Vol.2021-CSEC-95 No.18
Vol.2021-SPT-45 No.18
Vol.2021-EIP-94 No.18
2021/11/9

IoT マルウェア感染解析における通信形態及びアップリンク速度の推定手法

黄　緒平[1,2,a)]　望月　俊輔[3,b)]　吉岡　克成[2,c)]

概要：近年，IoT 機器を狙ったサイバー攻撃が増加している．マルウェアに感染した機器が踏み台として悪用され，不正な無線通信を大量に発生させる恐れがある．本研究はマルウェア感染による不正無線通信を抑制するため，収集した感染ログデータを解析し，IP アドレスから通信形態及びアップリンク通信速度を推定する手法を提案する．具体的に，GeoIP2 Connection Type DB による有線無線判定と自律システム（AS）番号を組み合わせてアドレスブロックを分割し，それぞれの範囲で Network Diagnostic Tool（NDT）の計測データと紐づけを行い，アップリンク通信速度を推定するメカニズムである．また，提案手法を IoT ハニーポットの実測値に適用し，接続形態を特定した後，無線通信の平均アップリンク推定速度が 40.6Mbps になり，プロバイダーの公称データと概ね合致した．

キーワード：IoT ハニーポットログ，無線削減，通信形態推定，アップリンク通信速度

IoT Malware Infection Analysis for Network Identification and Uplink Speed Estimation

Abstract: IoT botnets such as Mirai and its variants continue to evolve and keep consuming network resources, especially valuable radio resources. We show our approach to estimate the radio resources wasted by the IoT botnets. This paper proposes a cellular threats identification mechanism to identify the connection types, and to estimate the wireless uplink speed using IoT honeypot log, aiming developing a novel system to simulate the threat, by crosschecking Connection Type Database of Maxmind's GeoIP2, a well-known industrial resource for IP address related information, with uplink speed measurement data. Mobile Network Identification is approached by dividing IP addresses into IP ranges using Autonomous System (AS) numbers, combining the reverse DNS lookup solution. Network diagnostic tool database using IP ranges is used to calculate the uplink speed. To evaluate and verify the precision of the proposed method, we analyzed the IoT honeypot log as an alternation target application. The connection types are classified as "Cable/DSL", "Corporate", and "Cellular", and the maximum average uplink speed of cellular connection of the infected IPs is 40.6 Mbps, which is aligned with the average uplink speed of smartphone users of Softbank survey as 32 Mbps. Connection types and uplink speed is available using IP address as the input data by the proposed method, which may be used to estimate the attack scale.

Keywords: IoT Honeypot log, malware mitigation, mobile network identification, uplink speed estimation

1. Introduction

1.1 Background

With the high development of wireless devices and cellular communication techonologies, IoT connected devices are being daily used, and botnets have been a threat to cyber-security.From the first botnet that appeared, making advantageous of intelligent and stealthy propagation,

[1] Advanced Institute of Industrial Technology
10-40, Higashioi 1-chome, Shinagawa-ku, Tokyo 140-0011, Japan
[2] Yokohama National University
79-7 Tokiwadai, Hodogaya-ku, Yokohama, 240-8501, Japan
[3] NTT Data Mathematical Systems Inc.
1F Shinanomachi Brick Building, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-0016, Japan
[a)] huang-xuping@aiit.ac.jp
[b)] motiduki@msi.co.jp
[c)] yoshioka@ynu.ac.jp

distributed and large scale attacks, botnet has become one of the ideal platforms for malicious activities for cyber attacks. Symantec has reported more than 6,798,338 bots detected in 2019, and 85% of spam is from botnet [1]. Ubiquitous IoT devices enhance the daily live and improve the efficiency of business based on mobile communication services, however, they also probably be vulerability against DDoS attacks and cyber crimes [2,3]. Malware-infected devices can be used as a bastion host to generate a large amount of unauthorized wireless communication. There are increasing concerns regarding protecting IT infrastructure to mitigate the waste of cellular communication resources from these attacks.

