Assessing the Role of Online Banking Characteristics in the Target Selection of the Banking Malware

Samuel Natalius

Faculty of Technology, Policy and Management, Delft University of Technology,
Jaffalaan 5, 2628BX Delft, The Netherlands

Abstract. Understanding target selection is a step before making a suitable proactive measure to address the complex issue of banking malware in online banking landscape. Despite several previous studies, gaps in the research of target selection are still present like the lack of attention to the non-targeted entities, the presence of other potential factors and the change in the landscape itself. Seeking to address the gaps, this research is conducted to find out **what characteristics related to online banking services can affect the likelihood of the malware attack** to them. The research starts with literature review to identify characteristics which can potentially explain the target selection, in accordance to aspects of Routine Activity Theory (RAT). Next, data about malware attack and the list of banks as well as several external data like language and authentication factor of online banking were collected and processed for quantitative analysis. Several metrics to approach the actual attack count were proposed and other metrics were extracted from the data.
Some interesting findings were captured, like, within the period February 2014 - November 2017, from 5,039 banks in the EU, 1,188 banks were without any online banking services and from 3,851 banks with an online banking service, 1,802 banks were found targeted and 2,049 not targeted. Some malware variants were also seen performing targeted attacks. Meanwhile, it is found from explanatory analysis that some characteristics maintain their significance in explaining the likelihood of attack, like the presence of English and two-factor authentication. Services offering English language were seen to be more attacked. Contrarily, services which implemented 2-factor authentication were found to receive less attacks, although more entities with such authentication were targeted. Meanwhile, some other variables were getting less significant when more controlling factors are taken into account, indicating that some variables were relatively more or less important than others. Future work is needed in order to enhance the model so that more plausible conclusion can be obtained, such as improving and adding more data as well as including more factors, especially those that are financial and market related.

Keywords: target selection, banking malware, online banking, characteristics, cyber security

1 Introduction

Information technology (IT) has played a huge role in banking sector since the 1990s when banks started to utilise the Internet to access their computer systems [12]. Since then, IT has enabled banks to reach out to more customers, perform transactions with less cost and effort, and transform their products and services [20, 44, 8]. Digitalisation not only affects the technical aspect, but also makes social impact. For example, people are now becoming accustomed to accessing their account and performing financial transaction via their PC, laptop, or even mobile phone, rather than physically going to the bank office. Moreover, the view on IT in banking sector is currently shifting, from treating it as a tool to seeing it as a business opportunity.

While many benefits can be reaped from the penetration of IT in banking sector, it also raises a new type of security issues. Digital-related security issues are now prevalent and should be taken seriously as physical- and human-related security issues. Cyber criminals are interested in the banking sector since it has valuable assets and is heavily dependent on IT [31]. The fact that cybercrime accounts for almost USD 114 billion of financial losses in the banking sector globally [33] should see cyberattacks in the banking sector a priority for research, especially the malware-related attacks, as financial malware will potentially infect a large number of entities and causing them to suffer huge losses. [39].

Analysing only the technical aspect of the malware is not enough to explain the malware attacks in the banking sector. This is because the problem does not only lie in the technical layer but also socio-technical and governance layer [3]. There are multiple actors with their own interests, values, and goals in the landscape, making this problem complex. The financial sector is also constantly evolving [14] which could influence the related cyberspace. The malware itself, in the context of malware attack, is only a tool for performing the attacks [4, 17, 39, 45]. Its attack decision is based on the strategy and behaviour of the attackers.

Malware attacks may be defended in either a reactive or a proactive way. While a reactive approach tries to detect and stop an attack as it occurs, a proactive approach seeks to understand risks before an event happens [46]. Lu et al. [27] argued that the proactive approach can significantly reduce the negative impact of malware attacks more than the reactive defence since it makes the system more dynamic and harder to predict. Assessing the behaviour of attackers in general can be a part of the proactive approach. Asghari et al. [2] argued that incentives shape the behaviour of actors in the cyberspace including attackers, for example, attackers are likely to select targets and attack strategies based on their expected financial/political benefits and risks. Identifying the incentives may not be straightforward, but it could be approached by understanding the target selection. Target selection was once defined as attack choices by financial malware criminals as to which financial institution to attack, at which point of time and for how many weeks [9, 43]. In the case of online banking, the attacks were becoming more target-specific, requiring a conscious selection process on the criminals side [39, 36, 42]. Therefore, analysing the target selection can enable banking players to gain a clear understanding of the threat landscape, which is essential to develop cyber risk management in the end [43].

Many previous studies were carried out to understand the target selection within the context of online banking entities [38, 39, 43, 9, 40, 17]. Many of these studies focused on assessing the general characteristics, for example the size and type of banks that were targeted. However, until now, not much was known about those that were not targeted. Besides, other characteristics other than those that have been researched may also influence the target selection, for example the language and the authentication factor. Lastly, the previous studies limited their scope to one type of financial malware family, i.e. Zeus, while the similar research for other malware families is still lacking. Seeking to bridge these knowledge gaps, this research aims to understand factors within banks, either general or security measure related, that influence attackers to target or not target them. For this purpose, the following main research question (MRQ) is made: *What characteristics related to online banking services can affect the likelihood of the malware attack?*.

Due to the limited data availability, the research will consider only banks registered in the selected European Union (EU) countries. The attack data used for this research came from the attack database of Fox IT, a leading cyber security company in the Netherlands. The result of this research would be very useful for banks, as the potential victims, to revaluate their position in the malware threat landscape and make decisions around their proactive measures.

2 Methodology

In order to approach the main research question, a quantitative research will be conducted. A quantitative research uses data and statistical procedures to examine the relationship between variables [15], in this case was the relationship between characteristics of online banking services and the likelihood of malware attack. Characteristics which will be assessed in this research will be identified and selected in the literature study phase, which also provides the theoretical basis of this research.

Literature study

Literature study will be performed to identify characteristics of online banking services that could possibly influence the target selection of financial malware. For this purpose, the first research question (RQ1) is formulated as: *What characteristics of banks or their online banking services can potentially explain the likelihood of them to be targeted?*. Information, theories and frameworks to answer RQ1 will be obtained by collecting scientific and professional literature.

Quantitative analysis

Quantitative analysis will begin with collecting data and defining possible metrics and variables that are quantifiable and relevant to the selected characteristics. As these metrics and variables are important for the overall result of the analysis, the second research question (RQ2) is formulated for them as: *What quantifiable metrics and variables relevant to the characteristics can be defined for the analysis?*.

Using the collected data and defined metrics and variables, quantitative analysis can be done. The quantitative analysis will be descriptive and explanatory analysis using the available data to understand the landscape of online banking services in the context of malware attack and describe the target selection on banks' online banking and to what extend the selected characteristics could influence it. These lead to the third research question (RQ3): *How does the target landscape of online banks look like?*, and the fourth research question (RQ4): *Which of the selected characteristics could explain the extent the online banking were targeted by the malware.*

Qualitative analysis

Qualitative study will be performed in the last phase by interviewing experts in order to get their critical review and interpretation of the quantitative result.

3 Theoretical Background

3.1 The Concept of Cyberspace and Cybercrime

Cyberspace is defined by Singer & Friedman [37] as "the realm of computer networks (and the clients behind them) in which information is stored, shared, and communicated online". While technology is the enabler of the cyberspace, it cannot solely explain the complexity of the cyberspace that is known today. Human and society factors should also be considered and so should socio-technical and governance layer [3].

Cybercrime is a type of activities, which according to van den Berg et. al. [3], is enabled in the socio-technical layer of the cyberspace. Cybercrime can be defined as any activity that is illegal or considered illicit by certain parties, committed or facilitated by one or several entities using the capabilities of the electronic, computer and network technology, which may cause disadvantages to other parties or the society [28, 30, 33, 47, 41, 5, 7, 16, 9].

According to the Global Economic Crime and Fraud Survey [32], cybercrime accounts for 41 percent of the reported frauds in financial services, putting it on the second rank of the most reported frauds in the sector after consumer fraud. The same report also argued that the fraction of cybercrime in the financial services is also the highest among other sectors. Therefore, cybercrime is one of the most significant threats to the business continuation of financial firms and should be treated seriously by banks.

The economic cost of cybercrime can be grouped into four categories [1, 24]:

- Criminal revenue, which is the gain received by the criminal from a crime.
- Direct loss, which is loss, damage, or other suffering experienced by the victim because of a cybercrime. For example, asset loss or compensation given to the victims.
- Indirect loss, which is loss or opportunity cost felt by the victim due to the existence of a cybercrime, whether it is successful or not. For example, reputational cost and loss of trust by the customers.

- Defence cost, which can be further grouped into:
  - direct defence cost for the development, deployment and maintenance of its countermeasures (for example, the annual cost of implementing antivirus), and
  - indirect defence costs caused by the negative consequences and opportunity costs caused by the implemented measures (for example, the cost of employee training and change management).

3.2 Online Banking Services: Mechanism and Security

Nearly all banks offer online financial services and grant their clients access to such services [38]. Meanwhile, the use of mobile banking begins to rise following the growth of smartphones [29].

Nowadays, online banking services generally comprises these main components [38, 29, 43, 18]:

1. The client side, consisting of services and applications that enable customers to access their bank accounts and perform financial transactions,
2. The network infrastructure, which is required in order to connect the client-side services and applications to the bank-side, and
3. The bank side, more specifically, its web server and backend system which serves and processes requests coming from the client side. It is linked to the backend system which holds the main centralized database storing all the crucial data related to the bank and its customers.

As online banking are prone to cyberattacks, banks have implemented some security measures to their online banking services. Authentication is a measure performed so that ”one agent should become sure of the identity of the other” [26]. The authentication methods may range from the simple static username and password to two- or multi-factor authentications combining two or multiple factors such as something the user knows (knowledge), something the user physically has (possession) and something the user physically is or does (biometrics) [12, 23, 38]. So far, there are two forms of authentication for authorising financial transactions used in online banking services [21, 22]:

- entity authentication aiming to prove the identity of an online banking user, and
- transaction authentication aiming to prove that a certain transaction is intentionally performed and authorised by the right user.

3.3 Financial Malware

Many variants of financial malware today share the same capability, pattern of attack, and distribution. Many of them were derived from the previous variants, especially ZeuS which is said as ”the king of bank fraud trojan viruses” [19]. The characteristics shared by many financial malware variants include:

- The malware was coded in such way that it can be customised further by the attackers. Therefore, the attackers do not always have to build the malware from scratch and the malware kit can be easily traded for money.
- The presence of dynamic configuration file enables the malware to be easily configured and updated.
- It spreads via multiple ways, e.g. spam email, malicious website, external storage media, etc., and infected client machines when the victims execute a file containing the malware.
- A computer infected by the malware will become a bot, and it will initiate the communication with the C&C server and ask for a pre-generated dynamic configuration file [38].
- The malware mainly uses code injection techniques to manipulate victims so that they send their credential information to the drop server for the attackers to collect. It also steals other information available in the infected machines.

- The presence of money mules who offer their bank accounts as the transfer destination of stolen money. The purpose of money mules is to hide the trace of the attackers. A percentage of the stolen money will be reserved for the money mules as the service fee.

Some malware variants incorporate new features that make them more powerful than their predecessors. For example, Citadel has capability to capture video from its victim's camera and redirect DNS. Another example, Qadars, has capability to bypass mobile-based two-factor authentication [10].

Some malware variants may target a specific entity or be present in certain countries only. Qadars, for example, is known to target only several countries in Europe like the Netherlands, France and Italy [6].

4 Identification and Selection of Characteristics

This paper incorporates Routine Activity Theory (RAT) from Criminology field to identify characteristics of online banking services which potentially explain the target selection of banking malware. RAT stated that there are 3 (three) minimal elements for conducting a successful violation: a motivated offender, a suitable target for the offender and absence of guardians [13]. They also argued that the suitability of the target relates to VIVA, the value, inertia, visibility and access, of the target. However, multiple sources argued that not all parts of RAT are usable in explaining cybercrimes [47, 25]. Regarding target selection, the suitability of the target is the relevant element to look into. Among the aspects in that element (VIVA), only value, visibility and accessibility are feasible to focus on as inertia is difficult to transpose into the virtual environment within the cybercrime topic [47].

Characteristics are extracted from available literature and then mapped into the relevant aspects of RAT's suitability target (Value, Visibility and Accessibility). This research identified bank size and country location as characteristics belonging to value. Brand popularity, domain name visibility, banks' attack record, website domain popularity, ownership of the bank and language of the online banking belong to visibility, while bank authentication method, broadband quality, users' online awareness, rate of use of firewall/antivirus products and ease in securely performing criminal actions as characteristics belong to accessibility.

Due to the time limitation of the research as well as the added value this research could bring to the field, only some of the above characteristics are selected for the next phases of the analysis. The selection criteria were the feasibility of data collection (availability of data and the feasibility in collecting such data in a limited period) and whether previous researchers have analysed such characteristics. Based on these criteria, two characteristics were selected: the language offered by the banks' online banking services and the authentication method of the banks' online banking services. Other characteristics, like the country and the popularity of website domain, will also be considered in this thesis as control variables to the model.

5 Data

This section will describe the data collected and used for the analysis and the building up the statistical model.

5.1 List of EU Banks

The list of EU banks is obtained from the list of financial institutions provided by the European Central Banks (ECB) open data. The data used for the research was obtained on March 21st, 2018.

The data lists all financial institutions in Europe which is under the ECB's area of control. The dataset contains information about the bank name, the country in which it is registered, its address and its parent bank / headquarter (HQ), if applicable. There were initially 7166 financial institutions on the list, divided into several categories such as central banks, credit institutions, money market funds and other institutions. The research focuses on credit institutions as banks belong to this category, reducing the number of entities into 6234.

Furthermore, the data is filtered to the banks having the HQ within Europe, as there are also several banks on the list which are foreign branches and have the HQs and domains outside of Europe, therefore cannot be treated as European banks. The entities with the HQ outside of Europe are then removed, resulting in 6096 entities remaining on the list.

Next, clustering is performed to banks from the same country. The banks with the same type, management and e-banking domain are seen as one bank for the purpose of this research. This results in a list of 5039 bank entities which are ready to be merged with the attack target data.

5.2 Attack Target Data

The attack target data was obtained from the database provided by Fox IT, a leading cybersecurity company in the Netherlands, for Delft University of Technology. The data came from malware configuration files that have been collected by Fox IT over time by creating a system that resembled a bot to emulate Zeus malware.

