Mathematical Modeling for Network Attacks - Blockchain Economy

Nicholas Strange[1, 2, 4, †]

[1]Blockchain and Smart contract Systems Course Research Report
[2]MS in Financial Technology
[3]Computer Science Faculty
[4]University of Central Florida, USA
[*]Email correspondence to: ramya.akula@ucf.edu

July 2023

1 Mathematical Modeling of Network Attacks

1.1 Introduction

The fast growth of computational power and the internet of things has lead to a adjacent increase in network attacks. This has brought about the need for complex mathematical models to be trained and ran against real world network activity data in a controlled environment. Mathematical models have proven to be an invaluable asset in respect to analyzing and predicting data in a vast number of fields. The models work by identifying key relationships and trends in the data that then allow developers to gather insight into the data that were previously improbable to be found by even the most tenured analyst. These models are at their core just simplified representations of otherwise complex data insights and relationships of a selected data set. They allow for not only continuous monitoring on foreseen relationships but also bring about ideas and insights that were not originally scoped out. This sets in to motion a proverbial domino affect of further analysis' and more advanced models. The limitation of these models is by design a continuously moving target and only limited to the technical expertise and creativity of the developer. These mathematical models are often referred to within the machine learning world as the "opened black box" approach and can range from non-linear deterministic models to stochastic distributed models[13]. Each of these models are conducted in the for stages. Those stages are building, studying, testing and use [9]. The applications of these mathematical models have grown exponentially, and the network security industry is just one of the latest to be positively impacted by its seemingly limitless capabilities.

The field of network and cyber security has been in high demand as the number of network attacks and data in storage grow simultaneously. Companies wish to not only ensure that they can fend

off attackers, but that they are well equipped to minimize losses in event that one does invade the network. This increased interest in network security was in part fueled by the 1995 crime committed by Kevin Mitnick in which he had stolen eighty million dollars from multiple companies. This was to the date, the largest cyber-crime to occur within the United States and sparked great interest and fear in companies regarding their network security infrastructure [4]. In the past, internet protocols were not created to secure themselves and therefore left room for a lot of security and data breach vulnerabilities. Modern approaches and techniques have since evolved to offer companies more complex infrastructures and network connectivity requirements. It wasn't until recent years that the innovations in mathematical modeling were realized to have potential applications in network security. These mathematical models offer a more holistic approach to network security optimization. This approach utilizes complex mathematical formulas to create models for analyzing trends and understanding data relationships. In the case of network security, the mathematical models can be applied to better understand and prevent network attacks. The mathematical model applications range from predicting attackers next move based on those that occurred in the past to evaluating and stress testing current defense systems. By applying these models, companies hope to successfully ward off network attack and accept minimal loss if a nefarious agent breaches the network.

The functionality and structure of the internet by design opens up the users to large amounts of security risk. The open source and centralized network designs proves to be invaluable in operating a business, but come with increased exposure to network attacks. These network attacks are traditionally negated through a fairly inadequate reactive approach of adjusting security and protocols according to the network attacks that have occurred in the past. This approach is no longer sufficient, as even a small breach could lead to significant damages and hacker tactics are changing rapidly. The need for a deeper analysis of the attackers network infiltration path, thought process and companies data labeling system is needed to combat these breaches. It has become an industry agreement that the "building block" approach to cyber security is not enough to feel secure in today's environment. As networks increase in complexity and hackers grow exponentially more advanced, the need for a modern approach is increasingly in demand. This new concept is referred to as "Cyber Dynamics", which seeks to encompass the comprehensive mathematical modeling, quantification, and management of network security [24].

The objective of this paper is to provide insight and an in-depth analysis of mathematical modeling applications in the field of network security. Rather than proposing a new model, we will be exploring the effectiveness of previously established models. The application of each of these models will be described in the corresponding sections to provide the reader with real life examples of these complex mathematical model in production. We aim to explain these complicated topics in a manner that leaves the reader with a overview of their applications and a general understanding of their structure. We will not be delving into each of the models mathematical equations but rather how they are applied to network security. This paper is intended for readers that have a general understanding of computational modeling and network security.

The paper is structured to lead the reader through the applications of these mathematical models in the network security industry. The first section consists of relevant data sets and test networks used in building the mathematical models. This section is not all encompassing but rather includes

the most popular data sets that have proven to be most useful in this field. The data sets include those relevant financial information and non-financial data such as network paths. The second section will describe the methodology of these models within the network security realm. In this paper we have chosen to focus on five methodologies of mathematical modeling in combating network attacks. Those are the game theory model, attack graph, advanced statistics and probability, risk assessment and aversion, and machine learning. The third section will include and in depth review of the infrastructure and models established through each of the previously listed methods. The fourth section dives into the mathematical formulas applied to each of the models. We look at the selection of each of formulas and explain their relevancy and intention within the model. The fifth section provides an evaluation of the performance of each model and its ability to combat or prevent network attacks. The sixth section then establishes a quantitative analysis of the models in their performance against hackers and in relation to each other. The seventh section includes a review of successfully applied mathematical models in the network security industry. The eighth section aims to provide an unbiased analysis of the current mathematical models structure and performance. We discuss the limitations and the challenges faces in establishing a successful mathematical model. Finally, we present an overall review of the current mathematical models in their ability to combat network attacks. We also provide a list of further topics identified in our research.

1.2 Dataset

One of the greatest challenges faced when developing a mathematical model of network attacks is the availability and accessibility of real-world data sets. This is in large part due to the nature of network attacks to largely target privately held and personally identifying information. This information ranges from corporate strategies to private communications that are not intended to be publicly available. Those involved in these network attacks would be servery reluctant to release the data infected or stolen by the nefarious actors even if it would assist other in preventing similar attacks. There are also large amounts of legal issues surrounding the publication of the private held data and information regarding the network activity during the attack. Even if the researchers were to get a hold of the data, they would be mistaken to include it in their model, exposing themselves to unnecessary legal troubles. If these legal and privacy issue were not as relevant, the network security industry could avoid many repetitive network attacks. These barriers of data acquisition often steer researchers toward the route of creating their own manipulated data set and simulating a real world network attack. However, in recent years there have been efforts to share and expose this data to the public in an attempt to stay relevant in the cyber security realm and prevent similar attacks on other businesses and individuals from occurring as frequently.