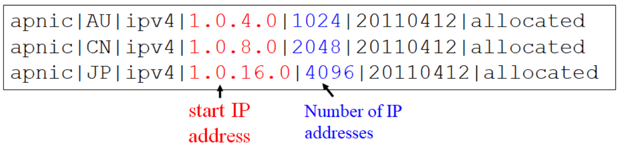
Due to the increasing popularity of the Internet of Things (IoT), which connects every piece of equipment to the internet to facilitate the communicationThe emergence of IoT malicious attacks introduces a vast ecosystem of new cellular communication with ubiquitous IoT devices. While ubiquitous IoT devices are as the prime targets as the cyber sabotage, such as DoS attack, which waste most of the IT communication resources. Most current cellular IoT services run over 4G cellular networks, and with the forecasted deployment of billions of IoT devices, the traffic characteristics of IoT devices are very different from those of smartphones and can enhance the risk for resource consumption.

Many industrial security schemes have been proposed to monitor the attacks from malicious attacks, and concerns IP spoofing and masquerade, such as the research results in NICT's work in Japan [4], which supplies a real-time monitoring of malicious attacks from variable devides, including the IP address of attack, detination IP, DNS. Using these monitering log, ATR Wave Engineering Laboratories developed a network type estimation method using RTT response time and port scanning [5,6]. Additionally, uplink speed of network is widely used to estimate the pattern of data trasmition [7], and to calculate the scale of malicious access to abuse the Internet resources [8], etc. As a reference, uplink speed is anlysis in alternative conventional works [9–16].

1.2 Scenario of usecase and contribution

We aim to develop a simulation against malicious attacks (DoS) mainly from ubiquitous wireless devides, which observes the wireless resources consumption through wireless attack.

As the priliminary process, we develop the mobile network identification method in this paper to identify the connection types of infected devices via IP addresses, and then to calculate the uplink speed of the traffic of attacks by analyzing the infection IoT honeypot log. The uplink speed is then used to simulate the ratio of mitigated radio resourses abused by unauthorized celullar communication caused by malware infection.



```
apnic|AU|ipv4|1.0.4.0|1024|20110412|allocated
apnic|CN|ipv4|1.0.8.0|2048|20110412|allocated
apnic|JP|ipv4|1.0.16.0|4096|20110412|allocated
```

図 1 IP addresses available in APNIC

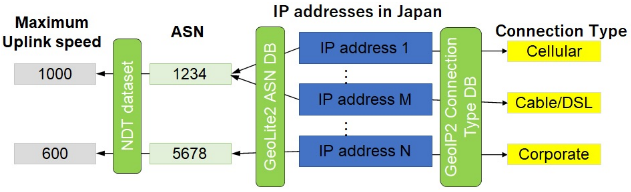


図 2 Illustration of crosschecking method

The main contribution of this paper is that: (1) we proposed a novel Mobile Network Identification to classify the connection types the infected devices in IoT honeypot log with a particular IP is from the Celullar or Cable/DSL by using reverse DNS lookup for DoS attack monitering; and (2) we estimate the uplink speed of the access by using the mobile access datasets, as a preliminary research for simulation of communication resource reduce by cutting off malicious access from DoS attacks.

The details of the proposed model will be discussed in Section 2. The results of celullar identification and uplink speed of communication are discussed in Section 3, and Section 4 concludes.

表 1 Mobile network identification results of attackers in IoT honeypot log

Connection Type	# of IP addresses	# of AS	# of Network Blocks
Cable/DSL	759 (76.3%)	225 (77.9%)	382 (72.6%)
Corporate	224 (22.5%)	56 (19.4%)	134 (25.5%)
Cellular	8 (0.8%)	4 (1.4%)	6 (1.1%)
Unknown	4 (0.4%)	4 (1.4%)	4 (0.8%)
Total	995	289	526

表 2 Correct ratio of mobile network identification

	Numbers of IP	Correct Ratio
Correct (Cellular)	39,655	99.5%
Incorrect (Cable/DSL)	212	0.5%

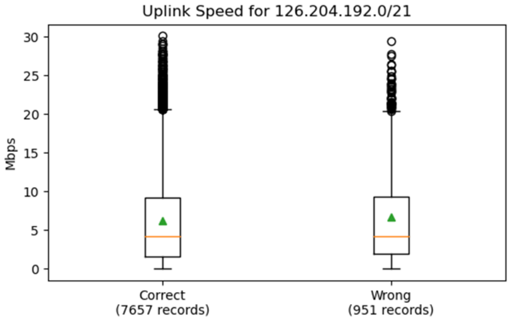
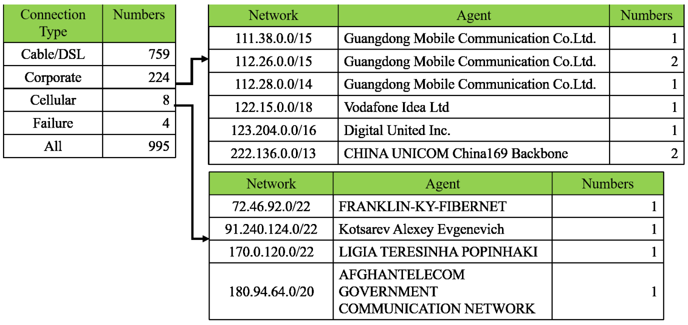


