Characterizing infrastructure of DDoS attacks based on DDoSDB fingerprints

Matthijs Vos
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands
m.vos@student.utwente.nl

ABSTRACT

Distributed Denial of Service (DDoS) attacks are a big problem in the current digital landscape. Many research is conducted on various sub parts of DDoS. However, little is known about the infrastructure behind the attacks. It can be of interest to know how attacker choose their infrastructure. It is possible that they choose their attacking nodes very specific based on some characteristics. This paper aims to characterize the infrastructure of a DDoS attack to gain more insights in the infrastructures and how attackers choose their attacking nodes. The paper will focus on seven different attack types and will analyze their infrastructure. We will show that DNS recursion is still enabled on a lot of DNS resolvers, that the non-RFC-compliant implementation of Chargen in Windows is widely misused and that small ISPs are the most common in DDoS attacks attacking nodes.

Keywords

DDoS, Infrastructure, Shodan

1. INTRODUCTION

Distributed Denial of Service (DDoS) attacks are a big problem in the current digital landscape. Sending multiple packets of data to a server from different devices can cause this server to be unreachable for legitimate visitors or users. This can cause a lot of impact, and potentially have economical consequences. Since a few years Booters [20] are introduced which makes it even easier for attackers without technical knowledge to attack a target. A Booter is a website on which an attacker can simply buy an attack on a target and have no need for technical background or a infrastructure to use for the DDoS attack. This made the world of DDoS so easy that even schools got attacked by their own students [20].

Attacks often make use of hosts which are not belonging to them. This makes them less traceable and can also increase the bandwidth of the attack, with up to 1.3 TB/s in the biggest attack seen at the moment [10]. How attackers choose this infrastructure is very interesting since this can help in understanding their behaviour and also expecting which devices are potential attacking nodes.

31^{th} *Twente Student Conference on IT* July 5^{th}, 2019, Enschede, The Netherlands.

The goal of the research is *to gain insights in the infrastructures of DDoS attacks.* To reach this goal the following main research question is defined; **What are the main characteristics of the infrastructure of a DDoS attack?** This general approach will be researched in different steps using the following sub research questions:

1. What kind of attack nodes exists?
2. What kind of data is available in the data sources?
3. What are the characteristics of attacking nodes for a specific attack type?
4. Are nodes used in attacks of multiple types? If yes, what are their characteristics?

This paper is organized as followed. In section 2 the needed background knowledge is explained, in section 3 the used data will be discussed. In section 4 the used methodology will be introduced, in section 5 the results will be shown, finally in 6 and 7 the discussion and conclusions which can be drawn from the results will be showed.

2. BACKGROUND

This section will provide background on the topic. The research questions 1 and 2 will be answered by a literature study. First the literature is studied on the existing node types, in the second part the available datasources will be analyzed.

2.1 Related work

Most research in the field of DDoS focuses on the mitigation of the problem [7, 13, 4]. However, not much is known about the choose of infrastructure by the attackers. The new world of Booters is researched in paper [20], this already gives a insight in the infrastructure of Booters. A theoretical and mathematical approach to find the attack source is researched by [5]. The identification of the originating attacker for amplification attacks by using honeypots is researched in [12]. The monitoring of amplification attacks is studied by AmpPot [11]. Mirkovic gave a general overview of the DDoS field and did a comparison between different defence systems in [14].

This paper distinguish itself from the other research by investigating the middle layer, the reflectors and botnets. The mitigation research mainly focus on the last layer, the target. The papers [5] and [12] mainly focus on attributing the original attacker, the first layer. However little/none research is done on the middle layer. This paper will focus on that part of the DDoS infrastructure and how specific characteristics are (mis)used.

2.2 Existing attacking nodes

In order to specify the characteristics of the infrastructure it is useful to state the possible attacking types and their used nodes. In general all DDoS attacks can be categorized by two features. The first feature is *semantic* (*vulnerability*) vs *brute-force* (*flooding*) attacks [14]. The semantic attack misuses a feature or implementation bug in a specific protocol or application on the target host. In the brute-force a high amount of traffic that is almost the same as legitimate traffic is used. This way the victims network gets overloaded and the service is not longer reachable by legitimate traffic. The second feature is *reflected* vs *direct* attacks. In a reflection attack a system is used to hide the identity of the attacker and in most cases also amplify the attack. With a direct attack a large number of systems owned by the attacker (e.g. botnet) will send request direct to the target, sometimes combined with spoofed IP addresses to hide the identity.

