BABELTOWER: How Language Affects Criminal Activity in Stolen Webmail Accounts

Emeric Bernard-Jones, Jeremiah Onaolapo, and Gianluca Stringhini
University College London
emeric.bernard-jones.15@ucl.ac.uk
{j.onaolapo,g.stringhini}@cs.ucl.ac.uk

ABSTRACT

We set out to understand the effects of differing language on the ability of cybercriminals to navigate webmail accounts and locate sensitive information in them. To this end, we configured thirty Gmail honeypot accounts with English, Romanian, and Greek language settings. We populated the accounts with email messages in those languages by subscribing them to selected online newsletters. We also hid email messages about fake bank accounts in fifteen of the accounts to mimic real-world webmail users that sometimes store sensitive information in their accounts. We then leaked credentials to the honey accounts via paste sites on the Surface Web and the Dark Web, and collected data for fifteen days. Our statistical analyses on the data show that cybercriminals are more likely to discover sensitive information (bank account information) in the Greek accounts than the remaining accounts, contrary to the expectation that Greek ought to constitute a barrier to the understanding of non-Greek visitors to the Greek accounts. We also extracted the important words among the emails that cybercriminals accessed (as an approximation of the keywords that they possibly searched for within the honey accounts), and found that financial terms featured among the top words. In summary, we show that language plays a significant role in the ability of cybercriminals to access sensitive information hidden in compromised webmail accounts.

KEYWORDS

Webmail; honeypot; information theft; language

1 INTRODUCTION

Online accounts provide many useful functionalities but also expose users to certain risks. For instance, we send emails, edit documents, and interact with colleagues via online accounts. Consequently, these accounts not only provide these capabilities, but also often become repositories of sensitive information such as passwords and financial information. Webmail accounts are particularly "susceptible" to this, since they mostly store private information by design. This makes them attractive to miscreants that seek to make a fortune from the content of such accounts.

Data breaches and unauthorized account accesses are commonplace nowadays, usually at high financial and reputation costs to victims and online service providers alike [1]. Cybercriminals often compromise online accounts by performing social engineering or phishing attacks on victims [10]. Other ways by which cybercriminals obtain credentials and compromise online accounts include database breaches,[1] information-stealing malware [24], and network attacks.[2]

After obtaining the credentials of online accounts, cybercriminals usually assess the value of such accounts by evaluating the content of the compromised accounts and searching for sensitive information [6]. Depending on the perceived value of the accounts, the miscreants then sell the account credentials on the underground black market [25] or use them privately. In some cases, cybercriminals carry out further attacks against the owners of such accounts, for instance by mounting blackmail attacks against them as seen in the Ashley Madison online dating website scandal.[3] In other cases, the compromised accounts are used to attack other online users, for instance by sending spam messages to the contacts of the account owner [25].

Existing literature on the use of compromised online accounts by cybercriminals is sparse. This is primarily because it is difficult to collect data on compromised accounts without being in control of a large online service. Bursztein et al. studied Gmail accounts that were compromised via phishing attacks, to understand the modes of operation of cybercriminals that gained illegitimate access to the accounts [6]. Similarly, [20] studied the modus operandi of miscreants accessing Gmail accounts leaked through multiple outlets. Lazarov et al. investigated the activity of miscreants on leaked online spreadsheets [13].

Online accounts often allow users to customize their accounts in various ways, for instance through language localization. A question then comes to mind – how do cybercriminals behave when they encounter accounts in an unfamiliar locale or language? How will this affect their activity? To the best of our knowledge, there is limited existing research on this theme. To close this research gap, we studied the impact of differences in account language on the activity of miscreants that connect to compromised Gmail accounts.

Thus, we employed the publicly-available infrastructure[4] and methodology proposed in [20]. We created and instrumented thirty Gmail accounts, and populated them with email messages in three languages, namely English, Greek, and Romanian. We seeded fifteen of the accounts with fake bank details containing keywords that are known to be appealing to cybercriminals. We then leaked credentials

WWW '18 Companion, April 23–27, 2018, Lyon, France

ACM ISBN 978-1-4503-5640-4/18/04.
https://doi.org/10.1145/3184558.3191529

[1] http://krebsonsecurity.com/2014/05/the-target-breach-by-the-numbers/
[2] http://crypto.stanford.edu/ssl-mitm
[3] https://blog.kaspersky.co.uk/cheating-website-hacked/
[4] https://bitbucket.org/gianluca_students/gmail-honeypot

to the accounts through paste sites on the Surface and Dark Webs, following the approach in [20], and recorded activity in the accounts.

We found that cybercriminals are more likely to discover fake bank account details hidden in the Greek accounts than the remaining accounts. This is contrary to our expectation that Greek ought to constitute a barrier to the understanding of non-Greek visitors to the Greek accounts. Previous work shows that cybercriminals usually assess the value of stolen accounts by searching for valuable information in them [6, 20]. Thus, we postulate that the cybercriminals possibly used online language translation tools to translate financial terms to Greek prior to searching the Greek accounts for such keywords. This would also explain the amount of time that they spent accessing the accounts: Greek accounts recorded longer access times than the rest, while English accounts recorded the lowest access times.