図 **3** Uplink speed of the correct and incorrect results for cellular connection by the proposed methods

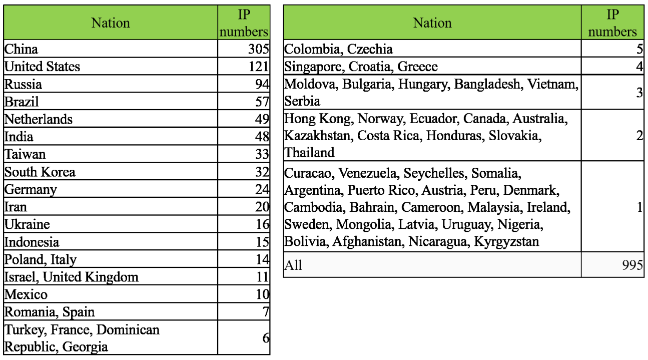


Connection Type	Numbers
Cable/DSL	759
Corporate	224
Cellular	8
Failure	4
All	995

Network	Agent	Numbers
111.38.0.0/15	Guangdong Mobile Communication Co.Ltd.	1
112.26.0.0/15	Guangdong Mobile Communication Co.Ltd.	2
112.28.0.0/14	Guangdong Mobile Communication Co.Ltd.	1
122.15.0.0/18	Vodafone Idea Ltd	1
123.204.0.0/16	Digital United Inc.	1
222.136.0.0/13	CHINA UNICOM China169 Backbone	2

Network	Agent	Numbers
72.46.92.0/22	FRANKLIN-KY-FIBERNET	1
91.240.124.0/22	Kotsarev Alexey Evgenevich	1
170.0.120.0/22	LIGIA TERESINHA POPINHAKI	1
180.94.64.0/20	AFGHANTELECOM GOVERNMENT COMMUNICATION NETWORK	1

図 **4** Connection types and ISP results of IoT honeypot log analylsis



Nation	IP numbers
China	305
United States	121
Russia	94
Brazil	57
Netherlands	49
India	48
Taiwan	33
South Korea	32
Germany	24
Iran	20
Ukraine	16
Indonesia	15
Poland, Italy	14
Israel, United Kingdom	11
Mexico	10
Romania, Spain	7
Turkey, France, Dominican Republic, Georgia	6

Nation	IP numbers
Colombia, Czechia	5
Singapore, Croatia, Greece	4
Moldova, Bulgaria, Hungary, Bangladesh, Vietnam, Serbia	3
Hong Kong, Norway, Ecuador, Canada, Australia, Kazakhstan, Costa Rica, Honduras, Slovakia, Thailand	2
Curacao, Venezuela, Seychelles, Somalia, Argentina, Puerto Rico, Austria, Peru, Denmark, Cambodia, Bahrain, Cameroon, Malaysia, Ireland, Sweden, Mongolia, Latvia, Uruguay, Nigeria, Bolivia, Afghanistan, Nicaragua, Kyrgyzstan	1
All	995

図 **5** Nations detected by IPs from IoT honeypot log

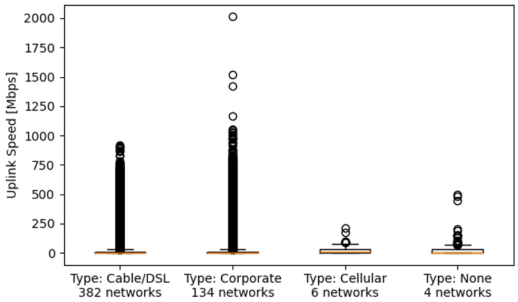


図 **6** Uplink speed detected by different types from IoT honeypot log

2. Mobile Network Identification and Uplink Speed Estimation

The proposed the crosschecking mechanism, aiming to moniter attacks from malware and malicious access using the ubiquitous IoTdevices, including the ISP and infected hosts identification for the network identification, uplink speed estimation, and IoT botnet's scope estimation from DoS attacks. Particularly, uplink speed estimation can be used to monitoring the infected hosts, which could be used to estimate the scope of the darknet. Furthermore, uplink speed is an essential parameter to be referred when we try to simulate the effectiveness of infection notification measurements. Thus, the proposed mechanism includes two functions for mobile network identification, and uplink speed estimation.