1.2.1 Security focus

Security focus is an open source bugtraq vulnerability notification database that provides insights to network security professionals or enthusiast[18]. The database has been available since 2009 and has been leveraged by security professionals and mathematical model developers alike. One of the large benefits of this platform is the ability of professionals to interact and offer advice regarding

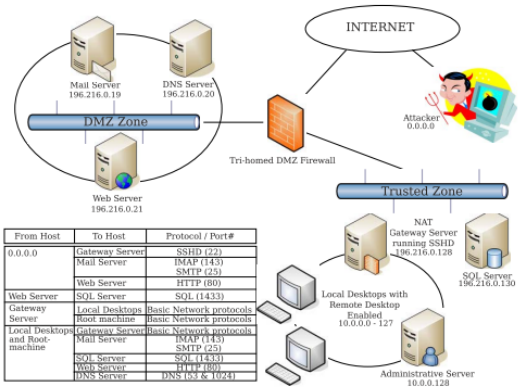

Figure 1: Test Network Attack

network security. This space allows professionals to interact, ask questions and share ideas. They are able to gain insight into possible vulnerabilities to combat and reduce the amount of network attacks throughout the industry. The database works by gathering reports from reputable sources regarding previous attacks and potential network exposures. The sources are pulled from a wide variety of professionals and companies ranging from security vendors, industry experts, institutional researchers and more. This creates a platform for innovation and exchange of ideas in an industry that was previously starved of both. This database enables security professionals to keep up to date with all the latest hacking tactics to reduce the effectiveness of any one hacking strategy. While the hackers are constantly innovating, this systems slows down their progress to give the professionals a competitive edge.

1.2.2 Attack Graph Test Network

Another solution for network security takes the form of a method called the Bayesian Attack Graph. This method was tested by establishing a test network that consisted of two sub-nets and eight local hosts. The network includes a tri-homed firewall that prevents any remote access privileges to the server. The servers were then supplied with multiple SQL databases that included private data to later use for simulating a network attack [15]). By setting up this server, the developers were able to simulate network attacks on the server without the need for external data sources. The test network also partially removes the restriction of testing against a single hacking technique as many data sets include limited real world data. Within this network, the developers can test as many network attack techniques as they please. Below is structure of the test network.

1.2.3 UNSW-NB15 and NIMS Botnet

As previously noted, one of the major challenges in creating mathematical models in the network security industry is the lack of a comprehensive network traffic data set. If the data sets do exist, they are usually severally outdated and do not reflect the current network traffic structure. This results in nonoperational or solely hypothetical models that cannot be put in production. The University

of New South Wales, Australia recognized this data gap and therefore created a database referred to as UNSW-NB-15 [22]. This database is categorized as a network intrusion detection database. The data includes a wide range of modern day network activities. One aspect that sets this data set apart from its peers is the range of network activity included. The data includes not only network attacks, but also normal traffic, making it ideal for building and testing machine learning models. The data is also publicly available, increasing is relevancy in the industry.

Another useful data set for building mathematical models for network attacks is the National Institute of Information and Communications Technology's Botnet data set. This was established using a collection of real-world network traffic activity [21]. The data set includes a large amount of real-world network traffic of bot-net infected host. Similar to the previously described data set, the advantage lies in the ability of users to access and train models against real-world data. And similar to the previous, the data provided in the data set was previously unavailable to the public. This data set is also far superior to its peers due to few aspects. Firstly, most botnet data sets suffer from a generality issue because they only include small amounts of niche network activity. Additionally, most other data sets are recorded in a controlled environment and therefore may not reflect real-world activity as well. Lastly, due to privacy concerns, most other botnet traffic data sets aren't able to provide network traffic traces or gather background data that may be extremely valuable when building the models [1].

1.2.4 CICDDoS2019

The CICDDo2019 data set is specific to distributed denial of service data attacks. These distributed denial of service attacks are a form of cyber attacks that involve compromising multiple systems within a network for the sole purpose of flooding the system with traffic, rendering its ability to validate users. The data set includes a wide range of distributed denial of service network attacks[19]. The data is distinguished between exploitation-based attacks and reflection-based attacks. The training and testing sets within the data set were captured on two consecutive dates. The test and train data sets consist of twelve and seven distributed denial of service network attack types respectively, each grouped in an individual file [7]. The effectiveness of this data set is reflected in it uniqueness and variance of network attack types. The data set also consists of more that seventy five flow features. These qualities have benefited developers and security professional in combating and identifying the characteristics of a distributed denial of service network attack. The data allows them to test defenses against real-life network threats while also injecting innovations to limit the affects of such attacks. The list of different attack can be viewed in the figure below.

1.3 Methodology

In this section we will be discussing the methodology used in reviewing and comparing the mathematical models used in network security. The intent of this study is to analyze how these mathematical models are used to negate network attacks. We also perform an in-depth analysis of the performance effectiveness of these models against the current network security methods as well as against themselves. Due to the vast range of applications in which these mathematical models have been tested

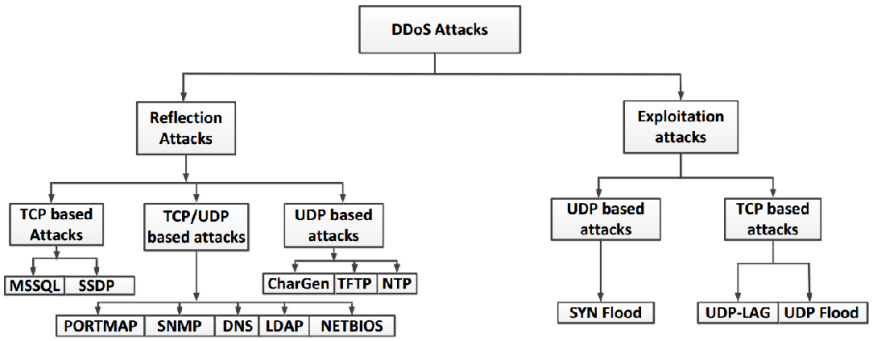

Figure 2: Network Attacks

in implemented in the network security industry, there is no one good way or set of performance metrics to apply to each of these models. For this reason, we chose to select a few of the most effective model applications based on the research of other peer reviewed papers. These papers were selected according to their credibility, number of citations and relevancy.

