Software Engineering Institute
Carnegie Mellon University

A Unique Approach to Threat Analysis Mapping: A Malware-Centric Methodology for Better Understanding the Adversary Landscape

Deana Shick
Kyle O’Meara

April 2016

TECHNICAL REPORT
CMU/SEI-2016-TR-004

CERT Division

http://www.sei.cmu.edu

This material is based upon work funded and supported by Department of Homeland Security under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center sponsored by the United States Department of Defense.

Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of Department of Homeland Security or the United States Department of Defense.

This report was prepared for the
SEI Administrative Agent
AFLCMC/PZM
20 Schilling Circle, Bldg 1305, 3rd floor
Hanscom AFB, MA 01731-2125

DM-0003537

Table of Contents

List of Figures

List of Tables

Abstract

Malware family analysis is a constant process of identifying exemplars of malicious software, recognizing changes in the code, and producing groups of “families” used by incident responders, network operators, and cyber threat analysts. With adversaries constantly changing network infrastructure, it is easy to lose sight of the tools consistently being used and updated by these various actors. Beginning with malware family analysis, this methodology seeks to map vulnerabilities, exploits, additional malware, network infrastructure, and adversaries’ using Open Source Intelligence (OSINT) and public data feeds for the network defense and intelligence communities. The results provide an expanded picture of adversaries’ profile rather than an incomplete story.

The goal of this document is to shift the mindset of many researchers to begin with the tools used by adversaries rather than with network or incident data alone for an “outside-in” approach to threat analysis instead of an “inside-out” method. We chose three malware families to use as case studies—Smallcase, Derusbi, and Sakula.

The results of each case study—any additional network indicators, malware, exploits, vulnerabilities, and overall understanding of an intrusion—tied to the malware families should be utilized by network defenders and intelligence circles to aid in decision making and analysis.

1 Introduction

The cyber security community currently is very attribution centric. When releasing reports, security vendors and other researchers typically employ an “inside-out” approach of hunting and analysis, using incident data of high-profile cyber attacks. This analysis centers on naming attackers and/or remediation efforts rather than on understanding every piece of the puzzle. While this model works for incident analysis, it does not provide a complete picture of adversarial operations for those in the network defense or intelligence communities. We present an “outside-in” methodology for using well-understood malware to provide analysis and context to intelligence circles and network defenders to satisfy this gap in reporting.

Our methodology uses five types of data to gather and expand information and indicators related to specific malware. We use data on malware families (or “knowns”), communication information used by those families, incidents related to the families in our group of interest, vulnerabilities used in those incidents, and the specific exploits that took advantage of the vulnerabilities. This is not a linear pivot from one data set to the next; instead, this methodology is more circular, as one data set can inform the rest of the cycle. This data can come from open source intelligence (OSINT), in-house efforts, and commercial sources. Our case studies use mostly OSINT and in-house data sources to illustrate the usefulness of the methodology at minimal cost.

To outline this methodology, we chose three malware families, Smallcase, Derusbi, and Sakula, to begin our analysis with the following criteria: adversaries used malware in recent attacks as published by other security vendors, we previously created a configuration dumper (decoder) for the malware, and YARA signatures were used to classify family members. Additionally, the malware files for these families are not easily found on the Internet.

2 Motivation

Rarely do cyber security organizations provide one single service to sponsors or clients—the Software Engineering Institute CERT Coordination Center (CERT/CC) is no different in this respect. The work that we do ranges from in-depth analysis of malware to network analysis efforts, scientific methodology construction data set, tool creation, vulnerability discovery and coordination, and threat analysis. Threat analysis efforts also vary from understanding and reporting on network data to understanding the blacklist ecosystem.

In the months preceding this analysis, we performed a survey of the data we have in-house to understand gap areas and where we may have "ground truth" for analysis. We determined that our static malware analysis efforts provided the starting point we needed to best map to all other data sets. While our malware analysis efforts have been a piece of threat analysis, they have not been the starting point for analysis and data fusion.

We wanted to best utilize all of the efforts listed above when creating a usable methodology for those we support. This method is an "outside in" hunting mechanism instead of an "inside out" approach, which begins with incident data.

3 Methodology

We created this methodology to enable network defenders and those in intelligence circles to better utilize the data they have in-house and the data that is available publicly on the Internet to better inform decision makers and/or aid in defense efforts of adversary tooling. Our goal was to cluster data sets in such a way that they inform each other better to create a more complete picture of adversarial attacks.

This method begins with a well-understood malware family, which adds connections to other data sets. A visual example is found in Figure 1. The methodology is non-linear, as an analyst can bounce from one data set to the next once a malware family is identified for analysis. We utilize indicator expansion, which is the process of pivoting across data to grow larger data sets, particularly when associating IP address to domain name or vice versa [Shick & Horneman 2014]. Once we discuss the Known Malware methodology, the remaining sections are in no particular order.

Network defenders should use this methodology to pivot from one data set to the next in order to derive additional indicators of compromise (IOC). They can then use the data in efforts to block network traffic, write new or updated Snort signatures, or deploy host-based protections using the attributes of the malware or exploit. These are just a few of the efforts towards which such data is useful. Intelligence or strategic analysts, on the other hand, should use this methodology as an inexpensive means for understanding how an adversary may operate in the form of its TTPs, tools, and infrastructure.

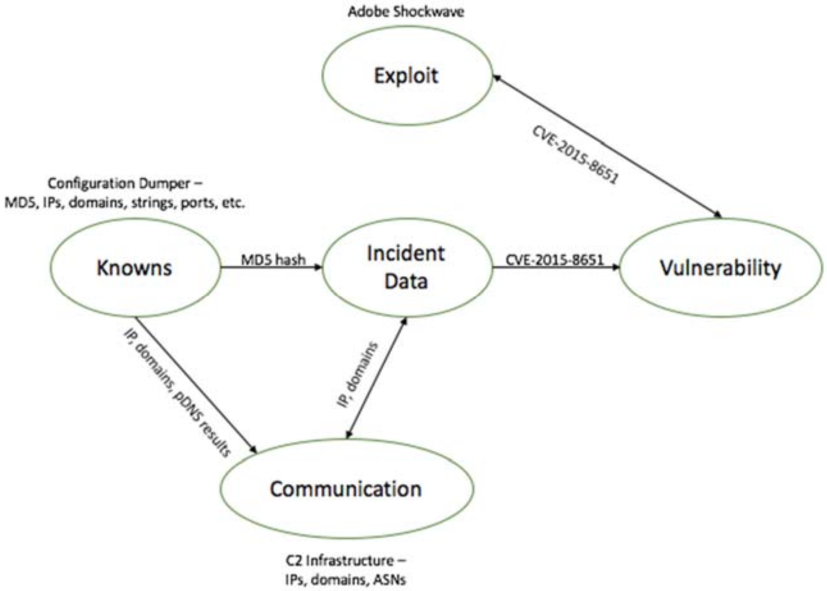

Figure 1: Threat Analysis Mapping—Example Using Smallcase Data

3.1 Pick Known Malware

Malware can be found in a myriad of places. We chose three malware families identified as complex, interesting, or destructive. Malware can be found in phishing emails, malware feeds (e.g., VirusTotal), and on the web. To be considered for this analysis the following must be present:

- static reverse engineering efforts
- the configuration information or the C2 information found in the malware, produced in the form of a configuration dumper, which can be written in a language of an analyst's choosing
- source code of the malware
- high-confidence YARA signatures

This analysis relies on having a set of malware in whose similarity or sharing of the same family the analysts have high confidence. The remaining steps in this methodology cannot be completed without a well-understood sample of malware.

3.1.1 Group Samples of Malware by using YARA

YARA is a tool used to identify similar malware samples given a known, malicious file. YARA rules can be found in OSINT or are written by analysts within an organization. If provided uncategorized malware, an analyst should use a vetted YARA rule to identify other files.

Once this step is completed, the analyst should have a collection of similar malware files for analysis. The message-digest algorithms (MD5s) of these malware samples will be used in the remaining parts of the analysis. Additional analysis using Fn2yara or other code comparison tools should be performed if there is ambiguity within the malware family or if a family has several names based on OSINT.

3.1.2 Use Code Comparison Tools if Needed

OSINT sources occasionally conflate malware based on arbitrary labeling. It helps to separate and appropriately label families so analysts can increase the fidelity of findings. Two code comparison tools were used during this analysis: Fn2yara and Malware Clone Mapping.

3.1.2.1 Fn2yara

Fn2yara is a tool used to create YARA rules for each function in an executable file to determine similarity of files. This tool was used to complete a systematic code comparison of the known malware from the use cases in the following way:

- Multiple YARA signature files were created from a known set of malware.
- YARA was then used to run the signature files generated by Fn2yara across a known set of clean, non-malicious files (cleanware):
  - for example, the Windows System Directory (i.e., System32). Any rules that alert on the known set of cleanware were removed from the YARA signature files.
- Following the removal of cleanware YARA rule alerts, frequency analysis was conducted to identify YARA rules that alert on all or a high frequency of the set of known sets of malware.

- The remaining rules with lower frequency hits can be removed at analyst's discretion. The final set of pared-down YARA rules was used to do code comparison of potentially similar malware.

