Darknet Traffic Analysis by Focusing on The Stability of Traffic

Napaphat Vichaidis[@#1], Toui Kanai[#2], Hiroshi Tsunoda[#3], Glenn Mansfield Keeni[*4]

[@#] *Faculty of Information Technology, Thai-Nichi Institute of technology, Bangkok, Thailand*
[1]napaphat@tni.ac.th
[#]*Graduate School of Engineering, Tohoku Institute of Technology, Miyagi, Japan*
[2]m142802@st.tohtech.ac.jp
[3]tsuno@m.ieice.org
[*]*Cyber Solutions Inc.Miyagi, Japan*
[4]glenn@cysols.com

Abstract— **Darknet is reachable but unused IP address space. Since legitimate hosts will generally have no reason to send packets to darknet, most of the packets seen in darknet are results of attacks, experiments or errors. Thus, darknet traffic analysis is a good candidate to understand the activities of attackers, worms, and infected hosts in the Internet. In this paper, we analyse darknet traffic by focusing on traffic stability. The concept of traffic stability is that the relative volume of dominant traffic components do not change drastically. We hypothesize that though the volume of darknet traffic is orders of magnitude smaller than Internet traffic, the stability principle holds and that the instabilities in traffic indicate the occurrence of some events in darknet. We categorize packets in darknet based on values of the fields in the packet header and calculate the volume of dominant components. We analysed two datasets of darknet traffic and found several significant instabilities. We analysed the causes of the instabilities and characteristics of the corresponding packet categories. Some of the detected events could be correlated with known and recorded network events. The analysis results show that traffic stability is a useful concept even for darknet traffic analysis.**

Keywords- **Network security, darknet, traffic analysis, traffic stability**

I. Introduction

With the growth in the number of Internet users and expanding scope of Internet applications, the number of security incidents in the Internet is also rapidly increasing. In one study, scan activity was 53.0%, website defacement 15.8%, phishing activity 13.9%, malware attacks 3.2%, targeted attacks 1.6%, DoS/DDoS attacks 0.6%, and others 12.0% [1]. Knowing and understanding the malicious activities and their mechanisms in the Internet is a challenge. Darknet traffic analysis is a good method to understand the activities of attackers, worms, and infected hosts.Darknet is reachable but unused IP address space. Since legitimate hosts will generally have no reason to send packets to darknet, most of the packets seen in darknet are results of attacks, experiments or errors. According to reference [2], darknet is one of the best method to monitor packets sent by malicious software. Darknet is often proposed as a means to monitor anomalous, external traffic sources, and require large, contiguous blocks of unused IP addresses [3].

In this paper, we analyse darknet traffic from the viewpoint of traffic stability. The concept of traffic stability assumes that the relative volume of dominant traffic components do not change drastically. Hence, we categorize packets in darknet based on values of the fields in the packet header and calculate the volume of dominant components.

We hypothesize that darknet traffic has the same stability characteristics as the normal (non-darknet) traffic. By observing the stability of darknet traffic, we will be able to detectevents and analyze activities of attackers, worms, and infected hosts on an internet scale.

The remainder of the paper is organized as follows: in Section 2 we review related works. In Section 3 we describe our analysis of darknet traffic. In section 4, we present the results. Finally, in Section 5 we give the conclusion of our work and discuss future works.

II. Related Works

This section introduces works related to darknet traffic monitoring and analysis.

A. Darknet monitoring systems

Several organizations have deployed darknet monitoring systems.

1) JPCERT/CC – TSUBAME [4]

TSUBAME is an Internet threat monitoring system operated by Japan Computer Emergency Response Team Coordination Center (JPCERT/CC). A number of traffic sensors are deployed in various address blocks in the Asia-Pacific region. They monitor the edge of xDSL lines and lines near Internet exchanges etc.

JPCERT/CC analyzes TCP, UDP, and ICMP packets monitored by TSUBAME and periodically reports trends of scans in the Internet [5]. Moreover, the observed data used as a basis of JPCERT/CC's activities *viz.* publishing alerts, advisories, and security awareness document.