The first step in order to get the data was to analyse the database and create a right query to extract the data out of the database. The extracted data was then stored in a CSV file. The dataset contains 14,198,778 attack entries from the period February 2014 to November 2017. Each entry represents a single attack target URL extracted from a configuration file of a variant of malware at a time. Not all entries belong to bank entities. Entries identified by Fox IT have information about the bank entity (domain, name and country) targeted by the malware. By filtering the data based on the country information, the entries belonging to EU banks could be obtained. 3,154,112 entries were extracted from this.

The bank entity information in the attack entry, especially the domain, is useful for the next steps of analysis because such information enables the data to be mapped into the list of banks. However, the identification process done by Fox IT was not perfect in identifying all entries belonging to banks. There were still many entries with bank-related URLs that were not identified, meaning that their information was not present. In order to enhance this, a manual observation of online banking domains of the EU banks was performed and a domain extraction algorithm was made and applied to the non-identified attack entries. The extracted domains were matched with the domains from the manual observation in order to detect additional entries which actually belong to EU banks but not identified in the previous mechanism. 138,912 additional entries were found and added to the entries from the previous approach, making the total of 3,293,024 entries selected for the next step.

Next, the bank information, especially the list of domains were extracted from the attack target data and mapped to the list of banks. The mapping was done manually. In short, this process checked which of the EU banks a domain from the list belongs to. The process is not straightforward in practice, due to complications such as multiple IDs that may refer to the same entity name and multiple banks that may be using the same entity name. In order to handle this, several intermediary tables need to be generated.

At the end of the process, each bank on the list which has domain(s) in the attack target data has a new field with its domain(s) as the value, enabling the list to be processed together with the attack target data.

5.3 The language of the online banking

Information about the site language of banks' online banking service was found by exploring the online banking service of each bank on the list manually. The exploration was done by simply accessing the Internet banking website to look for any language option offered there. For example, if a bank's online banking site is displayed in Dutch but it has an option to change the language into English, the online banking is perceived to be offering two languages: Dutch and English.

5.4 The authentication factor of the online banking

Similar to the language, information about the authentication factor of banks' online banking service was found by exploring the online banking service of each bank on the list manually. This research focuses on the authentication factor of the entity authentication (user login). The observation is done by checking the display and structure of the login form, the help page or Frequently Asked Questions (FAQ) page which describes the login procedure and/or the documents about the online banking services and/or the login instructions.

5.5 The domain popularity ranking

The ranking was obtained from Cisco Umbrella Popularity List, which reflects the relative Internet activity of a domain regardless of the invocation protocols and applications [11]. The list is updated daily. For this thesis, the list from May 30th, 2018 is used.

6 Metric and Variable Definition

6.1 Possible metrics extracted from the datasets

Several metrics can be extracted from the collected data as follows:

Raw attack count

A raw attack count refers to the number of attack URLs referring to a domain of an entity over a certain period. This is the easiest way to count the attack. However, Tajalizadehkhoob 2013 has argued that this basic definition may not represent the valid number of times a domain is actually attacked as the actual attack count may have actually been inflated. Therefore, other metrics which could approach the actual attack count better are needed.

Number of "week-interval" attack count

This definition is based on a rationale that a new attack appearing just after the similar attack (having a similar configuration file) is unlikely. The argumentation is, that kind of entries are only meant to update the previous configuration file. Aligned with the argument that it may take once per two days for a bot master to update a configuration file, or more days if the bot master is less active or if the bot is smaller or more stable [38], this definition assumes that new attacks that have very similar characteristics as one in the past may occur only 7 days (a week) or more after the previous similar attack. Hence, this definition tries to eliminate the entries that could be associated with updates to the previous attack configuration. The purpose of this metric is to eliminate entries which are perceived to be non-actual attacks so that it can estimate the actual attack count more accurately.

Suppose the assumption about attack interval is right, it can provide significantly more accurate count of the attack targeted to a domain since it will ignore any entry which is only meant to update the previous attack. On the other hand, the actual attack entries which are not aligned with the assumption would then be ignored.

In addition, it is argued before that the malware family heavily influences the assumption about the time interval. The current assumption, 7 days, was extracted from a study which uses ZeuS data from a period before 2013. This assumption may not be relevant any more to the current data and to other malware families. In order to make certain that the attack entry is actually the same as the previous one, one should check whether the corresponding C&C server is sending the related configuration file to the same bot at the same time. If not, this assumption cannot be applied.

Count of unique attack ID

This metric is created based on further investigation on the way malware attack data is stored, that is, a unique attack ID are present in the dataset which refers to a unique inject code and hence may indicate different attack attempts performed by criminals. This metric counts the number of unique attack ID which corresponds to a domain. This count may approach the actual attack count better than raw count because it only considers multiple attack entries which have the same inject code, represented by the same attack ID, as one entry. Multiple entries with the same inject code may happen due to updates of configuration files and they are the cause of over-counting problem in the raw attack count metric.

There is a huge advantage of using this metric to approach the actual attack count. Should the argument that different inject codes indicate different attack attempts is true, this metric can significantly eliminate the over-counting problem that the raw attack count metric encounters. Therefore, one could highly ensure that the count from this metric represents the actual number of attack targetting a bank. On the other hand, different attack attempts that share the same inject code may still be present although the chance is very small. However, this metric seems more acceptable than other attack count metrics because it is based on an assumption backed up by the technical properties of the dataset.

Number of weeks a bank is under threat

For this research, this metric is defined as the number of weeks when any attack entries are present, that is, the configuration file received by a botnet. This metric can be obtained by looking at the time dimension of the data. The attack entries are categorized into week groups. If there is an entry corresponding to an entity in a certain week, it means that the entity is attacked within that week and the value of the metric increases for that entity.

Number of unique URLs corresponding to the bank

The number of unique attack URLs can also become an indicator of the target rate of a bank. The rationale is it is unlikely for two completely different attackers to use exactly the same URL to target an online banking service. If they actually want to target an online banking service, they will target at least different parts of the service or use different ways to express the attack destination (for example, using regular expression), resulting in different URLs. Therefore, more unique URLs associated with an online banking could indicate higher actual target rate.

Number of different malware variants targeting an online banking service

As the attack target dataset contains information about the threat or the malware variant that corresponds to an attack entry, a metric involving the threat information can be extracted from the dataset.

There are 29 unique threats captured in the dataset. Each threat is used differently, targeting different entities in different countries. Some are used persistently overtime and some are used only for a specific attack. It is also possible that multiple threats target the same online banking service during the period. Like unique URLs, more threats targeting an online banking service could indicate that the service is more attractive to attackers since it implies that different malware users were looking at the same service as a prospective target.

The presence of a language in an online banking entity

This metric is extracted from the language of the online banking data. Each metric is created for each language present in the data and it will have true value if a language is present in an online banking service, false otherwise.

Number of languages offered by an online banking entity

As some of online banking sites offer more than one language, a metric able to capture the number of languages offered by an online banking service can be created.

The presence of a certain authentication factor in an online banking site

The metric can be extracted from the authentication factor data. Similar to the presence of language, the metric will have true value if an authentication factor is present in an online banking, otherwise false. Using this metric, the relationship between the presence of a particular authentication factor and the target rate can be assessed.

The ranking of an online banking domain

The metric can be extracted from the domain popularity data. Simply put, the metric will contain the ranking of the online banking domain as on the list.

6.2 Variables for the analysis

The variables are based on relevant metrics which were extracted earlier in this thesis. For the later regression analysis, both dependent variable(s) and independent variable(s)/predictor(s) are determined. They are summarized in table 1.

Table 1: Dependent and Independent Variables

Variable	Type	Description
Attack Count	Dependent Variable	Will use three versions of attack count: Raw attack count, 7-day interval attack count and unique Attack ID count.
The presence of a language A metric is created for each language: English, German, French, Dutch, Italian, Spanish, Portugese, Greek, Czech, Slovak, Slovenian, Polish, Hungarian, Romanian, Bulgarian, Danish, Swedish, Finnish, Latvian, Estonian, and Lithuanian.	Independent Variable	True if the language is offered in the online banking service, false if not
Language count	Independent Variable	A count number of languages offerd by an online banking
The presence of an authentication factor A metric is created for each authentication factor: 1-factor authentication (1FA) and 2-factor authentication (2FA).	Independent Variable	True if the authentication factor is offered in the online banking, false if not

Variable	Type	Description
Country	Independent Variable (Control)	The name of the country where it is located. The list comprises all countries in the European Union (EU)
Threat	Independent Variable (Control)	Threat / malware variants that are seen targeting the online banking
Unique URL count	Independent Variable (Control)	The number of unique attack URLs corresponding to the online banking
Popularity score of the online banking domain	Independent Variable (Control)	Inverse of the popularity ranking metrics, so that the first rank gets the highest score (1,000,000). For data points which have a missing value, "zero-coding" data imputation technique is applied

7 Analysis

7.1 Descriptive Analysis

Descriptive analysis was performed in order to describe and obtain interesting features of the observed data, mainly the attack target data.

There are 5,039 banks present throughout European Union (EU). The banks are spread across 28 EU countries[1]. The attack data are available from the period between February 2014 and November 2017. They consist of records associated with one of 30 unique threats / malware variants. From the merging between the attack target dataset and the list of EU banks, it is found that, out of 5,039 banks, 1,188 banks were without any online banking services and from 3,851 banks with an online banking service, 1,802 banks were found targeted and 2,049 not targeted. Banks without any online banking services were either corporate banks, private banks or small, local cooperative banks, which seem to only provide traditional channels to their customers like physical address, phone number or email address.

The first visualization to describe the data is the distribution of attack overtime per country within the scope period (February 2014 - November 2017). The visualization is based on the attack target data. Next, a similar visualization was also made, but instead of per country, the attack overtime is categorized per threat/malware variants. They are shown in figure 1.

It is seen from the plot that Germany had in general the highest attack count among other countries in EU region. This makes sense since Germany has much more banks than other countries. Second, the notable rise in attack activity can be seen in two periods of time: in April - August 2016 with the peak in May 2016, and September 2016 - January 2017 with the peak in December 2016. The rise in a malware activity did not always lead to the rise of attack count received by a country. The characteristic of the attack target URL itself also played a factor. The rise in attack activity in Germany as illustrated in the former graph was caused by the presence of generic URLs inserted in ZeuS-OpenSSL injection code. With such generic URLs, an injection code has a capability to target multiple banks. If the effect of generic URLs are removed, it is seen that three malware variants stood out during the period of analysis: Dyre in 2015, Citadel in 2016 and TheTrick in 2017 (see the second plot of figure 1). However, if the metric attack ID count is used, Citadel no longer dominated the year 2016 (see the fourth plot of figure 1). This may indicate that there were lots of updates of Citadel configuration files at that period which apparently did not introduce a new attack. Dyre and TheTrick, however, still dominate the year 2015 and 2017, consecutively. This suggests that a lot of new attacks were introduced using such malware variants in those periods.

[1] United Kingdom is still considered as part of EU

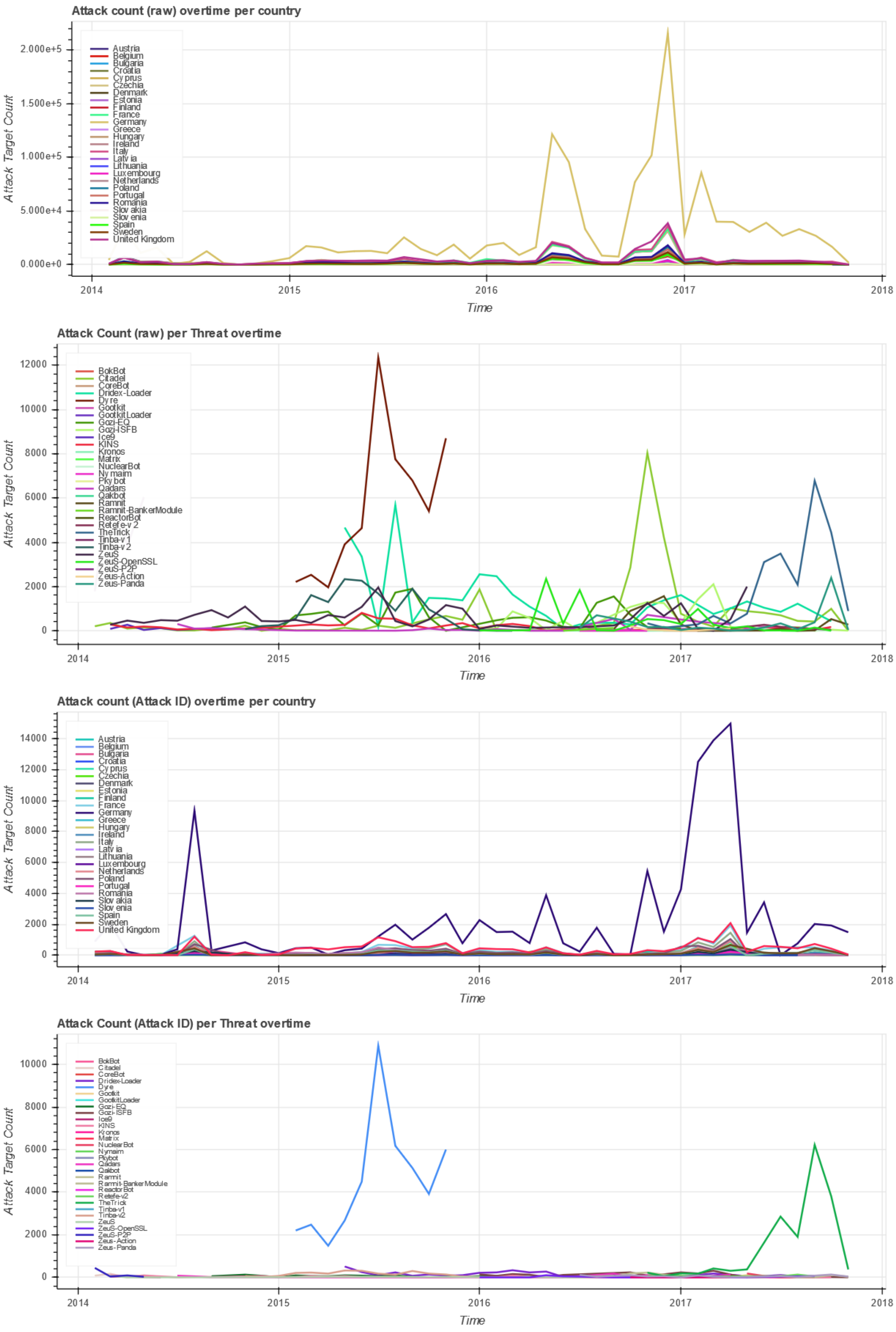

Fig. 1: From top to bottom: Raw attack count overtime per country, raw attack count overtime per threat, attack ID count overtime per country, attack ID count overtime per threat.