1.3.1 Architecture/Model

The approach taken within this paper was a mixed method blend between the most effective quantitative and qualitative metrics relevant to the network security industry. This method allowed us to not only identify the most effective uses of mathematical methods in network attacks but also analyze the performance of each of these models in a production environment. Had we have chosen to perform a solely quantitative or qualitative research approach; we would have neglected to fully measure the usefulness of mathematical models in network security. The qualitative research involved in our methodology focused on identifying the best uses of these models in network security. Each model was subject to an extensive filtration system that left only the most popular and industry approved models. We aimed to only include models that have been tested in production and show a significant process or performance improvement. These metrics include the model's ability to identify, combat or ward of network attacks in a corporate setting. The qualitative research within this paper is concentrated on measuring the effectiveness of each of the mathematical models identified.

1.3.2 Research Design

The approach taken within this paper was a mixed method blend between the most effective quantitative and qualitative metrics relevant to the network security industry. This method allowed us to not only identify the most effective uses of mathematical methods in network attacks but also analyze the performance of each of these models in a production environment. Had we have chosen to perform a solely quantitative or qualitative research approach; we would have neglected to fully measure the usefulness of mathematical models in network security. The qualitative research involved in our methodology focused on identifying the best uses of these models in network security. Each model was subject to an extensive filtration system that left only the most popular and in-

dustry approved models. We aimed to only include models that have been tested in production and show a significant process or performance improvement. These metrics include the model's ability to identify, combat or ward of network attacks in a corporate setting. The qualitative research within this paper is concentrated on measuring the effectiveness of each of the mathematical models identified.

Data Collection Methods The mathematical models selected to represent their impact on the network security industry were chosen very selectively. Through a thorough analysis of credible research papers, we were able to identify the most successful applications of mathematical models in network attacks. The process of selecting applications consisted of identifying the most researched and tested mathematical models in network security. After identifying these applications, we selected the most applicable and successful research conducted based on three equally weighted criteria. The first criteria was an analysis of the paper's credibility. The credibility of the research paper was quantified by the institution that supported the publishing, any sponsors of the paper as well as the academic experiences of each of the authors. The next criteria was regarding the number of accredited citations the paper received. The larger the number of citations, the more weight it received. The last criteria was focused on the relevancy of the article in regards to the network security industry.

Ethical Considerations In adherence with research best practices, we made sure to collect and cite past research in an ethically responsible manner. In order to do this consistently and effectively, we established a set of standards for our research. The guidelines we chose to follow were the six ethical "Hot Topics" as described in Hannah Farrimond's book published in 2012 titled "Doing Ethical Research". These topics are listed below. Although not all these topics are applicable to our type of study, we vetted each paper in terms of these guidelines as best we could. 1. Informed Consent. 2. Privacy, Anonymity and Confidentiality, 3. Assessment of Possible Harm 4. Vulnerable Groups and Sensitive Topics, 5. Ethics of Research and Young People 6. Internet Research and Ethics [8].

1.3.3 Technical Configuration

After establishing the top applications of mathematical models in preventing and combating network attacks we began an in-depth analysis of each of the models. As previously mentioned, the applications we identified were the game theory model, attack graph, advanced statistics and probability, risk assessment and aversion, and machine learning. In this section we briefly discuss the technical requirements and makeup of each of these methods.

Game Theory Game theory was originally publicized in a 1944 book "The Theory of Games and Economic Behavior" [20]. The theory essentially claims that the decision of an individual is determined by the actions of another in a strategic decision situation. In the context of network security the game at play is information warfare, in which each side aims to preserve or infiltrate information security [17]. Mathematical models built for this industry mostly involve building a

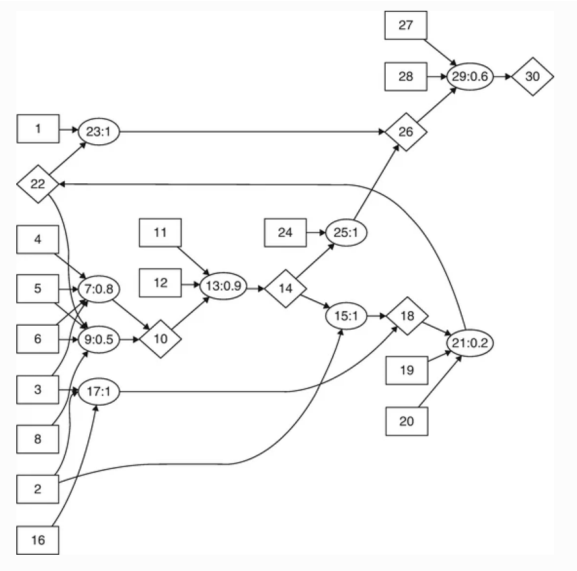

Figure 3: Probabilistic Attack Graph

model of interactions between network attackers and the security professionals. This information is then reworked and thoroughly analyzed to identify tracker trends and security professional's errors.

Attack Graphs Attack graphs are developed to represent all possible paths an attacker can take to attack a secure network. These graphs allow security professionals to scope out all possibilities of an attack to ensure the best defense system. One specific type of attack graph that involves complex statistics and probability calculation is called a probabilistic attack graph. This type of attack graph provides a quantitative representation of the likeliness each individual path has on being infiltrated. The risk for this model is calculated in the probability of the attacker's success given a selected path. Below is an example of an established probabilistic attack graph.[23].

Research Limitations Although we attempted to provide a thorough analysis of mathematical modeling in network security, the result of our research is not all encompassing. Our research consisted of publicized and accredited works that have been reviewed on the topic of mathematical modeling and network security. The nature of our research type is by design limited to the existence of successful models. The number of mathematical modeling applications in network security is fairly limited but growing substantially. Additionally, of the models that have been established and used in a production environment, most are reluctant to publicize the structure due to the sensitive nature of the data. This being said, we feel that we gathered sufficient information to provide an in-depth analysis of mathematical modeling in combating and preventing network attacks.