2.1 Mechanism for mobile network identification

We use Maxmind's GeoIP2 to decide which IP address ranges provide mobile communication services. More specifically, we use Connection Type Database that provides network connection types such as "Cable/DLS", "Corporate", "Cellular" for each IP address range. We also use GeoLite2 ASN Database to map IP address ranges to ASN. Finally, each address range is assigned ASN and connection type. Furthermore, we did the survey to check the coverage ratio of GeoIP database for IP addresses, which is being used in domastic Japan, by referring the IP addresses released in APNIC [17] as the preliminary step. The APNIC dataset is available as plotted in Fig. 1, which includes 190 million IP addresses, and GeoIP2 connection type database covers 85% of IP addresses in Japan.

Reverse DNS lookup scheme is used to identify the communication types. In the proposed scheme, we used Python geoip2 ver.4.1.0 library and GeoLite2-country datasets to identify the nations of the IP address in the IoT honeypot log data. Python dnspython ver.2.0.0 library and Google Public DNS Server (8.8.8.8) is specified for mobile network identification using reverse DNS lookup.

2.2 Mechanism for uplink speed estimation

We use the Network Diagnostic Tool (NDT) dataset in MLab project data, supplied in Google Cloud Platform (GCP) [18], the particular table "measurement-lab:ndt.unified_uploads", is used. The dataset is colleted during 2009/02/18-2020/08/18. We collected the IP addresses and upload speed from the dataset with a size of 947.4 GB. In order to verify the corectness of the proposed method, we compare the estimated uplink speed by collecting IP ranges of Softbanks users, in-

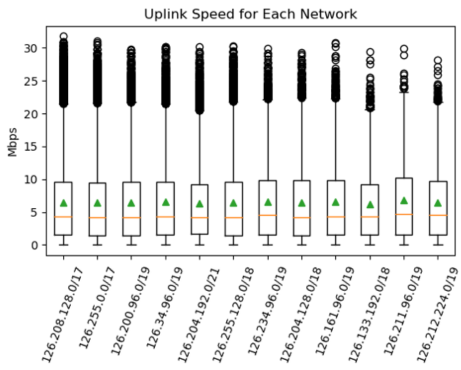


図 7 Uplink speed detected by different networks

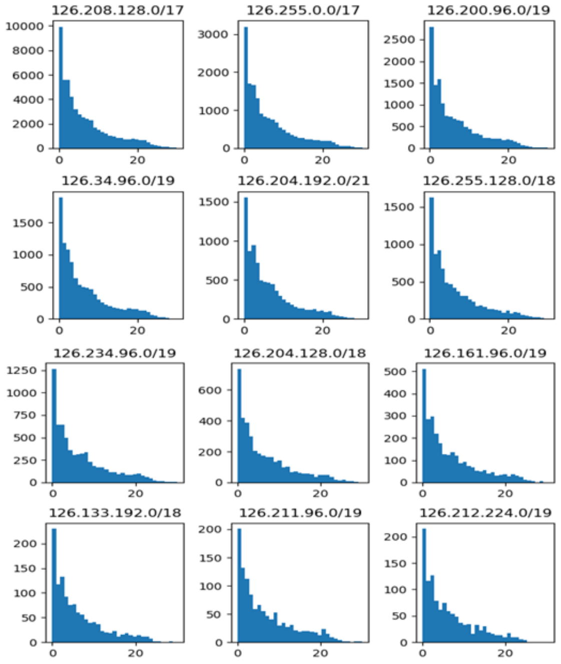


図 8 Histogram of uplink speed of cellular connection

cluding 107 networks, during 2019/4/1-2019/9/30 such as ”126.208.128.0/17”, etc, for comparation. We downloaded and analyzed the provided data and found that it covers 54.8% of ASNs in Japan.

The illustration is plotted in Fig. 2. The input data is the IP address of the infected devices, and the output data include the mobile network types, and the uplink speed.

