UNDERSTANDING TRAFFIC PATTERNS OF COVID-19 IOC IN HUGE ACADEMIC BACKBONE NETWORK SINET

Ruo Ando[1], Youki Kadobayashi[2], Hiroki Takakura[1], Hiroshi Itoh[3]

[1]National Institute of Informatics, 2-1-2 Hitotsubashi,
Chiyoda-ku, Tokyo 101-8430 Japan
[2]Nara Institute of Science and Technology 8916-5 Takayama,
Ikoma, Nara 630-0192 Japan
[3]National Institute of Information and Communications Technology 4-2-1 Nukui-Kitamachi, Koganei, Tokyo 184-8795, Japan

ABSTRACT

Recently, APT (Advanced Persistent Threats) groups are using the COVID-19 pandemic as part of their cyber operations. In response to cyber threat actors, IoCs (Indicators of Compromise) are being provided to help us take some countermeasures. In this paper, we analyse how the coronavirus-based cyber attack unfolded on the academic infrastructure network SINET (The Science Information Network) based on the passive measurement with IoC. SINET is Japan's academic information infrastructure network. To extract and analyze the traffic patterns of the COVID-19 attacker group, we implemented a data flow pipeline for handling huge session traffic data observed on SINET. The data flow pipeline provides three functions: (1) identification the direction of the traffic, (2) filtering the port numbers, and (3) generation of the time series data. From the output of our pipeline, it is clear that the attacker's traffic can be broken down into several patterns. To name a few, we have witnessed (1) huge burstiness (port 25: FTP and high port applications), (3) diurnal patterns (port 443: SSL), and (3) periodic patterns with low amplitude (port 25: SMTP) We can conclude that some unveiled patterns by our pipeline are informative to handling security operations of the academic backbone network. Particularly, we have found burstiness of high port and unknown applications with the number of session data ranging from 10,000 to 35,000. For understanding the traffic patterns on SINET, our data flow pipeline can utilize any IoC based on the list of IP address for traffic ingress/egress identification and port filtering.

KEYWORDS

COVID-19 related IoC, traffic patterns, data flow pipeline, port filtering, SINET, high port application.

1. INTRODUCTION

The outbreak of the COVID-19 pandemic crisis around the world had imposed an unexpected and great impact on everyday life. COVID-19 has become a global pandemic, which represents an opportunity for ATP groups to exploit the situation. Since January 2020, APT groups have done just that, targeting academic networks as well as government and non-government networks using COVID-19 as a lure. US-CERT releases the indicators of compromise concerning COVID-19-related security incidents [1]. In this paper, we present some traffic patterns unveiled from COVID-19 related IoC [2] in huge academic backbone network SINET. This paper is organized as follows: after the introduction and related work of section 1 and 2, we discuss the motivation in session 3 and background of our study in section 4. In section 5, we present the dataset. Our traffic observation is based on passive measurement and session data obtained by PaloAlto-7080

[3]. Methodology in section 6 is divided into three parts: architecture (dataflow pipeline), objectives (traffic discrimination and map-reduce) and components (Elasticsearch and Splunk). In section 7, we discuss traffic patterns unveiled by our system. In section 8, we discuss insights obtained in the observation of traffic flow in SINET. Then we conclude with some future works in section 9.

2. RELATED WORK

Favele et al. [4] presents the impact of the lockdown during pandemic on the Politecnico di Torino campus network. They discuss the robustness of the Internet for maintaining university operations. Baz et al. [5] present a research effort about the impact of COVID-19 pandemic on cybersecurity by ranking the various aspects of cybersecurity by using the multi-criteria decision-making problem-solving technique. In [6], COVID-19 pandemic is analyzed in the view of cyber-crime perspective. Besides, several cyber attacks experienced globally during the COIVID-19 pandemic is discussed. Yang et al. [7] presents an analysis of cybersecurity attacks on the smart home. Groenendaal et al. [8] conducts semi-structured in-depth interviews for analyzing how a global financial institution copes with the crisis of COVID-19.

In [9], they present a characterization of Amazon's Web Services(AWS) to unveil their infrastructure and pervasiveness of content. Drago et al. [10] presents a characterization of Dropbox based on passive measurements during 42 consecutive days. In [11], they aim at extracting and modelling traffic patterns of 9600 large scale towers deployed in the metropolitan city. Liu et al. [12] presents an analysis for five-month access trace of Tsinghua University. In [13], a dataset of 350 which consist of million HTTP requests from a large mobile cloud service are analyzed. Shafiq et al. [14] they present measurement and characterization of the spatial and temporal dynamics of mobile Internet traffic. In [14], they adopt a week-long aggregated flow level mobile device traffic data Ando et al [15] presents the task-decomposition based anomaly detection of massive and high-volatility traffic. In [16], they describe their experience with applying the Splunk log analysis tool in Sadia National Laboratories. Concerning detecting burstiness, Alsulaiman et al. [20] presents the evaluation of machine learning algorithms by Waikato Environment for Knowledge Analysis (WEKA) for the detection of DoS attack in wireless sensor network. In [21], markov chain is adopted for the early detection of DDoS attacks. In [22], a hybrid detection model of clustering and classification is proposed for IDS to work with STMP proxy.

