The Design of Cyber Threat Hunting Games: A Case Study

Md Nazmus Sakib Miazi*, Mir Mehedi A. Pritom†, Mohamed Shehab‡, Bill Chu§, and Jinpeng Wei¶
Department of Software and Information Systems
College of Computing and Informatics
University of North Carolina at Charlotte
Charlotte, NC, USA
Email: *mmiazi@uncc.edu, †mpritom@uncc.edu, ‡mshehab@uncc.edu §billchu@uncc.edu, ¶jwei8@uncc.edu

***Abstract*—Cyber Threat Hunting is an emerging cyber security activity. Recent studies show that, although similar actions like threat hunting are being actively practiced in some organization, security administrator and policy makers are far from being satisfied with their effectiveness. Most security professionals lack expertise in data analytics while most people with data analytics skills lack security knowledge. To understand the necessity of threat hunting education at university level, we organized a *Threat Hunting Competition* on campus with generated logs. In this paper, we identify skills needed for cyber threat hunting, describe the data generation process as well as the usage of logs to teach threat hunting at universities.**

I. INTRODUCTION

Cyber Threat hunting has recently emerged as a necessary activity for Cybersecurity [1], [2], [3]. Threat hunting is focused on finding threats and anomalies within the organization's networks and systems with monitoring and analyzing logs promptly both by automation and human analysis. Figure 1 shows the threat hunting process where Cyber Threat Intelligence sources, various logs, alerts from traditional IDS and firewalls works as inputs and verified intrusions are the output of the process. In many large enterprises, there are active security personnel or team deployed for threat hunting. However, hunting threats requires skills that are not part of existing cybersecurity education curriculum. One needs the right sets of analytical skills to search and probe for anomalies in a variety of datasets. Thus, we need to develop new education materials to prepare students with the necessary analytical skills for Threat Hunting.

SANS Institute's survey of 494 IT professionals conducted in April 2016 demonstrates the importance of threat hunting and its future [4]. About 86% of the respondents said affirmatively that their organization is involved in these activities while Around 75% responds that they had successfully minimized their attack surface as a consequence of threat hunting. Moreover, 59% attributed this method for faster and accurate incident response[5]. Though threat hunting is becoming common phenomena in organizational security, nearly 56% of organizations are still not satisfied, and about 22% are not sure about their hunting programs [4]. Moreover, traditional IDSs generate lots of false alarms while detecting intrusions [6], [7]. So, we need the proper analysis of those alarms with experienced analysts to make sure we do not miss any anomalies.

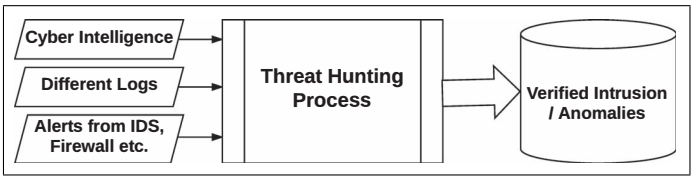

Fig. 1. Threat Hunting Process

In existing security courses at universities, learning is focused on understanding security technologies and applying tools to specific security problems such as cyber defense and digital forensics. Little effort is given to developing skills for students to analyze a wide variety of data and reach security conclusions. Experiences have clearly shown that attackers, both insiders as well as outsiders, can circumvent available cyber defense mechanisms [3]. They have proven to be able to evade security monitoring by disguising malicious actions as normal activities [3].

This paper describes our experiences in designing a threat hunting exercise. The remainder of this article is organized as follows. Section II discuss other available cyber security activities and how threat hunting can complement them. In Section III, we describe the setting for our threat hunting exercise and its objectives with use case scenario. Section IV presents the data generation process and our assumptions. Finally, in section V, we conclude this work with possible future directions.

II. THREAT HUNTING SKILLS

Successful threat hunting requires a broad set of technical skills and proper mindset [8]. At first, we need threat intelligence, which includes, gathering intelligence reports from various sources. The second skill is target and actor centric mindsets and well understanding of F3EAD system [9] for incident detection and analysis. Next skill is security data analysis that requires a thorough understanding of Data Science and several analytical models. Several other skills are also needed such as, Forensic analysis, malicious code

978-1-5090-2991-4/17/$31.00 ©2017 IEEE

analysis, penetration testing, and vulnerability analysis. Besides, vulnerability analysis includes attack vector analysis and network communication methods analysis.

