Quantifying the Effect of Incomplete Information in Denial of Service Detection

Mohamed Wasim Lorgat
Electrical Engineering Department
University of Cape Town
Cape Town, South Africa
lrgmoh002@myuct.ac.za

Alireza Baghai-Wadji
Electrical Engineering Department
University of Cape Town
Cape Town, South Africa
alireza.baghai-wadji@uct.ac.za

Andre McDonald
Council for Scientific and
Industrial Research
Pretoria, South Africa
amcdonald@csir.co.za

***Abstract*—The performance loss due to incomplete information in denial of service (DoS) detection is quantified in this paper. Volumetric DoS detection is formulated as a signal detection problem. Two detectors are defined: the first operates without knowledge of the attack model and the second operates as if the attack model were known. The performance loss is quantified by comparing the two detectors. Simulation results demonstrate that the performance loss is greatest for low intensity attacks and slowly diminishes as the attack intensity increases.**

***Index Terms*—Intrusion Detection, Anomaly Detection, Signal Detection, Denial of Service (DoS)**

I. Introduction

Anomaly-based intrusion detection operates by constructing a model of normal network traffic patterns and detecting deviations from the model as potential intrusions [1]. The anomaly-based approach operates without an attack model, thus allowing the detection of previously unknown attacks. However, this feature requires that assumptions be made about the attack. Depending on the problem conditions, the assumptions may lead to a performance decrease [2], [3].

In this paper, volumetric denial of service (DoS) detection is formulated as a signal detection problem; i.e., the detection of a signal (the attack traffic) embedded in noise (the background traffic). This formulation is similar to previous work utilizing sequential change-point methods [4], but differs in that it employs explicit models of network traffic rather than detecting distributional changes in network traffic. Signal detection theory provides the necessary tools to quantify the effects of incomplete information on detection performance. Two detectors are compared in this paper. The first detector operates without knowledge of the attack model and the second detector operates as if the attack model were known. By comparing the detectors, the performance loss due to the lack of an attack model is quantified. Simulations are performed on real network traffic from the Measurement and Analysis on the WIDE Internet (MAWI) archive [5] superposed with synthetically generated attacks.

The paper is organized as follows. The volumetric DoS detection problem is formulated in Section II, and the two detectors are defined in Section III. Section IV presents the simulation results and numerical analysis. Finally, Section V concludes and proposes future directions.

II. Problem Formulation

Formulated within the framework of signal detection theory, the problem is to detect whether a signal (the DoS attack traffic) is present in noise (the background traffic).

Modelling Network Traffic: The time series of interest is the number of packets arriving at the network within the n^{th} non-overlapping observation interval of duration Δ, denoted by $w[n]$ for $n = 1, \cdots, N_{\text{total}}$. It has been empirically demonstrated that the marginal distribution of $w[n]$ can be reasonably approximated by a gamma distribution over a wide range of observation interval lengths $10^{-3}\,\text{s} \leq \Delta \leq 10\,\text{s}$ [6]. The gamma marginal distribution model has been implemented to effectively detect network attacks in [6]–[8], and is assumed here. Formally, it is assumed that each $w[n]$ is a sample value of a random variable with gamma probability density function (PDF) given by

$$\gamma(w[n];\alpha[n],\beta[n]) = \frac{1}{\beta[n]\Gamma(\alpha[n])}\left(\frac{w[n]}{\beta[n]}\right)^{\alpha[n]-1} \times \exp\left(-\frac{w[n]}{\beta[n]}\right), \tag{1}$$

for $w[n], \alpha[n], \beta[n] > 0$. Here, $\Gamma(\cdot)$ denotes the gamma function, and $\alpha[n]$ and $\beta[n]$ denote the shape and scale parameters of the gamma distribution.

