ISSN 1346-5597

National Institute of Informatics

NII Technical Report

Profiling Internet Scanners: Spatial and Temporal Structures

Johan Mazel, Romain Fontugne, Kensuke Fukuda

NII-2016-008E
Dec. 2016

URL: http://www.nii.ac.jp/ Tokyo 101-8430, JAPAN +81-3-4212-2000 FAX: +81-3-3556-1916

Profiling Internet Scanners: Spatial and Temporal Structures

Johan Mazel [1], Romain Fontugne [2], and Kensuke Fukuda [3]

[1] NII/JFLI, [2]IIJ, [3] NII/Sokendai

Abstract. A great deal of effort has been dedicated to the study of network scanning. Nonetheless, previous studies focused on simple characteristics such as the number of scanning IPs (also called scanners) or targets, but usually neglected scanner behavior. We analyze 15 years of backbone traffic and propose a method for profiling scanning IPs. Our analysis first details evolution of targeted services, mass-scanning tool usage and scanning pattern. Then, we propose a new method to classify scanning IPs' spatial and temporal structure into three profiles that reveal vastly different intent. In particular, we find that 33% of scanners repeatedly target the same set of hosts. If unsolicited, this behavior provides an early warning to administrators regarding the malicious intent of scanners. Finally, we study publicly documented scanners' activities and show that security research-related scanning IPs behave differently than non-documented scanners. We also show that only 39% of scanning entities follow online documentation best practices.

1 Introduction

Scanners send packets to numerous destinations and analyze corresponding answers, or lack thereof, to acquire knowledge on remote hosts or networks. Probing may be defensive, to acquire knowledge on one's own network, or offensive, to assess attack surface of a targeted network [6].

A great deal of attention has been dedicated to scan detection and analysis. Previous works identify scanners as hosts with a high rate of unsuccessful connection [15], or map existing services in a network and consider hosts reaching unavailable services as scanners [2]. For a more complete account of scan detection techniques, we refer the reader to [3]. Other works provide analysis of scans in several stub network datasets, for example, 12 years of IDS logs from June 1994 until December 2006 [1], or, data from the Dshield repository: one month in 2001 and three months in 2002 [22], and the first fifteen days of 2005 [21]. Also, some studies provide simple description of scans in backbone traffic [11, 19] and darknet traffic [5, 13]. To evade detection, scanning mechanisms have evolved from simple sequential probing of the IP space to more complex probing schemes [7, 10, 14, 17]. In this paper, we study recent scanning trends and profile scanning IP behaviors that reveal potentially malicious intent.

The contribution of this paper is three-fold. First, we expose scanning recent evolution and show that the use of Internet-wide scanning tools and random

probing pattern are increasing. Second, we present a scanner behavior profiling method. This method defines three profiles: scanners either contact unrelated small network prefixes (41% of scanners), or randomly probe the same prefix for a long duration (8% of scanners), or repeatedly scan the same set of hosts (33% of scanners). This last behavior provides clues to administrators regarding the malicious intent of scanners that may attack the probed network. Third, we analyze security researchers' scanning activities (which represent 0.1% of all scanners) and show that: 1) their behavior is distinct from the others in terms of both targeted services and behavior profiles, and, 2) most identified scanning entities do not completely follow online documentation best practices.

2 Background

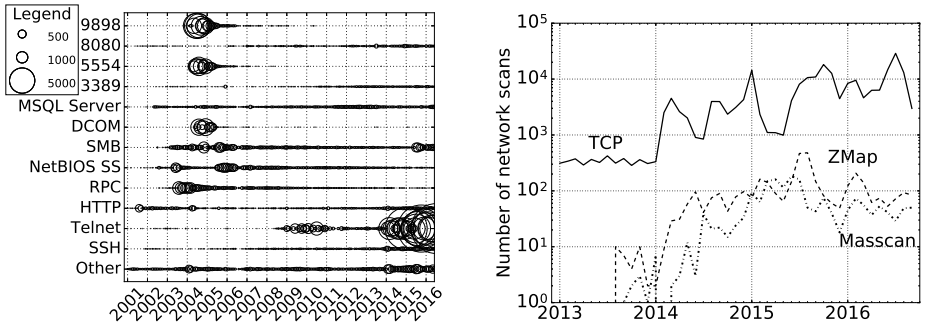
2.1 Dataset

We analyze network traffic traces from the MAWI repository, which is a collection of daily traces measured from 14:00 to 14:15 JST since January 2001 at a backbone link connecting Japanese universities and research institutions to the Internet. It mainly consists of international traffic between universities and commercial ISPs. Although the duration of each MAWI trace (i.e. 15 minutes) limits our study to a fraction of the daily traffic, the MAWI repository enables us to inspect scanning trends over 15 years. We also use several multi-day long trace captured on the same measurement point during the Day In The Life of Internet (DITL) event in 2012, 2013, 2014 and 2015. Unlike the publicly available MAWI traces where IP addresses are anonymized, our dataset contains original IP addresses to cross-reference our data with other dataset (DNS, Censys [8]).

