Proposal for Kelly Criterion-Inspired Lossy Network Compression for Network Intrusion Applications

Sidney C. Smith
Sidney.c.smith24.civ@mail.mil
Computational Information Sciences Directorate
U.S. Army Research Laboratory
Aberdeen Proving Ground, MD 21005, U.S.A

Robert J. Hammell II
rhammell@towson.edu
Department of Computer and Information Sciences
Towson University
Towson, MD 21252, U.S.A

Abstract

This paper describes a proposal for a Kelly criterion inspired compression algorithm to be used in distributed network intrusion detection applications. Kelly's algorithm instructs a gambler how much to bet based upon the chance of winning and the potential payoff. There has been a significant amount of research into anomaly detection algorithms that will provide some indications of the maliciousness of a network session. This work proposes to combine expert knowledge, data mining, and best of breed anomaly detection algorithms to determine the likelihood that a session is malicious. Further, this work proposes using a Kelly criterion inspired algorithm to select which sessions and how much of each session to transmit. The expectation is that this will minimize the total amount of traffic transmitted while maximizing the amount of malicious traffic transmitted.

Keywords: lossy compression, network intrusion detection, Kelly criterion, anomaly detection, data mining.

1. INTRODUCTION

Distributed Network Intrusion Detection Systems (NIDS) allow a relatively small number of highly trained analysts to monitor a much larger number of sites; however, they require information to be transmitted from the remote sensor to the central analysis system (CAS). Unless an expensive dedicated NIDS network is employed, this transmission must use the same channels that the site uses to conduct their daily business. This makes it important to reduce the amount of information transmitted back to the CAS to minimize the impact that the NIDS has on daily operations as much a practical.

One popular strategy for implementing a distributed NIDS is to do all of the intrusion detection on the sensor and send only alerts to the CAS. (Roesch, 1999) (Paxson, 1999) A second strategy might be to use lossless compression to reduce the size of the data returned to the CAS. A third strategy is to implement some form of lossy compression algorithm to send back relevant portions of traffic.

There are three problems with sending only alerts to the CAS. The first is that it has the potential to over burden the sensor's CPU and introduce packet loss. Smith et al. discovered that the impact of packet loss can sometimes be quite severe for even small rates of packet loss. (Smith, Hammell, Parker, & Marvel, 2016) (Smith & Hammell, An Experimental Exploration of the Impact of Sensor-Level Packet Loss on Network Intrusion Detection, 2015) (Smith, Wong, Hammell, & Mateo, 2015) (Smith, Hamell, Wong, & Mateo, 2016) The second problem is that the alerts by themselves often do not contain enough information to determine whether the attack was successful. The third problem is that these systems are most often implemented with signature based intrusion detection engines. Signature based systems may be tuned to produce few false positives; however, they are ineffective at detecting zero-day and advanced persistent threats. (Kemmerer & Vigna, 2002)

The second strategy is to use lossless compression; however, one of the most widely used is deflation which is a variation of the LZ77 algorithm described by Ziv and Lempel. (Ziv & Lempel, 1997) Compressing the 2009 Cyber Defense Exercise dataset (Sangster, et al., 2009) with GNU Zip provides a ratio of 56.4%. Years of providing computer network defense services has taught us that to minimize the impact of NIDS on day-to-day operations, compression ratios of less than 10% are required. Lossless compression alone will not provide a reasonable solution.

The third strategy is to use lossy compression to provide a solution. Network traffic may be considered to be composed of sessions that span spectrums from known to unknown and malicious to benign as illustrated in Fig. 1. Quadrant III, the known malicious quadrant, is the domain of intrusion prevention systems as described by Ierace, Urrautia, and Bassett. (Ierace, Urrutia, & Bassett, 2005) This research is most interested in quadrant II, the unknown malicious quadrant, because that is the quadrant where evidence of zero-day and advanced persistent threat attacks will be found. In 2004, Kerry Long described the Interrogator Intrusion Detection System Architecture. (Long K. S., 2004) In this architecture, remotely deployed sensors, known as Gators, collect network traffic and transmit a subset of the traffic to the analysis level. Interrogator employs "a dynamic network traffic selection algorithm called Snapper." (Long K. S., 2004) Long and Morgan describe how they used data mining to discover known benign traffic that they excluded from the data transmitted back to the analysis servers. (Long & Morgan, 2005)

This research proposes to combine expert knowledge, data mining, and best of breed anomaly based NIDS solutions to compute a maliciousness factor. It then proposes to feed this maliciousness factor into a Kelly criterion (Kelly, 1956) inspired algorithm to compute the amount of traffic in each session that will be transmitted to the CAS. This should produce a lossy compression of the network traffic designed to reduce the amount of benign traffic and maximize the amount of malicious traffic being sent to the CAS.

