Standardization of Intelligent Information of Specific Attack Trends

Ashima Rattan, Navroop Kaur and Shashi Bhushan

Abstract In recent days, cyber-attacks are rising rapidly by using various new techniques. These attacks have huge impact on organizational and an individual security. As many times an attack has been detected but it is too late to recover the damage perform by that attack. To study on previous attacks some organizations like Defense Advanced Research Project Agency (DARPA) provide offline dataset for researchers. KDD and DARPA dataset attributes was playing a good role in detection of many attacks and further useful in prevention of attacks also. But in recent days, dataset provided by them, become old one and not gives fruitful results. To keep in mind, the technique used in this research work is providing machine readable dataset attributes of specific attacks in a standard format which is CSV (Comma Separated Values) format. The attack data is captured by deploying various honeypot sensors. The achievement of this research work is "sharing of targeted attack data like Brute force Attack, Exploits etc., in machine readable form in standard format". This information is useful for security researchers, situational awareness programs and security communities. Security testing is another area, also needs some dataset attributes for security testing of the softwares or tools.

Keywords Attacks · Attack trends · Exploits · Brute force Scans · CSV format and honeypots

A. Rattan (✉) · S. Bhushan
Department of IT, Chandigarh Engineering College Landran, Mohali, Punjab, India
e-mail: rattan_ashima@yahoo.com

S. Bhushan
e-mail: shashibhushan6@gmail.com

N. Kaur
CSTD Center for Development of Advanced Computing (CDAC), Mohali, Punjab, India
e-mail: navroop_kohli@yahoo.com

C. R. Panigrahi et al. (eds.), *Progress in Advanced Computing and Intelligent Engineering*, Advances in Intelligent Systems and Computing 714,
https://doi.org/10.1007/978-981-13-0224-4_7

1 Introduction

In recent days, cyber-attacks are rising rapidly by using various new techniques. The existing security works with the focus on finding the traditional protection and detection methods [1]. However attacker performs lot of attacks in very short time. These attacks have huge impact on organization and an individual security [2]. The first response to such campaigns is to detect them and collect sufficient information regarding tools, techniques used to exploit the vulnerability [3]. Hence effective capturing of the attack data and its timely dissemination to defenders is required for the mitigation and prevention of the large-scale attacks [4]. In this paper we have established the need for such an automated attack data capturing and sharing mechanism. The cyber threat information sharing is very important and very useful in the field of cyber security, where organization can take protective measures on time by watching the previous attack information. Such type of information is useful for security researchers, security agencies, etc., in a standard structured format which is readily usable\actionable by them [5–7]. We have also highlighted the fact that the format for sharing attack data is very crucial and the data sharing format should be machine digestible to reduce the human intervention and increase the response time [8]. As the threat landscape is ever changing, so as the techniques used for mitigation of those threats needs to be dynamic in nature [9]. In this research we tend to look for the feasibility of using proactive approaches for the mitigation of the dynamic threat to the security. The first level of defense in any deficient security set up is the firewall hosted [10, 11]. This device/software is responsible for catalog all the communication to and from the organization and allows and disallows the IP address based upon their reputation [12]. In this research work, attack data is collected through capturing and the event database is created. It helps to detect the malicious traffic.

2 Research Scope

The main focus of our research work involves in three steps: Situational Awareness, Attack Attribution and Cyber Security Researchers.

2.1 Situational Awareness

As the threat landscape is ever changing, so as the techniques used for mitigation of those attacks needs to be dynamic in nature. Cyber-attacks are increasing day by day and their impact is likely to be very much disturbing and harmful for the users. The term Cyber Situational Awareness refers to monitor all the unusual events and

occurrence of bad activities which are specially performed by the attackers or the hackers [13]. The organization that works on Cyber Situational Awareness collects the current attack data, works on the collected attack data to find the refined and correct information about them as a result. Such information is helpful to aware the society about those new attacks and their possibilities by providing the refined data to harm over the cyber security network.

2.2 Attack Attribution

Attack attribution may be defined as in which it helps to provide the information of an attacker as well as attacker's channel [14]. The information includes the identity and the location of an attacker. Traditionally attack attribution is simply a process of trace back of an attacker. The identity of the attacker which is obtained by tracing includes the username, e-mail id, an account, an alias, password, an IP address, or geographic location [15]. The main principle behind the attack attribution technique depends upon the untrusted nature of IP protocol. If the source IP address is not authenticated then it is very easy to trace the location or address.