Next, a description showing the relationship between countries and malware variants was made. The plot was made in order to see how broad the area targeted by a malware variant is within the EU. Firstly, a plot is made to see the number of countries targeted by each malware that is present in the data. Secondly, a 2-dimensional heatmap relating the threat and the country is presented, where the color in a cell of the heatmap indicates the attack count for a particular threat and country. Both plots are shown in Figure 2.

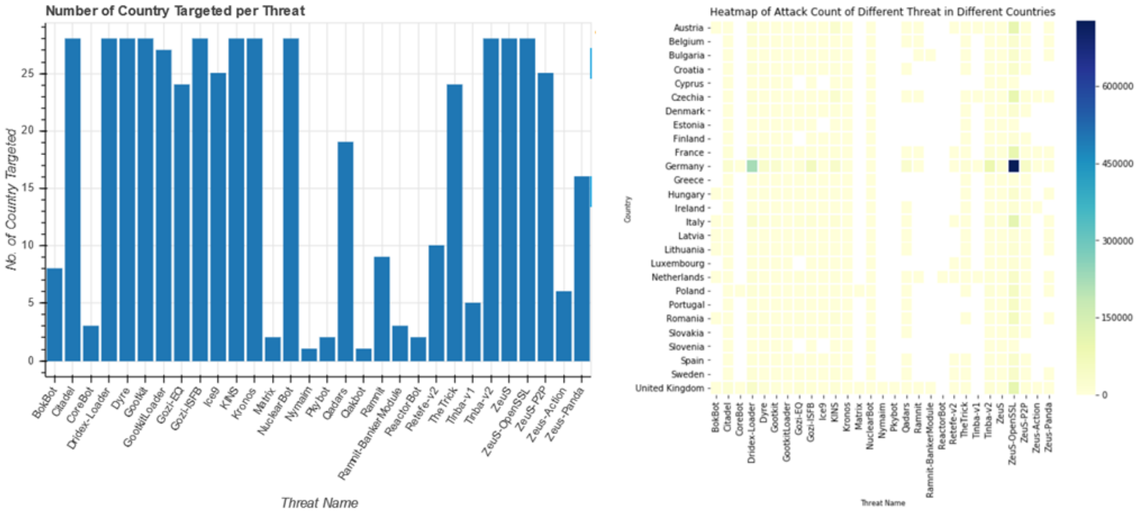

Fig. 2: Plot on number of country targeted per threat and heatmap showing the attack count of a threat in a country, based on the attack target data

It is seen from the plot that not all malware targeted the same number of country. There are even some malware variants that target only one country according to the data, namely Nymaim, Pkybot and ReactorBot. From the heatmap, it is shown that both Nymaim and PkyBot targeted only the United Kingdom (UK), while ReactorBot only targeted the Netherlands. Further analysis shows that Nymaim only targeted HSBC in the UK, while Pkybot targeted several entities in the UK i.e. The Cooperative Bank, Llyods Bank, Barclays United Kingdom, Halifax and Santander UK. ReactorBot itself only targeted ING Netherlands and Rabobank. This indicates that there is a relatively limited number of countries targeted by these threats and there is a tendency that these threats are specifically used to perform a targeted attack. This explanation is particularly acceptable for Pkybot and ReactorBot. Pkybot has been identified as a targeted threat focuses on banks in UK, Spain and Greece in 2015 although the web inject for Greece was removed during Greece financial crisis [34]. ReactorBot, on the other hand, was first discovered in 2015 and initially targeted banks and financial institutions in Netherlands and Germany [35].

The following description is the presence of attack per threat during the scope period. It shows when in the period the attack by a malware occured. It was created by looking at the presence of attack made by a certain threat on a daily basis. Figure 3 depicts the period of attack for each threat in the attack target dataset.

It is seen from the plot that, while many of the threats are persistent in doing attacks like ZeuS, KINS and Citadel, some of the malware, like ReactorBot and Pkybot, perform attack at a very specific time. This visualization makes these two threats interesting because, in the previous analysis, they were also found to only target a specific country. Highly is the chance that they were used for specific, targeted attacks.

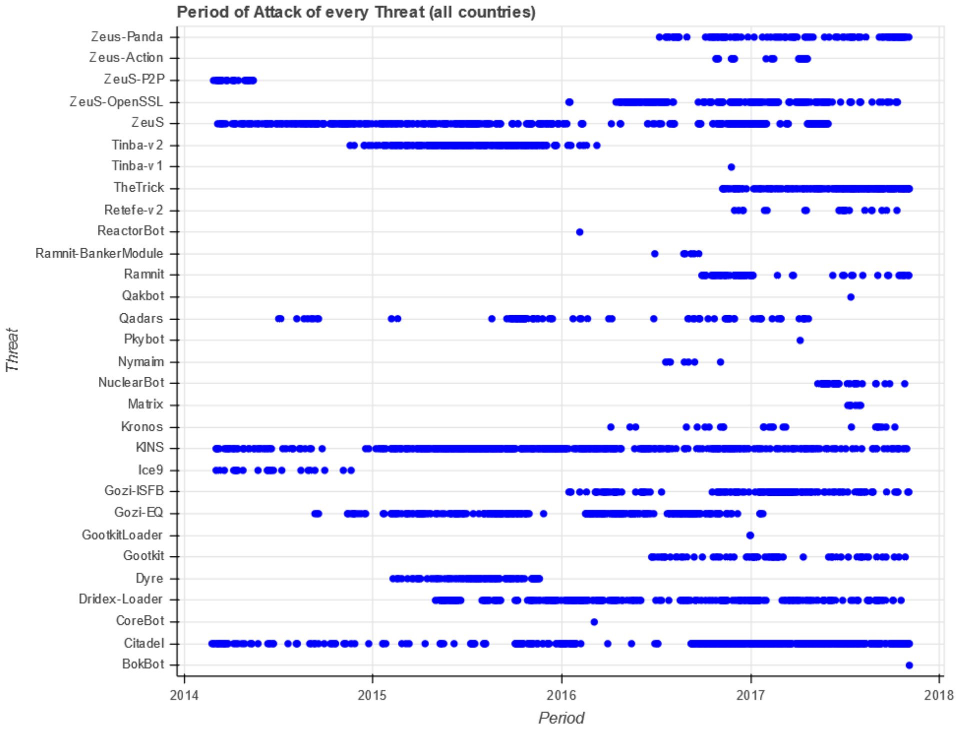

Fig. 3: Period of Attack per Threat

Based on the merge between attack target data and the list of bank, the proportion of targeted and non-targeted banks per country can also be obtained. Figure 4 provides the insight on the number of targeted and non-targeted banks per country in Europe. It is seen from the plot that Germany had the highest number of banks being targeted by the malware. However, a justification for this could be Germany has many banks. Yet, still, only less than 50% of the total number of banks in Germany were targeted. This is also true for many other countries. Looking deeper into the merged data, many of the banks not targeted in countries like Germany, Poland and Ireland are cooperative banks. They are banks organized on a cooperative basis and mostly covered only a limited local area in the country. Meanwhile, Croatia and Bulgaria have a high ratio of their banks being targeted by the malware, despite only having a few number of banks. More than 80% of banks in both countries were targeted.

Further analyses showed that there is a significantly higher proportion of online banks with the domain in the popularity list for those which are targeted, compared to those which are not targeted, which suggests that domain popularity has an effect on increasing the likelihood an online bank is targeted. It is also seen that the targeted group has a higher proportion of online banking entities offering English. However, the pattern is not as obvious as the popularity of an online banking domain. Meanwhile, like English, the proportion of the presence of 2 factor authentication in the targeted groups is also slightly higher than the proportion in the non-targeted groups. The relationship between these factors and the likelihood of the attack will be explained more in the explanatory analysis.

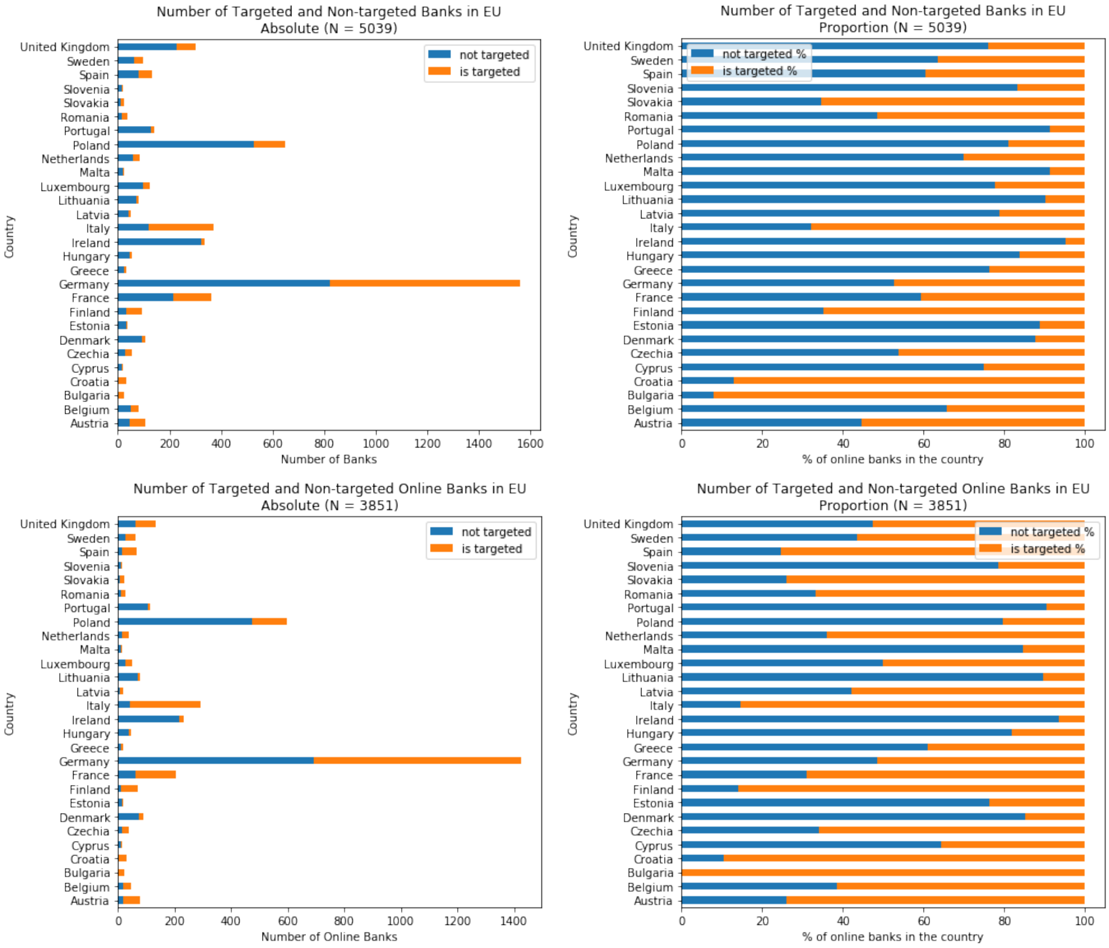

Fig. 4: Top: plot on number of targeted and non-targeted banks per country in EU (left: in absolute number, right: in proportion). Bottom: plot on number of targeted and non-targeted online banks (banks without online banking are removed) per country in EU (left: in absolute number, right: in proportion)

7.2 Explanatory Analysis

3,851 banks that have an online banking service will be taken into account for the explanatory analysis. Since the threat/malware variant and year factor are included for the model, the dataset was adjusted so that each datapoint refer to the attack count to an entity by a malware variant in a year. As a result, a dataset of 35,471 data points is ready for the analysis, with 33,384 data points of targeted entities.

The data distribution of the attack count was shown to follow a negative binomial distribution, where the standard deviation is higher than the mean. Since the data do not follow a normal distribution, normal linear regression is not suitable. Therefore, this research generates two regression models to approach it. First, logistic regression is used to model the probability of being targeted (having an attack count more than zero) or not targeted (having an attack count of zero). The logistic regression is useful to address potential issues caused by a lot of zero values in the data; singularity problem, for instance. Second, for the targeted entities, the data will be fitted by using a Poisson-family model, which is suitable for a dispersed count. The first attempt is to get the dispersion parameter of the data in order to justify which model is more suitable. Poisson GLM was run to check the dispersion parameter of raw attack count, 7-day interval attack count and the unique attack ID count. It is found that the dispersion parameters for all of them are 323.98, 36.80 and 20.26, respectively. The dispersion parameters are significantly higher than 1, indicating that the data is over-dispersed. This leads to the use of negative binomial regression, which is more suitable for an over-dispersed count data distribution. The use of any zero-inflated regression model is discouraged in this case since the number of zero values is still not excessive enough.

Model generation Four regression models were created to explain the relationship between the assessed characteristics and the likelihood of the attack as indicated by the attack count. The details and summary of the regression models can be seen in appendix A.

The first model is the logistic regression. It is made to explain the probability of an online banking entity being targeted or not targeted. A new variable is created whether an entity is targeted or not; true (1) if the attack count is more than zero for the entity and false (0) if the attack count is zero. This new variable becomes the dependent variable for the model. For the logistic regression, the presence of a particular language, the language count, the presence of an authentication factor, the domain popularity score and country are included as predictors. The other variables extracted from the attack target data are not included because they do not provide the variability required for performing the regression, for example, the number of unique threats attacking an online banking entity is also zero if the raw attack count is zero.

The model has a difference of 1,627.504 between null and residual deviance, with the associated p-value of 2.65e-307 (below 0.001), which indicates that the model with predictors is significantly better than the similar model with only intercept (null model). The model also has McFadden R-square value of 0.306. Further analysis using this model showed that around 70% of data points were correctly categorized, while around 25% were identified as false negative (actually targeted but not identified as targeted) and 5% as false positive (identified as targeted while actually not targeted). From the ROC curve, shown in Figure 5, it is seen that the area under curve (AUC) is around 0.831.