	Malicious	Benign
Unknown	II	I
Known	III	IV

Figure 1 Network Traffic Composition

The remainder of this paper is organized into the following sections. Section 2 provides background. Section 3 will outline the approach chosen to address this problem. Section 4 will provided expected and preliminary results. Finally, Section 5 will conclude by restating the goals and approach of this research.

2. BACKGROUND

This research is broken down into to 2 basic questions: 1) is it possible to combine expert knowledge with existing data mining and anomaly detection techniques to compute a useful maliciousness score for network traffic? and 2) is it possible to develop a Kelly criterion inspired algorithm that will take the maliciousness score from research question 1 and determine how much traffic from each session should be sent to the CAS in order to minimize the amount of traffic transmitted while maximizing the amount of malicious traffic which is available for analysis? The answer to the first question will be obtained by exploring expert knowledge, data mining and anomaly detection solutions. The answer to the second question will be obtained by exploring the application of the Kelly criterion. The review of the literature presented demonstrates a wealth of knowledge in each of these areas that may be leveraged for the maliciousness factor.

Session Rating

Data Mining

Lee and Stolfo used RIPPER (Cohen, 1995) on Tcpdump (Jacobson, Leres, & McCanne, 1989) data in their paper, "Data Mining Approaches for

Intrusion Detection." (Lee & Stolfo, 1998) The dataset they used from the Information Exploration Shootout (Grinstein, Laskowski, Wills, & Rogowitz, 1997) contained only the header information for the network traffic and no user data. Lee and Stolofo cooked the network traffic down into records that look very much like Cisco NetFlow (Claise, 2004) records. Then they were able to feed this information in to RIPPER to generate rules. Their initial efforts were unsuccessful; however, once they added a time window into their analysis they were able to achieve promising results. Since their data only contained Internet Protocol header information, and the positions of the exploits were not available to them, they were not able to assess the accuracy of their results.

While developing the Intelligent Intrusion Detection System at Mississippi State University, Bridges et al. integrated fuzzy logic, association rules, and frequency episodes data mining techniques to increase the flexibility of the system. (Luo, 1999) Genetic algorithms were employed to tune the membership functions of the fuzzy logic. (Bridges & Vaughn, 2000)

Dokas et al. addressed the problem of skewed class distribution in mining data for network intrusion detection. This problem exists because malicious activity compromises less than 2% of the network traffic. Their solution was to apply several boosting strategies to classification algorithms for rare classes as part of the Data mining in Minnesota Intrusion Detection System (MINDS). (Dokas, et al., 2002)

In the US Army Research Laboratory technical report, ARL-TR-4211 "Using Basic Data Mining Techniques to Improve the Efficiency of Intrusion Detection Analysis (Long & Morgan, 2005)**"**, Long and Morgan describe mining the Interrogator database to discover known benign traffic to be excluded from the traffic transmitted to the CAS. Their strategy was to exclude the most common day to day traffic flowing to and from the most popular trusted sites. (Long & Morgan, 2005)

Anomaly Based Network Intrusion Detection

In their history and overview of intrusion detection, Kemmerer and Vigna confirm a long standing belief that although anomaly detection techniques are capable of detecting unknown attacks, they pay for that capability with a high false positive rate. (Kemmerer & Vigna, 2002) In traditional NIDS, high false positive rates drain valuable time for the analysts.

In the computation of a maliciousness factor, false positives simply increase the amount of traffic transmitted. This is a cost to be considered; however, it is a much smaller price to pay than that paid by generating an alert for someone to analyze. This means that a significantly higher false positive rate can be tolerated in this application, making algorithms that would be unusable for detection attractive for rating the likelihood that traffic is malicious.

There has been a significant amount of work using anomaly detection in NIDS applications. Garcia-Teodoro et al. reviewed various types of anomaly-based detection techniques categorizing them as either statistics-based, knowledge-based, or machine-learning based. (Garcia-Teodoro, Diaz-Verdejo, Macia-Fernandez, & Vazquez, 2009)

In 1994 Mukherjee et al. provide a survey of intrusion detection technology titled, "Network Intrusion Detection." (Mukherjee, Heberlein, & Levitt, 1994) By today's standards the title is somewhat deceiving because almost all of the systems they surveyed are what would now be called host-based intrusion detection systems. These systems tend to examine the individual system's audit logs looking for intrusive activity. The notable exception is Network Security Monitor (NSM). NSM employs a System Description Language which is roughly modeled after a programming language and is used to describe the complex relationship which may be inferred from observable objects. These complex objects are analyzed using behavior-detection functions. NSM implements isolated object analysis and integrated object analysis. (Heberlein, et al., 1990) (Heberlein, Levitt, & Mukherjee, 1991) (Heberlein, Mukherjee, Levitt, Dias, & Mansur, 1991)

