Microtargeting or Microphishing? Phishing Unveiled

Bridget Khursheed[1], Nikolaos Pitropakis[1], Sean McKeown[1], and Costas Lambrinoudakis[2]

[1] School of Computing, Edinburgh Napier University, Edinburgh, United Kingdom
40311221@live.napier.ac.uk, {n.pitropakis,S.McKeown}@napier.ac.uk
[2] Department of Digital Systems, University of Piraeus, Greece, clam@unipi.gr

Abstract. Online advertisements delivered via social media platforms function in a similar way to phishing emails. In recent years there has been a growing awareness that political advertisements are being microtargeted and tailored to specific demographics, which is analogous to many social engineering attacks. This has led to calls for total bans on this kind of focused political advertising. Additionally, there is evidence that phishing may be entering a more developed phase using software known as *Phishing as a Service* to collect information on phishing or social engineering, potentially facilitating *microphishing* campaigns. To help understand such campaigns, a set of well-defined metrics can be borrowed from the field of digital marketing, providing novel insights which inform phishing email analysis. Our work examines in what ways digital marketing is analogous to phishing and how digital marketing metric techniques can be used to complement existing phishing email analysis. We analyse phishing email datasets collected by the University of Houston in comparison with Corporate junk email and microtargeting Facebook Ad Library datasets, thus comparing these approaches and their results using Weka, URL mismatch and visual metrics analysis. Our evaluation of the results demonstrates that phishing emails can be joined up in unexpected ways which are not revealed using traditional phishing filters. However such *microphishing* may have the potential to gather, store and analyse social engineering information to be used against a target at a later date in a similar way to microtargeting.

Keywords: Phishing · Email · NLP · URL Mismatch · Visual Placement Analysis.

1 Introduction

The effectiveness of Digital marketing has developed significantly with online delivery via social media channels. Marketing campaigns can now be targeted at very specific groups and deliver very specialised, tailored, messages. The campaign can even be customised and reactively modified in response to data gathered through this granular activity. This is known as *microtargeting*. Social awareness of microtargeting in social media, especially in the case of political advertisements, has caused increasing disquiet over its intentions [1].

Digital advertisements on social media operate in some ways that may be considered similar to phishing emails. Indeed, digital advertisements may themselves be used for phishing. For example after Brexit, it became clear that users had been encouraged to self-identify using an application named *thisisyourdigitallife* to receive direct messaging from the Leave campaign [2], which was intuited by clicking links tailored to their interests from previous interaction with data-gathering surveys. It is therefore prudent to explore existing marketing techniques and how their success is measured [3], and to explore if such analyses can be used as a preventative technique to stop users interacting with phishing emails [4]. Phishing damages and devastates business, both in terms of reputation and financial losses. If digital marketing operates in similar ways, it is essential to gain greater understanding of how online digital marketing operates and whether that understanding can be leveraged to reduce the success rate of phishing attacks.

Phishing email methodologies have made use of several approaches. Natural Language Processing (NLP) which uses data analysis to identify linguistic elements such as for example how grammar indicators can be used to assess phishing emails – use of imperatives, multiple verbs, intensifiers, time-related words, incorrect grammar, typos, lingo etc. [5]. Other social indicators can also be measured by NLP to identifying the sender of a *stereotypical phish* through, for example, foreign language identification, negative tone, demands and incorrect grammar [6]. URL mismatch is a key indicator that there is a likelihood of criminal intention in phishing email, which is often considered in terms of the destination website which may appear to be a well-known brand destination but links to a different site. This is known as domain squatting/combosquatting as criminals register domains, which, for example, are spelled similarly to a target legitimate site, acting as phishing destinations to entrap unsuspecting users [7].

To the best of our knowledge, our work is the first to expore the overlap between phishing and digital marketing. It compares online advertisements delivered via social media platforms to phishing emails and explores whether understanding digital marketing methods, such as microtargeting, can aid in phishing email analysis and identification. We also consider how phishing may work in similar ways to digital marketing in its use of stored user data to collect information and develop campaigns via *microphishing*. The contributions of our work can be summarised as follows:

- We explore the existing literature for a defined set of measurable visual placement marketing techniques and compare them with existing phishing analysis techniques which use NLP, URL mismatch and visual similarity.
- We set up an experiment using Weka [8] for our main analysis along with defined visual placement metrics derived from digital marketing. Our input datasets are Facebook Ad Library microtargeting advertisements [9] and phishing datasets from the University of Houston [10] and Corporate junk mail to compare these approaches and their results.
- We use the test results to compare the efficiency of selected phishing email identification techniques with those derived from digital marketing.

The rest of the paper is organised as follows: Section 2 explores the related literature and compares current digital marketing metrics, while Section 3 analyses our methodology and describes our experimental results. Section 4 evaluates and discusses our findings. Section 5, draws the conclusions giving some pointers for future work.