The second model is negative binomial regression, which is performed for data points which have the raw attack count of more than 0. Besides the predictors used in the logistic regression, other predictors that are extracted from attack data, such as threat, threat count and number of unique URLs, are included. The model has a difference of 82,806.56 between null and residual deviance, with the associated p-value of 0 (below 0.001), which indicates that the model with predictors is significantly better than the similar model with only intercept (null model). The model also has McFadden R-square value of 0.131.

The third model is the negative binomial regression which is performed towards week-interval attack count. The predictors used are the same as those for the regression towards raw attack count.

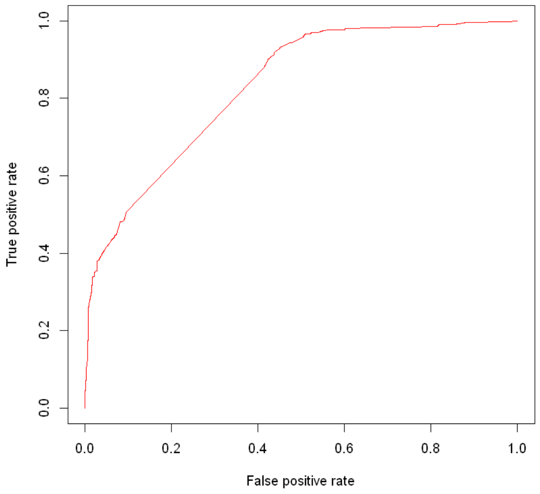

Fig. 5: ROC curve for logistic regression

The model has a difference of 52,671.1 between null and residual deviance, with the associated p-value of 0 (below 0.001), which indicates that the model with predictors is significantly better than the similar model with only intercept (null model). The model also has McFadden R-square value of 0.135.

The fourth model is the negative binomial regression which is performed for the attack ID count. The predictors used are the same as those for the regression towards raw attack count and 7-day interval attack count. The model has a difference of 87,265.09 between null and residual deviance, with the associated p-value of 0 (below 0.001), which indicates that the model with predictors is significantly better than the similar model with only intercept (null model). The model also has McFadden R-square value of 0.199.

As an illustration, the regression result for several language and authentication factor variables are presented in table 2. More detailed outcome and summary of the above regression models can be seen in appendix A.

Table 2: Short summary of regression models, for several variables

	Dependent variable:			
	is_targeted	raw_attack_count	week_attack_count	id_attack_count
	logistic	*negative binomial*	*negative binomial*	*negative binomial*
	(1)	(2)	(3)	(4)
Lang: English	0.629^{*}	0.379^{***}	0.124^{***}	0.193^{***}
	(0.314)	(0.029)	(0.025)	(0.025)
Lang: German	0.515	0.265^{***}	-0.150^{***}	0.005
	(0.385)	(0.038)	(0.032)	(0.033)

	Dependent variable:			
	is_targeted	raw_attack_count	week_attack_count	id_attack_count
	logistic	*negative binomial*	*negative binomial*	*negative binomial*
	(1)	(2)	(3)	(4)
Lang: Dutch	0.604 (0.561)	0.277*** (0.060)	0.364*** (0.051)	0.396*** (0.052)
Lang: Italian	0.847 (0.639)	−0.046 (0.058)	−0.279*** (0.049)	−0.256*** (0.050)
Lang: Spanish	−0.061 (0.625)	0.947*** (0.072)	0.537*** (0.061)	0.794*** (0.061)
Lang: Portugese	0.885 (1.022)	−0.228* (0.089)	−0.677*** (0.077)	−0.868*** (0.079)
Lang: Swedish	2.146** (0.793)	0.490*** (0.138)	0.318** (0.118)	0.379** (0.118)
Lang: Estonian	−3.439*** (0.916)	0.289* (0.135)	−0.557*** (0.119)	−0.249* (0.123)
2-factor Auth.	2.148*** (0.266)	−0.207*** (0.033)	−0.173*** (0.028)	−0.106*** (0.028)
Note:				*p<0.05; **p<0.01; ***p<0.001 Standard errors in brackets

Based on the model result, it is seen that the language variables, in general, do not show their significance in the logistic regression model. Only English, Swedish and Estonian are seen significant enough. However, the significance of Estonian should be treated with caution as this variable has a strong correlation with the 'Estonia' country variable and seen to be singular in the model. As multicollinearity exists between these variables, it is hard to determine whether the significance happened because of the Estonian language or because the banks are located in Estonia. Meanwhile, it is seen in English and Swedish that the coefficient is positive, which indicates that online banking entities which offer these languages have higher chance to be targeted.

More language variables are seen to be significant in the negative binomial models, indicating that the language may be able to explain why some banks are more or less targeted than others. English is again significant and has a positive coefficient, which indicates that its presence may cause the tendency of the online banking entity to be more targeted. This effect is also relevant for several other languages like Dutch and Spanish. The opposite effect seems to be present for Italian and Portugese. Based on the regression models above, despite the presence of control variables like the threat/malware variants, country, year and domain popularity, some languages are seen to still maintain their significance in explaining whether an online banking entity is more or less targeted.

It is important to note early that the result between models can be different, depending on the attack count metric used. This signals that the way the attack is counted affects the result and hence the quality of the analysis. The closer the metric to counting the actual attack is, the more acceptable the analysis. Although every metric approaching the actual attack count has its own advantages and limitations, this thesis evaluates that the metric that uses unique attack ID to count the attack seems to approach the actual attack count the closest among other metrics.

This is due to the metric being based on the legitimate explanation of how the malware data is stored instead of arguable assumptions that other metrics rely on.

It is also seen in the models that the presence of two-factor authentication (2FA) can explain the target selection of banking malware. This variable is seen to be significant in both logistic regression and negative binomial models. This finding seems to contradict Van Moorsel's argument that financial institutions with two-factor authentication are selected as much over the years as ones with one-factor authentication [43]. However, the interpretation of this variable is quite complicated. The variable has a positive coefficient in the logistic model, which means there are more banks implementing 2FA that are targeted (than banks which do not implement 2FA). On the contrary, it has a negative coefficient in all negative binomial models, which indicates that the presence of 2FA reduces the tendency of online banking to be more targeted. A possible interpretation to these findings is, since 2FA is becoming common nowadays, many criminals are still trying to attack the online banking. However, they may not or cannot perform a lot of attack, or maybe they tried to attack it in the first time and realized that the benefit was unsatisfactory. As the result, it discourages them to perform more attacks, causing a tendency the targeted bank to be less targeted.

In order to see how the significance level of variables changes as more factors were included in the model, several models were created out of the negative binomial regression model towards unique attack ID count. These models took into account different number of predictors: (1) the first model only considers language and authentication factor, (2) the second model adds unique attack URL count to the first model, (3) the third model add domain popularity to the second model, (4) the fourth model adds country variable to the third model, (5) the fifth model includes threat variable to the fourth model, and (6) the sixth model includes year variables, makes it similar to the model initially generated. Some independent variables are selected in order to show how their coefficient and standard deviation change overtime due to the addition of control variables. The summary is presented in table 3.

It is seen that the coefficient of the variables changed while more factors are taken into account. However, some variables were able to maintain their significance despite the presence of more factors, like the presence of English and 2-factor authentication. There are also some language variables which lose their significance after the addition of control variables. This indicates that some factors might be more important than others. It is also seen that the model gets better in explaining the variance of the data as more factors are included, indicated by lower Akaike Information Criterion (AIC). However, the AIC is still high. Referring back to the initial logistic and negative binomial models, the pseudo McFadden-R^2 of the models is considered low, around 0.3 for the logistic regression and 0.1 for the negative binomial regression. Although pseudo R^2 does not indicate the ability of the model to explain the data variance, like the real R^2, having a low pseudo R^2 suggests that the model is still far from the perfect fit and there might be many other factors out there which could improve the model in explaining the target selection.

The above inference is made based on the generated model and result, and it should be taken into account that the model itself has some limitations that may prevent plausible conclusions to be drawn. The models are still not able to include many more factors due to limited data and time. As an implication, the model cannot explain the influence of other factors that are not considered in the model on the target selection. Some non-considered factors may be more important than the factors considered in this model. Moreover, data quality could also be an issue. As explained previously, the model relies on attack target data from Fox IT and also Fox IT's mechanism in identifying attack URLs which are associated with EU banks. Although the improvement was applied, the unidentified bank-related attack URLs, or the false negatives, may still be present and they may cause the model to not be perfectly accurate. If the identification quality were better, the data would have been more robust and hence the model and the result would have been different to the current ones. Human error may have also occurred during manual observation of online banking entities, which would have affected the data quality.

Table 3: Negative binomial regression towards unique attack ID count

	Response Variable: Attack ID Count					
	Negative Binomial					
	(1)	(2)	(3)	(4)	(5)	(6)
Lang: English	1.108^{***} (0.032)	0.580^{***} (0.031)	0.522^{***} (0.031)	0.407^{***} (0.035)	0.188^{***} (0.026)	0.193^{***} (0.025)
Lang: Dutch	1.031^{***} (0.044)	0.496^{***} (0.043)	0.460^{***} (0.043)	0.586^{***} (0.072)	0.392^{***} (0.052)	0.396^{***} (0.052)
Lang: Spanish	1.735^{***} (0.065)	1.133^{***} (0.062)	1.119^{***} (0.062)	1.222^{***} (0.087)	0.776^{***} (0.061)	0.794^{***} (0.061)
Lang: Danish	0.614^{***} (0.091)	0.618^{***} (0.087)	0.542^{***} (0.087)	0.410^{*} (0.182)	0.039 (0.130)	0.045 (0.128)
2-factor Auth.	−0.043 (0.041)	-0.157^{***} (0.039)	-0.134^{***} (0.039)	-0.091^{*} (0.040)	-0.108^{***} (0.028)	-0.106^{***} (0.028)
Unique Attack URL count		0.019^{***} (0.0003)	0.016^{***} (0.0003)	0.016^{***} (0.0004)	0.012^{***} (0.0002)	0.012^{***} (0.0002)
Domain popularity			0.00000^{***} (0.00000)	0.00000^{***} (0.00000)	0.00000^{***} (0.00000)	0.00000^{***} (0.00000)
Country				True	True	True
Threat					True	True
Year						True
Log Likelihood	−111,203.800	−109,154.600	−109,084.700	−108,967.800	−93,033.010	−92,708.800
θ	0.634^{***} (0.005)	0.703^{***} (0.005)	0.706^{***} (0.005)	0.710^{***} (0.005)	1.729^{***} (0.015)	1.767^{***} (0.016)
Akaike Inf. Crit.	222,457.500	218,363.200	218,225.300	218,043.600	186,232.000	185,589.600

Note: *p<0.05; **p<0.01; ***p<0.001

Standard errors in brackets

8 Result Interpretation by the Experts

Several interview sessions with experts were conducted in order to get experts' knowledge and perspective about the target selection of banking malware as well as to obtain experts' interpretation of the models and analysis results. There are three experts interviewed, all of them are security experts in leading banks in the Netherlands. Therefore, they possess expertise and experience relevant to the topic of the research.

The interview is conducted in a semi-structured manner. It was divided into two phases. In the first phase, the questions about their opinion on the target selection are asked. The discussion also continues with explaining the potential factors, both financial and non-financial, from their point of view, which could affect the target selection. A slot is also given for them to argue about how the factors that become the highlights in this thesis, languages, authentication factors and domain popularity, could affect the target selection. The first phase is conducted before the model and the result are presented so that the experts perspective is not affected by the presence of the model and the result. In the second phase, the model and the result are presented. After that, questions are given to the experts in order to get their opinion about the model and the result. Their interpretation on the outcome of the model are also requested.

During the interview, experts expressed that there are many factors which can influence the target selection of banking malware. One important factor is something to gain for criminals. This relates with a positive business case the criminals have when they target a particular bank. It has to do mainly with financial and market profile of the banks. After the business case, the criminals will look at the easier target. This is where technical factors like the language and authentication factor become important. Experts argued that malware nowadays is becoming more personalised and therefore criminals started to look at language as a factor that could determine whether the attack attempt could be effective. In their opinion, English-speaking countries are more attacked as English is used widely and many people have an ability to speak it. However, they saw that malware is also getting more localised as criminals try to recruit local people to help with translating the phising page into other languages. Experts also believe that a good form of authentication is a good way to increase a sense of security and make an attack less successful. There is a concern, addressed by an expert, that the latest malware has a capability to circumvent multiple factor authentication. Security control is also seen as important by experts, especially the quality of detection system. Meanwhile, other characteristics like the record of successful attack attempts, the number of banks in a country, the maturity of online banking system of the country and the adoption rate of online banking by the society were argued to also influence the target selection. However, an expert believes that the financial factor is still the biggest deciding factor than others for the target selection.

In general, experts found that the regression models generated in this thesis look reasonable. Moreover, experts saw that this kind of model could be useful as a supplement to the risk model. However, they highlighted a limited number of factors considered in the models as something that may affect the quality of the models as well as the analysis. Experts expressed that it is a bit dangerous to make a conclusion from the outcome because it did not look at the full scope of the attack since the model only considered the language, the authentication factor and some of control variables. Experts believed that many other factors are more important than these factors in terms of explaining the target selection, especially, an expert highlighted the relative market share of the bank as a potential important factor. He argued that if the relative market share is compared together with the language factor, for example, it will diminish the significance of the language in explaining the target selection. Besides, taking into account factors about security control may also influence the result. In short, experts suggested that there are more factors that can explain the target selection and, only by considering all possible factors, one can draw a plausible conclusion about the target selection. Also, more robust data is needed in order to create a better model.

9 Discussion

There have been a lot of theoretical, quantitative and qualitative work that has been performed in order to answer the main research question. First of all, literature study were carried out in order to get the concept of cyberspace and cybercrime, the background, mechanism and security of online banking, the banking malware and actors in the financial malware attack. It also observed the criminology theory and argued why the Routine Active Theory (RAT) is still relevant to the cybersecurity cases. After that, characteristics which could potentially influence target selection of the banking malware, in accordance with relevant aspects of RAT, were defined by literature review.

In the next step, data collection was completed so that the statistical model can be generated. This process, in a big picture, consists of several activities, such as extracting and filtering of attack target data, merging the data with list of EU banks and collecting external data by finding a source or manual observation. From the collected data, possible metrics and variables related to the selected characteristics were defined. This research also proposed a metric to approach the actual attack count better than the traditional raw attack count.