Using Natural Language Processing (NLP) techniques, we extracted important words from the emails that cybercriminals accessed (as an approximation of the keywords that they searched for within the honey accounts), and found that financial terms featured among the top words. This is interesting because some of the sensitive words with which we seeded the honey accounts also showed up among those important words. This indicates that the cybercriminals paid particular attention to those sensitive emails.

In summary, we found that language indeed affects the ability of cybercriminals to locate sensitive information in webmail accounts. Our statistical tests show that there is a significant relationship between language and criminal activity in webmail accounts. We also corroborate previous findings that cybercriminals search for financial and other sensitive information in compromised webmail accounts [6, 20].

Contributions. We provide detailed statistical analyses showing that language differentiation affects the ability of cybercriminals to locate sensitive information in a compromised webmail account. To the best of our knowledge, this is the first study that explores the relationship between language and criminal ability.

2 BACKGROUND

In this section, we discuss the categories of cybercrime, webmail accounts, and the relationship between language and criminal ability. We also present our research questions and hypotheses.

2.1 Categories of cybercrime

Broadly speaking, cybercrime is a term used to describe a wide variety of instances within which technology is used or involved in the execution of a criminal act [8]. It embodies a variety of criminal activities (for instance, identity theft and fraud), many of which are among the most rapidly advancing crime types in developed countries [17]. We discuss the three distinct categories of cybercrime, namely cyber-assisted crimes, cyber-dependent crimes, and cyber-enabled crimes [28].

Cyber-assisted crimes. These are terrestrial crimes, such as burglary or theft, which incorporate the use of digital technologies into the execution of a criminal act [15]. An example of this is when a bicycle thief uses a mapping application to plan a route through the area they already intended to steal from. In cyber-assisted crime, the "cyber" element plays a tertiary role in the execution of the crime itself, that is, the crime would likely continue unaffected if the cyber element was removed. This type of cybercrime is not in the scope of this paper.

Cyber-dependent crimes. These are crimes that can be executed without the use of an Internet connection, but use technology as a force multiplier to commit terrestrial crimes within a "cybersphere" [18]. These crimes often take advantage of the global reach of the Internet but do not necessarily represent entirely new crime types. An example is bank fraud which existed before the Internet but has been greatly facilitated by the growth of the Internet.

Cyber-enabled crimes. These represent the "cybercrime archetype." These crimes cannot be committed without the use of an Internet connection or a computer network, for instance, a Distributed Denial of Service (DDoS) attack [7].

Although debate exists regarding small differences and measurement of these crime types [11], our intent in this paper is not to provide detailed insight into crime types or classifications. Instead, we focus on cyber-dependent and cyber-enabled crimes since we study the actions of criminals connect to webmail accounts.

2.2 Gmail accounts

Gmail accounts, like many other webmail accounts, allow users to send and receive text/multimedia messages to one another. However, beyond sending and receiving email messages, Gmail users can embed scripts in their accounts to automatically carry out other activities, for instance, to remind them about important emails that require attention. We leveraged this functionality to instrument the Gmail accounts that we used in our experiments, by configuring the scripts to send us notifications about changes in the accounts.

After authenticating to their accounts, Gmail users can access the email messages that other webmail users sent to them in the *Inbox* folder. While composing email messages in preparation for sending to other webmail users, those email drafts appear in the *Drafts* folder. Similarly, users can access the email messages that they previously sent to others in the *Sent* folder. They can mark emails for later reference by *starring* them. Gmail also provides a text search tool. Finally, users can change the display language of their Gmail interface so that menu items and options on Gmail pages will be displayed in the selected language.

2.3 Language and crime

Research suggests that criminal activities are carried out along familiar patterns of behavior, spatially, by crime type, or by the network of the actors [4]. This suggests that successful criminals rely heavily on a detailed understanding of the processes surrounding the crimes they commit and the areas within which they are committed [22]. Thus, we can safely assume that the ability to understand and interpret social cues, their environment, and the behavior of their victims has a knock-on effect on their ability to commit crime [5].

When attempting to study the behavioral patterns of criminals online, connecting to a webmail account and navigating through it can be considered "routine activity" since these are frequent actions by legitimate users. Changes in the composition, interface, layout, or language of the webmail account can therefore be considered a barrier to the execution of crime in the account – much like a

physical barrier (for instance, a fence) may deter terrestrial crime. This forms the thematic basis for our work.

Certain other aspects of criminal theory developed for terrestrial crime types have shown promise in their ability to be adapted to fit cybercrime types [29]. Even though ideas of locality or geographical nodes from crime pattern theories may need to be replaced with cyber equivalents, certain trends and routine online activities have been successfully attributed to specific online criminals [23].

There is a commonplace "truism" when discussing cybercrime: that cybercrime is somehow unrestricted by the same boundaries of time, space, and culture that may hinder traditional crime types [12]. However, the majority of contentions in previous work were made through logical inferences and assertions. In particular, after exploring previous work in the domains of crime sciences and language, we found very little research exploring the relationship between language and crime. This paper seeks to close that research gap and provide insights into whether the execution of a criminal act is indeed affected by language differences and comprehension, or not. To this end, we define our research question and hypotheses as follows.

Research question. Does language differentiation affect cybercriminal activity?

Hypothesis 0 (H_0). Language differentiation will not have a significant impact on the ability of cybercriminals to locate a sensitive item in a compromised webmail account.