1.4 Evaluation

Although the integration of mathematical models in the field of network security has been largely beneficial, not all methods are equally effective. Being that the introduction of these models in the industry is relatively recent, it is expected to see larger than normal performance curve tails. As the models are continuously tested and reworked, we assume to see the greater performance. Additionally, due to the constantly changing nature of network attacks, no one model unequivocally reins over the others. Many mathematical models are built for niche attacks and cannot be perfectly compared to another in terms of performance. For these reasons, we chose to evaluate the performance of the different mathematical models with multiple criteria. This approach allowed us to best compare and describe the performance of the models in regard to their written intention. Those are the models ability to prevent network attacks and the models ability to defend against a current network attack. We also provide other layers of the models' performance. Those additional performance metrics are the repeatability of the model, the complexity of the underlying formulas as well as the level of industry adoption.

1.4.1 Performance Preventing and Defending Network Attacks

Mathematical modeling applications in network security can take many forms and result in a wide range performance. However, all of these models are alike in their intention to prevent or defend network attacks. The first evaluation group we will discuss is the performance of the mathematical models in preventing network attacks. This performance metric is of particular use to network security teams because it is the most advantageous result of these models. By creating a model that can successfully prevent network attacks, money, information, and reputation can successfully be preserved. This measure can be further broken down into two categories. The first is the reduction in the number of network attacks to the company or individual. The second is the severity of the attacks upon a successful breach of the network. This performance metric is large majority of network security professionals focus and budget allocation. Successfully preventing network attacks is a far greater priority than defending them upon a breach.

Another metric for building a measuring the performance of a mathematical model in the field of network security is the models' ability to defend against active network attacks. These types of models are heavily sought after but are anticipated to never be activated. These models must not only identify a network attack but also restrict access and remove sensitive information simultaneously. Much of the error regarding these models is in providing timely identification of these nefarious actors while not bogging down servers or the performance of the company's technology. Often times these models are referred to as a last resort security protocol but remain a corporate wide necessity.

One particularly useful model in preventing network attacks involves the creation of a network attack graph. As we previously mentioned, these attack graphs layout all possible routes for network attacks and then run complex statistical analyses against the model to provide likelihood percentages for each path. In production, these models have proven effective in preventing the quantity and effectiveness of network attacks. By forecasting and analyzing the approaches of hackers into the network, companies are able to invest money and resources into the exposed network security

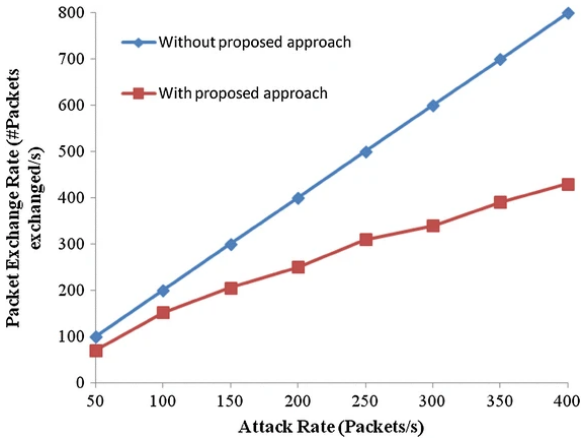

Figure 4: Performance Metrics

deficiencies.

Another network attack strategy that is very common and complicated to combat is a Distributed Denial of Service attack, commonly referred to as a DDOS attack. These attacks are extremely effective because they overload the server and restrict it from validating the user and performing effectively. For this reason, the attack has been a focus of many network security solution architects. One solution proposed for this network attack is a mathematical model. The evaluation of these models consists of the holding time, network attack rate and packet exchange rate. One experiment conducted by the Journal of Ambient Intelligence and Humanized Computing analyzed the performances of the models in combating distributed denial of service attacks. The paper proved that the mathematical models are extremely effective as well as efficient in these tasks. Below is performance metrics visualized with line graphs[2].

1.4.2 Additional Metrics

Although the accuracy of a model is perhaps the most important metric, it is not the sole determinant of a models effectiveness. Additional to the accuracy of the model, the repeatability of that model is extremely important. A model performing successfully against a small set of data in a controlled environment could prove useless in a larger, less structured production environment. For this reason, the models must undergo strenuous testing against large structured and unstructured data sets. Only after these additional tests have been completed and vetted can they have a good understanding of the model's performance. Another important metric for measuring model effectiveness is the reliability of the model structure. This metric determines whether the new model is more accurate than the previous techniques. The reliability of the new model must be increased by automation if data collection and the reduction of the human error element in processing the model. If the model is too prone to human errors, the model cannot be trusted to replace existing protocols. The last metric that is essential in measuring mathematical models in network security is the transparency of the model. A high transparency rating means that the steps within the model are well defined and documented. This is extremely important for the network security industry due to the highly

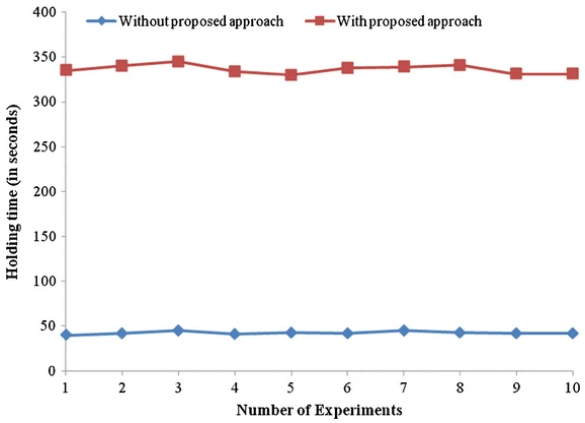

Figure 5: DDOS Performance

$$\begin{cases} R(x) = 0, & x = 0 \quad (a) \\ R(x) \leq R(y), & 0 \leq x \leq y \quad (b) \\ R(ax + by) = aR(x) + bR(y), & a, b \in \mathbb{R} \quad (c). \end{cases} \tag{6}$$

Figure 6: ROI Calculation

sensitive data involved. It is for this reason that the industry is reluctant to embrace "black box" models.