2) NICT - nicter darknet [6]

The National Institute of Information and Communications Technology (NICT) in Japan operates a threat monitoring system called nicter. This system monitors darknet covering hundreds of thousand IP addresses. The monitoring results are utilized in the DAEDALUS Security [7] alerting system.

3) CAIDA- Network telescope [8]

Network telescope is a darknet monitoring system operated by the Center for Applied Internet Data Analysis (CAIDA).This is a passive traffic monitoring system built on globally routed, but lightly utilized /8 network that carries almost no legitimatee traffic because most IP address in this prefix are not assigned to any hosts. After disregarding traffic to the few hosts with assigned IP addresses, the remaining packets represent a continuous sample of anomalous unsolicited traffic. CAIDA makes available a number of dataset for researchers who wish to study data. It has been open to the public. The dataset on denial-of-service attack, backscatter, and Internet worms resulted in the following publications by external researchers [9].

B. *Darknetanalysis*

Darknet traffic analysis is carried out in order to understand the overall characteristics and trends in the internet. A lot of work has been done using the traffic of different darknet monitoring systems. Darknet use for deploying backscatter detectors, packet sniffers, or IDS boxes and provides overview of darknet implementation for collecting malicious activities in the Internet. The goals of darknet analysis are to increase awareness, and provide easy mitigation [10] .In their work, network telescope [11] provides the opportunity to view remote network security events such as various forms of flooding denial-of-service attacks, infection of hosts by Internet worms, and network scanning. In Darknet Monitoring on Real-Operated Networks [12], theauthors deployed two types of sensors, darknet sensors and honeypots, on real-operated network, captured darknet traffic, conducted real darknet monitoring experiments and clarified what kind of information could be obtained.

Darknets can produce vast quantities of multi-dimensional measurement data. The authors in [13] demonstrate a method for building and operating darknet monitors which is simple and productive. Darknet traffic can be classified as infection attempts by worm, botnet, misconfigured application, backscatter from spoofed denial of service attacks, and network scanning problems. The authors in [14] focus on TCP sessions initiated by activities from botnets or worms in the Internet and propose a multi-dimensional malicious packet monitoring architecture for internet threat detection. Our study can detect and analyze network events on darknet traffic and developments of the darknet system for protection against threats that may arise in the future.

III. DARKNET TRAFFIC ANALYSIS METHODBY FOCUSING ONTRAFFIC STABILITY

This section describes the darknet analysis technique by focusing on stability of traffic.

A. *Stability of traffic*

Internet traffic contains several groups. Each group is characterized by a unique packet type. The volume of a group is the number of packets in the group. A group is dominant if it contributes more than some pre-specified threshold, Th%, of the total traffic seen at the traffic sensor. The concept of internet traffic stability is that the relative volume of dominant traffic components do not change drastically. We apply this idea to darknet traffic analysis. We categorize darknet traffic packets into groupsbased on values of the packet header fields and calculate the volume of the groups. Changes in relative volume of groups, and the number of dominant groups will indicate instability. We hypothesize that an instability is caused by one or more events in the darknet. The concept of event detection by focusing on traffic stability is illustrated in Figure 1.

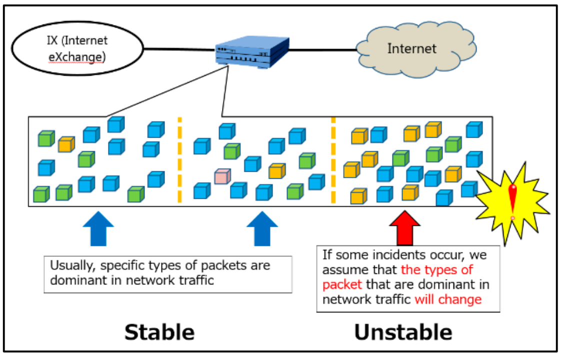
IX (Internet eXchange)
Internet
Usually, specific types of packets are dominant in network traffic
If some incidents occur, we assume that the types of packet that are dominant in network traffic will change
Stable
Unstable

Figure 1: Event detection by focusing traffic stability

As shown in Figure 1, specific types of packets are persistently dominant in network traffic when traffic is stable. If some events occur, the number of packets, type of packets and the number of packet types in dominant component will change. In the rest of this section, we explain the algorithm to judge traffic stability.

