Crime in the Cloud: An Analysis of the Use of Cloud Services for Cybercrime

Daan Kolthof
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands
d.kolthof@student.utwente.nl

ABSTRACT
With the rise of cloud computing services, cybercriminals discover new and improved ways of conducting cybercrime, using cloud services as their instrument of choice. This paper presents an overview of the misuse of cloud computing services for several malicious purposes, through a literature study and through studying real examples of misuse. Furthermore, a quantification of misuse of cloud services will be presented, for the specific case of email spam. Our study shows that there are major differences in the percentage of IP addresses operated by cloud providers that are listed in the PSBL blacklist.

Keywords
Cloud computing, security, cybercrime, DNS blacklist, Passive Spam Block List (PSBL)

1. INTRODUCTION
“Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction” [20] The use of cloud computing providers companies or individuals with more elasticity for their computing needs. Cloud computing services are services provided by so called cloud providers. Different service models are provided to individuals and companies. The most important difference between cloud service models is the level of responsibility over the underlying hardware. Service models range from providing a certain type of software to providing just the infrastructure and hardware, each of these models offering more responsibility to the customer [20].

With the rise of cloud computing services, cybercriminals discover new and improved ways of conducting cybercrime, using cloud computing services as their instrument of choice. The differences in responsibilities in the service models described earlier allow cybercriminals the opportunity to perform their activities while minimizing the required effort. Examples of malicious use of cloud computing services include sending massive amount of spam [16], using the reputation of cloud providers to deceive firewalls [13] and deploying botnet command-and-control servers [13].

Cybercrime itself is well-known and well-researched. However, information on the relation between the use of cloud computing services and cybercrime is scarce, although important for coming to an understanding of the effects of widely-available cloud computing services on modern-day cybercrime. This paper will present both an overview and a quantification of the current situation of misuse of cloud computing services for several malicious purposes. The overview, which has been researched through literature study and through researching real examples of misuse, will present information about the threats of misuse of cloud computing services. The quantification will provide information about a specific case of malicious use of cloud services, which is the sending of spam. The choice for this type of cybercrime depended on the information available, namely the Passive Spam Block List (PSBL) IP blacklist [5]. This blacklist contains a list of IP addresses that are known for sending spam to recipients controlled by the administrator of the blacklist. The IP addresses in the dataset will be compared to collected, known IP address ranges used by several cloud service providers.

The goal of the research is to answer the research question presented below:

To which extent are cloud services being used for cybercrime?

This main research question can be addressed by answering the following sub-questions.

- Which types of cybercrime are performed using cloud services?
- Can the dataset clarify whether cloud services are being misused for committing cybercrime?
- How does the malicious use of cloud services evolve over time?

The first sub-question will be answered by means of a literature study, which a focus on real examples of misuse of cloud services. The answers to this question will be the overview of the current situation of misuse of cloud services for cybercrime. The second and third sub-questions will be answered through analysis of the dataset. The answers to the second and third sub-questions will result in a quantification of the current situation of misuse of cloud services for cybercrime. In this paper, section 2 will discuss work related to this paper. In section 3, the different types of cybercrime performed using cloud services are discussed. Section 4 contains information about the dataset and the research methodology. The rest of the research questions
are answered in section 5, which describes whether cloud services are being misused for cybercrime, and section 6, which presents the evolution of malicious use of cloud services over time. Section 7 will discuss the findings and section 8 contains the conclusion of the research.

2. RELATED WORK

Research has been done in the area of cloud computing security. In this section, research that this paper builds on will be described and discussed, as well as research similar to the research conducted for this paper will be discussed. A collection of some noteworthy papers will be presented, including their relevance and importance to the research presented in this paper.

Providing an insight in the lifetime of IP addresses in the PSBL database is necessary to help understand the duration that an individual account or node in the cloud is performing malicious activities. The research on IP blacklists done by Dietrich and Rossow [14] provides an insight into the duration of the activity period of malicious hosts sending spam. Although this paper provides a recommendation of the duration an IP address should be listed in spam databases, information about the actual listing duration of an IP address in a spam database is not described and is discovered during the research.