3.1.2.2 Malware Clone Mapping

Malware clone mapping techniques allow researchers to understand the similarity of files belonging to one or more malware families. This is done at the micro-level by comparing the binaries of each file, unlike Fn2yara that provides the comparisons of function similarity.

We sampled malware families randomly to pick three samples from each family. We provided two tests in our analysis: the first is the Extra-Family, in which the samples are compared with files denoted as an independent family; the second is the Intra-Family, which determines how similar the file is compared to other files of the same family. We focused only on strings found in more than one sample.

3.1.3 Understand the Malware Compile Times

This analytic could help analysts determine when a piece of malware was first written and when the malware family grew in number. The compile time should be provided in the configuration output, and should be analyzed to build a timeline for situational awareness purposes.

3.1.4 Identify Common Themes

The configuration information can yield other interesting results beyond compile time and network information. Sometimes notable strings will be identified in the data, such as remote files requested by the malware, ports, and Uniform Resource Identifiers (URIs). This data may help analysts gain additional context around how a malware family operates.

3.2 Analyze Network Communications

3.2.1 Understand Address (A), IPv6 Address (AAAA), Name Server (NS), Mail Exchange (MX), and Start of Authority (SOA) Records and Trends

These equities are considered fragile, as adversaries can make a quick change to reach their objectives [Spring & Stoner 2015]. Despite this, they are nonetheless important to network defenders and intelligence communities, as they allow analysts to defend their networks and easily gather intelligence about adversary operations in a particular snapshot in time [Spring & Stoner 2015]. Analysis of the network communications was completed in a few steps using Bash and Python scripts:

- Pull the IP addresses and domain names from the configuration dumpers.
- Perform indicator expansion within the timeline chosen for analysis to find additional network information associated with the malware.
  - We ran IP addresses against the Farsight Passive DNS (pDNS) database to receive other domain names that are likely related to the malware, and vice versa [Shick & Horneman 2014].

  - Analyze A, AAAA, NS, SOA and MX records to discover trends such as shared infrastructure between malware families, particular infrastructure chosen for computer network operations (CNO), and so on. It is variable whether all record types are found in pDNS data. If certain record types were not useful, we did not provide it in the results sections.

3.2.2 Identify the Autonomous System Numbers (ASNs) and Organizations

If results showed that IP addresses used during operations were clustered within an ASN, it is reasonable to assume that the organization in question was targeted to be used in operations for a particular reason. Additionally, if a handful of ASNs were provided as a result of the analysis, it may indicate those organizations were targeted by adversary groups [Shick & Horneman 2014].

- System for Internet-Level Knowledge (SiLK) enabled us to create prefix maps (pmaps) to find the ASNs associated with the IP addresses collected.
- Once the ASN map was completed, we found the organizations responsible for the ASNs via a dictionary file based on potaroo.net data.
- We did not combine subsidiary companies or update mergers or acquisitions once we performed the analysis [Shick & Horneman 2014].

3.3 Research for Incident Data

Incident data can come from a variety of locations: vendor reports, internal networks, or Internet searches, which provide analysts with context about an intrusion. We performed this data search manually, and primarily used data released by anti-virus (AV) or security vendors easily found in public reporting. We also utilized the Malware Information Sharing Platform (MISP) of Circl.lu to find incidents associated with the malware. This tool easily culled several vendor and AV reports in one location.

- We searched by malware MD5 rather than IP address or domain name for this linkage.
- Once we found a report, we scraped it for relevant context to the intrusion.

One thing to note: security vendors typically rename attacker groups allowing one group to have several associated names. This aspect makes it ever more important to exhaust searches to understand the full context surrounding an intrusion. Additionally, we found that OSINT sources were not always correct in their assertions, so analysts must be vigilant and recognize this pitfall.

3.4 Analyze Vulnerabilities

Vulnerability information is necessary for understanding how the attackers may have gained access into a network. Unlike network indicators, which are more ephemeral and are only malicious during particular periods in time, older vulnerabilities are still in vogue due to machines remaining unpatched. Old vulnerabilities are exploited frequently by various types of actors.

Vulnerabilities can be found in all aspects of products, and some of the more common avenues of exploitation are found in Internet-facing applications. For example, these vulnerabilities have been linked to Microsoft Silverlight, Java, and Adobe Shockwave. Among phishing campaigns, the Incident Data or other OSINT searching revealed that specific vulnerabilities were targeted during actor CNO. This was also a manual process that required extensive internet searching. The two main databases that provide context around a vulnerability are the following:

- Common Vulnerabilities and Exposure (CVE) identifier found on MITRE’s CVE Database [Mitre 2016].
- CERT/CC Vulnerability notes [CERT Coordination Center 2016].

3.5 Research Exploits

Exploits are typically the front door into a network by way of vulnerable software, and are one overlooked analysis area. Exploits present an interesting problem to network defenders and other analysts, since time has no effect on the applicability of the usage of exploits, whereas it does for other types of data such as network indicators. For example, if systems are not patched, an exploit that attacks a vulnerability from 2012 can have similar impact as an exploit for vulnerability from 2016. By researching OSINT sources, we discovered a sample of each exploit mentioned in the use cases. Specially, we determined if the exploit was present in Exploit Database (Exploit-DB) and/or VirusTotal.

4 Results

This section outlines the results of the proposed methodology. There are three case studies using well-understood malware families: Smallcase, Derusbi, and Sakula. These malware families are not easily found on the Internet. Each case study begins with a Summary section denoting our findings at a high level followed by more granular analysis.

4.1 The Smallcase Malware Family

4.1.1 Summary of Findings

Smallcase malware was likely developed before 2008 and saw active improvements between 2010 and 2011. This toolkit contains a dropper, downloader, and remote access trojan (RAT) used to evade detection and exfiltrate information from victim networks. There were no indications that this malware was closely associated with other known families.

The network infrastructure used during CNO included 85 IP addresses and 254 domain names linked to the malware, 172 of which were not reported in OSINT. We found that attacks were carried out from 33 organizations, primarily hosting providers. A high percentage (85%) of the infrastructure, resolved back to twopiz[.]com, which is associated with the hosting company, Twopiz Hosting, based in Masai, Malaysia. Attackers using this network infrastructure likely used the same few name servers to serve malicious domains against multiple victims.

The Smallcase malware is associated with Kaspersky's DarkHotel group, which used this malware to exfiltrate sensitive information from those staying in high-end Asian hotels. OSINT suggests that DarkHotel is still active, as it was associated with a zero-day vulnerability and subsequent Adobe Flash exploit in late December 2015. We found at least two exploits associated with CVE-2015-8651, which is associated with DarkHotel activity. We believe APT groups chose to compromise Adobe Flash given the prevalence of the software on a variety of machines and will continue actively using vulnerabilities and exploits to compromise a high number of victims.

4.1.2 Malware Analysis

We have been tracking the toolkit, Smallcase, since early 2014. Reverse engineering efforts showed the malware contained at least a downloader, RAT, and obfuscation techniques for communications and processes. As of January 1, 2016, we collected 188 Smallcase files that are broken down into the following categories: Dropper, Downloader, and RAT, as seen in Table 1. We did not provide a code analysis for this case study, as OSINT suggested the malware family was not conflated or otherwise associated with any other families.

Table 1: Smallcase Breakdown by Type and Number of MD5s

Smallcase Type	**MD5/Type**
Dropper	70
Downloader	85
RAT	33

The configuration dumper provided us with the following as of January 1, 2016:

- 97 unique domain names used for command and control (C2)
- 35 specific uniform resource identifiers (URIs) relating to specific PHP files
- ten IP addresses
- five unique strings

The compile times for Smallcase are shown in Figure 2. This data suggests that Smallcase was developed before 2008, and several variants written in 2010 and 2011.

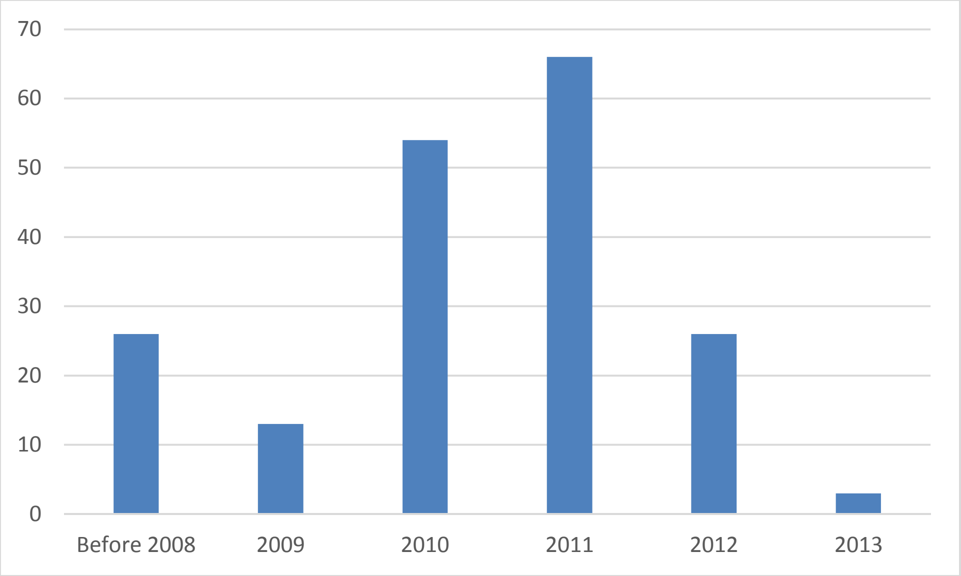

Figure 2: MD5 File Count Based on Smallcase Compile Times

By querying OSINT sources using MD5 hashes for the malware files, we found that Smallcase malware is associated with Kaspersky's DarkHotel group, which was still active into late 2015.