Abnormal events appearing in the MAWI repository are automatically reported in the MAWILab [11] database then classified and annotated with a taxonomy designed for network backbone anomalies [19]. In this paper, we make use of these results and study the characteristics of traffic annotated with *network scan* labels (i.e. labels with the prefix: *network_scan*). These labels ensure that corresponding traffic has a single source and a high number of destinations (> 20). Protocol header information (SYN, ACK, FIN flags for TCP and ICMP type Echo request, Netmask request and Timestamp request for ICMP) are also used to identify different types of network scan. For UDP network scans, the taxonomy ensures that on average a destination receives less than fifteen packets and traffic is sent to less than ten distinct destination port numbers. In this paper, we analyze TCP scans (56% of all scans) because flag-based signature reduce false positives.

To assess the reliability of MAWILab events, we compare the source IP address of events annotated as network scans in the MAWI traces with the IP addresses reported by the SANS Internet Storm Center (ISC) [1] from November 2014 to March 2015. 55% of probing events annotated as network scans in MAWILab are also present in ISC's suspicious domains. This shows that the majority

[1] https://isc.sans.edu/feeds/daily_sources



(a) TCP network scans (b) Scanning tools

Fig. 1: Port targeted by TCP network scans; radius represents the number of scans and number of scans exhibiting Zmap and Masscan fingerprints.

of IP addresses labeled as scans are also detected by the firewalls participating in the DShield project.

2.2 Macroscopic trends

Destination port Figure 1a depicts the TCP scans along the 15 years of analyzed traffic. For each event, we retrieve the dominant destination port. We observe two types of trends. In the first case, some ports or services quickly arise and then slowly decay. The most extreme example of this is ports linked to worm like ports 9898 (Dabber) or 1023 and 5554 (Sasser) in 2004. Other services such as RPC (port 135 linked to Blaster worm) experience similar surge but decrease more slowly. As noted by [1], the decay is likely due to disinfection activities. A sudden surge in Telnet scans occurs in March 2014. As exposed in [20], these scans targets Telnet-enabled Internet-of-Things (IoT) devices such as cameras. Contrary to all previous sudden surge, this one does not show any sign of decrease. We hypothesize that this scanning increase is due to the lack of security update on IoT devices and the regular addition of vulnerable devices on Internet. The second main trend is constituted of classic application or destination ports that were already present 15 years ago and that remain in use today. They are thus constantly scanned during the whole duration of our study. We can here quote SSH (port 22), HTTP (port 80), SMB (port 445), MS SQL Server (port 1433), and HTTP alternative (port 8080). Although not shown here, FTP (port 21) and HTTPS (port 443) exhibit the same behavior.

These observations are qualitatively consistent with past literature [1]. Furthermore, recent Telnet scanning shown in Figure 1a motivates the constant attention that researchers should hold on network scans.

Mass-scanning tools Leonard and Loguinov [16] proposes the first Internet-wide scanning tool and showed that their scanning pattern is as polite as possible [17]. Open source scanning tools ZMap [10] and Masscan [14] were then

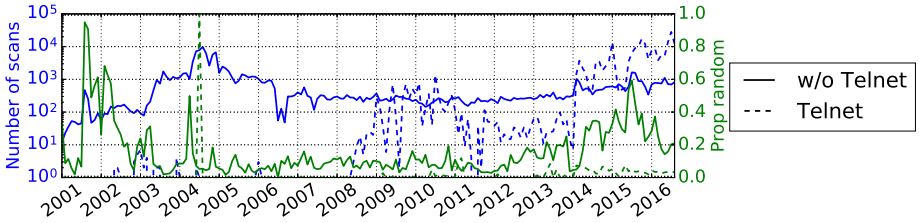


Fig. 2: Scanning patterns: number of scanners that use random patterns.

released later: in August 2013 for ZMap and September 2013 for Masscan. They are able to perform a wide variety of scans using TCP, UDP and ICMP protocols and implement specific packet fingerprints in the ID field in the IP header [9] that allow easy identification. If 95% of packets of an event match a tool's fingerprint, the scan is considered as having been performed with the considered tool. Figure 1b displays the total number of network scans along with the number of ZMap and Masscan scans. Following the release of both tools, the number of associated scans immediately arises but then almost disappears. This might be due to initial curiosity. The number of fingerprinted scans then re-increases in the beginning of 2014. Overall, ZMap is more prevalent than Masscan.

