Bachelor Degree Project

UNIVERSITY OF SKÖVDE

MACHINE LEARNING AND SYSTEM ADMINISTRATION

A Structured Literature Review

Bachelor Degree Project in Information Technology
Level ECTS 22,5 credits
Spring term 2020
Date: 2020-08-31

Karl Jonsson
c16karjo@student.his.se

Supervisor: Thomas Fischer
Examiner: Jianguo Ding

Contents

Sammanfattning

Denna literaturöversikt går igenom två olika system inom IT-säkerhet och hur de fungerar tillsammans med maskinlärningstekniker till en relativt ytlig nivå.

Syftet med denna rapport är att kunna sammanfatta dessa system och se hur de kan hjälpa med en systemadministratörs uppgifter, hur det kan användas för automatisera och vad för positiva och negativa förändringar det kan ha på en infrastruktur.

Maskinlärning kan vara ett kraftigt verktyg för systemadministratörer för att lätta på arbetsmängden som kan förekomma inom en organisation, vilket är också varför det är viktigt att diskutera när och var man ska utplacera en lösning. Den här studien ska diskutera användningen av maskinlärning och när och var det kan användas.

Nyckelord: Intrusion Detection systems, IDS, Machine Learning, Anti-virus, Anti-malware, malware

ABSTRACT

This literature review discusses two different systems within IT-security and how they work within machine learning to a relatively surface-level degree.

The purpose of this paper is to be able to summarize these systems and see how they can help a system administrator's assignments. how it can be used for automation and the positives and negatives.

Machine learning can be a powerful tool for system administrators to alleviate the workload which can exist within an organization, which is why it is important to discuss when and where to deploy a solution.

Keywords: Intrusion Detection systems, IDS, Machine Learning, Anti-virus, Anti-malware, malware

1. Introduction

Machine learning, also abbreviated to ML, is an important step for the future of automation, from self-driving cars to chatbots or deepfakes. Machine learning has added another level of automation that was harder to reach previously, for a system administrator automation is important to increase productivity if certain tasks require otherwise dedicated time and effort, from automated error-logging to automating entire systems.

System administrators often work with a lot of systems, eventually repeating work or, work that take long time can overwhelm the IT staff of a company. Instead of increasing the personnel, automation for tasks is often preferred, such as using scripts to automate small tasks. When it comes to the larger tasks, such as figuring out new attack vectors, machine learning could be a potential source of help.

The implementation of ML in a system administration environment could mean less time taken up by repeating daily tasks that should not need the full attention of a real person. This of course holds for other systems in the whole environment too. If we take something like a standard non-ML anti-virus software, which requires regular updates and only helps after the fact that a virus or malware has been identified, having an anti-malware machine learning could make the detection software have an easier time detecting malware than normal anti-virus, basically beating new viruses or malware without having to download a new update to your software whenever a new virus have been discovered.

Machine learning have always been about making things faster or making things accessible to humans that before required a tremendous amount of manual work. Machine learning has a large use today within certain automation, such as image recognition, that simply would not be as effective with “normal” programming.

This report will work to uncover if machine learning can make certain system administrative tasks require less management, if machine learning can help system administrators work easier, faster than before or alleviate the administrators workflow and what is necessary for a system environment to be able to deploy certain machine learning software.

2. Background

In this section I discuss some background to the work

2.1 Machine Learning

Machine learning is a type of technique that is often used in many different more advanced technologies, often used in image recognition for self-driving cars, as well as in certain countries to identify people on the streets. In the following image a machine learning algorithm shows what it believes to be cars and the perceived location. The different things in the image is categorized depending on how it has been programmed. Forson. (2017)

Figure 1: self-driving cars (Forson, 2017)

There are multiple uses of recognition software that make use of machine learning, another example is voice recognition.

