Abnormal Behavior-Based Detection of Shodan and Censys-Like Scanning

Seungwoon Lee[1], Seung-Hun Shin[2], and Byeong-hee Roh[1]
[1] Dept. of Computer Engineering, Ajou University, Suwon, Korea
[2]University College, Ajou University, Suwon, Korea
{swleeyg, sihnsh, bhroh}@ajou.ac.kr

***Abstract*—Shodan and Censys, also known as IP Device search engines, build searchable databases of internet devices and networks. Even these tools are useful for security, those also can provide the vulnerabilities to malicious users. To prevent the information disclosure of own IP devices on those search engines, a fundamental solution is blocking the access from the scanners of them. Therefore, it is needed to understand and consider their scanning mechanism. Therefore, we propose an abnormal behavior based scan detection of Shodan and Censys. To do this, several traditional scan detection approaches are combined and applied to satisfy their specification. Proposed idea is monitoring packets whether it is abnormal or not and adding on the suspicious list if it is. This is based on traditional threshold approaches. To figure out it is abnormal, stateful TCP stateful packet inspection is used. The response behavior during the connection can be identified with TCP flag and abnormal behavior can be classified with SYN Scan, Banner Grabbing, and Combined SYN and Banner Grabbing. Demonstration is simulated in a Censys-like environment and detected time variation per variance of distributed detectors and Threshold value is analyzed**.

Keywords—network security, IP Device search engine, Shodan, Censys, SYN Scan, Banner Grab

I. INTRODUCTION

The IP device search engines such as Shodan [1] and Censys [2] build searchable databases storing the information of the devices connected on the internet and provide the access to query for any user. They scan the group of IPv4 addresses and analyze the response by periods. Through this work, user can find the information such as states of ports, services, and operating system (OS). Even these tools are useful for monitoring and managing network to assure security, those also can provide the vulnerabilities to malicious users. Also, because of scanning which makes them traceable, malicious users are prefer to use search engines in information gathering step. For this reason, Shodan and Censys are generally known as hacker friendly search engines.

To prevent the information disclosure of own IP devices on those search engines, a fundamental solution is blocking the access from the scanners of Shodan and Censys. In a single host view, both engines use SYN scanning and banner grabbing method to collect the information of IP devices [3][4]. After open ports of target device are discovered with SYN scan, banner grab is activated against those ports. the scan result can consist of open-port list of the devices, service and its version on that port. Thus, it can make the device safe from Shodan and Censys to detect and prevent SYN scan and banner grab. In a topological view, Shodan and Censys operate several distributed servers and they scan the hosts separately and is aggregated in the database [1][4]. Also, they use horizontal scan which means scanning a single port on multiple hosts.

Those scans can be detected using the ideas of existing researches. Single host port scan including TCP SYN and Banner grabbing can be detected using misuse detection [5][6] and anomaly detection [7] in IDSs (Intrusion Detection System). Misuse detection is also called rule-based detection that misuse patterns are stored in the database. In Anomaly detection, scan is found out by recording unusual behavior of operations, so both needs to model the misuse and the unusual behavior. In sum, each method has its own advantages and vice versa. Therefore, deploying optimal scan detection of Shodan and Censys is possible with integrate their part on demand.

In this paper, we propose abnormal behavior-based scan detection to Shodan and Censys. To do this, several traditional scan detection approaches and methods are combined and optimized to satisfy the specification of Shodan and Censys. Proposed idea is monitoring packets whether it is abnormal or not and adding on the suspicious list if it is. This is base on traditional threshold approaches. To figure out it is abnormal, stateful TCP stateful packet inspection is used. The response behavior during the connection can be identified with TCP flag and abnormal behavior can be classified with SYN Scan, Banner Grabbing, and Combined SYN/Banner Grabbing. Also, we build models and simulate on Riverbed OPNET for experiments.

The rest of paper is organized as follows. Section 2 describes briefly background of Shodan and Censys, and their mechanism of information gathering, and Section 3 explains our proposed method for detecting Shodan and Censys. The modeling and simulation on OPNET are shown in section 4. Section 5 gives experiments and simulation results, and Section 6 concludes the paper.