Threat hunting differs from many traditional cyber security activities. It is a highly unstructured task that demands deep technical know-how, data analytics savvy, and out of the box thinking. We contrast threat hunting with popular Cybersecurity activities to illustrate its unique requirements.

Threat Hunting vs. Threat Detection: Threat hunting is based on finding threats at a stage when there are no specific signs of compromise. While threat detection [10], in general, is a process of identifying threats based on a structured process. Threat detection tools usually analyze network logs, application logs, data logs, and user activity logs for finding anomalies [11]. Traditional threat detection relies on sensors to detect threats and generate alarms, where threat hunting emphasizes on the proactive discovery of threats by involving skilled human for analyzing the dataset. A a successful hunt may codify a process for future threat detection.

Threat hunting vs. Cyber Defense: Cyber defense [12], [13] focuses on hardening systems against attacks. Cyber defense activities typically include security configurations and the use of cyber defense technology such as firewalls, intrusion detections and intrusion prevention systems. However, attackers can often evade cyber defense mechanisms by using new techniques unknown to cyber defenders at the time of the attack. Threat hunting can inform cyber defense once mechanisms used by attackers are found beforehand.

Threat hunting vs. Penetration Testing: Penetration testing [14], [15], [16] is concerned with finding vulnerabilities, either in software or system configurations. Usually, penetration testing is accomplished by doing static and dynamic code analysis of the associated software or system. In contrast, threat hunting aims at finding existing intruders in the system by analyzing logs, who may have penetrated the system before vulnerabilities are discovered.

Threat hunting vs. Forensics: Both activities involve data analysis. However, forensics [17], [18] start with a known intrusion and try to work backward to understand how the threat got there, collect evidence and assess the impact caused by the threat. Forensic is a useful procedure for taking legal actions for known cyber crimes. Threat hunting, on the other hand, is focused on finding unknown intrusions. Once an intrusion is found, forensics will be applied.

Threat hunting vs. Cyber Intelligence: Cyber intelligence [19], [20] is one possible input to the threat hunting process. For example, if the threat of a particular trojan is known, threat hunting may be focused on finding if the trojan has already gained food hold, using available cyber intelligence on the tactics and indicators of the Trojan during the hunting process.

Threat hunting vs. IDS: Broadly speaking, both are concerned with looking for anomalies, but IDS typically looks for network anomalies instead of authentication, authorization, access control anomalies in an organization [21]. On the contrary, threat hunting is not limited to network events and threats [22]. There are other kinds of threat hunting data such as process logs, system call logs, employee's authentication logs, employee's activity logs, DNS queries, and so on. Threat hunting often relies on analyzing correlations between events.

Fig. 2. C0mp@ny's Office Locations

According to SANS survey on Threat Hunting [4] most organizations practicing hunting in their environment use basic searching, statistical analysis, visualization techniques, aggregation, machine learning and Bayesian probability. However, less sophisticated strategy like searching is the most popular one. While more sophisticated techniques like, machine learning and advanced statistical analysis are not widely practiced among the security community. We find one of the important reasons behind this trend is the lack of skills required to apply such advanced analytics in the security field.

III. A CASE STUDY OF THREAT HUNTING

For the rest of the paper, we describe a threat hunting game we designed as a case study. We demonstrate a use case scenario in this section. In the next section we describe data generation process. This game can be used either in a competition or as a large course project in a class. Participants are given a number of log data sets. They are asked to identify threats based on the data. We conclude this study by presenting student experiences in the competition.

A. A Real Life Scenario

In order to understand the design of a hunting game better, we need to describe a real life scenario. In our case, we created employee authentication log data for a company named C0mp@ny. C0mp@ny is a medium sized company with their headquarter located in Charlotte, North Carolina, USA. It has its offshore offices in Paris, London and Luxembourg, as shown in Figure 2. The Charlotte office employs around 100 employees. The company has four departments: Human Resource (HR), Research, Information Technology (IT), and Finance. On every working day each employee logs onto their office machine. Employees can log on to their account either from home or office using the proper credentials and a secure connection. They can access documents shared with them or documents they have been given authorized access. They can use devices (printer, fax, telephone etc.) and other company resources available to them. After working hours, they need to

log out of the machines. The system keeps the logs of login-logout times, actions performed, accessed devices, and GPS co-ordinates from where the employee is logged on. So for a security analyst, the system log is a goldmine to hunt for mis-configurations, attacks, and malicious activities.