4.1.3 Network Communications

4.1.3.1 Address (A) Records

The conversion of IP address to domain name and vice versa via indicator expansion provided us with the following:

- 157 unique domain names pulled from pDNS data.
- Our analysis pool contained 254 unique domain names combining the configuration dumper information and the pDNS results.
- 75 unique IP addresses pulled from pDNS data.
- Our analysis pool contained 85 IP addresses combining the configuration dumper information and the pDNS results.

At least 93 domain names were still active and queried into December 2015.

4.1.3.2 Name Server (NS) and Mail Exchange (MX) Records

Our analysis suggests that the attackers chose a particular organization and name server for attacks. We deduced this by analyzing the NS and other record types for all of the A records collected during analysis.

- 81.7% of all NS records (3592 of 4396 records) are related to twopiz[.]com. This domain was registered by the hosting company, Twopiz Hosting, based in Masai, Malaysia.
- There were 69 unique MX records in the data. p[.]nsm[.]ctmail[.]com was shared by six domains while smtp[.]secureserver[.]net was shared by three domains. The other MX records were unique.

4.1.3.3 ASN Analysis

We found that IP addresses were clustered within 33 organizations as shown in Figure 3. Wildcard UK, Limited held the most IP addresses (13). Additionally, 10 IP addresses did not resolve to any ASN. The majority of organizations of adversary infrastructure were hosting providers.

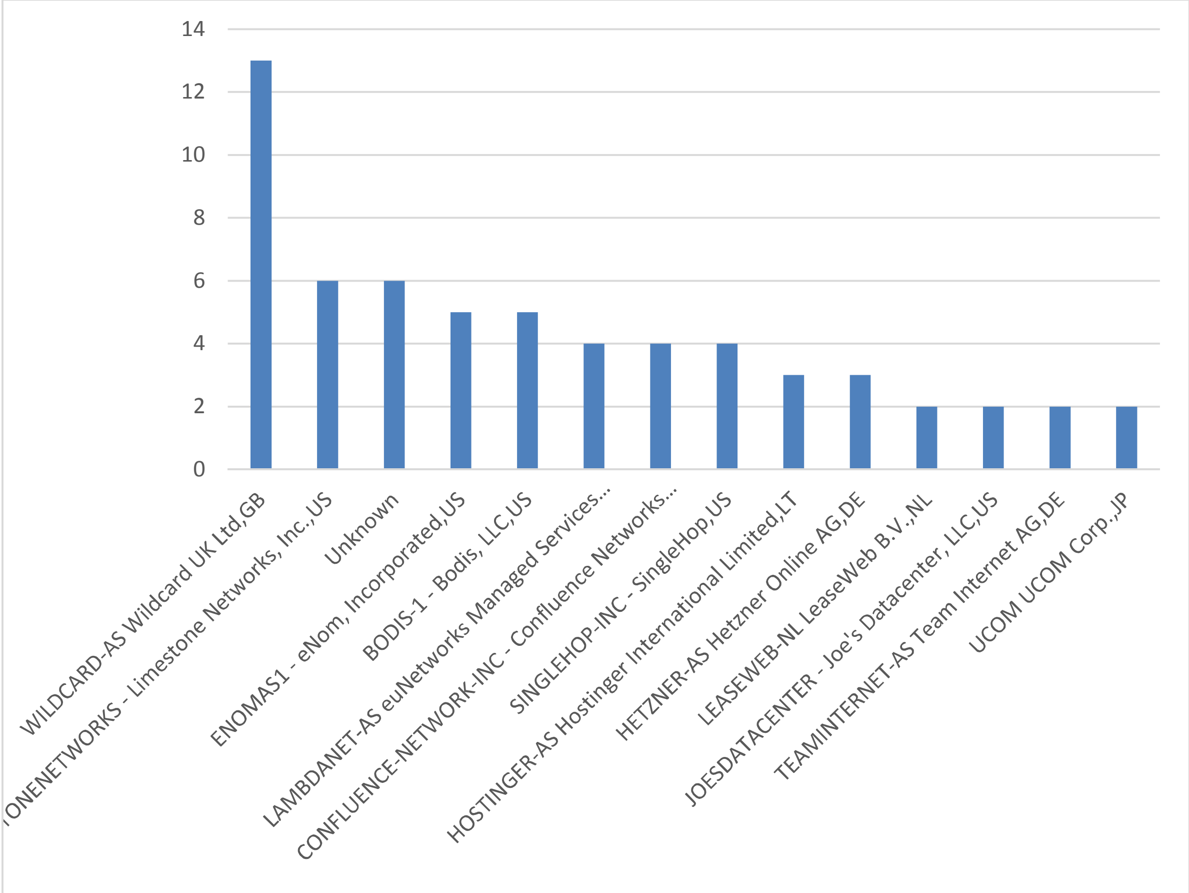

Figure 3: Organizations with Two or More IP Addresses

Interestingly, 130 domain names were associated with one IP address owned by TTNET-MY TIME dotcom located in Malaysia.

4.1.4 Incident Analysis

We do not know the providence of the malware, but based on OSINT research we found that Smallcase was used by at least one APT actor since 2014. We did not find any other linkages between the malware and other threat actors based on OSINT.

4.1.4.1 Kaspersky's "The Darkhotel APT: A Story of Unusual Activity"

In 2014, Kaspersky released a report outlining an APT actor compromising Wi-Fi networks in predominantly Asian hotels. The group has been in operation since at least 2007, and targets particular individuals staying in the hotels primarily to exfiltrate sensitive information. Darkhotel used several zero-day vulnerabilities, and a combination of wateringhole and spearphishing techniques to compromise potential victims. The group is known for its use of code-signing and other digital certificates used to obfuscate communication channels and malicious tools. This particular report released a series of IOCs relating to Darkhotel operations including MD5 hashes and domain names [Kaspersky Global Research and Analysis Team 2014]. In August 2015, Kaspersky reported that the group used an Adobe Flash zero-day exploit found in the Hacking Team leak, and targeted organizations and individuals in Europe, Africa, and Asia [Virus News 2015].

4.1.5 Vulnerability

Through the linkage of Smallcase to Kaspersky's DarkHotel group activity, we were able to examine the group's activity. We identified the vulnerability, CVE-2015-8651, used by the DarkHotel group in December 2015. The vulnerability was considered a zero-day at the time of discovery.

CVE-2015-8651 is an integer overflow found in Adobe Flash Player that would allow attackers to execute arbitrary code via unspecified vectors [Mitre 2016]. The footprint of potential victims is larger because of the widespread use of Adobe Flash. This suggests that the DarkHotel group attempted to exploit a large number of victims.

4.1.6 Exploit

News of vulnerabilities in a popular product such as Adobe Flash spreads quickly through OSINT; however, this trend is not an indication of exploit discovery. The providence of the exploit taking advantage of CVE-2015-8651 is unknown.

- Two files were found on VirusTotal; however, no Exploit-DB proof-of-concepts (PoC) exist at this time. Table 2 provides the results of our search.
- The potential large attack vector makes the use of this exploit very enticing. The analyst's understanding of the exploit's mode of operation becomes that much more important.

Table 2: Exploit MD5s Related to Smallcase Found in VirusTotal per CVE

CVE	Exploit MD5
CVE-2015-8651	06c75822159c48c405e74a3451961a44
CVE-2015-8651	89b865617046db9c68de78a9afd6dd22

4.2 The Derusbi Family

4.2.1 Summary of Findings

Derusbi malware has been involved in many high-profile breaches, including the Office of Personnel Management (OPM), Anthem Health, and Forbes.com compromises in 2014 [Threat Research Team at Threat Connect 2015, Symantec 2016]. The malware was likely developed around 2007 and since then has been significantly retooled by the malware authors, most likely to evade detection and to add functionality. While the malware may contain similar functions to other families such as Codoso and Briba, it is an independent family and should not be conflated with others without additional analysis.

Some vendors and researchers have asserted that Codoso and Derusbi (in particular) are specifically the same—even going so far to naming an operation using Derusbi malware as "Codoso" [Hardy 2012, Kovacs 2015]. Analysts must remain mindful and vet the analysis techniques used in OSINT reporting to draw conclusions. Using code comparison tools, such as Fn2yara, will allow an analyst to derive high-confidence conclusions about similar malware families and potentially similar operations.

Our analysis suggests we have several samples of the Derusbi dropper, which will call out to its C2 infrastructure for additional malware. The network infrastructure used during CNO included 50 IP addresses and 60 domain names linked to the malware and not necessarily found in OSINT. We found that attacks were carried out from 22 organizations, primarily internet service providers (ISPs) and hosting providers. China UNICOM was responsible for at least 12 IP addresses used in CNO. The group (or groups) using Derusbi also used the Taiwanese Academic Network during operations. Our analysis suggests that those using Derusbi cycled through 12 name servers.