II. BACKGROUND

A. IP Device Search Engine

1) Shodan

Shodan is known as Internet of Things search engine published by Matherly in 2009 [1]. This service collects information about 500 million devices per month and provides

them to users. Shodan checks whether target port is open or not with TCP SYN scan and if it is open and grabs target's banner string which includes specification such as operating system, name and version of the service. Unfortunately, Details about the tools used in Shodan are unknown. According to the official web site, distributed servers around the world scan 41 well known ports mainly and each banner grab executes separately and is aggregated in the database [1]. Each server scans only one port of many target hosts [7]. Also, Bodenheim et al. [3] analyzed that it takes under 4 days from SYN scan to uploading on the web but interarrival time of Scan probes is unknown.

2) Censys

Censys is a search engine that allows users to query the devices and networks on the internet [2]. Similar to Shodan, it collects data on hosts and websites through periodical and horizontal scan of the IPv4 address space with their own port scanning tools called Zmap and Zgrab [8]. Zmap is used to send the SYN scan probe to target and Zgrab is used for banner grab after SYN Scan. Censys can be inferred to scanning 35 ports based on its data collection and uses distributed scan using multiple scanners and scheduler [4].

TABLE 1. COMPARISON BETWEEN SHODAN AND CENSYS

	IP Device Search Engine	
	***Shodan* [1]**	***Censys* [2]**
Techniques to Collecting the Information	SYN Scan / Banner grab	SYN Scan / Banner grab
Scan Range	Horizontal Scan	Horizontal scan
SYN Scan Tool	Unknown	Zmap
Banner grab tool	Unknown	Zgrab
scan server	Distributed	Distributed
Target ports	41 ports	35 ports
Scan period	Unknown	Automatically scheduled

B. Mechanism of Information Gathering and detection

On their web sites, users can browse the open ports as SYN Scan result and their banner information. The characteristic of information gathering of them is as follows.

1) TCP SYN Scan

TCP SYN Scan [9] is a well-known scan and used as a default for major network scanning tools. It is also known as half open scan because it does not complete the three-way handshake. With sending SYN probe packet and receiving response packet, scanner can judge whether port is open or not. A port can be decided as open when ACK packet is received, while close when RST packet is received.

To detect TCP SYN scan, misuse detection [5][6] and anomaly detection [7] are used. To find misuse for other TCP Scan, TCP Flag is added on the misuse database. However, it is not effective for SYN Scan because SYN packet is frequently happened in typical network communication. In Anomaly detection, SYN scan is defined as no ACK Packet response after Sending SYN/ACK. It is very easy to model unusual behavior.

2) Banner Grab

Banner Grab is an application layer scan technique and activated when scanner connects to the target host with TCP 3-way hand shaking. in a narrow sense, Banner is the simple text usually includes signatures of service and displayed when the connection is established in several protocols such as FTP, SMTP, POP3 and telnet. In a broad sense, it includes every behavior gathering information from the open port (i.e. GET method in HTTP).

Process of banner grabbing is similar with TCP Connect Scan. Its detection is also based on misuse detection and anomaly detection.

3) Horizontal Scan

This mechanism is focused on a range of the targets not a behavior. horizontal scan means that scan a single port on multiple hosts. Both search engine is According to the result of Shodan, each scanned port has different time that cannot consider to vertical scan. On Zmap which Censys uses for port scan, only horizontal scan is implemented [4].

4) Distributed Scan

Those search engines cover a scale of whole IPv4 addresses. Thus, using multiple scanner is reasonable. Previously, it is mentioned that Shodan has distributed servers around the world and Censys uses scheduler and multiple scanners. Because each scanner has own IP Addresses unless they are behind NAT, it is one of challenges to detect and trace the distributed scan.

In Summary, Shodan Censys check horizontally whether the port of target is open or not using TCP SYN scan. After that, they grab the banners against the host with open port and upload on the web site.

III. ABNORMAL BEHAVIOR BASED SCAN DETECTION OF SHODAN AND CENSYS

We design abnormal behavior based scan detection of Shodan and Censys. In the previous section, we analyzed the scanning characteristics of Shodan and Censys and its detection methods. On the basis of those, it can be modifying the existing methods to satisfy specifications of Shodan and Censys

A. Thershold approach in Packet Level

This concept is traditionally used for scan detection but effective. This examines events across a time window to detect port scan attacks above certain thresholds [10]. In this environment, the suspicious hosts on the list by abnormal behaviors are granted the configurable threshold Th for counting value c and monitoring time window $\tau_{monitor}$. Because sometimes normal packets can be considered unusually to SYN scan by network failure, threshold Th must be configurable for user's policy. Figure 1 shows the flow chart Abnormal Behavior-Based Scan Detection of Shodan and Censys. The packet by suspicious host h_k are detected as an abnormal behavior, h_k is added on the suspicious list l_{sus}. Also,

Packet received time t_{rcv} is set to current time t_{cur} and count value c of h_k increases by one. When c reaches to the threshold Th, this host will be classified to scanners and user can choose block it or leave it. If there is no abnormal behavior after $\tau_{monitor}$ ($t_{cur} > t_{rcv} + \tau_{monitor}$), the host will be removed from l_{sus}.