Sekar et al. describe their experiences with specification-based intrusion detection. They created behavioral monitoring specification language that they compiled into detection engines (Sekar & Uppuluri, Synthesizing Fast Intrusion Prevention/Detection Systems from High-Level Specifications, 1999) (Uppuluri & Sekar, 2001) (Sekar, et al., 2002)**,** validating their approach using the DARPA dataset. (Lippmann, et al., 2000)

Eskin et al. describe an unsupervised anomaly detection framework where network connections are mapped to a feature space and either cluster-based, k-nearest, or support vector machine-based algorithms are used to find anomalies in the sparse spaces. One of the key advantages to their approach is that it does not required labeled or known normal data to train the engine. (Eskin, Arnold, Prerau, Portnoy, & Stolfo, 2002)

Kruegel et al. developed a service specific anomaly detection engine. This engine contained a packet processing unit and a statistical processing unit. The packet processing unit pulled packets from the network and reassembled them into service requests. The statistical processing unit measured the type of request, length of request, and content of the request. It then computed values that ranged from 1 to 15 for each of these aspects, such that greater deviation translated into higher numbers. These values were then combined to provide an anomaly score. This score was compared against a standard that the author suggested should be set, so that the system produces no more than 15 false positives a day. Because the deviation in type, length, and content varies significantly between services and even the types of requests, the statistical data must be partitioned by service and the length and content by type; however, the algorithms may be used without change by any service. Although the packet processing unit may need to be adjusted per service. (Krugel, Toth, & Kirda, 2002)

Ertoz et al. describe the MINDS. (Ertoz, et al., Detection and summarization of novel network attacks using data mining, 2003) (Chandola, Eilertson, Ertoz, Simon, & Kumar, 2007) (Ertoz, et al., Minds-minnesota intrusion detection system, 2004) MINDS uses Cisco NetFlow (Claise, 2004) data to collect statics for sixteen different features; half observed and half computed for each session. For each session the local outlier factor is computed. Sessions with features that contain very large local outlier factors are considered anomalous. These sessions then undergo associated pattern analysis which provides a summary of highly anomalous traffic for the security analyst. (Ertoz, et al., Detection and summarization of novel network attacks using data mining, 2003)

Munz et al. describe anomaly detection using K-means clustering. (Munz, Li, & Carle, 2007) Similar to Mukherjee et al. they separate the analysis for each service or port. Similar to Ertoz et al. they work with Cisco Netflow data. (Claise, 2004) Unlike the solutions mentioned above, this one requires both normal and attack training data to establish initial clusters. New traffic is then compared to the established clusters. (Munz, Li, & Carle, 2007)

Yassin et al. describe an approach which combines K-means clustering and naive Bayes classification called KMC+NBC. They were able to validate their algorithm against the ISCX 2012 Intrusion Detection Evaluation Dataset (Shiravi, Shiravi, Tavallaee, & Ghorbani, 2012) with strong positive results. (Yassin, Udzir, Muda, & Sulaiman, 2013)

These references show a considerable amount of research has been done using both data mining and anomaly detection to discover malicious network traffic. It is our intention of evaluate these techniques and use one or more to compute a maliciousness score for each session in the network traffic.

Session Selecting

In 1956 while working for Bell Telephone Laboratories, Kelly was developing a way to assign a value measure to a communication channel. He described a hypothetical illustration of a gambler who received advance notice about the outcome of an event through a communication channel with a non-negligible error rate. By doing this, Kelly was able to assign a cost value to the communication achieving his original goal. At the same time, he developed a formula based upon the probability of winning and the rate of pay off that would provide an amount to bet l that, if bet consistently over time would achieve and maintain greater wealth than any other value of l. This is shown in Eq. 1. where l is the fraction of wealth to bet, p is the probability of winning, and b is the net odds of the wager. (Kelly, 1956)

$$l = \frac{p(b+1)}{b} \qquad (1)$$

Breiman uses the Kelly's work while discussing optimal gambling systems. (Breiman, 2012) He considers the problem of how much to bet on a series of biased coin tosses. To maximize returns on each toss one would bet their entire fortune; however, this will ultimately ensure ruin. In order to maximize winning and avoid ruin, some fixed fraction of wealth will be bet at each iteration. He uses Kelly's work to discover that fixed fraction. (Breiman, 2012)