3. MOTIVATION

The fundamental purpose of a security job is to defend assets, which presupposes that someone wants to attack. Network security is no exception. What kind of security measure is necessary depends, in part, on the types of attackers you are defending against. A threat actor, a burglar, or a government agent - each requires different countermeasures to protect. Usually, a single attacker can undertake various and different approaches. Usually, a single type of attack can be launched by many various attackers. These combinations of attacker and attack pose a threat. It is commonly said that ``know your enemy'' is important. If we mischaracterize our threat actors, we are likely to misallocate our defenses. To make matters worse, you are also likely to consider nonexistent risks and ignore real ones. Mischaracterizing is not always a disaster, but it is certainly more likely to result in one. Therefore, understanding patterns of COVID-19 related IoC can be helpful to take appropriate protective measures.

4. BACKGROUND

4.1. COVID-19 (as a lure)

The outbreak of the COVID-19 pandemic crisis is imposing a significant impact, which results in that unprecedented measures are taken by governments. According to [1], threat actors and APT groups are likely to exploit the current situation of the COVID-19. Threats observed are as follows:

(1) Phishing email with the subject about COVID-19 as a rule.
(2) Distributing malware using COVID-19 related lures.
(3) Domain name registration containing tokens related COVID-19.
(4) Attacks against rapidly deployed remote access and teleworking infrastructure.

In this paper, we are focusing on threats observed (1) and (2) by exposing traffic patterns observed in SINET.

4.2. NII-SOCS and SINET

NII-SOCS, which stands for NII Security Operation Collaboration Services were founded by The National Institute of Informatics in Japan in 2016. NII-SOCS is launched with the Center for Cybersecurity Research and Development and the administrative officials of the Academic Infrastructure Division. Our center had developed a network security infrastructure to observe traffic and detect cyber attacks in SINET in Japan. SINET is a Japanese academic backbone network. SINET is responsible for handling the networking of more than 800 universities and research institutions. SINET manages a variety of research facilities in space science, seismology, high-energy physics, nuclear fusion, computing science. Currently, in 2021, SINET has over 2 million users. Our team of NII-SOCS has deployed a dataflow pipeline. The pipeline consists of PaloAlto-7080 firewall, ELK stack, Splunk, and NVidia Multi-GPU server [17].

4.3. IoCs and application

IoCs are the artifacts such as malicious file hashes, domain names, or IP addresses that indicate intrusion attempts or other malicious behaviour. In this paper, we use the list of IP addresses according to [2].

Table 1. Ranking of ingoing application traffic.

application	session counts
impcomplete	3662138
non-syn-tcp	29651
unknown-tcp	6628
ssl	2087
insufficient-data	1834
portmapper	612
web-browsing	578
bittorrent	243
netbios-ns	55

Tables 1 and 2 show the application breakdown of the traffic observed in SINET by the IOC in [1]. In both cases, incomplete, unknown-TCP, and unknown-UDP occupy the top positions. From this, it is clear that most of the traffic generated by the attacker is unidentifiable.

Table 2. Ranking of outgoing application traffic.

application	session counts
impcomplete	1189311
ssl	1149867
unknown-tcp	56398
quic	50560
evernote-base	47709
web-browsing	42864
google-app-engine	39386
dns	6962
udemy-base	2904

5. DATA SET

Our analysis is based on the passive measurement by PaloAlto 7080. The thrust of the PA-7000 series is a scalable architecture for the purpose of adopting the flexible and powerful processing of the tasks of firewall and security management. A firewall such as PaloAlto-7080 plays an essential role in network security. Traffic patterns of session data are extracted from the session data of PA-7080. Table 3 depicts the session data format. Column 1-9 is about TCP/IP packet header related items. Column 19-23 is used to generate statistics. Column 12 (application) and column (category) are essential for characterizing the traffic patterns.