B. Normal Behavior of People

To create a system log dataset, we need to think of general behaviors of employees in an office. At first, we looked at the routine of an employee in an office. A typical routine would be:

- Start working at morning (8:00AM - 9:00AM)
- Work for three to four hours
- Take a lunch break at afternoon (12:00PM - 1:00PM)
- Log back in an hour
- Work for two to three hours
- Take a break for coffee (2:45PM - 3:15PM)
- Log back in half an hour
- Work for a couple of hours
- Log out for the day (5:00PM - 6.00PM)

So, most of the employees will follow the same routine. However, some people may wake up late and start working late in the morning. They sign off from work late, maintaining 8 hours a day routine. A few of them may work on weekends too. In general, there are three types of employees based on the workplace: work at the office, work from home and, travel for work. The employees who work at office logs on to the office machines, hence the location remains the same as the office location. Sometimes, during lunch somewhere out of the office, an employee can log onto their machines. But, the most typical scenario is that an employee generally remains within an area close to their office. For the employees who work from home, the case is pretty similar. In general, an employee working from home stays at home, hence the location remains the same as their home location. But, sometimes they may go outside for doing several household chores, so their location remains within an area close to their home. For better understanding, we can encircle two areas surrounding the office and the employee's home having a certain radius. We denote them as *Working Circles*, see Fig. 3. Logging on from any place within these two circles denotes a regular phenomenon, not an anomaly. When an employee travels to any of the offshore offices, the scenario remains the same as office location changes to the offshore office location and home location changes to the hotel location where the employee is staying.

In addition, we needed to think of some other normal behaviors. In general, from an employee's personal machine there should not be any transfer of continuous bulk data packets instead there should be small bursts of packets to transfer regular documents. Sometimes an employee can be on leave for health issues. Moreover, an employee can only access the files inside their department and files he/she has access to.

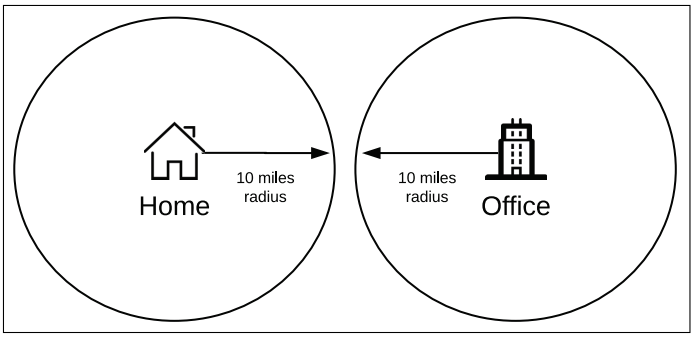

Fig. 3. Working Circles having radius of 10 miles

C. Anomaly Detection

Anomaly detection depends on the understanding of the normal behavior of the people. In our example, there are abnormal behaviors in authentication logs indicating probable malicious activities. We use example scenarios below to illustrate anomaly detection clearly.

Mike is an employee of @companyName and has an excellent working records. He maintains a consistent routine in his works. From the system logs, it has been found that he has been logging on to his laptop more than ten miles away from his office. It began three months ago and became a regular phenomenon. It indicates something interesting going on with Mike. He might be secretly meeting with people from rival companies and selling the confidential data out. Roman, one of Mike's colleagues works at the regular working hours. System logs show that, recently he has been logging on to the office workstations very late at night. This could indicate that he is doing something beyond normal activities. Cathy, who traveled to London last year attending a company meeting transferred bulk sized data packets for a suspiciously long time to an unknown IP. It could be an indication of potential anomaly. On the other hand, Bob, an employee with an excellent record was found logging on from an IP located in Eastern Europe. It denotes that Bob's login credentials may have been compromised. Lewis, a former employee, who has been fired recently, should not have records from his/her machine in the system logs. Surprisingly, the records show that there has been login attempts from his account. This is very suspicious and could mean two things: either his credentials got compromised, or he is trying to login with a bad intention. These are some of the example indications of anomalous behaviors in the dataset.

