

PhD Thesis Proposal

**Artificial Intelligence in Cyber
Security of Cloud Based Systems:
Detection, Repair and Defense**

by

Aytac Ozkan

Supervisor:
Co-Supervisor:

August, 2019

Abstract

The speed of process and the amount of data to be used in defending the cyber space cannot be handled by humans without considerable automation at the cloud base systems. However, it is difficult to develop software with conventional fixed algorithms (hard-wired logic on decision making level) for effectively defending against the dynamically evolving attacks in networks. This situation can be handled by applying methods of artificial intelligence that provide flexibility and learning capability to software.

Computer Security system providers are unable to provide timely security update. Most security systems are not designed to be adaptive to the increasing number of new threats. Companies lose considerable amount of time and resources when security attacks manifest themselves.

As a answer to these problems this research is aimed at developing security systems capable of learning and updating themselves.

The goal is to create security systems that will autonomously mature with expose to threats over time. To achieve this goal this research is proposing artificial intelligence based security systems with learning capability to perform intrusion, detection, repair and agnostics and defending network.

Contents

Abstract	iii
Contents	v
1 Introduction	1
1.1 AI: The next Frontier in IT Security	1
1.2 The Contribution of Intelligent Systems to Cyber Security	2
1.3 Literature review	4
1.4 Research Plan	5
Bibliography	7

Chapter 1

Introduction

In recent years, a large amount of research addressing the contribution of intelligent systems to cyber security has been conducted. One can hear more and more about artificial intelligence (AI), machine learning, and deep learning systems in cyber security. As often happens with buzz words, nonexperts tend to use them casually and loosely. Despite the fact that the borderlines between such terms are not always clear and these borderlines become even fuzzier with new practical and theoretical developments, it is important to try to characterize the focus of these systems in the field.

1.1 AI: The next Frontier in IT Security

AI in cybersecurity is a set of capabilities that allows organizations to detect, predict and respond to cyber threats in real-time using machine and deep learning.

AI-enabled cybersecurity is increasingly necessary, organizations face an urgent need to continually ramp up and improve their cybersecurity.

This is because the number of end user devices, networks and user interfaces continues to grow as a result of advances in cloud, the IoT, 5G and conversational interfaces.

The increases the difficulty involved with administering a computer network. Having artificially intelligent network infrastructure that can learn to detect, report and repair network security problems is an advantage to network administrators.

Adaptability is another key reason that brought AI and computer security close to each other. The malicious entities that generate computer security attacks have gained intelligence. The type of attacks evolved from a simple password guessing to a staged and distributed network attack that can be equipped with a stealth mode and attack that mutation. The compute security field has to adopt to rapid change and the evolution of security attacks. AI is well situated to answer this situation. AI is concerned with creating systems that evolve and react depending on the changes in the surrounding environment.

Fighting the Unknown. For many years, the efforts of the IT security industry were based on the assumptions that attacks were conducted on a large scale and major threats spread before arriving at a specific network or computer. Solutions were created on the basis of these assumptions and addressed as follows. A typical threat was identified as a virus, which was investigated by IT security company labs that identi-

fied the virus signature and sent this information to endpoints installed with leading antivirus software.

Internet of Things In the past, most Internet interactions were interactions between humans that communicated via different platforms over the global network. Communications between humans, as the primary users of online communications will change dramatically with the evolution of the IoT realm [1].

Most of the entities that will communicate on the Web in the future will be machines, and they will be used to initiate reports, make contact, or respond to requests. [2] The means to facilitate this is characterized by easy access to the Web and the ability for mass distribution. For example, there is a possibility to turn households into smart homes with intercommunicative technology instead of investing in designing a supporting infrastructure. Even scattering sensors along oil fields is the result of technological developments that led to a decrease in the price of materials, easy logistical operation, and resistant products that meet industrial standards. [3]

The cyber world has barely begun to focus on new types of dangers such as the autonomous world, since this process has just started. However, the severity level of the threat is clear. In coming years, the cyber defense community will be called upon to develop new and effective technologies to meet the challenges posed by our changing and increasingly autonomous world.

1.2 The Contribution of Intelligent Systems to Cyber Security

Artificial intelligence (AI), which was officially established in the late 1950s by scholars such as Marvin Minsky, is used as a generic name representing a wide variety of methods, tools, and techniques that mimic “cognitive” functions or tasks that people associate with the human mind, such as “learning,” “planning,” “reasoning,” or “problem solving” [4]. The importance of AI in cyber security is twofold and related to two opposing directions. The first direction focuses on AI-controlled systems as potential targets of cyber attacks, mainly due to their increasing role in controlling vital and complex systems. There are numerous examples of cyber risks related to AI-controlled systems, such as smart vehicles [5], smart grids and smart cities [6].