Hypothesis 1 (H_1). Language differentiation will have a significant impact on the ability of cybercriminals to locate a sensitive item in a compromised webmail account.

3 METHODOLOGY

3.1 Creating honey accounts

We created thirty honey accounts on Gmail across three languages, namely English (ten accounts), Romanian (ten accounts), and Greek (ten accounts). We chose those languages for linguistic reasons; English because it is an "international" language, Romanian because it is the only Latin-based Eastern European language, and Greek because it features a unique alphabet. In order to minimize potential biases in our dataset, we configured the fake personas of the honey accounts such that each linguistic group comprised five men and five women with birth dates ranging from 1960 to 2000.

To populate the accounts, we subscribed them to over fifty language-specific newsletters and mailing lists following certain themes. Those themes include fashion, law, and gardening, and were picked according to the genders and dates of birth of fake personas we chose for the accounts. We also changed the display language of each account to match the language of its content. Figure 1 shows the Gmail language configuration option that allows this.

Sensitive emails. In fifteen out of thirty honey accounts, we hid fake online banking information. The idea was to mimic the behavior of webmail users that store sensitive information in their accounts. To achieve this, we created screenshots of fake bank account details and online banking pages (see Figures 2 and 3), and sent emails containing those screenshots to the honey accounts themselves. For each honey account h_G designated to contain sensitive information, we sent the screenshots described earlier from h_G to itself. We used region-specific bank information while seeding the accounts,

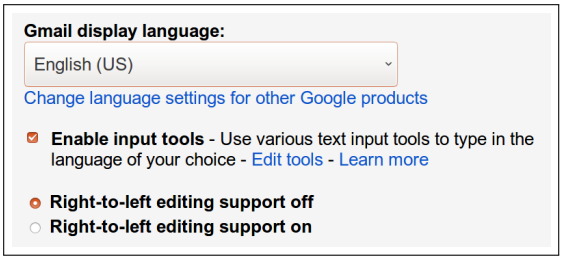

Figure 1: Gmail gives users the option to change the display language of its user interface. In addition to populating the honey accounts with language-specific newsletters, we also changed the display language of each account to match its contents.

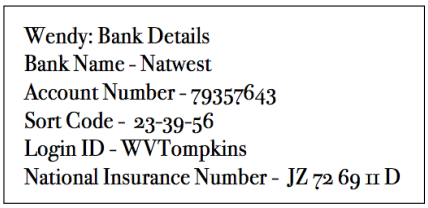

Figure 2: An example of fake banking details that we hid in English honey accounts.

for instance, fake Natwest and Santander information for English accounts, fake ING information for Romanian accounts, and fake Alpha Bank profiles for Greek accounts. We did this to ensure that the banks would be instantly recognizable in the designated countries of the honey accounts. We also included keywords such as *national insurance number*, *sort code*, and *account number* in the sensitive emails. Such keywords have been shown to be attractive to cybercriminals [6, 20]. Finally, we left the remaining fifteen accounts unseeded to enable comparisons between accounts containing sensitive information and those without sensitive information.

3.2 Monitoring honey accounts

To monitor illegitimate activity in the honey accounts, we used the publicly-available infrastructure in [20]. It comprises scripts embedded in the honey accounts, a sinkhole mail server, a notification store to receive activity notifications from honey accounts, a mail client to retrieve email messages from the notification store, and other monitor scripts (see Figure 4).

This infrastructure provides us with information about activity in honey accounts, specifically when emails are opened, sent, or starred. It also provides us with information about draft emails created by visitors to the honey accounts. In addition, we receive "heartbeat" messages daily from each honey account to notify us about accounts that are active. We cease to receive "heartbeat" messages from an account if it has been suspended by Google, or if it was hijacked completely by cybercriminals, that is, if they changed the account's password. Finally, the system provides us with information on accesses to the honey accounts; we receive IP addresses, location information, access times, and other details about visitors interacting with the honey accounts.

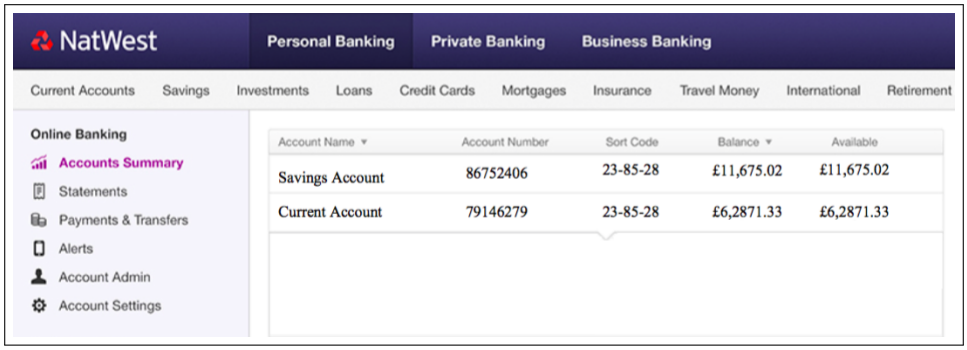

Figure 3: Fake online banking profile hidden in an English honey account.