3. Analysis results of IoT honeypot log

3.1 Mobile network identification results of attackers in IoT honeypot log

In order to crosscheck two different resources, connection types provided from GeoIP2 and uplink measurement data from MLab project, we use IoT honeypot log data [19], which provide IP addresses of infected host observed during 2020/11/18-2020/12/1 with 995 IPs. The classification of ASN and connection types by GeoIP2 is shown at Fig. 2. 759 IP addresses (76.3%) are classified as ”Cable/DSL”, 224 (22.5%) as ”Corporate”, and 8 (0.8%) as ”Cellular”. We then obtained uplink measurement data for each IP address range classified by ASN and connection types. The results are listed in Table .1. The connection types of the IP addresses of infected devices are classified properly. Furthurmore, as depicted in the table, uplink speed for ”Cellular” is much low compared to ”Cable/DSL” and ”Corporate”, with the average uplink speed of 40.6 Mbps.

In order to verify the effectiveness and precision of the proposed method, we apply the proposed method to the smartphone users of Softbank, Ltd in Japan. We analyzed and identified the types of communication towards 39,867 IPs. According to Softbank report [20], their average uplink speed is approximately 32 Mbps, which is aligned with the measurement. The correct ratio of mobile network identification by the proposed method is listed in Table 2. According to the result, 99.5% is identified to be ”Cellular” correctly, and only 0.5% is identified to be ”Cable/DSL” incorrectly. All of the incorrect IPs are belong to a particular IP range ”126.204.192.0/21”, and the uplink speed of the communication is plotted in Fig. 3. The uplink speed of the two results are approximately similar, thus we suggest that the 0.5% part of IPs be the result of false estimation.

Connection types and ISP results of IoT honeypot log analylsis is plotted in Fig. 4. According to this method, we can specify the ISP information of most attaks, and then estimate the scale of IoT attacks according to the ISP information.

Figure 5 lists the detection result of the nations from IoT honeypot log. According to the results, it suggest China, United States, and Russia ranks in top 3 nations with IoT attacks. As a target application, the proposed method can focuses on the nations of attack sources for monitering attacks during a specified events, for example, olympic games and bitcoin trading, etc.

3.2 Uplink speed for attackers in honeypot log

The results of uplink speed on networks of with more than 1000 IPs in each IP range are as follows. Figure. 6 plots uplink speed of different communication types, including Cable/DSL, Corporate, Cellular, and Unknown. The average uplink speed of cellular connection is 40.6 Mbps. Since the differences of the uplink speed between wire and wireless connections are obvious, it suggests the proposed mobile network identification method is proba-

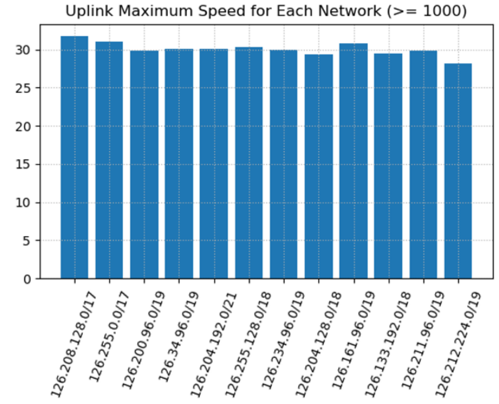


図 **9** Maximum values of uplink speed for each network with more than 1000 IPs

bly correct. Uplink speed categoried by the network connection types and nations/regions are available as the results.

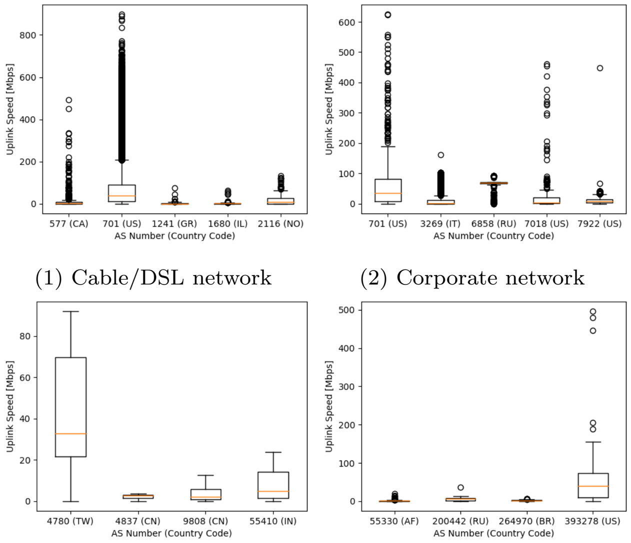
Figure. 7 plots the average uplink speed for each network for each IP range. Though the average uplink speed is various, the mainstream trend of each IP range is apprimately similar, which is conformity to the result of histogram of each network plotted in Fig. 8. Since the distribution feature of each network is similar, that we extract the maxmum uplink speed of each network with more than 1000 IPs to estimate the maximum uplink speed, which is plotted in Fig. 9. According to the results, a maximum of uplink speed of 32 Mbps is available, which is as the same as the open result of Softbank mobile users.

