# The Role of Cloud Services in Malicious Software: Trends and Insights

Xiao Han[1,2(✉)], Nizar Kheir[1], and Davide Balzarotti[2]

[1] Orange Labs, Issy Les Moulineaux, France
{xiao.han,nizar.kheir}@orange.com
[2] EURECOM, Sophia Antipolis, France
davide.balzarotti@eurecom.com

**Abstract.** In this paper we investigate the way cyber-criminals abuse public cloud services to host part of their malicious infrastructures, including exploit servers to distribute malware, C&C servers to manage infected terminals, redirectors to increase anonymity, and drop zones to host stolen data.

We conduct a large scale analysis of all the malware samples submitted to the Anubis malware analysis system between 2008 and 2014. For each sample, we extracted and analyzed all malware interactions with Amazon EC2, a major public cloud service provider, in order to better understand the malicious activities that involve public cloud services. In our experiments, we distinguish between benign cloud services that are passively used by malware (such as file sharing, URL shortening, and pay-per-install services), and other dedicated machines that play a key role in the malware infrastructure. Our results reveal that cyber-criminals sustain long-lived operations through the use of public cloud resources, either as a redundant or a major component of their malware infrastructures. We also observe that the number of malicious and dedicated cloud-based domains has increased almost 4 times between 2010 and 2013. To understand the reasons behind this trend, we also present a detailed analysis using public DNS records. For instance, we observe that certain dedicated malicious domains hosted on the cloud remain active for an average of 110 days since they are first observed in the wild.

## 1 Introduction

Public infrastructure-as-a-service (IaaS) clouds have rapidly expanded in the recent years, with almost half of US businesses now using cloud computing in some capacity [10]. IaaS offer a straightforward *pay-as-you-go* pricing model where users dynamically create virtual machines at will, provide them with public IP addresses and on-demand compute and storage resources, and then delete them without any sustainable cost. Major providers of IaaS clouds, such as Amazon EC2 [2] and Microsoft Azure [4], also propose scalable services and default configuration options that contributed to the wide adoption of cloud services.

Unfortunately, the rapid growth of cloud services has also attracted cyber-criminals, paving the way to an active underground economy. As a result,


M. Almgren et al. (Eds.): DIMVA 2015, LNCS 9148, pp. 187–204, 2015.
DOI: 10.1007/978-3-319-20550-2_10

Los et al. [17] list the abuse of cloud services among the top nine critical threats to cloud computing. In fact, public IaaS clouds provide users with virtually unlimited network, compute, and storage resources. These are coupled with weak registration processes that facilitate anonymity, and so anyone with a valid credit card can easily register and use cloud services. For example, an early case of cloud service abuse was publicly uncovered in 2009, where a Zeus command and control (C&C) server was found to be hosted on Amazon EC2 [11]. More recent examples include the SpyEye banking trojan that was found to be using Amazon S3 storage [7], Android malware that exploited the Google Cloud Message service [21], and more advanced persistent attacks that used Dropbox and Wordpress services as a cover [14]. Despite these multiple examples, we are unaware of any existing study that measures the extent at which public cloud services are being abused by cyber-criminals. Such study would advise the design and implementation of future cloud monitoring and accountability services. More precisely, we do not know if cyber-criminals use cloud-based servers only as redundant components of their malware infrastructure, or whether they *specifically* use cloud services to achieve a better sustainability. Besides, we do not know if public clouds add more resilience to malware infrastructures, what is the time it takes to detect a malicious server hosted on a public cloud, as well as the time required to take down this server after it was first discovered.

In this paper we present a framework to measure and analyze malicious activity that involves public cloud services. Unlike previous work that actively probed public cloud IP addresses [22] or only use passive DNS records [13], we directly collect malware communications by analyzing the network traffic recorded by the Anubis dynamic analysis system [6]. Anubis is a publicly accessible service that analyzes malware samples in an instrumented sandbox. As part of its analysis, the system also records which domains and IP addresses are contacted by each malware sample, and part of the data that is transferred through the connection. Unfortunately, malware can communicate with the cloud for multiple reasons, including malicious activities but also other innocuous connections which range from simple connectivity checks, to the use of public benign services. This greatly complicated the analysis, and required the development of several heuristics to discard malware samples using public services hosted on the cloud, as this cannot be considered an abuse of the cloud itself.

In our experiments, we analyzed the network communication of over 30 million samples, submitted between 2008 and 2014. Our system identified 1.08 million (roughly 3.6 %) that connected to at least one publicly routable Amazon EC2 IP address. These IPs were associated to 12,522 distinct cloud-based domains. Interestingly, we observed that over the same period, only 32,225 samples connected to Microsoft Azure. Due to the relatively low number of samples that interacted with Azure, we only focused our study on Amazon EC2.

To summarize, the paper makes the following contributions:

– We present the first systematic, large scale study on the use of public cloud services by malicious software.

- We perform a precise categorization of each cloud access, separating the cases in which the malware samples simply rely on legitimate services which happen to be hosted on the cloud, from the cases in which part of the malware infrastructure is hosted on the cloud.
- We study the evolution of cloud adoption over the past six years, and identify an increasing trend that affects many different categories of malware.
- We present some general observations and insights about the global picture. For instance, while the recent efforts towards enhancing malware detection capabilities have contributed to considerably reduce the detection time for malicious hosts and domains, we were unable to observe a similar effect in the domains pointing to machines hosted on the cloud.

The rest of the paper is structured as follows. Section 2 provides an overview of our approach. We then describe our experiments in Sect. 3, and summarize and discuss the major findings in Sect. 4. Finally, Sect. 5 presented an overview of the related work in the area and Sect. 6 concludes the paper.