Derusbi malware is notorious and linked to several groups in OSINT. In at least two cases, the group exploited zero-day vulnerabilities in Adobe products such as ColdFusion and Flash Player. We found at least 3 exploits related to CVE-2014-9163 and 14 related to CVE-2014-6271; both vulnerabilities were utilized by the actors associated with Derbusi. We believe those using the malware will continue compromising products with widespread use for the purposes of intelligence gathering.

4.2.2 Malware Analysis

OSINT has linked the Codoso malware and Briba as the same family, and in some cases, conflated the two with Derusbi [Hardy 2012, Kovacs 2015]. The outcome of this analysis—if the malware families turned out to be the same—would allow us to have a larger starting point for our analysis. This extra step is not necessary barring confrontation with this type of problem.

Through CERT's malware analysis efforts—reverse engineering, Fn2yara, and code comparisons—we determined these three malware families to be independent, and thus separate from each other. Additional results of this analysis are presented in Appendix A.

We have been tracking the Derusbi malware, since mid-2014. The malware is a RAT with obfuscation techniques to make it hard for network defenders to track. We used 112 Derusbi files as of January 1, 2016 for our analysis.

The configuration dumper provided us with the following as of January 1, 2016:

- 58 unique domain names used for command and control (C2)
- five IP addresses
- malware communicated over ports 53, 80, 443, 1426, 2515, 8080, and 8090

We found that the Derusbi files are less similar compared to the Codoso example, which may indicate polymorphism by the authors rewriting the code. Figure 4 shows the compile times of the Derbusi MD5s. It is likely that Derusbi was written on or before 2006, and the authors continued development of this tool well into 2014. The CERT/CC has been collecting variants monthly to add to the analysis pile.

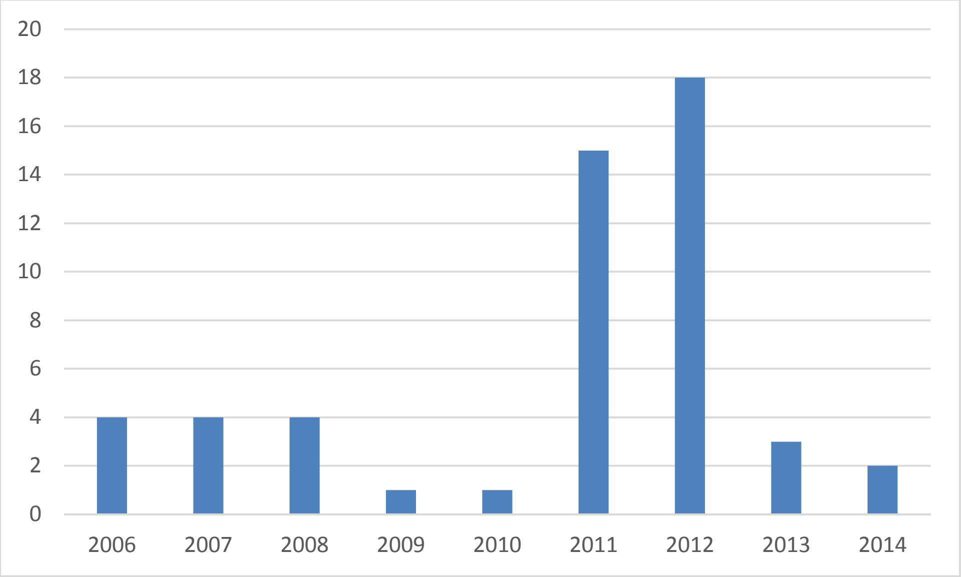

Figure 4: MD5 File Count Based on Derusbi Compile Times

By querying OSINT sources using MD5 hashes for the malware files, we found that Dersubi malware is associated with Crowdstrike's Deep Panda group, RSA's Shell_Crew, Palo Alto Networks' Codoso Team, and Symantec's Black Vine. It is likely that each company examined the intrusion. It is unknown whether each group is similar or different from the other associated with the malware. In 2014, the group was responsible for the Anthem intrusion and used the Derusbi and Sakula malware families to affect millions of healthcare patient records [Threat Research Team at Threat Connect 2015, Symantec 2016].

4.2.3 Network Communications

4.2.3.1 A Records

The conversion of IP address to domain name and vice versa via indicator expansion provided us with the following:

- Two unique domain names were pulled from pDNS data.
- Our analysis pool contained 60 unique domain names combining the configuration dumper information and the pDNS results.
- 45 unique IP addresses were pulled from pDNS data excluding non-routable IP space.
- Our analysis pool contained 50 IP addresses combining the configuration dumper information and the pDNS results.

4.2.3.2 NS and MX Records

Our analysis suggests that the attackers chose a group of name servers for attacks. We deduced this by analyzing the NS and other record types for all of the A records collected during analysis.

- The infrastructure used by Derusbi resolved to 12 unique name servers. At least two of the name servers resolve to cyberthreatssinkhole[.] com.

- There were three MX records in the data with some variation of yahoo[.]jp. This may be indicative of attackers using particular servers to send phishing emails with a legitimate address.

4.2.3.3 ASN Analysis

We found that IP addresses were clustered within 22 organizations as shown in Figure 5. China Unicom located in China held the most IP addresses (12). The majority of organizations of adversary infrastructure were hosting providers or internet service providers (ISPs).

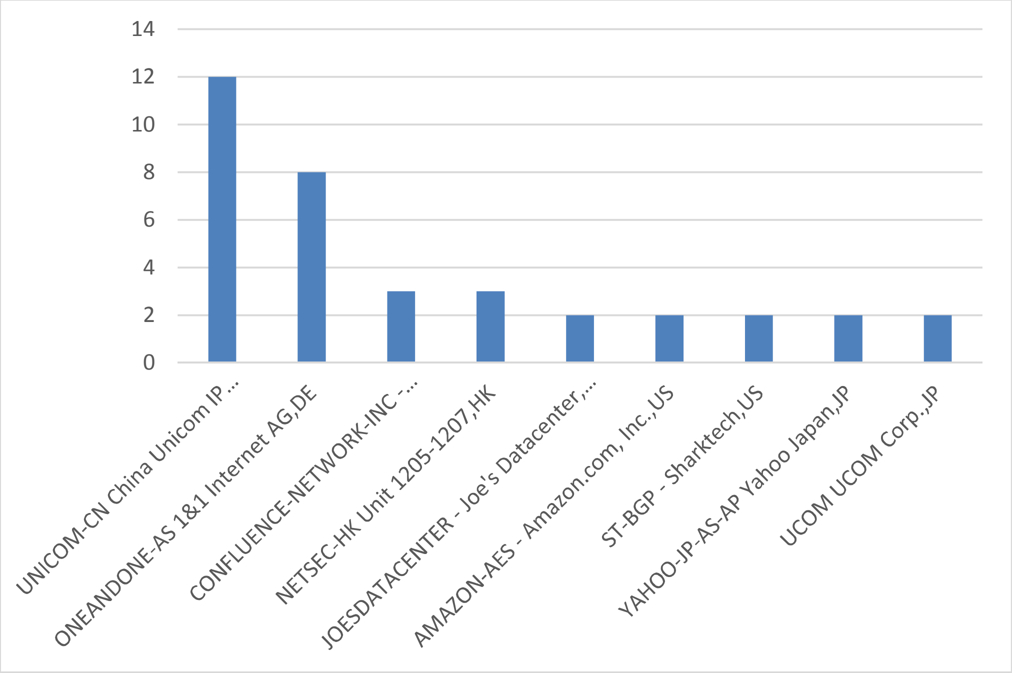

Figure 5: Organizations with Two or more IP Addresses from Derusbi Network Data

4.2.4 Incident, Exploit, and Vulnerability Mapping—Example 1

We do not know the providence of the malware, but based on OSINT research, we found that Derusbi was used by advanced threat actor/s since at least 2012. Security vendors gave the group (or groups) using this tool at least four different names, which can cause confusion for analysis.

4.2.4.1 Incident Analysis

Palo Alto Networks: “New Attacks Linked to Codoso Group”

- As early as 2012, unknown adversaries used a malware family titled Codoso (or Codoso) during CNO against human rights organizations, telecommunications, education, and manufacturing sectors [Hardy 2012, Gruntzweig & Lee 2016]. Some in the security community linked this malware with Briba and Derusbi families [Hardy 2012, Gruntzweig & Lee 2016]. The Codoso group is not attributed to a particular country, but is associated with a compromise of Forbes.com in 2014 [Gruntzweig & Lee 2016].

iSight Partners: “Codoso Team - Watering Hole Style Attack”

- iSight Partners have been tracking the operations of the Codoso Team as far back as 2010. In February 2015, iSight Partners released a report detailing operations of Codoso Team that targeted Forbes.com with Adobe Flash exploits; at the time they were zero-day exploits [iSight Partners 2015]. iSight Partners have seen the Codoso Team target a wide range of victims. The group tends to target its victims through watering hole intrusions as well as spearphishing campaigns [iSight Partners 2015].

