Vol.2021-DPS-186 No.43
Vol.2021-CSEC-92 No.43
2021/3/16

ハニーポットにより観測されるDRDoS攻撃の影響評価と要因分析

新谷 夏央[1,a)]　牧田 大佑[2,b)]　吉岡 克成[3,c)]　松本 勉[3,d)]

概要：これまで，DRDoS 攻撃の観測が可能なハニーポットにより，攻撃対象や頻度，攻撃継続時間などに関して多くの分析がなされている．しかしながら，DRDoS ハニーポットにより観測された攻撃が攻撃対象にどの程度の影響を与えているかについてはよく知られていない．そこで，本稿では DRDoS ハニーポットによって観測された攻撃対象の可用性に与える影響を ICMP パケットを送付することによって測定した．これによって，半数の測定対象の可用性が攻撃を受けていない時に比べて低下していることが確認された．特に，対象の 10%についてはほとんど通信不能となっていた．また，DRDoS 攻撃が可用性に与えた影響と，攻撃や攻撃対象の特徴との相関について分析した. その結果，攻撃の継続時間や時間当たりの攻撃パケット数，攻撃対象の国，AS, ドメインの累積名前解決件数が影響の程度と関連があることを確認した．さらに，ロジスティック回帰を用いて，攻撃を検知した直後に当該攻撃が対象に深刻な影響を与えるかどうかを判定する手法を提案する．

キーワード：DDoS 攻撃，DRDoS 攻撃，ハニーポット，影響度分析

Measurement and Factor Analysis of the Impact of Amplification DDoS Attacks Observed by Amppot

Natsuo Shintani[1,a)]　Daisuke Makita[2,b)]　Katsunari Yoshioka[3,c)]　Tsutomu Matsumoto[3,d)]

Abstract: A number of studies have utilized a honeypot to observe and analyze amplification DDoS attacks. However, little is known about the actual impact of these attack observed by the honeypot. In this study, we measured the impacts of the attacks observed by amplification DDoS honeypot, or AmpPot, by sending ICMP ping to the victims. We found that half of the victims was affected by the attacks. Moreover, about 10% of victims hardly responded during the attacks. We also show that various factors such as attacks duration, packets per second, victims' countries and ASes, the number of cumulative name resolutions, the number of related domains, are related to the attack impact. We then propose a method to decide the attack impact just in time of attack observation by AmpPot. Our method utilizes logistic regression to determine whether an observed attack will significantly affect the victims or not.

Keywords: DDoS, Amplification DDoS, Honeypot, Impact Analysis

1. Introduction

In recent years, amplification DDoS attacks have been threat on the Internet. Amplification DDoS attack is a type of Distributed Denial-of-Service (DDoS) attacks which puts stress on target network resources by sending

[1] Graduate School of Yokohama National University, Environment and Information Sciences 79–7 Tokiwa-dai, Hodogaya-ku, Yokohama,Kanagawa, 240–8501 Japan
[2] National Institute of Information and Communications Technology　Koganei, Tokyo 184–8795, Japan
[3] Graduate School of Environment and Information Sciences, Yokohama National University / Institute of Advanced Sciences, Yokohama National University
[a)] shintani-natsuo-yx@ynu.jp
[b)] d.makita@nict.go.jp
[c)] yoshioka@ynu.ac.jp
[d)] tsutomu@ynu.ac.jp

a large number of packets to target IP addresses from multiple devices existing on the Internet. In particular, amplification DDoS attacks abuse servers that provide NTP, Memcached, DNS, etc. as reflectors. In February 2020, Amazon Web Service was attacked by 2.3 Tbps amplification DDoS attack that abused CLDAP[1]. Even an attacker who does not have knowledge or skills about DDoS attacks can conduct attacks by using so-called ”Booter” services. With such circumstances, it is possible that DDoS attacks will increase and become a further threat on the Internet.

In order for effective countermeasures, it is necessary to analyze the situation and tendency of attacks, and one of the observation methods is AmpPot [2]. By preparing a server that imitates a reflector on the Internet, we observe how an attacker abuses it to conduct the attacks. AmpPot can observe attack packets sent by attackers. Then we observe the IP address of the attack target, the protocol used for the attack, the amount of packets sent to reflectors (honeypots), the duration of the attacks, etc. by analyzing the attack packets. However, we can not measure the effect on the target with AmpPot.

In this paper, we investigate the effect of the attack on the attack target. When an attack is observed by AmpPot, we send a measurement ping to the attack target. The responses from the target under attack is compared with those in normal situation measured after the attack is terminated. First, we analyze the measurement results to understand the actual situation of the targets. Further, we investigate relationship between impact and different factors of observed attacks. Secondly, we propose a classification model in order to quickly find a serious attack from the large number of attacks observed by AmpPot.