## 2 Approach

To identify malicious servers hosted on Amazon EC2, we first collected the range of IP addresses assigned to the cloud images, as reported by the Amazon website[1]. Moreover, to account for possible yearly changes, we also retrieved previous versions of the page from the web archive project[2]. We then extracted and analyzed the network traffic generated by all the malicious samples that have been collected and executed in Anubis, a popular malware analysis sandbox [6], over the past six years.

The main goal of our study is to verify the way miscreants make use of cloud services, whether they specifically target cloud infrastructures, and measure the time it takes for the provider to detect and drop malicious services hosted on EC2. To do so, our system tracks all domain names associated with the EC2 IP addresses that were contacted at least once by a malicious sample. Then, it further extracts and analyzes the DNS features and the content of network communications between the malware and the EC2 machines.

A major challenge in our study is that domain names extracted from the Anubis database do not only include dedicated malicious servers, and so we cannot simply mark as suspicious every connection toward a cloud-based IP address. In fact malware often contacts other public and benign cloud-based services, such as IP lookup services, advertisement websites, and URL shortening. These services are not part of the malicious activity and therefore need to be identified and discarded from our subsequent analysis.

On the other hand, real malicious domains may have been *sinkholed* by security organizations at the time the malware was analyzed in Anubis. Malware will be thus redirected towards sinkhole services that are sometimes hosted on EC2,

[1] http://docs.aws.amazon.com/general/latest/gr/aws-ip-ranges.html.
[2] https://archive.org/web/.

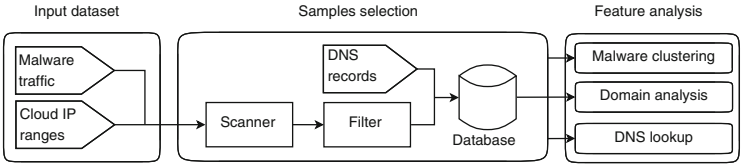



**Fig. 1.** Architecture of our platform

even though the original domains may have not been hosted on the cloud. Our system filters these cases and does not consider them as cloud-related malicious activities. Finally, in our experiments we discovered that many malware samples were *adwares* that leverage pay-per-install (PPI) services hosted on Amazon or other cloud providers. PPI services allow software publishers to pay affiliates by the number of installations that they perform on target machines. Caballero et al. analyze in [8] the *modus-operandi* of PPI services and measure their use for malware distribution. Although the use of PPI services to distribute malware still constitutes a malicious activity, PPI services are not malicious *per-se*, and so they need to be discarded from our dataset as we only focus in this paper on dedicated malicious services that were hosted on EC2.

### 2.1 Platform Description

To setup our experiments, we designed and implemented the platform illustrated in Fig. 1. Our system consists of two main components: the samples selection and the feature analysis modules. The first extracts from the Anubis database all malware samples that exhibited at least one network connection towards the Amazon cloud. During the period of our study, we identified 1.08 million malware samples that satisfied this criterion. The samples selection module further discards samples that have contacted benign public services hosted on EC2, and *keeps only dedicated malicious services* as input to the feature analysis module. Finally, the feature analysis module classifies the remaining malware samples and analyzes their dedicated malicious services hosted on cloud.

**Samples Selection.** The samples selection module aims at building a database of malware samples that, during their analysis, connected to malicious services hosted on EC2 – as well as the domain names or IP addresses that were associated with these services.

*Malware Scanner:* This module first extracts from Anubis all malicious samples that interacted with EC2 machines. We seed this module with the list of publicly routable IP ranges that were associated to the Amazon cloud in the year in which the analysis was performed. During the six years of our study, we identified $1,079,318$ distinct samples that connected to EC2.

The first thing we noticed in our experiments is that a large number of samples in our dataset were executables that leveraged pay-per-install (PPI)

**Table 1.** Top 20 PPI services in our dataset

| PPI Domain name | Samples | PPI Domain name | Samples |
|---|---|---|---|
| getapplicationmy.info | 116306 | torntv.net | 16578 |
| sslsecure1.com | 71965 | powerpackdl.com | 15586 |
| oi-imp1.com | 68255 | oi-config3.com | 15578 |
| secdls.com | 52857 | webfilescdn.com | 14050 |
| oi-config1.com | 43526 | torntvz.com | 12440 |
| ppdserver.com | 39434 | premiuminstaller.com | 11879 |
| optimum-installer.com | 38777 | ppdistro.us | 10463 |
| optimuminstaller.com | 35510 | bestringtonesmaker.com | 10136 |
| leadboltapps.net | 31918 | baixakialtcdn2.com | 9946 |
| xtrdlapi.com | 18615 | oi-config2.com | 9601 |

services hosted on EC2. PPI services have recently emerged as a key component of modern cybercrime and miscreants often refer to these services to outsource the global distribution of their malware. They supply PPI services with malware executables, which in turn charge them for successful installations based on the requested features for the desired victims. PPI service providers operate directly or through affiliate programs. They develop downloaders that retrieve and run the requested software (possibly malware) upon execution on the victim computer.

To identify PPI downloaders in our dataset, we refer to multiple public sources such as PPI forums [1] and public PPI web sites. The main challenge in our case was to identify the different PPI brands, since there are new brands that constantly appear over time. In order to address this challenge, we analyzed the public PPI services that were mostly contacted by the samples in our dataset, and we tried to infiltrate these services by supplying a small program we developed for distribution. By testing and manually reverse engineering the resulting installer we developed a set of 13 distinct network signatures that match the download URLs associated with different families of PPI services. By using these signatures on the malware traffic we could further discard their associated samples in our dataset. As illustrated in Fig. 2, we were able to discard $1,003,289$ PPI downloaders, which corresponds to up to $93.2\,\%$ of our initial dataset. Table 1 summarizes the top 20 PPI domain names that were contacted by malware in our dataset and the number of samples that were associated with each service.