4.2.4.2 Vulnerability

OSINT data revealed that the Codoso Team was targeting the vulnerability CVE-2014-9163 in 2014. This vulnerability is a stack-based buffer overflow in Adobe Flash Player that allows attackers to execute arbitrary code via unspecified vectors [Mitre 2016]. Again the wide landscape and usage of Adobe Flash Player made this zero-day vulnerability a critical target and thus easily chosen by attackers.

4.2.4.3 Exploit

Time plays in the favor of the security researcher during exploit discovery. Three exploit files were found on VirusTotal for CVE-2014-9163. However, as with CVE-2015-8651, no Exploit-DB PoC exists for CVE-2014-9163. Table 4 in Appendix B shows the CVE and exploit MD5 pairing.

4.2.5 Incident, Exploit, and Vulnerability Mapping—Example 2

4.2.5.1 Incident Analysis

RSA: “Incident Response Emerging Threat Profile: Shell_Crew”

- In 2014, RSA compiled a report detailing the operations of Shell_Crew, also known as Deep Panda, WebMasters, Kung-fu Kittens, and PinkPanther [RSA 2014]. The group was active well into late 2015 and was responsible for the Office of Personnel Management (OPM) breach [Hesseldahl 2015]. The group uses a series of web shells to obfuscate footprints on a network, utilizes code signing certificates for malware, and uses complex malicious code that evades forensic investigations [RSA 2014]. The group is loosely linked to an operational arm originating from China [Hesseldahl 2015]. Shell_Crew utilized a custom tool called Derusbi during CNO.

Crowdstrike: “Deep Panda”

- Following the receiving of data in December 2011, Crowdstrike produced an intelligence report on an actor that it named Deep Panda, also known as Shell_Crew. The malicious files highlighted in the report were potentially involved in an attack against a large Fortune 500 company [Crowdstrike Global Intelligence Team 2016]. As described by Crowdstrike, Deep Panda group uses a RAT, a post exploitation tool, and complex kernel-level tool as part of its attack platform. The kernel-level tool has been tied to the Derusbi Trojan [Crowdstrike Global Intelligence Team 2016].
- In 2014, Crowdstrike tracked attacks against the Defense Industrial Base (DIB), healthcare, government, and technology sectors [Dahl 2016]. It was determined that these campaigns

were using Derusbi and Sakula families and were eventually linked to Deep Panda [Dahl 2016].

4.2.5.2 Vulnerability

Comparison of our known malware and incident data of two separate findings enabled us to explore further into the operations of Shell_Crew/Deep Panda. The data revealed that the group was found targeting CVE-2010-2861 and CVE-2014-6271 in its operations. CVE-2010-2861 is an Adobe ColdFusion vulnerability that allows the attacker to traverse multiple administrator directories [Mitre 2016]. CVE-2014-6271, also known as Shellshock, was a vulnerability affecting the Unix Bash shell, and allowed attackers to execute arbitrary code via a crafted environment [Mitre 2016]. Both of these vulnerabilities allowed any attacker to target a wide range of victims due to the common use of the vulnerability applications and the lack of patching.

4.2.5.3 Exploit

Both of the above-mentioned vulnerabilities targeted popular applications. For CVE-2010-2861 and CVE-2014-6271, PoCs can be found in Exploit-DB. However, unlike CVE-2014-6271, no files can be found on VirusTotal that are specifically tagged as CVE-2010-2861. VirusTotal did have 14 exploit files for CVE-2014-6271. Figure 5 in Appendix B shows the pairings of CVE to exploit MD5.

4.3 The Sakula Malware Family

4.3.1 Summary of Findings

Sakula was likely developed around 2012. This toolkit contains at least a dropper and an implant used to put additional malware on a machine and exfiltrate information from victim networks. There were no indications that this malware was closely associated with other known families; however, the malware was used alongside Derusbi in high-profile compromises.

The network infrastructure used during CNO included 14 IP addresses and 47 domain names linked to the malware. We noticed that the domain names were binned into the following groups: those that resembled legitimate services, those that resembled the IP address the domain resolved, and a mix of English and Chinese words. We found that attacks were carried out from 13 organizations, primarily hosting providers and ISPs. At least one organization (JOESDATACENTER - Joe's Datacenter, LLC (U.S.)) is associated with other malicious activity. Our analysis suggests that those using Sakula cycled through nine name servers and made particular use of Chinese registrars.

The Sakula malware is associated with both CrowdStrike’s Deep Panda and Symantec’s Black Vine, which used this malware to exfiltrate sensitive information from organizations in the aerospace, healthcare, and energy sectors. This group is associated with the Anthem Health compromise in 2014 and is associated with activity into mid 2015 [Wagstaff 2015]. We found at the group used at least two use-after-free vulnerabilities in Microsoft Internet Explorer (IE) during CNO. We discovered at least three exploits related to CVE-2012-4792 (affecting IE 6 through 8) and eight related to CVE-2014-0322 (affecting IE 9 and 10), both of which are associated with

Black Vine activity. We believe APT groups chose to compromise Microsoft IE due to the prevalence of the software on a variety of machines and will continue actively using vulnerabilities and exploits to compromise a high number of victims.

4.3.2 Malware Analysis

We have been tracking the Sakula malware since mid-2014. The malware has also been referred to as Sakurel [Oleynikov & Jeet n.d.]. Sakula is a RAT, and we identified a dropper and implant. We used 357 Sakula files as of January 1, 2016 for our analysis.

The configuration dumper provided us with the following as of January 1, 2016:

- 14 unique domain names used as C2 infrastructure
- five IP addresses
- malware that communicated over ports 80 and 443
- malware that called out to files with a "media" theme such as newimage.asp, and newvideo.asp. Executables associated with this malware are MediaCenter.exe and AdobeUpdate.exe. Typically, these are stored in file path such as %Temp%\MicroMedia.
- processes disguised as legitimate through providing of names such as Microsoft, Intel, or Adobe

Figure 6 shows the compile times of the Sakula MD5s. It is likely that Sakula was written on or before 2012, and the authors continued development of this tool well into 2014. The CERT/CC has been collecting variants monthly to add to the analysis pile.

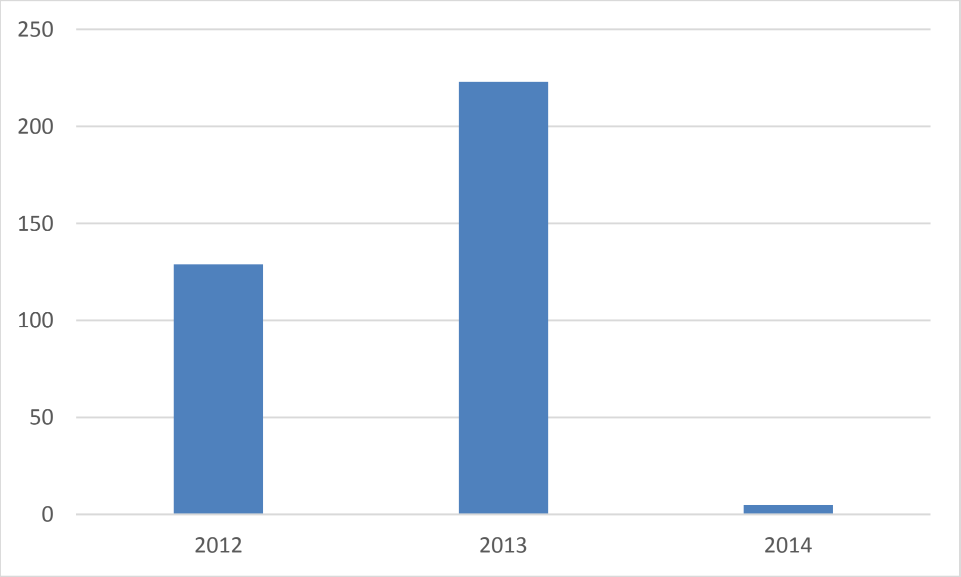

Figure 6: MD5 File Count Based on Sakula Compile Times

4.3.3 Network Communications

4.3.3.1 A Records

The conversion of IP address to domain name and vice versa via indicator expansion provided us with the following:

- 47 unique domain names were pulled from pDNS data.
- Our analysis pool contained 61 unique domain names combining the configuration dumper information and the pDNS results.
- 11 unique IP addresses were pulled from pDNS data, excluding non-routable IP space.
- One domain resolved to an IPv6 address.
- Our analysis pool contained 14 IP addresses combining the configuration dumper information and the pDNS results.

We noticed that a portion of domain names resembled those for legitimate goods or services, such as login[.]bitdefendor[.]com (BitDefender) and citrix[.]vipreclod[.]com (Citrix). Some domains also resembled the IP addresses they resolved to, such as 1111xf[.]66xxaa[.]us; 8777ygoudjg[.]com. There was also a mix of English and Asian words within the domains.

4.3.3.2 NS Records

We found that Sakula's domain infrastructure was associated with at least nine different name servers. The top name server is associated with GoDaddy (DomainControl). The NS records show use of Chinese registrars. Figure 7 shows the name servers for the Sakula malware per domain name.

