INDIA GOVERNMENT ICCR SCHOLARSHIP 2021-2022

APPLICANT: BENJAMIN MBUU MUTUA

UNIVERSITY: IIT KHARAGPUR

PHD RESEARCH STATEMENT

Proposed Research Title: Solving the Banking and Financial Fraud in Kenya: Developing an Adaptive Financial Security System using Machine Learning (Application of the Recursive Bayesian Estimation)

Background

In a typical business organization, network security system comprises of multiple levels of protection; the perimeter, endpoint and application/data security tools. We have hardware and software firewall applications that determine which communication is allowed and which is not. To remain positive, let's assume that our first level security that includes signature and rule-based firewalls with well configured white-lists is impenetrable. However, we all know that hackers have long devised ways of penetrating this wall. Then perhaps face our IDS/IPS tools. Past this level, an attacker encounters the organization's anti-malware and antivirus applications. The main question here is "what happens when an attacker penetrates these levels of protection undetected?, because the attacker will most likely find his way around the walls." Or still, an employee innocently or intentionally clicks a link they shouldn't, opening a leeway for attacks that might be undetectable to the existing security measures? Such inside abnormal behavior presents big challenge to the organization for they are often missed by the firewall configured rules on exfiltration of content. Well, we can hire more security experts, but which human being is able to (even if they are willing) sit and watch net flow or log data all day long to understand and super correlate cognitively when two log files look odd. That's just not a human's job. In today's era of cloud computing technology, things even become more complicated when it comes to keeping our networks and systems safe. The attack base widens. Let us not forget that the security industry is already facing an acute skill shortage, and increasingly sophisticated adversaries. The expansion of attack surface can only spell doom for us. That's where an Artificial Intelligence based security solution, which can work round the clock and respond in milliseconds to cyber-attacks that would take days or months for conventional security systems and humans to identify, come in handy. I am not at all suggesting that companies should replace the existing legacy security tools or sack their cyber-security analysts and professionals. According to thesslstore (A cyber-security research firm), Artificial intelligence can augment the

skills of security analysts and alleviate the talent shortage. An AI tool can help secure a network/system by keeping constant vigil over it, learning user patterns and traffic flows. This means that it learns what users log in at what times, what areas of the network they typically access and what credentials they have. That way if an account logs in at odd hours and starts accessing strange parts of the network, the AI tool can act quickly to mitigate a possible intrusion. A human could easily miss these details. In this project, I use unsupervised machine learning to develop a system security model that mimics the human adaptive immune system for an automated, efficient and timely response to threats/attacks.

Research Question

In an age when complexity of networks has grown beyond our ability to handle, when cyber criminals are relentless in launching increasingly sophisticated attacks, when the attack base is widening and the generated data is too much for a human eye to sort effectively and efficiently, how do we monitor our systems round the clock and detect, identify and respond to threats/attacks immediately?

Why Artificial Intelligence (AI) and Machine Learning (ML)

According to Encyclopedia Britannica, Artificial Intelligence involves the development of systems endowed with intellectual processes characteristic of humans, such as ability to reason, discover meaning, generalize, or learn from past experience. According to Wikipedia, Machine Learning is a subset of AI that deals with scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, but relying on patterns and inference instead. Due to the inability of humans to offer round the clock monitoring of systems and analyze big chunks of data (with the necessary speed) that characterizes the field of cyber-security, use of intelligent machines and well developed Machine Learning models/algorithms is the only option we have. According to Naveen Joshi, CEO and founder of Allerin (a company offering Big data, IoT, Digital Business and Cloud Computing consultancy services), AI systems can integrate with existing cyber security applications mainly to create more accurate, biometric-based login techniques, Detect threats and malicious activities using predictive analytics, and secure conditional authentication and access. If we do not have an adaptive learning tool that we can set rules on to monitor our systems, we are going to be reactive on our approach towards attacks, and what can be worse than that in today's cyber environment where a threat is always lurking somewhere waiting for an opportunity? With that in mind, a supervised machine learning model (machine learning algorithm that is signature-based, rule-based, threat intelligence-based and that specifies white-listed applications), as has

been implemented severally, will not help much. Having this is not different from having a human do the system/network monitoring. My intended approach in this project (which every researcher in machine learning and cyber-security will agree with), is to use unsupervised machine learning model. That is; allow the algorithm to learn what it needs to learn and model what it needs to model based on the system/network behaviors.

Adaptive Financial Security System

To use the analogy of a human immune system, an Adaptive Immune System is defined by Wikipedia as a subsystem of the overall human immune system that creates immunological memory after initial response to a specific pathogen, leading to enhanced response to subsequent encounters with that pathogen. From this paraphrased definition, it is clear that an Adaptive Immune System develops as it encounters the disease causing microorganisms, meaning that it does not depend on a predefined way of dealing with pathogens. We can rightly say that it is automated and adapts uniquely to each threat. In other words, it learns on the go. That is the essence of unsupervised machine learning algorithms, and in particular the Financial/Banking Security Software/System I intend to develop. It will eliminate the need of human supervision on our systems/networks, and we all know how limited a human being is. For instance, no one would watch over a cyber-security system for 24 hours in a day without a moment of slackness. Such moment, however short, can be very costly to the organization when it comes to the security of our systems. To eliminate such gaps, we need a system that monitors our perimeters and other security applications ceaselessly, and deals with any threat adaptively and automatically. That security system can only be based on an unsupervised machine learning algorithm.