In addition to PPI downloaders, our dataset also includes benign files that were submitted for analysis in Anubis. In fact Anubis is a public service where Internet users freely submit suspect files for analysis. These files may turn out to be benign files that connect to benign cloud-based services and so they also need to be discarded from our dataset as they do not belong to the malware category. Since our dataset covers a period where the most recent samples are few months old, we use anti-virus (AV) signatures to identify and discard benign

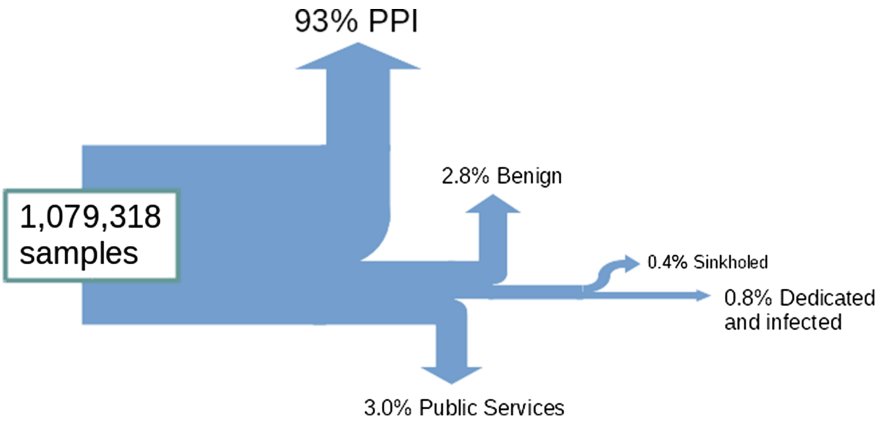



**Fig. 2.** Composition of our malware dataset

samples. We refer to public services such as VirusTotal[3] to scan our dataset, and we consider as benign files all samples that are detected by less than five AV editors. Our dataset finally includes 45,422 confirmed malicious malware samples. The remaining 30,607 samples (2.83 % of the initial malware dataset) were discarded as we do not have enough confidence about the malicious nature of these files.

*Domain Filter:* The domain filter module further discards from our dataset all domains that are associated with benign cloud-based services. Although these domains supply public Internet services that can be used by malware, they are not part of dedicated malicious services. Out of the initial set of 12,522 distinct EC2-based domain names or IP addresses, the malware scanner discarded 8,619 associated to PPI services or that were also contacted by benign programs. The domain filter classifies the remaining 3,903 domains into four categories, as illustrated in Table 2.

The first category includes public benign services that were contacted by malware. We found multiple examples in this category, including public IP resolvers (e.g. `hostip.info`), advertising and affiliate services, file sharing (e.g. `dropbox`), URL shortening (e.g. `notlong.com`), and multiple other free services (e.g. `about.me`, `spring.me`). To identify known public services in our dataset, the domain filter leverages multiple sources such as the Alexa list of top domains, public repositories that provide URL shortening services (e.g. `bit.do`) and file sharing[4]. We also refer to AV labels in VirusTotal in order to identify generic adwares. The domain filter module identifies as advertisement services all domains that were only contacted by Adwares samples. To be conservative, these domains were classified by our system into the public services category.

[3] http://www.virustotal.com.
[4] http://online-file-sharing-services-review.toptenreviews.com/.

**Table 2.** EC2-based service categories

| Service | | Domain Names | Malware Samples |
|---|---|---|---|
| Public Services | Advertising | 930 | 22,216 |
| | File sharing | 796 | 7,657 |
| | Domain redirection | 270 | 479 |
| | Others | 211 | 1,723 |
| Sinkholed | | 26 | 4,249 |
| Infected | | 22 | 231 |
| Dedicated | | 1,648 | 7,884 |
| | | N/A | 983 |
| **Total** | | **3,903** | **45,422** |

The second category includes domain names that have been sinkholed, and so they were redirected to sinkhole destinations that are hosted on the cloud. EC2 hosts multiple sinkhole destinations that are used to subvert BOT communications with their remote C&C domains. These domains were not originally hosted on EC2, and so they need to be discarded from our dataset. We leverage the `X-sinkole` HTTP header[5] in order to identify sinkhole destinations in our dataset.

The last two categories include both dedicated malware domains and domains that were once infected and temporarily used as part of the malicious infrastructure. The separation between these two categories is more difficult and more prone to errors. Our system relies on multiple empirical observations in order to discriminate between the two cases. First, we assume that dedicated malware machines that were hosted on EC2 more than one year ago have all been detected or abandoned at the time we run our experiments. We show in Sect. 3 that this is a reasonable assumption, consistent with the average lifetime of dedicated malicious domains hosted on EC2. Based on this assumption, the domain filter module actively probes all the domains and if the domain is still in use and points to a populated web page, we classify it as an infected host. Unfortunately, domain name vendors often return a HTML page to sell expired domains. Therefore, to correctly identify these domains, we parsed the HTML response page using multiple keywords (e.g. '`domain expired`', '`domain for sale`') and we removed these domains from the infected domains category. On top of this first heuristic, we also leveraged the history of DNS requests towards expired domains in order to assess the average lifetime of these domains. We use for this purpose DNS records that we extracted from DNSDB, a passive DNS duplication service[6]. In this case, our assumption is that infected domains usually have a longer turnover than other dedicated malicious domains. In other terms, infected domains are expected to appear in DNS records a long time before the

[5] http://www.iss.net/security_center/reference/vuln/HTTP_Malware_XSinkhole.htm.
[6] https://www.dnsdb.info/.

associated malware first appears in the wild. Dedicated domains instead, usually appear a short time before the malware is released and go offline a short time after the malware has been detected. By combining these two heuristics we were able to identify 22 infected services hosted on EC2 over the six years of observation. The remaining 1,648 domain names were identified by our system as being associated with dedicated malicious services.