Almost all reflection attacks are brute-force attacks, since they do not misuse a protocol on the target node but on the reflection nodes. The attacker will use two techniques to launch the attack [11]. First some kind of UDP protocol which runs on an open node is abused. An example is a DNS resolver to which they send a query. They will try to send a query which results in a much larger response, which is the amplification part of the attack. The second technique is IP address spoofing, they will identify themselves as the victim. This will result in the abused node to respond to the victim instead of the attacker. There is a RFC which will disallow this behaviour [6], however in order to completely block spoofing all AS need to implement this. According to CAIDA currently around 60% is blocking IP spoofing [1], which means that still around 40% of the internet allows spoofing.

2.3 Available data

Two data sources will be used to gain insights in the infrastructure. The first data source is DDoSDB[2]. DDoSDB is a project to make real DDoS attack data available to everyone. It can be used for improving mitigation and detection mechanisms, but also for comparison of attacks or even legal actions. It contains around 850 different fingerprints of DDoS attacks. Those fingerprints contains anonymous data of attacks, IP addresses, ports, protocol and date. For this paper mainly the IP addresses are relevant.

Only those IPs do not tell much about the characteristics of the infrastructure, so we need to enrich this data, This data will be retrieved from Shodan[3]. This is a web scanning platform which constantly scans the internet. By querying Shodan you can retrieve different (historical) metrics about the source. This data includes open ports, running software, geographical location and vulnerabilities. This data from Shodan will be used to create a characterization of the attacking nodes. The main reason to use Shodan is that is the common tool to use in most research (for example [3, 15, 21]).

3. DATA SET

We need to define which data of DDoSDB we will research, since not all the data can be researched. It is limited to the top 4 reflected (DNS, NTP, Chargen, SSDP) and top 2 directed (TCP, UDP) attacks of DDoSDB. This top is based on the amount of occurrences in the database, not their size. Since the TCP and UDP attacks contains too many IPs this is limited to the last 20 performed attacks. ICMP is a side effect of other types of attacks, so this type is left out.

In table 1 all those different attack types are listed, combined with their properties researched in 'Existing Nodes'.

Name	Semantic vs Flooding	Reflected vs direct	Used node types	% of DDoSDB
DNS [1, 19],	flooding	reflected	DNS servers	8.7%
NTP [1, 19]	flooding	reflected	NTP servers with monlist command activated	3.3%
Chargen [1, 19]	flooding	reflected	All kind of devices, with chargen (port 19) enabled	2.8%
SSDP [1, 19]	flooding	reflected	UPnP devices (Printers, ipcameras)	1.7%
TCP [2]	semantic	direct	Systems owned by attacker (e.g. botnet)	31.5%
UDP [8]	flooding	direct	Systems owned by attacker (e.g. botnet)	18.3%

Table 1: Different attack types and their used nodes

4. METHODOLOGY

The used methodology can be split up into four separate steps. Retrieve data from DDoSDB, Retrieve data from Shodan.io, Select correct data, Characterize. The following paragraphs will explain the different steps. This approach is visualized in figure 1.

In the first step we retrieve all the fingerprints from DDoSDB, those will be stored into a MongoDB so that we can easily query them for the following steps. The second step takes a fingerprint (or multiple) and retrieve the historical data from Shodan.io through their API. This data is also stored into the MongoDB.

When this data is retrieved we need to select the correct IP and Shodan data. This is different for RQ 3 and RQ 4. For RQ 3 we select all the IP addresses from a specific type of attack (DNS, Chargen, TCP). For RQ 4 we select IP addresses that are used in multiple attack types.

[1]https://spoofer.caida.org/summary.php
[2]https://ddosdb.org
[3]https://Shodan.io

Shodan contains a lot of so called 'services' data. This is information about the running services. All the historical data of Shodan is scraped. However, we want to have the data closest to the attack date. For this reason all services from 1 year before till 1 year after the attack will be queried. Of those services the closest entry will be used for the analysis.