Machine learning is "trained" by having datasets of whatever it is the algorithm needs to detect. in the above example the algorithm will be given a large amount of images with cars in different angles and colors, in different brands and similar, with the car marked with a square. When feeding this in the machine learning system will create an algorithm to be able to detect them. With image recognition more often than not, there will be closer to several thousand or million images of cars for the algorithm to properly start to detect things. Ray (2019)

2.2 Intrusion Detection Systems

The traditional intrusion detection system is a system that monitors a network or a number of computers, it is most often used within a corporate system to alert or stop unauthorized access to the network or computers, as well as serve as second wall of acceptance to a typical firewall. This is referred to as a Network intrusion detection system (NIDS).

The standard setup is to have the intrusion detection system in between the firewall and the router out to the internet as shown in the following image. Mitra. (2017):

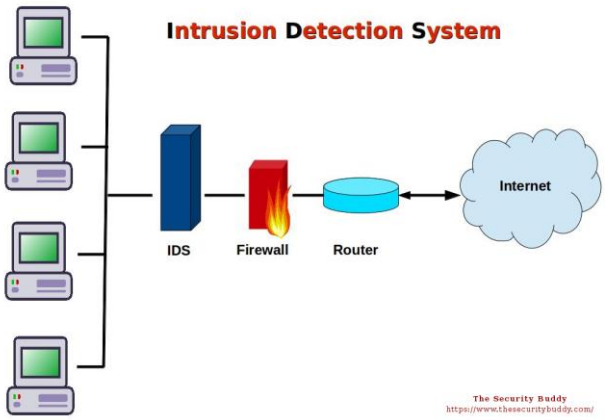

Figure 2: Intrusion detection system (Mitra, 2017)

An IDS requires deep understanding of normal behavior and network usage within the network that it is going to be set up. To do this an administrator with good knowledge of the system is required, they need to know the routines of everyone in the system and what programs that operate frequently or unfrequently within the network. This is what requires most time in a traditional IDS.

IDS operate with several set parameters, these parameters can also be called categories, as the program defines different operations within the set categories. Depending on the how it categorizes a certain operation it will either stop it, let it through, or message an administrator of suspicious action. These different actions are entirely customizable by the administrator and serve multiple functions. For example, if an administrator is doing live test runs, then setting up that the IDS sends messages is good to see whether the IDS is configured correctly.

When it comes to these parameters, there are several issues. One particular issue is that activity outside these parameters can be flagged incorrectly. For example if a company use several systems that connect outside of the local network, then mapping those can be troublesome especially if multiple new connections are established by several different programs, which can be the case in certain cloud solutions, this can cause false positives and false negatives. False positives and false negatives are basically when the IDS is not entirely sure whether to allow or stop and operation, however it tries it best to flag them according to the rules set. This can cause problems, such as letting through connections that should not be allowed and stopping perfectly fine operations. The more a company decide to use cloud services the more the manual work for an IDS increases. This is the reason that people have been working on implementing Machine Learning into IDS. Usama, et al. (2019)

The need for more advanced and automated types of intrusion detection has been severely needed to combat the rate of which intrusion attacks are developing, they are getting more complex and harder to detect by standard IDS. To combat this adversarial example have been introduced to make the system learn what it can expect to be a sort of attack. Kołaczek and Warzyński (2018).

Machine learning coupled with intrusion detection systems have only recently started to be used in practice. Machine learning has advanced significantly over the last years with

improvements in training time and techniques. The main cause of this is the need for more automation to repetitive tasks and not wanting to spend more money hiring system administrators to handle simple tasks or have them do long repetitive tasks. Salem, Taheri and Yuan (2018) explain that the machine learning implementation for IDS was needed to lift some weight off the network administrators and the network itself.