2 Related Literature

Phishing is a method of social engineering [11]. It exploits malicious email and/or websites to gain customer information. Emails are sent with the aim of encouraging users to follow URL links, with the result that the user gives away information such as credit card details or alternatively downloading files which may appear legitimate but have a malware payload. Phishing email is also used to gather information for social engineering attacks where the aim is a granular accumulation of data on a target company, potentially over a long time period. Phishing is a large-scale problem for companies as 55 percent of corporate emails in 2018 were identified as spam [12]. Spearphishing is a subset of phishing that contains genuine, targeted, information such as family details and/or visually authentic appearance to convince the target that they need to act [13]. Das et al. [14] recently published a survey which explored the existing phishing and spear phishing solutions e based on detection techniques for the known attack vectors (e.g., URLs, websites, emails).

Phishing software has become a commodity as phishing software kits sold by rogue developer are known as Phishing-as-a-service or PhaaS. PhaaS is a ready-made set of tools and resources to enable phishing often, featuring fake e-commerce sites and content for phishing campaigns over a variety of digital channels, including social networking posts, phishing email, sponsored social media network advertisements, etc. [15]. Ethical PhaaS has also been developed in response to phishing. It has two main purposes: *i)* pentesting an organisation; and *ii)* training users to recognise phishing emails. Many techniques are used to identify legitimate emails and separate them from phishing emails. These interventions can happen either as preventative measures or as offensive defence where, users are tested with phishing techniques as part of a controlled programme to identify phishing, where under-performing users are put on training courses. Table 1 details such techniques.

Digital marketing is online sales promotion and brand support often used in in social media campaigns. This channel very often employs, and some argue exploits, user data to promote their wares or message in an attractive way to a specific target audience. Digital marketing can take place in many forms: for example served as part of channel offerings on the likes of Facebook, YouTube, Twitter and most often targeted to a user's activity and likes etc.; but equally it can be seen within corporate websites where users can choose to click on promotional information to find out more about products or be served or suggested information based on cookies storing previous site activity. Online advertisements are measured in different ways depending on the aim of a marketing campaign, which are detailed in Table 2.

Table 1: Phishing Detection Techniques

Technique	Description	Effectiveness
NLP	Key words, parts of speech, jargon, typos, bad English etc. [10]	Successful if defined list is effective (cp whitelisting) but not zero day.
URL mismatch	Different URL to the one that is shown, URL that is almost correct, combosquatting; also used NLP [7]	Successful if implemented correctly, users can be trained to recognise this.
Sentiment analysis	Identifying phishing by picking up on emotion words e.g. urgency, intensity;also used NLP [6]	Provides a fine-tuning of NLP by picking out tone e.g. identifying rogue employees
Visual placement techniques	Looking at how the information is loaded in the email; breaking down the mail in terms of visible [16] sections/content ratio	Visual placement analysis and metrics techniques. Table 3
User training	Ethical PhaaS training to identify users who succumb to phishing and re-educate the [15]	Training an established part of corporate life but may train users to normalise phishing and underestimate its threat

Table 2: Online Advertisement Techniques

Approach	Concept	Techniques
UI granularisation	Positioning of information within the online ad, like UIs, online advertisement success can be improved depending on where on screen it is positione [17]	Breakdown of advertisement into constituent measurable parts. Eye-tracking studies can show how users interact with advertisements in great detail. Heat charts are often used to visualise activity.
Images	Effect of images, image size, colours etc.[18]	Use colour to highlight or diminish certain factors onscreen.
Power words	Words considered strong tools for advertising messages [19].	Certain words are shown to be more effective than others at capturing user attention.
Information diffusion	Desire to share marketing information and its effect [20].	Call to actions that encourage viral spread of advertisement Buttons/links e.g. Share option
Homophily	Wanting to be like other people and how that effects user online advertisement response [21].	Aspirational image Relatable models Nowadays this might be called the influencer-effect
Microtargeting	Repetition of advertisement to gain detailed picture of target group [2].	Repeated serving of online advertisements to acquire detailed information on target allowing the serving of tailored advertisements Popular with political parties

Microtargeting is a fine-grained online advertisement technique that is delivered across various digital channels alongside other more traditional methods such as targeted mail. Using information acquired from user data can enhance online advertisement success, for example by associating a particular individual's habits and data to enhance product placement. An example may be to analyse certain buying patterns using data mining techniques to identify that a user is pregnant, facilitating the promotion of associated items. Trust can be enhanced by making an ad's function explicit via onscreen messaging. This can be effective as long as it coincides with, rather than disrupts, a user's expectations and sense making [22]. There are inherent similarities between phishing and digital marketing: a) Phishing and digital marketing are developed and gather data using similar methods, such as PhaaS and microtargeting, and could therefore deliver what might be termed *microphishing* to build a dataset for social engineering purposes; b) Users interpret and choose to respond based on similar trust decisions; and c) Phishing or online advertisement product message and purpose appear similar. Digital marketing and phishing can therefore both be

considered as kinds of social engineering. Superficially, an online advertisement and a phishing email can look very similar as the latter often appears to sell a product or service.