Table 3. PA 7080 session data description

No	item name	value
1	capture_time	2018/01/01 00:00:00.000
2	generated_time	2018/01/01 00:00:00
3	start_time	2018/01/01 00:00:00
4	elapsed_time	3
5	source_ip	xxx.xxx.xxx.xxx
6	source_port	0
7	src_country_code	JP
8	destination_ip	yyy.yyy.yyy.yyy
9	destination_port	0
10	dest_country_code	NA
11	protocol	NA
12	application	NA
13	subtype	NA
14	action	NA
15	session_end_reason	NA
16	repeat_count	0
17	category	NA
18	packets	0
19	packets_sent	0
20	packets_received	0
21	bytes	0
22	bytes_sent	0
23	bytes_received	0
24	device_name	NA

6. METHODOLOGY AND ARCHITECTURE

6.1. Dataflow Pipeline

We have implemented the data flow pipeline between Elasticsearch and Splunk. Figure 1 depicts a brief description of the workflow of the basic design of the pipeline. The main objectives of the pipeline are as follows:

(1) Retrieving data from Elasticsearch.
(2) Dividing session data into outgoing/ingoing.
(3) Passing outgoing/ingoing data to Splunk.

The components of the pipeline are as follows:

(1) DB1: ElasticSearch is responsible for storing session data.
(2) Server: Multi GPU server divides session data into ingress/egress chunks in coordinated batch processing manner.
(3) DB2: Splunk is stores output from multi GPU server to generate time-series data and provide some analytics with SPL command.

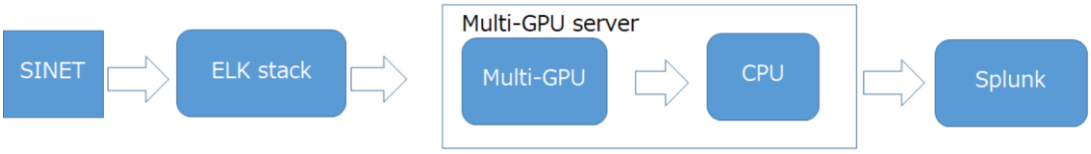

Figure 1. Our pipeline consists of three parts: ELK stack (Elasticsearch), Multi-GPU server and Splunk

Table 4. Parallelism

Objective	Parallelism	Algorithm	Granularity	Library
Dump from ELK Stack	task level	huge pagination	8 - 16 tasks	Intel SPDK
I/O identification	vector level	CIDR Matching	64 threads	CUDA Thrust
time series	thread level	Map Reduce	8 threads	Intel TBB

Our pipeline works sequentially from (1) to (3).

Table 4 shows parallel computing technologies adopted in three phases in our pipeline. First, retrieving data from ELK stack is leveraged by huge pagination in task level. Second, inside multi-GPU server, vector level parallelism is adopted for CIDR matching. Finally, Intel TBB is used for generating time series data by map-reduce algorithm in thread level. Ando et al. [18] discuss the effectiveness of the appropriate choice of parallelism in building the dataflow pipeline for handling huge traffic flow.

Algorithm 1. Dividing session data into ingoing/outgoing

Algorithm 1 session ingoing/outgoing discrimination

```
1: INPUT: chuncks of session data / CIDR list
2: OUTPUT: chuncks of ingoing session data
3: while chunk in workqueue is not EMPTY do
4:    CIDR notation = Y.Y.Y.Y/Z
5:    source IP notation = X.X.X.X/Z
6:    SB = bitmask (X.X.X.X, Z)
7:    CB = bitmask(Y.Y.Y.Y, Z)
8:    result = SB - CB
9:    if result == 0 then
10:      mark(chunk, outgoing)
11:   else
12:      mark(chunk, ingoing)
13:   end if
14: end while
```

6.2. Traffic Discrimination

Traffic discrimination is to identify traffic data into ingoing/outgoing. Traffic discrimination is essential to preprocess to analyze data in Splunk for time-series analysis to extract traffic patterns. Also, after diving session data into egress/ingress, we can proceed to the procedure of map-reduce to generate a histogram of traffic data. Algorithm 1 depicts the procedure step for traffic identification of ingoing/outgoing. At the first phase of this algorithm, session data have been divided into tens of chunks. The final goal of this algorithm is to mark ingoing/outgoing labels for each column of data. At lines 6 and 7, a bit sequence of source IP address session data (noted as SB) is obtained. Also, SINET address range (noted as CB) is provided for bitmasking. Then we proceed to line 9. If SD is equal to CB, the direction of the column is identified as outgoing. In the algorithm, instead of matching SD to CB directly, minus operator is used at line 8. Finally, if the result is 0, we add the flag of $outgoing$ at the head of the column. If the result equals 1, ingoing label is added to the column.

6.3. Map Reduce

The main objective of map-reduce is to generate a collection of key-value pairs of which the elements. Map-reduce consists of two steps:

(1) Taking a collection of data.
(2) Associating a value with each item in the collection.

The pair of keys and values should be independent of each input in the collection. The second phase of reduction is merging several different outputs from the previous (map) phase into a single output. The output of reduction is the representative data. The procedure of reducing should be repeated in order to yield the output down to a single value over an entire data set.

6.4. Elasticsearch

Elasticsearch is a distributed and full-text search engine based on the Lucene library. Elasticsearch is usually used as a NoSQL storage, indexing with unstructured documents.

