Investigating APT1

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Deana Shick and Angela Horneman

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

DM-0000813

Introduction

- We wanted to validate what Mandiant proposed in its APT1 report
- We wanted to utilize the Internet Census 2012 data
  - We wanted to understand how public sources can tell a story about a specific threat group
- We came to many similar conclusions as Mandiant with significantly less man power
- We were able to show that important information can be gathered, combined, and reported by not using private or otherwise sensitive data

Data Sources

Mandiant: APT1: Exposing One of China's Cyber Espionage Units

Joint Indicator Bulletins: INC260425, INC260425-2

Internet Census Data 2012

Security Information Exchange at the Internet System Consortium (SIE@ISC): passive DNS Data

Open Resolvers Data Set

Neustar GeoPoint Data for geo-location and routing data

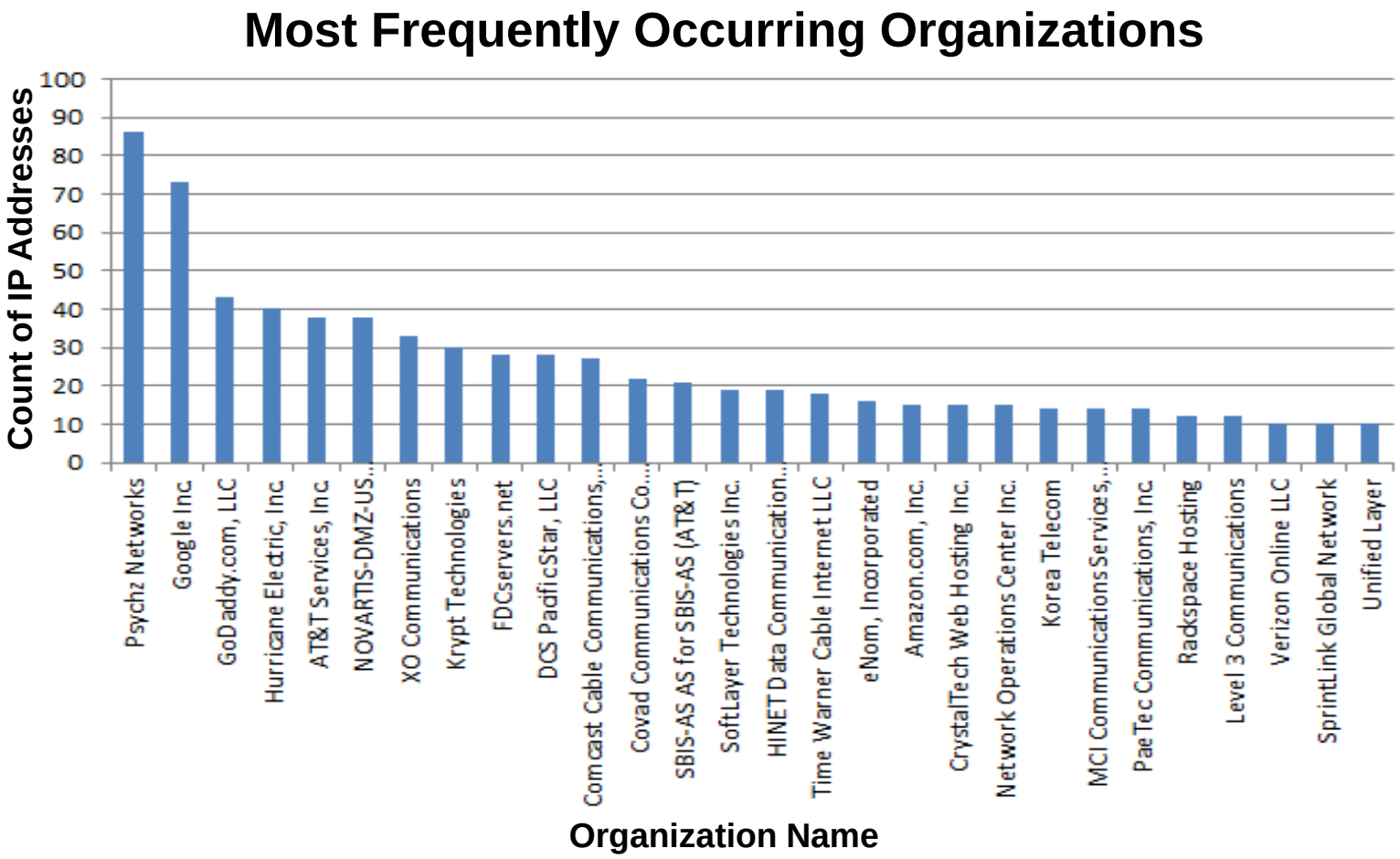
Internet Storm Center Data: Dshield

Indicator Expansion

- Address sets
  - **I_j**: IP addresses found in both JIBs
  - **DI_m**: IP addresses resolved from Mandiant FQDNs
  - **I_j+I_m**: Combined I_j and DI_m
- Network Attribution
  - **I_j+I_mA**: ASN information
  - **I_j+I_mR**: Routing type
  - **I_j+I_mC**: Country code, city and state (or province)
  - **I_j+I_mO**: Open resolvers
  - **I_jD:** Domain Names
  - **I_jDM:** Malicious Code
- Device Architecture
  - **I_j+$I_m$$F_{p75}$**: Fingerprints with minimum 75% match
  - **P**: Set of IP addresses having open ports
  - **I_j+I_mP** : Open ports
- IPv4 Random Sample
  - **SF_{p75}:** Fingerprints with minimum 75% match
  - **SP:** Open Ports