Information on the methods of querying the PSBL database, as described in the research done by Jung and Sit [15] will be used to briefly explain the working of the database. However, this information will not be used for conducting the analysis. The reason for this is that snapshots of the database are used, eliminating the need for DNS queries described in the discussed research.

Several papers about cloud service security and misuse that are similar to the research presented in this paper have been written. However, these papers do not address quantification of cloud service misuse, therefore not providing quantitative conclusions, which will be presented in this research. The research done by Mohd and Tarakji [12] provides information on risks, threats and vulnerabilities to execute malicious programs [17]. The advantage cloud computing brings to malware hosting is scalability of the capacity of the host, that is, the amount of data the host is able to serve. Another advantage, which is not necessarily a general property of cloud computing, is the abuse of reputation of popular cloud service providers. The abuse of the reputation of these cloud service providers can lead to a delayed listing in blacklists and can delude other reputation based systems.

3. TYPES OF CYBERCRIME PERFORMED USING CLOUD SERVICES

The properties of cloud computing give way to more and different types of cybercrime to be performed using cloud services. Many of these types of cybercrime existed before they were performed using cloud services. In the case of cloud computing, cybercriminals can abuse these services in two different ways. The first way is to use rented servers, provided by the cloud provider, to perform cyber-attacks. Another way is to compromise or in any other way misuse cloud services rented by others to perform their attack. Both of these methods result in the cybercriminal not being affected by the consequences of using these services for the malicious activities. The only parties affected are the cloud service providers, which own the physical machines [20] and the legitimate services users when their machines are compromised or in any other way misused. This section will discuss different types of cybercrime, including examples of real misuse if found during the research. Furthermore, influence of cloud computing will be discussed, including any advantages of performing the cybercrime by using cloud services.

3.1 Malware hosting

Malware hosting is a broad term that entails hosting of malicious programs, exploits and other types of malicious software. Examples of malicious programs are Trojans, hosted through websites promising users useful applications. These websites may dupe users into executing the hosted malware. Another example of malware hosting consists of hosting exploits to attack users without the need for the user knowingly interacting with the malicious website or host. The only thing the user needs to do is connect using a program that is vulnerable to an exploit used by a malicious host. Consider the example of drive-by downloads, these can be classified as an exploit because they make use of vulnerabilities to execute malicious programs [17].

The advantage cloud computing brings to malware hosting is scalability of the capacity of the host, that is, the amount of data the host is able to serve. Another advantage, which is not necessarily a general property of cloud computing, is the abuse of reputation of popular cloud service providers. The abuse of the reputation of these cloud service providers can lead to a delayed listing in blacklists and can delude other reputation based systems.

3.2 Non-malware malicious hosting

Malicious hosting is not restricted to hosting just malware. Malicious hosting also includes hosting of malicious websites that are not meant to infect its visitors. Examples of this type of malicious activity are hosting websites to scam people or phishing websites.

The advantages of the use of cloud computing for this type of cybercrime are mostly the same as the advantages described for malware hosting. This is because the technique required is (almost) the same. The attackers needs only to host a website or other way of communication a user can connect to [19].

3.3 Sending spam

One of the most known types of spam is email spam, which is discussed throughout this paper. Email spam involves sending messages to recipients by email. These emails may contain links sending users to phishing websites or websites hosting malware. Other spam emails may contain malware or unsolicited commercials [11].

To recognize emails as spam, many email clients and email services utilize so called spam filters. To a certain degree, these spam filters are able to filter out the unsolicited messages from incoming email. These spam filters often consist of statistical techniques for filtering [9], configuration of user preferences and reputation based filtering methods [7].

