Investigating Advanced Persistent Threat 1 (APT1)

Deana Shick
Angela Horneman

May 2014

TECHNICAL REPORT
CMU/SEI-2014-TR-001

CERT® Division

http://www.sei.cmu.edu
# Table of Contents

## Acknowledgments

ix

## Abstract

ix

1  Introduction

1

2  Data Sources

2

2.1  JIBs INC260425 and INC260425-2

2

2.2  Internet Census 2012

2

2.3  Security Information Exchange at the Internet System Consortium (SIE@ISC): rDNS Data

5

2.4  Mandiant’s Report on APT1: Exposing One of China’s Cyber Espionage Units

5

2.5  Open Resolvers Data

5

2.6  Neustar GeoPoint Data

6

2.7  Internet Storm Center Data

6

2.8  Other Data Sources

6

3  Data Issues

7

3.1  Available Data Is a Partial View of APT1 Operations

7

3.2  Timeline of Data

7

3.3  Internet Census Data Constraints

7

4  Tools

8

4.1  Linux Commands

8

4.2  SiLK

8

4.3  Excel

8

4.4  Python Scripts

9

5  Indicator Expansion

10

5.1  Defining Variables and Expansions

10

5.2  Defining the Same Type of Data from Two Sources

10

5.3  Defining the Range of a Data Set

10

6  Method

11

6.1  Variables Used

11

6.2  APT1 Address Set

11

6.2.1  \( I_j \): JIBs’ IP Addresses

11

6.2.2  \( I_D \): IP Address Set in JIB to Domain Names

12

6.2.3  \( I_m = D_m \cap I_j \): Mandiant IP Address Set

12

6.2.4  \( I_j \cap I_m \): Union of Mandiant and JIB IP Address Sets

12

6.2.5  Internet Census Data

13

6.2.6  \( I_j \cup I_m \cap A \): Autonomous System Numbers (ASNs)

15

6.2.7  \( I_j \cup I_m \cap R \) and \( I_j \cup I_m \cap C \): Routing Data and Country Code

15

6.2.8  \( I_j \cup I_m \cap O \): Open Resolvers

16

6.2.9  \( I_m \cap M \): Malicious Code

16

6.3  IPv4 Address Sample

16

6.3.1  \( S \): Sample Set

16

6.3.2  \( S \cap R \) and \( S \cap C \): Routing Data and Country Code

17

6.3.3  \( S_{75\%} \): Fingerprints of the S Set Given a 75% Match

17

6.3.4  \( S \cap P \): Port Analysis

17

6.3.5  \( S \cap R \) and \( S \cap C \): Routing Data and Country Code

18
7 Results: APT1

7.1 ID: JIB Domain Names
    7.1.1 Deliberately Deceptive Domain Names
    7.1.2 Pseudo-Random Alphanumeric Domain Names
    7.1.3 Malicious Domain Names

7.2 IDM Malicious Code

7.3 DI\textsubscript{m}: Mandiant IP Address Set

7.4 I\textsubscript{j}+I\textsubscript{m}: IP Addresses Concatenated from Both Working IP Address Sets

7.5 I\textsubscript{j}+I\textsubscript{m}F\textsubscript{p75}: TCP/IP Fingerprints of I\textsubscript{j}+I\textsubscript{m} Given a 75% Match
    7.5.1 (I\textsubscript{j}+I\textsubscript{m}F\textsubscript{p75})\textsuperscript{2}: TCP/IP Fingerprints with Multiple Instances
    7.5.2 I\textsubscript{j} to I\textsubscript{m} TCP/IP Fingerprint Comparison

7.6 I\textsubscript{j}+I\textsubscript{m}P: Port Analysis
    7.6.1 Most Common Open Ports in I\textsubscript{j}+I\textsubscript{m}
    7.6.2 I\textsubscript{j} to I\textsubscript{m} Port Comparison

7.7 I\textsubscript{j}+I\textsubscript{m}A: Autonomous System Numbers
    7.7.1 Types of Organizations Represented
    7.7.2 Targeted Organizations

7.8 I\textsubscript{j}+I\textsubscript{m}C: Location of IP Addresses

7.9 I\textsubscript{j}+I\textsubscript{m}R: Routing Data

7.10 I\textsubscript{j}+I\textsubscript{m}O: Open Resolvers

7.11 APT1 Analysis Summary

8 Results: IPv4 Sample

8.1 SF\textsubscript{p75}: TCP/IP Fingerprints of S Given a 75% Match
    8.1.1 Top TCP/IP Fingerprints

8.2 I\textsubscript{j}+I\textsubscript{m}F\textsubscript{p75} and SF\textsubscript{p75} Comparison
    8.2.1.1 I\textsubscript{j}+I\textsubscript{m}F\textsubscript{p75}

8.3 SP: Open Ports of S set
    8.3.1 Most Common Open Ports

8.4 I\textsubscript{j}+I\textsubscript{m}P and SP Comparison

8.5 SC: Location of IP Addresses for S Set
    8.5.1 I\textsubscript{j}+I\textsubscript{m}C and SC Comparison

8.6 SR: Routing Data for S set
    8.6.1 I\textsubscript{j}+I\textsubscript{m}R and SR Comparison

9 Discussion

9.1 Infrastructure Owners

9.2 Location

9.3 Fingerprint Data

9.4 Malicious Code

9.5 Port Data

9.6 Intermediary C2 Servers

9.7 Malware Servers

9.8 Hop Points and Inactive Servers

10 Future Work

11 Conclusion

Appendix: DNS Names Pulled from the Security Information Exchange Data

Bibliography
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fingerprints with Five or More Occurrences</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>Fingerprints with Two to Four Occurrences</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>Fingerprint Comparison of $I_i$ and $I_m$</td>
<td>24</td>
</tr>
<tr>
<td>4</td>
<td>Count of Open Ports per IP Address in $I_i$ and $I_m$</td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>Count of IP Addresses for Ports Occurring 20+ Times in $I_i$ and $I_m$</td>
<td>26</td>
</tr>
<tr>
<td>6</td>
<td>Port Comparison of $I_i$ and $I_m$</td>
<td>27</td>
</tr>
<tr>
<td>7</td>
<td>Organizations with 10 or More IP Addresses</td>
<td>28</td>
</tr>
<tr>
<td>8</td>
<td>U.S. States with Five or More Associated IP Addresses</td>
<td>29</td>
</tr>
<tr>
<td>9</td>
<td>Non-U.S. Countries with Two or More Occurrences</td>
<td>30</td>
</tr>
<tr>
<td>10</td>
<td>Neustar Routing Types in $I_i$ and $I_m$</td>
<td>31</td>
</tr>
<tr>
<td>11</td>
<td>Top 30 Fingerprints in S’s IP Addresses</td>
<td>34</td>
</tr>
<tr>
<td>12</td>
<td>Top 30 Fingerprints in S and Top 10 Overlap with $I_i$ and $I_m$</td>
<td>35</td>
</tr>
<tr>
<td>13</td>
<td>Count of IP Addresses per Top 30 Ports in S</td>
<td>36</td>
</tr>
<tr>
<td>14</td>
<td>U.S. States Containing More Than 180 IP Addresses</td>
<td>39</td>
</tr>
<tr>
<td>15</td>
<td>Countries Containing More Than 350 IP Addresses</td>
<td>39</td>
</tr>
<tr>
<td>16</td>
<td>Neustar Routing Types in S</td>
<td>41</td>
</tr>
</tbody>
</table>
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>SyncScan Examples</td>
<td>3</td>
</tr>
<tr>
<td>Table 2</td>
<td>Fingerprint Examples</td>
<td>3</td>
</tr>
<tr>
<td>Table 3</td>
<td>Indicator Expansion Variables</td>
<td>11</td>
</tr>
<tr>
<td>Table 4</td>
<td>Fingerprinting Script Results</td>
<td>13</td>
</tr>
<tr>
<td>Table 5</td>
<td>Recognizable Organizations Found in Domain Names and Frequency of Occurrence</td>
<td>19</td>
</tr>
<tr>
<td>Table 6</td>
<td>Top Five Fingerprints in I+Im</td>
<td>23</td>
</tr>
<tr>
<td>Table 7</td>
<td>Top 10 Fingerprint Comparison of I and Im</td>
<td>24</td>
</tr>
<tr>
<td>Table 8</td>
<td>Ports Representing 10+% of Open Ports</td>
<td>26</td>
</tr>
<tr>
<td>Table 9</td>
<td>Top Five Organizations in Registering IP Addresses in I+Im</td>
<td>28</td>
</tr>
<tr>
<td>Table 10</td>
<td>Organizations Associated with More Than One Open Resolver</td>
<td>31</td>
</tr>
<tr>
<td>Table 11</td>
<td>Top Five Fingerprints in S</td>
<td>33</td>
</tr>
<tr>
<td>Table 12</td>
<td>Operating System Comparison Between I+Im and S</td>
<td>34</td>
</tr>
<tr>
<td>Table 13</td>
<td>Top Fingerprint Comparison Between I+Im and S</td>
<td>35</td>
</tr>
<tr>
<td>Table 14</td>
<td>Ports Representing 5+% of Open Ports</td>
<td>36</td>
</tr>
<tr>
<td>Table 15</td>
<td>Port Data Comparison Between I+Im and S, All IP Addresses</td>
<td>37</td>
</tr>
<tr>
<td>Table 16</td>
<td>Port Data Comparison Between I+Im and S, Open Port IP Addresses Only</td>
<td>38</td>
</tr>
<tr>
<td>Table 17</td>
<td>Comparison of IP Location Between I+Im and S</td>
<td>40</td>
</tr>
<tr>
<td>Table 18</td>
<td>Comparison of Routing Type Between I+Im and S</td>
<td>41</td>
</tr>
</tbody>
</table>
Acknowledgments

We would like to thank Michael Hanley and Timur Snoke, who guided us through the process of compiling this report; Jonathan Spring and Edward Stoner, who pulled data from internal databases that we used in our analysis; and Rhiannon Weaver, who helped us with the sample and statistics.
Abstract

In February 2013, Mandiant uncovered Advanced Persistent Threat 1 (APT1)—one of China’s alleged cyber espionage groups—and provided a detailed report of APT1 operations, along with 3,000 indicators of the group’s activity since 2006. This report analyzes unclassified data sets in an attempt to understand APT1’s middle infrastructure: the system of hops, distribution points or relays, and the command and control (C2) servers that sit between APT1’s victims and main C2 servers located overseas. To build that infrastructure, APT1 chose and exploited particular organizations to obfuscate communications while remaining in plain sight.

This analysis, based on data from IP addresses known to be associated with APT1 and domain names provided by Mandiant, was conducted using a combination of System for Internet Level Knowledge (SiLK) tools, Microsoft Excel, and custom Python scripts. The study detailed in this report can be replicated easily using available sources and tools. By combining key unclassified information, the authors successfully described a large, malicious network used to steal important information.
1 Introduction

Since Mandiant released its report regarding China’s suspected cyber espionage unit, Advanced Persistent Threat 1 (APT1), in February 2013, there has been increased media attention on cyber capabilities in the international arena. Although the details of cyber missions are hidden under veils of secrecy, the footprints left behind during and after execution are increasingly apparent in various pieces of unclassified data. Our goal is to marry unclassified data sets to understand the infrastructure of APT1. We want to understand both what is necessary for APT1 to be successful in its mission and why the group chose particular machines as the vehicle for the exfiltration of sensitive data. This study uses IP addresses known to be associated with APT1, collected from the Joint Indicator Bulletins (JIBs) INC260425 and INC260425-2, and domain names provided by Mandiant. We analyzed this combined set against several unclassified data sets: the Internet Census 2012, Open Resolver list, Neustar GeoPoint data, and the Internet Storm Center data. We include indicator expansion algebra to best capture our process.

This study does not include targets; rather, it is aimed at the vast network of middle infrastructure, which we define as the pieces that sit between end targets and the home base. This middle infrastructure comprises intermediary command and control (C2) servers, malware servers, and hop points used to push sensitive information along to APT1. Our data begins with combining the JIBs and the reverse domain name server (rDNS) information from Mandiant’s report [Mandiant 2013]. The resulting 1,386 IP addresses form our analysis domain.
2 Data Sources

2.1 JIBs INC260425 and INC260425-2

On February 18, 2013 and February 26, 2013, the U.S. Department of Homeland Security (DHS) and Federal Bureau of Investigation (FBI) released two Traffic Light Protocol (TLP) Green documents of APT1 internet protocol (IP) addresses, hostnames, and second-level domains known to be associated with APT1 activity. Combining the two JIB IP sets [US-CERT 2013a, 2013b] into one was the first step of this study.

2.2 Internet Census 2012

The Internet Census 2012 is a project that scanned the IPv4 address space using the nmap Scripting Engine (NSE) between March and December 2012 [Unknown 2013]. The nmap open-source security tool is used to enumerate hosts on a network and compile a network map of them. As an nmap tool, NSE allows users to write unique scripts to use nmap for their individual needs [nmap.org 2013]. Instead of scanning a personal network, the creator of the Internet Census (identity unknown) used a personal NSE script to port scan the /0 address space. The creator conducted the scan by developing the Carna botnet that deployed a small binary onto a group of nonsecure sample machines [Unknown 2013]. These nonsecure machines were used to build a port scanner for the entire IPv4 address space. The botnet consisted of a central server for data collection and analysis, middle nodes to transmit large pieces of data to the central server, and many devices for data collection. The data collection machines scanned other machines via a slew of Python scripts from the binary on Telnet, port 23, which transferred data back to middle nodes. Those nodes would forward off the information to the main server for data collection and analysis [Unknown 2013]. The data was released into the public domain in early 2013 for further research. The Internet Census contains 9.6 terabytes of data. The following types of scans were performed and saved by the conductor of the Internet Census, but not necessarily used in our analysis:

- Internet Control Message Protocol (ICMP) ping: When an ICMP ping scan is sent to a device, it measures the time the device takes to return a response. The response will return one of the following: network unreachable, host unreachable, communication administratively prohibited, alive, alive from A.B.C.D, or unreachable from A.B.C.D. For the purposes of the IPv4 scan, the ICMP ping scan results tell if the address was reachable and, if not, what IP address was used for the scan. This scan sent an ICMP ping to every address every few days [Unknown 2013].

- reverse DNS (rDNS) scan: This scan queried every domain name for the IPv4 address space, sending queries to the top 16 DNS servers for each IP address. Once compiled, the DNS records were sent back to the collection server [Unknown 2013].

- service probes: Service probe scans look for responses on various ports that can be attributed to certain services. The probed IP address would return a state of 1 for open, 2 for open reset, 3 for open timeout, 4 for closed reset, and 5 for timeout. Only IP addresses returning a state of 1 provide any information in the request. Although 175 billion probes were saved, only a small minority of the data showed any useful information [Unknown 2013].

- host probe: Host probe scans relay to a researcher regardless of whether the IP address exists [Unknown 2013].
- SyncScan queries: SyncScan queries inform a researcher if a particular IP address has open, open-filtered, or closed ports. This scan contains every IP address in the IPv4 address space. The SyncScan data provided in the Internet Census is divided by CIDR /8 netblock; each /8 netblock contains its own unique file [Unknown 2013]. The data lists all the open, open-filtered, and closed ports associated with each IP address in the /8 netblock. In many cases, the IP address in question has all three statuses and therefore three different lines in the file. The SyncScan query also provides the type of packet sent to a particular IP address, a UNIX timestamp for when the data was received, and the type of packet sent to each port. Table 1 shows examples of an APT1 IP address listing in an Internet Census SyncScan file.