Knowledge-based systems and machine learning methods are two well-known classes of AI methods that contain valuable tools used in cyber security. In Knowledge-based systems, a huge amount of (experts’) Knowledge is uploaded to the computer memory (thus, it is also called expert systems). [7] The learning part in these systems is based on the reasoning related to this large body of knowledge, which is often obtained by programmed rules (such as “if-then” and “inference logic rules”).

Antivirus and antispam software packages [8] represent straightforward implementations of expert systems in cyber security. In this case, expert knowledge (accumulated based on a large amount of transactions) regarding the procedures used—such as applied protocols, network traffic (e.g., HTTP, HTTPS, VoIP, or email), and I/O interactions with the operating system—is organized systematically to protect the users from cyber breaches. There are several papers in this issue that serve as good examples of such expert systems [9].

Machine-learning methods, unlike knowledge-based systems, usually refer to applications in which the learning component is performed by the computers “themselves.” This is often done by extracting relevant patterns from the data and using them to derive predictions and smart recommendations. In other words, machine-learning algorithms are improved “automatically” through experience with the data, while giving the computers “the ability to learn without being explicitly programmed,” as suggested long before the field of cyber security was formally established [10]. There are many examples of cyber-security tasks that can be addressed by machine learning, including user monitoring, spam filtering, zero-day attack identification, risk analysis, and many more.

It is well known that machine-learning methods are divided into two main classes (and a hybrid class of methods that combines the two). The first class contains unsupervised learning methods, in which untagged data samples are introduced to the system in order to find significant patterns. Fraud detection in financial systems, anomaly detection in communication protocols, and segmentation of both users and software packages according to their risk potential are good examples of unsupervised machine-learning methods that apply techniques, such as anomaly detection and clustering, to identify both “positive” or “negative” deviations from the norm. These deviations are then mapped into actions that include, for example, risk assessment of new software packages, blacklisted websites, or blocking Internet connections with high rates of suspicious users. Two papers published in this issue are introduced in this section, which serve as good examples of unsupervised machine-learning methods [9] [11].

The second class of methods belongs to supervised learning, in which the data samples that are introduced to the system are tagged a priori. In other words, the sample data inputs are coupled with their desired outputs (thus the term supervised). The goal is to learn a general rule that maps inputs to outputs. Some examples from the cyber security domain are users’ risk scores [12], for which descriptive features of users—such as the communication volume, time, and the type of interaction—are tagged either as “risky” or “nonrisky” and are then learned by the system to predict high-risk users in advance [13].

Within the class of supervised-learning methods (that contain many other tools, such as decision trees, support vector machines, and regression models), there is a unique group of artificial neural networks, which are models that were inspired by the structure and functional aspects of biological neural networks in the human brain. Layers of nodes (“neurons”) are connected to each other via weighted edges that are fine-tuned by mapping inputs to tagged outputs. These models are often used to represent complex (nonlinear) relationships between inputs and outputs; they were considered to be very successful and got another boost recently from the deep-learning revolution.

Deep learning models initially emerged from this subset of methods, consisting of neural networks with multiple processing layers. These highly complex models require both high-speed processing units and a large amount of data, which became more available with the development of big-data technologies and cyber-security techniques. Deep-learning models also evolved to address unsupervised learning and were found to be extremely successful in signal processing (for which a lot of tagged data exists), specifically in image processing, which supports computer vision, and in speech recognition [14] [15], for which it has the primary task of identifying the user and providing

the user with the correct authorization level. These models gained quite a bit of attention recently when leading global companies decided to invest significantly in these models in order to develop new applications; examples include Apple (e.g., Siri), Alphabet/Google (e.g., self-driving cars), and Facebook (automated image processing). These models have been shown to generate excellent results with specific tasks, yet this performance is not always guaranteed, especially in noncontinuity cases [16]. Moreover, these models are not as descriptive as other analytic models, such as closed-form regression models, decision trees, or graphical networks. Therefore, their outputs pose challenges in terms of their interpretation or intuitive understanding by cyber security analysts (that play a vital role in cyber defense) [17] and by other decision makers (e.g., the Chief Information Officer) in cyber security. Thus, deep learning is a subset of machine learning, which is a central branch of AI. Deep learning should be viewed as an important tool in the cyber security toolkit, specifically when the analytic tasks involved require modeling a large amount of data by complex, often nonlinear, relations between the system's input and output.

To summarize, we believe that the role of intelligent systems in cyber security will continue to grow, while the development of these systems, like any other scientific development, poses both negative and positive effects on cyber security.