While spam itself is nothing new, the use of cloud computing services for sending spam can provide spammers with some advantages. An advantage of using cloud computing services for sending spam is its scalability. Spammers have the opportunity to instantly scale their spam sending operations using cloud services. Another impor-
A botnet is a collection of compromised computers, called bots, remotely controlled by its originator, known as a “bot master”. Controlling botnets can be as uncomplicated as utilizing an HTTP-based website or the Internet Relay Chat (IRC) protocol. Servers used to control these botnets are called command-and-control servers [23]. The advantage of using cloud services to perform this type of cybercrime is the scalability of cloud services. In this case, scalability refers to the number of devices, known as bots, that the command-and-control server is able to manage. Another advantage, which has also been mentioned earlier, is the ability to abuse the reputation of the cloud service providers. Abuse of the reputation of cloud service providers might lead to delayed blacklisted and might delude other reputation based systems. In the case of botnet command-and-control servers, the benefits of cloud computing outweigh the risks and dangers, if the botnet programmer took in account the risk of detection before sending spam, which led to blocks of IP addresses belonging to Amazon’s EC2 being blacklisted on multiple spam databases [16].

4. DATASET AND METHODOLOGY

Before an insight in the differences in malicious use of the cloud platforms and the insight in the evolution of the malicious use of cloud services is provided, an insight in the dataset of for the research is presented. Furthermore, the methodology used to analyze these datasets is presented.

4.1 Exploring the dataset

To provide an insight in the different datasets used for the research, an analysis of these datasets will be presented. The datasets are the collection of snapshots of the Passive Spam Block List (PSBL) [5] and a list of publicly known IP address range used by some of the cloud providers for their services. The PSBL is an online blacklist that lists IP addresses known to send email spam. Snapshots of this blacklist will be used for the research. The snapshots are hourly representations of the PSBL database. The IP address in this dataset have been used to send spam to recipients controlled by the administrator of the PSBL database. The second dataset used for this research are the known ranges of IP addresses of the chosen cloud service providers. These known ranges of IP addresses are also representations of the IP address ranges at a given time. Meaning that whenever these companies publish new or different IP address ranges, the changed IP addresses are included in the dataset. These IP address ranges are published on the websites of the cloud providers. These websites are checked and when there are changes in the IP address ranges, the new ranges are added to the dataset, including the range of dates on which this new IP range applies.

The cloud services providers chosen to analyse are Amazon, Microsoft and Google. The reason for analysing these cloud providers is that these are some of the biggest and most well-known companies and cloud service providers. Furthermore, these companies publish the range of IP addresses used for their cloud services. These IP addresses are needed for parts of the research. These companies are all providing public cloud services that present users with an opportunity to abuse these services for sending email spam. Other ways of abuse of these services can be the misuse of provided software services, called Software-as-a-Service, to send email spam [20].

In both of the presented datasets, the data is only valid for a certain amount of time. To compare these datasets, only the dates in which the datasets overlap can be used for the comparison. This period ranges from February 13th 2015 to April 28th 2015.

4.1.1 Contents of the datasets

The database snapshots consist of lists of IP addresses blacklisted through the PSBL. An indication of the size of the database snapshots, is now presented in Figure 1, displaying the number of IP addresses per database snapshot.

Figure 2 displays the total amount of IP addresses used at any particular time by the cloud providers. Only a change in the IP addresses is displayed.
4.1.2 Lifetime of IP addresses in PSBL

An important metric of the PSBL is the lifetime of a single IP address, which is the time between addition and removal of an IP address in the PSBL. Knowing the lifetime of an IP address in the database can help to understand the duration that the node connected using that IP address is performing the malicious sending of spam. Furthermore, information about the lifetime can provide an insight in the database itself, it can indicate how quickly IP addresses are removed from the database.

The research on IP black-lists done by Dietrich and Rossow [14] provides an insight into the duration of the activity period of malicious hosts sending spam. Furthermore, information about the lifetime can provide an insight in the database itself, it can indicate how quickly IP addresses are removed from the database. The PSBL website claims to remove IP addresses from the blacklist when the IP address has not sent any spam to the spamtraps for a few weeks. The PSBL also offers to possibly of manually removing any IP address listed, however less than 1% of the IP addresses are removed that way [6]. This claim does not match the measured average time until removal, suggesting that IP addresses are removed way faster. The reason for this might be a different configuration for the blacklisting system than claimed on the website. The blacklisting system using for PSBL is called Spamkaze and has the possibility of configuring the time a host is blacklisted after the first spam email arrived, and the time for each subsequent spam mail received [8]. These settings influence the average time until removal that was measured.