Descriptive and explanatory analysis was then conducted to derive insights about the data and also to find out which of the extracted variables have a significant relationship to the target selection. Statistical models were generated in the explanatory analysis. Finally, expert interview was conducted in order to obtain experts' opinion about the target selection as well as to get their comments and interpretations about the models.

In general, many interesting insights can be extracted from the data and the model. Some variables analysed were shown to have adequate significance with respect to the target selection. Some of them retain their significance despite the presence of control variables in the model. However, experts think that although some characteristics that were assessed in this research may be able to explain the target selection, it is too premature to conclude it that way since the argument was extracted from a limited model. Experts argued that in order to come up with plausible conclusions about the target selection, the model should be improved by adding many other factors into the mix. However, it should be taken into account that in order to add more factors in the model, more data is needed, which usually is difficult to gather. This suggests that the research of this topic should be iterative, meaning that future studies are required following this thesis so that more reliable models and more acceptable conclusions about the target selection could be drawn in the future.

10 Limitation and Future Analysis

The research has many limitations which, to an extent, may give implications to the outcome of the research. First, there are data and time limitations which enable the research to only include factors that become the focus of the research, i.e. language, authentication factor and some control variables. Many other factors, especially some financial and market-related factors which are deemed more determining according to the experts, cannot be included. This may create a problem since a limited model could lead to implausible conclusions. In order to come up with better explanation of the target selection, the model should consider additional factors. Second, there is also a limitation with data quality. The attack target dataset came from an external source and it was found that the initial mechanism used to identify the bank profile from the URL in the dataset was not perfect. Despite an improvement in place, some false negatives could still be present. Besides, some of other data were observed manually, which increases the risk of human errors coming into an effect. Third, there were several data points removed due to the lack of language and authentication factor information due to the banks not having an online banking service. This removal may have affected the model as well as the result. Finally, there is also a limitation with a metric formulated in this study. The week-interval attack count is heavily based on the assumption that the criminals follow the same interval periods when making the same actual attack, while the reality is more unpredictable than that. Therefore, the metric is only capable for guessing the actual attack, rather than making a precise count.

Based on the above limitations and also other inputs, some potential future studies are proposed:

- Enhance the research by adding financial and market factors to the model. Especially, experts mentioned a factor "relative market share of the bank" as another important factor that could explain the target selection.
- Extend this study with other studies that research other factors of the online banking to make a more comprehensive result so that more plausible conclusion can be drawn.
- Conduct a study to design and build much improved mechanisms or algorithms for identifying bank domains and entities out of the attack URLs.
- Execute the same quantitative analysis using a more robust data source in order to obtain a better model result and more reliable analysis.

11 Conclusion

With respect to the main research question, it is concluded that there are several characteristics of banks' online banking services that potentially can explain the likelihood of them being attacked by the banking malware. Several characteristics were assessed, and it is seen that some of them were significant for explaining the target selection, despite the presence of the control variables in the model. However, in order to make a plausible conclusion about their influence, according to the experts, improving the current statistical models by considering more factors, especially financial- and market-related ones, is required.

Bibliography

[1] Anderson R, Barton C, Böhme R, Clayton R, van Eeten MJG, Levi M, Moore T, Savage S (2013) Measuring the Cost of Cybercrime, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 265–300. DOI 10.1007/978-3-642-39498-0_12

[2] Asghari H, van Eeten M, Bauer JM (2016) Economics of cybersecurity. In: Handbook on the Economics of the Internet, Edward Elgar Publishing, chap 13, pp 262–287

[3] van den Berg J, van Zoggel J, Snels M, van Leeuwen M, Boeke S, van de Koppen L, van der Lubbe J, van den Berg B, Bos TD (2014) On (the Emergence of) Cyber Security Science and its Challenges for Cyber Security Education. NATO STO/IST-122 symposium in Tallin (c):1–10

[4] Bottazzi G, Me G (2014) The Botnet Revenue Model. In: Proceedings of the 7th International Conference on Security of Information and Networks, ACM, New York, NY, USA, SIN '14, pp 459:459—-459:465, DOI 10.1145/2659651.2659673

[5] Bougaardt G, Kyobe M (2011) Investigating the factors inhibiting SMEs from recognizing and measuring losses from cyber crime in South Africa. In: ICIME 2011-Proceedings of the 2nd International Conference on Information Management and Evaluation: ICIME 2011 Ryerson University, Toronto, Canada, 27-28 April 2011, Academic Conferences Limited, p 62

[6] Boutin JI (2013) Qadars a banking Trojan with the Netherlands in its sights. URL `https://www.welivesecurity.com/2013/12/18/qadars-a-banking-trojan-with-the-netherlands-in-its-sights/`

[7] Brenner SW (2006) At light speed: Attribution and response to cybercrime/terrorism/warfare. J Crim L & Criminology 97:379

[8] Campanella F, Della Peruta MR, Del Giudice M (2017) The Effects of Technological Innovation on the Banking Sector. Journal of the Knowledge Economy 8(1):356–368, DOI 10.1007/s13132-015-0326-8, URL `https://doi.org/10.1007/s13132-015-0326-8`

[9] Cheung R (2017) Targeting financial organisations: a multi-sided perspective. Master's thesis, Delft University of Technology

[10] Cimpanu C (2016) Qadars Trojan Returns Bigger and Badder than Ever Before. URL `https://news.softpedia.com/news/qadars-trojan-returns-bigger-and-badder-than-ever-before-508546.shtml`

[11] Cisco Umbrella (????) Cisco Popularity List. URL `http://s3-us-west-1.amazonaws.com/umbrella-static/index.html`

[12] Claessens J, Dem V, De Cock D, Preneel B, Vandewalle J (2002) On the Security of Today's Online Electronic Banking Systems. Computers & Security 21(3):253–265, DOI 10.1016/S0167-4048(02)00312-7, URL `http://www.sciencedirect.com/science/article/pii/S0167404802003127http://linkinghub.elsevier.com/retrieve/pii/S0167404802003127`

[13] Cohen LE, Felson M (1979) Social Change and Crime Rate Trends: A Routine Activity Approach. American Sociological Review 44(4):588, DOI 10.2307/2094589, URL `http://www.jstor.org/stable/2094589?origin=crossref`

[14] Craig D (2016) Five Cybersecurity Challenges Facing Financial Services Organizations Today. URL `https://securityintelligence.com/five-cybersecurity-challenges-facing-financial-services-organizations-today/`

[15] Creswel JW (2008) The Selection of a Research Approach. Research design: qualitative, quantitative, and mixed methods approaches pp 3–22, DOI 45593:01, `arXiv:1011.1669v3`

[16] Gordon S, Ford R (2006) On the definition and classification of cybercrime. Journal in Computer Virology 2(1):13–20

[17] Hutchings A, Clayton R (2017) Configuring Zeus: A case study of online crime target selection and knowledge transmission. eCrime Researchers Summit, eCrime pp 33–40, DOI 10.1109/ECRIME.2017.7945052

[18] Hutchinson D, Warren M (2003) Security for Internet banking: a framework. Logistics Information Management 16(1):64–73, DOI 10.1108/09576050310453750, URL http://www.emeraldinsight.com/doi/10.1108/09576050310453750

[19] Infosecurity Magazine (2010) Zeus is king of bank fraud trojan viruses - Infosecurity Magazine. URL https://www.infosecurity-magazine.com/news/zeus-is-king-of-bank-fraud-trojan-viruses/

[20] Jaleshgari RP (1999) Document trading online. Information Week 755(136):136

[21] Kiljan S, Simoens K, De Cock D, van Eekelen M, Vranken H (2014) Security of Online Banking Systems. Tech Rep TR-OU-INF-2014-01 (Open Universiteit)

[22] Kiljan S, Vranken H, van Eekelen M (2018) Evaluation of transaction authentication methods for online banking. Future Generation Computer Systems 80:430–447, DOI 10.1016/j.future.2016.05.024, URL http://dx.doi.org/10.1016/j.future.2016.05.024

[23] Kiljan SZ (2017) Exploring, Expanding and Evaluating Usable Security in Online Banking. Open Universiteit

[24] Lagazio M, Sherif N, Cushman M (2014) A multi-level approach to understanding the impact of cyber crime on the financial sector. Computers and Security 45(0):58–74, DOI 10.1016/j.cose.2014.05.006

[25] Leukfeldt ER, Yar M (2016) Applying Routine Activity Theory to Cybercrime: A Theoretical and Empirical Analysis. Deviant Behavior 37(3):263–280, DOI 10.1080/01639625.2015.1012409

[26] Lowe G (1997) A hierarchy of authentication specifications. In: Proceedings 10th Computer Security Foundations Workshop, pp 31–43, DOI 10.1109/CSFW.1997.596782

[27] Lu Z, Marvel L, Wang C (2015) To be proactive or not: a framework to model cyber maneuvers for critical path protection in MANETs. In: Proceedings of the Second ACM Workshop on Moving Target Defense, ACM, pp 85–93

[28] Nagurney A (2015) A multiproduct network economic model of cybercrime in financial services. Service Science 7(1):70–81

[29] Nand P, Astya R, Singh D (2015) An Add-on to Present Banking: m-banking. In: Proceedings of Fourth International Conference on Soft Computing for Problem Solving , Advances in Intelligent Systems and Computing, Springer India, vol 336, pp 377–388, DOI 10.1007/978-81-322-2220-0, URL http://link.springer.com/10.1007/978-81-322-2220-0

[30] Petee TA, Corzine J, Huff-Corzine L, Clifford J, Weaver G (2010) Defining ”cyber-crime”: Issues in determining the nature and scope of computer-related offenses. Futures Working Group 5:6–11

[31] PlugandPlay Tech Center (2017) The Cybersecurity Threats Facing Financial Institutions. URL http://plugandplaytechcenter.com/2017/06/26/cybersecurity-threats-financial-institutions/

[32] PwC (2018) Pulling fraud out of the shadows: Global Economic Crime and Fraud Survey 2018. Tech. rep., PricewaterhouseCoopers

[33] Raghavan AR, Parthiban L (2014) The effect of cybercrime on a bank’s finances. International Journal of Current Research and Academic Review 2(2):173–178

[34] Schwarz D (2015) Peeking at Pkybot. URL https://asert.arbornetworks.com/peeking-at-pkybot/

[35] Secureworks Counter Threat Unit Threat Intelligence (2016) Banking Botnets: The Battle Continues — Secureworks. URL https://www.secureworks.com/research/banking-botnets-the-battle-continues

[36] Sherstobitoff R (2012) Inside the world of the citadel trojan. Emergence 9

[37] Singer PW, Friedman A (2014) Cybersecurity: What everyone needs to know. Oxford University Press, Oxford

[38] Tajalizadehkhoob S (2013) Online Banking Fraud Mitigation: A Quantitative Study for Extracting Intelligence about Target Selection by Cybercriminals from Zeus Financial Malware Files. Master’s thesis, Delft University of Technology

[39] Tajalizadehkhoob S, Asghari H, Gan C, van Eeten M (2014) Why Them? Extracting Intelligence about Target Selection from Zeus Financial Malware. Workshop on the Economics of Information Security (WEIS) pp 1–26

[40] Tajalizadehkhoob S, Gan C, Noroozian A, van Eeten M (2017) The Role of Hosting Providers in Fighting Command and Control Infrastructure of Financial Malware. In: Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security, ACM, New York, NY, USA, ASIA CCS '17, pp 575–586, DOI 10.1145/3052973.3053023

[41] Thomas D, Loader BD (2000) Introduction - cybercrime: law enforcement, security and surveillance in the information age, Cybercrime: Law Enforcement, Security and Surveillance in the Information Age

[42] TrendMicro (2012) Security Threats to Business, the Digital Lifestyle, and the Cloud. Tech. rep., URL http://www.trendmicro.com/cloud-content/us/pdfs/security-intelligence/spotlight-articles/sp-trendmicro-predictions-for-2013-and-beyond.pdf

[43] Van Moorsel D (2016) Target selection regarding financial malware attacks within the Single Euro Payments Area. Master's thesis, Delft University of Technology

[44] Vrancianu M, Popa LA (2010) Considerations Regarding the Security and Protection of E-Banking Services Consumers' Interests. The AMFITEATRU ECONOMIC journal 12(28):388–403, URL http://ideas.repec.org/a/aes/amfeco/v12y2010i28p388-403.html

[45] Wyke J (2011) What is Zeus? Tech. rep., SophosLab, URL https://www.sophos.com/en-us/medialibrary/pdfs/technical%20papers/sophos%20what%20is%20zeus%20tp.pdf

[46] Xu Y, Bailey M, Vander Weele E, Jahanian F (2010) CANVuS: Context-aware network vulnerability scanning. In: International Workshop on Recent Advances in Intrusion Detection, Springer, pp 138–157

[47] Yar M (2005) The Novelty of 'Cybercrime': An Assessment in Light of Routine Activity Theory. European Journal of Criminology 2(4):407–427

