Characterizing Combatants of State-Sponsored APT in Digital Warfare by Reported Blocklist Database

Ruo Ando[1†] and Hiroshi Itoh[2††],
ruo@nii.ac.jp
National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, Japan, National Institute of Information and Communications Technology 4-2-1 Nukui-Kitamachi, Koganei, Tokyo, Japan

Abstract
Recently, the activity of the APT group has become organized and international. Unfortunately, the combatants in digital warfare are hybrid, and the distinction between types of combatants is hard to determine. In this paper, we present a new method for characterizing the combatants in state-sponsored APT by using the reported blocklist database. In the characterization, we use two open-source indicators of Grizzly Steppe and Hidden Cobra. We have obtained information from the reported blocklist database with the list of 877 and 633 IP addresses and analyzed the list of extracted country codes and IP address usage types. It turned out that two activities of APT can be well characterized by the distribution of countries and usage type.

Keywords:
Digital warfare, APT, Grizzly Steppe, Hidden Cobra, reported blocklist database.

1. Introduction

To conventional armed forces of army, navy, airforce from the early twentieth century to the last couple of decades, the capability in digital warfare has been added. Digital warfare is also called cyber warfare, which refers to any activities by a group or criminal organization usually sponsored by a state. Combatants in digital warfare use cyberspace to target another state.

An advanced persistent threat (APT) is executed by the nation-state or state-sponsored groups with stealth activities. The ATP group is aiming at unauthorized access vigilantly to an infrastructure network of the target organization. The access gained in compromised systems remains undetected for a long and extended period [1] [2]. To make matters worse, as cyber-attacks are commercialized, the term APT is used for non-state-sponsored grouping conducting the large and specified attacks [3]. For analyzing the activities of APT, in this paper, we cope with two open-source IoC of state-sponsored APT: Grizzly Steppe [4], and Hidden Cobra [5].

2. Combatants in Digital Warfare

According to Clausewitz, warfare is a concerted effort to impose one's will on another nation or organization. To this goal, digital warfare can be described as the ongoing Initiatives by using digital tools.

Governments are embarking on cyberwar big time. US, Russia, China, the major European countries, Israel, India, Brazil, Australia, New Zealand, Korea, Janan, and many Asian and African countries have divisions of cyber security.

Digital combatants can be classified into several categories.As we know, there have been considerably blurred boundaries between these categories. Among them, two categories are important for our analysis.

1) State-sponsored entities. Digital armed forces, intelligence services, and other kinds of agencies belong to state-sponsored entities. In some cases, they are organized as seemingly private security companies in which the state holds an influence stake.
2) Corporations. Private security firms, advertising or media outlets as a part of cyber warfare operation, public region agencies. Some groups are acting as mercenaries on behalf of states, unlike remaining on the right side of the law.

Corporate surveillance and government are usually intertwined. Mutual support between public-private surveillance is happening all over the world. Many countries adopt hybrid surveillance. Governments do not conduct surveillance alone. Digital combatants are supported by a vast public-private surveillance partnership. In this paper, we present the characterization of this kind of hybrid combatants in the view of partnership between the corporation and government surveillance.

Manuscript received March 5, 2022
Manuscript revised March 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.3.69

3. State-sponsored APT

3.1 Definition

Originally, the term APT was used by the United States Air Force (USAF) in 2006. In their discussion, APT was described as intrusive activities with their unclear civilian counterparts [1]. Activities of APT can be divided into three phases.

1) Advanced.The combatant has the proficient skill to evade detection and gain unauthorized access to a compromised system that has confidential information. The combatant has sophisticated skills and usually exploits opportunistic and zero-day vulnerabilities.
2) The term persistent means long-term. The nature of persistence causes the difficulty of detection and defense in cyberspace. Once the combatants compromise the computer network, it is difficult to detect and remove unauthorized access.
3) The information assets stolen are confidential in the targeted organization. Furthermore, the combatants bring about intellectual property theft resulting in serious economic damage.