Recursive Bayesian Estimation

The above stated approach (modeling an unsupervised machine learning algorithm) is not possible without a proper learning mechanism. Also known as Bayes Filter, Recursive Bayesian Estimation is a mathematical concept in Machine Learning, Statistics and Probability theory used to study events, cluster them into families and place a new event into any of the formed clusters. It has so far been used to model one industrial immune system by Darktrace (an AI cyber security company), and the results are interesting and encouraging more research into what it can do to create immune systems. With this mechanism, it becomes possible and easier for us to identify normal patterns of behavior and abnormal patterns of behavior, model them and compare them using recursive estimation, then classify any newly detected system/network behavior as either normal or abnormal.

Goals/Objectives of the Project

- To create a system security model that mimics human adaptive immune system
- To use Recursive Bayesian Estimation to efficiently & timely classify internal behavior as normal or abnormal
- To create a security model that will stand against adversarial machine learning (where the model can be tricked into misidentifying or misclassifying events due to intentionally modified inputs).

Methodology

I intent to first carry out this research project in a lab environment then later after testing and implementation, and if circumstances will allow, in one of the Kenyan banks premises. In the lab setup, the project will involve a thorough analysis and execution of the model using at least three cyber-security machine learning datasets of which one will be annotated (preferably *CIDDS* (Coburg Intrusion Detection Dataset), or any other that is available in the host institution and two unlabeled datasets (preferably the CICIDS2017 and the UNSW-NB15 Datasets), or as shall be available.

Impact

According to the 2019 African Cyber-security awareness report released by the popcorn training company on 10th of December 2019, Africa's economic growth is largely derailed by the rampant cases of cyber crime and the resulting financial loses. Africa has over the years become a safe haven for cyber criminals. This is mainly because many African governments need to attend to other pressing issues such as fighting poverty, unstable politics, violent crime and large youth unemployment, therefore regarding cyber-security as a luxury, not a necessity. This has led to unfathomable loses especially in the banking and financial sector through system hacking by criminals. Another research conducted recently states that Africa lost 4 billion US dollars to cyber-crime in that year. This figure is what has been uncovered during the research, meaning that the loss is higher because a lot of organizations (especially financial institutions) don't publicly announce whenever there is a system attack for fear of losing customers. Closer home in my country Kenya, on January 18th 2018, Kenyans woke up to disturbing news. Local news channels and newspapers relayed the headline: "The National Bank of Kenya losses more than Ksh.29 million to hackers." The bank management, after confirming that indeed there was a system attack leading to loss of about KSh. 29 million, aptly added "No customer account has

been affected", perhaps to avert an imminent backlash from their customers. First and foremost, a loss of Ksh. 29 Million is a very big loss to the Kenyan economy. While we still don't have the facts up to date, it has been established before that such institutions downplay losses to win the trust of their customers, else a panic ensues and clients move to competitors. However, the same has happened to almost all financial institutions in Kenya and the larger African continent at some point of time. Banks have become the leading target of cyber crime as people increasingly adopt the use of financial technology. According to a Cyber-security study carried out in 2016, Kenya's public sector lost more than Sh5 billion from cyber attacks, with the financial services sector losing a whopping Sh4 billion. We are not alone, according to scidev.net (a cyber-security research company), the world lost an estimate of 500 billion US dollars to cyber crime in 2019. Needless to mention, there is a desperate need of cyber-security professionals and research scientists. My research towards creating immune banking/financial systems will be of great reward not only to my country Kenya, but also Africa and the world at large.

References

1. B.J Copeland 2006, *Artificial Intelligence,* Encyclopedia Britannica, viewed 09th June 2020 <www.britannica.com/technology/artificial-intelligence>
2. Wikipedia 2012, *Machine Learning*, viewed 09th May 2021 <https://en.wikipedia.org/wiki/Machine_learning>
3. Wikipedia 2009, *Adaptive Immune System*, viewed 09th May 2021 <https://en.wikipedia.org/wiki/Adaptive_immune_system>
4. Naveen Joshi 2019, *Artificial Intelligence and Cyber-Security*, Allerin, viewed 10th May 2021 <https://www.allerin.com/Artificial Intelligence and Cyber-Security>
5. Erfan Ibranhim 2017, *Using Machine Learning for Next Generation ICS security*, National Renewable Energy Lab, viewed 10th May 2021 <https://www.nrel.gov/esif/text-industrial-immune-system.html>
6. Patrick Nohe 2017, *Artificial Intelligence and Cyber-Security: Multi-part Discussion*, The SSL Store, viewed 10th May 2021 <https://www.thesslstore.com/blog/artificial-intelligence-cyber-security-multi-part-discussion/>
7. Media Room 2019, *2019 African Cyber-security awareness report*, KnowB4 Africa, viewed 10th May 2021 < https://blog.popcorntraining.com/untitled/>
8. Jean Shiloh 2018, *Cyber-Crime in Africa: Facts and Figures*, SciDevNet, viewed 10th May 2021 <https://www.scidev.net/sub-saharan-africa/icts/feature/cybercrime-africa-facts-figures.html?>