There is also hardware/host intrusion detection system, Host intrusion detection is, as the name suggests, the detection of a threat by its attempts to gain access, maintain persistence, or spread across a system. As opposed to anti-malware it identifies the method of infiltration or intrusion rather than the threat. Tselios, Tsolis and Athanatos (2020)

Below shows a type of neural network, which can also be used as an IDS. Hodo, et al. (2017):

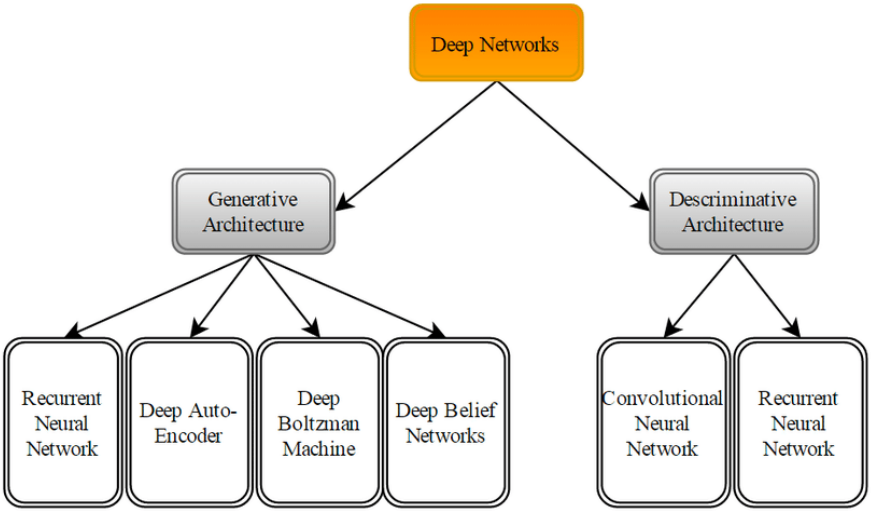

Figure 3: Deep-learning IDS (Hodo, et al. 2017)

2.3 Artificial neural networks

Artificial neural networks are almost self-explanatory, it is a network of connected machines that operate as a sort of neural system, the concept itself is simple to grasp but the implementation is often more complex.

> *"Artificial Neural Networks (ANNs) are large collections of interacting entities. Certain conditions of behavior and of nonlinear coupling between the entities enable self-organization of the system with emergent properties of associative memory, abstraction and generalization" Ligomenides. (2005)*

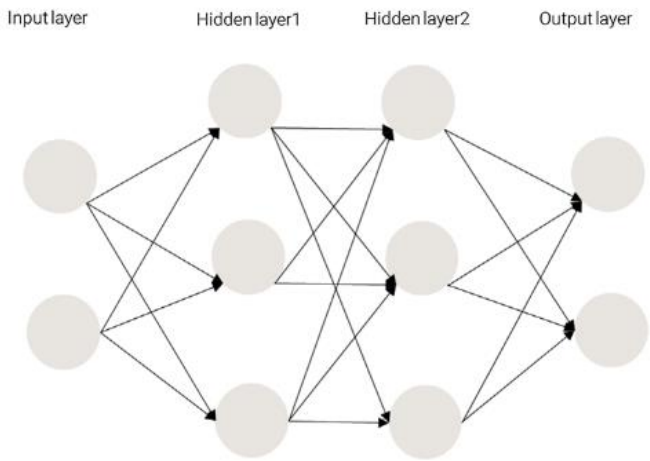

Figure 4: Neural network (Ligomenides,2005)

A neural network often consists of at least 3 parts; an input layer, a hidden layer, and an output layer. Simply put, the input layer takes in information in the form of variables or similar. The hidden layer transforms the input into a higher-level function. The output layer will output the final variable transformation. Ayyadevara. (2018)

2.4 Anti-Malware

Anti-malware or anti-virus software is an important part in a computer's safety. It helps the computer protect itself against malware or viruses that have made it onto the computer either through downloaded files, sketchy links, remote hacking, or social engineering.

Viruses or Malware can hurt and affect your computer in ways that is harmful for you as well. For example, someone could get private information off your computer.

Anti-virus software can use signature based checks on files, meaning that the software updates its signature list every so often, so that it can notice if a file contains a specific set of code that matches with the signature to then block it.

The mentioned method is the usual way signature based anti-virus/anti-malware works today.

3. Research Question

To explain and reach an understanding of what machine learning is, as well as discuss use cases within certain subjects or fields where machine learning can help alleviate the workload or automate it completely to give system administrators an easier time working.

I will be going through two systems that have machine learning solutions anti-malware and Intrusion detection systems

The main questions to answer are essentially,

How the ML can help in automation of systems?

