is the place that information is stored, for example VCs or DIDs that an entity has previously encountered and chosen to remember. Importantly digital wallets handle cryptographic key management, and should support secure backup and recovery which as the cryptocurrency space regularly illustrates is a notoriously challenging problem [219]. While an agent is the software component

that sends and receives messages, understands specific protocols and data formats, retrieves and stores information within a wallet, interfaces with a VDR and performs cryptographic operations. They have been compared to the browsers and servers of the decentralised identity ecosystem, handling VCs and their associated protocols instead of exchanging and rendering Web pages [180].

While the concept of digital wallets has been reincarnated within decentralised identity communities, the idea that entities could own and use digital technology to manage key material and engage in cryptographic protocols is a regular theme throughout the literature. David Chaum called for these types of technologies in 1985 when conceptualising cryptographic credential mechanisms [27], and Jøsang and Pope in a 2005 paper suggest Personal Authentication Devices (PADs), owned and controlled by users, could simplify a user's ability to manage identifiers and authentication credentials [20]. A key difference in 2021 is the proliferation of smart phones, often with built-in strong authentication mechanisms such as biometrics, within the wider population [179]. Although, it it is important to point out that not everyone has access to these devices, identification systems must factor the availability of certain technology within the population being asked to identify using it when designing these systems.

3.3.4 Machine-Readable Governance

Machine-readable governance is the final aspect of decentralised identification systems that this thesis is going to focus on. The standards developed and the systems they are intended to support intentionally avoided specifying a trusted central authority that would act as the arbiter of truth within these identification systems. Instead, it is the responsibility of each actor to make decisions about which entities, credential types and claims to trust when performing specific roles under any given interactive context. Different credential ecosystems will contain unique actors that will either issue, hold, require or present specific credential types and claims therein. The levels of assurance associated with specific credential types and the potential sensitivity of the claims data will impact these trust decisions. Governance is the process by which

the actors, rules, regulations and accountabilities surrounding these ecosystems are defined and need only be adapted to include new technological artefacts. However, in many cases, these governance frameworks within highly structured social systems such as healthcare are already defined. This section focuses primarily on the ways in which aspects of these governance policies can be made accessible to technological artefacts in order to support decision making processes of human users or in some cases facilitate automated decision making [220].

Questions of authority that need to be answered, and the mechanisms by which they are answered will depend on the context in which the identification system has been implemented for and the perspective of the actors asking the questions. Some generic examples of the types of questions the different roles might need to answer adapted from a webinar on *trust registries* include [221]:

- As a holder, what rules are an issuer offering to issue a certain credential operating under and based on these rules is this credential valuable to me and should they be trusted with the potentially sensitive personal information required to issue this credential.
- As a verifier, based on the presentation of a set of claims issued from a particular DID against a certain credential is this issuer authorised to have made these claims in the first place.
- As a prospective issuer, which governance framework should I be operating under and how do I join and become recognised as an issuer under this framework.
- As a verifier, which set of governance frameworks provide adequate governance over issuers attesting to the credentials and claims required for my purposes of identification.
- As a holder, based on a proof request to disclose some set of sensitive claims to a verifier how can I determine if this entity can be trusted to handle this information. Under what rules are they operating and how can they be held accountable.

- As a governance authority how do I convey the set of policies, authoritative issuers, legitimate schema and proof requests that compose of the identification system being governed.

While this is a new and emerging areas within decentralised identity, the Trust over IP Foundation [222] propose a solution that has similarities to CA except in a decentralised manner. Governance authorities, entities with public DIDs anchored to a DLT, act as roots of trust for ecosystems. In this role, they publish a human readable governance framework which defines the legal and business policies, actors, roles, credential types and identification processes [180, Chapter 11]. Additionally, they manage *trust registries*, verifiable lists of machine-readable data that digital agents can access and incorporate into their business processes and rules engines. These lists might include: the set of issuing DIDs, the set of credential types, which DIDs are authorised to issue which credential types, the set of verifier DIDs, a set of proof requests that a verifier should legitimately be asking for. Exactly how these lists are accessed by different entities is still being defined [221]. However, the idea is that entities can subscribe to multiple governance authorities receiving updates from the registries. They can also register and perform roles defined by multiple governance authorities. This allows for a more flexible, customisable approach to governance that can be designed to meet the needs of the specific ecosystem being governed. While each actor is free to determine which authorities and ecosystems they trust depending on their business requirements [220].

There is also recognition for the need for governance at all layers of the technology stack used when implementing an identification system. This includes the distributed ledger (or other Verifiable Data Storage used), the types of digital wallets and agents used and the authoritative actors within the system. As emphasised in the dual Trust over IP stack commonly referred to by a number of actors within this space (see Figure 3.4). It is important to recognise that governance is a fundamentally human activity with the rules, roles, ID artefacts and identification processes influenced by the context in which an identification system is realised.

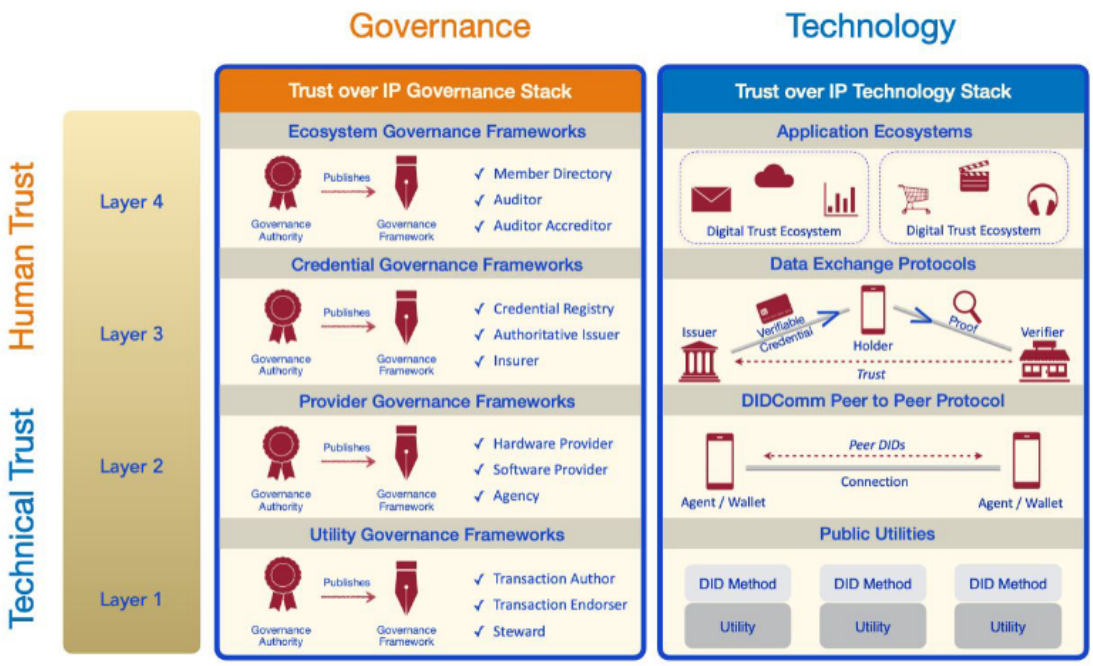

Figure 3.4: Trust over IP Stack (Taken from [222])

3.4 Lessons from Identification Systems for Development

Identification systems have existed long before the development of digital identification technologies. Especially to manage the provision of legal identity, *the right to recognition everywhere as a person before the law* [124], that became intertwined with nationality, at least in the western world, throughout the last two centuries [14]. States provisioned passports and maintained birth, death and voting registries of its citizens. However, the nation-state is historically a relatively new concept with many states only emerging after the two world wars and the following break up of colonial empires. Indeed, some states were created through arbitrarily drawn lines on a map [223]. Not all of these states are stable, nor do they acknowledge the existence of different citizen identities evenly, producing what is referred to as the *legal identity gap* with an estimated 1 billion individuals without the *official* means to prove who they are in the eyes of the law [14]. This can lead to basic rights, opportunities and access to services being withheld [22].

The sustainable development goals recognise the provision of proof of legal identity for all as critical and set an ambitious target to achieve this by 2030 [14]. With many

national, international and Non-Governmental Organisations (NGOs) turning to digital ICTs, which Gelb and Metz describe as driving an identification revolution, as a means realise this goal through the provision of digitally augmented identification systems [22]. This section reviews the rich literature available that documents the development of these systems, the challenges they have faced and the impact they have had on the lives of individuals being identified [23, 40]. The aim is to identify lessons that can be learnt from these real-world experiences about augmenting existing identification processes and introducing new ones.

It is important to emphasise that identification, especially in the context of development, is intended and widely recognised as a public good [22]. Systems of identification can create opportunities for civic participation, access to services, increased accountability and ensure people are visible as holders of rights [14, 23]. However, formal processes of identification should be recognised as an asymmetric power relationship in which the individual being identified must conform to the requirements for identification as defined by the entity performing the identification [224]. There are a myriad of ways in which this relationship can be intentionally abused or unintentionally introduce inherent risks for the individuals being identified and represented within these systems. Key themes identified in the literature include: Privacy and Surveillance, Exclusion, Accuracy, Complexity, Power and Governance [23]. Each is reviewed briefly in turn.

3.4.1 Privacy and Surveillance

Many digital identification systems being rolled out around world, collect and store a diverse range of data attributes about individuals to support their identification [23, 22]. These are often stored centrally, with the individual given means to authenticate themselves against this administrative representation. Increasingly, we are seeing these systems turn to biometrics as a mechanism to uniquely identify and authenticate individuals [172, 225]. For example, India's Aadhaar ID system collects a facial photograph, all ten fingerprints and a scan of both irises, highly sensitive personal information integral to the identified individual's body [226]. Autonomy and control over one's

biometric information is widely perceived as crucial for realising privacy now and into the future. Biometric information, when stored centrally, effectively empowers anyone with access to the dataset with the ability to perform non-consensual identifications of the individuals identified within the data. When considering that there is a non-zero risk of system compromise, a centralised architecture storing biometric data presents a substantial risk to individual privacy [227].

Digital identification systems have created rich data sets about identified individuals, both from the information collected during registration and that extractable as these individuals interface with digital systems in order to access services. The use of a single unique identifier across multiple systems enables the linkability of datasets across distinct contexts in ways not previously possible [228]. While foundational identification systems tend to be state led infrastructure projects, they are often developed in collaboration with private actors, who, in the case of the Indian Aadhaar system at least, are also building commercial services that integrate with these identification tools [229, 228]. Sriram uses the analogy of street hawkers selling wares on public roads, except with this new digital public infrastructure the street hawkers are now powerful commercial entities, often with intimate knowledge of the infrastructure, able to extract exorbitant rents by building services and applications on top of this infrastructure and the data it makes available [228]. Identification systems also present an enticing opportunity for state surveillance and citizen profiling in the name of national security [14]. How these systems are designed, the identifiers they assign, the personal information they collect and store, and the data produced when individuals interface with these systems all impact the perceived and realised privacy within these systems.

The design of Aadhaar has received much criticism for effectively creating a surveillance architecture spanning the entire population of India [227, 226]. However, challenges to this system resulted in groundbreaking ruling from the Indian supreme court which affirmed that privacy was an inalienable, fundamental right foundational to individual self-determination, autonomy and freedom enshrined in the Indian Constitution [230, 231]. Furthermore, the court ruling locates privacy concretely in the individual's interactions with the state and their struggle against coercive state power

[231]. While it remains to be seen how this ruling will be applied to specific challenges to the Aadhaar system, it highlights both the intrinsic value of privacy and the important role legal institutions have to play in protecting individual rights in this technologically augmented future.

3.4.2 Exclusion of Individuals and Groups

Formal identification systems mediate people's socio-economic interactions. They determine and control access to resources, locations, services, rights and entitlements [224]. These can include access to connectivity, civic participation, and financial services as well as basic human needs such as food and shelter [22, 232, 23]. There are many ways that identification systems designed for inclusion can just as easily end up excluding people. Analogue systems of identification were able to navigate the inconsistencies and edge cases that are unavoidable in the diversity and complexity within social systems using personal relationships and custom work arounds judged on a case-by-case basis [40]. Whereas digital systems tend to sharpen the edges of identification processes, as individuals are classified against machine-readable criteria and binary authentication mechanisms which are then applied to the population at scale [23]. A consequence of this is that some individuals fall through the cracks, with human agents performing the identifications often beholden to black box information system and unable to apply their own judgement [233].

It is important to remember that those most likely to fall through these cracks are disproportionately those individuals already structurally disadvantaged in society - those from rural areas, with low income, women and other minorities [14]. It is also these individuals who are often impacted the most by their inability to be seen by the state and hence access vital services [23]. Exclusion can be intentional, such as in the case of the Myanmar government attempting to force the Rohingya to accept ID cards that identified them as subclass citizens [224]. However, it can also be unintentionally introduced as new identification systems disrupt existing processes of identification, with some individuals unable to bear the burden of navigating these new systems [232].

These problems are exacerbated if inaccuracies are introduced into the identification systems during the process of transition.

3.4.3 Accuracy, Misrepresentation and Falsification

An individual's sense of identity is relational, dynamic and adaptive. It is in constant flux as individuals reinvent themselves and the way they wish to be presented over time [4]. In contrast identification is a process of representing a complex individual in a set of static, often unchanging, attributes [23]. The data types collected, and in some cases the acceptable data entry options are preconfigured by those performing the identification. This limits an individuals ability to participate in their self-construction [224]. Additionally, mistakes in the representation of an individual introduced during registration can impact their access to services. Biometrics are not free of these mistakes, with Indian officials admitting failure rates in biometric authentication occurred as often as $10-15\%$ of the time [232]. Fixing these mistakes takes up time and resources of the inaccurately represented individual [233]. Furthermore, a digital identification system as a source of truth about identified individuals does not prevent falsified representations entered into the system by those in positions of power [234]. In fact even criminal enterprises have been discovered in India selling ghost enrollment kits to enable entities to falsify Aadhaar entries [227].

3.4.4 Complexity of Identification Systems

Identification systems are complex by their very nature. They involve multiple stakeholders, each with different priorities, roles, processes and responsibilities [21]. Within the context of identification for development Sperfeldt identifies five distinct stakeholders archetypes each with different perspectives: 1) rights-based focused on emphasising the *multi-faceted nature of discrimination* as the root cause of a lack of legal identity, 2) governance actors that view provisioning legal identity as a technological challenge whose implementation can improve planning through better access to data, 3) those that view legal identity as foundational to achieving other sustainable development

goals, 4) actors that view these systems through their ability to enhance security and improve border protection and finally 5) commercial actors who see the market opportunities that provisioning legal identity can enable [14]. These actors all have different perspectives, access to resources and ability to exert influence on the design and implementation of the systems. Often the result is the actors with either financial resources, technical knowledge or holding positions of power within key institutions are able to exert significant influence over the project's direction. While those advocating for individual rights on behalf of the most marginalised are excluded from a seat at the table and are forced to challenge these systems from the outside [228, 230].

A key finding from the identities project that conducted qualitative research through interviews with over 150 identified individuals under the Aadhaar identification system was that no system is *greenfield*, rather they are layered on top of existing relationships and processes for managing identity and identification within a given context [40]. The introduction of digital identification systems can add additional complexity to the access of services, solidifying identification processes that were once flexible and adjusted to a locality. Additionally, complex technologies, such as biometrics, tend to be more fragile and come with increased risks [227]. This can turn these processes into a black box, with the reasons behind identification and authentication failures or delays opaque even to the agents performing the identification. As Chaudhuri writes, this can make the state appear *distant, opaque and seamful* [233]. Furthermore, where individuals must register with multiple organisations each with their own identification systems, as is often the case for refugees, this introduces additional complexity around understanding the purposes and governance surrounding each of these systems [235]. An analysis of refugee experiences within Lebanon, Jordan and Uganda indicates that individuals do not always understand why they are being registered or even remember all the agencies they have registered with [235]. Both the complexity of enrollment processes and lack of awareness of the benefits of identification systems is echoed in research into birth registration processes in India [234].

3.4.5 Power and Abuse

Identification systems convey power to those designing, implementing and governing the system, who are able to determine rules for access based on specific identification criteria [224]. Identification makes individuals visible to the state, which can be beneficial, allowing them to access services and entitlements previously inaccessible to them. However, being identified also makes an individual vulnerable with these systems often used to facilitate intrusive policing methods, discrimination and behavioural manipulation [14]. Displaced and marginalised populations are particularly vulnerable to these abuses of power [224]. The power conveyed by these systems must be appropriately constrained, with accountable institutions and robust legal frameworks that protect the individuals being identified [22].

3.4.6 Governance of the Identification System

Martin and Taylor state, *rights cannot be granted, they can only be claimed and accessed (or denied)* which places identification in the realm of politics, law and governance [224]. However, this means that existing injustices within political systems can seep into the design and implementation of identification technologies, which perpetuate and amplify their impact [23]. Furthermore, political systems are often fragmented, with institutions and laws at the regional and state level often conflicting with each other. The lack of political will to coordinate amongst these entities can result in sensitive information stored and managed in duplicated systems [225]. This creates an unnecessary risk that is compounded when you consider that these political systems of governance are not guaranteed to be stable. The situation in Afghanistan where identification systems using sensitive biometric information designed and managed by the US to provide aid and reduce fraud have now been re-purposed by the Taliban to identify dissidents provides a stark reminder of the need to factor in political instability into the design of these systems [235]. Individual protection against an uncertain and unpredictable future is paramount.

Three principles of identification for sustainable development relate to governance.

These are the safeguarding of data privacy, security and rights through comprehensive legal and regulatory frameworks, clear institutional mandates and accountabilities and independent oversight with defined pathways for adjudication of grievances [22]. Robust governance structures and processes can support trust and establish the legitimacy of systems of identification. The case of the legal challenge against the perceived privacy violations of the Aadhaar identification system that eventually resulted in the supreme court upheld the right to privacy as foundational to the Indian Constitution is an example of how existing legal structures can provide these pathways [230, 231].

3.5 Critical Discussion

While it is true that digital ICTs have increased the threat vectors surrounding identification systems, they have also led to the development of a novel suite of tools, standards and protocols that have transformed the possibilities for designing secure, robust and privacy-preserving digital identification systems. These have emerged from the functional requirement for ICTs to identify its users and virtually represent them within digital environments. Furthermore, these environments needed mechanisms to facilitate intelligent trust decisions placed in the representations of virtual entities they encounter. A number of approaches to meeting these requirements have been developed and iterated on, initially by enterprise actors who developed centralised, siloed solutions. Then as ICTs grew with complexity, and the internet became ubiquitous a community of *digital identity* practitioners formed. Initially they worked on user-centric identification systems and standards designed to simplify a user's interactions with internet services, then as Bitcoin and distributed ledgers grew in popularity focus switched to decentralised identification solutions. The premise of using censorship resistant verifiable data storage system that anyone could use to register self-authenticating identifiers, standardised at the W3C into DIDs is a powerful one. These identifiers were specifically designed to support identification systems in which any entity would be able to issue cryptographically signed, integrity assured claims about another entity in the form of a Verifiable Credential. Credentials would then be stored by the entity

to whom they were issued, who could then present the claims within the credential during future interactions. The core idea being to replicate paper-based identification systems within digital environments, removing the requirement for parties to store and maintain sensitive personal information that is increasingly regarded as a toxic asset [236].

3.6 Conclusion

This study of the literature surrounding identification systems has shown them to be powerful mechanisms to structure and constrain the field of action of the entity being identified, who must meet the rules defined by the system as enforced by the entity performing the identification. To be identified, is to be seen and recognised by some system for some purpose. For example within sustainable development, identification systems are a means to recognise individuals as holders of rights and provide access to key services and resources [23]. A well-intentioned justification for identification. However, to be seen by structures of power is also to be vulnerable, creating opportunities for discrimination and abuse [232]. This is especially true for those marginalised by society. Equally, to be unidentifiable under an identification system is to be excluded from the access and entitlements provided to those who can be identified.

The following chapter reviews the development of a largely equivalent concept within a different thought collective, cryptography. The idea of a credential mechanism designed to achieve security without unnecessary identification, which originates from a publication in 1985 and has evolved independently of the digital identity community [27]. This thesis aims to demonstrate that the ideas within these two thought collectives are not mutually exclusive but rather can be powerfully reinforcing.

Fourth Chapter

Security Without Identification

> *"The architecture chosen for these systems may have a long-term impact on the centralisation of our economic system, on some of our basic liberties, and even on our democracy."*
>
> Security without Identification, David Chaum [27]

In 1985, David Chaum outlined the concept of achieving security without identification and justified its importance for the future information society that he anticipated was being created [149, 27]. He accurately identified the worrying trend emerging as public and private sector organisations opted to collect and curate ever more pervasive, interlinked record keeping information systems designed to identify a unique human actor and correlate their actions across multiple contexts. These were intended as a security system to provide assurance about the entity an organisation was interacting with and reduce the vulnerability of these organisations to the abuses of bad actors. Security systems act as an important social pressure that constrain and (dis)incentivise certain behaviour, this is especially important in an advanced information society with its expanded attack surface and the increasingly valuable interactions that ICTs support [15]. However, living in the future that Chaum was anticipating, we experience the realities of these identification systems and the asymmetric power relationships that they have produced as our selves are rendered at increasing detail, beyond our influence and outside our awareness [74].

This chapter revisits the ideas of Chaum and presents the evolution and systematisation of cryptographic knowledge that these conceptual ideas, amongst others,

stimulated. Demonstrating that security without unnecessary identification within digitally mediated interactions is technically feasible, practically implementable and has a firm, scientific basis. Modern cryptography is the modular application of mathematical operations to information in order to define protocols that guarantee certain properties about the information and its flow between interacting entities. While the use of codes to secure information flows has been pursued throughout human history [11], it is only since the 1970s that this discipline, when combined with the capabilities of digital ICTs and the associated knowledge of information theory [18] and computational complexity theory [17], began to realise its modern scientific expression [26].

The structure of knowledge that has emerged to realise early conceptual ideas such as Chaum's is presented in this chapter as a hierarchical series of layers, with each layer constraining and enabling the possibilities of the layer above and directing focused perception on the layer below [32]. This model was developed after an in-depth analysis of the cryptographic literature and the relationships between key contributions, the workings that this chapter is based on can be seen in Figure 4.1. The layers identified in this chapter are shown in Figure 4.2 and a hypothesis that it was the conceptual ideas that directed the discovery and production of cryptographic knowledge in order to specify cryptographic protocols from the mathematical foundations of group theory is presented. Furthermore, it is important to recognise that each layer can be defined abstractly with idealised properties or instantiated concretely, with concrete instantiations attempting to realise the idealised properties of the abstractions (See Figure 4.3).

Group theory provides an abstract way to describe a mathematical universe, a mathematical setting selects one of these universes, a group, and demonstrates the existence of a computationally hard problem under a well-defined complexity assumption that can be used to construct one-way functions. A cryptographic primitive defines an equation in a mathematical setting that realises a set of idealised properties that characterise the primitive. Then the security of this primitive is proven through demonstrating a reduction to a computationally hard problem. A cryptographic protocol defines an algorithm, or collection of algorithms, composed of cryptographic primitives that realise a set of idealised properties derived from the conceptual framing of the solution space.

Cryptographic protocols include a model of actors, adversaries and their capabilities. Therefore there is need to define a model of security within this more realistic setting. It is important to note that it is possible to realise the same cryptographic primitives and protocols within multiple concrete mathematical settings. Changing the mathematical setting can have implications for the security and efficiency of these instantiations whilst functionally the protocols remain the same.

Throughout the early period of modern cryptography, from 1976-2001 in this chapter, these theoretical foundations were being discovered and understood from both an abstract and a concrete perspective. Cryptography existed primarily in the conceptual, virtual realm of science and mathematics. The discipline required cryptographic engineers able to synthesise their blueprints into technological artefacts, so that these mathematical constructions could be applied to practical applications in the real-world. This dependency between theory and practice, between the mathematician and the engineer, has been a recurring theme throughout the history of ICTs since at least the telephone [11]. It was not until the 2000s that the synthesis of cryptographic protocols began to be openly pursued. In part because the export of encryption software was prohibited in the US under the Munitions Act until 1996 [237]. This was only repealed thanks to a sustained effort from a group of individuals who believed in the value of privacy and saw cryptography an effective way to protect privacy in an information society. They identified themselves as cypherpunks and released, the now famous, Cypherpunks Manifesto [148]. Another bottleneck was the fact that limited concrete blueprints for cryptographic protocols demonstrated to be theoretically secure and practical had been produced. Only concrete instantiations can be practically implemented within software.

It was not until 2001 that Camenisch and Lysyanskaya (CL01) published a theoretically efficient credential system based on well-defined primitives with provable security under studied and established complexity assumptions [28]. The second part of this chapter focuses on how this protocol, and the layers upon which it relies, were gradually synthesised into software artefacts. First cryptographic libraries performing the protocol operations were developed [238, 239], then research was undertaken to under-

stand how these cryptographic engines could be integrated into a software engineering framework [240]. This took time, but the interplay between engineers and theoretical cryptographers has driven innovation and demonstrated the feasibility of these conceptual creations of scientific thought. Alongside this, cryptographic research into credential mechanisms continued apace. Including the transition to a more efficient mathematical setting, bilinear pairings over elliptic curves [241], the focus on universally composable security [242] and the definition and realisation of new desirable properties for credential mechanisms [243].

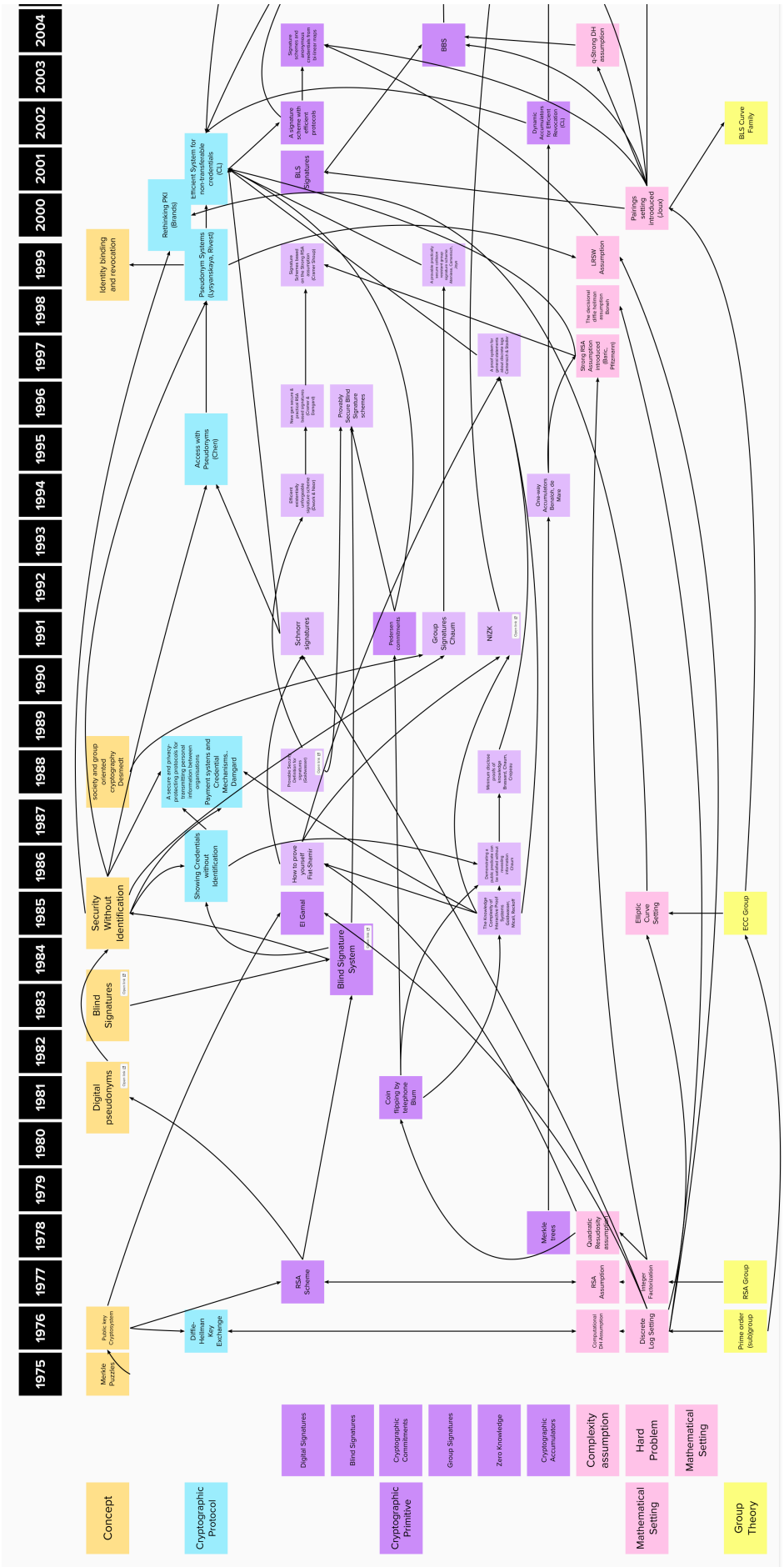

Figure 4.1: A timeline of the emergence of and relationships between cryptographic knowledge

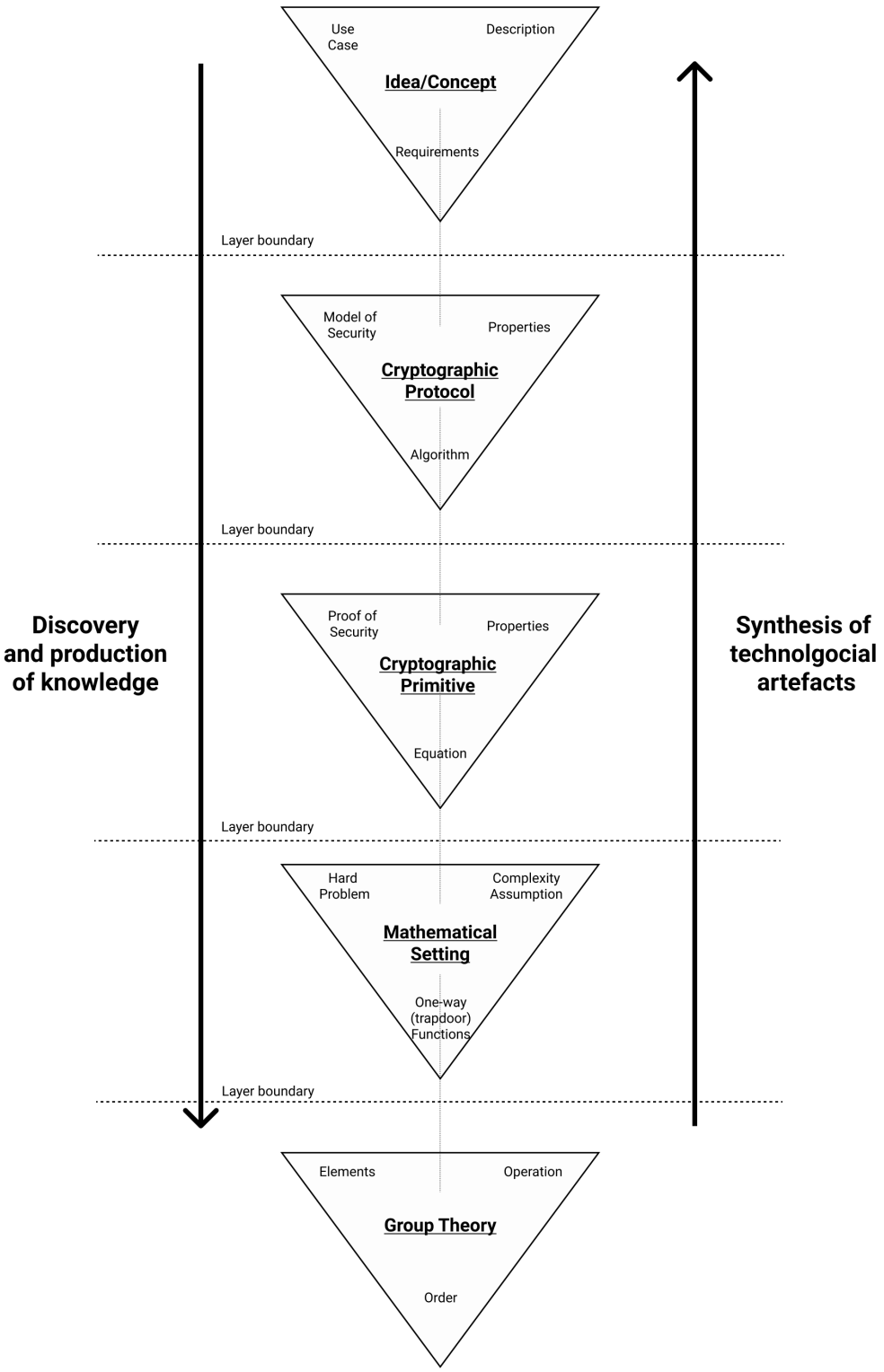

Figure 4.2: Structure and hierarchy of knowledge in cryptography

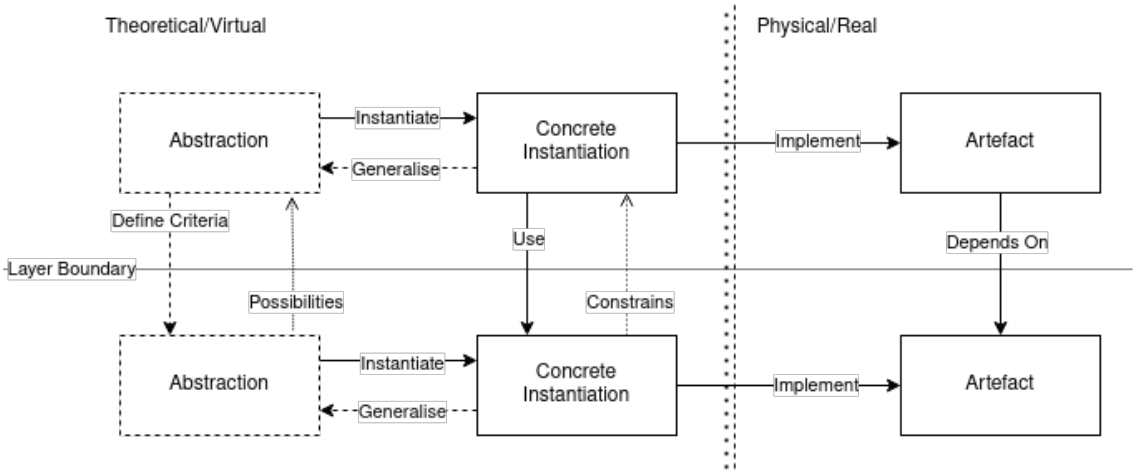

Figure 4.3: The relation between abstractions, concrete instantiations and practical implementations within cryptography

4.1 Laying the Foundations

This section reflects on the evolution and systematisation of cryptographic knowledge into the four layers identified in Figure 4.2 through the lens of Chaum's credential mechanism [27]. From Diffie and Hellman's seminal paper introducing public-key cryptosystems [26] up until Camenisch and Lysyanskaya defined an efficient system for non-transferable credentials in 2001 [28]. It is during this period that the foundations for modern cryptography, both in terms of the ideas and the techniques for realising them, were established. Furthermore, computers went from relative obscurity to ubiquity, the Internet was born, and the shape of our information society began to emerge. A credential mechanism is just one of the ideas to be conceptualised during this time.

4.1.1 Early Ideas and Concepts

Modern cryptographic research has been driven by visionaries, researchers who are able to synthesise the current state of the art, outline the future possible, justify its importance and indicate potential pathways for realising this future. Diffie and Hellman did this for public key cryptosystems [26]. Building on the work of Merkle, they provided an abstract mathematical definition in terms of a pair of families of algorithms representing invertible transformations [25]. By indexing these families with a key, k,

a pair of algorithms are selected (E, D) such that E is the inverse of D and can only be feasibly computed if k is known. Then they went on to show how any public key cryptosystem could be used to produce a one-way authentication system and justified the importance of such a system in computer-controlled communications networks. Highlighting entity authentication and digital document signing as key use cases [26].

This publication is widely acknowledged as the beginning of cryptography as a modern scientific discipline, indeed Diffie and Hellman state this themselves [26]. They theorised that cryptographic protocols could be demonstrated to be provably secure, by building on the established mathematical foundations of information theory and computational complexity theory [244, 17]. The goal of the cryptographer was to find trapdoor one-way functions whose inversion is assumed to be an NP-Complete problem, such that the computational complexity of the inversion grows exponentially in the size key space unless the secret trapdoor information is known.

An important contribution, and a pattern that is repeated throughout cryptographic literature, was to conceptualise the solution space for public key cryptosystems despite not having a concrete solution. This directed future research into concrete mathematical settings under which a public key cryptosystem could be realised and provided a criteria against which they could be identified. Indeed, two years after this Diffie-Hellman contribution, the Rivest, Shamir and Adelman (RSA) cryptosystem, based on the hardness of integer factorization of the product of two primes was proposed [245]. In addition to directing research into concrete mathematical instantiations, the concept of public key cryptosystem became an idealised primitive that unlocked new possibilities when envisioning and defining systems for other use cases.

The initial idea that public key cryptography could be used to design a system that facilitated the secure, verifiable exchange of third-party attestations (credentials) without leaking any additional identifiable information can be traced back to David Chaum's seminal paper *Security without Identification* published in 1985 [27]. This paper addresses the perceived threat to privacy and the centralisation of power that Chaum believed an information society that achieves security through the over identification, correlation and representation of individuals within information systems introduced.

The conceptual solution Chaum proposed involved using unlinkable pseudonyms for each relationship that enabled individuals and organisations alike to authenticate against these pseudonyms using a digital signature from an associated keypair. This developed the ideas proposed in earlier work by Chaum on untraceable electronic mail [149]. Across these pseudonymous relationships, individuals would be able to share signed attestations, issued by another organisations, linking only the information they chose to disclose to the specific pseudonym for that relationship. Chaum emphasised the importance that disclosure of the same attestation across multiple pseudonyms did not provide a mechanism for these pseudonyms to be linked [27, 246].

Chaum's paper describes a cryptographic system for realising a credential mechanism at a high level. From this paper, it is possible to extract many properties that Chaum proposed as desirable within such a system [27]:

- **Unlinkability**: Knowledge of one pseudonym should give no knowledge of other pseudonyms an individual uses in other relationships
- **Forgery resistant**: Protection from abuse from malicious entities
- **Minimal disclosure**: The ability to display only parts of a credential
- **Uniqueness of pseudonyms**: An organisation should be able to limit an individual to a single pseudonym within their domain
- **Multi-show**: Ability to transform credential issued under one pseudonym to the same credential but linked to another pseudonym while retaining unlinkability
- **Interaction Parity**: Both organisations and individuals protect their own secrets, manage credentials and create backups of their information
- **Asymmetric authentication**: Individuals can authenticate ownership (control) of their pseudonyms
- **Non-repudiation**: Provide a mechanism to prevent party falsely denying a message that they sent

Much like the publication *New Directions in Cryptography* [26], *Security without Identification* articulated the desirable properties of a cryptographic system without defining precisely how the system could be realised [27]. Chaum provides clear justification for the need for such a system in an information society, advocating for a system that protected privacy and promoted decentralisation by providing individuals with the digital equivalent to paper and card-based *tokens* that could be used in digital communications, payments and credentials. The paper also indicated mathematically how the system might be constructed, using cryptographic primitives from his earlier work on blind signatures and untraceable payments [150, 247] and the work on digital signatures [26, 245].

Through the process of constructing a cryptographic system to achieve security without identification, researchers over time both reformulated the desired properties and proposed new ones. Chen and Brands introduce the notion of single-use credentials, signatures that can only be shown once without compromising the unlinkability of the presenter [248, 36]. Although it is questionable whether this is more desirable than a credential that can be presented repeatedly across many different pseudonyms without being linked. Brands' thesis also provides a comprehensive review of desirable privacy properties of such a credential system [36]. Lysyanskaya et al introduce credential revocation, the requirement for issuing organisations to revoke credentials they have issued without compromising the privacy of credential holders [249]. This is an important aspect of a credential system, especially if the credentials are multi-show and long-lived. Lysyanskaya et al also introduce *identity binding*, identifying the need to prevent illegitimate credential sharing amongst holders [249].

The point that is worth emphasising is that the conceptualisation of the system describes its ideal properties as it interfaces with the external environment. How these properties concretely realise the idealised properties of the system is left unspecified and remains open to interpretation and innovation. As discussed later in this chapter, many concrete instantiations of a credential mechanism have been proposed since Chaum's initial publication. Each with their own set of realised properties and limitations.

4.1.2 Key Cryptographic Primitives

Cryptographic protocols are often composed of multiple smaller, well understood mathematical building blocks referred to as cryptographic primitives. There distinction from a protocol can be blurred, such as in the case of a digital signatures, which is both a protocol solving the use case for document signing and asymmetric authentication and a primitive that can be adapted and combined to produce other, more complex protocols. A credential mechanism is one of these, a signature on a set of claims that aims to achieve the properties initially outlined by Chaum [27]. Since then, researchers have attempted to realise these properties by constructing mathematically defined protocols from the available cryptographic primitives of the time [250, 251, 252, 248, 249, 36].

Cryptographic primitives are interdependent and mutually reinforcing, such that the discovery of a new primitive often leads to the development of further primitives and facilitates new and novel protocol constructions. For example, digital signatures were combined with a commitment scheme to produce blind signatures [150], commitment schemes are used extensively in Zero Knowledge Proof (ZKP) [253] and advances in ZKPs led to the development of new signature schemes [158]. Each low-level primitive had to be envisioned, abstractly defined, mathematically realised and proven secure before they could be used as concrete building blocks in higher-order protocols. Although there are also examples of research that uses abstract cryptographic primitives and demonstrates how, if they could be instantiated, they could be combined to produce a specific protocol [252].

This section reviews the cryptographic primitives that were instrumental in realising the Camenisch and Lysyanskaya (CL)01 credential mechanism, which is based on a group signature scheme on an arbitrary array of messages. Each message can optionally be blindly signed commitment to a value, such that the signer never learns the value signed or the signature they produced on this unblinded value [28]. Furthermore, this scheme supports efficient ZKPs of knowledge of signed message values enabling selective disclosure and unlinkable multi-show credentials [28]. Finally cryp-

tographic accumulators were added to this credential mechanism to provide a method for anonymous credential revocation [254].

4.1.2.1 Digital Signatures

Digital signatures were first mathematically specified in 1978 with the RSA signature scheme [245]. This introduced a new mathematical setting and assumption, the multiplicative RSA group and the assumption that factoring the product of two large primes was an NP-Complete problem. The RSA scheme demonstrated that it was mathematically possible to achieve the one-way authentication system described by Diffie and Hellman [26]. However, it was not until 1985 that El Gamal realised a public key cryptosystem using the discrete log problem [255]. Schnorr then demonstrated how the Fiat-Shamir transformation [253] could be applied to an interactive protocol for proving possession of a discrete logarithm developed by Chaum [256] to define a signature scheme whose signatures were a non-interactive zero-knowledge proof of knowledge [158]. A similar approach was first applied to a credential mechanism by Chen [248] and layer by Lysyanskaya [249].

There were also signature schemes developed that supported additional properties, other than the sign and verify operations specified in RSA, El Gamal and others [245, 255]. Chaum introduced the concept of a blind signature, a primitive that allowed a signer to sign a message without learning the contents of the message and subsequently demonstrated how this could be realised within a multiplicative RSA group [150, 247]. Specifically, a blinding factor is added to a message before it is signed, and then removed from the signature after signing, resulting in a signature on the unblinded message due to the homomorphic properties of RSA. This primitive was then applied by Chaum and others to propose protocols for credential mechanisms, untraceable electronic cash and election ballots [246, 250, 251]. Blind signatures provided a solution to the problem of getting a signed attestation from an entity, without the entity also getting access to that same signed attestation.

Group signatures are another type of signature scheme defined during this period that is closely connected with the development of credential mechanisms. First concep-

tualised by Desmedt when he outlined a set of cryptographic requirements for groups in society. A group signature is the ability for a member of a group to sign a statement on behalf of the group as opposed to as an individual actor [257]. However, it was Chaum and van Heyst's insight that a group signature scheme could be created by generalising credential authentication schemes that led them to specify four distinct group signature schemes. These schemes had their limitations, three required a trusted authority and the public key sizes were linear in the size of the group making them infeasible for large groups [154]. Such limitations provide insight into what was possible due to the state cryptographic research during the early 1990s.

Early in his career Camenisch worked on group signatures schemes [258, 259, 260], before adapting one of these schemes to produce a system for efficient, non-transferable anonymous credentials with Lysyanskaya [28]. They then subsequently generalised this into a group signature scheme with efficient protocols [29]. Group signature schemes have proven an effective primitive for realising credential mechanisms, as further emphasised by the development of the pairing-based BBS+ signature scheme which will be reviewed in detail later [214, 30].

4.1.2.2 Hash Functions

A cryptographic hash function is a foundational building block used within many other cryptographic primitives and protocols. They map an arbitrary length input to a fixed length output and were identified as a component of a digital signature scheme by Diffie and Hellman [26, 261]. Merkle first introduced the security properties of collision resistance, preimage resistance an second preimage resistance [262] and these were formalised through subsequent research [263, 261]. These security properties are defined in Appendix B.1.

Throughout the 80s numerous hash functions were proposed and many of these were subsequently demonstrated insecure [261]. The two that emerged were Message Digest Algorithm 5 (MD5) and Secure Hash Algorithm (SHA) 1. In 2004, Wang et al demonstrated how to find a collision for MD5 and significantly reduced the security of SHA-1 [264]. These have both been replaced, first by SHA-256 and now SHA-3 which

was selected from a NIST competition and has since been standardised [265].

Cryptographic hash functions are used within signature schemes to map an arbitrary message to a fixed sized input within the domain of the signature scheme (e.g. a 256-bit number). Assuming the hash function meets the formalised security properties then signing the hash of a message can be viewed as equivalent to signing the message. Hash functions increase efficiency of signature schemes by preventing the need for large messages to be chunked and individually signed. Furthermore, hash functions provide data integrity preventing modification attacks and existential forgeries within a signature scheme [266].

4.1.2.3 Cryptographic Commitments

A cryptographic commitment allows one party to commit to a secret value, such that the commitment can be shared with others for use in cryptographic protocols with confidence that only those who know the opening to the commitment can discover the secret value. Chaum developed a commitment scheme in the RSA group when constructing a blind signature scheme [247]. They are also used in the Fiat-Shamir transformation [253]. Commitment schemes should satisfy properties for both binding, the value committed to cannot be changed in the future, and hiding, from a commitment it is infeasible to reveal the secret value committed to. Pedersen defined a commitment scheme in the discrete log setting that was perfectly hiding - no amount of computing power would be able to reveal the correct secret value from the commitment without knowledge of the opening (See Appendix B.3). This powerful property is due to there being multiple indistinguishable solutions [155].

4.1.2.4 Zero-Knowledge Proofs of Knowledge

Zero-knowledge is another concept that emerged during this time period. It proposed that it should be possible for one party, a prover, to prove statements of knowledge to another party, the verifier, without revealing any more information than the validity of the statement the prover is attempting to prove [156]. These are probabilistic proofs that aim to demonstrating that the residue of doubt is provably negligible [267]. Gold-

wasser et al produced a seminal paper generalising the concept of an interactive ZKP system over a language, $\mathscr{L}$, of which numerous specific instances had previously been defined [267, 268, 269]. This publication provided a theoretical basis in computational complexity for determining the knowledge in a statement. Blum et al subsequently demonstrated that interaction, the exchange of messages between a verifier and prover, in any ZKP system could be removed if participants shared a common reference string, thus defining non-interactive ZKPs [270]. These general definitions of zero-knowledge protocols lacked efficient concrete instantiations apart from a small subset of languages, although both Damgård and Lysyanskaya devised a theoretical credential mechanism predicated on the existence of a general ZKP system [252, 249].

Zero-knowledge proofs of knowledge are a specific set of statements of knowledge that are derived from knowledge of a variable in a mathematical relationship, commonly discrete logarithms. They have proved useful and efficient for constructing credential mechanisms [29]. Demonstrating knowledge of some x in g^x without revealing anything to the prover other than the fact that the prover did indeed have knowledge of x and more complex relations were explored by Chaum and Camenisch throughout this early period [267, 256, 271, 272, 273]. In such settings the language was well-defined and understood, the mathematical operations possible in a finite cyclic group of known order. Fiat and Shamir demonstrated an efficient way to transform these interactive protocols into non-interactive ones by using a hash of the public protocol values as the challenge which both the prover and verifier could reproduce independently [253]. Schnorr then demonstrated that this transformation could be used to produce a signature scheme [158]. In 1997, Cramer and Damgård demonstrated that these ZKPs of discrete logarithms were an instance of a more general class of ZKPs of knowledge that exist for one-way group homomorphisms [274].

It was Chen that first applied these advances in zero-knowledge protocols to produce a credential mechanism, however the suggested mechanism only applied ZKPs during the issuance protocol with issuers producing a proof that they have correctly issued a credential [248]. Camenisch published a general proof system for proving linear relations between discrete logarithms and applied this to the proving protocol

of a credential mechanism. This allowed the holder of a credential, a signature on some messages, to use the signature to generate proofs of knowledge, which could subsequently be transformed using the Fiat-Shamir transformation, that attested to the fact that the set of messages disclosed were issued as part of some credential [272, 28].

4.1.2.5 Cryptographic Accumulators

A cryptographic accumulator, first conceptualised by Benaloh and De Mare in 1993 [275], although with parallels to Merkle's tree authentication [276], is a cryptographic primitive that supports the representation a large set of values as a single smaller value. Using this accumulator, along with a witness, it is possible to construct ZKPs that the witness has been accumulated in the accumulator, or that it has not [277]. As well as describing an instantiation in the RSA setting, a number of applications for this primitive were proposed to demonstrate how this might be applied in practice and why it is valuable [275]. These included the document time-stamping use case first introduced by Haber and Stornetta [278] and privacy-preserving membership testing without a dependency on a trusted party. In 2002 Camenisch and Lysyanskaya proposed how this primitive could provide an efficient, privacy-preserving revocation for a credential mechanism [28, 254].

4.1.3 Understanding the Mathematical Setting

Public key cryptography since its inception has been intrinsically linked to group theory. A branch of mathematics that abstractly defines a group as a set of elements and a well-defined operation, such that applying the operation to any two elements in the group meets certain rules (See Appendix A.2). Mathematicians have been studying and formalising knowledge about numbers and groups and their properties since the 16th century, with important contributions from Napier (1550-1617), Fermat (1607-1665), Euler (1707-1783), Lagrange (1736-1813) and Galois (1811-1832) establishing the principles of numbers and abstracting these from the decimal number system familiar to the general population. However, it is only since the 1970s that cryptographers

actively started applying group theory to realise and evaluate cryptographic protocols. The mathematical setting provides the fundamental elements a cryptographer uses to define cryptographic primitives and protocols. Only once these constructions have been defined in a concrete setting can they be judged to be theoretically practical and evaluated in terms of both its theoretical efficiency and security. Over time new mathematical settings are introduced, or new complexity assumptions are developed within an existing setting. This changes the available possibility space from which a cryptographer can construct a protocol and demonstrate its security.

During the early period of modern cryptography, there were primarily two classes of groups of interest to cryptographers: Multiplicative groups of prime order, p, and the RSA group, a multiplicative group modulo n, where $n = pq$, and both p and q are prime. Both groups contained integer elements. From finite, cyclic, prime order groups the Diffie-Hellman problem arises: take a generator of a group, g, and use the group operation ($\times \mod p$) to apply g to itself some secret value $x < p$ times, then g^x will produce a group element that for large groups, it is computationally infeasible to determine the secret x. Diffie and Hellman famously showed how this could be used to produce a key exchange protocol over an insecure channel [26]. The RSA group was the other group to receive attention due to it being the first mathematical setting a public key cryptosystem was realised [245]. Knowledge of p and q enables the order of the group to be determined $\phi(n) = (p-1)(q-1)$, which in group theory is the maximum number of times you can apply an element to itself. Doing so will always be equivalent to doing nothing to the element. From this Rivest et al showed that by computing the multiplicative inverse d of some value e modulo $\phi(n)$ (I.e. $e \circ d \equiv 1 \mod \phi(n)$) e and n could become a public key with $(d, \phi(n))$ a secret key. With knowledge of only n and e, it is thought to be computationally infeasible to calculate d for a large enough n, due to the assumed hardness of integer factorisation of n. Therefore $\phi(n)$, the order of the group, remains unknown to other parties.

The cryptographer's objective is then to show how, in some concrete mathematical setting, specific cryptographic protocols and primitives can be realised. They seek to hide information in computational complexity, such that knowledge of some secret

information allows the application or demonstration of a relationship to other information, that computationally are infeasible to produce without this secret information. These relationships must be discovered and described under a well-defined group and then mapped to some meaningful human context and purpose. For example, the RSA cryptosystem uses an invertible relationship between e and d modulo $\phi(n)$, from which it was shown possible to achieve both an encryption and a digital signature scheme in a group defined by multiplication modulo some large n [153]. It took almost ten years before El Gamal described a mathematical relationship to realise a digital signature scheme under the discrete log setting introduced by Diffie and Hellman [255]. The same idealised system, a digital signature scheme, realised under two distinct mathematical settings.

The choice and understanding of mathematical settings during this time impacted the protocols that could be designed. Not all cryptographic primitives had been well-defined under both finite cyclic groups of prime order and the RSA group. For example, ZKPs of knowledge combined with the Fiat-Shamir transformation used in Schnorr signatures and Chen's credential mechanism were, at the time of their inception, only defined under the discrete log assumption [253, 158, 248]. Brands, in his doctoral thesis, defines a credential mechanism that could be realised in either an RSA or discrete log setting and in doing so emphasised the similarities between the two settings in terms of functionality. Whilst also contrasting their differences in terms of practicality and efficiency, indicating a preference for the discrete log setting due to its faster computability and smaller storage space required to realise the same security properties of the RSA function [36, pp. 66-67]. This combination of the two settings was extended further by CL who developed a signature scheme for a credential mechanism which realised credential issuance under a safe RSA setting and a credential presentation protocol by adapting Schnorr signatures under the discrete log setting [28, 29].

The RSA and discrete log settings were the most widely studied and applied when designing cryptographic protocols in the period before the 2000s. However, during this time, new mathematical settings were being proposed as suitable for public key cryptography. Elliptic curves constrained by a finite field with a well-defined operation

for adding two elliptic curve points (elements of the group) were introduced to cryptography by Koblitz [279] and Miller [280]. Elliptic curves are appealing because the Diffie-Hellman problem and the knowledge surrounding how to apply this to design cryptographic protocols were found to be largely transferable, the difference being the elements and operation under which the group is specified. These differences meant that protocols were more efficient, both in terms of storage (if group elements were smaller) and computation (if group operations were faster to execute), while achieving the same level of security. Koblitz in a report produced in the year 2000 stated that a 160-bit representation of an elliptic curve point gives equivalent security as a 1024-bit prime modulus. Furthermore, he shows that for the same operation in both settings (an element operated on itself k times), the elliptic curve setting is at least eight times faster [281]. Elliptic curves subsequently led to the discovery and application of a bilinear pairing between elliptic curve groups in cryptography, a one way mapping of elliptic curve points in two groups G_1 and G_2 to a third group G_t [282, 283]. The pairing setting unlocked many new possibilities in the design of efficient cryptographic protocols and primitives as will be discussed later in this chapter.

Lattice cryptography is another mathematical setting explored in cryptography during this period. Lattices are an instance of a group that have different elements and group operations. They are applicable to cryptography because of two basic problems thought to be computationally hard; the shortest vector problem and the closest vector problem [284]. The details of lattice cryptography are outside the scope of this thesis. The important point is that different mathematical settings usable and useful to cryptographers exist. They include: a set of elements, a well-defined operation that forms a group over these elements and a problem assumed to be computationally hard. Different settings have different implications for security and efficiency. Another example of this is that protocols and primitives realised under lattice cryptography are widely believed to quantum resistant, in contrast to those based on the discrete log assumption which Shor showed could be broken with a quantum algorithm, assuming a quantum computer is realised in the future [285, 286].

It takes time for any mathematical settings and their assumptions to become es-

tablished, well understood and trusted within the cryptographic community. During the 90s the introduction of the Decision Diffie-Hellman problem [287] and the Strong-RSA problem [288] enabled the first provably secure signature schemes [289]. These two assumptions formed the basis of proofs of security for CL-RSA signatures and the credential mechanism constructed from them [28, 29].

4.1.4 The Security of Cryptographic Primitives and Protocols

The security model of the adversarial environment in which a protocol will exist within must be formally defined before any realisation of the protocol can be judged secure against this model. This includes classifying the types of attacks a protocol can handle and the types of breaks to the protocol that could occur. In 1988, after successive signature schemes had been shown to be insecure, Goldwasser et al formally defined the security of a signature scheme as a scheme that is existentially unforgeable under adaptive chosen plaintext attack [290]. This states that even when an adversary is able to adaptively request arbitrary signatures on messages from a signing oracle, they are unable to produce a single forged signature for a message they have not already requested from the signer. This is the most general security conceivable for a digital signature scheme.

Similar attempts to formalise the security model have been developed for other protocols. For example, the properties of group signatures, were formally defined by Bellare et al, first in the case of static groups [291] and later extended to consider dynamic groups [292]. These properties are:

- **Full Anonymity**: An adversary with access to all members signing keys and an opening oracle that returns a group members identity from a group signature is unable to trace the identity of a signer from a signature that they have not queried the opening oracle for.

- **Full Traceability**: A group manager is always able to trace a signature back to the signer. This holds for sets colluding members including those with knowledge of the group manager secret, where tracing identifies some member from the set.

If these two properties hold, then signatures can only be produced by group members. These signatures will be unlinkable to a member of the group unless opened using the group manager's secret key, whereby they will by fully traceable to the signing member. These properties cover the framing and coalition resistance discussed in earlier work [260, 293]. When group signatures are used to construct a credential mechanism, the group member producing a signature is equivalent to a credential presentation with full anonymity protecting the member from linkability across presentations. Full traceability is often removed from the scheme, Camenisch and Lysyanskaya refer to this as optional anonymity revocation in the credential mechanism they propose based on a group signature scheme [28].

Another important aspect of security for a specific cryptographic primitives and protocols involves specifying the assumptions under which the security of the protocol holds. This includes identifying the mathematical problems that are assumed to be computationally hard, for example the Decision Diffie-Hellman Problem [287], as well as assumptions about the access to idealised mathematical operations. The Random Oracle Model (ROM) is used to prove security based on the assumption that the actors completing the protocol have access to a secure random number generator [294]. This, in practical implementations, is then replaced by a cryptographic hash function. This is common throughout cryptography, in particular any protocol that uses the Fiat-Shamir transformation is only provably secure in the ROM [253]. The implications of this abstraction have since been questioned by researchers, showing that just because a scheme is provably secure in the ROM does not guarantee the security of a practical implementation under known assumptions [295]. Nevertheless, the authors do concede the practical benefits of using such a model.

4.1.5 Credential Mechanisms 1976-2001

The credential mechanisms developed up until the early 2000s demonstrate how cryptographic knowledge evolved and became increasingly systematised over this period. A credential mechanism started off as an idea [27], it was then specified both mathemat-

ically under the RSA assumption [250] and in the discrete log setting [248, 249], as well as shown theoretically possible using abstract generic constructions [252, 249]. This is similar to public key cryptosystems, which when initially introduced by Diffie and Hellman were shown possible if one-way functions existed [26]. Each of these mechanisms came with proofs of security, although it wasn't until the work of both Brands and Camenisch and Lysyanskaya that a concrete protocol was proven secure under standard assumptions [36, 28]. Specifically, the Decision Diffie-Hellman assumption [287], the Strong-RSA Assumption and the ROM [288, 289]. This is largely because these assumptions themselves, as well as a definition for the security of signatures, were still being formalised in cryptographic thought during this period [290].

Over time, the fund of cryptographic knowledge that researchers could draw on, in terms of available cryptographic primitives, known and understood mathematical settings and both models and proofs of security of realisations within these concrete mathematical settings increased. Table 4.1 lists the different credential mechanisms developed during this period. All mechanisms developed after the introduction of ZKPs of knowledge and their non-interactive transformation due to Fiat and Shamir [253], made use of this primitive in some manner in their construction [248, 36, 249, 28]. Brands and Camenisch both described protocols for credentials with more than one attribute, before this, credentials had been conceived of as a single piece of information deterministically mapped to a number (group element) within the domain of the protocol. Furthermore, both publications made use of complex ZKPs of knowledge to prove linear relationships between these attributes [36, 28]. Brands was also the first to specify a credential mechanism under elliptic curve groups $E(\mathbb{Z}_p)$, bringing significant efficiency gains that made his constructions suitable for smartcards [36]. In 2001, Camenisch and Lysyanskaya were able to create an efficient protocol that could benefit from these advances in ZKPs of knowledge to create new mechanisms to realise many of the properties first outlined in security without identification [28, 27]. Including for the first time, a practical construction of credentials that could be presented multiple times without being linkable across presentations. As Brands noted in his thesis, such multi-show credentials require a mechanism for revocation which itself can be a vector

Publication	Year	Mathematical Setting	Security Assumption	Single/Mutli Show	Attributes	Comments
Chaum & Evertse [250]	1986	RSA Group	RSA Assumption	Single Show	No	Semi-trusted third party
Damgård [252]	1988	Abstract	Existence of one-way functions and ZKPs	Multi Show	No	Not practically implementable. Does not protect against credential sharing
Chen [248]	1995	DH Group ($\mathbb{Z}_p^*$)	Computational Diffie-Hellman and ROM	Single Show	No	Does not protect against credential sharing
Lysyanskaya et al (Generic) [249]	1999	Abstract	Existence of one-way functions and ZKPs	Multi Show	No	No known realisation
Lysyanskaya et al (DL) [249]	1999	DH Group ($\mathbb{Z}_p^*$)	LRSW Assumption	Single Show	No	Non standard security assumption
Brands DLREP [36]	2000	DH Group ($\mathbb{Z}_p^*$ or $E(\mathbb{Z}_p)$)	Computational Diffie-Hellman and ROM	Single Show	Yes	Efficient, but no way to support unlinkable multi-use credentials
Brands RSAREP [36]	2000	RSA Group	RSA Assumption and ROM	Single Show	Yes	Efficient, but no way to support unlinkable multi-use credentials
Camenisch & Lysyanskaya [28]	2001	Safe RSA Group (n the product of two safe primes)	Strong-RSA, Decision Diffie-Hellman and ROM	Multi Show	Yes	Use of ROM in showing protocol now considered to invalidate provable security [295]

Table 4.1: Table of Credential Mechanisms 1976-2001

for compromising privacy. Camenisch and Lysyanskaya proposed how this could be achieved using cryptographic accumulators a year later [254].

4.2 Protocols for a Credential Mechanism from a Signature Scheme with Efficient Protocols

This section reviews the core protocols that form the basis of a credential mechanism constructed from a signature scheme with efficient protocols first practically described in CL01 [28]. These are pseudonym generation, credential issuance and credential presentation. The concrete mathematical definitions are replaced with abstract cryptographic primitives and a short functional definition is provided. This demonstrates how a concept (a credential mechanism) can be realised with an abstract cryptographic protocol (a signature scheme with efficient protocols) constructed from a combination of abstract cryptographic primitives (See Figure 4.2). If a signature scheme with efficient protocols can be concretely instantiated in **any** mathematical setting, then it should be possible to use this primitive to construct a cryptographic protocol realising a credential mechanism. Evidence for this can be seen with the transition from the RSA setting to bilinear pairings [29, 30]. Appendix C reviews the mathematics underpinning the BBS+ protocol, which is a concrete instantiation of a signature scheme with efficient protocols.

All three protocols presented are interactive, meaning they require communication between two entities [296]. Additionally a demo walking through these protocol interactions and the messages that get communicated between the relevant parties in a production implementation using the Hyperledger Indy/Ursa/Aries stack has been made available on YouTube[1] and the code is open source on GitHub[2].

All participants within the credential system have a (single) master secret, following Camenisch, this is a large random number selected from the mathematical setting referred to as S_u throughout [296]. It is used to bind credentials together, creating a disincentive for credential sharing referred to as *all-or-nothing sharing* [28]. Furthermore, it is used to produce pseudonyms, authenticate against them and prove they are linked to the credentials presented.

[1] `https://www.youtube.com/watch?v=rWv4HS5Xhag`
[2] `https://github.com/wip-abramson/aries-jupyter-playground`

4.2.1 Definitions

- $Comm(x, bf) \rightarrow C$: A commitment scheme that takes in an element, x, and a blinding factor bf and outputs a commitment C to x. The Pedersen commitment scheme is often used (see Appendix B.3) [155].

- $OpenComm(C, bf) \rightarrow x$: A commitment scheme can be opened, effectively removing the blinding factor applied. This is used in the context of credential issuance, where the signature received is on a credential is a signature on a set of attributes with one being a commitment to the master secret. The credential holder removes the blinding factor from this signature value to receive a signature on the set of attributes including the master secret, S_u known only to them.

- $PK\{(a, b) : eqn1 \wedge eqn2\} \rightarrow \pi$: A proof of knowledge for elements a and b such that the mathematical equations $eqn1$ and $eqn2$ hold. All other elements within the equations are known to the verifier. These are typically Schnorr proofs (sigma protocols) [272] (see Appendix B.4). Although other proof systems are now available, for example Bulletproofs [297].

- $SPK(\pi) \rightarrow \sigma_\pi$: A non-interactive transformation of a proof of knowledge through the Fiat-Shamir heuristic to produce a signature of knowledge [253]. The prover produces a challenge, c, by hashing the problem instance such that the verifier can reproduce the same c independently.

- $VerifySPK\{\sigma_\pi : (w_1, w_2)\} \rightarrow 0, 1$: Verification of a signature proof of knowledge against a set of witnesses. The verifier must first reproduce the challenge, c, by hashing the proof instance.

- $HashMap(M) \rightarrow M'$: Signature schemes work on messages within a specified mathematical setting, typically a finite group. This means that inputs to the signature protocols must be within that domain, which is especially relevant for any messages being signed. The $HashMap(M)$ function turns an array of M human meaningful messages (the actual attribute values within a credential) and

maps each message into the domain required for the specific signature scheme using a hash function, producing M' (see Appendix B.1).

- $Store(v)$: Store value v in persistent storage so it can be retrieved at a later point.
- $Fetch(v)$: Retrieves a value from persistent storage

4.2.2 Pseudonym Generation

A pseudonym is a cryptographic identifier that enables an entity in control of its associated, cryptographically bound key to (re)authenticate against the pseudonym. They were designed to act as a replacement for username and password-based identification, enabling parties to identify and authenticate virtual entities represented within an information system against identifiers and authentication mechanisms the entities themselves provide. Essentially entities generate pseudonyms which they then mutually exchange and use to authenticate each other and the messages they send during an interaction. An important property of a pseudonym that Chaum identified is that they should be perfectly unlinkable [149]. Such that even for an adversary with infinite computing power, it should be infeasible for them to determine if two pseudonyms were produced by the same entity. This property was theoretically shown to be possible in some of the earliest protocols in the literature, although many of these focused only on pseudonyms for individuals [250, 248]. This overlooked the value and level playing field that the mutual exchange of pseudonyms can achieve. The doctoral thesis of Lysyanskaya revisited these ideas showing how the exchange of pseudonyms could be mutually achieved and suggesting they could be used to instantiate secure communication channels between actors [249]. The DIDComm specification being worked on at the Decentralized Identity Foundation is in many ways a standardisation of these cryptographic ideas, making them available to a wider community of practice [216].

A pseudonym generation protocol should also support the ability for scope-limited pseudonyms. This property is designed to prevent a party from generating more than one pseudonym within defined scope, such as the URL for a website. Scope limited pseudonyms provide a mechanism to limit malicious and fraudulent behaviour first

identified in Chaum's conceptual introduction [27]. Camenisch and Lysyanskaya described a protocol for achieving this property [28]. In the cryptographic literature, pseudonym generation involves a commitment to a (master) secret and authentication against this pseudonym is a signature proof of knowledge of this secret and the blinding factor that produced the commitment. By randomly selecting the blinding factor, an entity can produce an arbitrary amount of unlinkable pseudonyms. Furthermore, by limiting the blinding factor to a specific value, the pseudonym can be scoped to a domain such that for a given master secret, only one pseudonym can be produced [296]. A general protocol description is provided in Figure 4.4. It shows two entities Alice and Bob, where Alice is registering a pseudonym with Bob and then at a later time identifying themselves by providing the pseudonym and authenticating against it by producing a signature proof of knowledge. The protocol does not show Bob registering with Alice for simplicity, although, as Chaum identified, the mutual exchange of pseudonyms is preferable [27]. Pseudonyms provide a foundation upon which additional cryptographic protocols can be built. They provide a means to identify entities and authenticate them over time. Actors, or their software agents, must be able to remember which pseudonyms they used in each of the relationships they have established.

Alice	**Out of Band**	**Bob**
S_a		S_b
$r \leftarrow\$ \mathbb{Z}_q$		
$nym_b \leftarrow Comm(S_a, r)$		
$\pi_1 \leftarrow PK\{S_a, r : nym_b = Comm(S_a, r)\}$		
$\sigma_{\pi_1} \leftarrow SPK(\pi_1)$		
$Store(Bob : (nym_b, r))$		
	$\xrightarrow{(nym_b, \pi_1)}$	
		$b \leftarrow VerifySPK\{\sigma_{\pi_1} : nym_b\}$
		$b \stackrel{?}{=} 1$
		$Store(nym_b)$
$(nym_b, r) \leftarrow Fetch(Bob)$		
$\pi_2 \leftarrow PK\{S_a, r : nym_b = Comm(S_a, r)\}$		
$\sigma_{\pi_2} \leftarrow SPK(\pi_2)$		
	$\xrightarrow{(nym_b, \sigma_{\pi_2})}$	
		$b \leftarrow VerifySPK\{\pi_2 : nym_b\}$
		$b \stackrel{?}{=} 1$
		$Fetch(nym_b)$

Figure 4.4: Alice registering a pseudonym with Bob and later authenticating against it

4.2.3 Credential Issuance

Credential issuance is a cryptographic protocol involving two entities in the respective roles of issuer and holder. The issuer signs a set of statements that they attest to about their relationship with the holder. The signature, along with the set of signed statements, are then given to the holder to store as a credential. These credentials can then be used in future interactions with other entities, either in presentation or group signature protocols [28, 29]. The signature that gets issued should be compatible with these cryptographic protocols, supporting the selective disclosure of attributes and removing unintended correlatable information across multiple presentations. Additionally, there

needs to be a mechanism to enable the holder to prove that the credential was issued to them and no one else. Even the issuer should be unable to use the credentials they issue. Since a signature scheme with efficient protocols was first defined, there have been numerous concrete cryptographic protocols instantiated that realise the properties of the protocol in multiple mathematical settings. These include RSA-based [29], bilinear pairing under the LSRW assumption [298, 31], bilinear pairing under the q-Strong Diffie-Hellman assumption [299, 30] and Lattice cryptography [300].