Appendices

A Regression Result

A.1 Logistic Regression

```
Call:
glm(formula = is_targeted ~ langEnglish + langGerman + langFrench +
    langDutch + langItalian + langSpanish + langPortugese + langGreek +
    langCzech + langSlovak + langSlovenian + langPolish + langHungarian +
    langRomanian + langBulgarian + langDanish + langSwedish +
    langFinnish + langLatvian + langEstonian + langLithuanian +
    lang_count + auth1FA + auth2FA + pop_score + Country, family = "binomial",
    data = data_logit)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.1156  -0.8775  -0.2468   1.1215   3.5663

Coefficients: (1 not defined because of singularities)
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)          -8.561e-01  4.860e-01  -1.762 0.078113 .
langEnglishTrue       6.286e-01  3.142e-01   2.001 0.045420 *
langGermanTrue        5.152e-01  3.851e-01   1.338 0.180985
langFrenchTrue        3.548e-01  4.551e-01   0.780 0.435605
langDutchTrue         6.035e-01  5.615e-01   1.075 0.282397
langItalianTrue       8.465e-01  6.395e-01   1.324 0.185564
langSpanishTrue      -6.052e-02  6.252e-01  -0.097 0.922880
langPortugeseTrue     8.850e-01  1.022e+00   0.866 0.386715
langGreekTrue         1.160e+00  1.211e+00   0.958 0.337857
langCzechTrue         9.680e-02  1.106e+00   0.087 0.930285
langSlovakTrue        2.075e+00  1.074e+00   1.932 0.053340 .
langSlovenianTrue    -1.987e+00  1.817e+00  -1.094 0.274011
langPolishTrue       -1.573e+00  1.623e+00  -0.970 0.332247
langHungarianTrue     1.987e+00  1.561e+00   1.273 0.203008
langRomanianTrue     -1.735e+01  3.917e+02  -0.044 0.964662
langBulgarianTrue     1.196e+01  1.485e+03   0.008 0.993572
langDanishTrue       -5.425e-01  1.197e+00  -0.453 0.650465
langSwedishTrue       2.146e+00  7.926e-01   2.708 0.006777 **
langFinnishTrue      -1.894e+00  1.335e+00  -1.419 0.155779
langLatvianTrue       2.482e-01  1.568e+00   0.158 0.874244
langEstonianTrue     -3.439e+00  9.157e-01  -3.756 0.000173 ***
langLithuanianTrue    1.597e+01  2.400e+03   0.007 0.994690
lang_count           -3.465e-02  2.579e-01  -0.134 0.893119
auth1FATrue           4.577e-01  2.861e-01   1.600 0.109644
auth2FATrue           2.148e+00  2.658e-01   8.084 6.26e-16 ***
pop_score             4.466e-06  2.956e-07  15.111  < 2e-16 ***
CountryBelgium       -1.844e+00  7.015e-01  -2.629 0.008574 **
CountryBulgaria       1.413e+01  1.370e+03   0.010 0.991771
CountryCroatia        7.971e-01  8.065e-01   0.988 0.322969
CountryCyprus        -1.879e+00  8.693e-01  -2.161 0.030681 *
```

```
CountryCzechia          -3.515e-01  1.204e+00  -0.292 0.770389
CountryDenmark          -4.040e+00  1.265e+00  -3.193 0.001407 **
CountryEstonia                  NA         NA      NA       NA
CountryFinland           1.552e+00  1.358e+00   1.143 0.253238
CountryFrance           -6.775e-01  5.497e-01  -1.233 0.217755
CountryGermany          -9.497e-02  2.994e-01  -0.317 0.751092
CountryGreece           -2.406e+00  1.277e+00  -1.883 0.059636 .
CountryHungary          -5.971e+00  1.576e+00  -3.790 0.000151 ***
CountryIreland          -3.671e+00  5.120e-01  -7.171 7.45e-13 ***
CountryItaly            -8.979e-01  7.032e-01  -1.277 0.201631
CountryLatvia           -1.435e+00  1.623e+00  -0.884 0.376778
CountryLithuania        -1.819e+01  2.400e+03  -0.008 0.993952
CountryLuxembourg       -2.082e+00  5.290e-01  -3.936 8.27e-05 ***
CountryMalta            -3.517e+00  1.032e+00  -3.410 0.000650 ***
CountryNetherlands      -1.343e+00  6.935e-01  -1.937 0.052804 .
CountryPoland           -1.455e-01  1.632e+00  -0.089 0.928978
CountryPortugal         -6.810e+00  1.073e+00  -6.349 2.17e-10 ***
CountryRomania           1.688e+01  3.917e+02   0.043 0.965613
CountrySlovakia         -1.136e+00  1.035e+00  -1.098 0.272076
CountrySlovenia         -7.328e-01  2.003e+00  -0.366 0.714491
CountrySpain             3.515e-01  7.017e-01   0.501 0.616451
CountrySweden           -3.840e+00  8.665e-01  -4.431 9.37e-06 ***
CountryUnited Kingdom   -1.060e+00  4.437e-01  -2.390 0.016847 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5319.3  on 3850  degrees of freedom
Residual deviance: 3691.7  on 3799  degrees of freedom
AIC: 3795.7

Number of Fisher Scoring iterations: 15
```

A.2 Negative Binomial Regression - Raw Attack Count

```
Call:
glm.nb(formula = raw_attack_count ~ langEnglish + langGerman +
    langFrench + langDutch + langItalian + langSpanish + langPortugese +
    langGreek + langCzech + langSlovak + langSlovenian + langPolish +
    langHungarian + langRomanian + langBulgarian + langDanish +
    langSwedish + langFinnish + langLatvian + langEstonian +
    langLithuanian + lang_count + auth1FA + auth2FA + pop_score +
    Country + threat_name + year + unique_attackurl_count ,
    data = data_notzero , init.theta = 1.033310472, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.2969  -0.9894  -0.2593   0.1431   9.0820
```

```
Coefficients: (1 not defined because of singularities)
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)             -2.704e+00  2.842e-01  -9.513  < 2e-16 ***
langEnglishTrue          3.788e-01  2.923e-02  12.957  < 2e-16 ***
langGermanTrue           2.651e-01  3.750e-02   7.070 1.55e-12 ***
langFrenchTrue           9.960e-03  4.865e-02   0.205 0.837788
langDutchTrue            2.771e-01  6.014e-02   4.607 4.08e-06 ***
langItalianTrue         -4.624e-02  5.796e-02  -0.798 0.424969
langSpanishTrue          9.468e-01  7.194e-02  13.161  < 2e-16 ***
langPortugeseTrue       -2.281e-01  8.913e-02  -2.560 0.010475 *
langGreekTrue            4.187e-03  1.631e-01   0.026 0.979521
langCzechTrue            3.085e-01  9.932e-02   3.106 0.001898 **
langSlovakTrue          -4.878e-01  1.075e-01  -4.537 5.71e-06 ***
langSlovenianTrue       -6.783e-01  1.473e-01  -4.606 4.11e-06 ***
langPolishTrue           5.567e-01  1.711e-01   3.254 0.001139 **
langHungarianTrue        4.117e-01  9.996e-02   4.119 3.81e-05 ***
langRomanianTrue         3.695e-01  1.936e-01   1.909 0.056247 .
langBulgarianTrue        2.753e-01  1.728e-01   1.593 0.111202
langDanishTrue           2.139e-01  1.515e-01   1.411 0.158183
langSwedishTrue          4.905e-01  1.382e-01   3.549 0.000387 ***
langFinnishTrue          3.895e-01  1.808e-01   2.155 0.031184 *
langLatvianTrue         -1.212e+00  8.624e-01  -1.405 0.159997
langEstonianTrue         2.894e-01  1.351e-01   2.142 0.032161 *
langLithuanianTrue      -7.331e-01  8.625e-01  -0.850 0.395345
lang_count              -1.369e-01  2.170e-02  -6.309 2.80e-10 ***
auth1FATrue              1.929e-03  3.637e-02   0.053 0.957694
auth2FATrue             -2.074e-01  3.281e-02  -6.321 2.59e-10 ***
pop_score                7.971e-07  2.954e-08  26.987  < 2e-16 ***
CountryBelgium          -2.713e-01  8.357e-02  -3.247 0.001168 **
CountryBulgaria         -5.882e-01  1.678e-01  -3.506 0.000455 ***
CountryCroatia           5.577e-03  6.748e-02   0.083 0.934128
CountryCyprus           -6.416e-01  1.823e-01  -3.519 0.000433 ***
CountryCzechia           6.510e-01  1.062e-01   6.128 8.93e-10 ***
CountryDenmark          -1.151e-01  1.560e-01  -0.738 0.460465
CountryEstonia                  NA         NA      NA       NA
CountryFinland           1.660e-02  1.734e-01   0.096 0.923726
CountryFrance           -5.627e-01  5.654e-02  -9.953  < 2e-16 ***
CountryGermany          -7.734e-01  3.397e-02 -22.768  < 2e-16 ***
CountryGreece           -9.466e-01  1.699e-01  -5.572 2.51e-08 ***
CountryHungary           5.615e-02  1.211e-01   0.464 0.642784
CountryIreland          -5.111e-01  7.790e-02  -6.560 5.36e-11 ***
CountryItaly            -1.569e-01  6.648e-02  -2.360 0.018282 *
CountryLatvia            1.179e+00  8.673e-01   1.359 0.174004
CountryLithuania         1.179e+00  8.673e-01   1.359 0.174004
CountryLuxembourg       -3.818e-01  6.788e-02  -5.625 1.86e-08 ***
CountryMalta            -1.228e-01  3.951e-01  -0.311 0.756028
CountryNetherlands      -5.509e-01  8.356e-02  -6.593 4.32e-11 ***
CountryPoland           -8.796e-01  1.706e-01  -5.157 2.51e-07 ***
CountryPortugal          5.701e-01  1.047e-01   5.445 5.19e-08 ***
CountryRomania           1.695e-01  1.914e-01   0.886 0.375804
CountrySlovakia         -9.691e-04  1.011e-01  -0.010 0.992353
CountrySlovenia          5.794e-01  2.081e-01   2.784 0.005365 **
```

```
CountrySpain                      -9.721e-01  7.502e-02 -12.957  < 2e-16 ***
CountrySweden                     -5.838e-01  1.398e-01  -4.176 2.96e-05 ***
CountryUnited Kingdom             -4.766e-01  5.153e-02  -9.249  < 2e-16 ***
threat_nameCitadel                 4.146e+00  2.771e-01  14.963  < 2e-16 ***
threat_nameCoreBot                 1.141e+00  7.360e-01   1.550 0.121116
threat_nameDridex-Loader           5.480e+00  2.767e-01  19.804  < 2e-16 ***
threat_nameDyre                    4.506e+00  2.783e-01  16.195  < 2e-16 ***
threat_nameGootkit                 3.607e+00  2.768e-01  13.032  < 2e-16 ***
threat_nameGootkitLoader           2.107e+00  2.784e-01   7.569 3.76e-14 ***
threat_nameGozi-EQ                 4.544e+00  2.789e-01  16.292  < 2e-16 ***
threat_nameGozi-ISFB               5.040e+00  2.770e-01  18.193  < 2e-16 ***
threat_nameIce9                    4.262e+00  2.862e-01  14.895  < 2e-16 ***
threat_nameKINS                    5.393e+00  2.775e-01  19.434  < 2e-16 ***
threat_nameKronos                  3.029e+00  2.779e-01  10.901  < 2e-16 ***
threat_nameMatrix                  1.815e+00  7.757e-01   2.339 0.019320 *
threat_nameNuclearBot              1.185e+00  2.780e-01   4.260 2.04e-05 ***
threat_nameNymaim                  1.232e+00  1.057e+00   1.166 0.243749
threat_namePkybot                 -7.069e-01  5.311e-01  -1.331 0.183199
threat_nameQadars                  3.800e+00  2.771e-01  13.713  < 2e-16 ***
threat_nameQakbot                  5.985e-01  1.174e+00   0.510 0.610269
threat_nameRamnit                  2.128e+00  2.837e-01   7.499 6.42e-14 ***
threat_nameRamnit-BankerModule     2.090e+00  4.435e-01   4.712 2.45e-06 ***
threat_nameReactorBot              1.812e+00  5.443e-01   3.328 0.000874 ***
threat_nameRetefe-v2               4.241e+00  2.888e-01  14.687  < 2e-16 ***
threat_nameTheTrick                4.458e+00  2.777e-01  16.053  < 2e-16 ***
threat_nameTinba-v1                2.225e+00  2.805e-01   7.933 2.14e-15 ***
threat_nameTinba-v2                4.080e+00  2.770e-01  14.729  < 2e-16 ***
threat_nameZeuS                    4.770e+00  2.781e-01  17.155  < 2e-16 ***
threat_nameZeus-Action             3.865e+00  2.925e-01  13.212  < 2e-16 ***
threat_nameZeuS-OpenSSL            7.311e+00  2.768e-01  26.415  < 2e-16 ***
threat_nameZeuS-P2P                6.381e+00  2.806e-01  22.745  < 2e-16 ***
threat_nameZeus-Panda              4.229e+00  2.775e-01  15.239  < 2e-16 ***
year2015                           1.736e+00  2.455e-02  70.706  < 2e-16 ***
year2016                           7.935e-01  2.347e-02  33.805  < 2e-16 ***
year2017                           1.301e+00  2.501e-02  52.025  < 2e-16 ***
unique_attackurl_count             9.244e-03  2.960e-04  31.232  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(1.0333) family taken to be 1)

    Null deviance: 118979  on 33383  degrees of freedom
Residual deviance:  36173  on 33298  degrees of freedom
AIC: 313666

Number of Fisher Scoring iterations: 1


              Theta:  1.03331
          Std. Err.:  0.00737

 2 x log-likelihood:  -313492.18500
```