Elasticsearch handles schema-free JSON documents. One of the most important advantages of Elasiticsearch is horizontal scalability. The session data obverse in SINET per day ranges from hundreds of millions to billions. To handle such huge session data, we adopt Elasticsearch with the powerful horizontal scaling. In our system, we are running 36 data nodes with 36 shards in 4 servers.

6.5. Splunk

Splunk is a semi-structured time-series database. Splunk can handle index, search and analyze heterogeneous datasets. Splunk is a feature-rich tool with over 140 commands of search processing language (SPL). SPL helps explore a massive amount of data to discover the needle in the haystack and the root cause of incidents. Mainly we use two SPL commands as follows:

(1) The time chart command is for a statistical aggregation to produce a chart with time plotted as the X-axis. It creates a chart with aggregated values against time shown in X-axis.
(2)
(3) The streamstats command calculates statistics at the time the event focused is seen, you can sum up the running total for a particular field in a cumulative manner. It figures out the total of the values in the specified field for every event, up to the current event. Figure 2 shows output table of Splunk streamstats which are defined as outlier from September 3 to 12.

_time	count	isOutlier	lower_bound	upper_bound
2021-09-04 04:30:00	37849	1	-18607.07523428202	25026.908567615355
2021-09-03 16:00:00	37045	1	-16794.65184262862	33728.65184262862
2021-09-03 15:50:00	33618	1	-12064.566730032479	23463.06673003248
2021-09-12 20:50:00	12261	1	-5873.607589120193	8142.440922453527
2021-09-06 07:40:00	11811	1	-6125.406924476041	9617.573591142707

Figure 2. Output table of Splunk streamstats commands

7. TRAFFIC PATTERNS

Traffic flow in SINET has high volatility, which may obscure the patterns of a lower range. As with the ingress traffic, there are often large spikes seen, but we can see a pattern of small but periodic fluctuations. The characterization of the observed traffic can be divided into the following four categories. The four categories are:

(1) patterns of bursts.
(2) patterns of daily cycles.
(3) volatility that fall within a specific range.
(4) patterns with bursts within periodic variations.

In the following subsections, we discuss burstiness and daily patterns.

7.1. All COVID-19 related IoC traffic

Figure 3 and Figure 4 show the time series data of the IOC traffic for COVID-19. As for the inward traffic, the number of sessions is usually only a few hundred, but more than 25,000

sessions were observed with the onset of bursts. This is about 25 times larger than normal traffic. Ingress traffic can be characterized as high volatility. For outgoing traffic, there are a few scattered bursts in the daily cyclic pattern. The cyclic pattern shows a range of traffic volume between about 1000 and 2500, but the burst observed on September 9 was about 4 to 10 times larger, more than about 11,000.

Figure 3. Timeseries of all ingress traffic of COVID-19 IoC on SINET. Traffic bursts are observed on September 3 and 4. During peak hours, more than 15 times the number of sessions than normal were observed. The third largest peak is due to FTP sessions

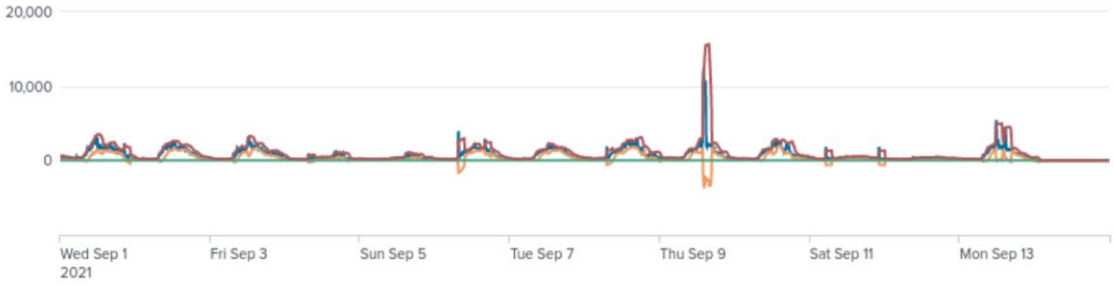

Figure 4. Timeseries of all egress traffic of COVID-19 IoC on SINET. The largest peak was observed around September 3. This was caused by the high port application communicating with outside SINET

7.2. Burstiness

Burstiness means a drastic traffic increase in time-series data. Coping with burstiness appropriately is important because of two reasons:

(1) Traffic burst is generated by scanning, attacks, and information leakage.
(2) Huge burstiness makes other important traffic patterns invisible.