Furthermore, $w[n]$ is assumed to be locally stationary (in accordance with [9]); i.e., the parameters $\alpha[n]$ and $\beta[n]$ are assumed to vary slowly with respect to n. A standard approach for dealing with locally stationary time series is to construct a sliding window, such that $w[n]$ observed within the window are stationary ($\alpha[n]$ and $\beta[n]$ are approximately constant for n within the window). A window of length $N_{\text{ref}} + N_{\text{test}}$ sample points is constructed, and shifted across the time series by S sample points in each iteration. For each position of the window $m = 0, \cdots, M-1$, the first N_{ref} sample points comprise the reference subsequence ($n = mS + 1, \cdots, mS + N_{\text{ref}}$) and the next N_{test} sample points comprise the test subsequence ($n = mS+N_{\text{ref}}+1, \cdots, mS+N_{\text{ref}}+N_{\text{test}}$). Fig. 1 on the next page provides a schematic description of the windowing procedure, demonstrated for an example time series of 10 sample points. Selection of appropriate values for N_{ref}, N_{test}, and S will be guided in Section IV.

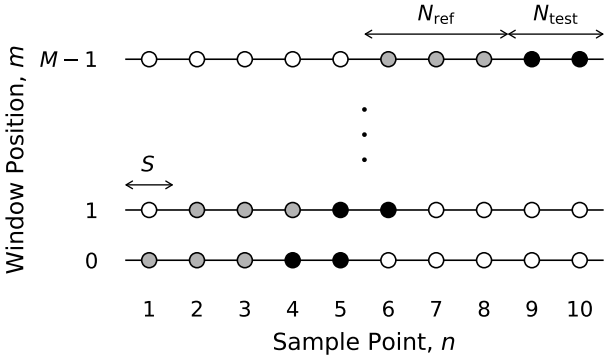


Fig. 1. A schematic description of the problem setup for a time series of length $N_{\text{total}}(= 10)$ sample points. A sliding window is started at the initial position $m = 0$ (not shifted). The first $N_{\text{ref}}(= 3)$ sample points of the window comprise the reference subsequence, represented by the grey circles; and the next $N_{\text{test}}(= 2)$ sample points comprise the test subsequence, represented by the black circles. White circles represent sample points that are not processed for a given window position. The window is shifted by $S(= 1)$ sample points to give the next position $m = 1$ (shifted once). The process is repeated for a total of M window positions.

Detection Problem: The subsequences corresponding to each consecutive window position are assumed to be statistically independent, therefore the problem is statistically invariant to the choice of the window position m. Without loss of generality, the case $m = 0$ will be considered here for notational simplicity. The detection problem is to distinguish whether an attack of amplitude $A > 0$ appears in background network traffic $w[n]$. Denoting, the observed time series by $x[n]$, the problem is to distinguish between the two hypotheses:

$$\mathcal{H}_0 : x[n] = w[n], \tag{2}$$

$$\mathcal{H}_1 : x[n] = A + w[n], \tag{3}$$

for $n = N_{\text{ref}} + 1, \cdots, N_{\text{ref}} + N_{\text{test}}$. Due to the assumed local stationarity, and provided N_{test} is chosen appropriately, $w[n]$ are identically gamma-distributed with parameters α and β; i.e., the parameters are no longer n-dependent within the window. The parameters α and β are unknown. However, it is assumed that the reference subsequence $(x[1], \cdots, x[N_{\text{ref}}])$, consists only of noise, from which α and β can be estimated provided N_{ref} is sufficiently large. The attack amplitude A plays an important role in this paper and will be discussed in the next section, after defining the two detectors.

III. Detection Methods

Signal detection problems are typically solved by optimizing a specified criterion. Two fundamental criteria include the Bayes risk and the Neyman-Pearson (NP) criterion. The Bayes risk assigns a prior probability to each hypothesis and a cost to each possible decision outcome. When it is difficult to assume prior probabilities or error costs, the NP criterion is appropriate, which attempts to maximize the probability of detection, P_{D}, for a given probability of false alarm, P_{FA}. It is well known that optimizing either criterion leads to detectors of the same general form; i.e., the likelihood ratio test (LRT) [10].