In Sect. 2, we briefly explain about amplification DDoS and AmpPot. In Sect. 3, we first explain how to measure the impact of amplification DDoS attacks detected by the AmpPots. In Sect. 4, we explain our dataset. In Sect. 5, we analyze result and investigate the possibility of just-in-time estimation of impacts on victims in Sect. 6. In Sect. 7, we introduce related works and summarise our conclusion in Sect. 8.

2. Amplification DDoS & AmpPot

2.1 Amplification DDoS Attacks

Amplification DDoS attack puts pressure on network resources by concentrating packets on the attack target, using servers that provide services such as NTP and Memcached on the Internet as reflectors. In particular, amplification DDoS attack exploits a server that meets the following two properties:

Amplification. This is the property that the length of the response packet from the server is larger than the length of the request packet sent to the server. By exploiting this, it is possible for attackers to amplify the attack packet and generate a large amount of communication.

Reflection. The reflection effect is that the server sends a response packet without verification, even though it receives a packet with a spoofed source IP address. This property exists in protocols that use UDP at the transport layer. An attacker can reflect a response packet to any IP address by exploiting the property. Due to this property, servers that are abused in amplification attacks are called reflectors.

Next, we explain the attack procedure of amplification DDoS attacks. As a preparation, an attacker scans the Internet for reflectors that satisfy the above properties. When conducting an actual DoS attack, the attacker sends a large number of request packets to the reflectors with the source IP address spoofed as the attack target. Then, the response packets are amplified at the reflectors and get concentrated on the attack target.

2.2 AmpPot

The Amplification DDoS honeypot (AmpPot) is a decoy server that mimics a reflector. Attacks can be observed when an attacker abuses it as a reflector. The information obtained by the AmpPot includes IP addresses of the attack target, communication protocols used for the attack, payload of the request packets, and amount of request packets sent to the AmpPot.

3. Method

In this section, we explain a method to measure the impact of amplification DDoS attacks observed by AmpPots. The procedure is divided into four steps, which are shown below. Fig 1 shows an overview of the system.

(1) Attacks are detected in real time by AmpPot.

(2) We send measurement packets to the attack destination IP address immediately after the attack is detected by AmpPot and measure the responses sent from the victim under the attack.

(3) After a sufficient time has passed and the attack is terminated, we send measurement packets to the above-mentioned IP address and measure the re-

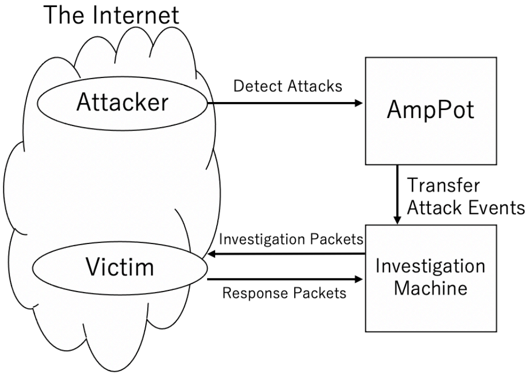


Fig. 1 Overview of the measurement system

sponses from the victim in a normal situation.

(4) We evaluate the impact of the amplification DDoS attack by comparing the responses measured when the victim is under attack and in a normal situation.

Next, we explain the details of each procedure.

3.1 Attack detection and sending measurement packets

We use AmpPots to detect attacks. In this paper, we consider an arrival of 100 or more consecutive packets with intervals no more than 60 seconds as an attack event, by following the literature [3]. Next, the attack information from the AmpPot is transferred to the investigation machine. We send ICMP ping packets to those targets. We send one packet per target at 1 minute intervals and continue this for 90 minutes for each target. Immediately after the probe starts, we check if the target IP address has corresponding domains using a passive DNS historical database. In addition, we resolve the domains at 5 minutes intervals to confirm if the domains are still resolved to the target's IP address. Thirdly, after 24 hours have passed since the detection of the attack, we send measurement packets to the above-mentioned IP address and measure the responses from the victim in a normal situation. The probe rate is the same as the measurement when under attacks. Simultaneously, we check if the IP address and the associated domain name have not changed during and after the attacks to make sure that we are measuring the same host. We do not consider load balancing (Anycast, CDN, DNS round robin, etc.) of the measurement target.

3.2 Calculating attack impact

For calculating attack impact, we focus on the attack targets whose IP address and corresponding domain names have not changed during and after the detected attacks. In this paper, the impact of an attack is defined using the response rate below:

$$ResponseRate = \frac{the\ number\ of\ responses}{the\ number\ of\ scan\ packets} \quad (1)$$

We calculate and compare this response rate for each IP address during and after an attack. We define the impact rate of an attack as follows, where RR_{attack} is the response rate during an attack and RR_{normal} is the response rate during normal times (after attack).

$$ImpactRate = \frac{RR_{normal} - RR_{attack}}{RR_{normal}} \quad (2)$$

If RR_{normal} is greater than RR_{attack}, Impact Rate becomes positive value. In contrast, if RR_{attack} is greater than RR_{normal} , Impact Rate becomes negative value. If Impact Rate is zero, it means there is no impact on victim by the attack. In contrast, the higher the value, the more availability of the target is lost. We focus on targets that provide a stable response at normal times as it is difficult to determine whether or not a target that is unstable at normal times is impaired by the amplification DDoS attack. We consider targets with response rate of more than 0.9 in normal times as stable targets.

