"Using Machine Learning to Solve Security Issues in Smart Healthcare Systems"
CSE 543 - Information Assurance and Security
Individual Student Report

Ramanan Durairaj (Group 18)
ASU ID: 1219512242
rduraira@asu.edu

OVERVIEW:
The study's purpose was to investigate various security concerns in a Smart HealthCare System (SHS) that incorporates new information technologies such as IoT, cloud, AI, and Big Data to make the advanced medical care framework "smart." The study also considered different machine learning algorithms to build an intrusion detection system to protect the SHS from unauthorized access or malicious attacks such as false data injection, DDoS attack, User to Remote(U2R), Remote to User (R2L), probe, and various malware attacks by regularly monitoring the computer network traffic and alerting users beforehand.

Intrusion Detection System:

Types of IDS:
SBIDS (Signature Based Intrusion Detection System) analyzes packets and checks whether the signature sent to one of the SHS databases matches the packet received. It works in a similar way to antivirus software. These signatures must be updated and maintained on a regular basis.

Anomaly-based IDS examines network traffic for anomalies and uses machine learning methods to detect them. The system is constantly detecting new threats and requires a lot of fine tuning.

Algorithms for IDS:

Logistic Regression:
The method of modeling the probability of a discrete result given an input variable is known as logistic regression. The most frequent logistic regression models have a binary outcome, which might be true or false, yes or no, and so forth.

Logistic regression is a handy analysis tool for determining if a fresh sample fits best into a category in classification tasks. Because components of cyber security, such as threat detection, are classification problems, logistic regression is a valuable analytic tool.

Decision Tree:
The most powerful and widely used tool for categorization and prediction is the decision tree. A Decision Tree is a Supervised Machine Learning technique that separates data depending on a parameter on a regular basis. It has a flowchart-like structure, with an internal node representing a Test on attribute, a leaf node representing a Class Label, and a branch representing a test result.

Random Forest:
As the name suggests, a random forest is made up of a huge number of individual decision trees that work together as an ensemble. Each tree in the random forest generates a class prediction, and the class with the most votes become the prediction of our model.

The wisdom of crowds is the basic principle behind random forest, and it's a simple yet effective one. Any of the individual constituent models cannot outperform many reasonably uncorrelated models (trees) working as a committee.

ANN:
Artificial neural networks (ANN) are computational networks that are inspired by biology. An ANN is a computer model made up of many processing components that receive inputs and produce outputs depending on predetermined activation functions.

Observations:
In conclusion, we find that the Artificial Neural model performs best in terms of accuracy, precision, recall, and F1 scores, whereas the Logistic Regression model performs worse. However, due to the higher calculation

of data in nodes, the ANN model requires the most time in terms of time complexity. Using GPUs for parallel processing helps reduce the time complexity of ANN. So, if building the system takes a long time, the most efficient approach is random forest, and for high accuracy, ANN is employed.

The observations made were:

- Protecting a network against assaults requires the use of a network intrusion detection system. The ability of an Artificial Intelligence-based intrusion detection system to detect attacks and produce accurate findings is dependent on the dataset being trained and updated on a regular basis.
- Intelligent feature engineering and DL algorithm complexity reduction using only the important features, the detection accuracy will be roughly similar to using the entire collection of features. This fact will eventually lead to a simplified model and the utilization of fewer resources in the real world.
- As a hybrid technique, we can use ML for classification and DL for feature extraction. The model will be simplified by combining ML and DL.
- the use of artificial intelligence in the creation of Cyber-Physical Systems like UAV-enabled networks has the potential to be a promising research area, but further research is needed.

Individual Contributions:
My main contributions to this study project were conducting research by reading journal publications and conference papers on all potential cyber-attacks that could disrupt the healthcare system and how the system we're designing will handle them. I also did extensive research into the Logistic Regression technique and how it could be used to create an IDS. I attended weekly meetings/discussions in addition to writing and studying the topic. The following are a few security attacks that a couple of us working on this project thoroughly researched and found to be potentially harmful to a SHS:

Distributed Denial of Service:

Financial organizations, government infrastructure, and cloud service providers are usually prominent targets for distributed denial of service attacks. But DDoS attacks have recently become increasingly common in healthcare sector. Denial of service attacks are used by cybercriminals to take down networks and apps by overloading them.

DDoS assaults are used by cybercriminals for a variety of objectives, including extortion and to divert security teams' attention away from more malicious actions like data theft or ransomware infection. DDoS attacks can be disruptive, as they impede patients from booking appointments online and doctors from sending or receiving critical information. In the worst-case situation, systems might be compromised, and patient data could be lost. Healthcare institutions should be able to tell the difference between fact and fiction when it comes to DDoS.