Thorp first wrote about applying mathematical theory to the game of Black Jack in the 1960 paper, "Fortune's Formula: The Game of Blackjack." (Thorp E. O., Fortune's formula: The game of blackjack, 1960) Later Thorp published the book, *Beat the Dealer*, where he referred to what he called, "The Kelly Gambling System." (Thorp E. O., Beat the dealer, 1966) Although he mentions using the Kelly criterion as the optimal way to bet in his research for *Beat the Dealer* in his later work, (Thorp E. O., Understanding the Kelly Criterion, 2012) he mentions it only once in passing in this book. (Thorp E. O., Beat the dealer, 1966) The bulk of this book discusses the rules of Blackjack and methods to determine when one has an advantage over the dealer and how great that advantage might be. The Kelly criterion would be used to calculate how large of a bet to place based upon the size of the advantage. Instead of directly using the Kelly criterion, he talks about placing big bets and little bets. (Thorp E. O., Beat the dealer, 1966)

In his paper "Understanding the Kelly Criterion", Thorp mentions the application of the Kelly criterion to the stock market and his previous book *Beat the Market* (Thorp E. O., Understanding the Kelly Criterion, 2012); however, the Kelly criterion is not mentioned at all in *Beat the Market*. Instead Thorp concentrates on how the market works, what short selling and warrants are all about, and how to determine the relative value of a stock or a warrant. (Thorp & Kassouf, Beat the Market: A Scientific Stock Market System, 1967) Thorp goes into greater detail about how the Kelly criterion would be used in Blackjack and the stock market in "Optimal Gambling Systems for Favorable Games." (Thorp E. O., Optimal gambling systems for favorable games, 1969) Thorp goes into even greater detail in his later work, "The Kelly Criterion in Blackjack, Sports Betting, and the Stock Market" where he graphically illustrates how the log for wealth is maximized to maximize the growth of wealth over time. (Thorp E. O., 1998) He specifically applies the criterion to the stock market in "The Kelly Criterion and the Stock Market." (Rotando, 1992) Studying Thorp's works, it appears that although having a formula to calculate the optimum bet is useful, clearly understanding the game is far more important.

Nekrasov created a formula for implementing the Kelly criterion in multivariate portfolios as seen in Eq. 2. Consider a market with n correlated stocks S_k with stochastic return r_k and a riskless bond with return r. An investor puts a fraction u_k of his capital in S_k and the rest is invested in bonds. The following formula may be used to compute the optimum investments where $\hat{\vec{r}}$ and $\hat{\Sigma}$ are the vector of the means and the matrix of 2nd mixed noncentral moments of the excess returns. (Nekrasov, 2014)

$$\overrightarrow{u^*} = (1+r)\left(\hat{\Sigma}\right)^{-1}\left(\hat{\vec{r}} - r\right) \qquad (2)$$

The interest of Thorp and others in the Kelly criterion indicate its usefulness is selected how much of the available resources to invest. Nekrasov's work extends this across multiple options in a collection that might resemble sessions in network traffic. Although the differences between our specific requirements are different enough from the requirement of those cited that it will be necessary to start from first principles to create our Kelly criterion inspired formula, their work is close enough to demonstrate the feasibility of our approach.

Summary

Given the wealth of data mining and anomaly detection techniques published in the literature, selecting the best or combine two or three promising ideas should be relatively straight forward. This is especially true since our tolerance for false positives is significantly higher than the applications for which they were designed. Given the long history of applying the Kelly criterion in finance, its effectiveness has been well established. Even though it will be necessary to develop the Kelly criterion inspired formula from first principles, the success of the Kelly criterion supports the feasibility of the approach.

3. APPROACH

This research effort breaks down into 2 research questions and 2 phases. The first question, which will be addressed in phase 1, is the following: is it possible to combine expert knowledge with existing data mining and anomaly detection techniques to compute a useful maliciousness score for network traffic? The second question, which will be addressed in phase 2, is the following: is it possible to develop a Kelly criterion inspired algorithm that will take the maliciousness score from research question 1 and determine how much traffic from each session should be sent to the CAS in order to minimize the amount of traffic transmitted while maximizing the amount of malicious traffic which is available for analysis?