Table 1: SyncScan Examples

<table>
<thead>
<tr>
<th>IP Address</th>
<th>Timestamp</th>
<th>Status</th>
<th>Reason</th>
<th>TCP/UDP</th>
<th>Open Ports</th>
</tr>
</thead>
<tbody>
<tr>
<td>38.104.203.222</td>
<td>1334544300</td>
<td>open</td>
<td>syn-ack</td>
<td>tcp</td>
<td>1723,3389</td>
</tr>
<tr>
<td>38.104.203.222</td>
<td>1335746700</td>
<td>closed</td>
<td>reset</td>
<td>tcp</td>
<td>20,21,22,23,25,53,80,110,111,143, 443,993,995, 3306,5900,8080</td>
</tr>
</tbody>
</table>

- TCP/IP fingerprint: This scan attempts to gather information that will identify the type of device and the operating system (OS) running on the machine [Unknown 2013]. It appears that the nmap tool used in the scan was modified to store the response used in the determination of the fingerprint. The standard nmap program captured the OS results but did not keep the actual response used in fingerprinting. The OS results were not provided in the Internet Census data. Like the SyncScan queries, the fingerprint data is divided by CIDR /8 netblock; each /8 netblock contains its own unique file.

Table 2 shows examples of APT1 IP address listings in a fingerprint file.

Table 2: Fingerprint Examples

<table>
<thead>
<tr>
<th>IP Address</th>
<th>Timestamp</th>
<th>Fingerprint</th>
</tr>
</thead>
<tbody>
<tr>
<td>161.58.177.111</td>
<td>1346465700</td>
<td>SCAN(V=6.01%E=4%D=1/2%OT=21%CT=22%CU=%PV=N%DC=I%G=N%TM=19DB3%P=mipsel-openwrt-linux-gnu), SEQ(SP=105%GCD=1%ISR=10%TI=1%CI=1%II=1%SS=S%TS=0), OPS(O1=M5B4NW0NNT00NNS%O2=M5B4NW0NNT00NNS%O3=M5B4NW0NNT00NNS%O4=M5B4NW0NNT00NNS%O5=M5B4NW0NNT00NNS%O6=M5B4NW0NNT00NNS%W1=4000%W2=4000%W3=4000%W4=4000%W5=4000%W6=4000), ECN(R=Y%DF=N%TG=80%W4=4000%O=M5B4NW0NNT00NNS%CC=N%Q=), T1(R=Y%DF=N%TG=80%S=O%A=S+%F=AS%RD=0%Q=), T2(R=Y%DF=N%TG=80%S=Z%A=S+%F=AR%O=RD=0%Q=), T3(R=Y%DF=N%TG=80%S=O%A=S+%F=AS%O=M5B4NW0NNT00NNS%RD=0%Q=), T4(R=Y%DF=N%TG=80%S=0%A=S+%F=AS%RD=0%Q=), T5(R=Y%DF=N%TG=80%S=Z%A=S+%F=AR%O=RD=0%Q=), T6(R=Y%DF=N%TG=80%S=Z%A=S+%F=AR%O=RD=0%Q=), T7(R=Y%DF=N%TG=80%S=Z%A=S+%F=AR%O=RD=0%Q=), T8(R=Y%DF=N%TG=80%S=Z%A=S+%F=AR%O=RD=0%Q=), T9(R=Y%DF=N%TG=80%S=Z%A=S+%F=AR%O=RD=0%Q=), U1(R=N), IE(R=Y%DFI=S%TG=80%CD=Z)</td>
</tr>
</tbody>
</table>
Our analysis focuses on the TCP/IP fingerprints and SyncScan data from the Internet Census. We did not look into host probe or ICMP ping scans because the SyncScan query data contains the same information. If an IP address was unreachable as per the ICMP ping scan, the SyncScan query data of that IP address would return only closed ports. Additionally, the time it took for the packet to return a response was not necessary for our analysis. For purposes of our data collection, we were only interested in open ports, further making the ICMP ping and host probe data unnecessary.

Although we did expand our APT1 indicators to include domain names, we did not use the rDNS records found within the Internet Census. The information in that source only used the top 16 name servers; we wanted to ensure we had a full picture for our APT1 data sets and instead chose a different source for this information.

The service probe scan files contain the most information in the Internet Census. Unfortunately, only a handful of our IP addresses returned information when probed. After going through the results of our probed IP addresses manually, we found no interesting information that either confirmed other pieces of our data or led us down new paths.

The traceroute information did not return any results for the APT1 IP address set, either as start/endpoints or hop points within other traceroutes.

There are some ethical considerations in using the Internet Census 2012 data. Since it was obtained through a botnet doing network scans without permission, should researchers use it? In answering this question, it is important to understand which data is lacking and where else it could be found.

The Internet Census 2012 data set contains machine information, not personally identifiable information (e.g., Social Security numbers, birthdates, places of birth). While some names, email addresses, and phone numbers occur in rDNS and service probe results, they are not considered private information. The email addresses and phone numbers all appear to be valid and point to IT helpdesks; the service probes stored all responses, which were occasionally error messages from the probed machines or their proxies or firewalls. Furthermore, the information in the Internet Census data set is not an up-to-date picture of the internet as a whole or its individual components. The devices connected to the internet change frequently from being connected and disconnected, installing software patches...
and upgrades, and implementing firewall configuration changes. This means the data set only provides an accurate picture of its components at the time of the nmap scans.

Whether the information from these scans is public or private is another question. Some of the information is public so the internet can work as it should, while other parts could be obtained in other manners. The main difference between the Internet Census 2012 data set and other similar data sets is the breadth of the information released.

2.3 Security Information Exchange at the Internet System Consortium (SIE@ISC): rDNS Data

The Internet Census data did not provide us with comprehensive rDNS information associated with each IP address because information was only pulled from the top 16 name servers. To get a wider range of results, we used the passive domain name server (pDNS) records collected by the Security Information Exchange at the Internet System Consortium (SIE@ISC) for 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 Internet Census.

2.4 Mandiant’s Report on APT1: Exposing One of China’s Cyber Espionage Units

On February 18, 2013, Mandiant published its report exposing APT1, which has been operational since 2006 and successfully compromised a slew of U.S. targets in sectors privy to Chinese government interests [Mandiant 2013]. Mandiant analyzed patterns in the security breaches of its clients and tracked APT1’s activities to four networks in the Pudong New Area of Shanghai. APT1 maintained a stronghold in its target’s infrastructure for months and, in some cases, for years. Its success is attributed to being a state-sponsored secret of the Chinese government, who provide the group ample funding for cyber missions [Mandiant 2013]. In its report, Mandiant provided the APT1 attack lifecycle, the types of software used to exploit the targets, and reoccurring identities found in the company’s data collection [Mandiant 2013]. Along with that report, Mandiant released 3,000 indicators of APT1 activity that included descriptions of APT1’s malware, MD5 hashes of the malware, self-signed secure sockets layer (SSL) certificates, indicators of compromise (IOCs), and fully qualified domain names (FQDNs) [Mandiant 2013].

The Mandiant report sparked our interest in APT1 and was an excellent resource when we questioned APT1’s motives and attack strategy. The information provided in the report’s appendices was invaluable and gave us many additional data points in our research. By using the FQDNs, in particular, we expanded our APT1 IP address set and, thus, were able to get a fuller picture of APT1 activity.

2.5 Open Resolvers Data

Open resolvers, which are best known for amplifying distributed denial of service (DDoS) attacks, were recently in the public eye when on March 27, 2013, Spamhaus—an organization that tracks internet spammers—was hit in the biggest DDoS attack in history [Leyden 2013]. The Open Resolver data set, provided by the Open Resolver Project at openresolverproject.org, is the list of DNS resolvers that have been configured to allow recursive DNS requests [Open Resolver Project 2013]. Since the end of March 2013, new lists have been released weekly. We used the April 7, 2013 list in this study.
2.6 Neustar GeoPoint Data

Neustar GeoPoint provides two types of data published monthly: geo-location information and routing data for IP addresses [Neustar 2013]. We used the available files closest to the Internet Census time frame to determine the country and state (or province if outside the United States) of origin for analysis.

The Neustar data uses the term *route type* as others use *media type*: the physical device that connects machines to the wider internet, not the logical application of internet routing. The Neustar route types include the connection method to the internet, like a fixed or satellite connection. Some route types are also associated with connection types, which indicate the connection medium, such as fiber optic, digital subscriber line (DSL), or leased line. A category is defined as unknown for both routing and connection type, indicating that Neustar did not have data on a particular IP address. For purposes of the study, we maintained consistency with the Neustar terms for both the geo-location and routing information.

2.7 Internet Storm Center Data

The Internet Storm Center DShield data provides a sample of internet traffic containing source IP addresses and destination ports that were blocked on any particular day [Internet Storm Center 2013]. We used information from January 1, 2012 through December 31, 2012 for our analysis. This data set did not provide information that was useful in determining a profile, but it was interesting to note that the most frequently blocked ports were also those that were most often open in the Internet Census. Since the Internet Storm Center data is just a daily sampling of occurrences, it might have been useful if the JIBs contained information for more of the IP addresses. Also, preserving the destination IP address in the data set might have enabled us to better determine how each IP address from the bulletins was being used.

2.8 Other Data Sources

Several different data sources went into determining who owned the IP addresses in the lists. Those addresses were associated with autonomous system numbers (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). We used information from potaroo.net to determine ASN ownership and several search engines to identify the line of business for those owners.
3 Data Issues

3.1 Available Data Is a Partial View of APT1 Operations

Although it includes information from the JIBs and the pDNS records that are linked to Mandiant’s domain names, our data is not a full representation of APT1 activities; rather, it is only a snapshot of how the IP address devices looked at a particular point in time. We also believe there are many more compromised IP addresses from around the world that we do not know about.

3.2 Timeline of Data

APT1 has been operational since 2006. We do not have enough information from our collective resources to establish a timeline for the IP addresses found in the Mandiant report or the JIBs. The combined IP address set could represent active IP addresses that APT1 used continually since 2006, or it could be a set active for only a period in time anywhere between January 2006 and February 2013.

The Internet Census data gave us a time frame of several months for our analysis: March to December 2012. To mitigate uncertainty in our analysis, we used the available data in each data source that corresponded most closely with the March to December time frame.

3.3 Internet Census Data Constraints

Although the Internet Census data provided great insight into the IPv4 address space, we are limited to only several months of data. We cannot track changes over time with this data. We confront the same issue with the timeline of the JIB and Mandiant IP addresses; we assume the APT1 IP addresses were actively used during the Internet Census time frame.

The Internet Census is very rich in content, but we found some issues while working with it. Many scans repeated during the Internet Census, providing us with conflicting information. This was especially true in regards to the TCP/IP fingerprint data. While analyzing that data, we found that a minority of APT1 IP addresses contained two unique fingerprints. In order to resolve these conflicts, we chose the fingerprint with the highest percentage match of the two to use in our analysis. The Internet Census is not perfect, but it does provide very rich data in the public domain.
4 Tools

Our analysis was aided by various Linux commands, the System for Internet Level Knowledge (SiLK), Excel, and several Python scripts. The Linux commands grep, awk, and cut helped extract the needed data from the Internet Census data files, as well as the Neustar GeoPoint file. SiLK provided the ability to efficiently combine, intersect, and difference the data, as well as to quickly create IP address sets. Excel was very useful in sorting and visualizing data and deriving statistics. Python scripts were created to turn TCP/IP fingerprint scan results into the OS or device those fingerprints represent. We chose SiLK, Linux, and Python for their robustness and availability to others wanting to do similar analysis. We chose Excel because it is used extensively, gives us the ability to quickly create tables and charts, and allows us to combine data from various sources in an easily readable format.

4.1 Linux Commands

Various Linux commands provided the ability to retrieve the relevant APT1 data from the large text files making up much of our data sets. The grep, awk, and cut commands were especially helpful in preprocessing the data files. Because the Internet Census data files are so large (ranging from several hundred MB to several GB), we used these commands to pull out and format the relevant entries for analysis or further preprocessing with SiLK, Excel, or Python scripts. Occasionally, we ran into memory-usage problems with the grep command that required us to run certain commands outside of normal business hours.

4.2 SiLK

SiLK is a set of network-traffic, flow-analysis tools created by the Network Situational Awareness team, which is part of the Carnegie Mellon® Software Engineering Institute’s CERT Division [SEI 2013]. We used SiLK for its speed and ability to translate IP lists into flow sets for comparisons and combinations. The commands rwsetbuild, rwsetcat, rwsettool --intersect, and rwuniq --pmap were most helpful. We used the rwuniq --pmap command with several custom pmap files built to allow efficient attribution of IP addresses to other data, such as geo-location and routing types.

4.3 Excel

Excel’s VLOOKUP, CONCATENATE, and various sorting and filtering functions enabled quick concatenation of the data from the various data sources. The COUNTIF, COUNTIFS, and SUMIFS functions allowed us to efficiently summarize data. Installing the Data Analysis add-in enabled us to correlate some of the data. Excel had some memory issues when creating tables for port analysis using the VLOOKUP function. As mitigation, we did sections of the data at a time, and then copied and pasted the results as values into a different table. That way, we kept the number of cells containing functions in any one workbook to a minimum.

* Carnegie Mellon is a trademark owned by Carnegie Mellon University.
4.4 Python Scripts

We used Python to code a script comparing the TCP/IP fingerprints from the Internet Census data to the nmap-os-db file to determine the OS identity for each available IP address. The script opens a tab-delimited fingerprint file and outputs a tab-delimited identity file. Due to differences in how Microsoft Windows and Linux/UNIX machines handle carriage returns and line feeds, we created one script for use on Windows machines and one for Linux/UNIX-based machines. We stayed consistent with the nmap fingerprint labels and used those terms in our analysis.
5 Indicator Expansion

5.1 Defining Variables and Expansions

Indicator expansion allows us to capture how we arrived at conclusions given our data sets and provides readers the best step-by-step process of our data manipulations in hopes they can recreate our data elsewhere.\textsuperscript{1} The algebra first defines the variables, which are uppercase letters. For example, IP addresses are denoted as $I$. To go from IP addresses to domain names, $D$, the process is denoted by two variables side by side. The algebra is read from left to right.

\begin{align*}
I \quad D
\end{align*}

5.2 Defining the Same Type of Data from Two Sources

To denote data sets that contain the same type of data from different sources, the variables will have a subscript for the data source. For example, if we have two IP address sets, one from $x$ and another from $y$, the variables will be expressed as the following:

\begin{align*}
I_x \\
I_y
\end{align*}

To denote that the IP sets are conjoined, the variables will be connected with a plus sign, $\oplus$. All the transformations to the conjoined IP address set are done to both sets as if the sets were concatenated:

\begin{align*}
I_x \oplus I_y
\end{align*}

5.3 Defining the Range of a Data Set

At times, our data requires excluding a particular percentage of the data from the data set. To allow this transformation, our data must reflect a range that is included, 75%-100%, along with denoting the data as being a percentage by using $p$ instead of $\%$. For example, if we wanted to exclude any IP addresses with fingerprints, $F$ that contains less than a 75% match, the algebra would be the following:

\begin{align*}
F_{p<75}
\end{align*}

A similar method is used for non-percentage ranges. For example, if we want to show fingerprints that occur between two and four times, the algebra would be

\begin{align*}
F_{2}^{4}
\end{align*}

\textsuperscript{1} For more information, see An Algebra for Describing the Steps in an Indicator Expansion, written by J. M. Spring and scheduled to be published in October 2013 as part of the eCrimes Research Summit proceedings.
6 Method

Each step of our analysis is reflected using indicator expansion.\(^2\) We saved all our source data and data manipulations onto a shared server. Each expansion explained in the section below is shown in the order of how we analyzed the data and is an independent analysis.