Screen image placement has been researched in relation to software User Interfaces (UIs), websites and other online UIs such as mobile phones using testing techniques, such as eye tracking and screen tracking, to gather hotspot information on where users' eyes are focused on a screen during testing. Gathering eye tracking information showed a strongly consistent screen read and advert handling behaviour by both experienced and new users. Users have an intuitive sense of where the core page content resides, with their eyes barely straying to the footer, or bottom area, of a page, forcing relatively major design changes to draw their attention to a particular area [23]. When image analysis is focused especially on digital marketing and phishing emails, visual placement and use of image metrics may have some specialized applications and goals. Methods used to identify visual placement depend on factors such as the level of matching required, the aim of the match ranging from exact duplicate to rough comparison, computational resources required, and the match speed required. The latter two factors will affect choices made regarding the analysis depth as there must be a trade-off between investment and possible results. Common analysis techniques depicted in more detail in Table 3.

As previously discussed, digital marketing can be viewed in some ways as analogous to phishing. Our work describes the methodology used to explore how digital marketing metric techniques are used to enhance phishing email analysis. Our approach differs to previous work by exploring connections between digital marketing and phishing email by considering microtargeting and how it works and how that might be connected to PhaaS for example *textmicrophishing* where digital marketing and phishing overlap.

3 Experimental Design and Results

Our work explores and measures how online marketing microtargeting based techniques and metrics could be used as a complementary identification method for phishing. Three datasets were used in our experiments. The first, a popular email phishing dataset, was supplied by Professor Rakesh M. Verma from the University of Houston [10] and is available to everyone under request. The second dataset includes Corporate junk mail collected from a single company, received by the main author as part of an email group on Microsoft Outlook from June to August in 2019. The last dataset came from Conservative Party advertisements found in the Facebook Ad Library [9] which was active from 1st of August 2019 to the 14th of September 2019.

In relation to each dataset, further specific features were refined. Regarding the University of Houston phishing dataset [10] the features were confirmed and improved through reference to previous work and subsequently the Corporate junk mail set, for example, derived stopwords to improve results. The full list of our features is described in Table 4. For the Conservative Party advertisements we used features coming from the work of Fu et al. [26], as well as histograms.

Table 3: Visual Placement Analysis and Metrics Techniques

Source	Technique	Paper	Techniques	Limitations
Phishing	HTML analysis & similarity score	Wenyin et al.[16]	Identifying pages based on breaking HTML into 3 categories block level layout and overall style, a similarity calculation based on each element within each category provides alerts based on ranges.	Webpage rather than email; won't work if graphic is used instead of code.
	Cascading StyleSheet analysis & similarity score	Mao et al.[24]	Chrome extension measuring and alerting on key CSS analysis points text pieces including style, embedded and overall browser-rendered visual page appearance.	Webpage rather than email; won't work if graphic is used instead of code.
	Signature-based method using selected pages attributes	Medvet et al.[25]	Visual comparison using three page features namely text pieces and their style, images embedded in the page, and the overall visual appearance of the page as rendered by the browser.	Webpage rather than email; won't work if graphic is used instead of code.
	Low-spec rendering of images and measurement using EMD	Fu et al.[26]	Web pages are converted into low resolution images and images signatures are created using colour and coordinate features. EMD is used to pinpoint differences and is able to recognise and measure similarity rather than direct similarity.	Webpage rather than email
	HTML page check & machine learning	Fette et al.[27]	Spam filter approach based on machine learning technique measuring HTML elements	Minimal visual elements
Digital marketing	Online advertisement metrics to establish relative success	Rzemieniak et al.[28]	How online advertisements are measured to define relative success of strategy e.g. CPM, CPA, CPC, FF and combination approaches	Minimal visual specific elements considered
	Static versus dynamic creative content and its effectiveness	Bruce et al.[29]	Creative format, message content, and targeting on digital advertisement performance of static (GIF) and animated (Flash) display advertisement formats of various sizes analysing results from a major retailer and creating a dynamic model.	Performance based analysis
	Using ghost advertisements or non-advertisements to judge online advertisement effectiveness and improve metrics	Johnson et al.[30]	Ghost advertisements, which are also compared to Public Service Announcements which have been used for the same purpose, are used to measure what users would do had they not seen the brand ad; in order to establish a baseline user metric of interaction outside a marketing campaign.	Minimal visual elements
S.M Images	Visual comparison and techniques trained for usein for example sentiment analysis	You et al.[31]	Machine learning is used with the following process where images were cropped to uniform size, trained using Convolutional Neural Networks progressively in datasets from Flickr and then tested Twitter to analyse political images to predict election results	High time and resource commitment

With regards to the Corporate junk mail dataset, a combination of the above methods would be used to experiment along with adapted methodologies using recurring blocks of HTML or style features. Our experimentation methodology in steps is illustrated in Figure 1 and described as follow: 1) Apply NLP on University of Houston dataset; 2) Apply image metrics to the Facebook Ad Library Conservative party ads; and 3) Apply the same training test model and further visual metrics to the Corporate junk mail dataset (HTML and image source).