4.2 Finding listed hosts of cloud providers in PSBL blacklist

The core method for obtaining the results in the research is the matching of IP addresses used by cloud providers to the IP addresses in the PSBL blacklist. This algorithm will have as input the PSBL database snapshots and the known IP address ranges of different cloud providers. The algorithm will iterate over the IP addresses listed in each database snapshot. The IP addresses will then be compared to the IP address ranges of the cloud providers. The output of this algorithm is an overview of the amount of IP address matches per cloud provider per snapshot.

This algorithm is used in the section discussing whether cloud services are misused for cybercrime and the section on the evolution of misuse over time.
5. CLOUD SERVICES MISUSED FOR CYBERCRIME

This section will clarify whether cloud services are being misused for committing cybercrime at all. This is done through an analysis of the PSBL database and the list of known IP address ranges for the chosen cloud providers. This analysis is described in the methodology part of the research.

After performing the comparison between the database snapshots of the PSBL and the collection of IP addresses of the cloud providers, as described in the methodology, matches were found between all of the cloud providers’ IP addresses and some of the listed IP addresses in some of the snapshots. Because matches have been found, it can be concluded that at some point in time, spam has been sent from the machines used by the examined cloud service providers, assuming the database is reliable.

On the first snapshot taken, Amazon had 63 IP addresses listed in the PSBL database, accounting for 0.08% of the total number of IP addresses in the dataset. There were no matches found for IP addresses belonging to Google. Microsoft had 164 matches, which is 0.32% of the total snapshot size.

6. EVOLUTION OVER TIME

Comparing the snapshots of the PSBL database to the IP address ranges can provide an overview of the misuse of cloud services and its evolution. The method for comparing these parts of the dataset is described in the methodology part of the research.

Presenting an overview of the amount of IP address matches per cloud provider relative to the total amount of IP addresses used by cloud providers can give an objective overview of the misuse of the services of these providers. This overview of the relative number of IP address matches is presented now.

The overview presented in Figure 4 can be used to compare the different cloud service providers. Figure 4 and Table 1 show that Amazon and Microsoft have on average 0.1% and 0.2% of their total amount of IP addresses listed in the database during the measured period. To give an idea of the percentage of IP addresses used by cloud services in the complete blacklist, another overview is presented. This overview presents the number of IP address matches per provider, divided by the total amount of IP addresses in each blacklist snapshot. This overview is presented in Figure 5. Furthermore, the absolute number of IP matches in the blacklists snapshots is presented in Figure 6. In the presented figures, some gaps can be seen. The reason for these gaps is the absence of some database snapshots. The largest gap of missing data is the one between March 27th and April 2nd. Apart from this large gap, gaps of a few hours appear in the dataset.
Figure 6 shows the absolute number of IP addresses of the cloud providers that are listed in the PSBL blacklist at the given times. As can be seen in the figure, there are no listings of any machines belonging to the cloud service of Google, meaning no spam has been sent from machines belonging to cloud services provided by Google to the PSBL spam traps. The reason for the absence of any sent spam is that Google restricts sending email from their machines, unless the Google cloud support team is contacted first [3]. Amazon and Microsoft apply the same kind of methods for blocking spam being sent from their machines. Amazon has a permission-based system in place, which means that there is a limit in the amount of emails that can be sent from their machines. After this limit is reached, customers have to contact the Amazon support before being allowed to send any more email from their machines [1]. The services offered by Microsoft do not support sending email from their machines, however, Microsoft does offer Infrastructure-as-a-Service, enabling customers to set up their own email server [4].