6.1 Variables Used

Table 3 contains a list of the variables used in this study and their descriptions.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>autonomous system number</td>
</tr>
<tr>
<td>C</td>
<td>country</td>
</tr>
<tr>
<td>D</td>
<td>domain name</td>
</tr>
<tr>
<td>F</td>
<td>fingerprint</td>
</tr>
<tr>
<td>I(j)</td>
<td>IP addresses in the JIB</td>
</tr>
<tr>
<td>I(m)</td>
<td>IP addresses associated with Mandiant domain names</td>
</tr>
<tr>
<td>I(n)</td>
<td>IP addresses in the non-routable IP space</td>
</tr>
<tr>
<td>M</td>
<td>malicious code</td>
</tr>
<tr>
<td>N</td>
<td>service probes returning a state of 1</td>
</tr>
<tr>
<td>O</td>
<td>open resolvers</td>
</tr>
<tr>
<td>P</td>
<td>open ports</td>
</tr>
<tr>
<td>R</td>
<td>routing data</td>
</tr>
<tr>
<td>S</td>
<td>IP address set of IPv4 sample</td>
</tr>
</tbody>
</table>

6.2 APT1 Address Set

We provided an explanation of how we analyzed our data along with the commands used for our analysis. Note that the dollar symbol (\$) used at the beginning of a command indicates the bash prompt and a new command.

6.2.1 I\(j\): JIBs' IP Addresses

The JIBs dated February 18, 2013 and February 26, 2013 contained 855 IP addresses known to be associated with APT1 malicious activity. To streamline our analysis, we first combined the IP addresses found in both JIBs into one text document called I\(j\).list and then turned that document into a SiLK set for use later in analysis through this command:

```bash
$ rwsetbuild Ij.list Ij.set
```

\(^2\) For more information, see An Algebra for Describing the Steps in an Indicator Expansion, written by J. M. Spring and scheduled to be published in October 2013 as part of the eCrimes Research Summit proceedings.
From this large, concatenated document, we created a second file, putting each IP address into regular expression notation, ^127\.0\.0\.1[:blank:], in order to use the `grep -f` command later in our analysis. The file was called Ij_reg.list.

We manually divided the concatenated list into smaller, individual files based on the Classless Inter-Domain Routing (CIDR) /8 netblock for each IP address. This step was necessary because the Internet Census data was divided already by /8 netblocks and searching through each file in some of the scan data for all the IP addresses was inefficient and caused memory issues.

### 6.2.2 IjD: IP Address Set in JIB to Domain Names

To execute this expansion, we ran a pDNS search on Ij to get a list of domain names. Once we received the results, we associated the IP address with its domain name and saved the combination into a single Excel document. We used the Excel `SORT` command and its filtering functionality to analyze the domain names.

### 6.2.3 Im = DmI-In: Mandiant IP Address Set

Mandiant published a list of FQDNs associated with APT1 activity as an appendix to its report. We performed an rDNS search via the SIE@ISC to retrieve the IP addresses, saving the associated IP addresses with their corresponding domain names into a single Excel document. Any non-routable IP address returned from the search was not included in the final list. The retrieved IP addresses were saved into a file called Im.list that we turned into a SiLK set for use later in analysis through this command:

```
$ rwsetbuild Im.list Im.set
```

There were 622 IP addresses in total after the rDNS lookup.

- From the Im.list document, we created a second file, putting each IP address into regular expression notation, ^127\.0\.0\.1[:blank:], in order to use the `grep -f` command later in our analysis. The file was called Im_reg.list.
- We manually divided the concatenated list into smaller, individual files based on the CIDR /8 netblock for each IP address. This step was necessary because the Internet Census data was divided into individual files for each /8 netblock for each data set. We put each IP address into regular expression notation to use the `grep -f` command later in our analysis.

### 6.2.4 Ij+Im: Union of Mandiant and JIB IP Address Sets

We took the 855 IP addresses from the Ij IP address set and the 622 from the Im IP address set and combined them into one file with the associated domain names for each IP address, making sure duplicates were removed.

- There are 1,386 IP addresses in the Ij+Im set.
- We made several different files from this data to use with the different tools and commands used for analysis.
  - We retained a file with the IP addresses in standard notation in case we needed to refer back to the original set for any reason. This file was called Ij+Im.list.
  - Using SiLK’s `rwsetbuild` command, we created a set of these IP addresses to aid in our analysis and use with other SiLK commands:
    ```
    $ rwsetbuild Ij+Im.list Ij+Im.set
    ```
We created a file called Ij+Im_reg.list that contained the IP addresses in regular expression notation to use with `grep -f` commands.

### 6.2.5 Internet Census Data

The bulk of our analysis came from the Internet Census and its TCP/IP fingerprint and SyncScan data. We needed to augment certain types of the data in order to return the information we needed. These changes are outlined in the sections below, along with how we performed the analysis.

#### 6.2.5.1 Ij+ImFp75: Fingerprints of the Ij+Im Set Given a 75% Match

The Ij+Im IP addresses separated into CIDR /8 netblocks were used to find the fingerprints. We used the following command to run the Ij+Im netblock sets against the corresponding netblock found in the fingerprint data:

$$ \text{grep -f Ij+Im_reg.list X > Ij+ImF}_X\text{.txt} $$

The X is the netblock in the Internet Census data, which is how the fingerprint files are named (the files have no file extensions). Upon finding the fingerprints for all the netblocks, we concatenated them into a single file using the command below:

$$ \text{cat * >> all_Ij+ImF.txt} $$

We ran the file through our Python script to give us usable data as shown in Table 4. We put the corresponding fingerprint data into our working Excel document containing the Ij+Im IP addresses and domain names. Using primarily the `SORT` and `COUNTIF` commands in Excel, we categorized and counted the various firewall, switch, gateway, and OS types.

Table 4: Fingerprinting Script Results

<table>
<thead>
<tr>
<th>IP Address</th>
<th>%Match</th>
<th>Description of Operating System Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>161.58.177.111</td>
<td>0.825049702</td>
<td>Microsoft Windows Server 2003 Enterprise Edition SP2</td>
</tr>
<tr>
<td>71.8.243.14</td>
<td>0.926865672</td>
<td>Linux 2.6.32</td>
</tr>
</tbody>
</table>

It is important to note that the resulting fingerprint match string comes directly from the nmap database file. This means there may be some ambiguity in the resulting match. For instance, some fingerprints will match the string “Microsoft Windows Server 2003 SP2,” while others may match “Microsoft Windows Sever 2003 Service Pack 1 or 2.” For the second instance, the fingerprint was not specific enough to match the particular SP2 template, but it did match a template that belongs to either service pack 1 or service pack 2.

#### 6.2.5.2 P: Internet Census Data Containing Only Open Ports

The SyncScan data found in the Internet Census was one of the most important pieces to this study. The data contains lists of open, open-filtered, and closed ports for all IP addresses in the IPv4 address space. To help us extract the APT1 data, we found it helpful to manipulate the SyncScan data before using it. We created a file listing each IP address in the IPv4 space containing only open ports to streamline analysis using the following command:

$$ \text{grep "[:blank:]open[:blank:]" * > syncscan_open.list} $$

Then, we used the command below to remove the timestamp and the syn-ack TCP/UDP fields associated with the open port list:

$$ \text{cut -f1,6 syncscan_open.list > P.list} $$
6.2.5.3 Ij+ImP: Port Analysis

We put the Ij+Im IP addresses into regular expression notation and concatenated them into one document to find open ports. We used the command below to pull the list of open ports associated with each IP address in Ij+Im:

```
$ grep -f Ij+Im_reg.list P.list > Ij+ImP.txt
```

- Once all the open ports were returned, we changed the delimiters in the file to be semicolons instead of commas using the Replace function in a normal text editor. We did that to ensure that any program we used for analysis would interpret the ports as individual values and not one long number.

- Using Excel, we manipulated the data to repeat the IP address with each associated open port to best analyze the data:
  - Ij+ImP$_{20}$, Ij+ImP$_{10}^{15}$, Ij+ImP$_{5}^{9}$, Ij+ImP$_{1}^{2}$ broke down the instances of IP addresses having various ports running by 20+ times, 10-15 times, 5-9 times, and 1-3 times.
  - The Excel VLOOKUP, COUNT, and COUNTIF commands aided our analysis.
  - Pivot tables helped us visually understand the data.
  - We plotted the distribution of open ports.

- Once we found all the open ports, we ran pair correlations to identify any unusual port pairing that occurred frequently in the data set.

6.2.5.4 N: Service Probes Returning a State of 1

The service probe data that we found in the Internet Census was very intriguing for our APT1 analysis. We hoped to find interesting data buried in the response to the service probes for the Ij+Im set. To find that data easily, we created a SiLK IP address set that contained all the IP addresses returning a response of 1 to a service probe. The commands were

```
$ grep "^[[:blank:]]1[[:blank:]]" */* > serviceprobe_1.list
$ cut -f1 serviceprobe_1.list > serviceprobe_1_IPS.list
$ rwsetbuild serviceprobe_1_IPS.list serviceprobe_1_IPS.set
```

6.2.5.5 IjN: Service Probe Data

Using SiLK, we found the intersection between Ij and N using the command below:

```
$ rwsettool --intersect Ij.set serviceprobe_1_IPS.set | rwsetcat | sort \n> Ij_sp1_intersection.txt
```

Once the intersection returned, we broke these IP addresses into CIDR /8 netblocks, resulting in four different files. We used the command below to pull out the responses associated with each IP address from the service probe data. The X is the netblock in the Internet Census data, which is how the service probe files are named (there are no file extensions). The Y represents the names of each of the four separate files created from the intersection.

```
grep -f Y */X > Ij_sp1.list
```

- 17 IP addresses returned information from the service probe scans.
- We analyzed the resulting search data manually in Excel.
6.2.6 Ij+ImA: Autonomous System Numbers (ASNs)

SiLK enabled us to make pmap files and then use the `rwtuc` command to find the ASNs associated with each IP address. The pmap files contained just the ASNs associated with IP addresses.

$ rwpmapbuild --in=IP ASN --out=IP ASN.pmap

The IP ASN file was created based on information from the Oregon Route Views Project and the RIPE Routing Information Service, and contains associations for the advertised IP address space. Once the pmap file was created, we used the command below to associate the APT1 IP address list with the appropriate ASN:

$ rwtuc --field=sip Ij+Im.list | rwuniq --pmap-file=asn:IP ASN.pmap -- field=src-asn --no-col > Ij+ImA.txt

- After determining the ASN for each IP address, we mapped the numbers to the registered companies via a dictionary file created from potaroo.net data.
- We did not combine subsidiary companies into the parent company or acquired companies into their new ownership.
- We manually researched the services provided by the ASN owners, such as internet service, hosting provider, education, or government.
- We used Excel to analyze the IP addresses and their corresponding ASNs.
- We used the SORT and FILTER tools, and the COUNTIF and VLOOKUP commands as needed to analyze the data.
- We used Pivot tables to examine visually how many IP addresses were associated to the same ASN.

6.2.7 Ij+ImR and Ij+ImC: Routing Data and Country Code

Instead of using ASN registration to determine country code, we decided to use Neustar data to determine the routing type and country code because it is more accurate. The Neustar data includes an address range, country code, city, and state where this address range is located, and a routing and connection type. The routing type is associated with the connection type (fiber, DSL, etc.). If there was no routing type but a connection type was listed in the file for an IP address range, we ignored the connection type in our analysis, cut the IP address ranges and the routing data, and then used SiLK to make a pmap. Next, we replicated the process containing city and state, and once again used SiLK to create a pmap. We used these commands:

$ gzip -dc neustar.csv.gz | cut -d"," -f1,2,5 | awk -F"," '{print $1 " " $2 " " $3}' | grep -v "start" | rwpmapbuild --in=-- --out=routingConnection.pmap

$ gzip -dc neustar.csv.gz | cut -d"," -f1,2,5 | awk -F"," '{print $1 " " $2 " " $3}' | grep -v "start" | rwpmapbuild --in=-- --out=countryState.pmap

$ rwtuc --field=sip Ij+Im.list | rwuniq --pmap-file=routingConnection.pmap --field=src-route --no-col > Ij+ImR.txt
$ rwtuc --field=sip Ij+Im.list | rwuniq --pmap- \
   file=country:countryState.pmap --field=src-country --no-col > \n   Ij+ImC.txt

Upon completion, we concatenated the IP addresses and the pmap data into a single tab-separated document for analysis. We used Pivot tables, and Excel’s SORT and FILTER tools and COUNTIF command for the data analysis.

6.2.8 Ij+ImO: Open Resolvers

The Open Resolver data was provided in a gzip file. Using these commands, we unpacked the file, created an IP address list, and then built a SiLK set:

$ gzip -dc open_resolvers.out.gz | grep ":0:1:\[0-9\]" | cut -d".:" -f3 \
   > open_resolvers.list
$ rwsetbuild open_resolvers.list open_resolvers.set

Next, we used SiLK to find the intersection between the Ij+Im set against the Open Resolver List and marked in our working document which IP addresses were open resolvers:

$ rwsettool --intersect open_resolvers.set Ij+Im.set | rwsetcat \
   | sort > Ij+ImO.txt

6.2.9 IjDM: Malicious Code

Using runtime analysis efforts from CERT, we associated malicious code hashes with the domain names from the Ij set:

- We took the domain names provided in Ij and ran them against metadata from CERT’s runtime analysis efforts to get the domain names known for hosting malicious code on their networks.
- Then, we reviewed these results and found the intersection between the malicious code domains provided by the metadata and the malicious code hashes in the Mandiant report.

6.3 IPv4 Address Sample

We created an IPv4 sample to use as a baseline for comparison with the Ij+Im data set findings. If the findings were similar to what we found in the IPv4 sample, we would have to assume that our hypothesis (APT1 has specific requirements for its infrastructure) is incorrect. Differences in the findings between the two data sets would bolster our theory. Our sample did not include private, multicast, or loopback IP addresses. We found that between 5.5% and 7.5% of our sample IP addresses were not advertised at any given time during the March to December 2012 time frame.

6.3.1 S: Sample Set

We first made an IP address set of the entire IPv4 address space with SiLK called all_ips.set. This set excluded any reserved IP address space. We then used SiLK to create a random sample set of IP addresses from the entire IPv4 IP set. The sample size was 100,000 to give us a good picture of the IPv4 space and to ensure that our findings were a representative portion of the address space and acceptable for making comparisons and drawing conclusions. We used this command:

$ rwsettool --sample --size 100000 all_ips.set | rwsetcat | sort > S.list
As with the I\textsubscript{1}+I\textsubscript{m} data set, we saved the data into different formats to suit our needs:

- We retained the S.list file with just the random sample IP addresses in standard notation in case we had to refer to the original set for any reason.
- Using SiLK, we used this command to create a SiLK set of these IP addresses to aid in our analysis and called the file S.set:
  
  $$\texttt{rwsetbuild S.list S.set}$$

- Next, we created a file containing the IP addresses in regular expression notation, called S\_reg.list. Then, using the IP addresses in regular expression notation, we broke down the IP addresses into CIDR /8 netblocks in accordance to the format of the Internet Census data.