Most of the connections were initiated using a domain name, but 983 malware samples directly connected to EC2-based IP addresses that were hard-coded in the malware itself, without any prior DNS request. To summarize, 8,867 of the 45,422 samples used at least one dedicated server hosted on Amazon EC2. For these, we also analyzed the content of their network communications. Almost 90.3 % of these samples used the standard HTTP protocol (either GET, POST, or HEAD methods). Few samples were IRC bots (19 distinct samples) and spam bots (136 distinct samples) that connected to malicious IRC and SMTP servers hosted on EC2. The remaining samples belonged to the Zeus version 3 and the Sality peer to peer (P2P) malware families, and were using the UDP protocol to connect to malicious P2P services hosted on EC2.

**Feature Analysis.** The analysis module processes the output dataset provided by the feature extraction module in order to extract main trends. First, it clusters malware families according to their antivirus labels, in order to figure out whether there exists a general trend towards moving malware infrastructures into the cloud, or whether this phenomenon is limited to some specific malware families. Second, it analyzes the network activity of each malware sample, computing the distribution of IP addresses and the domain flux to tell if miscreants specifically target cloud services, or if they use these services as part of their redundant malware infrastructure. Third, the feature analysis module observes the average duration a dedicated malicious server remains publicly accessible on the cloud. This can be used to estimate how effective are cloud providers in detecting abuse of their services, and whether malware writers sustain long lived malicious activities through the use of public cloud services. The following section provides the details and the main results of our experiments.

## 3 Experiments

The dataset provided by the feature extraction module (as described in Table 2) allows us to analyze both the malware families that are using EC2 cloud services in some capacity, as well as the distribution and lifetime of malicious domains that are hosted on EC2. Therefore, a first question that we would like to address in this section is whether the use of public cloud services is still limited to a small set of malware families, or whether it can be generalized to different families of malware. A straightforward approach to answer this question is to analyze the 8,867 distinct malware samples that we found to be connecting to dedicated malicious EC2 machines.

Since our dataset includes malware samples that are at least few months old at the time we run our analysis, we believe it is reasonable to use AV labels

**Table 3.** Top 20 malware family

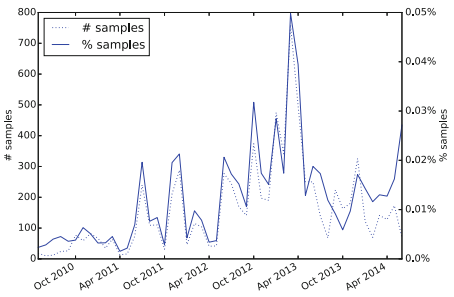
| AV label | # samples | AV label | # samples |
|---|---|---|---|
| Downloader Fosniw | 1249 | Trojan Kryptik | 160 |
| Worm Vobfus | 909 | Ramnit | 129 |
| Android DroidAp/SmsSend | 634 | Downloader Banload/Zlob | 128 |
| Downloader Murlo/Renos | 567 | Trojan Kazy | 127 |
| Backdoor QQRob | 528 | Downloader Virut/Virtob | 127 |
| Downloader Small BKY | 208 | Zbot | 117 |
| Delf Downloader | 196 | Malware SoftPulse | 108 |
| Trojan Injector | 194 | Downloader Karagany | 90 |
| Downloader 8CCBF09D99CF | 186 | Trojan Krap | 89 |
| Clicker Agent | 172 | Downloader Cutwail | 80 |

as a reference to understand and classify our dataset. More complex behavioral clustering mechanisms, as proposed for instance by Bayer et al. [5] and Perdisci et al. [19], could be applied to refine the classification. However, since we only need a broad understanding of the major malware families that use cloud services and we can tolerate few misclassification errors, a simple AV-based solution is better suited for our study.

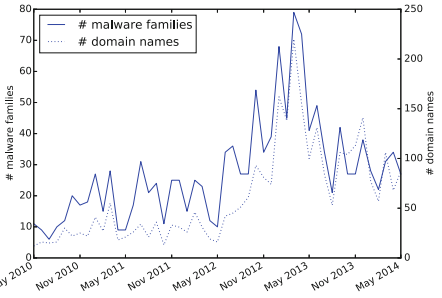
It is well known that different AV vendors assign different labels for the same malware sample. For example, the SpyEye malware can be identified by Kaspersky as `Trojan-Spy.Win32.SpyEyes`, and by McAfee as `PWS-Zbot.gen.br`. To limit the impact of such inconsistencies, we applied a majority voting to assign the labels to our dataset. In order to do so, we pre-process each label by splitting it in multiple elementary keywords according to non-alphanumeric characters We then discarded common prefixes such as `W32`, `Mal` and `Trojan`, as well as Generic malware identifiers, such as `Heur`, `Worm`, `Gen`, and `malware`. To handle malware aliases, we referred to multiple public sources such as the spywareremove website[7] to group together all aliases of a given malware family. For example, the labels `win32.spammy` by Kaspersky and `W32/Sality` by McAfee were identified as aliases for the same sality malware, and therefore grouped as part of the same family.