To present a comparison of the cloud providers, Figure 4 has to be examined. The average percentage of IP addresses blacklisted and the standard deviation listed in Table 1 can be used to compare the cloud providers. As seen, the percentage of IP addresses, used by Microsoft, listed in the PSBL at any given time is on average two times as much as Amazon's percentage of blacklisted IP addresses.

The difference between the amount of blacklisted IP addresses of Amazon and Microsoft can be explained by looking at the policies of these two companies on sending email from their clouds. Amazon has a clear policy on sending email from their machine. Furthermore, there is a clear limit on the amount of email that may be sent before support has to be contacted. Microsoft does not have a clear policy, except that their services do not support sending email by default.

Another reason for Microsoft's higher blacklist percentage might be the reaction of these providers. Amazon has been criticized in the past for not reacting swiftly on abuse notices [16], which might have resulted in a change on their approach on clients sending spam. This could result in Amazon being less interesting to spammers compared to Microsoft.

Another explanation for the presented difference can be the price difference in use of services of these cloud providers. Furthermore, there might be a difference in the use of these cloud services among the legitimate users. Legitimate users can have compromised applications or services on their machines, which may send emails spam and thus result in blacklisting of IP addresses of these users.

8. CONCLUSION

With the rise of cloud computing services, cybercriminals discover new and improved ways of conducting cybercrime, using cloud computing services as their instrument of choice. An overview and quantification of the use of cloud services for cybercrime has been presented in this paper.

The main research questions answered in this paper is: “To which extent are cloud service being used for cybercrime?”. This question is answered by looking at the sub-questions. These sub-questions, which are answered through literature study, study of real examples of misuse and quantification of misuse of cloud services, are listed now.

The question on which types of cybercrime are performed using cloud services is answered by providing a literature study and a study of real examples of misuse. The analysis presented to answer this question lists some of the advantages and disadvantages of using cloud services to perform some of the discussed types of cybercrime.

The second sub-question, which aims to answer the question whether the dataset used for this research, the PSBL database, can be used to clarify whether cloud services are being misused for committing cybercrime. The relevant section of this paper shows that traces of malicious use, in this case IP listings on the blacklist, can be found in the datasets.

The last question, which is about the evolution and differences between cloud services providers, is answered in the section on the evolution of blacklisted IP addresses. There is a noticeable difference in the percentage of IP addresses of cloud providers that are listed in the PSBL database, as seen in Figure 4. In both the absolute and the relative number of matches, Microsoft on average has more listings than Amazon and Google. Google has no listings at all in the spam database during the given dates. The different figures are able to quantify the situation of misuse of cloud services for cybercrime in different ways. Depending on which information is needed, the different figures can be inspected.

The major differences in the email sending policies of the cloud providers were unforeseen, as was the influence of these policies on malicious use of their platforms.

To gain an insight in the causes of the discovered email spam, research has to be done on whether the blacklisted hosts were rented for malicious purposes or were compromised prior to being misused for cybercrime. Suggestions for further research also include analysing more and different IP blacklists, which can provide more information about the sending of spam, as well as information about other types of cybercrime.

9. REFERENCES

APPENDIX

A. ALGORITHMS

A.1 Lifetime of IP addresses in the PSBL

The following algorithm is used to determine the lifetime of the IP addresses in the PSBL blacklist. The input of this algorithm is the collection of snapshots of the database. This algorithm outputs, for each IP address found in the snapshots, the individual number of hours the IP address is listed before removal. These listings before removal can happen multiple times and thus the algorithm can output multiple lifetimes per IP address.

\[ C \leftarrow \emptyset; \]

**for every snapshot except the very first**

**for every IP address in the snapshot**

**if IP address has been seen previous snapshot**

**then**

**if IP \subseteq C**

\[ C[IP].\text{count} \leftarrow C[IP].\text{count} + 1; \]

**else**

\[ IP \leftarrow IP \cup \{C\}; \]

**end**

**else**

save count of IP address;

\[ C[IP].\text{count} \leftarrow 0; \]

**end**

**end**

**for (\forall IP: IP \subseteq C)**

save count of IP if there is still counting going on;

output IP address and all its counts;

**end**