### 6.3.2 SR and SC: Routing Data and Country Code

Like for our APT1 data, we decided to use Neustar data to determine the routing type and country code of the sample IP addresses. We used the same pmap files that we created above for both the routing/connection type and country code in these commands:

$$\texttt{rwtuc S\_list --field=sip | rwuniq --pmap-file=route: \ routingConnection.pmap --field=sip,src-route --no-col > SR.txt}$$

$$\texttt{rwtuc S\_list --field=sip | rwuniq --pmap- \ file=country:countryState.pmap --field=sip,src-country --no-col \ > SC.txt}$$

We used the Excel \texttt{SORT}, \texttt{FILTER}, and \texttt{COUNTIF} commands, as well as Pivot tables for the data analysis.

### 6.3.3 S\textsubscript{p75}: Fingerprints of the S Set Given a 75% Match

We used the S IP addresses separated into CIDR /8 netblocks to find the fingerprints. We used this command to run the S netblock sets against the corresponding netblock found in the fingerprint data:

$$\texttt{grep –f S\_reg.list X > S\_X.txt}$$

The \textit{X} is the netblock in the Internet Census data, which is how the fingerprint files are named. Upon finding the fingerprints for all the netblocks, we concatenated them into a single file using this command:

$$\texttt{cat * > all\_SF.txt}$$

We ran the file through our Python script to give us usable data. Then, we put the corresponding fingerprint data into our working Excel document containing the S IP addresses. For our analysis, we primarily used the Excel \texttt{SORT} and \texttt{FILTER} tools, and \texttt{COUNTIF} and \texttt{VLOOKUP} commands.

### 6.3.4 SP: Port Analysis

We used the S IP addresses in regular expression notation to find open ports. This command pulled the lists of open ports associated with each IP address in S:

$$\texttt{grep –f S\_reg.list P.list > SP.txt}$$

- Once all the open ports were returned, we changed the delimiters in the file to be semicolons instead of commas using the Replace function in a normal text editor. Doing so ensured that any
program we used for analysis would interpret the ports as individual values and not one long num-
number.

- We manipulated the data to repeat each IP address with each of its associated open ports. The
  result was a file with an IP address and single port on each line, which helped our data analysis.
  - $SP_{20}$, $SP_{15}$, $SP_{5}$, $SP_{3}$ broke down the instances of IP addresses having various ports running
    by 20+ times, 10-15 times, 5-9 times, and 1-3 times.
  - The Excel SORT, FILTER, COUNT, and COUNTIF commands helped our analysis.
  - Pivot tables helped us visually understand the data.
  - We plotted the distribution of open ports.

- Once we found all the open ports, we ran pair correlations to identify any unusual port pairing that
  occurred frequently in the data set.

### 6.3.5 SR and SC: Routing Data and Country Code

Like the $I_{j} + I_{m}$ IP set, we decided to use Neustar data to determine the routing type and country code
because it is more accurate. We used the pmaps created above for the $I_{j} + I_{m}$ set to analyze our sample,
replicated the process containing city and state for our sample, and once again used SiLK to create a
pmap. We used these commands:

```bash
$ rwtuc --field=sip S.list | rwuniq --pmap- 
  file=route:routeConnection.pmap --field=src-route --no-col 
  > SR.txt
$ rwtuc --field=sip S.list | rwuniq --pmap- 
  file=country:countryState.pmap --field=src-country --no-col 
  > S.txt
```
7 Results: APT1

7.1 I_D: JIB Domain Names

The pDNS search connects IP addresses to domain names. While the names from the I_D set did not have many exact matches to those found in the Mandiant report, many of them were similar, often having the same second-level domain name. There were 3,573 domain name entries for the I_D expansion. We broke those down into four categories: (1) deliberately deceptive domain names, (2) seemingly random alphanumeric domain names, (3) domain names that appear to belong to legitimate organizations but are known to be malicious, and (4) domain names that appear to be legitimate and do not have a reputation for malware or phishing. Categories 1-3 are described below.

7.1.1 Deliberately Deceptive Domain Names

Many of the domain names appeared to be deliberately deceptive and mimic legitimate domains. For example, many contain forms of “Apple Soft Update,” “CNN Daily,” “Yahoo Daily,” “Firefox Update,” “Symantec Online,” “AOL Daily,” or “NY Times News.” Below is a list of takes on several well-known domain names and the number of occurrences.

<table>
<thead>
<tr>
<th>Organizations Used Deceptively in Domains</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple (as Apple Soft)</td>
<td>18</td>
</tr>
<tr>
<td>NY Times</td>
<td>15</td>
</tr>
<tr>
<td>MSN</td>
<td>10</td>
</tr>
<tr>
<td>CNN</td>
<td>8</td>
</tr>
<tr>
<td>Google</td>
<td>7</td>
</tr>
<tr>
<td>Symantec</td>
<td>6</td>
</tr>
<tr>
<td>Yahoo</td>
<td>6</td>
</tr>
<tr>
<td>McAfee</td>
<td>3</td>
</tr>
<tr>
<td>Microsoft</td>
<td>2</td>
</tr>
<tr>
<td>Firefox</td>
<td>2</td>
</tr>
</tbody>
</table>

Below is a small sample of the deliberately deceptive domains found in the pDNS results for the I_D set:
- epa.yahoodaily.com
- weather.yahoodaily.com
- fashion.cnnnewsdaily.com
- news.cnndaily.com
- flash.aoldaily.com
- update8.firefoxupdate.com
- download.applesoftupdate.com
- update.symentec.org
7.1.2 Pseudo-Random Alphanumeric Domain Names

Instead of containing company names that many target users would know, other domains contained strings appearing to be pseudo-random alphanumeric sequences. This type of domain name occurs for approximately 25% of the entire IjD set. Examples include:

- 29ig82qwpxkfsaeg4gapfw3jigmdoe4vn6otcpivtuowr5v3a.41.cgfde.com
- ae4en558r8gcmj9sbsra7u5e8d8v45rt7froujsbrvkumrep.45.okjyu.com
- 9urdkpdnd33t2uvaq7hkkejq6eawvb6b32dumgf4ldd6sbgve.41.nhtyd.com

At times, these names are associated with another second-level domain name that is not random, for example:

- q10765919.299.vpszuyong.com
- q1811815333.69vps.vpszuyong.com
- ns-00275-11412.irl-dns.info
- xn--12c4b5a2g9a5b.dmc.tv
- ns.44691d0d_cdc4b217_18521.dns.irl.cs.tamu.edu

7.1.3 Malicious Domain Names

Many of the other domain names found appear to belong to legitimate businesses but either currently host malware or have hosted it in the past. The owners of these domains may not know malware is being hosted on their sites, while others may be hosting malicious content deliberately. Examples of malicious domain names include:

- lanteckstudios.com
- pietros.com
- johnspizzerianyc.com

7.2 IjDM Malicious Code

After analyzing the domain names in IjD, the next logical step was to discover malicious code associated with the domain names found with the pDNS queries. The Mandiant hashes available to us contained over 1,000 unique hashes.

- 266 unique malicious code hashes were found in 2012 in the metadata from CERT’s runtime analysis efforts linked to IjD.

- Of the hashes found in the metadata, only seven were also found in the Mandiant report:
  - Those seven hashes occurred on eight different domains.
  - Three of those eight domains were found in the Mandiant list:
    - flash.cnndaily.com
    - japan.yahoodaily.com
    - flash.cnndaily.com
The remaining five of those eight domains were not found in the Mandiant list:
- engineer.lflinkup.org
- news.hoenjet.org
- news.hqrls.com
- report.crabdance.com
- www.freelanceindy.com

The metadata identified 80 unique domains with malware hashes.
- 15 hashes occurred two times each; the remaining 251 occurred only one time.

7.3 \( D_{Im} \): Mandiant IP Address Set

The Mandiant report provided many domain names associated with APT1 activity. We performed an rDNS search to retrieve a set of IP addresses associated with these domains.
- A total of 622 IP addresses were found.
- 91 overlapping IP addresses were already found in \( I_j \).
- 531 unique IP addresses made up the \( I_m \) set.

7.4 \( I_j + I_m \): IP Addresses Concatenated from Both Working IP Address Sets

For the purposes of our analysis, we found it beneficial to combine the \( I_j \) and \( I_m \) sets, but we did not create a separate variable for comparing them. In this report, we include our overall analysis of the combined set that contained 1,386 IP addresses.

7.5 \( I_j + I_m F_{75} \): TCP/IP Fingerprints of \( I_j + I_m \) Given a 75% Match

When we began this project, we speculated that APT1 would be targeting low-hanging fruit while building its infrastructure. Many times this comes in the form of using old OSs that still have exploitable vulnerabilities. Although these machines are not targets, APT1 needs to control hop points to exfiltrate data from the targets. Out of the 1,386 IP addresses, 451 were fingerprinted, constituting 32.5% of the combined set:
- 27.9% were Linux machines.
- 65.4% were Microsoft Windows machines.
- The remaining 6.7% ran an OS other than Linux or Windows.

7.5.1 \((I_j + I_m F_{75})_2\): TCP/IP Fingerprints with Multiple Instances

As we suspected, the top fingerprints were versions of Windows and Linux from 2009 or before. The top Windows fingerprints were from OSs released before 2005. Likewise, the top Linux kernels fingerprinted were versions from 2010 and earlier, but most of them date to around 2009. There is a time frame of 5-6 years between the Windows machines and Linux kernels. This data point challenged our initial hypothesis that all OSs would be dated around the same time and may suggest that APT1 changed operations to include more Linux machines after initially targeting Windows. The top five fingerprints made up 50% of the data. All fingerprints with five or more occurrences are listed in Figure 1. Figure 2 represents the fingerprints with two to four occurrences. Table 6 lists the top five most frequently occurring fingerprint matches with the total occurrences and percentage of the overall data.
set. Note that many of the fingerprints do not match exact versions or service packs but rather match an OS range.

Figure 1: Fingerprints with Five or More Occurrences
Figure 2: Fingerprints with Two to Four Occurrences

Table 6: Top Five Fingerprints in $I_j + I_m$

<table>
<thead>
<tr>
<th>Fingerprint</th>
<th>Count</th>
<th>% Total of IPs with Fingerprint Data in $I_j + I_m$</th>
<th>% Total of $I_j + I_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microsoft Windows Sever 2003 Service Pack 1 or 2</td>
<td>97</td>
<td>21.5%</td>
<td>7.0%</td>
</tr>
<tr>
<td>Microsoft Windows Server 2003 SP2</td>
<td>42</td>
<td>9.3%</td>
<td>3.0%</td>
</tr>
<tr>
<td>Linux 2.6.32 - 3.2</td>
<td>34</td>
<td>7.5%</td>
<td>2.5%</td>
</tr>
<tr>
<td>Microsoft Windows XP SP2 or Windows Server 2003 SP1 or SP2</td>
<td>29</td>
<td>6.4%</td>
<td>2.1%</td>
</tr>
<tr>
<td>Linux 2.6.32</td>
<td>24</td>
<td>5.3%</td>
<td>1.7%</td>
</tr>
</tbody>
</table>

7.5.2 $I_j$ to $I_m$ TCP/IP Fingerprint Comparison

We compared the fingerprints in $I_j$ and $I_m$. The sets were very similar in their frequency and distribution but differed in two ways:

1. Linux 2.6 versions occurred twice as often in the $I_m$ set.
2. The Microsoft Xbox Game Console occurred five times in the $I_m$ set but not at all in the $I_j$ data set.

Figure 3 categorizes the various fingerprints for each data set. Table 7 represents the top 10 fingerprint comparisons between $I_j$ and $I_m$ as a percentage of the corresponding IP address set.
Table 7: Top 10 Fingerprint Comparison of \( I_j \) and \( I_m \)

<table>
<thead>
<tr>
<th>Fingerprints</th>
<th>( I_j ) Count</th>
<th>( I_m ) Count</th>
<th>( I_j )%</th>
<th>( I_m )%</th>
</tr>
</thead>
<tbody>
<tr>
<td>unknown</td>
<td>610</td>
<td>376</td>
<td>71.3%</td>
<td>60.5%</td>
</tr>
<tr>
<td>Microsoft Windows Server 2003</td>
<td>100</td>
<td>92</td>
<td>11.7%</td>
<td>14.8%</td>
</tr>
<tr>
<td>Linux 2.6.x</td>
<td>44</td>
<td>82</td>
<td>5.2%</td>
<td>13.2%</td>
</tr>
<tr>
<td>Microsoft Windows XP</td>
<td>26</td>
<td>25</td>
<td>3.0%</td>
<td>4.0%</td>
</tr>
<tr>
<td>Microsoft Windows XP or Windows Server 2003</td>
<td>23</td>
<td>19</td>
<td>2.7%</td>
<td>3.1%</td>
</tr>
<tr>
<td>Microsoft Windows 2000</td>
<td>10</td>
<td>6</td>
<td>1.2%</td>
<td>1.0%</td>
</tr>
<tr>
<td>Microsoft Windows Server 2008</td>
<td>7</td>
<td>2</td>
<td>0.8%</td>
<td>0.3%</td>
</tr>
<tr>
<td>Microsoft Windows 7</td>
<td>4</td>
<td>3</td>
<td>0.5%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Microsoft Windows 2000 or Windows XP</td>
<td>4</td>
<td>1</td>
<td>0.5%</td>
<td>0.2%</td>
</tr>
<tr>
<td>Microsoft Xbox game console</td>
<td>0</td>
<td>5</td>
<td>0.0%</td>
<td>0.8%</td>
</tr>
</tbody>
</table>
7.6 $I_j+I_m$: Port Analysis

Along with the TCP/IP fingerprint information, the open ports allow us to best categorize what APT1 found necessary in its infrastructure. We looked at the open ports to understand what types of pathways are needed for APT1 operations:

- 791 IP addresses (or 51.1%) in $I_j+I_m$ had open ports.
- 4 IP addresses (possibly proxy servers) had more than 100 ports open.
- 609 unique open ports are represented in the data.

Figure 4 reflects the count of open ports per IP address. Most IP addresses in $I_j+I_m$ have fewer than 16 ports open. Running correlations between ports did not present any high-correlation pairs. We surmise there would be high correlations if we could break down the IP addresses into groups based on the exact role they play in the infrastructure. However, this was not possible with the data available for this study.

![Figure 4: Count of Open Ports per IP Address in $I_j+I_m$](image)

7.6.1 Most Common Open Ports in $I_j+I_m$

Table 8 reflects the most common open ports, each occurring in more than 10% of the IP addresses having open ports in $I_j+I_m$.

Figure 5 shows the distribution of open ports and the frequency of occurrence for the open ports with more than 20 occurrences.

![Figure 5](image)
Table 8: Ports Representing 10+% of Open Ports

<table>
<thead>
<tr>
<th>Port</th>
<th>Count</th>
<th>% Total IP Addresses with Open Ports in $I_j + I_m$</th>
<th>% Total IP Addresses in $I_j + I_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>655</td>
<td>82.8%</td>
<td>47.3%</td>
</tr>
<tr>
<td>3389</td>
<td>434</td>
<td>54.9%</td>
<td>31.3%</td>
</tr>
<tr>
<td>443</td>
<td>312</td>
<td>39.4%</td>
<td>22.5%</td>
</tr>
<tr>
<td>21</td>
<td>305</td>
<td>38.6%</td>
<td>22.0%</td>
</tr>
<tr>
<td>25</td>
<td>217</td>
<td>27.4%</td>
<td>15.7%</td>
</tr>
<tr>
<td>135</td>
<td>212</td>
<td>26.8%</td>
<td>15.3%</td>
</tr>
<tr>
<td>53</td>
<td>165</td>
<td>20.9%</td>
<td>11.9%</td>
</tr>
<tr>
<td>22</td>
<td>162</td>
<td>20.5%</td>
<td>11.7%</td>
</tr>
<tr>
<td>1025</td>
<td>156</td>
<td>19.7%</td>
<td>11.3%</td>
</tr>
<tr>
<td>3306</td>
<td>146</td>
<td>18.5%</td>
<td>10.5%</td>
</tr>
<tr>
<td>139</td>
<td>145</td>
<td>18.3%</td>
<td>10.5%</td>
</tr>
<tr>
<td>445</td>
<td>137</td>
<td>17.3%</td>
<td>9.9%</td>
</tr>
<tr>
<td>1723</td>
<td>123</td>
<td>15.5%</td>
<td>8.9%</td>
</tr>
<tr>
<td>110</td>
<td>117</td>
<td>14.8%</td>
<td>8.4%</td>
</tr>
<tr>
<td>1026</td>
<td>101</td>
<td>12.8%</td>
<td>7.3%</td>
</tr>
<tr>
<td>143</td>
<td>86</td>
<td>10.9%</td>
<td>6.2%</td>
</tr>
</tbody>
</table>

Figure 5: Count of IP Addresses for Ports Occurring 20+ Times in $I_j + I_m$
7.6.2 Ij to Im Port Comparison

When comparing the number of open ports in the Ij and Im sets, we noticed a similar distribution. However, Im had a higher percentage of open ports:

- 51.7% of the Ij set were open (442 IP addresses).
- 77.2% of the Im set were open (410 IP addresses).