Weka [8] was used to explore the datasets. Several algorithms were evaluated in a preliminary study, where poorly performing algorithms such as REPTree were subsequently abandoned in favour of the best performing: IBk, J48, and

Fig. 1: Experiment Workflow

NaiveBayes. These are commonly used as baselines for email spam classification [32, 33], each taking a different approach: **IBk** is a lazy implementation of the k-nearest neighbours algorithm, where output classification is determined via consensus of the nearest node classes; **J48** is a Java implementation of the C4.5, entropy based, decision tree, where the splitting attributes are chosen based on the information gain ratio; and **NaiveBayes** is a simple probabilistic classifier which assumes feature independence. Bayes' theorem is applied to calculate the probability of the data matching each class, with the highest probability match resulting in the final classification.

Our experiments have a combination of results from NLP processing (of which URL mismatch was a subset) and visual placement analysis. The NLP model used a simplified breakdown of an established model [10] with significant additions taken from other factors such as image presence, sentiment analysis, urgency of request, foreign language usage and an emphasis on the imperative verb form. The attributes are described in more detail in Table 4.

Table 4: NLP Attribute model Breakdown

Attribute	Description	Type
Filename	Filename	Nominal
Word count	Amount of words in email	Numeric
Stopword count	Non-key words in email. Adapted from the Natural Language Toolkit	Numeric
Link y/n	Boolean	Numeric
Link number	Number of links	Numeric
URL mismatch	Boolean	Numeric
Spelling mistake	Number of spelling mistakes	Numeric
Grammar mistake	Number of grammar mistakes	Numeric
Punctuation mistake	Number of punctuation mistakes	Numeric
Code y/n	Visible code in email -Boolean	Numeric
Language OE	Email partially or entirely in a language other than English - Boolean	Numeric
Swear	Boolean	Numeric
Sentiment analysis	A – angry N – neutral H – happy	Numeric
Imperative verb	Instance of imperative verb form e.g. Click here or View now Boolean	Numeric
Urgency	Response urgency measured 0-5: 5 – immediate response required 4 - contact requested soon – e.g. meeting, appointment, action tomorrow or next week 3 – contact requested – e.g. meeting, appointment, action 2 – action suggested but no pressure 1 – descriptive e.g. affirmation or congratulatory email 0 – no response required as information only	Numeric
Image present	Image present or not. Only relevant to Corporate junk mail dataset	Numeric
Named entities	Count of proper nouns e.g. names of people, street names etc.	Numeric
Legit	Phishing email or not. Class attribute - Boolean	Nominal

URL mismatch is commonly considered in relation to spoofed sites. However for the scope of our work, the focus is more on a mismatch between the address as shown in the email sender field and the email that it is actually sent from or similar mismatch in any links in the body of the email. This is seen as a key indicator that the email is spam or phishing as described previously. URL mismatch via link or email address is largely pre-identified in the University of Houston dataset where the dataset is clearly divided into Legit and Phish, with the URL and sender details mostly stripped out. The Corporate junk mail dataset contained live junk mail and potential malware and it was therefore prudent to avoid following links and limit them to manual inspection. Our investigation became more interesting when further connections between emails were explored, as several emails were observed to be directly connected. These were not obvious mismatched URLs emails from the addressee; although they were all marked as spam by the spam filter using other heuristics.

Image, look, and feel are key to both digital marketing and phishing email effectiveness. Having identified microtargeting as the main area of interest because of its relationship to the social engineering of political campaigns, and its analogous behaviour to PhaaS output, the experiment centres on interpreting visual similarity. Our methodology examines three visual analysis techniques in more detail in the context of the datasets as follows: a) Signatures; b) Histogram comparison and c) HTML sections/code id using NLP techniques. A selection of advertisements was extracted from the Facebook Ad Library; the baseline focuses on a subset of these which centre on the former Conservative, now Lib Dem MP, Sara Wollaston. These were examined to provide an overview of what a microtargeting image set looks like. There was no accessible background provided by the Facebook Ad Library on the intended strategy changes behind each targeted advertisement. However it is possible to analyse the Facebook Ad Library political advertisements and draw some conclusions. In this case this is done by focusing on one set of advertisements on a similar message.

The experiment aimed to find similarities as potential evidence of *microphishing*. These results focus on the superficially similar footers in the following Corporate junk mail files. Footer areas can often be overlooked when a phishing email modifies an existing template, which, as noted in the literature [34], is often assumed by users (and developers) to contain no important information. In the author's own experience of developing an early company website for Digitext in 1995, the footer was discovered on a Canadian company who had borrowed the HTML source but not updated the footer; this then appeared in a Web search for company mentions for Digitext online as evidence of this adaptation.

The file footers were tested in the following ways: **Baseline MD5**: To emulate images the footer areas were cropped to exactly the same size. **Histogram Comparison**: Although the collected images were not suitable for providing a signature they can produce histograms which show a close similarity. Histogram comparison produces clear results for the very similar microtargeted advertisements and can be refined further; techniques such as Earth Mover Distance measurement can also be used to compare histograms by looking at the distance

required to move the pixels of the image or "earth" from one position to another to measure difference.