From another perspective, Bodmer et al. [6] describes the objectives of APT as long-term unprivileged access to target computer network. The threat remains undetected for several months and sometimes years. Actually, the activities of Grizzly Steppe and Hidden Cobra have been continuing for tens of months. A brief description of these two APTs is as follows.

3.2 Grizzly Steppe

Grizzly Steppe is the alias of the malicious activity by purported Russian intelligence services [4]. In December 2016, The Department of Homeland Security (DHS) and the Federal Bureau of Investigation (FBI) in the United States firstly released the Joint Analysis Report (JAR) as the result of analytic efforts. According to this document, the Russian civilian and military intelligence Services (RIS) has compromised and exploited networks and endpoints associated with the the U.S. election. The activities of Grizzly Steppe also include the political and private sector entities of the U.S. Government.

3.2 Hidden Cobra

Hidden Cobra is the reference name of malicious cyber activity of the North Korean government, which is officially known as the Democratic People Republic of Korea (DPRK). DPRK employs malicious cyber activity for the purpose of collecting intelligence, conducting attacks, and gaining revenue [8][9]. Recently, DPRK's cyber program poses growing espionage according to the 2021 Annual Threat Assessment of the U.S. Office of the Director of National Intelligence 2021 [10]. It is presumed that Hidden Cobra is involved with cybertheft against financial institutions and cryptocurrency exchange worldwide. The amount of damage is estimated to be hundreds of millions of dollars which is probably funded to government priorities, including nuclear and missile programs [11]. The IoC of Hidden Cobra is available at [12].

4. Reported blocklist database

AbuseIPDB [13] is a reported blocklist database available in Intenet. AbuseIPDB is supported by a project for coping with hackers, spammers, and abusive activity on the Internet. AbuseIPDB has a central blacklist for network administrators, webmasters, and others stakeholders. They're working together to discover IP addresses associated with malicious parties online.
Also, we can report an IP address associated with malicious activity. In this paper, we use AbuseIPDB to extract features of an IP address list of Grizzly Steppe and Hidden Cobra.

4.1 API

Web REST API is provided by AbuseIPDB for reporting and checking IP addresses. There are various kinds of activities for checking, including spamming, hacking, vulnerability scanning, and so on. API queries reported blocklist database for protecting the network by inspecting IP addresses. Also, we can contribute by reporting malicious IP addresses to the database. API Endpoints. Both GET, and POST methods may be used.

https://www.abuseipdb.com/check/
[IP]/json?key=[API_KEY]&days=[DAYS]

4.2 Response

AbuseIPDB provides the desired data regarding the IP address queried, including version, country of origin, usage type, ISP, and domain name. We can get comments of inspecting IP as valuable abusive reports. The whitelist checks whether the IP address is spotted in any of the whitelists of database. The abuseConfidenceScore is an important indicator for action because this property is nonbinary and allows for nuance. AbuseIPDB calculates the abuseConfidenceScore as the evaluation on how abusive the checking IP is based on the users' reports. A hard minimum of 25% is placed on the abuseConfidenceScore. Table 1 shows the output table of AbuseIPDB.

AbuseIPDB returns a JSON array that contains the IP address, confidence of abuse score, and the timestamp of the last report. A JSON array is ordered by abuseConfidenceScore descending and then by lastReportedAt descending.

Table 1: AbuseIPDB outputs

IP address	X.X.X.X
isPublic	True
ipVersion	4
isWhitelisted	false
abuseConfidenceScore	100
countryCode	JP
countryName	Japan
usageType	Data Center/Web Hosting/Transit
isp	*** Co. Ltd
domain	***.com
hostname	TotalReports
numDistinctUsers	1
lastReportedAt	”2021-12-20T20:55:14+00:00”

The usage type is an indicator for describing the general usage of checking IP addresses. Values of usageType are: Commercial, Organization, Government, Military, University/College/School, Library, Content Delivery Network, Fixed Line ISP, Mobile ISP, Data Center/Web Hosting/Transit, Search Engine Spider, Reserved.