Figure 5 shows the ongoing traffic of high port applications where the port number is higher than 1024. Figure 6 depicts the outgoing traffic with a port number is higher than 1024. As usual, session data volume is around a few hundred sessions. On September 3 and 4, the traffic volume exceeded over 30,000. The number of numbers observed is scattered around 60,000. It is worth noting that the burst timing is not consistent for egress/ingress high port application traffic. Based on this observation, it can be inferred that if attacks with the ingoing session are successful, the outward traffic as a response is not generated immediately and intentionally. It also became clear that at the time of the bursts of high port applications in Figures 5 and 6, the high port application traffic accounts for more than 95% of the total traffic.

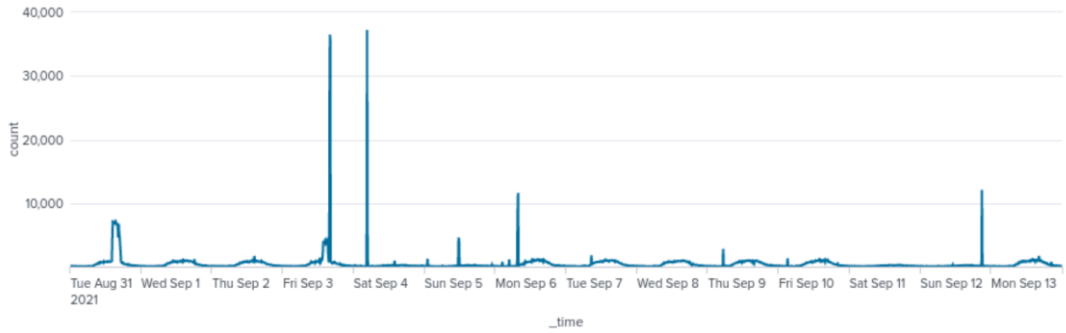

Figure 5. Time series of ingress traffic genearted by highport ($portNo > 1024$) application in SINET. Traffic is the accumulation of small sessions. The traffic is an accumulation of small sessions, and the connections are concentrated on 60,000 ports

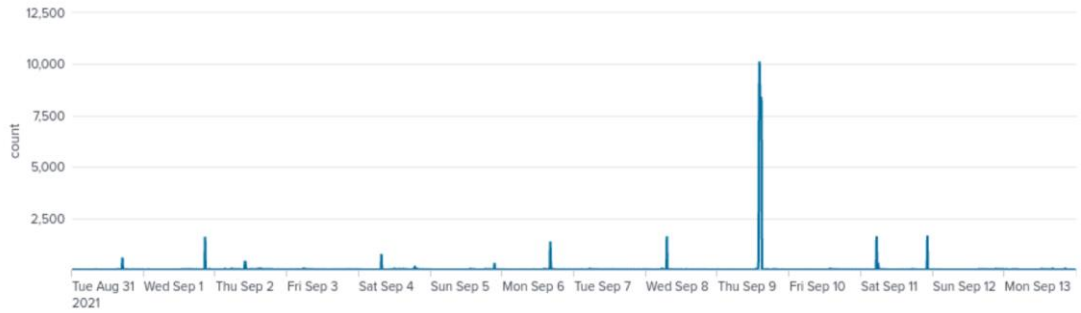

Figure 6. Time series of ingress traffic generated by highport ($portNo > 1024$) application in SINET.A spike of about 10000 sessions was observed on September 9. When analyzed in conjunction with Figure 5, this spike is not an immediate response to the inward spike and requires further analysis

Figure 7 depicts the ongoing traffic time-series using port 21 (FTP). Port filtering of 21 successfully reveals the patterns of burstiness with the number of sessions around 10,000. In daily observation, no FTP sessions are observed. Some other burstiness with the session number ranging from 500 to 2,200 have been observed. From a macroscopic point of view, the number of sessions during bursts is about 5,000, which is quite small compared to the entire SINET traffic volume. This suggests that the attackers have narrowed down the IP address range to be investigated in advance. It can be inferred that the threat actors might research the address range of SINET in detail.

Figure 7. Time series of ingress traffic of port-21(FTP). Other than the spikes observed on September 5 and 13, there are no FTP sessions under normal circumstances, so we assume that the attackers are systematically conducting reconnaissance to port 21

7.3. Diurnal Patterns

Figure 8 shows the outward traffic on port 443 from SINET. We can see a pattern of daily cycles, with session counts ranging from about a few dozen to about 2000. In addition, the amplitude of the cycle is different between weekdays and weekends. It is presumed that IP addresses of COVID-19 related IoC include the destination of normal traffic where SINET user accesses in daily usage. Figure 9 shows the time series of egress traffic port 25. STMP traffic shows a slight cyclical variation, though not a daily cycle. Since some kind of outgoing STMP traffic has been observed on the IP address identified as IoC, further investigation is required in some cases.