Potential advantages/disadvantages of a ML implementation?

Mostly the point is not to give an advanced understanding of machine learning, but rather to give a comprehensive look into the use cases.

It is also a point to try and objectively draw a conclusion on the use cases and the effectiveness of such an implementation.

To also discuss the requirements for setting up machine learning software in an environment, and how it will affect an already deployed environment etcetera.

This will be in the form of a structured literature review; I will be gathering the information around the main questions in each category and trying to reach a conclusion.

4. Methodology

I used keywords to find relevant research papers together with thematic coding to mark text that is relevant into different subjects. The subjects are based on level of information, system, usability in the report, and so on.

For databases, I used IEEExplore, ACM and Springer Link to get the information, where I also will be using keywords “Intrusion detection system” and “adversarial attacks” as well as limiting the publishing years to 2015-2020 because of the developments of artificial intelligence and machine learning during those years.

The choice of keywords is based on search queries, for example at the start as I was searching for examples to help me select subjects under machine learning I simple used “Machine learning” and “system administration” however this gave different answers that it did not work well, I decided to eventually search for different subjects individually, for example “Machine learning” and “anti-virus” or “anti-malware”

Exclusion terms, though not many it was based of the criteria of:

1. Being free to read (non-paid)

As I am a student, I do not have the economical luxury to buy papers, fortunately the university allows for free access to a majority of articles on the aforementioned databases.

2. Being in English

It being in English is mostly preference, though it is most relevant since the review is in English as well.

3. Has been peer-reviewed.

Extremely important to make sure the quality of the paper, all papers on IEEExplore are however peer-reviewed. The papers in ACM are peer-reviewed too

4. Would be relatively relevant to the topic.

Meaning that the papers can go into other tangents and can be about a completely different point but can still add to my report. If a paper does not add anything which another paper has already described well it will not be included.

After trying to narrow down the amount of papers in my search terms, eventually I ended up with two different search queries.

IEEE: ((((((“All Metadata”:Machine Learning) AND “All Metadata”:antivirus) NOT “All Metadata”:mobile) NOT “All Metadata”:android) NOT “All Metadata”:poison) NOT “Publication Title”:Attack)

They have filtered out a couple of terms that just simply did not give good results.

ACM: [All: "machine learning"] AND [All: "intrusion detection"] AND [All: ids] AND [All: allfield(analysis)] AND NOT [All: android] AND NOT [All: mobile] AND NOT [All: approach] AND NOT [All: augmented] AND NOT [All: sql] AND [Publication Date: (01/01/2015 TO 12/31/2020)]

5. The Review

In this chapter I will go through the findings I have found. They will be separated depending on the subject, so anti-virus, and IDS. They will have similar sub-headers, like Functionality, necessities and comparatively.

5.1 Anti-virus and Machine Learning

The use of machine learning within anti-virus is nothing new, and many already existing anti-virus software use machine-learning together with the traditional way of signature-based detection. However, implementation of this design can vary, and the performance will also suffer.

5.1.1 Functionality

When anti-viruses use machine learning it can do so in separate categories, Cloud or local. When using a cloud infrastructure, the computers quickly uploads the entire file to either an internal cloud anti-virus service or a public one for review by that server. The software will go through the entire programs line after line and try to see if there are any harmful lines of code that could affect a system in any malicious way. Dev, et al. (2016).

If the code is malicious it will send out an order to the systems with that file that it cannot be run and removes access to it from any machine in the system. It will determine maliciousness on multiple factors. These factors can include several things, from program sending out information to another computer, to if it directly affects your computers other files. This is all decided by the learning method used to create the system. Dev, et al. (2016).