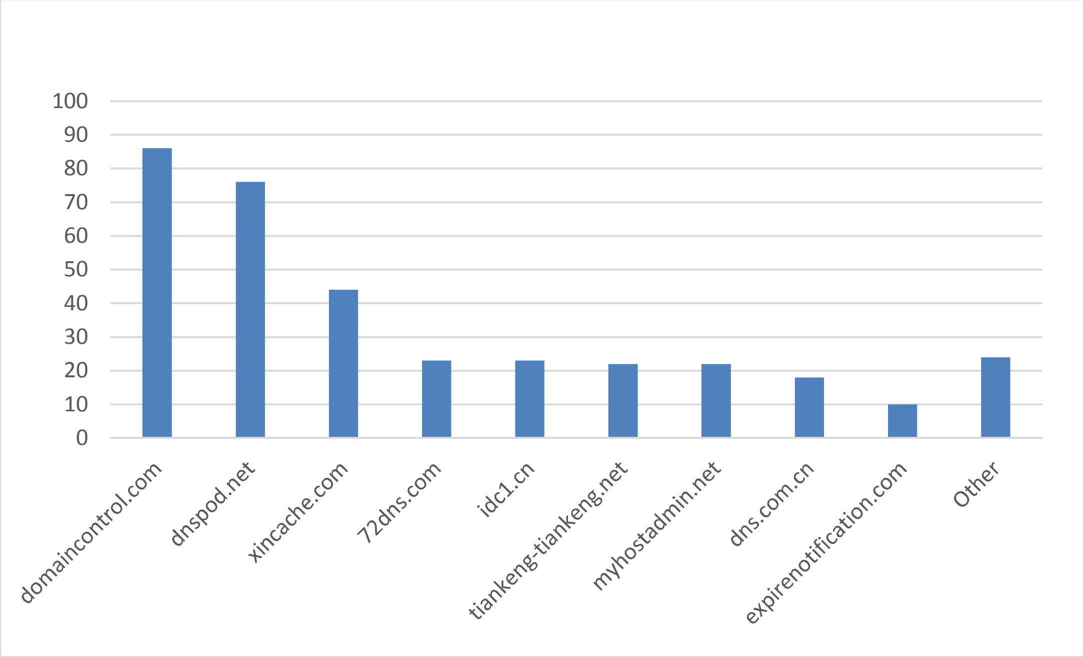

Figure 7: Name Servers per Domains Found in Sakula Infrastructure

4.3.3.3 ASN Analysis

The Sakula network infrastructure is associated with 13 organizations primarily located within the United States. Only AS-26496-GO-DADDY-COM-LLC - GoDaddy.com, LLC, US was responsible for more than one IP address. The organizations and IP count are found in Table 3. The organizations targeted for adversary infrastructure were primarily ISPs or hosting providers.

Table 3: Organizations with IP Addresses from Sakula Network Data

Organization	IP Count
JOESDATACENTER - Joe's Datacenter, LLC, US	1
WEHOSTWEBSITES-COM - WeHostWebSites.com, US	1
RMH-14 - Rackspace Hosting, US	1
SPARKSTATION-SG-AP 10 Science Park Road, SG	1
DIGITALOCEAN-ASN - Digital Ocean, Inc., US	1
IOFLOOD - Input Output Flood LLC, US	1
SBN-FBB-AS-AP Fixed Broadband Network, TH	1
LEASEWEB-NL LeaseWeb B.V., NL	1
AS-26496-GO-DADDY-COM-LLC - GoDaddy.com, LLC, US	2
HURRICANE - Hurricane Electric, Inc., US	1
HANARO-AS Hanaro Telecom Inc., KR	1
ALCHEMYNET - Alchemy Communications, Inc., US	1
EGIHOSTING - EGIHosting, US	1

JOESDATACENTER - Joe's Datacenter, LLC (U.S.) is associated with other malicious activity, including Zusy and Shiz along with other malicious code.

4.3.4 Incident Analysis

Sakula was linked to the Derusbi malware family via CrowdStrike's analysis of Deep Panda. A description is found in 4.2.5. We believe Sakula has been used by actor groups since at least 2012.

4.3.4.1 Symantec's "The Black Vine Cyber Espionage Group"

Black Vine group, dubbed by Symantec, has been conducting cyber operations since 2012 [Symantec 2016]. Its targets include aerospace, healthcare, and energy sectors, which included the breach of the healthcare insurance company Anthem that exposed 80 million patient medical records [Symantec 2016]. Zero-day exploits are used to compromise victims. If the exploits are successful, two variants of custom malware, Hurix and Sakurel, are dropped on the system [Symantec 2016]. The backdoors then created by the successful installation of the malware are leveraged by the Black Vine group to steal information from the target's computer systems.

4.3.5 Vulnerability

The vulnerabilities, CVE-2012-4792 and CVE-2014-0322, were actively used by the Black Vine group. CVE-2012-4792 is a use-after-free vulnerability in Microsoft Internet Explorer versions 6 through 8 that allows remote attackers to execute arbitrary code [Mitre 2016]. This vulnerability was exploited in the wild (ITW) in watering hole campaigns in late 2012. CVE-2014-0322 is similar to CVE-2012-4792 in that it is also a use-after-free vulnerability in Microsoft Internet Explorer affecting versions 9 and 10 [Mitre 2016]. The attack landscape of these vulnerabilities was considered high because of common use of Microsoft Internet Explorer, which is found on Microsoft Windows Operating Systems.

4.3.6 Exploit

Samples of exploits targeting CVE-2012-4792 and CVE-2014-0322 can be found on VirusTotal and Exploit-DB; however, this was not the case in 2012 and 2014, as these exploits were initially considered zero-days. Typically, this means that few individuals have immediate access to the exploit. However, because of the popularity of the targeted application, (Microsoft Internet Explorer), there is an increased interest in exploit discovery and an urgency for vendor patches for the vulnerable application. Table 4 in Appendix B shows the results of the CVE and Exploit MD5 pairing.

5 Data Sources

We used any easily accessible sources such as blogs, news articles, vendor products, Tweets, and other information to aid in our analysis. Exploit information, in particular, was the most difficult information type to find. We also utilized in-house data sources created and cultivated for CERT/CC analysis.

5.1 Malware Family Analysis

CERT/CC has a Malware Catalog containing files suspected of being used in computer security incidents. Our large repository aids in the malware trending, analysis, and reverse-engineering efforts that we publish about and report on.

We analyze and collect malware samples from a variety of sources. A majority of the analysis is rooted in the “knowns” process, or the process of reverse engineering and grouping malware samples by similarity in the code. Once a family of malware is identified, an analyst writes a configuration dumper, to extract the network indicators, strings, and other data from the families. Families can range from a few files to hundreds.

The CERT/CC names malware families based on characteristics in the files or open source data. Each situation is different as to whether analysts choose their own name or not.

5.2 Farsight’s Passive Domain Name Server (pDNS)

To get a wider range of network indictors, we used the pDNS records collected by Farsight, primarily to aid in indicator expansion, the transformation of IP addresses to domain names and vice versa [Ziegast 2010]. These domain names or IP addresses were pulled during a time frame similar to that of the OSINT reporting [Shick & Horneman 2014]. Those who join (i.e., become a sensor) can get access to this data source.

We used the first_seen and last_seen dates in the pDNS data to help bolster a timeline of operations for our analysis. The first_seen field denotes when a response was first seen in the data, while last_seen is the last time the query was made.

5.3 Blacklist Analysis

Since 2013, the CERT/CC analyzed the Blacklist Ecosystem—blacklists of IP addresses and domain names, both public and private—of the security community for network defense and threat intelligence. The work focused first on 25 blacklists, but grew to nearly 100 as of 2016. Lists are compared directly and indirectly, based on data type [Metcalf & Spring 2014].

- The list contents are compared to determine if any list shared indicators before another list. The lists are compared again to determine if there were patterns in the indicator collections [Metcalf & Spring 2014].
- The comparison indicates a range for how often a list provides an indicator with unique information. The comparison also indicates the value to CND [Metcalf & Spring 2014].

For this methodology, we used the blacklists provided from the Blacklist Ecosystem studies to determine if network infrastructure showed up in any of the feeds. Results could provide additional context to an incident and the overall operations of an adversary. We used 35 domain-name and 88 IP-address blacklists. The blacklists are anonymized to preserve consistency with past work, and to avoid favoring of one list over another.

We found that the overlap between malicious domain names and IP addresses was insignificant, and thus we did not include the results. Few to none of the indicators found in the expansion process were found in blacklists.

5.4 Common Vulnerability and Exposure Database (CVE)

To best capture information surrounding vulnerabilities used in attacks, we used the CVE database maintained by MITRE. This data source provides targeted systems/applications and gives idea into who is likely to be impacted and the likelihood of that occurring. We combined this information with open source documents easily found on the web to gain additional context as needed [Mitre 2016].

5.5 Twitter

Often security researchers publish findings on personal blogs and use Twitter as the vehicle to attract attention. Twitter is an excellent resource for finding context around a particular cyber event.

5.6 Exploit Database (Exploit-DB)

Offensive Security maintains the Exploit-DB, which provides researchers exploits and proof of concepts leveraging particular CVEs. It is one of the few existing open source databases that maintain data on known exploits. This database was a primary source for exploit retrieval; however, there is a noticeable lag from when researchers publish an exploit (or when it is found on Twitter) to when it is found on Exploit-DB [Exploit Database 2016]

5.7 Malware Information Sharing Platform (MISP) of Circl.lu

The Computer Incident Response Center Luxembourg (CIRCL) coordinates the release of IOCs for private and public entities within the country. It also offers the MISP, a tool that can be accessed by other country-level incident response teams around the world for dissemination and discussion of IOCs and other information of targeted attacks [Luxembourg CERT 2016].