We next discuss techniques to identify common types of anomalies. First, we discuss the anomaly of locations. Any activity outside the Working Circles should be suspicious. This might be a coincident, but the recurrence of such phenomena should be investigated. Secondly, we talk about the probability of an employee being sick. An employee perhaps can take a leave for sickness, but being absent without informing the office should be considered as an anomaly. Also, we can think of suspicious activity of accessing unauthorized files. For example, an employee from the HR department tries to access an unauthorized file from IT department. Indeed, he/she cannot

open the file, but we can get the evidence from the system log. The next abnormal activity is trying to log on from an account repeatedly with different passwords. It is a common phenomenon that an employee wants to log on to his machine and misspelled his password couple of time. However, in general, he does not need to type in his password too many times. So, if the system log says that he tries several times repeatedly with different passwords, it means that there might be another person who is trying to hack into his machine. Moreover, there can be system logs that include records related to previous employees. Also, there can be suspicious activity from unknown IPs from countries other than the countries where the company has offices.

IV. DATA GENERATION PROCESS

In this section, we describe how we generated artifical, yet realistic dataset for the purpose of practicing threat hunting, authentication logs for C0mp@ny [23]. First, we focus on generating data with all normal activities. Then we inject the abnormal data into the logs. The abnormal data points are the target for threat hunters to discover.

A. Assumptions

The key assumptions for the log generation are listed below-

- Typical work hour duration for the employees are more or less 8 hours a day. We described a typical routine of an employee in section III. We consider that some employees start working late and work late at night.
- An employee can log in from either office or home. Normally, an employee will not login to office account so far away from the office or his home. If she is traveling offshore, she should not log in so far away from her living place or office. We assume that the radius of her Work Circle should not exceed 10 miles.
- An employee can be absent for several days for sickness.
- An employee should not try to access the files from the other departments.
- He/she does not attempt to log in several times with different passwords. He/she may misspell the password on a few occasions.
- A former employee should not try to login after he/she left the company.
- An employee transfers short bursts of data packets, not a long burst of data for a long time.

B. Generating Normal Behavioral Dataset

We create a basic structure of C0mp@ny. We have already known that it has its head quarters at Charlotte with three off-shore offices in Paris, London and, Luxembourg respectively. We generate 4 locations for the respective offices. We generate personal details for 100 employees with their employee ID, name, department they work in, joining date, resigning date (if applicable) and, location of home. We divide 100 employees among the four departments. Then, we assign several resources like files, printers, servers etc to each of the departments.

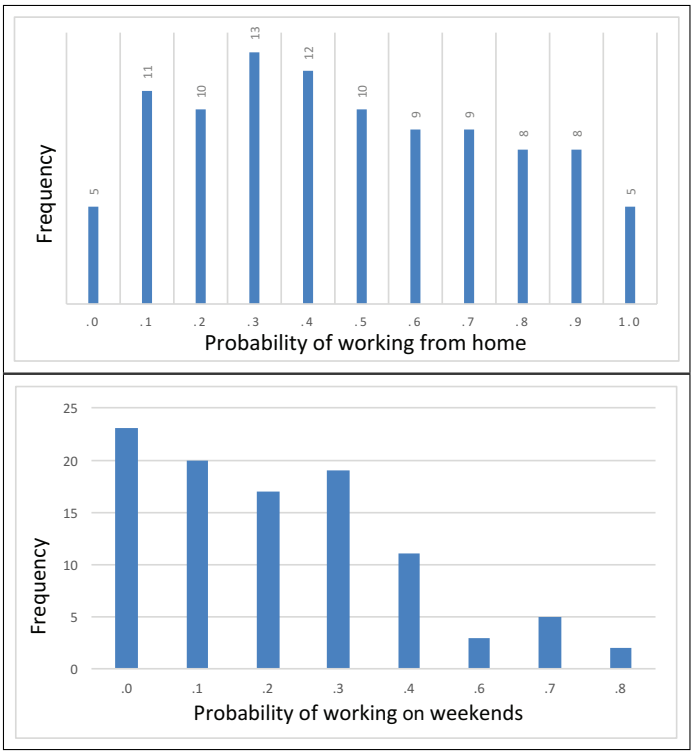

Fig. 4. Probability distribution of the employees working from home and working on weekends