1.3 Literature review

There have been several similar works done in IoT fields. Still, researchers are working in this area. Pahl et al. [18] have mainly developed a detector and firewall for an anomaly of IoT microservices in IoT site. Clustering methods like K-Means and BIRCH have been implemented [19] for different microservices in this work. In clustering, different clusters were grouped in the same if the center is in the three times of standard deviation distance. The clustering model has been updated using an online learning technique. With the algorithms implemented, the overall accuracy obtained by the system is 96.3%. A detailed description of a smart home system where security breaches were detected by deep learning method Dense Random Neural Network (DRNN) [20] have been introduced in [20]. They have mainly described Denial of Service attack and Denial of Sleep attack in a simple IoT site.

Liu et al. [21] proposed a detector for On and Off attack by a malicious network node in industrial IoT site. By On and Off attack they meant that IoT network could be attacked by a malicious node when it is in an active state or On state. Furthermore, the IoT network behaves normal when its malicious node is in the inactive or off state. The system was developed using a light probe routing mechanism with the calculation of trust estimation of each neighbor node for the detection of an anomaly. Diro et al. [22] discussed the detection of attack using fog-to-things architecture. The authors of the paper gave a comparison study of a deep and a shallow neural network using open source dataset. This work's primary focus was to detect four classes of attack and anomaly. For four class the system got the accuracy of 98.27% for deep neural network model and accuracy of 96.75% for shallow neural network model. Usmonov et al. [23] described the recent security problem when developing embedded technologies for the IoT. Preserving data transfer between the physical, logical and virtual components of

the IoT system was also challenging. For these problems, the use of digital watermarks was proposed by the authors of this work. Ukil et al. [24] discussed the detection of anomalies in healthcare analytics based on IoT. A model of cardiac anomaly detection through a smartphone was also introduced in this paper. For the anomaly detection in healthcare; IoT sensors, medical image analysis, biomedical signal analysis, big data mining, and predictive analytics were used. Pajouh et al. [25] presented a model for intrusion detection based on a two-layer dimension reduction and two-tier classification module. This model was also designed to identify malicious activities such as User to Root (U2R) and Remote to Local (R2L) attacks. For dimension reduction, component analysis and linear discriminate analysis have been used. NSL-KDD dataset was used to carry out the whole experiment. For detecting suspicious behaviors with the two-tier classification module, Naive Bayes and Certainty Factor version of K-Nearest Neighbor were applied. D'Angelo et al. [15] applied Uncertainty-managing Batch Relevance-based Artificial Intelligence (U-BRAIN) on binary NSL-KDD dataset and Real Traffic Data (from Fredrico II University of Napoli). The U-Brain is a dynamic model operated on multiple machines which can handle missing data. The NSL-KDD dataset contains 41 features. From 41 features 6 features

1.4 Research Plan

As an aims of the research and We provided a concrete plan for completion of the research including the design and methods.

- Analyzing previous AI solutions which is developed for cybersecurity and figure out the deficiencies of the application.
- Investigating "exploit databases" and collecting data related the kind of the threats.
- Analyzing Cloud Based Systems' network infrastructures for collecting data.
- Researching common and recent threats for the computer networks.
- Investigating the literature of the machine learning algorithms.
- Analyzing the vulnerabilities of the cloud systems.
- Defining the constraints of the problem.
- Modeling the problem.
- Creating the fittest machine learning models for AI application
- Developing the AI application that can detect, repair and defense the cloud based systems to against the cyber threats and attacks.
- Creating real-world test environment and creating real-time cyber-attacks scenarios for test the developed AI supported security applications.

- Regarding the test results of the application, optimize the AI algorithms with heuristic and metaheuristic approaches.