The general protocol to achieve this, based on a signature scheme with efficient protocols is outlined in Figure 4.5. The issuing party has previously created keypair (ipk, isk) for the signature scheme being used and intends to issue credentials against a set of attributes A. A credential offer, CO, contains an array of attributes, A, and an array of messages, M, values the issuer is offering to sign as part of the credential. The holder and the issuer have mutually authenticated each other against their pseudonyms and used this to establish a secure communications channel across which protocol interaction takes place.

Issuer	**Secure Channel**	**Holder**
$S_i, (isk, ipk)$	$(nym_{h,i}, nym_{i,h})$	S_h
$A \leftarrow (a_1, a_2, ...a_n)$		
$M \leftarrow (m_1, m_2, ...m_n)$		
$CO \leftarrow (A, M)$		
	$\xrightarrow{CO}$	
		$bf \leftarrow_{\$} \mathbb{Z}_q$
		$C \leftarrow Comm(S_h, bf)$
		$[\pi \leftarrow PK\{(S_h, bf) :$
		$C = Comm(S_h, bf) \wedge (S_h \in nym_{h,i} \wedge C)\}]$
		$\sigma_\pi \leftarrow SPK(\pi)$
	$\xleftarrow[\text{Accept}]{(C, \sigma_\pi)}$	
$b \leftarrow VerifySPK(\sigma_\pi : C, nym_{h,i})$		
$b \stackrel{?}{=} 1$		
$M'_i \leftarrow HashMap(M)$		
$\sigma' \leftarrow \mathsf{Sign}([...M'_i, C], isk)$		
	$\xrightarrow{(M, A, \sigma')}$	
		$\sigma \leftarrow OpenComm(\sigma', bf)$
		$M'_h \leftarrow HashMap(M)$
		$b \leftarrow Verify(\sigma, [...M'_h, S_h], ipk)$
		$b \stackrel{?}{=} 1$
		$Cred \leftarrow (A, M, \sigma)$
		$Store(Cred)$

Figure 4.5: Overview of Credential Issuance protocol

1. The issuer offers a Credential containing a set of n messages against a set of A attribute names to the holder $M = (m_1, m_2, ...m_n)$, $A = (a_1, a_2, ...a_n)$

2. If the holder accepts, they create a commitment to their master secret $C = Comm(S_h)$ and send this to the issuer along with a proof π that the commitment is correctly formed.

3. The issuer verifies the proof, then transforms all messages in $HashMap(M) \rightarrow M'$ into the message domain, elements of a group that can be used in the signature scheme.

4. The issuer appends blinded master secret C to the set of messages M'.

5. The issuer then signs using their secret issuer key creating a partially blinded signature: $Sign([...M', C], isk) \rightarrow \sigma'$.

6. The issuer sends the signature along with the messages that were signed to the holder.

7. The holder unblinds the signature by removing the blinding factor they added: $\sigma \leftarrow OpenComm(\sigma', bf)$

8. The holder verifies the unblinded signature σ' against the set of messages in the credential and their master secret. Again raw messages must be mapped into the domain specified by the mathematical setting of the protocol: $Verify(\sigma, [...M', sk], ipk)$

9. The holder stores $Cred = (\sigma, M)$ in their system.

4.2.4 Credential Presentation

Credential showing or presentation is an interactive protocol whereby a holder that has previously been issued, and has stored, any number of credentials in their storage system ($Ds = (C_1, C_2)$) can use these credentials to prove any subset of the statements were signed by a specific issuance key. Each credential, C, contains a signature, σ, against an array of messages, M, with each message representing the value of an attribute specified by an array of attributes, A, that give meaning to the messages. In addition to this, the messages signed by the signature includes the holders master secret S_h. The presentation protocol involves the holder disclosing any subset of message values and proving with a signature of knowledge that they know a signature, signed by a specific key pair (isk, ipk), created against the disclosed message values. Furthermore, the holder proves that all signatures they use to attest to these message values contain the same master secret, S_h, which the holder has knowledge of. Optionally, the holder can prove that the pseudonym they are identified by was also produced by the master secret contained within the credentials. The verifier is sent the messages the holder

wishes to disclose along with an aggregated Schnorr proof, proving this combination of statements which can all be represented as relationships between discrete logarithms [272, 28]. The representation of the production and verification of the Schnorr proof within the protocol outline (Figure 4.6) has been simplified.

The protocol outline in Figure 4.6 shows the interaction between two parties in the role of verifier and holder, respectively. Both parties have identified and authenticated themselves by establishing pseudonyms for each other. The verifier first defines a set of attributes they wish the holder to disclose, A_p, as well as any constraints, C_s on the presentation such as the public key they require to have signed the credential. This is then communicated to the holder along with a random nonce, r_n, which the holder is expected to include in their proof to prevent replay attacks. The holder then queries their storage system, retrieves the credential objects and produces the relevant Schnorr proof, assuming they can meet the verifier's request. Upon verification of this proof, the verifier learns the following with confidence:

- The signature was created in that moment because it includes the nonce, they communicated at the start of the protocol. With this added liveness nonce, replay attacks are considered infeasible.
- The messages disclosed were signed by the public keys used to verify the proof (the issuers) and these messages have not changed since they were originally signed.
- The holder knows the master secret included in each of the credentials used to construct the presentation and that this secret is the same across all credentials.
- They can optionally learn that the master secret these credentials were issued under is the same one used to construct the pseudonym that they identify this actor against (not included in Figure 4.6).

Verifier	**Secure Channel**	**Holder**
S_v	$(nym_{h,v}, nym_{v,h})$	S_h
		$Ds = (C_1, C_2)$
$A_p \leftarrow (a_1, a_3, a_n)$		
$Cs \leftarrow (cs_1)$		
$r_n \leftarrow_{\$} \mathbb{Z}_q$		
	(A_p, Cs, r_n) → Request Proof	
		$A_p \stackrel{?}{\in} Ds$
		$C_1 \leftarrow Credentials(A_p \in Ds)$
		$C_1 = (A, M, \sigma)$
		$M_p \leftarrow Messages(A_p \in C_1)$
		$M'_p \leftarrow HashMap(M_p)$
		$r \leftarrow_{\$} \mathbb{Z}_q$
		$C_{S_h} \leftarrow Comm(S_h, r)$
		$[\pi \leftarrow PK\{S_h, r, \sigma :$
		$M'_p, ipk, nym_{v,h}, r, C_{S_h}\}]$
		$\sigma_\pi \leftarrow SPK(\pi, r_n)$
	$(\sigma_p i, M_p, C_{S_h})$ ← Present	
$M'_v \leftarrow HashMap(M_p)$		
$b \leftarrow VerifySPK\{\sigma_p i : M'_v, ipk, nym_{v,h}, C_{S_h}\}$		
$b \stackrel{?}{=} 1$		
	→ Accept	

Figure 4.6: Overview of the Credential Presentation protocol

This presentation protocol was designed specifically not to reveal linkable information. Each time a credential is shown, the data that is communicated contains no uniquely correlatable information across repeat presentations apart from the attributes themselves. Whereas the need to reveal a public key which the credential was issued against each time the credential is disclosed such as in identity certificates, or some Verifiable Credential implementations, would, over time, completely compromise the privacy of credential holders within the system [36]. Each invocation of the protocol, even if exactly the same information is disclosed, creates a fresh and unlinkable presentation object. The credentials themselves never leave the holders system and

even the issuers do not know the signature values on these credentials. Nevertheless, as holders reveal an increasing amount of information about themselves to a specific verifier overtime, the attribute values themselves clearly may be correlating. Cryptographic literature acknowledges this, even back in 1985, and suggests regular rotation of pseudonyms could be used to avoid unwanted correlation over time [27].

4.2.5 Supporting Protocols

As credential mechanisms matured alongside the systematisation of cryptographic knowledge, researchers began to suggest additional desirable properties for such a system and realise them using cryptographic primitives. These include revocation, delegation and various mechanisms for achieving accountability amongst the different parties in the system. These can be thought of as distinct, but composable cryptographic building blocks that can be optionally included within a credential system to realise specific properties.

4.2.5.1 Revocation

When a credential is issued to a holder, the issuer is providing them with a signed set of attributes that they can present by invoking the showing protocol to any other actors within the system. The issuer does not know when, where, or how often this credential is used. This is an important privacy-preserving feature [28]. The challenge in such a system arises when an organisation wishes to invalidate a credential that it previously issued to an actor, for example, if they misbehave or the credential becomes compromised [301]. Revocation protocols are the process by which the issuer invalidates a credential they have previously issued such that when the credential is presented, the verifier can learn if it has been revoked. A naive approach would be to maintain a database for the verifier to query [254], or to force the user to request a new credential after a certain period of time [36]. Both solutions suffer from high communication burdens. Furthermore, forcing the verifier to query the issuer's hosted service is a privacy concern that allows the issuer to know which users used their credentials with which

verifier and when.

A different approach is to use cryptographic accumulators to manage revocation [28, 254, 302, 301]. A cryptographic accumulator, introduced by Benaloh and De Mare [275], is a way to represent a large set of values as a single smaller value. Using this accumulator, with a witness, it is possible to construct zero-knowledge proofs that the witness has been accumulated in the accumulator, or that it has not [277]. Since their introduction, accumulators have been implemented in a number of different settings such as Merkle Hash trees [277]; RSA [275, 254, 303, 301]; and bilinear maps [302, 214]. Camenisch et al introduced the concept of a *dynamic accumulator* [254], an accumulator that made it possible to add and remove values over time. However, initially, it was only possible to create efficient proofs of membership of the accumulator, not of non-membership. This led to researchers defining *universal accumulators*, which enable efficient proofs to be computed for both membership and non-membership [303]. Another weakness of the dynamic accumulator outlined by Camenisch was the need for a trusted accumulator manager to update the accumulator [254]. Camacho et al then defined *strong accumulators*, which removes the need for this trusted manager, although this came with an efficiency cost [277].

4.2.5.2 Delegation

The ability to delegate a credential is a useful feature in many systems, for example, allowing the valet to drive your car or your secretary to sign a document on your behalf. Delegatable credentials can be used to reduce information leakage from a hierarchy of credential issuers, where being able to trace the chain of issuers could reveal personal information about the individual presenting the issued credential, such as their local council area [243]. An implementation for Delegatable Anonymous Credentials (DACs) was first outlined in the literature by Chase and Lysyanskaya [304], however, the size of the credential increased exponentially with the number of delegations, due to the use of generic non-interactive proofs. Belenkiy et al [305] formally defined DACs, outlining cryptographic credentials that follow a similar delegation model to the real-world. The solution they proposed is practical in that all operations are linear in both time

and space complexity. The solution uses the Groth-Sahai method for non-interactive zero-knowledge proofs [306], where each delegator provides the delegatee with a non-interactive proof of the signature on the credential they are delegating. The delegatee is then able to randomise this proof every time to either present the credential or delegate it further. The chain of delegation is hidden from each delegatee.

This work was built on by Fuchsbauer [307], who developed a DAC system twice as efficient as Belenkiy et al that also supports non-interactive issuance and delegation of credentials. The assumption that the delegator and delegatee should not learn the identity of each other is kept. Camenisch et al [243] further added to the body of research around delegatable credentials, outlining a system that supports attribute-based delegation of credentials. These are then useful for implementing more fine-grained access control. It has also been argued that the same functionality could be implemented efficiently using revocation registries for each attribute [308].

The DAC system proposed by Camenisch [243] makes a weaker assumption on the privacy requirement between the delegator and delegatee than in previous work [305, 307]. Specifically, they assume that knowledge of the delegation chain does not need to be hidden from entities within that chain. The privacy of the delegation chain is only required for the presentation of a delegated credential. While this is a weaker assumption, for most practical use cases, that is, within an organisation, the delegator and delegatee will already know each other. This assumption reduces the need for complex zero-knowledge proofs along the delegation chain and enables the delegatee to use more efficient zero knowledge proofs, Schnorr proofs [158], when creating a presentation token as they have knowledge of all the values in the delegation chain.

4.2.5.3 Accountability

Accountability is not explicitly a cryptographic protocol, rather a collection of properties that can be introduced into existing protocols within a credential system that help to hold different actors to account. It is important to convey that while unlinkability by default is a key property of these systems, it does not mean that actors are anonymous and unaccountable. In fact, the literature covers a wide array of mechanisms that can

be included to add accountability while retaining the privacy of honest users [309]. The approaches discussed include the following:

- Inspection - A holder verifiably encrypts certain attributes from their credentials under the keys of a trusted Inspector. Under some well specified conditions the inspector can decrypt these attributes and thus reveal the identity of the holder [310, 311].

- N-times authentication schemes - Limiting the number of times a credential can be used to authenticate against a service. Constructions for global limits to authentication such as e-Cash schemes [312], locally specified by an application [214] and periodic n-times authentication schemes have been proposed [313].

- Traceable - group signatures schemes often include a trusted manager with knowledge of specific information enabling it to trace the signatures created by members of the group [291]. Tracing is also possible to add to schemes conditionally, such as when an entity authenticates more than the specified number of times [312].

- *Identity* escrow - a holder registers with a trusted authority (perhaps in person) and receive a credential against a master secret attesting to this verification. The holder must prove that all credentials and pseudonyms they create use the same master secret as the one bound to the credential the trusted authority issued [249, 28]. Optionally, verifiers can specify conditions under which anonymity revocation can take place for the interactions. If these conditions are met, the verifier can prove this to the trusted authority who can then reveal the identity-related attributes held in escrow. Note, this can also work with financial escrow.

There is also a focus on ways that other roles within a credential system can be held to account. Issuers need to be prevented from fraudulently issuing credentials and from revoking credentials without cause. Any trusted authority such as a tracing authority, if included in the system, needs to be accountable for their decisions to trace entities. The two common approaches pursued in the cryptographic literature are:

- Removing trusted authorities wherever possible. It is common in cryptography to see protocols first defined using a trusted authority before being improved with its removal. This can be seen throughout the early literature on credential mechanisms [250, 252, 248, 28].

- Where trusted authorities are required, approaches to distribute this trust among multiple entities using threshold cryptography is considered [314]. In 2020 new protocols for fully distributed group signature schemes were published [315]. These supported distribution of issuing and tracing authority among sets of entities, with invocation of the protocols requiring t of n signatures.

4.2.5.4 Linkability

Credentials mechanisms aim to remove unnecessary and unintended points of correlation for digital information exchanges. However, in some circumstances the ability to link identifiers for virtual entities that represent the same physical person across multiple domains is desirable and justifiable. The common solution is to simply use the same unique identifier across these domains, which totally compromises privacy protections. A number of cryptographic protocols have been proposed to address this concern whilst enabling selective linkability of identifiers across domains [316, 317, 318].

Camenisch & Lehmann first proposed a system that enabled individuals to produce pseudonyms through an interaction with trusted converter, who could then perform conversions from these across domains for these identifiers [316]. The converter is modelled as *honest-but-curious* and does not learn anything about the identifiers it is requested to perform conversions of. In 2017 this work was extended to include an important transparency feature, whereby the converter published an encrypted log of all conversion requests it had received. Each individual could then download and decrypt the conversions related to them, providing a mechanism to challenge inappropriate requests. This was further generalised by Garms and Lehmann and applied to create a signature scheme for selectively linkable group signatures, note that a credential

presentation is essentially a type of group signature [318]. This selectively linkable signature scheme has been further refined, with a provably secure proof of security and weaker trust assumptions [319].

4.3 From Theory to Practice

The synthesis of technological artefacts able to execute complex actions by performing mathematical operations on information (bits) held in the memory of computing hardware has a history dating back to Charles Babbage [11]. However, modern computing software written in programming languages recognisable today began in earnest during the 70s and 80s. The C programming language was released in 1971. Microsoft and Apple were founded in 1975 and 1976 respectively. In 1982 Time magazine named the computer their first machine of the year. And in 1984 an advert for Apple Macintosh was broadcast during the Superbowl, stating why this machine would be the reason 1984 would not be like Orwell's famous book 1984[3]. The era of personal computing had begun.

The 90s brought the World Wide Web, an interconnected system of publicly accessible webpages build on top of the Transmission Control Protocol/Internet Protocol (TCP/IP) protocol. Along with new programming languages and tools for interfacing with and developing software; Python, Java, JavaScript and Hypertext Markup Language (HTML) were all developed in the 90s. The introduction of the Web began a wave of innovation, as software developers and entrepreneurs learnt how to develop applications, explore possibilities and create value within this new medium. The transformative potential of networked ICT was becoming clear.

Despite this wave of innovation at the start of the 21st century, thought products from cryptography largely remained theoretical and conceptual existing only in minds and publications of those participating in this highly specialised scientific discipline. Throughout the 90s the introduction of the World Wide Web, the gradual adoption of the Internet and the infrastructure, communications and commerce that it grew

[3] `https://www.youtube.com/watch?v=VtvjbmoDx-I`

to support created an imperative for solutions to use cases that cryptographers had anticipated. The Diffie-Hellman key exchange protocol and its integration into SSL and subsequently TLS is an early example of a conceptual idea, mathematically specified and then implemented and integrated into a real-world system [320]. Apart from this though, cryptographic protocols were simply not in the toolkit of typical software developers of the time. The tools had not been developed. The success of Bitcoin and the subsequent innovation in distributed ledgers has changed this. Funnelling resources and talent towards the synthesis of cryptographic protocols within software artefacts and their application within real-world systems. This has included a Cambrian explosion of available software libraries that implement different layers of cryptographic knowledge. The process and challenges of taking a theoretical, mathematical construct and realising it within a real-world system are discussed in the context of credential mechanisms in this section.

All theoretical protocols have the benefit of being able to make assumptions about the computing power and capabilities of the parties involved. Including assumptions about agreed semantics for data models, secure key management and other vital implementation details. Rightly so, while important these details are not the responsibility of theoretical cryptographers. Cryptographers aim to demonstrate that protocols are feasible, formally defined, practical and secure. Once that is achieved, it is the responsibility of cryptographic engineers and system architects to understand and implement the algorithms presented in the literature and integrate them into a functional, secure systems architecture. This is a complex process that requires the coordination of a number of highly specialised individuals from across multiple disciplines.

4.3.1 Implementing Cryptographic Algorithms

To implement a cryptographic system, you need to implement the cryptographic algorithms that make up that system. Before you can do this, you must be able to perform operations in the mathematical setting upon which the algorithm has been theoretically specified. Meaning, to implement cryptography somebody has to implement the under-

lying number system and operations on this number system. This includes considering the binary size of the integers required by the algorithm to achieve an appropriate level of security. For example, implementations of RSA cryptography today require key sizes of at least 2048 bits whereas for the equivalent security an Elliptic Curve algorithm requires only 224-bit keys [321]. Importantly though, both key sizes are larger than the 32 or 64-bit integers operating systems are designed to process. Therefore the capability to represent these large cryptographic elements must be programmed in whatever language is being used.

Once the elements of the mathematical setting can be represented within a system, its group operation must be efficiently implemented. This could be modular arithmetic for a large prime, or more recently this involves constructing an elliptic curve-based number system. Whereby the element, P, is an (x_1, y_1) point on the curve, with each of x_1 and y_1 being represented by the underlying BIG number. Elliptic curve addition must then be programmed into the code such that $P + Q = R$ can be evaluated using the correct definition for elliptic curve addition within the code (see Appendix A.4). Also note, the (x_r, y_r) coordinates of the result, R, must be within the domain of the finite field used to constrain the elliptic curve. Furthermore, algorithms can rely on different primes and curve equations. How the elements of this number system are represented and the operation is performed directly affects the practical efficiency of a scheme and can introduce potential security vulnerabilities.

Once the mathematical setting has been implemented, the algorithm for a specified protocol can be implemented. This involves combining elements together in the code in a way specified within the cryptographic literature, for example g^x requires a function that can evaluate element g operated on itself x times. The protocol implementation must then be exposed through an API so that applications can call and invoke these protocols [296].

All of these stages are challenging and implementation bugs can and do compromise the security of the entire system. This has led to the popular meme, *don't roll your own crypto*. In other words, unless you're an expert you should not be attempting to implement cryptographic protocols that you intend to use in production systems. The

notorious Heartbleed bug introduced into OpenSSL indicates that even experts can struggle to engineer secure cryptographic systems [322].

Credential mechanisms were first practically implemented around 2009. In fact two different sets of cryptographic protocols were implemented in different libraries:

- uProve [239] a library implemented by Microsoft based on protocols published by Stephan Brands in his doctoral thesis [36].
- Idemix, a library first prototyped in 2002 [296] based on Camenisch and Lysyanskaya schemes using RSA primitives [28, 29]. This work was further refined into a technical specification and implementation at IBM Zurich [238].

With these implementations and technical specifications, a wider audience of researchers and practitioners could begin to experiment with these technologies. Examples include developing applications like an anonymous petition system [323], analysing the real-world implications of complex protocols such as revocation [324] and evaluating the feasibility of these protocols based on actual resource consumption in specific contexts such as the Internet of Things [325]. Gradually, these theoretical ideas became tangible and accessible.

4.3.2 Key Management

The theoretical security of cryptographic protocols is tied to the assumption that only the appropriate entities have access to specific key material. Private keys are assumed to be kept secret, whilst the correct public keys are made available to the entities that require them. In practice, keys must be generated, stored, used, rotated and recovered by technological artefacts and the human actors that interface with them. The management of cryptographic keys required to be kept secret is often the place in which the security of cryptographic protocol is most vulnerable, both to adversarial actors and human error [266]. Best practice includes using key hierarchies, storing keys in secure hardware modules, enforcing key separation and backing up important keys. Although key backup must be at least as secure as the mechanism employed for

securing the original key. Key management should be considered within the context of the wider system the keys are used and governance procedures for the entire lifecycle should be defined.

Practices for managing public keys have a slightly different set of requirements. They are, by definition, designed to be published such that they can be accessed by relevant parties in order to encrypt messages or verify signatures. They must be highly available and have a mechanism to assess the trustworthiness of the key in relation to its application. This requires strong assurance in the binding of the public key to its controller and any additional information that characterises the key (such as expiry dates) and the entity that controls this key [266]. An example of public key management is CAs, whereby entities register with a CA who then signs a public key certificate attesting to the entities public key and any additional detail about the key or the entity who registered it [36, Chapter 1.1]. DIDs offer another approach to public key management without a dependency on a few centralised root authorities [202].

In the case of credential mechanisms, every actor needs the ability to generate key pairs and securely manage private keys and other secret information [36]. Credentials themselves become private information whose signatures should be protected and mechanisms for recovery should be considered. The experiences of those managing keys for cryptocurrencies such as bitcoin indicates some of the usability challenges this introduces. First, almost half of the people asked within a survey preferred to this responsibility to a centralised service provider. Second, over 20% of respondents had irrevocably lost important keys controlling cryptocurrency [326]. Additionally, credential mechanisms require public key management. Issuers of credentials must be able to manage publish their public issuance keys and these keys must be accessible and trusted by the verifiers of any credential presentations. An infrastructure to support this must form part of the system architecture for credential mechanisms when deployed in practice.

4.3.3 System Architecture

Another important step towards realising these protocols in real-world applications is to develop a system architecture for these protocols and convey the architecture in a language understandable to those not well versed in cryptography. The European Union (EU) funded project Attribute Based Credentials for Trust (ABC4Trust) contributed extensively to this work [240]. They identified the differences in complexity between cryptographic protocols as key factor limiting real-world adoption of these technologies [311]. As a solution, the project created a language framework that abstracted the underlying cryptographic method using Extensible Markup Language (XML) Schema objects (note the now popular JSON was still being standardised during this time) and defined a generic set of actors and interactions within the system (see Figure 4.7). This language extended earlier work such as the credential-based authentication language requirements for specifying the attributes a Verifier wishes to authenticate and provides semantics for the entire privacy-enhancing Attribute Based Credential (ABC) system [327, 311].

The ABC4Trust project published a full system architecture [240] as well as an open source Java implementation for a credential engine designed to be interoperable with both uProve and Idemix cryptographic protocol implementations [238, 239]. In theory, the architecture was designed to enable a common interface to cryptographic protocols, so that the underlying cryptography could be updated as alternative protocols became available. Diagrams such as the one shown in Figure 4.8, illustrated how these cryptographic protocols could be combined into a full system, in which the cryptographic engine is just a small part.

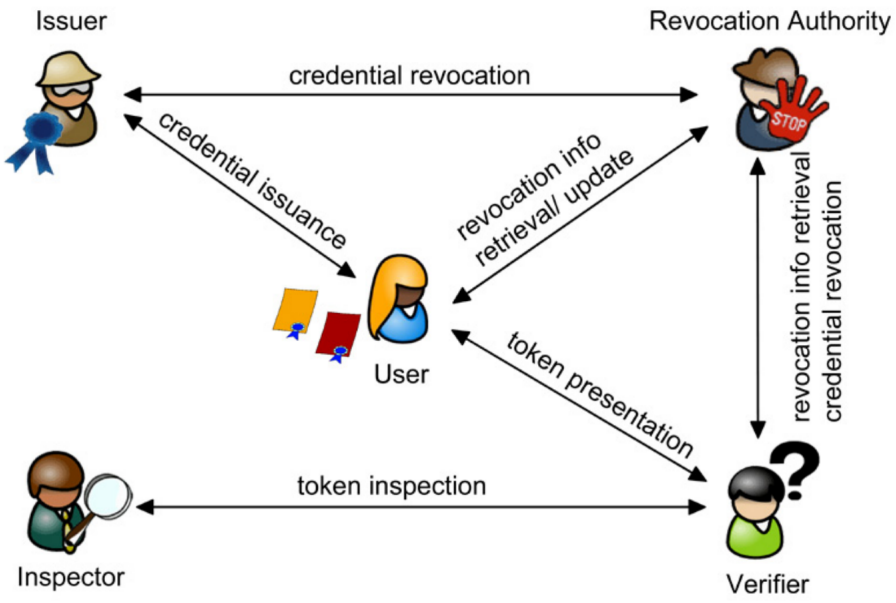

Figure 4.7: Actors and interactions in Attribute-Based Credential system [30]

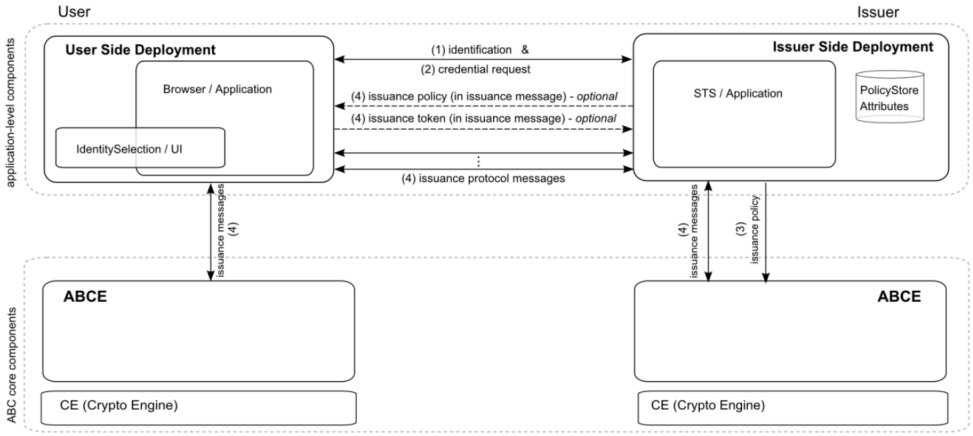

Figure 4.8: Credential Issuance Architecture Diagram (ABC4Trust) [240]

An interesting aspect of the ABC4Trust project was the application of privacy-preserving ABCs to real-world scenarios in two pilot studies [328]. One pilot used ABCs to create an online course evaluation system at Patras University in Greece [329]. The implementation allowed students to anonymously fill out online evaluation forms if they could prove they were a student registered on the course at the university and they had attended the required number of lectures. The other pilot integrated ABCs into a school community platform with different areas specifying different access policies for both children and their guardians [330]. This demonstrated the technology could

work with a younger user group, with results showing they trusted and understood the system and would continue using it given the option. Both tested advanced features including inspection and key binding [328].

A quantitative threat modelling approach based on both security and privacy by design principles was developed and applied to the course evaluation pilot. This research created attack trees for the system, and contrasted the threats and their likelihood to a course evaluation with and without privacy ABCs. Showing that these technologies can help to mitigate some important privacy threats like correlation risk, while acknowledging the new threat vectors introduced to the system [331].

Trust relationships within a privacy ABCs identity management system answer the question of *who needs to trust whom on what?* by identifying 19 distinct trust relationships, with the *user* placing the most trust in other roles within the system [332]. This analysis is key to understanding the implications of deploying this technology in real-world scenarios.

Benenson et al were able to use the university class (30 students) to perform a qualitative study on the user acceptance of these technologies [333]. This study suggested that the perceived usefulness of the primary task and an interface that provides awareness of the data being transmitted as more important factors than an understanding of the underlying technology.

A framework for benchmarking privacy ABCs against four main categories (Functionality, Efficiency, Security Assurance and Practicality) was put forward by Veseli et al [334]. This was subsequently used to evaluate the storage, communication and computational efficiency of the uProve and Idemix protocols [335, 336]. This work identified Idemix as more efficient at the majority of operations, although uProve was shown to be faster for credentials with many attributes and in general for the proving protocols. Idemix also supports unlinkable multi-show credentials whereas uProve requires credentials to be reissued each time. In both systems, revocation and inspection protocols incurred constant costs to the presentation of credentials [335, 336].

The ABC4Trust project represents a maturing in the general understanding of the capabilities of the underlying cryptographic protocols and how to build systems using

them. It took theoretical concepts from cryptography and demonstrated how they could be integrated into a broader framework for software development framework. This included developing a full systems architecture, running a number of pilot studies and an analysis of trust, privacy and practicality of these systems in realistic settings [240, 331, 332, 334]. It also included the first qualitative analysis of actual users interacting with these systems [333].

4.4 State-of-the-Art

At the start of the new millennium Camenisch and Lysyanskaya demonstrated how a practical credential mechanism, achieving many of the properties first identified by Chaum [27], could be realised by adapting a group signature scheme to support efficient protocols for proving discrete logarithm relationships with zero-knowledge protocols and blind signing of committed messages [28]. The previous section reviewed how this cryptographic protocol, and others, became practically implemented in cryptographic libraries, integrated into a technology stack for software development and applied to realistic use cases [238, 296, 240]. Alongside this period of implementation, cryptographic knowledge has continued to be produced and refined. Indeed, the interaction between engineers and theoretical cryptographers, and the application of cryptographic ideas to real-world systems has created a fertile ground for innovation. New ideas for protocols and primitives, realised in more efficient mathematical settings and provable secure under well understood complexity assumptions have been proposed. The last section of this chapter presents the state of the art in cryptographic research in relation to privacy-preserving cryptographic credentials.

It is important to recognise the time in which ideas are conceived. When Chaum proposed the initial idea for a credential mechanism in 1985 [27], it was before society became truly digitised and you could even argue that Camenisch and Lysyanskaya's 2001 publication was before the Internet became ubiquitous and its potential realised [28]. Furthermore, in 2001 pairing-based cryptography had only just been discovered [282, 337]. The state of the art of cryptography presented in this chapter exists in a

different world, one where increasingly powerful, networked computing devices are being integrated into everything [181]. This has created a plethora of new use cases that have different demands from cryptographic systems. They often need to run on constrained devices with limited storage and computation capacity, for example, the Internet of Things. A common modern use case is Vehicle to Vehicle/Infrastructure communication, or more generally sensing and information processing systems in our built environment [338]. Additionally, systems in 2021 exist in unpredictable, highly complex adversarial environments, leading to new models of security that attempts to define the meaning of security in these contexts [242].

4.4.1 Pairing Cryptography

Bilinear pairings, as previously mentioned, provide an efficient setting under which cryptographic protocols based on the discrete log assumption can be realised. Type-3 pairings (see Appendix A.5) offers the best efficiency properties [241] and are widely considered the most useful mathematical setting, replacing the Type-1 setting used in early pairing constructions such as the pairing-based CL signatures [298]. Furthermore, the Decision Diffie-Hellman assumption holds under Type-3 pairings for both groups $\mathbb{G}_2$ and $\mathbb{G}_2$ but not under Type-1 pairings [241, 339].

Pairings have been used to realise many cryptographic primitives including: Accumulators [302, 299], Non-Interactive Zero Knowledge Proofs (NIZKs) [306, 340, 341], structure-preserving signatures where all messages, public keys and signatures are elements from the same group [342, 343], group signature schemes [339, 214, 30, 31] and short signature schemes [344, 345, 346, 31]. All of these have proved useful when proposing new credential mechanisms and improving existing ones.

To emphasise the efficiency benefits of pairing cryptography, Table 4.2 shows the different signature sizes for RSA [245], Elliptic Curve Digital Signature Algorithm (ECDSA) [347] and Boneh-Lynn–Shacham (BLS) [344] signature schemes that achieve the same (128-bit) security level. RSA takes up a substantially larger number of bits than both other schemes, but BLS - a pairing signature scheme - produces signatures 50% smaller

than the ECDSA scheme. Furthermore, analysis has shown that the BLS scheme is twice as efficient as the ECDSA scheme at producing signatures [344].

Signature Scheme	Size of Signature (bits)
RSA [245]	3248
ECDSA [347]	512
BLS [344]	256

Table 4.2: Comparison of digital signature sizes that achieve 128-bit security level (adapted from [241])

Implementation of pairing cryptography required the identification of pairing friendly curves with a known, computable pairing function [348]. While the understanding of pairing cryptography and its application to the design of cryptographic protocols has matured, it remains challenging to produce efficient implementations [349]. The two families of curves widely used today in practical implementations are BLS and Barreto-Naehrig (BN) [350, 351]. Further indication of the pace of innovation can seen by looking at the BLS12-381 curve [352]. This pairing friendly elliptic curve was specifically designed by Bowe, to speed up cryptographic operations in the production ZCash protocol [352, 164, 163]. It has since been implemented in a number of existing and emerging open source cryptographic libraries and has been included in an Internet Engineering Task Force (IETF) draft for pairing friendly curves [352]. Subsequently, Bowe demonstrated how to achieve faster subgroup checks in the BLS12-381 curve [353].

4.4.2 Universally Composable Security

Universally composable (UC) security is an approach to modelling the security of a cryptographic protocol when it is used as a component within any arbitrary system. This model was first proposed by Canetti, who identified the weakness of existing definitions of security that analysed primitives and protocols independently as opposed to as parts of complex and unpredictable environments where the protocol might be run concurrently alongside many other instances [242]. Universal composability provides an approach to guarantee individual protocols can be securely composed, by

demonstrating that the protocol securely realises some idealised functionality defined for the protocol.

This model of security is clearly relevant for credential mechanisms in the real world, whereby an arbitrary number of holders can engage in pseudonym generation, credential issuance, presentation and revocation protocols. Furthermore, these protocols are likely only a small part of a larger software system [240]. Definitions for UC-Secure credential protocols were published by Camenisch et al, alongside a practical realisation of a set of protocols meeting these definitions [354]. This model has also been applied to accumulators [355], credential delegation [243], direct anonymous attestation [356] and revocation [308].

A key challenge to achieving universal composability in a protocol arises when the protocol uses the ROM, which when realised with a cryptographic hash function, as is common practice when applying the Fiat-Shamir transformation [253], does not provide the same security guarantees as the idealised ROM. This makes the composability of such protocols hard to guarantee [295, 357]. Note that a signature scheme with efficient protocols, is only efficient because it makes use of the ROM [29]. Canetti proposed a UC global ROM as an alternative, where there is only a single instance of the random oracle which is made available to all protocol executions [357]. However, this model did not allow the random oracle to be programmable meaning many protocols proven secure under the traditional ROM would not be provably secure in this UC model. Camenisch explore different formulations of the global ROM proposed by Canetti, demonstrating that practical realisations of signature schemes that are existentially unforgeable under adaptive chosen message attack in the original ROM can be proven secure in the UC model under these new formulations [358]. This is a promising result for credential mechanisms that rely on the ROM for security proofs.

4.4.3 Credential Mechanism Constructions

Alternative constructions for achieving credential mechanisms have been proposed in the literature. Some protocols are based on group signature schemes realised under

bilinear settings, while other constructions propose novel approaches.

4.4.3.1 A Signature Scheme with Efficient Protocols

The construction of a credential mechanism from a signature scheme with efficient protocols has been integrated into a number of practical applications [28]. Since first outlined, a number of distinct signature schemes realising these properties have been proposed [298, 214, 31]. Camenisch et al defined the equivalent signature scheme to their original RSA construction in a Type-1 pairing setting [29, 298]. Boneh et al adapted BLS signatures [344] into a group signature scheme (Boneh-Boyen-Shacham (BBS)) which Au et al then extended to BBS+ signatures and applied it to the use case of an n-times authentication system [214]. Camenisch proved the security of the BBS+ signature scheme under the more efficient Type-3 pairing setting [30] and emphasised its applicability for designing credential mechanisms.

PS signatures are another promising signature scheme with efficient protocols first published in 2016 [31]. The signatures are short, consisting of only 2 group elements independent of the number of messages, l, being signed, whereas the size of CL-signatures are linear in l. Furthermore, PS signatures are randomisable such that for any party with a signature σ on a set of messages M, can easily be transformed to σ^t for some random $t \leftarrow\$ \mathbb{Z}_n$ whilst remaining verifiable against the messages M and indistinguishable from a fresh signature. This means that during a proving protocol in a credential mechanism, the signature does not need to be committed to and proven in zero-knowledge instead it can be randomized and disclosed to the verifier. BBS+ signatures are not randomizable [30]. PS Signatures have since been applied to the Coconut credential system designed to work in a blockchain environment with constrained resources and support threshold issuance [359]. Camenisch et al generalise threshold issuance and opening in a group signature scheme based PS signatures [315].

Both PS and BBS+ signatures are short signature schemes defined under a Type-3 pairing setting and they both support the efficient protocols required to construct a credential mechanism. However, both schemes when applied to a credential mechanism require the ROM for proofs of security which is thought to invalidate complexity

assumptions [294, 295]. Although as discussed, recent work introducing the universal composability of the global ROM provides confidence in the ROM and it remains a widely used tool throughout cryptographic protocols [357, 358].

4.4.3.2 Message Authentication Codes

Chase [360], suggests that instead of building anonymous credentials using public key cryptography, in certain situations, it is possible to use algebraic Message Authentication Codes (MACs) [361] to issue Keyed-Verification Anonymous Credentials (KVAC). These lend themselves to situations where the issuer and verifier are the same entity, such as issuing a bus pass or a credential to access a website. Chase et al point out that these can be combined with existing public-key credentials, as in [28], with the initial registration process verifying a public-key credential and then issuing a KVAC for future use to access the system [360]. They showed that KVAC's were more efficient than the existing anonymous credential implementation, Idemix [238]. Camenisch et al [362] expand this work, developing a KVAC scheme designed for smart cards, stating efficiency improvements of at least 44% in precious technical implementations. This KVAC implementation from Camenisch et al [362] requires only $u+2$ scalar multiplications to present an attribute proof, compared with $2u+3$ in an implementation from Couteau [363] and $u+12$ in the original work of Chase et al [360]. Where u is the number of undisclosed attributes in the credential. This KVAC approach has been implemented in practice and forms the basis of the Signal private messaging system [364].

4.5 Critical Discussion

Identification within information systems is a functional requirement that enables these systems to *recognise, remember and respond* to people and things that interact with them [365]. However, when Chaum published Security without Identification, he was really calling for security without facilitating the pervasive over identification and correlation of people as they interacted with and across ICTs in multiple distinct contexts [27]. In response to this perceived threat, Chaum outlined a conceptual cryptographic

system that aimed to remove unnecessary points of correlation from digital interactions, whilst enabling the presentation of authentic, integrity-assured information attributes characterising the virtual entity interacting with an information system. Therefore enabling a physical entity to present themselves as multiple virtual entities in the different contexts they engage in without these representations being linkable by default. Note that linkability is trivial to introduce into these interactions, Chaum's point was that unique identifiers for individuals such as public keys, or even static signatures on credentials, should not form a part of the underlying protocol.

This chapter has presented the evolution and systematisation of cryptographic knowledge through the lens of a credential mechanism to achieve security without identification. In doing so a conceptual model for the layered, modular structure of cryptographic thought has been proposed (see Figure 4.2). Evidence from the literature validates this model, demonstrating how early ideas, such as security without identification, directed the focus of early cryptographic research. Knowledge was developed about group theory and computational complexity in an attempt to construct concrete, mathematically specified, provably secure cryptographic protocols realising the idealised properties of these ideas. In the process, smaller, discrete cryptographic primitives such as zero-knowledge proofs of knowledge and blind signatures were both abstractly defined and concretely realised in a well-defined mathematical setting. These then became composable building blocks that could be used to construct higher order cryptographic protocols (See Figure 4.3). The transition to a different mathematical setting, for example with bilinear pairings over elliptic curves [283], did not change the idealised primitives and protocols found in cryptographic thought, they simply provided a new substrate in which these theoretical concepts could be realised.

Cryptography, since it emerged as a modern scientific discipline in the 70s, has become a highly systematised area of human knowledge built from the same mathematical foundations that underpins many other scientific disciplines. Furthermore, cryptography is no longer purely theoretical and is rapidly becoming intertwined with software engineering as cryptographic protocols are synthesised into software artefacts and integrated into real-world systems. This chapter has demonstrated that achieving

security without identification within digital interactions is now possible in practice. Although the implementation of complex cryptographic protocols remains challenging.

4.6 Conclusions

Chapter 2 provides ample evidence from within the social sciences that justify Chaum's conceptual idea and its importance within an information society. Individuals hold multiple identities at any one time, whilst never being wholly defined by them. They present a version of themselves they deem appropriate for the context of the interaction, whilst intentionally keeping private that which does not fit [5]. The correlation and compilation of informational records on individuals based on their digital footprint has created huge asymmetries of power, enabling physical human beings to be identified, categorised and targeted with informational input at ever greater precision [74]. Our identities effectively constructed for us, by unknown others with motives not necessarily aligned to our own. Our ability to place trust intelligently has been impacted by the *context collapse* created by our correlatable and correlated digital interactions [16], and further distorted by the different information bubbles that we each exist now exist within. Not to mention our inability to identify the sources of information so that we might judge their trustworthiness.

Large organisations have sought to replace trust, a relationship with uncertainty that is inherently part of the human condition, with tools of influence and prediction as a means to provide security by herding future behaviour. If, following Luhmann, we accept that trust increases social complexity allowing society, and the individuals within it, to perceive a richer set of possibilities from which to select actions from [2]. Then the current approach, broadly categorised as Surveillance Capitalism, can be understood as the reduction in social complexity as possibility is calculated as probability and nudged towards certainty by a powerful few. That our privacy has been compromised as new and inappropriate information flows have been introduced into our interactions is a key enabler for the abuses of power we experience in the digital world today [7].

Chapter 3 emphasised that systems of identification existed long before digital ICTs

as a mechanism to structure a social context. These systems have created asymmetries of power between those identified and those defining, implementing and governing the processes of identification. The augmentation of human systems of identification with digital ICTs can further amplify these asymmetries, creating new avenues for surveillance, discrimination and exploitation. It is widely acknowledged that the privacy of those identified and represented within an identification system is fundamental to their protection from abuse [22]. Privacy-preserving credential mechanisms make it possible for digital ICTs to protect and enhance privacy, by removing points of correlation and eliminating a dependency on centralised databases whilst retaining the integrity of identity-related attributes as they are presented during a process of identification.

It is for these reasons that privacy-preserving credential mechanisms have been selected by this thesis to form the cryptographic foundations of a digitally augmented identification system. The scheme identified for use by this thesis is a signature scheme with efficient protocols. This abstract cryptographic protocol has been realised and prove secure in the RSA [29], Bilinear [298, 214, 31, 30] and even the quantum secure Lattice-based mathematical settings [300]. It is a mature and well understood cryptographic protocol, that Camenisch demonstrated could be used to instantiate a credential mechanism realising security without identification as first set out be Chaum [28, 27]. In contrast to standard digital signature schemes, which while simpler to implement, embed linkability and correlation into any credential-based identification system realised using these schemes. Such identification systems would make it trivial for the logic of surveillance capitalism, which Zuboff demonstrated to be both pervasive and voracious [74], to be applied to these interactions.

Fifth Chapter

Defining a Technical Architecture

This chapter describes a set of design principles and technical requirements which have been used to evaluate existing software architectures for credential-based identification systems. Following this, the three existing architectures are selected from the available solutions for further analysis against the identified technical requirements and the selection of the HVIEP is justified. This selected architecture is reviewed in depth.

5.1 Design principles for digital identification systems

The success of any identification system depends not only on the technical feasibility, but also user acceptance and the trust placed in the system. In this thesis the term *users* refers to all stakeholders (entities) within a specific context that will use the system of identification.

There are many papers that refer to the *Laws of Identity* (coined by Cameron in 2005) [193] as the foundation for design principles for digital identification systems. These laws explain the dynamics causing digital identity systems to succeed or fail within various contexts. Although written before the era of Self-Sovereign Identity (SSI), Cameron himself finds the laws still relevant for identification systems that use distributed ledgers and decentralised identity standards [193]. He points out, for instance, that the first four of the laws are also requirements within GDPR.

In 2016, Christopher Allen [203] wrote 10 principles inspired by (amongst others) the work of Cameron. His aim was to ensure that user control is at the heart of SSI.

Allen pointed out that identity can be a double-edged sword: it can be used for both beneficial and maleficent purposes. Therefore, he states: *an identity system must balance transparency, fairness, and support of the commons with protection for the individual.*

The Sovrin Foundation [366] adopted Allen's principles and arranged them in three sections, but this causes some confusion as they used one principle twice and made another principle a section above other principles. Other researchers and developers have used Cameron's or Allen's principles for inspiration and adapted them to their own lists of features. However, testing of SSI systems against these features and design principles is still rare.

Dunphy & Petitcolas [367] evaluated three identity management solutions (uPort, Sovrin, and ShoCard) against Cameron's laws of identity. Their overview demonstrated that none of the solutions meets all seven laws, and none of them meets the law of human integration: usability; user understanding; and user experience. They state that none of the schemes they evaluated are accompanied by an evidence-based vision of user interaction. One of the limitations that remains unaddressed is usable end user key management for nontechnical users. Furthermore, Dunphy & Petitcolas express concern about tightening regulation, such as with GDPR, as this sometimes contradicts the transparency of data storage in these solutions. Finally, most solutions provide only ad-hoc trust, as trust relies on integration between participating entities and methods to achieve trust in the context of identity-related attributes are still evolving.

Ferdous et al. [368] elaborated on the principles of Allen and designed a taxonomy of essential properties for SSI. Then, they compared four blockchain-based SSI systems (uPort, Jolo, Sovrin, Blockcerts) against the properties and through desk research they found that most of the systems satisfy most of the properties. Similar work was done in a student project [369] where students compared eight blockchain-based methods (IDchainz, Uport, EverID, Sovrin, LifeID, Selfkey, Shocard, and Sora) and three non-blockchain based SSI systems (PDS, I Reveal My Attributes (IRMA), and reclaimID) against each of Allen's principles, with one additional principle [370]. They concluded that some of the blockchain-based solutions fulfil all properties, but that some of the

non-blockchain-based implementations meet most of the criteria, as well. Interestingly, their conclusions as to whether the properties are met do not always match the conclusions of [368] for two systems (Uport and Sovrin) that both projects evaluated. Toth and Anderson-Priddy [206] validated nine properties from earlier sources (e.g. Allen and Sovrin Foundation) and added new properties. They applied these properties to their architecture for digital identification and reasoned how these apply to their solution (NexGenID).

It has been difficult to find published evaluations of values and principles developed through interaction with users in the context of a specific identification system is rare. One project that focused on citizens and digital identification systems, in general, was the Digital Identity lab in the Netherlands. In several interactive sessions with citizens, they found which *values* matter the most when developing digital identification systems [371]. The research methods included interviews in the streets, meet-ups, expert sessions and design sprints. The results include evaluation quadrants to plot digital identity providers and an overview of values that citizens find important. These values can be used as input for ethical design and trust of digital identity systems.

Another project focusing on user experience is the IRMA Made Easy project [372], IRMA is a SSI solution with a digital wallet. The IRMA Made Easy project works on the design of the app and website with a focus on accessibility. The developers of IRMA point out that user experience design affects how users handle the control over their information. From their experience they share three lessons:

1. In order for new technology to be adopted, users require a smooth user experience.

2. User experience design for privacy is not the same as general user experience design.

3. A system that puts people in control over their data does not always lead to people using that control to protect their privacy: it can even lead to the opposite when they are tricked by others.

The literature review demonstrates there is a gap in academic research that includes evaluation of proposed identification solutions from a user perspective with domain knowledge of the ecosystem. Furthermore, the most commonly used design principles have not been validated by users for importance and priority. Projects that included consumers focus on identity management in general, and studies on decentralised identity systems tend to focus on the evaluation by technical experts through desk research. Furthermore, when decentralised identity design principles and features indeed are evaluated, researchers often re-use existing frameworks or lists of principles without user elicitation for principles that technology experts have not imagined yet. If we do not understand the requirements of end users, we run the risk of creating digital tools that no one wants to use, or worse introducing unintended consequences through the deployment of these systems to domains with poorly understood requirements.

We compared the different lists of principles, features and values that we found, and created an overview of different and overlapping principles. The results are presented in Table 5.1. From these a condensed set of eight principles were selected to evaluate different software architectures against and subsequently to validate against a specific domain, healthcare, to address the gap in current research. These principles are defined in Table 5.2.

5.2 Requirements for Credential-based Identification Architectures

There are numerous technical architectures designed to facilitate credential-based identification systems. This section defines a technical requirements against which these architectures can be evaluated. There are many ways that a software development framework can satisfy these criteria and these choices both enable and constrain the possible design space and resulting functionality of artefacts and systems synthesised using these frameworks. This includes the ability to interface, interact and interoperate with software systems instantiated under a different architecture.

Cameron [193]	Allen [203]	Ferdous et al [368]	De Waag [371]	Toth Anderson Priddy [206]
	Existence	Existence		
User Control and Consent	Control and Consent	Consent	Control	Control and consent
	Access	Access	Access	Access
	Transparency	Transparency	Transparency	
Pluralism of operators and technologies	Portability	Portability		portability
Consistent experience across contexts	Interoperability	Interoperability		Interoperability
	Persistence	Persistence		Persistence
Minimal disclosure for a constrained use	Minimalization	Minimalization	Data-minimalization	
	Protection	Protection	Security	secure transactions and identity transfer
		Autonomy	Autonomy	
Justifiable parties		Choosability		
Human integration			Ease of use	usability
		Disclosure		
		Ownership		
Directed identity		Single source		
		Standard		
		Cost		
		Availability		
			Trust	
			Privacy	
			Integrity	
			Decentralization	
			Inclusivity	
			Reliability	
				Counterfeit prevention
				Identity verification
				Disclosure
				Identity assurance

Table 5.1: List of design principles

Design Principle	Definition
Autonomy	An individual is recognised as independent of the digital representations of them persisted within information systems. They should be capable of create and manage as many of these representations as they require, without depending on any third party.
Consent and Consent	The person has agency over the personal data they choose to share and whom they choose to share it with. This data should only be released only after the individual has consented to do so.
Transparency	Systems and algorithms are transparent and anyone should be able to examine how they work.
Flexibility	Identification systems should be flexible, adaptive and extensible such that an identified individuals virtual representation is useful and usable within many different services and interactions.
Interoperability	Digital identification systems are continually available and as widely usable as possible. Where possible they should be based on open standards in order to prevent vendor lock-in and foster open innovation.
Minimal disclosure	When data is disclosed, that disclosure should involve the minimum amount of data necessary to accomplish the task at hand. For example, if only a minimum age is called for, then the exact age should not be disclosed, and if only an age is requested, then the more precise date of birth should not be disclosed.
Protection	The rights of individuals must be protected. When there is a conflict between the needs of the identification system and the rights of identified individuals, then the network should err on the side of preserving the freedoms and rights of the individuals over the needs of the network.
Justifiable Parties	Information flows within systems of identification should be justified against the context within which the identification system is applied. Steps should be taken to prevent unjustified parties having access to identity-related information about subjects identified by the system.

Table 5.2: The list of design principles distilled from the literature

5.2.1 The Credential Data Model

A data model defines the structure of data that must be communicated between distinct software systems and therefore need to be mutually understood by the respective software systems that are interacting. A data model must encapsulate all the information necessary to enable the intended use of this data within software systems. Within credential-based identification systems, the important data models to define are credentials and their derivative presentations as these are the objects that are passed between interacting actors. Specifically, credentials issued by Issuers and communicated to Holders and presentations produced by Holders and presented to Verifiers.

A data model for credential-based identification systems, the VC Data Model, has been undergoing standardisation at the W3C since 2015 and in 2019 it became an official W3C recommendation. This standard defines two data models, a Verifiable Credential and a Verifiable Presentation, each of which can be expressed in a number of formats [34, 213]. Earlier attempts to produce a credential data model can be seen in academic projects such as the EU ABC4Trust project, which used XML to express these objects [328, 311]. Although, this work never became a formal standard.

The majority of architectures considered in this chapter claim to be implementing the VC Data Model, however as Young identified in a technical report different implementations express this using different data formats [213]. The report specifically identified three, JSON JWT, JSON Linked Data and format to represent ZKP CL signatures. If an implementation cannot understand and interpret the data format, then they are unable to parse and interpret the data model even if both implementations follow the same specification.

5.2.2 Identifiers

Identifiers are required in any information system that wishes to identify, correlate, remember and respond to entities performing actions within the system [373]. Credential-based identification systems clearly have a strong requirement for identifiers to identify credential issuers and subjects aswell as being a prerequisite for interaction between

two identified entities. For example, when a software artefact interfaces with another in order to issuer or present a credential.

Early cryptographic research into credential mechanisms has long recognised this fact, often referring to identifiers as pseudonyms derived from an asymmetric cryptographic function which enables cryptographic authentication by the controlling party [149, 27, 249]. Another W3C specification in the final stages of standardisation is for DIDs, which add a layer of indirection between the identifier and its cryptographic control mechanisms [35]. These are held within a DID document, which is rooted to a VDR such as a distributed ledger.

The DID specification defines an abstract structure and interface for a DID which is intended to facilitate interoperability between DIDs. However, there are currently over 130 concrete DID methods at various stages of development and with varying degrees of compliance with the specification [212]. Evaluating, selecting or designing a DID method for a specific system is an important choice that will affect that types of credentials that can be issued and the other software systems that it will be possible to interoperate with. Additionally, DID methods root their documents to different VDRs each of which come with their own assumptions and dependencies. The W3C Credentials Community Group (CCG) released a rubric to help implementers with these decisions [374, 375].

5.2.3 Protocols

A protocol defines how two distinct software systems can interface and engage in a meaningful exchange of information. The following identifies a set of protocols that an architecture for credential-based identification systems might support.

5.2.3.1 Transport and Messaging

Before software artefacts can exchange structured messages according to a protocol, they need a protocol to determine how they will communicate. On the web the transport mechanism is typically Hypertext Transfer Protocol (HTTP) with messages communic-

ated through a Representational State Transfer (REST) based Application Programming Interface (API). This can either be a bespoke API, or follow an emerging standard. These emerging standards include CHAPI, the VC API or DIDComm messaging. Unless two implementations use common transport and messaging protocols they will not be able to interface with each other.

5.2.3.2 Credential Issuance

Credential issuance is a negotiation between a Issuer and a Holder that takes place through the exchange of a series of messages in which parties agree on the type and contents of the credential being issued as well as exchanging necessary intermediary cryptographic data such as a nonce. A successful protocol execution results in the Holder receiving and storing a credential payload that has been cryptographically signed by the Issuer's key. There are many different approaches to achieving this, including the Aries Request for Comment (RFC) defining a credential issuance protocol over DIDcomm [376], the issuance server being defined in the VC API and the DIF Wallet and Credential Interactions (WACI).

5.2.3.3 Credential Verification

Credential verification is a negotiation between a Holder and a Verifier in which the Holder learns the requirements for verification that the Verifier is enforcing and determines if they are capable of, and wish to, respond with a presentation that meets these requirements. Attempts are also underway to standardise this interaction, although most implementations today have emerged from necessity of implementers for this functionality.

5.2.3.4 Credential Revocation

Credential revocation enables an Issuer to revoke a credential that they previously issued to a Holder without having to interact with the Holder. Once revoked, a credential can still be presented by its Holder but the Verifier will learn that the issuer has revoked it. Not all implementations support revocation yet.

5.2.4 Key Management

Cryptographic key management is integral to any technical architecture to support credential-based identification systems. Asymmetric key pairs are used to issue credentials, prove control over identifiers and produce VPs. They are also used to establish secure sessions across which authenticated, encrypted communication between entities can occur. Holders, Issuers and Verifiers all need to be able to generate key pairs, store private keys securely and exchange public keys with the relevant parties. DIDs and DID documents held on a VDR are often used to store public key material required throughout the system of identification, in particular credential issuance public keys which are required to verify a signature on a credential payload. Storage and management of private key material and cryptographic secrets is typically custom to the implementation, and an important distinction is whether the key management is provided by a web service or only accessible internally to the system of the actor to whom the keys belong. The Web Key Management System (Web KMS) is a CCG draft report being led by Digital Bazaar that may become a standard in the future [377].

5.2.5 Cryptographic Signature Suites

A signature suite defines the interfaces, data model and data format used when implementing a particular cryptographic signature scheme for the purposes of credential issuance and verification. While the literature contains examples of credential-based systems that do not rely on signature schemes [360], the software architectures considered in this thesis all use digital signatures to issue and ensure the integrity and authenticity of credentials. Implementations must be able to understand and execute a signature suite in order to issue or verify credentials under that scheme. The difference in supported signature suites is a strong limiting factor on interoperability. Suites and signature schemes in use today include: ECDSA on both the Ed25519 and Secp256k1 curve which have associated CCG draft reports [378, 379]; CL-RSA signature scheme preferred by the Hyperledger Aries implementations for its privacy features and currently in the initial stages of becoming standardised [380]; and JSON Web Signatures

(JWS) used by implementations that represent the VC Data Model using JWTs. The BBS+ signature suite is another suite defined in a CCG community report led by Mattr, which attempts to leverage the privacy-preserving properties of a signature scheme with efficient protocols whilst being compatible with JSON Linked Data representations of credentials.

5.2.6 Data Storage

Within credential-based identification systems all actors need access to some form of storage. Credentials, including the data encapsulated within them and the signature that can demonstrate this data's integrity and authenticity need to be stored and accessible to those to whom control and present these credentials within future interactions. Issuers may also wish to store credential schema defining the credentials they are issuing, whilst Verifiers might store policies against which they authenticate Holders and lists of identifiers for Issuers that they trust.

5.2.7 Recovery and Backup

Recovery and backup are important for the resilience of any credential-focused identification system where credentials and key material is held and managed by individuals using their own personal devices. The different mechanisms that are used by implementations include the ability to export encrypted backups and the use of a seed phrase to recover cryptographic material. Due to the immaturity of the space, not all implementations support effective recovery mechanisms and the user experience is often limited.

5.2.8 Schema Management

Schema are used to define the semantics of a credential, specifying a set of attribute names which the Issuer would populate with information when issuing a credential. Schema help Holders and Verifiers interpret the credentials they are issued and subsequently are request requested and presented. A common pattern is define schema

either using JSON Linked Data retrievable through the @context property within a VC enabling them to leverage the extensive existing schema already available from schema.org. Those using CL-RSA signatures within the Aries community, store there schema on a VDR as these are used to compute issuance keys specific to this schema. Another option is to define schema directly within the VC using the credentialSchema property of the VC Data Model specification. However schema are defined, they must be accessible by the relevant actors within the system who can gain assurance that the schema is bound to the VC they are interpreting.

5.2.9 Complexity / Ease of Use

The complexity of software architectures and associated libraries and tooling available for an implementer intending to implement a credential-based digital identification system is an important factor. Some architectures contain full-stack open source libraries designed to be integrated into existing applications with minimal effort. Others provide modular simpler libraries that can be easily customised but require more effort to get to an end to end working system. Others still provide only proprietary solutions built to open standards with the aim of interoperating with other market solutions. All have their own mental model and language that must be internalised before the architecture can be effectively applied in practice.

5.2.10 Adoption

Different technical architectures have differing levels of adoption and maturity. This can be judged from a number of factors including public PoCs and pilots and activity on GitHub. Overall the ecosystem of standards, libraries and codebases that make up the different technical architectures analysed in this chapter are still emerging and the levels of adoption are difficult to judge accurately [381].

5.2.11 Transparency and Governance

This criteria analyses the governance mechanisms associated with a technical architecture. This looks at the organisations and processes around how codebases are manged, including whether they are proprietary or open source. As well as the DID methods used by different architectures and the rules governing the VDR that these DID are rooted to.

5.3 Evaluation of Existing Implementations

The inception of SSI and its associated emerging open technical standards has led to a loose constellation of individuals, organisations, communities working on decentralised technologies for identification, authentication and authorization. The VC Data Model and DID specifications have begun to provide a degree of standardisation between small, atomised building blocks for constructing identification systems. This has led to an array of different software architectures, libraries and APIs that are attempting to align their implementations with these standardised specifications and some other subset of proto specifications being collaborated on in spaces such as the W3C CCG and DIF. An implementer of a credential-based digital identification system must decide, at a point in time, whether existing frameworks can be used to meet their systems needs or whether to develop an independent, custom architecture with the aim of being interoperable with implementations that use other frameworks. A study in 2021 evaluated 19 different software frameworks from this space [381], and there are currently over 130 DID methods registered in the W3C DID spec registries repository [212]. However, despite this apparent surfeit of options for the prospective implementer it is widely acknowledged that many of these systems are still immature and actively under development. In particular, a common protocol for the exchange of credentials between actors has not been standardised beyond individual communities of practice [381].

The decision of which framework to use within this thesis was simplified by two factors. First, the intention to implement a credential-based identification system that

leveraged the cryptographic protocols derived from a signature scheme with efficient protocols. As discussed in Chapter 4, this scheme was designed to have the strong privacy-preserving features first identified by Chaum [27, 28]. Furthermore, this signature scheme is theoretically mature having first been defined in 2001 and has been demonstrated to be concretely realisable in a number of distinct mathematical settings [214, 300]. Second, this thesis is focused on decentralised architectures that support, or at least aspire to support, interoperability and permissionless innovation through open standards including the use of DIDs and VC. The justification for this is that the ability for multiple distinct implementations to interact in a standardised way prevents vendor lock-in and technological centralisation which would otherwise add powerful intermediaries into the digital interactions [176]. These types of digital intermediaries are typically not bound by the roles and responsibilities of the social context whose interactions they mediate [70].

The software development frameworks that were attempting to leverage a signature scheme with efficient protocols for credential issuance and presentation include; HVIEP, IRMA, Mattr and Dock.io. However, when this research was undertaken neither Mattr or Dock.io had a fully integrated this scheme into their frameworks making them unsuitable for the production of a technical PoC. This left Hyperledger Aries and IRMA. Both were reviewed against the technical requirements identified in Section 5.2 along with the Serto / Veramo platform which was included for comparison. The full reviews of these frameworks can be found in Appendix D. Ultimately, the HVIEP platform was selected. The HVIEP offers the richest set of libraries across multiple distinct programming languages, including both open-source and proprietary code bases and was clearly attempting to follow the W3C standards. Whereas IRMA provides a single set of libraries that all implementers had to use and defined custom data models only understood by these libraries. IRMA is governed and developed by a single non-profit organisation, which also acts as the root certificate authority for all IRMA based credential issuers, while the HVIEP code is maintained by the linux foundation and actively developed by numerous distinct organisations. Furthermore, the code is distinct from its instantiation whose governance can be determined separately. For example, the Sovrin

MainNet is an instantiation of a Hyperledger Indy-based distributed ledger governed by the Sovrin Foundation. All three frameworks evaluated have a requirement for public, highly available verifiable data storage; IRMA chose to use a Github repository managed centrally by the Privacy-by-Design foundation, the HVIEP can use any Indy-based distributed ledger whose governance rules can be determined (the Sovrin MainNet is a public permissioned network) and Serto / Veramo use the public permissionless Ethereum blockchain. The flexibility of the HVIEP to use different ledgers, as well as the established governance of the Sovrin MainNet made it an attractive choice.

Numerous other frameworks could have been evaluated if this thesis had chosen not to require the use of a signature scheme with efficient protocols. These include: The framework being worked on by Digital Bazaar who have championed a set of different specifications within the W3C, have their own open-source and proprietary codebases for verifiable credentials and have developed their own DID method and associated VDR - the Veres One blockchain; Mattr's custom platform for VC interactions using DID methods ion,web and key which is provided to implementers as a service; and Microsoft's Verified ID platform. These frameworks and others are all actively under development in this rapidly maturing decentralised identity ecosystem.

5.4 The Hyperledger Verifiable Information Exchange Platform (HVIEP)

The three projects in the HVIEP are:

- **Hyperledger Indy** - Provides tools, infrastructure and libraries for issuing verifiable credentials from DIDs rooted to an Indy ledger. It is made up of two main code repositories:
  - **indy-node** - Code used to run a node and join and maintain consensus within an Indy distributed ledger network. The Sovrin network is an example of this.

  - **indy-sdk** - Containing c-callable libraries for use by both indy-node and application development. Libindy is the most important, providing interfaces to the cryptographic protocols in Ursa and the ability to write to and read from Indy ledgers. The indy-sdk also includes wrappers for libindy in most popular programming languages.

- **Hyperledger Ursa** - A shared cryptographic library designed to provide a common interface to protocols making it easier for the underlying cryptography to be swapped out. This project has similarities to the cryptographic engine in the ABC4Trust project [240]. It includes an implementation of the CL-RSA signature scheme [29] currently signing for Indy-based credentials as well as more performant pairing-based signatures schemes with efficient protocols such as BBS+ [30] and PS signatures [31]. Additionally, it includes support for revocation and other useful cryptographic primitives [302, 243].

- **Hyperledger Aries** - Defines specifications in the form of Aries RFCs for creating a reusable, interoperable tool kit for creating, transmitting and storing verifiable credentials. There are currently implementations underway in numerous different languages, each at various stages of conformance with the RFCs. Implementations include:

  - Aries Cloud Agent Python (ACA-Py)
  - Aries Framework Go
  - Aries framework for .NET
  - Aries Framework JavaScript

It is worth pointing out that Hyperledger Indy was created first, originally by the for-profit company Evernym and then donated to the Hyperledger Foundation in 2016; at this time it contained early implementations of both Ursa and Aries functionality. Over time this code was refactored into distinct projects, first Ursa in 2018 then Aries in 2019. This process is still ongoing, with most Aries implementations still dependent on libindy to interface with the cryptography and the ledger (See Figure 5.1).

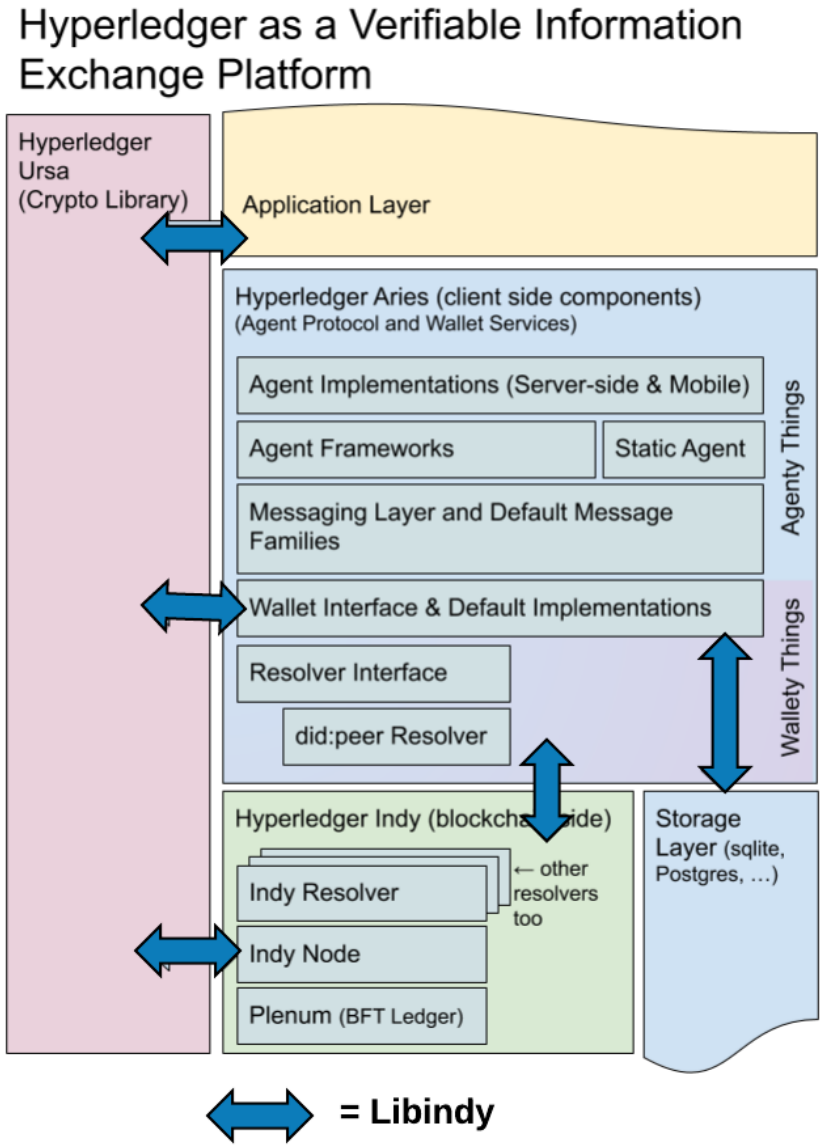

Figure 5.1: Hyperledger Verifiable Information Exchange Platform (adapted from hyperledger.org)

5.4.1 Indy-based Distributed Ledgers

Indy-node is an open-source codebase to enable a distributed set of entities to maintain consensus on the state of a distributed ledger by each running a node instance. The ledger has been designed specifically to provide a public key infrastructure for privacy-enhancing attribute-based credential systems. Currently it supports the creation and maintenance of public identifiers, DIDs, able to issue and revoke cryptographic creden-

tials using a CL-RSA signature scheme first published by [28, 29]. This data is written to the ledger in a number of different transaction types. These are:

- **NYM** - These transactions write a new DID and related DID Document containing an associated public key to the ledger.
- **ATTRIB** - Transactions that update existing DID Documents on the ledger, such as rotating public keys or changing service endpoints. These must be authored and signed by the DID that identifies the DID Document being updated.
- **SCHEMA** - These transactions define a schema name, version and list of attribute names for a specific credential. The schema name must be unique on the ledger, but can be altered by writing a schema with the same name and different version number. Versioning a schema must be done by the original author of the schema transaction.
- **CLAIM_DEF** - Often referred to as a credential definition, these transactions write the public key from a generated key pair of an CL-RSA signature for a specific credential schema [29]. Only DIDs with CLAIM_DEF transactions for specific schema included in the ledger can issue credentials of this schema that are publicly verifiable. Many DIDs can author CLAIM_DEF transactions referencing the same schema.
- **REVOC_REG_DEF** - Transactions that define a revocation registry for a certain credential definition transaction (CL-RSA public key) meaning that credentials signed by this public key can be revoked. Currently, these registries use cryptographic accumulators defined in a 2009 paper by Camenisch et al [302].
- **REVOC_REG_ENTRY** - Whenever an issuer issues or revokes a credential they must author a transaction that updates the revocation registry keeping them up-to-date so they can be used to construct and verify proofs of non-revocation.