A.3 Negative Binomial Regression - 7-day interval Attack Count

```
Call:
glm.nb(formula = week_attack_count ~ langEnglish + langGerman +
    langFrench + langDutch + langItalian + langSpanish + langPortugese +
    langGreek + langCzech + langSlovak + langSlovenian + langPolish +
    langHungarian + langRomanian + langBulgarian + langDanish +
    langSwedish + langFinnish + langLatvian + langEstonian +
    langLithuanian + lang_count + auth1FA + auth2FA + pop_score +
    Country + threat_name + year + unique_attackurl_count,
    data = data_notzero, init.theta = 1.622533869, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.5714  -0.8813  -0.2708   0.2325  15.1667

Coefficients: (1 not defined because of singularities)
                                 Estimate Std. Error z value Pr(>|z|)
(Intercept)                    -1.560e+00  2.507e-01  -6.222 4.91e-10 ***
langEnglishTrue                 1.235e-01  2.500e-02   4.941 7.75e-07 ***
langGermanTrue                 -1.499e-01  3.210e-02  -4.668 3.04e-06 ***
langFrenchTrue                 -2.843e-01  4.156e-02  -6.841 7.87e-12 ***
langDutchTrue                   3.644e-01  5.137e-02   7.094 1.31e-12 ***
langItalianTrue                -2.789e-01  4.898e-02  -5.694 1.24e-08 ***
langSpanishTrue                 5.370e-01  6.100e-02   8.803  < 2e-16 ***
langPortugeseTrue              -6.766e-01  7.682e-02  -8.807  < 2e-16 ***
langGreekTrue                  -2.434e-01  1.424e-01  -1.710 0.087333 .
langCzechTrue                  -2.650e-02  8.510e-02  -0.311 0.755506
langSlovakTrue                 -5.051e-02  9.256e-02  -0.546 0.585255
langSlovenianTrue              -3.487e-01  1.251e-01  -2.787 0.005312 **
langPolishTrue                  1.747e-01  1.446e-01   1.209 0.226704
langHungarianTrue               6.233e-02  8.448e-02   0.738 0.460634
langRomanianTrue                7.239e-01  1.646e-01   4.397 1.10e-05 ***
langBulgarianTrue               1.606e-01  1.495e-01   1.074 0.282707
langDanishTrue                  1.137e-01  1.288e-01   0.883 0.377438
langSwedishTrue                 3.180e-01  1.184e-01   2.687 0.007211 **
langFinnishTrue                 1.393e-01  1.542e-01   0.903 0.366327
langLatvianTrue                -5.707e-01  9.031e-01  -0.632 0.527449
langEstonianTrue               -5.573e-01  1.189e-01  -4.685 2.80e-06 ***
langLithuanianTrue             -6.297e-01  9.032e-01  -0.697 0.485670
lang_count                      1.604e-02  1.857e-02   0.864 0.387858
auth1FATrue                    -1.504e-02  3.107e-02  -0.484 0.628313
auth2FATrue                    -1.733e-01  2.791e-02  -6.208 5.37e-10 ***
pop_score                       4.416e-07  2.516e-08  17.553  < 2e-16 ***
CountryBelgium                 -2.115e-01  7.130e-02  -2.966 0.003014 **
CountryBulgaria                -4.521e-01  1.452e-01  -3.114 0.001844 **
CountryCroatia                 -1.857e-01  5.787e-02  -3.209 0.001334 **
CountryCyprus                  -5.356e-01  1.615e-01  -3.316 0.000914 ***
```

```
CountryCzechia                  4.342e-01  9.057e-02   4.795 1.63e-06 ***
CountryDenmark                 -1.959e-01  1.327e-01  -1.477 0.139720
CountryEstonia                         NA         NA      NA       NA
CountryFinland                 -1.784e-01  1.477e-01  -1.208 0.227237
CountryFrance                  -2.521e-01  4.827e-02  -5.223 1.76e-07 ***
CountryGermany                 -3.096e-01  2.907e-02 -10.649  < 2e-16 ***
CountryGreece                  -4.623e-01  1.476e-01  -3.132 0.001736 **
CountryHungary                  1.581e-01  1.021e-01   1.549 0.121441
CountryIreland                 -1.463e-01  6.565e-02  -2.228 0.025880 *
CountryItaly                   -3.646e-03  5.623e-02  -0.065 0.948306
CountryLatvia                   3.082e-01  9.068e-01   0.340 0.733946
CountryLithuania                3.082e-01  9.068e-01   0.340 0.733946
CountryLuxembourg               2.310e-02  5.786e-02   0.399 0.689718
CountryMalta                   -3.742e-01  3.755e-01  -0.997 0.318996
CountryNetherlands             -4.581e-01  7.125e-02  -6.430 1.28e-10 ***
CountryPoland                  -6.368e-01  1.440e-01  -4.421 9.81e-06 ***
CountryPortugal                 2.221e-01  9.020e-02   2.462 0.013817 *
CountryRomania                 -3.233e-01  1.630e-01  -1.984 0.047273 *
CountrySlovakia                -3.020e-01  8.734e-02  -3.457 0.000546 ***
CountrySlovenia                 2.293e-02  1.793e-01   0.128 0.898224
CountrySpain                   -6.121e-01  6.411e-02  -9.547  < 2e-16 ***
CountrySweden                  -4.172e-01  1.197e-01  -3.487 0.000489 ***
CountryUnited Kingdom          -3.232e-01  4.376e-02  -7.387 1.50e-13 ***
threat_nameCitadel              2.288e+00  2.449e-01   9.344  < 2e-16 ***
threat_nameCoreBot              5.620e-01  6.751e-01   0.832 0.405143
threat_nameDridex-Loader        3.799e+00  2.445e-01  15.538  < 2e-16 ***
threat_nameDyre                 3.135e+00  2.457e-01  12.761  < 2e-16 ***
threat_nameGootkit              2.415e+00  2.446e-01   9.874  < 2e-16 ***
threat_nameGootkitLoader        9.520e-01  2.467e-01   3.859 0.000114 ***
threat_nameGozi-EQ              2.619e+00  2.463e-01  10.634  < 2e-16 ***
threat_nameGozi-ISFB            3.594e+00  2.447e-01  14.686  < 2e-16 ***
threat_nameIce9                 2.040e+00  2.532e-01   8.057 7.81e-16 ***
threat_nameKINS                 2.965e+00  2.452e-01  12.092  < 2e-16 ***
threat_nameKronos               2.097e+00  2.456e-01   8.538  < 2e-16 ***
threat_nameMatrix               2.161e-01  7.327e-01   0.295 0.768016
threat_nameNuclearBot           6.094e-01  2.461e-01   2.477 0.013267 *
threat_nameNymaim               1.068e+00  8.717e-01   1.225 0.220482
threat_namePkybot              -1.574e+00  5.113e-01  -3.079 0.002078 **
threat_nameQadars               1.968e+00  2.450e-01   8.031 9.68e-16 ***
threat_nameQakbot              -5.242e-01  1.295e+00  -0.405 0.685654
threat_nameRamnit               1.098e+00  2.513e-01   4.370 1.24e-05 ***
threat_nameRamnit-BankerModule  1.262e+00  3.859e-01   3.270 0.001075 **
threat_nameReactorBot           7.411e-01  4.955e-01   1.496 0.134737
threat_nameRetefe-v2            1.286e+00  2.562e-01   5.020 5.18e-07 ***
threat_nameTheTrick             3.228e+00  2.453e-01  13.162  < 2e-16 ***
threat_nameTinba-v1             1.067e+00  2.493e-01   4.282 1.86e-05 ***
threat_nameTinba-v2             2.849e+00  2.448e-01  11.640  < 2e-16 ***
threat_nameZeuS                 2.791e+00  2.457e-01  11.361  < 2e-16 ***
threat_nameZeus-Action          6.100e-01  2.634e-01   2.316 0.020552 *
threat_nameZeuS-OpenSSL         3.630e+00  2.446e-01  14.844  < 2e-16 ***
threat_nameZeuS-P2P             3.677e+00  2.476e-01  14.854  < 2e-16 ***
threat_nameZeus-Panda           2.389e+00  2.452e-01   9.739  < 2e-16 ***
```

```
year2015                          1.100e+00  2.076e-02  52.976  < 2e-16 ***
year2016                          2.607e-01  2.028e-02  12.858  < 2e-16 ***
year2017                          6.794e-01  2.163e-02  31.409  < 2e-16 ***
unique_attackurl_count            1.087e-02  2.449e-04  44.386  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(1.6225) family taken to be 1)

    Null deviance: 84802  on 33383  degrees of freedom
Residual deviance: 32131  on 33298  degrees of freedom
AIC: 216737

Number of Fisher Scoring iterations: 1


              Theta:  1.6225
          Std. Err.:  0.0135

 2 x log-likelihood:  -216563.4640
```

A.4 Negative Binomial Regression - Unique Attack ID Count

```
Call:
glm.nb(formula = id_attack_count ~ langEnglish + langGerman +
    langFrench + langDutch + langItalian + langSpanish + langPortugese +
    langGreek + langCzech + langSlovak + langSlovenian + langPolish +
    langHungarian + langRomanian + langBulgarian + langDanish +
    langSwedish + langFinnish + langLatvian + langEstonian +
    langLithuanian + lang_count + auth1FA + auth2FA + pop_score +
    Country + threat_name + year + unique_attackurl_count,
    data = data_notzero, init.theta = 1.767001301, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.9040  -0.6102  -0.1625   0.2683  11.4126

Coefficients: (1 not defined because of singularities)
                                   Estimate Std. Error z value Pr(>|z|)
(Intercept)                      -4.946e-01  2.454e-01  -2.016 0.043834 *
langEnglishTrue                   1.931e-01  2.544e-02   7.590 3.19e-14 ***
langGermanTrue                    4.822e-03  3.277e-02   0.147 0.883021
langFrenchTrue                   -9.358e-02  4.218e-02  -2.219 0.026518 *
langDutchTrue                     3.957e-01  5.181e-02   7.638 2.20e-14 ***
langItalianTrue                  -2.563e-01  4.995e-02  -5.131 2.88e-07 ***
langSpanishTrue                   7.938e-01  6.095e-02  13.025  < 2e-16 ***
langPortugeseTrue                -8.685e-01  7.881e-02 -11.020  < 2e-16 ***
langGreekTrue                     1.816e-01  1.455e-01   1.248 0.212025
langCzechTrue                     4.732e-02  8.721e-02   0.543 0.587400
langSlovakTrue                    4.928e-02  9.546e-02   0.516 0.605659
```

```
langSlovenianTrue              -4.320e-01  1.351e-01  -3.197  0.001386 **
langPolishTrue                  1.703e-01  1.438e-01   1.184  0.236409
langHungarianTrue               1.579e-03  8.600e-02   0.018  0.985346
langRomanianTrue                7.145e-01  1.680e-01   4.252  2.12e-05 ***
langBulgarianTrue               1.708e-01  1.513e-01   1.129  0.258784
langDanishTrue                  4.537e-02  1.285e-01   0.353  0.723981
langSwedishTrue                 3.792e-01  1.177e-01   3.223  0.001269 **
langFinnishTrue                 3.299e-01  1.536e-01   2.147  0.031764 *
langLatvianTrue                 8.135e-02  7.989e-01   0.102  0.918888
langEstonianTrue               -2.490e-01  1.227e-01  -2.028  0.042512 *
langLithuanianTrue             -2.012e-02  7.990e-01  -0.025  0.979910
lang_count                     -2.782e-03  1.900e-02  -0.146  0.883602
auth1FATrue                    -3.747e-02  3.150e-02  -1.190  0.234206
auth2FATrue                    -1.064e-01  2.816e-02  -3.779  0.000158 ***
pop_score                       3.636e-07  2.566e-08  14.169   < 2e-16 ***
CountryBelgium                 -2.739e-01  7.330e-02  -3.737  0.000186 ***
CountryBulgaria                -3.293e-01  1.464e-01  -2.249  0.024491 *
CountryCroatia                 -1.481e-02  5.969e-02  -0.248  0.804077
CountryCyprus                  -3.486e-01  1.670e-01  -2.088  0.036784 *
CountryCzechia                  3.153e-01  9.274e-02   3.399  0.000676 ***
CountryDenmark                  6.379e-02  1.322e-01   0.482  0.629511
CountryEstonia                         NA         NA      NA        NA
CountryFinland                  1.466e-01  1.461e-01   1.004  0.315450
CountryFrance                  -6.957e-02  4.922e-02  -1.413  0.157540
CountryGermany                 -4.475e-02  2.984e-02  -1.500  0.133605
CountryGreece                  -1.195e-01  1.511e-01  -0.791  0.429107
CountryHungary                  3.159e-01  1.034e-01   3.054  0.002259 **
CountryIreland                  1.089e-01  6.613e-02   1.646  0.099706 .
CountryItaly                   -1.093e-01  5.748e-02  -1.902  0.057178 .
CountryLatvia                  -1.758e-01  8.034e-01  -0.219  0.826822
CountryLithuania               -1.758e-01  8.034e-01  -0.219  0.826822
CountryLuxembourg               1.371e-01  5.867e-02   2.337  0.019443 *
CountryMalta                    1.252e-01  3.664e-01   0.342  0.732492
CountryNetherlands             -2.335e-01  7.165e-02  -3.258  0.001121 **
CountryPoland                  -3.769e-02  1.435e-01  -0.263  0.792794
CountryPortugal                 5.311e-01  9.313e-02   5.703  1.18e-08 ***
CountryRomania                 -2.226e-01  1.664e-01  -1.338  0.180871
CountrySlovakia                -1.563e-01  9.032e-02  -1.730  0.083606 .
CountrySlovenia                 3.100e-01  1.913e-01   1.621  0.105046
CountrySpain                   -5.444e-01  6.517e-02  -8.354   < 2e-16 ***
CountrySweden                  -2.828e-01  1.188e-01  -2.380  0.017305 *
CountryUnited Kingdom          -1.469e-01  4.445e-02  -3.304  0.000953 ***
threat_nameCitadel              1.146e+00  2.393e-01   4.790  1.67e-06 ***
threat_nameCoreBot              8.650e-02  6.570e-01   0.132  0.895246
threat_nameDridex-Loader        2.515e+00  2.388e-01  10.532   < 2e-16 ***
threat_nameDyre                 3.814e+00  2.400e-01  15.893   < 2e-16 ***
threat_nameGootkit              1.730e+00  2.389e-01   7.242  4.41e-13 ***
threat_nameGootkitLoader        6.176e-01  2.409e-01   2.564  0.010353 *
threat_nameGozi-EQ              1.503e+00  2.407e-01   6.243  4.30e-10 ***
threat_nameGozi-ISFB            3.905e+00  2.389e-01  16.344   < 2e-16 ***
threat_nameIce9                 2.122e-01  2.496e-01   0.850  0.395207
threat_nameKINS                 2.346e+00  2.395e-01   9.797   < 2e-16 ***
```

```
threat_nameKronos               1.775e+00  2.398e-01   7.399 1.37e-13 ***
threat_nameMatrix              -5.130e-01  7.694e-01  -0.667 0.504866
threat_nameNuclearBot           8.166e-01  2.400e-01   3.403 0.000667 ***
threat_nameNymaim              -1.061e+00  1.060e+00  -1.001 0.316865
threat_namePkybot              -1.645e+00  5.002e-01  -3.288 0.001009 **
threat_nameQadars               1.566e+00  2.393e-01   6.545 5.96e-11 ***
threat_nameQakbot              -6.011e-01  1.274e+00  -0.472 0.637139
threat_nameRamnit               8.668e-01  2.458e-01   3.527 0.000420 ***
threat_nameRamnit-BankerModule  1.181e+00  3.688e-01   3.202 0.001366 **
threat_nameReactorBot           1.361e+00  4.437e-01   3.067 0.002160 **
threat_nameRetefe-v2            2.498e+00  2.477e-01  10.084  < 2e-16 ***
threat_nameTheTrick             3.715e+00  2.394e-01  15.520  < 2e-16 ***
threat_nameTinba-v1             1.424e+00  2.419e-01   5.887 3.93e-09 ***
threat_nameTinba-v2             8.064e-01  2.393e-01   3.370 0.000751 ***
threat_nameZeuS                 2.329e-01  2.407e-01   0.968 0.333225
threat_nameZeus-Action          2.327e-01  2.588e-01   0.899 0.368588
threat_nameZeuS-OpenSSL         2.303e+00  2.389e-01   9.642  < 2e-16 ***
threat_nameZeuS-P2P             8.582e-01  2.428e-01   3.534 0.000409 ***
threat_nameZeus-Panda           2.355e+00  2.394e-01   9.839  < 2e-16 ***
year2015                       -1.282e-01  2.220e-02  -5.775 7.70e-09 ***
year2016                       -5.313e-01  2.184e-02 -24.332  < 2e-16 ***
year2017                       -3.456e-01  2.290e-02 -15.095  < 2e-16 ***
unique_attackurl_count          1.214e-02  2.433e-04  49.872  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(1.767) family taken to be 1)

    Null deviance: 117087  on 33383  degrees of freedom
Residual deviance:  29822  on 33298  degrees of freedom
AIC: 185590

Number of Fisher Scoring iterations: 1


              Theta:  1.7670
          Std. Err.:  0.0158
Warning while fitting theta: alternation limit reached

 2 x log-likelihood:  -185415.5960
```