3.3 Discussion

Uplink speed of different networks by nations/regions and connection types are avalable from IP addresses by the proposed method. Figure. 10 plots the results partially from the IoT honeypot log that we collected. The detected uplink speed values depend on the period and data monitored in the honeypot log. This mechanism can be applied to any malware log with IP address of infected devices and timestamps. Specific cyber security incident occured from particular nations/regions can be observed, and furthermore, the scale of attack can be estimated by the proposed network identification and uplink speed, which could be a reference to make measurement against cyber pandemic.

表 **3** Details of data for analysis of uplink speed for communication types using IP address

Conection type	# of IP addresses	# of IP NDT dataset	Ratio of IPs in in NDT dataset	# of records
Cellular	7,925,504	181,556	2.29%	415,813
Cable/DSL	144,285,648	970,378	0.67%	3,675,924
Dialup	627,200	2,129	0.34%	4,247
Corporate	8,722,809	4,511	0.05%	34,749
None	28,567,223	21,738	0.08%	70,914
Total	190,128,384	1,180,312	0.62%	4,201,647



(1) Cable/DSL network

(2) Corporate network

(3) Cellular network

(4) Unknown

図 **10** Uplink speed for variable networks and nations/regions from IoT honeypot log (partially)

Furthermore, all of these results above are estimated and analyzed by AS numbers in the same IP range. Besides, in order to verify whether the proposed method can identify and divide the cable and celullar traffic precisely, we also experimentally analyze the uplink speed by each IP independently. The dataset we used is plotted in Table 3 from GeoIP2 connection type database, including 190,128,384 IPs. The results of uplink speed of each IP address by different network types are plotted in Fig. 11. According to the results, the uplink speed of cable, cellular, corporate are totally different, which matches the theoretical values approximately and makes cellular traffic, including IoT access distinguishable from all of the network communications using the proposed method, with only IP address as the input data. It suggests our proposed method could be applied to almost honeypot log to moniter infected devices in Japan.

4. Conclusion

In this paper, we proposed a novel mobile network identification method to identify the communication types (Cable/DSL, Corporate, Cellular, or Unknown using IP addresses from the collected IoT honeypot log data, by crosschecking multiple databases for IoT Botnet attack monitering. As a result, 99.5% is successfully identified by verifying the results referring to smartphone access

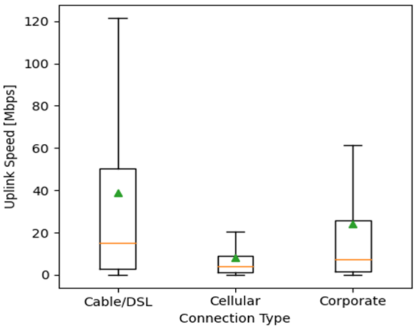


図 11 Uplink speed for different network types for IP addresses

IPs from Softbank, Ltd. Furthermore, we estimated the uplink speed of cellular communication by crosschecking mobile access datasets as a preliminary research. We also verified the precision of our method, by comparing the uplink speed of specified IP addresses in IP ranges specified in Softbank's report, partially. According to the statistical analysis results, the average of the maximum of uplink speed is 40.6 Mbps, which is align with the result released in Softbank, Ltd reported in 2019, which suggested the average uplink speed of smartphone users is approximately 32 Mbps. Thus, the effectiveness of the proposed celullar identify model has been verified.

The disadvatage of this method, which is also the challenge of this research field, that the proposed method can not detect or trace the IP masquerade, using proxy servers. Additionally, the verification of the precious of celullar identification is challenging, and the solution is challenge. Comparing the results to multiple real monitoring database towards cellular communication traffic, including IP addresses and communication types is as listed as a future work.