Durumeric et al. [9] observed ZMap and Masscan usage in darknet data. Their results are difficult to compare to ours because they use a different network scan definition: scanning events need to reach more than 100 destinations at a minimal rate of 10 packets per second. It is important to note that ZMap and Masscan being open-source tools, it is easy for a malicious actor to remove the fingerprinting mechanism. We may thus underestimate these tools' usage.

Scanning patterns Studying scanning patterns can provide insights into scanners' sophistication and intent. A SIP scan [7] using byte-reverse order permutation has previously been observed. Byte-reversed permutation is less aggressive than naive sequential patterns. The use of this pattern shows that the attacker wanted to avoid detection. It is however not as efficient as pseudo random pattern used by ZMap and Masscan [16]. While Section 2.2 provides an assessment of Internet-wide scanning tool usage, it however does not address overall occurrence of random scanning. We here test the monotonicity of probed destination IP addresses using the Mann–Kendall test, a nonparametric trend test [18]. We use a significance level of 0.5% to avoid false positive as suggested by Li et al. [18]. Scans that do not exhibit any trend are considered as random.

Figure 2 displays the number of scans and proportion of random ones. We here consider Telnet scans separately because they constitute the overwhelming majority of scanning after March 2014. The overall tendency for non-Telnet scans is an increase in random pattern use. Proportion values are high for the early years but those values are not reliable due to the small number of scans. The increase of random scanning starts at the beginning of 2012. Telnet scans however remain massively non-random across the dataset. The purpose of random scanning is to spread the probe load uniformly across the targeted IP range.

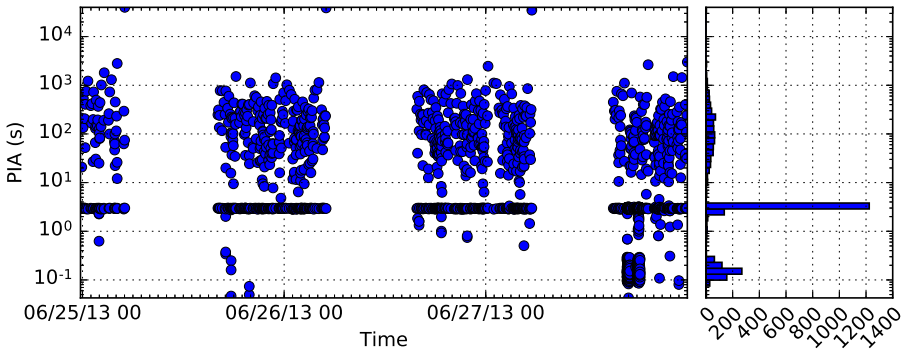


Fig. 3: Packet Inter-Arrival (PIA) for a single scanning IP observed in the DITL 2013 that targets port 445 (SMB).

This makes detection inside subsets of the whole IP range (e.g. through stub network monitoring) much more difficult because of the small number of packets that probe each subnetwork in a given time window [17]. Our results thus show that scanners increasingly use scanning patterns that aim at avoiding detection.

Existing probing patterns analyses in darknet snapshot data support our results. Bou-Harb et al. [4] found that 57% of scans in 2013 are random. Similarly, Fukuda et al. showed that, in November 2006, 10-15% are randomly behaved [12].

3 Profiling scanners

Previous section only focuses on spatial (IP address-wise) aspects of scanning. In this section, we inspect the temporal and spatial patterns of scanners.