The last step to do is to find characterizations. The first that is done here is checking the distributions of specific attributes on the running services. For example the open port, the running software or the vulnerabilities that are found. We also used the 'Autonomous System Taxonomy Repository' from CAIDA[4]. This repository contains data with classifications of Autonomous Systems. By using this we can identify which kind of Autonomous Systems where used. The labels of the categories in the graphs could be interpreted as followed, t1 is a large ISP, t2 is a small ISP, edu is a University, ix is an Internet Exchange, nic for Network Information Centers, comp for customers and unknown for no prediction or no data.

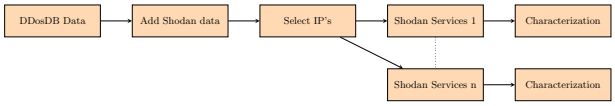


Figure 1: Methodology overview

5. RESULTS

In this part the results for the used methodology are explained. First the research into the characterization for specific attack types is performed and explained. In the second subsection the nodes that are used in multiple attack types are researched.

5.1 Specific attack types

In this subsection we will focus on the characterization of a specific attack type. For this the data is split into 6 groups, based on their used protocol. This part will answer research question 3: *What are the characteristics of attacking nodes for a specific attack type?*

5.1.1 DNS

We have analyzed 79 DNS DDoS attacks from DDoSDB. Those attacks contained in total 140664 distinct IP Addresses, of which 68% was available on Shodan. In total 11371 distinct IP have use-able services data (8% of all fingerprint IPs), all percentages are relative to this amount.

The first observation is that 83% of the nodes have port 53 accessible (Figure 2a), this is expected since we would expect to see DNS servers. This is also confirmed by the fact that the most used running services are the DNS-UDP and DNS-TCP (Figure 2c). However, 17% do not have port 53 accessible, which are 1927 nodes, of all those nodes 668 do have a DNS service on Shodan but no within the one year offset. Of 807 of those nodes the first time that it was added was after the attack (but before the one year offset), this could mean that the DNS server was already enabled during the attack. The other 452 nodes are unclear, they are on Shodan but they never have a DNS service enabled. So in general this fact is probably due to missing/wrong data on Shodan.

The port that is open on 55% of all the nodes is port 80 and also the http service is running on those nodes. This means that those nodes do not only run a DNS server but also a HTTP server.

[4]https://www.caida.org/data/as-taxonomy/

When analyzing the 'data' field in each service item we found that 79% of the nodes have a 'Recursion: enabled' message. This means that most nodes have the DNS recursion setting turned on. When this behaviour is turned on and a DNS server does not have the answer to a specific query in its cache it will query it for the requester at another DNS resolver. When this is turned off the DNS server replies with a message telling the requester that it does not know and that it should ask it at another resolver. When this is turned off and the specific query is not in the cache of the server the amplification effect is lost. This is the reason that attacks could like DNS servers with the recursion enabled. This behaviour is already researched and discouraged by ICANN [9]. In their report is stated that around 75% of all DNS servers have recursion enabled, which is very close to our outcome.

After applying the CAIDA AS classification to the DNS dataset it appeared that 44% of the attack nodes are located in a AS belonging to a small ISP (Figure 2b). However, in the dataset of CAIDA only 29% of the ISPs is a small ISP. This difference means that small ISPs are attractive for attacks or that vulnerable systems are mostly are placed within small ISPs.

5.1.2 NTP

We have analyzed 30 NTP DDoS attacks from DDoSDB. Those attacks contained in total 3739 distinct IP Addresses, of which 74% was available on Shodan. In total 1429 distinct IP have use-able services data (38% of all fingerprint IPs), all percentages are relative to this amount.

The expected port for NTP, port 123, is open on most nodes (Figure 3a). On 50% of the nodes port 80 is open, which means that also a HTTP server is running.

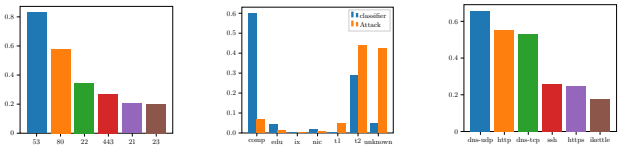
The distribution of used transport layers is different that with the other attack types. A relative high amount of UDP services is running (Figure 7). This could be expected since NTP is a UDP based protocol.