謝辞 This research was conducted under a contract of "MITIGATE" among "Research and Development for Expansion of Radio Wave Resources(JPJ000254)", which was supported by the Ministry of Internal Affairs and Communications, Japan.

参考文献

[1] Symantec Security Response team. Internet Security Threat Report, vol. 24, Feb 2019. https://docs.broadcom.com/doc/istr-24-executive-summary-en

[2] B. Stock, G. Pellegrino, F. Li, M. Backes, and C. Rossow, "Didnt You Hear Me Towards More Successful Web Vulnerability Notifications," The Network and Distributed System Security Symposium (NDSS), 2018.

[3] Y. Shen, P. A. Vervier, and G. Stringhini, "Understanding Worldwide Private Information Collection on Android", in The Network and Distributed System Security Symposium (NDSS 2019), pp. 1-16, 2019.

[4] O. Cetin, C. Ganan, L. Altena, D. Inoue, T. Kasama, K. Tamiya, Y. Tie, K. Yoshioka, M. van Eeten, "Cleaning Up the Internet of Evil Things: Real-World Evidence on ISP and Consumer Efforts to Remove Mirai", The Network and Distributed System Security Symposium (NDSS), 2019.

[5] R. Teng, et al, "Identification of IoT Network Type Based on The Response Delay Property," in Proc. ICACT2020, pp. 319–324, Feb. 2020.

[6] R. Teng, et al, "Statistical study on response delay characteristics of network scan to IoT devices - the base for estimation of IoT device attributes," IEICE Tech. Rep., NS2019-142, pp.45–49, Dec. 2019.

[7] E. Eyceyurt, Z. Josko, "Uplink Throughput Prediction in Cellular Mobile Networks", Internation Journal of Electronics and Communication Engineering, vol. (16)-6, 2020

[8] Z. W. Tan, J. H. Zhao, et, al, "Device-Based LTE Latency Reduction at the Application Layer", 18th USENIX Symposium on Networked Systems Design and Implementation, pp. 471-486, 2020.

[9] Nicola Bui, Foivos Michelinakis, and Joerg Widmer, "Fine-grained lte radio link estimation for mobile phones. Pervasive and Mobile Computing", vol.(49), pp.76–91, 2018.

[10] Otero, Y. Egi C., "Machine-Learning and 3D Point-Cloud Based Signal Power Path Loss Model for the Deployment of Wireless Communication Systems," IEEE Access, vol. 7, pp. 42507-42517, 2019.

[11] C. Yue, R. Jin, K. Suh, Y. Qin, B. Wang and W. Wei, "LinkForecast: Cellular Link Bandwidth Prediction in LTE Networks," IEEE Transactions on Mobile Computing, vol. 17, pp. 1582-1594, 2018.

[12] Y. Liu and J.Y.B. Lee, "An Empirical Study of Throughput Prediction in Mobile Data Networks," in IEEE Global Communications Conference, San Diego, 2015.

[13] D. Lee, D. Lee, M. Choi and J. Lee, "Prediction of Network Throughput using ARIMA," in International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Fukuoka, Japan, 2020.

[14] I. Oussakel, P. Owezarski, P. Berthou, "Cellular Uplink Bandwidth Prediction Based on Radio Measurements," in MobiWac, Miami, 2019.

[15] L. Liu, S. Zhang and R. Zhang, "Exploiting NOMA for Multi-Beam UAV Communication in Cellular Uplink," ICC 2019 - 2019 IEEE International Conference on Communications (ICC), pp. 1-6, 2019. doi: 10.1109/ICC.2019.8761670.

[16] S. K. Jha, R. Rokaya, A. Bhagat, A. R. Khan and L. Aryal, "LTE Network: Coverage and Capacity Planning — 4G Cellular Network Planning around Banepa," 2017 International Conference on Networking and Network Applications (NaNA), pp. 180-185, 2017. doi: 10.1109/NaNA.2017.23.

[17] "https://www.apnic.net/", accessed on Sep. 26, 2021.

[18] Network Diagnostic Tool, "https://www.measurementlab.net/tests/ndt/", Google, Inc.

[19] IoTPOT project of Yoshioka Lab: "https://sec.ynu.codes/iot/"

[20] Speed survey, "https://www.softbank.jp/mobile/network /service/speed-survey/", SoftBank Group Corp.