When analyzing the common but less frequent ports, we noticed that ports 22, 445, and 111 occurred more frequently in Im than in Ij. Figure 6 highlights the comparisons between the two data sets.

![Figure 6: Port Comparison of Ij and Im](image)

7.7 Ij+Im: Autonomous System Numbers

We mapped the IP addresses in Ij+Im to their respective ASNs for two reasons. First, we wanted to see what types of companies were represented. Second, we wanted to understand if the infrastructures of any particular companies were being targeted:

- All the IP addresses were mapped (1,386).
- In the Ij+Im list, there were 388 unique ASNs, belonging to 362 different organizations.
- Organization types included internet service providers (ISPs), hosting providers, universities, and state governments. We could not determine a type for a small number of the ASN owners.

7.7.1 Types of Organizations Represented

The majority of the ASNs represented either hosting or internet service providers. The minority represented educational institutions, or local or state government. The majority of the hosting and ISPs are fairly well known; of the top 20 hosting companies, approximately 50% are represented in the Ij+Im set.
7.7.2 Targeted Organizations

If IP addresses are clustered within ASNs, it is reasonable to assume the companies owning those ASNs have been targeted for some particular reason. IP addresses that are distributed across many ASNs have likely been chosen randomly. Approximately half of the IP addresses in $I^+_{1+m}$ appear clustered in a small number of ASNs.

Figure 7 reflects the 28 organizations with 10 or more IP addresses in the data set. Table 9 reflects the top five companies and the percentage of IP addresses that resolve to them:

- 28 organizations (or 51.9%) were associated with 10 or more IP addresses.
- 208 organizations were associated with only one IP address.
- 127 organizations were associated with 2 to 9 IP addresses.

![Figure 7: Organizations with 10 or More IP Addresses](image)

**Table 9: Top Five Organizations in Registering IP Addresses in $I^+_{1+m}$**

<table>
<thead>
<tr>
<th>Company</th>
<th>Associated IP Addresses</th>
<th>Percentage of Total IP Addresses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychz Networks</td>
<td>86</td>
<td>6.2%</td>
</tr>
<tr>
<td>Google Inc.</td>
<td>73</td>
<td>5.3%</td>
</tr>
<tr>
<td>GoDaddy.com, LLC</td>
<td>43</td>
<td>3.1%</td>
</tr>
<tr>
<td>Hurricane Electric, Inc.</td>
<td>40</td>
<td>2.9%</td>
</tr>
<tr>
<td>AT&amp;T Services, Inc.</td>
<td>38</td>
<td>2.7%</td>
</tr>
</tbody>
</table>
Some of the companies found in $I_I^{+I_m}$ are known to be malicious outside of APT1 operations. Many have contained malware used by other malicious groups during various operations of their own. For instance, Psychz Networks and Krypt Technologies popped up very frequently, more frequently than the biggest hosting sites. Research on these two companies indicates they both provide bulletproof hosting services. Psychz Networks in particular has a reputation of hosting a slew of malicious websites.

### 7.8 $I_I^{+I_m}$: Location of IP Addresses

The $I_I^{+I_m}$ IP addresses are associated with 54 different countries, 45 U.S. states, and the District of Columbia. Figure 8 reflects the distribution of companies with more than five resolving IP addresses by U.S. state. Figure 9 shows the non-U.S. countries with resolving IP addresses.

<table>
<thead>
<tr>
<th>Company</th>
<th>Associated IP Addresses</th>
<th>Percentage of Total IP Addresses</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOVARTIS-DMZ-US (Qwest/CenturyLink)</td>
<td>38</td>
<td>2.7%</td>
</tr>
</tbody>
</table>
Figure 9: Non-U.S. Countries with Two or More Occurrences

7.9 I_{j+m}R: Routing Data

The routing data allows us to identify how the IP addresses are connected to the internet, whether through a fixed, dynamic, or other connection. Figure 10 shows the types of routing data found for I_{j+m}:

- **Fixed, dedicated** lines are lease lines or fiber connections, which indicate this is a connection from a mid-sized to large organization.
- **Fixed, other** lines constitute all other fixed connections such as DSL or cable.
- **Pop** represents a dial-up connection.
- **Other** is any other type of connection, for example, a connection through an international proxy or gateway.
In the figure, note that

- 1,000 IP addresses have routing types.
- 50% of those addresses have fixed connections through leased lines or optical fiber cable, indicating they likely belong to organizations and not individuals.
- 984 IP addresses have connection types, which means their routing type is fiber, leased line, DSL, cable, or dial-up.

**7.10 I_{j+m}O: Open Resolvers**

An open resolver is a misconfigured DNS server that answers recursive queries for hosts outside its network. We wanted to see whether open resolvers appeared to be an important factor for the APT1 infrastructure. Several open resolvers occur in the data set, and several of the companies represented in the set have more than one open resolver in our list:

- There are a total of 43 open resolvers in I_{j+m}.
- They constitute 3.1% of I_{j+m}.
- They belong to 30 different organizations.
- The biggest offenders with more than one open resolver are listed in Table 10.

<table>
<thead>
<tr>
<th>Company Name</th>
<th>Number of Open Resolvers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comcast Cable Communications, Inc.</td>
<td>6</td>
</tr>
<tr>
<td>CrystalTech Web Hosting, Inc.</td>
<td>5</td>
</tr>
<tr>
<td>MegaPath Networks, Inc.</td>
<td>3</td>
</tr>
<tr>
<td>Charlotte Colocation Center, LLC</td>
<td>2</td>
</tr>
<tr>
<td>NOVARTIS-DMZ-US (Qwest/CenturyLink)</td>
<td>2</td>
</tr>
</tbody>
</table>
7.11 APT1 Analysis Summary

Within the APT1 data set, all IP addresses had information for at least ASN and country. Almost 24% (329) of the IP addresses also have a fingerprint, at least one open port, and routing data. Less than 13% (177) of the IP addresses lacked a fingerprint, open ports, and routing data.
8 Results: IPv4 Sample

To confirm our analysis from the I_j+I_m data set, we had to know what the IPv4 address space looked like in contrast to the data associated with APT1. Then, we had a baseline of what we would expect to see with TCP/IP fingerprint information and SyncScan data in the APT1 data set if there was nothing special about the IP addresses and the infrastructure they represent. We found the sample to differ considerably from the I_j+I_m data, which confirmed our speculation that APT1 has special requirements for its infrastructure.

8.1 SF_{p75}: TCP/IP Fingerprints of S Given a 75% Match

Of the 100,000 IP addresses in our sample, 11,132—or 11.13% of the entire sample set—were fingerprinted:

- 686 of them were unique.
- 35.2% of the fingerprinted machines ran Linux, 12.9% ran Microsoft Windows, and 51.9% ran neither.

8.1.1 Top TCP/IP Fingerprints

Unlike the I_j+I_m set, none of the Microsoft Windows OSs breaks the top five occurring fingerprints. Instead, those top five contain two Linux OSs and three other, unrelated identities. Table 11 reflects the top five fingerprints, the number of occurrences, the percentage of the fingerprinted data, and the percentage of the sample.

Table 11: Top Five Fingerprints in S

<table>
<thead>
<tr>
<th>Fingerprint</th>
<th>Count</th>
<th>% Total of IP Addresses with Fingerprints in S</th>
<th>% Total IP Addresses in S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linux 2.6.32</td>
<td>734</td>
<td>6.6%</td>
<td>0.7%</td>
</tr>
<tr>
<td>VxWorks</td>
<td>691</td>
<td>6.2%</td>
<td>0.7%</td>
</tr>
<tr>
<td>Linux 2.6.18</td>
<td>463</td>
<td>4.2%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Linksys WRT610Nv3 WAP</td>
<td>451</td>
<td>4.0%</td>
<td>0.5%</td>
</tr>
<tr>
<td>AVM FRITZ!Box FON WLAN 7170 WAP (Linux 2.6.13)</td>
<td>418</td>
<td>3.8%</td>
<td>0.4%</td>
</tr>
</tbody>
</table>

Figure 11 shows the top 30 fingerprint occurrences for the IPv4 sample and the number of times each particular fingerprint occurs.
The IPv4 sample is crucial because it tells us what we should expect to see throughout the IPv4 space. If APT1 was not targeting something specific, we should expect similar results to the sample:

- Of the IPv4 sample, only 11.13% of the IP addresses returned fingerprints, compared to 32.5% in the $I_{j+I_m}$ set.
- No single fingerprint constituted more than 1% of the total IP addresses in $S$.
- No single fingerprint constituted more than 7% of the total fingerprinted IP addresses in $S$.

While the Linux percentages in the $I_{j+I_m}$ and $S$ sets were close, the Windows OSs diverged considerably as shown in Table 12. The $S$ set had a greater percentage of non-Windows and non-Linux machines than the $I_{j+I_m}$ set: We would expect these differences if APT1 was targeting specific architecture.

<table>
<thead>
<tr>
<th>Operating System</th>
<th>$I_{j+I_m}$</th>
<th>$S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linux</td>
<td>30.8%</td>
<td>35.2%</td>
</tr>
<tr>
<td>Windows</td>
<td>71.6%</td>
<td>12.9%</td>
</tr>
<tr>
<td>Other</td>
<td>7.3%</td>
<td>51.9%</td>
</tr>
</tbody>
</table>
8.2.1.1 \( \text{I}_{j+Im}^{*}F_{p75} \) and \( SF_{p75} \) Top TCP/IP Fingerprint Comparison

Although we found the \( \text{I}_{j+Im}^{*} \) set’s top fingerprints in the sample, they occurred in a much lower percentage of the IP addresses. Additionally, we found that three of the top five most frequently occurring OSs in the sample were nonexistent in the \( \text{I}_{j+Im}^{*} \) set. Microsoft Windows Server 2003 machines, in particular, had a much higher percentage of occurrences in the \( \text{I}_{j+Im}^{*} \) data. This further supports the theory that APT1 was targeting specific architecture. Table 13 compares the top fingerprints of the \( \text{I}_{j+Im}^{*} \) and S sets.

Table 13: Top Fingerprint Comparison Between \( \text{I}_{j+Im}^{*} \) and S

<table>
<thead>
<tr>
<th>Fingerprint</th>
<th>% Total IP Addresses with Fingerprints in ( \text{I}_{j+Im}^{*} )</th>
<th>% Total IP Addresses with Fingerprints in S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microsoft Windows Server 2003 Service Pack 1 or 2</td>
<td>21.5%</td>
<td>0.6%</td>
</tr>
<tr>
<td>Microsoft Windows Server 2003 SP2</td>
<td>9.3%</td>
<td>3.0%</td>
</tr>
<tr>
<td>Linux 2.6.32 - 3.2</td>
<td>7.5%</td>
<td>1.8%</td>
</tr>
<tr>
<td>Microsoft Windows XP SP2 or Windows Server 2003 SP1 or SP2</td>
<td>6.2%</td>
<td>0.2%</td>
</tr>
<tr>
<td>Linux 2.6.32</td>
<td>5.3%</td>
<td>6.6%</td>
</tr>
<tr>
<td>VxWorks</td>
<td>0%</td>
<td>6.2%</td>
</tr>
<tr>
<td>Linux 2.6.18</td>
<td>4.2%</td>
<td>4.2%</td>
</tr>
<tr>
<td>Linksys WRT610Nv3 WAP</td>
<td>0.2%</td>
<td>4.1%</td>
</tr>
<tr>
<td>AVM FRITZIBox FON WLAN 7170 WAP (Linux 2.6.13)</td>
<td>0%</td>
<td>3.8%</td>
</tr>
</tbody>
</table>

Figure 12 shows the top 30 fingerprints found in the IPv4 sample data. The fingerprints shown in black also occur in the top 10 fingerprints of the \( \text{I}_{j+Im}^{*} \) data. Only 5 of the top 10 fingerprints in \( \text{I}_{j+Im}^{*} \) are in the top 30 of the S data set.
8.3 SP: Open Ports of S set

Along with the TCP/IP fingerprint information of the sample, open port analysis allows us to confirm what pathways APT1 found necessary to its infrastructure:

- 33,573 IP addresses (33.6%) in the S set have open ports.
- 222 IP addresses had more than 100 ports open.
- One IP address had 249 ports open.
- 14,024 unique open ports are represented in the data.

Figure 13 reflects the count of open ports per IP address. About 95% of the IP addresses have three or fewer open ports at once. About 99% of the IP addresses have 13 or fewer open ports.

![Figure 13: Count of IP Addresses per Top 30 Ports in S](image)

8.3.1 Most Common Open Ports

Table 14 reflects the most common open ports occurring in at least 5% of the IP addresses having one or more open ports in the S data set.

<table>
<thead>
<tr>
<th>Port</th>
<th>Count</th>
<th>% Total IP Addresses with Open Ports in S</th>
<th>% Total IP Addresses in S</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>21140</td>
<td>63.0%</td>
<td>21.1%</td>
</tr>
<tr>
<td>443</td>
<td>5897</td>
<td>17.6%</td>
<td>5.9%</td>
</tr>
<tr>
<td>21</td>
<td>5557</td>
<td>16.6%</td>
<td>5.6%</td>
</tr>
<tr>
<td>23</td>
<td>5184</td>
<td>15.4%</td>
<td>5.2%</td>
</tr>
<tr>
<td>22</td>
<td>3808</td>
<td>11.3%</td>
<td>3.8%</td>
</tr>
</tbody>
</table>
Like with the fingerprint data, comparing open ports between the \( I_j + I_m \) and SP data sets reveals noticeable differences:

- Of the IPv4 sample, only 33.6% of the IP addresses had open ports, compared with 57.1% in the \( I_j + I_m \) set.
- Only 5 different ports were open on more than 10% of the IP addresses in the sample, compared to 16 in the \( I_j + I_m \) set.

### 8.4.1.1 \( I_j + I_m \) and SP: Top Open Port Comparison

The top fingerprints for the \( I_j + I_m \) set were also found in the sample; however, the rate at which they occur is very different between the two sets. The top 15 open ports in the \( I_j + I_m \) set occur at a rate of between 2 and 17 times as frequently as in the S set, with the exception of port 80. The rates of occurrence of ports 3389, 443, 25, 135, 1025, and 445 especially stand out.