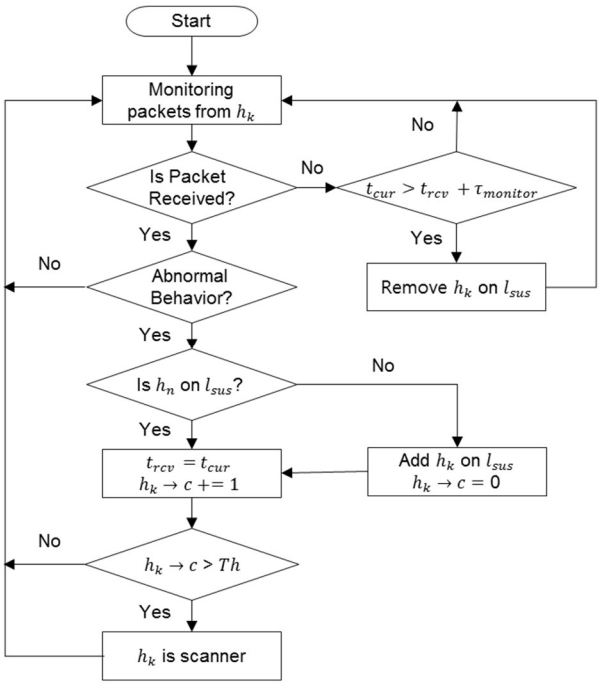

Figure 1. Flow Chart for Abnormal Behavior-Based Scan Detection

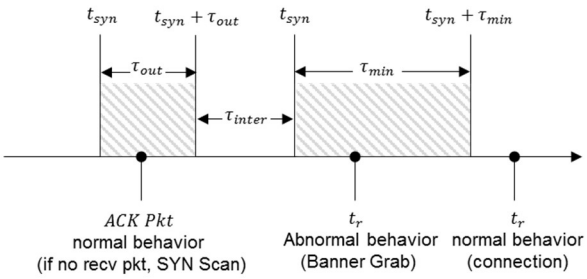

Figure 2. Time Diagram for Abnormal Behavior Detection

B. Abnormal Behavior Detection

To detect an abnormal behavior alike Shodan and Censys, TCP stateful packet inspection is used [11]. Packets are identified refer to stored TCP Flag information of the previously received packet. The time diagram for each detection is illustrated in Figure 2.

1) TCP SYN Scan Detection

When the detector receives the SYN packet to the open port, it responds with SYN/ACK packet to the scanner. However, the scanner does not reply ACK packet because the information scanner looking for is already acquired. Assume the SYN packet arrival time is t_{syn} and τ_{out} is a waiting time until time out for receiving ACK packet. If there is no ACK packet response until $t_{syn} + \tau_{out}$, the sender would be added on the list of suspicious host.

On the contrary, when the port is closed, detector sends RST/ACK packet and scanner responses nothing and banner grab is not executed. Because Sending SYN packet to the closed port is generally uncommon [12], it can be considered as an abnormal behavior by the strength of policy user configured.

2) Banner Grab Detection

In banner grab, scanner makes full connection with the host. Typical banner grab is manually activated by human makes long-term connection, but interconnection time of automated banner grab only takes less than a second. Let assume t_{syn} is the SYN packet arrival time, t_r is RST/ACK packet arrival time which means abnormal disconnection time and τ_{min} is the minimum time for disconnection. If t_r is between t_{syn} and $t_{syn} + \tau_{min}$, it would be regarded to abnormal behavior.

3) Combined SYN+Banner Grab Detection

If Banner grab is detected right after SYN scan detection during τ_{inter}, this situation can be defined as "apparent scanning behavior". It also can be considered as scanner directly or a single abnormal behavior depending on the strength of policy user configured.

C. Distributed Detection

Distributed detection is a countermeasure of horizontal and distributed scans. In the distributed detector, the detection method is exactly same with the abovementioned detection. Main difference between them is that detectors share their suspicious lists. The aggregated list can be assigned on any one of the detectors or extra node or gateway. On the contrary to single host, deploying detection and sharing on every node costs but is more effective.