*Cloud-Based Malware Families:* We mainly focus in this paper on malware that uses dedicated malicious services hosted on EC2. Therefore, we build clusters of malware families for our dataset including 8,867 distinct samples that belong to this category. Using our approach, we are able to identify 377 distinct malware families. As clearly illustrated in Table 3, which provides the list of top 20 malware families, we were not able to identify a predominant malware family that uses dedicated malicious cloud services. More interestingly, our dataset includes malware that uses different topologies, including also decentralized peer-to-peer networks such as the Sality malware. Clearly the use of dedicated malicious

[7] http://spywareremove.com/.




(a) Number and ratio of distinct malware samples that connected to dedicated malicious domains hosted on EC2




(b) Number of distinct malware families that used dedicated EC2-based malicious services and number of their associated malicious domain names

**Fig. 3.** Malware dataset analysis

cloud services is not limited to a small set of malware families, but it could be generalized to all categories of malware.

*Time Evolution:* Since the hosting and usage of malicious services on public cloud infrastructures such as EC2 is not limited to specific malware families, our next goal is to identify if there is a clear trend on the amount of malicious software that make use of cloud services. Figure 3a illustrates the number of distinct samples that connected to dedicated malicious domains hosted on EC2 during the period of our observation. To account for changes in the overall number of submissions, the figure also shows the percentage of distinct samples compared to the total number of samples submitted to Anubis in the same period. Figure 3b shows instead the number of distinct malware families, and the number of their associated malicious domains that were found to be hosted on EC2 over the same period.

On average, the number of malware that uses dedicated cloud-based malicious services has grown by almost 4 times between 2010 and 2013. The overall trend also includes multiple peaks, that after a manual analysis resulted to be associated with multiple instances of malicious servers found to be temporarily hosted on Amazon EC2. While the fast growing number of malware samples that use cloud-based services may appear as a natural consequence of the general increase in the number of malware attacks [3], Fig. 3a shows that this is not the case and that the ratio between these malware samples and the total number of malware submitted to Anubis has been increasing at the same rate. As illustrated in Fig. 3b, this trend can be generalized to all malware families, which means there is a growing appetite towards using cloud infrastructures to host malicious servers. This could be due to multiple elements, including the fact that cloud services have been rapidly expanding in the past few years, and the fact that they are ease to access and still lack a strict accountability and control over their hosted machines [15].

### 3.1 Role of Public Cloud Services in Malware Infrastructures

In this section we describe the different ways malicious software makes use of public cloud services. In particular, we are interested in understanding whether miscreants specifically target cloud services or whether they use these services as small parts in a much larger redundant infrastructure.

For this purpose, we measured the ratio of remote malicious destinations that were hosted on EC2, compared to all malicious destinations contacted by the malware during the analysis. Then, for those malicious services that were hosted on EC2, we determined if they were hosted on EC2 only as part of a redundant mechanism. In this case, we extracted the DNS requests executed by the malware and we monitored the DNS records history using DNSDB service in order to compute the ratio of IP addresses that belong to the EC2 IP range, compared to all IP addresses associated with that malicious domain in other moment in time. This technique works particularly well in the presence of round-robin DNS and DNS fast-flux techniques that are often adopted by botnet herders. For instance, miscreants can associate different IP addresses with the same malicious domain name, where only some of these IPs may be hosted on EC2.

Figure 4 presents the average distribution of the ratio of remote malicious destinations that were hosted on EC2, compared to all malicious destinations contacted by all malware samples. We present our findings as a box plot where malware samples are classified according to their submission date to Anubis. The Y-axis characterizes the ratio of dedicated malicious domains that were hosted on EC2, with respect to all malicious domains contacted by malware. Since we included in this experiment only malware samples that used dedicated malicious services on the Cloud, the percentage is always greater than 0 %. On the other hand, a malware would fit into the 100 % category in case all dedicated malicious domains that were contacted by the malware were strictly found to be hosted on EC2.

As shown in Fig. 4, miscreants mostly use public cloud services in order to host only certain components of their malware infrastructures. Note that while in 2010, and for malware that uses EC2 to host its dedicated malicious services, only few components of its malware infrastructures were found to be hosted on EC2 (less than 40 % of remote malicious domains in average); the use of public clouds to host dedicated malicious services has rapidly evolved in the recent years, including malware samples that were found to be exclusively communicating with dedicated cloud-based malicious domains in years 2013 and 2014. In other terms, miscreants have been recently referring to public cloud services in order to setup and manage their entire malware infrastructure. Therefore, although the use of public cloud services is still limited to only specific components of malware infrastructures, we observe an increasing appetite for miscreants towards using public cloud services to setup and manage additional components of their malware infrastructures.

Since most miscreants refer to public cloud services to host only certain components of their malware infrastructure, the second question we would like to answer is whether they specifically refer to public cloud services for this purpose,