*Table 15: Port Data Comparison Between \( I_j + I_m \) and S, All IP Addresses*

<table>
<thead>
<tr>
<th>Port</th>
<th>% Total IP Addresses in ( I_j + I_m )</th>
<th>% Total IP Addresses in S</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>47.3%</td>
<td>21.1%</td>
</tr>
<tr>
<td>3389</td>
<td>31.3%</td>
<td>1.5%</td>
</tr>
<tr>
<td>443</td>
<td>22.5%</td>
<td>5.9%</td>
</tr>
<tr>
<td>21</td>
<td>22.0%</td>
<td>5.6%</td>
</tr>
<tr>
<td>25</td>
<td>15.7%</td>
<td>3.3%</td>
</tr>
<tr>
<td>135</td>
<td>15.3%</td>
<td>1.1%</td>
</tr>
<tr>
<td>53</td>
<td>11.9%</td>
<td>2.8%</td>
</tr>
<tr>
<td>22</td>
<td>11.7%</td>
<td>3.8%</td>
</tr>
<tr>
<td>1025</td>
<td>11.3%</td>
<td>0.5%</td>
</tr>
<tr>
<td>3306</td>
<td>10.5%</td>
<td>1.4%</td>
</tr>
<tr>
<td>139</td>
<td>10.5%</td>
<td>0.5%</td>
</tr>
<tr>
<td>445</td>
<td>9.9%</td>
<td>0.3%</td>
</tr>
<tr>
<td>1723</td>
<td>8.9%</td>
<td>1.2%</td>
</tr>
<tr>
<td>110</td>
<td>8.4%</td>
<td>1.7%</td>
</tr>
<tr>
<td>1026</td>
<td>7.3%</td>
<td>0.5%</td>
</tr>
<tr>
<td>143</td>
<td>6.2%</td>
<td>1.6%</td>
</tr>
<tr>
<td>8080</td>
<td>5.4%</td>
<td>2.6%</td>
</tr>
</tbody>
</table>
Port data comparison between IP addresses in the entire set is shown in Table 16.

Table 16: Port Data Comparison Between I_{j}+I_{m} and S, Open Port IP Addresses Only

<table>
<thead>
<tr>
<th>Port</th>
<th>% Total IP Addresses with Open Ports in I_{j}+I_{m}</th>
<th>% Total IP Addresses with Open Ports in S</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>82.8%</td>
<td>63.0%</td>
</tr>
<tr>
<td>3389</td>
<td>54.9%</td>
<td>4.3%</td>
</tr>
<tr>
<td>443</td>
<td>39.4%</td>
<td>17.6%</td>
</tr>
<tr>
<td>21</td>
<td>38.6%</td>
<td>16.6%</td>
</tr>
<tr>
<td>25</td>
<td>27.4%</td>
<td>9.8%</td>
</tr>
<tr>
<td>135</td>
<td>26.8%</td>
<td>3.2%</td>
</tr>
<tr>
<td>53</td>
<td>20.9%</td>
<td>8.2%</td>
</tr>
<tr>
<td>22</td>
<td>20.5%</td>
<td>11.3%</td>
</tr>
<tr>
<td>1025</td>
<td>19.7%</td>
<td>1.4%</td>
</tr>
<tr>
<td>3306</td>
<td>18.5%</td>
<td>4.2%</td>
</tr>
<tr>
<td>139</td>
<td>18.3%</td>
<td>1.4%</td>
</tr>
<tr>
<td>445</td>
<td>17.3%</td>
<td>1.0%</td>
</tr>
<tr>
<td>1723</td>
<td>15.5%</td>
<td>3.4%</td>
</tr>
<tr>
<td>110</td>
<td>14.8%</td>
<td>5.2%</td>
</tr>
<tr>
<td>1026</td>
<td>12.8%</td>
<td>1.5%</td>
</tr>
<tr>
<td>143</td>
<td>10.9%</td>
<td>4.8%</td>
</tr>
<tr>
<td>8080</td>
<td>9.6%</td>
<td>2.6%</td>
</tr>
<tr>
<td>23</td>
<td>4.2%</td>
<td>5.2%</td>
</tr>
</tbody>
</table>

8.5 SC: Location of IP Addresses for S Set

The S data set contained 193 different countries; North Korea, Iran, and the Virgin Islands (British) were not represented. All 50 U.S. states were represented, along with the District of Columbia. Figure 14 reflects the distribution of U.S. states with more than 180 resolving IP addresses.

Figure 15 shows the countries with more than 350 resolving IP addresses. These countries represent 92.3% of the data.
Figure 14: U.S. States Containing More Than 180 IP Addresses

Figure 15: Countries Containing More Than 350 IP Addresses
8.5.1 Ij+ImC and SC Comparison

There are disparities between the percentages of countries represented in the Ij+Im set and what we would expect to see in the IPv4 address space. The biggest difference is the group of IP addresses resolving to the United States. The U.S. IP addresses in the Ij+Im data set occurred three times more frequently than in the S set. There were also many more IP addresses resolving to China in S than in Ij+Im. Table 17 provides a comparison between the most frequently occurring countries in the Ij+Im and S data sets.

Table 17: Comparison of IP Location Between Ij+Im and S

<table>
<thead>
<tr>
<th>Country</th>
<th>Percentage of APT1 IP Set (1,386)</th>
<th>Percentage of Sample IP Set (100,000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>79.73%</td>
<td>25.21%</td>
</tr>
<tr>
<td>Taiwan, Province of China</td>
<td>2.45%</td>
<td>1.66%</td>
</tr>
<tr>
<td>Korea, Republic of</td>
<td>2.02%</td>
<td>4.84%</td>
</tr>
<tr>
<td>Japan</td>
<td>1.37%</td>
<td>5.51%</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>1.23%</td>
<td>2.67%</td>
</tr>
<tr>
<td>Thailand</td>
<td>0.94%</td>
<td>0.41%</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>0.79%</td>
<td>0.33%</td>
</tr>
<tr>
<td>Singapore</td>
<td>0.79%</td>
<td>0.28%</td>
</tr>
<tr>
<td>Ireland</td>
<td>0.72%</td>
<td>0.21%</td>
</tr>
<tr>
<td>Canada</td>
<td>0.65%</td>
<td>1.72%</td>
</tr>
<tr>
<td>Germany</td>
<td>0.65%</td>
<td>5.44%</td>
</tr>
<tr>
<td>Portugal</td>
<td>0.65%</td>
<td>0.38%</td>
</tr>
<tr>
<td>India</td>
<td>0.58%</td>
<td>1.62%</td>
</tr>
<tr>
<td>Poland</td>
<td>0.51%</td>
<td>0.79%</td>
</tr>
<tr>
<td>China</td>
<td>0.43%</td>
<td>15.76%</td>
</tr>
<tr>
<td>Romania</td>
<td>0.43%</td>
<td>0.91%</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.43%</td>
<td>1.14%</td>
</tr>
<tr>
<td>Russian Federation</td>
<td>0.36%</td>
<td>2.15%</td>
</tr>
</tbody>
</table>

8.6 SR: Routing Data for S set

We found the routing types of the sample in the same way we did for the Ij+Im IP data set. We broke the data into four categories: (1) fixed, dedicated, (2) fixed, other, (3) pop, and (4) unknown. Figure 16 reflects the types of Neustar routing data found for S:

- 65,440 IP address have routing types.
- 65,595 IP addresses have connection types, which means the routing type is fiber, leased line, DSL, cable, or dial-up.
- Of the IP addresses with routing types, only 11% have fixed, dedicated lines through either leased lines or optical fiber cable, indicating they belong to medium- to large-sized organizations, not individuals.
The routing type indicates how well connected certain machines are to the internet. The “fixed, dedicated” routing type suggests the most stable infrastructure. This category appears almost five times as frequently in I$_j$+I$_m$ than in S. The “fixed, other” category is less stable than “fixed, dedicated” and occurs twice as frequently in S than in I$_j$+I$_m$. There is also a higher occurrence of “other” (various proxies or gateways) routing types in S than in I$_j$+I$_m$. Table 18 provides a comparison between the I$_j$+I$_m$ and S data sets by percentage.

**Table 18: Comparison of Routing Type Between I$_j$+I$_m$ and S**

<table>
<thead>
<tr>
<th>Routing Types</th>
<th>Percent of I$_j$+I$_m$ (1,386 IP Addresses)</th>
<th>Percent of S (100,000 IP Addresses)</th>
</tr>
</thead>
<tbody>
<tr>
<td>fixed, other</td>
<td>20%</td>
<td>40%</td>
</tr>
<tr>
<td>fixed, dedicated</td>
<td>50%</td>
<td>11%</td>
</tr>
<tr>
<td>other</td>
<td>1%</td>
<td>12%</td>
</tr>
<tr>
<td>pop</td>
<td>2%</td>
<td>1%</td>
</tr>
<tr>
<td>other</td>
<td>27%</td>
<td>36%</td>
</tr>
</tbody>
</table>
9 Discussion

Given the analysis we conducted, we made some generalizations about the attributes of the devices used in APT1’s middle infrastructure. These generalizations reflect the Mandiant report’s description of how APT1 works. Our data is representative of machines already involved in APT1 activity; the data reflects what APT1 was using in its infrastructure, not necessarily what it was looking for when it chose which machines to exploit. We can only provide a high-level generalization of its operations, not detailed analysis. We first discuss each of the attribute points: infrastructure owners, location, fingerprints, and open ports. We then provide generalizations for the infrastructure: intermediary C2 servers, malware servers, hop points, and inactive servers.

9.1 Infrastructure Owners

The research based on the ASNs and domain names suggests that APT1 was choosing stable organizations to build its middle infrastructure. This is also confirmed by the Neustar routing analysis: APT1 was targeting fixed, dedicated routing types as part of its infrastructure instead of something less stable. APT1 was hidden in plain sight; it was not using obscure ISPs or hosting providers. Rather, most of the companies chosen are very well known and easily found during a web search. After analyzing the routing data, we can also conclude that APT1 targeted fixed connections such as fiber, T1, and DSL. This makes sense because it would be costly and inefficient for APT1 to use infrastructure that was unstable and potentially unresponsive for active C2 servers.

We found that APT1 seemed to target certain ASNs, with over half the IP addresses in the $I_I^{+}I_m$ set clustered within a small number. These may have been exploited, but it is also possible that APT1 was purchasing the hosting services of some of these organizations. The remaining IP addresses occurred in a more distributed manner across many ASNs. These may have been used as anonymizing hop points or belonged to clusters where the other IP addresses were just not present in the $I_I^{+}I_m$ set.

9.2 Location

The majority of the IP addresses are located in the United States, though some come from around the world. Taiwan and other Asian countries account for four out of the top five non-U.S. countries associated with $I_I^{+}I_m$ IP addresses. Most other non-U.S. countries have fewer than 10 IP addresses in the data set. In the United States, a majority of the IP addresses resolve to California, which holds almost five times the IP addresses found in Arizona—the second highest. This may be deliberate or could simply be the result of California being the location of both Psychz Networks and Krypt Technologies.

9.3 Fingerprint Data

The data suggests that APT1 was targeting Windows machines dating to the time period when it began operations in 2006. The top fingerprints resolve to Microsoft Windows Server 2003, Service Packs 1 and 2, and Microsoft Windows XP Service Pack 2. Although these machines seem outdated in 2013, they were very common in 2006 and would have been likely targets for either zero day exploits or unpatched machines. The high occurrence of Linux kernels in the range of 2.6.32 - 3.2 suggests that APT1 may have switched OS preferences around 2010 [Linux Kernel Organization 2013].
One fingerprint that surprised us was the occurrence of Microsoft Xbox game consoles. All these reside on fixed, optical cable connections, with three in France and two in the United States. All five IP addresses belong to an ASN registered to Google Inc. Since these fingerprints are just within our cutoff range of a 75% fingerprint match, we suspect they may be misidentified.

9.4 Malicious Code

The 259 unique malicious code hashes found in the meta data from CERT’s runtime analysis efforts suggest there is a potentially larger infrastructure of APT1 operations than what was released to the public.

9.5 Port Data

Since we are looking at the middle infrastructure, we believe that some of these machines in our data were used purely for hop points, while others were used to directly compromise the target. Mandiant suggested that APT1 C2 servers are located in Shanghai and gather information from the servers located in the United States. Our data suggests that part of the middle infrastructure was used as “intermediary C2 servers,” which pushed off information to the main C2 servers located in Shanghai. Other machines in the infrastructure are used purely as vehicles for exploitation. There is depth between the C2 servers and the ultimate targets; however, our data does not suggest how much depth exists.

9.6 Intermediary C2 Servers

Two distinctions in C2 servers were identified. Most of the C2 servers appear to use port 443, which is used for several of the malware programs identified by the Mandiant report, but there is also a smaller set that uses port 111. In general, intermediary C2 servers

- are webservers
- have fixed connections, either leased lines or optical fiber
- reside on IP addresses assigned to well-known, well-connected organizations
- are located in the United States or Asia

Those relying on port 443

- also often use port 3389 once the infrastructure has been compromised, as identified by the Mandiant report
- or have three or more other open ports, including one or more of the following
  - 21
  - 25
  - 135
  - 53
  - 22
  - 1025
  - 3306
  - 139
  - 445
Those using port 111 (ONC RPC or Sun RPC)

- are Linux 2.6.32 or higher
- often also have port 22 (SSH) open

9.7 Malware Servers

These machines are used, for instance, as malware distribution points or mail relays. They

- likely are also web servers that contain malware that can be pushed to target machines in some manner
- either have ports 21 or 3389 open, or have three or more open ports, including one or more of the following
  - 1723
  - 135
  - 139
  - 25
  - 110

9.8 Hop Points and Inactive Servers

The available data sets did not allow determination of a profile for IP addresses with no open ports or those with only port 80 open. These are assumed to be hop points, inactive C2 servers, or inactive distribution points. The hop points were used to anonymize traffic and do not necessarily require capabilities beyond handling web traffic.
10 Future Work

Many other aspects of APT1 are worth exploring, such as understanding the role of bulletproof hosting providers in the APT1 data. Because at least two of the ASNs in the data set provide bulletproof hosting services, it would be interesting to determine if APT1 was purchasing services or exploiting those organizations. Another question to explore is if the low occurrence of an ASN that occurs in this data set indicates an instance of random exploitation, incomplete information (for instance, more IP addresses of that ASN have actually been compromised but were not flagged or released as being used by APT1), a single exploited company that was using multiple ISPs, or some other unrelated reason. Exploring the role of the few dial-up connections may also provide additional insight into APT1 operations in the future.
11 Conclusion

This study sought to understand what was necessary for APT1 infrastructure. JIBs INC260425 and INC260425-2, along with a pDNS search of Mandiant-released domain names provided us with 1,386 IP addresses associated with APT1 activity. The Internet Census 2012, Open Resolver list, SIE@ISC rDNS data, and Neustar GeoPoint data contributed information associated with the APT1 IP addresses that was helpful in analysis. Actual owners of the middle infrastructure and its location, fingerprints, and open ports provide insight into what appears to be used for the APT1 intermediary C2 servers, malware servers, and hop points.

APT1 uses stable, well-connected infrastructure from either hosting providers or ISPs in the United States. Most of the middle infrastructure runs Windows 2003, Windows XP, or the range of Linux kernels 2.6.18 - 2.6.32. Additionally, APT1 infrastructure may be evolving because the Windows machines appear to be much older than the Linux ones.