A.5 Summary of Regression Models

Table 4: Summary of logistic model and negative binomial model towards different metrics

	Dependent variable:			
	is_targeted	raw_attack_count	week_attack_count	id_attack_count
	logistic	*negative binomial*	*negative binomial*	*negative binomial*
	(1)	(2)	(3)	(4)
langEnglishTrue	0.629*	0.379***	0.124***	0.193***
	(0.314)	(0.029)	(0.025)	(0.025)
langGermanTrue	0.515	0.265***	−0.150***	0.005
	(0.385)	(0.038)	(0.032)	(0.033)
langFrenchTrue	0.355	0.010	−0.284***	−0.094*
	(0.455)	(0.049)	(0.042)	(0.042)
langDutchTrue	0.604	0.277***	0.364***	0.396***
	(0.561)	(0.060)	(0.051)	(0.052)
langItalianTrue	0.847	−0.046	−0.279***	−0.256***
	(0.639)	(0.058)	(0.049)	(0.050)
langSpanishTrue	−0.061	0.947***	0.537***	0.794***
	(0.625)	(0.072)	(0.061)	(0.061)
langPortugeseTrue	0.885	−0.228*	−0.677***	−0.868***
	(1.022)	(0.089)	(0.077)	(0.079)
langGreekTrue	1.160	0.004	−0.243	0.182
	(1.211)	(0.163)	(0.142)	(0.146)
langCzechTrue	0.097	0.308**	−0.026	0.047
	(1.106)	(0.099)	(0.085)	(0.087)

	Dependent variable:			
	is_targeted	raw_attack_count	week_attack_count	id_attack_count
	logistic	*negative binomial*	*negative binomial*	*negative binomial*
	(1)	(2)	(3)	(4)
langSlovakTrue	2.075	−0.488***	−0.051	0.049
	(1.074)	(0.108)	(0.093)	(0.095)
langSlovenianTrue	−1.987	−0.678***	−0.349**	−0.432**
	(1.817)	(0.147)	(0.125)	(0.135)
langPolishTrue	−1.573	0.557**	0.175	0.170
	(1.623)	(0.171)	(0.145)	(0.144)
langHungarianTrue	1.987	0.412***	0.062	0.002
	(1.561)	(0.100)	(0.084)	(0.086)
langRomanianTrue	−17.352	0.370	0.724***	0.714***
	(391.660)	(0.194)	(0.165)	(0.168)
langBulgarianTrue	11.961	0.275	0.161	0.171
	(1,484.644)	(0.173)	(0.149)	(0.151)
langDanishTrue	−0.543	0.214	0.114	0.045
	(1.197)	(0.152)	(0.129)	(0.128)
langSwedishTrue	2.146**	0.490***	0.318**	0.379**
	(0.793)	(0.138)	(0.118)	(0.118)
langFinnishTrue	−1.894	0.389*	0.139	0.330*
	(1.335)	(0.181)	(0.154)	(0.154)
langLatvianTrue	0.248	−1.212	−0.571	0.081

	Dependent variable:			
	is_targeted	raw_attack_count	week_attack_count	id_attack_count
	logistic	*negative binomial*	*negative binomial*	*negative binomial*
	(1)	(2)	(3)	(4)
	(1.568)	(0.862)	(0.903)	(0.799)
langEstonianTrue	-3.439^{***}	0.289^{*}	-0.557^{***}	-0.249^{*}
	(0.916)	(0.135)	(0.119)	(0.123)
langLithuanianTrue	15.970	−0.733	−0.630	−0.020
	(2,399.545)	(0.862)	(0.903)	(0.799)
lang_count	−0.035	-0.137^{***}	0.016	−0.003
	(0.258)	(0.022)	(0.019)	(0.019)
auth1FATrue	0.458	0.002	−0.015	−0.037
	(0.286)	(0.036)	(0.031)	(0.032)
auth2FATrue	2.148^{***}	-0.207^{***}	-0.173^{***}	-0.106^{***}
	(0.266)	(0.033)	(0.028)	(0.028)
pop_score	0.00000^{***}	0.00000^{***}	0.00000^{***}	0.00000^{***}
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
CountryBelgium	-1.844^{**}	-0.271^{**}	-0.211^{**}	-0.274^{***}
	(0.702)	(0.084)	(0.071)	(0.073)
CountryBulgaria	14.129	-0.588^{***}	-0.452^{**}	-0.329^{*}
	(1,369.951)	(0.168)	(0.145)	(0.146)
CountryCroatia	0.797	0.006	-0.186^{**}	−0.015
	(0.806)	(0.067)	(0.058)	(0.060)

	Dependent variable:			
	is_targeted	raw_attack_count	week_attack_count	id_attack_count
	logistic	*negative binomial*	*negative binomial*	*negative binomial*
	(1)	(2)	(3)	(4)
CountryCyprus	-1.879^{*}	-0.642^{***}	-0.536^{***}	-0.349^{*}
	(0.869)	(0.182)	(0.162)	(0.167)
CountryCzechia	-0.351	0.651^{***}	0.434^{***}	0.315^{***}
	(1.204)	(0.106)	(0.091)	(0.093)
CountryDenmark	-4.040^{**}	-0.115	-0.196	0.064
	(1.265)	(0.156)	(0.133)	(0.132)
CountryEstonia				
CountryFinland	1.552	0.017	-0.178	0.147
	(1.358)	(0.173)	(0.148)	(0.146)
CountryFrance	-0.678	-0.563^{***}	-0.252^{***}	-0.070
	(0.550)	(0.057)	(0.048)	(0.049)
CountryGermany	-0.095	-0.773^{***}	-0.310^{***}	-0.045
	(0.299)	(0.034)	(0.029)	(0.030)
CountryGreece	-2.406	-0.947^{***}	-0.462^{**}	-0.119
	(1.277)	(0.170)	(0.148)	(0.151)
CountryHungary	-5.971^{***}	0.056	0.158	0.316^{**}
	(1.576)	(0.121)	(0.102)	(0.103)

	Dependent variable:			
	is_targeted	raw_attack_count	week_attack_count	id_attack_count
	logistic	*negative binomial*	*negative binomial*	*negative binomial*
	(1)	(2)	(3)	(4)
CountryIreland	-3.671^{***}	-0.511^{***}	-0.146^{*}	0.109
	(0.512)	(0.078)	(0.066)	(0.066)
CountryItaly	-0.898	-0.157^{*}	-0.004	-0.109
	(0.703)	(0.066)	(0.056)	(0.057)
CountryLatvia	-1.435	1.179	0.308	-0.176
	(1.623)	(0.867)	(0.907)	(0.803)
CountryLithuania	-18.188	1.179	0.308	-0.176
	(2,399.545)	(0.867)	(0.907)	(0.803)
CountryLuxembourg	-2.082^{***}	-0.382^{***}	0.023	0.137^{*}
	(0.529)	(0.068)	(0.058)	(0.059)
CountryMalta	-3.517^{***}	-0.123	-0.374	0.125
	(1.032)	(0.395)	(0.376)	(0.366)
CountryNetherlands	-1.343	-0.551^{***}	-0.458^{***}	-0.233^{**}
	(0.694)	(0.084)	(0.071)	(0.072)
CountryPoland	-0.145	-0.880^{***}	-0.637^{***}	-0.038
	(1.632)	(0.171)	(0.144)	(0.143)
CountryPortugal	-6.810^{***}	0.570^{***}	0.222^{*}	0.531^{***}
	(1.073)	(0.105)	(0.090)	(0.093)
CountryRomania	16.885	0.169	-0.323^{*}	-0.223

	Dependent variable:			
	is_targeted	raw_attack_count	week_attack_count	id_attack_count
	logistic	*negative binomial*	*negative binomial*	*negative binomial*
	(1)	(2)	(3)	(4)
	(391.660)	(0.191)	(0.163)	(0.166)
CountrySlovakia	-1.136	-0.001	-0.302^{***}	-0.156
	(1.035)	(0.101)	(0.087)	(0.090)
CountrySlovenia	-0.733	0.579^{**}	0.023	0.310
	(2.003)	(0.208)	(0.179)	(0.191)
CountrySpain	0.351	-0.972^{***}	-0.612^{***}	-0.544^{***}
	(0.702)	(0.075)	(0.064)	(0.065)
CountrySweden	-3.840^{***}	-0.584^{***}	-0.417^{***}	-0.283^{*}
	(0.867)	(0.140)	(0.120)	(0.119)
CountryUnited Kingdom	-1.060^{*}	-0.477^{***}	-0.323^{***}	-0.147^{***}
	(0.444)	(0.052)	(0.044)	(0.044)
threat_nameCitadel		4.146^{***}	2.288^{***}	1.146^{***}
		(0.277)	(0.245)	(0.239)
threat_nameCoreBot		1.141	0.562	0.087
		(0.736)	(0.675)	(0.657)
threat_nameDridexLoader		5.480^{***}	3.799^{***}	2.515^{***}
		(0.277)	(0.245)	(0.239)
threat_nameDyre		4.506^{***}	3.135^{***}	3.814^{***}

	Dependent variable:			
	is_targeted	raw_attack_count	week_attack_count	id_attack_count
	logistic	*negative binomial*	*negative binomial*	*negative binomial*
	(1)	(2)	(3)	(4)
		(0.278)	(0.246)	(0.240)
threat_nameGootkit		3.607***	2.415***	1.730***
		(0.277)	(0.245)	(0.239)
threat_nameGootkitLoader		2.107***	0.952***	0.618*
		(0.278)	(0.247)	(0.241)
threat_nameGozi-EQ		4.544***	2.619***	1.503***
		(0.279)	(0.246)	(0.241)
threat_nameGozi-ISFB		5.040***	3.594***	3.905***
		(0.277)	(0.245)	(0.239)
threat_nameIce9		4.262***	2.040***	0.212
		(0.286)	(0.253)	(0.250)
threat_nameKINS		5.393***	2.965***	2.346***
		(0.278)	(0.245)	(0.239)
threat_nameKronos		3.029***	2.097***	1.775***
		(0.278)	(0.246)	(0.240)
threat_nameMatrix		1.815*	0.216	−0.513
		(0.776)	(0.733)	(0.769)
threat_nameNuclearBot		1.185***	0.609*	0.817***
		(0.278)	(0.246)	(0.240)

	Dependent variable:			
	is_targeted	raw_attack_count	week_attack_count	id_attack_count
	logistic	*negative binomial*	*negative binomial*	*negative binomial*
	(1)	(2)	(3)	(4)
threat_nameNymaim		1.232	1.068	−1.061
		(1.057)	(0.872)	(1.060)
threat_namePkybot		−0.707	-1.574^{**}	-1.645^{**}
		(0.531)	(0.511)	(0.500)
threat_nameQadars		3.800^{***}	1.968^{***}	1.566^{***}
		(0.277)	(0.245)	(0.239)
threat_nameQakbot		0.599	−0.524	−0.601
		(1.174)	(1.295)	(1.274)
threat_nameRamnit		2.128^{***}	1.098^{***}	0.867^{***}
		(0.284)	(0.251)	(0.246)
threat_nameRamnit-BankerModule		2.090^{***}	1.262^{**}	1.181^{**}
		(0.443)	(0.386)	(0.369)
threat_nameReactorBot		1.812^{***}	0.741	1.361^{**}
		(0.544)	(0.495)	(0.444)
threat_nameRetefe-v2		4.241^{***}	1.286^{***}	2.498^{***}
		(0.289)	(0.256)	(0.248)
threat_nameTheTrick		4.458^{***}	3.228^{***}	3.715^{***}
		(0.278)	(0.245)	(0.239)

	Dependent variable:			
	is_targeted	raw_attack_count	week_attack_count	id_attack_count
	logistic	*negative binomial*	*negative binomial*	*negative binomial*
	(1)	(2)	(3)	(4)
threat_nameTinba-v1		2.225***	1.067***	1.424***
		(0.280)	(0.249)	(0.242)
threat_nameTinba-v2		4.080***	2.849***	0.806***
		(0.277)	(0.245)	(0.239)
threat_nameZeuS		4.770***	2.791***	0.233
		(0.278)	(0.246)	(0.241)
threat_nameZeus-Action		3.865***	0.610*	0.233
		(0.293)	(0.263)	(0.259)
threat_nameZeuS-OpenSSL		7.311***	3.630***	2.303***
		(0.277)	(0.245)	(0.239)
threat_nameZeuS-P2P		6.381***	3.677***	0.858***
		(0.281)	(0.248)	(0.243)
threat_nameZeus-Panda		4.229***	2.389***	2.355***
		(0.278)	(0.245)	(0.239)
year2015		1.736***	1.100***	−0.128***
		(0.025)	(0.021)	(0.022)
year2016		0.794***	0.261***	−0.531***
		(0.023)	(0.020)	(0.022)

	Dependent variable:			
	is_targeted	raw_attack_count	week_attack_count	id_attack_count
	logistic	*negative binomial*	*negative binomial*	*negative binomial*
	(1)	(2)	(3)	(4)
year2017		1.301^{***}	0.679^{***}	-0.346^{***}
		(0.025)	(0.022)	(0.023)
unique_attackurl_count		0.009^{***}	0.011^{***}	0.012^{***}
		(0.0003)	(0.0002)	(0.0002)
Constant	-0.856	-2.704^{***}	-1.560^{***}	-0.495^{*}
	(0.486)	(0.284)	(0.251)	(0.245)
Log Likelihood	−1,845.873	−156,747.100	−108,282.700	−92,708.800
θ		1.033^{***} (0.007)	1.623^{***} (0.014)	1.767^{***} (0.016)
Akaike Inf. Crit.	3,795.746	313,666.200	216,737.500	185,589.600
Note:				*p<0.05; **p<0.01; ***p<0.001
				Standard errors in brackets