All transactions include the time they were authored and a unique identifier that can be used to reference and resolve the data contained within the transaction from the

ledger. Transactions must be signed by the public key associated with a DID already stored on the ledger before they are accepted by the nodes maintaining the ledger state. This leads to a hierarchical structure whereby all DIDs must first be authored to the ledger in a NYM transaction signed by the key of another DID before they can themselves write transactions to the ledger. Any Indy-based ledger is instantiated with a number of genesis NYM transactions and all other NYM transactions can be traced to a NYM transaction signed by one of these DIDs. This structure of signed transactions allows any entity to verify the validity of the ledger state by starting from these genesis transactions. It also ensures rules around which DID has the authority to update a DID Document, schema, or revocation registry can be cryptographically enforced by node operators [42].

The indy-node codebase allows a collection of entities to come together to instantiate and manage an Indy network. There are a number of networks in existence today, most established are the ones managed by the Sovrin Foundation. This thesis focuses on two, the Sovrin MainNet and Sovrin StagingNet, both of which have been around since 2017. The major difference between the Sovrin MainNet and other Indy networks is that it is a public-permissioned network governed by the Sovrin Governance Framework that defines the roles and responsibilities of different actors within the network [382]. A permissioned network adds additional constraints around who can write to the ledger. In the Sovrin MainNet only DIDs with the role of transaction endorser are able to write NYM transactions to the ledger and all subsequent transactions these DIDs author must be additionally signed by a transaction endorser [383].

The nodes within the Sovrin MainNet are run by Sovrin Stewards, organisations that volunteer time and resources to maintain the network. The network is administered and managed by the Sovrin Foundation which also acts as a Governance Authority [382]. Stewards are selected by the Governance Authority to ensure maximal distribution across hardware, business domain and geographic location, limiting the threat vector of malicious takeover and promoting resilience. All stewards agree to the requirements specified by the Governance Framework and sign the Sovrin Stewards Agreement [384, 382]. Nodes accepted into the network then engage in a consensus protocol named

plenum based on redundant Byzantine fault tolerance [385]. As such, assuming the Sovrin Foundation and a subset of the stewards are trustworthy, then the transactions stored within the ledger can be trusted with a high degree of confidence.

5.4.2 Aries Agent Architecture

For the implementations considered in this thesis we focus primarily on the ACA-Py, a codebase developed by the Government of British Columbia and is actively maintained and used in a number of production systems today [386]. Despite this, much of what is discussed will be applicable to any implementation of the Hyperledger Aries specifications. Before going into the use-case specific implementation details, a mental model for applications built using this tooling is introduced.

5.4.2.1 Aries Agents Facilitate Peer to Peer Interaction

Hyperledger Aries implementations are designed to facilitate interaction between other Aries agent instances. An important requirement for any credential-focused identification system architecture is to facilitate the exchange of messages between actors as they engage in specific protocols, for example, the interactive credential issuance protocol (Section 4.2.3) [328]. Aries agents implement a protocol called DIDComm for this purpose [216]. First agents must exchange identifiers and associated public keys and communication endpoints through some out of band process. These are DIDs, following the did:peer specification [387]. Then using this information, an authenticated encryption communication channel is established. All future messages are sent across this channel.

An overview of the DIDComm protocol can be seen in Figure 5.2, in which Alice and Bob want to communicate securely and privately. Alice signs and encrypts a message for Bob. Alice's then sends the encrypted, signed message to Bob's endpoint. Bob first decrypts the message using his secret key, then verifies the message's integrity by resolving Alice's DID and checking the signature on the message. All the associated information required for this interaction are defined in each person's respective DID

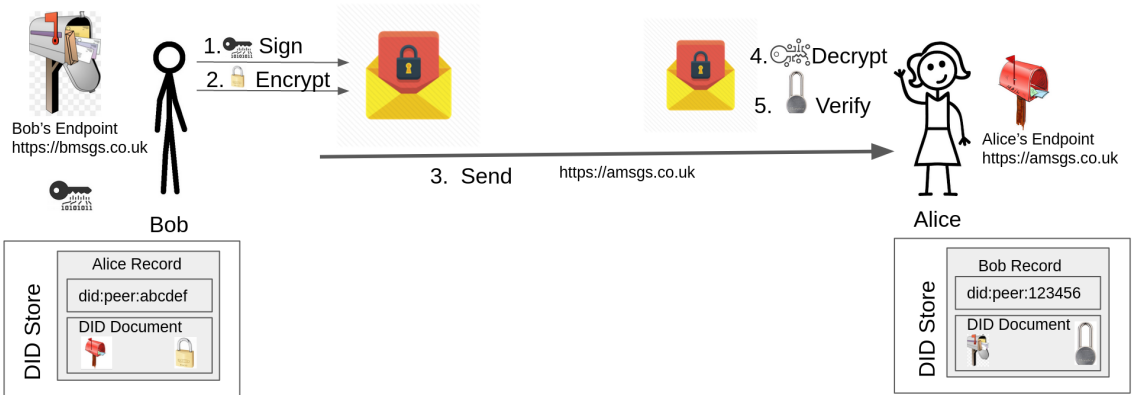

Figure 5.2: DID Communication Between Alice and Bob

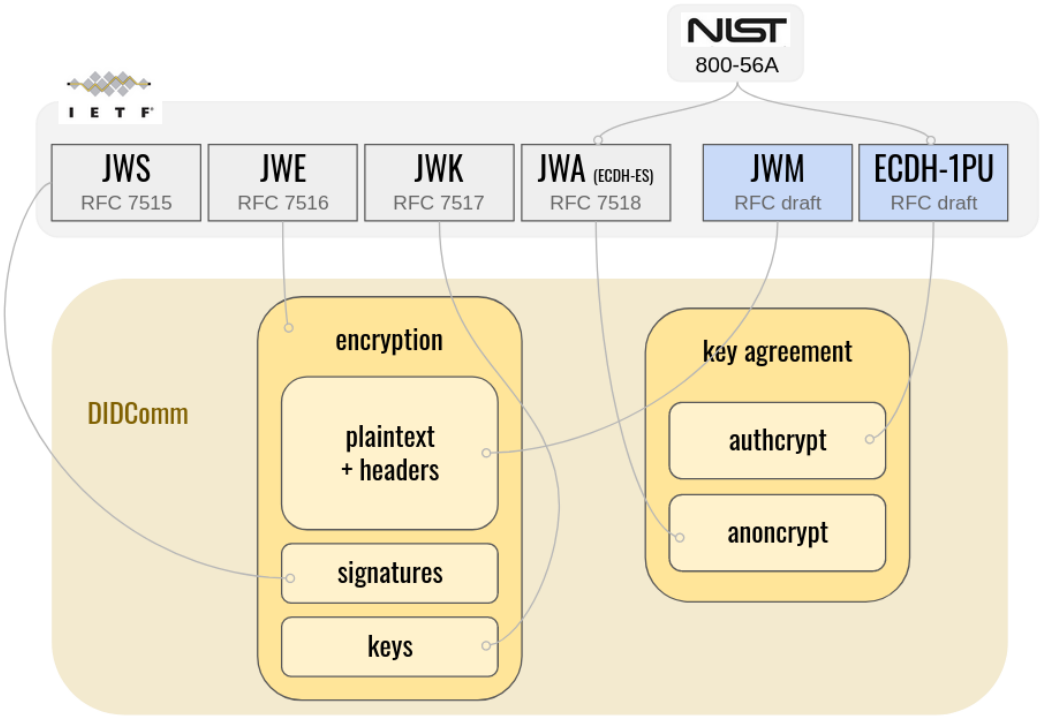

Figure 5.3: DIDComm dependencies [388]

document. Additionally, a recent presentation from Hardman at the Internet Identity Workshop emphasised that DIDComm v2, currently being standardised at the Decentralised Identity Foundation, is composed of existing, established standards (see Figure 5.3 [388]).

5.4.2.2 Aries Agents Understand A Specific Set of Protocols

Aries RFCs define the specifications for protocols. Protocols define the semantics and sequence of messages that should be exchanged to between agents to progress the protocol state. They can be defined for invoking specific cryptographic protocols

such as credential issuance [376], but they are not limited to these. Any interaction that can be modelled by a sequence of messages can be represented as a DIDComm protocol and implemented within an Aries agent framework. Both agents interacting must implement the protocol in line with the specification in order to understand the messages being exchanged and successfully engage in the protocol.

5.4.2.3 Aries Agents are Event-Driven

Aries agents know how to construct, send, receive, verify and understand messages based on the protocols that they implement. They do not know when to send these messages, who to send them to or how to respond to messages they receive. This is handled by the business logic of an application, allowing customisation to specific use-cases without changing the underlying Aries implementation.

ACA-Py is an open-source Aries implementation being developed primarily by the Government of British Columbia[1]. ACA-Py agent instances receive instructions over an administrative Swagger-API at a specified endpoint and forward messages posting them to a defined webhook server endpoint (See Figure 5.4). The application business logic defines what instructions to send and what conditions or actions in the application cause them to be sent. For example, a human actor clicking the issue button in an applications user interface. Furthermore, the business logic specifies how the application will respond to certain messages they are forwarded from their agent. This could include automatically sending further instructions to the agent, saving information in a database or requesting human input.

5.4.2.4 Aries Agents Interface with Ursa and Indy

The open-source implementation of the Aries RFCs[2], ACA-Py in this thesis, handles a lot of the associated complexity when applying this technology. It interacts with a distributed ledger, encrypts and decrypts messages and invokes complex cryptographic protocols. As shown in Figure 5.4, an application developer creating business logic to

[1] `https://github.com/hyperledger/aries-cloudagent-python`
[2] `https://github.com/hyperledger/aries-rfcs`

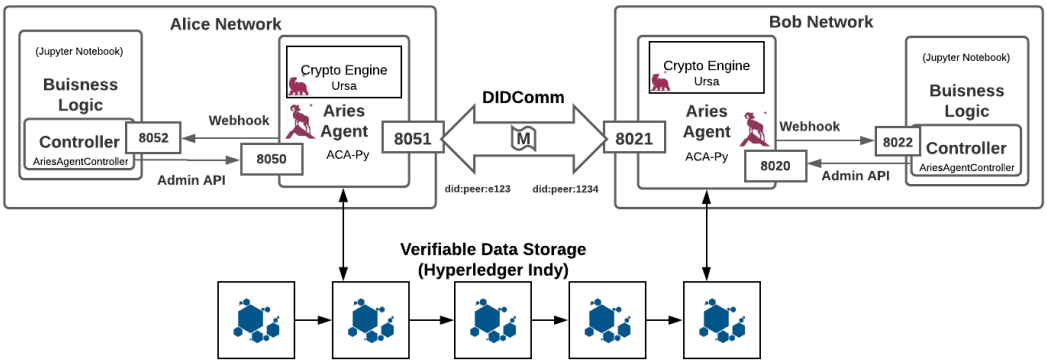

Figure 5.4: Application Architecture Hyperledger Aries Based DIDComm Interaction

apply the HVIEP to their use-case need only send HTTP requests to defined endpoints exposed internally to initiate certain interactions with other agents. This does require that the user of the application places a degree of trust in the underlying codebases they are dependent on. That ACA-Py is open-source, well maintained by a public entity, used in production applications and part of the Hyperledger Foundation (a subsidiary of the Linux Foundation) provides some evidence of the trustworthiness of the code.

The typical flow between two interacting entities Alice and Bob is as follows. Alice, through some user interface and application-specific business logic, sends a request to her ACA-Py instance. The request identifies a protocol, message and external connection to send this message to. The ACA-Py instance constructs and encrypts and signs the message before sending it over the identified DIDComm channel. Bob's ACA-Py instance receives the message, verifies and decrypts it using the relevant connection keys and then forwards a notification to Bob's business logic over an internal webhook endpoint. Bob's business logic determines how to handle this, maybe it pushes a notification to Bob's user interface to request human input. Bob might decide to send a message back to Alice through a similar process. Each message forms a sequence of a particular protocol advancing the state closer to successful completion.

5.5 Conclusion and Critical Discussion

The decision about the software development framework, tooling and libraries to use when implementing any software application is a important design choice. It both

constrains and expands the available possibility space within which the application can be realised. In credential-based identification systems that facilitate the exchange of verifiable information between distinct, decentralised entities each running their own application of separate technical artefacts the choice of technologies and the standards and specifications that they implement become even more important. Distinct software systems must be able to connect with, identify and exchange information across mutually understood protocols. The information exchanged must be represented in a mutually understood data format, structure and integrity assured under a common signature scheme. If each entity were to run exactly the same software, this is trivial to achieve, however decentralisation requires choice over the technologies any entity chooses to use. Otherwise these systems will recentralise around the technology provider that manages to capture the largest network of entities. Open standards play an important role in preventing this kind of vendor lock-in that is prevalent in earlier identification systems such as Facebook Connect [198]. However, the effort to standardise an framework for credential-based identification systems will take time with few formal standards currently fully developed and many disparate companies and organisations implementing these standards in their own way. Before practical interoperability can truly be achieved many additional standards need to be refined and adopted. Whether this will happen and how long it will take remains to be seen.

This chapter has identified a set of requirements that all software frameworks for credential-based identification systems need to meet through their implementation. Based on these requirements and the desire to use a signature scheme with efficient protocols the ACA-Py framework from the HVIEP was ultimately selected to produce a practical PoC in the following chapter. It is important to acknowledge that selecting the specific signature scheme constrained the available frameworks for the PoC, it is also a limiting factor in terms of the implementations with the practical interoperability can be achieved. At this point, ACA-Py agents can only interoperate with other libraries that are aligned under the HVIEP and associated RFCs of which there are numerous. Furthermore, at the time of this analysis ACA-Py was one of the few libraries available that was open-source and fully featured enabling a comprehensive PoC to be produced

including credential revocation. This allowed the focus to shift from how to implement all the features of a credential-based identification framework towards how to apply that framework to a specific use case where these technologies could add value.

Sixth Chapter

An Identification System for Professionals Within the Scottish Healthcare Ecosystem

---

> *"There is an irreducible minimum level of bureaucracy in any complex system. Rules and processes are essential means to manage risk and, in the context of health and care, keep people safe and help ensure a consistent level of quality care and outcomes across the country. But excess bureaucracy reduces the time that staff have for care and hinders staff and leaders from deciding how to manage risk, being creative, innovating to fix problems, empowering others and being flexible. This negatively impacts staff well-being, morale and retention while hindering the very outcomes the processes aim to support."*
>
> UK Department of Healthcare and Social Care Report, Busting bureaucracy [24]

Identification systems involve a complex, and highly contextual, set of actors, processes, information flows, governance procedures and justifications for identification. This thesis has selected the identification of healthcare professionals within the Scottish Healthcare Ecosystem (SHE) as a case study in order to provide the realistic constraints of a real-world identification system for this research. This chapter first introduces the healthcare context and reflects on its requirement for high assurance identification processes as well as some of the challenges identified with existing identification systems that lead to increased administrative burdens. Then focus is turned towards the SHE, a specific instance of a healthcare context with a well-defined set of institutions, ID artefacts and identification processes. The current system of identification within the

SHE has been understood and modelled from the perspective of a healthcare professional as they progress through their career. These models were produced using input from an interactive workshop with a number of key stakeholders at the Royal College of Physicians of Edinburgh (RCPE) [41].

This chapter then describes a PoC implementation that demonstrates how these existing processes could be augmented using technological artefacts that combine the open, interoperable standards work of the decentralised identity community with strong privacy and security guarantees found in cryptographic credential mechanisms. The PoC is built using the open-source HVIEP and has been realised in a set of interactive Jupyter notebooks. These notebooks are open-source and available on Github[1] such that anyone should be able to walk through these technologically augmented identification processes. Furthermore, the underlying Jupyter environment and its potential for describing digital identification processes within Jupyter notebooks is presented as a mechanism to disseminate this work to key stakeholders [43]. Jupyter notebooks expose the right level of complexity of these processes and can easily be customised for specific audiences, either technical or domain specific, making them ideal for facilitating meaningful discussions about a proposed augmentation to existing identification processes.

6.1 The Healthcare Context

The medical profession is one of the oldest in human history. The Hippocratic Oath dates back to Ancient Greece, and is still taken in some form by the majority of medical professionals, is a commitment to uphold a set of ethical standards while practicing medicine [389]. Furthermore, evidence suggests some form of medical licence has been required since at least the middle ages [390]. These licences, issued by authorities of their time and locality, attested to the successful completion of medical training and the eligibility of an individual to practice medicine. This enables a degree of control over the conditions under which someone could perform in the role of medical doctor.

[1] `https://github.com/blockpass-identity-lab/scottish-healthcare-ecosystem-poc`

The licence can became an ID document that could be used in formal identification processes through which the legitimacy of a performance could be verified.

Using the lens and language of identity theory presented in Section 2.1, we see that the role identity of a healthcare professional is constructed through the Hippocratic Oath, among other things This sets expectations and motivates the behaviour and actions of those that hold this role during an interaction [389, 4]. Those interacting with an actor performing the role of doctor also apply their understanding of this role identity to the interactive context. This influences the expectations they have of the actor as well as their own actions, such as the willingness to truthfully disclose sensitive medical information about oneself. Trust is clearly fundamental to these interactions [94]. The patient must place trust in persons within the medical profession to fulfill a reasonable set of expectations, and those being trusted must perform their responsibilities trustworthily. Hawley warns us that misplaced trust in an unreasonable set of expectations is risky for both parties [94]. Through this lens, the purpose of the licence can be viewed as a mechanism to assess the trustworthiness of an actor, providing assurance that they are competent to act in the role they are performing. It creates an avenue for trust to be established outwith a dependency on personal interaction. Trust is placed in this system of identification, which includes the authority issuing this licence and the processes and constraints under which they are issued [2].

Throughout the 20th Century and into the 21st, societies and their healthcare systems throughout the world have complexified. This has included hospitals, surgeries, integrated care networks, the public and private sector, licensing bodies, specialist medical knowledge and their associated education pathways, legislation, government regulation, globalisation and, of course, advanced ICTs. Luhmann stated over 50 years ago, that social complexity requires increasing levels of trust, as individuals become increasingly interdependent on the actions of impersonal others while going about their lives [2, 83]. This creates risk that needs to be managed at a systems-level by institutional entities. A UK Government report echoed this sentiment stating that *there is an irreducible minimum level of bureaucracy in any complex system*, these processes manage risk, provide assurance and keep people safe [24].

Within healthcare, the primary mechanism to mitigate risk has been providing assurance that a medical professional has the education and skills required to perform in the role associated with the position they are employed to fulfill. This is achieved through an evidence mosaic that the doctor must collate throughout their careers. Healthcare providers and licensing bodies are then required to verify this evidence alongside strong identity verification checks to ensure that their employees are who they claim to be and that they have the required skills and training for the role [391, 392, 393].

Over the years, healthcare service providers have increased the minimum standard for identity verification and pre-employment checks in line with new regulations [394, 391]. As a result, the time spent on these processes has increased. A British House of Lord's report, for example, estimated that 25,000 junior doctor days a year were currently being spent on these administrative tasks [395]. Another major challenge is presented by doctors working across both public health care and private practice [396].

The administrative burden that these credential checks entail takes time away from treating patients, which has been shown to be a key factor in burnout and low job satisfaction [397, 398]. Furthermore, not every place of employment is capable of completing identification processes to the same standard. In healthcare organisations with more funding and a higher turnover of staff, checks are likely to be undertaken by trained *identity managers*, often with the use of digital tools for document verification such as a passport scanner [399]. Whereas, in smaller healthcare facilities, this role can be attached to the job role of a member of staff such as receptionist or to IT services.

Unfortunately, there are numerous examples throughout the world of under-qualified doctors practising without licences putting patient's lives at risk and reducing the trust in the profession as a whole. Examples include; the UK General Medical Council (GMC) having to recheck credentials of 3,000 doctors after a fraudulent psychiatrist was found to have practiced for 23 years without proper credentials [400]. Also we have the case of a social worker in Canada involved in more that 100 child protection cases [401] and the notorious US case of Christopher Duntsch - also known as Dr Death [402]. Along with this, there are also cases which highlight the burden these identification processes

place on the institutions and organisations enforcing them. For example, the case of the Royal College of Psychiatrists who failed to carry out over 350 background checks on their staff [403].

These assurance processes are clearly necessary within the healthcare profession. The demand and opportunity for streamlining these processes is widely acknowledged by healthcare professionals and researchers alike [41, 24, 392]. The US Federation of State Medical Boards [392] analysed the use of digital credentials in healthcare, looking at the potential for both current and future technologies to streamline the process and enhance trust in the system. It highlights the movement to *disintermediate the creation and management of credentials* as a key time reduction strategy. A report from the UK Government commissioned in response to the COVID-19 pandemic presented digital systems as an approach to reduce *resource-intensive, inefficient and time-consuming* processes [24]. It also emphasised that digital systems can often add to the administrative burden, something frequently reported in the literature [404, 405]. The report recommends the design of these systems be driven by healthcare professionals and *any changes must lead to greater trust being placed in frontline staff, delivering better outcomes and experiences for people that use health and care services* [24].

6.1.1 The Scottish Healthcare Ecosystem

The SHE has been selected to provide a concrete, geographically-bound instantiation of a healthcare context. It provides healthcare services for a population of around 5.5 million people and approximately 160,000 healthcare professionals [406]. It is predominantly part of the public sector, with infrastructure and coordination of services managed through a hierarchical organisation structure that is overseen by the National Health Service (NHS) Scotland. NHS Scotland is itself one of four governing bodies that are part of the UK NHS. Despite this, NHS Scotland operates largely independently from the wider UK context with health and social care decisions devolved to the Scottish Government.

The SHE provides an ideal case study for this thesis. As Coiera points out, large

top-down approaches to implementing healthcare information technologies usually fail due to their scale. Instead, he recommends a *middle-out approach*, with Government focusing on standards for interoperability [407]. NHS Scotland has a well defined organisational structure, responsible for provisioning a full range healthcare services for an entire population. This suggests that convincing stakeholders throughout the healthcare ecosystem, something identified as crucial for the success of a digital credentialing solution for staff movement [41], is a real possibility. Perhaps NHS Scotland could utilise the *middle-out approach* for instigating wider changes in the UK NHS [407].

6.2 Engaging With Healthcare Professionals through an Interactive Workshop

In order to gain a detailed understanding of the SHE and ensure the digital credentialing system proposed by this thesis was driven by the professionals whom it is intended to serve, a workshop was hosted at the RCPE [41]. The aim of this workshop was to validate, refine and prioritise the set design principles for identification systems, initially distilled from the technical literature presented in Section 5.1 as defined in Table 5.2. Additionally, the workshop aimed to identify key identification processes, institutions, organisations and ID artefacts that make up the system of identification for healthcare professionals within the SHE. This included an attempt to quantify the time spent on these processes and rank the importance of the ID artefacts.

The workshop involved 14 participants with a wide ranging experience of different aspects of the healthcare system. The participants for the workshop were selected through consultation with the RCPE, and who were able to use their contacts to invite a diverse range of attendees. This included clinicians, RCPE trainees, and RCPE staff involved in data management, digital transformation and education, as well as a representative from the GMC. While no personal data was captured during the workshop and all attendees remain anonymous, explicit consent was obtained and the research aims were explained at the beginning of the workshop.

6.2.1 Workshop Methodology

After a brief introduction, participants were asked to complete a warm up exercise where they recorded different identification processes that occur throughout a typical day in their life. These included using a Radio Frequency Identification (RFID) card, logging into a digital system with a username and password, authenticating to a mobile device, bank card payments. Anything that involved some form of identification and authentication. Participants were also asked to record times when authentication failed, for example through a rushed or forgotten password attempt.

The aim of this initial exercise was to get attendees thinking about how often they interact with digital systems, how many different username and passwords they currently manage and the number of different authentication devices that they have to carry. The majority of participants recorded over 25 separate identification interactions, all in a single day.

6.2.1.1 Healthcare Ecosystem Process Mapping

The next stage of the workshop focused on eight core identification processes that captured at a high level the typical experiences of a healthcare professional throughout their career. Participants were asked a create process maps identifying key organisations involved in each of these stages and the identity-related information and other evidence that a healthcare professional would be required to present to them. Additionally, participants were asked to capture frustrations that a healthcare professional might experience while navigating these interactions. The general identification processes identified and validated prior to the workshop to provide some structure were as follows:

- Doctor graduating from university.
- Doctor applying for a job.
- Doctor joining a hospital.
- Doctor training.

- Doctor rotation.
- Doctor begins RCPE accreditation.
- Doctor qualifies as a Physician.
- Doctor moves abroad.

The workshop participants were split into groups, each group focused on four of the identification interactions. The results were then presented back to the group providing a detailed overview of each of the stages in a doctor's career, including recurring and trusted ecosystem entities such as the GMC. These maps were combined and synthesised into a Gantt chart, showing the time burden and repetition associated with a healthcare professionals as they progress through their career, see Figure 6.1. This was developed through follow up communication with a final year trainee at the RCPE who attended the workshop.

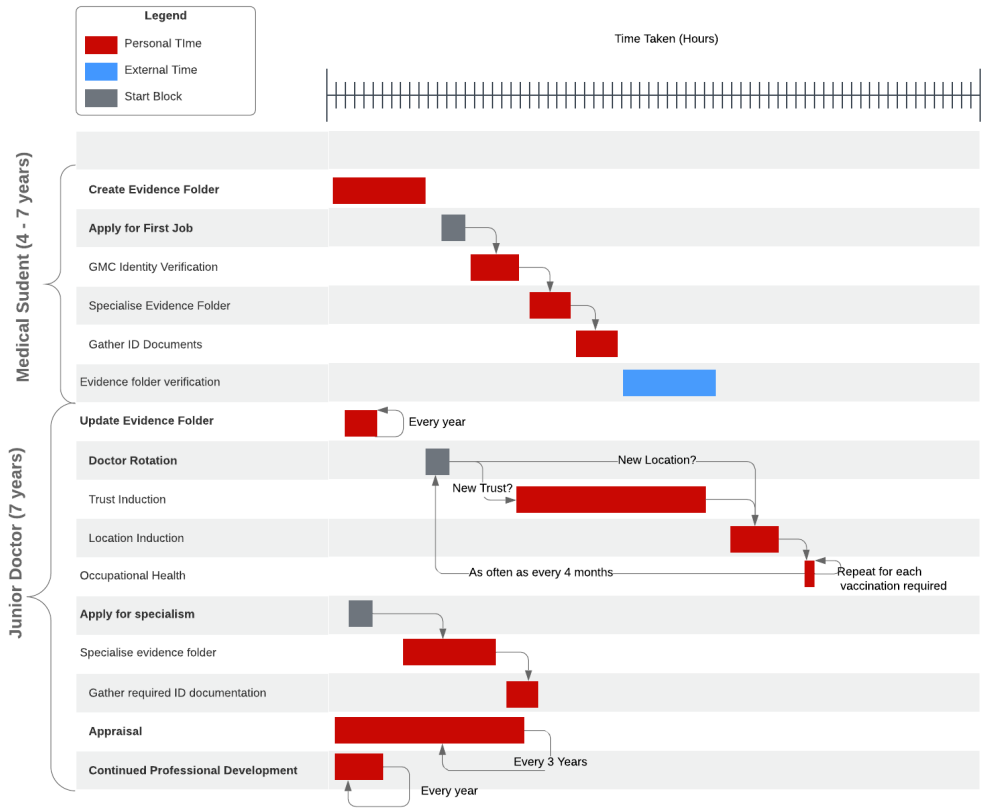

Figure 6.1: Time Estimates for a Healthcare Professional's Identification Interactions

6.2.1.2 *Re-imagining Identification Processes using Decentralised Identification Technologies*

The next stage of the workshop involved an interactive session on the technical architecture identified in Chapter 5 and the capabilities it could provide healthcare professionals when applied to the key identification interactions that participants previously mapped out. The goal was to give attendees a high level understanding of how this technology works and where it might fit into existing processes within healthcare.

Physical props were used to represent different aspects of the system and a number of workshop attendees were asked to play roles within the healthcare ecosystem. Specifically, six participants acted as the key organisations and institutions identified in the process mapping stages:

- **A Medical School** - Before becoming a licensed doctor, individuals must first complete a degree at a medical school.
- **The GMC** - The doctor licensing body in the UK
- **The RCPE** - Royal colleges are involved with training and examination procedures for junior doctors as they gradually specialise in a medical discipline.
- **Edinburgh Hospital** - This hospital was used as the initial place of employment once the fictional doctor in our scenario graduated.
- **Glasgow Hospital** - This entity represented the doctor rotation process within the scenario modelled.
- **National Health Service Education for Scotland (NES)** - A body involved with continuous training and education of doctors.

The initial setup of the SHE was represented by asking each actor in the scenario to *generate a pubic/private key pair*. A red (private) and green (public) card was used to show the two halves of a public/private key pair. Actors then attached their *public key* to a white card, which was used to represent a DID. All actors were asked to place

their DID onto the wall, representing the act of registering a public DID on a distributed ledger such that the public keys for these trusted entities could be resolved by anyone.

After the initial setup, a member of the research team played the role of doctor and walked through each of the identification interactions discussed and mapped in the process mapping session. This interactive approach was used for a couple of purposes:

- To educate workshop attendees about the capabilities that the identified technical architecture enables.
- To illustrate how this technical architecture might be applied to the SHE to streamline identity interactions.

Throughout this interactive session, a flipchart was used to represent the doctors digital wallet. The wallet gradually collected VCs represented as large sticky notes and digital relationships, formed through establishing peer DID connections, were represented as small blue post-its within the wallet. Figure 6.2 is a digital representation of the resulting flipchart after the session.

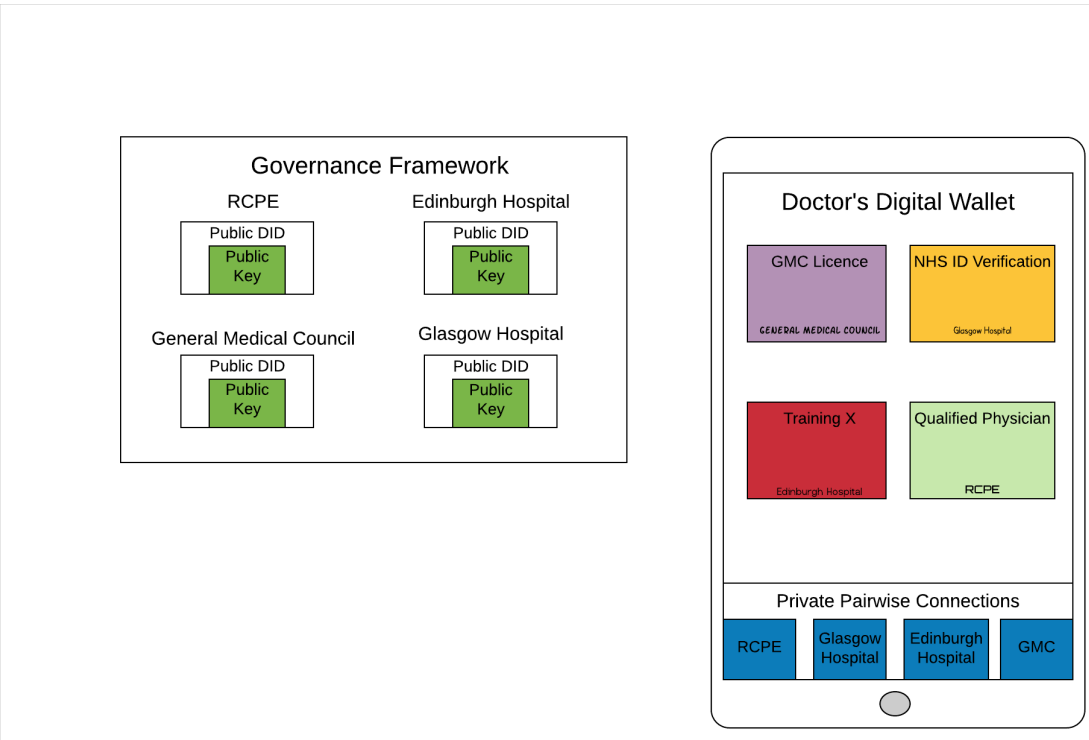

Figure 6.2: RCPE Workshop Representation of Healthcare Professional's Digital Wallet after completing a number of career interactions [41]

6.2.1.3 Evaluating the Design Principles

The last session of the workshop, involved a discussion the design principles that they believed to be important to realise a trustworthy instantiation of this technical architecture within healthcare.

Before introducing the design principles from the literature to our audience, we asked them what they thought was the most important feature of future technology that would help them trust it. The list included the following:

- Use all over the ecosystem, all entities need to participate.
- Attention for end users, usability, convenience, workable.
- Buy-in from government and NHS.
- Future proof.
- Resilient, reliable, fraud resistant, protection, security.
- Control.
- Transparent data sharing, clear, clarity.

Comparing this list with the list of design principles that we distilled from the literature demonstrates that our audience adds two specific principles to the generic list. The first is that they find it important to know that all entities will be involved in the ecosystem, including the government and the NHS. The success of the system depends on buy-in of all of the involved entities, and that is something that should be developed from the start. Second is the attention for usability and convenience. User engagement is important from the start and throughout the development process.

Then the participants were presented with eight design principles selected from the literature and their definitions were explained. Then we asked them to rank the principles in order of importance. The majority selected protection as the most important, followed by control & consent and interoperability. Then we asked to rank importance for each individual principle. Again, the highest scores were for: protection; control and

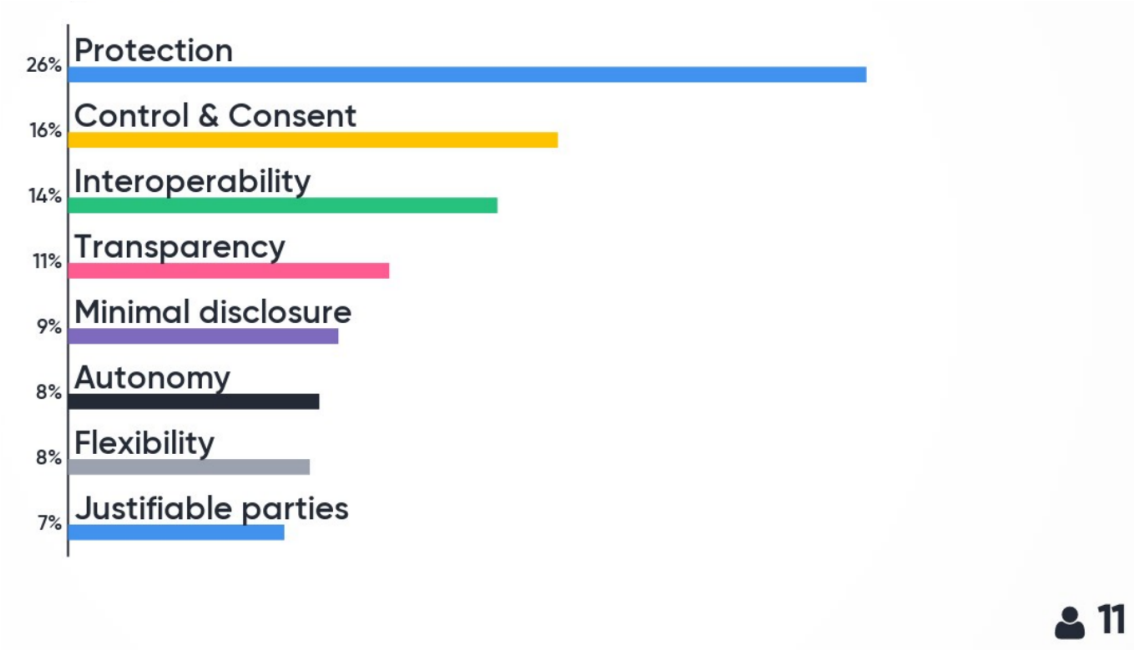

Figure 6.3: Principles ranked by importance

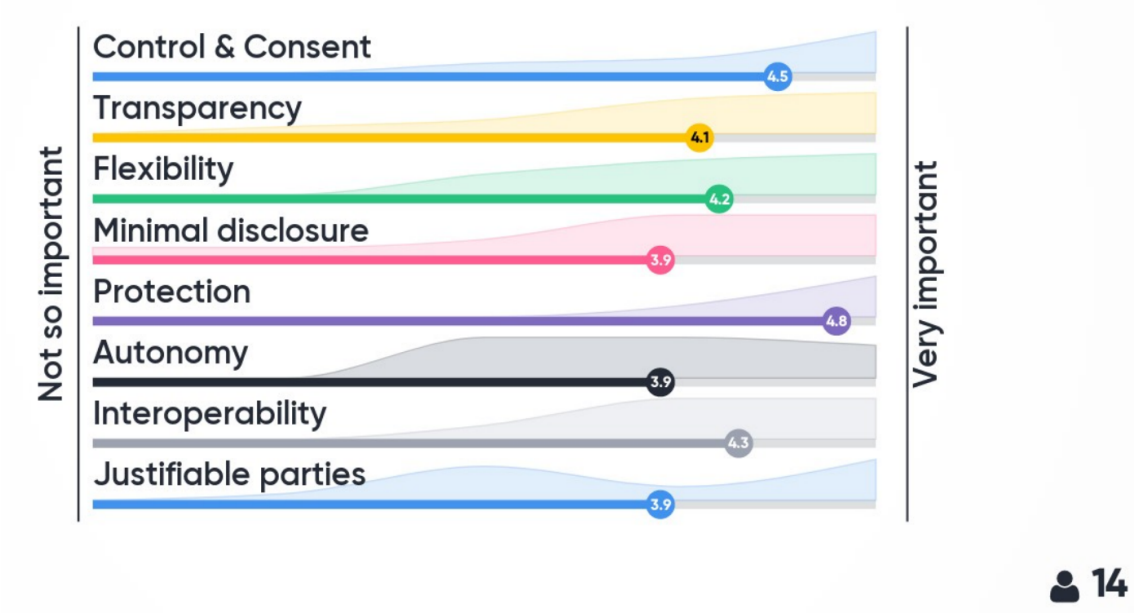

Figure 6.4: Principles individually rated

consent; and interoperability. Figures 6.3 and 6.4 show the results of the Mentimeter polls.

6.2.2 Results

The workshop led to a number of important findings about the system of identification for healthcare professionals within the SHE. The existing system was found to be

complex, fragmented; and time consuming in its current form. Through the process mapping session an increased understanding of the actors, roles, institutions, organisations and ID artefacts was developed. This included identifying an additional important identification interaction, appraisal and re-validation, that had not been considered initially. This is a process whereby healthcare professionals must prove to relevant institutions that they have gone through the required training to keep their skills up to date such that they can continue practice. This is a repetitive process that occurs every three years. It was interesting to find out that even the top level professionals in attendance typically spent a couple of days every three years getting their documentation in order for this procedure.

Another point of clarification was that while we positioned the eight identification interactions as chronological, a lot of them happen in parallel. For example, a medical student typically completes identity verification procedures with the GMC prior to graduating, they also spend their final year applying for jobs so that on graduation they are ready to begin their career immediately. It was pointed out that there is no strict temporal relationship between the eight stages initially identified and a large part of the frustrations arise from the fact that the majority of these stages are repeated over the course of a doctors career. Each time requiring the same time consuming procedure.

A good example of this comes from what is broadly termed doctor rotation, something commonly experienced within the healthcare system. Especially for junior doctors who can rotate to a new location and role as often as every four months. While this is reduces significantly as doctors progress in their career, it still a common occurrence. Every time a doctor moves to a new hospital there are a number of tasks that the doctor needs to complete in order to on-board into the institution:

- They must complete a full identity verification check to the NHS standards [391].
- They must demonstrate they are licensed to practice in the UK by the GMC.
- They must provide evidence supporting all the claims they made in their application. This is typically verified by a consultant and their were estimates it takes up to a full day of their time.

- They must complete an induction session, taking between one and two days. This induction is required for both new locations (e.g. hospitals) and new trusts and generally includes repeated content due to lack of standardisation across locations and trusts.

- They must organise an appointment with occupational health, to prove they have up to date vaccinations. If they don't or are unable to prove they do then doctors must have their vaccinations refreshed. Often leading to needless re-vaccinations.

In summary, the process mapping and ensuing discussions highlighted numerous frustrations experienced by healthcare professionals just to meet the requirements for managing their career. A phrase that came up was the need to professionalise the digital experience throughout a doctors career. Furthermore, the GMC licence was identified as the foundational ID artefact used throughout a healthcare professionals career.

The workshop additionally surfaced a number of challenges that attendees thought would need be to overcome in order to roll this out within a healthcare system. Many participants were senior professionals within healthcare, so had experiences of other attempts to digitise aspects of healthcare. The three core challenges identified were:

- **Funding and business model**: Who is paying for these tools and what is in it for them. There is no clear path to monetisation of the system, however for this to work it needs to be well funded in order to produce something that can scale. It was suggested that part of doctors annual fees and membership to organisations like the GMC and the Royal Colleges could be allocated towards a system like this.

- **Adoption**: A decentralised identification system works only when it achieves large scale adoption. In order for this to happen it needs to be led on a national level and show the benefit to the entire ecosystem. It was also pointed out that while benefit to doctors is relatively easy to show it needs to be able to show benefit to the individual trusted entities. They need to be the ones advocating for this system not the doctors. All stakeholders must be clear on why this shift is happening and what the benefit is too them.

- **Overreaching**: This technology is new and relatively untested at scale. An interesting point was made that it is important to take small steps to prove its value and build human trust before expanding the scope. Attempting anything too large too quickly and failing could be disastrous for the trust placed in the system and underlying technology.

Through this engagement, key institutions and interactions requiring the exchange of evidence were identified and described. Each takes place at different stages of the Physician's career, with changing frequencies and associated administrative costs.

Findings indicated that throughout their career, a healthcare professional collects increasing amounts of evidence from a disparate set of authoritative sources that attest to their education, eligibility to practice, their professional skills and their career performance. However, this evidence is itself an administrative burden to all involved. Healthcare professionals must collect, manage and sort this paper-based evidence. It must be then presented in a format determined by the verifying entity, which might mean scanning the paper document and uploading it into some digital system, or it might not. This is especially prevalent in the case of yearly appraisals and the regular licence re-validation procedures. A result corroborated in a UK Government report into busting bureaucracy, which quoted a comment from one consultant claiming gathering evidence for appraisals can take at least a day of their time [24]. Furthermore, the discussions with an RCPE trainee indicated that often evidence can be as low assurance as a printed-out email confirming registration to a certain training event. Verifying the legitimacy of this evidence, which in appraisal and re-validation meetings is the responsibility of another medical professional, is time-consuming and the evidence itself is of a poor quality [41].

In addition to career-based evidence, the workshop at the RCPE highlighted a key cause of administrative burden they experienced arose from rotating through numerous employing organisations [41]. Each time they interact with a new organisation they are required to navigate repetitive, time-consuming identification processes and employee onboarding procedures. These include identity verification checks to the required level

of assurance as defined by the NHS [391], checking the healthcare professional's licence to practice, a right to work check and a Disclosure and Barring Service (DBS) check. As well as proving immunology status and demonstrating the successful completion of an NHS compulsory basic training course. If either of these cannot be satisfactorily proven then often the healthcare professional will be required to retake them, leading frustrating and time consuming repetition. Another point echoed in the busting bureaucracy report [24].

These identification processes are key mechanisms by which the health service mitigates organisational risk by increasing assurance about the individuals fulfilling healthcare roles. Without an effective solution for different organisations within NHS Scotland to communicate in a secure, integrity-assured and authenticated manner about the assurance processes a particular healthcare professional has completed, each organisation must enforce these processes independently. It is the only way they can feel confident enough to employ an individual in a role within their healthcare organisation. Were they not to complete these checks and the individual was found to be under-qualified, unlicensed or not eligible to work in Scotland, they would be liable. However, when you realise that a junior doctor might rotate places of employment every 4-6 months and an average of two days of their time a year might be spent manually presenting credentials to employers it seems clear there are opportunities, as well as demand, for improvement [408].

6.3 Modelling the Scottish Healthcare Ecosystem

A model of the identification processes within the SHE from the perspective of a healthcare professional's interactions with different entities as they take on different positions and associated roles throughout their career was created. The aim was to create a complex, realistic picture of the institutional structure, the organisations, and the information exchanges that a healthcare professional must complete before they can take on roles required by the system. Information gathered from the RCPE workshop, as well as discussions with healthcare professionals and published reports, informed the

development of this model.

The iStar 2.0 goal-oriented modelling language was used to produce the models. The language provides an approach to represent actors, roles, intentions, resources and strategies that has been applied to many areas of academia including healthcare and security systems [409]. Additionally, Barclay et al attempted to apply this language to model Self-Sovereign Identity systems - recognising the ability for models created using this language to *promote effective communication between domain experts, software architects and developers* [410]. Something identified in both the literature and the RCPE workshop as crucial if effective multi-stakeholder collaboration within a system as complex as healthcare is to be enabled [41] (see Section 3.4).

The iStar language defines a set of components that can be used to describe a system. The default set can be seen in Figure 6.5. The language contains two main views. The strategic dependency (SD) view only shows dependencies between actors within the system modelled, for example, for tasks or resources they require from others. Whereas, the strategic rationale (SR) view creates a representation of the internal goals, tasks and resources relevant to a specific actor enclosed by an actor boundary. As such, the SR view creates a more complete, but also complex, representation of the system being modelled and the rationale behind the dependencies between actors [409].

The approach taken by this thesis has been to create multiple models of the healthcare system from different perspectives relevant to the interactions between a healthcare professional and the institutions and organisations they participate within. Each model will be reviewed in turn, with the different perspectives leading to specific customisation's of the language to more usefully represent the relevant and important characteristics of that perspective. These models are necessarily complex and it is worth reiterating that their purpose is to produce a realistic representation of a healthcare professional's interactions throughout the SHE as they progress through their career. Patient interactions are considered out of scope.

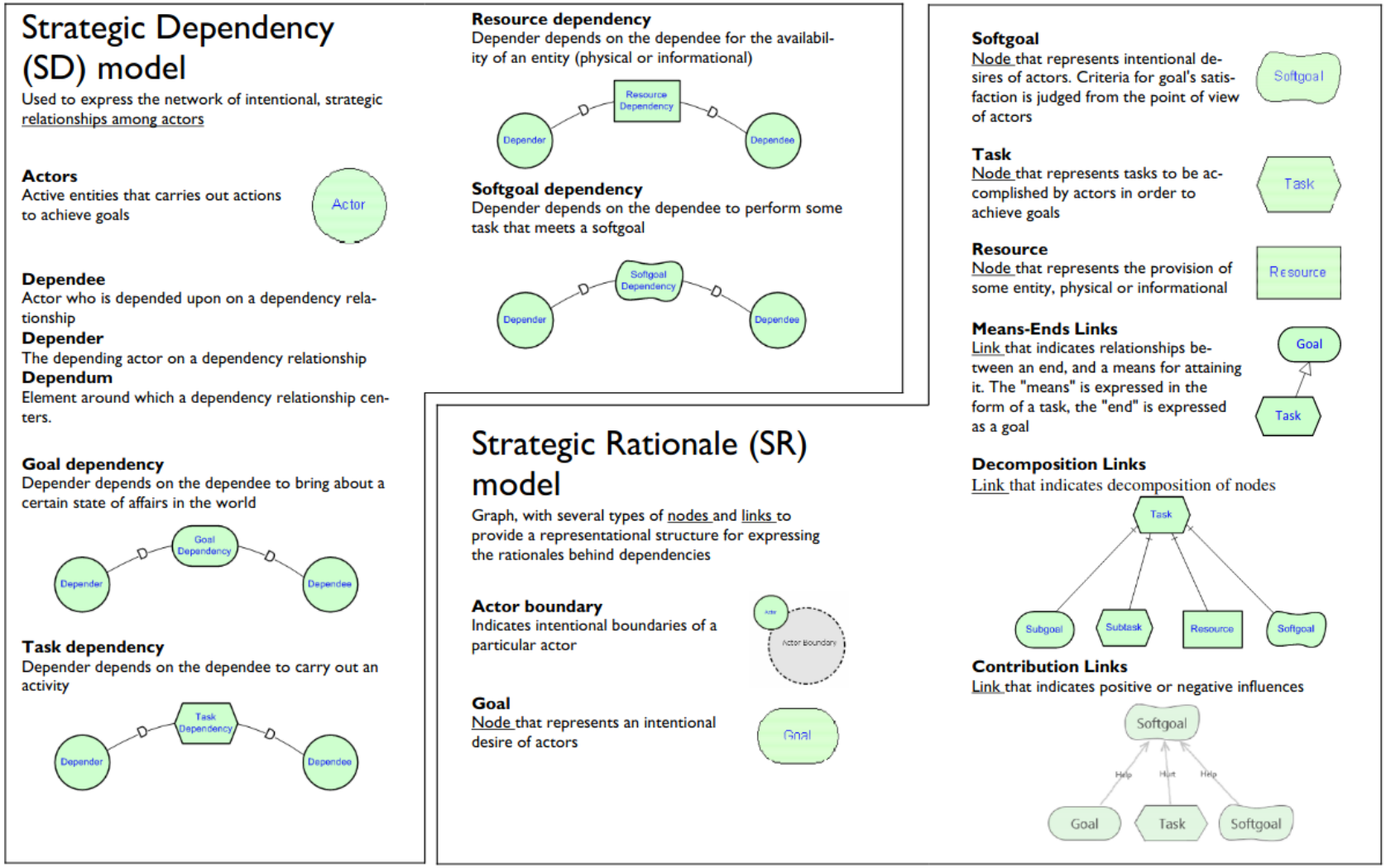

Figure 6.5: iStar 2.0 Modelling Language Components [411]

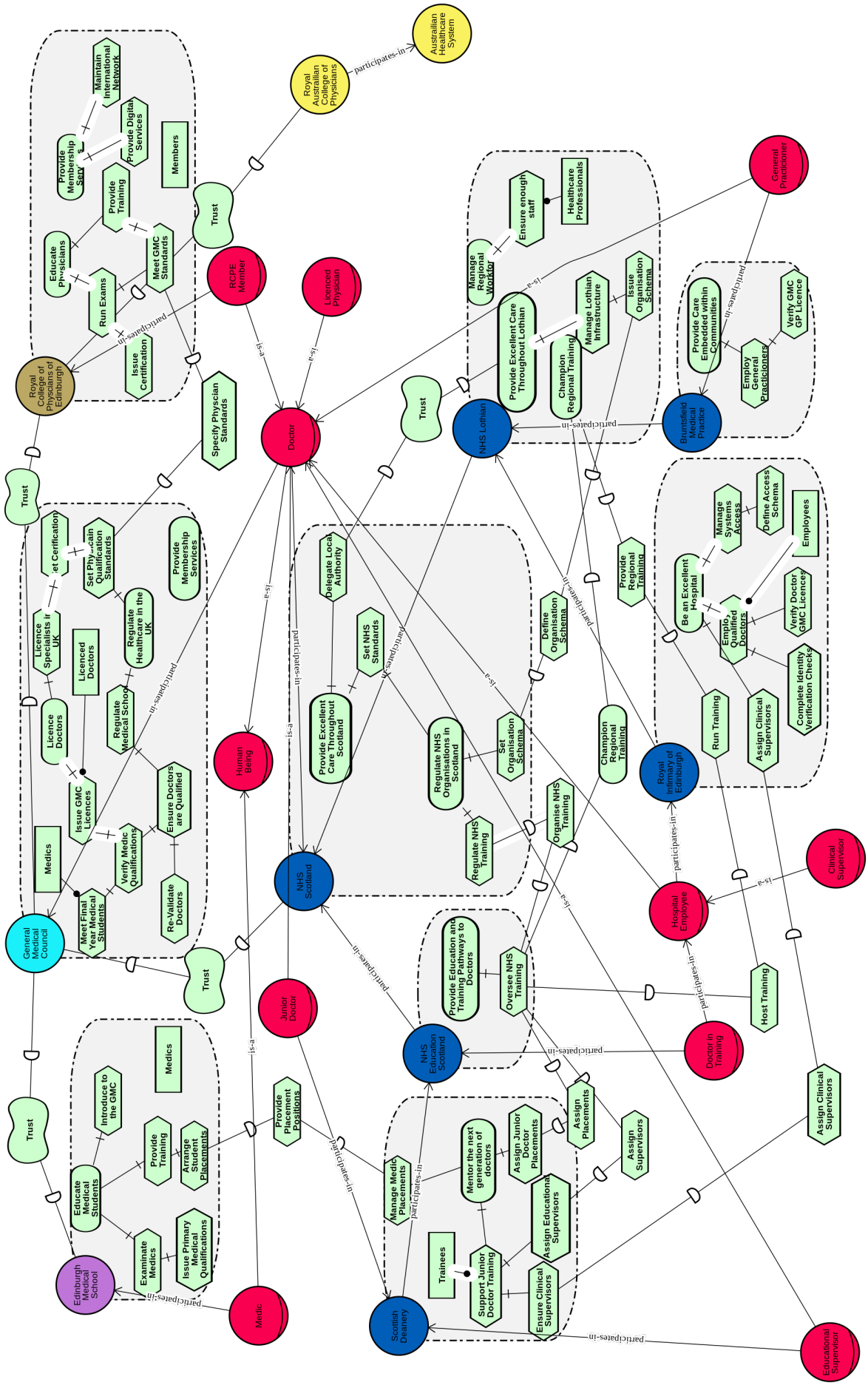

Figure 6.6: Overview of the SHE

6.3.1 High-level Healthcare System

The first model introduces the SHE as a coherent whole (Figure 6.6), representing the key institutions and organisations that it encompasses and the different roles a healthcare professional takes throughout their career. The perspective is from the system and its institutional structures managing the workforce (a resource) of medics and healthcare professionals at different stages of a Physician's career who are vital to providing quality care throughout any healthcare ecosystem [412].

The NHS organisations - represented in dark blue - participate in an organisational hierarchy. NHS Scotland is the umbrella organisation for the National Health Service within Scotland, although it itself is a subsidiary of the UK National Health Service (not represented in the model). Within NHS Scotland there are fourteen regional health boards and seven specialist health boards. Each regional health board is responsible for providing healthcare to their region, employing staff and managing the services including hospitals, pharmacies and general practices. This is clearly region-dependent, for example the Lothian board is responsible for the health of approximately 800,000 people, whereas the Shetland board covers only 24,000 people and has to rely on visiting consultants. The model in Figure 6.6 represents a single health board, NHS Lothian, and only two services this regional board is responsible for, a hospital and a general practice surgery. It is predominantly these organisations that employ healthcare professionals.

The specialist health boards cover the entirety of Scotland. This thesis focuses on a single specialist board, NES, responsible for ensuring medical training at all levels is provided. An important part of NES is the Scottish Deanery, an organisation largely focused on facilitating and supporting junior doctors training, placements and career progression.

Alongside these NHS organisations there is the GMC, the UK wide licensing body for healthcare professionals. It is illegal to practice without a licence. The GMC works with the NES to determine standards for training and re-validation of healthcare professionals. Every healthcare professional must re-validate every five years, demonstrating they have been working to the required standard using evidence collected at yearly

appraisals with senior staff. The GMC will only licence healthcare professionals after verifying the relevant information about them, such as who they are and what their education level is. In addition to the GMC, the UK NHS includes independent specialist colleges that doctors can become members of as they decide how they want to specialise. Figure 6.6 models only one, the RCPE. The RCPE sets a curriculum and training required to become a Physician; these must be approved by the GMC.

The aim of this model is to show how many different organisations, some within the same organisational hierarchy (the NHS), others outside it, participate within a single ecosystem. As well as the dependencies they have on each other in order to maintain high-quality care by ensuring there are enough healthcare professionals at different stages of their career to provide care. NHS Scotland needs healthcare professionals of all different levels and specialties, it delegates employment to regional bodies who further delegate this to specific services. All NHS bodies are dependent on the GMC to licence and re-validate healthcare professionals, they must trust the processes by which this takes place. The GMC is itself dependent on medical schools to educate medics to the required standards and specialist colleges to train and support healthcare professionals as they specialise, eventually becoming licensed specialists on the GMC Registry. Importantly all of these actors are dependent on healthcare professionals fulfilling positions within the system and taking on the associated roles and responsibilities.

6.3.2 Healthcare Professionals ID Artefacts and Dependencies

Within identification systems, ID artefacts act as a container for integrity-assured attestations from authoritative sources that can be presented within an interaction to provide evidence of trustworthiness. In highly structured systems such as healthcare, ID artefacts and their authoritative issuers are formally defined. These artefacts must then be collected and presented by healthcare professionals as they as they take on different roles throughout their career. Figure 6.6 reviewed this ecosystem from an institutional perspective, introducing the organisations and their goals and dependencies. Figure 6.7 focuses on the ID artefacts a healthcare professional collects as they navigate their

career, and the dependencies between them.

This model adapts the iStar 2.0 modelling language, with the orange colored elements representing pieces of evidence as opposed to representing an agent as in the classic language [409]. Each piece of evidence, or ID artefact, is dependent on an institutional actor to perform an issuance task and has an associated healthcare role that a professional can assume once they have collected this evidence. Furthermore, each representation of the evidence might have resource dependencies on other pieces of evidence. These are prerequisites that a healthcare professional is required to collect and present to an issuing authority before being issued a specific piece of evidence. For example, a medical student can only receive a GMC Licence if they have successfully completed their medical training as attested to by a Primary Medical Qualification (PMQ) issued by an authoritative medical school. Therefore, in the model GMC Licence has a resource dependency on the PMQ (see Figure 6.7).

Figure 6.7 represents the important evidence-based interactions in a healthcare professional's career as identified and validated in a workshop at the RCPE [41]. It is recognised that this model is overly complex and difficult to parse, however it is included to give a sense of the administrative complexity that healthcare professionals have to manage. It is important to highlight that this model is a simplification of the real world, with only one specialist college, one regional health board, one hospital and one specialist health board represented. Whenever a healthcare professional changes roles, or interacts with a different organisation they are forced to rebuild this evidence jigsaw from the ground up. This is a time-consuming, repetitive and error-prone process for both the professional and administrative staff involved. Increased evidence can lead to an increased burden of proof, as opposed to providing stronger assurances of a professional's trustworthiness.

In order to simplify the system of study, this thesis focused on two phases of a healthcare professional's career. First, that of a medical student graduating from medical school and receiving a PMQ, which they then use as prerequisite evidence in order to become licensed to practice in the UK and receive a GMC Licence. Second, that of a healthcare professional receiving a role assignment and associated access credential

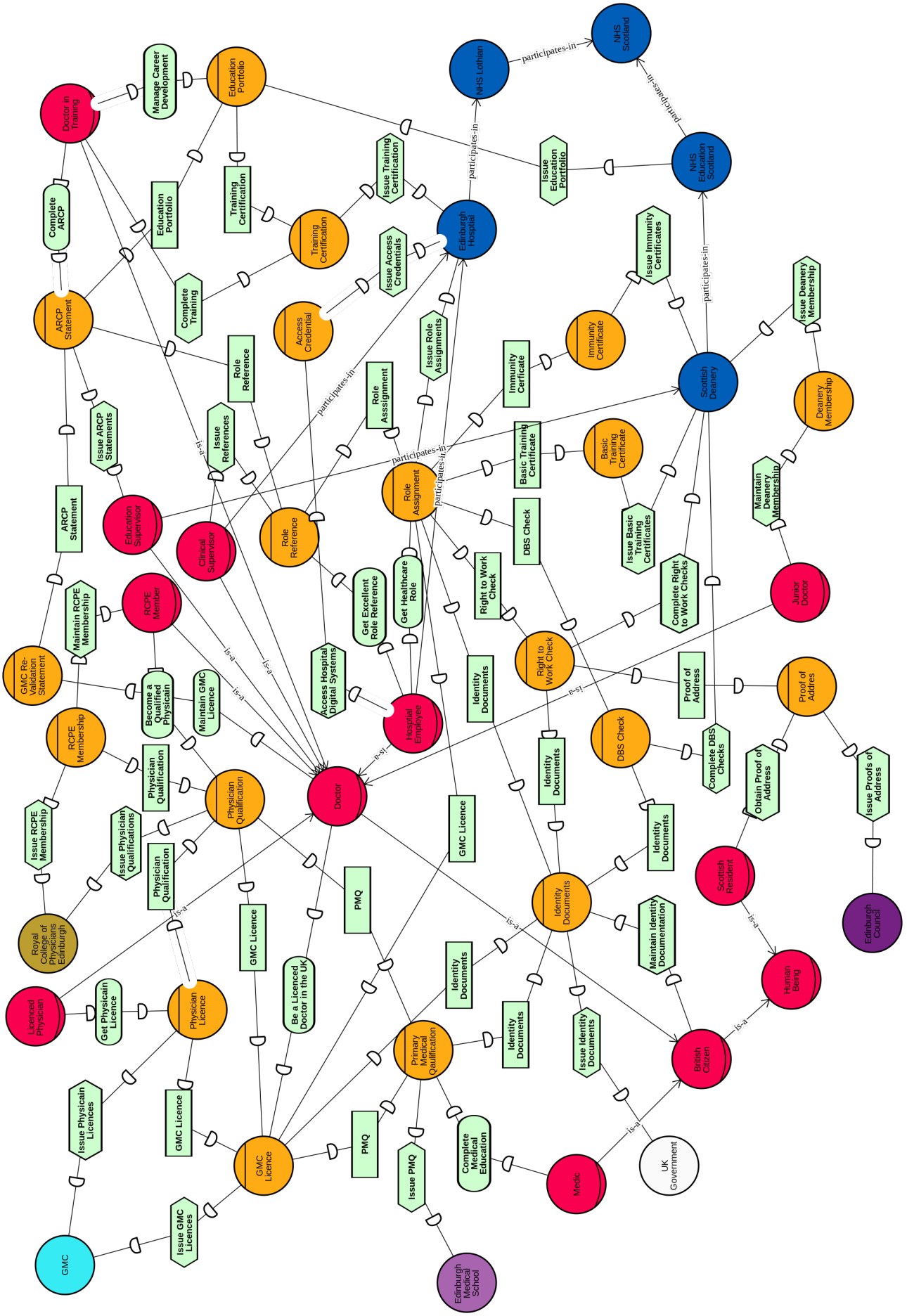

Figure 6.7: Healthcare Professional Credential Dependencies

from an employing organisation such as a hospital. Before they can be onboarded as a hospital employee and assigned a role, they must complete formal identity verification and pre-employment checks. These include a DBS check, a right to work check, verification of license to practice and identity documents. Doctors must also be able to demonstrate they have the relevant immunisations and have completed standard basic training. A focused version of Figure 6.7 can be seen in Figure 6.8.

The next two models concentrate on the two identified interactions from the perspective of the different actors involved, the dependencies they have on each other and their own rationale for performing actions. Specifically focusing on the resource dependencies for verifiable information exchanges, with each piece of evidence modelled as a resource rectangle and colored in orange. For each interaction modelled (a medical student becoming a junior doctor and doctor onboarding), both the strategic dependency and strategic rationale views are shown.

6.3.3 A Medical Student becoming Junior Doctor

A medical student becomes a junior doctor eligible to practice in the UK only when they have received a licence to practice from the GMC. The GMC is highly motivated to licence medical students but needs to have confidence that they are qualified to practice in the UK. A prerequisite requirement for a medic, before they can receive their licence, is the successful completion of their medical education which is attested to by a PMQ from a trusted medical school. The GMC, at least in this scenario, is responsible for the regulation of medical schools and applies institutional pressure to ensure medical schools are providing education to an adequate standard. This regulation gives the GMC confidence that medics in possession of a PMQ from a regulated school are ready to begin a career as a healthcare professional.

In addition to verification of a medical student's PMQ, the GMC administrative staff need to verify the student is who they are claiming to be by performing formal identity verification to the specified level of assurance [391]. They also need to check the medical student has the right to work in the UK and perform a DBS check. As such, the

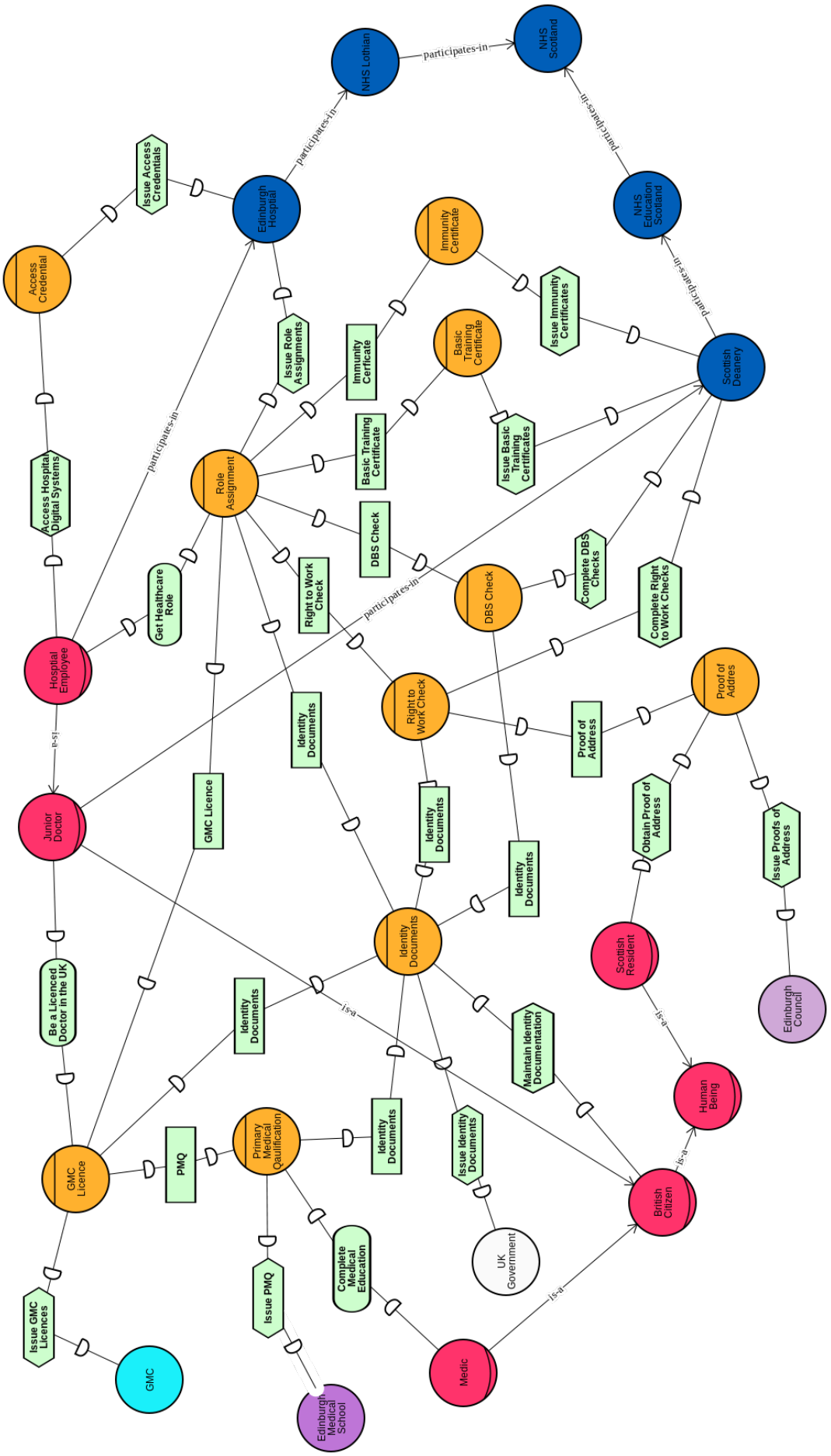

Figure 6.8: Healthcare Professional Credential Dependencies Focused on Medical Licensing and Employee Onboarding Evidence

administrative staff are dependent on the medical student to provide adequate identity documents and proof of address.

The medical student is dependent on the medical school for their education and the subsequent PMQ on completion. They are dependent on the GMC to provide them with a valid GMC licence. The RCPE workshop identified an additional dependency connecting all three actors in the model. The medical student is dependent on their medical school to introduce them to the GMC, this typically happens in their final year of education. The medical school arranges a meeting for each of their final year students and is dependent on the GMC admin team to attend [41]. This provides an opportunity to collect the required information necessary to perform the identity verification and employment checks. As we will revisit when discussing implementation, it is also an excellent opportunity to bootstrap a digital relationship between the medical student and the GMC.

The iStar strategic dependency model for the interaction described can be seen in Figure 6.9. Resource dependencies between actors model the exchange of evidence, when the medical student is dependent on a resource from another actor this is evidence that this actor attests to about the medic. When the GMC is dependent on a resource from the medic, this represents evidence they expect the medical student to be able to provide. The internal rationale for each actor is represented in Figure 6.10, this provides a more complex picture of the same set of interactions.