Useful data for enumerating a particular infrastructure does not need to be classified or proprietary. Our study illustrates that unclassified data from multiple sources, when used together, can draw, at a relatively low cost, many of the same conclusions of an expensive study. Our analysis is easily replicated, and our process can be used to easily evaluate the infrastructure of other malicious groups.
Appendix: DNS Names Pulled from the Security Information Exchange Data

0007722.com
0016.qq.bjxidebao.com
001htys.com
00234.com
0104.qq.bjxidebao.com
023boys.com
029zh.info
03877.com
0410dj.com
0471.qq.bjxidebao.com
0551dm.com
0575.me
06kv.com
0755ktxs.com
0755mdkt.com
0776.qq.bjxidebao.com
0893.qq.bjxidebao.com
0dwx.com
10.258982.com
1069sc.com
110.bm110.com.cn
110id.cn
1124.qq.bjxidebao.com
12.258982.com
1205.info
123050.com
12341090.cn
1234abcd.com
125801069.com
128556.net
132l.com
1337legendenbude.zapto.org
133wg.w99.1860php.com
1367.qq.bjxidebao.com
1379.xmhost.ludaidc.com
144654.com
14ts.com
14vd.com
152330.com
1546.qq.bjxidebao.com
155.258982.com
15dushi.com
1639762.com
164.258982.com
1674.qq.bjxidebao.com
173-163-133-177-centralpennsylvania.hfc.comcastbusiness.net
173-252-255-52.take2hosting.com
177456.com
1775.qq.bjxidebao.com
179444.com
188.258982.com
18renti.info
192.258982.com
19466.com
1staab.com
1staffonline.com
1staffonline.net
1staffonline.org
1stmob.com
1svc1.exchange.1-service.com
1xtx.com
2.79dbb44ec739a693.griddns.global.sonicwall.com
2000365.com
2001365.com
201.258982.com
2012jerseyswholesale.org
202-176-81-175.static.asianet.co.th
204-14-142-210.utilitytelephone.net
204-74-218-145.take2hosting.com
208.37.108.211.ptr.us.xo.net
208.43.154.7-static.reverse.softlayer.com
210.249.196.216.ded-dsl.fuse.net
210.ylbanqian.com
2253.qq.bjxidebao.com
2255.qq.bjxidebao.com
225577.co.cc
225577.co.cc
227227.com
228royal.com
235158.com
2352.ylbanqian.com
2365.qq.bjxidebao.com
246899.com
25.258982.com
258982.com
262999.com
266488.com
266488.com
2705.qq.bjxidebao.com
2791.999hhh.net
2795.999hhh.net
29ig82wpkasaeg4gapf3jigmdoe4vn6otcpivtuowr5v3a.41.cgfde.com
311345.com
3233.qq.bjxidebao.com
entech-group.com
entheos.com.sg
entrustm.com
environment.us.to
environment.mooo.com
environment.uk.to
epa.yahoodaily.com
equipic.com
equiv.ipsecsl.net
equiv.ocry.com
eri.serveuser.com
erniesgallery.com
ershou.saomen.com
es.pc-free-games.com
e-shoppingguide.com
esku.eilatinter.co.il
esoterism.org
etonnant.com
eum.businessconsults.net
eurosatory.mit-info.com
evolvemotorsports.com
exa.coastmaritime.org
exactearth.info.tm
exchange.ashtonplace.com
exchange.babelroad.net
exchange.bostonheartlab.com
exchange.golfnaperville.com
exchange.itsystemworks.com
exchange.thomsclan.com
exchange.vestarcapital.com
express.pmac.com
f.i7life.com
f.vaone.info
f.vpcup.com
f1f.ns06.net
faahsai.com
faan.cn
fairwaves.com
fangchan.saomen.com
fangqi.6600.org
fansex.mynetav.org
faq.bird-technologies.com
fasa.bigish.net
fasa.marsbrother.com
fasa.newsonet.net
fasa.purpledaily.com
fashion.cnnnewsdaily.com
fastbygast.com
fbrshop.com
fccmv.com
fcop.net
feaceblog.mefound.com
feedback.changeip.net
feedback.dynamic-dns.net
feedback.itemdb.com
feedback.justdied.com
feedback.longmusic.com
feedback.qpoe.com
feedback.youdontcare.com
feedback.zyns.com
feigou.cc
fel.labutes.com
fengqiuixian.com
fengziyu.net
ffeafitz.k12.mi.us
fhc.alphagammadelta.org
fijishaadi.com
file.myftp.info
file.nytimesnews.net
file.serveusers.com
files.downloadsnews.net
filmat36.com
finance.cnnnewsdaily.com
finance.thehealthmood.net
finance.todayusa.org
finance.usnewssite.com
finekl.worthhummer.net
firebirdgifts.com
firegvt1.firebirdrestaurant.com
firenapolitano.com
first.voiceofman.com
fishingreels.savebigsale.com
fitworks-anderson.com
fjddz.com
fjiao.net
flagulfbeachvacations.com
flash.aoldaily.com
flash.aunewsonline.com
flash.cnndaily.com
flash.jobsadvanced.com
flash.livemysm.com
flash.mcafeepaying.com
flash.themafia.info
flash.usnewssite.com
flash.yahoodaily.com
flash4ui.com
flex.ns06.net
floradelmar.net
floydleeband.com
fqqz.net
flucare.worthhummer.net
fmslaw.net
fni.newsonet.net
fom-prokuplje.co.rs
gps.aidtax.com
graceislandresort.com
graduateslandmarkeducation.com
graduateslandmarkforum.com
grandelabel.com
gravitybound.org
greasetrapmaintenance.com
greatesoft.net
grifonenyc.com
grillsmokers.savebigsale.com
groupware.cas-architekten.ch
gsftpt.dtdtech.net
gsxrzone.com
gszone.biz
gtaautoparts.ca
guage.ns06.net
guage.vpnhouse.net
guigang.saomen.com
gunstorage.letgoonsale.com
guynight.com
gylehua.com
gz.saomen.com
gzcityhotels.com
h1.hw247.net
h69-11-244-91.mdsnwi.dedicated.static.tds.net
habad.eilatinter.co.il
hack568.com
hackerstyler.no-ip.biz
hagermanparts.com
haibei.saomen.com
haipiy.com
handicraftvilla.com
handsperformance.com
haogw7.com
haoyun.xmhost.ludaoidc.com
hapaxsearch.com
haran.dougpoland.com
haran.dougpoland.net
haran.heathermariedesigns.com
haran.polandconsulting.net
haran.polands.org
harcoalgrillsbest.letgoonsale.com
harris.info.tm
hayabusazone.com
hayneedle.seotool.com
hb12369.org
hc7.beotel.net
hcbi.org
hcgdgs.com
hdg.cc
hdhfc.com
hdgdxcb.com
hdmmall.com
hdzdjj.com
health.jobsadvanced.com
healthbelieve.com
healthcareinteriorsdigest.com
hearingbest.letgoonsale.com
heavypayments.com
heb.2schina.net
hebja.com
heirloomrosedays.com
hektargroup.com
hektargroup.com.my
hektargroup.comwww.hektargroup.com
hello.usposters.net
help.dns-dns.com
help.dynamic-dns.net
help.thehealthmood.net
hen.2schina.net
hendrickseng.com
herculesafe.com
hexianguan.com
heyjude.eilatinter.co.il
hezuo18.com
hg93.com
hh2.1526.25u.com
hh2.5042.25u.com
hh2.58946.25u.com
hh2.woa-8927.25u.com
hh3.5278.25u.com
hh3.58946.25u.com
hh3.8692.25u.com
hh3.h02-xi03.25u.com
hh3.tlww-089.25u.com
hideawaybingo.com
hifo.net
highlandsranch.usswim.net
hikingbootsrated.letgoonsale.com
hill.arrowservice.net
hill.businessformars.com
hill-billyboggers.com
hk.4ren.net
hk-cpa.hkwww.com
hm-cca.com
hn.2schina.net
hnhks.com
holidayadventureslsa.com
holidayways.com
holleman.rs
holyzhou.com
home.brandywinema.com
home.sctzzj.com
home.staycools.net
ip-72-167-34-212.ip.secureserver.net
ip-72-167-37-238.ip.secureserver.net
ipserver.ee.ntu.edu.tw
iris.strangled.net
is.miyays.com
isabelmarantsoldeparis.com
isatap.jackmont.com
ischool.fitz.k12.mi.us
iss.businessconsults.net
issuenews.strangled.net
issues.ignorelist.com
istpl.com
it.newsonlinesite.com
it.pc-free-games.com
item.taobao.com.kolim12.tk
item.taobao.com.kolim18.tk
item.taobao.com.kolim2.tk
item.taobao.com.kolim3.tk
item.taobao.com.kolimj8.tk
item.taobao.com6h.tk
it-leaked.com
itri.auestrap.com
itrmall.com
itsystemworks.com
jackmont.com
jackstraw.coapt.com
jackypow.com
jandstours.com
jandvp.com
japan.yahoo.daily.com
jasonschultz.ca
jaszz.net
jaxd168.com
jbigdeal.com
jchbjc.com
jennettespier.net
jennettespier.org
jericho.orderhq.com
jeromedownes.info
jiandadn.com
jianxi.tk
jianyu.com.sg
jiaozuo.saomen.com
jieyang.saomen.com
jinanyanchu.com
jingxuanwu.com
jiuaidu.com
jiuyangshop.info
jiuzhabi.com.cn
jjtristate.com
jm-dc2.jackmont.com
jm-sql2.jackmont.com
jmtools.cn
jn9988.cn
jnepay.com
jnzjd.37.600dns.info
job.dima.ac.kr
jobbigger.com
jobskillsonline.com
joeasy.info
johnspizzerianyc.com
joyyard.cn
jp57.info
jstcyh001.host.znearly.com
juliteg.com
junglejewelsreptiles.com
junzhifu.com
jxwh.net
jxyichun.saomen.com
k211.com
kaiyangroup.com
katos.info.tr
katos.it.cx
kayauto.net
kb.seotool.com
kbadmin-train1.kb.net
kbapi-train1.kb.net
kdzhijia.com
kearnyalumni.org
kentstotz.com
kentstotzracing.com
keplerresearch.dynet.com
keruite.net
kesam.eilatinter.co.il
kesolutions123.nameserverpool.com
kettlekross.com
kfoassoc.com
kgartsindia.com
kiattana.co.th
kidsoutdoorfurniture.letgoonsale.com
kiem.ipsecsi.net
kindadi.com
kingboll.com
kingworlds.com
kiosk05.kis-kiosk.com
kitchenfaucetswwi.letgoonsale.com
kk2.5042.25u.com
kk2.58946.25u.com
kk2.safe-110.25u.com
kk3.13213.25u.com
kk3.5278.25u.com
kk3.58946.25u.com
klati.newsonet.net
luxuryairrsvp.com
luxuryairrsvp.htoademos.com
lvunwired.com
lvunwired.net
lw.infobusinessus.org
lwwave.arrowservice.net
m.it-leaked.com
mach5auto.com
magazine.yahoodialy.com
magnismunchies.com
mail.3nj.com
mail.3njpharma.com
mail.7perhour.co.uk
mail.absotech.com
mail.acdragon.com
mail.adbarker.com
mail.affm.info
mail.ahidistribution.com
mail.aidincorporated.org
mail.a-i-m.ie
mail.aiscea.com
mail.arrowservice.net
mail.athomecontracting.com
mail.audiobible.com
mail.avowners.com
mail.balearics.com
mail.bareart.com
mail.barkingcrow.com
mail.belmapangratz.com
mail.bent-severin.com
mail.bestcasuals.com
mail.bevlyne.com
mail.bfl-llc.com
mail.biblebible.com
mail.biohealth.com.sg
mail.blaze-inc.com
mail.bluestarredimix.com
mail.boseathorsham.com
mail.bossoshosshelp.com
mail.boyer-boyer.net
mail.boyerfamilypa.org
mail.brantleycustomhomes.com
mail.bretonhouse.ca
mail.broadtechsolutions.com
mail.btsnetworks.net
mail.buniqe.com
mail.businessconsults.net
mail.callawayfinancial.com
mail.canonunskincare.com
mail.carmela.com
mail.carticacapital.com
mail.carzrus1.com
mail.catalystgroupjax.com
mail.ccagency.com
mail.cdpa-architects.com
mail.cguslaw.com
mail.chileexe77.com
mail.choointhang.com
mail.clsinc.net
mail.colatitude.com
mail.computerpossibilities.com
mail.conour.com
mail.convexoptical.com
mail.creativecivilization.com
mail.crestwood.com
mail.csworship-pa.net
mail.customorderpolicespecialties.com
mail.dalipchand.com
mail.datatek-inc.com
mail.daten-loeschen.org
mail.diamond.sti.com
mail.dickdyermercedes.com
mail.dillemunery.com
mail.dotthemes.net
mail.drrzingaro.com
mail.dunnbenefit.com
mail.eadsmurraypugh.com
mail.echampions.org
mail.edifyconsultant.com
mail.elevation.com.tw
mail.emerge-systems.net
mail.emersonlimited.com
mail.emmarossosteopathypractice.com
mail.empowerednotary.com
mail.essexpropertylettings.co.uk
mail.fabonline.net
mail.fantasyhusband.com
mail.fastpack.net
mail.fresearch.org
mail.fijishaadi.com
mail.fountaininjurylaw.com
mail.frankwilliams.com
mail.freesexdoor.com
mail.galiservice.com
mail.genbukan.ca
mail.ghsscsa.net
mail.glacctg.com
mail.gmtca.com
mail.goaskdavid.com
mail.goldmanlawoffices.com
mail.good-dogs.co.uk
mail.googlemart.com
mail.hektargroup.com
mail.hendrickseng.com
mail.hetsco.com
mail.hgburacker.com
mail.hideawaybingo.com
mail.homemaidssolutions.com
mail.hometown-insurance.net
mail.hudson.org
mail.hudsondc.org
mail.icu.com.sg
mail.idealauto.ca
mail.indykiwi.com
mail.instantlinksites.com
mail.insulation-tubing.com
mail.internetdataexpress.com
mail.internetdesignconcepts.com
mail.lorl.net
mail.it-leaked.com
mail.jackmont.com
mail.jennettespier.net
mail.josievenable.com
mail.kaiyangroup.com
mail.kedco.us
mail.lamonsprinting.com
mail.langeed.com
mail.lastwordedits.com
mail.latitudecg.com
mail.lksoftvc.net
mail.locumrelief.co.uk
mail.lotnet.com
mail.lotsoshup.com
mail.m.it-leaked.com
mail.madsen-howell.com
mail.mariescatering.com
mail.mercies.com
mail.midwestcat.com
mail.mrduffy.com
mail.musicvisionsound.com
mail.nab.org.za
mail.nabr.org
mail.nbcir.net
mail.neocreative.net
mail.newwayedu.com
mail.nfms.net
mail.nodenine.net
mail.nodeninestudios.com
mail.northfieldcorp.com
mail.nuoglobal.com
mail.onelifewouthatl.com
mail.oriennradio.com
mail.oryanlawfirm.com
mail.oshkoshchristian.com
mail.osmb.bz
mail.othersellers.com
mail.oviédosafetylights.com
mail.pangratz.net
mail.pc-free-games.com
mail.pcpponet.com
mail.pcpponet.net
mail.pdllc.net
mail.penncorp.net
mail.pentagonsi.com
mail.persol.cn
mail.philliasrecords.com
mail.pixeintl.com
mail.planningpod.com
mail.planningpodevents.com
mail.planningpodpersonal.com
mail.pocketcaddie.co.za
mail.pocketsvp.co.za
mail.powz.com
mail.praxi.gr
mail.profinancialservices.com
mail.prohoists.com
mail.projectfile.com.sg
mail.raj-enterprises.com
mail.rcsweb2.com
mail.rcswebmail.com
mail.rdconsultingllc.com
mail.reallysmile.com
mail.redseasports.co.il
mail.reloadinteractive.com
mail.rfrick.com
mail.rginy.com
mail.richmondcencials.com
mail.ronish.ca
mail.ronishcomputers.com
mail.rsdu.org
mail.serenitysalonilc.com
mail.shawnaboyer.com
mail.sheridanpatterson.com
mail.sheridanpatterson.org
mail.shipsinaday.com
mail.silversurfersguide.com
mail.solution.com.sg
mail.somervilleinc.com
mail.sonfishgear.com
mail.sonshineriders.com
mail.sorenen.com
mail.soundnotes.com
mail.spiritofjefferson.com
mail.ssa.ukchiltracker.org
mail.starliteleasing.net
mail.statusholidays.com
mail.stevelma.net
mail.storeinuk.com
<table>
<thead>
<tr>
<th>ns-05374-11412.irl-dns.info</th>
<th>ns-05500-05586.irl-dns.info</th>
</tr>
</thead>
<tbody>
<tr>
<td>ns-05541-11412.irl-dns.info</td>
<td>ns-05731-11412.irl-dns.info</td>
</tr>
<tr>
<td>ns-06723-11412.irl-dns.info</td>
<td>ns-06845-11412.irl-dns.info</td>
</tr>
<tr>
<td>ns-07746-11412.irl-dns.info</td>
<td>ns-08066-11412.irl-dns.info</td>
</tr>
<tr>
<td>ns-08123-11412.irl-dns.info</td>
<td>ns-08644-11412.irl-dns.info</td>
</tr>
<tr>
<td>ns-08773-11412.irl-dns.info</td>
<td>ns-09155-11412.irl-dns.info</td>
</tr>
<tr>
<td>ns-09307-11412.irl-dns.info</td>
<td>ns-09712-11412.irl-dns.info</td>
</tr>
<tr>
<td>ns-09123-11412.irl-dns.info</td>
<td>ns-09712-11412.irl-dns.info</td>
</tr>
<tr>
<td>ns-08644-11412.irl-dns.info</td>
<td>ns-09009-11412.irl-dns.info</td>
</tr>
<tr>
<td>ns-08773-11412.irl-dns.info</td>
<td>ns-09155-11412.irl-dns.info</td>
</tr>
<tr>
<td>ns-09712-11412.irl-dns.info</td>
<td>ns-10252-11412.irl-dns.info</td>
</tr>
<tr>
<td>ns-11123-11412.irl-dns.info</td>
<td>ns-12506-11412.irl-dns.info</td>
</tr>
<tr>
<td>ns-13083-11412.irl-dns.info</td>
<td>ns-14371-11412.irl-dns.info</td>
</tr>
<tr>
<td>ns-15190-11412.irl-dns.info</td>
<td>ns-15780-11412.irl-dns.info</td>
</tr>
<tr>
<td>ns-15974-11412.irl-dns.info</td>
<td>ns-16094-11412.irl-dns.info</td>
</tr>
<tr>
<td>ns-16170-11412.irl-dns.info</td>
<td>ns-16453-11412.irl-dns.info</td>
</tr>
<tr>
<td>ns-16575-11412.irl-dns.info</td>
<td>ns-16757-11412.irl-dns.info</td>
</tr>
<tr>
<td>ns-16958-11412.irl-dns.info</td>
<td>ns-17125-11412.irl-dns.info</td>
</tr>
<tr>
<td>ns-17148-11412.irl-dns.info</td>
<td>ns-17429-11412.irl-dns.info</td>
</tr>
<tr>
<td>ns-18084-11412.irl-dns.info</td>
<td>ns-18086-11412.irl-dns.info</td>
</tr>
<tr>
<td>ns-18870-11412.irl-dns.info</td>
<td>ns2.absolutelyhosting.com</td>
</tr>
<tr>
<td>ns2.absolutelyhosting.com</td>
<td>ns2.atchurch.org</td>
</tr>
<tr>
<td>ns2.broadtechsolutions.com</td>
<td>ns2.cdmsinc.net</td>
</tr>
<tr>
<td>ns2.civis.com</td>
<td>ns2.daten-loeschen.org</td>
</tr>
<tr>
<td>ns2.daten-loeschen.org</td>
<td>ns2.dmt-portal.com</td>
</tr>
<tr>
<td>ns2.etonnant.com</td>
<td>ns2.freepdfpaper.com</td>
</tr>
<tr>
<td>ns2.freepdfpaper.com</td>
<td>ns2.freexdooor.com</td>
</tr>
<tr>
<td>ns2.fxtrades.info</td>
<td>ns2.gadgetshopgirl.com</td>
</tr>
<tr>
<td>ns2.gadgetshopgirl.com</td>
<td>ns2.hkwww.com</td>
</tr>
<tr>
<td>ns2.it-leaked.com</td>
<td>ns2.it-leaked.br</td>
</tr>
<tr>
<td>ns2.it-leaked.br</td>
<td>ns2.polandconsulting.net</td>
</tr>
<tr>
<td>ns2.polandconsulting.net</td>
<td>ns2.poland.org</td>
</tr>
<tr>
<td>ns2.poland.org</td>
<td></td>
</tr>
</tbody>
</table>
- oempartsgarage.com
- oewarehouse.com
- officiagroup.com
- ofurn.com
- oilamericagroup.com
- old.hisd.us
- old1.brocksperformance.com
- oliver.arrowservice.net
- omegalogos.org
- omegawatch.replicagarden.com
- once.downloadsiteme.net
- onecalleddearth.com
- onecourselearning.net
- oneabs.com
- onedaysweet.com
- one-world-cargo.com
- online.livemymsn.com
- online.mcafeepaying.com
- online.quickcert.com
- on-reflection.org
- onyxweb.streamguys.com
- o-o.preferred.telkom-cpt1.v11.lscache3.c.youtube.com
- o-o.preferred.telkom-cpt1.v11.lscache4.c.android.clients.google.com
- o-o.preferred.telkom-cpt1.v2.lscache1.c.youtube.com
- o-o.preferred.telkom-cpt1.v2.lscache4.c.youtube.com
- o-o.preferred.telkom-cpt1.v20.lscache2.c.youtube.com
- ooocar.com
- ope.coastmaritime.org
- opisolutions.com
- opp.coastmaritime.org
- or-69-69-94-3.sta.embarqhsd.net
- orderhelper.com
- orderparts.smythautomotive.com
- oregonunwired.com
- oscar.goddns.com
- otto.rs
- outdao.com
- outlook.somervilleinc.com
- owa.itsystemworks.com
- owa.softsolutionbox.net
- pacific.worthhummer.net
- paddi.ns06.net
- paddi.vphenhouse.net
- page-me.net
- paladin-ent.com
- pandawn.com
- pantechinc.com
- papper.booksonlineclub.com
- parker.creativecivilization.com
- parts.alsautomotive.com
- pasco.ujheadph.com
- pashe.rs
- patternshop.biz
- pay.freshreaders.net
- paydayloanshut1b.com
- pc-free-games.com
- pcpronet.com
- pdllc.net
- pdxservers.com
- peakwaves.usswim.net
- peixun.saomen.com
- penncorp.net
- penndelswim.org
- penndelswim.usswim.net
- pentagonsi.com
- perfectmulchproducts.com
- performax.co.th
- pfo.globalsecurity.org
- phb.arrowservice.net
- philippi.pmcc4w.org
- phoenix2.csceng.com
- photoshop-eilat.com
- pickmeup.tk
- pick-nick-frituur.be
- pietros.com
- girlsandiego.net
- pistonon.us
- planningpod.com
- planningpodevents.com
- plasamusic.com
- play.reload.com.sg
- pmcc4thwatch.com
- pmcc4thwatch.eu
- pmcc4thwatch.sg
- pmcc4thwatchtruth.info
- pmcc4w.org
- pmcc4wme.org
- podbeaver.com
- poolheatersccrr.letgoonsale.com
- pop.amk-services.com
- pop.csceng.com
- pop.richmondcommercialsvcs.com
- pop3.amk-services.com
- pop3.csceng.com
- post.ghsscsa.net
- powermax.kr
- power-public-relations.com
- pprmail.net
- praxi.gr
- prc.newsonet.net
- prefix.usapappers.com
- press.gift186.com
tetratech.zyns.com
thaieasyweb.com
thaifoam.com
thales.info.tm
thales.moo.com
thales.uk.to
the-apparel.com
thebarsprogram.com
thecatalystcompany.com
thecrownsgolf.org
thelawmatrix.com
thepinkhousebrighton.co.uk
theraodifference.com
thgfmt.com
thinkdesign.co.th
thomsclan.com
threefriends.info
t5juwes8skjqka49n4phksnsu6e4xi6opkxgg14v3scn4k8.43.vuytd.com
tiandimei.com
tiaonline.americanunfinished.com
tidalwave.usswim.net
tieling.saomen.com
tielu110.com
tiger.hw247.com
timboyer.net
time.issnbgkit.net
times.nytimesnews.net
timhaileyworld.com
timvitaman.com
tj.2schina.net
tk2.trkhosting.com
tlwkd.com
tm.sektori.org
tnets.net
tnniketnpascher.com
tnrequintnfr.com
tolovana.net
topcadet.com
topeagle-th.com
topmoney.purpledaily.com
topoo.cn
tortsurfer.com
totalrecall.entrustrm.com
tower.celetex.com
tplinkwuxianwang.mannasoft.cn
tpt.com
trade.daev.ca
tradeforexparttime.com
trading-post.ca
trainwiththepain.com
trans.vpnhouse.net
trans-personnelcorp.com
trends.signalfcc.com
trust.ns06.net
ts.dstech.net
ts02.dstech.net
ts2.safavieh.com
tsg-tech.com
tt22699.com
ttieurope.co.uk
tti-global.com
tti-india.com
ttinao.com
tpx.xsoldier.com
tutorcare.co.uk
tv.361fy.com
tv.wwl.net
tw.emasweb.com
twistedsanta.com
twodaydiets.com
twowheelextreme.com
tx-67-76-57-77.sta.embarqhsd.net
bxx.com
tzits.net
u15208005.onlinehome-server.com
u15286567.onlinehome-server.com
uc277.com
udb.in
ug-asg.hugesoft.org
ug-nema.hugesoft.org
ug-opm.hugesoft.org
uiu.crabdance.com
ukfexmail5.co.uk
uktraveldirectory.net
un linersd.org
ungdungvui.com
unionjackspeedway.com
uniqtek.com
uniqueambuilding.com
uniquegroup.net
uniqueindustrialsolutions.com
unitycambridge.org
unityintheocity.org
unitymass.com
unseenplanet.com
update.advancbusiness.com
update.dns05.com
update.dns-dns.com
update.freshreaders.net
update.ghssc.ca.net
update.ikwb.com
update.jetos.com
update.livemymsn.com
webinsyte.com
webmail.blackcake.net
webmail.btsnetworks.net
webmail.esgsats.com
webmail.nationalboiler.com
webmail.sunnytools.com
webmail.sunriseofficesystems.com
webmail.tnslaw.com
websence.com
webservice.coe-dmha.org
wefada.com
weidlerlodge.com
weightsolution.com
welcomehomeloans.com
weliveinthefridge.com
wellnesscentre.ky
wesinco.com
wesinstallation.com
west.steeringgearsuperstore.com
westbrookcapital.com
westkl.worthhummer.net
westmail.com.sg
wewillplay.oncecalleddearth.com
wgw.businessconsults.net
what.arrowservice.net
whatisdlandmarkeducation.cn
whatislandmarkforum.co.il
whatsthedeadaboulandmark.com
wheelies.ca
whir.li
wholesalecarkeys.net
wide4.worldwidemailing.org
width.vpnhouse.net
wifiserver.coppernet.net
wildlifememoriesgallery.com
willisdragracing.com
windycitymetals.com
wjo83.quickfunds4u.in
wlanb2b1.no-ip.org
wm.bostonheartlab.com
wmins.com
wmp.businessconsults.net
wnam.businessconsults.net
wnew.businessconsults.net
wolfone.cn
woodhavenrange.com
worldbestpriceautoparts.com
worldwidebearings.com
www.sg
wpad.csceng.com
wpad.jackmont.com
wpot.businessconsults.net
wpsn06.net
wstat.vpnhouse.net
wuiningya.com
wundercarparts.com
wunfccc.businessconsults.net
www.mmouu.com
www.comrepair.net
www.00234.com
www.029zh.info
www.03877.com
www.0410dj.com
www.0475woool.com
www.0543xxw.com
www.06kv.com
www.0755kxs.com
www.0755mdkt.com
www.0797top.com
www.0dwx.com
www.1000net.org
www.1069sc.com
www.110ld.cn
www.123050.com
www.12341090.cn
www.1234abcd.com
www.123lk.cn
www.132l.com
www.133335.com
www.138bt.com
www.14ts.com
www.14vd.com
www.15dushi.com
www.169444.com
www.177456.com
www.179444.com
www.17jeyan.info
www.17sexs.info
www.188799.com
www.193338.com
www.19466.com
www.1perfectday.com
www.1xtx.com
www.2000365.com
www.2001365.com
www.2011y2011.tk
www.20356.com
www.22222pk.com
www.228royal.com
www.235158.com
www.246899.com
www.266488.com
www.2schina.net
Bibliography