3.3 System Implementation

First, we used the AmpPots to detect attacks implemented by the following literature [3]. Secondly, we implemented investigation system on Ubuntu OS [8] using Python [9]. We used a scan tool ZMap [4] to send the probe packets. We limited the amount of packets sent from our system to 100 pps. In addition, we used tcpdump [5] to record packets. We utilized DNSDB [7], a passive DNS historical database, to check if the target IP address has corresponding domains. Moreover, we used nslookup[10] and a public DNS server (8.8.8.8) provided by Google[11] to investigate the correspondence between domain name and IP address at the time of measurement.

4. Data Set

This section describes data set used in this paper. Our honeypot is installed in addresses managed by multiple ISPs in Japan. We have operated 11 proxied honeypot (protocol compliant honeypot) and 8 agnostic honeypot (protocol non-compliant) that responds back to any requests on any UDP ports with random strings of size 8K bytes. Currently, proxied honeypot supported seven types of protocols: QotD, CharGen, DNS, NTP, SNMP, SSDP, and Memcached. If AmpPot receives a large number of request packets, AmpPot reduces the amount of response

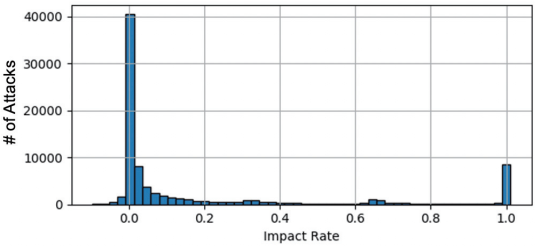


Fig. 2 Impact Rate Distribution

packets sufficiently to the same target so that it does not participate in the attack. We observed 1,619,345 attacks during the 119 days for which we sent measurement probes. The observation periods are as follows, April 1st to May 31st, Jun 6th to 30th, October 7th to 28th, November 15th to 30th ,2019 (119 days). Of these, we extracted 140,750 cases in which victims' IP addresses and corresponding domains were unchanged before and after the attacks. Finally, we checked victims' communication stability at normal times. Then we selected 85,890 cases whose response rate is above 0.9 at normal time for further analysis.

5. Result

5.1 Overview of the measured impacts

We show the distribution of the impact rate of the observed 85,890 attacks in the Figure 2. The x-axis is the impact rate of amplification DDoS attacks and the y-axis is the number of attacks. Peaks are formed around impact rate of 0 and 1.0. About 60% of the attacks have the impact rate between -0.05 and 0.05, and 10% of them are around the impact rate of 1.0, which means the targets were unable to respond at all. In the next section, we analyze what factors of attacks and targets are related to the attack impact.

5.2 Analysis of Impact Rate

We analyze Impact Rate from perspectives of victims' and attacks' characteristics. The characteristics of the targets include target's country, AS, the number of name resolutions for the target IP addresses in DNSDB, and type of hosting services (dedicated or shared). We obtained the number of unique second level domains associated with the victim's IP address to determine if the victim was operated by dedicated or shared hosting service. In this study, the victim with one associated domain is treated as operated by dedicated hosting service. The characteristics of attacks include average packet per seconds (pps) received by the AmpPots and

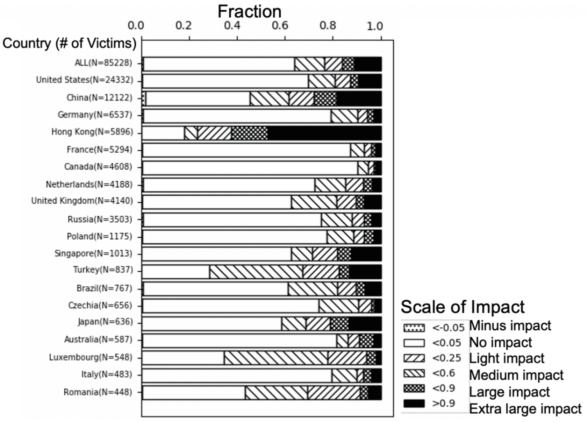


Fig. 3 Scale of Impact by Country

attack duration. In the following, we analyze these characteristics and the impact rates by categorizing them into 6 levels: Minus impact (less than -0.05), No impact (-0.05 or more and less than 0.05), Light impact (0.05 or more and less than 0.25), Medium impact (0.25 or more and less than 0.6), Large impact (0.6 or more and less than 0.9), Extra large impact (0.9 or more).