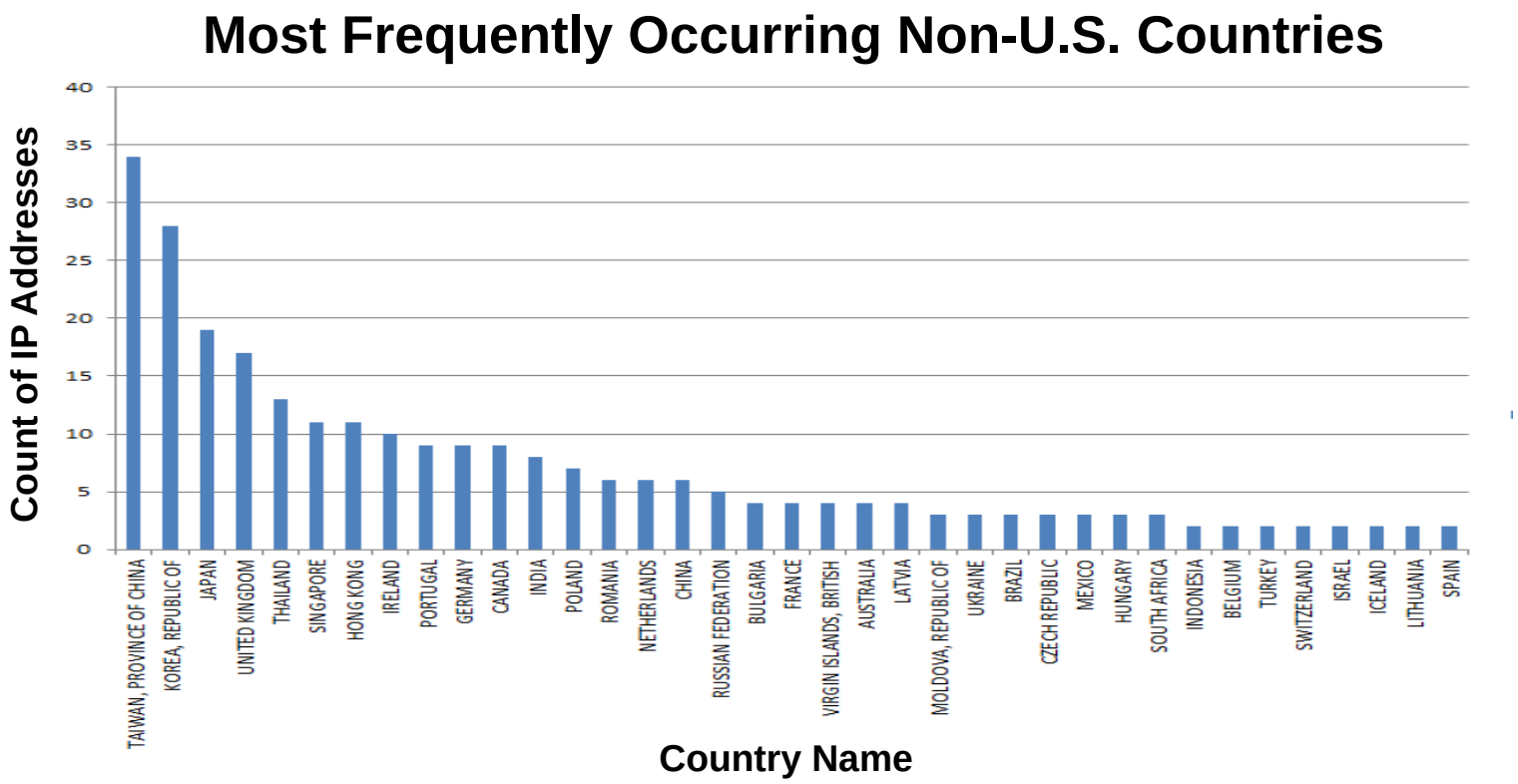
I_j+I_mA: Autonomous System Numbers

There are 28 organizations with 10 or more IP addresses which comprise 51.9% of the data.



I_j+I_mC: Location of IP Addresses

- 79% of I_j+I_m IPs are located in the United States
  - 45 U.S states and the District of Columbia are represented