B. *Calculating TopN*

$TopN$ indicates the top groups, sorted by traffic volume, which contribute to network traffic above a threshold. It is determined by the traffic percentage of that group. Figure 2 illustrates $TopN$ calculation method. In this figure, packets are categorized based on the destination port number in TCP and UDP header fields.

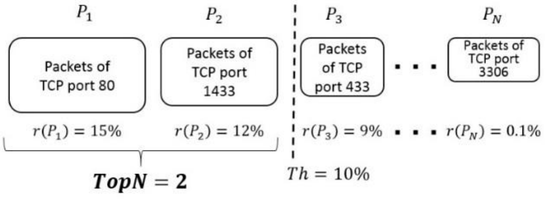

Figure 2:Calculation of TopN

In Figure 2, $P_i$represents i-th packet group sorted in descending order by volume of the group.$n(P_i)$ denotes the number of packets in $P_i(n(P_i) \geq n(P_{i+1}))$. $r(P_i)$, the occupancy of P_i, is calculated by Eq. (1). $TopN$ is the number of packet groups for which$r(P_i)$ is larger than a pre-defined threshold Th. In Figure 2, only $r(P_1)$ and $r(P_2)$ are larger than $Th = 10\%$. Hence$TopN$ becomes 2. In the proposed method, we evaluate the traffic stability based on $TopN$ and detect events in darknet traffic by investigating $TopN$ packet groups.

$$r(P_i) = \frac{n(P_i)}{\sum_{x=1}^{N} n(P_x)} \times 100\% \qquad (1)$$

By calculating and focusing on $TopN$, we can dynamically change the investigation target. Conventional methods often select the investigation target statically. Typically, $Top5$or $Top10$groups of packets are selected. However, focusing on a fixed number of packet groups sometimes leads to misunderstanding of events in traffic. For instance, when $Top10$ packet groups arefocussed on, a scan of 11 or more ports will be detected but its characteristics will not be correctly understood. On the other hand, in the case of a DDoS attack against a single port, investigation of all $Top10$ packet groups is not required. Focusing on $TopN$ groups is useful for efficiently selecting investigation target.

C. Evaluating Stability

The proposed method evaluates stability of traffic on a daily basis by observing $TopN$ groups of the day. Let $TopN_t$ denote $TopN$ value of a times lot t. Timeslot t is considered stable if $TopN_t$satisfiesEq.(2).

$$\mu - 2\sigma < TopN_t < \mu + 2\sigma \qquad (2)$$

In this formula, μ and σ denotes an average and standard deviation of $TopN$ for the last T consecutive timeslots from the timeslot t respectively. If $TopN_t$satisfies the condition below, traffic in timeslot t is considered stable.

IV. Analysis Result

A. Dataset

In this research, we analyse two datasets: traffic data monitored at JPCERT/CC TSUBAME and traffic data monitored at a local darknet. In this section, we give an overview of each dataset and then discuss the detected instabilities and their reason.

1) Dataset1: Tsubame

This dataset contains darknet traffic seen by the JPCERT/CC TSUBAME system over a period of 5 years (2010/1/1~2014/12/31). With 18,948,793 packets in the dataset the average number of packets/day is 10,365. Packets are categorized according to the destination port in TCP/UDP header. The contiguous data is split into slots of appropriate time duration for stability analysis TopN is calculated for$Th$values of 1.0 %, 0.75% and 0.5%. We show the variation of daily $TopN$ in Figure 3. The stability of each day was evaluated using 1-day slots$T$=7. We found 76 unstable days (76slots).