Denoting an observed test subsequence by the vector $\mathbf{x} \coloneqq (x[N_{\text{ref}} + 1], x[N_{\text{ref}} + 2], \cdots, x[N_{\text{ref}} + N_{\text{test}}])$, the LRT decides that $\mathcal{H}_1$ was true if

$$\Lambda(\mathbf{x}) \coloneqq \frac{p_{\mathbf{X}|\mathcal{H}_1}(\mathbf{x}|\mathcal{H}_1)}{p_{\mathbf{X}|\mathcal{H}_0}(\mathbf{x}|\mathcal{H}_0)} > \eta, \tag{4}$$

and decides that $\mathcal{H}_0$ was true otherwise. In (4), $\Lambda(\mathbf{x})$ denotes the likelihood ratio, η denotes the threshold value, and $p_{\mathbf{X}|\mathcal{H}_0}(\mathbf{x}|\mathcal{H}_0)$ and $p_{\mathbf{X}|\mathcal{H}_1}(\mathbf{x}|\mathcal{H}_1)$ denote the conditional PDFs of $\mathbf{x}$ under each hypothesis. In the Bayesian approach, η is defined in terms of the assumptions about error costs and prior probabilities, whereas in the NP approach, η is defined such that a given P_{FA} is achieved. Here, the choice of criterion will be left arbitrary by referring in general to η.

In practice, it is rare that the attack amplitude A will be known before the occurrence of the attack. This challenge is a primary reason for the growing interest in anomaly-based detection, which operates without an attack model. In this paper, a comparison is drawn between anomaly-based detection and signal detection without a signal-plus-noise model; i.e., without knowledge of $p_{\mathbf{X}|\mathcal{H}_1}(\mathbf{x}|\mathcal{H}_1)$. Therefore, anomaly-based detection is modelled here as a detector that decides $\mathcal{H}_1$ was true if

$$\frac{1}{p_{\mathbf{X}|\mathcal{H}_0}(\mathbf{x}|\mathcal{H}_0)} > \xi, \tag{5}$$

and decides that $\mathcal{H}_0$ was true otherwise, where ξ is a threshold value. In contrast with (4), there is no optimality associated with (5). However, Bishop [11] has shown that under certain assumptions about $p_{\mathbf{X}|\mathcal{H}_1}(\mathbf{x}|\mathcal{H}_1)$, (4) and (5) are equivalent detectors. The first assumption being that $p_{\mathbf{X}|\mathcal{H}_1}(\mathbf{x}|\mathcal{H}_1)$ is equal to a constant on the support of $p_{\mathbf{X}|\mathcal{H}_0}(\mathbf{x}|\mathcal{H}_0)$. And, the second assumption being that $p_{\mathbf{X}|\mathcal{H}_1}(\mathbf{x}|\mathcal{H}_1)$ is of the form $F(p_{\mathbf{X}|\mathcal{H}_0}(\mathbf{x}|\mathcal{H}_0))$, where $F(\cdot)$ is a strictly decreasing function. The details of the assumptions are contained within ξ.

These assumptions are not often appropriate in real problems. Thus follows the main contribution of this paper: to quantify the performance loss of (5), hereafter referred to as the anomaly-based (AB) detector, under more realistic conditions. This is accomplished by comparing the performance of the suboptimal AB detector with the optimal LRT detector for a simple class of volumetric DoS attacks, via the simulations presented in the next section.

IV. Simulations

The simulations presented here were performed with network traffic data from the MAWI archive. Details of the MAWI dataset and the simulation parameter settings will be provided next. Thereafter, the simulation results will be presented and discussed.

A. Simulation Setup

The MAWI Dataset: The publicly available MAWI working group archive samplepoint-F [5] connects several Japanese research institutes and universities to the Internet. The archive offers 15 minutes of traffic from 14:00 to 14:15 (Japanese