5. Numerical result and visualization

5.1 Country Code

AbuseIPDB provides area code information of most of the countries. There are four values of the availability: full availability, partial availability, no availability, or not applicable. Full availability means that all IP addresses in the country include an Area Code). Currently, the number of countries in full availability is 159. From the JSON file, we have obtained from AbuseIPDB; we generate the unique number list of country codes with alphabetical order as follows.

 # cat grizzly-steppe-all | jq .data.countryCode | sort | uniq

Here we use the JQ command, which is useful to extract data from JSON documents. We obtain the JSON document from the database by issuing a CLI command and getting the result of a REST API call.

From the JSON file, we have obtained from AbuseIPDB, we generate the unique number list of countries as follows.

 # cat grizzly-steppe.out | jq .data.countryName | sort | uniq -c | sort -k 1nr,1

By using the JQ command, we have generated a unique list that consists of 70 countries in alphabetical order.

Table 2: Frequency of country code found in the Ioc of Grizzly Steppe sorted by \# of lines (top 5)

Table 3: Frequency of country code found in the Ioc of Hidden Cobra

# CC	Code	# of lies (freq)	Country Name
67	US	125	United States of America
47	NL	90	Netherlands
20	FR	89	France
57	RU	78	Russian Federation
13	DE	63	Germany

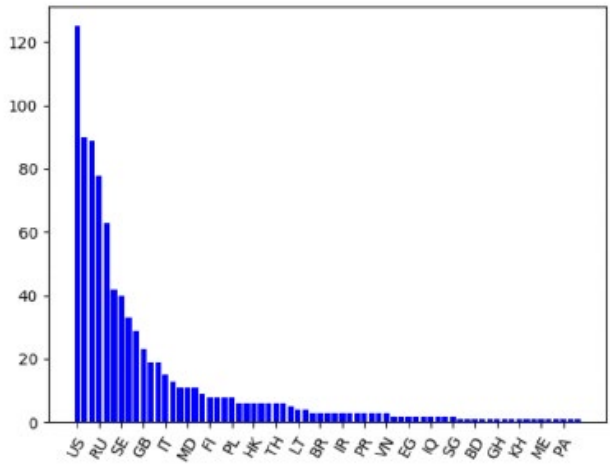
sorted by # of lines (top 5)

# CC	Code	# of lies (freq)	Country Name
57	RU	96	Russian Federation
1	AE	80	United Arab Emirates
20	FR	52	France
42	SA	51	Saudi Arabia
18	HR	31	Croatia

For the comparison of the number of countries listed in IoC of Grizzly Steppe and Hidden Cobra, we use the one-linear as follows:

 # cat apt29.out-all | jq .data.countryCode | sort | uniq -c | sort -k 2,2 > 1; cat hiddenCobra.out-all | jq .data.countryCode | sort | uniq –c | sort -k 2,2 > 2; join -j 2 1 2

Here we use the JOIN command, which is a command-line utility for joining lines on a common field. JOIN command is useful for joining two files by selecting fields within the line and joining the files. Tables II and III depict the top five active countries that appeared in IoC of Grizzly Steppe and Hidden Cobra. Two IoCs have their own characteristics. The only country they appear to have in common is Russia.



Also, Figures 1 and 2 show the frequency of country code appearing in IoC of Grizzly Steppe and Hidden Cobra. In both figures, the distribution of country code has a similar shape to the power distribution. It can be concluded that the groups in several countries are working vigorously as the main force.