2.3 Security Researchers

As the threat landscape is ever-changing, so as the techniques used for mitigation needs to be dynamic in nature. Hence we are providing latest dynamic attacked data (i.e. IP reputation) and providing it in a standard structured format which is in a CSV format which is machine digestible. This format can be directly input for machine learning. Therefore, this kind of data provided in standard structured format and is hence useful for researchers in the domain of detection of malicious traffic. Standard sharing format can be used as CSV format and can be directly input into machine learning. Hence, this data is extremely useful for researchers. You will get the latest trends, reputation latest attack trends and IP reputation.

2.4 Security Testing

Security testing may be defined as the software testing which is useful to uncover the vulnerabilities (holes) of the system. Security testing is the technique to secure the data and information so that attackers could not able to attack or steal the important and personal data of the system. To protect and maintain data properly, it is the easiest way. It helps to find out the all means of escape and fault of the system

which may results to the loss of information. It finds the way out to protect the mislay information.

3 Design Principles of Event Database

The principal behind event database is to provide readily available attack data in machine digestible form and provide broad prospective to researchers in the field of security. This basically covers specific types of attacks like Brute force attack, Exploits, etc., [16]. This is a big contribution to nation in cyber security by providing specific attack enrich data.

3.1 Network Architecture

In network architecture three machines are used, these are servers, each one is consisting of two ports, i.e., eth0 and eth1. Almost every server has two ports for various communications. The three servers are named as (Fig. 1):

- Broadband relational server.
- Elasticsearch server, i.e., Event DB is deployed on OS Ubuntu.
- Web server having OS Ubuntu—used for portal to access the data.

3.2 Repository Architecture

3.2.1 Honeypot

Honeypot is a system to trace the attacks by fooling the attackers and to get the information about how attackers exploit vulnerabilities in IT system [17–19].

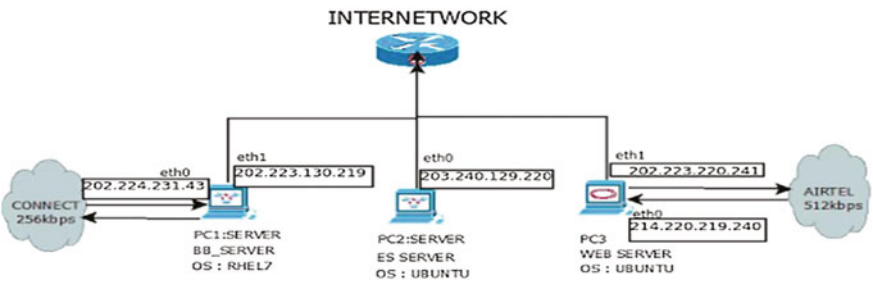

Fig. 1 Network architecture

Honeypot sensors do not let the attacker know about the system legitimacy. Attackers do not know that somebody has kept eye on them and are being monitored secretly.

- **Active Honeypot**—The term active (client) honeypot describes an advanced honeypot system. In contrary to traditional honeypots that undergo passively all attack attempts, active honeypot systems actively react to them.
- **Passive Honeypot**—Passive (server) honeypots offer services and wait for attacker to exploit the vulnerabilities.

3.2.2 Broadband Server

OS: RHEL7 having broadband connection to find the vulnerability in broadband network (Fig. 2).

3.2.3 Relational Database

In this research the relational database is used to store the data which is captured through various honeypots. Relational database is mandatory for establishing a relation between data captured from various honeypots and hence play a big role in specific types of data collection. The collected data further refined to provide the information about attacks such as Specific Attack Trend, example of collected data from various honeypot sensors, data capturing date and time, Attacker IP, Connection established, Services exploited, Malwares Downloads, Malware_virustotal_results, Generation_of_network_traffic, Events_detected.