5.2.1 Scale of Impact by Country and AS

First, Impact Rate by country is shown in the Figure 3. The x-axis is fraction of attacks and the y-axis is target's country. In addition, countries are excerpted and arranged in descending order of number of attacks. The darker the color mapped to the bar graph indicates the greater the impact rate. We found that victims in Hong Kong and China have been affected with large impacts, followed by Turkey and Luxembourg. In particular, about half of the attacks on Hong Kong have extra large impact. On the other hand, we can see that 80 to 90% of the targets in France and Canada were not affected by the attack. The difference among countries may come from communication infrastructures, communication bandwidth of the telecommunications carriers, and deployment of DDoS countermeasures.

Second, the impact rate by AS is shown in the Figure 4. The x-axis is fraction of attacks, and the y-axis is AS of the targets. In addition, ASes are excerpted and listed in descending order of number of attacks. The impact rates are significantly different among ASes. In particular, Alibaba Technology and PEG TECH INC are significantly affected by the attacks while 90% or more targets in Cloudflare, OVH SAS, Google or i3D net are not affected. In particular, only a few percent of the targets in Cloudflare are affected by the attacks. As mentioned in the previous section, impact rate could differ due to the communication bandwidth of each AS and ASes' service

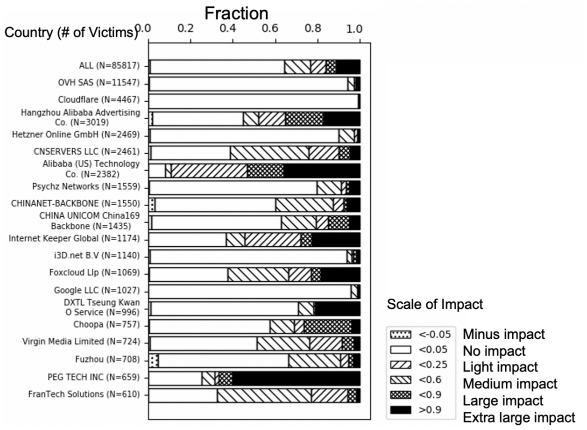


Fig. 4 Scale of Impact by AS

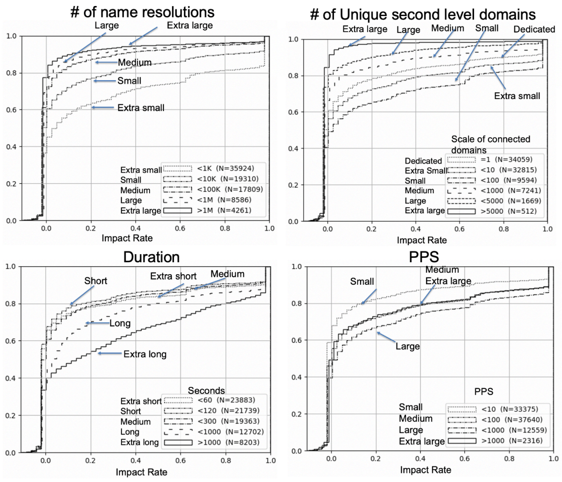


Fig. 5 Impact CDF by scale of victims' and attacks' characteristics

quality and policy against DoS attacks. One example we found is that Alibaba has a policy of dropping entire traffic to the targets when they are under attack, which results in a significant impact rate. [12].

5.2.2 Relationship between impact and characteristics of targets and attacks

In this section, we analyze the relationship between the impact and characteristics of the targets and the attacks. In particular, we look at the number of name resolutions and the number of domains resolved to the victim IP address as characteristic of victims. We also analyze the attack duration and pps (packet per seconds) as characteristic of attacks.

We show impact CDF by scale of victims' and attacks' characteristics in the Figure 5. First, we explain relationship between impact and the number of name resolutions. The number of name resolutions is from the DNSDB[7] database. We categorize the number of name resolutions into 5 levels and analyze their relationship with the attack impact. The levels are as follows: Extra small (less than 1K), Small (1K or more and less than 10K), Medium (10K or more and less than 100K), Large (100K or more and less than 1M), Extra large (1M or more). The CDF of the impact of attacks by the scale of the number of name resolutions is shown in the graph on the upper left of the Figure 5. The x-axis is impact rate of attacks and y-axis is the fraction of attacks. The scale of the number of name resolutions is distinguished by the plot type. It can be seen that the impact tends to be smaller as the number of domain name resolutions increases. In particular, we confirmed that the impact of Large and Extra large was almost 0 for about 80% of attacks. On the other hand, about 20% of attacks with impacts of around 1.0 were confirmed for Extra small. Assuming that a large number of name resolutions indicates the high popularity of the targets, we can say that popular sites are less affected, presumably because they could afford deployment of costly DoS protection.