We collected our results using the NLP model within Weka using the URL mismatch and visual similarity information. In the first part of our experiment, as depicted in Table 5 (a), we had as input the University of Houston [10] sample dataset. Overall, the IBk algorithm performed best with 99% Precision and 91% Recall while J48 came second with 99% Precision and Recall and Naive Bayes ranked third with 95% Precision and 96% Recall. It can be observed that the performance difference between the lesser performing algorithms is not large. Results for the Corporate junk mail dataset are depicted in Table 5 (b). This time both Naive Bayes and IBk ranked first with aproximately 95% Precision and Recall, while J48 came second with 93% Precision and Recall.

Table 5: Comparative Results of Individual Algorithms (10-fold cross-validation)

(a) Houston Dataset

Algorithm	**Precision**	**Recall**	**F Score**
IBk	0.990	0.990	0.989
J48	0.951	0.960	0.954
Naive Bayes	0.948	0.910	0.926

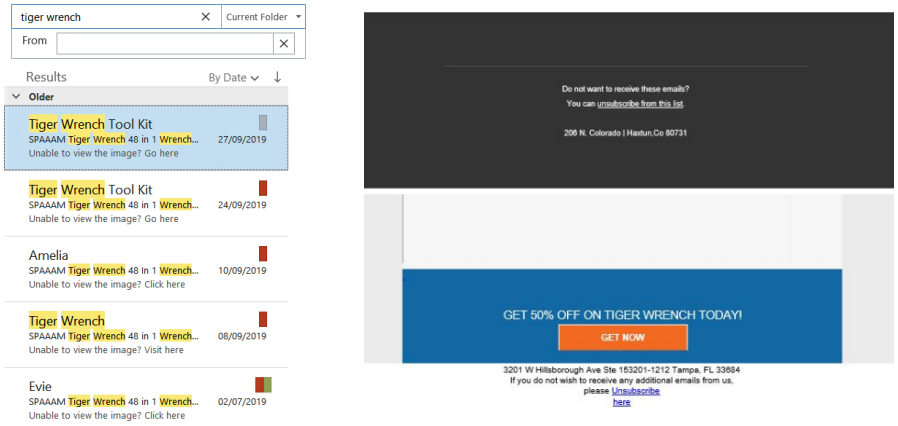
(b) Corporate Junk Mail Dataset

Algorithm	**Precision**	**Recall**	**F Score**
IBk	0.951	0.950	0.950
J48	0.931	0.930	0.930
Naive Bayes	0.953	0.950	0.950

While investigating URL mismatches on the Corporate junk mail dataset, we noticed that an email with the same, or similar, subject line containing similar text and title in its body arrived from 5 different email addresses, as illustrated in Figure 2 (a). Other correlations in apparently different spam email include the use of repeat footer addresses. This could be a regular spam server but potentially could also have the PhaaS capacity to record any responses in a more granular way that could allow results analysis. This could be comparable to the way data is recorded and stored in microtargeted adverts for example to allow the sender to build a bigger picture of the company they are targeting.
The Sarah Wollaston Facebook Ad Library subset was explored in more detail. Microtargeting data are being measured in very small variations in text, colour weight and tone as, illustrated in Figure 3 (a). The comparable metrics were set using histograms. Example histogram difference can be observed on Figure 3 (b), which show how histograms highlight small distinctions in the microtargeting data. Histogram comparison produces clear results for the very similar microtargeted advertisements and can be refined further.

4 Evaluation And Discussion

The primary aim was to develop a working NLP model that identifies phishing methods based on the content of the email, which was shown to be effective in Table 5 above. This section will discuss the feature and dataset performance in more detail. Following this, discussion pivots to the evaluation of visual elements of the email, via direct histogram comparison, cropped footer analysis, and NLP processing of the HTML code itself.



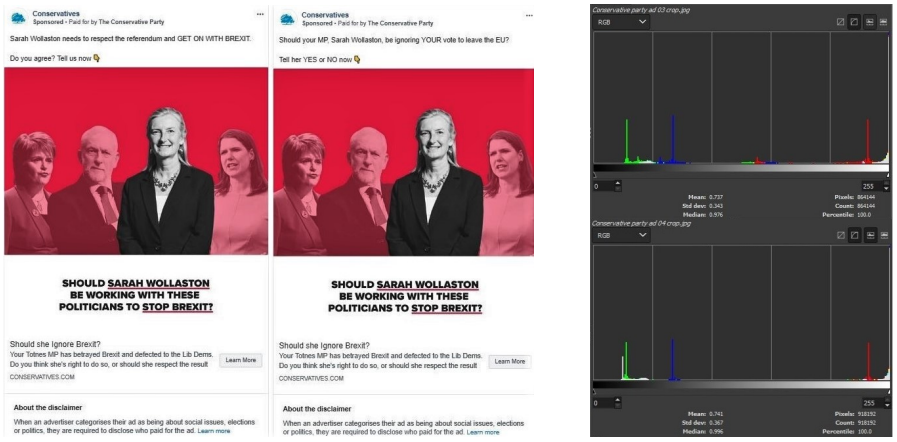
(a) Tiger Wrench Emails (b) Footers

Fig. 2: URL Mismatch Examples

The NLP model was successful in that it was demonstrated to still be an effective method of categorising phishing and non-phishing emails. In a real life example, a build-up of match word knowledge can allow filtering out alongside tweaking known factors like a correlation between imperative words, such as Click and Unsubscribe, and the urgency of phishing emails can be used to alert to heightened risk. The visual metrics and the microtargeting example gave more scope to identify, perhaps in parallel in a real life system, what the implications of such persistent connections could be, i.e. some kind of *microphishing* technique run from a PhaaS software console gathering data on the target, to be analysed at leisure by the acquirers whether pentesters or criminals.