5.8 Other Data Sources

Several different data sources went into determining who owned the IP addresses or domain names in our analysis and allowing us to understand the infrastructure that may have been chosen to engage in CNO. Those addresses were associated with ASNs using a combination of data from the University of Oregon Route Views Project and the Réseaux IP Européens (RIPE) Network Coordination Centre Routing Information Service (RIS) [Shick & Horneman 2014].

6 Tools

Our analysis relies on the following tools and a few of bash scripts to best extrapolate data from sources.

6.1 YARA

YARA is a tool that enables researchers to identify and classify malware samples in a meaningful way based on text or binary patterns in the code. The tool can be used in a variety of environments including Windows, Linux, and Mac OSX [Manuel n.d.].

6.2 Fn2yara

Fn2yara is a tool built by CERT/CC that enables analysts to generate YARA signatures for matching functions in an executable program [Threat Analysis Team 2016]. This allows for a macro-level code base analysis to compare malware files or families.

6.3 Malware Clone Mapping

Malware clone mapping is a tool and methodology created by the CERT/CC that compares the binaries of two or more files to understand a percentage of similarity. This method rules out accidental relation and can aid in identifying how the code was written by an author. This allows a micro-level comparison of malware as compared to the Fn2yara tool.

6.4 System for Internet-Level Knowledge (SiLK)

SiLK is a set of network flow analysis tools developed by the Software Engineering Institute at Carnegie Mellon University. It allows analysts to quickly dissect network data with a variety of commands. This tool allowed us to easily compare IP sets (`rwsetbuild, rwsetcat, rwsettool --intersect`), and build prefix maps, or pmaps (`rwuniq --pmap`) for analysis [SEI 2013].

6.5 Linux System

A Linux system contained the built-in command line tools for easy data analysis. It also provided simple integration of YARA and Fn2yara for performing the necessary analysis. This is also where we wrote several bash and python scripts.

7 Future Work

This document presents only one methodology for exploring how analysts can think about cyber threats differently. Additional methodologies should be created to identify the strengths and weaknesses of starting or ending with particular points (vulnerabilities, exploits, etc.). Additional investigation into analytic pivoting should be conducted by other researchers.

We identified a gap area in the community's understanding of exploits. While Exploit-DB and VirusTotal are good resources for finding malicious code, they do not necessarily provide a holistic repository of the exploits seen in the wild. We did not explore any avenues to trend or discover exploits; however, this would present an interesting challenge problem for the field.

8 Conclusion

This study sought to examine what is necessary and fruitful for threat analysts to understand particular groups based on the malware used during operations. This method maps, in a non-linear manner, vulnerabilities, exploits, network infrastructure, and adversary behavior beginning with a well-understood malware family. This methodology should be used as an outside-in approach

The goal of this document is to encourage researchers to begin with well-understood malware to create a threat mapping instead of incident data alone. The results of this exercise can be applied to blocking or defense efforts or for intelligence purposes to better understand the adversary.

The results of each case study—any additional network indicators (IPs, domains, autonomous system numbers, malware, exploits, vulnerabilities, and overall understanding of an intrusion tied to the malware families—should be utilized by network defenders and intelligence circles to aid in decision making and analysis.

Appendix A: Malware Code Comparison Findings

8.1 Fn2yara Results for Codoso, Derusbi, and Briba

We compared 244 Briba files, 25 Codoso files, and 183 Derusbi files. The comparative pairings were Derusbi and Codoso, Derusbi and Briba, and Codoso and Briba.

The results show that both Derusbi and Codoso and Derusbi and Briba are in fact their own unique malware family. Codoso and Briba do share several functions; however, this does not indicate that they are the same malware family.

- Derusbi and Codoso shared two functions.
  - One function was found in 4 of the 183 Derusbi files and the another function was found in 11 of the 183 Derusbi files.
- Derusbi and Briba shared seven functions.
  - Five functions were only found in 1 file each of the Derusbi files.
  - Two functions were only found in 2 files each of the Derusbi files.
  - This demonstrates that these 2 malware families are not closely related.
- Codoso and Briba share 10 functions.
  - Two functions are found in 17 files each of the Briba files.
  - One function is found in 209 Briba files.
  - One function is found in 186 Briba files.
  - One function is found in 96 Briba files.
  - One function is found in 132 Briba files.
  - One function is found in 154 Briba files.
  - One function is found in 37 Briba files.
  - One function is found in 114 Briba files.
  - One function is found in 115 Briba files.

8.2 Malware Clone Mapping Results

We found a possible provenance link between one Codoso and one Derusbi file. One fourth of the code is identical. While this is significant, it is not enough of an overlap to determine that the malware families are the same. The Extra-family relationship is expressed in Figure 8.

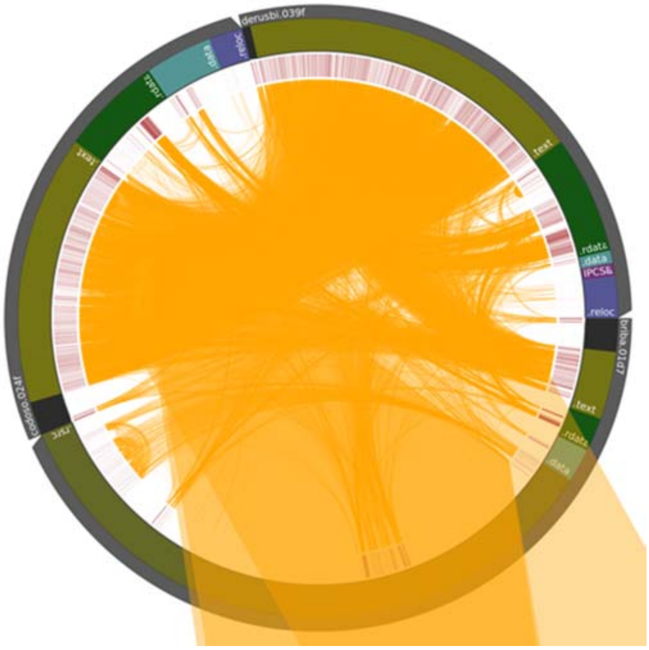

Figure 8: Extra-Family Relationship Among Codoso, Derusbi, and Briba

Appendix B: Exploit and CVE Tables

Table 4: Exploit MD5s per CVE Related to Smallcase Found in VirusTotal

CVE	Exploit Hash
CVE-2014-9163	ca5a35d71a01aaecc28877d316230d20
CVE-2014-9163	f81e20c5059fe1d364080e51974418d8
CVE-2014-9163	faa74be286c58be616470558d78a137f

Table 5: Exploit MD5s per CVE Related to Derusbi Found in VirusTotal

CVE	Exploit MD5
CVE-2014-6271	a4e3a74a1096a5d3b7429b65a4988ac3
CVE-2014-6271	2e9035888dc073d1b0491a20c6c1b7b6
CVE-2014-6271	58e3ac8313bf53e9c5c83c9ae11535a0
CVE-2014-6271	f042552026a68b44f0afbec1a996a9a8
CVE-2014-6271	066d5f0c33f731d1c5b1e832c27ae426
CVE-2014-6271	44b691803534e18416cbe556c0df3c1a
CVE-2014-6271	591acece5004ca64f03249a58a3a8e05
CVE-2014-6271	bad8248397050a6cc9f03f6635fd5fdb
CVE-2014-6271	a5b8af8bb047cad57bab684da35582fb
CVE-2014-6271	7d0b0015920a4898e6d2c178d25c1afd
CVE-2014-6271	7ba71e1b4ce3d582d532021df0e6eeae

CVE	Exploit MD5
CVE-2014-6271	e5855dfcf2e2b1524e6040246679ee3c
CVE-2014-6271	6bc895478c5925efd98b47027da23437
CVE-2014-6271	13263896c57b153946d693c03e08cb87

Table 6: Exploit MD5s per CVE Related to Sakula Malware Found in VirusTotal

CVE	Exploit MD5
CVE-2012-4792	098b4f6e66ca3e6ddc363b8a08ad474f
CVE-2012-4792	dc459fddc87aca57e0635b748813cfbd
CVE-2012-4792	08a23b76f184aaa0f656e6e2d64926ad
CVE-2014-0322	3654c907ce3a1098f29d43e4431a8a7f
CVE-2014-0322	70551f3bfb454c2344a8ae700a83d47e
CVE-2014-0322	cc9a1052ea161719e32cff23bd1575c7
CVE-2014-0322	83d478b6c609f47c75b23eb24971edb0
CVE-2014-0322	242f805a1ddf6610622ef8d920071433
CVE-2014-0322	01aaae4ba3260c55a7c0889d665b47d5
CVE-2014-0322	8beb88a76b45bb6c5ed73083f2cd3184
CVE-2014-0322	9cc4f65a2bff4973ec265040bd15e603
CVE-2014-0322	edfc8feedd43d6b7b335065536d0a42d

Bibliography

URLs are valid as of the publication date of this document.