Phase 1

In phase 1, expert knowledge, data mining, and best of breed intrusion detection will be combined in order to compute a maliciousness rating. The first step of this phase will be to discover the relevant facts that may be gleaned from expert knowledge (e.g. when the Heart Bleed vulnerability was discovered, an expert could have caused the system to rate secure socket layer traffic higher; and when a known malicious internet protocol address or domain is discovered, an expert could cause the system to rate traffic including that IP or domain higher.) The second step of this phase will be to discover the relevant facts that may be mined from the Interrogator data store (e.g. Long and Morgan mined

Interrogator to develop a white list of web servers to be excluded and instances of new servers to be included. (Long & Morgan, 2005) This could be expanded to rate traffic more malicious which contains addresses and ports associated with alerts or incidents.) The third step of this phase will combine best of breed anomaly detection algorithms to form a maliciousness rating (e.g. MINDS collected, computed, and assigned a local outlier factor to 16 different features (Chandola, Eilertson, Ertoz, Simon, & Kumar, 2007) (Ertoz, et al., Detection and summarization of novel network attacks using data mining, 2003) (Ertoz, et al., Minds-minnesota intrusion detection system, 2004) KMC+NBC uses K-Means clustering and Naïve Bayes Classification to detect anomalies in network traffic. (Yassin, Udzir, Muda, & Sulaiman, 2013) Again a measure of abnormality could factor into the session rating. The fourth step of this phase will be to develop a formula to combine all of these into a single score. Phase 1 corresponds to the top half of Fig. 2 where unrated sessions are captured by the sensor and flow into the session rater which uses expert knowledge, mined data, and anomaly algorithms to rate each session. The green sessions are known benign, the red sessions are known malicious, and the other colors are meant to represent the continuum in between.

Phase 2

In phase 2, a Kelly criterion based formula that takes the scores generated from phase 1 as input and produces as output a fraction of the available network traffic that should be invested in each session will be developed. Kelly proved that there exists an amount to bet l being some portion of the total wealth G, that if the gambler bets it consistently, G will obtain and maintain a level greater than any other possible value for l. (Kelly, 1956) This may be seen in Eq. 1 where *l* is the fraction of wealth to bet, *p* is the probability of winning, and *b* is the net odds of the wager. Thorp applied the Kelly criterion to the game of blackjack. (Thorp E. O., Beat the dealer, 1966) Smoczynski and Tomkins applied the Kelly criterion to horse racing. (Smoczynski & Tomkins, 2010) Separately Thorp and Nekrasov applied the Kelly Criterion to the stock market. (Thorp & Kassouf, Beat the Market: A Scientific Stock Market System, 1967) (Nekrasov, 2014) Using this generalization, one would consider network flows to be stocks and rate of return to be the maliciousness score of the session. Phase 2 corresponds to the bottom half of Fig. 2 where the rated sessions flow into the algorithm and the session selector feeds those ratings into the Kelly criterion (Kelly, 1956) inspired formula to determine how much traffic to invest in each session. The fatter sessions represent more traffic being invested in the session and the skinnier sessions represent less traffic being invested in the session.

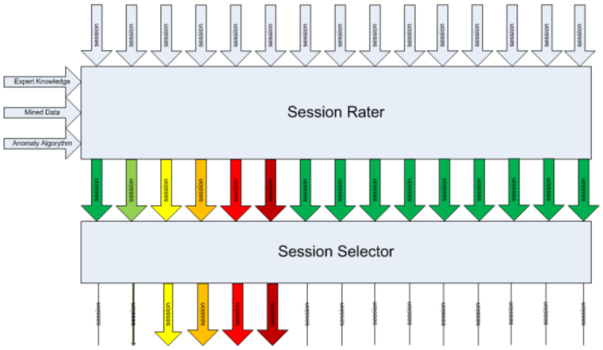

Figure 2 Kelly Compressor Diagram

Nekrasov's formula will be used in Eq. 2 to illustrate how this might work. This will be applied to our problem by substituting the returns for the maliciousness score and the investment for the amount of available traffic to assign to each session. Since a riskless bond makes no sense in our problem, it will be set the value to zero simplifying the equation shown in Eq. 3. This leaves us with only one variable because the 2nd noncentral moment is a function of the maliciousness rating over time. Remember it is unlikely that Nekrasov's formula will work as given. This is because of the fixed nature of our investments. Correctly selecting malicious sessions does not increase the bandwidth available, and incorrectly selecting benign session does not decrease the bandwidth available. Further, there is no chance of ruin. It will be necessary to start from the same starting point that Kelly did to retrace his steps to construct a formula for this specific application.

$$\overrightarrow{u^*} = (\hat{\Sigma})^{-1}(\hat{\vec{r}}) \qquad (3)$$

Once the session rater and session selector algorithms are developed, they will be incorporated into a prototype which will be tested against open sources datasets to include those used by Smith et al. in their theoretical exploration. (Smith, Hammell, Parker, & Marvel, 2016)

4. RESULTS

Many of the data mining and anomaly detection techniques have settings that will increase the sensitivity creating more false positives and fewer false negatives or decrease the sensitivity creating fewer false positives and more false negatives. As the research is completed, these settings will be tuned until the appropriate amount of compression and an acceptable level of false negatives are obtained. This will be illustrate this by applying entropy to remove compressed and encrypted data.