After applying the CAIDA AS classification to the NTP dataset it appeared that 54% of the attack nodes are located in a AS belonging to a small ISP (Figure 3c). Also 16% is located in a Large ISP, but in the dataset only a few AS are marked as large ISP. This difference means that small and large ISPs are attractive for attacks or that vulnerable systems are mostly are placed within small and large ISPs.

5.1.3 Chargen

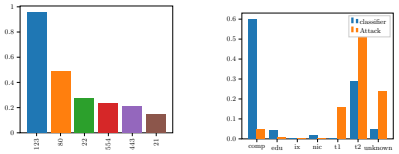
We have analyzed 25 Chargen DDoS attacks from DDoSDB. Those attacks contained in total 4487 distinct IP Addresses, of which 73% was available on Shodan. In total 312 distinct IP have use-able services data (7% of all fingerprint IPs), all percentages are relative to this amount.

Port 19 is expected to be open, since this port is needed for this attack (Figure 4a). However, most nodes also have port 17 (Quote of the Day Protocol [18]), 7 (Echo Protocol [17]) and 13 (Daytime Protocol [16]) available. All those protocols are part of the Windows Simple TCP/IP services, which would mean that most of the nodes are Windows machines. As shown by [20] Windows has a non-RFC-compliant implementation of Chargen. It sends a random size of 0-6956 bytes of data, and by the RFC it should only send at most 512 bytes. This means that the amplification factor of the Windows chargen implementations is 10 times bigger that a RFC-compliant implementation. This makes using Windows machines more powerful for Chargen attacks.



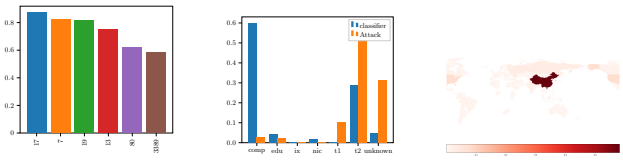
(a) Open ports (b) AS classification (c) Running services

Figure 2: DNS attack results



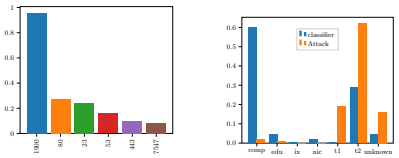
(a) Open ports (b) AS classification

Figure 3: NTP attack results



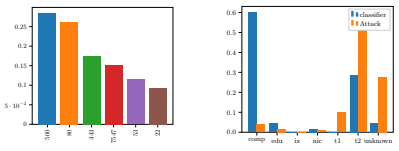
(a) Open ports (b) AS classification (c) Country distribution

Figure 4: Chargen attack results



(a) Open ports (b) AS classification

Figure 5: SSDP attack results



(a) Open ports (b) AS classification

Figure 6: TCP attack results

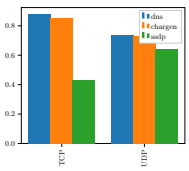


Figure 7: Transport between different attack types

After applying the CAIDA AS classification to the Chargen dataset it appeared that 53% of the attack nodes are located in a AS belonging to a small ISP (Figure 4b). As seen before this is a lot higher than the amount of AS classified as small ISP.

Most of the Chargen attacks appeared to be originated from China, 58% of all the requests are originated from nodes located there (Figure 4c). There is no clear reason for this, however there is a specific Windows version made for China. It could be the case that this is compromised or vulnerable. However, we did not prove that in this paper.

5.1.4 SSDP

We have analyzed 14 SSDP DDoS attacks from DDoSDB. Those attacks contained in total 151847 distinct IP Addresses, of which 68% was available on Shodan. In total 9063 distinct IP have use-able services data (6% of all fingerprint IPs), all percentages are relative to this amount.

The expected port for NTP, port 1900, is open on most nodes (Figure 5a). On 50% of the nodes port 80 is open, which means that also a HTTP server is running.

The distribution of used transport layers is different than with the other attack types. A relative low amount of TCP services is running (Figure 7). This could be the caused by the fact that SSDP is a UDP only protocol.

After analyzing the 'data' field in the Shodan data we found that 45% was running the IGD protocol or the MiniUPnP project (which is a implementation of IGD). This protocol is used to manage port forwarding on a router. However, this should only be exposed to the internal network, since there is no reason to manage this from an external network (at least not for the average user). That 45% of all attack nodes have this enabled and reachable from the external network could indicate that consumer routers have wrong default settings.