> *"Machine learning based detection system requires training dataset of malware attributes and according that machine learning algorithm detects the malware. There are different machine algorithms such Decision Tree, Support Vector Machine, Random Forest, Boosting, etc. The obfuscation techniques such as Dead code insertion, Register Renaming, Code Transportation, Instruction Substitution hinder the malware detection mechanism." Matin, Rahardjo (2016).*

Using datasets, the algorithm is trained to detect malicious code out of any example, meaning that if it is trained properly it can detect viruses and malwares that are "new". A well-executed training of the machine learning system can mitigate the need to update it. Meaning that, compared to image recognition, even if a certain line of code or program has not been seen before on any device publicly available, it can still detect it as malicious. Matin, Rahardjo (2016).

Local infrastructure would work in the same way, it could also work faster however it could also affect performance of the computer quite heavily depending on the level of machine learning. Of course, the biggest positive in this type is that it does not require constant internet connection, or direct connection to the server.

"Malware is one of the threats to information security that continues to increase. In 2014 nearly six million new malware was recorded. The highest number of malware is in Trojan Horse malware while in Adware malware is the most significantly increased malware. Security system devices such as antivirus, firewall, and IDS signature-based are considered to fail to detect malware." Matin, Rahardjo(2019).

As Matin and Rahardjo explain in their paper the number of malwares has increased exponentially over the last couple of years, the current outlook of signature based anti-malware protection is not keeping up with the influx of new malware. Which is why some type of machine learning system would function as the primary stopgap in the future.

5.1.2 Necessities

The first part of any machine learning system requires the knowledge to train it. There are many different methods described earlier.

"A discriminative learning model function is used to predict if an unknown malware belongs to a specific family. 70% of malware instances not identified by Antivirus software could be classified using discriminative learning (ibid). Some existing work is related to speeding up the classification of malware, a pressure point, with a ML algorithm." Kargaard, et al. (2018)

The quote implies that many training methods used to train the machine learning algorithm exist, however some methods affect the usability of the algorithm. To employ a proper machine learning anti-malware/virus infrastructure the proper knowledge on what training to use is required.

As mentioned, Machine learning works in datasets. When it comes to anti-virus in particular it is important to know that it is a very arduous task to teach it the correct things, this is because the number of different exploits or viruses are plenty. Therefore, a dedicated server for isolated execution will most likely be the safest option. First feed it datasets that is related to actions that happen to the computer, for example, a program starts to remove files on the computer, this is obviously malicious, so you tell it that it is not an approved action. If the machine learning algorithm knows this, it will start to associate certain parts of programs better with malicious behavior, this is done through test feeding it viruses in an isolated environment, this is the main learning process, which is the most important. The whole process will take a relatively long time, but it can be worth the effort.

However, there is also the case of commercial ML anti-virus software. They offer either the service or the software required to set up an easy to use environment. Of course, the training method for these are often under NDA, so the actual effectiveness is also questionable. However, they supply a relatively easy to deploy service which many companies can easily adopt. This can also be coupled with a standard anti-virus software for extra protection.

Often case with machine learning, there is a required use of a dedicated server depending on what type of environment you have. If you are a company, then a dedicated server should be employed to take care of the analysis of the files or programs. The server needs to have the hardware required to actively use machine learning, which is not a high standard in modern computing, but can strain individual work computers.

5.1.3 Comparatively

Signature based detection and machine learning based detection, each come with their own criteria and use areas.

The classic signature base detection works well enough for the standard end-user, someone who is not prone to targeted attacks such as mail scams or similar. The strain of a ML software running on a personal system outweighs the usefulness of the increased security unless the user peruses the lesser safe areas of the internet.

However, the average company or it-personality do get target to some extent. Any company email or any real company employee have the tendency to be exposed to files that are harmful for the personal computer or the entire company environment. This is the area in which signature-based detection does not reach the standard that is required to keep a company safe. That is where machine learning steps in and covers the cracks. The machine learning algorithms have a general advantage when it comes to detect malicious files.

A company can also have the resources to deploy a dedicated server, the use of a machine learning solution does benefit a company to a large extent because of previously explained reasons. Even if policies exist within companies everyone also makes mistakes, the existence of social engineering and several other types of specific attacks cannot be prevented with only a signature based anti-virus solution.