1.4.3 Return on Investment

Perhaps that most important metric involved in scoping out whether a model is worth investing time and money into is the return on investment of that model. This calculation can be particularly difficult in the context of network security because there is likely no projected revenue for the projects. However, with some adjustment and creative thinking we can provide a potential revenue loss calculation to replace the value. One article written by multiple professors across the world in 2013 proposed an executable return on investment function for determining the value of a distributed denial of service attack defensive model. The function is shown below, with x referring to CPU, IO and bandwidth [25].

1.5 Qualitative Analysis

The field of mathematical modeling has been advancing at a rapid pace. As the world grows more complex and computation power becomes more capable, the business applications of these complex models grow seemingly endlessly. The models can help in automating, predicting, defending and

many more otherwise costly and time-intensive business functions. However, like most all things in the commercial world, these mathematical models are also subjected to the realities of inflated applications. For this reason, we held in mind the principles of the Gartner hype cycle when analyzing each of these network security applications. The Gartner life cycle model was publicized nearly twenty years ago and still holds value in today's ever-changing landscape. The model maps the progression of technological innovations as they transition through various stages characterized by the peak of excitement, subsequent disillusionment, and eventual recovery of expectations [5]. With this in mind, we aimed to serve as an unbiased third party to analyze the usefulness of these models in the network security industry, regardless of how innovated or cutting edge they seemed. This enabled us to focus on the performance and usefulness of each model application. In order to best provide an unbiased and in-depth analysis of mathematical modeling's application in network attack, we followed closely the industry standards set by Michael Quinn Patton in the 1999 article titled "Enhancing the quality and credibility of qualitative analysis" [14].

1.5.1 Previous Works

The integrations of these mathematical models are a relatively new tool in detecting and combating network attacks. These models and therefore not as well researched and tested as the previously used network security tactics. However, there have been multiple papers and tests conducted in recent years regarding the performance of these models in the network security industry. One research paper on the matter that was well received and cited was a 2018 paper published in the IEEE Internet of Things Journal. The article titled "An Ensemble Intrusion Detection Technique Based on Proposed Statistical Flow Features for Protecting Network Traffic of Internet of Things" was written by Nour Moustafa, Benjamin Turnbull and Kim-Kwang Raymond Choo [12]. The article proposes an ensembled approach to identifying an intrusion technique applicable for safeguarding network traffic in an internet of things environment. The new approach involves statistical flow features that have increased the effectiveness and accuracy in relation to the previously used methods of intrusion detection. The model also performs more accurately due to the ensemble approach of integrating the most successful methods into a single tool. The selection of the features to be included in the model is very important in building a successful model that can be replicated outside the test environment. This method involves a complex math problem that was used to identify the features that would have the highest return on input for the model. Below we have attached a figure that shows the mathematical functions used. The researchers identified that the best ensemble model for intrusion detection is a combination of the decision tree, naive bayes, and artificial neural network techniques.

1.5.2 Findings and Interpretation

Through our research and analysis of the mathematical models' applications in the world of network security, we determined them to be overall sufficient and effective in replacing previous industry methods. This transition has been more intense and fast paced due to the accelerated need for more advanced techniques in combating network attacks. These network attackers have become increasingly advanced and efficient; therefore, the industries have no choice but to adapt at a similarly

The CC of the features f_1 and f_2 is calculated by

$$\mathrm{CC}(f_1, f_2) = \frac{\mathrm{cov}(f_1, f_2)}{\delta_{f_1} . \delta_{f_2}}$$

$$\mathrm{CC}(f_1, f_2) = \frac{\sum_{i=1}^{N} (a_i - M_{f_1})(b_i - M_{f_2})}{\sqrt{\sum_{i=1}^{N} (a_i - M_{f_1})^2} . \sqrt{\sum_{i=1}^{N} (b_i - M_{f_2})^2}}.$$

Figure 7: Feature Selection Formula

fast pace.

Each of the models we researched and even those that did not make it through the filtration system showed a sustained qualitative or quantitative improvement. These improvements range from identifying malicious actors more accurately to predicting an attacker's next move. Every model tested in a production model showed some level of process improvement. Every company can benefit from instituting a mathematical model into their network security framework.

Although the models unobjectively provide better performance than the average network security protocol, this does not mean that each company should implement each and every mathematical model shown to provide value. As mentioned earlier, there are many additional considerations to take into consideration when deciding whether or not to implement a model. For example, not all networks are prone to distributed denial of service attacks (DDOS), and therefore they might benefit from allocating funds to a more common hacking technique in their industry.

1.5.3 Validity and Reliability

Our research not only analyzed the performance of the mathematical models in combating and/or defending against network attacks, but also in terms of validity and reliability of the model. The validity of a model refers to the accuracy of a chosen measure. In terms of the research conducted by our team, the validity of the research papers was defined by additional research into the measures selected to quantify the results. This led to a large number of papers being discarded for favoring self-promoting model measures. The next aspect is the model's reliability, which refers to the consistency of the measure. This also filtered out many papers because they did provide sufficient evidence of repeated success and averaged performance. The papers that did pass both the validity and reliability test showed strong well tested and repeatable functionality.

1.5.4 Implications and Recommendations

Overall, most companies would benefit from instituting some form of mathematical model in their effort to defend or prevent network attacks. After conducting our research on the topic, we would

suggest a few tips to those looking to implement one of these models. First, take into account any previous network attacks or attempts made on your network and categorize them according to the method used. Next, look up the most common network attacks happening today. Start locally, within your industry and then branch out. After you have identified the most common attacks, identify the functionality you would like from the model. For example, you might be looking to identify attackers before they infiltrate the network. Now you can start identifying the most popular mathematical models according to your needs. Once a model is selected, set up a test environment with real network activity and data for testing. If the model provides repeated sufficient results in testing, the production version can be stood up.

1.6 Application

The implementation or integration of a mathematical model in the network security industry is unobjectively advantages. These models have been proven to increase process efficiency, detect and defend against network attacks and much more. They have also resulted in substantial cost savings for those companies that have implemented the models in production. Although there are countless means of applying these models, in this paper we discuss a few that have proven to be the most widely accepted. Those applications involve the game theory approach, attack graphs, advanced statistics and probability, general risk assessment and aversion and machine learning.