3.1 Temporal and spatial structure example

Figure 3 depicts the Packet Inter-Arrival (PIA) times of the packets sent by a previously identified scanner in the 3-day long 2013 DITL trace. We observe a periodic repeated scan toward the same network prefix. This scanner exhibits two specific PIA values. The small PIA values (2-3 seconds) separate two probes sent to the same destination address and port. The high PIA values (between 5s and 1 hour) are due to inactive periods between couple of probes. By analyzing the destination IP address sequence, we observe that all probes target the same /1 network prefix on the destination port 445 but that successive couple of packets do not reach adjacent IP addresses. Namely the pattern does not exhibit any trend and is thus random (see Section 2.2). This example suggests that the considered scanner performs several scans or probing events that share some characteristics. We name these groups of related scans: *activity periods*.

3.2 Activity period classification

In this section, we investigate the behavior of scanners across their detected scanning events in the DITL trace. Our goal is to identify *activity period(s)* of

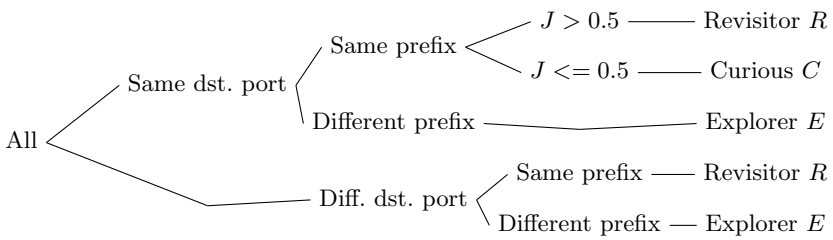
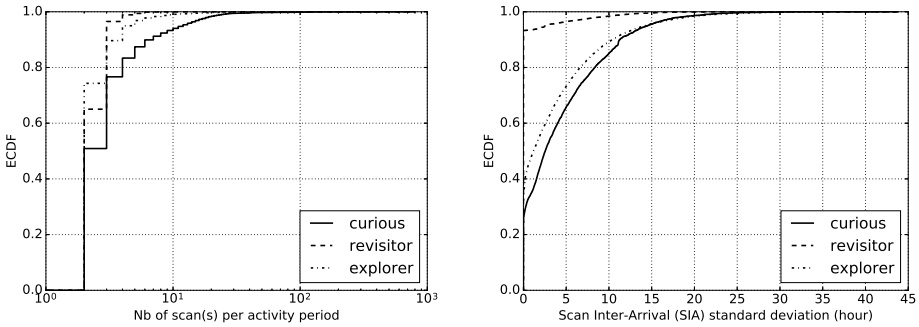


Fig. 4: Classification of activity period. J: Jaccard index-based overlap.

Labels / **Year**	**2012**	**2013**	**2014 w/o Telnet**	**2015 w/o Telnet**	**2014 Telnet**	**2015 Telnet**	**Total**
All	1416	1483	1654	11860	21293	63882	101362
1 scan	20%	49%	26%	59%	23%	29%	31%
Iso. (I)	29%	22%	36%	9%	26%	22%	22%
Expl. (E)	62%	38%	66%	20%	52%	40%	41%
Cur. (C)	3%	17%	17%	20%	1%	8%	8%
Rev. (R)	34%	22%	7%	4%	37%	38%	33%
1 AP type	42%	46%	40%	30%	42%	42%	41%

Table 1: Number and percentage of IPs that perform each type of activity periods. 1 AP type means that scanner are either E or C or R.

scanners, i.e. period(s) of time when they perform several scans that share some characteristics. We first discard scanners that perform only one scan event. Our method classifies scanner activities into the categories presented in Figure 4 from the top to the bottom using the destination port and targeted network prefix as criteria. We use destination in order to understand if scanners target several port or just a single one. We thus first check if scanning events target the same destination port and reach hosts located in the same network prefix (see Figure 3). For each scanner, we thus check sequences of successive probing events and investigate whether they target similar network prefix. The targeted network prefix of a single scanning event is defined as the CIDR prefix that contains all the destination addresses of the considered event. Two prefixes are considered as similar if they are equal, or if one is included in the other and the difference of the prefixes length is not greater than 1. Activity periods with scans that target similar prefix using the same destination port are labeled as "revisitor" or "curious". Scans in "revisitor" activity periods visit the same set of destination hosts several times, hence the "revisitor" label. This repeated behavior either intends to capture the dynamic of the probed hosts. Scans in "curious" activity periods do not exhibit any spatial overlap but instead incrementally acquire knowledge on the IP address space. They thus do not target a specific network but instead target a specific characteristic (e.g. vulnerability) on a large number of hosts. In order to separate these two behaviors, we analyze how successive scans overlap between each other in terms of destination addresses. To this end, we generalize the Jaccard index $J(A)$ of an activity period A that contains n probing events as: $J(A) = \frac{|\bigcap(S_i)|}{|\bigcup(S_i)|}, i \in 1 \dots n$ where S_i is the set of destination addresses of the