The results of different attributes used by the NLP model and their efficacy in the model will be discussed. *URL mismatch* was a constant factor in phishing emails and was weighted highly in the model. For *Spelling/Grammar/Punctuation mistakes*, there was a more noticeable correlation between phishing and language mistakes in the University of Houston dataset than in the contemporary Corporate junk mail dataset. Visible *code* in the email occurred in both datasets but did not appear to have a strong correlation with phishing email; emails with malicious code would be filtered out before hitting the Corporate junk mail mailbox. *Sentiment analysis* was explored firstly in terms of Happy/Neutral/Angry tones in the University of Houston dataset, but this category worked less well for the Corporate junk mail dataset which ended up being defined as largely neutral in tone. The *imperative verbs* facilitate a quick user response and were always present in phishing emails. They were usually in combination with the next attribute Urgency to push a quick unconsidered response. This was a weighted attribute to heighten the effectiveness of the model. There was also a strong correlation between phishing emails and *Urgency*.

J48 was noticeably more effective in the Corporate junk mail dataset compared to the University of Houston dataset samples; this is likely to be due to a more balanced sampling. URL mismatch was factored into the Weka model



(a) Facebook Advertisement Pictures (b) Histograms

Fig. 3: Visual Similarity Examples

originally developed using the University of Houston dataset. There is an almost complete correlation between phishing and URL mismatch in the instances used from the Houston dataset as identified in the benign and phishing separation within the dataset. Having this clear identification made it easier to train the model. However the Corporate junk mail dataset had access to all header information and the experiment had to use resulting evidence to identify phishing email. Although URLs were only explored through manual inspection, as this dataset contained live potentially harmful emails, URL mismatch is a significant factor in assessing phishing email. For example, the clearest phishing with perhaps the most obvious malevolent intent spoofed the leading author's company.

Looking at URL mismatch on the basis of developing the model using the corporate junk mail dataset, and using visual similarity, it became clear that there were significant unexpected connections through almost 50% of the dataset contents coming from the same effective second level domain. Footer addresses reveal a connection between emails as they showed similar subject lines. Consequently, a simple assumption of URL mismatch as an indicator that an email was non-legitimate became more nuanced.

Visual Placement

Digital marketing relies heavily on images and that informed the selection of techniques and the way data was collected. A set of results were collected for visual placement using hashes, code snippets, and histograms from two of the datasets (excluding the University of Houston dataset which is text only and therefore could not be evaluated in this way). By examining the Sara Wollaston dataset information, it was possible to see microtargeting in action.

The voters targeted lived mostly in England but with small numbers in Wales and Scotland. The target of the advertisements is overwhelmingly statistically older men and women as illustrated in Figure 4. Less than £100 was paid for

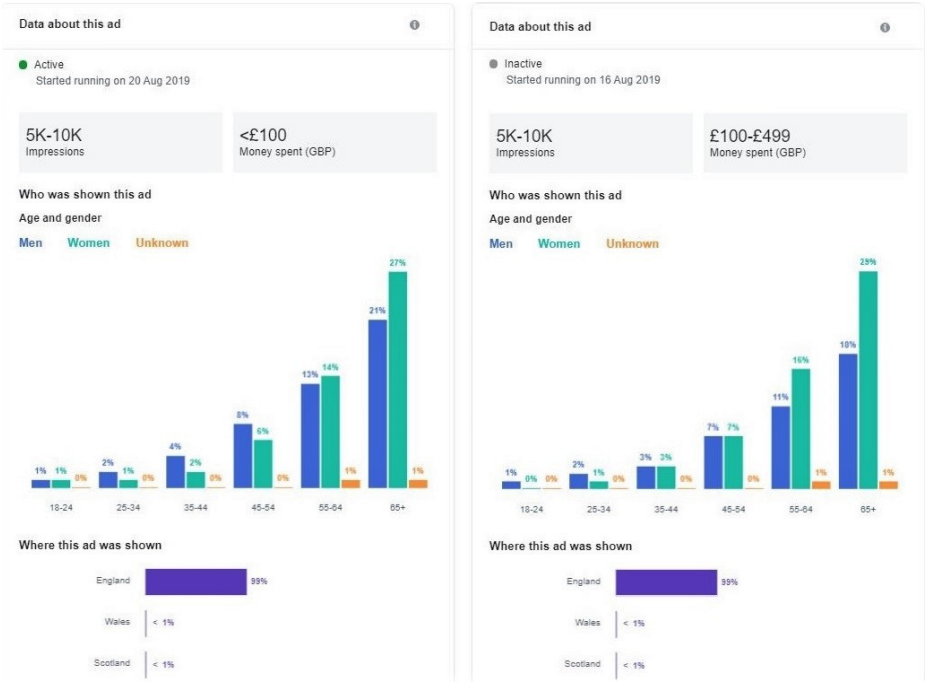


Fig. 4: Conservative Party Advertisement Statistics

all the advertisements targeting, with the exception of one which cost less than £499. Ad 23 had more money spent on it and was aimed predominantly at older women over 65 and achieved between 5000 and 1000 views. Only Conservative Party Ad 07 had similar impression numbers although less than £100 had been invested. It was primarily displayed to women over the age of 55 years old. Both of the most popular advertisements show the same image with the same text and calls to action. Conservative Party Ad 23 started showing on the 16th August and was inactive when the images were gathered; whereas Conservative Party Ad 07 started showing on the 20th August and was still active.