[@PhysicalDrive0 2016]

@PhysicalDrive0. @PhysicalDrive0 Twitter Profile Page. *Twitter.* March 4, 2016. https://twitter.com/PhysicalDrive0

[CERT Coordination Center 2016]

CERT Coordination Center. Vulnerability Notes Database. *CERT/CC.* 2016. https://www.kb.cert.org/vuls/

[Crowdstrike Global Intelligence Team 2016]

Crowdstrike Global Intelligence Team. *Crowdstrike*. February 2016. Report was found via OSINT and immediately from Crowdstrike's website. The report is footnoted as 'Sensitive.' http://www.crowdstrike.com

[Dahl 2016]

Dahl, Matt. I am Ironman: DEEP PANDA Uses Sakula Malware to Target Organizations in Multiple Sectors. *Crowdstrike.* March 23, 2016 [accessed]. http://www.crowdstrike.com/blog/ironman-deep-panda-uses-sakula-malware-target-organizations-multiple-sectors/

[Exploit Database 2016]

Exploit Database. Who We are. *Exploit-db.* March 16, 2016 [accessed]. https://www.exploit-db.com/about/

[FireEye 2014]

FireEye. APT28: A Window into Russia's Cyber Espionage Operations. *FireEye.* 2014. https://www2.fireeye.com/rs/fireeye/images/rpt-apt28.pdf

[Gruntzweig & Lee 2016]

Gruntzweig, Todd & Lee, Bryan. New Attacks Linked to Codoso Group. *Palo Alto.* Januray 22, 2016. http://researchcenter.paloaltonetworks.com/2016/01/new-attacks-linked-to-Codoso-group/#more-11940

[Hardy 2012]

Hardy, Seth. iExplore Rat. *Citizen Lab Technical Brief.* August 2012. https://citizenlab.org/wp-content/uploads/2012/09/IEXPL0RE_RAT.pdf

[Hesseldahl 2015]

Hesseldahl, Erik. FireEye Identifies Chinese Group Behind Federal Hack. *Re/Code.* June 9, 2015. http://recode.net/2015/06/19/fireeye-identifies-chinese-group-behind-federal-hack/

[iSight Partners 2015]
iSight Partners. Codoso Team Watering Hole Style Attack. *iSight Partners.* February 10, 2015. http://www.isightpartners.com/2015/02/codoso/

[Kafeine 2016]
Kafeine. Malware Don't Need Coffee. *Malware Don't Need Coffee.* March 4, 2016. http://malware.dontneedcoffee.com/

[Kaspersky Global Research and Analysis Team 2014]
Kaspersky Global Research and Analysis Team. *The Darkhotel APT: A Story of Unusual Hospitality*. November 11, 2014. https://securelist.com/files/2014/11/darkhotel_kl_07.11.pdf

[Kovacs 2015]
Kovacs, Eduard. New Attacks Attributed to Chinese Group "Codoso." *Security Week.* January 26, 2015. http://www.securityweek.com/new-attacks-attributed-chinese-group-codoso

[Krebs 2011]
Krebs, Brian. Homegrown: Rustock Botnet Fed by U.S. Firms. *Krebs on Security.* March 2011. http://krebsonsecurity.com/2011/03/homegrown-rustock-botnet-fed-by-u-s-firms/

[Luxembourg CERT 2016]
Luxembourg CERT Mission Statement. *Computer Incidence Responce Center Luxembourg.* February 29, 2016 [accessed]. http://circl.lu/mission/

[Manuel n.d.]
Manuel, Victor. Yara. *Github - VirusTotal.* https://plusvic.github.io/yara/

[Metcalf & Spring 2014]
Metcalf, Leigh & Spring, Jonathan. *Blacklist Ecosystem Analysis Update: 2014.* Software Engineering Institute, Carnegie Mellon University, 2014. http://resources.sei.cmu.edu/library/asset-view.cfm?assetid=448123

[Mismo 2015]
Mismo, Michael. Relentless Sofacy APT Attacks Armed With Zero Days, New Backdoors. *Kaspersky Labs.* December 2015. https://threatpost.com/relentless-sofacy-apt-attacks-armed-with-zero-days-new-backdoors/115556/#sthash.LrQ5lblh.dpuf

[MITRE 2016]
Mitre. Common Vulnerabilities and Exposures. *CVE.* https://cve.mitre.org

[Oleynikov & Jeet n.d.]
Oleynikov, Maksim & Maksim, Jeet. Trojan.Sakurel. *Symantec Security Response.* https://www.symantec.com/security_response/writeup.jsp?docid=2014-022401-3212-99

[RSA 2014]

RSA. Incident Response Emerging Threat Profile: Shell_Crew. *RSA*. January 2014. https://www.emc.com/collateral/white-papers/h12756-wp-shell-crew.pdf

[Shick & Horneman 2014]

Shick, Deana & Horneman, Angela. *Investigating Advanced Persistent Threat 1 (APT1).* (*CMU/SEI-2014-TR-001).* Software Engineering Institute, Carnegie Mellon University, 2014. http://resources.sei.cmu.edu/library/asset-view.cfm?assetid=90426

[SEI 2013]

Software Engineering Institute,Carnegie Mellon University. CERT NetSA Security Suite. 2013. http://tools.netsa.cert.org/silk/index.html

[Spring & Stoner 2015]

Spring, Jonathan & Stoner, Edward. *CND Equities Strategy*. Software Engineering Institute Carnegie Mellon University, 2015. http://resources.sei.cmu.edu/library/asset-view.cfm?assetid=442305

[Symantec 2016]

Symantec. The Black Vine Cyber Espionage Group. *Symantec.* March 24, 2016. http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/the-black-vine-cyberespionage-group.pdf

[Threat Analysis Team 2016]

Threat Analysis Team. cmu-sei pharos. *Github.* 2016. https://github.com/cmu-sei/pharos

[Threat Research Team at ThreatConnect 2015]

Threat Research Team at ThreatConnect. The Anthem Hack: All Roads Lead to China. *Threat Connect.* February 27, 2015. https://www.threatconnect.com/the-anthem-hack-all-roads-lead-to-china/

[Virus News 2015]

Virus News. “Darkhotel” Cyberespionage Group Boosts Attacks with Exploit Leaked from Hacking Team. *Kaspersky Labs.* August 2015. http://www.kaspersky.com/about/news/virus/2015/Darkhotel-Cyberespionage-Group-Boosts-Attacks-with-Exploit-Leaked-from-Hacking-Team

[Wagstaff 2015]

Wagstaff, Jermey. Hunt for Deep Panda intensifies in trenches of U.S.-China cyberwar. *Reuters.* June 2015. http://www.reuters.com/article/us-cybersecurity-usa-deep-panda-idUSKBN0P102320150621

[Ziegast 2010]
Ziegast, Eric. Introduction to SIE. *Flocon 2010.* Cited: February 19, 2016 [accessed].
http://www.cert.org/flocon/2010/proceedings.htmls

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE April 2016	3. REPORT TYPE AND DATES COVERED Final

4. TITLE AND SUBTITLE

A Unique Approach to Threat Analysis Mapping: A Malware-Centric Methodology for Better Understanding the Adversary Landscape A Unique Approach to Threat Analysis Mapping: A Malware-Centric Methodology for Better Understanding
the Adversary Landscape

5. FUNDING NUMBERS

FA8721-05-C-0003

6. AUTHOR(S)

Deana Shick; Kyle O'Meara

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

8. PERFORMING ORGANIZATION REPORT NUMBER

CMU/SEI-2016-TR-004

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFLCMC/PZE/Hanscom
Enterprise Acquisition Division
20 Schilling Circle
Building 1305
Hanscom AFB, MA 01731-2116

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

n/a

11. SUPPLEMENTARY NOTES

12A DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified/Unlimited, DTIC, NTIS

12B DISTRIBUTION CODE

13. ABSTRACT (MAXIMUM 200 WORDS)

Malware family analysis is a constant process of identifying exemplars of malicious software, recognizing changes in the code, and producing groups of "families" used by incident responders, network operators, and cyber threat analysts. With adversaries constantly changing network infrastructure, it is easy to lose sight of the tools consistently being used and updated by these various actors. Beginning with malware family analysis, this methodology seeks to map vulnerabilities, exploits, additional malware, network infrastructure, and adversaries' using Open Source Intelligence (OSINT) and public data feeds for the network defense and intelligence communities. The results provide an expanded picture of adversaries' profile rather than an incomplete story.

The goal of this document is to shift the mindset of many researchers to begin with the tools used by adversaries rather than with network or incident data alone for an "outside-in" approach to threat analysis instead of an "inside-out" method. We chose three malware families to use as case studies—Smallcase, Derusbi, and Sakula.

The results of each case study—any additional network indicators, malware, exploits, vulnerabilities, and overall understanding of an intrusion—tied to the malware families should be utilized by network defenders and intelligence circles to aid in decision making and analysis.

14. SUBJECT TERMS

Threat Analysis, Malware Analysis, Data fusion, Network Defense, Threat Intelligence

15. NUMBER OF PAGES

42

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89) Prescribed by ANSI Std. Z39-18
298-102