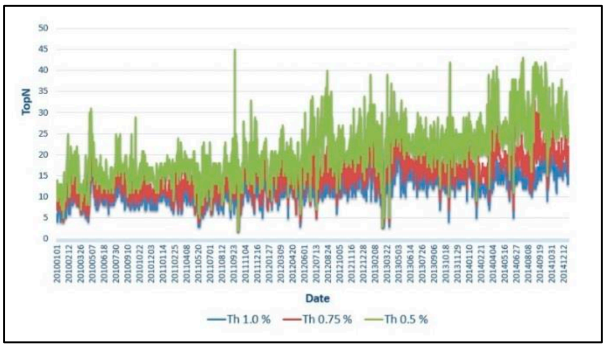

Figure 3: Variation daily TopN (Dataset1)

2) Dataset 2: A local darknet

This dataset contain s traffic seen over a period of 3 months (2015/12/13 ~ 2016/02/14) at a local darknet. With 12,277,716 packets, the average number of packets/day is 188,887. Packets are categorized according to the destination port in TCP/UDP header and $TopN$ is calculated forTh values of 1.0 %, 0.75% and 0.5%.The variation of daily $TopN$is shown in Figure 4. The stability of each day is evaluated using 1-day slots and T=7. We found 6unstable days (6 slots).

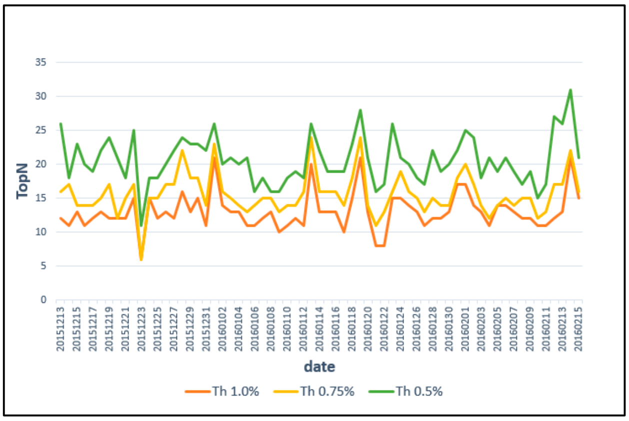

Figure 4: Variation of daily TopN (Dataset2)

B. Case studies from Dataset1

In this subsection, we show several anomalous events detected by careful investigation of traffic data in 76 unstable days.

1) Case 1: Instability on 2013/03/06

There was instability on this day due to the sudden decrease in $TopN$ value. $TopN$ decreased to 3from the range of 14to 19in the previous 7 days. To explore the probable event that caused the instability we examine the traffic in the $TopN$ groups in detail.

Table I shows the number of source IP addresses, number of destination ports, number of packets to each destination port, and type of packet in $TopN$ groups. A large number of packets are sent to UDP port 2849 on a specific sensor during relatively short period. Figure 5shows the number of packets per minute, sent to UDP port 2849. Those packets are sent from 233 hosts in a distributed manner. Therefore, we conclude that the instability was due to a Distributed Denial of Service

(DDoS) attacks. In DDoS attacks, many devices send a large number of packets all at once.

TABLE I

THE NUMBER OF PACKETS IN TOPN GROUPS OF 2013/03/06

Number of src hosts	Number of dst ports	Number of pkts	Dst port	Type of pkts
233	1	229627	2849	UDP
4488	1	19341	6667	TCP SYN
1001	1	3349	3128	TCP SYN

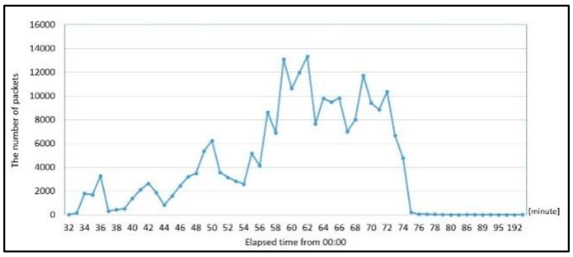

Figure 5: Number of packets sent to UDP port 2849

2) *Case 2: Instability on 2013/03/31*

There was instability on this day due to the sudden decrease of $TopN$to 3 from a range of 12 to 22in the previous 7 days.