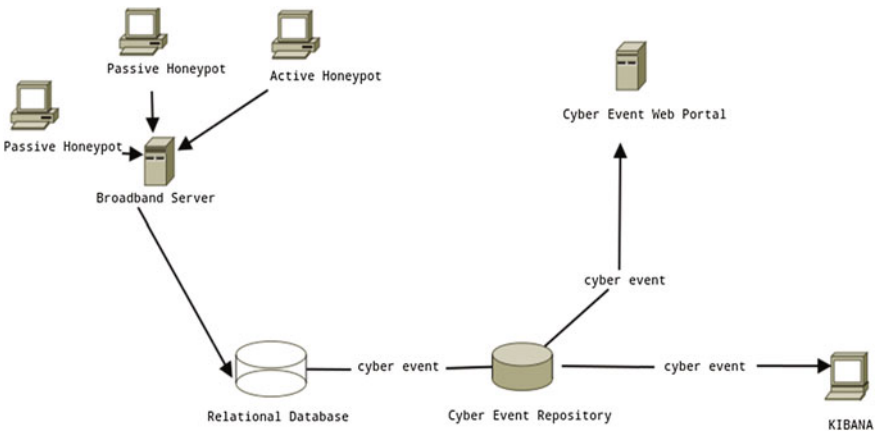

Fig. 2 Architecture of repository

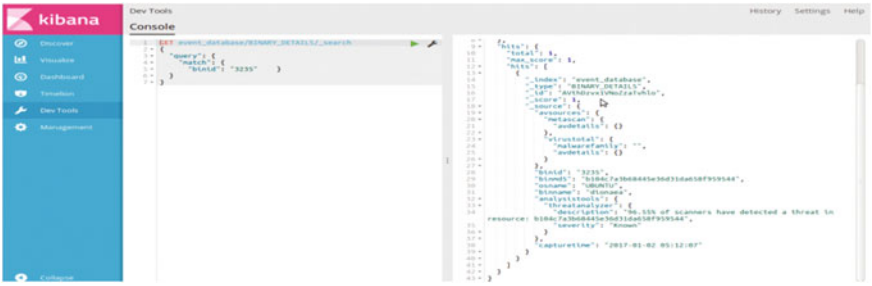

Fig. 3 Kibana output

3.2.4 Cyber Event Repository

Cyber Event Repository is a repository that contains latest information about specific Attack Trends. This has been designed by keeping all the aspects of security where this type of information can be useful.

3.2.5 Elasticsearch, Logstash and Kibana (ELK)

In this research paper, the results are obtained by using the combination of Elasticsearch, Logstash and Kibana (ELK stack). Elasticsearch uses Apache Lucenewhich helps to generate and govern the inverted index. Elasticsearch is an approach to provide the fast responses according to user's search. It works on the real time platform; therefore it is the easiest and fastest approach to obtain the accurate results [20]. Logstash is an open-source tool which logs are collected, parsed, and stored for future use. Kibana is the web-based interface, the logs are indexed through Logstash which is helpful to display the results. Elasticsearch, Logstash and Kibana, when used together are known as an ELK stack. Therefore in this research ELK combination is used for better responses. Few of the kibana commands used to extract SMTP scans (Fig. 3).

4 Attack Data Results

Capturing and sharing specific attack trends is a big challenge which come up with many new problems like selection of standard format for sharing, to find out the most prominent features which attackers left with us, type of attack attacker prefer, what situational awareness we can provide [21]. So keeping all this in mind the attack data results gives various types of information like top 10 attacker IP along with the list of attacking IP's captured by our honeynet sensors, most top attacked port along with the details of various attacking port captures now a days, specific types of attacks

captured using honeynet sensors, etc. [22]. Based on type of attack attacker is doing we have categorized and providing details of basically 3 types of attacks, i.e., Exploits, Scans and Brute force, etc. Details of which are mention below.

TOP 10 Attacker IP
See Fig. 4.

4.1 Specific Attack Trend Captured

As already discussed above about to find and segregation of the attack trends is a big challenge, when new attacks are coming day by day. Here shown graphical representation is shown of three specific attacks such as: Exploits, Brute force, and scans (Fig. 5).

4.2 Exploits

An exploit is an attack, when the attacker finds the vulnerability in the system then it takes the advantage to exploit the particular vulnerability. We keep check on

