International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 16 (2017) pp. 5900-5904
© Research India Publications. http://www.ripublication.com

Port Scanning Attack Analysis with Dempster-Shafer Evidence Theory

Rajni Ranjan Singh

Department of Computer Science and Engineering,
Maulana Azad National Institute of Technology,
Bhopal, M.P., India.

Orcid Id: 0000-0002-8524-2556

Deepak Singh Tomar

Department of Computer Science and Engineering,
Maulana Azad National Institute of Technology,
Bhopal, M.P., India.

Orcid Id: 0000-0001-9025-1679

Abstract

Port scanning is a process of probing networks, finding vulnerabilities and than infiltrate IT recourses. It is often the fundamental method utilized by intruder prior to initiate a targeted cyber attack. Port scan attack traffic does not contain any specific signature, therefore IDS based detection may suffer by generating many/false alerts. Manual examination is an error prone, labor intensive and time consuming process.

This work presented an approach to detect port scanning attack based on the entropy and failed connection attempt made by each host. To analyze and prioritize the observed evidence, Dempster-Shafer theory is utilized to calculate combined belief of each host in support of the proposed hypothesis.

A proof of concept prototype has been implemented using open source SNORT IDS system which uses, internet traffic data injected with crafted scans to validate the system. It is observed that the proposed approach correctly identifies and prioritize the crafted scans injected into real traffic.

Keywords: Intrusion Detection System, Network Forensics, Dempster-Shafer Evidence Theory, Port scanning Attack

INTRODUCTION

Port scanning is a facility to check whether a given port is open or not. Network administrator utilize port scanning facility to troubleshoot the networking errors, however attacker can use same facility to identify open ports, firewall rules, which operating system is running on the remote system etc. Once these information has been extracted than attacker can find out the associated vulnerability of the remote system that can be further exploited to enter in to the system.

Port scanning can be active and passive. In passive port scanning, attacker passively observe the traffic generated by the remote system by sniffing the network. It is very difficult to trace passive scanner because attacker do not send any packet to the target system therefore it is very difficult to identify attacker. On the other hand in active port scanning, attacker send network packets to the target system and on the basis of response attacker will identify open and closed ports. Active port scanner can be easily caught by observing firewall/IDS logs.

Port scanning techniques can be classified in four categories. Categories are Vertical, horizontal, strobe and block scans. In vertical scanning attacker scan all the ports of a single machine. In horizontal scanning attacker scan same port of all the machines in the given network. In strobe scanning attacker scans multiple ports of many machines and in block scanning attacker scans all the ports of all the machines in a given network.

DEMPSTER SHAFER EVIDANCE THEORY

Depster shafer theory is an efficient method to combine degree of belief derived from independent item of evidence. D-S theory deal with uncertain information based on the evidences and combination of them.

D-S evidence theory includes the frame of discernment. usually frame is represented by θ which contains mutually exclusive facts (events).

Basic belief probability assignment (BPA), allocate the belief over the power set of the frame of discernment and is defined as:

$$m: 2^{\Theta} \rightarrow [1, 0]$$

let θ be a frame of discernment and m_{Θ} is a BPA function. The belief function is defined as.

$$\text{For } x \subseteq \theta \;\; \text{Bel}(x) = \sum_{y \subseteq x} m_{\theta}(y)$$

The belief function shows how much confidence we have in that one of the hypothesis contained in x hold.

Dempster-Shafer has a combination method, the goal of which is to combine evidence for a proposed hypothesis from multiple independent sources and calculate an overall belief for the hypothesis. In general we have following rule of combination known as the Dempster Rule.[8-10]

$$m_{1,2}(h) = \frac{1}{1-K} \sum_{h_1 \cap h_2 = h} m_1(h_1).m_2(h_2)$$

$$K = \sum_{h_1 \cap h_2 = \{\}} m_1(h_1).m_2(h_2)$$

SYSTEM ARCHITECTURE

Figure 1 presents the proposed system architecture. The acquisition system consists of two sensor modules. Sensor modules can be software or hardware devices attached in front of the router through an in-line network cable and capture all packets streams transmitted to and from the network. Two independent sensors are deployed to collect evidences (packets) from the same live networks during same time window. This provides more complete data collection. In future both parties can show the same set of captured traffic this association verifies the correctness of evidence collection. The logs gathered by the two sensors are transmitted to hash calculation module and a copy of Meta data are preprocessed and stored in database for further investigation.

A hash value of collected data is calculated using most common hash functions MD5 and SHA1. This gives guarantees that digital evidence has not been changed since it was acquired and investigator will be capable to prove same when the similar process has been repeated on the original data.