Second, we analyze how the number of domains resolved to the target IP address is related to the attack impact. Again, we obtain the number of associated domains from DNSDB. The number of associated domains to the victim IP address is categorized into 6 stages: Dedicated (1 domain), Extra small (2 or more and less than 10), Small (10 or more and less than 100), Medium (100 or more and less than 1K), Large (1K or more and less than 5K), Extra large (5K or more). The CDF of the impact of attacks by scale of the number of connected domains is shown in graph on the upper right of the Figure 5. The x-axis is the impact rate of attacks and the fraction of attacks.

Overall, we can see the trend that targets associated to a large number of domains are less affected by attacks. If services of multiple domains are deployed on one server, all the services running on the server will stop at the same time as the server goes down, leading to a large economic and credit loss. However, this trend becomes opposite as the number of associated domains decrease. Namely, dedicated hosting is less affected than Small or Extra small categories. We have no clear explanation on this trend. One possible reason is that dedicated hosting can also serve for selected important services, such as financial and e-commerce services that should also be well protected.

Third, we analyze the relationship between attack duration and the impact of attacks. The duration of the attack is also categorized into 5 stages: Extra short (less than 60 seconds), Short (60 or more and less than

120 seconds), Medium (120 or more and less than 300 seconds), Long (300 or more and less than 1000 seconds), Extra long (1000 seconds or more). Next, the CDF of the impact of attacks by attack duration is shown in the graph on the bottom left of the Figure 5. One can see that the longer the attack continues, the greater the impact becomes. The trend is extreme for Extra long attacks whose impacts are obviously larger than other categories.

Finally, we analyze the relationship between average pps of the attacks observed by AmpPot and the impact of attacks. The pps of the attack was also categorized into 4 stages: Small (less than 10 pps), Medium (10 or more and less than 100 pps), Large (100 or more and less than 1000 pps), Extra large (1000 pps or more). The CDF of the impact of attacks by attack average pps is shown in graph on the bottom right of the the Figure 5. It is intuitive that the impact of small pps attacks is small. On the other hand, it is counter-intuitive that the impact of Extra large pps attack does not have the biggest impact. One possible explanation is that such extra large pps attack could actually clog the reflectors and the attacks become less effective although we need further investigation to support this explanation.

6. Just-in-time detection of serious attacks

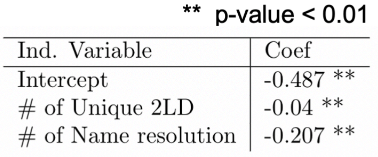
In this section, we propose a model for determining whether the target will be seriously affected by an observed attack just in time of its detection.

6.1 Classification model

In order to classify if the detected attack would significantly affect the target or not, we could only use the features of attacks and victims that can be measured at the time of attack detection. We use the following features as independent variables: the number of unique second level domains, the number of name resolutions, abused protocols, Countries, and ASes, where abused protocols, countries and ASes are category variables and the number of unique second level domains and the number of name resolutions are quantitative variables.

The impact rate is used as the dependent variable, where the attacks with the impact of more than 0.1 are treated as serious attacks, and the attacks with the impact rate of less than 0.1 were treated as non-serious attacks. We assign serious attacks to 1 and non-serious attacks to 0.

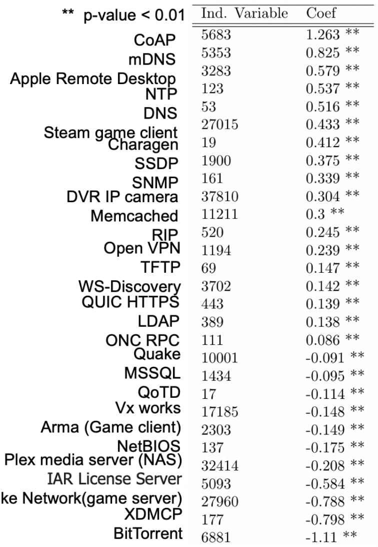
We use natural logarithm of these quantitative variables to calculate the model.



** p-value < 0.01

Ind. Variable	Coef
Intercept	-0.487 **
# of Unique 2LD	-0.04 **
# of Name resolution	-0.207 **

Table. 1 Coefficients of Unique 2LD and name resolutions



** p-value < 0.01

	Ind. Variable	Coef
CoAP	5683	1.263 **
mDNS	5353	0.825 **
Apple Remote Desktop	3283	0.579 **
NTP	123	0.537 **
DNS	53	0.516 **
Steam game client	27015	0.433 **
Charagen	19	0.412 **
SSDP	1900	0.375 **
SNMP	161	0.339 **
DVR IP camera	37810	0.304 **
Memcached	11211	0.3 **
RIP	520	0.245 **
Open VPN	1194	0.239 **
TFTP	69	0.147 **
WS-Discovery	3702	0.142 **
QUIC HTTPS	443	0.139 **
LDAP	389	0.138 **
ONC RPC	111	0.086 **
Quake	10001	-0.091 **
MSSQL	1434	-0.095 **
QoTD	17	-0.114 **
Vx works	17185	-0.148 **
Arma (Game client)	2303	-0.149 **
NetBIOS	137	-0.175 **
Plex media server (NAS)	32414	-0.208 **
IAR License Server	5093	-0.584 **
Quake Network(game server)	27960	-0.788 **
XDMCP	177	-0.798 **
BitTorrent	6881	-1.11 **