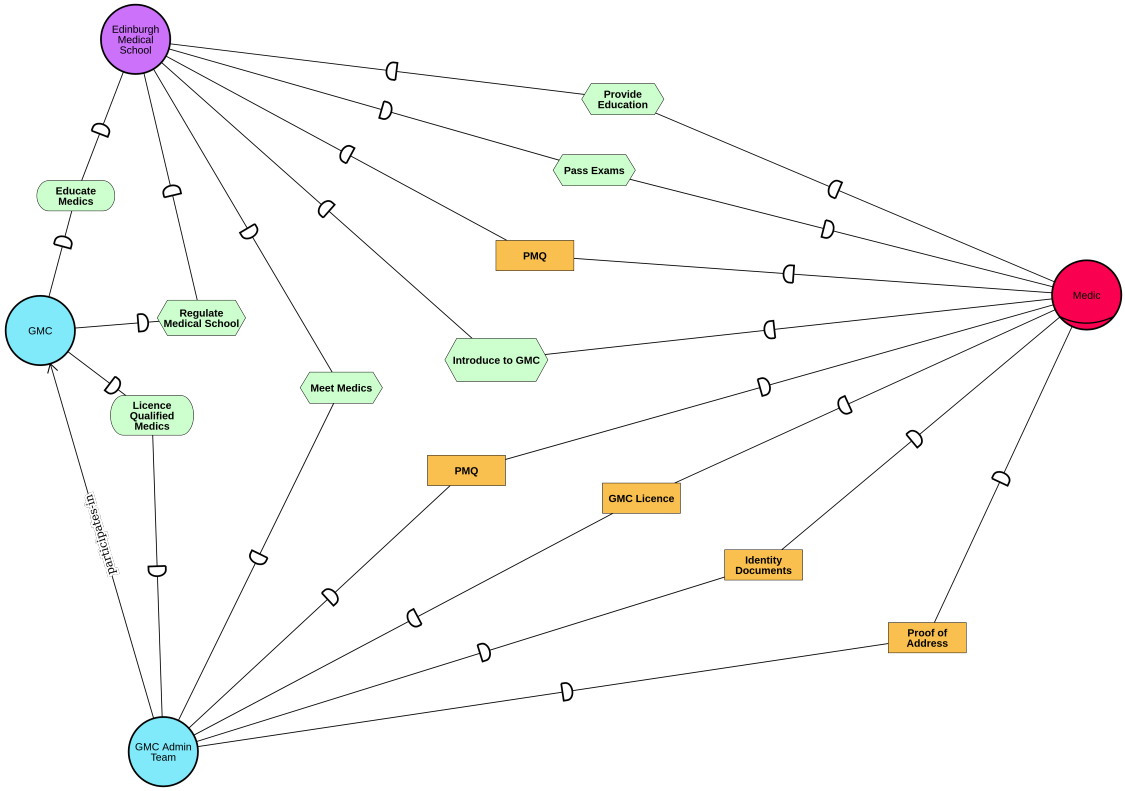

Figure 6.9: SD View of Medical Student Becoming a Junior Doctor

6.3.4 Doctor Onboarding

A medical doctor, within the NHS at least, begins their career as a junior doctor after around seven years of medical studies. This period of their career involves rotating through a number of positions within a number of different hospitals and boards. The author of the autobiographical book *This Is Going To Hurt* [413] rotated through a total of nine posts during his six year career. hospital rotation is not unique to Dr Kay's experience. A video on the UK Health Education Service website states that over 53,000 doctors in training rotate between 66 medical specialisms across 230 plus NHS organisations [414]. They claim that an average of 2 days of a doctor's time per year is spent on manually presenting credentials to employers, leading to 20 million pounds in lost time per year. This is further complicated by the frustration these doctors feel from the repetition of the processes [408]. The NHS released an advisory report on how to increase Staff Mobility [415], in an attempt to simplify these processes. One of the challenges involved with staff mobility stems from the requirement that each new

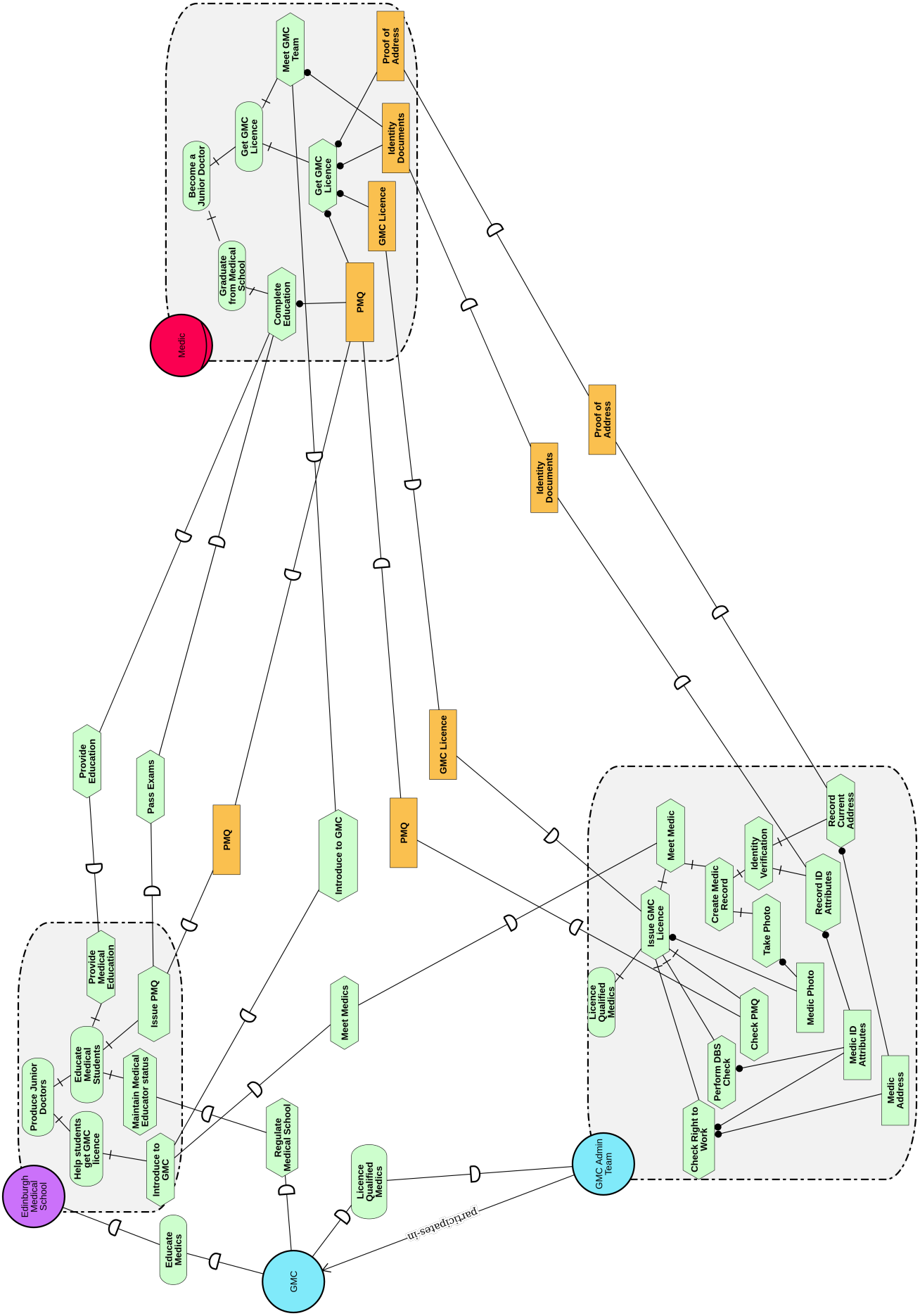

Figure 6.10: SR View of Medical Student Becoming a Junior Doctor

place of employment completes an identity verification check to the required level of assurance as defined by the NHS [391].

The pre-employment checks that a healthcare professional typically goes through, as identified through the workshop at the RCPE [41] and subsequent attendance at a Staff Access hackathon organised by the NHS, are show in the Figure 6.11 and Figure 6.12. These include proving they are licensed, disclosing the required identity documents to pass identity verification, completing a DBS check and a right to work check. Identity verification can be especially complex, depending on the documents a healthcare professional has available. In addition to pre-employment checks, the employing organisation needs to check the healthcare professional has completed basic training and can pass a health screening to verify their immunisation status. Once satisfied, the employing organisation, through a human agent, assigns the healthcare professional a role within the organisation and provisions them with the relevant systems access.

These checks are an important and necessary institutional pressure that provides assurance that only eligible individuals receive role assignments within a healthcare organisation [15]. However, a junior doctor might rotate through placements assigned to them by the Scottish Deanery as often as every 4 months. Despite these organisations all being part of NHS Scotland, and in some cases, part of the same regional health board, each time a doctor rotates to a new organisation they often repeat these lengthy processes. Employing organisations have no mechanism by which to gain assurance that these checks had successfully carried out and so are forced to repeat them. Discussion at the RCPE workshop indicated that in many cases this can mean healthcare professionals completing the same basic training and receiving duplicate immunisations because of their inability to prove that these took place at a previous organisation [41].

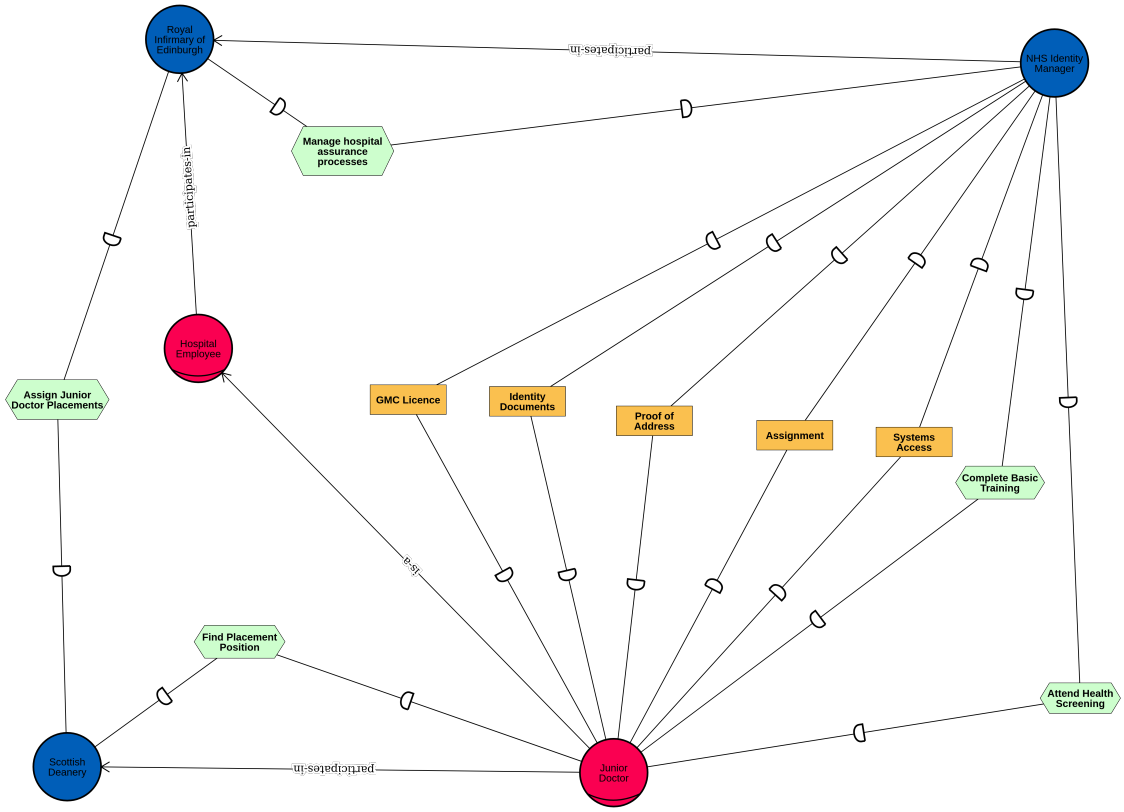

Figure 6.11: SD View of Junior Doctor Onboarding at a Hospital

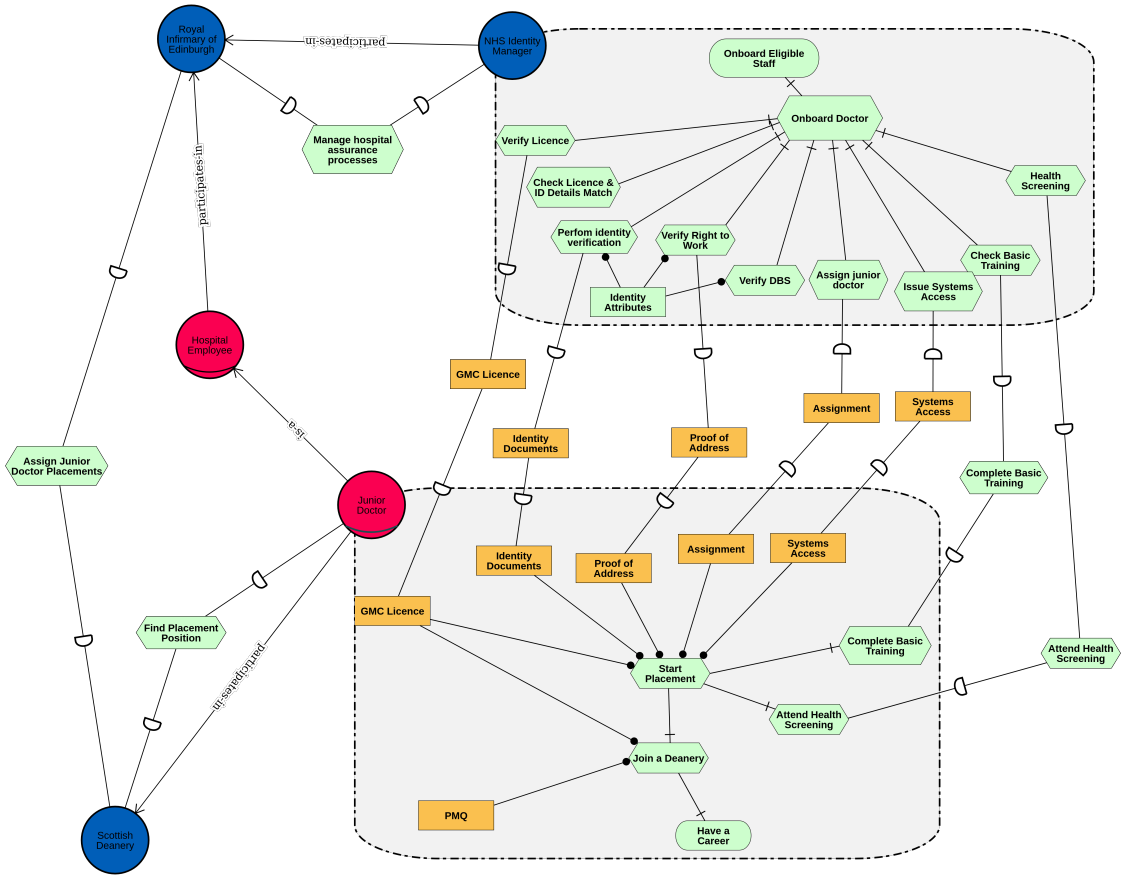

Figure 6.12: SR View of Junior Doctor Onboarding at a Hospital

6.4 Implementation

This thesis produced a technical implementation to demonstrate and explore the potential to augment existing identification processes with cryptographically signed, verifiable digital evidence. This evidence, in the form of verifiable credentials, can be issued by domain-specific, trusted actors within the SHE identifiable under a public key infrastructure based on decentralised identifiers. However, the evidence itself can be stored, managed and presented by the healthcare professional to whom the evidence pertains. This empowers these professionals to accumulate digital evidence from multiple sources through the interactions and relationships they have with actors throughout the SHE. Healthcare professionals can then use this evidence to strongly authenticate themselves to the required level of assurance, whilst reducing the burden that achieving this level of assurance using paper-based evidence verification currently

places on them.

The implementation presented in this thesis is not meant as a concrete set of changes the SHE should implement immediately. Rather it is used to illustrate the design decisions that could be made, demonstrate the practicality of the underlying technology and to provide a realistic system that can then be evaluated. To further this objective, the implementation has been produced within an interactive Juypter notebook environment customised specifically to support decentralised identity interactions using the Hyperledger platform [43]. The aim is to contribute an artefact that can be used in future research to facilitate further discussions with the relevant stakeholders within the SHE.

6.4.1 Scottish Healthcare Ecosystem Proof of Concept

A PoC has been produced that demonstrates how the assurance processes modelled in Section 6.3 could be augmented using the HVIEP . The business logic for each actor has been implemented within Jupyter notebooks and uses a custom class, the AriesAgentController, to interface with their specific agents. For simplicity, the healthcare professional is also represented through a Jupyter notebook despite the fact that they would likely use a mobile application in a more realistic scenario. The notebooks are self-documenting with an emphasis on describing the context surrounding the specific interactions. All code created for this thesis has been open sourced and is available on GitHub[2].

6.4.1.1 Medical Student Becoming a Junior Doctor

In order to practice medicine within the UK, all doctors require a licence from the GMC. This PoC shows how the GMC can become an issuer of integrity-assured digital GMC licences, signed by an Aries agent acting on their behalf. The agent is publicly identifiable through a decentralised identifier registered and resolvable on an Indy network. In order to issue these digital GMC licences, the GMC must sign and author a

[2] `https://github.com/blockpass-identity-lab/scottish-healthcare-ecosystem-poc`

SCHEMA transaction to the ledger specifying the attributes that the licence contains. As well as a CLAIM_DEF transaction which creates and publishes their public issuing key for this schema to the ledger (currently a CL-RSA keypair [29]). The transactions written as part of the PoC can be seen on the Sovrin StagingNet - SCHEMA[3], CLAIM_DEF[4].

It is not the aim of this thesis to precisely specify this schema, such a task should be undertaken with consultation throughout the healthcare system and the professionals that will use this credential for the purposes of identification. However, there is some information that can be drawn on. The current GMC Register is a searchable database of licensed healthcare professionals in the UK which includes, their name, licence number, medical degree, training and specialisation. All of this would seem useful to include within a credential [416]. Another point of reference is the GMCs expressed desire to include additional sensitive information in this public register in 2016, including a photograph, which after a consultation and significant backlash over safety and privacy concerns was rejected [417]. While the GMC register is currently public, a digital, VC that leverages privacy-preserving credential mechanisms would be explicitly private while at the same time empowering the credential holder, the healthcare professional, to present any subset of attributes within the credential in interactions they deem appropriate [240]. For example, when onboarding at a new hospital. It remains to be demonstrated if healthcare professionals would accept this more sensitive information within a credential format, but it would appear to satisfy both parties and has the benefit of providing strong identity verification as well as licensing [391]. We take this approach in the scenarios presented, and show that a Base64 encoded image can easily be included as an attribute of a verifiable credential under this architecture.

Once the GMC has an Aries agent with a DID, GMC licence schema and associated issuing key on a public Indy network, they can technically issue verifiable GMC licence credentials to other Aries agents they establish connections with. However, the human processes surrounding the issuance of these credentials are key to providing the necessary assurances required for other actors to place trust in these credentials when

[3] https://indyscan.io/tx/SOVRIN_STAGINGNET/domain/223371
[4] https://indyscan.io/tx/SOVRIN_STAGINGNET/domain/223372

presented by a healthcare professional. This includes understanding how connections are established, the requirements that must be met before the licence is issued and how attribute values are populated. The iStar models in Figures 6.9 and 6.10 are instructive here, because they show that these human processes are already in place.

The GMC Admin staff are introduced to all final year medical students by their respective medical schools. Through this meeting, the typical pre-employment checks are initiated, including identity verification and right to work checks. This face-to-face meeting provides an opportunity for the medical student and GMC to exchange private, pairwise DIDs and establish a secure, private communications channel that uses DIDComm [216]. Furthermore, all attribute values for the GMC licence including a photograph of the medical student can be collected and recorded against the DID identifying the medical student's agent in this relationship. The GMC has a high degree of confidence that they only establish connections with medical students, through the trust they place in the medical school's introduction. Furthermore, they do not issue licences until the medical student has completed their education and received a PMQ. In the PoC, the medical school is also represented through an Aries agent with a public identifier able to issue digitally verifiable PMQ credentials to medical students upon qualification. This allows the medic, upon qualification, to digitally present their PMQ credential across the secure connection they established with the GMC previously. The GMC resolves the relevant cryptographic information from the ledger, verifies the integrity of the presented attributes and the authenticity of the issuing DID. Once they are confident that the PMQ is valid, they can check the attributes presented against the details recorded when the connection with the medical student was established. If they match then the GMC licence is issued and the medical student becomes a licensed doctor.

Under the PoC developed for this thesis, the GMC licence is the only revocable credential considered. Using the CKS scheme [302] implemented within the Hyperledger platform, all GMC licences are able to be revoked by the GMC Aries agent publishing an update to the revocation registry to the Indy ledger. Once revoked, the healthcare professional who has that licence is no longer able to create valid proofs of non-revocation.

This is important because a doctor might retire, or might be removed from the registry for bad practice. Furthermore, every five years a doctor must go through a re-accreditation process with the GMC. While this interaction is not considered here in-depth, this thesis believes that a secure, digital relationship between a healthcare professional and the licensing body presents an opportunity to re-imagine this interaction and indeed the entire relationship.

6.4.1.2 Doctor Onboarding

When onboarding at a hospital, a healthcare professional goes through standard pre-employment checks. As doctors, and particularly junior doctors, regularly move around different organisation these manual checks become repetitive and time consuming for all involved. The evidence identified in the modelling (see Figure 6.11), includes identity documents, right to work status, DBS check, GMC licence and immunity certificates. Before these can be turned into digitally verifiable forms of evidence useful to healthcare professionals throughout the SHE it is important to understand which actors would be trusted to attest to this information. Ideally these would be primary source attestations, however that would require engaging with actors outside of healthcare such as local councils or government bodies. Instead, this thesis identified the Scottish Deanery as a candidate issuer ideally suited to provide digital evidence that can be used within these pre-employment checks. Particularly for junior doctors, which the Deanery is already responsible for assigning placements throughout Scotland. The GMC would be another likely candidate issuer, as it is already trusted throughout the healthcare system, has a continuous relationship with the doctor throughout their career, and already performs these pre-employment checks when licensing doctors. Alternatively, there could be a role for a specialised identification team throughout the system whose role is to perform physical, in person identification processes and provide trustworthy, verifiable evidence in a digital format that can be used throughout the SHE.

As discussed, in the PoC implemented for this thesis, the Scottish Deanery acted as the issuer of Right to Work Status, DBS Check, Compulsory Basic Training and Immunity Certificate credentials. These pieces of digital evidence would only be issued

after manually performing verification through practices already in place today. These credentials all included an expiry date, upon which the healthcare professional would need to seek fresh versions of this evidence after another physical check. The length of time the digital evidence should be valid for, would need to be discussed with the stakeholders of the SHE, however, this thesis believes it would be less frequent than the 4-6 month rotations that junior doctors typically go through. Furthermore, by centralising the entity responsible for these employment checks it would be possible to ensure they are effectively and consistently undertaken by actors specifically trained for these purposes.

Once a junior doctor has graduated, received their PMQ and GMC Licence, they establish a relationship with the Scottish Deanery. Through this, they would receive the relevant digital evidence required to onboard at a hospital for their placement. Then at the beginning of their placement, pre-employment checks would be greatly simplified. The healthcare professional would meet the administrative staff at the hospital and during that meeting establish a secure connection with the hospitals information system. This is likely to be achieved by scanning a QR Code presented by the administrative staff to initiate the identification process. Across this connection they would be able to prove they have the right to work in the UK, they have successfully passed a DBS check and they have the adequate immunisations. All of which would be evidence that these checks took place within the defined time as attested to by the Scottish Deanery. They would also present their digital GMC Licence, including the photograph biometric which enables the licence to act as an identity document. The hospital admin staff would only require a single physical ID document with claims that match information held within the digital GMC Licence, for example a passport or driving licence to complete the identity verification to the required level of assurance [391].

Once the hospital's administrative staff is satisfied, the healthcare professional is issued an assignment credential and systems access credential. These digital credentials can be used to prove current employment as well as having the possibility of being integrated into hospital information systems. This would enable a simplified, secure

login experience without the requirement for the healthcare professional to remember additional account credentials. This is out of scope for this thesis, although it has been demonstrated feasible to use verifiable credentials with other account management standards such as Open ID Connect [418]. This could be especially powerful if account access additionally required the disclosure of a non-revoked attribute from the GMC Licence, as it would allow continuous authentication of the individual as a licensed healthcare professional.

Finally, this thesis emphasises that the verifiable credentials issued to and presented by the healthcare professional within the PoC only augment current assurance processes. Ultimately, it is a human decision that determines whether the evidence presented by an individual is adequate for identification processes determined by their organisation. The PoC illustrates how making this evidence digital and verifiable, by identifying appropriate issuers, can simplify the repetitive processes and make it easier for a healthcare professional to establish and manage evidence required for the identification processes they must engage in, throughout their career. Verifiers, assuming they know and trust the issuers of this evidence and the processes governing how it is issued, can have confidence in the information presented thanks to the cryptographic properties of the underlying technology stack.

6.5 The Aries Jupyter Playground

The Aries Jupyter Playground [43] is an open source Github repository and codebase that aims to help researchers and developers experiment and quickly deploy Hyperledger Aries based credential systems that can be easily customised to specific use cases. Specifically, within an academic setting, the aim is to simplify the process by which researchers can spin up a set of agents and develop business logic around how these agents interact and exchange verifiable information. Allowing researchers to design and validate specific use cases involving a set of actors, roles, purposes, and information exchanges within an interactive environment.

This environment was used to realise the SHE PoC and has since been extracted

into a standalone GitHub repository that anyone can clone and customise accordingly. This work has been published in the Software Impacts Journal [43]. The playground uses Docker Compose [419] to manage an environment containing the relevant Docker services required for each actor modelled within the playground. Each actor has the following services:

- An ACA-Py instance - This uses a Docker image published and maintained by the open-source ACA-Py project. The agent instance is configurable through an environment file, including defining the Indy network they interact with and a cryptographic seed. An Indy network can be run locally, or a test network such as the Sovrin StagingNet can be used.

- A PostgreSQL database [420] - This stores and persists the state of the agent across multiple instantiations of the environment. Specifically, any cryptographic keys and credential objects they have received.

- A Jupyter Notebook server [421] - This is where custom business logic can be written to model an actor controlling their agent to engage in protocols with other agents within the environment. The ACA-Py agent exposes a Swagger interface [422] and posts events (such as when it receives a message) to a definable endpoint. A custom "pip installable" Python library, the Aries Cloud Controller [423], has been developed to simplify this process within the notebooks. Additionally, each notebook's Docker service has the same recipes folder mounted as a volume providing templates for common interactions, further reducing the time it takes to get a working demo.

- An Ngrok server [424] (optional): Ngrok can be used to tunnel the HTTP port that the ACA-Py instance exposes to receive messages from other agents to the public internet. This is useful if one wishes to interact with agents not running locally on their computer, for example, mobile agents downloadable from app stores.

The resulting architecture is illustrated in Figure 5.4, with two example actors Bob and Alice, as modelled in the default Aries Jupyter Playground. The key innovation

here is the ability for any set of actors within a system to easily be modelled utilising this environment in a way that minimises the prerequisite knowledge of the different components of the HVIEP and the way they integrate together, which can be complex. Researchers can focus on writing domain-specific business logic within a familiar Jupyter notebook interface. This allows them to explore research related to this technology quickly. Furthermore, any experiments can easily be made self-documenting using Markdown cells to be straightforward for other researchers to replicate, challenge, or extend. An example of the resulting notebooks produced within this environment to model the SHE can be found in Figures 6.13 and 6.14.

6.6 Conclusion

This chapter has applied one of the emerging software engineering frameworks for credential-focused identification systems, the HVIEP, demonstrating how it provides a toolkit that can be used by application developers to instantiate ICT artefacts capable of engaging in privacy-preserving cryptographic credential protocols customised towards a contextually bound interactive setting. The HVIEP moves a credential mechanism based around a signature scheme with efficient protocols discussed in Chapter 4 from theory to practice, absorbing much of the complexity associated with engaging in these cryptographic protocols. This includes the communication protocols, data models and storage of cryptographic material both private and public. Furthermore, the HVIEP attempts to combine these cryptographic protocols with the open standards of Decentralised Identifiers and Verifiable Credentials being developed within the decentralised identity community. This process is ongoing, but is intended to ensure interoperability across implementations and prevent centralisation through vendor lock-in.

This chapter has demonstrated how advanced ICTs can be used to augment existing identification processes within a social context. The steps taken included; understanding the social context, its requirements for identification and existing identification system; interacting with key stakeholders of the system [41]; mapping the ID artefacts, their dependencies and dependents (see Figure 6.7); defining how existing ID artefacts

Meet Medic Face to Face through Medical School Introduction

GMC and Medical Student Establish Connection

A GMC Administrative Staff would be introduced to each Medical student in a physical meeting organised by the medical school. As part of this meeting **which already happens**, the GMC Admin staff could establish a DIDComm connection between the ACA-Py agent representing the GMC (which they would have to be authorised to use) and the medic's agent.

Both parties can therefore have a high degree of confidence that the digital connection they just established represents the entity they think it does. E.g. a medical student, or the GMC.

In [13]:
```
# Alias for invited connection
alias = "Medic"
auto_accept = "true"
# Use public DID?
public = "false"
# Should this invitation be usable by multiple invitees?
multi_use = "false"

invitation_response = await agent_controller.connections.create_invitation(alias, auto_accept, public, multi_use)
# Is equivalent to above. Arguments are optionally
# invitation_response = await agent_controller.connections.create_invitation()



# You will use this identifier to issue a credential across this connection
connection_id = invitation_response["connection_id"]
```

```
------------------------------------------------------------
Connection Webhook Event Received
Connection ID :  a849ff01-e859-445d-9889-36d572a38b86
State :  invitation
Routing State :  none
Their Role :  invitee
------------------------------------------------------------
```

Share Invitation Object with Medic

Again this would likely happen as part of a face to face interaction.

In [14]:
```
invitation = invitation_response["invitation"]
## Copy this output
print(invitation)
```

```
{'@type': 'did:sov:BzCbsNYhMrjHiqZDTUASHg;spec/connections/1.0/invitation', '@id': 'c75a7b88-3bea-4b67-a201-00839e36ebb8', 'serviceEndpoin
t': 'http://592e-86-18-68-143.ngrok.io', 'recipientKeys': ['Fb3REpx3yJVdd6zqrNNRrWsQBjq65dV8CpE52JSsdFQC'], 'label': 'General Medical Coun
cil'}
------------------------------------------------------------
Connection Webhook Event Received
Connection ID :  a849ff01-e859-445d-9889-36d572a38b86
State :  request
Routing State :  none
Their Role :  invitee
------------------------------------------------------------
------------------------------------------------------------
Connection Webhook Event Received
Connection ID :  a849ff01-e859-445d-9889-36d572a38b86
State :  response
Routing State :  none
Their Role :  invitee
------------------------------------------------------------
------------------------------------------------------------
Connection Webhook Event Received
Connection ID :  a849ff01-e859-445d-9889-36d572a38b86
State :  active
Routing State :  none
Their Role :  invitee
------------------------------------------------------------
Connection ID: a849ff01-e859-445d-9889-36d572a38b86 is now active.
```

Figure 6.13: GMC Notebook to Establish Connection with Healthcare Professional

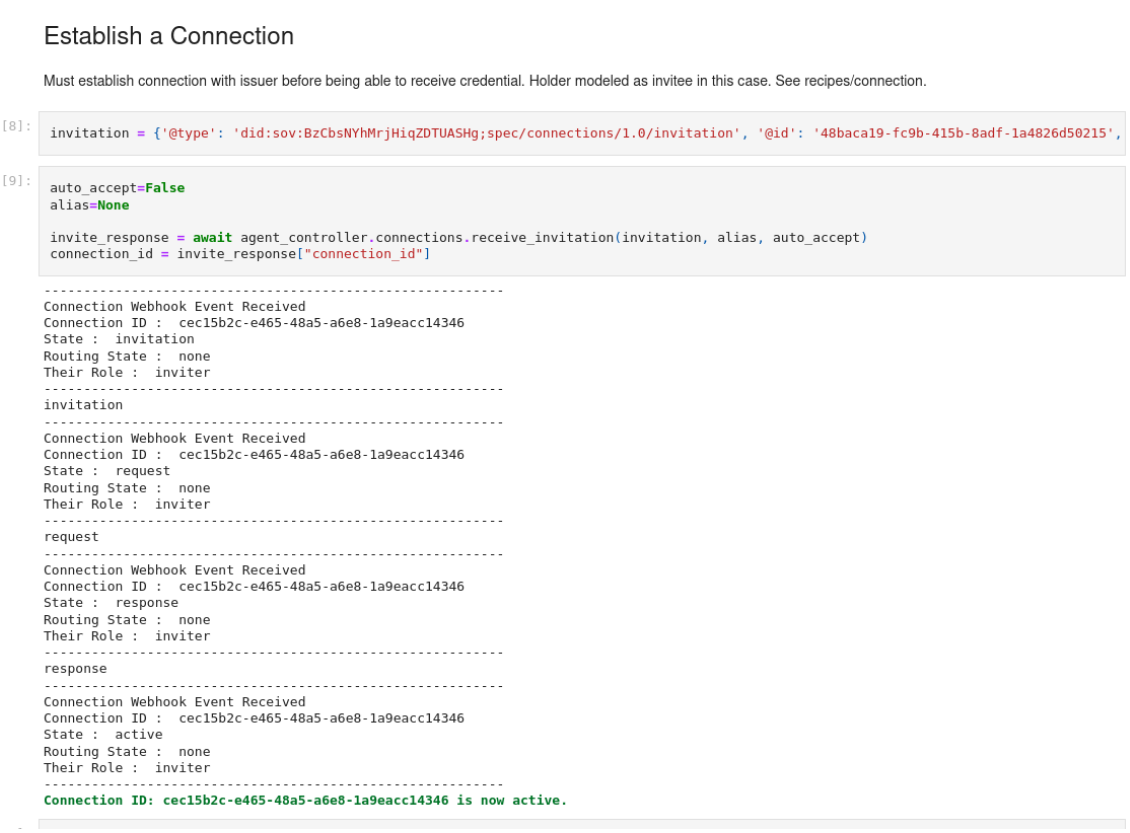

Figure 6.14: Healthcare Professional Accepting Connection with GMC

would be both issued and presented when technologically augmented and finally implemented these flows within a Jupyter notebook environment to demonstrate their feasibility. It should be possible to apply a similar approach to produce a PoC for any social context with an existing system of identification. Certainly this thesis argues strongly that any attempt to introduce new identification technologies into existing identification systems should involve dialogue with those impacted by these technologies for the beginning and throughout the process. Echoing the busting bureaucracy report, any changes must lead to greater trust being placed in those identified and increase their ability to effectively perform their role [24].

This thesis intentionally focused on healthcare for a number of reasons. Advanced ICTs are widely recognised as having transformative potential for healthcare services, however, it is consistently reported that technology hinders healthcare professionals' ability to provide effective care [24]. Secondly, the purpose of identification is clearly justified and legitimate, however, existing identification systems have placed an increasingly high administrative burden on healthcare professionals to collect, record, manage

and present the set evidence required to meet the levels of assurance of different organisations and institutions. Finally, healthcare is a highly structured subsystem of society and as such the governance of this subsystem is already well-defined and understood. While important work is required to understand how these technical systems of identification interface with, and are constrained by, human systems of governance, the PoC left this out of scope.

In conclusion, this chapter has demonstrated how the HVIEP makes privacy-preserving credential mechanisms technically feasible and functionally capable of augmenting a complex and concrete, modelled identification system, the Scottish Healthcare Ecosystem, through the instantiation of a set of ICT artefacts. It follows that the internal constraints and limitations placed on these artefacts, and in turn the interactions with which they are applied, by the HVIEP and the cryptographic protocols it operationalises should be analysed and evaluated. This is set out in the next chapter.

Seventh Chapter

Evaluation

> *"Often we shall have to be satisfied with meeting the design objectives only approximately. Then the properties of the inner system will 'show through.' That is, the behaviour of the system will only partly respond to the task environment; partly, it will respond to the limiting properties of the inner system."*
>
> Herbert A. Simon, Sciences of the Artificial [10]

The PoC developed for this thesis synthesised a set of technical artefacts that modelled agents of different actors within the Scottish Healthcare Ecosystem using the HVIEP. The artefacts engaged in identification processes from important stages of a healthcare professionals career. The implemented flows demonstrated how existing evidence and ID artefacts used within the Scottish Healthcare Ecosystem could be also represented as a set of cryptographically signed, integrity assured attributes that are then able to be presented and verified over a digital medium. Furthermore, this evidence is held at the edge, by the healthcare professionals to whom it pertains, as opposed to within centralised services that must be queried only after the healthcare professional has been uniquely identified to a high degree of assurance.

This thesis makes the case that this approach can reduce the burden formal processes of identification currently place on identified professionals while retaining the levels of assurance required within a healthcare setting. The chapter evaluates the internal technical structure of these synthesised artefacts and the constraints this places on the identification processes they are used to augment. It aims is to demonstrate that identification processes augmented using privacy-preserving cryptographic protocols

operationalised within the HVIEP are both practical and secure. Furthermore, from this analysis of the underlying technology this chapter identifies important considerations that should be taken into account when architecting an identification system using the HVIEP.

The distributed ledger and the resulting footprint from this architecture is analysed as a root of trust that enables control over which actors can make specific attestations and under what authority. The publication of this information within a verifiable data registry accessible to all actors, providing integrity-assured information that can be incorporated into decision making processes (both human and machine). The design of the schema and the identification processes that credentials issued against this schema are intended to support are considered from the perspective of the overall performance of these interactions. This includes the compute time required by each interacting entity participating in the protocol and the communication time required send messages between entities across the network. A set of experiments were developed within the custom Jupyter environment to benchmark these interactions under varying schema and interaction properties within a production setting. Finally, the cryptographic protocols themselves are benchmarked independently from the Aries agent layer. This is intended to evaluate how the transition from CL-RSA signatures [29] to BBS+ signatures impacts the time and space complexity of these schemes. Thus allowing inference of the impact on the overall performance of artefacts that use these signature schemes as a credential mechanism within identification systems. This is highly relevant as this transition is already underway within the Aries codebases and the wider community of practitioners [215]. All the technical evaluation within this chapter is considered within the context of healthcare, however, it is also applicable more generally to the design and implementation of credential-focused digital identification systems using the HVIEP.

7.1 Ledger Footprint

The Hyperledger Indy-based distributed ledger is a custom ledger designed specifically to facilitate a cryptographic credential system that integrates the cryptographic

protocols of Camenisch and others within a software development framework [28, 29, 302]. The ledger is designed to be a highly available, public permissioned verifiable data storage system for the necessary cryptographic information required to verify signatures on credentials. It acts as a public key infrastructure supporting issuance, verification and revocation of Verifiable Credentials based on privacy-enhancing attribute-based credential cryptography, currently using CL-RSA signatures [29]. The details of the transactions types, and permissions on this data storage system are covered in Section 5.4.1. This section evaluates how the transaction footprint on the ledger scales in relation to the actors and credentials modelled in the ecosystem. The size and type of information that is required to be stored on a ledger to support an implementation, affects the complexity and cost of that implementation, as well as having implications for the privacy of actors and information flows within the ecosystem. Building on this analysis, this section demonstrates how the footprint on the ledger, as a historical chain of signed transactions maintained by a distributed set of actors, can be used as an input to assess the robustness of an identification system and the assurance that can be placed in the credentials issued within it.

The first important point to note is that within credential ecosystems rooted in Indy networks the only actors required to author transactions to the ledger are those wishing to issue verifiable credentials. All other actors within the system - the holder and verifiers of credentials - only require read access to this ledger. Issuers must first get a NYM transaction authored to the ledger - this is the current mechanism by which issuers under the Indy network publish a public DID. The NYM transaction contains an identifier for the actor and a public key which can be used to authenticate the actor against the identifier by challenging them to produce a signature that can be verified using the public key contained within the NYM transaction. Within Indy networks the identifier is referred to as a NYM, while the public key is called a verkey. Currently this is based on an Ed25519 key pair [425], with NYM and verkey values generated from a private seed value known only to the issuer. Indy networks are permissioned such that an actor can't author their own NYM transaction, rather they must get another actor, already identified on the ledger, and authorised with write access, to author this

transaction on their behalf.

A NYM transaction acts as the genesis event for the actor in respect to being publicly identifiable within an Indy-based verifiable credential ecosystem. Using this key pair, the actor's agent signs all future transaction they wish to author to the ledger, producing a digital footprint correlated with their public identifier and verifiably authored by the corresponding key pair. Furthermore, Indy ledgers are permissioned, meaning that for transactions to be included within the ledger, they must either be authored or endorsed by an actor using a NYM with the appropriate role (TX ENDORSER). This footprint is persisted within the Indy ledger, with the distributed set of nodes first authenticating the signature of the transaction author, then verifying the appropriate write permissions of the transaction before it is included in the ledger. This requires that the NYM that signed the transaction has the role of endorser, or the transaction has an additional signature from a NYM with the appropriate role (see Figure 7.1). This produces a historical, time-ordered set of transactions with a verifiable audit trail indicating how each transaction came to be included within the ledger. The transaction endorser essentially acts as a gatekeeper, providing an additional layer of assurance about the information that gets added into the ledger. Which, while not suited to some use cases, this is ideal for a high assurance professional context such as healthcare. These have a clear hierarchy of authority and responsibilities as can be seen in Figure 6.6.

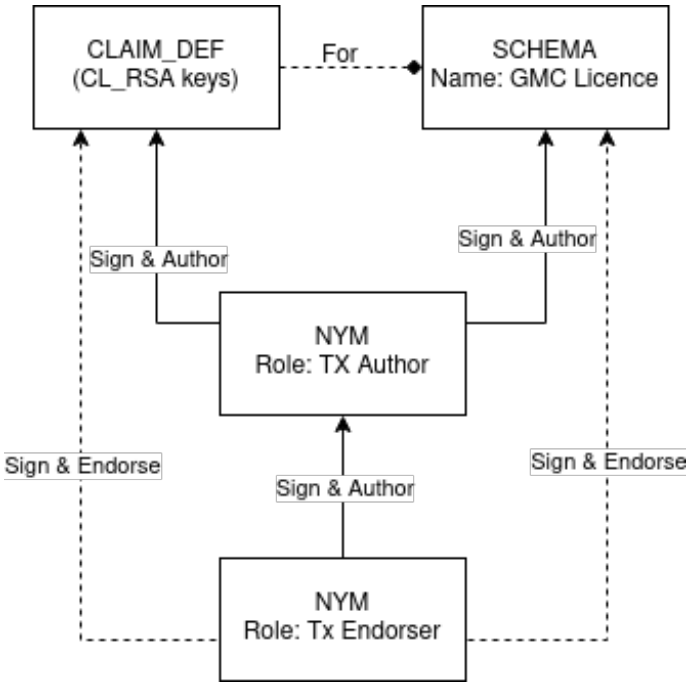

Figure 7.1: Indy Transaction Relationships Between NYM, SCHEMA and CLAIM_DEF

Once an actor has a public DID, through an authored NYM transaction and is either in the endorser role or has identified a suitable endorser they may begin to write further transactions. These include ATTRIB transactions which updates the serviceEndpoint for the DID document associated with their DID. SCHEMA transaction specifying a name, version and set of attribute names that a credential of this type must fulfil. Once a SCHEMA has been authored, only the transaction author can version the schema on the ledger. Schema names must be unique to a ledger. Then actors who intend to issue credentials against this schema create a CL-RSA key pair and author a CLAIM_DEF transaction which publishes their public key from this pair. Note actors do not need have authored the SCHEMA, to be able to write a CLAIM_DEF against that SCHEMA. This is a potential weakness of the ledger, assuming an actor can get their transactions endorsed, they can become issuers of verifiable credentials against any schema objects that exist on the ledger. With high assurance credentials such as those used to identify healthcare professionals this is clearly not acceptable, further emphasising the importance of the endorser role. When authoring a CLAIM_DEF an actor must decide if they wish the credentials they issue to be revocable, if so they must initialise a revocation registry for this CLAIM_DEF from which credential holders can create proofs that their

credential has not been revoked. To do so, they must author a further five transactions, a REVOC_REG_DEF and four REVOC_REG_ENTRY transactions to initialise the registry. Every time an issued credential is revoked an additional REVOC_REG_ENTRY must be authored. A registry has a fixed, definable capacity defaulting to 1000 entries before a new one must be authored. The current revocation implementation within Indy also requires a large static file containing cryptographic constants used within the protocol. This must be hosted at a public url identified by the REVOC_REG_DEF, without access to this location holders and verifiers are unable to create or verify proofs of non-revocation. It is also important to note that actors with public DIDs are not necessarily issuers, instead they may have a public DID to act in a governance role responsible for endorsing the transactions of others onto the ledger.

The number of ledger transactions a credential ecosystem would be required to support is dependent on a number of factors; public DIDs, schema, credential issuers per schema and the revocability of each schema. Revocable credentials have the greatest impact on the ledger with transactions required to publish revocations of issued credentials. However, this analysis highlights that the number of transactions on the ledger is invariant to the number of verifiers and credential holders within the ecosystem; they have no ledger footprint. Furthermore, apart from revocations issuing a credential has a minimal impact on the ledger. Each issuer needs to author at most 4 transactions, a NYM, an ATTRIB, the SCHEMA and the CLAIM_DEF. In cases where they are issuing against multiple scheme, the cost is only an additional CLAIM_DEF transaction assuming the SCHEMA has already been authored.

The ledger is also important for the robustness of identification systems augmented using verifiable credentials. It acts as a highly available source of truth for issuer public keys and other necessary cryptographic information required to verify a credential presentation. While on its own the transaction data held within the ledger is not enough to judge the authority of a credential issuer, it provides key inputs for the design of mechanisms to support verifiers making these judgements. Specifically, the public DIDs stored within a ledger must be correlated to real-world actors whose authority to issue credentials can be judged against the context in which the credential is used. Such

inputs could be used to implement machine-readable governance, whereby a verifying agent first checks an issuing identifier against a list of trusted issuing identifiers for the identification system. Once this check has been satisfied the cryptographic material from the ledger should be resolved and used to verify the integrity of the presentation. Machine-readable governance is discussed in Section 3.3.4.

The requirement for all transactions to be signed by a DID with the role of Transaction Endorser further increases the robustness of these identification systems. Within high assurance identification systems such as healthcare, identifying the actors and their DIDs trusted to endorse further transactions onto the ledger provides the ability to limit bad actors fraudulently issuing credentials within the healthcare ecosystem. For example, a DID representing NHS Scotland could be the only NHS organisation in the role of Transaction Endorser written to the ledger. They would then author all DIDs for the subsequent NHS actors within the system, and endorse any of the transactions these DIDs subsequently wished to author. This enables NHS Scotland to essentially act as the root Certificate Authority for a subsection of the Scottish Healthcare ecosystem. Assuming verifiers had knowledge of NHS Scotland's DID, they could easily increase their confidence in the authority of a credential issuer by checking if their DID and CLAIM_DEF transactions were indeed endorsed by the NHS Scotland DID. The GMC might be another actor within the Scottish Healthcare ecosystem in control of a Transaction Endorser DID.

This produces a hierarchical relationship between transactions based on both authorship and endorsership that is clearly visible from transaction data taken from the Sovrin MainNet. Figure 7.2 visualises the transaction author relationships for a pilot project to issue NHS Staff Passport credentials currently being undertaken within NHS England by a consortium of partners including Truu.id and Evernym [426]. It shows a single Transaction Endorser authoring many NYMs, who each then go on to author a CLAIM_DEF that references the same SCHEMA transaction. Furthermore, it shows one additional NYM in the role of Endorser authoring a CLAIM_DEF for the same SCHEMA. This appears as an anomaly and highlights how using the ledger footprint in combination with domain-specific knowledge could help identify fraudulent issuers

and increase the assurance verifiers can place in credential presentations [42]. This analysis is further backed up by viewing the visualisation from the perspective of the Endorser of each transaction (see Figure 7.3). Logic for performing these additional checks programatically could be included within a verifiers application logic and combined with domain-specific governance information such as the identifiers for SCHEMA used within the identification system. Finally, it is important to point out that this does create centralisation and the potential for a single point of failure should a trusted Transaction Endorser behave maliciously, or have their keys compromised. This could be mitigated by requiring multiple endorsers of transactions before they are authorised and written to the ledger.

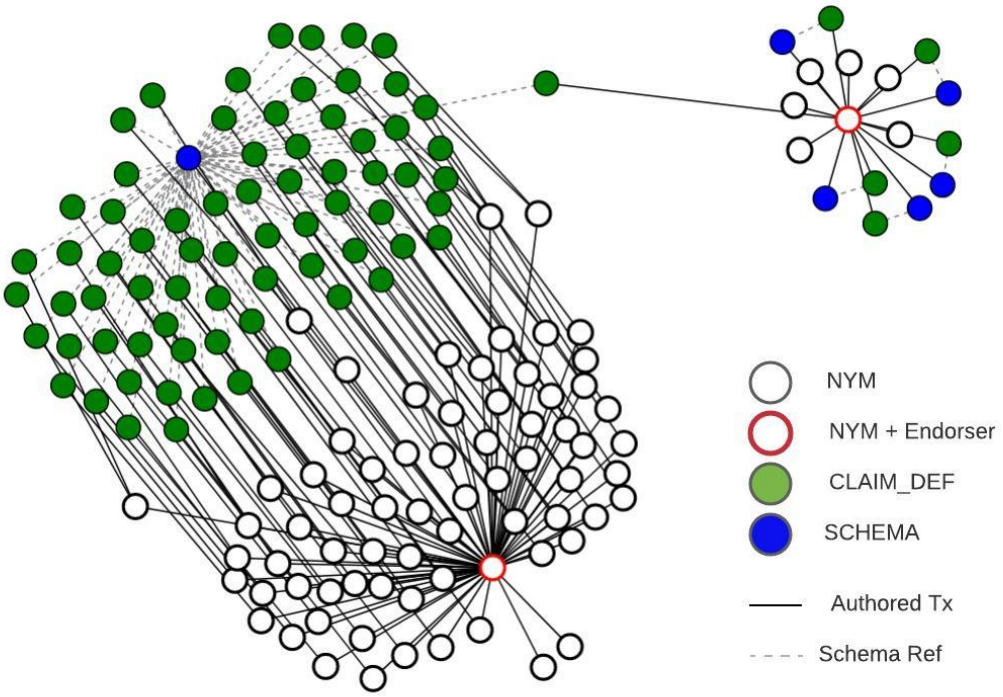

Figure 7.2: Transaction Author Ledger Footprint for an NHS Staff Passport Pilot Project (Sept 2020) [42]

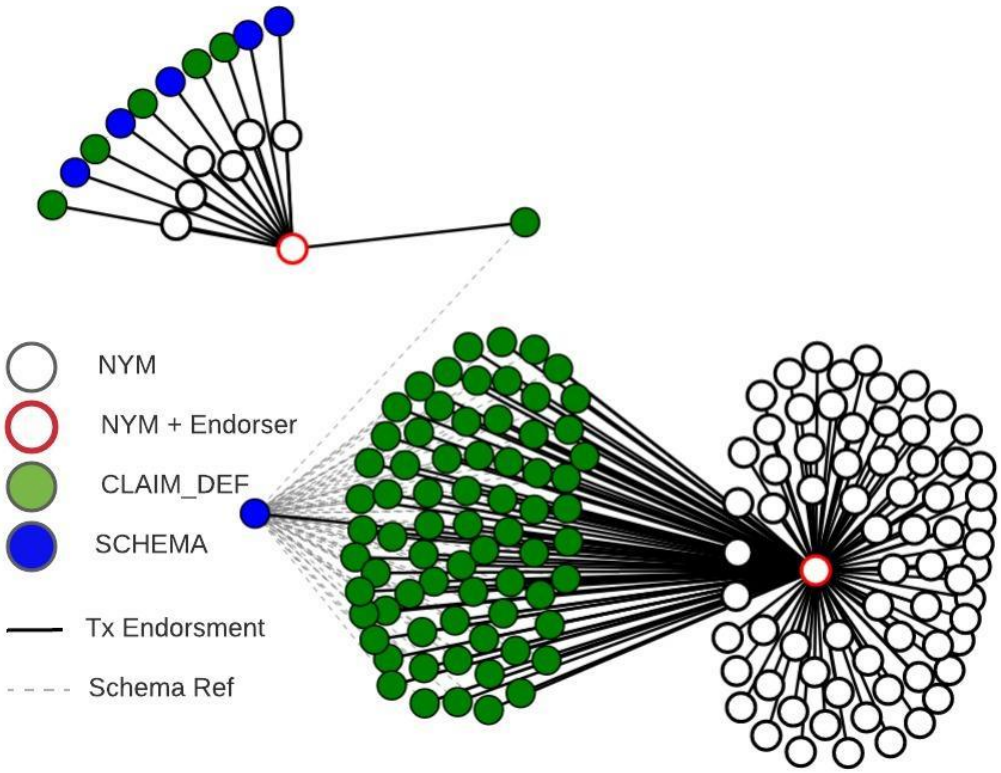

Figure 7.3: Transaction Endorser Ledger Footprint for an NHS Staff Passport Pilot Project (Sept 2020) [42]

It is important to recognise what properties a distributed ledger is providing to a credential-focused digital identification system such as the healthcare system studied in this thesis. It provides a highly available, append-only verifiable data storage that isn't dependent on any single centralised authority to maintain. Instead transactions, and their data, are included onto the ledger, based on a set of rules independently evaluated by actors operating nodes and participating in consensus with the aim of increasing the assurance that can be placed in this transaction data. Hyperledger Indy, as a public permissioned network, contains another chain of transaction data. All transactions signed and authored by a single entity represented by a DID stored on the ledger. This chain of provenance can be verified by any actor participating in the identification system by reading from the ledger. The implications of this are that even with malicious, colluding nodes, it would be infeasible to fraudulently impersonate an actor by including transactions into the ledger that that actor had not signed themselves.

The worst attack they would be able to mount is a denial of service whereby they fail to include legitimately authored transactions into the ledger. Such an attack would be noticed by the transaction author immediately, and since the actors running nodes within the network are identifiable and bound by a legal agreement, it should be easy to hold them to account - further disincentivising this attack vector [384].

7.2 Aries Agent Performance

The PoC identification system for the Scottish Healthcare ecosystem modelled a healthcare professional interacting with four other actors and receiving six Verifiable Credentials which they later presented to other actors to complete necessary identification processes. The design of the schema for these credentials was largely arbitrary, acknowledging that such a design should be produced in close consultation with key stakeholders of the ecosystem in order to meet their requirements. However, from this experience it was possible to identify a number of design decisions that are expected to be common across any ecosystem attempting to augment their identification processes with digital credentials. These were the number of attributes contained within a single credential, the expected size of an attribute's value (and by implication the credential payload) and whether or not the credential should support revocation. Furthermore, when requesting presentations of attributes from a credential holder, it is possible to require only a subset of attributes contained within a credential as well as requesting attributes from across multiple credentials. This section runs performance benchmarks against the ACA-Py agent framework in a production setting to assess the implications these design choices have on the performance of both the issue-credential and present-proof protocols. Both use interactive cryptographic protocols currently based on RSA cryptography [28, 29].

The experiments all followed the same generic setup. ACA-Py agents were deployed to separate Ubuntu 20.04 LTS virtual machines with dedicated 4GB Random Access Memory (RAM) running in the same region (UK). Each machine exposed the port associated with the HTTP endpoint for the agent in order for them to receive and

send Aries protocol messages to other agents. Agents were configured with flags such that they would automatically respond to protocol messages without the need for intervention where possible. To replicate as close to a production example as possible, agents were configured to use the Sovrin StagingNet and a revocation registry server hosted by the Government of British Columbia for development purposes. As with the PoC healthcare implementation, agents were controlled through a Jupyter notebook interface where the business logic for each experiment was implemented [43]. Each experiment consisted of a series of tests that varies a single aspect of the schema or attributes used within the protocol to judge its impact on performance. Each test is executed 100 times. Furthermore, each experiment is repeated for both Revocable and Non-Revocable credentials. Code for these experiments is open-source and available on Github [1].

The results for each experiment are presented in a series of graphs and tables. First, each test in an experiment is plotted on a box plot to show the spread of the data including outliers which are classified as such if they are below/above the lower/upper quartile by 1.5 times the interquartile range. All box plots for the tests of an experiment are displayed on a single discrete graph to show the change over time. Then a table presents the standard deviation and a trimmed mean for each test in seconds. The trimmed mean ignores results that are $+,-$ three standard deviations from the mean to exclude any outliers. The table also shows the relative trimmed mean against the result of the first test. This more clearly illustrates how each test impacts the result. Finally, the adjusted means for both Non-Revocable and Revocable versions of the experiments are plotted on the same continuous line graph.

[1] `https://github.com/wip-abramson/aries-jupyter-playground/tree/project/aries-performance`

7.2.1 Issue Credential

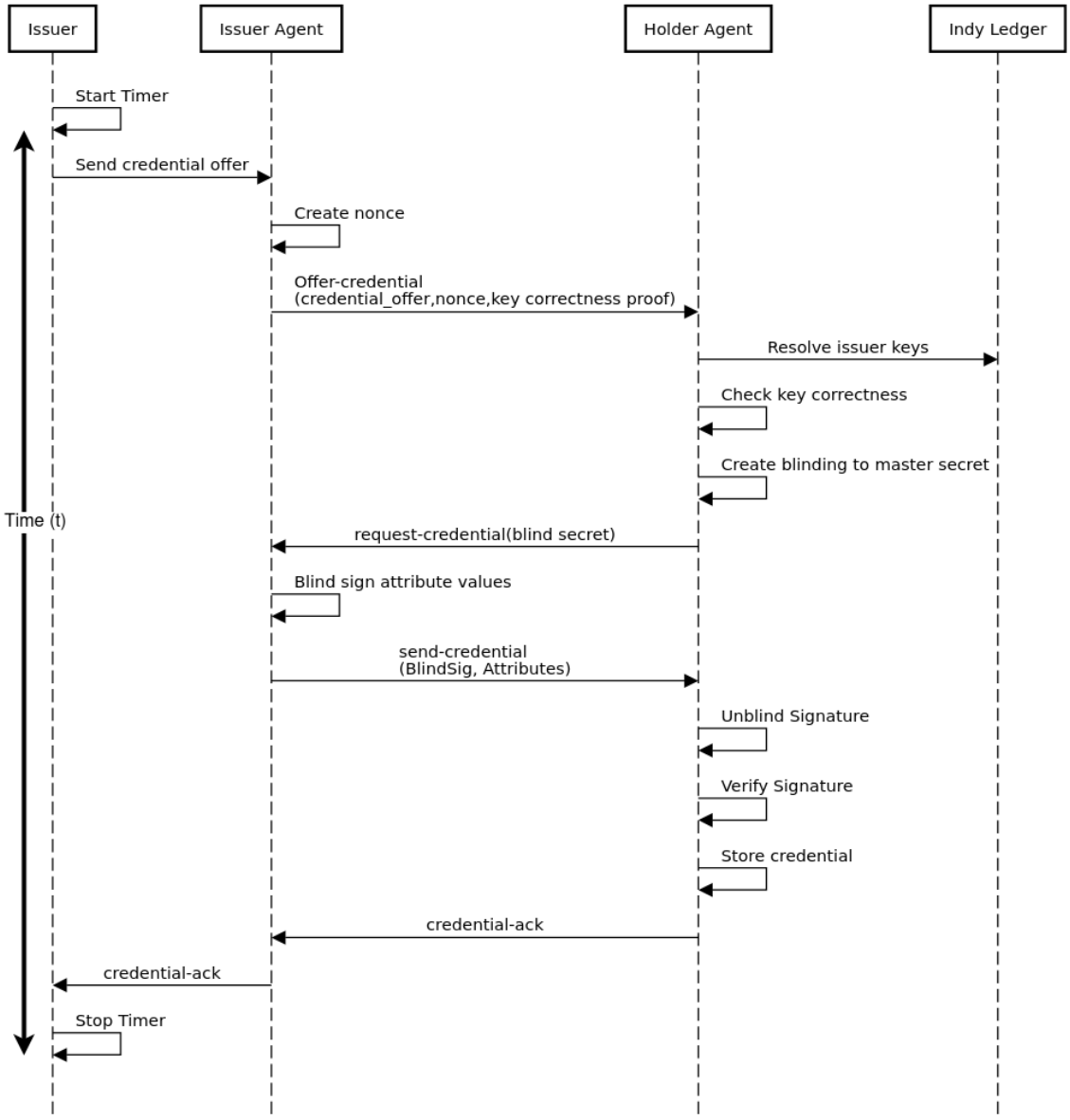

Figure 7.4: Interactions benchmarked between Issuer and Holder while engaging in the issue-credential protocol (Aries RFC 0036) [376]

The issue-credential protocol as defined in the Hyperledger Aries RFC 0036 [376] - involves a series of messages being sent between an issuer and a holder whereby they propose, negotiate, accept, issue and store a cryptographically signed verifiable credential. In the benchmarking tests ran against this protocol, timings were taken just before a credential offer was sent from the issuer up until the protocol transitioned to the final protocol state - *credential_acked*. This occurs once the holder has stored

the credential they have been issued and acknowledges this through a message sent to the issuer. As agents had been configured to automatically respond to messages, this time represents a minimum time to complete the issuance protocol in a deployment closely replicating a production environment, including the cryptographic signatures, message encryption, ledger reads and agent communication (see Figure 7.4). SCHEMA and CLAIM_DEF transactions were authored to the ledger during configuration, timing for this was not considered as an issuer is only likely to author a limited number during an initialisation phase (as discussed in Section 7.1). For each schema used in the experiment, two CLAIM_DEF transactions were authored, one that supported revocable credentials using CKS signatures and revocation registries [302] and one that did not support cryptographic revocation.

Two separate experiments were used to benchmark the issue credential protocol under varying conditions realistic for issuers. The first experiment aimed to understand how varying the credential size affected credential issuance times. A credential with a single attribute was issued, the attribute size was varied from 128 B to 256 KB. While it is possible to issue larger credentials, the experiment ran into a limitation due to the HTTP server implementation used setting a max size for requests by default. This was to do with the ACA-Py implementation and interface with the agent's business logic and not with the underlying protocols and agent to agent communication. The code could be adapted to increase the size to support megabytes of data, however it was judged that the range was adequate to show the impact of issuing credentials with larger amounts of data. Attribute size is directly relevant to the Scottish Healthcare ecosystem PoC due to the intention for the GMC Licence to contain a Base64 encoded image string, so understanding the impact of attribute size on the performance of the protocols is important.

Box plots showing the spread of the results for each test in the experiment are shown in Figure 7.5, with tests for non-revocable credentials shown on the left and revocable ones on the right. Issuance times for both credential types follow a similar pattern, increasing as the size of the credential increases. This is emphasised clearly in the trimmed mean values shown in Table 7.1 and plotted on a line graph in Figure 7.6.

After 1kB credential size, results show that issuance times increase at a steady rate across both credential types as you increase the size of the credential. For credential sizes less than 1kB, variance in issuance time appears to be negligible. Finally, the box plots in Figure 7.5 show a substantial number of outliers, indicating that the mean is skewed high. These have been removed by considering a trimmed mean, which has been plotted on the line graph in Figure 7.6. This graph shows that both revocable and non-revocable credential issuance times increase at the same rate as the attribute size increases, however revocable issuance times appear to be larger by a fixed amount (0.6s) across all tests in the experiment giving an indication of the cost of the cryptographic operations required to support revocation. Results from Veseli et al who developed a framework to benchmark the cryptographic protocols also confirmed this fixed cost of revocation [335].

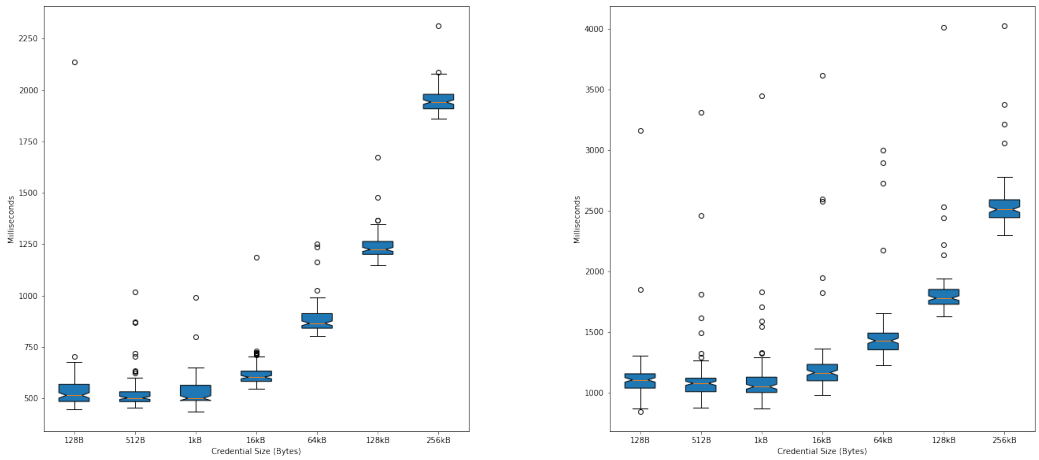

Figure 7.5: Box Plots of Times to Issue Non-Revocable (left) and Revocable (right) Credentials with Varying Attribute Size

In addition to changing the size of the data contained within the credential, issuers can issue credentials against schema containing any number of attributes. Each attribute can then represent a different value within the credential and a holder can later choose to selectively disclose any subset of the attributes within credentials they hold. For example, in the Scottish Healthcare PoC the GMC Licence schema contained seven distinct attributes including name, licence number and expiry date. Furthermore, the

Credential Size			128B	512B	1kB	16kB	64kB	128kB	256kB
Relative Size			1	4	8	128	512	1024	2048
Time	Non Revocable	Relative Mean	1	0.97	0.98	1.15	1.65	2.32	3.67
		Trimmed Mean (ms)	532	515	521	614	876	1233	1949
		Std (σ)	169	83	72	72	74	69	61
	Revocable	Relative Mean	1	0.99	0.99	1.07	1.30	1.65	2.30
		Trimmed Mean (ms)	1097	1085	1083	1176	1430	1805	2519
		Std (σ)	235	290	281	340	273	256	218

Table 7.1: Trimmed Mean Issuance Times for Credentials with Varying Attribute Size

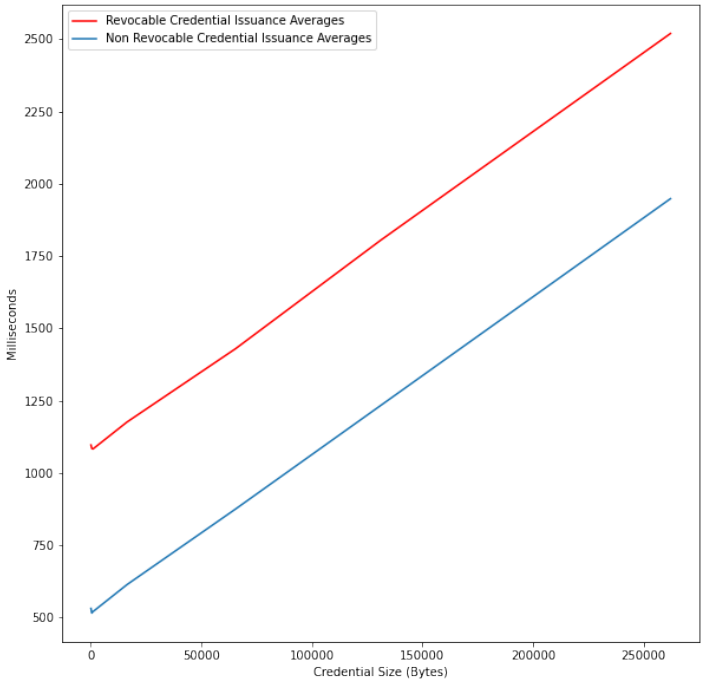

Figure 7.6: Trimmed Mean Issuance Times for Credentials with Varying Attribute Size

Staff Passporting pilot project in NHS England contained 36 attributes in a single credential schema [426]. As each attribute had to be individually signed by the protocol, the second issuance experiment aimed to determine how varying the number of attributes within a schema affected credential issuance times. Tests were run for credentials with 1,5,10,20,50 and 100 attributes and each attribute was populated with a fixed size 32 Byte value.

The box plots in Figure 7.7 show the spread of the results for each test. While there

is a small trend upwards as attribute numbers are increased this is not as pronounced as in Figure 7.5. Again, there are a substantial number of outliers skewed high, with some iterations taking around two seconds higher than the reach of the whiskers of the plots. Table 7.2 shows the standard deviation and trimmed mean results for each test in this experiment. The relative mean in this table highlights that even 100 times as many attributes in a credential does not lead to double the issuance time. The line plot of the trimmed mean test results for both revocable and non-revocable experiments in Figure 7.8 indicates a similar fixed cost to revocation independent of the number of attributes within the credential as was shown in Figure 7.6. It is also important to point out that a cause of the increase in issuance times must be in part attributed to the increase in credential size; a credential with one 32 B attribute is smaller than one with 100 32 B attributes.

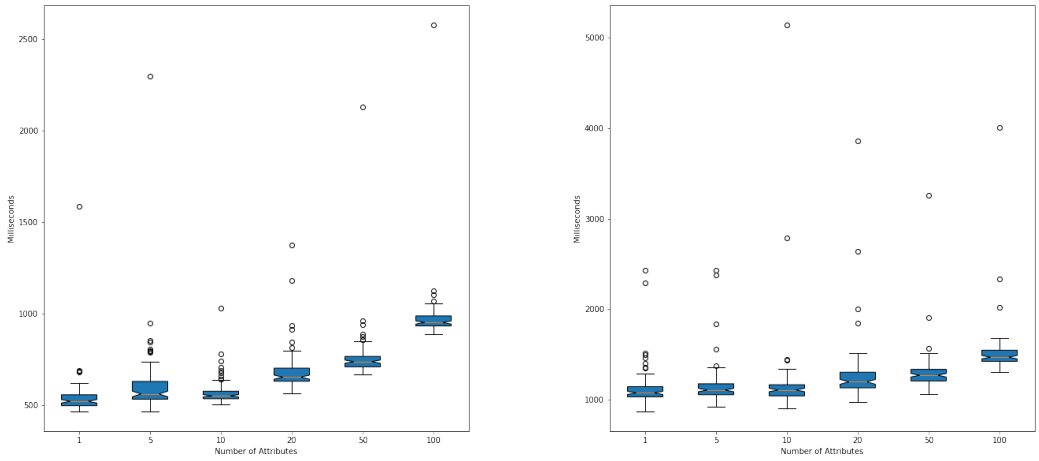

Figure 7.7: Box Plots of Timed to Issue Non-Revocable (left) and Revocable (right) Credentials with Varying Attribute Number

The results from both experiments show that the most significant cost associated with credential issuance times is enabling revocation, which adds a constant fixed cost to issuance for credentials of any size or number of attributes. This makes sense because of the additional cryptographic operations that need to take place [302, 335]. Results also indicate that credential size has a larger impact on issuance times than the number of attributes, although as the attributes within a credential schema increases

Number of Attributes			1	5	10	20	50	100
Time	Non Revocable	Relative Mean	1	1.11	1.06	1.26	1.41	1.81
		Trimmed Mean (ms)	531	592	563	671	747	962
		Std (σ)	115	191	68	108	148	166
	Revocable	Relative Mean	1	1.02	1.01	1.10	1.15	1.34
		Trimmed Mean (ms)	1099	1121	1113	1224	1276	1494
		Std (σ)	212	217	444	333	230	282

Table 7.2: Trimmed Mean Issuance Times for Credentials with Varying Attribute Number

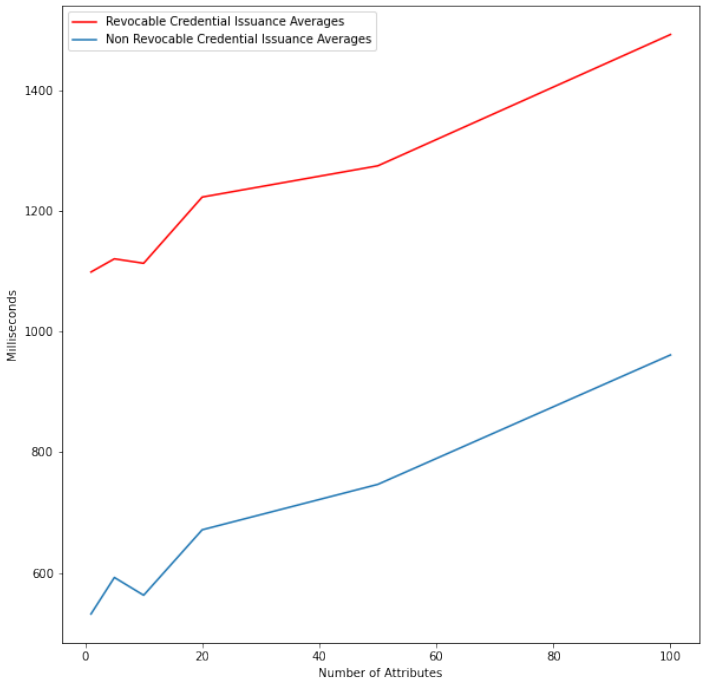

Figure 7.8: Trimmed Mean Issuance Times for Credentials with Varying Attribute Number

the size of the credential issued against that schema is also likely to increase. Limiting credential sizes to 128 kB gives reasonable issuance times, 1.1s for non-revocable credentials and 1.7s for revocable, while providing plenty of space for a biometric such as a photograph as with the GMC Licence in Section 6.3.3. Different use cases will have different performance demands on this technology, however in many use cases the performance focus is likely to be on the presentation of proof that certain attributes were issued within specific credentials. As this is an action that is likely to occur in an order of magnitude more often than issuance. Again the GMC Licence from our PoC is

instructive here; this credential is likely to be issued once every five years in line with when the licensed doctor must re-validate. However, it could conceivably be presented hundreds of times throughout this period, depending on which identification processes this digitally augmented licence is integrated into.