(a) Number of scanning events (b) Standard deviation of SIA time

Fig. 5: ECDF of the number of scanning events in activity periods (a) and standard deviation of Scan Inter-Arrival (SIA) time in activity periods (b) for explorer (solid line), curious (dashed line) and revisitor (dotted line).

i^{th} event in A. A is labeled as "revisitor" if $J(A) > 0.5$, "curious" otherwise. Successive scanning events that target unrelated network prefixes on the same destination port are labeled "explorer". We then follow a similar procedure to build activity periods out of scans that target distinct destination ports. Successive scans that always target the same prefix are classified as "revisitor". We did not observe scans that target the same prefix in a non-overlapping fashion using distinct destination ports and thus we did not introduce a label similar to "curious". Then, successive scans that target unrelated prefix and destination port are labeled as "explorer". After extracting these three types of activity periods, remaining scans are labeled as "isolated".

We apply this method to 101,362 scans found in four multi-day-long traces captured during DITL from 2012 to 2015. Results are displayed in Table 1. Overall, we observe that 69% of scanners perform several scans. 41% of scanners perform a single type of activity periods. 41% have explorer activity periods, 8% contain curious ones and 33% perform revisitor activity periods. Revisitors (resp. explorer) that target a single destination port represent 98.8% (resp. 34%) of all revisitors (resp. explorer). Due to the Telnet scanning surge since 2014, we actually consider Telnet and non-Telnet probing separately. We first consider non-Telnet scanners. Revisitor activity periods are steadily decreasing while curious ones are increasing. We hypothesize that this increase is due to the rise of mass-scanning tools (Section 2.2). Explorer activity periods do not exhibit any specific trend. Telnet scanners contain less curious activity periods and more explorer and revisitor than non-Telnet scanners. This is due to the fact that many Telnet scans target several /24 prefixes.

3.3 Activity period characteristics

We then analyze the characteristics of activity period belonging to the three previously defined types of scanners: explorer, curious and revisitor. Figure 5

presents the Empirical Cumulative Distribution Functions (ECDF) of activity period characteristics. Figure 5a depicts the number of scans inside activity periods. Curious activity periods contain a higher number of scans than revisitor and explorer activity periods. Figure 5b displays the ECDF of the Scan Inter-Arrival time for activity periods that contain at least three probing events. Scans in revisitor activity periods occur in more regular interval. This is consistent with an automated periodic probing of a specific set of hosts.

By further analyzing activity period characteristics, we note that revisitors activity periods and scanning events have a shorter duration than explorer ones which are in turn shorter than curious. The same ranking applies to network prefix size of activity periods. Scans in revisitors and explorer activity periods exhibit similar network prefix coverage (i.e. they reach the same proportion of destination addresses inside their targeted prefix). Scans in curious activity periods have a much lower coverage. Similarly, scans of revisitor and explorer activity periods have a higher packet rate than curious ones. These observations are consistent with their activity periods roles: curious activity periods perform slow incremental scans and acquire knowledge one scan at a time, while revisitor and explorer ones quickly gather information on a smaller scope.

From a general point of view, we note that curious scanners seem to target very large prefixes in a random manner. Their goal is likely to scan a wide part of the Internet for a specific purpose. However, revisitor scanners exhibit an intent to acquire knowledge on a specific target. Our method helps administrators to understand the actual intent of scanners, and thus, provide a clear picture of the active threats towards their network.

4 Publicly documented scanning entities

By analyzing DNS name of scanning IPs using ZMap and Masscan, we noticed several security researchers, from both universities and companies, such as University of Michigan or Shadow Server. We extended this analysis on our whole dataset, 15 years of MAWI traces and DITLs, and identified 18 entities.