Interviews with experts in network intrusion detection revealed that there is very little value in transmitting encrypted or compressed data back to the CAS. Encrypted data is not very valuable because decrypting it is prohibitively expensive and beyond the capabilities of most network defense analysts. Compressed data is of little value because it is very difficult to decompress the file unless every packet of the session containing the compressed file is available. Network file carving is more efficiently done on the sensor and a cryptographic hash is sufficient for most network intrusion detection applications. The entropy of data may be used to detect if data are encrypted or compressed because this data has a much higher entropy than clear text data (Shannon, 2001).

The following experiment illustrates the kinds of results that may be expected. An experiment was conducted where packets with entropy values greater than a given threshold were dropped and the abridged data was processed by Snort. (Roesch, 1999) This experiment was repeated lowering the entropy values from 7.9 to 4.0 in increments of 0.1. Fig. 3 plots the size of the datasets for each iteration and the alert loss rate for each iteration. Notice that at an entropy value of 7.0 the data has been compressed to 27% of its original size, but has only lost 0.6% of the alerts.

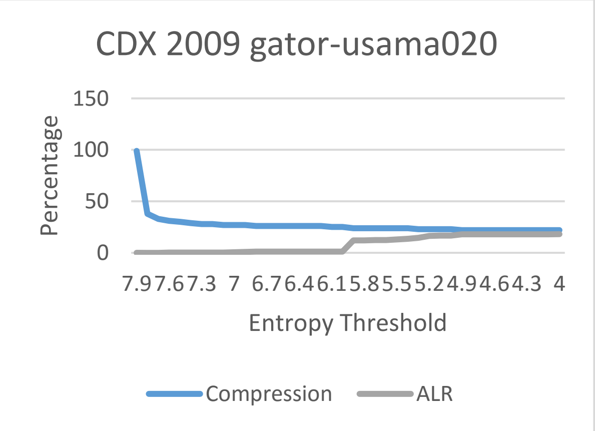

Figure 3 Lossy Compression Using Entropy

An interesting property of low entropy data is that it compresses very well. Applying GNU Zip lossless compression to the dataset that has been compressed using the entropy based lossy compression results in a file that is 4% of the original size of the dataset which is well within our bandwidth budget of 10%. These results are anecdotal and certainly may not be typical, but they do illustrate the feasibility of the approach.

5. CONCLUSIONS

In a distributed NIDS environment, it is necessary to transmit the right data back to the central analysis servers to provide analysts with the information necessary to detect and report malicious activity. Bringing back all of the data would double the bandwidth requirements of the site and require that the analysis servers have massive bandwidth available to receive it all. Standard lossless compression is not sufficient to reduce this traffic to an acceptable level. The goal of this research is to develop a lossy compression algorithm that will ensure that the traffic lost is the least likely to contain malicious activity. The approach is to use an algorithm based upon the Kelly criterion to allocate the limited bandwidth available, coupled with best of breed anomaly detection, to assess the maliciousness of the traffic. These two technologies will be combined into a packet capture tool which will produce data compliant with the standards used by existing NIDS tools. Preliminary results show a compression ratio of 96%. Although these results were obtain from a dataset that is unlikely to reflect real word traffic, they demonstrate the feasibility of the approach.

6. REFERENCES

Breiman, L. (2012). Optimal gambling systems for favorable games. In *The Kelly Captial Growth Investment Criterion: Theory and Practice* (pp. 47-60). New Jersey NJ: World Scientific.

Bridges, S. M., & Vaughn, R. B. (2000). Fuzzy data mining and genetic algorithms applied to intrusion detection. *Proceedings of the 12th Annual Canadian Information Technology Security Symposium.* Ottawa, Canada.

Chandola, V., Eilertson, E., Ertoz, L., Simon, G., & Kumar, V. (2007). MINDS: Architecture & Design. In *Data Warehousing and Data Mining Techniques for Cyber Security* (pp. 83-108). New York, NY: Springer.

Claise, B. (2004). *Cisco Systems NetFlow Services Export Version 9.* Fremont, California, United States: Internet Engineering Task Force (IETF).

Cohen, W. W. (1995). Fast Effective Rule Induction. *Proceedings of the Twelfth International Conference on Machine Learning* (pp. 115-123). Lake Taho, CA: Morgan Kaufman.