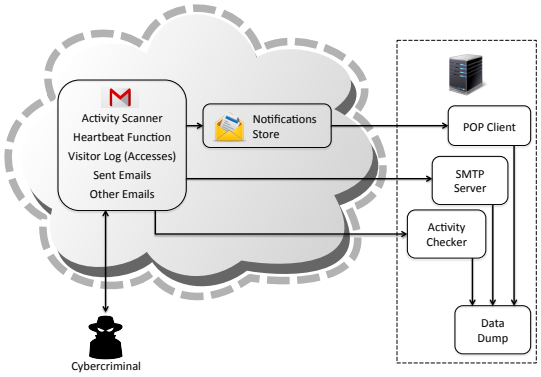

Figure 4: Overview of the honeypot system.

To minimize the risk of abuse, we configured the honey accounts' default send-from addresses to point to a mail server which is part of the monitor infrastructure described earlier. Hence, all emails sent from the honey accounts would be delivered to our mail server and not to the outside world, since our mail server is a sinkhole server (it does not forward emails to the intended destination).

3.3 Leaking honey accounts

After instrumenting the honey accounts, we leaked their credentials via paste sites on the Surface Web and the Dark Web, namely on `Pastebin`, `Insertor`, and `Stronghold` (all credentials were leaked across these outlets evenly). `Insertor` and `Stronghold` are Dark Web paste sites that are accessible only through special software, such as TOR browser. `Pastebin` is accessible via any common web browser, for instance Firefox, Chrome, or Safari. In each leak, we included sets of honey account credentials and captions indicating that the credentials were obtained from hacked accounts. Each set contained between 10 to 15 credentials. We then recorded accesses made to the honey accounts by miscreants.

3.4 Threats to validity

It is important to mention that the monitoring infrastructure we used in this study can only detect if an email was opened; not necessarily if it was read. For the purpose of this study, we assume that opened emails were also read by the person that opened them. In addition, we currently lack a way to determine the exact words that were searched for in the honey accounts by cybercriminals. Instead, we approximate those search terms by evaluating important words in the emails that were opened by cybercriminals. We consider this the main threat to internal validity of this study. To minimize the impact of this threat, we seeded the accounts with email messages containing sensitive content (fake financial information) and hid the emails, such that finding them would require some effort by the cybercriminals. We then focused our analysis on those sensitive emails. In future work, we hope to find a more accurate way to determine search terms in honey accounts. Another threat to internal validity is that many of the honey accounts were hijacked at least once by cybercriminals, that is, the passwords of such accounts were changed. Recall that we are unable to collect access and activity information from a honey account when that happens. However, we were able to recover some of the accounts and continue the experiments. Finally, we leaked account credentials through paste sites only, therefore our results may not necessarily reflect what happens when accounts are compromised via other outlets.

3.5 Ethics

Due to the sensitive nature of our study, we ensured that experiments were carried out in an ethical manner. Since the experiments require releasing account credentials to cybercriminals, there is the risk of abuse. We minimized this risk by configuring the honey accounts to send all outgoing emails to a mail server under our control which does not deliver emails to their intended destinations. Thus, we were able to prevent the accounts from being used for spamming. Also, we seeded the honey accounts with financial information such as bank accounts and online banking information. To avoid harming anyone, we ensured that all financial details loaded in the accounts were fake (we generated them randomly). Finally, since our experiments involve deceiving cybercriminals to engage with fake accounts, we obtained ethics approval from our institution.

4 DATA ANALYSIS

A Gmail account keeps records of each unique access and labels the access with a unique identifier also known as a *cookie*, along with other information such as access time, IP address, and location. We extracted this information from the honey accounts via the honeypot infrastructure described earlier. We also evaluated the

actions corresponding to those accesses (for instance, email opening, sending, starring, or draft creation). In other words, each data unit encapsulates an *access–action* (for instance, *COOKIE–OPEN*).

During a period of fifteen days, we recorded 650 data units across 29 honey accounts, from nineteen countries. 210 of those data units originated from miscreants that carried out a lot of unusual actions in the accounts, for instance, reading all emails in the accounts that they connected to — in other words, those data units are outliers. Also note that some aspects of our analysis require comparing accessed emails to emails that were not accessed. Hence, we excluded those 210 outliers since they were not useful for our analysis.

In this section, we present the results of statistical tests and textual analysis on the data. We establish the relationship between language and cybercriminal ability, and also show some keywords that cybercriminals were interested in.

4.1 Statistical tests

Recall that we set out to understand the relationship between the language of an account and activity/ability of criminals on that account. Hence, to determine whether such a relationship exists or not, we conducted a chi-squared (χ^2) test [21] to assess any possible associations between discrete languages variables (Greek, Romanian, and English), and the ability of the cybercriminal to access sensitive items in the accounts.

The Pearson χ^2 test (see Table 1) shows that there is indeed a significant association between language and the ability of a cybercriminal to locate sensitive items within an email account ($\chi^2(2) = 15.3097$, $p < 0.001$). Due to the risk of inflation, we also generated a Cramer's V statistic [9] to reveal further information about the strength of the association. This confirmed that there was a weak, yet significant association between language and cybercriminal ability ($V = 0.1865$). However, it must be noted that χ^2 tables are relatively unable to provide more substantive information regarding the interactions between the variables or the fit of the implemented model. Thus, we carried out logistic regression to further explore if a substantive relationship exists among the three language variables and criminal activity (see Table 2). We found that the language variables, in combination, significantly affected the ability of a cybercriminal to find a sensitive item ($\chi^2(3) = 19.77$, $p < 0.001$), with the model accurately predicting 81.59% of criminal action. Note that we dropped the Romanian data points from the analysis in name due to collinearity, and henceforth refer to them as *Cons* in subsequent analyses (see Tables 2, 3, and 4).