  - 28% of IPs resolve to California
- 54 different countries are represented

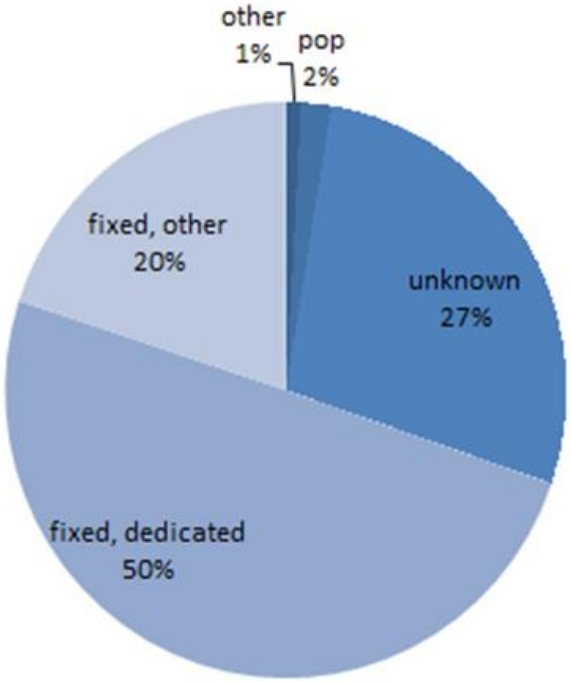


I_j+I_mR: Routing Data

- 1,000 I_j+I_m IPs have a routing type

- 984 IP addresses having connection types, which classifies if the routing type is fiber, leased line, DSL, cable or dialup.