Original collected data with hash value are preserved on the read only write once backup media.[11-21]

Some specialized sensors like SNORT can able to collect relevant packets based on the supplied rules and a metadata is created in text format. As per the ACPO guidelines [22] selective captured evidences are now permission in the court of law.

A copy of the meta data are converted in to suitable format (CSV is mostly acceptable format) and imported into any open source database for further query driven manual examination.

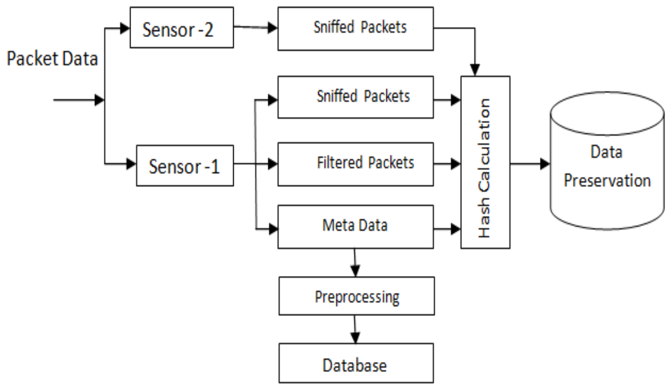

Figure 1: Proposed Forensic Architecture

PORT SCAN ATTACK CHARECTORIZATION

A port scan involves the remote host trying to connect to a large number of destination ports. it is observed that out of 65535 ports only few port are active(open) at a time therefore it is sure some connections attempts are targeted to closed ports and they will be judged abnormal.

Suppose if there are H hosts in a network and the probability of one host being active is S1.the probability of finding an live host after trying only once is S2.

S2 = H.S1…………… (1)

Each host having 65535 ports. Generally only few ports are active. if there are Hp active ports in the host and the probability that an attacker finds an active ports after trying only once is S3,

S3=Hp / 65535.............. (2)

It is observed that Hp is commonly less than 15 i.e Hp < 15, so

S3<0.000228

If the probability of finding an active port of an active host after trying only once is S4,

S4= S3.S2……………….(4)

S4 < 0.000228 P2

S4 < 0.000228

If scanner does not know the present inner information of network. The probability to find active port of an active host after trying once is very small and its proximately less than 1%.

Therefore remote host scan large number of ports in order to get open ports and this process generate huge amount of failed connection.

Following two hypotheses are drawn from the above mention theory.

Hypothesis 1(h_1): As the host which scans larger no of different destinations Ip addresses and ports, is probably a port scanners.

We can computer the entropy of each host, which reflects the distribution of its destinations Ip addresses and ports.

Suppose x={x_1,x_2---x_m} are the set of host observed while analyzing traffic in a given network. The entropy of a particular host x is defined as H(x).

P(x_i) = Total no. of unique connection/Total connection.

$$H(X) = -\sum_{i=1}^{M} p(x_i)\log_2 P(x_i)$$

In vertical port scanning attempt, a node perform port scanning to many ports of a single host, therefore the entropy of destination port number field is high. However in horizontal port scanning attempt, a node perform port scanning to the large no of computers with same port information, therefore entropy of destination IP address should be high.

Therefore in this work we are going to calculate two different entropy values for a particular host. In our experiment we consider only maximum of two entropies.

Hypothesis 2(h_2): As the host which scans large no of different destinations Ip addresses and ports, it is sure that, some connections are targeted to closed ports results many failed connections. Therefore host which attempt may failed connections is probably the port scanner.

Failed connections are identified by TCP Reset packets or ICMP errors. In this work we are going to calculate total number of failed connection attempt made by each host.

A node is a port scanner if an only if it satisfies both hypotheses. Suppose a node, which perform many connection attempts to many destinations however it does not attempt any failed connection therefore it cannot be a port scanner.

Similarly a node which performs many failed connection attempts. However it does not send packets to many destinations. Therefore it cannot be treated as a port scanner.

PORT SCAN ATTACK ANALYSIS WITH DEPTSTER-SHAFER EVIDANCE THEORY

In this work we utilize Dempster-Shafer model to calculate a numeric confidence score(combined belief) for both proposed hypothesis and prioritize the results based on the scores.

Here let Θ be a frame of discernment is a disjoint set of host machines represented by their IP addresses.

$$\Theta = \{Ip_1, ip_2, ip_3 \ldots\ldots ip_n\}$$

Two BPA function h_1 and h_2 allocates the belief over the set of the frame of discernment after normalization. Here $m_1(h_1)$ and $m_2(h_2)$ are the numerical values observed by the hypothesis 1 and 2 for each IP addresses.