IV. Modeling and Simulation

We model scan method used in Shodan and Censys and its detection on OPNET Network simulator for an experiment. Unlike other network simulators, it supports graphical user interface(GUI), therefore it is easy to consider scalability of the topology in this virtual network. In other words, we can scale the number of scanners and detectors per the experiment.

Scanner model is revised simple Ethernet client model by adding SYN scan and banner grab. Abnormal behaviors caused by SYN scan and banner grab are considered to the procedures between nodes in this abstracted simulation. For SYN scan, when scanner sent SYN scan packet and received SYN/ACK Packet, it should not reply ACK packet and any others to target. For automated banner grab, scanner makes full connection with the target host and send a small amount of packet after SYN Scan

The n_{sc} scanners process each periodic scan following Poisson distribution $P_{sc}(k, \tau_{sc})$ with parameter λ_k ($k =$

$1,2,3,\dots,n_{sc}$). The probability that m scan events of kth scanner occurs, where τ_{sc} is a scan period and n_{port} is the number of target ports, can be as follows:

$$P_{sc}(k,\tau_{sc}) = \frac{(\lambda_k \tau_{sc})^m e^{-\lambda_k \tau_{sc}}}{m!} \quad where\ \lambda_k = n_{port} \qquad (1)$$

Detector model follows the proposed method, abnormal behavior based scan detection and is revised normal Ethernet server model.

The topology adopts simple mesh networks as shown in Figure 3 and Figure 4 shows the environment with 40 detectors. ISP Backbone network makes a connection between scanner and detectors. The number of detectors increases and they connect to the mesh routers per each experiment.

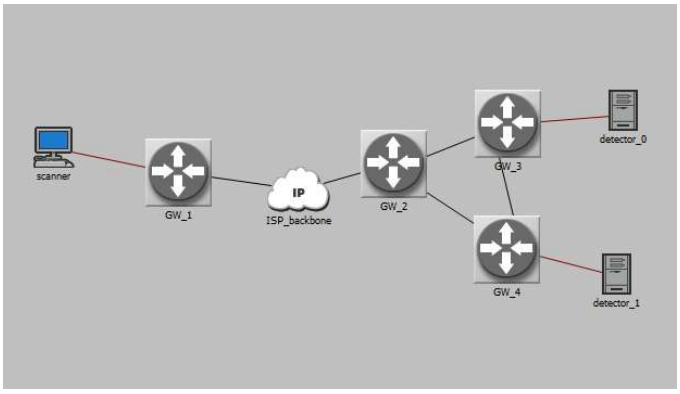

Figure 3. Simulation Topology with Two Detectors

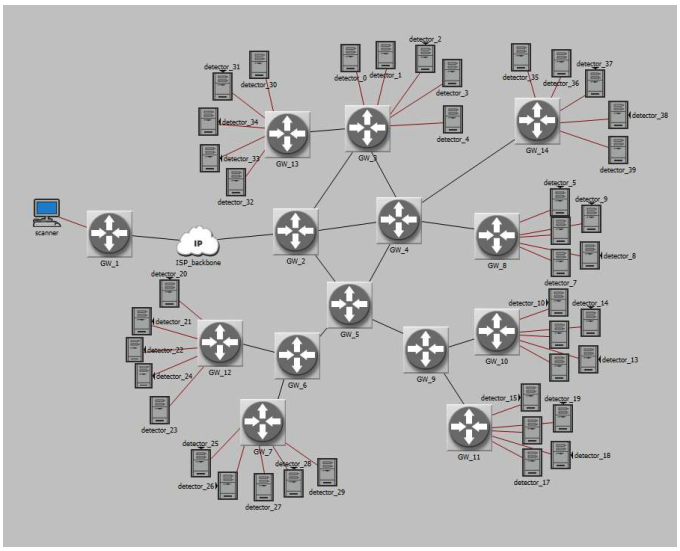

Figure 4. Simulation Topology with 40 Detectors

V. Experiment and Analysis

As an experiment, detected time variation per variance of distributed detectors and Threshold value is analyzed through simulation result.

In the configuration setup of the simulation environment, scan period imitates Censys analyzing its data history data. Thus, n_{port} is set to 35 and τ_{sc} is set to three days. Parameters of detector model is configured as follows: τ_{out} for timeout period set to is 30 seconds calculated based on Windows default retransmission timeout value which is 5. The number of detector $n_{detector}$ increases from one to 40 and Threshold to count the behavior of suspicious host increases by one to 30.