Although prior to testing the visual placement metric techniques seemed comparable to NLP, the results indicate that they appear less flexible and discriminative for longer-term use. This could mean it would be harder to evolve results as different factors are important in a fast-moving corporate analysis setting. The visual placement results show that some measurable activity can occur to identify phishing that may be connected to a larger, non-malicious, probing social engineering attack. However, without a large scale investment in image recognition it may be difficult to extract visual placement metrics from the ever-changing and Protean phishing email. However, visual similarities may be a factor that is usable to raise a low-level or complementary alert that may be of use to companies in facing a more general preparatory rogue activity that is hard to attack or pinpoint using traditional NLP methods. Visual similarity techniques could allow a more fuzzed approach that could centre on identifying

multiple similar images produced by e.g. PhaaS software-based attack for the purposes of social engineering probing.

5 Conclusions and Future Work

A review of existing literature showed that there is a clear and identified correlation between digital marketing on social media and social engineering and also to phishing and social engineering. We also identified and gave greater significance to PhaaS, the organised data-gathering and analysed delivery of phishing. The experimental setup was able to compare the PhaaS method of phishing delivery to microtargeting, the digital marketing technique that allows a super-refinement of targets based on returned data. This comparison was explored in terms of how digital marketing metrics, as seen in online advertisement data-gathering techniques, could function in phishing analysis. So in effect a wider role of phishing as social engineering might be seen as, in its early stages, more interested in probing likely targets and understanding their behaviour prior to directly attacking them with phishing emails or other social engineering attacks. Using digital marketing metrics enables a measurement of how this works.

We were also able to observe through our experiments how the data gathering and reporting activities in digital marketing microtargeting can be seen to be mirrored in the reporting and analysis behaviour of ethical PhaaS and potentially in *microphishing*. In political advertisement terms the call to action would be the use of the microtargeted information to get the target to vote or perhaps not vote. In the same way basic and even already identified spam is a possible vehicle for data gathering. Both can be considered a kind of social engineering although one is currently most of the time still legitimate.

There is now continuing and urgent ongoing debate as to whether political advertisements should be permitted on Facebook and Google that has arisen since this work was carried out. Twitter has already banned such advertisements. Our work provides further evidence that collated results from phishing analogous to microtargeting are used in PhaaS which strengthens its connection to the negative association of social engineering; it also shows that connections between phishing emails that may be missed using for example an NLP model alone or in combination with URL mismatch are more likely to be picked up by adding visual placement analysis techniques. The comparison of the three techniques and experiment delivers a potential enhancement to phishing identification in that further information could now be gathered and a greater understanding of how phishing works and its wider aims and strategy can be gained.

The next steps for this work would be to explore the results by looking at more real world datasets or of phishing email collected in a corporate environment with the aim of establishing how prevalent connections within phishing emails are, and whether there are discernible patterns of microphishing targets that reveal a development of a social engineering campaign analogous to microtargeting in digital marketing. As PhaaS extends the possibilities for an organised campaign and its associated risk to a company or organisation, a methodology could also then be developed to provide the architecture for the

automation of the processes used in our work. The visual analysis techniques could then be developed as part of an existing tool or as a plugin addon. As understanding of microtargeting, for example in the areas of its use in 2019 United Kingdom general election is only now coming to the attention of legislators, politicians and the general public, further exploration analysing microtargeting data mining goals and techniques especially in the area of granular level analysis metrics. In this context, similarity ratings could then become more pertinent as evidence of larger scale probing by rogue actors.