Further analysis revealed a significant positive relationship between the ability to locate a sensitive item and Greek language sets ($z = 2.52$, $p < 0.01$) with an odds ratio of 2.316176, meaning that accounts established in Greek are more than twice as likely to have a sensitive item accessed than either of the other language sets. English language as a variable was not significant ($z = -0.63$, $p = 0.530$) with an odds ratio of 0.8123249. This means that an account being constructed in English actually lessens the chance of a miscreant accessing a sensitive item in it. We obtained similar results for the Romanian account set which was significant ($z = -6.30$, $p < 0.01$) with an odds ratio of 0.1888889. This indicates that there is a significant negative relationship between emails written in Romanian and the ability of a criminal to locate sensitive items in them.

We further introduced *access duration* as a variable into logistic regression (see Table 3). This is because we observed that the mean of the average access rates for accounts across languages varied: Greek accounts had the highest access times on average while English accounts had the lowest. This might indicate further activity such as text translation to facilitate navigation through the honey accounts. Logistic regression with access duration included among the discrete language variables was significant, accurately predicting 82.05% of criminal activity, and accounting for a small level of variance within the model ($z = 2.17$, $p < 0.01$). The access time variable also had a slight positive effect on the significance levels represented by Greek and English variables, with an English odds ratio of 0.8618208 ($z = 0.45$, $p = 0.656$) and a Greek odds ratio of 2.345972 ($z = 2.53$, $p < 0.01$). However, the Romanian variable suffered a corresponding decrease (odds ratio 0.1589789) while still remaining significant ($z = -6.56$, $p < 0.01$).

To re-affirm our findings, we mean-centred the access duration values before running the model again to ensure that the logistic model was not centring the access duration values at an intercept with a value of 0, but rather a value integral to the rest of the model (see Table 4). Mean-centring had no effect on the fit of the model, other than marginally improving the significance of the Romanian language variable ($z = -6.40$, $p < 0.01$), resulting in the final odds ratio of 0.1084737.

Since these results clearly demonstrate that there is a significant relationship between language and cybercriminal ability, we reject our null hypothesis H_0. In the next section, we present our findings on the sensitive items that cybercriminals searched for within the honey accounts.

4.2 Digging for webmail "gold"

We wanted to understand the themes and words that interest cybercriminals when they connect to compromised webmail accounts. Previous research has shown that one of the first steps of cybercriminals after compromising an online account is to assess its value by perusing its contents [6]. This implies that they run search queries to isolate email messages of interest. However, we did not have access to search terms in the honey accounts since there is currently no API to retrieve such information from honey accounts. To overcome this limitation, we approximated the search terms by analyzing the opened emails and extracting important words from them, relative to all emails in the honey accounts. To achieve this, we relied on Term Frequency–Inverse Document Frequency (TF-IDF) analysis following the method outlined in previous work [20].

For each language set (English, Greek, Romanian), define d_R as the corpus of opened emails in the honey accounts of that language. Similarly, define d_A as the corpus of all emails in the inboxes of those accounts. Note that d_R is a subset of d_A. During preprocessing, we removed all words that had less than five characters from the corpus, and also removed signaling and header information. We obtained $\texttt{tfidf}_R$ and $\texttt{tfidf}_A$ as the resulting vectors of words and their probabilities after performing TF-IDF analysis on the text corpus $[d_R, d_A]$. We further computed the vector $\texttt{tfidf}_R - \texttt{tfidf}_A$. The idea is that words with higher $\texttt{tfidf}_R - \texttt{tfidf}_A$ values have higher importance in the set of emails opened by miscreants, relative to the

Table 1: Chi-squared (χ^2) analysis showing the differences between expected and actual criminal access to a sensitive item.

	Not Sensitive		Sensitive		Total	
Language	Frequency	Expected Frequency	Frequency	Expected Frequency	Frequency	Expected Frequency
English	189	177.9	29	40.1	218	218
Greek	80	93.8	35	21.2	115	115
Romanian	90	87.3	17	19.7	107	107
Total	359	359	81	81	440	440

Table 2: Logistic regression assessing the relationship between language and criminal ability to locate a sensitive item.

Sensitive	Odds Ratio	Std. Err.	z	$P > \|z\|$	95% Confidence Interval	
Lang–Eng	0.8123249	0.2690604	-0.63	0.530	0.4244168	1.554773
Lang–Gre	2.316176	0.7716938	2.52	0.012	1.205513	4.450116
Cons	0.1888889	0.049952	-6.30	0.000	0.1124876	0.3171816

Table 3: Logistic regression including access durations.