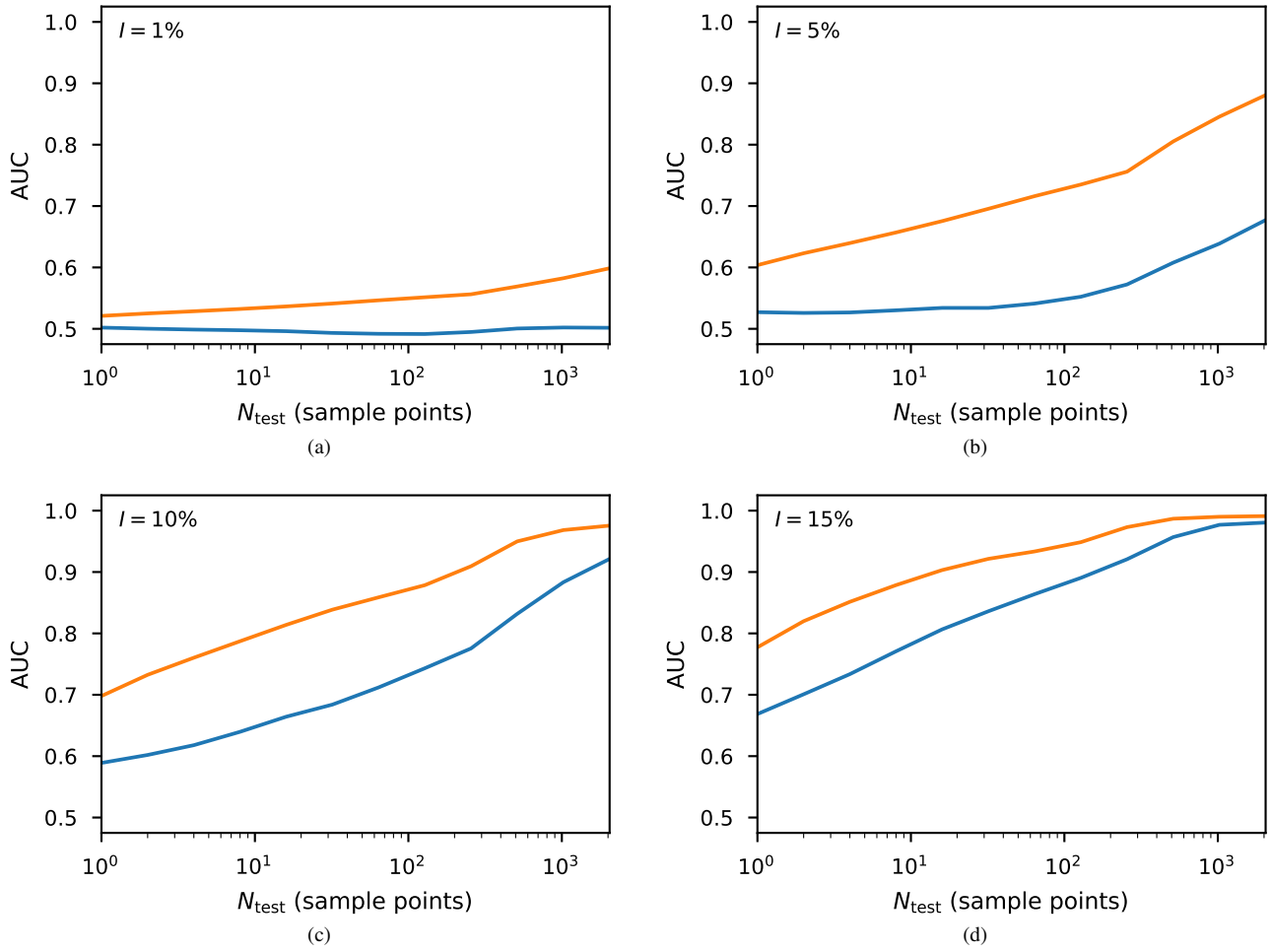


Fig. 2. Plots of the accuracy of each detector, in terms of AUC, versus logarithmically-scaled test subsequence length, N_{test}, for a fixed attack intensity of (a) $I = 1\%$, (b) $I = 5\%$, (c) $I = 10\%$, and (d) $I = 15\%$. Orange curves represent the LRT detector and blue curves represent the AB detector.