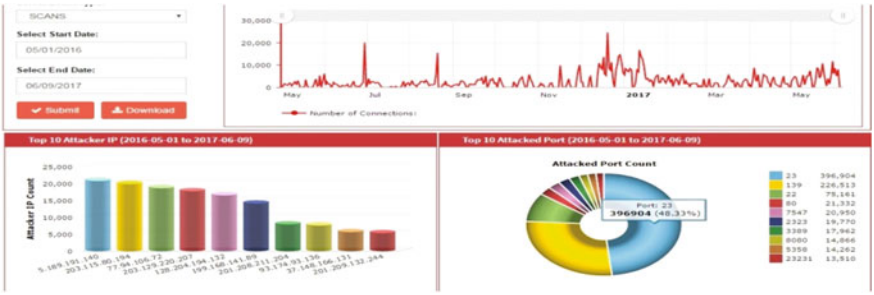

Fig. 4 Top Attacker IP

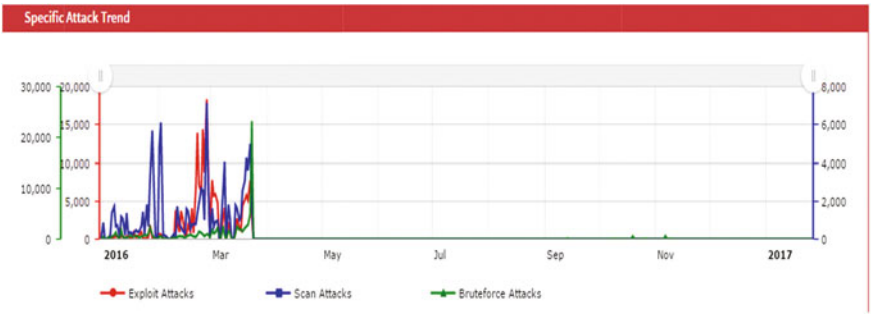

Fig. 5 Graph showing specific attacks

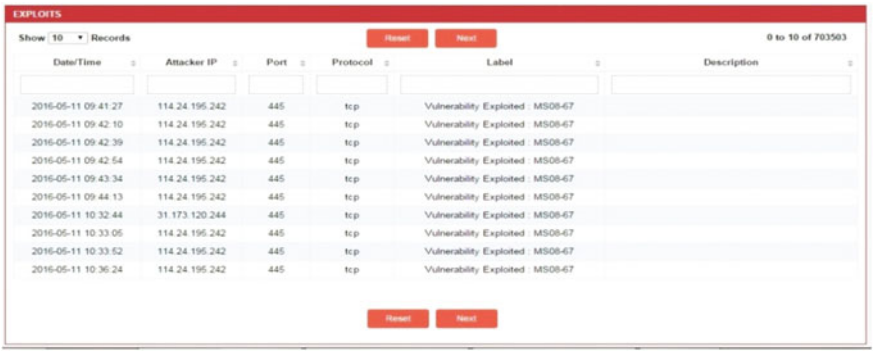

Fig. 6 Exploit attacks

types of vulnerability exploited with the aim of finding the exploits. And at the end we are showing the count of exploits we have captured on daily and providing the record accordingly (Fig. 6).

4.3 Brute Force Attacks

Brute force attack may be defined as the attack when an attacker wants to steal the user's password or personal identification by attacking. When any type of authentication like if the user is asked for username or password on any website then user must be aware of it, that he is going to be a target of attacker. Therefore when this type of data (username and password) is found here in the payload, then it is clear that attacker is trying to do the brute force attack (Fig. 7).