1.6.1 Game Theory

Game theory is a branch of mathematics and economics that studies how decision making is affected in situations in which a player's decisions are based on the choices of another player. This theory is relatively new to the network security industry and has offered very significant contributions to the framework. Within the field of network security, the players are generally the hackers and cyber security networks, the payoff is the access of, or retention of confidential information and funds and the strategies vary by attack type.

A paper written by School of Engineering and Information Technology, Australian Defense Force Academy and the University of New South Wales Canberra researched the applications of this Game theory in network security. The paper describes the game as information warfare in which the goals consist of the following six accomplishment:

1. To increase the availability of information to an attacker (offensive); 2. To decrease the availability of information to a defender(offensive); 3. To decrease the integrity of information (offensive); 4. To protect information from an attacker (defensive); 5. To protect the availability of information to a defender (defensive); 6. To protect the integrity of information (defensive). [11].

The qualitative results of implementing a game theory are further described in the paper and all provide substantial value to the companies bottom line and public branding. However, the applications of these models are not only production level information security tools. Many of these game theory models are used to train and test employees' responses to network attacks.

$$r = \frac{1-\eta}{\eta} \sum_{m=1}^{\infty} \eta^m \sum_{n=0}^{m} P^n s$$

Figure 8: Reachability Probability

1.6.2 Attack Graphs

Attack Graphs, like game theory models aim to detect network intrusions and predict their movements. The introduction of these attack graphs has far proven invaluable to the network security industry as a whole. These models rely on advanced statistics and probabilities to visually represent the path of a hacker within a model as well as identify and quantify the most likely route. One thing that sets these models apart from their peers is the ability of the model to quantify the movements and thought process of the hackers. Once complete, security professionals have a better understanding of any vulnerabilities in the network. This can lead to bug fixes, patches or perhaps a more informational approach to researching the relevant mathematical models most advantageous to the companies' network. Before attack graphs, companies relied on model checking to enumerate through attack sequences, as described in the paper written by Ronald Ritchey [16].

A conference paper analyzing the usefulness of these attack graphs managed to quantify and compare the performance of these complex visualizations [10]. The article uses two algorithms to determine the probability of an attacker reaching a specified state. The first algorithm is similar to Google's page rank algorithm, while the second calculated the likelihood of an attacker reaching a randomized state. Each ranking system showed significant evidence of the importance of these attack graphs in the network security industry.

1.6.3 Advanced Statistics and Probability

Most mathematical models are applied to a specific model that has previously been tested and proved to improve network security, such as the attack graphs and game theory models. However, these models can be applied in a more general manner such as explorations and testing functions. What makes these models so beneficial to the instituting companies is the ability to streamline and automate extremely difficult statistic and probability functions. The statistical applications of these models can range from quantifying network vulnerabilities to testing the efficiency of departments to identify and categorize a network attack. Advanced probability formulas can also be incorporated into the network security to benchmark and monitor network attack preparedness. The implementation of the advanced statistic and probability models saves companies time, money and removes the human error deficiency from otherwise tedious processes.

1.6.4 Defense and Prevention

The main applications of mathematical models in the network security industry fall within two categories of the models' purpose. The first is a model that can effectively defend against an ongoing network attack. These models require great resources and continuous upgrades to their operating

Manhattan norm

Example 3.1 (Manhattan Norm)
The *Manhattan norm* on $\mathbb{R}^n$ is defined for $\boldsymbol{x} \in \mathbb{R}^n$ as

$$\|\boldsymbol{x}\|_1 := \sum_{i=1}^{n} |x_i|\,, \tag{3.3}$$

where $|\cdot|$ is the absolute value. The left panel of Figure 3.3 shows all vectors $\boldsymbol{x} \in \mathbb{R}^2$ with $\|\boldsymbol{x}\|_1 = 1$. The Manhattan norm is also called ℓ_1 *norm*.

ℓ_1 norm

Euclidean norm

Example 3.2 (Euclidean Norm)
The *Euclidean norm* of $\boldsymbol{x} \in \mathbb{R}^n$ is defined as

$$\|\boldsymbol{x}\|_2 := \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}} \tag{3.4}$$

Figure 9: Analytic Geometry

systems to keep up with the innovations in cyber attackers. This type of modeling is commonly referred to as a last resort defense that companies hope to never initiate. The next commonly sought after model type is one that can successfully prevent network attack gaining access to the system. These are the most commonly referenced models in the network security industry because they provide a proactive solution. Almost all companies have some sort of network monitoring system that involves a high degree of back end mathematical formulations. Some models have performed effectively in both defending against and preventing network attacks.

1.6.5 Machine Learning

The last application of mathematical models in the network security industry is machine learning. Machine learning has taken the world by storm with the help of modern innovations that substantially increased performance and decreased the technical barriers to entry. These machine learning models can be applied to defend against and prevent network attacks. There are many types of machine learning techniques, many of which require extensive mathematical formulations. In a book written by Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong the importance of mathematics in machine learning is clearly and effectively defined [6]. The categorizes the applications of mathematical modeling by linear algebra, analytic geometry, and matrix decomposition. Examples of analytical geometry in machine learning are the Manhattan Norm and the Euclidean Norm. These are both very commonly used in machine learning. Another aspect of machine learning is that aids in its performance is the ability to combine multiple models together in an ensemble model to ensure the best possible performance and prevent overfitting.

1.7 Challenges/Limitations

As with the majority of studies, our research on the application of mathematical modeling for network attacks is subject to challenges and limitations. While we stand strongly by our conviction that the study was meticulously conducted and yielded import results, we are acutely aware of the presence of certain limitations within our research. By thoroughly revealing and explaining the limitations, we hope to provide a comprehensive understanding of the current state of mathematical models in the network security industry. We believe that the approach reinforces the integrity of our research and supports our readers in understanding the topic. We also hope that being transparent with our study will help promote improvements and additional studies to be conducted by fellow researchers. The challenges of the study are a combination of general research related issues and those specific to mathematical modeling and network security. We have categorized the challenges by research limitations and application limitations.