After applying the CAIDA AS classification to the SSDP dataset it appeared that 61% of the attack nodes are located in a AS belonging to a small ISP (Figure 5b). Also 19% belongs to a AS of a large ISP. As seen before this is a lot higher than the amount of AS classified as small and large ISPs.

5.1.5 TCP

We have analyzed 20 TCP DDoS attacks from DDoSDB. Those attacks contained in total 500108 distinct IP Addresses, of which 28% was available on Shodan. In total 128791 distinct IP have use-able services data (26% of all fingerprint IPs), all percentages are relative to this amount.

Since this is a directed attack there is not an expected port. There is not a common port that is open on almost every node, the port that is open the most often is port 500 and 80 but this is only 28% and 26% (Figure 6a).

The AS classification matches the previous outcomes. The small ISPs are over represented by 56% of all attacks using a small ISP (Figure 6b).

5.1.6 UDP

We have analyzed 20 TCP DDoS attacks from DDoSDB. Those attacks contained in total 242132 distinct IP Addresses, of which 19% was available on Shodan. In total 37281 distinct IP have use-able services data (15% of all fingerprint IPs), all percentages are relative to this amount.

Since this is a directed attack there is not an expected port. The same behaviour as with TCP is seen, the ports

	DNS	Chargen	SSDP	NTP	TCP	UDP
DNS		26	25	12	10	40
Chargen			0	0	72	968
SSDP				0	14	7
NTP					0	0
TCP						43
UDP						

Table 2: Amount of nodes used in two different attack types

that are open the most are 80 and 500 (27% each). The same counts for the AS classification, small ISPs are the largest (51%).

5.2 Multiple type nodes

We did an analysis on the nodes that are used in multiple attack types. In table 2 the amount of IPs used in both attack types are listed. Those are used to answer research question 4: *Are nodes used in attacks of multiple types? If yes, what are their characteristics?*

5.2.1 UDP/TCP combinations

The TCP and UDP attacks should be directed attacks. Which would mean that those machines should be compromised. The protocol which has the biggest intersection with TCP/UDP is Chargen. This could explain why the Chargen attacks have the simple TCP/IP services enabled (which are disabled by default). The machine was compromised first, the services turned on and then used for Chargen and UDP attacks.

5.2.2 DNS combinations

The combinations with DNS could indicate to home routers. Many consumer home routers do have a build in DNS server. Some of the IPs which are in both DNS and SSDP have a IGD (Internet Gateway Device) UPnP service running, which indicates to a home router. This was also a finding in the SSDP part, so this would support this idea. However, the groups are too small to create a general characterization.

6. DISCUSSION

During the analysis we found out that on most of the attack types only around 10% of data was available on Shodan. This was partly due to the fact that we used an offset of one year around the attack and partly because Shodan did not have data at all. This made the researched groups smaller, but still big enough to do research on. In future research the availability could be extended by using Censys.io as a second data source.

We used the CAIDA data to classify the used Autonomous Systems. We compared the relative amount of used AS in a specific group against the relative amount of AS in the classifier in that group. In this comparison we did not correct for the group size (for example there could be a lot of really small AS in specific group and only a few really big in the other). Information about actual usage within an AS is hard to find, however there is information about the amount of available IPs within an AS, for example at ipinfo.io. Within the limited time of this research there was no time to add this to the project but this could be done in further research.

In the last research question we investigated IP addresses that are used for multiple attack types. We found some interesting hypothesis there. However, those groups are quite small so we do not have enough confidence to make hard conclusion in that subpart.

7. CONCLUSION

In this paper we have shown that it is possible to find characteristics of a DDoS attack. By combining the data of Shodan and DDoSDB it is possible to identify the common characteristics of a specific protocol attack. Our main research question was **What are the main characteristics of the infrastructure of a DDoS attack?**. In all attacks we have seen that the small ISPs are over represented, which means that most hosts are located in such networks. Furthermore we have seen that misconfiguration or -implementation of protocols and devices have a big impact on the choose of attack hosts.