4.1 General results

We present our results in Table 2. These entities' scans (resp. scanning IPs) represent 0.44% (resp. 0.1%) of all the observed scans (resp. IPs), 22% (resp. 18%) of ZMap ones and 32% (resp. 3.7%) of Masscan ones. Some entities perform many scans during a long period of time (e.g. Shadow Server) while others are only active punctually (e.g. SBA research). 55% of entities only scan /1 prefix while other entities target prefix lengths between 1 and 24. We only display the number of scans that target SSH, HTTP, HTTPS and POP. Some entities such as Michigan, 360.cn, Shadow Server or Shodan actually probe many other ports. The proportion of Telnet scans from identified entities is much smaller than that of all observed scans (see Figure 2). Entities mainly use ZMap and Masscan although ZMap is more prevalent. The increase in activities between

Entity type	Entity name	Dataset	# scans	# IP	Prefix len. range	# dest. port					Tool			Act. per.			
						22	80	443	8080	Other	Other	ZMap	Masscan	I	E	C	R
Academic	Berkeley	M15	3	1	1-18		3					3		-	-	-	-
		D15	6	1	1		6					6		1			
	Camb.	M15	3	1	1-18		3					3		-	-	-	-
		D15	12	1	1-18		12					12		2			
	Michigan	M14-16	25	24	1-18	6	3	5		7	4	25		-	-	-	-
		D15	416	170	1-18	32	36	81		251	19	413		3		91	34
	TUM	M13-15	10	2	1-18	1	1	4		3	9	2		-	-	-	-
Company	Cymru	M13-15	6	1	1			6			6			-	-	-	-
		D15	6	1	1			6								1	
	Eddie	M14	3	2	1			3			3			-	-	-	-
	Cornejo	D14	8	1	1			8			8					1	
	Errata	M13-16	8	4	1		4	1	2	1			8	-	-	-	-
	Security	D14	16	1	1	8	8						16	10		2	
	IPredator	M14	2	1	1		2						2	-	-	-	-
	ISP	M14	2	1	1		2						2	-	-	-	-
	Labs	M14-16	61	27	1	1				60		61		-	-	-	-
	Rapid7	D14	12	11	1		12					12				1	
	NS All.	M13-16	19	2	1-8				1	18		1	18	-	-	-	-
	PLC	M14-16	18	1	1					18		18		-	-	-	-
	Scan	D14	33	1	1					33		33				1	
		D15	2	1	1					2		2				1	
	P. 25499	M14-15	17	5	1		2	5		10		17		-	-	-	-
	PS	M15	7	1	1				7	7		7		-	-	-	-
	SBA	M15-16	1	1	1			1	1				1	-	-	-	-
	360.cn	M14-16	655	7	1	29	7	5	9	598	21	144	504				
		D14	1135	3	1-2	87	31	29		980	8	1135		2		25	
		D15	19	4	1	1				18			19	3		4	1
		M13-16	462	135	1-24		3	186		273	13	449		-	-	-	-
	Shadow	D13	72	1	1-24			72			72			10	3	21	
	Server	D14	279	31	1			279				279				31	
		D15	3374	139	1-18		1251	425		1698		3374		125	34	685	4
	Shodan	M13-16	80	8	1	1	19	12	1	38	71			-	-	-	-
		D13	5	2	1			1		4	5					2	
		D14	35	1	1			17			53						2
		D15	803	10	1	5	43	53	73	627	805			9		208	4

Table 2: Entities activity in MAWI traces. Entities short names: Camb.: Cambridge, TUM: Technische Universität München, ISP: Internet Scanning Project, NS All.: Network Security Alliance, P. 25499: Project 25499, PS: Proxy Scan, SBA: SBA Research. Dataset: M: MAWI (ex M14-16 → MAWI 2014-2016), D: DITL (ex: D14 → DITL 2014). Activity periods: I: Isolated ; E: Explorer ; C: Curious ; R: Revisitor.

Entity type	Entity name	# IP	IP DNS PTR prop	IP webpage prop	PTR webpage prop	Webpage information		
						Email contact	Optout	IPs/ Prefix available
Academic	Cambridge	1	1.0	1.0	1.0	Yes	Yes	Yes
	Michigan	181	1.0	1.0	0.92	Yes	Yes	Yes
	Berkeley	1	1.0	1.0	1.0	Yes	Yes	No
	TUM	2	1.0	1.0	1.0	Yes	Yes	Yes
Company	Cymru	1	1.0	1.0	1.0	Yes	Yes	Yes
	Eddie Cornejo	3	1.0	0.0	0.0	-	-	-
	Errata Security	4	0.75	0.5	0.5	No (ND)	No (ND)	No (ND)
	IPredator	1	1.0	0.0	0.0	-	-	-
	IS Project	1	1.0	1.0	1.0	Yes	Yes	No
	Labs Rapid7	27	0.93	0.93	0.93	Yes	Yes	Yes
	NS Alliance	2	0.0	0.5	1.0	No	Yes	No
	PLC Scan	1	1.0	0.0	1.0	Yes	Yes	No
	Project 25499	5	1.0	1.0	1.0	Yes	Yes	Yes
	Proxy Scan	1	1.0	1.0	1.0	No	Yes	No
	SBA Research	2	1.0	1.0	1.0	Yes	Yes	Yes
	360.cn	7	1.0	0.0	0.0	-	-	-
	ShadowServer	140	0.98	0.97	0.97	Yes	No	No
	Shodan	11	1.0	0.0	0.5	No	No	No