Table. 2 Coefficients of abused protocols

We utilize Logistic Regression [14] of Scikit-learn [13], a machine learning library of python, to build the model. In this model, we used L2 penalty for the regularization of logistic regression and set C = 1.0. The ratio of training data to test data is 0.9: 0.1. In training data, the variance of rare categorical variables can be zero. Therefore we treat these categories as ” other” .

6.1.1 Coefficients of Logistic Regression

Logistic regression coefficients of the number of unique 2LD and the number of name resolutions are shown in the Table 1. The coefficient of the number of name resolutions is a negative value, indicating that the impact decreases as the number of name resolutions increases. On the other hand, the coefficient of the number of unique second level domains is close to 0, which shows that the dependent variable cannot be explained well.

Next, logistic regression coefficients of abused protocols are shown in the Table 2. This table shows data with a p-value of less than 0.01. One can see that the coefficients differed considerably depending on the protocol. The protocols known to be particularly well abused in amplification DDoS: mDNS, NTP, DNS, Charagen, SSDP, SNMP, etc. have large coefficients. We confirm that the coefficients of the CoAP and DVR IP camera protocols implemented in IoT devices are also high.

Ind. Variable	Coef
Thailand	1.644 **
Republic of Moldova	1.469 **
Turkey	1.1 **
Malaysia	1.092 **
Vietnam	0.909 **
Indonesia	0.891 **
Hong Kong	0.796 **
Iran	0.766 **
China	0.742 **
Romania	0.656 **
United Kingdom	0.536 **
Taiwan	0.492 **
Singapore	0.454 **
Argentina	0.442 **
India	0.422 **
Bulgaria	0.405 **
Republic of Lithuania	0.387 **
Spain	0.345 **
Belgium	0.315 **
Israel	0.292 **
Sweden	0.276 **
Portugal	0.224 **
Japan	0.154 **
Brazil	0.147 **
South Africa	0.127 **
Ukraine	0.116 **
United States	0.097 **
Canada	0.019 **
Germany	-0.057 **
Poland	-0.081 **
France	-0.105 **
Serbia	-0.16 **
Italy	-0.196 **
Australia	-0.215 **
Czechia	-0.241 **
Netherlands	-0.304 **
Luxembourg	-0.367 **
Finland	-0.414 **
Russia	-0.421 **
Norway	-0.433 **
South Korea	-0.484 **
Belize	-0.617 **
Switzerland	-0.662 **
Ireland	-0.81 **
Denmark	-0.81 **
Belarus	-0.855 **
Austria	-0.891 **
Kazakhstan	-1.126 **
Pakistan	-1.158 **
Hungary	-1.375 **
Latvia	-1.836 **

Table. 3 Coefficients of Country

Ind. Variable	Coef
Anchnet Asia Limited	3.737 **
HENGDA NETWORK LIMITED	3.634 **
Hostinger International Limited	3.512 **
LinkChina Telecom Global Limited.	3.457 **
Cloudie Limited	3.432 **
Sun Network (Hong Kong) Limited - HongKong Backbone	3.291 **
Dimension Network & Communication Limited	3.163 **
ICIDC NETWORK	3.151 **
Cloud Iv Limited	3.108 **
DataClub S.A.	3.105 **
NETSEC	3.087 **
POWER LINE DATACENTER	2.993 **
Gigabit Hosting Sdn Bhd	2.84 **
Itace International Limited	2.727 **
AS number for New World Telephone Ltd.	2.595 **
FEDERAL ONLINE GROUP LLC	2.57 **
SAKURA Internet Inc.	2.5 **
eSited Solutions	2.418 **
Shenzhen Katherine Heng Technology Information Co.	2.401 **
Alibaba (US) Technology Co.	2.38 **
VpsQuan L.L.C.	2.331 **
SonderCloud Limited	2.291 **
Dennis Rainer Warnholz trading as active-servers.com	2.251 **
EAGLE SKY CO LT	2.245 **
Accelerated IT Services & Consulting GmbH	2.233 **
IT7 Networks Inc	2.182 **
LEMON TELECOMMUNICATIONS LIMITED	2.13 **
SS-Net	2.121 **
Enzu Inc	1.749 **
Liteserver Holding B.V.	1.473 **
QuadraNet Enterprises LLC	1.462 **
Online S.a.s.	1.408 **
HostDime.com	1.289 **
PEG TECH INC	1.289 **
Akamai International B.V.	-1.752 **
Sucuri	-1.768 **
Valve Corporation	-1.815 **
i3D.net B.V	-1.852 **
Microsoft Corporation	-1.855 **
Hetzner Online GmbH	-1.857 **
Highwinds Network Group	-1.887 **
AT&T Mobility LLC	-1.887 **
Akamai Technologies	-1.891 **
UK-2 Limited	-1.932 **
Chernyshov Aleksandr Aleksandrovich	-1.939 **
Aruba S.p.A.	-1.949 **
Hurricane Electric LLC	-2.002 **
Iomart Cloud Services Limited	-2.019 **
DMIT Inc.	-2.056 **
Diqing	-2.058 **
Internap Corporation	-2.062 **
Nuclearfallout Enterprises	-2.105 **
Anson Network Limited	-2.151 **
Fastly	-2.159 **
Proxy Pipe	-2.213 **
OVH SAS	-2.344 **
Anpple Tech Enterprise	-2.365 **
Domain names registrar REG.RU	-2.519 **
Google LLC	-2.735 **
Alibaba.com Singapore E-Commerce Private Limited	-2.738 **
Guangdong Mobile Communication Co.Ltd.	-2.798 **
Foxcloud Communications Srl	-2.955 **
Hydra Communications Ltd	-2.962 **
Cloudflare	-3.572 **