Sensitive	Odds Ratio	Std. Err.	z	$P > \|z\|$	95% Confidence Interval	
Lang–Eng	0.8618208	0.2878145	-0.45	0.656	0.4478668	1.658384
Lang–Gre	2.345972	0.7901058	2.53	0.011	1.212396	4.539428
Access	1.008337	0.0038651	2.17	0.030	1.00079	1.015941
Cons	0.1589789	0.0445502	-6.56	0.000	0.091793	0.27534

Table 4: Logistic regression with mean centralised access durations.

Sensitive	Odds Ratio	Std. Err.	z	$P > \|z\|$	95% confidence Interval	
Lang–Eng	0.8618208	0.2878145	-0.45	0.656	0.4478668	1.658384
Lang–Gre	2.345972	0.7901058	2.53	0.011	1.212396	4.539428
C–Access	1.008337	0.0038651	2.17	0.030	1.00079	1.015941
Cons	0.1804734	0.0482639	-6.40	0.000	0.1068506	0.3048244

entire corpus. Thus, such words reveal the themes that miscreants were likely searching for.

Tables 5, 6, and 7 show the results of TF-IDF analysis on English, Greek, and Romanian honey accounts respectively. They show that those who accessed the Greek and Romanian accounts attempted to search for words outside the linguistic confines of the accounts. For instance, the word "posted" appeared to be the most searched word in the Greek and Romanian accounts. The terms searched in the Romanian accounts did not include any financial or banking indicators, whereas the TF-IDF search approximation for the Greek accounts includes words such as τράπεζας (bank) and κωδικός (code). Both words are among the sensitive terms that we used to seed the accounts beforehand. On a related note, financial terms such as "banking" and "investment" appear among the top TF-IDF words in the English accounts (see Table 5). These findings show that cybercriminals indeed searched for financial terms in the honey accounts. This result is further strengthened by the observation that the terms found to be important in the entire email text d_A are not important in the corpus of opened emails d_R (as shown by the low $\texttt{tfidf}_R - \texttt{tfidf}_A$ values, some of which are negative). This is a strong indicator that the opened emails were not selected randomly by cybercriminals; they were opened deliberately after searches were conducted for those terms. This corroborates findings in [6, 20].

5 DISCUSSION

Summary of findings. Contrary to our expectations, our findings show that cybercriminals are more likely to locate sensitive information in the Greek accounts than accounts in the other languages. This is rather intriguing since only two of the accesses originated from Greece or Greek-speaking countries. We recognize that some accesses to the accounts may have been made through proxy servers. However, it is clear that those who visited the accounts were not solely Greek-speaking individuals. These findings run contrary to the ideas espoused in theories of language comprehension and understanding which suggest that individuals should be significantly hindered in their comprehension if they do not understand the language of the object they are interacting with. Thus, we postulate that the cybercriminals possibly used online language translation tools to translate financial terms to Greek prior to searching the Greek accounts for such keywords. This would also explain the amount of time that they spent accessing the accounts: Greek accounts recorded

Table 5: TF-IDF results for the English language variant.

Searched Words	tfidf_R	tfidf_A	$\text{tfidf}_R - \text{tfidf}_A$	Common Words	tfidf_R	tfidf_A	$\text{tfidf}_R - \text{tfidf}_A$
written	0.4371	0.04322	0.3938	unsubscribe	0.109	0.1833	-0.0743
question	0.447	0.0678	0.3796	click	0.0953	0.1671	-0.0718
answer	0.2283	0.0377	0.1907	please	0.0931	0.1597	-0.0666
commission	0.2224	0.0386	0.1838	about	0.0761	0.1279	-0.0518
union	0.2273	0.0565	0.1708	service	0.0394	0.1248	-0.0854
european	0.2508	0.088	0.1628	twitter	0.0257	0.1193	-0.0936
source	0.2267	0.0663	0.1604	trump	0.0399	0.1085	-0.0685
banking	0.1599	0.0394	0.1205	london	0.2158	0.1017	-0.1141
london	0.2158	0.1017	0.1141	contact	0.0465	0.1001	0.0536
investment	0.0548	0.0122	0.0425	health	0.0717	0.0983	-0.026

Table 6: TF-IDF results for the Greek language variant.

Searched Words	tfidf_R	tfidf_A	$\text{tfidf}_R - \text{tfidf}_A$	Common Words	tfidf_R	tfidf_A	$\text{tfidf}_R - \text{tfidf}_A$
posted	0.1233	0.0002	0.1230	alpha	0.0830	0.4820	-0.3990
βιβλίο,	0.1182	0.0003	0.1179	αγόρασέ	0.1358	0.0809	0.0549
ίδρυμα	0.0906	0.0007	0.0899	ekdromi.gr	0.1258	0.0624	0.0634
κωδικός	0.0830	0.0079	0.0751	hotel	0.0704	0.0608	0.0096
τράπεζας	0.0830	0.0001	0.0829	newsletter	0.0453	0.0560	-0.0107
όνομα,	0.0830	0.0006	0.0825	εικόνα	0.0629	0.0483	0.0146
γιάννης	0.0805	0.0014	0.0791	έκδοση	0.0528	0.0470	0.0058
subscribed	0.0780	0.0013	0.0767	διαθέσιμη	0.0478	0.0454	0.0024
states	0.0755	0.0001	0.0754	column	0.0453	0.0392	0.0061
united	0.0755	0.0001	0.0753	outlook	0.0428	0.0322	0.0106

Table 7: TF-IDF results for the Romanian language variant.