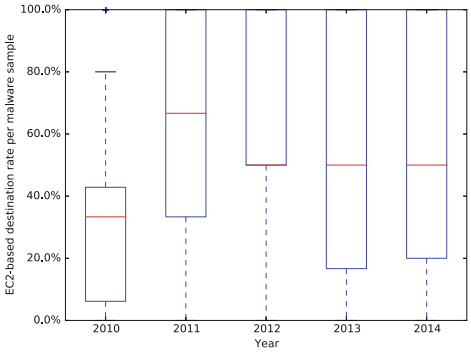



**Fig. 4.** Rate of dedicated malicious EC2-based domains contact per malware sample

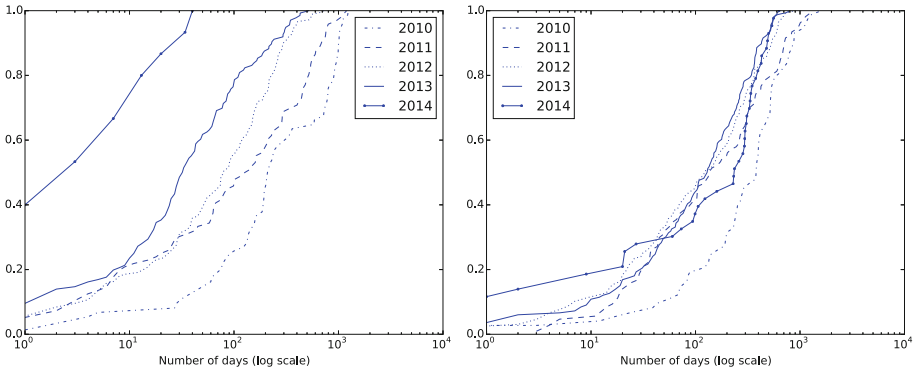
or whether they use these services as redundant or failover components. We observed for this purpose the history of DNS records for all dedicated malicious EC2-based domains in our dataset until they were blacklisted, then we identified all IP addresses that were associated with these domains and their registrars in the DNSDB service. The results of our investigation were compelling. Out of the initial 1,648 dedicated malicious EC2-based domains that constitute our dataset, 1,620 domains (almost 98.3 % of our dataset) were exclusively associated with IP addresses that belong to the EC2 IP range. Note that while 87.5 % of dedicated malicious domains were associated with only a single EC2-based IP address, another 10.8 % where found to be associated with multiple IP addresses that all belong to the EC2 IP range. In other terms, miscreants were specifically using public cloud infrastructures such as EC2 to host their dedicated malicious services.

While the use of public cloud services to host dedicated malicious domains is still limited to only certain components of today's malware infrastructures, miscreants appear to be specifically targeting cloud infrastructures for this purpose, and do not use public clouds only as redundant components of their malware infrastructures.

### 3.2 Dedicated Domains Lifetime Estimation

In the last part of our study, we tried to estimate the average time that malicious domains persist on EC2 cloud. Our approach leverages the lifetime of the EC2-based malicious domains in order to estimate whether the use of public cloud providers such as EC2 adds more resilience to malware infrastructures.

In the following, we refer to the lifetime of a EC2-based malicious domain as the duration when it was consecutively associated with EC2 cloud IP address in the passive DNS records. Note that the use of passive DNS service only provides an estimation of the real lifetime of these domains but this is an approximation that is often used for this type of measurements [16]. Since domains first appear in the passive DNS services when they are actively requested on the Internet,




(a) Time elapsed until a dedicated malicious EC2-based domain was first observed in Anubis

(b)Time until a edicated malicious domain was no longer hosted on EC2, after it was observed in Anubis

**Fig. 5.** Lifetime of dedicated malicious EC2-based domains

we consider that the use of this service provides a reliable estimation of the real duration in which a given domain remained active and accessible on the wild. In this section, we observe only dedicated malicious domains that were hosted on EC2, and that were contacted by our malware dataset collected over a period that ends by June 2014. Hence, we only consider historical DNS records associated with malicious servers that are no longer hosted on EC2 cloud at the time of writing. Note that these domains may be still accessible but no longer associated with any EC2 IP address.

We defined two metrics for our experiments. First of all, we measured the time between the domain first appeared in the passive DNS service and the time the malware was analyzed by Anubis. Second, we extract the time when a dedicated malicious domain is no longer associated with an EC2 IP address, after the malware was first submitted to the Anubis service.

The results of our experiment are summarized by the cumulative distributions that are illustrated in Fig. 5. The graphs separately illustrate the results of our experiments for the last five years since 2010, in order to extrapolate some trends and assess the efficiency of security measures implemented by Amazon over time. The first graph shows that the distribution is clearly moving toward the top-left corner, meaning that each domain was observed in Anubis soon after it first appeared in the wild. For instance, while in 2011 around 50 % of the domains were already present in the passive DNS service (and therefore active in the wild) for 100 days before some malware in Anubis contacted them, in 2014 they had only been active for two days. In other words, the security community became very efficient to promptly detect, collect, and submit malware samples.

Unfortunately, Fig. 5b shows that the time these domains were hosted on EC2 *after* the malware was analyzed remained stable over the same period. While many factors are involved in this process, this seems to suggest that Cloud

providers did not improve their ability to detect and report abusive behaviors. In other words, our observations suggest that the security mechanisms implemented by public cloud service providers have not contributed to reducing the lifetime of malicious domains hosted by these providers.

In order to confirm these findings, and since cloud providers may take-down malicious IPs and not their associated domain names, we analyzed the way malicious EC2 domains resolve to different IP addresses over time. We wanted to evaluate how long malicious machines remain active on EC2 before they are taken down by the cloud provider, and so miscreants may be forced into migrating their malicious domains towards other IP addresses. We monitored for this purpose all DNS records in the DNSDB service, searching for different IP addresses that were associated with every malicious domain in our dataset. Confirming our hypothesis, we found multiple instances of malicious machines that remained active on EC2 even for several months before they were migrated towards other IP addresses in the cloud.