Table II shows the number of source IP addresses, number of destination ports, number of packets to each destination port, and type of packet, in $TopN$ groups. A large number of packets are sent to UDP port 52290 during relatively short period. Figure 6 shows the number of packet per minute, sent to UDP port 52290. Those packets are sent from specific host but no specific service is defined on this port. Thus, we conclude that misconfiguration at source host is a probable cause of the instability.

TABLE II

THE NUMBER OF PACKETS IN TOPN GROUPSOF 2013/03/31

Number of src hosts	Numberof dst ports	Numbero f pkts	Dst port	Type of pkts
1	1	27079	5229 0	UDP
548	1	1101	445	TCP SYN
98	1	465	1433	TCP SYN

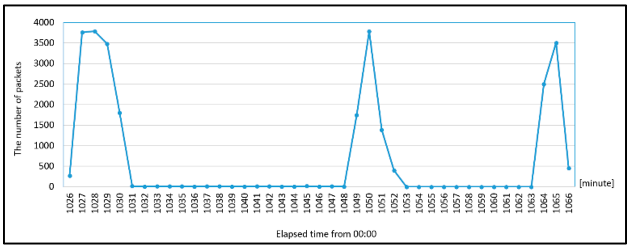

Figure 6: Number of packets sent to UDP port 52290

3) *Case 3: Instability on 2013/10/30*

There was instability on this day due to the sudden decrease of $TopN$ value to 4 from a range of 12 to 18 in the previous 7 days.

Table III shows the number of source IP addresses, number of destination ports, number of packets to each destination port, and type of packet, in $TopN$ groups. A large number of packets are sent to UDP port 3391 on a specific sensor during relatively short period. Figure 7shows the number of packets sent to UDP port 3391. UDP port 3391 is used for Remote Desktop Protocol (RDP) over Datagram Transport Layer Security (DTLS). Those packets are sent from a specific host. We conclude that the instability is caused by a misconfiguration at source host or scanning for potential reflectors of RDP over DTLS for a reflection attack.

TABLE III

THE NUMBER OF PACKETS IN TOPN GROUPS OF 2013/10/30

Number of src hosts	Number of dst ports	Number ofpkts	Dst port	Type of pkts
1	1	21429	3391	UDP
542	1	1085	445	TCP SYN
79	1	491	1433	TCP SYN
109	1	276	22	TCP SYN

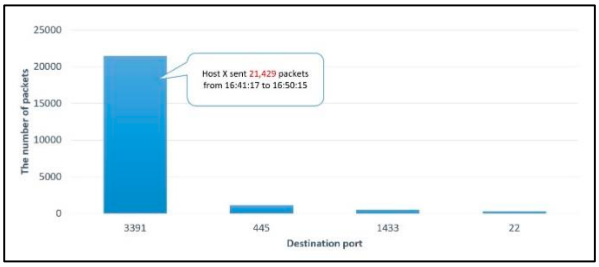

Figure 7: Number of packets sent to UDP port 3391

4) *Case 4:Instability on 2014/01/02*

There was instability on this day due to the sudden decrease of $TopN$ to9from a range of 13 to 19 in the previous 7 days.

Table **IV** shows the number of source IP addresses, number of destination ports, number of packets to each destination port, and type of packets in the $TopN$ groups. A large number of packets are sent to UDP port 3395 on a specific sensor during relatively short period. Figure 8 shows the number of packets sent to UDP 3395. UDP port 3395 is sometimes used for Remote Desktop Protocol (RDP) over Datagram Transport Layer Security (DTLS).

We also conclude that instability is caused by misconfiguration or scanning for a potential reflector of RDP over DTLS for a reflection attack.