8. REFERENCES

[1] S. Arukonda and S. Sinha. The Innocent Perpetrators: Reflectors and Reflection Attacks. In *Advances in Computer Science : an International Journal*, volume 4, pages 94–98. ACSIJ, jan 2015.

[2] M. Bogdanoski, T. Shuminoski, and A. Risteski. Analysis of the SYN Flood DoS Attack. In *International journal of computer network and information security.*, pages 1–11. MECS Publisher, 2013.

[3] J. Bugeja, D. Jonsson, and A. Jacobsson. An Investigation of Vulnerabilities in Smart Connected Cameras. In *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 537–542. IEEE, mar 2018.

[4] V. T. Dang, T. T. Huong, N. H. Thanh, P. N. Nam, N. N. Thanh, and A. Marshall. SDN-Based SYN Proxy—A Solution to Enhance Performance of Attack Mitigation Under TCP SYN Flood. *The Computer Journal*, 62(4):518–534, apr 2019.

[5] O. Demir and B. Khan. Finding DDoS attack sources: Searchlight localization algorithm for network tomography. In *2011 7th International Wireless Communications and Mobile Computing Conference*, pages 418–423. IEEE, jul 2011.

[6] P. Ferguson and D. Senie. Network ingress filtering: Defeating denial of service attacks which employ ip source address spoofing. BCP 38, RFC Editor, May 2000.

[7] P. Gulihar and B. B. Gupta. Cooperative Mitigation of DDoS Attacks Using an Optimized Auction Scheme on Cache Servers. pages 401–412. 2019.

[8] U. Gurusamy, H. K, and M. MSK. Detection and mitigation of UDP flooding attack in a multicontroller software defined network using secure flow management model. *Concurrency and Computation: Practice and Experience*, page e5326, may 2019.

[9] ICANNSecurity and Stability Advisory Committee. SSAC Advisory SAC008DNS Distributed Denial of Service (DDoS) Attacks. Technical report, ICANN, 2006.

[10] S. Kottler. February 28th DDoS Incident Report, 2018.

[11] L. Krämer, J. Krupp, D. Makita, T. Nishizoe, T. Koide, K. Yoshioka, and C. Rossow. AmpPot: Monitoring and Defending Against Amplification DDoS Attacks. pages 615–636. Springer, Cham, 2015.

[12] J. Krupp, M. Backes, and C. Rossow. Identifying the Scan and Attack Infrastructures Behind Amplification DDoS Attacks. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security - CCS'16*, pages 1426–1437, New York, New York, USA, 2016. ACM Press.

[13] T. Lukaseder, K. StOlzle, S. Kleber, B. Erb, and F. Kargl. An SDN-based Approach For Defending Against Reflective DDoS Attacks. In *2018 IEEE 43rd Conference on Local Computer Networks (LCN)*, pages 299–302. IEEE, oct 2018.

[14] J. Mirkovic and P. Reiher. A taxonomy of DDoS attack and DDoS defense mechanisms. *ACM SIGCOMM Computer Communication Review*, 34(2):39, apr 2004.

[15] M. Patton, E. Gross, R. Chinn, S. Forbis, L. Walker, and H. Chen. Uninvited Connections: A Study of Vulnerable Devices on the Internet of Things (IoT). In *2014 IEEE Joint Intelligence and Security Informatics Conference*, pages 232–235. IEEE, sep 2014.

[16] J. Postel. Daytime protocol. STD 25, RFC Editor, May 1983.

[17] J. Postel. Echo protocol. STD 20, RFC Editor, May 1983.

[18] J. Postel. Quote of the day protocol. STD 23, RFC Editor, May 1983.

[19] C. Rossow. Amplification Hell: Revisiting Network Protocols for DDoS Abuse. 2014.

[20] J. J. Santanna, R. Van Rijswijk-Deij, R. Hofstede, A. Sperotto, M. Wierbosch, L. Z. Granville, and A. Pras. Booters - An analysis of DDoS-as-a-service attacks. In *Proceedings of the 2015 IFIP/IEEE International Symposium on Integrated Network Management, IM 2015*, pages 243–251. IEEE, may 2015.

[21] R. Williams, E. McMahon, S. Samtani, M. Patton, and H. Chen. Identifying vulnerabilities of consumer Internet of Things (IoT) devices: A scalable approach. In *2017 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 179–181. IEEE, jul 2017.