**Table 4.** Examples of domains that rotated their IP addresses on EC2 over time

| Id | Domain | IP address | First seen | Last seen | Duration (months) |
|---|---|---|---|---|---|
| 1 | 09sp.co.tv | 174.129.222.176 | August 2010 | October 2010 | 3 |
| | | 174.129.242.247 | January 2011 | November 2012 | 22 |
| 2 | 47gr.co.tv | 174.129.222.176 | July 2010 | July 2010 | 1 |
| | | 174.129.242.247 | February 2011 | November 2012 | 21 |
| 3 | dl.ka3ek.com | 107.20.206.69 | January 2013 | November 2013 | 11 |
| | | 54.209.129.218 | January 2014 | January 2014 | 1 |
| 4 | hightool.com | 107.20.206.69 | January 2013 | December 2013 | 12 |
| | | 54.209.168.250 | March 2014 | September 2014 | 7 |
| | | 54.208.247.222 | September 2014 | September 2014 | 1 |
| 5 | hzmksreiuojy.com | 54.241.7.53 | April 2013 | April 2013 | 1 |
| | | 50.18.179.196 | April 2013 | October 2013 | 7 |
| | | 50.17.195.149 | July 2014 | July 2014 | 1 |

As illustrated by the examples in Table 4, the first two malicious domains were associated with the same EC2 IP address for up to twenty consecutive months before they went out from the EC2 IP ranges. Interestingly, certain malicious domains, such as domains 1 and 2, as well as domains 3 and 4 in Table 4, were associated with the same IP address during the same period of time, which seems to indicate that miscreants may associate different domain names with their malicious machines in the cloud in order to obtain a better resilience against domain blacklisting. Moreover, they also seem to benefit from the flexibility offered by the cloud in order to migrate towards new IP addresses in

EC2 as soon as their current IP addresses have disappeared from the active DNS records, which may suggest that their malicious machines have been identified and taken down by the cloud provider (EC2 in our study). In total, we observed similar behaviors in over 240 malicious domains in our dataset.

## 4 Discussion

When we started our experiments, we were surprised to discover that over 3.5 % of the malware samples in our dataset exhibited at least one network connection with a machine hosted on the Amazon cloud. However, as clearly depicted in Fig. 2, the vast majority of these connections had nothing to do with the fact that criminals were intentionally using the Cloud as part of their infrastructure. In fact, once PPI and other benign services were filtered out, we discovered that less than 1 % of the traffic toward Amazon involved a malicious EC2 machine.

Even though this number may seem incredibly small (roughly one every 3200 malicious samples), it is still relevant when scaled to the entire dataset containing tens of millions of malicious samples. Moreover, our experiments show that the use of public cloud services by malicious software has been increasing over the last six years, despite the measures taken by certain cloud providers to limit the extent of these abuses. It also seems that the use of dedicated malicious cloud services is not limited to a small set of malware families, but it can be generalized to most of the malware categories – as summarized by Table 3.

The final observation of our study is related to how the cloud providers, Amazon in our case, respond to this threat. Even though we did not have a direct way to observe their reaction, we were able to measure for how long – after the malware was publicly known – the malicious domains it contacted were resolving to IPs hosted on EC2. While the absolute value is not very important, the fact that it remained constant over the past four years seems to indicate that the cloud provider did not make any substantial improvement in detecting and taking down malicious machines.

## 5 Related Work

Prior abuse cases of public cloud providers have attracted a lot of interests in the recent years [7,11,14,21]. For instance, cloud services are listed by Solutionary [20] among the major components of modern cybercrime, and attackers seem to use these services the same way and for the same reasons as legitimate customers. Despite this popularity, we are aware of only few research studies that managed to evaluate the real extent of this phenomenon.

Hamza et al. [12] presented a survey of possible techniques to abuse cloud services in modern cybercrime. They provide interesting insights on the way cyber-attacks are perpetrated from within cloud platforms, including examples such as host hopping attacks and abuse of privileges. However, this survey only focuses on strong attack signals, and does not consider other weak signals that

determine the way cloud services are being used as part of the attackers command and control infrastructures.

In [18], Nappa et al. analyze drive-by download attacks and exploit servers that are managed by the same organizations. They found that 60 % of the exploit servers are hosted by public cloud service providers. More interestingly, they evaluated the abuse report procedures implemented by public cloud service providers. They realized that out of 19 abuse reports they have submitted, only 7 were investigated by cloud providers. Moreover, the authors computed that it takes on average 4.3 days for a cloud provider to take down an exploit server after it has been reported. It is important to note that the authors of this study only focus on drive-by-download attacks that involve cloud services. Although drive-by-download servers constitute a major component of a modern malware infrastructures, we go beyond this unique use case in order to provide in this paper a more comprehensive assessment about the way cloud services are being integrated in malware infrastructures in modern cybercrime. We also try to understand whether clouds constitute core elements of the malware structure, or whether they are only used as redundant or failover components.

In [9], Canali et al. propose an active approach to evaluate the security mechanisms implemented by web hosting providers. They installed vulnerable web services on 22 distinct hosting providers, and triggered multiple attacks to leverage the reaction capabilities of these providers. To test the security mechanisms implemented by cloud service providers we adopt a less intrusive approach where we only observe malware interactions with the cloud. In our study we only focus on Amazon EC2. While this choice may limit the extent of our observations, at the same time eliminate as much as possible the impact of rogue or other hosting providers that do not guarantee minimal security SLA requirements to their users. We believe that focusing only on the biggest cloud providers in terms of market share also shed light on the limits of current security and accountability mechanisms implemented by today's cloud providers.