TABLE IV

THE NUMBER OF PACKETS IN TOPN GROUPS OF2014/01/02

Number of src hosts	Number of dst ports	Number ofpkts	Dst port	Type of pkts
1	1	7514	3395	UDP
402	1	824	445	TCP SYN
126	1	499	22	TCP SYN
75	1	389	1433	TCP SYN
86	1	286	80	TCP SYN
96	1	265	3389	TCP SYN
7	1	153	1234	TCP SYN/ACK
58	1	143	8080	TCP SYN
30	1	137	5900	TCP SYN

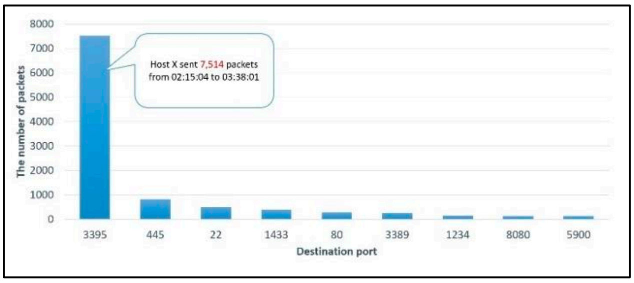

Figure 8: Number of packets sent to UDP port 3395

5) *Case 5: Instability on 2014/06/02*

There was instability on this day due to the sudden decrease of $TopN$ to 9from a range of 11 to 16 in the previous 7 days. Figure 9shows the number of packets sent to each destination port. A single host sends a large of packets sent to UDP port 4614 on a specific sensor. We assume that instability is caused by misconfiguration or scan activity to port4614 [15].

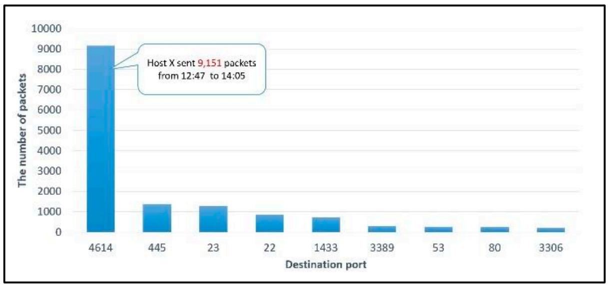

Figure 9: Number of packets to each destination port in **TopN** groups

6) *Case 6: Instability on 2014/06/16*

This day is considered as instable due to sudden decrease of $TopN$ to 5from a range of 12 to 17in the previous 7 days. Table V shows the number of source IP addresses, number of destination ports, number of packets to each destination port, and type of packet in $TopN$ groups. Three hosts sent a large number of packets sent to UDP port 4614 on specific host during relatively short period. Figure 10 shows the number of packets sent from each source host. We consider that instability is cause by misconfiguration or scan attack.

TABLE V

THE NUMBER OF PACKETS IN TOPN GROUPS OF 2014/06/16

Number of src hosts	Number of dst ports	Number of pkts	Dst port	Type of pkts
3	1	23092	4614	UDP
247	1	1207	22	TCP SYN
568	1	1173	445	TCP SYN
159	1	763	1433	TCP SYN
152	1	598	23	TCP SYN

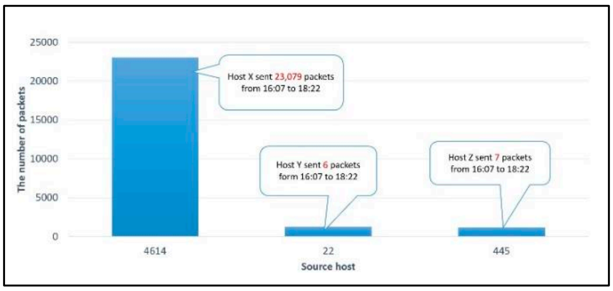

Figure 10: Number of packets from each source host

7) *Case 7: Instability on 2014/10/15*

There was instability on this day due to the sudden decrease of $TopN$ to 9from a range of 18 to 23in the previous 7 days. We conclude that the instability was likely due to misconfiguration of P2P software. Table VI shows the number of source IP addresses, number of destination ports, number of packets to each destination port, and type of packet in $TopN$ groups. A large number of packets sent to UDP port 12543 on specific host during relatively long period. Those packets are sent from 4445 hosts. JPCERT/CC also detected this suspicious activity and concluded that this is due to the misconfiguration of P2P software [16].