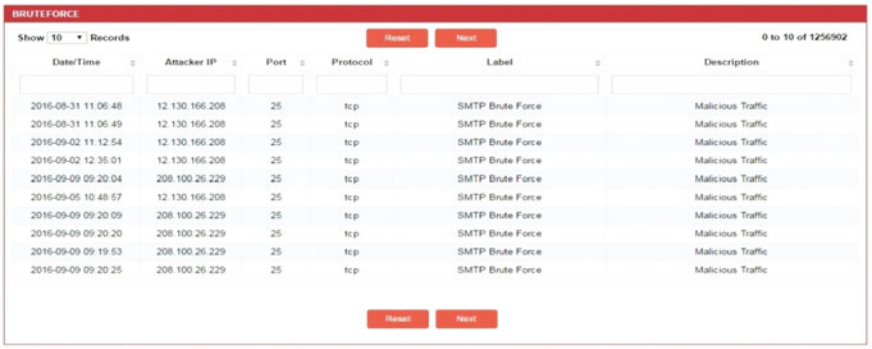

Fig. 7 Brute force attacks

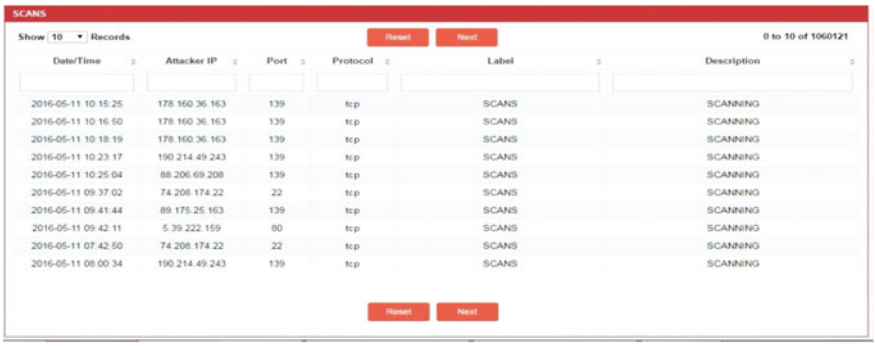

Fig. 8 Scan attacks

4.4 Scans

Scans identify the hosts who are active on the network for network security assessment. It is a way to recognize the running network services on the targeted hosts. The network services include User Datagram Protocol (UDP), Transmission Control Protocol (TCP), Operating System (OSs), and TCP sequence number predictability, etc. It is a method to tighten the system security and also an effective way to troubleshooting the system. It is a technique which is used to detect the known vulnerabilities computing system that are available on the network. So based upon this, we are trying to find the attacks which are executed through scanning (Fig. 8).

5 Standard Format for Researchers

Attack data is required to detect unauthorized activities to positively identify all true attacks and negatively identify all non-attacks, monitoring and analyzing user and system activities, to recognize known specific types of attacks and alerts, for statistical analysis of abnormal behavior model. In this research work we are providing the standard sharing format for researchers. They require this type of data for better response. The standard structured sharing format is provided in this research work which can be used as CSV format and can be directly input into the machine learning [23]. Hence the information provided in this format is extremely refined, qualitative, and useful in manners of latest attack trends for IP reputation.

One of the formats is CSV format (Comma Separated Values) is a file format for data storage which looks like a text file [24–26]. The information is organized with one record on each line and each field is separated by comma. CSV is human readable and easy to edit manually, simple to implement and parse, provides straight forward information schema (Figs. 9 and 10).

	A	B	C	D	E	F	G	H	I
1	date_time	attacker_ip	protocol	source_port	destination_port	label	others	description	payload
2	5/3/2016 14:28	222.186.21.57	tcp	6786	1433	Malicious MSSQL TRAFFIC		MSSQL Bruteforce	
3	5/3/2016 14:28	222.186.21.57	tcp	10254	1433	Malicious MSSQL TRAFFIC		MSSQL Bruteforce	
4	5/3/2016 14:41	222.186.3.52	tcp	5773	1433	Malicious MSSQL TRAFFIC		MSSQL Bruteforce	
5	5/3/2016 14:41	222.186.3.52	tcp	8520	1433	Malicious MSSQL TRAFFIC		MSSQL Bruteforce	
6	5/3/2016 15:45	124.193.177.29	tcp	6224	1433	Malicious MSSQL TRAFFIC		MSSQL Bruteforce	
7	5/3/2016 15:45	124.193.177.29	tcp	7358	1433	Malicious MSSQL TRAFFIC		MSSQL Bruteforce	
8	5/3/2016 15:54	222.186.56.21	tcp	16900	1433	Malicious MSSQL TRAFFIC		MSSQL Bruteforce	
9	5/3/2016 15:14	123.249.34.132	tcp	2194	3306	Malicious MySQL TRAFFIC		MySQL Bruteforce	
10	5/3/2016 15:33	123.249.45.166	tcp	3532	3306	Malicious MySQL TRAFFIC		MySQL Bruteforce	
11	5/3/2016 15:41	221.194.44.173	tcp	2302	3306	Malicious MySQL TRAFFIC		MySQL Bruteforce	
12	5/3/2016 15:59	173.254.236.104	tcp	1289	3306	Malicious MySQL TRAFFIC		MySQL Bruteforce	
13	5/3/2016 16:26	173.254.236.104	tcp	1567	3306	Malicious MySQL TRAFFIC		MySQL Bruteforce	
14	5/3/2016 16:58	120.24.177.101	tcp	50481	3306	Malicious MySQL TRAFFIC		MySQL Bruteforce	
15	5/3/2016 17:21	222.186.34.204	tcp	1353	3306	Malicious MySQL TRAFFIC		MySQL Bruteforce	
16	5/3/2016 17:54	23.88.177.135	tcp	4295	3306	Malicious MySQL TRAFFIC		MySQL Bruteforce	
17	5/3/2016 15:04	118.193.213.172	tcp	1921	3306	Malicious MySQL TRAFFIC		MySQL Bruteforce	
18	5/3/2016 15:05	173.254.236.104	tcp	4057	3306	Malicious MySQL TRAFFIC		MySQL Bruteforce	