Routing Types



I_jD : Domain Names

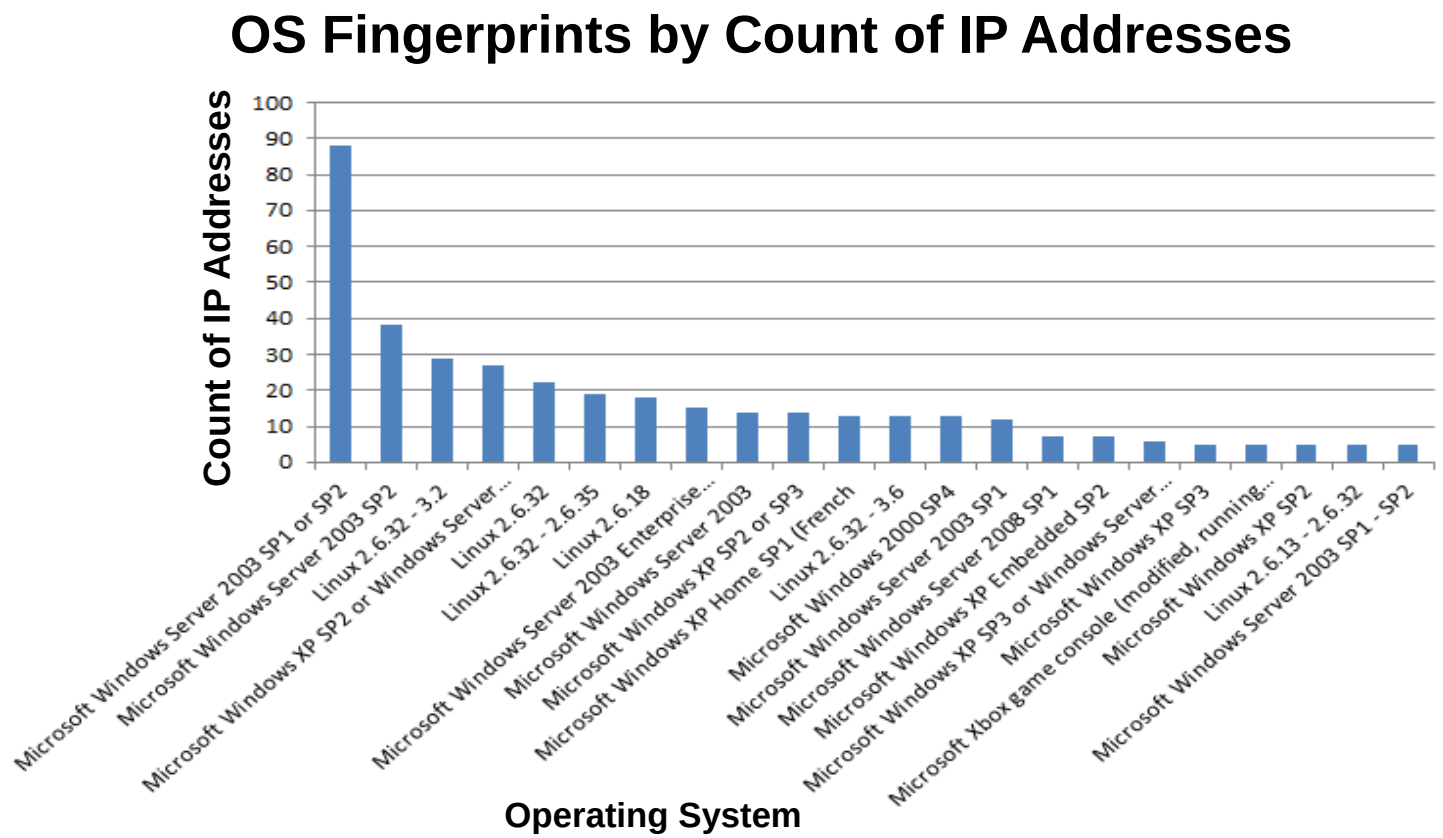
- Deceptive Domains; “off by 1”
- Pseudo-random Alphanumeric Strings
- Malicious Domains

I_jDM: Malicious Code

- 266 unique malicious code hashes found in CERT's Malicious Runtime Analysis efforts
- 7 of 266 appeared in the Mandiant report in 8 different domain names
- Only 3 of these 8 domain names were listed by Mandiant
- 15 hashes occurred more than once, all others occurred only once

$(I_j+I_mF_{p75})_5$: Top TCP/IP Fingerprints

- 32.5% of I_j+I_m was fingerprinted
  - Linux made up 27.9 % of the fingerprinted machines.
  - Microsoft Windows accounted for 65.4% of the fingerprinted machines.