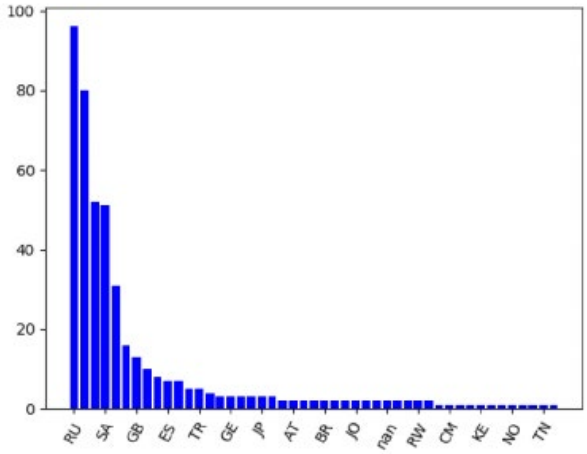


Table 4: Frequency of UsageType appeared in Grizzly Steppe

Usage Type #	Count
Data Center / Web Hosting Transit	630
N.A.	109
FixedLineISP	89
University / College School	19
Commercial	18

AbuseIPDB classifies IP addresses into 12 usage types based on the function of the organization/business unit. We can use usageType for filtering certain ranges of IP addresses for our needs. For example, if we are to check the IP address originates from a university, college, or school, we can check the EDU IP address.

(1) Commercial
(2) ContentDeliveryNetwork
(3) DataCenterWebHostingTransit
(4) FixedLineISP
(5) Government
(6) MobileISP
(7) N.A.
(8) SearchEngineSpider
(9) UniversityCollegeSchool

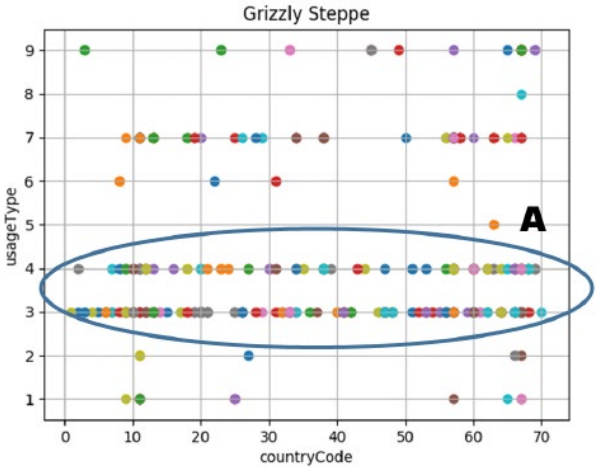
We have generated a list of frequency of Usage type in both Grizzly Steppe and Hidden Cobra by issuing the one-linear as follows:

 # cat $1 | jq -j '.data.ipAddress,",",.data.usageType,"\n"' | tr -d "\/" | tr -d " " | sort | uniq –c | sort -k 2,2

Tables IV and V show the frequency of UsageType appearing in Grizzly Steppe and Hidden Cobra. The items of Data Center / Web Hosting Transit, FixedLineISP, and Commercial are entered in both tables.

Table 5: Frequency of UsageType appeared in Hidden Cobra

Usage Type #	Count
N.A	269
FixedLineISP	125
Data Center / Web Hosting Transit	29
Commercial	12
MobileISP	12



Figures 3 and 4 show the 2D characterization of Grizzly Steppe and Hidden Cobra usage type. In both figures, the X-axis is country code ranging from 1 to 70. Y-axis is the number of Usage Type of IP. In Figure 3, area A is characterized by a high density of points. The area consists of two usage types: Fixed Line ISP and Data Center / Web Hosting Transit. In Figure 4, similarly, there is a concentration of points in area B. In addition, the area of A is highly dense. Area A is N.A. It is common that two activities (grizzly steppe and hidden cobra) frequently uses two usage types: Fixed Line ISP and Data Center / Web Hosting Transit.