Fig. 9 Machine digestible data in CSV format

811	5/3/2016 17:13	27.54.248.124	tcp	5081	1433	alicious MSSQL TRAFF		MSSQL Bruteforce		
812	5/3/2016 16:29	46.172.71.249	SSH	null	22	SSH BruteForce	user_name	Malicious SSH Traffic	cd ..	
813	5/3/2016 15:52	183.3.202.88	SSH	null	22	SSH BruteForce	user_name	Malicious SSH Traffic	wget http://104.223.72.179:258/hlvip	
814	5/3/2016 14:29	74.208.174.22	SSH	null	22	SSH BruteForce	user_name	Malicious SSH Traffic	wget http://104.223.72.179:258/hlvip	
815	5/3/2016 14:29	74.208.174.22	SSH	null	22	SSH BruteForce	user_name	Malicious SSH Traffic	chmod 0777 hlvip	
816	5/3/2016 15:16	183.3.202.88	SSH	null	22	SSH BruteForce	null	Malicious SSH Traffic		

Fig. 10 Result showing extracted payloads

6 Conclusion

Cyber security is a broad area and everything cannot be secured at the same time, when lots of attacks are propagating day by day with different intention of attacks [27]. In this research work the used technique is providing the specific attack trends in a machine digestible form, i.e., CSV (Comma Separated Value). For researchers, finding attack detection methodology and traditional prevention, "intelligent dataset attributes can be very useful". But nowadays, the dataset attributes provided by other standard organizations is not so fruitful as there is lack of research environment, privacy issues, and specific types of attack data available in industry or any other reason [28]. Taking this as a problem, our work starts from capturing the attacks from various vulnerable honeypot sensors deployed, refine the data using various technologies such as Snort, Wireshark, Sandbox, etc., along with our knowledge, further making the repository (in Elasticsearch, Logstash and Kibana) of attack data and at last result come up with dataset attributes of specific attacks in a standard format [29, 30]. In future work the researchers can expand the capturing strength to capture more of Ransomware and analysis strength to analyze more of Ransomware and also provide data to the researchers and academician's, with a challenge to clean the nation from Ransomware.