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

[Internet Storm Center 2013]

[Leyden, 2013]
http://www.theregister.co.uk/2013/03/27/spamhaus_ddos_megaflood/ (March 27, 2013).

[Linux Kernel Organization 2013]

[Mandiant 2013]
Mandiant. APT1: Exposing One of China’s Cyber Espionage Units.

[Neustar 2013]

[nmap.org 2013]

[Open Resolver Project 2013]

[SEI 2013]

[Unknown 2013]

[US-CERT 2013a]

[US-CERT 2013b]

[Ziegast 2010]
# Investigating Advanced Persistent Threat 1 (APT1)

**Title and Subtitle:** Investigating Advanced Persistent Threat 1 (APT1)

**Authors:** Deana Shick & Angela Horneman

**Performing Organization:** Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

**Performing Organization Report Number:** CMU/SEI-2014-TR-001

**Sponsoring/Monitoring Agency:**
AFLCMC/PZE/Hanscom
Enterprise Acquisition Division
20 Schilling Circle
Building 1305
Hanscom AFB, MA 01731-2116

**Abstract:**
In February 2013, Mandiant uncovered Advanced Persistent Threat 1 (APT1)—one of China’s alleged cyber espionage groups—and provided a detailed report of APT1 operations, along with 3,000 indicators of the group’s activity since 2006. This report analyzes unclassified data sets in an attempt to understand APT1’s middle infrastructure: the system of hops, distribution points or relays, and the command and control (C2) servers that sit between APT1’s victims and main C2 servers located overseas. To build that infrastructure, APT1 chose and exploited particular organizations to obfuscate communications while remaining in plain sight.

This analysis, based on data from IP addresses known to be associated with APT1 and domain names provided by Mandiant, was conducted using a combination of System for Internet Level Knowledge (SiLK) tools, Microsoft Excel, and custom Python scripts. The study detailed in this report can be replicated easily using available sources and tools. By combining key unclassified information, the authors successfully described a large, malicious network used to steal important information.

**Subject Terms:**
APT1, Fingerprinting, Internet Census Data, Open Ports, Analysis, IPv4 Sample

**Number of Pages:** 100

**Distribution/Availability Statement:** Unclassified/Unlimited, DTIC, NTIS

**DISTRIBUTION CODE:** Unclassified/Unclassified/Unclassified/Unlimited

**SECURITY CLASSIFICATION OF REPORT:** Unclassified

**SECURITY CLASSIFICATION OF ABSTRACT:** Unclassified

**LIMITATION OF ABSTRACT:** UL