TABLE VI

THE NUMBER OF PACKETS IN TOPN GROUPS OF 2014/10/15

Number of src hosts	Number ofdstports	Number of pkts	Dst port	Type of
4445	1	23747	12543	UDP
2	1	1716	65328	TCP SYN
507	1	1023	445	TCP SYN
279	1	926	22	TCP SYN
1	1	852	50206	TCP SYN
162	1	700	1433	TCP SYN
229	1	667	23	TCP SYN
6	1	603	53	TCP SYN
145	1	454	80	TCP SYN

C. *Case studies from Dataset2*

We show two case studies from the second dataset.

1) *Case 1: Instability on 2015/12/23*

There was instability on this day due to the sudden decrease of $TopN$to 6 from a range of 11 to 15 in the previous 7 days.

Figure 11 shows the number of packets sent to each destination port. A large number of packets are sent to UDP port 64157 and 64066 on a specific sensor for a relatively long period. These packets are sent from UDP port 53 of a specific host. It is likely that this event is the result of backscatter from an attack against a specific DNS server.

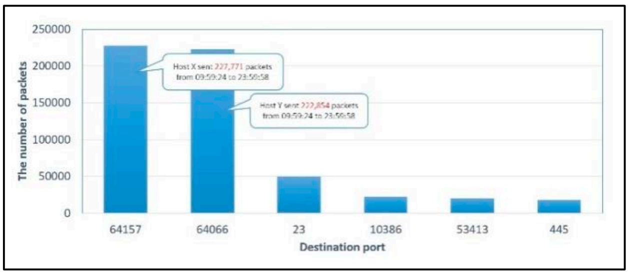

Figure 11: Number of packets sent to destination port in**TopN** groups

2) Case 2: Instability on2016/01/01

There was instability on this day due to the sudden increase of $TopN$to 21 from a range of 11 to 16 in the previous 7 days. Table VII shows the number of source IP addresses, number of destination ports, number of packets to each destination port, and type of packet in $TopN$ groups. A large number of packets sent to TCP port 23. However, we focus on number of packet sent to UDP port 64187 on specific host during relatively long period. Those packets are sent from UDP port 53 on a specific host. It is considered that this event is backscatter of the attack against specific DNS server.

Moreover, several host sends UDP packets to several ports during relatively short period. Thus, we assume that some scan activity affected to the instability on this day.

Thus, we conclude that the instability was likely due to the combination of backscatter from an attack on a DNS server and a scan.

TABLE VII

THE NUMBER OF PACKETS IN TOPN GROUPSOF2016/01/01

Number of src hosts	Number of dst ports	Number of pkts	Dst port	Type of pkts
3759	1	43438	23	TCP
1	1	33806	64187	UDP
5740	1	15586	445	TCP
35	1	11726	53413	UDP
273	1	6303	22	TCP
400	1	4098	80	TCP
205	1	3653	8000	TCP
1	1	3616	23387	UDP
1	1	3595	29630	UDP
1	1	3573	16635	UDP
1	1	3519	32516	UDP
1	1	3472	27998	UDP
1	1	3454	5919	UDP
1	1	3248	45703	UDP
1	1	3186	11252	UDP
540	1	2927	3389	TCP
19	1	2906	5060	UDP
265	1	2591	3306	TCP
300	1	2373	53	UDP
62	1	2324	123	UDP

V. CONCLUSIONS

In this paper, we have applied stability based traffic analysis to darknet traffic

We have proposed that stability based traffic analysis can be used for event detection and categorization on on internet-scale. The concept of traffic stability is that the relative volume of dominant traffic components does not change drastically. We have used real-life datasets for experimentation. We calculated $TopN$ groupsand evaluated stability of traffic on a daily basis by using $TopN$ of the day. The result of this analysis, we have found 76 unstable days in traffic monitored at JPCERT/CC TSUBAME and 6 unstable days from traffic monitored at a local darknet system. Due to the time limitation, we have examined a part of the detected instabilities and traffic, and have shown that the instabilities are due to malicious events such as Denial-of-Service attacks, scans, backscatters etc.