References

1. Masato Terada: Work on Cyber Security Measures for Collaboration between Organizations: Vol. 65, No. 1 in 2016.
2. Ashima Rattan, Navroop Kaur, Saurabh Chamotra and Shashi Bhusan: Attack Data Usability and Challenges in its Capturing and Sharing In the 3rd International Conference on Cyber Security (ICCS-2017) at Rajasthan Technical University Kota (Rajasthan), Published in "International Journal of Advanced Studies in Computer Science and Engineering" (IJASCSE): Vol-6-theme-based-issue-9.
3. http://www.icasi.org/cvrf/.
4. Vijay Varadharajan: On Malware Characterization and Attack Classification: Proceedings of the First Australasian Web Conference (AWC '13), Vol. 144, 43–47 in 2013.
5. Ashima Rattan and Shashi Bushan: IP Reputation Engine Based upon Malicious Events In the proceedings of the 11th INDIACom 2017 in the IEEE 4th International conference on "Computing for Sustainable Global Development", March 2017.
6. Sean Barnum: Standardizing Cyber Threat Intelligence Information with the Structured Threat Information expression (STIX™): Version-1.1, Revision-1 in Feb 20, 2014.
7. Panos Kampanakis: Security automation and threat information-sharing options: co-publish by the IEEE computer and reliability societies: Vol. 12, Issue-5, 42–51 in September/October 2014.
8. Kutub Thakur Meikang Qiu Keke Gai and Md Liakat Ali: An Investigation on Cyber Security Threats and Security Models in the IEEE 2nd International Conference on Cyber Security and Cloud Computing: 978-1-4673-9300-3/15, pp. 307–311, 2015.
9. Komal K. More and Prof. Pramod B. Gosavi: A Real Time System for Denial of service Attack Detection Based on Multivariate Correlation Analysis Approach in IEEE International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT): 978-1-4673-9939-5/16/, pp. 1125–1131, 2016.
10. Saoreen Rahman, Muhammad Ahmed and M. Shamim Kaiser: ANFIS Based Cyber Physical Attack Detection System in IEEE 5th International Conference on Informatics, Electronics and Vision (ICIEV): 978-1-5090-1269-5/16/, pp. 944–948, 2016.
11. D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, and N. Weaver: Inside the slammer worm: In Proceedings of IEEE Security and Privacy: Vol. 1, Issue: 4, 33–39 in June 2003.
12. Dhanashri Ashok Bhosale and Vanita Manikrao Mane: Comparative Study and Analysis of Network Intrusion Detection Tools: International Conference on Applied and Theoretical Computing and Communication Technology (ICATCCT), 312–315, in 2015.
13. Ulrik Franke, Joel Brynielsson: Cyber situational awareness A systematic review of the literature, Computers & Security, Volume 46, Pages 18–31 in October 2014.
14. Guodong Zhao, Ke Xu, Lei Xu, and Bo Wu; "Detecting APT Malware Infections Based on Malicious DNS and Traffic Analysis", IEEE 20 July 2015, pp. 1132–1142.
15. Jessica Steinberger, Anna Sperottoz, Mario Gollingy and Harald Baier: How to Exchange Security Events? Overview and Evaluation of Formats and Protocols in Biometrics and Internet Security: IEEE International Symposium on Integrated Network Management (IM2015), Darmstadt, Germany 2015.
16. M. Dacier, F. Pouget, and H. Debar: Attack processes found on the internet: NATO Research and technology symposium IST-041 "Adaptive Defence in Unclassified Networks", 19 April 2004, Toulouse, France.
17. Honeynet.org.
18. Dikshant Gupta, Suhani Singhal, Shamita Malik and Archana Singh: Network Intrusion Detection System Using various data mining techniques in IEEE International Conference on Research Advances in Integrated Navigation Systems (RAINS - 2016), April 06–07, 2016, R. L. Jalappa Institute of Technology, Doddaballapur, Bangalore, India: 978-1-4673-8819-8/16/, 2016.

19. Logrhythm Labs Embedded Expertise on Security Analysis Suite-Honeypot.
20. Sanjeev Kumar, Rakesh Sehgal and J.S. Bhatia: Hybrid Honeypot Framework for Malware Collection and Analysis in IEEE 7th International Conference on Industrial and Information Systems (ICIIS-2012), August 6–9, 2012, IIT Chennai, Published in IEEE Xplore.
21. Daniel Ramsbrock: Profiling Attacker Behavior Following SSH Compromises: Department of Computer Science University of Maryland, College Park in 2007.
22. Eric Ziegast, Paul Vixie: Domain Name Service Based block List in 1997.
23. CERT Polska and European Union Agency for Network and Information Security (ENISA) team: Standards and tools for exchange and processing of actionable information in November 2014.
24. Nazmul Shahadat, Imam Hossain, Anisur Rohman and Nawshi Matin: Experimental Analysis of Data Mining Application for Intrusion Detection with Feature reduction in International Conference on Electrical, Computer and Communication Engineering (ECCE), February 16–18, 2017, Cox's Bazar, Bangladesh, pp. 209–216.
25. Zhang, Xiaoming, and Guang Wang. "Hadoop-Based System Design for Website Intrusion Detection and Analysis." 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity), IEEE, 2015.
26. El Mostapha Chakir, Mohamed Moughit And Youness Idrissi Khamlichi: An Efficient Method for Evaluating Alerts of Intrusion Detection Systems in the conference of IEEE 978-1-5090-6681-0/17/ in 2017.
27. V. Yegneswaran, P. Barford, and D. Plonka: Design and use of internet sinks for network abuse monitoring: Lecture Notes in Computer Science book series (LNCS), Vol. 3224, Springer, Berlin, Heidelberg 2004.
28. Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani: A Detailed Analysis of the KDD CUP 99 Data Set in the conference of IEEE in 2009.
29. Mike Schiffman: Cisco Systems on The Common Vulnerability Reporting Framework An Internet Consortium for Advancement of Security on the Internet (ICASI) Whitepaper in 2011.
30. Abdul Razzaq, Ali Hur, H Farooq Ahmad, Muddassar Masood: Cyber Security: Threats, Reasons, Challenges, Methodologies and State of the Art Solutions for Industrial Applications 2013 in the conference of IEEE Eleventh International Symposium on Autonomous Decentralized Systems (ISADS), 1–6. 2013.