Searched Words	tfidf_R	tfidf_A	$\text{tfidf}_R - \text{tfidf}_A$	Common Words	tfidf_R	tfidf_A	$\text{tfidf}_R - \text{tfidf}_A$
posted	0.2307	0.0011	0.2296	click	0.1567	0.2693	-0.1127
charm	0.1481	0.0038	0.1443	multe	0.1253	0.2238	-0.0984
dimensiune	0.1424	0.0024	0.1401	É^{TM}te	0.0541	0.1470	-0.0928
greutate	0.1424	0.0045	0.1379	adresa	0.0741	0.1436	-0.0696
numar	0.1339	0.0093	0.1245	romania	0.0427	0.1161	-0.0734
cutiuta	0.1253	0.0017	0.1237	online	0.0627	0.1118	-0.0491
livreaza	0.1253	0.0019	0.1234	video	0.0968	0.1085	-0.0117
argint	0.1310	0.0103	0.1207	dintre	0.0826	0.1037	-0.0211
material	0.1253	0.0068	0.1185	dezabonare	0.0370	0.0992	-0.0622
produsul	0.1253	0.0089	0.1164	iulie	0.0826	0.0991	-0.0165

longer access times than the rest, while English accounts recorded the lowest.

Miscreants spent more time on average going through the Greek and Romanian accounts. This indicates a number of possibilities. As earlier stated, cybercriminals may spend more time on the accounts to incorporate the use of online translation services to improve their limited understanding of email content. Alternatively, it may be because individuals are more readily able to assess the contents of a webmail account whose language is English, and consequently disregard such an account if it appears to have limited value.

Finally, the ability to search for keywords in the content of an email account may be a key factor in the ability of a criminal to traverse a compromised webmail account, as seen in our TF-IDF evaluation which highlighted words such as "bank" and "code." This suggests that it might be possible for webmail service providers to hamper criminal elements from finding sensitive information in compromised accounts by obfuscating or removing banking and financial keywords.

Limitations. First, we were able to leak the honey accounts through paste sites only. Hence, our results may not reflect what happens to accounts that are compromised via other outlets. Second, our approach relies on TF-IDF to approximate search terms in the honey accounts. As a result, we only have insight into searches whose results were opened by the miscreants. We are unable to assess

searches that did not return results and searches that returned results which the miscreants did not open.

Future work. In the future, we intend to explore the use of compromised online accounts in other scenarios, for instance, targeted attacks. We also intend to study the impact of language differentiation on cybercriminal activity on other platforms, for instance online social networks, cloud storage accounts, and online banking accounts.

6 RELATED WORK

Bursztein et al. [6] studied the use of compromised Gmail accounts in the wild with specific focus on spearphishing as a way by which cybercriminals obtain account credentials. They deployed Gmail honeypots and collected data from them. [20] used a similar honeypot approach to investigate the use of compromised Gmail accounts but explored more outlets, namely paste sites, underground forums, and malware. Other researchers have used honeypot systems to study the use of compromised online accounts as well. Liu et al. [16] placed honey credentials (inside honey files) in P2P shared spaces to study illegitimate accesses. Nikiforakis et al. [19] also studied privacy issues in file hosting systems using honeyfiles. Stringhini et al. [26] deployed honeypot profiles to study social spam. Other studies exploring the misuse of online accounts include [2, 3, 14, 27]. They focus on the abuse of online accounts while we focus on the effect of language differentiation on the ability of cybercriminals that attempt to abuse webmail accounts and steal sensitive information.

7 CONCLUSION

In this paper, we studied the impact of language differentiation on the activity of cybercriminals accessing compromised webmail accounts. We created, deployed, and leaked thirty honey accounts spanning three languages, namely English, Greek, and Romanian. We collected and analyzed data on accesses and activity from the honey accounts for fifteen days. Our tests revealed a significant relationship between language and the ability of a cybercriminal to access a sensitive item (that we seeded the account with). We also found that cybercriminals searched for sensitive financial information in the accounts. Our findings will help the research community to gain deeper insight into the relationship between language and cybercriminal activity, and potentially provide insight into ways to develop effective techniques towards detecting illegitimate activity in online accounts.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their comments. This work was supported by EPSRC grant EP/N008448/1 and a Google Faculty Award. Jeremiah Onaolapo was supported by the Petroleum Technology Development Fund (PTDF) of Nigeria.

REFERENCES

[1] Ross Anderson, Chris Barton, Rainer Böhme, Richard Clayton, Michel JG Van Eeten, Michael Levi, Tyler Moore, and Stefan Savage. 2013. Measuring the Cost of Cybercrime. In *The Economics of Information Security and Privacy*. Springer, 265–300.

[2] Fabricio Benevenuto, Gabriel Magno, Tiago Rodrigues, and Virgilio Almeida. 2010. Detecting Spammers on Twitter. In *Conference on Email and Anti-Spam (CEAS)*.

[3] Yazan Boshmaf, Ildar Muslukhov, Konstantin Beznosov, and Matei Ripeanu. 2011. The socialbot network: when bots socialize for fame and money. In *Annual Computer Security Applications Conference (ACSAC)*.