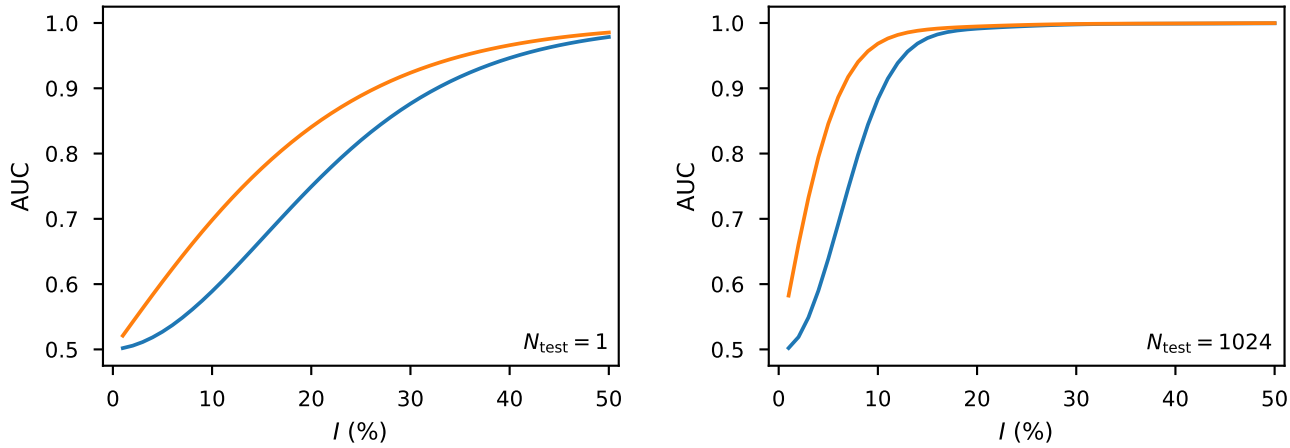
Standard Time) every day of the year since the year 2000. Packet payloads are removed and IP addresses are anonymized. The present experiments were performed on the traffic trace from Tuesday 25th July 2017, the latest available at the time. The traffic trace contains roughly 121 million packets and has an average traffic rate of 974 Mbit/s.

Synthetic attacks were used, similar to [6] and [7]. In particular, the authors of [7] simulated flooding attacks using iPerf[1] on a quiet network, and superposed the resulting pure attack traffic on real recorded background traffic. Those authors acknowledge that this approximation does not take into account the changing behaviour of the underlying network protocols under the effect of the attack. However, they demonstrated that the marginal distribution of the superposed time series reasonably approximates that of a real attack time series [7, Fig. 4]. In this paper, attacks have been synthesized as a constant-valued time series, rather than simulated via tools such as iPerf. While this approximation allows to demonstrate the main contributions of the paper, performing the analysis with real attack tools on a real or simulated network is preferable. Work is currently underway in that direction.

Parameter Settings: The fixed parameters were set as follows: $\Delta = 1$ ms, $N_{\text{ref}} = 1$ min $= 60\,000$ sample points, and $N_{\text{shift}} = 100$ ms $= 100$ sample points. Given the short duration (15 min) of the traffic trace, the sample period Δ was chosen to give a large total sample size of $N_{\text{total}} \approx 900\,000$ sample points. The reference subsequence length N_{ref} was selected to attain a balance between avoiding non-stationarity within each window position and providing a sufficient number of samples so that model parameter estimation accuracy did not affect the results. The shift length S ensured $M \approx 9000$ window positions per experiment, therefore roughly 9000 detection opportunities from which to reliably compute P_{FA} and P_{D}.

The test subsequence length was dyadically increased, $N_{\text{test}} = 2^k$ for $k = 0, \cdots, 11$, until further increases resulted in erratic detection accuracy (which the authors suspect is due to introducing non-stationarities). The attack intensity was uniformly increased, $I = 0.01l$ for $l = 1, \cdots, 50$, such that any smaller lead to barely detectable attacks, and any larger resulted in too easily detectable attacks.

[1] iPerf is a network bandwidth measurement tool. Available: https://iperf.fr/



(a) (b)

Fig. 3. Plots of the accuracy of each detector, in terms of AUC, versus attack intensity, I (%), for a fixed test subsequence length of (a) $N_{\text{test}} = 1$ sample point and (b) $N_{\text{test}} = 1024$ sample points. Orange curves represent the LRT detector and blue curves represent the AB detector.