Table 3: Scanning entity identifiability for famous scanners. ND means not directly (actually documented in blog posts).

DITL 2013 and 2014 clearly emphasizes the rise of both tools (see Section 2.2): except Shodan, all identified entities now use ZMap or Masscan. Using the classification presented in Section 3.2, we observe that until 2013, these entities' activity periods are isolated or explorer or curious. As they switch to use ZMap and Masscan, their activity periods almost completely become curious. Curious activity periods contain scans that target the same prefix but different IP addresses which is consistent with the probing behavior of ZMap and Masscan. As curious represent only 8% of all scanners (see Table 1) identified entities are thus behaving differently from other scanners. This further shows that the method proposed in Section 3 provides accurate insights on scanner behavior.

4.2 Deployed online documentation infrastructure

ZMap [10] documentation proposes several guidelines regarding scanning[2]. They especially emphasize three aspects that are relevant for network administrators. First, researchers must state the benign nature of the traffic through DNS PTR record and webpages. Second, probing purpose must be explained. Third, one must provide contact and the possibility to opt-out from probing. Beyond the obvious interest of informing administrators, appropriate documentation may also demonstrate good faith in case of lawsuit[3]. We here analyze how entities docu-

[2] `https://zmap.io/documentation.html#bestpractices`

[3] https://community.rapid7.com/community/infosec/sonar/blog/2013/10/30/legal-considerations-for-widespread-scanning

ment their scanning and whether they propose opt-out mechanisms and contact information. Table 3 presents our results. We gather IP-reachable webpage using Censys.io [8] and automate crawling of scanning IPs' DNS PTR record and associated webpage.

All entities, except Network Security Alliance, provide PTR records for the majority of their scanning IPs. Some entities do not actually have any webpages accessible with either IP or PTR record that document their scanning: Eddie Cornejo, 360.cn and IPredator. Shodan redirects some PTRs to their homepage. Errata Security sets up webpages but only lists previous scanning results. Errata Security however documents its activities through blog posts[4][5][6][7] but this makes it difficult for operators to understand that the activity they investigate is innocuous scanning. We then manually analyzed every setup webpage. All entities provide contact email address except Network Security Alliance and Proxy Scan. These entities actually both scan port 8080 and use a form for opt-out request. We thus hypothesize that they actually cooperate. Shadow Server webpages do not propose opt-out. Finally, many entities do not provide the IP addresses or prefixes that they use to perform scans: Berkeley, Internet Scanning Project, Network Security Alliance, Proxy Scan and Shadow Server.

Some entities that do not scan anymore or change their scanning IP along time may have removed DNS PTR record and/or webpages. We thus may miss some infrastructure that may have been set up in the past. We thus contacted entities that do not follow guidelines and asked them about the state of their infrastructure in the past. We chose to update Table 3 accordingly despite the fact that we cannot verify their statements. From a general point of view, only 39% of entities completely follow ZMap guidelines. Existing online scanning documentation is thus not sufficient.

5 Concluding remarks

Scanning is a pervasive component of network traffic. We provide new insights into recent such as the rise of Telnet scanning, and, the increase of mass-scanning tool and random scanning patterns usage. We propose a new method that profiles scanners' behavior, and discover that 33% of scanners repeatedly target the same hosts along time. These scanners show intent to acquire knowledge on a specific target and update this knowledge along time. Their occurrence can alert administrators that their network is under scrutiny from an attacker. Publicly documented scanners are different from other scanners. For example, while a Telnet scanning surge is occurring, documented scanners only marginally probe Telnet. Furthermore, they mainly perform spread random scans. This further show that our profiling method is efficient to discriminate scanners' behavior. Finally, only 39% of these scanners follow online documentation best practices.