Dempster-Shafer method calculates the overall combined belief of both hypotheses for each element of frame of discernment (for each IP address).

Combined belief helps the investigator to prioritize the further analysis.

EXPERIMENT UNDER REAL NETWORK TRAFFIC

To test the usefulness of proposed work an experiment environment has been setup consists of four machines connected via switch (layer 2) as shown in the figure 2.

As shown in the figure 2. Scanner machines 1 & 2 perform port scan attack to the target system using well known port scanning tool nmap[23]. Here scanner 1 & 2 perform TCP SYN and Connect scanning. Table-1 shows configuration details of test bed.

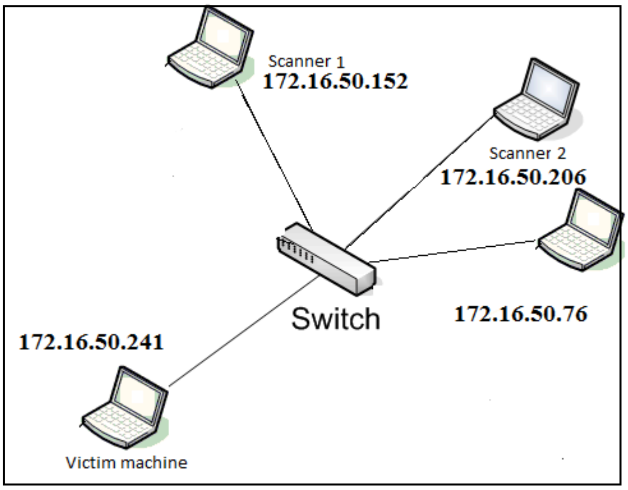

Figure 2: Test bed Topology

Table 1: Test-bed machines configuration details

Machine Name and IP Address	Operating system, Hardware configuration	Applications	Application Configuration/ Command
Victim machine 172.16.50.241	Microsoft Win. 7 professional Intel dual 2 Duo 2.40 GHz 3 GB RAM	Snort 2.9.5.3 Wireshark 1.10.1	SNORT configured in Network Intrusion detection mode and generate alert in Fast alert mode.
Scanner 1 172.16.50.152		nmap 6.47(an windows version of Nmap)	**Nmap –sT 172.16.50.241**
Scanner 2 172.16.50.206			**Nmap –sS –P 172.16.50.241**
Scanner 3 172.16.50.76			Perform normal operation

To capture the network traffic as a evidence an Intrusion detection system SNORT[24] is utilized to carry out sniffing at the victim machine.

SNORT is a signature based intrusion detection system able to capture selected packets based on the rules. Snort captures the packet in TCPDUMP format and generates alerts in text format. Traffic has been captured by both snort and wireshark[25] acquisition system from the same live network source during the same acquisition time window. Snort alerts are converted into CSV format and uploaded to open source MYSQL[26] database for further Query driven analysis.

Scanner 1, 2 performing port scanning simultaneously to the victim machine (in the presence of normal internet traffic). A relevant packet capture run using the snort IDS.

Table 2: Observed Traffic Information

IP address	**Entropy (Destination Port)**	**Entropy (Destination IP address)**	**Failed Connection Attempts**
17.16.50.152	0.5271	0.00052	2968
17.16.50.206	0.00144	0.00052	989
17.016.50.76	0.1066	0.2825	0

Table 3: Basic belief value of each scrutinize host

	172.16.50.152	**17.16.50.206**	**17.16.50.76**
$\boldsymbol{m_1(h_1)}$	0.64	0.001153	0.34
$\boldsymbol{m_2(h_2)}$	0.75	0.2499	00

Combined belief of each IP addresses has been calculated by D-S theory that shows the belief on hypothesis h_1 and h_2 for each host machines.

Table 4: Combined belief of each host

	Combined belief
$m_{h1,h2}$(172.16.50.152)	0.72
$m_{h1,h2}$(172.16.50.206)	0.000576
$m_{h1,h2}$(172.16.50.76)	00

It is observed that combined belief of 172.16.50.152 is the highest. Host 172.16.50.76 belief is 00 means, it is not a port scanner. Combined belief is useful to prioritize the further analysis.

CONCLUSION

This work focused on the detection and prioritization of port scanning attack evidences. Here two hypotheses are introduced to carry out attack detection and dempster shafer theory is utilized to prioritize the further investigation.