Finally, Wang et al. [22] propose a system that measures the churn rates in EC2 and Azure in order to evaluate the efficacy of IP blacklists for malicious activity in the cloud. The authors actively probed the EC2 and Azure IP ranges, and proposed a clustering mechanism that groups together IP addresses implementing the same services. They also observed all web services hosted by cloud providers, spanning both benign and malicious activities. The results of their experiments show only small amounts of malicious activity (mostly phishing and malware hosting) by comparing data from their system with public blacklists. We propose in this paper a complementary approach that observes only malware interactions with the cloud in order to leverage the true extent of the malicious activity hosted by public cloud providers.

## 6 Conclusion

Public cloud services have rapidly expanded in recent years, yet they have attracted cyber criminals because of the wealth of resources they make available,

and the lack of accountability over the usage of these resources. In order to measure the extent at which public cloud services are being abused by cyber-criminals, we conducted a longitudinal study of malware in order to better understand the way it interacts with public cloud services.

In particular, in this paper we study several characteristics of the traffic observed between malicious samples and the Amazon EC2 cloud. Based on our measurements, we discuss the evolution of this phenomenon over the past six years, and we present few key observations and insights into this growing problem.

We hope that our study can shed some light on a key component of the modern cyber crime infrastructure, and would provide useful input to devise appropriate mitigation strategies.

**Acknowledgments.** We would like to thank the reviewers for their valuable comments that allowed us to improve the quality of this paper. This research was partly funded by the French Ministry of education and research under Cifre grant given to Xiao Han, and by the European Unions Horizon 2020 project SUPERCLOUD under grant agreement 643964.

## References

1. Best pay-per-install affiliate program reviews. http://pay-per-install.com. Accessed December 2014
2. Amazon, E.: Amazon elastic compute cloud (amazon ec2) (2010). http://aws.amazon.com/ec2/
3. AVTest Institute. Malware statistics & trends report. http://www.av-test.org/en/statistics/malware/
4. Azure, W.: Microsofts cloud platform (2013). http://azure.microsoft.com/
5. Bayer, U., Comparetti, P.M., Hlauschek, C., Kruegel, C., Kirda, E.: Scalable, behavior-based malware clustering. NDSS **9**, 8–11 (2009)
6. Bayer, U., Kruegel, C., Kirda, E.: Ttanalyze: a tool for analyzing malware. In: 15th European Institute for Computer Antivirus Research (EICAR 2006) Annual Conference (2006)
7. Bestuzhev, D.: Financial data stealing malware now on amazon web services cloud (2011). http://www.securelist.com/en/blog/208188099/Financial_data_stealing_Malware_now_on_Amazon_Web_Services_Cloud. Accessed 15 May 2014
8. Caballero, J., Grier, C., Kreibich, C., Paxson, V.: Measuring pay-per-install: the commoditization of malware distribution. In: 20th USENIX Conference on Security (2011)
9. Canali, D., Balzarotti, D., Francillon, A.: The role of web hosting providers in detecting compromised websites. In: Proceedings of the 22nd International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, pp. 177–188 (2013)
10. Cohen, R.: The cloud hits the mainstream: more than half of u.s. businesses now use cloud computing. In: Forbes Magazine (2013)
11. Ferrer, M.C.: Zeus in the cloud. http://community.ca.com/blogs/securityadvisor/archive/2009/12/09/zeus-in-the-cloud.aspx

12. Hamza, Y.A., Omar, M.D.: Cloud computing security: abuse and nefarious use of cloud computing. Int. J. Comput. Eng. Res. **3**, 22–27 (2013)
13. He, K., Fisher, A., Wang, L., Gember, A., Akella, A., Ristenpart, T.: Next stop, the cloud: understanding modern web service deployment in ec2 and azure. In: ACM Internet Measurement Conference (IMC) (2013)
14. Higgins, K.J.: Dropbox, wordpress used as cloud cover in new apt attacks (2013). http://www.darkreading.com/attacks-breaches/dropbox-wordpress-used-as-cloud-cover-in-new-apt-attacks/d/d-id/1140098?. Accessed 15 May 2014
15. Ko, R.K., Jagadpramana, P., Mowbray, M., Pearson, S., Kirchberg, M., Liang, Q., Lee, B.S.: Trustcloud: a framework for accountability and trust in cloud computing. In: 2011 IEEE World Congress on Services (SERVICES), pp. 584–588. IEEE (2011)
16. Li, Z., Alrwais, S., Xie, Y., Yu, F., Wang, X.: Finding the linchpins of the dark web: a study on topologically dedicated hosts on malicious web infrastructures. In: 2013 IEEE Symposium on Security and Privacy (SP), pp. 112–126. IEEE (2013)
17. Los, R., Shackleford, D., Sullivan, B.: The notorious nine cloud computing top threats in 2013. In: Cloud Security Alliance (2013)
18. Nappa, A., Xu, Z., Rafique, M.Z., Caballero, J., Cyberprobe, G.Gu.: Towards internet-scale active detection of malicious servers. In: Network and Distributed System Security Symposium (NDSS) (2014)
19. Perdisci, R., Lee, W., Feamster, N.: Behavioral clustering of http-based malware and signature generation using malicious network traces. In: USENIX Symposium on Networked Systems Design and Implementation (NSDI) (2010)
20. Solutionary. Security Engineering Research Team (SERT) Quarterly Threat Intelligence Report (2014). http://www.solutionary.com/_assets/pdf/research/sert-q4-2013-threat-intelligence.pdf. Accessed 15 May 2014
21. Unuchek, R.: Gcm in malicious attachments (2013). http://www.securelist.com/en/blog/8113/GCM_in_malicious_attachments. Accessed 15 May 2014
22. Wang, L., Nappa, A., Caballero, J., Ristenpart, T., Akella, A.: Whowas: a platform for measuring web deployments on iaas clouds. In: ACM Internet Measurement Conference (IMC) (2014)