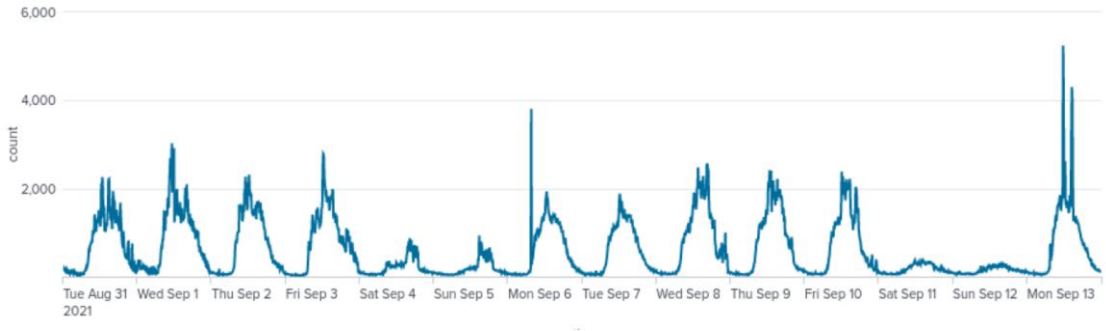

Figure 8. Time series of egress traffic of port 443. A cyclical pattern can be seen between weekdays and weekends. Perhaps some of the IoCs are generating normal traffic, for example, they may be providing some kind of web service. In addition, sessions of around 4,000 were observed on September 6 and 13, but this requires further investigation

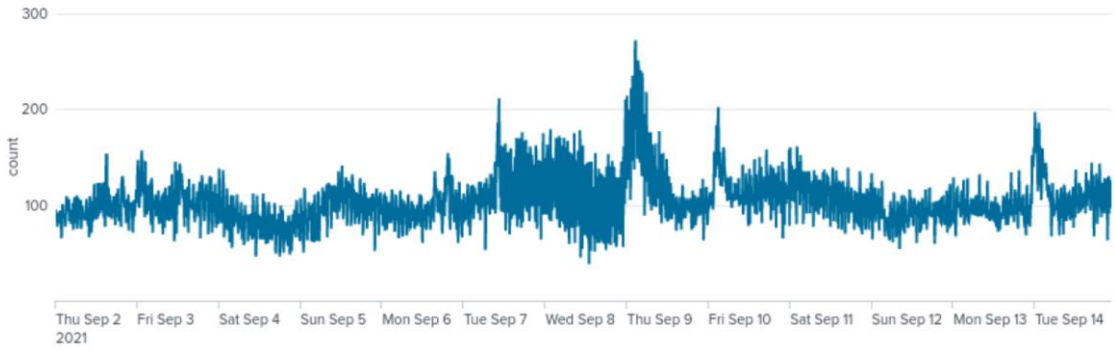

Figure 9. Time series of egress traffic port 25. Although the number of SMTP sessions has been hovering around 100, further investigation may be necessary in terms of communication to the servers listed as IoCs

7.4. Traffic Breakdown by application

Figures 9 and 10 depict the ratio of application traffic. Figure 9 shows the applications of outgoing traffic of SINET around traffic burst on September 9. The list of applications tends to be changed frequently according to the period of observation. From a long-term perspective, “incomplete” usually accounts for 60% of the total.

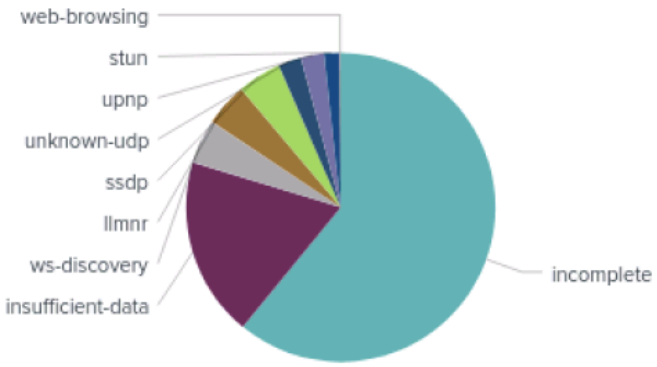

Figure 9. Traffic breakdown of application of outgoing session from SINET

Figure 10 shows the breakdown of ingoing application traffic of SINET around traffic burst on September 3. Compared with Figure 9, "incomplete" traffic accounts for a larger percentage. This means the traffic flow in the observation was not well recognized.