REFERENCES

[1] S. Panjwani, S. Tan, K.M. Jarrin, and M. Cukier. An experimental evaluation to determine if port scans are precursors to an attack. In Proc. Int. Conf. Dependable Systems and Networks, 2005. DSN 2005., pages 602 – 611, june-1 july 2005 IEEE.

[2] Elias Bou-Harb, Mourad Debbabi, and Chadi Assi. Cyber Scanning: A Comprehensive Survey. IEEE Communications Surveys & Tutorials, Vol. 16, No. 3, Third Quarter 2014.

[3] Manowar H Bhuyan,D K Bhattacharyya and J K Kalita" Surveying port scans and their Detection methodologies" The Computer Journal (2011) 54 (10): 1565-1581 April 20, 2011 ACM.

[4] Whitepaper by dethy@synnergy.net "Examining port scan methods - Analyzing Audible Techniques"

[5] Brenden Claypool, Stealth port scanning methods, SANS Institute 2000 - 2002. As part of GIAC practical repository. Claypool – 2002 v1.4.

[6] Port Scanning Techniques, http://nmap.org/book/man-port-scanning-techniques.html

[7] Atul Kant Kaushik, Emmanuel S. Pilli, R.C. Joshi, Network Forensic System for Port Scanning Attack. 2010 IEEE 2nd International Advance Computing Conference.

[8] Loai Zomlot, Sathya Chandran Sundaramurthy, Prioritizing Intrusion Analysis Using Dempster-Shafer Theory, Proceedings of the 4th ACM workshop on Security and artificial intelligence, Chicago, Illinois, USA October 21 - 21, 2011 ACM Digital Library.

[9] TIAN Zhihong,JIANG Wei,LI Yang,DONG Lan,A digital evidence fusion method in network forensics systems with Dempster-shafer theory, China Communication Volume: 11, Issue: 5, May 2014 IEEE.

[10] Guo-lin Shao, Xing-shu Chen,Xue-yuan Yin and Xiao-ming Ye, A Fuzzy detection approach toward different speed port acan attacks based on Dempster-Shafer Evidance theory, Security and Communication Networks, Wiley online library 2016.

[11] Bruce J. Nikkel, A portable network forensic evidence collector, digital investigation 3 (2006)127–135 2006 Elsevier Ltd.

[12] M.I. Cohen, PyFlag – An advanced network forensic framework , digital investigation 5 (2008) S112 – S120 2008 Digital Forensic Research Workshop. Published by Elsevier Ltd.

[13] Emmanuel S. Pilli*, R.C. Joshi, Rajdeep Niyogi, Network forensic frameworks: Survey and research challenges digital investigation 7 (2010) 14 – 27 , 2010 Elsevier Ltd.

[14] Vicka Corey, Charles Peterman, Sybil Shearin, Michael S.Greenberg, and James Van Bokkelen, Network Forensics Analysis, IEEE Internet Computing, November December 2002

http://computer.org/internet/ 1089-7801/02/$17.00 ©2002 IEEE

[15] Eoghan Casey, Network traffic as a source of evidence: tool strengths, weaknesses, and future needs, digital investigation (2004)1 28 – 43 , 2004 Elsevier Ltd.

[16] Bruce J. Nikkel, Generalizing sources of live network evidence , digital investigation (2005)2 193 – 200 , 2005 Elsevier Ltd

[17] Mark Solon, Penny Harper, Preparing evidence for court, digital investigation (2004)1 279–283, 2004 Elsevier Ltd

[18] Florian Buchholz,Eugene Spafford, On the role of file system metadata in digital forensics, digital investigation (2004)1 298–309, 2004 Elsevier Ltd.

[19] Erin E. Kenneally ,Digital logs-proof matters, Digital investigation (2004)1 94–101, 2004 Elsevier Ltd.

[20] Bruce J. Nikkel, Improving evidences acquisition from live network sources, Digital investigation 3 (2006) 89–96, 2006 Elsevier Ltd.

[21] Chris Boyd,Pete Forster, Time and Data issues in forensic computing – a case Study, Digital investigation (2004)1 18–23, 2004 Elsevier Ltd.

[22] APCO Good Practice Guide for Digital Evidance http://www.digital-detective.net/digital-forensics documents/ACPO_Good_Practice_Guide_for_Digital_Evidence_v5.pdf

[23] Nmap, www.nmap.org.

[24] Snort, https://www.snort.org/

[25] Wireshark, https://www.wireshark.org/

[26] MYSql,www. http://www.mysql.com/