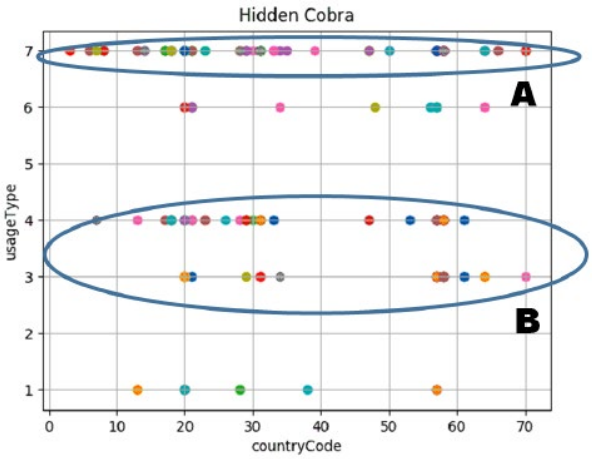


Fig. 4 Scatter diagram of 2D characterization of usage type of Hidden Cobra

6. Related Work

Peng et al. [14] present a series of measurements Virustotal by setting up their own phishing websites by imitating Paypal and IRS. Zhu et al [15]. propose a data-driven approach of online anti-malware engines by surveying 115 academic papers and collecting daily snapshots of Virustotal for more than 14000 files. Lewis [16] et al. developed the Automated IP Reputation Analyzer Tool (AIPRA) for analyzing many reliable blacklist databases. They also integrate the geolocation-based machine learning approach. Karadi[17] reports the Democratic National Committee's (DNC) operations in the past election season. In [18], the crucial links that played a role in the incidents of the Lazarus Group, a.k.a. Hidden Cobra. Sinha et al. [19] performs a preliminary study of a type of reputation-based blacklists. PhishFarm [20] deploys 2380 live phishing sites with subsets of these sites to 10 distinct anti-phishing entities.

7. Key Insights

Espionage is older than the Internet and has a big financial impact. It is hard to estimate precisely; the cost is easily the tens of billions of dollars. Such a large-scale activity, digital combatant in ATP is both organized and international. Our insights obtained from the reported blocklist database are as follows:

(1) ATP is being commercialized. Commercial and private companies are developing and providing the service concerning cyber-attack with different types of digital weapons: spyware, trojans, and scanning services working on contracts into arsenals preparing for cyber attacks. These activities recruit digital combatants by using private ISP lines which have previously worked in military units or security forces in some cases.

(2) Corporation among the different types of combatants. From the numerical results in the previous section, some countries are cooperating with private hacker groups. They are acting as affiliates and sometimes operating independently. In some cases, they are part-time combatants in the private sector. In others, they are state-sponsored digital combatants.

8. Conclusions

Due to the nature of digital warfare, combatants in APT are both organized and international. In this paper, we present the characterization of the combatants in state-sponsored APT by using AbuseIPDB. In characterization, we use two open-source IP address lists of Grizzly Steppe and Hidden Cobra. We have obtained JSON documents by queuing 877 IP addresses of Grizzly Steppe and 633 IP addresses of Hidden Cobra For both 877 and 633 IP addresses, we reduce the values in JSON documents in the view of country code and usage type. We can conclude that the two activities of Grizzly Steppe Hidden Cobra are distinguished by the distribution of country code and IP usage type. We have obtained the insights that combatants in digital warfare have the following characteristics: a highly decentralized and international distribution over the Internet, a lack of hierarchical authority of conducting countries, and hybrid of the public-private sector. Our further work is to inspect the structure of digital combatants in the view of lack of hierarchical authority, lack of formal structure.

References

[1] Chen P, Desmet L, Huygens C (2014) A study on advanced persistent threats. In: ifip International Conference on Communications and Multimedia Security, pp. 63-72

[2] Jeun I, Lee Y, Won D (2012) A practical study on advanced persistent threats. Computer applications for security, control and system engineering. Springer, Berlin, Heidelberg, pp. 144-152

[3] Moon D, Im H, Lee JD, Jong Park H (2014) MLDS: multi-layer defense system for preventing advanced persistent threats. Symmetry 6(4):997-1010

[4] GRIZZLY STEPPE - Russian Malicious Cyber Activity https://www.cisa.gov/uscert/GRIZZLY-STEPPE- Russian-Malicious-Cyber-Activity

[5] North Korea Cyber Threat Overview and Advisories https://www.cisa.gov/uscert/northkorea