The result is obtained with logarithmic-scaled graph, as shown in Figure 5. It shows that detected time per the number of detector and detection threshold. When threshold Th is one, scanners can detect the scanner as even they are not distributed. It also means that false positive of detection might be too frequent. Increasing the threshold value gradually, detected time also is getting late linearly. The results of distributed detectors are formed like step curve, because the count of same host is reaching faster when they share their list of suspicious host. This result suggests that scan and detection are activated as intended. Scanner model imitates the behavior of Shodan and Censys and detector model can detect SYN scan and banner grab. Also, it is confirmed that distributed nodes are sharing their suspicious list. In other word, detecting horizontal scan is available and effective with abnormal behavior based detection.

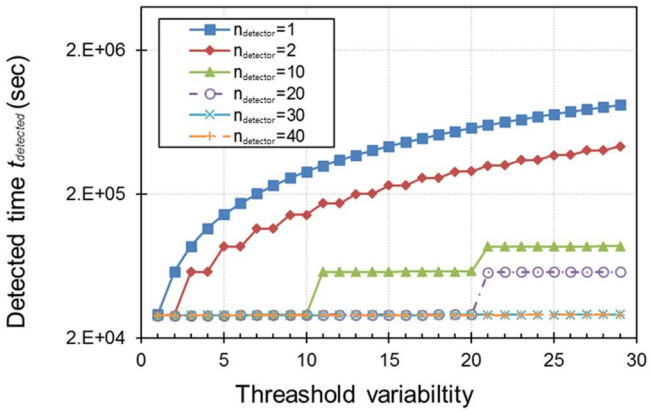

Figure 5. detected time variation per the number of detector and threshold

VI. Conclusion

This paper presents an abnormal behavior based scan detection of Shodan and Censys. Oure method processes monitoring packets until abnormal behavior is detected and counts abnormal behaviors of each host using threshold approach. Our abnormal behavior divides into three categories: SYN Scan, Banner Grabbing, and Combined SYN and Banner Grabbing. To identify them as an abnormal, stateful TCP stateful packet inspection is used. Our proposed method is shown to be effective with demonstration in a Censys-like environment. Also, With the experiment of detected time variation per variance of

distributed detectors and Threshold value, horizontal scan detection is shown to be detected and its effect.

ACKNOWLEDGMENT

This work was supported partially by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIP) (no. NRF-2015R1A2A2A01005577).

REFERENCES

[1] Shodan, Available: https://www.shodan.io/

[2] Censys, Available: https://www.censys.io/

[3] R. Bodenheim, J. Butts, S. Dunlap, B. Mullins, "Evaluation of the ability of the Shodan search engine to identify Internet-facing industrial control devices," International Journal of Critical Infrastructure Protection, Vol.7, No. 2, pp. 114-123, Jun. 2014

[4] Z. Durumeric, D. Adrian, A. Mirian, M. Bailey, "A Search Engine Backed by Internet-Wide Scanning," ACM CCS' 2015, Oct. 2015

[5] R. Singh, D. Tomar, "Network Forensics: Detection and Analysis of Stealth Port Scanning Attack," International Journal of CNCS, vol. 3, No.2, Feb. 2015

[6] D. Kang, "Learning Classifiers for Misuse and Anomaly Detection Using a Bag of System Calls Representation", In Proceedings of the 6th IEEE Workshop on Information Assurance and Security United States Military Academy, West Point, NY,2005.

[7] M. Bhuyan, D. Bhattacharyya, Kalita, J, "Surveying Port Scans and Their Detection Methodologies," The Computer Journal, BXR035, 2011

[8] Z. Durumeric, E. Wustrow, J. Halderman, "ZMap: Fast Internet-wide Scanning and its Security Applications, " In 22nd USENIX Security Symposium, Aug. 2013

[9] M. De Vivo, E. Carrasco, G. Isern, and G. O. de Vivo, "A review of port scanning techniques," ACM SIGCOMM Computer Communication Review, vol. 29, no. 2, pp. 41–48, 1999.

[10] S. Bahk, H. Kim, I. Kang, "Method of improving security performance in stateful inspection of TCP connection,". U.S. Patent Application No 11/129,774, 2005

[11] C. Gates, "Co-ordinated port scans: a model, a detector and an evaluation methodology," PhD Thesis, Dalhousie University Halifax, Nova Scotia., 2006

[12] S. Staniford, J. Hoagland, J. McAlerney, " Practical Automated Detection of Stealthy Portscans," Journal of Computer Security, vol. 10, no. 1-2, pp. 105-136, 2002