7.2.2 Present Proof

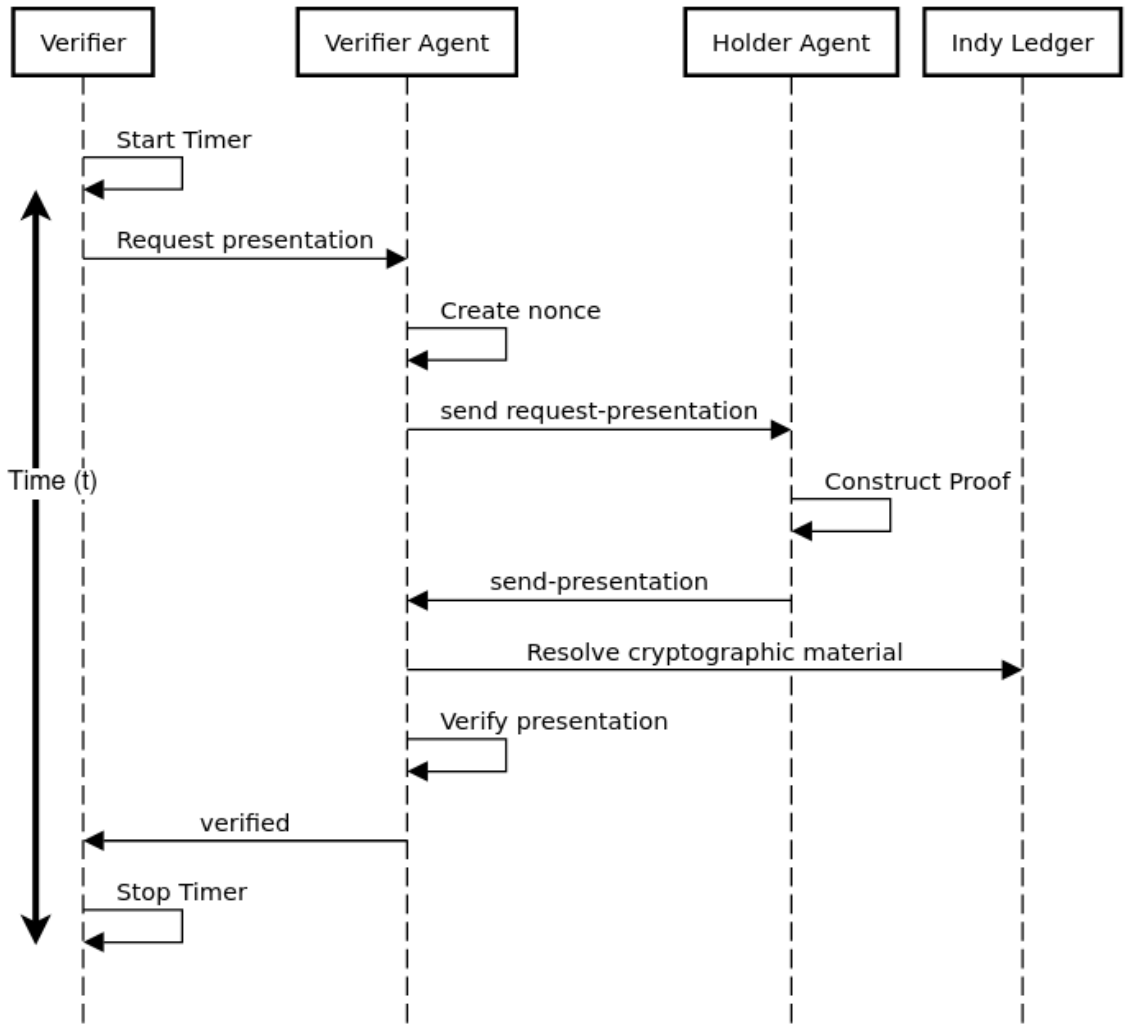

Figure 7.9: Interactions benchmarked between a Verifier and Holder while engaging in the present-proof protocol (Aries RFC 0037) [427]

The present-proof protocol specifies an interaction between a holder and a verifier, such that the verifier can request a proof presentation of a set of attributes under certain constraints and the holder can respond with the appropriate attribute values and a zero-knowledge proof attesting to their integrity and authenticity. The verifier is able to check this attestation using the public key material for the issuer stored on an Indy ledger. The

protocol was timed from the point the verifier requested a proof presentation from the holder, up until they verified the holder's presentation (see Figure 7.9). This protocol operationalises the cryptographic protocol discussed in Section 4.2.4. Again all ACA-Py agents were configured to automatically progress the protocol state to completion. The holder, a separate agent to the holder used for the issue credential experiments, was issued the relevant credentials required in order for it to successfully respond to the proof requests used by the verifier as part of the experiments. The verifier was also a separate agent from the issuer with knowledge of the relevant identifiers required to request proofs from the holder, this was done to reflect a likely scenario where the issuer is not the same entity as the verifier.

Four separate experiments were run, each time varying a different aspect of the presentation while keeping the rest fixed in order to determine how much that particular aspect affected the performance of the overall protocol. The first experiment varied the size of the attribute presented from 128B to 258kB. The attribute always originated from a credential with a single attribute schema. Both revocable and non-revocable credentials were tested and all presentation requests stated that the credentials must not be revoked.

The box plots showing the spread of the 100 results for each test are shown in Figure 7.10. These results were then used to calculate the standard deviation and using this, the trimmed mean for each test in the experiment; Table 7.3 shows these processed results. For non-revocable credentials, increasing the size from 128 B to 256 KB, an increase of 2048 times, causes the proof presentation times to more than triple. With revocable credentials the impact is only 1.5 times slower for the same range of attribute sizes. However, the line graph plotting the trimmed mean values for each test in the revocable and non-revocable versions of this experiment show an equivalent rate of change across both experiments (See Figure 7.11). Revocable proof presentation experiments just appear offset by around 1 second, which, as in the issuance experiments, can be attributed to the fixed cost of producing a valid cryptographic proof of non-revocation [335].

The next experiment tested the impact on presentation times of the number of

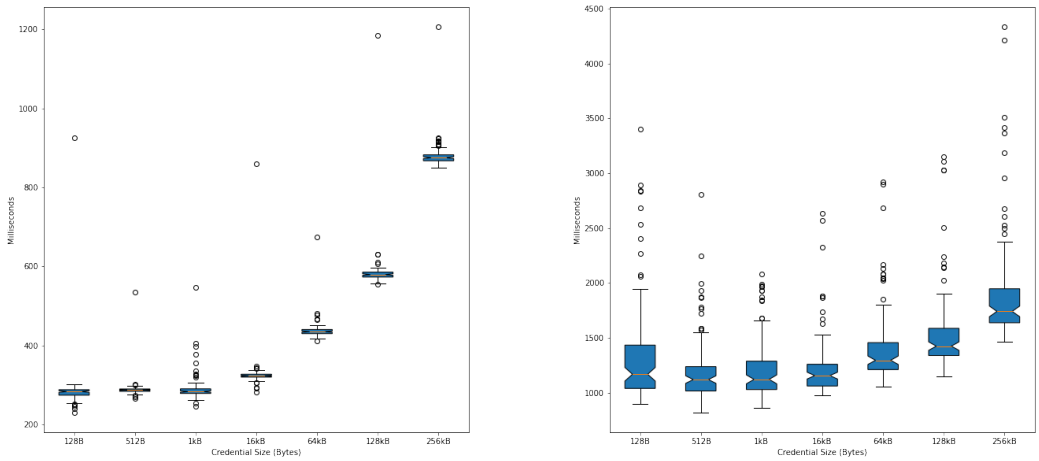

Figure 7.10: Box Plots of Times to Present Attribute from Non-Revocable (left) and Revocable (right) Credentials with Varying Attribute Size

Credential Size			128B	512B	1kB	16kB	64kB	128kB	256kB
Relative Size			1	4	8	128	512	1024	2048
Time	Non Revocable	Relative Mean	1	1.03	1.03	1.15	1.56	2.07	3.14
		Trimmed Mean (ms)	280	287	287	323	436	580	878
		Std (σ)	65	25	35	54	26	61	36
	Revocable	Relative Mean	1	0.93	0.95	0.94	1.08	1.17	1.47
		Trimmed Mean (ms)	1276	1181	1216	1197	1374	1488	1881
		Std (σ)	488	308	291	295	360	392	530

Table 7.3: Trimmed Mean Presentation Times for Credentials with Varying Attribute Size

attributes within a credential schema. A single, 32 B attribute was requested from the holder for each test but the schema the attribute was requested from was changed. Schema containing 1,5,10,20,50 and 100 attributes were tested. The experiment is aimed to evaluate the performance cost of designing schema with a large number of attributes, even if the holder is only ever requested to present proof of a subset of these attributes.

The spread of the results for each test from both revocable and non-revocable experiments are show in Figure 7.12. As with most experiments in this analysis, outliers are skewed high. These have been removed from the mean, and the trimmed mean for each test is shown in Table 7.4. Trimmed means for both revocable and non-revocable experiment results have then been plotted on the line graph in Figure 7.13. Results

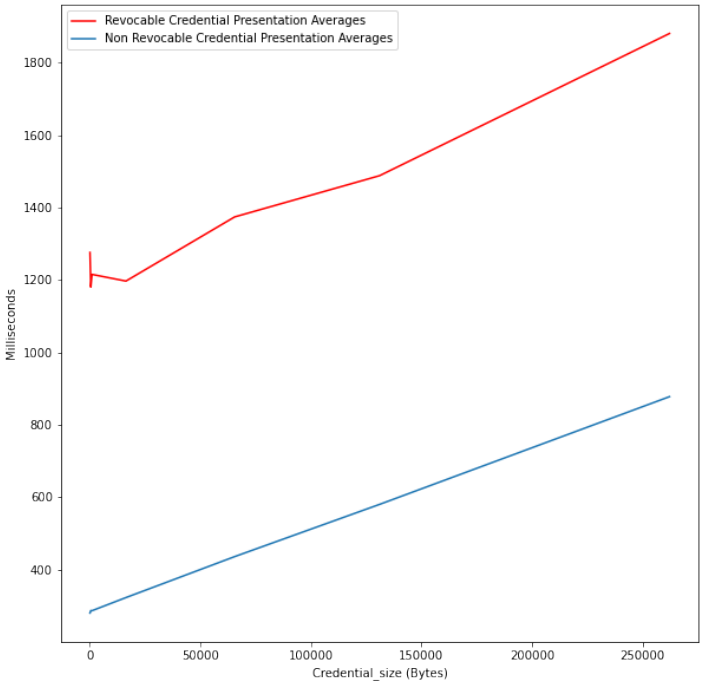

Figure 7.11: Trimmed Mean Presentation Times for Credentials with Varying Attribute Size

indicate that the number of credentials within a schema, whether disclosed or not, will impact the proof presentation times, schema with 100 attributes in comparison to 1 are three times slower to produce presentations from despite in both instances only a single attribute was disclosed from the presentation. Even a schema containing 20 attributes almost doubles the proof generation time. These relative increases are less strong within the revocable experiment, which again can be attributed to the constant offset from generating the proof of non-revocation. For instance, under the environment used for these experiments, the presentation of proof of a single attribute from a schema with only one attribute takes on average 308 ms for non-revocable credentials whereas, it takes 1068 ms for a revocable credential, almost 3.5 times slower. Again, from the line graph in Figure 7.13, this appears as a constant cost for revocation, invariant to the varying number of attributes within the schema.

The next experiment varied the number of attributes disclosed from the same 20 attribute credential. This tested whether requesting more attributes be disclosed

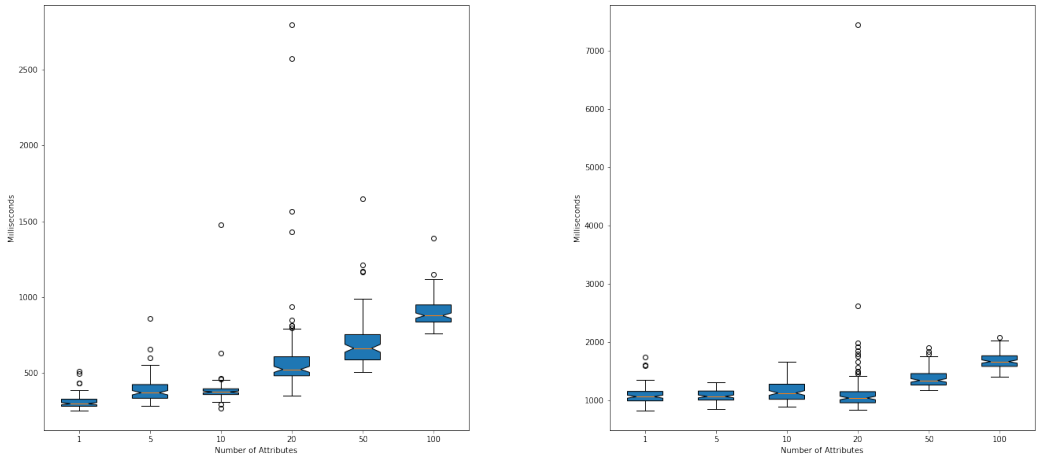

Figure 7.12: Box Plots of Times to Present a Single Attribute from Non-Revocable (left) and Revocable (right) Credentials with Varying Attribute Number in Schema

Schema Attributes			1	5	10	20	50	100
Time	Non Revocable	Relative Mean	1	1.24	1.23	1.83	2.22	2.92
		Trimmed Mean (ms)	308	381	380	564	684	898
		Std (σ)	46	86	117	344	166	96
	Revocable	Relative Mean	1	1.01	1.07	1.04	1.29	1.57
		Trimmed Mean (ms)	1068	1083	1140	1114	1377	1676
		Std (σ)	149	100	165	689	155	139

Table 7.4: Trimmed Mean Presentation Times for a Single Attribute Presented from a Credential with Varying Attribute Number in Schema

within a single proof presentation had any impact on the protocol execution time. The number of attributes disclosed were 1,2,5,10 and 20. Again all attribute values were fixed 32 B strings and the same experiment was run for both revocable and non-revocable credentials.

The box plots in Figure 7.14 shows that the revocable credential experiment had a wider spread, with many outliers skewed high. This is further confirmed by comparing the standard deviation of the hundred data points for each test between the revocable and non-revocable experiments (see Table 7.5). The relative mean results for each test Table 7.5 show that increasing the number of attributes disclosed within a presentation has a substantial impact on the protocol; when all 20 attributes were requested to be disclosed within the presentation, the execution time was 2.35 times larger for non-

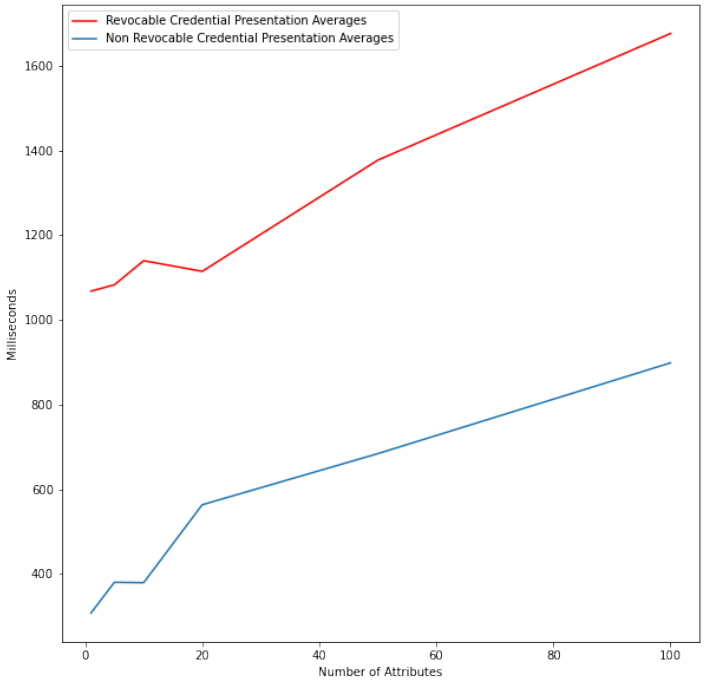

Figure 7.13: Trimmed Mean Presentation Times for a Single Attribute Presented from a Credential with Varying Attribute Number in Schema

revocable credentials (and 1.4 times for revocable) than when only a single attribute was requested. Despite these apparent differences in relative impact on execution times, the line graph of trimmed means in Figure 7.15 shows that the effect of varying the number of attributes disclosed, as indicated by the gradient of the plotted lines, is approximately equivalent. By calculating the difference in milliseconds between the test that disclosed just one attribute and the one that disclosed all 20 for both experiments we get a change of 507 ms for the non-revocable experiment and 514 ms for the revocable one, further confirming this analysis.

The final experiment sought to understand how changing the number of credentials used to construct a proof presentation affected presentation protocol execution time. Under the HVIEP, it is possible to construct a single presentation of attributes from multiple distinct credentials, the includes proving a common link secret was signed within all credentials (see Camenisch and Lysyanskaya [28]). For example, a medical professional might prove their GMC number and photograph from their GMC Licence

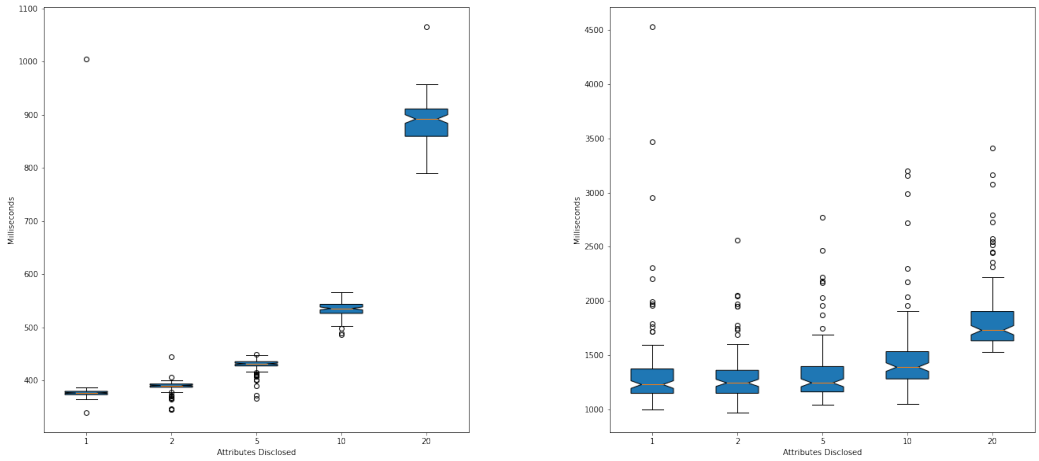

Figure 7.14: Box Plots of Times to Disclose a Varying Number of Attributes from a Non-Revocable (left) and Revocable (right) Credential

Attributes Disclosed			1	2	5	10	20
Time	Non Revocable	Relative Mean	1	1.03	1.14	1.42	2.35
		Trimmed Mean (ms)	376	389	430	535	883
		Std (σ)	63	12	13	13	37
	Revocable	Relative Mean	1	1	1.02	1.10	1.40
		Trimmed Mean (ms)	1295	1290	1315	1428	1808
		Std (σ)	483	261	302	385	370

Table 7.5: Trimmed Mean Presentation Times for Varying Number of Attributes Disclosed

credential alongside a proof of immunity, right to work and completion of basic training all from their own distinct credentials within a single presentation. A presentation of five attributes was requested during each test, with the number of credentials these attributes were requested from varying from one to five credentials. All credentials had five attributes in their schema, and all attributes contained a 32 B value. This design ensures that the impact on protocol execution times can be confidently attributed to the changing number of credentials used to construct the proof presentation.

The box plots showing the spread of the results from each test in Figure 7.16 again show that revocable credentials have a much greater spread with many outliers skewed high. As the number of revocable credentials included in the proof presentation increases, so does the spread of the results, with the standard deviation changing from 392 ms for a single revocable credential up to 984 ms for five revocable credentials

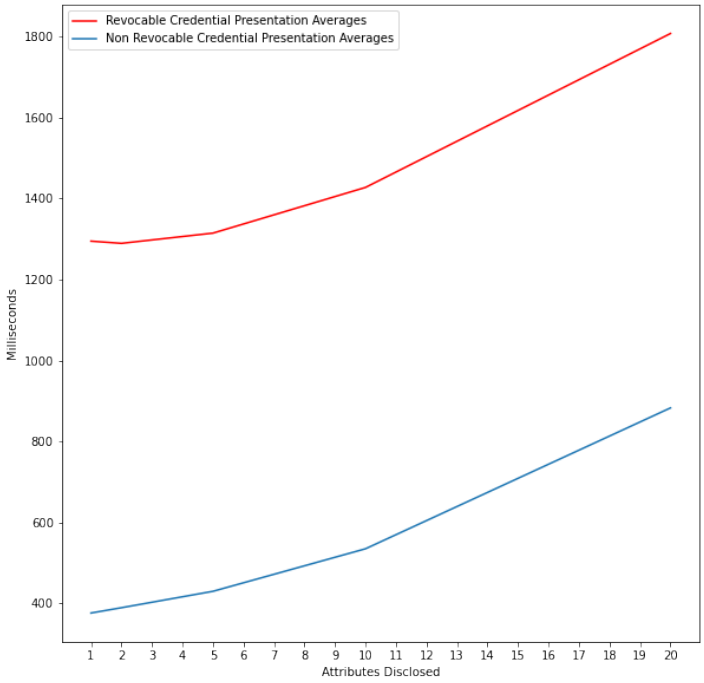

Figure 7.15: Trimmed Mean Presentation Times for Varying Number of Attributes Disclosed

(see Table 7.6). This highlights that the calculation of the cryptographic proof of non-revocation is non-deterministic and, at times, can take significantly longer than the average. The trimmed and relative mean values for these experiments - as shown in Table 7.6 and plotted on the graph in Figure 7.17 - are the most striking of the proof presentation experiments. Requiring proof of attributes from multiple revocable credentials has a significant impact on the protocol execution times, with each additional credential increasing the time by approximately 75 percent relative to a proof presentation of attributes from a single revocable credential. For proof presentations containing attributes from multiple non-revocable credentials are less costly, but still increase the time by around 22 percent per credential.

7.2.3 Discussion

The results from the experiments on the performance of both the credential issuance and proof presentation protocols as currently implemented within the open-source

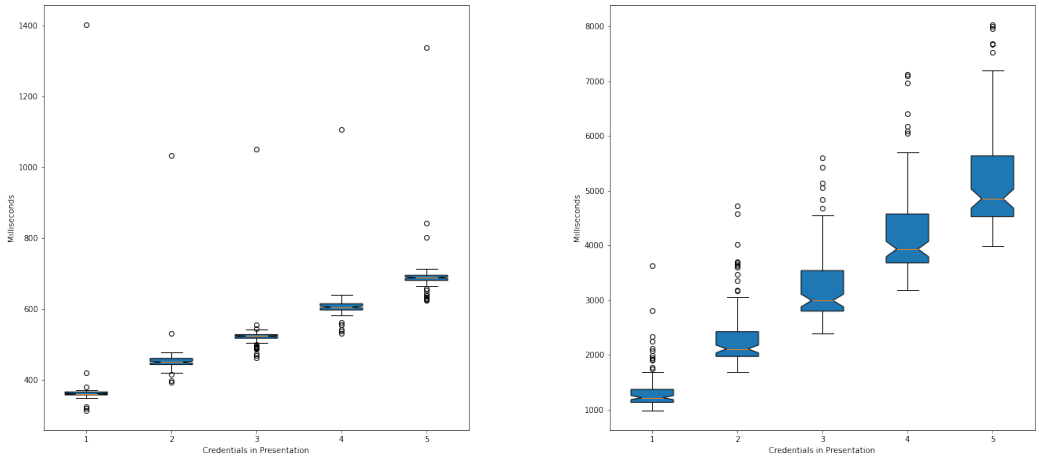

Figure 7.16: Box Plots of Times to Present 5 Attributes from Varying Number of Non-Revocable (left) and Revocable (right) Credentials

Credentials in Presentation			1	2	3	4	5
Time	Non Revocable	Relative Mean	1	1.25	1.44	1.68	1.91
		Trimmed Mean (ms)	360	450	520	603	686
		Std (σ)	104	60	55	53	70
	Revocable	Relative Mean	1	1.75	2.43	3.16	4
		Trimmed Mean (ms)	1312	2296	3193	4146	5250
		Std (σ)	392	613	665	845	984

Table 7.6: Trimmed Mean Presentation Times for 5 Attributes from a Varying Number of Credentials

ACA-Py code base demonstrates that the privacy-preserving cryptographic protocols integrated into the software development framework of the HVIEP are practical. These experiments were undertaken in a realistic production setting, with each actor running on a separate VM, communicating over HTTP and interacting with a public ledger - the Sovrin StagingNet. Although it is important to recognise that different deployments, using different infrastructure, with different resource constraints, would produce different results. For example, implementations that use agents running on mobile phones or IoT devices. However, the relative impact from varying each of the aspects of the protocol are expected to be consistent across infrastructure.

Each experiment isolates a particular property of the protocol that is likely to vary use case to use case whilst keeping all other properties the same. This was intended to

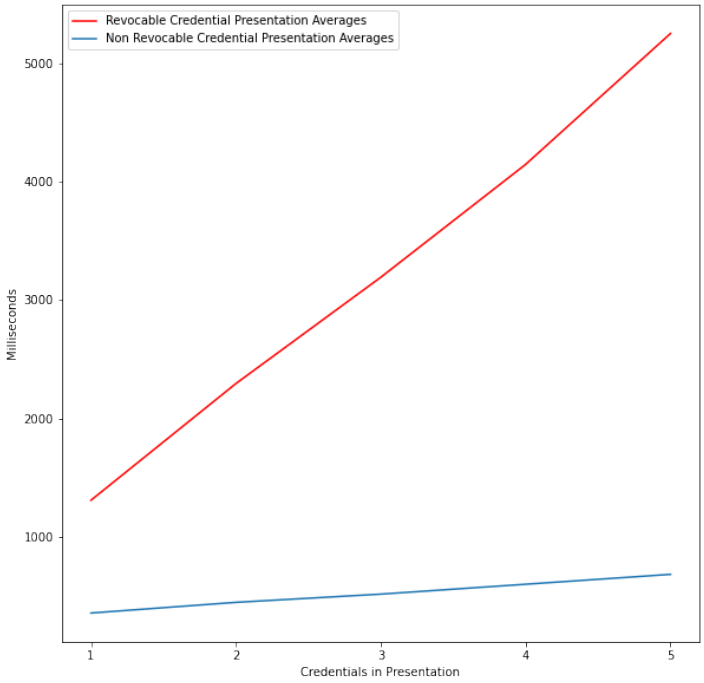

Figure 7.17: Trimmed Mean Presentation Times for 5 Attributes from a Varying Number of Credentials

evaluate the impact of that property independently of all other aspects of the protocol that could affect its performance. The results show how designing the credentials used within a system can impact the performance of the system. They highlight the importance of understanding the credentials issued and attributes they contain, from the perspective of how they will be presented. In particular, they show the cost of presenting revocable credentials especially if multiple revocable credentials might be combined in a single credential (see Figure 7.17). While these experiments aimed to vary these properties independently, within a realistic credential system, they are likely to be mutually reinforcing and hence further impact performance times. An example of this would be the doctor presenting their GMC Number and photograph from a revocable credential for a 20 attribute schema. Here the size of the photograph, the number of attributes within the schema and the number of attributes presented all increase the performance of the proof presentation protocol. Results from these experiments show how careful design of the schema used to issue and verify credentials within an

identification system can bring significant performance gains.

Different use cases will have different requirements for the performance of the interactions within their identification system, both in terms of credential issuance and proof presentation and verification. However, when considering these interactions from a socio-technical perspective with human identification processes augmented, but not replaced, by technology, the mean protocol execution times appear more than reasonable. A healthcare professional would be able to join a new hospital, anywhere within Scotland, and as part of their onboarding process with the administration team they could identify themselves by presenting proof of their GMC Licence including the photograph to strongly identify them as the holder of that licence and any number of other verifiable pieces of evidence required. The technical execution times of the protocols evaluated in this section are unlikely to be a limiting factor in these interactions, as they did not take into account the human interactions that would typically initiate, respond to and complete these credential exchanges. Furthermore, these technologically augmented identification processes could reduce these interaction times from hours or days down to minutes. In addition to this, there would be a reduction in administrative burden that could be achieved through the simplified process by which a healthcare professional can store, manage and maintain the professional evidence they need to interact within the SHE.

Finally, to emphasise the earlier discussion about the ledger footprint (see section 7.1), during these experiments the issuer issued 2,636 credentials and the verifier verified 4,600 presentations. The ledger footprint to support these interactions required only 129 transactions authored only by the issuer represented as DID Rip3JjGGzq2kmk4sxgqhtL [2]. These can be seen on the IndyScan transaction explorer. Furthermore, 72 of these transactions were to support revocation, 18 authored SCHEMA, 36 were CLAIM_DEF transactions, one was a NYM and two were ATTRIB transactions. This shows that credential issuance and verification is a peer-to-peer interaction that depends on the distributed ledger primarily as a source of integrity-assured information. Ledger writes

[2] `https://indyscan.io/txs/SOVRIN_STAGINGNET/domain?page=1&pageSize=50&filterTxNames=[]&sortFromRecent=true&search=Rip3JjGGzq2kmk4sxgqhtL`

are only required by to provision or manage an issuers public information, which scales independently from the number of issuance or presentation interactions that depend on this information.

7.3 Cryptographic Protocols

The technology stack used for the PoC and the Aries performance evaluations (Section 7.2) used a signature scheme with efficient protocols based on RSA cryptography, CL-RSA, published in 2002 [29]. The signature scheme remains provably secure under the Strong-RSA assumption and Decisional Diffie-Hellman assumption under the Random Oracle Model, well established cryptographic assumptions that provide confidence that the scheme is existentially unforgeable [290, 294, 288, 287]. As processing power continues to increase, the size of the RSA modulus required to ensure that an instance of the signature scheme is computationally infeasible to crack must also increase. The PoC used RSA with a 2048-bit modulus, which NIST state provides the equivalent of 112-bit security [428]. The size of this modulus impacts the speed of the cryptographic operations as well as the size of the payloads that need to be exchanged between interacting entities. For example, an issuer and a holder engaging in credential issuance (see Figure 7.4).

Since the early 2000s many cryptographic protocols been reproduced within elliptic curve or pairing cryptography mathematical settings. Within the context of cryptographic credentials, BBS+ (a group signature scheme with efficient protocols) was first proposed building on BLS signatures and then shown to have efficient protocols before being proven secure in the more efficient Type-3 pairing setting [344, 339, 214, 30]. Functionally, it is equivalent to the CL-RSA signature scheme however as it has been realised under a pairing-based mathematical setting, it is more efficient and is proven secure under different assumptions, specifically the q-Strong Diffie-Hellman assumption [346, 30]. BBS+ signatures are currently in the process of being adopted throughout the decentralised identity space [215]. This brings the privacy-preserving benefits specifically designed into a signature scheme with efficient protocols for cre-

dential mechanisms to decentralised identity systems, whilst reducing the performance costs associated with such a scheme realised in the RSA setting.

This section benchmarks the KeyGen, Credential Issuance and Presentation protocols for both CL-RSA and BBS+ signature schemes using the Hyperledger Ursa Rust implementations and the Criterion statistical benchmarking Rust crate [429]. One hundred samples were taken for each experiment after a three second warm-up. The CL-RSA signature scheme benchmarked used a 2048-bit modulus as in the PoC, and the BBS+ signature scheme benchmarked was instantiated in the BLS12-381 curve. As with the ACA-Py performance tests, benchmarks are taken on a dedicated 4GB Random Access Memory (RAM) virtual machine. The difference is that these benchmarks focus on pure cryptographic implementations without the communication complexity associated with separate agents interacting over a secure messaging channel.

7.3.1 Key Generation

Key generation is an operation that must be completed before signatures can be produced and verified. It is an algorithm that, when executed, generates a public and private key pair for either CL-RSA signatures or BBS+ signatures that can subsequently be used in signing and verification algorithms for their respective signature schemes. Both signature schemes allow signatures on an array of messages and support efficient privacy-preserving protocols, as opposed to the single message found in signature schemes such as Ed25519 [425].

The results in Table 7.7 clearly indicate the superior efficiency of the BBS+ key generation in comparison to CL-RSA with an average reduction in speed of over 99.9%. This is to be expected. The key generation algorithm for the BBS+ signature scheme involves selecting at random an integer $x \leftarrow \mathbb{Z}$, calculating $w = g_2^x$ and selecting $L+1$ generators at random from $\mathbb{G}_1$, where L is the number of messages being signed [30]. Whereas CL-RSA key generation is a non-deterministic algorithm in which a suitable RSA modulus n is selected, where n is the product of two safe primes $p = 2p' + 1$ and $q = 2q' + 1$ and both p' and q' are primes. Once n has been selected, a set of

$L+2$ quadratic residues modulo n are chosen uniformly at random [29]. It is the non-deterministic generation of the RSA modulus from safe primes, which must be selected and tested, that impacts the efficiency of the CL-RSA key generation algorithm as can be seen by the results in Table 7.7 in which the impact of varying the number of attributes is not clearly discernible. Furthermore, the spread of the results from the CL-RSA key generation algorithm is large, making confidence in the averages presented low. In contrast to BBS+ key generation results, which have a low spread of results and increasing the number of attributes clearly correlates to increased algorithm execution time. Generating a key for 100 attributes is more than eight times slower than for a single attribute with the BBS+ algorithm. However, when considering the concrete execution times under the virtual machine environment used for this experiment, it becomes apparent that this increase is insignificant to the performance cost. To further emphasise this point, key generation was run for a 1000 attribute BBS+ key pair resulting in an average time of only 128 ms. In comparison to the CL-RSA scheme, where even for a single attribute takes around the five second mark.

Messages			1	2	5	10	20	50	100
Time (ms)	CL-RSA	Relative Mean	1	1.06	1.20	0.88	1.11	1.06	1.43
		Mean	5089.4	5370.4	6121.3	4480.6	5674.2	5405.0	7259.9
		Std (σ)	4211.6	5667.7	4062.2	3506.7	3646.2	3855.2	3864.1
	BBS+	Relative Mean	1	1.12	1.41	1.88	2.47	4.71	8.35
		Mean	1.7	1.9	2.4	3.2	4.2	8.0	14.2
		Std (σ)	0.06	0.04	0.03	0.13	0.14	0.06	0.18
Percentage Decrease: CL-RSA → BBS+			99.97	99.96	99.96	99.92	99.93	99.85	99.80

Table 7.7: Comparison of CL-RSA and BBS+ Key Generation Times

BBS+ key sizes are smaller than CL-RSA, although the exact size is dependent on the concrete mathematical setting the protocols are constructed in. With CL-RSA realised under a 2048-bit RSA modulus, the key sizes are a 256 B (2048-bit) private key and 771 B + 257 B per attribute public key. In contrast BBS+ signatures under the BLS12-381 has a 48 B (384-bit) private key with a public key of 293 B + 97 B per attribute. Sizes are more than 2.5 times smaller in every instance. In addition to this, the BBS+ implementation in Hyperledger Ursa supports the deterministic creation of the $L+1$ $\mathbb{G}_1$ generators from w

(a $\mathbb{G}_2$ element) required for a to sign or verify L message signature. This reduces the public key to a fixed compressed size of 96 B. From this single deterministic key pair (x, w) the issuer is able to sign an arbitrary number of messages, such that any verifier with knowledge of only w can verify by deterministically generating the full public key for the number of messages that were signed. This innovation reduces the size of any ledger supporting the issuance of privacy-enhancing attribute-based credentials by storing the issuer's public keys. The requirement for CLAIM_DEF transactions, whereby issuers publish public keys for specific credential schema generated based on the number of attributes within that schema is no longer necessary from a performance standpoint. Although, this transaction serves an additional purpose as a public statement of intent such that verifiers can see which credential a particular issuer is issuing (see Figure 7.2) [42].

7.3.2 Credential Issuance

The cryptographic operations used during credential issuance were ran in a single benchmarking test. These include generating a nonce, verifying the key was correctly produced (for CL-RSA only), populating the attributes, creating a blind signature on the attributes and unblinding the signature (see Section 4.2.3). These tests do not take into account the different roles, issuer and holder, and the communication between them, instead they produce an overall benchmark for the cryptographic operations required to issue a credential. This is then used to compare the two signature schemes.

The first difference apparent from inspecting the code that implements these two protocols for benchmarking is that CL-RSA requires an additional proof to be verified. For security, CL-RSA requires that issuers generate an RSA modulus from the product of two safe primes, holders should check this is the case and Camenisch provided a protocol to create and verify this fact in zero-knowledge [29, 273]. BBS+ uses elliptic curve based pairing cryptography such that this operation is not required. This means for a CL-RSA based credential signature scheme with 2048-bit RSA modulus an additional 320 B + 288 B per attribute will need to be communicated for the key correctness proof.

Messages			1	2	5	10	20	50	100
Time (ms)	CL-RSA	Relative Mean	1	1.07	1.31	1.70	2.46	4.74	8.29
		Mean	141.7	151.9	185.4	241.1	348.8	671.5	1174.9
		Std (σ)	42.7	44.8	41.5	40.9	46.8	63.4	40.2
	BBS+	Relative Mean	1	1.07	1.25	1.50	2.03	3.10	5.19
		Mean	10.0	10.7	12.5	15.0	20.3	31.0	51.9
		Std (σ)	0.12	0.18	0.19	0.41	0.26	0.39	1.20
Percentage Decrease: CL-RSA → BBS+			92.9	92.8	93.3	93.8	94.2	95.4	95.6

Table 7.8: Comparison of CL-RSA and BBS+ Credential Issuance Times

7.3.3 Credential Presentation

Credential presentation using the BBS+ and CL-RSA signature schemes with efficient protocols involves using a signature to create a non-interactive proof of knowledge attesting to the integrity of a set of messages. This signature of knowledge is verifiable against the public issuing key. The cryptographic operations benchmarked include: constructing the zero-knowledge proof; converting it into a signature of knowledge using the Fiat-Shamir transformation [253]; and verifying the signature of knowledge against a set of message values (the disclosed attributes) and issuing public key. In this experiment all attributes within the credential were revealed in the presentation apart from the master secret.

The results in Table 7.9 show that BBS+ presentations are more performant than CL-RSA, although both signature schemes appear to be efficient enough to be considered practical for most use cases. Especially those involving face to face verification's, such as the identification system for the Scottish Healthcare ecosystem detailed in Chapter 6. Here the limiting factor is likely to be the human verification of attributes presented against other supporting evidence rather than the cryptographic verification benchmarked here. According to results from Ursa, the sizes for this proof object are 696 B + 74 B per attribute for CL-RSA and 312 B + 104 B per attribute for BBS+. So again, BBS+ is more efficient in terms of space, although the difference is less pronounced and, for credentials with a large number of attributes, a BBS+ proof could grow larger than CL-RSA.

Messages			1	2	5	10	20	50	100
Time (ms)	CL-RSA	Relative Mean	1	1.01	1.07	1.18	1.40	2.04	3.12
		Mean	42.0	42.5	45.1	49.7	58.6	85.6	131.1
		Std (σ)	1.01	0.73	0.70	0.54	0.65	0.64	1.22
	BBS+	Relative Mean	1	1.09	1.25	1.50	2.05	3.10	5.14
		Mean	17.7	19.3	22.1	26.6	36.3	54.8	91.0
		Std (σ)	0.37	0.14	0.32	0.52	1.09	0.87	1.07
Percentage Decrease: CL-RSA → BBS+			59.3	54.6	51.0	45.6	38.1	36.0	30.1

Table 7.9: Comparison of CL-RSA and BBS+ Credential Presentation Times

7.3.4 Discussion

This section has demonstrated the efficiency benefits, both in terms of time and space complexity, that can be gained by implementing a cryptographic protocol within a different mathematical setting. Specifically this section focused on a Signature Scheme with Efficient Protocols, first constructed in the RSA setting under the Strong RSA assumption and since realised in the newer pairing setting under the q-Strong Diffie Hellman assumption [28, 288, 346, 30]. This section has presented benchmarked performance results run against the open-source implementations of both of these protocols that can be found within the Hyperledger Ursa code base [3]. That these implementations exist, are open-source and written in Rust, a programming language that only realised a stable version in 2015, is strong evidence that theoretical cryptographic protocols are increasingly being implemented within software libraries and available to be integrated into technological artefacts. Cryptography is no longer only a science about theoretical possibilities for mediating information flows through mathematical transformations, these ideas are being picked up by engineers and translated into real-world software tools. The difference, in contrast to 2001, when Camenisch and Lysyanskaya first proposed CL-RSA as a practical provably-secure approach to realising a version of Chaum's credential mechanism is the number of years further into the information age humanity [28, 27]. Each of these years has increased humanity's collective experience in designing, implementing, applying and interfacing with software and cryptography in practice.

[3] `https://github.com/hyperledger/ursa`

7.4 Evaluation Limitations

The series of experiments presented in this chapter have a number of limitations that should be taken into account. First, it should be recognised that this evaluation focuses solely on the technical aspects of digital identification systems developed from the HVIEP using the ACA-Py codebase. While the results of these experiments increases the confidence that can be placed in the practicality of this architecture and software framework for use in production systems, they remain detached from the human context of any specific identification system. This thesis has argued throughout that digital technologies should not be considered in isolation, but as elements of social systems and the specific interactive settings in which they are applied. The evaluation in this chapter falls short of this, providing only a means for an implementer to appraise the HVIEP after its appropriateness and applicability to a social context has been justified. This important area of research is left to future work and a suggestion for how this might be achieved is provided in the following section.

The analysis of the transactional footprint authored to a Hyperledger Indy based ledger shown in Section 7.1 provides valuable insight into the performance cost incurred by these systems, as well as the usefulness of this information in providing trust assurance. However, only the footprint produced by a single identification system, albeit from a high assurance ecosystem in production on the Sovrin MainNet, was analysed. Further research is required to validate these results by contrasting them with the ledger footprints of other ecosystems. Additionally, these results could be contrasted with ecosystems that choose to root their identifiers to different distributed ledgers.

The performance of Aries agents provided a useful benchmark for the two common interactions within a credential-based identification system: credential issuance; and credential presentation. While the Aries agents used in these experiments were deployed to virtual machines in a production-like configuration, the experiments did not evaluate these agents at the scale of a production deployment. Rather than focusing on the implications agents processing multiple requests at once, the experiments looked at

single credential issuance or presentation interaction times. Furthermore, as mentioned previously, the experiments were set up to test two ACA-Py agents interacting without human intervention or input required. A more realistic scenario, especially considering the healthcare use case, would be a human administrator authorising different stages of these interactions and a healthcare professional accepting them on a mobile device.

7.5 Future Work: A Participatory Evaluation Engaging with Healthcare Professionals

The evaluation presented in this chapter has demonstrated the technical feasibility of augmenting the identification processes of healthcare professionals within the SHE through a series of experiments. However, these experiments are unable to determine if the proposed architecture fully meets the contextual requirements of a healthcare environment. How people are identified, the factors that shape trust and influence trustworthiness and the information flows deemed appropriate, are all aspects of an identification system that impact. These are perceived by the individuals fulfilling different roles throughout this system over time. A logical next step is thus to build on the analysis from the RCPE workshop in order to engage with a diverse set of stakeholders throughout the SHE, and leverage their direct first-hand experiences of these identification processes. This thesis proposes adopting a participatory evaluation approach, empowering stakeholders with ownership and active involvement in shaping the process of evaluation and indicators of success for a digital identification system within healthcare [430]. It is suggested that the environment in which the PoC was realised provides an ideal interface to facilitate these interactions, allowing a concrete demonstration of the technical possibility space for digital identification. This environment, as a tool to facilitate the evaluation process, could also be evaluated by the participants.

Within healthcare there are many stakeholders, each with a different set of needs, experiences, priorities and perspectives. For any technological solution to be successful within healthcare it needs *buy-in* from all the relevant stakeholders across an ecosystem,

including those whose voices which are not always heard [41]. The aim of a participatory evaluation approach is to bring in these stakeholders and make them full partners, who are able to influence the direction of the project and determine its indicators for success. The technical PoC demonstrates that it is possible to introduce digital credentials into the SHE, however there are many more stages of development that this PoC must iterate through before it could be adopted and integrated into the actual identification processes of healthcare professionals. The participatory evaluation approach provides an opportunity to continuously evaluate these iterations and to do so against criteria defined and refined by the stakeholders of the ecosystem.

The identified stakeholders within a healthcare environment listed in Table 7.10 are:

- Individuals being identified (medical students and junior doctors).
- Individuals from the institutions and organisations that both provide and require specific ID artefacts from healthcare professionals within identification processes.
- Institutions that define and regulate the identification system.

Participants from each of these stakeholder groups could be selected through the existing relationships established with the RCPE that ensures a diversity of representation across different group identities, such as gender, ethnicity and age [41].

A participatory evaluation process could initially involve running a series of semi-structured interviews with representatives from each of the stakeholder groups individually, in order to give an opportunity to speak openly from their unique perspective, and without being influenced by the power differentials between different stakeholder groups. The aim of these interviews would be to define the perceived purpose of a healthcare identification system and determine the indicators of impact and effectiveness against which this system could be evaluated, as well as identifying areas of concern and challenges that would need to be overcome. These interviews could also seek to understand the factors that influence trust for each of the different stakeholders, and how the different ID artefacts and identification processes fit into these factors.

Stakeholder	Description
Final Year Medical Students	These are individuals that are just about to complete their PMQ and are beginning to interact with the GMC and the NES in order to become licensed professionals and secure their first place of employment. They will have experienced healthcare settings through student placements only.
Junior Doctors	Recently qualified and licensed professionals that regularly transition between different places of employment on placements to gain experience.
GMC Staff	Employees at the GMC knowledgeable in the existing processes of identity verification and assurance currently in place at the GMC before licences are issued to healthcare professionals.
NES Staff	Responsible for managing junior doctors education and career progression.
NHS-X	An organisation for the digital transformation of NHS services.
NHS Administration Staff	People responsible for onboarding new employees and verifying their eligibility for employment to the required standards.
Department of Health and Social Care	The UK government department responsible for regulating the healthcare service, including the policy requirements around identity verification and assurance of healthcare professionals.
NHS Systems Administrators	Individuals experienced with the information systems currently in use within NHS organisations who can speak to the challenges of integrating new technologies.

Table 7.10: Stakeholders of the Scottish Healthcare Ecosystem

Additionally, these interviews would attempt to identify any stakeholders missing from Table 7.10 and aim to include them in this evaluation process.

Following these discussions, all participants would be invited to a facilitated workshop that leverages the PoC and the interactive Jupyter environment to walk through the scenarios currently implemented for a medical student becoming a junior doctor

(Figure 6.9) and a healthcare professional onboarding at a hospital (Figure 6.11). Participants would be invited to download a mobile application able to interface with the Jupyter notebooks in order to store and present the different ID artefacts necessary to complete the modelled identification flows. The aim of this would be to provide a tangible sense of the ways in which digital credential technologies work and how they might change existing healthcare identification processes. Using this direct experiential knowledge of the PoC, participants would be split up into smaller groups with representatives from each identified stakeholder and asked to envision a future where digital credentials have been adopted throughout the SHE. Prompts for this conversation would be taken from the initial semi-structured interviews. Insights would be fed back and discussed with the whole group.

Time could then be spent discussing the evaluation process in order to understand what worked well and which areas could be improved in a future session. Part of the goal of this participatory approach would be to create a repeatable, educational tool that could be used to engage with individuals throughout the SHE.

7.6 Conclusion

This chapter has evaluated the limitations of technological artefacts synthesised to support identification processes using the HVIEP and specifically the Hyperledger Aries ACA-Py agent framework.

Artefacts whose purpose is to facilitate credential issuance must pre-define and author transactions for public identifiers (NYMs), credential schema (SCHEMA) and public keys for specific schema (CLAIM_DEFs) to an identified Indy-based ledger as part of an instantiation phase. All other artefacts must have read access to this same VDR. The results from this chapter have indicated a number of design considerations and technical constraints that should be taken into account during this phase.

Visualisation of the relationships between transactions on the Indy-based ledgers taken from data held within the Sovrin MainNet demonstrates the value of identifying strong roots of trust within the identification system. All transactions are signed by

their author as identified by a NYM transaction, however, to be included in the ledger these transactions should also be endorsed by NYM's with a specific role. Defining rules, governance and assurance processes around an identified, contextually authoritative and accountable set of identifiers in the role of endorser on the ledger provides a strong root of trust for the system.

The next aspect that is important to consider during the instantiation phase is the structure and type of the credentials that will be issued by identified issuers. This includes the name of the schema, number and names of the attributes, the data that would typically fill these attributes and whether the credential should be revocable by the issuer. This is clearly influenced by the identification system, through which the requirements for the ID artefacts and the information they should encapsulate can be determined.

While digitally verifiable credentials can, in theory, hold information of any size, format and type, these decisions have implications on the performance of the interactions that they will be used within. Furthermore, if information for an attribute is of a specific format then everyone who interfaces with this credential should have the means to understand and meaningfully render the information.

Lastly, benchmarking and contrasting CL-RSA and BBS+ signature schemes independent from the ACA-Py agent infrastructure emphasised a number of important points. First, it is the signature scheme that places the majority of the performance constraints on the protocol execution times. Particularly striking are the slow key generation times for CL-RSA. Second, by realising the same abstract cryptographic protocol (a signature scheme with efficient protocols) in a different mathematical setting these protocols can realise significant performance improvements. As the protocols are functionally equivalent, swapping CL-RSA for BBS+ should be modular and the credential-issuance and present-proof protocols implemented in ACA-Py need not be re-architected from the ground up. This work is already underway.

Eighth Chapter

Conclusion

> *"Where there is trust there are increased possibilities for experience and action, there is an increase in the complexity of the social system and also in the number of possibilities which can be reconciled within its structure, because trust constitutes a more effective form of complexity reduction."*
>
> Trust and Power, Niklas Luhmann [2]

The overarching focus of this thesis has been on human systems of identification, the framework of interaction within which they are embedded and the evolution of these systems that advanced digital ICTs have initiated [22]. This evolution is ongoing. The information society we are embedded within today is still in its infancy and the pace of technological change shows no sign of abating. Our world is complexifying and the decisions we make today about the socio-technical systems we introduce into this world will have long term ramifications on the possible futures we might experience. This is especially true where these systems involve the identification of human beings as a mechanism to provision, but also deny, access to resources and the recognition of rights [14]. When taking into account the current, and historical, abuses of power that identification systems have enabled and amplified, as well as the uncertainty and instability that the climate crisis means we must acknowledge the future holds, it is not hard to imagine digitally augmented identification systems being used to apply opaque, illegitimate and unjustifiable constraints and controls over individual freedoms. At the same time, this thesis has argued that digital ICTs, when applied appropriately to legitimate systems of identification, can create new possibilities for structuring the

information flows within identification systems in ways that respect an individuals right to self-presentation, protect individual privacy, distribute trust and balance and constrain power.

The aim of this thesis has been to demonstrate how technical artefacts that leverage privacy-preserving cryptographic protocols can augment existing formal systems of identification found in highly assured social contexts. The approach taken has been to view identification systems from three distinct, but mutually supporting perspectives: 1) the practical instantiation of a technologically augmented identification system, 2) the technical possibility space for implementing such a system based on the knowledge and understanding developed within two distinct thought collectives; digital identity and cryptography, and 3) a sociological framing that situates identification systems within human systems of interaction from which an understanding of identity, trust and privacy has been distilled. It is within this framework that identification systems are made meaningful and can be justified. This thesis argues that this interdisciplinary appreciation enlarges our understanding of identification systems and their implications as they are applied to structure human systems of interaction. This novel interdisciplinary treatment of identification systems is in itself a contribution of this thesis. Additionally, the diagrams presented in Appendix F depict a series of different views of these human systems of interaction that provide a unique visual commentary to this thesis.

8.1 Technologically Augmenting an Existing Identification System

8.1.1 Modelling the System

This thesis has developed a PoC demonstrating how formal systems of identification with a well-defined set of identification processes, ID artefacts, organisations and institutions could be augmented with advanced digital ICT. The identification of healthcare professionals as they interact within the SHE throughout the course of their career was selected to provide a concrete identification system to study. A detailed understanding

of the existing identification system was developed through a facilitated workshop that engaged with key stakeholders and healthcare professionals with lived experience of this system [41]. This understanding was then expressed as a network of intentional, strategic relationships and dependencies between entities that participate within the SHE using iStar diagrams. These diagrams provide a comprehensive overview of the ecosystem with a particular focus on formal processes of identification that a healthcare professional would likely negotiate during their career (See Figure 6.6). The dependencies between ID artefacts within this ecosystem were also modelled, identifying the GMC licence as a foundational ID artefact required within all other identification processes (See Figure 6.7).

From this analysis of the SHE, two phases of a healthcare professional's career were selected to demonstrate how these identification processes could be augmented with digital technologies. These were the transition from a medical student to a qualified and licensed healthcare professional eligible to practice within Scotland and the pre-employment checks that take place before a healthcare professional becomes employed within an hospital or other healthcare organisation. A contextualised set of iStar diagrams for these two phases were produced and used to inform the PoC (See Figures 6.10 & 6.12). Within the PoC, a number of ID artefacts were issued as digitally verifiable credentials cryptographically signed using a privacy-preserving digital signature scheme, CL-RSA [28]. These ID artefacts were stored and controlled by the software system representing the healthcare professional, which could then present them over encrypted, authenticated communication channels to meet the necessary requirements of identification.

This PoC repeatedly emphasised that these digital interactions should not replace physical credentials or displace existing human processes of identification, but rather they can augment existing processes by providing a digital channel for integrity-assured information to flow. This can even include biometric information identifying the subject such as a photograph as demonstrated in the PoC within the GMC Licence credential. Binding of information from ID artefacts presented over a digital channel and correlated with an identifier to a physical person should be done through a face-to-face

interaction. This removes the layer of indirection introduced by technological artefacts which enables strong assurance to be achieved in the identification of the person as the legitimate holder of the ID artefact. This thesis has shown these interactions already exist within healthcare and likely many other highly structured contexts. Augmenting these interactions with digital credentials provides the ability to contrast information contained within multiple ID artefacts from distinct sources and formats, increasing the level of assurance that can be placed in the identification. This also reduces the burden these processes of identification place on those being identified. Furthermore, the fact that healthcare professionals, organisations and institutions can establish pairwise, authenticated, encrypted communication channels presents an exciting possibility space for future research. For example, adjacent research demonstrated that healthcare organisations could form authenticated communication channels amongst themselves and use these channels to engage in privacy-preserving federated machine learning flows [431, 432].

8.1.2 The Hyperledger Verifiable Information Exchange Platform

The PoC leveraged the open-source Hyperledger Indy/Ursa/Aries technology stack, that integrates advanced privacy-preserving cryptographic protocols within a software development framework that facilitates the implementation of application specific software artefacts. Whilst it is legitimate to argue that this technology stack is not standards compliant, achieving standards-based interoperability is a continuous process. This thesis takes the view that the Hyperledger community are actively participating in these processes. Examples of this include their efforts to replace did:sov with did:indy, ensuring all ledger objects are in line with the DID core specification [35, 433] and the transition of DIDComm from Hyperledger to the Decentralised Identity Foundation [216].

At the time of implementation the Hyperledger stack provided the only libraries within the decentralised identity space attempting to apply a signature scheme with efficient protocols to issue verifiable credentials. The Hyperledger community, along

with Evernym and the Sovrin foundation deserve much credit for evangelising and operationalising these more complex cryptographic protocols, demonstrating to the decentralised identity community that it is possible to embed privacy into the lowest levels of the credential exchange protocols.

8.1.3 Schema Design and its Impact on Performance

This thesis evaluated the overall performance of the PoC by deploying a number of Hyperledger Aries agents, instantiated in production-like settings on independent virtual machines. Against this infrastructure a series of experiments were run to stress test the issue-credential and present-proof Aries protocols under varying conditions including; credential size, attribute number and credentials presented. Furthermore, each experiment was run for both revocable and non-revocable credentials.

Results showed that engaging in these protocols is unlikely to be a limiting factor for face-to-face healthcare interactions. However, generalising these results does indicate that careful design of credential schema and presentation requests can have a significant impact on protocol execution. In particular, the use of revocable credentials incurs a significant fixed cost to both credential issuance and presentation so limiting the number of revocable credentials within a identification system is recommended (See Figure 7.17). Within the SHE PoC only the GMC Licence was issued as a revocable credential. These experiments modelled a realistic production setting, with both Aries agents engaging in the protocols were running on a virtual machine as an ACA-Py instance. However, it is likely in many scenarios, including those involving a healthcare professionals, that one agent would be running on an edge device such as a mobile. Further research is needed to determine the impact this would have on protocol execution.

8.1.4 Analysing the Ledger Footprint

This thesis analysed the ledger footprint produced by digital identification systems that make use of Indy-based ledgers by fetching and visualising transaction data for a

known pilot project being undertaken by NHS England that uses the Sovrin MainNet (See Figures 7.2 and 7.3). From these visualisations a clear hierarchical structure can be seen whereby one DID authored many other DIDs to the ledger, these DIDs then wrote a CLAIM_DEF transaction identifying an existing credential schema on the ledger. The chain of signatures that author and endorse transactions on Indy-based ledgers can be traced and independently verified, giving confidence in the information within the ledger. In this manner, from knowledge of a single (or multiple) root DID(s), a representation of the credential ecosystem can be produced and assurance in its issuers and schema can be increased. This is effectively a decentralised, domain specific certificate authority chain, accept that the certificates entities are capable of issuing are arbitrarily configurable to the context in which they are being issued. That the Sovrin MainNet is a public-permissioned ledger, requiring transactions authored to the ledger be endorsed by DIDs with specific roles supports this architecture and is ideal for high assurance identification systems that need confidence that ID artefacts were issued by authorised actors [42].

8.1.5 Cryptographic Constraints

It is important to acknowledge that the PoC was implemented at a moment in time, built using the technologies and libraries that were available, as all technological artefacts are. However, the tools, libraries, protocols and standards within the decentralised identity space are experiencing a period of rapid innovation. A notable illustration of this is the adoption and integration of BBS+ signatures, a signature scheme with equivalent protocols to the CL-RSA scheme but realised under a pairing-based mathematical setting. This makes it more efficient in terms of both space and time complexity [28, 299]. BBS+ was only demonstrated to be theoretically secure under the Type-3 pairing setting in 2016 by Camensich et al [30].

The BBS+ signature scheme was then implemented within the Hyperledger Ursa cryptographic engine in 2018, however it was not until late 2020 that the scheme began to be integrated into software development frameworks such as the Aries Cloud Agent

Python (ACA-PY). Additionally, Mattr has produced an open-source NodeJs wrapper for the underlying BBS+ scheme implemented within Hyperledger Ursa[1]. Latest developments include a distinct open-source cryptographic implementation of BBS+ by Dock.io[2]. This combined with an initial draft specification for the BB+ scheme within the W3C Credentials Community Group provides a clear indication that the value of privacy-preserving cryptographic schemes are being recognised within the wider decentralised identity space [213, 215]. However, a current limitation with both the specification and existing integration of this scheme within frameworks for software development is the requirement to include a public DID identifying the subject of an issued verifiable credential rather than a zero-knowledge proof of knowledge of a signed master secret. While this approach enables selective disclosure, it irrevocably compromises the privacy of the credential holder by forcing them to reveal a correlatable identifier across all credential presentations. Additionally, the current integration of BBS+ signatures within the ACA-Py library does not include a mechanism for privacy-preserving credential revocation. Both of these limitations are expected to be addressed in the near future.

To understand the implications of this transition to the more efficient pairing-based BBS+ signature scheme, both BBS+ and CL-RSA signature schemes were benchmarked using the implementations from the Hyperledger Ursa library. This included benchmarking the following; Key Generation, Credential Issuance and Present Proof. In all instances, BBS+ led to a significant reduction in both time and space complexity. Of particular note was the contrast between Key Generation, with BBS+ reducing execution times by more than 99.9% (See Table 7.7). This improvement has been translated directly into implementation, such that a single shorter key can deterministically generate a BBS+ public key to verify signatures on an arbitrary number of messages at a negligible cost. Whereas, CL-RSA key generation takes seconds and hence it would have an unacceptable impact on credential verification times. This is why CLAIM_DEF transactions, containing public keys for a specific number of attributes as identified by a specific

[1] https://github.com/mattrglobal/node-bbs-signatures
[2] https://github.com/docknetwork/crypto

SCHEMA are published to the Indy ledger. Credential verification involves fetching this public key from the VDR rather than generating it. Here we see constraints imposed on the architecture of a system from the limitations of the mathematical setting, RSA under the Strong-RSA assumption, which the cryptographic protocol has been realised within. Transitioning to a pairing setting removes these constraints whilst functionally achieving the same cryptographic properties.

8.1.6 Aries Juypter Playground

To support the practical aspects of this thesis, a custom Juypter notebook environment for experimenting with Hyperledger Aries agent using the ACA-Py library was created in collaboration with the open-source Open Mined community. Jupyter notebooks allow users to run individual execution cells that can break down the steps within an application into its constituent parts, furthermore the notebooks can be made self-documenting by including Markdown cells for custom text. These notebooks provide a developer friendly, configurable interface to interact with Aries agents running within a docker container and orchestrated through docker-compose. Together they form the Aries Juypter Playground, an open-source, generic and customisable environment simplifying the arbitrary configuration of systems leveraging Aries agents[3]. This has been recognised as a contribution in its own right within the Software Impacts journal [43].

The SHE PoC and the ACA-Py performance benchmarking experiments were both developed within this custom Juypter environment and are available on Github[4,5]. This means that in the future researchers will be able to independently run and verify these results, and easily extend and build on the PoC. Furthermore, these notebooks enable these research artefacts be self documenting, transparent and interactive [43]. This creates a valuable tool that can be used to facilitating dialogue with key stakeholders, both technical and domain specific, of the identification system modelled. These

[3] https://github.com/wip-abramson/aries-jupyter-playground

[4] https://github.com/blockpass-identity-lab/scottish-healthcare-ecosystem-poc

[5] https://github.com/wip-abramson/aries-jupyter-playground/tree/project/aries-performance

interactions are vital to ensuring the technologies we introduce into these systems are understood, trusted and fit for purpose.

Finally, the Aries Jupyter Playground has been used to develop educational material around public key infrastructures, decentralised identity and privacy-preserving cryptographic credentials that was published as a section of Open Mined's publicly available course - the Foundations of Private Computation[6]. Producing and disseminating these research artefacts in such a manner embraces the infancy of these new technologically enabled possibilities by reducing the burden on future researchers wishing to explore and contribute to this space. Evidence that this has already been successful can be seen in a fork of the Aries Juypter Playground Github repository that explores a sovereign data exchange use case for vehicle emission data combining Secure Multi-Party Combination with Verifiable Credentials[7]. This forms the practical element of a masters thesis undertaken at the University of Postdam in Berlin [434].

8.2 The Genesis, Development and Standardisation of Digital Identification

The literature from within the cryptographic and digital identity communities have been analysed in order to understand the design space for digital identification systems. The first point to again emphasise is the difference in origins of the two communities of thought studied in this thesis and the collective knowledge they have produced, maintained and evolved since their inception.

8.2.1 Cryptographic Thought

Cryptographic thought is constrained by the fundamental laws of mathematics, and many of its ideas surrounding the application of this math to information flows originated before networked, digital ICT had been fully realised. It was, and remains, a highly theoretical, conceptual science in many ways continuing in the footsteps of Turing and

[6] https://courses.openmined.org/courses/foundations-of-private-computation
[7] https://github.com/kajaschmidt/aries-sympc-jupyter

Shannon [17, 244]. The transition of cryptography into a highly systematised scientific discipline began with the introduction of public key cryptosystems and asymmetric authentication by Ralph Merkle, Whitfield Diffie and Martin Hellman [26, 25]. Concepts and constructs that remain a part of cryptographic thought to this day were first defined by cryptographers anticipating future information societies and the tools and systems that might be needed to protect privacy, encourage and support decentralisation and empower individuals. This thesis reflected on the evolution of cryptographic thought through the lens of the conceptual idea of credential mechanism first described by Chaum in 1985 [27].

By tracing the history of a credential mechanism through the literature, a loosely coupled hierarchical structure of cryptographic thought emerged (See Figure 4.2). Although this structure requires further validation, it provides a lens through which cryptographic knowledge can be situated and understood, emphasising the important distinctions between abstractions, concrete instantiations and practical implementations (See Figure 4.3). This framing shines a light on the highly scientific nature of the cryptographic discipline and the carefully defined foundations and building blocks from which cryptographic protocols are constructed. In doing so, this thesis demonstrates that secure credential mechanisms have now been theoretically realised by multiple concreted cryptographic protocols, without encoding the ability for pervasive over identification into our digital interactions [28, 36, 30, 31]. Furthermore, from the 2000s these protocols have increasingly been implemented and integrated into software development frameworks so that they are available as practical tools for software engineers and system architects to apply when designing and implementing real-world systems [238, 239, 240].

8.2.2 Digital Identity Practitioners

In contrast, the origin of the digital identity community can be traced to the pragmatic requirement for identification processes functional within digital environments. During the 80s and much of the 90s this was a requirement defined primarily from the

perspective of enterprise, government or academic organisations who had to manage employee access to this new technical infrastructure [33]. The modern inception of digital identity began to coalesce around 2005. The bi-annual Internet Identity Workshop was established and Kim Cameron published his widely cited *Laws of Identity* [193]. Decentralised identity, characterising this communities latest iteration of thought, has led to the emergence of further groups such as Rebooting the Web of Trust, the W3C Credentials Community Group and the Decentralized Identity Foundation. Taken as a whole, this loosely coupled community aims to achieve standards-based interoperability in digital identification systems, recognising this as a vital mechanism to protect against vendor lock-in and encourage open, permissionless innovation [176].

The initial focus of this community revolved around developing user-centric systems and standards, although these have now shifted to decentralised, or Self-Sovereign identification systems and standards. The pursuit of open standards means this community of thought is broadly constrained by group consensus, with different parties attempting to shape the narrative and influence the contents of the specifications that become standardised. Standardisation becomes a place to wield power and exert control. Furthermore, it should be pointed out that this is an internet native community, with many of its founding members participating in and shaped by the innovation surrounding the dot com era. For example, it was during this time that the W3C standards body was established. Finally, this community is highly practical in contrast with, at least the perception of, cryptography and its thought products. They want systems that work.

8.2.3 The Intellectual Interaction between Cryptography and Digital Identity

This thesis makes the case that we are experiencing a significant moment in the history of this conceptual space. The digital identity and cryptographic communities have both arrived at essentially the same system for digital identification. Specifically, the use of public key cryptography to ensure the integrity and authenticity of ID artefacts, which are able to be held within an information storage system controlled by the

entities to whom the claims pertain. Cryptographic protocols can guarantee strong privacy and security properties about these systems, while open standards enable interoperability and encourage competition. Both, together, can foster decentralisation in digital infrastructure, distribute trust and protect individuals from abusive power differentials that systems of identification tend to facilitate.

There is also a tension that exists between these two thought collectives which revolves around how many desirable, functional properties of specific interactions should be realised within mathematically specified cryptographic protocols. The challenge, accurately identified within the digital identity community, is that math alone cannot realise a human system. The mathematical protocols must be meaningfully grafted to human processes and operationalised within software artefacts. To realise interoperable artefacts synthesised using distinct implementations requires shared semantic meaning and mutually executable cryptographic protocols. Where the cryptographic protocols are more complex, they necessarily constrain certain aspects of the specifications, standards and protocols required to operationalise them. These constraints may not be acceptable to all entities participating in standards organisations and hence make standards-based interoperability more difficult.

While this is unfortunate, this thesis has argued that cryptographic protocols add real value by seeking to make the realisation of specific protocol properties, such as unconditional unlinkability, a matter of science. Cryptographers work on the basis that interaction between ICT artefacts involves the exchange of information which can be represented as a binary string and hence a number that can be incorporated into mathematical algorithms whose properties and assumptions can be formally studied. Closer intellectual interaction and appreciation of the role and value of both communities of thought can ensure standards do not constrain cryptographic possibility while facilitating interoperability and well defined semantic meaning. There is evidence in both the academic literature and the standards organisations that this is beginning to happen [37, 215].

8.2.4 A Pivotal Moment in History

The intellectual interaction between the digital identity and cryptographic communities of thought stimulated by Bitcoin and the subsequent innovation around distributed ledgers is fascinating. However, there is another reason this is a pivotal moment in the history of digital identification systems. These systems are now practical, implementable and are increasingly being rolled out in practice to support identification processes that are no longer constrained to digital environments [22]. The literature on identification systems within developing countries provides the clearest example of this trend, including a strong warning of the potential impacts of these systems are having on the individuals being identified [40, 23, 14, 435]. The global COVID-19 response provides another, where technologies have been introduced to paper over the cracks created by societal injustices rather than addressing their root causes [182]. We can all learn from the real-world experiences of individuals as they interface with these technologies and should be cautious whenever advocating for technological solutions to complex societal problems. Especially considering identification technologies are designed to improve a historically asymmetric power relationship between the entity being identified and the entities defining and performing the means and processes of identification.

8.3 Identification Systems in an Information Society

This thesis situates identification systems within a framework of human interaction synthesised from a rich and interdisciplinary array of literature found within the social sciences. Identity, trust and privacy are all understood within this framing, providing a scientific language to describe an aspect of life that each of us have a front row seat to. Our lived experience moment to moment as human beings entangled in a messy, dynamic web of interactions embedded within social, cultural and historical contexts [436, 437] (See Figure F.7). From systems of interaction, complex social organisations emerge and their structures are perceived. Interaction holds us in relation with each

other, with social structures and increasingly with our technologies [52]. It is through interaction that information flows imperfectly around human systems, shaping the meanings that individuals apply as they attempt to understand their present experience (See Figures F.1 & F.3).