[6] Bodmer S, Kilger M, Carpenter G, Jones J (2012) Reverse deception: organized cyber threat counter-exploitation. McGraw Hill Education.
[7] Reports and Publications of ODNI https://www.odni.gov/index.php/newsroom/reports-publications/reports-publications-2021
[8] U.S. Department of Defense Military and Security Developments Involving the Democratic People's Republic of Korea 2013 URL: https://fas.org/irp/world/dprk/dod-2013.pdf
[9] Reuters North Korea took 2 billion in cyberattacks to fund weapons program: U.N. report 05_AUG-2019 https://www.reuters.com/article/us-northkorea-cyber-un/north-korea-took-2-billion-in-cyberattacks-to-fund-weapons-program-u-n-report-idUSKCN1UV1ZX
[10] U.S. Office of the Director of National Intelligence 2021 Annual Threat Assessment April 9, 2021 https://www.dni.gov/files/ODNI/documents/assessments/ATA-2021-Unclassified-Report.pdf
[11]https://www.mhprofessional.com/details.php?isbn=0071772499. Accessed 24 June 2016
[12] Hidden Cobra https://www.cisa.gov/uscert/ncas/alerts/TA17- 318A
[13] AbuseIPDB https://www.abuseipdb.com/
[14] Peng Peng, Limin Yang, Linhai Song, Gang Wang: Opening the Blackbox of VirusTotal: Analyzing Online Phishing Scan Engines. Internet Measurement Conference 2019: pp.478-485
[15] Shuofei Zhu, Jianjun Shi, Limin Yang, Boqin Qin, Ziyi Zhang, Linhai Song, Gang Wang: Measuring and Modeling the Label Dynamics of Online Anti-Malware Engines. USENIX Security Symposium 2020: pp.2361-2378
[16] Jared Lee Lewis, Geanina F. Tambaliuc, Husnu S. Narman, Wook-Sung Yoo: IP Reputation Analysis of Public Databases and Machine Learning Techniques. ICNC 2020: pp.181-186
[17] Gouthum Karadi, Russian Fingerprints on the DNC: OSINT FOR RUSSIAN INFILTRATION OF THE DNC Preprint December 2017 doi: 10.13140/RG.2.2.12549.60649
[18] Peter KálnaiMichal, PoslušnýMichal Poslušný Lazarus Group: a mahjong game played with different sets of tiles Virus Bulletin International Conference, Montreal, Canada, October 2018
[19] Sushant Sinha, Michael Bailey, Farnam Jahanian: Shades of grey: On the effectiveness of reputation-based "blacklists". MALWARE 2008: pp.57-64
[20] Adam Oest, Yeganeh Safaei, Adam Doupé, Gail-Joon Ahn, Brad Wardman, Kevin Tyers: PhishFarm: A Scalable Framework for Measuring the Effectiveness of Evasion Techniques against Browser Phishing Blacklists. IEEE Symposium on Security and Privacy 2019: 1344-1361

Ruo Ando received Ph.D. from Keio University in 2006. He is now associate professor by special appointment of National Institute of Informatics since 2016. Before joining NII, he worked as senior researcher of National Institute of Information and Communications Technology since 2006. His research interests focus on network security, information security and big data mining technologies.He received Outstanding Leadership Award in the 8th IEEE International Conference on Dependable, Autonomic and Secure Computing (DASC-09) at China in 2009. He is the member of Trusted Computing Group JRF (Japan Regional Forum) in 2008-2015. He worked in project "Next Generation Security Info-Security R&D" METI (FY2008-10). He was engaged in project "Unknown malware detection using incremental malware detection" MEXT FY(2012-2015).

Hiroshi Itoh received the Master degrees from Keio University in 1980. He has Doctor of engineering from Keio University. He is now Executive Researcher in National Institute of Information and Communications Technology in Japan. He was working in Symantec Japan, Inc. from 2007 to 2010. He was Deputy Director-General for Cybersecurity and Information Technology Management in Ministry of Economy, Trade and Industry.