Dokas, P., Ertoz, L., Kumar, V., Lazarevic, A., Srivastava, J., & Tan, P.-N. (2002). Data mining for network intrusion detection. *Proc. NSF Workshop on Next Generation Data Mining.* Baltimore, Maryland, United States.

Ertoz, L., Eilertson, E., Lazarevic, A., Tan, P., Dokas, P., Srivastava, J., & Kumar, V. (2003). *Detection and summarization of novel network attacks using data mining.* Minneapolis, Minnesota: Army High Performance Computing Research Center.

Ertoz, L., Eilertson, E., Lazarevic, A., Tan, P.-N., Kumar, V., Srivastava, J., & Dokas, P. (2004). Minds-minnesota intrusion detection system. In *Next Generation Data Mining* (pp. 199-218). Cambridge, MA: MIT Press.

Eskin, E., Arnold, A., Prerau, M., Portnoy, L., & Stolfo, S. (2002). A geometric framework for unsupervised anomaly detection. In *Applications of data mining in computer security* (pp. 77-101). New York, NY: Springer.

Garcia-Teodoro, P., Diaz-Verdejo, J., Macia-Fernandez, G., & Vazquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. *Computer & Security, 28*(1), 18-28.

Grinstein, G., Laskowski, S., Wills, G., & Rogowitz, B. (1997). Information Exploration Shootout Project and Benchmark Data Sets (Panel): Evaluating How Visualization Does in Analyzing Real-world Data Analysis Problems. *Proceedings of the 8th Conference on Visualization '97.* Los Alamitos, CA, USA.

Heberlein, L. T., Dias, G. V., Levitt, K. N., Mukherjee, B., Wood, J., & Wolber, D. (1990). A network security monitor. *Research in Security and Privacy, 1990. Proceedings., 1990 IEEE Computer Society Symposium on.* Oakland, CA.

Heberlein, L. T., Mukherjee, B., Levitt, K., Dias, G., & Mansur, D. (1991). *Towards detecting intrusions in a networked environment.* Davis, CA: U. of Calif., Davis.

Heberlein, L., Levitt, K., & Mukherjee, B. (1991). A method to detect intrusive activity in a networked environment. *Proceedings of the 14th National Computer Security Conference.* Washington, DC, USA.

Ierace, N., Urrutia, C., & Bassett, R. (2005). Intrusion Prevention Systems. *Ubiquity*, 1530-2180.

Jacobson, V., Leres, C., & McCanne, S. (1989). The tcpdump manual page. *Berkley (CA): Lawrence Berkley Laboratory*.

Kelly, J. L. (1956). A New Interpretation of Information Rate. *IRE Transactions on Information Theory, 2*(3), 185-189.

Kemmerer, R. A., & Vigna, G. (2002). Intrusion detection: A brief history and overview (supplement to computer magazine). *Computer*, 27-30.

Krugel, C., Toth, T., & Kirda, E. (2002). Service specific anomaly detection for network intrusion detection. *Proceedings of the 2002 ACM symposium on Applied computing.* Madrid, Spain.

Lee, W., & Stolfo, S. J. (1998). Data Mining Approaches for Intrusion Detection. *Proceedings of the 7th USENIX Security Symposium.* San Antonio, TX.

Lippmann, R., Fried, D., Graf, I., Haines, J., Kendall, K., McClung, D., . . . Zissman, M. (2000). Evaluating intrusion detection systems: the 1998 DARPA off-line intrusion detection evaluation. *DARPA Information Survivability Conference and Exposition* (pp. 12-26). DISCEX '00.

Long, K. S. (2004). *Cathing the Cyber Spy: ARL's Interrogator.* Army Research Laboratory, Computational Information Systems Directory. Aberdeen Proving Ground: DTIC.

Long, K. S., & Morgan, J. B. (2005). *Using Data Mining to Improve the Efficiency of Intrusion Detection Analsysis.* Aberdeen Proving Ground (MD): Army Research Laboratory (US).

Luo, J. (1999). *Integrating fuzzy logic with data mining methods for intrusion detection.* Starkville, Mississippi, United States: Mississippi State University.

Mukherjee, B., Heberlein, L. T., & Levitt, K. N. (1994). Network Intrusion Detection. *IEEE Network, 8*(3), 26-41.

Munz, G., Li, S., & Carle, G. (2007). Traffic anomaly detection using k-means clustering. *GI/ITG Workshop MMBnet.* Hamburg, Germany.

Nekrasov, V. (2014). *Kelly Criterion for Multivariate Portfolios: A Model-Free Approach.* Rochester, NY: Social Science Research Network.