We conclude that stability of traffic is a potential candidate to understand the activities of attackers in the Internet. In future works, we plan to analyze stability of traffic using header fields other than destination port.

ACKNOWLEDGMENT

The authors would like to thank Japan Computer Emergency Response Team Coordination Center for providing traffic data seen at their Internet threat monitoring system.

REFERENCES

[1] "JPCERT/CC Incident Handing Report," Japan Computer Emergency Response Team Coorditional Center, JPCERT-IR-2015-04, Jul.–Sep. 2015.

[2] A. Shimoda, T. Mori, and S. Goto, "Extended Darknet: Multi-Dimensional Internet Threat Monitoring System," *IEICE Transactions on Communications*, vol. E95-B, no. 6, pp. 1915–1923, 2012.

[3] L. Miao, W. Ding, and H. Zhu, "Extracting Internet Background Radiation from raw traffic using greynet," in *2012 18th IEEE International Conference on Networks (ICON)*, 2012, pp. 370–375.

[4] "TSUBAME Info," Japan Computer Emergency Response Team Coordination Center, 2015.

[5] "Quarterly Report," Japan Computer Emergency Response Team Coordination Center, 2015.

[6] "National Research and Development Institute of Information and Communications technology, NICT Cyber security Laboratory," 2014. [Online]. Available: http://www. nicter. jp/nw_public/scripts/index.php#nicter.

[7] D. Inoue, M. Eto, K. Suzuki, M. Suzuki, and K. Nakao, "DAEDALUS-VIZ: Novel Real-time 3D Visualization for Darknet Monitoring-based Alert System," in *Proceedings of the Ninth International Symposium on Visualization for Cyber Security*, New York, NY, USA, 2012, pp. 72–79.

[8] "The UCSD Network Telescope," *Center for Applied Internet Data Analysis*, 2015. [Online]. Available:http:// www. caida. org/projects/network_telescope/.

[9] Z. Zhang, B. Wang, and J. Lan, "Identifying elephant Flows in internet backbone traffic with bloom filters and LR U," *Computer Communications*, vol. 61, pp. 70–78, May 2015.

[10] "Team Cymru Darknet Project," *Team Cymru*, 2014. [Online]. Available: http://www.team-cymru.org/darknet.html.

[11] D. Moore, C. Shannon, G. M. Voelker, and S. Savage, "Network Telescopes: Technical Report," CS2004-0795, 2004.

[12] S. Mizoguchi, Y. Fukushima, Y. Kasahara, Y. Hori, and K. Sakurai, "Darknet Monitoring on Real-Operated Networks," in *Proceedings of the 2010 International Conference on Broadband, Wireless Computing, Communication and Applications*, Washington, DC, USA, 2010, pp. 278–285.

[13] M. Bailey, E. Cooke, F. Jahanian, and A. Myrick, "Practical Dark net Measurement," in *40th Annual Conference on Information Sciences and Systems*, 2006, pp. 1496–1501.

[14] A. Shimoda, T. Mori, and S. Goto, "Extended Darknet: Multi-Dimensional Internet Threat Monitoring System," *IEICE Transactions on Communications*, vol. 95, pp. 1915–1923, 2012.

[15] C. Rossow, "Amplification Hell: Revisiting Network Protocols for DDoS Abuse." NDSS Symposium, 22-Feb-2014.

[16] "JPCERT/CC Internet Threat Monitoring Report," Japan Computer Emergency Response Team Coordination Center, JPCERT-IA-2015-01, Dec. 2014.