[4] http://blog.erratasec.com/2011/10/scanning-internet.html
[5] http://blog.erratasec.com/2013/07/scanning-internet.html
[6] http://blog.erratasec.com/2013/09/we-scanned-internet-for-port-22.html
[7] http://blog.erratasec.com/2016/05/doing-full-scan-of-internet-right-now.html

References

1. Allman, M., Paxson, V., Terrell, J.: A brief history of scanning. In: Proc. of IMC'07. pp. 77–82 (2007)
2. Alsaleh, M., van Oorschot, P.C.: Network scan detection with lqs: A lightweight, quick and stateful algorithm. In: Proc. of AsiaCCS'11. pp. 102–113 (2011)
3. Bhuyan, M.H., Bhattacharyya, D.K., Kalita, J.: Surveying port scans and their detection methodologies. Computer Journal 54(10), 1565–1581 (2011)
4. Bou-Harb, E., Debbabi, M., Assi, C.: On fingerprinting probing activities. Computers & Security 43, 35 – 48 (2014)
5. Brownlee, N.: One-way traffic monitoring with iatmon. In: Proc. of PAM 2012. pp. 179–188 (2012)
6. Czyz, J., Kallitsis, M., Gharaibeh, M., Papadopoulos, C., Bailey, M., Karir, M.: Taming the 800 pound gorilla: The rise and decline of ntp ddos attacks. In: Proc. of IMC'14. pp. 435–448 (2014)
7. Dainotti, A., King, A., Claffy, K., Papale, F., Pescapè, A.: Analysis of a /0 stealth scan from a botnet. Transactions on Networking 23(2), 341–354 (2015)
8. Durumeric, Z., Adrian, D., Mirian, A., Bailey, M., Halderman, J.A.: A search engine backed by Internet-wide scanning. In: Proc. of CCS'15 (2015)
9. Durumeric, Z., Bailey, M., Halderman, J.A.: An internet-wide view of internet-wide scanning. In: Proc. of USENIX Security'14. pp. 65–78 (2014)
10. Durumeric, Z., Wustrow, E., Halderman, J.A.: Zmap: Fast internet-wide scanning and its security applications. In: Proc. of USENIX Security'13. pp. 605–620 (2013)
11. Fontugne, R., Borgnat, P., Abry, P., Fukuda, K.: MAWILab: combining diverse anomaly detectors for automated anomaly labeling and performance benchmarking. In: Proc. of CoNEXT'10. pp. 1–12 (2010)
12. Fukuda, K., Fontugne, R.: Estimating speed of scanning activities with a hough transform. In: Proc. of ICC'10. pp. 1–5 (2010)
13. Glatz, E., Dimitropoulos, X.: Classifying internet one-way traffic. In: Proc. of IMC'12. pp. 37–50 (2012)
14. Graham, R.D.: Masscan: Mass ip port scanner. `https://github.com/robertdavidgraham/masscan`, accessed: 2015-10-16
15. Jung, J., Paxson, V., Berger, A., Balakrishnan, H.: Fast portscan detection using sequential hypothesis testing. In: Proc. of SP 2004. pp. 211–225 (2004)
16. Leonard, D., Loguinov, D.: Demystifying internet-wide service discovery. Transactions on Networking 21(6), 1760–1773 (2013)
17. Leonard, D., Yao, Z., Wang, X., Loguinov, D.: Stochastic analysis of horizontal ip scanning. In: Proc. of INFOCOM'12. pp. 2077–2085 (2012)
18. Li, Z., Goyal, A., Chen, Y., Paxson, V.: Towards situational awareness of large-scale botnet probing events. Transactions on Information Forensics and Security 6(1), 175–188 (2011)
19. Mazel, J., Fontugne, R., Fukuda, K.: Taxonomy of anomalies in backbone network traffic. In: Proc. of TRAC'14. pp. 30–36 (2014)
20. Pa, Y.M.P., Suzuki, S., Yoshioka, K., Matsumoto, T., Kasama, T., Rossow, C.: Iotpot: Analysing the rise of iot compromises. In: Proc. of WOOT'15 (2015)
21. Wahid, A., Leckie, C., Zhou, C.: Characterising the evolution in scanning activity of suspicious hosts. In: Proc. of NSS'09. pp. 344–350 (2009)
22. Yegneswaran, V., Barford, P., Ullrich, J.: Internet intrusions: Global characteristics and prevalence. In: Proc. of SIGMETRICS'03. pp. 138–147 (2003)