TABLE I
DISTRIBUTION OF THE PROBABILITIES GENERATED FOR DEFINING EMPLOYEE BEHAVIORS

	Probability Distribution	**Mean**	**Standard Deviation**	**Shape**
Working from home	Normal	0.5	0.287	Normal
Being Sick	Normal	0.005	0.003	Normal, Skewed to left
Working on weekends	Normal	0.22	0.205	Normal, Skewed to left
Working late	Normal	0.26	0.164	Normal, Skewed to left
Misspelling Password	Normal	0.35	0.296	Normal

Next, we generate the probabilities for each employee for defining different behaviors. The first probability metric is the probability of working from home. We have assumed that, the company is very flexible and let employees to work from home. So, for employees, the probability distribution of working from home should be a normal distribution. We create the probability of working from home for every employee having the mean 0.5 and standard deviation of 0.287. Similarly, in Table I, we present the distribution of probabilities we have used to generate the activity logs. Figure 4 illustrates two probability distributions of employees working from home and working on weekends respectively.

6/1/15	10163	Maisie	1:25	logout	AuthServer	192.168.1.49	35.228719	-80.82172 success
6/1/15	10159	Lenore	1:56	logout	AuthServer	192.168.1.25	35.175927	-80.999608 success
6/1/15	10161	Linda	8:08	login	AuthServer	192.168.1.38	35.2360225	-80.556678 success
6/1/15	10104	Amanda	8:09	login	AuthServer	192.168.1.8	35.228719	-80.82172 success
6/1/15	10124	Claire	8:12	login	AuthServer	192.168.1.21	35.228719	-80.82172 success
6/1/15	10174	Quinlan	8:21	login	AuthServer	192.168.1.60	35.228719	-80.82172 success

Fig. 5. A snapshot of the System logs file

C. Authentication and Activity Log Generation

We have used the normal behavioral data set as the input of log data generator program to generate 12-months of system authentication and activity logs for all hundred employees. The output was a 29 Megabyte CSV file composed of $333,952$ log events. Figure 5 shows a small snapshot of the data we have generated. Every entry in the generated log file contains the date, time-stamp, employee ID, employee nickname, action (e.g., login, logout, access, etc.), resource (e.g., servers, files, etc.), IP, location latitude, location longitude and the status (e.g., success, fail). We have injected $2,259$ entries in the normal behavioral dataset (approximately 0.67% of the total data) that represent anomalous data. Finally, this generated log file was given to the participants of the competition to find the out the anomalies from the data set.

V. PARTICIPANTS' EXPERIENCES IN FINDING ANOMALIES

Participants of the hunting competition analyzed the data based on several key points, such as location, access frequency, access date and time, and role. In general, they could identify the anomalous IDs related to access role, access frequency, and access date. They struggled to find anomalies regarding access time. They could determine the anomalies related to login from unknown locations, mostly remote resource access attempts.

Firstly, we can talk about the location based anomalies. To recall, we set the *Working Radius* of 10 miles from home and offices of an employee, but we did not state that in the problem scenario of the competition. As a result, nobody could identify geographically closer abnormal locations. The most prominent solution was to generate a world map of access attempts using the log. In the map, it was clearly visible that, from several places in eastern Asia, somebody was trying to access organization's resources. Participants identified those events as anomalies. In addition to the places in eastern Asia, we generated some fake anomalous data for eastern Europe too. However, from participant's analysis, they could not identify them as anomalies, as those were closer to the two office locations in Europe. The main reason behind their failure to detect anomalies is that they did not use the access and authentication log as training data. They need to find a regular pattern from the data to make a prediction of the Working Radius.

Most of them were successful to find out anomalies related to access frequency. The common sense behind this is the prediction that a person normally should not be accessing more than a certain number of resources in a very short time (e.g., 1 minute). It is evident that by counting the frequency of access requests from certain employees one could find out these anomalies. Moreover, to seek anomalies in access dates was easier for them. We have generated data in which some of the previous employees have been trying to access the resources remotely. By checking the access roles of the users, participants successfully identified these anomalies. They figured four employees who tried to access the files for which they do not have accesses.