Bibliography

- [1] I.-Y. Ko, H.-G. Ko, A. J. Molina, and J.-H. Kwon, *SoIoT: Toward A User-Centric IoT-Based Service Framework*, ACM Trans. Internet Technol. **16**, 8:1–8:21 (2016).
- [2] N. Komninos, M. Pallot, and H. Schaffers, *Special Issue on Smart Cities and the Future Internet in Europe*, Journal of the Knowledge Economy **4** (2012).
- [3] X. Chen, D. Zhang, L. Wang, N. Jia, Z. Kang, Y. Zhang, and S. Hu, *Design Automation for Interwell Connectivity Estimation in Petroleum Cyber-Physical Systems*, Trans. Comp.-Aided Des. Integ. Cir. Sys. **36**, 255–264 (2017).
- [4] S. Russell and P. Norvig, *Artificial Intelligence - A modern Approach*, Prentice Hall International Editions, Upper Saddle River, NJ (1995).
- [5] L. T. Berger, A. Schwager, and J. J. Escudero-Garz s, *Power Line Communications for Smart Grid Applications*, JECE **2013**, 3:3–3:3 (2013).
- [6] A. AlDairi and L. Tawalbeh, *Cyber Security Attacks on Smart Cities and Associated Mobile Technologies*, Procedia Computer Science **109**, 1086 – 1091 (2017).
- [7] E. A. Felgenbaum. *The Art of Artificial Intelligence: Themes and Case Studies of Knowledge Engineering*. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI’77, pages 1014–1029, San Francisco, CA, USA, (1977). Morgan Kaufmann Publishers Inc.
- [8] E. Blanzieri and A. Bryl, *A Survey of Learning-based Techniques of Email Spam Filtering*, Artif. Intell. Rev. **29**, 63–92 (2008).
- [9] A. Maltinsky, R. Giladi, and Y. Shavitt, *On Network Neutrality Measurements*, ACM Trans. Intell. Syst. Technol. **8**, 56:1–56:22 (2017).
- [10] A. L. Samuel, *Some Studies in Machine Learning Using the Game of Checkers*, IBM J. Res. Dev. **3**, 210–229 (1959).
- [11] Y. Harel, I. B. Gal, and Y. Elovici, *Cyber Security and the Role of Intelligent Systems in Addressing Its Challenges*, ACM Trans. Intell. Syst. Technol. **8**, 49:1–49:12 (2017).
- [12] M. B. Neria, N.-S. Yacovzada, and I. Ben-Gal, *A Risk-Scoring Feedback Model for Webpages and Web Users Based on Browsing Behavior*, ACM Trans. Intell. Syst. Technol. **8**, 53:1–53:21 (2017).

- [13] A. Gruber and I. Ben-Gal, *Using Targeted Bayesian Network Learning for Suspect Identification in Communication Networks*, Int. J. Inf. Secur. **17**, 169–181 (2018).
- [14] K. He, Z. Xiangyu, R. Shaoqing, and S. Jian. *Deep Residual Learning for Image Recognition*. In *CVPR*, pages 770–778. IEEE Computer Society, (2016).
- [15] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, et al., *Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups*, IEEE Signal Processing Magazine **29**, 82–97 (2012).
- [16] J. Schmidhuber, *Deep learning in neural networks: An overview.*, Neural Networks **61**, 85–117 (2015).
- [17] R. Ganesan, S. Jajodia, and H. Cam, *Optimal Scheduling of Cybersecurity Analysts for Minimizing Risk.*, ACM TIST **8**, 52:1–52:32 (2017).
- [18] M.-O. Pahl and F.-X. Aubet. *All Eyes on You: Distributed Multi-Dimensional IoT Microservice Anomaly Detection*. In S. Salsano, R. Riggio, T. Ahmed, T. Samak, and C. R. P. dos Santos, editors, *CNSM*, pages 72–80. IEEE Computer Society, (2018).
- [19] C. C. Aggarwal, P. S. Yu, J. Han, and J. Wang. - *A Framework for Clustering Evolving Data Streams*. In J.-C. Freytag, P. Lockemann, S. Abiteboul, M. Carey, P. Selinger, and A. Heuer, editors, *Proceedings 2003 VLDB Conference*, pages 81 – 92. Morgan Kaufmann, San Francisco, (2003).
- [20] O. Brun, Y. Yin, and E. Gelenbe, *Deep Learning with Dense Random Neural Network for Detecting Attacks against IoT-connected Home Environments*, Procedia Computer Science **134**, 458 – 463 (2018).
- [21] X. Liu, Y. Liu, A. Liu, and L. T. Yang, *Defending ON-OFF Attacks Using Light Probing Messages in Smart Sensors for Industrial Communication Systems.*, IEEE Trans. Industrial Informatics **14**, 3801–3811 (2018).
- [22] A. A. Diro and N. Chilamkurti, *Distributed attack detection scheme using deep learning approach for Internet of Things*, Future Generation Computer Systems **82**, 761 – 768 (2018).
- [23] B. Usmonov, O. Evsutin, A. Iskhakov, A. Shelupanov, A. Iskhakova, and R. Meshcheryakov. *The cybersecurity in development of IoT embedded technologies*. In *2017 International Conference on Information Science and Communications Technologies (ICISCT)*, pages 1–4, (2017).
- [24] A. Ukil, S. Bandyopadhyay, C. Puri, and A. Pal. *IoT Healthcare Analytics: The Importance of Anomaly Detection*. In *2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA)*, pages 994–997, (2016).

- [25] H. H. Pajouh, R. Javidan, R. Khayami, A. Dehghantanha, and K. R. Choo, *A Two-Layer Dimension Reduction and Two-Tier Classification Model for Anomaly-Based Intrusion Detection in IoT Backbone Networks*, IEEE Transactions on Emerging Topics in Computing **7**, 314–323 (2019).