Paxson, V. (1999). Bro: a System for Detecting Network Intruders in Real-time. *Computer Networks, 31*(23), 2435-2463.

Roesch, M. (1999). Snort: Lightweight Intrustion Detection for Networks. *Proceedings of the 13th System Administration Conference LISA '99. 99*, pp. 7-12. Seattle WA, US: USENIX.

Rotando, L. M. (1992). The Kelly criterion and the stock market. *American Mathematical Monthly, 99*(10), 922-931.

Sangster, B., O'Connor, T., Cook, T., Fanelli, R., Dean, E., Adams, W. J., . . . Conti, G. (2009). Toward instrumenting network warfare competitions to generate labeled datasets. *Proc. of the 2nd Workshop on Cyber Security Experimentation and Test (CSET09).* Montreal, Canada.

Sekar, R., & Uppuluri, P. (1999). Synthesizing Fast Intrusion Prevention/Detection Systems from High-Level Specifications. *Proceedings of the 8th USENIX Security Symposium.* Washington, DC.

Sekar, R., Gupta, A., Frullo, J., Shanbhag, T., Tiwari, A., Yang, H., & Zhou, S. (2002). Specification-based anomaly detection: a new approach for detecting network intrusions. *CCS '02: Proceedings of the 9th ACM Conference on Computer and Communications Security.* Washington, DC, USA.

Shannon, C. E. (2001). A Mathematical Theory of Communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 3-55.

Shiravi, A., Shiravi, H., Tavallaee, M., & Ghorbani, A. A. (2012). Toward developing a systematic approach to generate benchmark datasets for intrusion detection. *Computer & Security, 31*(3), 357-374.

Smith, S. C., & Hammell, R. J. (2015). *An Experimental Exploration of the Impact of Sensor-Level Packet Loss on Network Intrusion Detection.* Army Research Laboroatory (US). Aberdeen Proving Ground: Army Research Laboroatory (US).

Smith, S. C., Hamell, R. J., Wong, K. W., & Mateo, C. J. (2016). An Experimental Exploration of the Impact of Multi-level Packet on Network Intrusion Detection. *Proceedings of the 14th IEEE/ACIS International Conference on Software Engineering Research, Management and Applications (SERA 2016)* (pp. 23-30). Towson, MD, U.S.A: IEEE.

Smith, S. C., Hammell, R. J., Parker, T. W., & Marvel, L. M. (2016). A Theoretical Exploration of the Impact of Packet Loss on Network Intrusion Detection. *International Journal of Networked and Distributed Computing*, 1-10.

Smith, S. C., Wong, K. W., Hammell, R. J., & Mateo, C. J. (2015). *An Experimental Exploration of the Impact of Network-Level Packet Loss on Network Intrusion Detection.* Aberdeen Proving Ground: Army Research Laboratory (US).

Smoczynski, P., & Tomkins, D. (2010). AN EXPLICIT SOLUTION TO THE PROBLEM OF OPTIMIZING THE ALLOCATIONS OF A BETTOR'S WEALTH WHEN WAGERING ON HORSE RACES. *Mathematical Scientist, 35*(1).

Thorp, E. O. (1960). Fortune's formula: The game of blackjack. *Notices of the American Mathematical Society, 7*(7), 935-936.

Thorp, E. O. (1966). *Beat the dealer.* New York, NY: Random House.

Thorp, E. O. (1969). Optimal gambling systems for favorable games. *Revue de l'Institut International de Statistique, 37*(3), 273-293.

Thorp, E. O. (1998). The Kelly Criterion in Blackjack, Sports Betting, and the Stock Market. *Finding the Edge: Mathematical Analysis of Casino Games, 1*(6).

Thorp, E. O. (2012). Understanding the Kelly Criterion. In *The Kelly Capital Growth Investment Criterion: Theory and Practice* (pp. 511-525). New Jersey, NJ: World Scientific.

Thorp, E. O., & Kassouf, S. T. (1967). *Beat the Market: A Scientific Stock Market System.* New York, NY: Random House.

Uppuluri, P., & Sekar, R. (2001). Experiences with specification-based intrusion detection. *Recent Advances in Intrusion Detection.* Davis, CA.

Yassin, W., Udzir, N. I., Muda, Z., & Sulaiman, M. N. (2013). Anomaly-Based Intrusion Detection through K-Means Clustering and Naives Bayes Classification. *Proceedings of the 4th International Conference on Computing and Informatics (ICOCI).* Sarawk, Malaysia.

Ziv, J., & Lempel, A. (1997). A Universal Algorithm for Sequential Data compression. *IEEE Transactions on Information Theory*, 337-343.