1.7.1 Research Limitations

The first set of challenges and limitations that we faced is regarding the research conducted. These challenges are similar to many other studies and many of which were unavoidable given the circumstances and topic at hand. One of the main issues surrounding our research was the availability and relevancy of data. Due to the highly sensitive and ever-changing nature of the network security industry, finding sufficient data sets was extremely challenging. Many data sets that appeared relevant were severely outdated or niche in nature. We overcame this by increasing our attention on the results of successful mathematical models and adding supporting evidence from relevant data sets when applicable. Another challenge faced in our research was the time constraint of the research project. The project was to be conducted over the course of a couple of months and therefore resulted in relatively limited research. Given additional time, the study would and could include a deeper understanding of the mathematical formulations and why they are optimal over the results of similar methods. The restraint of human capital was also a large issue. The number of researchers available on the topic of mathematical modeling was very limited and possibly resulted in a personal bias or expertise constraint. Additionally, the ethical considerations of presenting relative private information and company specific network vulnerabilities restricted the amount of applicable model representations in that paper. The last important limitation that we experienced is a product of the previously stated challenges within the field of network security. This was the limitation of relevant and up to date peer reviewed papers on the application of applied mathematical models that have been successful in preventing and/or combating network attacks.

1.7.2 Application Limitations

The second set of challenges and limitations we identified in our research are those specific to the implementation of these mathematical models in the network security industry. Although the models are beginning to be more widely accepted, they do face a large amount of resistance. The network security industry is notoriously resistance to new and untested technologies. This is due to the highly sensitive and fragile structure of the underlying data and infrastructure. One mistake or

a technology implemented without sufficient due diligence could result in a substantial loss. The losses are both financial losses as well as trust and reputation losses. This is far more concerning for consumer companies who require consumer trust for market adoption. For these reasons, the implementation of mathematical models has been relatively slow in this market relative to the other industries. These models also require constant research, development, and testing. Without large market adoption, the models will enhance slowly, perhaps too slow to keep up with the attackers' innovative strategies. In order to guarantee that these mathematical models are able to keep up to date with the latest hacking trends, they also require relevant network data. These data sets are few and far between and most that are available are outdated or niche in nature. The problem consists of two issues. The first I that there is not sufficient public data to train on and that most organizations are reluctant to release their data to the public due to privacy and vulnerability concerns. A private blockchain network could solve this problem and send the network security industry to the next level of cyber defense.

Mathematical models are by nature very complicated and require a high degree of skill to build and modify. These highly narrow and in demand skills make it very difficult and expensive to build a mathematical model. One large aspect of this challenge is finding qualified candidates. Most of the people qualified to build these models hold a STEM degree from an accredited university. According to an article written on the role of STEM degrees, "The proportion of students who graduate with degrees in science and engineering has consistently oscillated around 16 percent of college graduates for the past 10 years" [3]. This percentage is even smaller when you narrow down math related majors. This relatively small talent pool has made it difficult for companies to build these models. Even if a company is able to find qualified engineers, they must be willing and able to sink large amounts of money into the research and development of these models. This skill requirement challenge has been eased as large companies release versions of mathematical models and open-source software's continue to develop. This enables companies to utilize and build off the complex mathematical models built by larger organizations, saving on cost, labor, and time.

The last major challenge facing the successful adoption of these mathematical models in the network security realm is the success of the prior security protocols. The development of network security structures is constantly analyzed but stay relatively the same. This is in part due to the nature of the network security industry to protect the current state of the business and data at all costs. Many companies that have not recently been a victim of a network attack assume that their systems must be sufficient in deferring the hacker. This is a very reactive and insufficient approach and as the number of attacks has increased, so has the desire of companies to incorporate more modern approaches to network attack defenses.

1.8 Conclusion

As the corporate world becomes more digital and companies shift towards virtual storage, the need for advanced network security models becomes evermore clear. The shift of companies to the cloud offers great advantages such as decreased fixed cost and quicker innovation. However, this structure further exposes businesses to network attacks. The need for investment in a strong and malleable

network security tool is in dire need. This is why we propose that all companies invest in some variable of a mathematical model for network attack defense. These systems have been proven to provide value and resilience in an ever-changing market. The implementation of one of these models through a reliable vendor or in-house department can provide considerable value to a company or organization. Our research into the relevancy and successful implementation of these models in the network security industry resulted in a favorable analysis of the models. When successfully implemented, these complex mathematical models are able to identify hackers, prevent attacks and much more. Although the barrier of entry into developing and thoroughly testing these models is relatively height the results soundly justify the effort and investment.

Due to the complexity of the mathematical models and the slow market adoption within the network security industry, we felt our effort was best allocated to providing an overview of the models and how they have been performing. An in-depth analysis of the mathematical functions within each of these models would likely have deterred adoption further. Our intention within this paper is to provide sufficient evidence of the relevancy of these models within the network security industry and promote industry adoptions. This process involved identifying the most reputable data sets and research papers regarding mathematical models and network attacks. Although the availability of this information was limited due to the complex and private nature of the network security industry, we found ample resources for our research. The datasets and research papers were also subjected to a high degree of scrutiny to ensure only the most relevant were included in our findings.

The application of mathematical models in network security has proven to be very successful in combating and preventing network attacks. The development of these models requires a large investment in time, salaries and other research and development costs, but remains justified. In order to build and test these models, you need access to a large amount of network history data. This data needs to be up to date and include a wide range of network traffic, both nefarious and general. The obtainment of such a data set has been a barrier for most companies looking to implement a mathematical model into their network security infrastructure. Relying on one's own data requires an attack on the system, which companies are aiming to avoid. The data also creates an overfitted model to the company's past structure and attack type history. However, this data availability has been improved due to open-source sites and third-party network security vendors.

The companies that are fortunate enough to afford the labor and technology necessary for implementing these complex mathematical models have reaps tremendous benefits. Although the possibilities that these models provide for the network security industry are countless, mass adoption has been concentrated in a few areas. These areas differ greatly in structure and complexity, but all have the same accomplishment target of preventing and/or defending against network attacks.