However, identification of anomalies regarding working hours was not easy, and unfortunately, no participants in the competition were able to spot those. One interesting observation was that working hour anomalies might vary from organization to organization as it mostly depends on company policy. However, for working hours, some common assumptions can make the difference. For example, the norm for working hours in business is more or less 8 hours a day. Even if the employees of C0mp@ny do not exactly follow a certain routine like 8AM to 5 PM, they still work 8 hours a day. Hence, focusing on the duration of work could find regular routines of employees from the data set. So, it indicates that, we need more expertise in this field.

VI. POSSIBLE SOLUTION STRATEGIES

There is no specific standard solution framework to find anomalies quintessentially from raw logs till date. However, there are many possible solution techniques by which one can get to near-perfect solutions. Researchers have already applied several techniques starting from basic manual searching to statistical and/or probabilistic models [24], visualizations [25] on logs to detect anomalies. Even more sophisticated approaches such as machine learning (e.g., decision tree, neural networks, and so on.) are also taking place to uncover new threats [26], [27]. In this section, we are proposing a potential hybrid solution composed of statistical analysis, visualization, and machine learning followed by expert analysis to verify intrusions. Figure 6 shows our step-wise solution strategies that can be adopted by analysts for threat hunting from logs.

In the first step, we apply statistical analysis and visualization techniques to find out possible outliers. Next, we refine the raw data set deducting all possible outliers from the data set. In the second step, we will train our machine learning system to build up a normal behavior model for further testing. The third phase is the anomaly detection step. We use all of

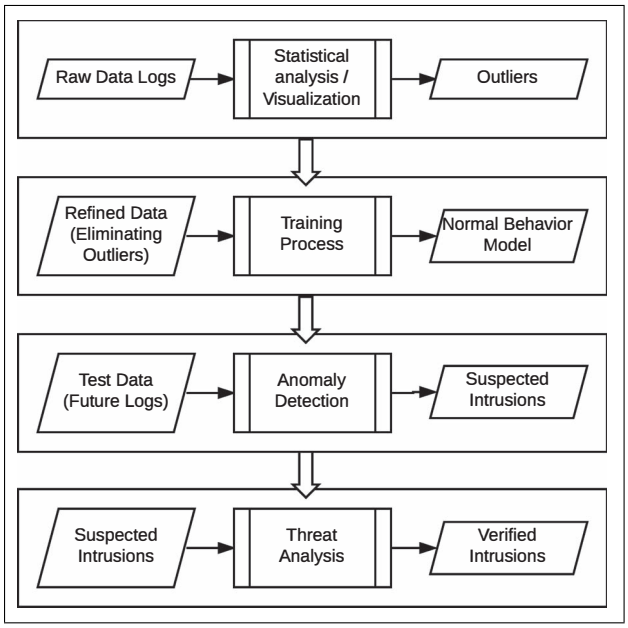

Fig. 6. Threat Hunting Solution Steps

the future or current data as inputs in our trained anomaly detection system. Based on our normal behavior model, we can find out suspected intrusions. The final stage is the threat analysis stage. As we already get a set of alleged intrusions by now, it becomes easy for the analyst to verify the actual intrusions using popular and effectual data analysis methods.

VII. CONCLUSION

To effectively carry out cyber hunting, one must have a variety of skills such as working with large data sets, techniques for data analysis, correlations among different types of data, and a thorough understanding of cyber security fundamentals. Current cyber security curriculums in universities do not prepare students for cyber hunting. The lack of realistic practice data sets is one of the challenging issues to develop these skill sets for future students in current security academia. We believe this competition can act as a preliminary beginning to practice threat hunting in the academic community. Moreover, cyber security educators can develop similar enterprise standard validated logs (e.g. process logs, NetFlow, DNS logs, and so on.) to minimize the dataset challenges faced by future analysts.

ACKNOWLEDGEMENT

The authors would like to thank the University of North Carolina at Charlotte for organizing Cyber Threat Hunting Competition 2016 at the campus which creates the opportunity to develop such dataset.

REFERENCES

[1] https://en.wikipedia.org/wiki/Cyber_threat_hunting. Last accessed on 28th March, 2017.

[2] https://sqrrl.com/solutions/cyber-threat-hunting/. Last accessed on 28th March, 2017.

[3] Robert M. Lee and Rob Lee. A SANS Whitepaper: The Who, What, Where, When, Why and How of Effective Threat Hunting, February 2016.