8.3.1 Identity

By identifying ourselves, and others, as belonging to, associated with, or in relation to, a collection of meanings we create and apply identities to make sense of experienced information [53]. Identities, under this definition, are not a static collection of attributes, but dynamic systems of meaning from which purposes, intentions and expectations can be derived [39, 4]. These systems are continuously re-calibrated in response to information. Furthermore, while identities are uniquely perceived and applied by individuals they are better understood as emergent at all scales of human social organisation. Stets and Burke categorise these as persons, role and group identities, highlighting that many identities of each classification are typically aggregated and applied to a single entity during an interaction [4] (See Figure F.4).

8.3.2 Trust

Trust adds another layer to the human experience that is required, secured and applied to navigate uncertainty in the present. This uncertainty is primarily introduced by human actors capacity for unpredictable actions [438]. Trust is placed in the expectations of others to uphold commitments to fulfill certain obligations in the future that they are judged to have made based on the identities applied to them during interaction (See Figure F.2). For trust to be placed intelligently, identities and their binding to entities must be stabilised by systems of trustworthy information flows that provide assurances in the constancy and continued applicability of these identities within future presents. In doing so, trust structures possibilities from a meaningful experience reducing the complexity of action selection in relation to an uncertain future populated with unpredictable entities [2].

Trust, like identity, is not static or fixed, but contextually dependent, individually applied and constantly changing in response to information. Trust requires information, which is then applied to the present when making decisions to place trust. Information originates from events that occurred in the past, either directly experienced and remembered or communicated through interaction with others (see Figure F.2). Where events were not personally experienced, the trustworthiness of the information about the event must be assessed, which in turn requires information about the source whose trustworthiness must also be judged [70]. The selection of actions and the processing of experience continuously produces new informational input which must be taken into account when placing trust in a future present [2].

Trust is a fragile, intangible asset that is fundamental to a functioning society [16, 2, 15]. As society has scaled and specialised into increasingly complex, interdependent configurations, mechanisms for placing trust impersonally in the actions of unknown others were required. Institutions arose to structure interactions within specific social contexts, regulating, defining and enforcing specific meanings from which interacting parties could base their world-view, independent from the requirement for intimate personal knowledge about the other. As Schneier puts it, institutions *induced* trust within these social contexts [15].

8.3.3 Institutions and Identification

Formal systems of identification have become an important component of institutionalised, highly structured social contexts [15]. They provide static, tamper-resistant attributes characterising an identified subject, encapsulated within an ID artefact and issued by an identifiable source. These ID artefacts can be requested, presented and evaluated within an interactive setting, binding entities to labels and the formalised collections of meanings that characterise this labelled identity (See Figure F.11). This in turn informs decisions to place trust, grant access to resources and recognise identified subjects as holders of rights and responsibilities [14]. Processes of identification provide a mechanism for distrust to be applied impersonally at the boundaries of social struc-

tures, mitigating the risk of placing trust in the actions of actors occupying positions and associated role identities within these structures.

Over time, institutions became the authoritative sources of information defining and regulating role identities with their associated rights, responsibilities and accountabilities from which expectations for behaviour and action selection could be judged. This ability to structure social interactions through the control of information flows conveyed and centralised power, which institutions and the systems they created to maintain and regulate these structured interactions had to be trusted to wield legitimately [14]. This thesis has emphasised throughout that the process of identification has always been an asymmetric power relationship in which an entity being identified must meet the requirements of identification set by some other entity and enforced by its agents. Unfortunately, both past and present human experience is littered with examples of this trust being misplaced and abused by those in positions of power [14].

8.3.4 Advanced Digital ICTs

Advanced digital ICTs add an additional layer of complexity to human interaction. As artefacts capable of recording, retrieving and processing information, ICTs have become active participants within our interactions able to react and respond to informational input [9, 7]. Human actors interface with these technological artefacts, which augment our capacity for action and mediate an increasing number of societies information flows. An analysis of the impact of these changes from a complex systems perspective is presented in Appendix E.

In order to realise many of the possibilities enabled by digital ICTs, mechanisms to place trust and judge trustworthiness of actors and their actions virtually represented within these new informational environments were required. This spawned the development of *identity management systems*, designed to manage the boundaries of informational environments by identifying and authenticating human actors against virtual entities characterised by a collection of static attributes. Information systems then used these *digital identities* to determine access and authorisation [33]. Purposefully

constructed information systems replaced humans as the agents performing identifications in order to structure and control the possibilities for action available to virtual entities on behalf of a controlling entity.

8.3.5 Asymmetries of Knowledge and Power

Asymmetric power relationships can be clearly identified within information systems. Individuals have their identities *managed* for them within the domain of an information system and are provided with means to authenticate against these virtual representations. This approach to digital identity originates from academic and corporate contexts with justifiable requirements to exert control at a time when information systems were understandably viewed as separate from the real-world. However, the increasing computational power combined with the popularisation of the Internet as a space that facilitates both social and economic activity proliferated these approaches to *identity management* into the personal lives of private citizens. Such that today we find ourselves managing a plethora of account credentials, giving us the ability to authenticate against virtual entities through which we mediate our digital interactions. In almost all instances, these entities and the information that characterises them are held on information systems beyond our influence and are used for purposes without our consent and outside of our awareness [74].

The ability to abuse this asymmetric relationship is made worse by the assumption that in order to place trust in virtual entities, information that can identify and characterise the physical human actor executing actions through that entity is required. The lack of trustworthy information sources, led to the collection and aggregation of rich informational dossiers that correlated the actions of virtual entities across distinct contexts with a unique and identified human individual [7, 27]. This has produced vast asymmetries of knowledge and power within society. Increasingly organisations have sought to replace trust as a means to navigate uncertainty, with predictive capabilities that strive for certainty as our actions are nudged and herded towards purposes decided by others beyond our awareness [72].

This thesis argues strongly that prediction is a poor substitute for trust when navigating a future with orders of magnitude more possibilities than could ever be realised in the present [2]. Prediction treats humans as if they were predictable, when it is our ability to change and adapt in response to information as we explore possibilities in our environment that identify us as a complex living beings [51, 52, 439]. Furthermore, predictive capabilities require information that has been repeatedly shown to encode human biases, which are then algorithmically enforced and difficult to challenge [81, 146]. Trust enables us to respond creatively and unpredictably to an uncertain future. As Luhmann states, *trust constitutes a more effective form of complexity reduction* [2].

The risk of the proliferation of information systems collecting and maintaining dossiers of sensitive personal information on human individuals executing actions within these new environments was identified by cryptographers in the 80s. In an attempt to mitigate this risk, David Chaum conceptualised a solution that could achieve the necessary security within these new environments without the pervasive over identification and correlation of humans [27]. This thesis has demonstrated that this conceptual system has been theoretically realised through a rigorous scientific process based on strong mathematical foundations and is consistent with the now highly systematic field of cryptographic knowledge. Furthermore, these theoretical mathematical constructions have now been implemented within software libraries and are available to be integrated into software systems that require identification processes.

8.3.6 Structured Transparency

Cryptography has unlocked new possibilities for structuring digital interactions in ways that achieve security and ensure the appropriate transparency whilst preserving privacy. Whilst on the surface, privacy in our information society might seem far away, in fact ensuring the contextual integrity of societies information flows through carefully structured transparency has never been more achievable. The cryptographic credential system studied in this thesis is just one example, there are many others emerging from theoretical research that are beginning to be practically applied to information flows.

Our expectations of privacy within these informational spaces need to be updated to match these new possibilities for realising it. While the value of privacy is hard to pin down, it is widely acknowledged as foundational for a free and Democratic society that respects freedom of expression, encourages creativity and contains strong, trusting relationships [7]. Whilst privacy must be balanced with other needs of society, privacy-enhancing technologies offer the tantalising prospect of carefully structured transparency within our socio-technical interactions [117].

Privacy-preserving credential systems enable individuals to represent themselves virtually as a constancy without a reliance on third parties. Trust can then grow naturally in this virtual entity over time, based on their actions which can be cryptographically bound to them by all who digitally *experience* them. In many cases this should be enough. We must move away from the desire to bind virtual entities to human individuals, unless such a requirement can be clearly justified. Empowering virtual entities with the ability to prove their uniqueness, or humanity, within specific contexts would support a broad range of interactions without introducing the requirement to uniquely identify a human individual [440]. This layer of indirection between physical and virtual environments should be seen as a positive, enabling rich human diversity of expression and presentation in the different contexts they choose to interact. However, within certain social contexts interacting with a stranger and allowing them to demonstrate trustworthiness over time is not enough. The identification of individuals as holders of formalised identities and the structuring of interactions on the basis of these identities has been used as a mechanism to institutionalise trust since long before the introduction digital ICTs. Privacy-preserving credential systems present an opportunity to replicate and extend existing institutionalised trust into virtual environments without further centralising informational power within the hands of institutions.

8.4 Limitations and Future Work

8.4.1 Identification Systems within Professional Contexts

Focusing of the identification of healthcare professionals within SHE provided a concrete and complex case study upon which to base the PoC. While this ecosystem was understood through a workshop that engaged with healthcare professionals and key stakeholders [41], the resulting PoC was not validated through a similar workshop. This would provide opportunities to evaluate the iStar diagrams and the Juypter notebooks as providing an accurate and useful model of the SHE. Furthermore, this could be used to produce a more generic evaluation of the application of these tools when designing a credential ecosystem for a specific social context.

The PoC produced and evaluated within this thesis focused primarily on its technical characteristics, demonstrating that privacy-preserving cryptographic credentials could be applied to augment existing identification processes in a technically secure and performant system. Further work is required to refine and systematise the performance metrics developed within this thesis, so they can be used compare distinct software frameworks for implementing credential-based digital identification systems. Furthermore, throughout this thesis it has been made clear that technical systems are embedded within social systems and it is only through understanding these social systems that a more meaningful, real-world analysis of these technical systems can be undertaken. Assessing how augmenting existing identification processes within healthcare impacts risk within this context is important when considering whether such a system should be introduced. A cursory analysis is that requiring both digital and physical ID artefacts from multiple sources can provide mutual assurance about the validity of the information presented. The ability to bind digital ID artefacts to their subjects using biometrics such as a photograph and the cryptographically verifiable integrity of these digital ID artefacts further increases the assurance within the identification process. For example, requiring the presentation of digital GMC licence containing a photograph identifying the subject alongside a physical ID document such as passport

both to be presented during an in-person interaction should be considered at least as strong as the requirement for two photographic identification documents currently in place within the NHS [391]. The Trust over IP foundation have released a suite of tools to support a systematic risk assessment conformant with industry standards within context specific credential ecosystems [441], applying these tools to the SHE is left to future work.

The PoC intentionally focused on the existing human processes of identification that healthcare professionals must navigate as they interact within the SHE throughout the course of their career. This captures the foundational pieces of evidence they must collect, and the key interactions that this evidence is presented. However, as ICT proliferate into our social systems, professionals are increasingly having to manage a plethora of account credentials to access different informational systems that are being introduced into their working environment. A logical next step for this research is to evaluate how empowering healthcare professionals with digitally verifiable domain specific credentials can transform and simplify the identification systems managing the boundaries between professionals and the information systems they are required to interface with. While this presents challenges with respect to integrating the new technology into legacy systems [171], it would allow information systems to plug into a common identification system as verifiers of domain specific ID artefacts. Rather than, as is common today, each system introducing an additional digital identification system provisioning account credentials to authenticate healthcare professionals against virtual entities trusted only within the boundaries of the information system.

Expanding the analysis of the healthcare context offers many possibilities for future work. Examples include: how existing well established governance frameworks and processes could be adapted to include and support technologically augmented identification processes; exploring the role digital verifiable credentials could play in supporting international movement of professionals between geographic regions; further research into the implications of establishing, flexible, private, authentic, peer to peer communication channels between different entities within healthcare (the GMC to healthcare professional relationship and hospital to hospital connections both

seem interesting places to start); the implications of digitally empowered healthcare professionals in relation to patient care. Establishing a strong foundation that enables healthcare professionals to represent themselves as trustworthy virtual entities opens up a new possibility space that we are only just beginning to explore.

Finally, the approach taken in this thesis to understand and augment a specific highly structured professional context - healthcare - could be applied to other similar social contexts with existing identification systems. Examples include; Law, Accounting, Teaching, Architecture and Engineering. This thesis suggests that there is value in empowering professionals, trusted within a specific social context, with the means to bring this established institutional trust into specific, contextually relevant virtual environments.

8.4.2 The Structure of Cryptographic Thought

This thesis produced a conceptual framework emphasising the layered and loosely coupled nature of cryptographic knowledge (Chapter 4). This was depicted within Figures 4.2 and 4.3. This framework was produced as a result of a detailed analysis of the systematisation of cryptographic knowledge since the inception of public key cryptography through the lens of a credential mechanism [26, 27]. Influenced by the work of Fleck [32], this analysis indicates how early abstract conceptual ideas such as a credential mechanism stimulated interest and directed focus within cryptographic research. This ultimately resulted in these early ideas being refined into concrete cryptographic protocols with well understood mathematical properties. It is only once this had occurred that these ideas could be realised within software artefacts and applied to real-world use cases. While this conceptual framing proved useful in the presentation of cryptographic knowledge as a highly systematised, coherent, academic discipline built on strong mathematical foundations, further research is needed to validate and refine this framework.

The presentation of cryptography in this manner was intended to emphasise that conceptual ideas that have been refined over decades of academic research are now

maturing into practical, implementable and implemented cryptographic protocols. Further research, education and collaboration is needed to ensure these innovations are recognised and built on within wider software engineering disciplines. Cryptography places mathematical constraints on information flows, providing mechanisms for these flows to be carefully structured to decentralise power, distribute trust and protect privacy. Privacy-preserving credential mechanisms are just one of the protocols to have emerged from this scientific discipline. Others include threshold cryptography [314], group signatures [154, 339], secure multi-party computation [152], homomorphic encryption [161], zero-knowledge proof systems [156] and more. Together they create a new suite of tools that can be applied when designing socio-technical systems within society. We are only beginning to understand the practical implications of these new tools.

8.4.3 Social Sciences

This thesis synthesised literature on identity, trust and privacy from within the social sciences because researchers working on technical systems that augment social systems have an ethically responsibility to view these holistically as human systems. Identity, trust and privacy are human concepts that have meaning within these systems, this thesis has attempted to integrate these meanings throughout. As ICTs are increasingly interwoven into the fabric of society, it is imperative we recognise the impact of these technologies have on ourselves as human beings and the social systems we participate in. Starting from an understanding of identity, trust and privacy as emergent properties of human systems of interaction provides a basis for this analysis. However, this thesis remains a technical thesis undertaken by a computer scientist and a software engineer not a sociologist. It is through this lens that the sociological literature presented in this thesis has been filtered and understood. This thesis hopes to contribute to, and encourage, wider interdisciplinary research within computer science and especially with respect to identity and identification in an information society.

8.5 Closing Remarks

Digital identification systems offer immense potential in the future. However, it is easy to get caught up in this potential for streamlined, authentic, efficient, verifiable information exchanges that we overlook the complex, unpredictable, highly contextual human actors that must interface with and interact across across these systems. Identification systems manage the boundaries between spaces, determining access, affordances and trust that are conveyed to the identified entity. Control over identification systems, is control over these boundaries and the spaces that they surround. Digital identification systems managing informational spaces bring this into sharp relief, encoding boundaries into rules that are then evaluated by a deterministic software system. There are few open, public digital spaces, instead there are privatised virtual environments within which we have no choice but to be identified if we wish to participate. These informational spaces are controlled and configured beyond our awareness or influence by private entities pursuing motives not necessarily aligned with our own.

As existing human systems of identification become technologically augmented we should be careful not to replicate these patterns within real-world interactions. We do not want all our interaction spaces to be surrounded by boundaries that enforce some *authorities* identification processes, no matter how easy technology makes this to achieve. Furthermore, this thesis argues strongly that identification systems capable of identifying unique human individuals without their consent or knowledge are an insidious, and dangerous, development that has proliferated our information society [226]. Tools like facial recognition, digital fingerprinting and location tracking all fall into this category. To be identifiable, is to be vulnerable and while vulnerability helps to cultivate trust it also creates the possibilities for abuse. When augmenting our existing human systems of identification with digital ICTs, protecting individual privacy through carefully structured transparency should underpin the design these systems [117]. This can go a long way towards limiting the potential abuses of power these systems create, whilst still enabling identification when required and justified. But we must move away

systems that enable the identification and correlation of individuals across contexts by default, merely as a side effect of our interaction with the socio-technical systems we have introduced.

Whilst respecting privacy within identification systems is paramount, alone it is not enough. Control over an identification system conveys power. The power to influence meaning and through carefully constructed meaning, direct and control the selection of actions. What information ID artefacts encode, which entities are authorised to issue them under what conditions and to whom. Once issued, under which circumstances can these be required within an identification process. Formalised systems of identification can provide an approach to consistently and fairly apply rules to an identified population beyond the influence and manipulation of interpersonal relationships. However, it must be acknowledged that they can just as easily encode, solidify and amplify existing injustices found within social systems. The legitimacy of an identification system should be carefully established and continuously justified against the social context within which it is applied. Nissenbaum's theory of contextual integrity would seem applicable here, after all an identification system fundamentally changes the information flows within society [7].

Healthcare and similar professional contexts have evolved systems of identification alongside the complexification of their social organisations. Professionals have occupied increasingly specialised roles in society, with knowledge and competence beyond that of the average individual. As a result they must be trusted to select actions based on their unknown and unknowable perception of experience. Formal systems of identification provide the means by which individuals can identify themselves as holders of specific role identities regulated by and accountable to authoritative institutions and as such can be trusted within specific interactive settings. These systems of identification applied to a specific subset of individuals occupying positions of power within social contexts can clearly be justified.

This thesis has demonstrated that technological artefacts can augment and simplify existing identification processes. When considering the future, the creep of digital ICTs into all aspects of our social systems seems inevitable and will undoubtedly bring many

benefits. However, to fully realise these benefits this thesis argues that digital ICTs should aim to augment rather than displace existing institutionalised trust. The desire for trustlessness advocated for by some working with distributed ledger technologies is misplaced. This narrative can obfuscate rather than remove those in positions of power, who must implicitly be trusted [108]. Institutional trust is important within our social organisations, especially within highly assured professional contexts, but it must adapt the information society we all now inhabit. Trusted and trustworthy professionals require effective tools to enable them to securely interface with technological artefacts, empowering them to act as stewards of any socio-technical systems introduced into their domains. Designing a decentralised, standards-based, privacy-preserving identification system for healthcare professionals can act as a powerful enabler for the design of legitimate, trustworthy and beneficial use technology within healthcare, and potentially society in general, more widely.

References

[1] Lewis Dartnell. *Origins: How the Earth shaped human history*. Random House, 2019.

[2] Niklas Luhmann. *Trust and power*. John Wiley & Sons, 2018.

[3] Frens Kroeger. "Unlocking the treasure trove: How can Luhmann's theory of trust enrich trust research?" In: *Journal of Trust Research* 9.1 (2019), pp. 110–124.

[4] Jan E Stets, Peter J Burke, Richard T Serpe and Robin Stryker. "Getting identity theory (IT) right". In: *Advances in Group Processes*. Emerald Publishing Limited, 2020.

[5] Erving Goffman. *The presentation of self in everyday life*. Penguin Psychology, 1990.

[6] Michael J Carter and Celene Fuller. "Symbols, meaning, and action: The past, present, and future of symbolic interactionism". In: *Current Sociology* 64.6 (2016), pp. 931–961.

[7] Helen Nissenbaum. *Privacy in context*. Stanford University Press, 2020.

[8] Gregory Bateson. "A theory of play and fantasy". In: *The game design reader: A rules of play anthology* (2006), pp. 314–328.

[9] Luciano Floridi. *The fourth revolution: How the infosphere is reshaping human reality*. OUP Oxford, 2014.

[10] Herbert A Simon. *The sciences of the artificial*. MIT press, 2019.

[11] James Gleick. *The information: A history, a theory, a flood*. Vintage, 2011.

[12] Marshall McLuhan. *Understanding media: The extensions of man*. MIT press, 1994.

[13] Gregory Bateson. "Redundancy and coding". In: *Animal communication: Techniques of study and results of research* (1968), pp. 614–626.

[14] Christoph Sperfeldt. "Legal identity in the sustainable development agenda: actors, perspectives and trends in an emerging field of research". In: *The International Journal of Human Rights* (2021), pp. 1–22.

[15] Bruce Schneier. *Liars and outliers: enabling the trust that society needs to thrive*. John Wiley & Sons, 2012.

[16] Rachel Botsman. *Who can you trust?: How technology brought us together and why it might drive us apart*. Hachette UK, 2017.

[17] Alan Mathison Turing. "On computable numbers, with an application to the Entscheidungsproblem". In: *Proceedings of the London mathematical society* 2.1 (1937), pp. 230–265.

[18] Claude E Shannon. "A mathematical theory of communication". In: *The Bell system technical journal* 27.3 (1948), pp. 379–423.

[19] Onora O'Neill. "Reith lectures 2002: a question of trust. Lecture 4: trust and transparency". In: *BBC Reith Lect* (2002).

[20] Audun Jøsang and Simon Pope. "User centric identity management". In: *AusCERT Asia Pacific information technology security conference*. Citeseer. 2005, p. 77.

[21] Jonathon Donner. *The difference between digital identity, identification, and ID*. 2018. URL: https://medium.com/caribou-digital/the-difference-between-digital-identity-identification-and-id-41580bbb7563.

[22] Alan Gelb and Anna Diofasi Metz. *Identification revolution: Can digital ID be harnessed for development?* Brookings Institution Press, 2018.

[23] Silvia Masiero and Savita Bailur. *Digital identity for development: The quest for justice and a research agenda*. 2021.

[24] *Busting Bureaucracy*. Tech. rep. UK Department of Health and Social Care. URL: https://www.nhsemployers.org/-/media/Employers/Documents/Retain-and-improve/Kings-College-Hospital-identity-scanning.pdf.

[25] Ralph C Merkle. "Secure communications over insecure channels". In: *Communications of the ACM* 21.4 (1978), pp. 294–299.

[26] Whitfield Diffie and Martin Hellman. "New directions in cryptography". In: *IEEE transactions on Information Theory* 22.6 (1976), pp. 644–654.

[27] David Chaum. "Security without identification: Transaction systems to make big brother obsolete". In: *Communications of the ACM* 28.10 (1985), pp. 1030–1044.

[28] Jan Camenisch and Anna Lysyanskaya. "An efficient system for non-transferable anonymous credentials with optional anonymity revocation". In: *International conference on the theory and applications of cryptographic techniques*. Springer. 2001, pp. 93–118.

[29] Jan Camenisch and Anna Lysyanskaya. "A signature scheme with efficient protocols". In: *International Conference on Security in Communication Networks*. Springer. 2002, pp. 268–289.

[30] Jan Camenisch, Manu Drijvers and Anja Lehmann. "Anonymous attestation using the strong diffie hellman assumption revisited". In: *International Conference on Trust and Trustworthy Computing*. Springer. 2016, pp. 1–20.

[31] David Pointcheval and Olivier Sanders. "Short randomizable signatures". In: *Cryptographers' Track at the RSA Conference*. Springer. 2016, pp. 111–126.

[32] Ludwik Fleck. *Genesis and development of a scientific fact*. University of Chicago Press, 2012.

[33] Alan H Karp, Harry Haury and Michael H Davis. "From ABAC to ZBAC: the evolution of access control models". In: *Journal of Information Warfare* 9.2 (2010), pp. 38–46.

[34] Manu Sporny, Grant Noble, Dave Longley, Dan Burnett and Brent Zundel. *Verifiable Credentials Data Model 1.0*. 2021. URL: https://w3c.github.io/vc-data-model/.

[35] Manu Sporny, Dave Longley, Markus Sabadello, Drummond Reed, Orie Steele and Christopher Allen. *Decentralized Identifiers (DIDs) 1.0*. 2021. URL: https://w3c.github.io/did-core/.

[36] Stefan Brands. *Rethinking public key infrastructures and digital certificates: building in privacy*. Mit Press, 2000.

[37] Jesús García-Rodríguez, Rafael Torres Moreno, Jorge Bernal Bernabé and Antonio Skarmeta. "Towards a standardized model for privacy-preserving Verifiable Credentials". In: *The 16th International Conference on Availability, Reliability and Security*. 2021, pp. 1–6.

[38] Paul A Grassi, Michael E Garcia and James L Fenton. "NIST Special Publication 800-63-3. Digital Identity Guidelines". In: *National Institute of Standards and Technology, Los Altos, CA* (2017).

[39] Peter J Burke. "Identities and social structure: The 2003 Cooley-Mead award address". In: *Social psychology quarterly* 67.1 (2004), pp. 5–15.

[40] Caribou Digital. *Identities: New Practices in a connected age*. 2017. URL: https://identitiesproject.com.

[41] Will Abramson, Nicole E van Deursen and William J Buchanan. "Trust-by-Design: Evaluating Issues and Perceptions within Clinical Passporting". In: *Blockchain in Healthcare Today* (2020).

[42] W Abramson, N Hickman and N Spencer. "Evaluating Trust Assurance in Indy-Based Identity Networks Using Public Ledger Data". In: *Front. Blockchain 4: 622090. doi: 10.3389/fbloc* (2021).

[43] Will Abramson, Pavlos Papadopoulos, Nikolaos Pitropakis and William J Buchanan. "PyDentity: A playground for education and experimentation with the Hyper-

ledger verifiable information exchange platform". In: *Software Impacts* 9 (2021), p. 100101.

[44] Gregory Bateson. *Steps to an ecology of mind: Collected essays in anthropology, psychiatry, evolution, and epistemology*. University of Chicago Press, 2000.

[45] Randy Connolly. "Why computing belongs within the social sciences". In: *Communications of the ACM* 63.8 (2020), pp. 54–59.

[46] Stephen Wilson. "Identities Evolve: Why Federated Identity is Easier Said than Done". In: *Available at SSRN 2163241* (2011).

[47] Philip Sheldrake. *Generative identity — beyond self-sovereignty*. 2019. URL: https://akasha.org/blog/2019/09/02/generative-identity-beyond-self-sovereignty.

[48] Andreas Pfitzmann and Marit Hansen. *A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management*. 2010.

[49] George Herbert Mead. *Mind, self and society*. Vol. 111. Chicago University of Chicago Press., 1934.

[50] Sheldon Stryker. "Identity salience and role performance: The relevance of symbolic interaction theory for family research". In: *Journal of Marriage and the Family* (1968), pp. 558–564.

[51] Lynn Margulis and Dorion Sagan. *What is life?* Univ of California Press, 2000.

[52] Margaret J Wheatley and Myron E Rogers. *A simpler way*. Berrett-Koehler Publishers, 1998.

[53] Kwame Anthony Appiah. *The lies that bind: Rethinking identity*. Profile Books, 2018.

[54] Herbert A Simon. "Rationality as process and as product of thought". In: *The American economic review* 68.2 (1978), pp. 1–16.

[55] Herbert A Simon. "A behavioral model of rational choice". In: *The quarterly journal of economics* 69.1 (1955), pp. 99–118.

[56] Herbert H Clark and Susan E Brennan. “Grounding in communication.” In: (1991).

[57] Gregory Bateson. “Problems in cetacean and other mammalian communication”. In: *Prehistory* (1966), p. 303.

[58] University of Oklahoma. Institute of Group Relations and Muzafer Sherif. *Intergroup conflict and cooperation: The Robbers Cave experiment*. Vol. 10. University Book Exchange Norman, OK, 1961.

[59] Susan A Gelman. “Psychological essentialism in children”. In: *Trends in cognitive sciences* 8.9 (2004), pp. 404–409.

[60] Sheldon Stryker. “Identity theory: Developments and extensions.” In: (1987).

[61] Jan E Stets and Peter J Burke. “Identity theory and social identity theory”. In: *Social psychology quarterly* (2000), pp. 224–237.

[62] Peter J Burke. “Identity control theory”. In: *The Blackwell encyclopedia of sociology* (2007).

[63] Lee Freese and Peter J Burke. “Persons, identities, and social interaction”. In: *Advances in Group Processes*. 1994.

[64] Peter J Burke. “Identity processes and social stress”. In: *American sociological review* (1991), pp. 836–849.

[65] Christian Morgner and Michael King. “Trust and power”. In: John Wiley & Sons, 2018. Chap. Niklas Luhmann’s Sociological Enlightenment and it’s Realization in Trust and Power, p. 57.

[66] Robert Axelrod. “An evolutionary approach to norms”. In: *The American political science review* (1986), pp. 1095–1111.

[67] Peter J Burke, Jan E Stets and Scott V Savage. “Punishments and the dominance identity in networks”. In: *Social Science Research* 93 (2021), p. 102489.

[68] Peter J Burke. “Identity change”. In: *Social psychology quarterly* 69.1 (2006), pp. 81–96.

[69] Graeme S Cumming and John Collier. “Change and identity in complex systems”. In: *Ecology and society* 10.1 (2005).

[70] Onora O’Neill. “Trust and accountability in a digital age”. In: *Philosophy* 95.1 (2020), pp. 3–17.

[71] Carl Shapiro, Shapiro Carl, Hal R Varian et al. *Information rules: A strategic guide to the network economy*. Harvard Business Press, 1998.

[72] Shoshana Zuboff. “The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power”. In: Profile Books, 2019. Chap. Make Then Dance, pp. 293–328.

[73] Hal R Varian. “Beyond big data”. In: *Business Economics* 49.1 (2014), pp. 27–31.

[74] Shoshana Zuboff. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile Books, 2019.

[75] *Disinformation Then and Now*. podcast. 2021. URL: https://your-undivided-attention.simplecast.com/episodes/disinformation-then-and-now-93q8tqTq.

[76] Ariel Perkins. “The US Capitol insurrection shows we need to take conspiracy and distrust seriously”. In: *USApp–American Politics and Policy Blog* (2021).

[77] Hal Berghel. “Malice domestic: The Cambridge analytica dystopia”. In: *Computer* 51.5 (2018), pp. 84–89.

[78] Jeff Buechner and Herman T Tavani. “Trust and multi-agent systems: applying the “diffuse, default model” of trust to experiments involving artificial agents”. In: *Ethics and information Technology* 13.1 (2011), pp. 39–51.

[79] Cathy O’neil. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown, 2016.

[80] Cher Tan. “Film: Coded Bias-slave to the Algorithm”. In: *Big Issue Australia* 617 (2020), pp. 30–31.

[81] Ruha Benjamin. “Race after technology: Abolitionist tools for the new jim code”. In: *Social Forces* (2019).

[82] Katherine Hawley. “Trust, distrust and commitment”. In: *Noûs* 48.1 (2014), pp. 1–20.

[83] Niklas Luhmann. *Vertrauen*. Enke Stuttgart, 1968.

[84] Katherine Hawley. “Trustworthy groups and organizations”. In: *The philosophy of trust* (2017), pp. 230–250.

[85] Onora O’Neill. *A question of trust: The BBC Reith Lectures 2002*. Cambridge University Press, 2002.

[86] Onora O’Neill. “Linking trust to trustworthiness”. In: *International Journal of Philosophical Studies* 26.2 (2018), pp. 293–300.

[87] J Adam Carter and Mona Simion. “The ethics and epistemology of trust”. In: *Internet Encyclopedia of Philosophy* (2020).

[88] David DeSteno. *The truth about trust: How it determines success in life, love, learning, and more*. Penguin, 2014.

[89] J David Lewis and Andrew J Weigert. “The social dynamics of trust: Theoretical and empirical research, 1985-2012”. In: *Social forces* 91.1 (2012), pp. 25–31.

[90] Andrew J Weigert. “Pragmatic trust in a world of strangers: trustworthy actions”. In: *Comparative Sociology* 10.3 (2011), pp. 321–336.

[91] Kari Chopra and William A Wallace. “Trust in electronic environments”. In: *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the*. IEEE. 2003, 10–pp.

[92] Annette Baier. “Trust and antitrust”. In: *ethics* 96.2 (1986), pp. 231–260.

[93] PJ Nickel and K Vaesen. “Risk and trust”. In: *Handbook of risk theory: epistemology, decision theory, ethics and social implications of risk* (2012), pp. 857–876.

[94] Katherine Hawley. “Trust and distrust between patient and doctor”. In: *Journal of evaluation in clinical practice* 21.5 (2015), pp. 798–801.

[95] Margaret Urban Walker. *Moral repair: Reconstructing moral relations after wrongdoing*. Cambridge University Press, 2006.

[96] Yaobin Lu, Ling Zhao and Bin Wang. “From virtual community members to C2C e-commerce buyers: Trust in virtual communities and its effect on consumers’ purchase intention”. In: *Electronic Commerce Research and Applications* 9.4 (2010), pp. 346–360.

[97] Susan P Shapiro. “The social control of impersonal trust”. In: *American journal of Sociology* 93.3 (1987), pp. 623–658.

[98] Sven Ove Hansson. “Philosophical perspectives on risk”. In: *Techné: Research in Philosophy and Technology* 8.1 (2004), pp. 10–35.

[99] Catherine C Eckel and Rick K Wilson. “Is trust a risky decision?” In: *Journal of Economic Behavior & Organization* 55.4 (2004), pp. 447–465.

[100] Florian Hawlitschek, Benedikt Notheisen and Timm Teubner. “The limits of trust-free systems: A literature review on blockchain technology and trust in the sharing economy”. In: *Electronic commerce research and applications* 29 (2018), pp. 50–63.

[101] Esther Keymolen. “Trust and technology in collaborative consumption. Why it is not just about you and me”. In: *Bridging distances in technology and regulation* 135 (2013), pp. 135–150.

[102] Richard Wilson. “Cambridge analytica, Facebook, and Influence Operations: A case study and anticipatory ethical analysis”. In: *European conference on cyber warfare and security*. Academic Conferences International Limited. 2019.

[103] Camille François. “Actors, Behaviors, Content: A Disinformation ABC”. In: *Algorithms* (2020).

[104] Primavera De Filippi. “What blockchain means for the sharing economy”. In: *Harvard Business Review* 15 (2017).

[105] Florian Glaser. “Pervasive decentralisation of digital infrastructures: a framework for blockchain enabled system and use case analysis”. In: *Proceedings of the 50th Hawaii international conference on system sciences*. 2017.

[106] Phil Champagne. “The book of Satoshi”. In: *Lexington, KY: e53 Publishing* (2014).

[107] Caitlin Lustig and Bonnie Nardi. “Algorithmic authority: The case of Bitcoin”. In: *2015 48th Hawaii International Conference on System Sciences*. IEEE. 2015, pp. 743–752.

[108] Gili Vidan and Vili Lehdonvirta. “Mine the gap: Bitcoin and the maintenance of trustlessness”. In: *New Media & Society* 21.1 (2019), pp. 42–59.

[109] Florian Hawlitschek, Benedikt Notheisen and Timm Teubner. “A 2020 perspective on “The limits of trust-free systems: A literature review on blockchain technology and trust in the sharing economy””. In: *Electronic Commerce Research and Applications* 40 (2020), p. 100935.

[110] Benedikt Notheisen, Florian Hawlitschek and Christof Weinhardt. “Breaking down the blockchain hype–towards a blockchain market engineering approach”. In: (2017).

[111] Constantin Cătălin Drăgan and Mark Manulis. “Bootstrapping online trust: Timeline activity proofs”. In: *Data Privacy Management, Cryptocurrencies and Blockchain Technology*. Springer, 2018, pp. 242–259.

[112] Yifan Yang et al. “TAPESTRY: a de-centralized service for trusted interaction online”. In: *IEEE Transactions on Services Computing* (2020).

[113] Robert Axelrod. *The complexity of cooperation: Agent-based models of competition and collaboration*. Vol. 3. Princeton university press, 1997.

[114] Martin A Nowak and Karl Sigmund. “Tit for tat in heterogeneous populations”. In: *Nature* 355.6357 (1992), pp. 250–253.

[115] Barbara J Grosz, Sarit Kraus, David G Sullivan and Sanmay Das. “The influence of social norms and social consciousness on intention reconciliation”. In: *Artificial Intelligence* 142.2 (2002), pp. 147–177.

[116] David Vincent. *Privacy: a short history*. John Wiley & Sons, 2016.

[117] Andrew Trask, Emma Bluemke, Ben Garfinkel, Claudia Ghezzou Cuervas-Mons and Allan Dafoe. “Beyond Privacy Trade-offs with Structured Transparency”. In: *arXiv preprint arXiv:2012.08347* (2020).

[118] Susan Whyman. *The pen and the people: English letter writers 1660-1800*. Oxford University Press, 2009.

[119] Samuel D Warren and Louis D Brandeis. “Right to privacy”. In: *Harv. L. Rev.* 4 (1890), p. 193.

[120] Alan F Westin. “Privacy and freedom”. In: *Washington and Lee Law Review* 25.1 (1968), p. 166.

[121] Tom Gerety. “Redefining privacy”. In: *Harv. CR-CLL Rev.* 12 (1977), p. 233.

[122] Jeffrey H Reiman. “Privacy, intimacy, and personhood”. In: *Philosophy & Public Affairs* (1976), pp. 26–44.

[123] Ruth Gavison. “Privacy and the Limits of Law”. In: *The Yale law journal* 89.3 (1980), pp. 421–471.

[124] UN General Assembly et al. “Universal declaration of human rights”. In: *UN General Assembly* 302.2 (1948), pp. 14–25.

[125] William A Parent. “Privacy, morality, and the law”. In: *Philosophy & Public Affairs* (1983), pp. 269–288.

[126] Raymond Wacks. *Personal information: Privacy and the law*. Oxford: Clarendon Press, 1989.

[127] James B Rule. “Contextual Integrity and its Discontents: A Critique of Helen Nissenbaum’s Normative Arguments”. In: *Policy & Internet* 11.3 (2019), pp. 260–279.

[128] Jeffrey H Reiman. “Driving to the panopticon: A philosophical exploration of the risks to privacy posed by the highway technology of the future”. In: *Santa Clara Computer & High Tech. LJ* 11 (1995), p. 27.

[129] Julie E Cohen. “Examined lives: Informational privacy and the subject as object”. In: *Stan. L. Rev.* 52 (1999), p. 1373.

[130] Charles Fried. “Privacy “A moral analysis”. 77 Yale LJ 475 (1969), reprint in Ferdinard D”. In: *Philosophical Dimensions of privacy* 1 (1984), p. 984.

[131] James Rachels. “Why privacy is important”. In: *Philosophy & Public Affairs* (1975), pp. 323–333.

[132] Priscilla M Regan. *Legislating privacy: Technology, social values, and public policy*. Univ of North Carolina Press, 2000.

[133] Darrell O’Donnell. “The Current and Future State of Digital Wallets”. In: *Continuum Loop Inc* (2019), p. 83.

[134] Krishna Bharat, Stephen Lawrence and Mehran Sahami. *Generating user information for use in targeted advertising*. US Patent 9,235,849. 2016.

[135] Lauren H Rakower. “Blurred line: zooming in on Google Street View and the global right to privacy”. In: *Brook. J. Int’l L.* 37 (2011), p. 317.

[136] *Investigations of Google Street View*. URL: `https://epic.org/privacy/streetview/`.

[137] Arnold Roosendaal. “Facebook tracks and traces everyone: Like this!” In: *Tilburg Law School Legal Studies Research Paper Series* 03 (2011).

[138] Jennifer Golbeck, Cristina Robles and Karen Turner. “Predicting personality with social media”. In: *CHI’11 extended abstracts on human factors in computing systems*. ACM. 2011, pp. 253–262.

[139] Jennifer Golbeck, Cristina Robles, Michon Edmondson and Karen Turner. “Predicting personality from twitter”. In: *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*. IEEE. 2011, pp. 149–156.

[140] Michal Kosinski, David Stillwell and Thore Graepel. “Private traits and attributes are predictable from digital records of human behavior”. In: *Proceedings of the National Academy of Sciences* 110.15 (2013), pp. 5802–5805.

[141] Wu Youyou, Michal Kosinski and David Stillwell. “Computer-based personality judgments are more accurate than those made by humans”. In: *Proceedings of the National Academy of Sciences* 112.4 (2015), pp. 1036–1040.

[142] Victoria R Brown and E Daly Vaughn. “The writing on the (Facebook) wall: The use of social networking sites in hiring decisions”. In: *Journal of Business and psychology* 26.2 (2011), p. 219.

[143] Yanhao Wei, Pinar Yildirim, Christophe Van den Bulte and Chrysanthos Dellarocas. “Credit scoring with social network data”. In: *Marketing Science* 35.2 (2015), pp. 234–258.

[144] Lyria Bennett Moses and Janet Chan. “Algorithmic prediction in policing: assumptions, evaluation, and accountability”. In: *Policing and society* 28.7 (2018), pp. 806–822.

[145] Frederik Zuiderveen Borgesius, Judith Möller, Sanne Kruikemeier, Ronan Ó Fathaigh, Kristina Irion, Tom Dobber, Balazs Bodo and Claes H de Vreese. “Online political microtargeting: Promises and threats for democracy”. In: *Utrecht Law Review* 14.1 (2018), pp. 82–96.

[146] Joy Buolamwini and Timnit Gebru. “Gender shades: Intersectional accuracy disparities in commercial gender classification”. In: *Conference on fairness, accountability and transparency*. PMLR. 2018, pp. 77–91.

[147] David Chaum. “Security without identification: transaction systems to make big brother obsolete”. en. In: *Communications of the ACM* 28.10 (Oct. 1985), pp. 1030–1044. ISSN: 00010782. (Visited on 06/11/2018).

[148] Eric Hughes. “A cypherpunk’s manifesto”. In: *Crypto anarchy, cyberstates, and pirate utopias* (1993), pp. 81–83.

[149] David L Chaum. “Untraceable electronic mail, return addresses, and digital pseudonyms”. In: *Communications of the ACM* 24.2 (1981), pp. 84–90.

[150] David Chaum. “Blind signatures for untraceable payments”. In: *Advances in cryptology*. Springer. 1983, pp. 199–203.

[151] David Chaum, Amos Fiat and Moni Naor. “Untraceable electronic cash”. In: *Conference on the Theory and Application of Cryptography*. Springer. 1988, pp. 319–327.

[152] Andrew C Yao. “Protocols for secure computations”. In: *23rd annual symposium on foundations of computer science (sfcs 1982)*. IEEE. 1982, pp. 160–164.

[153] Ronald L Rivest, Len Adleman, Michael L Dertouzos et al. “On data banks and privacy homomorphisms”. In: *Foundations of secure computation* 4.11 (1978), pp. 169–180.

[154] David Chaum and Eugène Van Heyst. “Group signatures”. In: *Workshop on the Theory and Application of of Cryptographic Techniques*. Springer. 1991, pp. 257–265.

[155] Torben Pryds Pedersen. “Non-interactive and information-theoretic secure verifiable secret sharing”. In: *Annual international cryptology conference*. Springer. 1991, pp. 129–140.

[156] Shafi Goldwasser, Silvio Micali and Charles Rackoff. “The knowledge complexity of interactive proof systems”. In: *SIAM Journal on computing* 18.1 (1989), pp. 186–208.

[157] Adi Shamir. “Identity-based cryptosystems and signature schemes”. In: *Workshop on the theory and application of cryptographic techniques*. Springer. 1984, pp. 47–53.

[158] Claus-Peter Schnorr. “Efficient signature generation by smart cards”. In: *Journal of cryptology* 4.3 (1991), pp. 161–174.

[159] Helen Nissenbaum. “Contextual integrity up and down the data food chain”. In: *Theoretical Inquiries in Law* 20.1 (2019), pp. 221–256.

[160] Cynthia Dwork. “Differential privacy: A survey of results”. In: *International conference on theory and applications of models of computation*. Springer. 2008, pp. 1–19.

[161] Craig Gentry et al. *A fully homomorphic encryption scheme*. Vol. 20. 9. Stanford university Stanford, 2009.

[162] Qinbin Li, Zeyi Wen, Zhaomin Wu, Sixu Hu, Naibo Wang, Yuan Li, Xu Liu and Bingsheng He. “A survey on federated learning systems: vision, hype and reality for data privacy and protection”. In: *arXiv preprint arXiv:1907.09693* (2019).

[163] Eli Ben Sasson, Alessandro Chiesa, Christina Garman, Matthew Green, Ian Miers, Eran Tromer and Madars Virza. “Zerocash: Decentralized anonymous payments from bitcoin”. In: *2014 IEEE Symposium on Security and Privacy*. IEEE. 2014, pp. 459–474.

[164] Daira Hopwood, Sean Bowe, Taylor Hornby and Nathan Wilcox. “Zcash protocol specification”. In: *GitHub: San Francisco, CA, USA* (2016).

[165] Johnson Bobbie. “Privacy no longer a social norm, says Facebook founder”. In: *The Guardian* (11th Jan. 2010). URL: https://www.theguardian.com/technology/2010/jan/11/facebook-privacy.

[166] Paul Voigt and Axel Von dem Bussche. “The eu general data protection regulation (gdpr)”. In: *A Practical Guide, 1st Ed., Cham: Springer International Publishing* (2017).

[167] Elizabeth Liz Harding, Jarno J Vanto, Reece Clark, L Hannah Ji and Sara C Ainsworth. “Understanding the scope and impact of the California Consumer Privacy Act of 2018”. In: *Journal of Data Protection & Privacy* 2.3 (2019), pp. 234–253.

[168] Michael Seadle. “The great hack (documentary film). Produced and directed by Karim Amer and Jehane Noujaim. Netflix, 2019. 1 hour 54 minutes”. In: *Journal of the Association for Information Science and Technology* 71.12 (2020), pp. 1507–1511.

[169] Shuili Du. “Reimagining the Future of Technology:“The Social Dilemma” Review”. In: *Journal of Business Ethics* (2021), pp. 1–3.

[170] Sidney Perkowitz. “The Bias in the Machine: Facial Recognition Technology and Racial Disparities”. In: *MIT Case Studies in Social and Ethical Responsibilities of Computing* (2021).

[171] Edgar A Whitley, Uri Gal and Annemette Kjaergaard. *Who do you think you are? A review of the complex interplay between information systems, identification and identity*. 2014.

[172] Reetika Khera. *Dissent on Aadhaar: Big data meets big brother*. Orient BlackSwan Hyderabad, 2019.

[173] David-Olivier Jaquet-Chiffelle, Emmanuel Benoist, Rolf Haenni, Florent Wenger and Harald Zwingelberg. "Virtual persons and identities". In: *The Future of identity in the Information Society*. Springer, 2009, pp. 75–122.

[174] Tony Collings. "Some thoughts on the underlying logic and process underpinning Electronic Identity (e-ID)". In: *Information security technical report* 13.2 (2008), pp. 61–70.

[175] Jorge Bernal Bernabe, Martin David, Rafael Torres Moreno, Javier Presa Cordero, Sébastien Bahloul and Antonio Skarmeta. "ARIES: Evaluation of a reliable and privacy-preserving European identity management framework". In: *Future Generation Computer Systems* 102 (2020), pp. 409–425.

[176] Bennett Cyphers and Cory Doctorow. *Privacy Without Monopoly: Data Protection and Interoperability*. Tech. rep. 2021. URL: `https://www.eff.org/wp/interoperability-and-privacy`.

[177] Rogers Brubaker and Frederick Cooper. "Beyond" identity"". In: *Theory and society* 29.1 (2000), pp. 1–47.

[178] Roland Schlöglhofer and Johannes Sametinger. "Secure and usable authentication on mobile devices". In: *Proceedings of the 10th International Conference on Advances in Mobile Computing & Multimedia*. 2012, pp. 257–262.

[179] Rui A Martins, Alexandre B Augusto and Manuel E Correia. "A Potpourri of authentication mechanisms The mobile device way". In: *2013 8th Iberian Conference on Information Systems and Technologies (CISTI)*. IEEE. 2013, pp. 1–6.

[180] Alex Preukschat and Drummond Reed. *Self-sovereign identity*. Manning Publications, 2021.

[181] Laura DeNardis. *The Internet in Everything*. Yale University Press, 2020.

[182] Linnet Taylor, Gargi Sharma, Aaron Martin and Shazade Jameson. "Data justice and COVID-19". In: (2020).

[183] Audun Jøsang and Stéphane Lo Presti. "Analysing the relationship between risk and trust". In: *International conference on trust management*. Springer. 2004, pp. 135–145.

[184] Audun Jøsang, John Fabre, Brian Hay, James Dalziel and Simon Pope. "Trust requirements in identity management". In: *Proceedings of the 2005 Australasian workshop on Grid computing and e-research-Volume 44*. Australian Computer Society, Inc. 2005, pp. 99–108.

[185] Santosh Chokhani, Warwick Ford, Randy Sabett, Charles R Merrill and Stephen S Wu. "Internet X. 509 Public Key Infrastructure Certificate Policy and Certification Practices Framework." In: *RFC* 3647 (2003), pp. 1–94.

[186] Tariq Fadai, Sebastian Schrittwieser, Peter Kieseberg and Martin Mulazzani. "Trust me, I'm a Root CA! Analyzing SSL Root CAs in Modern Browsers and Operating Systems". In: *2015 10th International Conference on Availability, Reliability and Security*. IEEE. 2015, pp. 174–179.

[187] Jake A Berkowsky and Thaier Hayajneh. "Security issues with certificate authorities". In: *2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*. IEEE. 2017, pp. 449–455.

[188] *Twitter says bug exposed user plaintext passwords*. URL: `https://www.zdnet.com/article/twitter-says-bug-exposed-passwords-in-plaintext/`.

[189] Catalin Cimpanu. *US government releases post-mortem report on Equifax hack*. URL: `https://www.zdnet.com/article/us-government-releases-post-mortem-report-on-equifax-hack/`.

[190] Jason Hong. "The state of phishing attacks". In: *Communications of the ACM* 55.1 (2012), pp. 74–81.

[191] Matteo Dell'Amico, Pietro Michiardi and Yves Roudier. "Password strength: An empirical analysis". In: *2010 Proceedings IEEE INFOCOM*. IEEE. 2010, pp. 1–9.

[192] Steven Furnell. "Assessing website password practices–over a decade of progress?" In: *Computer Fraud & Security* 2018.7 (2018), pp. 6–13.

[193] Kim Cameron. "The laws of identity". In: *Microsoft Corp* 12 (2005), pp. 8–11.

[194] David Recordon and Drummond Reed. "OpenID 2.0: a platform for user-centric identity management". In: *Proceedings of the second ACM workshop on Digital identity management*. 2006, pp. 11–16.

[195] Barry Leiba. "Oauth web authorization protocol". In: *IEEE Internet Computing* 16.1 (2012), pp. 74–77.

[196] Wanpeng Li and Chris J Mitchell. "Analysing the Security of Google's implementation of OpenID Connect". In: *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*. Springer. 2016, pp. 357–376.

[197] Felix Richter. *https://www.weforum.org/agenda/2019/02/how-facebook-grew-from-0-to-2-3-billion-users-in-15-years*. 2019. URL: https://www.weforum.org/agenda/2019/02/how-facebook-grew-from-0-to-2-3-billion-users-in-15-years.

[198] Scott Gilbertson. "OpenID: The Web's Most Successful Failure". In: (2011). ISSN: 1059-1028. URL: https://www.wired.com/2011/01/openid-the-webs-most-successful-failure/.

[199] Moo Nam Ko, Gorrell P Cheek, Mohamed Shehab and Ravi Sandhu. "Social-networks connect services". In: *Computer* 43.8 (2010), pp. 37–43.

[200] Satoshi Nakamoto. "A Peer-to-Peer Electronic Cash System". In: (2008). URL: https://bitcoin.org/bitcoin.pdf.

[201] D Madavi. "A comprehensive study on blockchain technology". In: *International Research Journal of Engineering and Technology* 6.1 (2019), pp. 1765–1770.

[202] Christopher Allen et al. "Decentralized public key infrastructure". In: *Group Report. Rebooting the Web of Trust (RWOT)* (2015).

[203] Christopher Allen. “The path to self-sovereign identity”. In: *Life with Alacrity* (2016).

[204] Matthew Schutte. *Schutte’s Critique of the Self-Sovereign Identity Principles*. 2016. URL: http://matthewschutte.com/2016/10/25/schuttes-critique-of-the-self-sovereign-identity-principles.

[205] *Principles of SSI*. URL: https://sovrin.org/principles-of-ssi/.

[206] Kalman C Toth and Alan Anderson-Priddy. “Self-Sovereign Digital Identity: A Paradigm Shift for Identity”. In: *IEEE Security & Privacy* 17.3 (2019), pp. 17–27.

[207] Frederico Schardong and Ricardo Custódio. “Self-Sovereign Identity: A Systematic Map and Review”. In: *arXiv preprint arXiv:2108.08338* (2021).

[208] Georgy Ishmaev. “Sovereignty, privacy, and ethics in blockchain-based identity management systems”. In: *Ethics and Information Technology* 23.3 (2021), pp. 239–252.

[209] Abhilasha Bhargav-Spantzel, Jan Camenisch, Thomas Gross and Dieter Sommer. “User centricity: a taxonomy and open issues”. In: *Journal of Computer Security* 15.5 (2007), pp. 493–527.

[210] N Neubauer and L Liu. “Guardianship and self-sovereign identity: implications for persons living with dementia”. In: *Innovation in Aging* 5.Suppl 1 (2021), pp. 718–718.

[211] Alexander Mühle, Andreas Grüner, Tatiana Gayvoronskaya and Christoph Meinel. “A survey on essential components of a self-sovereign identity”. In: *Computer Science Review* 30 (2018), pp. 80–86.

[212] W3C Credential Community Group. *DID Method Registry*. Tech. rep. 2019. URL: https://w3c-ccg.github.io/did-method-registry/.

[213] Kaliya Young. *Verifiable Credential Flavors Explained*. Tech. rep. 2021.

[214] Man Ho Au, Willy Susilo and Yi Mu. “Constant-size dynamic k-TAA”. In: *International conference on security and cryptography for networks*. Springer. 2006, pp. 111–125.

[215] Tobais Looker and Orie Steele. *BBS+ Signatures 2020*. 2020. URL: https://w3c-ccg.github.io/ldp-bbs2020/.

[216] Daniel Hardman. *DIDComm Messaging*. 2021. URL: https://identity.foundation/didcomm-messaging/spec/.

[217] Dave Longley and Manu Sporny. *Credential Handler API 1.0*. 2021. URL: https://w3c-ccg.github.io/credential-handler-api/.

[218] Daniel Buchner, Brent Zundel and Mark Riedel. *Presentation Exchange v1.0.0*. URL: https://identity.foundation/presentation-exchange/spec/v1.0.0/.

[219] Artemij Voskobojnikov, Oliver Wiese, Masoud Mehrabi Koushki, Volker Roth and Konstantin Beznosov. "The U in Crypto Stands for Usable: An Empirical Study of User Experience with Mobile Cryptocurrency Wallets". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021, pp. 1–14.

[220] Rieks Joosten, Sterre den Breeijen and Drummond Reed. *Decentralized SSI Governance, the missing link in automating business decisions*. 2021.

[221] Darrell O'Donnell. *Trust Registries*. 2021. URL: https://www.youtube.com/watch?v=b_xG2un5WaI.

[222] *Introducing the Trust over IP Foundation*. 2020. URL: https://trustoverip.org/wp-content/uploads/2020/05/toip_introduction_050520.pdf.

[223] Tim Marshall. *Prisoners of geography: ten maps that explain everything about the world*. Vol. 1. Simon and Schuster, 2016.

[224] Aaron Martin and Linnet Taylor. "Exclusion and inclusion in identification: Regulation, displacement and data justice". In: *Information Technology for Development* 27.1 (2021), pp. 50–66.

[225] John Effah and Emmanuel Owusu-Oware. "From national to sector level biometric systems: the case of Ghana". In: *Information Technology for Development* 27.1 (2021), pp. 91–110.

[226] Shyam Divam. “Aadhaar’s Biometric Tsunami”. In: *Dissent on Aadhaar: Big data meets big brother*. Orient BlackSwan Hyderabad, 2019, pp. 124–136.

[227] Sunil Abraham. “Surveillance Project”. In: *Dissent on Aadhaar: Big data meets big brother*. Orient BlackSwan Hyderabad, 2019, pp. 86–94.

[228] M. S. Sriram. “Public Investments and Private Profits”. In: *Dissent on Aadhaar: Big data meets big brother*. Orient BlackSwan Hyderabad, 2019, pp. 187–202.

[229] Usa Ramanathan. “Aadhaar - From Welfare to Profit”. In: *Dissent on Aadhaar: Big data meets big brother*. Orient BlackSwan Hyderabad, 2019, pp. 173–186.

[230] Prasanna S. “Aadhaar-Constitutionally Challenged”. In: *Dissent on Aadhaar: Big data meets big brother*. Orient BlackSwan Hyderabad, 2019, pp. 137–148.

[231] Gautam Bhatua. “The Privacy Judgement”. In: *Dissent on Aadhaar: Big data meets big brother*. Orient BlackSwan Hyderabad, 2019, pp. 149–165.

[232] Anumeha Yadav. “On the Margins of Aadhaar”. In: *Dissent on Aadhaar: Big data meets big brother*. Orient BlackSwan Hyderabad, 2019, pp. 46–58.

[233] Bidisha Chaudhuri. “Distant, opaque and seamful: seeing the state through the workings of Aadhaar in India”. In: *Information Technology for Development* 27.1 (2021), pp. 37–49.

[234] Amiya Bhatia, Elizabeth Donger and Jacqueline Bhabha. “‘Without an Aadhaar card nothing could be done’: a mixed methods study of biometric identification and birth registration for children in Varanasi, India”. In: *Information Technology for Development* 27.1 (2021), pp. 129–149.

[235] Emrys Schoemaker, Dina Baslan, Bryan Pon and Nicola Dell. “Identity at the margins: data justice and refugee experiences with digital identity systems in Lebanon, Jordan, and Uganda”. In: *Information Technology for Development* 27.1 (2021), pp. 13–36.

[236] Bruce Schneier. *Click here to kill everybody: Security and survival in a hyper-connected world*. WW Norton & Company, 2018.

[237] B Clinton. “Executive Order 13026–administration of export controls on encryption products November 15, 1996.” In: *Weekly Compilation of Presidential Documents* 32.46 (1996), pp. 2399–2400.

[238] *Specification of the Identity Mixer Cryptographic Library*. Version 2.3.0. IBM Research - Zurich. Apr. 2010.

[239] Christian Paquin and Greg Zaverucha. “U-prove cryptographic specification v1. 1”. In: *Technical Report, Microsoft Corporation* (2011).

[240] Patrik Bichsel et al. “D2. 2 Architecture for attribute-based credential technologies-final version”. In: *ABC4TRUST project deliverable. Available online at https://abc4trust. eu/index. php/pub* (2014).

[241] Sébastien Canard and Jacques Traoré. “Pairing-Based Cryptography”. In: *Guide to Pairing-Based Cryptography*. Chapman and Hall/CRC, 2017, pp. 1–1.

[242] Ran Canetti. “Universally composable security: A new paradigm for cryptographic protocols”. In: *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*. IEEE. 2001, pp. 136–145.

[243] Jan Camenisch, Manu Drijvers and Maria Dubovitskaya. “Practical UC-secure delegatable credentials with attributes and their application to blockchain”. In: *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. 2017, pp. 683–699.

[244] Claude E Shannon. “Communication theory of secrecy systems”. In: *The Bell system technical journal* 28.4 (1949), pp. 656–715.

[245] Ronald L Rivest, Adi Shamir and Leonard Adleman. “A method for obtaining digital signatures and public-key cryptosystems”. In: *Communications of the ACM* 21.2 (1978), pp. 120–126.

[246] David Chaum. “Showing credentials without identification”. In: *Workshop on the Theory and Application of of Cryptographic Techniques*. Springer. 1985, pp. 241–244.

[247] David Chaum. “Blind signature system”. In: *Advances in cryptology*. Springer. 1984, pp. 153–153.

[248] Lidong Chen. “Access with pseudonyms”. In: *International Conference on Cryptography: Policy and Algorithms*. Springer. 1995, pp. 232–243.

[249] Anna Lysyanskaya, Ronald L Rivest, Amit Sahai and Stefan Wolf. “Pseudonym systems”. In: *International Workshop on Selected Areas in Cryptography*. Springer. 1999, pp. 184–199.

[250] David Chaum and Jan-Hendrik Evertse. “A secure and privacy-protecting protocol for transmitting personal information between organizations”. In: *Conference on the Theory and Application of Cryptographic Techniques*. Springer. 1986, pp. 118–167.

[251] David Chaum. “Elections with unconditionally-secret ballots and disruption equivalent to breaking RSA”. In: *Workshop on the Theory and Application of of Cryptographic Techniques*. Springer. 1988, pp. 177–182.

[252] Ivan Bjerre Damgård. “Payment systems and credential mechanisms with provable security against abuse by individuals”. In: *Conference on the Theory and Application of Cryptography*. Springer. 1988, pp. 328–335.

[253] Amos Fiat and Adi Shamir. “How to prove yourself: Practical solutions to identification and signature problems”. In: *Conference on the theory and application of cryptographic techniques*. Springer. 1986, pp. 186–194.

[254] Jan Camenisch and Anna Lysyanskaya. “Dynamic accumulators and application to efficient revocation of anonymous credentials”. In: *Annual international cryptology conference*. Springer. 2002, pp. 61–76.

[255] Taher ElGamal. “A public key cryptosystem and a signature scheme based on discrete logarithms”. In: *IEEE transactions on information theory* 31.4 (1985), pp. 469–472.

[256] David Chaum, Jan-Hendrik Evertse and Jeroen Van De Graaf. “An improved protocol for demonstrating possession of discrete logarithms and some generalizations”. In: *Workshop on the Theory and Application of of Cryptographic Techniques*. Springer. 1987, pp. 127–141.

[257] Yvo Desmedt. “Society and group oriented cryptography: A new concept”. In: *Conference on the Theory and Application of Cryptographic Techniques*. Springer. 1987, pp. 120–127.

[258] Jan Camenisch. “Efficient and generalized group signatures”. In: *International Conference on the Theory and Applications of Cryptographic Techniques*. Springer. 1997, pp. 465–479.

[259] Jan Camenisch and Markus Stadler. “Efficient group signature schemes for large groups”. In: *Annual International Cryptology Conference*. Springer. 1997, pp. 410–424.

[260] Giuseppe Ateniese, Jan Camenisch, Marc Joye and Gene Tsudik. “A practical and provably secure coalition-resistant group signature scheme”. In: *Annual International Cryptology Conference*. Springer. 2000, pp. 255–270.

[261] Bart Preneel. “The first 30 years of cryptographic hash functions and the NIST SHA-3 competition”. In: *Cryptographers’ track at the RSA conference*. Springer. 2010, pp. 1–14.

[262] Ralph Charles Merkle. *Secrecy, authentication, and public key systems*. Stanford university, 1979.

[263] Ivan Bjerre Damgård. “Collision free hash functions and public key signature schemes”. In: *Workshop on the Theory and Application of of Cryptographic Techniques*. Springer. 1987, pp. 203–216.

[264] Xiaoyun Wang and Hongbo Yu. “How to break MD5 and other hash functions”. In: *Annual international conference on the theory and applications of cryptographic techniques*. Springer. 2005, pp. 19–35.

[265] Morris J Dworkin. “SHA-3 standard: Permutation-based hash and extendable-output functions”. In: (2015).

[266] Keith M Martin. *Everyday cryptography*. Oxford University Press, 2017.

[267] David Chaum. “Demonstrating that a public predicate can be satisfied without revealing any information about how”. In: *Conference on the Theory and Application of Cryptographic Techniques*. Springer. 1986, pp. 195–199.

[268] László Babai and Shlomo Moran. “Arthur-Merlin games: a randomized proof system, and a hierarchy of complexity classes”. In: *Journal of Computer and System Sciences* 36.2 (1988), pp. 254–276.

[269] Gilles Brassard and Claude Crepeau. “Non-transitive transfer of confidence: A perfect zero-knowledge interactive protocol for SAT and beyond”. In: *27th Annual Symposium on Foundations of Computer Science (sfcs 1986)*. IEEE. 1986, pp. 188–195.

[270] Manuel Blum, Alfredo De Santis, Silvio Micali and Giuseppe Persiano. “Noninteractive zero-knowledge”. In: *SIAM Journal on Computing* 20.6 (1991), pp. 1084–1118.

[271] David Chaum and Torben Pryds Pedersen. “Wallet databases with observers”. In: *Annual International Cryptology Conference*. Springer. 1992, pp. 89–105.

[272] Jan Camenisch and Markus Stadler. “Proof systems for general statements about discrete logarithms”. In: *Technical Report/ETH Zurich, Department of Computer Science* 260 (1997).

[273] Jan Camenisch and Markus Michels. “Proving in zero-knowledge that a number is the product of two safe primes”. In: *International Conference on the Theory and Applications of Cryptographic Techniques*. Springer. 1999, pp. 107–122.

[274] Ronald Cramer and Ivan Damgård. “Zero-knowledge proofs for finite field arithmetic, or: Can zero-knowledge be for free?” In: *Annual International Cryptology Conference*. Springer. 1998, pp. 424–441.

[275] Josh Benaloh and Michael De Mare. “One-way accumulators: A decentralized alternative to digital signatures”. In: *Workshop on the Theory and Application of of Cryptographic Techniques*. Springer. 1993, pp. 274–285.

[276] Ralph C Merkle. “Protocols for public key cryptosystems”. In: *1980 IEEE Symposium on Security and Privacy*. IEEE. 1980, pp. 122–122.

[277] Philippe Camacho, Alejandro Hevia, Marcos Kiwi and Roberto Opazo. “Strong accumulators from collision-resistant hashing”. In: *International Conference on Information Security*. Springer. 2008, pp. 471–486.

[278] Stuart Haber and W Scott Stornetta. “How to time-stamp a digital document”. In: *Conference on the Theory and Application of Cryptography*. Springer. 1990, pp. 437–455.

[279] Neal Koblitz. “Elliptic curve cryptosystems”. In: *Mathematics of computation* 48.177 (1987), pp. 203–209.

[280] Victor S Miller. “Use of elliptic curves in cryptography”. In: *Conference on the theory and application of cryptographic techniques*. Springer. 1985, pp. 417–426.

[281] Neal Koblitz, Alfred Menezes and Scott Vanstone. “The state of elliptic curve cryptography”. In: *Designs, codes and cryptography* 19.2 (2000), pp. 173–193.

[282] Antoine Joux. “A one round protocol for tripartite Diffie–Hellman”. In: *International algorithmic number theory symposium*. Springer. 2000, pp. 385–393.

[283] Alfred Menezes. “An introduction to pairing-based cryptography”. In: *Recent trends in cryptography* 477 (2009), pp. 47–65.

[284] Jacques Stern. “Lattices and cryptography: An overview”. In: *International Workshop on Public Key Cryptography*. Springer. 1998, pp. 50–54.

[285] Peter W Shor. “Algorithms for quantum computation: discrete logarithms and factoring”. In: *Proceedings 35th annual symposium on foundations of computer science*. Ieee. 1994, pp. 124–134.

[286] Chris Peikert. “A decade of lattice cryptography”. In: *Foundations and Trends® in Theoretical Computer Science* 10.4 (2016), pp. 283–424.

[287] Dan Boneh. “The decision diffie-hellman problem”. In: *International Algorithmic Number Theory Symposium*. Springer. 1998, pp. 48–63.

[288] Niko Barić and Birgit Pfitzmann. “Collision-free accumulators and fail-stop signature schemes without trees”. In: *International conference on the theory and applications of cryptographic techniques*. Springer. 1997, pp. 480–494.

[289] Ronald Cramer and Victor Shoup. “Signature schemes based on the strong RSA assumption”. In: *ACM Transactions on Information and System Security (TISSEC)* 3.3 (2000), pp. 161–185.

[290] Shafi Goldwasser, Silvio Micali and Ronald L Rivest. “A digital signature scheme secure against adaptive chosen-message attacks”. In: *SIAM Journal on computing* 17.2 (1988), pp. 281–308.

[291] Mihir Bellare, Daniele Micciancio and Bogdan Warinschi. “Foundations of group signatures: Formal definitions, simplified requirements, and a construction based on general assumptions”. In: *International Conference on the Theory and Applications of Cryptographic Techniques*. Springer. 2003, pp. 614–629.

[292] Mihir Bellare, Haixia Shi and Chong Zhang. “Foundations of group signatures: The case of dynamic groups”. In: *Cryptographers’ Track at the RSA Conference*. Springer. 2005, pp. 136–153.

[293] Lidong Chen and Torben P Pedersen. “New group signature schemes”. In: *Workshop on the Theory and Application of of Cryptographic Techniques*. Springer. 1994, pp. 171–181.

[294] Mihir Bellare and Phillip Rogaway. “Random oracles are practical: A paradigm for designing efficient protocols”. In: *Proceedings of the 1st ACM conference on Computer and communications security*. 1993, pp. 62–73.

[295] Ran Canetti, Oded Goldreich and Shai Halevi. “The random oracle methodology, revisited”. In: *Journal of the ACM (JACM)* 51.4 (2004), pp. 557–594.

[296] Jan Camenisch and Els Van Herreweghen. “Design and implementation of the idemix anonymous credential system”. In: *Proceedings of the 9th ACM conference on Computer and communications security*. 2002, pp. 21–30.

[297] Benedikt Bünz, Jonathan Bootle, Dan Boneh, Andrew Poelstra, Pieter Wuille and Greg Maxwell. “Bulletproofs: Efficient range proofs for confidential transactions”. In: *Cryptology ePrint Archive, Report 2017/1066, Tech. Rep.* (2017).

[298] Jan Camenisch and Anna Lysyanskaya. “Signature schemes and anonymous credentials from bilinear maps”. In: *Annual International Cryptology Conference*. Springer. 2004, pp. 56–72.

[299] Man Ho Au, Patrick P Tsang, Willy Susilo and Yi Mu. “Dynamic universal accumulators for DDH groups and their application to attribute-based anonymous credential systems”. In: *Cryptographers’ track at the RSA conference*. Springer. 2009, pp. 295–308.

[300] Benoît Libert, San Ling, Fabrice Mouhartem, Khoa Nguyen and Huaxiong Wang. “Signature schemes with efficient protocols and dynamic group signatures from lattice assumptions”. In: *International Conference on the Theory and Application of Cryptology and Information Security*. Springer. 2016, pp. 373–403.

[301] Foteini Baldimtsi, Jan Camenisch, Maria Dubovitskaya, Anna Lysyanskaya, Leonid Reyzin, Kai Samelin and Sophia Yakoubov. “Accumulators with applications to anonymity-preserving revocation”. In: *2017 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE. 2017, pp. 301–315.

[302] Jan Camenisch, Markulf Kohlweiss and Claudio Soriente. “An accumulator based on bilinear maps and efficient revocation for anonymous credentials”. In: *International Workshop on Public Key Cryptography*. Springer. 2009, pp. 481–500.