Man in the middle:

Hackers can use Man in the Middle attacks to put themselves between two communicating parties and intercept, send, and receive data that was never intended for them. One of the most important aspects of MITM assaults is that both the victim and the person with whom they are seeking to contact are unaware that the "man in the middle" is present - until it is too late.

Hackers insert themselves between communicating parties — either two healthcare providers or a patient and their provider — in MITM attacks on healthcare systems. This can also be done by intercepting communications between medical devices in the same hospital or care facility, implying that the industry is becoming more vulnerable to cyber-attacks as AI-based technologies, particularly in the cloud, become more widely used.

Probes:

To detect vulnerable systems, probes collect data. IP Sweep is an example of a Probe that utilizes ping to examine a range of IP addresses to see which ones are still active. Another method for obtaining a list of potential targets is to spoof a zone transfer request to a DNS server, as done with the ls command in NSLOOKUP. Other types of Probes include port scans, fingerprinting, and inside sniffing.

Remote to Local:

While TCP/IP protocols may be used by probes and DOS attacks to gain control of the target, R2L assaults always use application protocols. Password guessing, server vulnerabilities, configuration errors, and backdoors are all things to be aware of.

User to Root:

The attacker first obtains a user session on the remote system, ideally in the form of an interactive shell or TELNET window, in a User to Root attack. Using several traditional methods, the attacker seeks to gradually enhance his privileges until he obtains super-user capabilities. Such example of an escalation strategy is stack smashing, which feeds a packet to a set-UID-to-root program that corrupts its address space so that a return from-subroutine instruction results in the construction of a set UID-to-root command shell.

Learning Outcomes:

The following are my primary learning results from this project:

- The importance and advantages of a smart healthcare system (SHS) and the technology that go into it.
- Gaining knowledge of the data in SHS's security vulnerabilities, risks, and issues.
- Classification techniques based on machine learning and deep learning, such as Logistic Regression, Decision Tree, Random Forest, and Artificial Neural Networks.
- The benefits and drawbacks of adopting various algorithms for an intrusion detection system in a SHS
- Different types of existing intrusion detection systems, as well as the design of an intrusion detection system.
- Characteristics and metrics while designing an optimal intrusion detection system like:
    - False positive prevention
    - Expertise in a given field
    - Knowing existing signatures
    - Hidden from attackers
    - Adaptability
    - Real time monitoring and filtering.

References:

[1] A. A. Hady, A. Ghubaish, T. Salman, D. Unal and R. Jain, "Intrusion Detection System for Healthcare Systems Using Medical and Network Data: A Comparison Study," in IEEE Access, vol. 8, pp. 106576-106584, 2020, DOI: 10.1109/ACCESS.2020.3000421.

[2] P. A. Williams and A. J. Woodward, “Cybersecurity vulnerabilities in medical devices: a complex environment and multifaceted problem,” Medical Devices (Auckland, NZ), vol. 8, p. 305, 2015.

[3] H. Pandey and S. Prabha, "Smart Health Monitoring System using IOT and Machine Learning Techniques," 2020 Sixth International Conference on Bio Signals, Images, and Instrumentation (ICBSII), 2020, pp. 1-4, doi: 10.1109/ICBSII49132.2020.9167660.

[4] M. A. Al-Shaher, R. T. Hameed and N. Ţăpuş, "Protect healthcare system based on intelligent techniques," 2017 4th International Conference on Control, Decision and Information Technologies

(CoDIT), 2017, pp. 0421-0426, doi: 10.1109/CoDIT.2017.8102628.

[5] P. Illavarason and B. Kamachi Sundaram, "A Study of Intrusion Detection System using Machine Learning Classification Algorithm based on different feature selection approach," 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2019, pp. 295-299, doi: 10.1109/I-SMAC47947.2019.9032499.

[6] Pant, A., 2021. Introduction to Logistic Regression. [online] Medium. Available at: [Accessed 31 October 2021].

[7] Patel, D., 2021. Logistic Regression Classification | Math. [online] Medium. Available at: [Accessed 31 October 2021].

[8] N.s.chandolikar & v.d.nandavadekar,International Journal of Computer Science and Engineering (IJCSE),Vol.1, Issue 1 Aug 2012 81-88 —comparative analysis of two algorithms for intrusion attack classification using kdd cup dataset