[4] Eric Cole. A SANS Survey: Threat Hunting: Open Season on the Adversary, April 2016.

[5] 'Threat Hunting' On The Rise. http://www.darkreading.com/endpoint/threat-hunting-on-the-rise/d/d-id/1325144. Last accessed on 19th March, 2017.

[6] Klaus Julisch and Marc Dacier. Mining intrusion detection alarms for actionable knowledge. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 366–375. ACM, 2002.

[7] Dingbang Xu and Peng Ning. Correlation analysis of intrusion alerts. *Intrusion Detection Systems*, 38:65–92, 2008.

[8] https://sroberts.github.io/2015/04/14/ir-is-dead-long-live-ir/. Last accessed on 11th April, 2017.

[9] Charles Faint and Michael Harris. F3ead: Ops/intel fusion feeds the sof targeting process. *Journal Article— Jan*, 31(6):54pm, 2012.

[10] Kenny Tidwell, Kumar Saurabh, Debabrata Dash, Hugh S Njemanze, and Pravin S Kothari. Threat detection in a network security system, August 21 2007. US Patent 7,260,844.

[11] What is Threat Hunting? The Emerging Focus in Threat Detection. https://digitalguardian.com/blog/what-threat-hunting-emerging-focus-threat-detection. Last accessed on 20th March, 2017.

[12] Keith A Repik. Defeating adversary network intelligence efforts with active cyber defense techniques. Technical report, DTIC Document, 2008.

[13] Ruzena Bajcsy, Terry Benzel, Matt Bishop, B Braden, C Brodley, Sonia Fahmy, Sally Floyd, W Hardaker, A Joseph, George Kesidis, et al. Cyber defense technology networking and evaluation. *Communications of the ACM*, 47(3):58–61, 2004.

[14] Pen testing vs. threat hunting: Understanding the differences. https://www.cybereason.com/blog-pen-testing-vs-threat-hunting-understanding-the-differences/. Last accessed on 28th March, 2017.

[15] Bruce Potter and Gary McGraw. Software security testing. *IEEE Security & Privacy*, 2(5):81–85, 2004.

[16] Matt Bishop. About penetration testing. *IEEE Security & Privacy*, 5(6), 2007.

[17] Vicka Corey, Charles Peterman, Sybil Shearin, Michael S Greenberg, and James Van Bokkelen. Network forensics analysis. *IEEE Internet Computing*, 6(6):60–66, 2002.

[18] Harold F Tipton and Micki Krause Nozaki. *Information Security Management Handbook, Volume 6*. Auerbach Publications, 2012.

[19] Sarah Brown, Joep Gommers, and Oscar Serrano. From cyber security information sharing to threat management. In *Proceedings of the 2nd ACM Workshop on Information Sharing and Collaborative Security*, pages 43–49. ACM, 2015.

[20] Charles Edwards, Samuel Migues, Roger Nebel, and Daniel Owen. System and method of data collection, processing, analysis, and annotation for monitoring cyber-threats and the notification thereof to subscribers, September 13 2001. US Patent App. 09/950,820.

[21] Roberto Di Pietro and Luigi V Mancini. *Intrusion detection systems*, volume 38. Springer Science & Business Media, 2008.

[22] How threat hunting is different from an intrusion detection system. https://www.cybereason.com/how-threat-hunting-is-different-from-an-intrusion-detection-system/. Last accessed on 28th March, 2017.

[23] 2016 cyber hunting competition. http://cybersecurity.uncc.edu/cyber-hunting-competition. Last accessed on 29th March, 2017.

[24] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15, 2009.

[25] Hossein Siadati, Bahador Saket, and Nasir Memon. Detecting malicious logins in enterprise networks using visualization. In *Visualization for Cyber Security (VizSec), 2016 IEEE Symposium on*, pages 1–8. IEEE, 2016.

[26] Robin Sommer and Vern Paxson. Outside the closed world: On using machine learning for network intrusion detection. In *Security and Privacy (SP), 2010 IEEE Symposium on*, pages 305–316. IEEE, 2010.

[27] Eleazar Eskin. Anomaly detection over noisy data using learned probability distributions. In *In Proceedings of the International Conference on Machine Learning*. Citeseer, 2000.