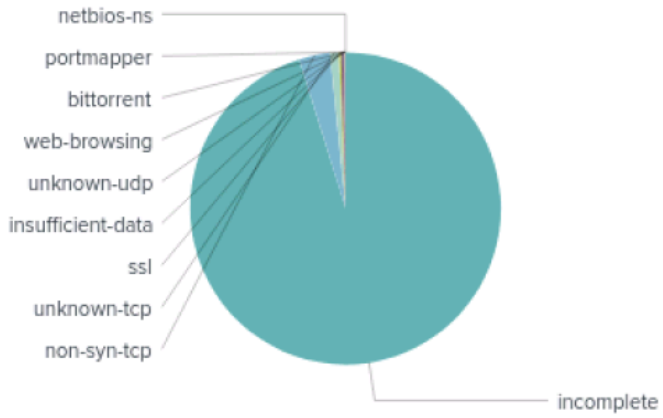

Figure 10. Traffic breakdown of application of ingoing session from SINET

8. INSIGHTS AND DISCUSSION

Both APT groups and cybercriminals continue to exploit the COVID-19 pandemic. Threats observed include:

(1) Phishing, using COVID-19 as a lure.
(2) Malware distribution.
(3) Malformed domain registration.
(4) Compromising remote access and teleworking infrastructure.

As for (1), no large-scale SMTP or webmail traffic was observed on SINET. This may be because the mail server is outside the group of IP addresses listed in the IOC. That is, if ATP groups send phishing emails, they may send it by the mail server outside SINET. As for (2), a large amount of high-port application traffic was observed during the observation period, and it is highly likely that malware sessions were included in the ingoing/outgoing traffic of SINET. Another possible scenario is that a large number of mail transmissions are occasionally observed

within SINET. Consequently, it can be inferred that the mail server inside SINET is being hijacked to send spam mail. There are several possible reasons why no significant spam mail traffic has been observed in the short term. One reason is that the sender's mail server is not located in the IoC; it is possible that a server inside SINET has been hijacked. Another reason may be that this IoC group is not in charge of Japanese spam mail due to the problem of Japanese localization of spam mail. As is commonly believed, no reconnaissance activities targeting the RDP (3389) were observed in SINET. There were many scans targeting high-port applications and FTP (21). In addition, a deeper analysis of the high-port application traffic is required.

Anomaly detection is an effective way to detect security incidents on the network. As we saw in Tables 1 and 2, as long as most of the traffic from the IoC is either unknown or incomplete, we have no choice but to use anomaly detection. However, in order for anomaly detection algorithms to be effective, they need to be filtered in advance with crafted traffic filtering. If the filtering as a pre-processing is not done well, the anomaly detection will not be successful. At this point, the problem lies in how to do efficient filtering for large-scale data.

In 2009 researchers from Google wrote a paper with the title The Unreasonable Effectiveness of Data [19]. In [19], they argue that a highly effective machine learning algorithm with a trillion lines of a dataset cannot work well with a clean dataset with a few million lines. The concept of our system is partly based on this paper. That is, first of all, the amount of observed session data is an issue to be considered. What kind of algorithm implementation and its tuning is proper often depends on the size of traffic data.
Concerning the countermeasures, the NII-SOCS team analyses the anomaly high port traffic and The NII-SOCS team will alert each university of the incident. It became clear that APT groups exploiting the current situation of COVID-19 are active in the use of high port traffic. BGP flowspec can be adopted for the mitigation of traffic bursts.

9. CONCLUSION

Recently, APT groups are exploiting the COVID-19 pandemic as part of their cyber operations. In response to cyber threat actors, IoC (Indicator of Compromise) is being provided to help us take some countermeasures. However, there have been few things understood concerning traffic patterns of COVID-19 related IoCs in the academic backbone network. In this paper, we analyze how the coronavirus-based threat actors behave on the academic infrastructure network SINET. Our analysis is based on the passive measurement with IoC. To extract and analyze the traffic patterns of the COVID-19 attacker group, we implemented a data flow pipeline for handling huge session traffic data observed on SINET. With the help of our dataflow pipeline, we have obtained some important insights. From the output of our pipeline, it is clear that the attacker's traffic can be broken down into several patterns. To name a few, we have witnessed (1) huge burstiness (port 25: FTP and high port applications), (3) diurnal patterns (port 443: SSL), and (3) periodic patterns with low amplitude (port 25: SMTP). Concerning the application breakdown on ingoing burst traffic around September 3 and outgoing burstiness on around September 9, a lot of incomplete application traffic was observed. At this point, COVID-19 IoC-related traffic was not well recognized and therefore needs to be inspected in more detail. It can be concluded that some unveiled patterns by our pipeline are informative to handling security operations of the academic backbone network. Future work includes a qualitative study of the traffic of high-port applications. As far as we can see, most of the traffic is homogeneously distributed over 60,000 port numbers, and we believe that anomaly detection can be effective using this feature.

ACKNOWLEDGMENTS

The authors would like to thank the NII-SOCS team.