Table. 4 Coefficients of AS

We show logistic regression coefficients of the countries excerpted in Table 3. We confirm that the coefficient varies greatly by country. Characteristically, the coefficients of countries in Asia are relatively large, and the coefficient of countries in Europe tends to be small. It is presumed that the factors that cause the difference in countries may be the quality of the communication infrastructure and telecommunications carriers owned by each country.

Finally, we show logistic regression coefficients of AS excerpted in the Table 4. We confirmed that the coefficient varies greatly by the AS. In addition, the absolute the coefficient is also large, suggesting that it is strongly related to the seriousness of the attack. The coefficient of Cloudflare, Google, OVH SAS, etc. are very small, confirming they have high resistance to DDoS. It is presumed that the reason why the coefficient varies by each AS is that the quality of infrastructure and DoS protection technology deployed by telecommunications carriers and cloud service providers are different.

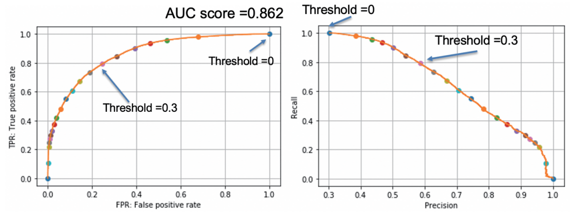


Fig. 6 ROC curve and Precision-Recall curve

		Predicted		
		Non affected	Affected	Total
Actual	Non affected	4480	1491	5971
	Affected	534	2019	2553
	Total	5014	3510	8524

Accuracy	0.762
Precision	0.575
Recall	0.790
F-measure	0.665

Table. 5 Confusion Matrix and Evaluation Scores

6.2 Model evaluation

Figure 6 shows the ROC curve (left graph) and Precision-Recall curve (right graph) obtained from the model constructed in the previous section. The AUC score of the model is 0.862, when TP is 0.8, FP is 0.25 from the figure. We also confirmed when recall 0.8, precision is 0.58.

Next, for this model, we show the confusion matrix when the threshold for judgment by logistic regression is set to 0.3. Table 5 shows the confusion matrix when the test data is applied to the model. Comparing the Precision-Recall curve and Confusion Matrix of test data, there is not much difference in precision and recall scores, so it seems that overfitting hardly occurred. As a future work, we will examine the factors for improving the accuracy of the model.

7. Related Work

Many investigations have been conducted to analyze amplification DDoS attacks. Noroozian et al. profiled victimization patterns of DDoS attacks through data collected by AmpPots [15]. Jonker et al. analyzed four independent data sources and demonstrated a macroscopic characterization of DoS ecosystem[16]. Welzel et al. measured the effect of L7 DDoS attacks launched by DDoS botnets[17]. Zebari et al. analyzed the impact of HTTP and SYN flood DDoS attacks on web servers from the perspectives of response time, error rate, and CPU usage[20]. Sassani et al. analyzed the impact of NTP amplification DDoS attacks and investigated attack mitigation implemented on routers[21]. It has been reported that most attack packets can be blocked by blocking reflectors with high attack capability [18]. Shadowserver continuously check the activity status of the reflectors in the wild [19]. A method for estimating the amount of DoS packets that actually reach the target using a darknet has been reported[22].

8. Conclusion

We measured and investigated actual situation of the impact on the availability of victims using AmpPot. We also examined the relationship between the impact of attacks and the characteristics of attacks and targets. Furthermore, we examined a model to determine whether the attack has a serious effect on the target immediately after detecting the attack. We will also continue measurement to collect more data and consider developing a highly accurate alert that can detect serious attacks.