[4] Patricia Brantingham and Paul Brantingham. 1995. Criminality of place. *European Journal on Criminal Policy and Research* 3, 3 (1995), 5–26.

[5] Patricia L Brantingham and Paul J Brantingham. 1993. Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology* 13, 1 (1993), 3–28.

[6] Elie Bursztein, Borbala Benko, Daniel Margolis, Tadek Pietraszek, Andy Archer, Allan Aquino, Andreas Pitsillidis, and Stefan Savage. 2014. Handcrafted Fraud and Extortion: Manual Account Hijacking in the Wild. In *ACM Internet Measurement Conference (IMC)*.

[7] Armin Büscher and Thorsten Holz. 2012. Tracking DDoS Attacks: Insights into the Business of Disrupting the Web. In *USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET)*.

[8] Jonathan Clough. 2011. Cybercrime. *Commonwealth Law Bulletin* 37, 4 (2011), 671–680.

[9] Harald Cramér. 1946. Mathematical Methods of Statistics. *Princeton: Princeton University Press* (1946).

[10] Rachna Dhamija, J. Doug Tygar, and Marti Hearst. 2006. Why phishing works. In *ACM Conference on Human Factors in Computing Systems (CHI)*.

[11] Steven Furnell, David Emm, and Maria Papadaki. 2015. The challenge of measuring cyber-dependent crimes. *Computer Fraud & Security* 2015, 10 (2015), 5–12.

[12] Peter Grabosky. 2004. The Global Dimension of Cybercrime. *Global Crime* 6, 1 (2004), 146–157.

[13] Martin Lazarov, Jeremiah Onaolapo, and Gianluca Stringhini. 2016. Honey Sheets: What Happens to Leaked Google Spreadsheets?. In *USENIX Workshop on Cyber Security Experimentation and Test (CSET)*.

[14] Kyumin Lee, James Caverlee, and Steve Webb. 2010. The social honeypot project: protecting online communities from spammers. In *World Wide Web Conference (WWW)*.

[15] Michael Levi, Alan Doig, Rajeev Gundur, David Wall, and Matthew Leighton Williams. 2015. The Implications of Economic Cybercrime for Policing. (2015).

[16] Bingshuang Liu, Zhaoyang Liu, Jianyu Zhang, Tao Wei, and Wei Zou. 2012. How many eyes are spying on your shared folders?. In *ACM Workshop on Privacy in the Electronic Society (WPES)*.

[17] Lynn M LoPucki. 2001. Human identification theory and the identity theft problem. *Tex. L. Rev.* 80 (2001), 89.

[18] Mike McGuire and Samantha Dowling. 2013. Cyber crime: A review of the evidence. *Home Office Research Report* 75 (2013).

[19] Nick Nikiforakis, Marco Balduzzi, Steven Van Acker, Wouter Joosen, and Davide Balzarotti. 2011. Exposing the Lack of Privacy in File Hosting Services. In *USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET)*.

[20] Jeremiah Onaolapo, Enrico Mariconti, and Gianluca Stringhini. 2016. What Happens After You Are Pwnd: Understanding the Use of Leaked Webmail Credentials in the Wild. In *ACM SIGCOMM Internet Measurement Conference*. Association for Computing Machinery (ACM).

[21] Karl Pearson. 1900. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5* 50, 302 (1900), 157–175. https://doi.org/10.1080/14786440009463897 arXiv:http://dx.doi.org/10.1080/14786440009463897

[22] Jesenia M Pizarro, Nicholas Corsaro, and Sung-suk Violet Yu. 2007. Journey to Crime and Victimization: An Application of Routine Activities Theory and Environmental Criminology to Homicide. *Victims and Offenders* 2, 4 (2007), 375–394.

[23] Martijn Spitters, Femke Klaver, Gijs Koot, and Mark van Staalduinen. 2015. Authorship Analysis on Dark Marketplace Forums. In *Intelligence and Security Informatics Conference (EISIC), 2015 European*. IEEE, 1–8.

[24] Brett Stone-Gross, Marco Cova, Lorenzo Cavallaro, Bob Gilbert, Martin Szydlowski, Richard Kemmerer, Christopher Kruegel, and Giovanni Vigna. 2009. Your Botnet is My Botnet: Analysis of a Botnet Takeover. In *ACM Conference on Computer and Communications Security (CCS)*.

[25] Brett Stone-Gross, Thorsten Holz, Gianluca Stringhini, and Giovanni Vigna. 2011. The underground economy of spam: A botmaster's perspective of coordinating large-scale spam campaigns. In *USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET)*.

[26] Gianluca Stringhini, Christopher Kruegel, and Giovanni Vigna. 2010. Detecting Spammers on Social Networks. In *Annual Computer Security Applications Conference (ACSAC)*.

[27] Kurt Thomas, Chris Grier, Dawn Song, and Vern Paxson. 2011. Suspended accounts in retrospect: an analysis of Twitter spam. In *ACM Internet Measurement Conference (IMC)*.

[28] David S Wall. 2003. Mapping Out Cybercrimes in a Cyberspatial Surveillant Assemblage. (2003), 112–36.

[29] Majid Yar. 2005. The Novelty of 'Cybercrime' An Assessment in Light of Routine Activity Theory. *European Journal of Criminology* 2, 4 (2005), 407–427.