References

1. Cosentino, G.: The post-truth world order. In: Social Media and the Post-Truth World Order. Springer (2020) 1–31
2. Cadwalladr, C., Graham-Harrison, E.: Revealed: 50 million facebook profiles harvested for cambridge analytica in major data breach. The guardian **17** (2018) 22
3. Gordon, B.R., Zettelmeyer, F., Bhargava, N., Chapsky, D.: A comparison of approaches to advertising measurement: Evidence from big field experiments at facebook. Marketing Science **38**(2) (2019) 193–225
4. Goldman, M., Rao, J.: Experiments as instruments: Heterogeneous position effects in sponsored search auctions. EAI Endorsed Trans. Serious Games **3**(11) (2016) e2
5. Park, G., Taylor, J.M.: Poster: Syntactic element similarity for phishing detection
6. Kandias, M., Stavrou, V., Bozovic, N., Mitrou, L., Gritzalis, D.: Can we trust this user? predicting insider's attitude via youtube usage profiling. In: 2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing, IEEE (2013) 347–354
7. Kintis, P., Miramirkhani, N., Lever, C., Chen, Y., Romero-Gomez, R., Pitropakis, N., Nikiforakis, N., Antonakakis, M.: Hiding in plain sight: A longitudinal study of combosquatting abuse. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, ACM (2017) 569–586
8. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: an update. ACM SIGKDD explorations newsletter **11**(1) (2009) 10–18
9. Silva, M., de Oliveira, L.S., Andreou, A., de Melo, P.O.V., Goga, O., Benevenuto, F.: Facebook ads monitor: An independent auditing system for political ads on facebook. arXiv preprint arXiv:2001.10581 (2020)
10. Egozi, G., Verma, R.: Phishing email detection using robust nlp techniques. In: 2018 IEEE International Conference on Data Mining Workshops (ICDMW), IEEE (2018) 7–12
11. McDowell, M.: Avoiding social engineering and phishing attacks. URL: http://www. us-cert. gov/cas/tips/ST04-014. html (2004)
12. Symantec, I.: Internet security threat report. Broadcom (2019)
13. Krombholz, K., Hobel, H., Huber, M., Weippl, E.: Advanced social engineering attacks. Journal of Information Security and Applications **22** (2015) 113–122
14. Das, A., Baki, S., El Aassal, A., Verma, R., Dunbar, A.: Sok: A comprehensive reexamination of phishing research from the security perspective. IEEE Communications Surveys & Tutorials (2019)

15. Meijdam, K.: Phishing as a service: Designing an ethical way of mimicking targeted phishing attacks to train employees. (2015)
16. Wenyin, L., Huang, G., Xiaoyue, L., Min, Z., Deng, X.: Detection of phishing webpages based on visual similarity. In: Special interest tracks and posters of the 14th international conference on World Wide Web. (2005) 1060–1061
17. Vanderdoncktf, J., Ouedraogo, M.: A comparison of placement strategies for effective visual design. People and Computers (1994) 125
18. An, D.: Advertising visuals in global brands' local websites: a six-country comparison. International Journal of Advertising **26**(3) (2007) 303–332
19. Myers, G.: Words in ads. Edward Arnold London (1994)
20. Stieglitz, S., Dang-Xuan, L.: Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. Journal of management information systems **29**(4) (2013) 217–248
21. Halevi, T., Lewis, J., Memon, N.: Phishing, personality traits and facebook. arXiv preprint arXiv:1301.7643 (2013)
22. Kim, T., Barasz, K., John, L.K.: Why am i seeing this ad? the effect of ad transparency on ad effectiveness. Journal of Consumer Research **45**(5) (2019) 906–932
23. Djamasbi, S., Siegel, M., Skorinko, J., Tullis, T.: Online viewing and aesthetic preferences of generation y and the baby boom generation: Testing user web site experience through eye tracking. International Journal of Electronic Commerce **15**(4) (2011) 121–158
24. Mao, J., Li, P., Li, K., Wei, T., Liang, Z.: Baitalarm: detecting phishing sites using similarity in fundamental visual features. In: 2013 5th International Conference on Intelligent Networking and Collaborative Systems, IEEE (2013) 790–795
25. Medvet, E., Kirda, E., Kruegel, C.: Visual-similarity-based phishing detection. In: Proceedings of the 4th international conference on Security and privacy in communication netowrks. (2008) 1–6
26. Fu, A.Y., Wenyin, L., Deng, X.: Detecting phishing web pages with visual similarity assessment based on earth mover's distance (emd). IEEE transactions on dependable and secure computing **3**(4) (2006) 301–311
27. Fette, I., Sadeh, N., Tomasic, A.: Learning to detect phishing emails. In: Proceedings of the 16th international conference on World Wide Web. (2007) 649–656
28. Rzemieniak, M.: Measuring the effectiveness of online advertising campaigns inthe aspect of e-entrepreneurship. Procedia Computer Science **65** (2015) 980–987
29. Bruce, N.I., Murthi, B., Rao, R.C.: A dynamic model for digital advertising: The effects of creative format, message content, and targeting on engagement. Journal of marketing research **54**(2) (2017) 202–218
30. Johnson, G.A., Lewis, R.A., Nubbemeyer, E.I.: Ghost ads: Improving the economics of measuring online ad effectiveness. Journal of Marketing Research **54**(6) (2017) 867–884
31. You, Q., Luo, J., Jin, H., Yang, J.: Robust image sentiment analysis using progressively trained and domain transferred deep networks. In: Twenty-ninth AAAI conference on artificial intelligence. (2015)
32. Shi, L., Wang, Q., Ma, X., Weng, M., Qiao, H.: Spam email classification using decision tree ensemble. Journal of Computational Information Systems **8**(3) (2012) 949–956
33. Li, W., Meng, W., Tan, Z., Xiang, Y.: Design of multi-view based email classification for iot systems via semi-supervised learning. Journal of Network and Computer Applications **128** (2019) 56–63
34. Schroeder, W.: Testing web sites with eyetracking. Eye for design (1998)