Acknowledgement

This work has been enabled through the support of Web-based Attack Response with Practical and Deployable Research InitiatiVE, National Institute of Information and Communications Technology.

References

[1] ”AWS Shield” ,(https://aws-shield-tlr.s3.amazonaws.com/2020-Q1_AWS_Shield_TLR.pdf),last visited 2021/01/18

[2] Lukas Krämer, Johannes Krupp, Daisuke Makita, Tomomi, Nishizoe, Takashi Koide, Katsunari Yoshioka, Christian Rossow, “Amppot: Monitoring and Defending Against Amplification DDoS Attacks,” RAID 2015,2015.

[3] Daisuke Makita, Tomomi Nishizoe, Katsunari Yoshioka, Tsutomu Matsumoto, Daisuke Inoue, Koji Nakao, ”DRDoS Attack Alert System for Early Incident Response,” Information Processing Society of Japan, Vol.57 No.9 1974–1985 (2016)

[4] Zmap, “ZMap:The internet Scanner”,(https://github.com/zmap/zmap) last visited 2020/12/10

[5] Tcpdump, ”TCPDUMP & LIBPCAP”,(http://www.tcpdump.org/#latest-release) last visited 2020/01/13

[6] farsight security,” FARSIGHT SECURITY”,(https://www.farsightsecurity.com/) last visited 2020/01/13

[7] DNSDB, “Welcome to DNSDB”,(https://www.dnsdb.info/) last visited 2020/12/10

[8] Ubuntu,”Enterprise Open Source and Linux”,(https://ubuntu.com/), last visited 2020/02/10

[9] Python, ”Welcome to Python.org” (https://www.python.org/),last visited 2020/02/10

[10] nslookup,”Linux man page”,(https://linux.die.net/man/1/nslookup) last visited 2021/01/17

[11] google public dns,”Google Public DNS” (https://developers.google.com/speed/public-dns/) last visited 2021/01/13

[12] ”Blackhole filtering policy of Alibaba Cloud”,(https://www.alibabacloud.com/help/doc-detail/40032.html),last visited 2021/01/18

[13] ”Scikit-learn”,(https://scikit-learn.org/stable/index.html),last visited 2021/01/20

[14] ”sklearn.linear_model.LogisticRegression”,(https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html),last visited 2021/1/20

[15] Arman Noroozian, Maciej Korczynski, Carlos Hernandez Ganan, Daisuke Makita, Katsunari Yoshioka, Michel van Eeten, “Who Gets the Boot? Analyzing Victimization by DDoS-as-a-Service,” Proceedings. Research in Attacks, Intrusions, and Defenses , Lecture Notes in Computer Science, RAID 2016, 2016.

[16] Mattijs Jonker, Alistair King,Johannes Krupp, Christian Rossow, Anna Sperotto, Alberto Dainotti, ” Millions of targets under attack: a macroscopic characterization of the DoS ecosystem,” Proceedings of the 2017 Internet Measurement Conference,Pages 100-113,IMC ’17,2017.

[17] Arne Welzel, Christian Rossow, Herbert Bos, ” On measuring the impact of DDoS botnets,” Proceedings of the 7th European Workshop on System Security,Article No. 3,EuroSec ’14, 2014.

[18] Jumpei Urakawa, Kosuke Murakami, Akira Yamada, Ayumu Kubota, ”Empirical Analysis of Reflectors in DRDoS attacks based on Large-Scale Network Data ,” IEICE technical report ,vol. 117, no. 481, ICSS2017-82, pp. 199-204

[19] Shadowserver, “The shadowserver foundation “,(https://www.shadowserver.org/) last visited 2021/01/10

[20] Rizgar R. Zebari, Subhi R. M. Zeebaree, Karwan Jacksi,”Impact Analysis of HTTP and SYN Flood DDoS Attacks on Apache 2 and IIS 10.0 Web Servers,” International Conference on Advanced Science and Engineering, pp.156-161, 2018.

[21] Bahman A. Sassani, Charly Abarro, Ivan Pitton, Craig Young, Farhad Mehdipour, ”Analysis of NTP DRDoS attacks’ performance effects and mitigation techniques,” 2016 14th Annual Conference on Privacy, Security and Trust, pp.421-427,2016

[22] Norbert Blenn, Vincent Ghiëtte, Christian Doerr, ”Quantifying the Spectrum of Denial-of-Service Attacks through Internet Backscatter,” Proceedings of the 12th International Conference on Availability, Reliability and Security, Pages1-10,ARES 2017

[23] Arne Welzel, Christian Rossow, Herbert Bos, ”On Measuring the Impact of DDoS Botnets,” EuroSec ’14: Proceedings of the Seventh European Workshop on System Security, No. 3, Pages 1-6.