5.2 Intrusion Detection Systems (IDS) and Machine Learning

Intrusion Detection Systems are essential in real working environments, they will help document intrusion attempts on your systems and potentially automate some type of defenses. However, they work solely on how the system administrator have programmed it to respond, and a lot of IDS do have a large amount of customizability to help with such types of automation work.

5.2.1 Functionality

An IDS typically measures network activity within a network, IDS is primarily for use within companies that have an extensive network for multiple machines or locations. The IDS then categorize the behavior of users or programs within the network and decides if it should be blocked or not. Typically, the accepted behavior is “Hardcoded” meaning that the system administrators manually decides what is accepted behavior and what is not. This can lead to problems when introducing new software or new methods for working within the network and can include a lot of manual work.

Similarly, to any other type of machine learning the use of it within and IDS can learn from use and training, this allows for a system administrator to insert datasets for the machine learning algorithm to categorize. This type of IDS is sometimes referred to as Anomaly-based IDS.

An IDS is usually set up between the firewall and the company or private network to act as a sort of extended firewall as explained in the background.

Some IDS sometimes be referred to as a neural network, this can be because of the particular infrastructure of the network, by using nodes to help detect activity an IDS can be referred to

as a sort of primitive neural network. This is true for many machine learning systems that work over a larger area or infrastructure. An IDS specifically for use on a network are sometimes also referred to as NIDS meaning “Network Intrusion Detection System”.

> *“NIDS monitor traffic packets and analyze its headers and contents to determine if this packet is normal or malicious.” Pengju. (2019)*

> *“Intrusion detection is a technique in computer network which play an important role in detecting different type of attacks. Intrusion detection system (IDS) is a device application, both software and hardware that can detect suspicious activity in a system or network. IDS systems can also mean a combination of software and hardware that work together to detect attacks and intrusions on the network. IDS basically have the ability to record and analyze data in real-time in the detection process, log and stop abuse and attack.” Salman, et al. (2018)*

5.2.2 Necessities

For a deployment of an IDS in general, an extensive knowledge about the basic tasks performed over a network is needed. What should be categorized as bad behavior and what should be categorized as a normal behavior, this question is the first hard question when designing the system. The network or system administrators should know this before starting training or deploying of any type of IDS. The architecture of the network needs to be in consideration when placing the IDS for maximum effectiveness as well.

An IDS requires dedicated sensors placed in effective areas as mentioned earlier. A ML IDS specifically needs dedicated hardware for maximum work capacity as well.

Similarly, to other types of ML it is necessary to know how to train it. Using similar methods as mentioned in the anti-malware section, an IDS can be properly trained. In general, the same rules when training the ML system as any other ones.

When it comes to training, the general idea is pretty simple, feed the machine learning algorithm some datasets that will be similar to the expected input, such as network activity from a certain program or just basic user network activity actions, such as accessing shared folders or admin activity such as admins shares. Then coupled with this, the expected outcome. For example, an administrator is trying to reach an admin share folder on a server from his home computer, depending on company policies this may be acceptable, so in the dataset this must be set as a “legal” action.

When doing test deployment, it is important to document over time to see if it works as intended.

5.2.3 Comparatively

> *"Signature-based IDS […] detect attacks by comparing new traffic with extracted signatures from known attacks. It has a high accuracy and low false accept rate for detecting known attacks, but it cannot detect unknown or new attacks." Liu. (2019)*

A basic signature-based IDS does still work incredibly well *as long as* the system gets constantly updated to account for new updated attack vectors. The standard company should seriously consider this option before a machine learning solution.

> *"[…] because of well-known deficiencies in artificial intelligence and classification algorithms, anomaly IDSs are slow and produce a lot of false positives and negatives." Aldwairi, Alshboul, Seyam. (2018)*

The fact still stands that machine learning is a resource demanding process, and that poses a problem when talking efficiency over performance.

When working with an IDS in general a network infrastructure will slow down, however machine learning adds another factor that can slow down the entire process should the dedicated hardware not be able to keep up with the influx of network packages.

6. Results

The point of the literary review was to see if machine learning will be a viable or possible more effective way to implement certain systems, specifically for Intrusion Detection systems and Anti-virus.