REFERENCES

[1] US-CERT - COVID-19 Exploited by Malicious Cyber Actors
https://us-cert.cisa.gov/ncas/alerts/aa20-099a
[2] GitHub list of IOCs used COVID-19-related cyberattack campaigns gathered by GitHub user Parth D. Maniar
https://github.com/parthdmaniar/coronavirus-covid-19-SARS-CoV-2-IoCs
[3] Palo Alto Networks Enterprise Firewall PA-7080
https://www.paloguard.com/Firewall-PA-7080.asp
[4] Thomas Favale, Francesca Soro, Martino Trevisan, Idilio Drago, Marco Mellia: Campus traffic and e-Learning during COVID-19 pandemic. Comput. Networks 176: 107290 (2020)
[5] M. Baz, H. Alhakami, A. Agrawal, A. Baz and R. A. Khan, "Impact of covid-19 pandemic: a cybersecurity perspective," Intelligent Automation and Soft Computing, vol. 27, no.3, pp. 641-652, 2021.
[6] Harjinder Singh Lallie, Lynsay A. Shepherd, Jason R. C. Nurse, Arnau Erola, Gregory Epiphaniou, Carsten Maple, Xavier J. A. Bellekens: Cyber security in the age of COVID-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic. Comput. Secur. 105: 102248 (2021)
[7] Xin-She Yang, Simon James Fong, Segundo Moises Toapanta, Ion Andronache, Niko Phillips, "Cybersecurity attacks on smart home during Covid-19 pandemic", in Proc of the World Conference on Smart Trends in Systes, Security and Sustainability, WS4 2020, pp398-404, July 2020
[8] Jelle Groenendaal,Ira Helsloot, "Cyber resilience during the COVID-19 pandemic crisis: A case study", Journal of Contingencies and Crisis Management, May 2021
[9] LIgnacio Bermudez, Stefano Traverso, Marco Mellia, Maurizio M. Munafò: Exploring the cloud from passive measurements: The Amazon AWS case. INFOCOM 2013: 230-234
[10] Idilio Drago, Marco Mellia, Maurizio M. Munafò, Anna Sperotto, Ramin Sadre, Aiko Pras: Inside dropbox: understanding personal cloud storage services. Internet Measurement Conference 2012: 481-494
[11] Fengli Xu, Yong Li, Huandong Wang, Pengyu Zhang, Depeng Jin: Understanding Mobile Traffic Patterns of Large Scale Cellular Towers in Urban Environment. IEEE/ACM Trans. Netw. 25(2): 1147-1161 (2017)
[12] Songbin Liu, Xiaomeng Huang, Haohuan Fu, Guangwen Yang: Understanding Data Characteristics and Access Patterns in a Cloud Storage System. CCGRID 2013: 327-334
[13] Zhenyu Li, Xiaohui Wang, Ningjing Huang, Mohamed Ali Kâafar, Zhenhua Li, Jianer Zhou, Gaogang Xie, Peter Steenkiste: An Empirical Analysis of a Large-scale Mobile Cloud Storage Service. Internet Measurement Conference 2016: 287-301
[14] Muhammad Zubair Shafiq, Lusheng Ji, Alex X. Liu, Jia Wang: Characterizing and modeling internet traffic dynamics of cellular devices. SIGMETRICS 2011: 305-316
[15] Ruo Ando, Youki Kadobayashi, Hiroki Takakura, "Task-decomposition based anomaly detection of massive and high-volatility session data on huge academic backbone network", International Journal of Distributed and Parallel systems 12(1/2):pp1-9, 2021/3
[16] Jon Stearley, Sophia Corwell, Ken Lord: Bridging the Gaps: Joining Information Sources with Splunk. SLAML 2010
[17] Ruo Ando. Multi-gpu accelerated processing of timeseries data ofhuge academic backbone network in ELK stack. In 33rd LargeInstallation System Administration Conference (LISA2019), Portland, OR,October2019. USENIX Association.
[18] Ruo Ando, Youki Kadobayashi, Hiroki Takakura, "Choice of parallelism: multi-GPU driven pipeline for huge academic backbone network", International Journal of Parallel, Emergent and Distributed Systems", Volume 36, Issue4, 2021/6
[19] Alon Y. Halevy, Peter Norvig, Fernando Pereira: The Unreasonable Effectiveness of Data. IEEE Intell. Syst. 24(2): 8-12 (2009)

[20] Lama Alsulaiman and Saad Al-Ahmadi, "Performance Evaluation of Machine Learning Techniques for DoS Detection in Wireless Sensor Network", International Journal of Network Security & Its Applications (IJNSA) Vol.13, No.2, March 2021, pp21-29

[21] Chin-Ling Chen and Jian-Ming Chen, "Use of Markov Chain for early detection of DDoS Attacks", International Journal of Network Security & Its Applications (IJNSA) Vol.13, No.4, July 2021, pp1-11

[22] Ruo Ando, Yoshiyasu Takefuji, "Two-Stage Orthogonal Network Incident Detection for the Adaptive Coordination with SMTP Proxy", MMM-ACNS 2003: pp.424-427