The first area that has received a large amount of attention is the integration of a game theory approach to network security models. The game theory design allows security professionals to better understand the though process of the hackers. This information assist companies in preparing for network attacks and blocking off attackers before the progress through the network. The game theory model has also been applied to training network security professionals by running scenarios test and identifying vulnerabilities. The implementation of these models requires a large degree of

mathematics, specifically advanced statistics, and probability.

Another successful implementation of a mathematical model utilizes attack graphs to identify an attacker's path through the network. With the help of our advanced statistics and data visualization tools, these attack graphs allow security professionals to visualize their infrastructure and identify the most likely route of an attacker within their network. This provides a previously unimaginable advantage to the network security industry over their highly advanced opponents.

Each of the applications that we chose to analyze provided the implementing organization with a substantial technological advantage and performance improvement. The integration of a mathematical model in the network security protocols is advantageous to companies of any size and structure. The models can not only defend and identify network attacks far more effectively than the current methods, but a well-tested model can also save a company large network infrastructure and labor costs.

Although the success of these models in combating network attacks has been proven consistently effective, they are not limitless or without challenges. As we previously stated, the implementation of such a model is severely complicated. The key challenges that an organization will face are affording and identifying qualified staff, gaining access to relevant data sets, and gaining executive approval. However, these challenges are becoming less resistant as models move to open-source platforms and market adoption accelerates.

Upon completion of our research, we came across a large number of research gaps on the topic of mathematical modeling for network attacks. In the consideration of limited time and resources, we were not able to provide analyses on the identified discrepancies, but rather hope to spark the interest of future researcher to the topic. One gap we identified was the lack of a well-researched study on identifying the reasoning behind a developer's selection of a mathematical formula over another. Another area that could benefit from additional research is the risk and payoff of a company releasing network history for developers to create, train and test network security models. The last suggestion we have for future research is a study regarding the development of a private blockchain system between banks to share network history in an effort to better develop mathematical network security models.

References

1. Beigi EB, Jazi HH, Stakhanova N, and Ghorbani AA. Towards effective feature selection in machine learning-based botnet detection approaches. In: *2014 IEEE Conference on Communications and Network Security*. IEEE. 2014:247–55.
2. Bhushan K and Gupta BB. Distributed denial of service (DDoS) attack mitigation in software defined network (SDN)-based cloud computing environment. Journal of Ambient Intelligence and Humanized Computing 2019;10:1985–97.
3. Bottia MC, Stearns E, Mickelson RA, and Moller S. Boosting the numbers of STEM majors? The role of high schools with a STEM program. Science Education 2018;102:85–107.

4. Daya B. Network security: History, importance, and future. University of Florida Department of Electrical and Computer Engineering 2013;4.

5. Dedehayir O and Steinert M. The hype cycle model: A review and future directions. Technological Forecasting and Social Change 2016;108:28–41.

6. Deisenroth MP, Faisal AA, and Ong CS. Mathematics for machine learning. Cambridge University Press, 2020.

7. Elsayed MS, Le-Khac NA, Dev S, and Jurcut AD. Ddosnet: A deep-learning model for detecting network attacks. In: *2020 IEEE 21st International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*. IEEE. 2020:391–6.

8. Farrimond H. Doing ethical research. Bloomsbury Publishing, 2012.

9. Marion G. An Introduction to Mathematical Modelling. Research 2008:1–35.

10. Mehta V, Bartzis C, Zhu H, Clarke E, and Wing J. Ranking attack graphs. In: *International Workshop on Recent Advances in Intrusion Detection*. Springer. 2006:127–44.

11. Merrick K, Hardhienata M, Shafi K, and Hu J. A survey of game theoretic approaches to modelling decision-making in information warfare scenarios. Future Internet 2016;8:34.

12. Moustafa N, Turnbull B, and Choo KKR. An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of internet of things. IEEE Internet of Things Journal 2018;6:4815–30.

13. Myung IJ and Pitt MA. Mathematical modeling. Stevens' handbook of experimental psychology 2002;4:429–60.

14. Patton MQ. Enhancing the quality and credibility of qualitative analysis. Health services research 1999;34:1189.

15. Poolsappasit N, Dewri R, and Ray I. Dynamic security risk management using bayesian attack graphs. IEEE Transactions on Dependable and Secure Computing 2011;9:61–74.

16. Ritchey RW and Ammann P. Using model checking to analyze network vulnerabilities. In: *Proceeding 2000 IEEE Symposium on Security and Privacy. S&P 2000*. IEEE. 2000:156–65.

17. Roy S, Ellis C, Shiva S, Dasgupta D, Shandilya V, and Wu Q. A survey of game theory as applied to network security. In: *2010 43rd Hawaii International Conference on System Sciences*. IEEE. 2010:1–10.

18. security focus bugtraq vulnerability notification database. 2009. URL: `http://www.securityfocus.com/archive`.

19. Sharafaldin I, Lashkari AH, Hakak S, and Ghorbani AA. Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy. In: *2019 International Carnahan Conference on Security Technology (ICCST)*. IEEE. 2019:1–8.

20. Simon HA. Theory of Games and Economic Behavior. 1945.

21. The-NIMS-Dataset. 2018. URL: `https://projects.cs.dal.ca/projectx/Download.html`..

22. The-UNSW-NB15-Dataset. 2018. URL: https://www.unsw.adfa.edu.au/australian-centre-for-cybersecurity/cybersecurity/ADFA-NB15-Datasets/..

23. Wang L, Jajodia S, Singhal A, Singhal A, and Ou X. Security risk analysis of enterprise networks using probabilistic attack graphs. Springer, 2017.

24. Xu S. Cybersecurity Dynamics: A Foundation for the Science of Cybersecurity. In: *Proactive and Dynamic Network Defense*. Ed. by Wang C and Lu Z. Cham: Springer International Publishing, 2019:1–31. DOI: 10.1007/978-3-030-10597-6_1. URL: https://doi.org/10.1007/978-3-030-10597-6_1.

25. Yu S, Tian Y, Guo S, and Wu DO. Can we beat DDoS attacks in clouds? IEEE Transactions on parallel and distributed systems 2013;25:2245–54.