The results are divided between the different systems.

6.1 Machine learning

With anti-malware/anti-virus software machine learning automates detection with the help of datasets involving code paired with a classification of this code. Then it can be trained in an isolated environment with viruses or malware, but also normal programs, to learn by itself what is ok and not ok to allow to run. This allows for the detection of malware that is new or targeted malware, things that traditional anti-malware can struggle with.

The positives of a machine learning anti-malware solution would be that you have a system that can, if trained properly, stop many future malware/viruses without need constant updates or management.

The downsides are the time it can take to train the algorithm and the resources required.

The time it takes to train an algorithm is not something you can find a reliable source on however estimate a couple of weeks to get the solution ready for deployment at least.

A machine learning solution also requires dedicated hardware, this implies setting up a virtual machine or buying a small, dedicated server with the required specifications. This will have to be able to launch an isolated virtual environment to be able to train as well.

6.2 Intrusion detection systems

With intrusion detection systems machine learning has a different use. As explained, a traditional IDS can be set to only allow a few actions over the network, this deny all rule is useful for internal traffic. However, when a company does need to accommodate outside network activity with other companies, this can create problems.

This is where the use case can be seen, to as explained feed the machine learning algorithm with datasets of normal network activity and train it is using traffic simulations similar to a company's expected activity. This potentially allows for the network to be protected without administrators having to add exceptions and rules whenever a ne

Similarly, to anti-malware the downsides are training times and resources. It can potentially take longer time to train an IDS ML algorithm than it does for an anti-malware because of the need to make sure that it does work as intended. The ML IDS also require dedicated hardware similarly to a traditional IDS, however it requires more performance than the latter.

Another potential problem is the performance, as mentioned, the performance over the internal network will be slowed down, even traditional IDS affect the networks performance however an ML IDS have a higher performance decrease.

7. Conclusions

When conduction this report 15 reports and papers where analyzed to help explain the different positives and negatives of machine learning in an infrastructure environment. Each reference adds some new information to further help illustrate or explain the different scenarios in which the systems operate. The information is then divided up into positive and negative traits, which conditions or weight these traits have on different infrastructures.

The need for machine learning within either anti-virus or IDS is seldom useful for the single person in their home systems. The need for it comes for larger small-medium to larger companies that cannot risk new attack vectors to reach their systems.

In the case of Anti-virus ML, it can be useful for most organizations that have multiple servers or computers, as updating the software is often no as important compared to a standard anti-virus, which makes management of the servers easier.

When it comes to IDS ML the case stays similar to anti-virus. However, as mentioned, IDS puts a certain amount of strain on the network and can cause slowdown of the network speeds.

To take full advantage of these Machine learning methods dedicated servers are required as well as personnel that have knowledge in how to operate the systems to some extent. The solutions can also be bought from other companies as a service, but that is another layer of cost to consider.

The main issues lead down to resources and performance over cost difference. Any organization or person can deploy machine learning equipment. However, it machine learning is not an all around solution, and can be considered excessive for smaller organizations or for home networks.

Conclusively, it all depends on the financial situation of the organization or person wanting to create a safer infrastructure and how important it is to increase security within the infrastructure.

8. Reflections

As I am relatively inexperienced with actually working with machine learning and the different systems that have been mentioned here, the only real experience I have is the machine learning anti-virus software that my employer have in the backend. Even then I have no experience in how to train these myself. That is also why I wanted to do this review, to learn at least the basics in machine learning, to make people understand how it works.

This paper goes through what I would call surface level information, it is not meant to sell the idea that machine learning is some sort of solution to every problem, but that it can have its potential positives and negatives, and depending on those points whether or not it can be useful to even use.

8.1 Future work

What I would like to continue this would be to go in and learn first or second hand in how these machine learning systems are trained, because reading can only give you so much comprehension. For example, I describe how datasets are used to train machine learning algorithms, but the technical details about the actual programming of the datasets and even the algorithm is a whole other report in of itself, and I do not believe that I have the education to comprehend its complexity.