[303] Jiangtao Li, Ninghui Li and Rui Xue. “Universal accumulators with efficient nonmembership proofs”. In: *International Conference on Applied Cryptography and Network Security*. Springer. 2007, pp. 253–269.

[304] Melissa Chase and Anna Lysyanskaya. “On signatures of knowledge”. In: *Annual International Cryptology Conference*. Springer. 2006, pp. 78–96.

[305] Mira Belenkiy, Jan Camenisch, Melissa Chase, Markulf Kohlweiss, Anna Lysyanskaya and Hovav Shacham. “Randomizable proofs and delegatable anonymous credentials”. In: *Annual International Cryptology Conference*. Springer. 2009, pp. 108–125.

[306] Jens Groth and Amit Sahai. “Efficient non-interactive proof systems for bilinear groups”. In: *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer. 2008, pp. 415–432.

[307] Georg Fuchsbauer. “Commuting Signatures and Verifiable Encryption and an Application to Non-Interactively Delegatable Credentials.” In: *IACR Cryptol. ePrint Arch.* 2010 (2010), p. 233.

[308] Jan Camenisch, Maria Dubovitskaya and Alfredo Rial. “UC commitments for modular protocol design and applications to revocation and attribute tokens”. In: *Annual International Cryptology Conference*. Springer. 2016, pp. 208–239.

[309] Zuoxia Yu, Man Ho Au and Rupeng Yang. “Accountable Anonymous Credentials”. In: *Advances in Cyber Security: Principles, Techniques, and Applications*. Springer, 2019, pp. 49–68.

[310] Michael Backes, Jan Camenisch and Dieter Sommer. “Anonymous yet accountable access control”. In: *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*. 2005, pp. 40–46.

[311] Jan Camenisch, Maria Dubovitskaya, Anja Lehmann, Gregory Neven, Christian Paquin and Franz-Stefan Preiss. “Concepts and languages for privacy-preserving attribute-based authentication”. In: *IFIP Working Conference on Policies and Research in Identity Management*. Springer. 2013, pp. 34–52.

[312] Jan Camenisch, Susan Hohenberger and Anna Lysyanskaya. “Compact e-cash”. In: *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer. 2005, pp. 302–321.

[313] Jan Camenisch, Susan Hohenberger, Markulf Kohlweiss, Anna Lysyanskaya and Mira Meyerovich. “How to win the clonewars: efficient periodic n-times anonymous authentication”. In: *Proceedings of the 13th ACM conference on Computer and communications security*. 2006, pp. 201–210.

[314] Yvo G Desmedt. “Threshold cryptography”. In: *European Transactions on Telecommunications* 5.4 (1994), pp. 449–458.

[315] Jan Camenisch, Manu Drijvers, Anja Lehmann, Gregory Neven and Patrick Towa. “Short threshold dynamic group signatures”. In: *International Conference on Security and Cryptography for Networks*. Springer. 2020, pp. 401–423.

[316] Jan Camenisch and Anja Lehmann. “(Un) linkable Pseudonyms for Governmental Databases”. In: *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. 2015, pp. 1467–1479.

[317] Jan Camenisch and Anja Lehmann. “Privacy-preserving user-auditable pseudonym systems”. In: *2017 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE. 2017, pp. 269–284.

[318] Lydia Garms and Anja Lehmann. “Group signatures with selective linkability”. In: *IACR International Workshop on Public Key Cryptography*. Springer. 2019, pp. 190–220.

[319] Ashley Fraser, Lydia Garms and Anja Lehmann. “Selectively Linkable Group Signatures-Stronger Security and Preserved Verifiability”. In: *Cryptology ePrint Archive* (2021).

[320] Tim Dierks, Christopher Allen et al. *The TLS protocol version 1.0*. 1999.

[321] Dindayal Mahto, Danish Ali Khan and Dilip Kumar Yadav. “Security analysis of elliptic curve cryptography and RSA”. In: *Proceedings of the world congress on engineering*. Vol. 1. 2016, pp. 419–422.

[322] Marco Carvalho, Jared DeMott, Richard Ford and David A Wheeler. “Heartbleed 101”. In: *IEEE security & privacy* 12.4 (2014), pp. 63–67.

[323] Kristof Verslype, Jorn Lapon, Pieter Verhaeghe, Vincent Naessens and Bart De Decker. “Petanon: A privacy-preserving e-petition system based on idemix”. In: *CW Reports* (2008).

[324] Jorn Lapon, Markulf Kohlweiss, Bart De Decker and Vincent Naessens. “Analysis of revocation strategies for anonymous Idemix credentials”. In: *IFIP International Conference on Communications and Multimedia Security*. Springer. 2011, pp. 3–17.

[325] Ibou Sene, Abdoul Aziz Ciss and Oumar Niang. “I2PA, U-prove, and Idemix: An Evaluation of Memory Usage and Computing Time Efficiency in an IoT Context”. In: *International Conference on e-Infrastructure and e-Services for Developing Countries*. Springer. 2019, pp. 140–153.

[326] Katharina Krombholz, Aljosha Judmayer, Matthias Gusenbauer and Edgar Weippl. “The other side of the coin: User experiences with bitcoin security and privacy”. In: *International conference on financial cryptography and data security*. Springer. 2016, pp. 555–580.

[327] Jan Camenisch, Sebastian Mödersheim, Gregory Neven, Franz-Stefan Preiss and Dieter Sommer. “A card requirements language enabling privacy-preserving access control”. In: *Proceedings of the 15th ACM symposium on Access control models and technologies*. 2010, pp. 119–128.

[328] Ahmad Sabouri and Kai Rannenberg. “ABC4Trust: protecting privacy in identity management by bringing privacy-ABCs into real-life”. In: *IFIP International Summer School on Privacy and Identity Management*. Springer. 2014, pp. 3–16.

[329] Vasiliki Liagkou, George Metakides, Apostolis Pyrgelis, Christoforos Raptopoulos, Paul Spirakis and Yannis C Stamatiou. “Privacy preserving course evaluations in Greek higher education institutes: an e-Participation case study with the empowerment of Attribute Based Credentials”. In: *Annual Privacy Forum*. Springer. 2012, pp. 140–156.

[330] Ahmad Sabouri, Souheil Bcheri and Kai Rannenberg. “Privacy-Respecting School Community Interaction Platform”. In: *Cyber Security and Privacy Forum*. Springer. 2014, pp. 108–119.

[331] Jesus Luna, Neeraj Suri and Ioannis Krontiris. “Privacy-by-design based on quantitative threat modeling”. In: *2012 7th International Conference on Risks and Security of Internet and Systems (CRiSIS)*. IEEE. 2012, pp. 1–8.

[332] Ahmad Sabouri, Ioannis Krontiris and Kai Rannenberg. “Trust relationships in privacy-ABCs’ ecosystems”. In: *International Conference on Trust, Privacy and Security in Digital Business*. Springer. 2014, pp. 13–23.

[333] Zinaida Benenson, Anna Girard and Ioannis Krontiris. “User Acceptance Factors for Anonymous Credentials: An Empirical Investigation.” In: *WEIS*. 2015.

[334] Fatbardh Veseli, Tsvetoslava Vateva-Gurova, Ioannis Krontiris, Kai Rannenberg and Neeraj Suri. “Towards a framework for benchmarking privacy-abc technologies”. In: *IFIP International Information Security Conference*. Springer. 2014, pp. 197–204.

[335] Fatbardh Veseli and Jetzabel Serna Olvera. “Benchmarking Privacy-ABC technologies-an evaluation of storage and communication efficiency”. In: *2015 IEEE World Congress on Services*. IEEE. 2015, pp. 198–205.

[336] Fatbardh Veseli and Jetzabel Serna. “Evaluation of privacy-ABC technologies-a study on the computational efficiency”. In: *IFIP International Conference on Trust Management*. Springer. 2016, pp. 63–78.

[337] Dan Boneh and Matt Franklin. “Identity-based encryption from the Weil pairing”. In: *Annual international cryptology conference*. Springer. 2001, pp. 213–229.

[338] Gregory Neven, Gianmarco Baldini, Jan Camenisch and Ricardo Neisse. “Privacy-preserving attribute-based credentials in cooperative intelligent transport systems”. In: *2017 IEEE Vehicular Networking Conference (VNC)*. IEEE. 2017, pp. 131–138.

[339] Dan Boneh, Xavier Boyen and Hovav Shacham. “Short group signatures”. In: *Annual International Cryptology Conference*. Springer. 2004, pp. 41–55.

[340] Jens Groth. “Short pairing-based non-interactive zero-knowledge arguments”. In: *International Conference on the Theory and Application of Cryptology and Information Security*. Springer. 2010, pp. 321–340.

[341] Shuichi Katsumata, Ryo Nishimaki, Shota Yamada and Takashi Yamakawa. “Compact NIZKs from Standard Assumptions on Bilinear Maps.” In: *IACR Cryptol. ePrint Arch.* 2020 (2020), p. 223.

[342] Masayuki Abe, Jens Groth, Kristiyan Haralambiev and Miyako Ohkubo. “Optimal structure-preserving signatures in asymmetric bilinear groups”. In: *Annual Cryptology Conference*. Springer. 2011, pp. 649–666.

[343] Masayuki Abe, Jens Groth, Markulf Kohlweiss, Miyako Ohkubo and Mehdi Tibouchi. “Efficient fully structure-preserving signatures and shrinking commitments”. In: *Journal of Cryptology* 32.3 (2019), pp. 973–1025.

[344] Dan Boneh, Ben Lynn and Hovav Shacham. “Short signatures from the Weil pairing”. In: *International conference on the theory and application of cryptology and information security*. Springer. 2001, pp. 514–532.

[345] Dan Boneh and Xavier Boyen. “Short signatures without random oracles”. In: *International conference on the theory and applications of cryptographic techniques*. Springer. 2004, pp. 56–73.

[346] Dan Boneh and Xavier Boyen. “Short signatures without random oracles and the SDH assumption in bilinear groups”. In: *Journal of cryptology* 21.2 (2008), pp. 149–177.

[347] Cameron F Kerry and Patrick D Gallagher. “Digital signature standard (DSS)”. In: *FIPS PUB* (2013), pp. 186–4.

[348] David Freeman, Michael Scott and Edlyn Teske. “A taxonomy of pairing-friendly elliptic curves”. In: *Journal of cryptology* 23.2 (2010), pp. 224–280.

[349] Michael Scott. “Pairing Implementation Revisited.” In: *IACR Cryptol. ePrint Arch.* 2019 (2019), p. 77.

[350] Paulo SLM Barreto, Ben Lynn and Michael Scott. “Constructing elliptic curves with prescribed embedding degrees”. In: *International Conference on Security in Communication Networks*. Springer. 2002, pp. 257–267.

[351] Paulo SLM Barreto and Michael Naehrig. “Pairing-friendly elliptic curves of prime order”. In: *International Workshop on Selected Areas in Cryptography*. Springer. 2005, pp. 319–331.

[352] Sean Bowe. “BLS12-381: New zk-SNARK elliptic curve construction”. In: *Zcash Company blog, URL: https://z. cash/blog/new-snark-curve* (2017).

[353] Sean Bowe. “Faster Subgroup Checks for BLS12-381.” In: *IACR Cryptol. ePrint Arch.* 2019 (2019), p. 814.

[354] Jan Camenisch, Maria Dubovitskaya, Kristiyan Haralambiev and Markulf Kohlweiss. “Composable and modular anonymous credentials: Definitions and practical constructions”. In: *International Conference on the Theory and Application of Cryptology and Information Security*. Springer. 2015, pp. 262–288.

[355] Foteini Badimtsi, Ran Canetti and Sophia Yakoubov. “Universally composable accumulators”. In: *Cryptographers’ Track at the RSA Conference*. Springer. 2020, pp. 638–666.

[356] Jan Camenisch, Manu Drijvers and Anja Lehmann. “Universally composable direct anonymous attestation”. In: *Public-Key Cryptography–PKC 2016*. Springer. 2016, pp. 234–264.

[357] Ran Canetti, Abhishek Jain and Alessandra Scafuro. “Practical UC security with a global random oracle”. In: *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. 2014, pp. 597–608.

[358] Jan Camenisch, Manu Drijvers, Tommaso Gagliardoni, Anja Lehmann and Gregory Neven. “The wonderful world of global random oracles”. In: *Annual*

International Conference on the Theory and Applications of Cryptographic Techniques. Springer. 2018, pp. 280–312.

[359] Alberto Sonnino, Mustafa Al-Bassam, Shehar Bano, Sarah Meiklejohn and George Danezis. “Coconut: Threshold issuance selective disclosure credentials with applications to distributed ledgers”. In: *arXiv preprint arXiv:1802.07344* (2018).

[360] Melissa Chase, Sarah Meiklejohn and Greg Zaverucha. “Algebraic MACs and keyed-verification anonymous credentials”. In: *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. 2014, pp. 1205–1216.

[361] Yevgeniy Dodis, Eike Kiltz, Krzysztof Pietrzak and Daniel Wichs. “Message authentication, revisited”. In: *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer. 2012, pp. 355–374.

[362] Jan Camenisch, Manu Drijvers, Petr Dzurenda and Jan Hajny. “Fast keyed-verification anonymous credentials on standard smart cards”. In: *IFIP International Conference on ICT Systems Security and Privacy Protection*. Springer. 2019, pp. 286–298.

[363] Geoffroy Couteau and Michael Reichle. “Non-interactive keyed-verification anonymous credentials”. In: *IACR International Workshop on Public Key Cryptography*. Springer. 2019, pp. 66–96.

[364] Melissa Chase, Trevor Perrin and Greg Zaverucha. “The signal private group system and anonymous credentials supporting efficient verifiable encryption”. In: *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*. 2020, pp. 1445–1459.

[365] Joe Andrieu. *A Primer of Functional Identity*. Tech. rep. Rebooting the Web of Trust, 2019.

[366] Andrew Tobin and Drummond Reed. “The inevitable rise of self-sovereign identity”. In: *The Sovrin Foundation* 29.2016 (2016).

[367] Paul Dunphy and Fabien AP Petitcolas. “A first look at identity management schemes on the blockchain”. In: *IEEE security & privacy* 16.4 (2018), pp. 20–29.

[368] Md Sadek Ferdous, Farida Chowdhury and Madini O Alassafi. “In search of self-sovereign identity leveraging blockchain technology”. In: *IEEE Access* 7 (2019), pp. 103059–103079.

[369] Dirk van Bokkem, Rico Hageman, Gijs Koning, Luat Nguyen and Naqib Zarin. “Self-sovereign identity solutions: The necessity of blockchain technology”. In: *arXiv preprint arXiv:1904.12816* (2019).

[370] Quinten Stokkink and Johan Pouwelse. “Deployment of a blockchain-based self-sovereign identity”. In: *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. IEEE. 2018, pp. 1336–1342.

[371] Job Spierings and Tom Demeyer. *Digitale Identiteit: een nieuwe balans?* nl. Tech. rep. Amsterdam: De Waag Technology & Society, 2019, p. 42.

[372] Hanna Schraffenberger. *IRMA made easy*. https://irma.cs.ru.nl/. 2020. (Visited on 01/03/2020).

[373] Joe Andrieu. “A Primer on Functional Identity”. In: (2019). URL: https://github.com/WebOfTrustInfo/rwot9-prague/blob/master/topics-and-advance-readings/functional-identity-primer.md.

[374] Joe Andrieu, Ryan Grant and Daniel Hardman. *DID Method Rubric v1*. 2021. URL: https://www.w3.org/TR/did-rubric/.

[375] Walid Fdhila, Nicholas Stifter, Kristian Kostal, Cihan Saglam and Markus Sabadello. “Methods for Decentralized Identities: Evaluation and Insights”. In: *International Conference on Business Process Management*. Springer. 2021, pp. 119–135.

[376] Nikita Khateev. *Issue Credential Protocol 1.0*. Github Requests for Comments. RFC. 2019. URL: https://github.com/hyperledger/aries-rfcs/blob/main/features/0037-present-proof/README.md.

[377] Manu Sporny and Dave Longley. *WebKMS v0.7*. 2021. URL: https://w3c-ccg.github.io/webkms/.

[378] Orie Steele. *Ed25518 Signature 2020*. 2020. URL: https://w3c-ccg.github.io/lds-ed25519-2020/#dfn-signature-suite.

[379] Orie Steele. *Ecdsa Secp256k1 Signature 2019*. 2021. URL: https://w3c-ccg.github.io/lds-ecdsa-secp256k1-2019/.

[380] Stephen Curran and Matt Raffel. *AnonCreds Specification*. 2022. URL: https://anoncreds-wg.github.io/anoncreds-spec/.

[381] Nikesh Lalchandani, Frank Jiang, Jongkil Jay Jeong, Yevhen Zolotavkin and Robin Doss. "Evaluating the Current State of Application Programming Interfaces for Verifiable Credentials". In: *2021 18th International Conference on Privacy, Security and Trust (PST)*. IEEE. 2021, pp. 1–7.

[382] Sovrin Foundation. *Sovrin Governance Framework V2*. URL: https://sovrin.org/wp-content/uploads/Sovrin-Governance-Framework-V2-Master-Document-V1.pdf.

[383] Foundation Sovrin. "Sovrin Transaction Endorser Agreement". In: (2019). URL: https://sovrin.org/wp-content/uploads/Transaction-Endorser-Agreement-V2.pdf.

[384] Foundation Sovrin. "Sovrin Stewards Agreement". In: (2019). URL: https://sovrin.org/wp-content/uploads/Sovrin-Steward-Agreement-V2.pdf.

[385] Pierre-Louis Aublin, Sonia Ben Mokhtar and Vivien Quéma. "Rbft: Redundant byzantine fault tolerance". In: *2013 IEEE 33rd International Conference on Distributed Computing Systems*. IEEE. 2013, pp. 297–306.

[386] Hyperledger. *Hyperledger Aries Cloud Agent - Python*. 2019. URL: https://github.com/hyperledger/aries-cloudagent-python.

[387] Daniel Hardman. *Peer DID Method Specification*. Tech. rep. 2019. URL: https://openssi.github.io/peer-did-method-spec/index.html.

[388] Daniel Hardman. *DIDComm Mythconceptions (IIW Presentation)*. 2021. URL: https://www.youtube.com/watch?v=rwvQdRyMeY4.

[389] Peter Tyson. "The Hippocratic oath today". In: *Nova* 27 (2001).

[390] Elizabeth Scoville and James Newman S. "A very brief history of credentialing". In: *ACP Hospitalist* (2009).

[391] Melissa Ruscoe. "Identity Verification and Authentication Standard for Digital Health and Care Services". en. In: (2018), p. 19.

[392] Federation of State Medical Boards. *Healthcare and Digital Credentials: Technical, Legal and Regulatory Considerations*. en. Tech. rep. Federation of State Medical Boards, June 2019.

[393] *Identity Checks*. Tech. rep. NHS Employers, Apr. 2019.

[394] *Regulatory Overload*. Tech. rep. American Hospital Association, Oct. 2017.

[395] Christopher Holmes. *Distributed ledger technologies for public good: leadership, collaboration and innovation*. 2018.

[396] C Peisah, E Latif, K Wilhelm and B Williams. "Secrets to psychological success: why older doctors might have lower psychological distress and burnout than younger doctors". In: *Aging and Mental Health* 13.2 (2009), pp. 300–307.

[397] Christine A Sinsky, Rachel Willard-Grace, Andrew M Schutzbank, Thomas A Sinsky, David Margolius and Thomas Bodenheimer. "In search of joy in practice: a report of 23 high-functioning primary care practices". In: *The Annals of Family Medicine* 11.3 (2013), pp. 272–278.

[398] Mark W Friedberg et al. "Factors affecting physician professional satisfaction and their implications for patient care, health systems, and health policy". In: *Rand health quarterly* 3.4 (2014).

[399] *The Benefits of ID Scanning*. Tech. rep. King's College Hospital. URL: https://www.gov.uk/government/consultations/reducing-bureaucracy-in-the-health-and-social-care-system-call-for-evidence/outcome/

busting-bureaucracy-empowering-frontline-staff-by-reducing-excess-bureaucracy-in-the-health-and-care-system-in-england.

[400] Clare Dyer. *GMC checks 3000 doctors' credentials after fraudulent psychiatrist practised for 23 years*. 2018.

[401] "Child Protection Fraud Case Ontario". In: (2019). URL: https://www.thestar.com/news/gta/2019/07/31/expert-who-gave-more-than-100-assessments-in-ontario-child-protection-cases-lied-about-credentials-for-years-judge-finds.html.

[402] Matt Goodman. *Dr Death*. 2016. URL: https://www.dmagazine.com/publications/d-magazine/2016/november/christopher-duntsch-dr-death/.

[403] Matt Discombe. *Royal college failed to carry out hundreds of background checks*. Feb. 2020. URL: https://www.hsj.co.uk/workforce/royal-college-failed-to-carry-out-hundreds-of-background-checks/7026895.article.

[404] Rajesh N Keswani, Tiffany H Taft, Gregory A Coté and Laurie Keefer. "Increased levels of stress and burnout are related to decreased physician experience and to interventional gastroenterology career choice: findings from a US survey of endoscopists". In: *The American journal of gastroenterology* 106.10 (2011), p. 1734.

[405] Tarja Heponiemi, Hannele Hyppönen, Tuulikki Vehko, Sari Kujala, Anna-Mari Aalto, Jukka Vänskä and Marko Elovainio. "Finnish physicians' stress related to information systems keeps increasing: a longitudinal three-wave survey study". In: *BMC medical informatics and decision making* 17.1 (2017), pp. 1–8.

[406] NHS Education Scotland. *NHSScotland Workforce 31 December 2020*. 2021. URL: https://turasdata.nes.nhs.scot/workforce-official-statistics/nhsscotland-workforce/publications/02-march-2021/.

[407] Enrico Coiera. "Building a national health IT system from the middle out". In: *Journal of the American Medical Informatics Association* 16.3 (2009), pp. 271–273.

[408] Health Education England. *Doctors in Training Programme*. URL: https://www.hee.nhs.uk/our-work/doctors-training.

[409] Fabiano Dalpiaz, Xavier Franch and Jennifer Horkoff. "istar 2.0 language guide". In: *arXiv preprint arXiv:1605.07767* (2016).

[410] Iain Barclay, Maria Freytsis, Sherri Bucher, Swapna Radha, Alun Preece and Ian Taylor. "Towards a Modelling Framework for Self-Sovereign Identity Systems". In: *arXiv preprint arXiv:2009.04327* (2020).

[411] Oscar Pastor, Sergio España, Marcela Ruiz, Dolors Costal and Xavier Franch. *i* Quick Guide*. 2014. URL: http://hci.dsic.upv.es/istar2ca_exp/instruments/I5-iStar_cheat_sheet/iStarcheatsheet_v1.1.pdf.

[412] Thomas Bodenheimer and Christine Sinsky. "From triple to quadruple aim: care of the patient requires care of the provider". In: *The Annals of Family Medicine* 12.6 (2014), pp. 573–576.

[413] Adam Kay. *This is going to hurt: secret diaries of a junior doctor*. Pan Macmillan, 2017.

[414] Health Education England. *Enabling staff movement - doctors in training*. Oct. 2019. URL: https://www.youtube.com/watch?v=GhplwTximTc.

[415] *System working - staff mobility / portability guidance for employers*. Tech. rep. Feb. 2019. URL: https://www.nhsemployers.org/-/media/Employers/Publications/System-working---Staff-mobility-guidance.pdf.

[416] *What's on the medical register?* URL: https://www.gmc-uk.org/registration-and-licensing/the-medical-register/a-guide-to-the-medical-register/whats-on-the-medical-register.

[417] Richard Marchant. *Changes to the medical register – what people have told us*. 2016. URL: https://gmcuk.wordpress.com/2016/12/21/changes-to-the-medical-register-what-people-have-told-us/.

[418] Zoltán András Lux, Dirk Thatmann, Sebastian Zickau and Felix Beierle. "Distributed-Ledger-based Authentication with Decentralized Identifiers and Verifiable Credentials". In: *2020 2nd Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS)*. IEEE. 2020, pp. 71–78.

[419] Randall Smith. *Docker Orchestration*. Packt Publishing Ltd, 2017.

[420] Bruce Momjian. *PostgreSQL: introduction and concepts*. Vol. 192. Addison-Wesley New York, 2001.

[421] Thomas Kluyver et al. *Jupyter Notebooks-a publishing format for reproducible computational workflows*. Vol. 2016. 2016.

[422] Vijay Surwase. "REST API modeling languages-a developer's perspective". In: *Int. J. Sci. Technol. Eng* 2.10 (2016), pp. 634–637.

[423] DIDx. *Aries Cloud Controller - Python*. Available online at `https://pypi.org/project/aries-cloudcontroller/`. Last accessed 01 June 2021. 2019.

[424] Ngrok. *Ngrok Service*. Available online at `https://ngrok.com/`. Last accessed 01 June 2021.

[425] Daniel J Bernstein, Niels Duif, Tanja Lange, Peter Schwabe and Bo-Yin Yang. "High-speed high-security signatures". In: *Journal of cryptographic engineering* 2.2 (2012), pp. 77–89.

[426] *COVID-19 Digital Staff Passporting*. URL: `https://beta.staffpassports.nhs.uk/`.

[427] Nikita Khateev. *Aries RFC 0037: Present Proof Protocol 1.0*. Github Requests for Comments. RFC. 2019. URL: `https://github.com/hyperledger/aries-rfcs/blob/master/features/0036-issue-credential/README.md`.

[428] Elaine Barker and Quynh Dang. "Nist special publication 800-57 part 1, revision 4". In: *NIST, Tech. Rep* 16 (2016).

[429] *Criterion - Rust*. Version 0.3.5. URL: `https://docs.rs/criterion/0.3.5/criterion/`.

[430] J Bradley Cousins and Lorna M Earl. “The case for participatory evaluation”. In: *Educational evaluation and policy analysis* 14.4 (1992), pp. 397–418.

[431] Will Abramson, Adam James Hall, Pavlos Papadopoulos, Nikolaos Pitropakis and William J. Buchanan. “A Distributed Trust Framework for Privacy-Preserving Machine Learning”. In: *Trust, Privacy and Security in Digital Business*. Ed. by Stefanos Gritzalis, Edgar R. Weippl, Gabriele Kotsis, A. Min Tjoa and Ismail Khalil. Cham: Springer International Publishing, 2020, pp. 205–220. ISBN: 978-3-030-58986-8.

[432] Pavlos Papadopoulos, Will Abramson, Adam J Hall, Nikolaos Pitropakis and William J Buchanan. “Privacy and trust redefined in federated machine learning”. In: *Machine Learning and Knowledge Extraction* 3.2 (2021), pp. 333–356.

[433] Stephen Curren, Paul Bastian and Daniel Hardman. *Indy DID Method*. Tech. rep. 2022. URL: https://hyperledger.github.io/indy-did-method/.

[434] P. F. F. Silva Filho. *Mitigating Sovereign Data Exchange Challenges: A Conceptual Framework to Apply Privacy- and Trust-Enhancing Technologies*. 2022.

[435] Emrys Schoemaker. *Digital Identity for Development — and protection*. 2021. URL: https://medium.com/caribou-digital/digital-identity-for-development-and-protection-d92716f24bb6.

[436] Erving Goffman et al. *The presentation of self in everyday life*. Harmondsworth London, 1978.

[437] Merlin Sheldrake. *Entangled life: how fungi make our worlds, change our minds & shape our futures*. Random House, 2020.

[438] Gregory Bateson. “Why do things have outlines?” In: *ETC: A review of general semantics* (1953), pp. 59–63.

[439] Gregory Bateson. “About games and being serious”. In: *ETC: A Review of General Semantics* (1953), pp. 213–217.

[440] Divya Siddarth, Sergey Ivliev, Santiago Siri and Paula Berman. “Who watches the watchmen? a review of subjective approaches for sybil-resistance in proof of personhood protocols”. In: *Frontiers in Blockchain* 3 (2020), p. 46.

[441] *Trust over IP Foundation Issues Its First Tools for Managing Risk in Digital Trust Ecosystems*. 2021. URL: https://trustoverip.org/blog/2021/09/23/trust-over-ip-foundation-issues-its-first-tools-for-managing-risk-in-digital-trust-ecosystems/.

[442] *Modular Elliptic Curve*. URL: https://en.wikipedia.org/wiki/Modular_elliptic_curve.

[443] *What is so special about elliptic curves?* URL: https://crypto.stackexchange.com/questions/11518/what-is-so-special-about-elliptic-curves.

[444] *E is for Elliptic Curves*. URL: https://www.maths.ox.ac.uk/about-us/life-oxford-mathematics/oxford-mathematics-alphabet/e-elliptic-curves.

[445] Jeremy Kun. *Elliptic Curves as Algebraic Structures*. 2014. URL: https://jeremykun.com/2014/02/16/elliptic-curves-as-algebraic-structures/.

[446] Dan Boneh and Victor Shoup. “A graduate course in applied cryptography”. In: (2015).

[447] Nadia El Mrabet and Marc Joye. *Guide to pairing-based cryptography*. CRC Press, 2017.

[448] Gergely Alpár, Fabian van den Broek, Brinda Hampiholi, Bart Jacobs, Wouter Lueks and Sietse Ringers. “IRMA: practical, decentralized and privacy-friendly identity management using smartphones”. In: *HotPETs 2017* (2017).

[449] D Daniel Ostkamp. “IRMA and Verifiable Credentials What is their relation?” In: ().

[450] Jelle C Nauta and Rieks Joosten. *Self-Sovereign Identity: A Comparison of IRMA and Sovrin*. Tech. rep. Technical Report TNO2019R11011, 2019.

[451] Ivar Derksen, Bart Jacobs, Hanna Schraffenberger and Timen Olthof. “Backup and Recovery of IRMA Credentials”. In: (2019).

[452] Donella H Meadows. *Thinking in systems: A primer*. chelsea green publishing, 2008.

[453] W Ross Ashby. *An introduction to cybernetics*. Chapman & Hall Ltd, 1961.

[454] Sander van der Leeuw. *Social sustainability, past and future: undoing unintended consequences for the Earth's survival*. Cambridge University Press, 2020.

[455] Ilya Prigogine. “From being to becoming”. In: (1982).

[456] Gregory Bateson. “Morale and national character.” In: (1942).

[457] Alex Haley. *Roots: The saga of an American family*. House of Majied, 2017.

[458] David Bohm, Chris Jenks et al. *Thought as a System*. Psychology Press, 1994.

[459] Edward Burnett Tylor. *Primitive culture: Researches into the development of mythology, philosophy, religion, art and custom*. Vol. 2. J. Murray, 1871.

[460] Yuval Noah Harari. *Sapiens: A brief history of humankind*. Random House, 2014.

[461] Gregory Bateson. “Effects of conscious purpose on human adaptation”. In: *Steps to an Ecology of Mind* (1972), pp. 440–447.

[462] Gregory Bateson. “Bali: The value system of a steady state”. In: (1963).

[463] Michio Kaku. *The future of humanity: terraforming Mars, interstellar travel, immortality, and our destiny beyond Earth*. Anchor, 2018.

[464] George Herbert Mead and Herbert Mind. “Self and society”. In: *Chicago: University of Chicago* (1934), pp. 173–175.

[465] Donald R Kelley. “Natalie Zemon Davis. The Return of Martin Guerre. Cambridge, Mass.: Harvard University Press, 1983.” In: *Renaissance Quarterly* 37.2 (1984), pp. 252–254.

[466] Herbert A Simon. “Bounded rationality in social science: Today and tomorrow”. In: *Mind & Society* 1.1 (2000), pp. 25–39.

[467] Elinor Ostrom. “Tragedy of the commons”. In: *The new palgrave dictionary of economics* 2 (2008).

[468] Gregory Bateson. “The cybernetics of “self”: A theory of alcoholism”. In: *Psychiatry* 34.1 (1971), pp. 1–18.

[469] David Wallace-Wells. *The uninhabitable earth: A story of the future*. Penguin UK, 2019.

[470] Bill Sharpe. “Three Horizons: the patterning of hope.” In: *Journal of Holistic Healthcare* 12.1 (2015).

[471] William Abramson, William J Buchanan, Sarwar Sayeed, Nikolaos Pitropakis and Owen Lo. “PAN-DOMAIN: Privacy-preserving Sharing and Auditing of Infection Identifier Matching”. In: *arXiv preprint arXiv:2112.02855* (2021).

[472] Iain Barclay and Will Abramson. “Identifying Roles, Requirements and Responsibilities in Trustworthy AI Systems”. In: *arXiv preprint arXiv:2106.08258* (2021).

[473] Adam James Hall et al. “Syft 0.5: A Platform for Universally Deployable Structured Transparency”. In: *arXiv preprint arXiv:2104.12385* (2021).

Appendix A

Number Theory

Public key cryptography applies number theory to generate finite groups with certain, well defined properties that are used to create cryptographic protocols and prove their security. This section introduces the underlying algebraic structures and notation for them used throughout the paper.

A.1 Sets

A set is a collection of well defined distinct elements or members of the set. Elements are well defined by axioms (rules), that can be used to determine if an object is an element within a specific set. A set can be any collection of objects, defined by any number of rules. A simple example is the set of whole integers less than 10.

Set notation is as follows:

- $x \in N$: The element x is in some set N
- $x \notin N$: The element x is not in some set N
- $\emptyset$: The empty set {}.
- $S \subseteq N$: The set S is a subset of set N, if and only if S contains every element of N. The empty set is a subset of all other sets $\emptyset \subseteq N$.
- $|N|$: Denotes the order of set N, the number of elements that are members of the set N.

A.2 Groups

A group is a set of elements, $\mathbb{G}$, and a mathematical operator $\circ$ which can be applied to to all elements in the set. E.g. $(x, y \in \mathbb{G} : x \circ y)$. For $(\mathbb{G}, \circ)$ to be classified as a group it must satisfy four axioms:

- **Identity Element**: There must exist an element $I \in \mathbb{G}$ such that, for every element $A \in \mathbb{G}$ the following equation holds: $I \circ A = A \circ I = A$. The identity element has no effect when combined under the operation $\circ$ with any other element in $\mathbb{G}$.
- **Inverse Element**: For all $A \in \mathbb{G}$ there exists an element $B \in \mathbb{G}$ such that $A \circ B = B \circ A = I$ where I is the **Identity element**. B is the inverse of A in the group $\mathbb{G}$.
- **Closure**: For all $A, B \in \mathbb{G}$, the result of an operation $A \circ B$ must also be in $\mathbb{G}$.
- **Associativity**: For all $A, B, C \in \mathbb{G}$, $(A \circ B) \circ C = A \circ (B \circ C)$. The order of the operations has no effect on the result.

These are extremely attractive properties when designing cryptographic protocols, which perform operations on information. The inverse element, ensures that the application of one element A, can always be reversed if you know the inverse B. The closure property means the universe of possible elements is only ever those defined by the finite set, under the group operation $\circ$.

A group that satisfies and additional property, commutativity, is called an Abelian Group.

- **Commutativity**: For all $A, B \in \mathbb{G}$, $A \circ B = B \circ A$.

An example of a finite group is the group that results from applying arithmetic modular n. That is, every integer i can be expressed as $i = q \times n + r$ for some quotient q and remainder r. Modular arithmetic is only interested in the remainder and not the quotient. Take the example finite group made up of the set of remainders modular 7 - $\mathbb{Z}_7 = \{0, 1, 2, 3, 4, 5, 6\}$, this has order $n = 7$, and an identity element $I = 0$.

A group is called cyclic if all the members in the group can be generated by a single entity, g, in that group through repeated operations on itself. A finite cyclic group is always an Abelian group (the commutative property holds). The very simplest example of this can be show in the additive group modulo 5, where $\mathbb{Z}_5$ has a generator element g. The elements $\{0,1,2,3,4\}$ are generated by repeatedly performing the group operation (addition) on the generator. Groups can have many different generators, each generating the elements of the group in a different order. See the Table A.1.

Operation	0g	1g	2g	3g	4g	5g
Element (g=1)	0	1	2	3	4	0
Element (g=3)	0	3	1	4	2	0

Table A.1: Generators for Additive Group Modulo 5

A more complex illustration of a cyclic group can be seen in Z_{13}^* the multiplicative group modulo 13. This group has a generator $g = 2$ with successive operations of $g * g$ equivalent to exponentiation, g^2.

Operation	g^0	g^1	g^2	g^3	g^4	g^5	g^6	g^7	g^8	g^9	g^{10}	g^{11}	g^{12}	g^{13}
Element in $\mathbb{Z}_{13}^*$	1	2	4	8	3	6	12	11	9	5	10	7	1	2

Table A.2: Multiplicative Group Modulo 13 with g=2

More generally, a finite cyclic group $\mathbb{G}$ of order n can be represented as $\{I, g, g^2, ..., g^{n-2}, g^{n-1}\}$ given any generator $g \in \mathbb{G}$.

All prime order groups are cyclic, and cyclic groups are used throughout cryptography because of a number of powerful properties. A generator can be used to index elements in the group, with the exponents of the generator creating an alternate finite number system, $(0, 1, ..., n)$ where n is the order of the group, in which mathematical operations can be performed. For example if group element $X = g^5$ and $Y = g^1 1$ then $Z = X \circ Y$ can either be calculated by applying the group operation $\circ$ to X and Y or it can be represented as $Z = g^{5+11} = g^1 6$. If 16 is larger than the order of the group then it can be reduced modulo the order. This only works if the generator, g, is the same for both elements. Changing the generator, changes the order in which the group elements are generated.

A subgroup $\mathbb{S}$ of a group $\mathbb{G}$ is a group that satisfies the group axioms under the group operation of $\mathbb{G}$ and some subset of elements of $\mathbb{G}$ including the identity element. The most trivial subgroup possible is the group containing the identity element. Under the Lagrange theorem, if $\mathbb{S}$ is a subgroup of $\mathbb{G}$ then the order of $\mathbb{S}$ must divide the order of $\mathbb{G}$. This is theorem is used to select groups that will have prime order, hence cyclic, subgroups.

A group is **isomorphic** with another group if all the elements in the group have a one to one mapping to each other. Two groups $(\mathbb{G}, \circ)$ and $(\mathbb{H}, *)$ are isomporphic if there exists a bijective group homomorphism. A function $f : \mathbb{G} \to \mathbb{H}$ such that for all elements u, and v in $\mathbb{G}$:

$f(u \circ v) = f(u) * f(v)$

This can be thought of intuitively as, for every element $g \in \mathbb{G}$ there is an equivalent element $h \in \mathbb{H}$. An important property to note is that if cyclic groups are isomorphic then there exists a mapping from a generator in one group, to a generator in the other. It has been proven that if a group of order n is cyclic then it is isomorphic over the additive group of integers modular n ($\mathbb{Z}_n$) with the one to one mapping $f(g^x) = x$.

A.3 Fields

Fields are a set of elements that satisfy certain rules when defined over two mathematical operations $(\mathbb{F}, +, *)$. A simple way to think about a field is a group of elements under which the laws of mathematics make sense. That is, you know some operations $(\circ, *)$ under which the rules for for addition and multiplication respectively hold over the elements in the group. The set of all real numbers $\mathbb{R}$ is an infinite field under the well defined operations $(+, \times)$.

The following axioms define a field:

1. The group $\mathbb{F}$ forms a finite abelian group, with an identity element denoted 0, under the operation +

2. The group $\mathbb{F}^* = \mathbb{F} - \{0\}$, ie the group $\mathbb{F}$ without the 0 element, forms a finite abelian group with identity denoted 1 under the operation $*$

3. The field follows the **Distributive Law**: For all $a, b, c \in \mathbb{F}, (a+b)*c = (a*b)+(a*c)$

Finite fields, fields with a finite set of elements, are used throughout public key cryptosystems to constrain groups to a set of positive integers such that the group operation on these integers only ever create positive integers in the defined field. This allows cryptographic algorithms to keep the required precision for them to work as expected. Understanding finite fields is key to understanding the cryptosystems built on top of these algebraic structures. Again cryptography is only interested on finite fields, a typical example is the set of $\mod p$ remainders where p is any given prime number. That is the operation + is modular p arithmetic and the operation $*$ is the multiplication modular p.

It is possible to prove that there exists a finite field $\mathbb{F}_p^m$ for every prime p and positive integer $m \geq 1$. Furthermore there are no finite fields that exist with $q = p^m$ elements where p is not prime. This then shows that integers $\mod p$ for a prime field $\mathbb{F}_p$ where $q = p^1$. Any two fields with p^m elements are isomorphic. The prime number p here is termed the **characteristic** of the field $\mathbb{F}$. Fields also have the concept of subfields such that a subfield of $\mathbb{F}$ forms a field under the same operands $(+, *)$ as $\mathbb{F}$.

A.4 Elliptic Curve Groups

An elliptic curve is an equation that defines an infinite set of (x, y) coordinates, elliptic curve points that satisfy the equation (see Equation A.1). This can be draw on a graph as shown in Figure A.1.

$$E_q : y^2 = x^3 + ax + b \tag{A.1}$$

These equations are used to construct groups which can then be applied to cryptography. To achieve this, first the set of elements defined by the elliptic curve equation

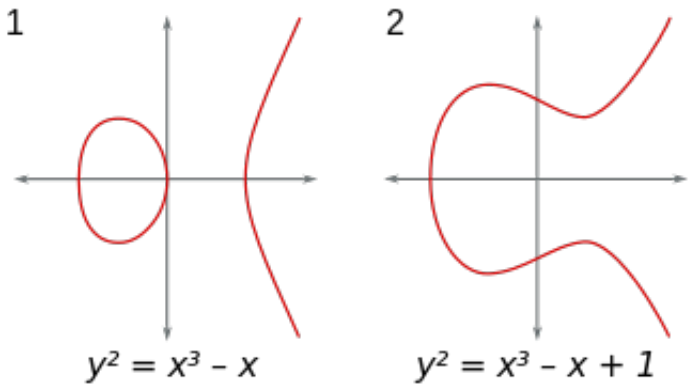

Figure A.1: Elliptic Curve Graphs - taken from [442]

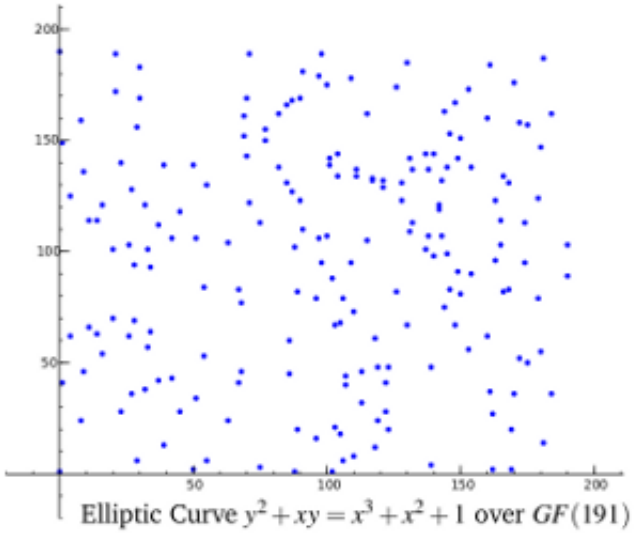

Figure A.2: Elliptic Curve over finite field 191 - taken from [443]

must be made finite. The is achieved by defining the elliptic curve group over the a finite field number system. Typically this is a prime field $\mathbb{F}_p$ or a binary field $\mathbb{F}_{2^m}$. An elliptic curve over finite field is then limited to only (x, y) coordinates that satisfy the Equation A.1 and both x and y are members of the finite field the curve is constrained by, $\mathbb{F}_p$. That is they must be whole numbers less than p and greater than 0. See Figure A.2, where the blue points on the graph represents these group elements.

For an elliptic curve to form a finite abelian group needed for cryptography algorithms, it must satisfy the group axioms including commutativity outlined in Section A.2. For this to be possible we have to define the group operation over these elements. The following questions need to be answered:

1. What will act as zero, the identity element?

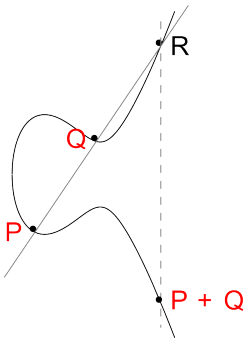

Figure A.3: Elliptic Curve point addition - taken from [444]

2. What is the group operation, i.e. how do you actually add two elliptic curve group elements together?

3. How do you calculate the additive inverse of an elliptic curve group element?

The identity element of an additive elliptic curve group is an *invented* point on the curve that has the following constraints:

1. This curve point is the intersection of all vertical lines through the curve.

2. Reflecting this point over the x-axis has no effect on it.

This point is often termed the point at infinity and the fact that all vertical lines go through the same point can be explained mathematically by representing an elliptic curve using a protective coordinate system.

To calculate the addition of two points P and Q on an elliptic curve, the following algorithm is defined (See Figure A.3 for a visual example):

- Form a line, $y = L(x)$, between the two points on the curve
- Find a third point, R, on the elliptic curve that intersects this line.
- Reflect this point across the the x-axis to get the result of $P + Q$

Where P and Q are unique points on the curve you will always be able to find a third point, unless the line between P and Q is vertical. In this case the result of adding P and Q is the identity element.

Calculating the coordinates of this third point on the curve means solving these two equations:

$$y = L(x) \tag{A.2}$$

$$y^2 = x^3 + ax + b \tag{A.3}$$

This is the same as solving the equation:

$$L(x)^2 = x^3 + ax + b \tag{A.4}$$

Using the knowledge of the two roots to this equation already known (the points defined by P and Q) it is possible to calculate the third solution to the curve. A more in depth explanation can be found here [445]. Cryptographic protocols that use elliptic curve groups carefully select the equation of the curve (see equation A.1) and the finite field ensure the order of the elliptic curve groups produced is prime, or has prime subgroups. This means they are cyclic such that repeated elliptic curve additions on a generator cycles through every element in the group.

A.5 Bilinear Maps

A bilinear map is a function that combines elements from two groups and maps them to an element in a third. Cryptography focuses on bilinear maps between finite cyclic groups, $e : G_1 \times G_2 \rightarrow G_t$ such that for the mapping e, the following properties hold:

- **Bilinearity:** For all $X, Y \in G_1, \hat{Z} \in G_2, e(X+Y, \hat{Z}) = e(X, \hat{Z})e(Y, \hat{Z})$.
- **Non-degeneracy:** $e(R, \hat{R}) \neq 1$. Every element is not mapped to the identity element.
- **Computability:** the function e must be efficiently computable.

This is referred to as a pairing function, that pairs elements from G_1 and G_2 with an element in G_t. Each of these groups are isomorphic, they all have the same order and are cyclic. This means for generators g_1 and g_2 of groups G_1 and G_2 respectively, the map $e(g_1, g_2)$ is a generator of the group G_t.

In cryptography, the groups G_1 and G_2 are typically elliptic curve groups and G_t is a multiplicative group. The ability to move, in one direction from elements in G_1 and G_2 to and element in G_t has proved useful when constructing cryptographic protocols. Note that from an element in G_t, the map provides no way to deconstruct into an element in either of the groups G_1, or G_2.

Bilinear pairings for cryptography are commonly categorised into three distinct types: [283].

- **Type 1:** $G_1 = G_2$.
- **Type 2:** $G_1 \neq G_2$ and there is no efficiently computable homomorphism $\phi : G_1 \rightarrow G_2$.
- **Type 3:** $G_1 \neq G_2$ and there is an efficiently computable homomorphism.

A.6 Integer Factorization and the RSA Group

Integer factorization, the process of spitting a composite number n into its prime factors, is a known hard problem used to construct cryptographic protocols. Specifically, n is generated such that it is the product of two prime numbers p and q. The semi-prime number n is then used to parameterize a multiplicative group Z_n^* with order defined by Euler's totient $\phi(n) = (p-1)(q-1)$. This means that $\mathbb{Z}_n^*$ forms a group with $\phi(n)$ elements that satisfies the group axioms under multiplication, thus for any element $u \in \mathbb{Z}_n^*$ when raised to an exponent e, the exponent can always be reduced $\mod \phi(n)$.

Euler's theorem states that if a and n are relatively prime, then $a^{\phi(n)} \equiv 1 \mod n$. This is used to calculate the inverse of a function $(a^e)^d = a$, such that e and d are multiplicative inverses modulo $\phi(n)$. Those with knowledge of the prime factors of n

can calculate the order of the group $\phi(n)$ and use the extended euclidean algorithm to compute the inverse d for some exponent e.

A very simple example would be $p = 5, q = 11$ giving $n = 55$ and $\phi(n) = (p-1)(q-1) = 40$. This tells us that there are 40 elements that satisfy the group axioms for the multiplicative group $\mathbb{Z}_{55}^*$. For some exponent $e = 7$, the multiplicative inverse $\mod \phi(n)$ can be calculated as follows:

First compute the Euclidean algorithm:

$$\begin{aligned} 40 &= 5 \times 7 + 5 \\ 7 &= 1 \times 5 + 2 \\ 5 &= 2 \times 2 + 1 \end{aligned} \tag{A.5}$$

Then extend it to get something in the form $1 = (a \times 40) + (b \times 7)$, such that a is removed $\mod 40$ leaving b and the multiplicative inverse of $7 \mod 40$.

$$\begin{aligned} 1 &= 5 - 2 \times 2 \\ &= 5 - 2 \times (7 - 1 \times 5) \\ &= 3 \times 5 - 2 \times 7 \\ &= 3 \times (40 - 5 \times 7) - 2 \times 7 \\ &= 3 \times 40 - 17 \times 7 \end{aligned} \tag{A.6}$$

Here we see that -17 is the multiplicative inverse of $7 \mod \phi(n)$ and $-17 \equiv 23 \mod 40$. So $7, 23$ are multiplicative inverses modulo 40. Indeed this can easily be verified as $23 \times 7 = 161 \equiv 1 \mod 40$.

Appendix B

Cryptographic Primitives and Protocols

B.1 Hash Functions

A hash function is type of function that maps arbitrary size inputs to fixed size outputs, often 128 or 256-bit strings. Hash functions, like one-way functions, can be indexed by keys, separating the generic algorithm from a specific keyed instance: $H(k,x) \rightarrow t$. These functions are widely used throughout cryptography.

Hash functions must satisfy a number of properties [266]:

- **Deterministic**: Every input must always map to the same output.
- **Collision Resistant**: Computing a collision whereby $H(k,m_1) = H(k,m_2)$ should occur with negligible probability.
- **Preimage Resistant**: Given an output from a hash function $H(x) \rightarrow z$ it is hard to find any input that hashes to z.
- **Second Preimage Resistant**: Given an input, y, and its hash, $H(y)$, it is hard to find a different input q such that $H(q) == H(y)$
- **Efficient**: There exists a polynomial algorithm for computing H.

B.2 Public Key Cryptosystems

A public key cryptosystem is a system that provides a generic method for encryption and decryption of a message such that $D(E(M))) = M$. The general method can be instantiated by selecting a keypair (sk, pk) from the keyspace which is finite but large enough that there is a negligible probability of collisions. The specific method for an entity with a keypair becomes $D(sk, E(pk, M)) = M$. M represents a finite set, the message space under which the cryptosystem has been defined [245, 26].

More formally a public key cryptosystem consists of a triple of algorithms (G, F, I) [446]:

- G is a probabilistic polynomial time algorithm that selects a random public and secret keypair from the keyspace when invoked: $G() \rightarrow (pk, sk)$.
- F is a deterministic polynomial algorithm that takes a public key and an element in the message space as inputs and returns an element in the output space: $F(pk, x) \rightarrow y$.
- I is a deterministic polynomial algorithm that takes a secret key and an element from the output space and returns an element from the input space: $I(sk, y) \rightarrow x$.

The cryptosystem must satisfy **correctness**. That is for every keypair, (pk, sk) within the keyspace, $I(sk, F(pk, M))$ holds for all elements in M.

B.3 Pedersen Commitments

A commitment is a way to commit to value in a protocol without revealing what that value is. The Pedersen Commitment scheme is as follows [155]: A commitment to $x \in \mathbb{Z}_q$ can be created in some group G of prime order q with generators g and h. First choose at random $r \leftarrow \mathbb{Z}_q$, then set the commitment $C = g^x h^r$. A Pedersen commitment is perfectly hiding as there a multiple x, r pairs that satisfy the equation $C = g^x h^r$, and they are binding under the discrete log assumption.

B.4 Zero Knowledge Proofs of knowledge

Zero knowledge proofs (ZKPs) have usecases throughout the cryptography space. A useful ZKP used throughout cryptographic protocols is the ability to prove knowledge of the discrete log problem. Which as discussed in Appendix B.2 forms the backbone of public key cryptography. A ZKP is a way for a prover to prove knowledge of a witness to a certain relationship or statement without disclosing any information about the witness.

The key terminology used when discussing zero knowledge proofs of knowledge are:

- **Witness**: The value that knowledge is being proven of.
- **Instance**: The other elements of the relation in the proof.
- **Prover**: The entity that is proving knowledge of the witness.
- **Verifier**: The entity the prover wants to convince of knowledge of a witness, this entity verifies the proof from the Prover.

A protocol for producing ZKPs must satisfy the following three properties assuming interacting parties follow the protocol [156]:

- **Completeness**: If the statement is true the verifier will accept the proof provided by the prover
- **Soundness**: If the statement is false, the prover is unable to produce a proof that will convince the verifier
- **Zero-knowledge**: The verifier will learn no additional knowledge from a proof other than the validity of the statement it proves

The notation PK(a,b) : $P = g^a$ & $Q = h^b$ denotes a zero knowledge Proof of Knowledge of integers a, b such that the instance $P = g^a$ and $Q = h^b$ holds where P, Q, g and h are elements of groups $\mathbb{G} = \langle g \rangle = \langle h \rangle$. The formal definition states that given a proof in this notation, it is easy to derive a protocol for implementing such a proof [272].

B.4.1 The Sigma Protocol

A common protocol for generating and verifying a zero knowledge proof is called a Sigma protocol, an interactive three step protocol between the Prover and the Verifier. Interactive meaning that the Prover and the Verifier must be online and able to interact for the duration of the protocol. Although, as Schnorr showed in this paper [158] it is possible through the Fiat-Shamir transformation [253] to turn this into a non-interactive protocol. These non-interactive protocols have the additional benefit of being verifiable by anyone at any point in the future.

The protocol itself consists of three phases:

1. **Commitment**: The Prover generates a random number, $r \leftarrow\$ Z_n$, and makes a commitment $t = g^r$. This commitment is sent to the Verifier. The Prover must use a new random number for every proof they generate or the witness becomes extractable.

2. **Challenge**: After receiving the commitment, the Verifier generates a random challenge number, c, and sends it to the prover. They must wait until they have received the commitment t, or the Prover will be able to generate proofs regardless of whether they know the witness. The Fiat-Shamir transformation removes the need for this step by simulating the random choice of a Verifier with a hash of the problem instance that both the Prover and Verifier can produce independently.

3. **Response**: The Prover uses the challenge, c, to create a response along with the random number, r, from the commitment phase and the witness to the proof. This response is sent to the Verifier who verifies the proof through a computation and is convinced, or not, that the Prover has knowledge of the witness.

B.4.2 An Example

A Prover wants to prove they have knowledge of the witness x to the statement $g^x = y$, where g is a generator to an elliptic curve, x is an integer exponent and y is a point on the curve. This is a common statement in cryptography, and can be thought of as

proving knowledge of the solution to the discrete log problem. Most obviously used to prove knowledge of your private key.

1. Prover generates a random r and $t = g^r$. Sends t to the Verifier.
2. Verifier stores t and then generates a random challenge, c, which is sent to the Prover.
3. When the Prover receives the challenge, they can generate a response $s = r + x^c$. The response, s, is sent to the Verifier.

The Verifier now needs to make a decision about whether, based on the proof s, they believe the Prover has knowledge of of the witness x. To do this they check $g^s = y^c \cdot t$:

$$\begin{aligned} g^s &= y^c \cdot t \\ &= g^{x^c} \cdot g^r \\ &= g^{x^c + r} \\ &= g^s \end{aligned} \tag{B.1}$$

Appendix C

BBS+: A Signature Scheme with Efficient Protocols

A signature scheme with efficient protocols is an abstract class of signature schemes with important properties that can be used to instantiate a privacy-preserving credential mechanism [28, 28]. These properties are:

- The ability to sign an array of messages
- The ability to blindly sign messages
- The ability to create non-interactive zero-knowledge proofs of knowledge of a signature on a message

A number of concrete instantiations of this abstract scheme have been realised in the literature. This chapter details the mathematical algorithms of the BBS+ signature scheme. The maths presented in this chapter has been adapted from Au et al [299], Camenisch et al [30] and the cryptography book *A Guide to Pairing-Based Cryptography* [447].

C.1 Key Generation

Let $(p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_t, g_1, g_2, e)$ be a Type-3 bilinear environment as discussed in Section A.5. p is the prime order of groups $\mathbb{G}_1$, $\mathbb{G}_2$ and $\mathbb{G}_t$. $\mathbb{G}_1$, and $\mathbb{G}_2$ are both elliptic curves and

$\mathbb{G}_1 \neq \mathbb{G}_2$. The group $\mathbb{G}_t$ is a multiplicative group generated by the bilinear map $e(g_1, g_2)$ and e is efficiently computable.

For a key pair that can create signatures over L messages:

- Take $(h_0, ..., h_L) \leftarrow\$ \mathbb{G}_1^{L+1}$ (Select $L+1$ random ECC group elements from $\mathbb{G}_1$)
- $x \leftarrow\$ \mathbb{Z}_p^*$ (Select a random integer x the multiplicative prime order group $\mathbb{Z}_p^*$)
- $w \leftarrow g_2^x$ (Apply the generator, g_2, of elliptic curve group G_2 to itself x times to produce ECC group element w)
- Set $sk = x$, the random integer x is the secret key and should be held and managed securely.
- Set $pk = (w, h_0, ..., h_L)$. The public key is one ECC point from group G_2 - w - and $L+1$ random ECC group elements from $\mathbb{G}_1$. That is a random ECC point for every message, plus an additional one - h_0. The public key must be published and accessible to all who need to verify signatures created by sk.

C.2 Signing

A signature is produced a set of messages $(m_1, ...m_L) \in \mathbb{Z}_p^L$ (note each message has been mapped into the domain of the protocol, $\mathbb{Z}_p$ using a hash function (See Section B.1)) and a secret key x as follows.

- $e, s \leftarrow\$ \mathbb{Z}_p$. (Select random integers e and s from the prime order group $\mathbb{Z}_p$)
- Compute $A \leftarrow (g_1 h_0^s \prod_{i=1}^{L} h_i^{m_i})^{\frac{1}{e+x}}$. This produces a single group element in group $\mathbb{G}_1$ that is the result of adding different ECC points together. I.e $h_0^s, h_1^{m_1}, ..., h_L^{m_l}$. The integer $\frac{1}{e+x} \in \mathbb{Z}_p$ can be thought of as applying to each of these elements.

C.3 Verification

On input of a public key $(w, h_0, ..., h_L) \in \mathbb{G}_2 \times \mathbb{G}_1^{L+1}$, message $(m_1, ...m_L) \in \mathbb{Z}_p^L$ and a possible signature $(A, e, s) \in \mathbb{G}_1 \times \mathbb{Z}_p^2$ check $e(A, w g_2^e) = e(g_1 h_0^s \prod_{i=1}^{L} h_i^{m_i}, g_2)$.

This works because of the following:

$$
\begin{aligned}
e(A, wg_2^e) &= e((g_1 h_0^s \prod_{i=1}^{L} h_i^{m_i})^{\frac{1}{e+x}}, wg_2^e) && \text{Substitute } A \\
&= e((g_1 h_0^s \prod_{i=1}^{L} h_i^{m_i})^{\frac{1}{e+x}}, g_2^x g_2^e) && \text{Substitute w} \\
&= e((g_1 h_0^s \prod_{i=1}^{L} h_i^{m_i})^{\frac{1}{e+x}}, g_2^{e+x}) && \text{Combine } g_2^x g_2^e \\
&= e((g_1 h_0^s \prod_{i=1}^{L} h_i^{m_i})^{\frac{e+x}{e+x}}, g_2) && \text{Move scalar (x+e) from } \mathbb{G}_2 \text{ to } \mathbb{G}_1 \\
&= e(g_1 h_0^s \prod_{i=1}^{L} h_i^{m_i}, g_2) && \text{Cancel out } \frac{e+x}{e+x}
\end{aligned}
\tag{C.1}
$$

The signature scheme works because only an entity with knowledge of the secret key x can produce valid (A, e, s) signatures on messages, but anyone with the public key $w = g_2^x$ can verify the validity of these signatures.

However, when used as a credential mechanism some additional protocols are required. The signature (A, e, s) acts as a credential against a set of L attributes and a holder needs to be able to prove they know this signature without revealing what it is every time they wish to disclose the attributes in the credential to a verifying party. To achieve this, the holder contributes a commitment to a master secret which is blindly signed by the issuer. Then subsequently proves knowledge of a signature on a master secret that they know when they which to disclose attributes from the credential to a verifier.

C.4 Signing Committed Messages

A general protocol for a signer to blindly sign a set of J committed messages $M_c = (m_1, ...m_J) \in \mathbb{Z}_p^J$ is as follows (Adapted from Au et al [299]):

First the entity committing to a set of message computes:

- Generate a binding factor: $s' \leftarrow\$ \mathbb{Z}_p$
- Compute Pedersen commitment on M_c using the blinding factor s': $C_m = g_1^{s'} g_2^{m^1} g_3^{m^2} ... g_{J+1}^{m^J}$

- Prove in zero knowledge that C_m is correctly formed: $\sigma_{C_m} \leftarrow PK\{(s', m_1, ..., m_J) : C_m = g_1^{s'} g_2^{m^1} g_3^{m^2} ... g_{J+1}^{m^J}\}$.
- Send C_m and σ_{C_m} to the signer.

Then the signer to create a signature on the committed messages M_c along with L known messages $M = (m_1, ...m_L) \in \mathbb{Z}_p^L$ completes the following steps:

- Verifies the ZKP σ_{C_m}
- Selects $s'', e \leftarrow_\$ \mathbb{Z}_p$
- Compute $A \leftarrow (g_1 h_0^{s''} C_m \prod_{i=1}^{L} h_i^{m_i})^{\frac{1}{e+x}}$
- Send (A, s'', e) to the entity that contributed the committed messages along with any plaintext messages that were signed M

On receipt of (A, s'', e), $s = s' + s''$ can be calculated such that a signature unblinded messages M_c along with the plaintext messages M is (A, e, s). This can be verified as shown in Section C.3. s' completely hides the information about the messages M_c that are signed from the signer.

C.5 Proof of Knowledge of A Signature

This protocol allows for an entity in possession of a BBS+ signature (A, e, s) created by public key $(w, h_0, ..., h_L)$ on an array of L messages $M = (m_1, ..., m_L)$ can prove in zero knowledge that they know this signature and that any subset of messages from M was signed within this signature by the identified public key. This can use any protocol for proving relations amongst components of a discrete-logarithm representations of group elements such as the Sigma protocol discussed in Section B.4.1. This has been adapted from Au et al [299].

- $r_1, r_2 \leftarrow_\$ \mathbb{Z}_p^*$. Select two random integers, r_1, r_2 from the multiplicative prime order group $\mathbb{Z}_p^*$

- Compute $A_1 = g_1^{r_1} g_2^{r_2}$ and $A_2 = A g_2^{r_1}$
- Compute $\delta_1 = r_1 e$ and $\delta_2 = r_2 e$
- Using $A_1, A_2, \delta_1, \delta_2$ compute the signature proof of knowledge Π as shown in Equation C.2. This is a proof of knowledge that is made non-interactive through the Fiat-Shamir transformation [253].
- Send $\Pi, A_1, A_2, (w, h_0, ..., h_L)$ to the verifier, where $(w, h_0, ..., h_L)$ is the public key associated with the secret key x used to produce the signature (A, e, s)

Note that this protocol does not include disclosing or hiding any messages signed as part of the signature.

$$\begin{aligned} &\Pi : SPK\{(r_1, r_2, \delta_1, \delta_2, e, s, m_1, ..., m_L) : \\ &A_1 = g_1^{r^1} g_2^{r_2} \wedge A_1^e = g_1^{\delta_1} g_2^{\delta_2} \\ &\wedge \frac{e(A_2, w)}{e(g_0, h_0)} = e(A_2, h_0)^{-e} e(g_2, w)^{r_1} e(g_2, h_0)^{\delta_1} e(g_2, h_0)^{m_0} ... e(g_{L+1}, h_0)^{m_L}\} \end{aligned} \tag{C.2}$$

Appendix D

Evaluation of Software Frameworks for Credential-based Identification Systems

D.1 HVIEP

The HVIEP emerged from the work of a for-profit company, Evernym, who developed a framework for credential-based identification system drawing on the cryptographic literature surrounding credential mechanisms. Recognising the need for an ecosystem of implementers and implementations, Evernym donated code to the Hyperledger Foundation in what became Hyperledger Indy. At the time Indy comprised of both the code to run a node in an Indy-based VDR and the code to instantiate a software artefact, referred to as an agent, able to interface with the VDR and other agents to engage in credential-based interactions. Additionally, the Sovrin Foundation was formed to run a reliable, governed instance of an Indy-based VDR the Sovrin MainNet as a public utility. Subsequently, Indy was broken up into two further projects under the Hyperledger foundation. Hyperledger Ursa containing the underlying cryptographic engine and Hyperledger Aries which defined RFCs outlining how software agents participating in this ecosystem should interface and interact with each other.

Hyperledger Aries spawned a number of distinct open source implementations of the RFCs in various different languages including, .NET, python, JavaScript and Go. On top of these frameworks, proprietary solutions provided software as a service to simplify

an implementers ability to design credential-based identification systems. For example, the Evernym and Trinsic platforms. The latest iteration of this ecosystem has been the instantiating of numerous Indy-based VDR including Bedrock Consortium, the Finnish based Findy network and Indicio. Additionally, cheqd are attempting to apply a similar model for a VDR using a cosmos application chain.

The analysis presented here is largely applicable to all implementations under the HVIEP, however practical experience with the ACA-Py framework has shaped this evaluation.

D.1.1 Data Model

The HVIEP aspires to use the W3C VC Data Model for VCs and VPs, however the current data model designed for credentials issued and presented using a more complex ZKP signature scheme is not currently conformant with the W3C standard. This is in part because this was defined before the VC Data Model standard was finalized. Promising steps have been made to move towards conformance, with the addition of non-Hyperledger VCs to the HVIEP and the willingness to adopt BBS+ signatures represented in a conformant JSON Linked Data structure.

D.1.2 Identifiers

The HVIEP uses DIDs specifically designed for their custom VDR - an instantiation of Indy-based distributed ledger. Initially DIDs rooted to this VDR were did:sov, referring to the Sovrin Foundation's Indy instance. This how now been generalised to did:indy which can refer to identifiers on any Indy ledger. In addition to DIDs rooted to a public VDR, the HVIEP makes use of pairwise peer DIDs (did:peer) to enable software artefacts and their controllers to form and maintain relationships with others.

D.1.3 Protocol support

D.1.3.1 Transport & Messaging

HVIEP uses DIDComm messaging, a standard initially defined in a Aries RFC and subsequently adapted to a version 2.0 standard in DIF. DIDComm is transport agnostic and can work over HTTP, WebSockets and WebRTC to name a few. The emphasis within the HVIEP is on asynchronous communication between peers able to engage in secure interactions through the ordered exchange of structured messages defined by protocols.

D.1.3.2 Credential Issuance

HVIEP defines a DIDComm credential issuance protocol in a RFC. In version 1.0 this only supported issuing credentials using CL-RSA signatures using public issuance keys stored on Indy-based VDRs and is widely supported by different software frameworks within the Hyperledger ecosystem. A 2.0 version has since been developed, specifying the issuance of credentials using both BBS+ and ECDSA signatures with key material represented within did:key identifiers.

D.1.3.3 Credential Presentation

Credential presentation is supported by the HVIEP through a DIDComm protocol between a Holder and a Verifier. Again it has two versions and both specifications are defined in Aries RFCs. Version 1 supports the presentation and verification of a credential signed using a CL-RSA signature. The presentation is a zero knowledge proof of knowledge of a valid signature on a set of disclosed messages. Version 2.0 adds presentation support for other W3C credential formats signed using different signature schemes. Most software frameworks have support for 1.0 and are in the process of adding support for the 2.0 RFC.

D.1.3.4 Credential Revocation

Credential revocation is supported by most implementations in the HVIEP ecosystem. Revocation makes use of a cryptographic accumulator to enable credential issuers to published revocation registries and subsequently update these registries in order to revoke a specific credential. Through the use of accumulators, revocation retains strong unlinkability of credentials [302]. The registries and associated updates are stored on a public Indy-based VDR which Holders and Verifiers can access to produce and verify proofs of non-revocation.

D.1.4 Key Management

Secret keys including link secrets used to bind credentials to Holders, private keys able to control and authenticate against DIDs and private issuance keys was originally handled by the indy-sdk. A dependency that the majority of implementations of the HVIEP relied on. The indy-sdk supports a pluggable database configuration defaulting to sqlite and the contents of this database is encrypted under a key defined when instantiating an aries agent. Secure storage has since been factored out of the indy-sdk library into, aries-askar with improved performance and cryptographic support. In addition to secret keys, public credential issuance keys must be made available to the necessary actors in the system. This is achieved by publishing them to an Indy-based VDR in a transaction, which must be signed by a DID already stored within the VDR. Finally, public keys must be exchanged between interacting peers before they can establish secure connections. This is assumed to happen out of band, for example through a QR code on a website or a face to face interaction.

D.1.5 Cryptographic signature suites

The HVIEP makes use of a more complex signature scheme with efficient protocols designed specifically to preserve unlinkability within credential-based identification systems. This scheme, CL-RSA has not yet been standardised into a signature suite with most implementations relying on the same cryptographic engine - Hyperledger Ursa.

However, work is underway to standardise this approach under the AnonCreds specification. In addition to CL-RSA signatures, the HVIEP uses Ed25519 ECDSA signatures for public key pairs which control DIDs and some implementations support the use of these key pairs in credential issuance. The JSON-LD BBS+ signature suite is another that is gradually being adopted for credential issuance within this ecosystem.

D.1.6 Data Storage

Credentials are stored within a database, that can be optionally encrypted using a seed phrase provided when initialising the aries agent. This database can either be managed on the edge device of a specific actor, or maintained by a cloud agent that is capable of partitioning the database into different tenants for each of the actors data that they manage.

D.1.7 Recovery and Backup

Recovery and backup remains complex in the HVIEP. It is possible to export and import the encrypted data of an agent, including key material and credential data although this is not well supported. Promising progress in this area occurred when two mobile agents, Lissi and Trinsic, demonstrated how the associated data from one application could be exported and imported into the other application. Although it should be noted that both mobile applications depend on the same Aries .NET framework.

D.1.8 Schema Management

Credential schema are defined as custom JSON by Issuers and published to an Indy-based VDR. One published any Issuer with a DID on that VDR is capable of issuing credentials against that schema. To do so they must publish a public CL-RSA key stored within a credential definition that binds the issuance key to a specific schema.