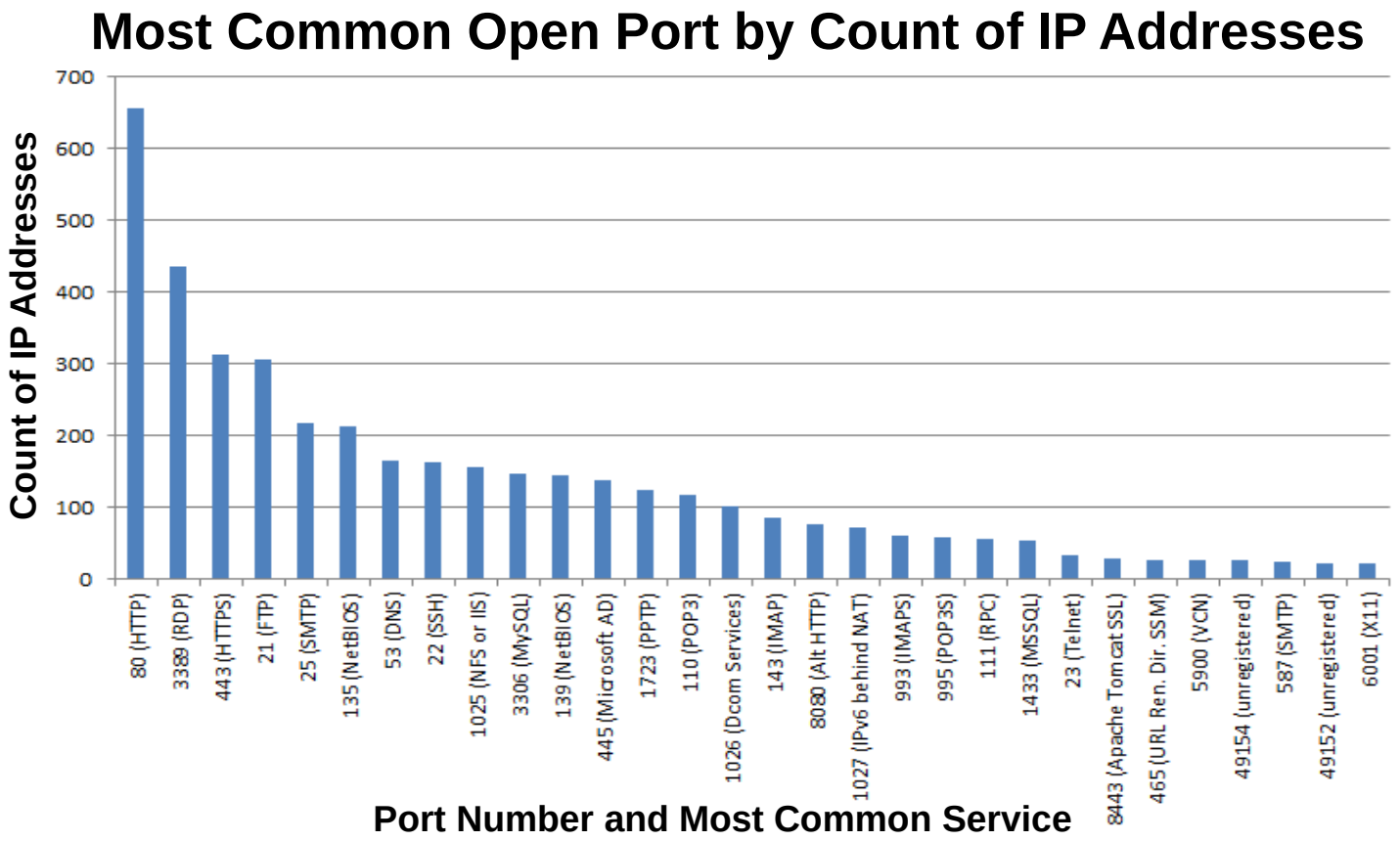
SF_{p75}:TCP/IP Fingerprints of S given a 75% Match