Something I was also thinking about was maybe a more thorough investigation on the factual costs from a small company perspective to a larger one, to deploy a system similar to the ones discussed, and if it is more profitable to look into buying it as a service. As I explain in the report, the costs will be larger for a machine learning solution, to try and expand on that idea and see if there can be a middle ground for something like an “aaS”.

8.2 Ethical

In this report there is not much in way of ethical thinking that should be able to affect future work or considerations, as this is not an experiment that will have consequences on the future of the machine learning.

References

Aldwairi M, Alshboul M A, Seyam A. (2018) Characterizing Realistic Signature-based Intrusion Detection Benchmarks. ICIT 2018: Proceedings of the 6th International Conference on Information Technology: IoT and Smart City. ACM

Ayyadevara V.K. (2018) Artificial Neural Network. In: Pro Machine Learning Algorithms. Apress, Berkeley, CADev M, Gupta H, Mehta S, Balamurugan B (2016) Cache implementation using collective intelligence on cloud based antivirus architecture. 2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT). IEEE

Forson E. (2017). Self-driving Cars — OpenCV and SVM Machine Learning with Scikit-Learn for Vehicle Detection on the Road. https://towardsdatascience.com/teaching-cars-to-see-vehicle-detection-using-machine-learning-and-computer-vision-54628888079a [2020-02-20]

Hodo E, Bellekens X, Hamilton A W, Atkinson R C, Tachtatzis C (2017). Classification-of-deep-learning-IDS. https://www.researchgate.net/figure/Classification-of-deep-learning-IDS_fig5_312170608 [2020-03-04]

Kargaard J, Drange T, Kor A. Twafik H, Butterfield E. (2018) Defending IT systems against intelligent malware. 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT). IEEE

Kołaczek G, Warzyński A. (2018). Intrusion detection systems vulnerability on adversarial examples. 2018 Innovations in Intelligent Systems and Applications (INISTA). IEEE.

Ligomenides P.A. (1991) Cooperative computing and neural networks. In: Prieto A. (eds) Artificial Neural Networks. IWANN 1991. Lecture Notes in Computer Science, vol 540. Springer, Berlin, Heidelberg

Liu P. (2019). An Intrusion Detection System Based on Convolutional Neural Network. ICCAE 2019: Proceedings of the 2019 11th International Conference on Computer and Automation Engineering. ACM

Matin L M, Rahardjo B (2019) Malware Detection Using Honeypot and Machine Learning. 2019 7th International Conference on Cyber and IT Service Management (CITSM). IEEE

Milad S, Shayan T, Jiann Shiun Y. (2018) Anomaly Generation Using Generative Adversarial Networks in Host-Based Intrusion Detection. 2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). IEEE

Mitra A. (2017) What is IDS or Intrusion Detection System and how does it work? https://www.thesecuritybuddy.com/data-breaches-prevention/what-is-ids-intrusion-detection-system-how-does-it-work/ [2020-03-01]

Ray S, "A Quick Review of Machine Learning Algorithms," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 2019, pp. 35-39, doi: 10.1109/COMITCon.2019.8862451.

Salman M, Husna D, Apriliani S G, Pinem J G. (2018). Anomaly based Detection Analysis for Intrusion Detection System using Big Data Technique with Learning Vector Quantization (LVQ) and Principal Component Analysis (PCA). AIVR 2018: Proceedings of the 2018 International Conference on Artificial Intelligence and Virtual Reality. ACM

Tselios C, Tsolis G, Athanatos M. (2020) A Comprehensive Technical Survey of Contemporary Cybersecurity Products and Solutions. In: Fournaris A. et al. (eds) Computer Security. IOSEC 2019, MSTEC 2019, FINSEC 2019. Lecture Notes in Computer Science, vol 11981. Springer, Cham

Usama M, Asim M, Latif S, Qadir J, Fuqaha A. (2019). Generative Adversarial Networks for Launching and Thwarting Adversarial Attacks on Network Intrusion Detection Systems. 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC). IEEE

Yewale A, Singh M. (2016) Malware detection based on opcode frequency. 2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT). IEEE