D.1.9 Complexity / Ease of use

The HVIEP is complex in that it is not a single library, codebase, organisation or project. Rather it is a coordination effort to ensure interoperability amongst implementations using CL-RSA signatures to issue credentials as first demonstrated by Evernym. This has led to information scattered throughout numerous distinct Github repositories and getting to a basic level of understanding how the pieces fit together can be challenging. That said, a number of service providers exist that aim to reduce this complexity by providing a hosted service and easy to use API for a cost. Furthermore, once the mental model of the HVIEP has been internalised, using open source code bases such as ACA-Py which provide a complete full-stack framework for instantiating and configuring aries agents for specific credential-based identification systems is relatively easy. Developing an independent framework for the HVIEP remains a daunting task.

D.1.10 Adoption

The HVIEP has been adopted internationally and is being applied in numerous proof of concepts and pilot studies. Notably mentions include the Government of British Columbia which developed the open source ACA-Py framework, which they have applied to a number of government use cases. The NHS has been exploring this platform using Evernym as a service provider. There are a number of mobile applications available from the app store, which can be used by individuals to receive, hold and present these credentials. A reason for the HVIEP's success can be put down to it's maturity. It is possible for an implementer to adopt and use a complete framework that supports credential issuance, presentation and revocation with compatible mobile applications without requiring much additional tooling. Implementers can focus realising a specific set of credential-based interactions relevant to their use case. It should also be noted that the HVIEP is a broad classification containing a interconnected series of service providers, libraries, frameworks and tooling. Each is in various stages of maturity and interoperability when compared to another. Furthermore, interoperability beyond the HVIEP remains challenging and is a key barrier to adoption for some.

D.1.11 Transparency / Governance

The HVIEP itself is a part of the Hyperledger Foundation, a non-profit under the Linux Foundation. This maintains standards for code contributions to any of the open-source projects that make up the HVIEP. Projects often organise weekly contributor calls open to anyone and must produce quarterly reports. The Hyperledger Foundation assigns a status to each of the projects it manages identifying their readiness for production use. There are also proprietary closed source codebases typically produced by service providers that leverage different components of the HVIEP to offer easy to use services to implementers. These have there own internal rules and relations.

A Indy-based VDR consisting of a set of nodes running the indy-node code base to maintain consensus over the state of the VDR is also subject to governance mechanisms. The Sovrin Foundation was the first example of this, maintaining the Sovrin MainNet and authorising the stewards which participated by running nodes. Today, there exists a number of other Indy-based VDRs each with their own set of governance rules.

D.2 IRMA

IRMA is a project and open-source codebase developed in the Netherlands under the Privacy by Design Foundation from 2016. It aims to make privacy-preserving ABCs usable within identification processes, such that individuals can choose the attributes they reveal across the different digital relationships they participate in. It is distinct from the other implementations evaluation in this section in that it was developed independently from the influence of the W3C standards and bases its design on the academic literature around privacy-preserving ABCs [311]. Due to this, it does not use a standard data model for credentials, decentralised identifiers or distributed ledgers in its design.

The IRMA technology stack comprises of an JavaScript library, *irma.js*, an Android mobile application, the IRMA App, and the IRMA API server. Credential issuers and verifiers are expected to host their own IRMA API servers, although issuance and verific-

ation capabilities could also be provided by a third-party service. The *irma.js* and the IRMA App are required by those acting in the role of Holder and either receiving a signed credential or presenting a proof. Through a QR code presented by the *irma.js* library, an individual can begin a session with a IRMA API server to engage in a credential presentation of issuance protocol.

Issued credentials are stored in the app, while successful presentations are recorded in a signed JWT that is stored in the client website allowing the individual to continue as an authenticated user [448].

D.2.1 Data Model

The IRMA architecture has a custom data model for credentials which is understood by the existing open-source IRMA software. An issued is specified using JSON which references a specific scheme using a Uniform Resource Locator (URL) and a set of attributes. The scheme for the credential is defined in XML and can be used to interpret the credential attributes. It is doubtful that an interoperable implementation of this architecture could be achieved. Although a masters thesis did explore how the existing IRMA data model could be made conformant with the W3C VC Data Model specification [449].

D.2.2 Identifiers

IRMA does not make use of DIDs, instead credentials reference issuer information through URLs. Additionally, credentials do not contain public identifiers for Holders. This is because of the cryptographic signature scheme used, which enables Holders to blindly contribute a master secret to all credentials without revealing it. The master secret is a large random number known only to the Holder, which they can use to prove a credential was issued to them [28].

D.2.3 Protocol Support

D.2.3.1 Transport and Messaging

The IRMA architecture uses HTTP to communicate between Issuer, Holder and Verifier applications. The exact protocol specifying these interactions is custom to the internal IRMA architecture. Issuers and Verifiers run the IRMA API server and include the *irma.js* package in the client of websites they host. Through a QR code presented by the *irma.js* library, an Holders can begin a session with a IRMA API server to engage in specific protocols.

D.2.3.2 Credential Issuance

Credential issuance is a custom protocol contained within the open-source implementation of the IRMA architecture. Interoperability with other architectures is unlikely to be feasible without significant effort. The Issuer displays a QR code on the client interface a Holder is visiting, the Holder scans this using their IRMA mobile application, this initiates a connection between the IRMA API server across which the Issuer sends a credential offer which the Holder can choose to accept or reject. If they accept, the IRMA mobile app replies with a commitment to their master secret which is the included in the credential which is signed and returned to by the Issuer. This credential is then stored in the Holder's application. There are similarities with the HVIEP, because both use the same underling cryptography - CL-RSA signatures. However, HVIEP explicitly defines an RFC for credential issuance and has multiple distinct implementations whereas the IRMA protocol is only defined in the code that implements it.

D.2.3.3 Credential Presentation

Credential presentation is a custom protocol contained within the open-source implementation of the IRMA architecture. Interoperability with other architectures is unlikely to be feasible without significant effort. As with credential issuance, the Verifier displays a QR code to the Holder who scans this with their IRMA app to initate a connection with

the Verifier's IRMA API server. Across this connection the Verifier requests a specific presentation, which the Holder can respond to. Once the presentation has been successfully verified, the IRMA API server assigns a JWT attesting to this successful presentation to the *irma.js* client that initially displayed the QR code that initiated the credential presentation protocol with the Holder.

D.2.3.4 Credential Revocation

IRMA only supports revocation by the Holder, currently Issuers are not able to revoke credentials they have issued [450]. It is recommended to include expiration dates within credentials to mitigate the risk of Issuers being unable to revoke credentials as this would require Holders to periodically get their credentials reissued.

D.2.4 Key Management

Private keys for credential issuance can be passed in to an IRMA API server using XML files with a specific structure. These keys can also be generated using the command line interface provided by IRMA. Public issuance keys are made available to other actors through a centralised scheme manager. Currently, this is a Github repository managed by the Privacy by Design Foundation.

D.2.5 Cryptographic Signature Suites

IRMA implements the idemix cryptographic specification developed by IBM [238], apart from domain-specific pseudonyms and range proofs to reduce the complexity of the system [448]. This is theoretically the same cryptographic signature used by the HVIEP, the CL-RSA signature scheme, although implementations are distinct and neither are yet standardised making it likely that they are incompatible.

D.2.6 Data Storage

Credentials are stored by Holders within their own IRMA mobile application. The IRMA API server is not designed to store credentials, only issue, request and verify them from

Holders with IRMA mobile applications.

D.2.7 Recovery and Backup

IRMA does not currently support recovery and backup. However, a masters thesis from Radbound University demonstrated how credentials held within IRMA mobile applications could be backed up and restored within another device. This work emphasised the importance of strong encryption and two factor authentication, enforcing the participation of a trusted third party in the recovery process [451].

D.2.8 Schema Management

Schema within IRMA based identification systems are managed by a centralised scheme manager. This role could theoretically be done by any entity trusted by all parties in the system. At the moment the Privacy by Design Foundation fulfills this position by maintaining a github repository containing schema definitions [450].

D.2.9 Complexity / Ease of Use

IRMA is open-source and relatively well documented. The work to use IRMA involves configuring a IRMA API server, the getting a scheme manager to add the necessary scheme and issuer public keys into the repository and then make sure users have the IRMA application on the phones. Additionally, the IRMA API server is made available as a service simplifying the implementation process for those willing to trust an external party [448].

D.2.10 Adoption

The IRMA project is arguably more mature than the HVIEP and other credential-based identification architectures. It has had a working mobile application for some time and has run a number of projects within the Netherlands. This includes government issued credentials from the municipality of Nijmegen [450]. However, adoption beyond

the Netherlands is limited not least because of the lack of open standards making interoperability with other service providers infeasible.

D.2.11 Transparency / Governance

The IRMA project originated within academia and all of its code is open source and available on Github. The Privacy by Design Foundation was established as a non-profit responsible for maintaining the code and performing in the role of scheme manager. The foundation effectively acts as the root CA for IRMA credential-based identification systems. While it is intended for other actors to fulfill this role in the future, this transition is yet to occur.

D.3 Serto / Veramo

Serto and Veramo are two projects that evolved out of an early SSI project, uPort. Veramo is an open source, modular, extendable framework for verifiable data an decentralised identity. Serto is a tool suite build using Veramo to streamline the developer experience; including deploying agents, managing schema and evaluating the trustworthiness of other actors. Together the represent an architecture and technical stack based around the use of the Ethereum public blockchain as a VDR supporting credential-based identification systems.

D.3.1 Data Model

This architecture structures credentials and presentations using the VC Data Model specification recommended by the W3C. The proof that encapsulates a signature on this credential uses the JWT format as opposed to JSON Linked Data signatures. Other software systems must be able to parse and understand JWTs to verify these credentials [213]. Fortunately, JWTs are widely supported across many different programming languages and software architectures.

D.3.2 Identifiers

The Veramo framework uses DIDs for the identifiers of credential subjects and issuers. The DID methods currently supported are did:web, did:key and did:ethr, although it is possible to add support for more through plugins to the Veramo framework. did:key is a deterministic encoding of a public key making it simple to use but limited in functionality. did:web uses the .wellknown property of specific websites to manage DID documents, binding an identifier to a specific domain. Finally, did:ethr identifiers are rooted to the Ethereum blockchain. It is up to the implementer and end users to determine which identifiers to use.

D.3.3 Protocol Support

D.3.3.1 Transport and Messaging

Veramo messaging uses an implementation of the DIDComm v2 specification. This means in theory, software artefacts instantiated using the Veramo framework should be able to exchange messages with those instantiated using the HVIEP. Although, not all frameworks within the HVIEP have transitioned to the v2 DIDComm specification.

D.3.3.2 Credential Issuance

Veramo framework has a plugin that adds the credential issuance protocol to Veramo agents. This is a series of messages that get exchanged between two agents as they negotiate the type and contents of a credential to be issued. Currently, this plugin only supports W3C credentials encoded as JWT.

D.3.3.3 Credential Presentation

Credential presentation support is provided by the same plugin that adds credential issuance to the Veramo framework. Again this is a DIDComm protocol defining the series of messages that need to be exchanged to facilitate a credential presentation. As with credential issuance only JWT encodings of VPs are currently supported.

D.3.3.4 Credential Revocation

Credential revocation is not obviously supported by Serto or Veramo, although expiry dates could be added to credentials where this is necessary. The original uPort project did have a mechanism for revocation using a credential status registry managed on the Ethereum blockchain. This may not yet be supported, or just poorly documented, in the Veramo framework.

D.3.4 Key Management

Veramo defines a plugin that can be used to define the Key Management Service (KMS) for a specific Veramo agent instance. Currently supported KMS include storing keys in memory and a libsodium backed KMS solution. Additionally, Serto enterprise agents make use of the KMS solution provided by Amazon Web Services. It should be possible to define additional KMS solutions through the plugin interface.

D.3.5 Cryptographic Signature Suites

Veramo primarily supports ECDSA signatures over the secp256k1 curve as this is the curve used by the Ethereum blockchain. It also supports did:key identifiers that use the Ed25519 curve for ECDSA signatures as a verification method.

D.3.6 Data Storage

Veramo includes a data store plugin for object relation mapping databases. This enables Veramo agents to store of keys, DIDs and credentials. Public issuance keys are either made available through a did:ethr resolvable against the Ethereum VDR or through did:web identifiers that can be resolved to verification mechanisms using the domain name system. Serto is a service built on top of the Veramo framework designed for production use cases. It is recommended that it is hosted on Amazon Web Services, including configuring separate remote database.

D.3.7 Recovery and Backup

Veramo does not explicitly support recovery and backup, although Serto demonstrates how this can be achieved. Specifically, if a Serto agent is configured to use a remote Amazon Web Services (AWS) hosted database then backup and subsequent recovery to a new agent instance is easily enabled. However, such data backups can only be used by other Serto agent instances. Furthermore, recovery in the case of a compromised keys is not well documented. The verification methods of DIDs would need to be rotated, or credentials reissued to new identifiers.

D.3.8 Schema Management

Veramo does not come with explicit schema management, but supports the issuance of credentials defined using either JSON Linked Data of plain JSON schema. The Serto tool suite includes a schema management application, that enables the creation and discovery of different credential schema that can be used within issuance by either Serto of Veramo instantiated agents. Schema hosted by the Serto schema application can then be referenced with credentials using a URL. The Serto schema application currently acts as a trusted third party, enabling URLs to be resolved to their schema.

D.3.9 Complexity / Ease of Use

The Veramo framework is well documented and written in TypeScript, a popular typed programming language that enables the framework to support agents deployed in different environments, including mobile and cloud deployments. The aim of the Veramo framework is to simplify the developer experience, whilst enabling flexibility through a modular plugin architecture that encourages interoperability. Serto further simplifies the process of instantiating an agent, by providing a out-of-the-box enterprise ready agent that can be easily deployed on AWS.

D.3.10 Adoption

It is difficult to judge the adoption of the Serto / Veramo agent framework due its recent evolution from uPort. There is limited information on the Serto website about existing PoCs and pilots using this technology. However, the Veramo GitHub repository has been starred 197 times and forked 67 times indicating interest from the developer community.

D.3.11 Transparency / Governance

Veramo is fully open source and available on Github and Serto also has many open source repositories. Both projects are supported by ConsenSys, a large for-profit blockchain technology provider in the Ethereum ecosystem. They also have a Discord server and encourage permissionless innovation on the technologies and tools they have developed. Governance is not explicitly considered and left to the application specific context that these tools are applied.

Appendix E

The Technological Augmentation of Complex Adaptive Human Systems

E.1 Introduction

In the following analysis, the impact of advanced digital Information and Communication Technologies (ICTs) on human social systems is considered from the perspective of complex systems. A system refers to a set of elements, that are interconnected and interrelated to form a larger whole that can be described in terms of its structure and purpose. Systems can interact with other systems, they can form a nested hierarchy of systems within systems. They can exhibit dynamic, adaptive, self-preserving, goal seeking and evolutionary behaviour that will often surprise an observer focused only on events [452].

Systems science shifts the focus away from individual parts, aiming to understand elements within the context of their relationships and the overall functioning of the combination of elements and relationships that makes up the system. It is a study of change, of difference, either between distinct elements or between a single element over discrete time intervals [453]. The objective is to identify the balancing and reinforcing feedback loops driving the behaviour of the system by modelling the information pathways relevant to the area of study.

A system is an incomplete representation of a more complex reality based on ar-

bitrary boundaries that should be reevaluated within the context of the purpose of the study [452]. It is the study of *the difference that makes a difference* [44], with a focus on patterns and relationships that make it a naturally applicable across multiple disciplines.

The study of complex systems built on these foundations of systems theory with the specific aim of understanding systems whose relationships are nonlinear with behaviour that exhibits emergence, adaptation and evolution over time. These systems are open and influenced by the environment that surrounds them. Complex systems theory recognises that the phenomenon and system dynamics observed are impossible to predict, but rather can only be described in terms of a set of possible future states that might occur [454]. It represents a shift from *being to becoming* whereby the historical states of the system have affect on the possibilities in the present [455]. When interactions are mutually adaptive towards both elements in a relationship these systems are referred to as Complex Adaptive Systems (CAS) [454]. These encompass all living systems in interaction with their environment, in these systems it is the relationship that stays relatively constant while the elements adapt [44].

E.2 Human Systems

Society can be viewed through the lens of systems science. Any group of people can be modelled as a single self-organising entity such that all aspects of the interconnections and relationships are mutually modifiable and mutually interacting [456]. In other words a complex adaptive system, that is goal-pursuing, self-organising and evolutionary. While it is possible to draw the boundary of a system around any group of individuals, this is a common system trap [452]. All individuals are embedded within the wider context of our global society which influences the elements, the individuals, within any human system.

CAS emphasises the importance of both the history of the system and its inherent unpredictability that gives rise to its emergent properties that are often greater than the sum of their parts [454]. In human society, the importance of history and its impact on

the present can be seen throughout the system. For example, the Black Lives Matter movement and the systemic failures that created it can only wholly be understood by tracing the historical roots of these issues [457]. Patterns of thought and ways of being that have sunk into our conscious such that we often struggle to recognise the fault at all [458].

This history of the system can be thought to be analogous to the culture described by Tylor, *that complex whole which includes knowledge, belief, art, morals, law, custom, and any other capabilities and habits acquired by man as a member of society* [459] and is often recognised as the distinguishing feature between our species and others [53, 460, 44]. Our ability to evolve culturally, to build on the ready made ideas preserved within the system has propelled us to the technically advanced species we are today. However, it has been argued that the learning and conditioning individuals receive from this cultural heritage distorts our perception of reality and could be responsible for many of the misconceptions and problems we experience in the world around us [461, 458]. Certainly the errors we make in the application modern ideas of identity are full of historical misconceptions are so embedded into the system of thought prevalent in our society that we struggle to break free of them [53, 458].

E.3 Individuals

The individual, the irreducible element of any human system are themselves a complex biological system that is self-organising, self-correcting and evolutionary [452]. All of which happens largely beyond our awareness and comprehension. Additionally as a mammal we been shown to have a primary value system described as multidimensional and non-maximising [462]. However, the individual Homo Sapien evolved self-awareness, the ability to think in terms of the future and to put ourselves into that model [463]. To define our own values to maximise. To control our own destiny.

Within the wider context of society, this notion of control falls away. Individual exists within society, emeshed within social relationships and conditioned through the context of the space and time that they have inhabited. They are in constant negotiation

with society, a process in which they continuously redefine the roles they play, the values they maximise and the purposes they pursue [464, 436]. This process can be thought of as discovering the rules embodied by the current system, rules that are open for interpretation, constantly changing and are never fully decipherable [439].

Systems thinking teaches us the elements of a system are its least important component. They can often be replaced without affecting the overall function of a system [452]. The story of a 16th century peasant Martin Guerre replaced by an imposture for three years is evidence of this fact within human social systems. The relationships remained and were upheld by both parties, so Arnaud du Tilh effectively became Martin Guerre within the context of the the social system of that small Pyrenean village [465]. This story challenges our interpretation of identity being about the individual and highlights that it is the relationships created and maintained through interaction that matter.

E.4 Interaction

Interaction, the active exchange of information between elements, is a crucial aspect of any system. These exchanges are governed by the rules of the system and can sometimes be represented through non linear equations describing the relationship between elements [453, 452]. In human systems, this mechanisms and rules governing this exchange of information has evolved alongside us with language being only the latest iteration.

When we interact we encode information into our messaging and signals that provide comment on the relationship and the trustworthiness of the information encoded [57]. These are generally semi-voluntary, applied reflexively by the individual implying an evolutionary emphasis on the importance placed on honesty within the system [13]. Within this communication system that humans and other mammalian animals exhibit, it is recognised that signals can be trusted, distrusted, falsified, denied, amplified and corrected, hence they are redundantly applied to messaging giving a greater context to the information [8].

E.5 Environment

The physical environment underpins all living systems. Ecology developed specifically to focus on this relationship between the physical environment and living systems. It is the structure upon which interaction occurs that provides the constraints and limitations forming the basis of all rules of a system. The impact of our environment on the evolution of human history is clearly articulated in this publication [1]. The evolutionary advantage of the thinking man arising from the unpredictable conditions in East Africa, the distribution of edible food grain and ungulate mammals throughout the globe, the trade winds across the oceans and the location of energy reserves can all be attributed to the features of the planet we call our home.

It is important to recognise that the rules inherent in the structure of the environment are not static. Rather like all rules, they are open to interpretation within the context of the living systems that inhabit them. Marshall provides an excellent illustration of this interpretation taking place within the geopolitics of today's nation states [223].

E.6 Technological Impact on the System

Successive information and communication technologies have transformed the properties of interaction within society. The complexity of the interrelationships between elements and the rules that determine the possibilities for information exchange are no longer mediated solely by the mechanisms of redundant signalling, observation and language that we evolved over millennia. Even without considering all the other influences on society, with changes this dramatic to the structure of the system it is little wonder that the future seems so uncertain and our human systems so unbalanced.

Donella Meadows, a pioneering systems thinker, identified and ranked twelve leverage points for instigating system change by their effectiveness. The technological transformation and its affect can be mapped to multiple of these leverage points [452].

Balancing feedback loops help a system keep aligned with its goal, whatever that is

defined to be. These have been built into societies through our cultural heritage that is emergent in the thought of individuals, both conscious and unconscious [458]. These loops are contingent on the systems capacity to recreate this information within its constituent parts. Digital technology now means we can record, analyse and reinterpret more information with fewer delays by more parts of the system than at any other time in history. Indeed this digital capability itself appears to be increasing exponentially.

Systems also include self-reinforcing feedback loops that drive the system behaviour in certain directions. With the increasing proliferation of surveillance capitalism and the misrepresentation and amplification of information there appears evidence of new reinforcing feedback loops available within society [74]. Those able to curate the information that gets distributed within the system have the ability to reinforce the behaviour of the system towards their goals.

Another key leverage point within systems Meadows identified is the ability to restructure who has access to certain information about the system. Missing, incomplete or misrepresented information can lead to bounded rationality, where individuals make rational decisions within the context of the information available [466]. This has been identified as a key cause of systemic problems such as the tragedy of the commons [467]. Clearly technology has transformed the information available to all elements of the system and our ability to interpret meaning from this information. However, we are no longer able to filter this information for honesty through the redundant, conscious signals of face to face interaction [13].

Finally, the highest ranked leverage point relevant to the technological transformation of the human system (5th) are the rules of a system. We have already discussed how the rules of natural systems are determined in the first instance by their environment. The possibilities, constraints and incentives that in turn shape the living systems that populate that environment. Well successive application of technologies from the papyrus paper to penny post, to the telegraph all the way up to the digital era of ubiquitous, multi application, multi modal communication we exist in today have transformed the rules of the human system and the flows of information within it [116]. Individuals and organisations experiment with these new possibilities and test their limitations by

creating new virtual environments for individuals to inhabit. These new environments are no longer inanimate, they are carefully designed and developed by us and for us. It is important to realise that this process conveys power, power over the rules governing human interaction within these environments. There can be little doubt that these digital systems present powerful mechanisms for exerting leverage within the human social systems [452].

E.7 Digital Identity Systems

Digital identity as a discipline is focused primarily on designing mechanisms for the identification, authentication of entities and the authorisation of their actions within digital systems. In other words it the study of a set of tools that enable designers of digital systems to construct the rules, define the incentives and enforce constraints. Self-sovereign identity is the latest iteration within this problem space, and undoubtedly it has value to add. However, at a time when the world is experiencing unprecedented global challenges and the lines between the physical and digital systems are becoming increasingly blurred is it time we re-evaluated the implicit assumption within digital identity. That the world needs better, more effective, more precise mechanisms for enforcing the rules? In some cases this is desirable, but is this something we should aspire for within our human system as a whole?

Any digital identity system that mediates human interaction, is an extension of the complex adaptive system of society. A new environment for humanity to inhabit. They interact in a mutually adaptive way, human systems evolving digital ones which in turn adapt the systems that created them. And on and on. These systems thrive on unpredictability and the emergence of adaptive and evolutionary system properties that this brings. If we are not careful in our design of these systems, the rigidity of the rules within digital systems could end up limiting the unpredictability and evolutionary flexibility within human society.

Furthermore, it is important to remember that Meadows only ranked power over the rules as the 5th most influential leverage point in systems. With the capacity for self-

organisation, the goals of the system, the paradigm under which the system operates and the ability to transcend paradigms all ranked higher [452].

This raises some important questions about the design of digital identity systems:

- Who makes the rules? How are they challenged? And how are they changed?
- How do we build virtual environments that support the capacity for those that populate these environments to evolve the structure and rules of the system?
- If we create digital identity systems that are so effective at enforcing the rules, must we not consider the paradigms that provide the context within which the goals of the system are determined?
- How might we ensure that these systems are designed to evolve beyond the current paradigms within which they were created?

E.8 Looking to the Future

The COVID-19 crisis we are collectively experiencing, the wildfires in Australia, the US and countless other anomalous and often disastrous events happening throughout the globe are symptoms of a larger problem. The affect of a technically advanced species acting to reconfigure its environment towards some self-defined purposes without their members recognising themselves as part of the ecological systems they have knocked out of balance [461]. In 1971 Gregory Bateson doubted whether such a species could endure [468], and yet almost 50 years later and our patterns of thought remain largely unchanged. To restate Wallace-Wells, we have done more damage to our environment knowingly than we ever did in ignorance [469].

For all the advances in digital technology over the last 30 years, there is a case to make that we are more divided than ever before. Thoughts of necessity are inflamed by the lies and misconceptions we hold about identity often intentionally amplified by information technology to keep us apart [458, 53]. Digital identity must be considered within this context if we want to avoid creating further tools for those with knowledge

and power to strengthen the structures we played no part in designing and have little means to change [53].

Humanity must learn to see ourselves as entangled within a complex web of interconnected systems, feedback loops and non linear relationships with the planet and everything in it [452]. There is a need to develop a future consciousness, to extend our time horizons beyond the next election cycle or business year. Bill Sharpe encourages us to see the future as a part of the system we exist in today, the actions we all take can both create and limit possibilities of the future [470]. Not just for me or you, but for future generations who must inhabit this earth after we are gone.

Appendix F

Identity and Interaction in Complex Human Systems

The following diagrams all began as drawings in my notepad produced throughout the course of my studies in an attempt to produce a visual representation of the different elements of human systems of interaction. All diagrams are heavily influenced by the literature presented in this thesis, especially Chapter 2. They are intended to be rich in detail, playful and thought provoking. While most diagrams have been referenced throughout the thesis, they have been included in this appendix without accompanying descriptive text leaving them open to the interpretation of the subjective observer. Hopefully they stimulate some interesting patterns of thought for you.

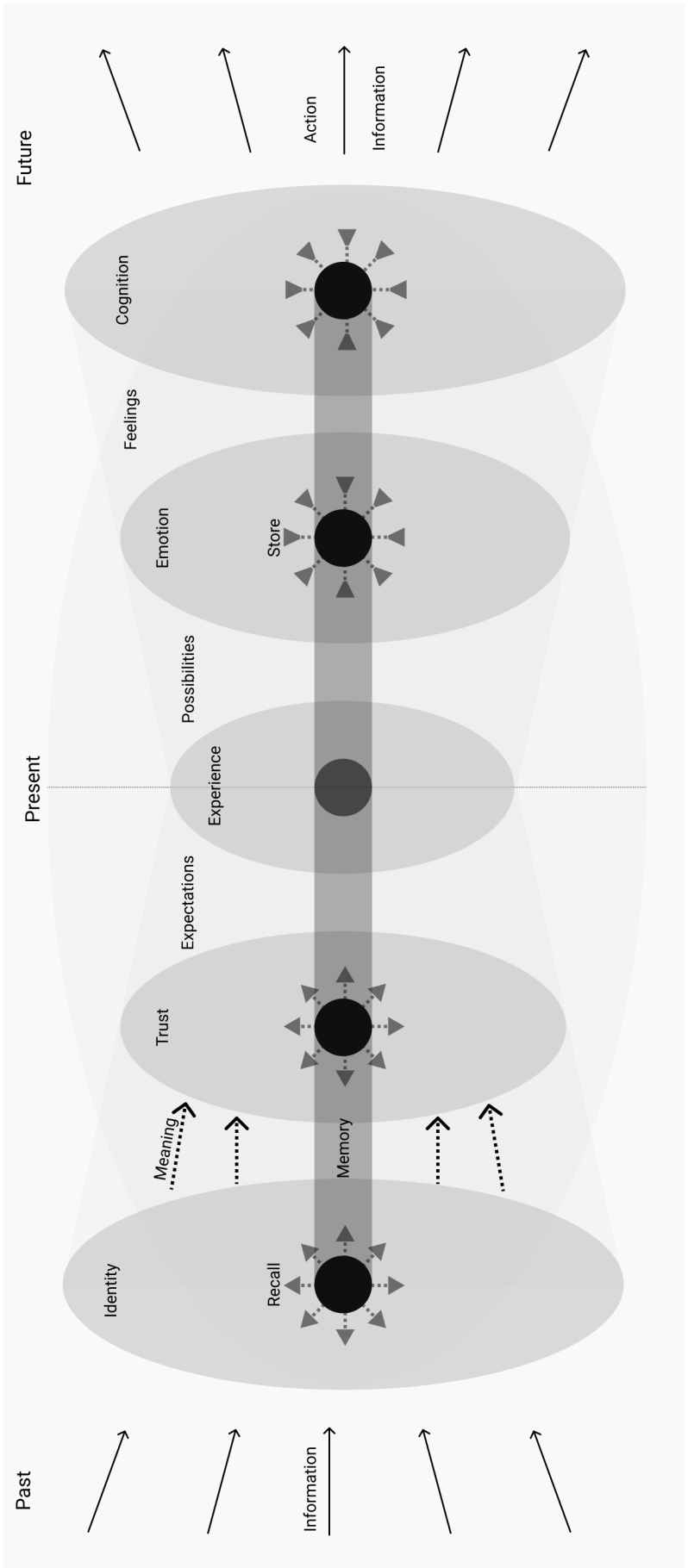

Figure F.1: Process of Experience

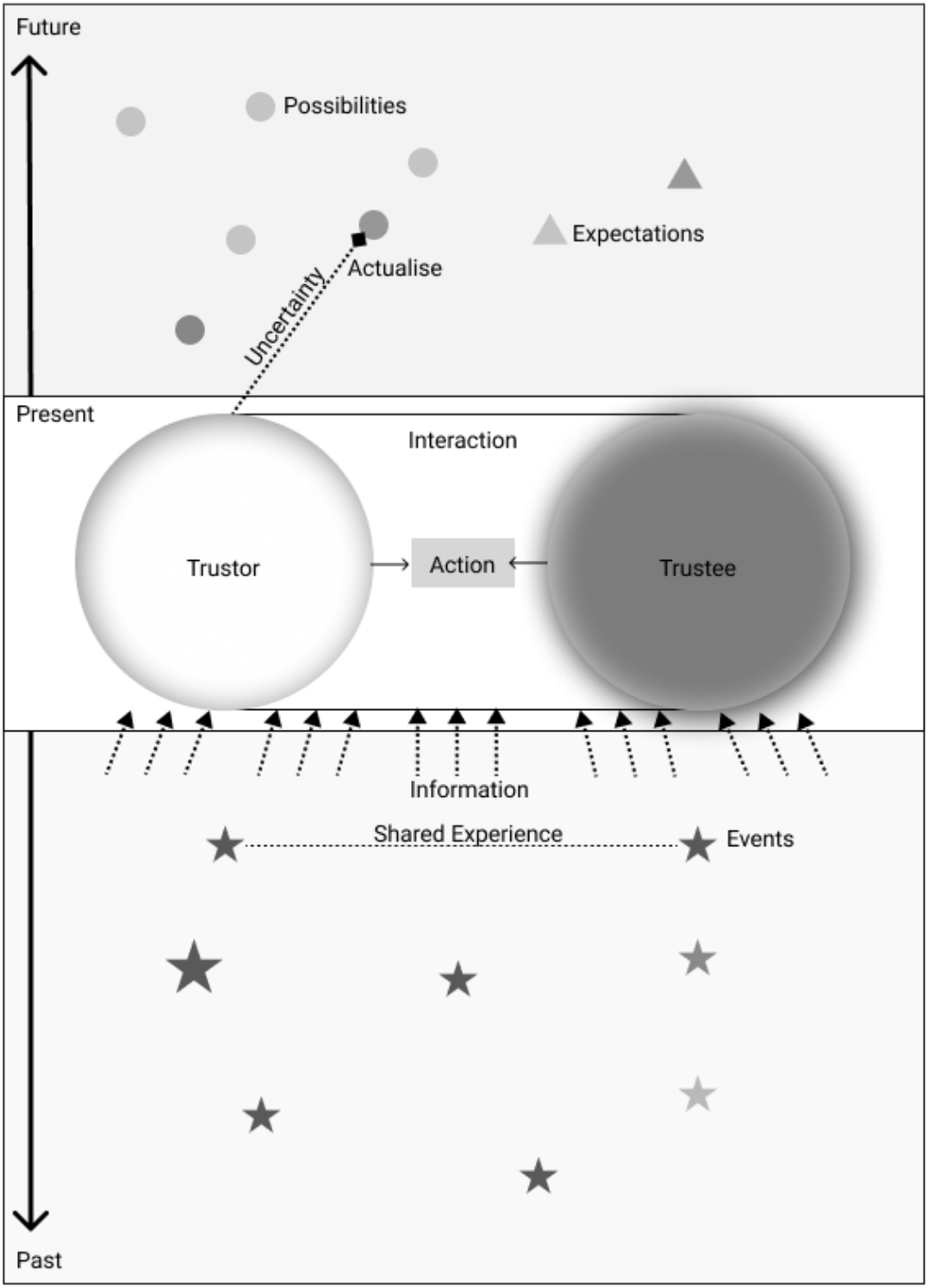

Figure F.2: Trust Placed in The Present (Based on Luhmann's Constancies and Events [2, pp. 12-20])

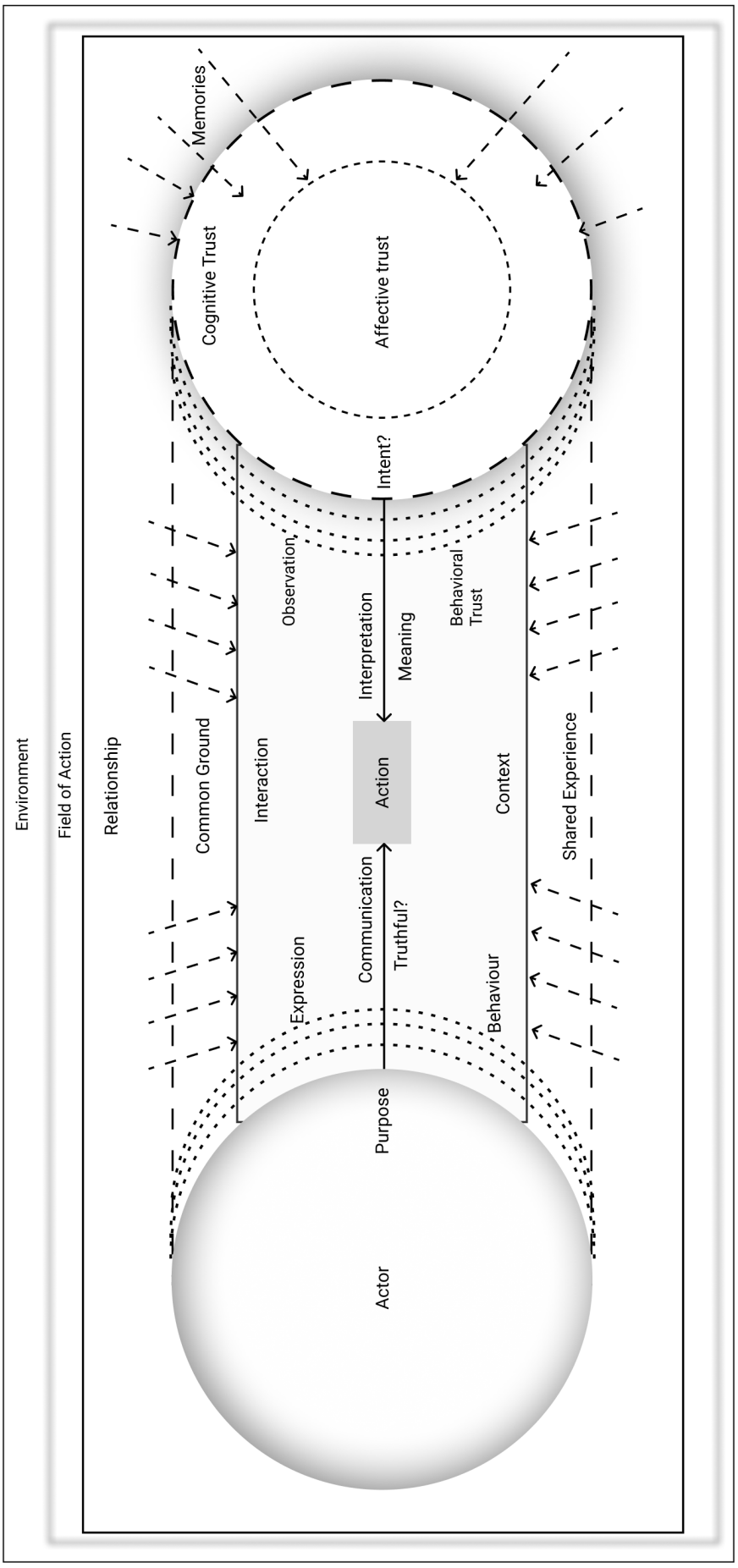

Figure F.3: Human Interaction

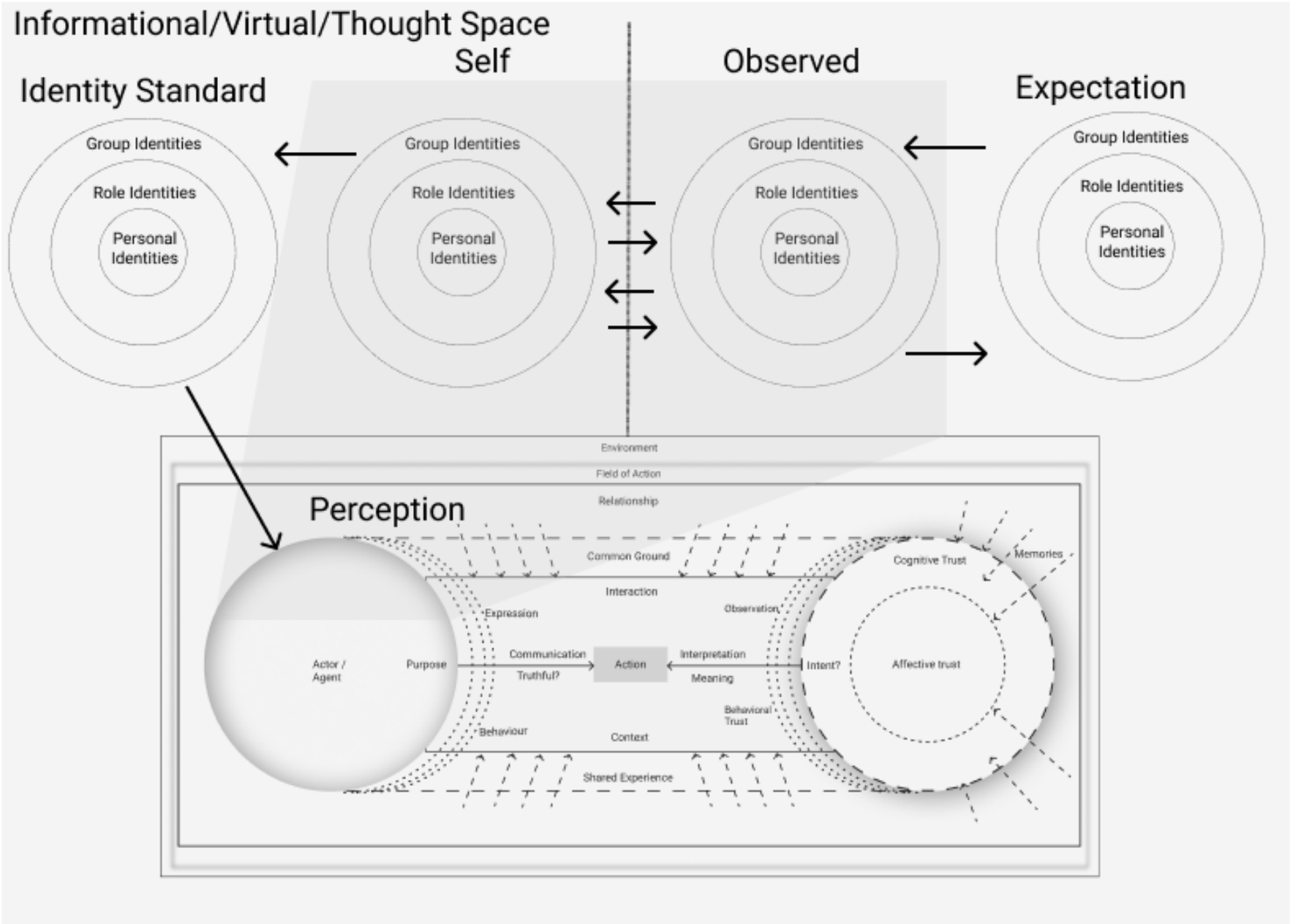

Figure F.4: Identities and Interaction

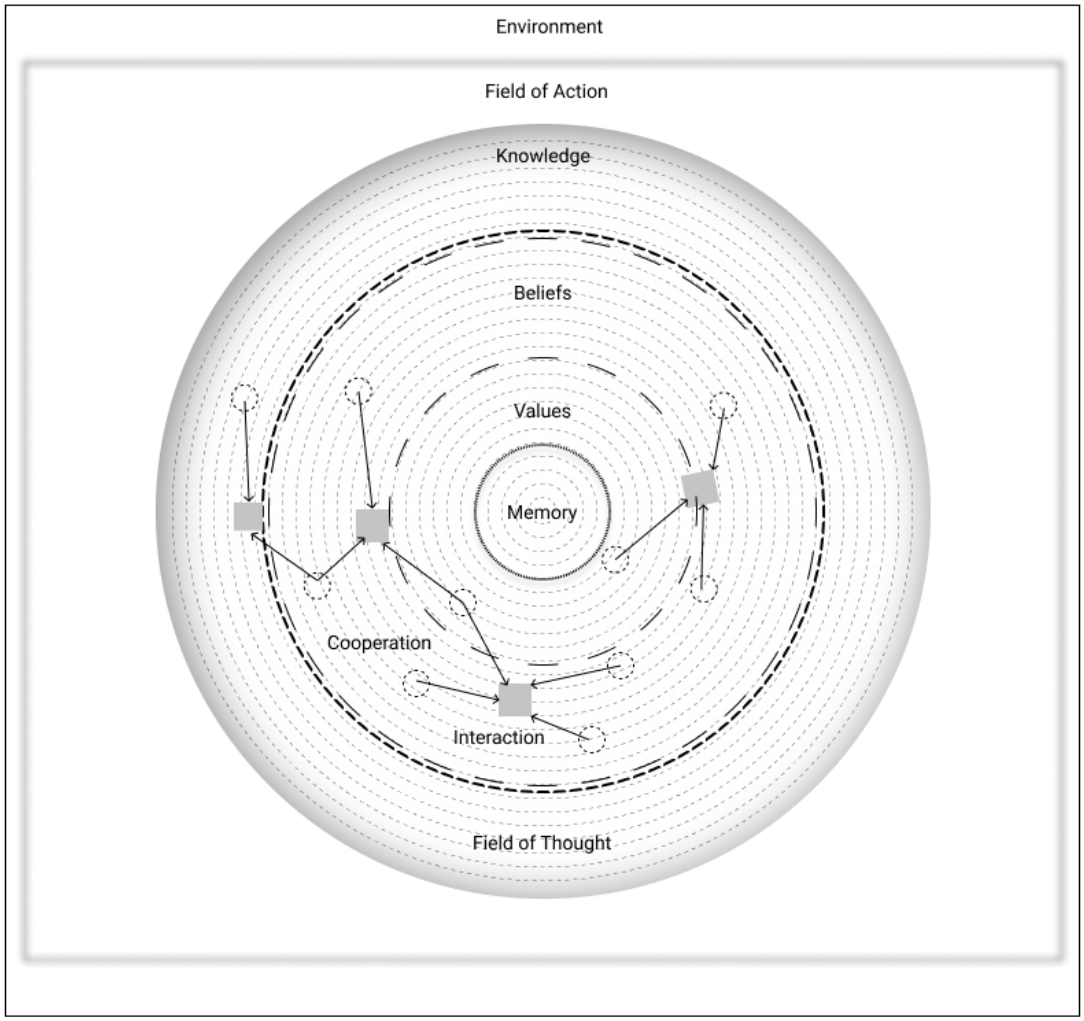

Figure F.5: Collective

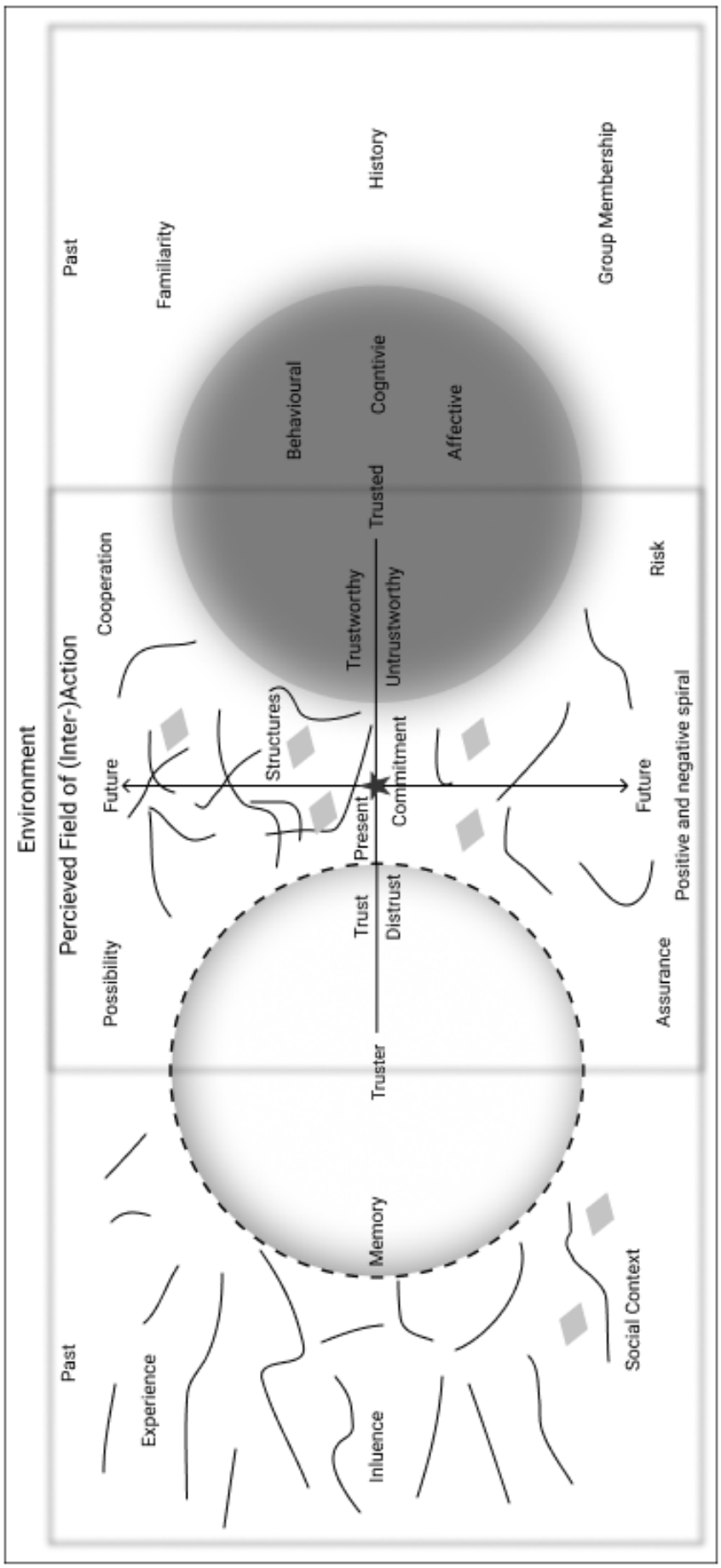

Figure F.6: Trust and Distrust

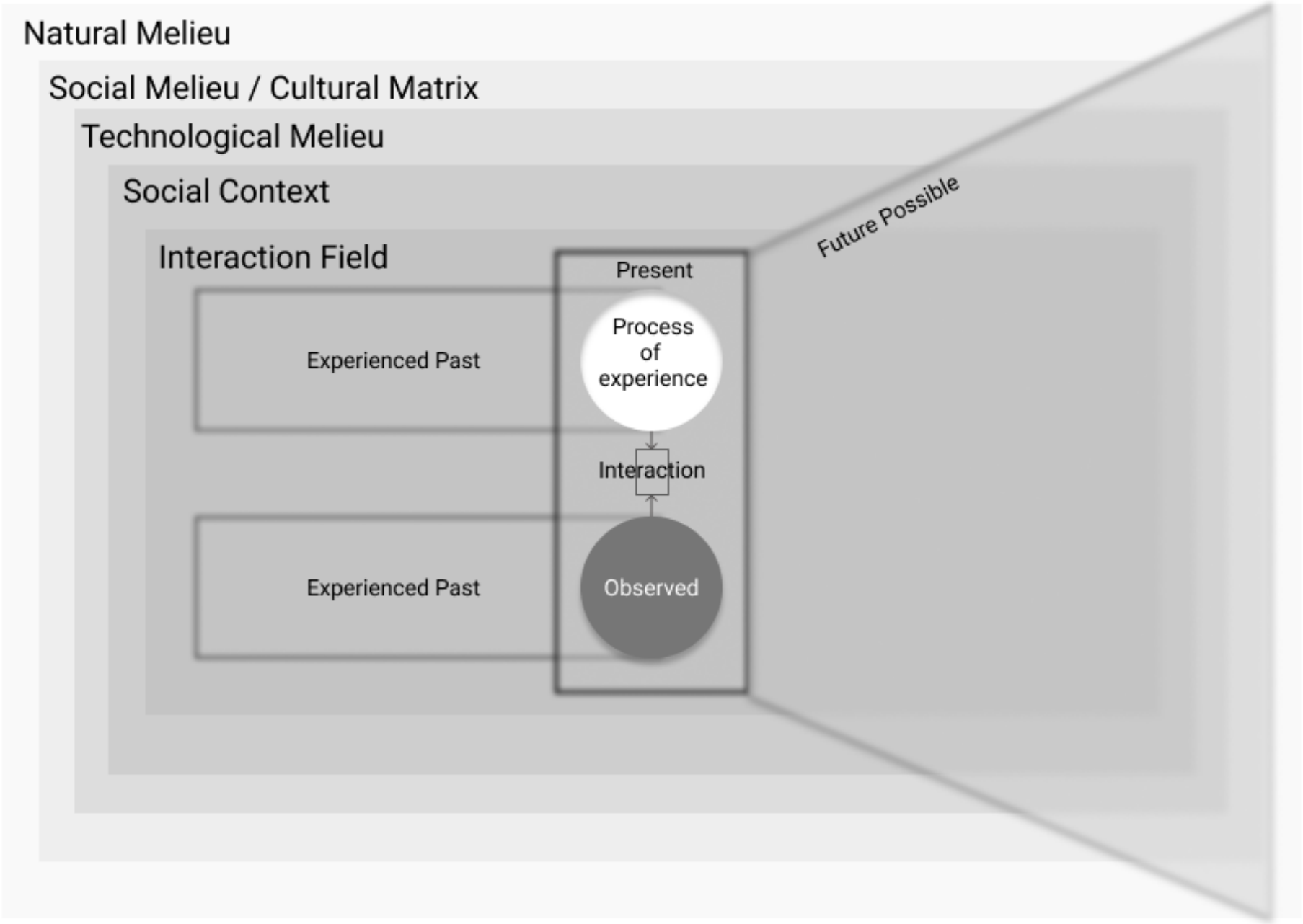

Figure F.7: The Social and Historical Context that Influences Interaction

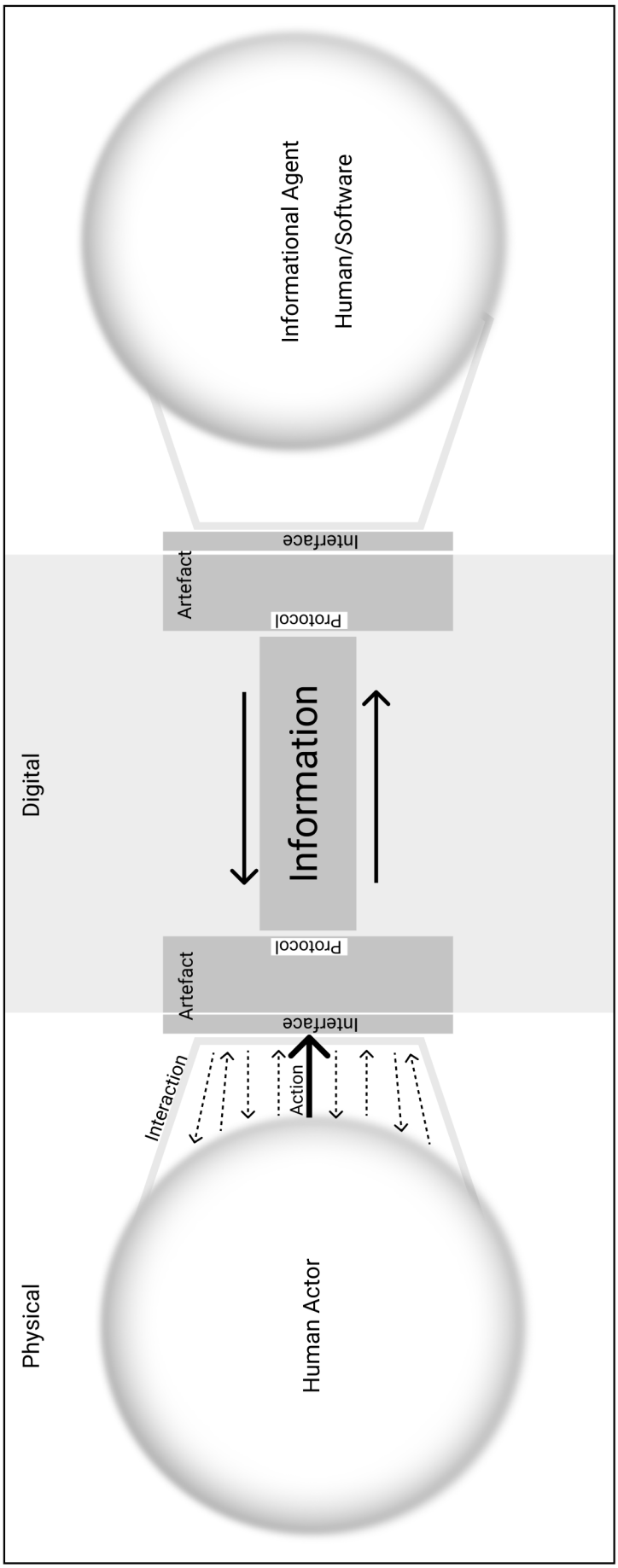

Figure F.8: Digital Mediation

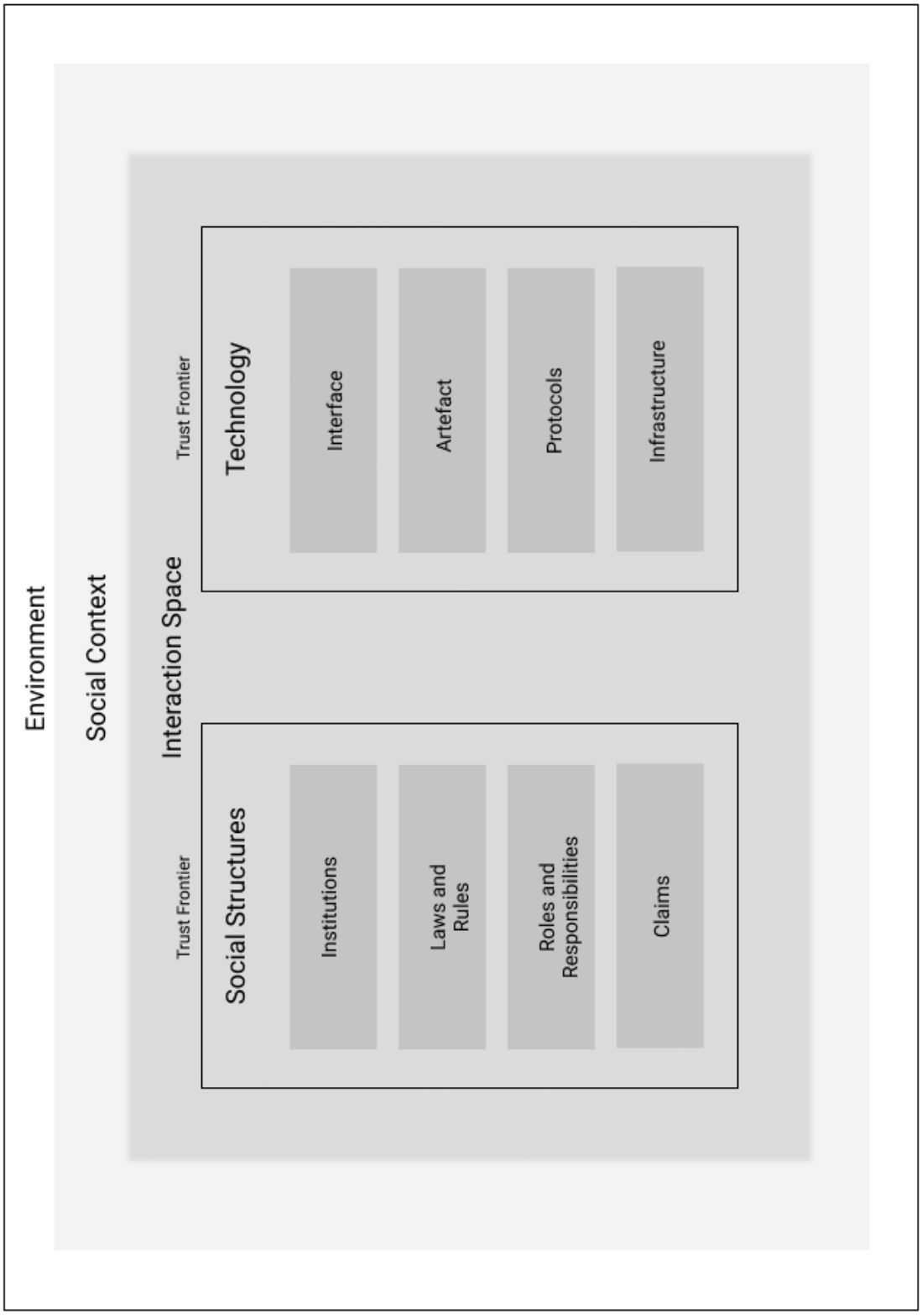

Figure F.9: Interaction Design Space

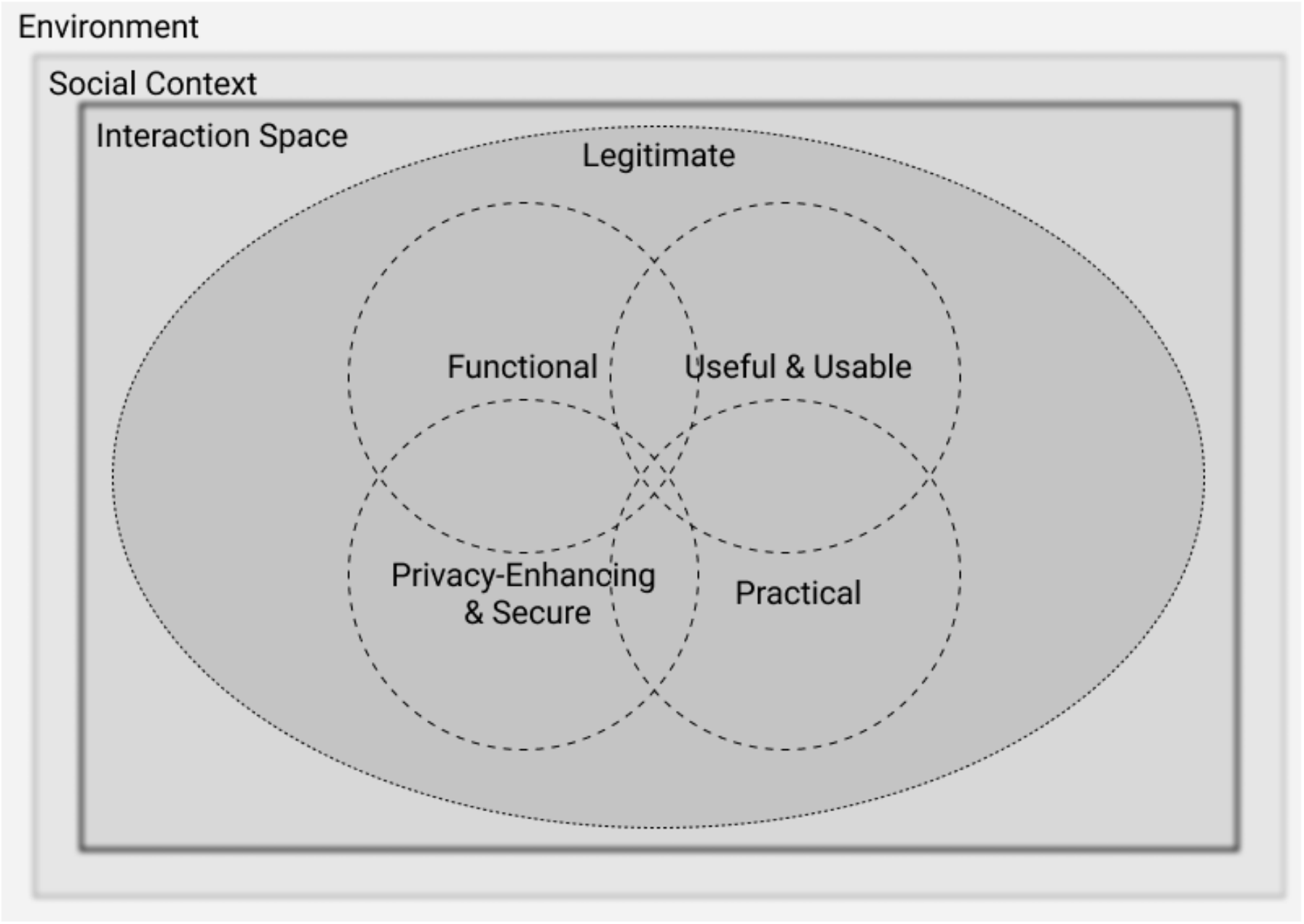

Figure F.10: Technology Design Space

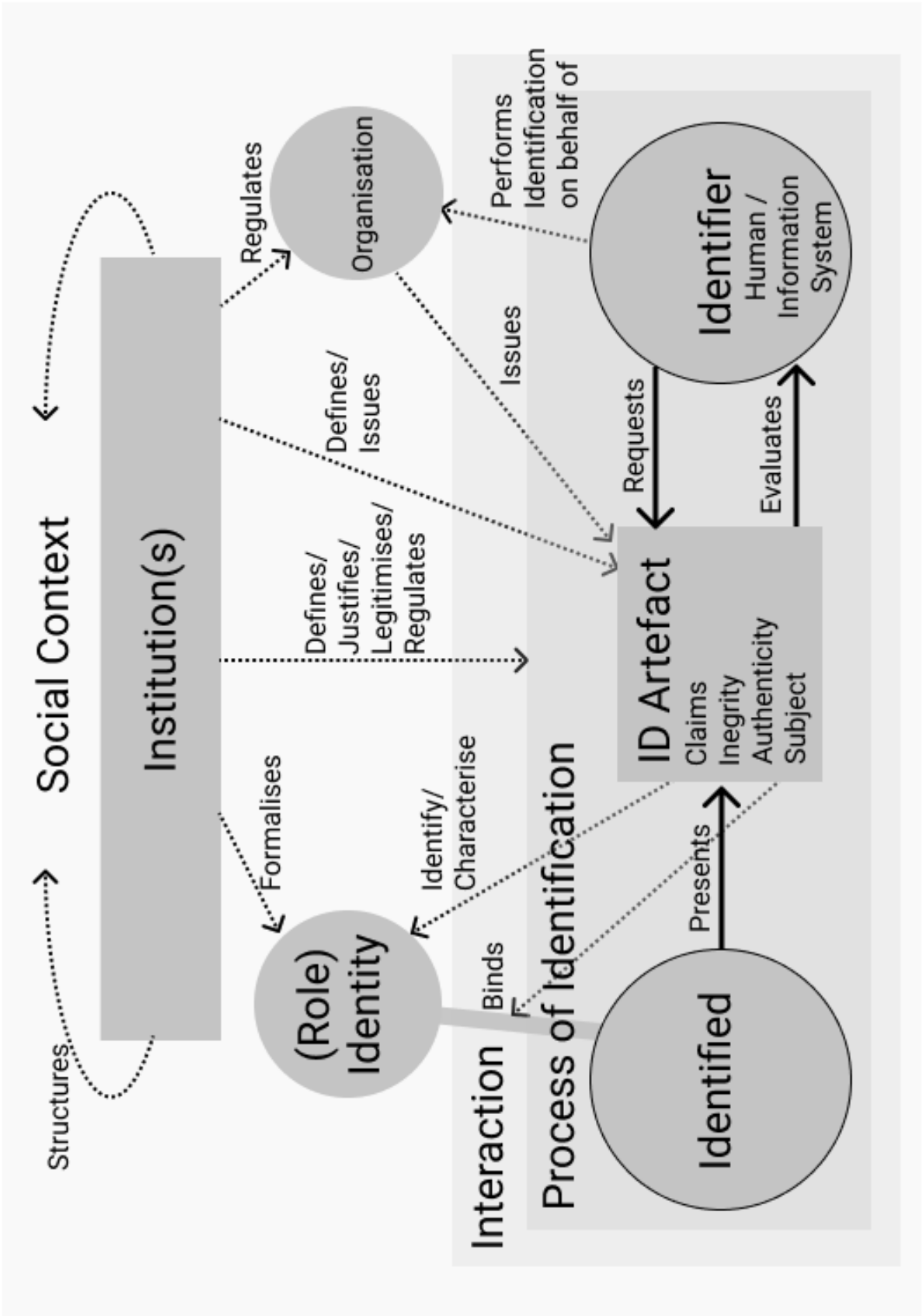

Figure F.11: Formal System of Identification

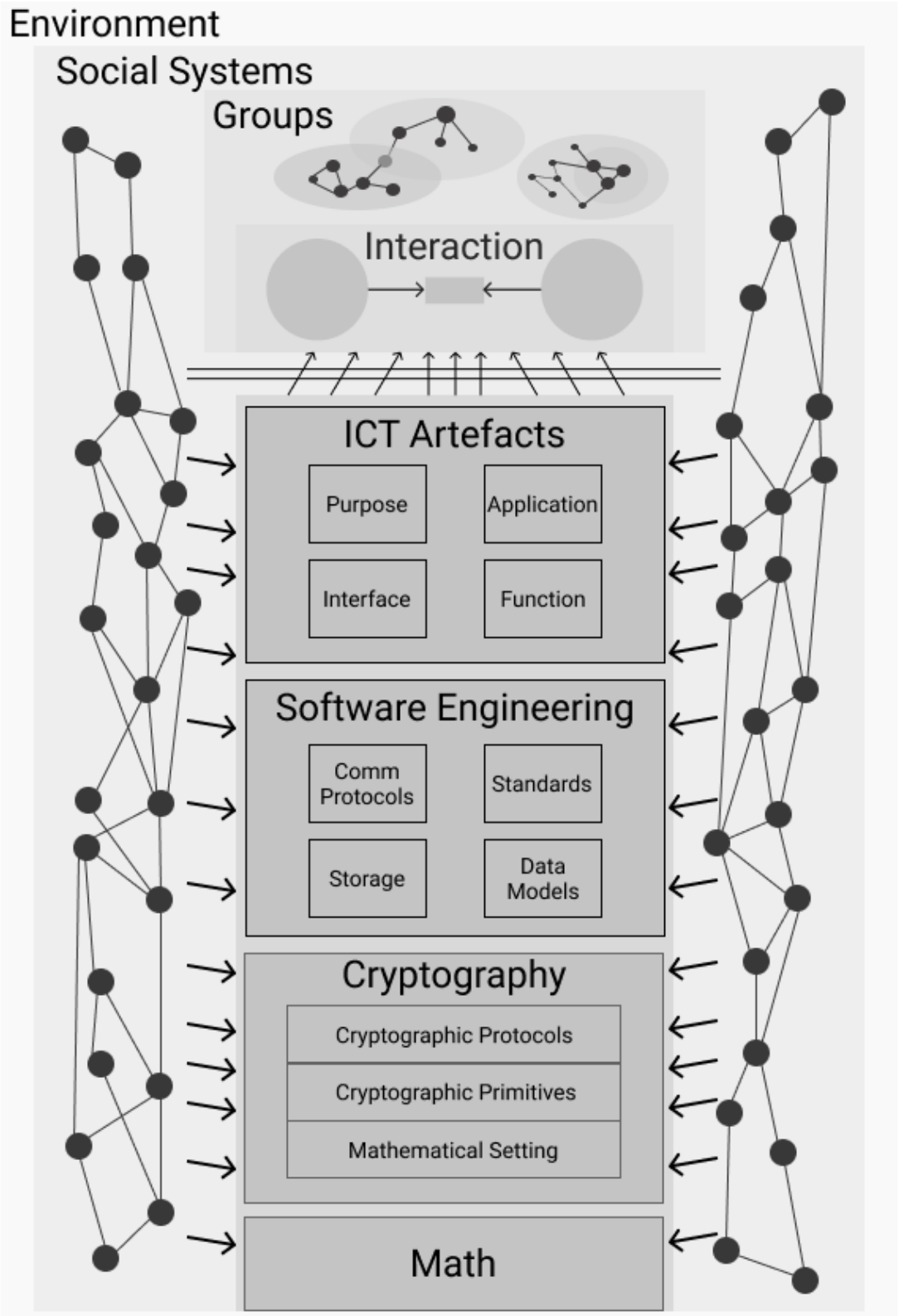

Figure F.12: Synthesis of Technologies Built on Scientific Fundamentals, Produced By and Embedded Within Human Systems of Interaction

Appendix G

Publications

This is a list of publications that have contributed to the completion of this thesis.

- A distributed trust framework for privacy-preserving machine learning [431]
- Trust-by-Design: Evaluating Issues and Perceptions within Clinical Passporting [41]
- Evaluating trust assurance in Indy-based identity networks using public ledger data [42]
- PyDentity: A playground for education and experimentation with the Hyperledger verifiable information exchange platform [43]
- PAN-DOMAIN: Privacy-preserving Sharing and Auditing of Infection Identifier Matching [471]
- Privacy and trust redefined in federated machine learning [432]
- Identifying Roles, Requirements and Responsibilities in Trustworthy AI Systems [472]
- Syft 0.5: A Platform for Universally Deployable Structured Transparency [473]