- 11.13% of the sample set was fingerprinted
  - Linux made up 35.2% of the fingerprinted machines.
  - Windows made up 12.9% of the fingerprinted machines.
  - 51.9% were other device types

Fingerprint	% Total IPs I_j+I_m	%Total IPs S
Microsoft Windows Sever 2003 Service Pack 1 or 2	21.5%	0.6%
Microsoft Windows Server 2003 SP2	9.3%	3.0%
Linux 2.6.32 - 3.2	7.5%	1.8%
Windows XP SP2 or Windows Server 2003 SP1 or SP2	6.2%	0.2%
Linux 2.6.32	5.3%	6.6%
VxWorks	-	6.2%
Linux 2.6.18	4.2%	4.2%
Linksys WRT610Nv3 WAP	0.2%	4.1%
AVM FRITZ!Box FON WLAN 7170 WAP (Linux 2.6.13	-	3.8%

I_j+I_mP: Port Analysis

- 51.1% of the I_j+I_m had open ports
  - 609 unique open ports
  - Four IP addresses had more than 100 ports open



SP: Open Ports

- There are 33,573 IP addresses found in the S having open ports.
  - This constitutes 33.6% of the sample data.
- There are 14,024 unique open ports represented in the data.

Port	% Total of IPs I_j+I_m	% Total IPs S
3389 (RDP)	**54.9%**	**4.3%**
443 (HTTPS)	**39.4%**	**17.6%**
21 (FTP)	**38.6%**	**16.6%**
25 (SMTP)	27.4%	9.8%
135 (NetBIOS)	**26.8%**	**3.2%**
53 (DNS)	20.9%	8.2%
22 (SSH)	20.5%	11.3%
1025 (NFS or IIS)	19.7%	1.4%
3306 (MySQL)	18.5%	4.2%
139 (NetBIOS)	**18.3%**	**1.4%**
445 (Microsoft AD)	**17.3%**	**1.0%**
1723 (PPTP)	**15.5%**	**3.4%**
110 (POP3)	14.8%	5.2%
1026 (DCom Services)	12.8%	1.5%
143 (IMAP)	10.9%	4.8%
8080 (Alt HTTP)	9.6%	2.6%
23 (Telnet)	4.2%	5.2%

I_j+I_mO: Open Resolvers

- There are 43 open resolvers total in the I_j+I_m .
  - This constitutes 3.1% of the I_j+I_m.
  - These belong to 30 different organizations

- Companies associated with more than 1 open resolver

Company Name	Number of Open Resolvers
Comcast Cable Communications, Inc.	6
CrystalTech Web Hosting Inc.	5
MegaPath Networks Inc.	3
Charlotte Colocation Center, LLc	2
NOVARTIS-DMZ-US (Qwest/CenturyLink)	2

Overall Findings

- Our available unclassified data gives a snapshot in time of what APT1 was using

- APT1 uses stable, well connected infrastructure, mostly in the US
  - Windows 2003 or XP, Linux ~ 2.6.32
  - Mostly ISPs or hosting providers
  - Especially from California

- The APT1 infrastructure may be evolving
  - Time frame of Windows vs. Linux versions
  - Open port differences (including absences) across IP addresses even within same ASN

- Malware hashes indicate there is a much bigger network for APT1 than what was released

Conclusion

- We were able to validate at a high level some of what Mandiant proposed
  - Command and control servers communicate over port 443
  - Malware communicated using RDP or similar protocols
  - APT1 used compromised mail servers in its operations
- In some cases we found more data then it provided
  - Exploited devices are predominately in the U.S. and in particular California
  - Found 259 previously unattributed pieces of malware
- In other cases what we were able to investigate did not directly corroborate Mandiant's findings
  - Exploited devices are hosting providers
- We were able to find important information by not using private or otherwise sensitive data