B. Simulation Results

The purpose of the simulations is to evaluate the performance, in terms of detection accuracy, of two detectors for a constant-rate volumetric DoS attack (as detailed in Sections II and III). The AB detector operates without knowledge of the attack-plus-noise PDF, and the LRT detector operates as if the attack-plus-noise PDF were known. By comparing the two detectors, the performance loss due to the incomplete information utilized by the AB detector is assessed. The numerical analysis to follow serves to investigate how the performance loss varies with respect to the test subsequence length, N_{test}, and the attack intensity, I. A complete measure of performance, in terms of detection accuracy, is given by the receiver operating characteristic (ROC) curve; i.e., P_{FA} versus P_{D} with varying detector threshold. The ROC curve allows to disentangle the assessment of a detector's accuracy from problem-specific assumptions, such as prior probabilities and error costs. Performance is measured here by the area under the ROC curve (AUC), ranging from 0.5, meaning that the detector is no better than flipping a coin, to 1, meaning that the detector has perfect accuracy.

AUC Versus Test Subsequence Length: Each subfigure in Fig. 2 fixes I and plots the AUC as a function of N_{test} on a semi-logarithmic axis. The blue curve corresponds to the AB detector, and the orange curve corresponds to the LRT detector. For small intensities $I < 5\%$ (Fig. 2a), the AB detector performance is roughly constant, whereas the LRT detector performance is increasing and piecewise linear with a breakpoint at $N_{\text{test}} \in (200, 300)$. Note that the piecewise linearity is observed on a semi-logarithmic axis. The figure demonstrates that the AB detector performance remains poor, with negligible improvement despite a substantial increase in the number of samples available for detection.

At $I = 5\%$ (Fig. 2b), the AB detector performance begins to increase with N_{test}, however, the performance difference still increases with N_{test}. The AB detector performance also shows approximate piecewise linearity (on the semi-logarithmic axis) with the breakpoint at $N_{\text{test}} \in (200, 300)$. When the intensity reaches $I = 10\%$ (Fig. 2c), LRT performance begins to saturate for $N_{\text{test}} \in (500, 600)$; i.e., further increasing N_{test} yields minimal increase in the AUC. This observation is to be expected as detection accuracy draws nearer to perfect. The saturation has the effect that at this intensity, the performance difference decreases for large N_{test}.

Whereas for $I < 15\%$, both curves have positive curvature, for $I \geq 15\%$ (Fig. 2d), both curves have negative curvature. The authors are not yet able to explain structural change. At this intensity, AB performance begins to saturate for $N_{\text{test}} \in (500, 600)$. For greater intensities than have been plotted here, no further structural changes were observed; the curves draw increasingly nearer to each other as they asymptotically approach perfect accuracy (AUC ~ 1).

AUC Versus Attack Intensity: Each subfigure in Fig. 3 fixes N_{test} and plots the AUC as a function of I. As before, the blue curve corresponds to the AB detector, and the orange curve corresponds to the LRT detector. While the LRT curve has positive curvature throughout, the AB curve has negative curvature for small intensities (roughly $I < 5\%$) and positive curvature thereafter. The negative curvature demonstrates a structural defficiency in the AB detector for small intensity attacks. For each fixed N_{test}, both detectors' performances saturate with increasing I and the performance difference diminishes. As N_{test} increases, the curves appear to be increasingly compressed along the I-axis, giving the effect that the performance difference diminishes at lower intensities.

Performance Loss Measure: The performance difference is characterized in another way in Fig. 4. In order to describe the function $\bar{R}(I)$ plotted in the figure, the reader is first referred to Fig. 2c. For $I = 10\%$, the LRT detector achieves an AUC of approximately 0.7 at $N_{\text{test}} = 1$. Since the curve is monotonically increasing, it can be said that $N_{\text{test}} = 1$ is the minimum required value for the LRT detector to achieve an

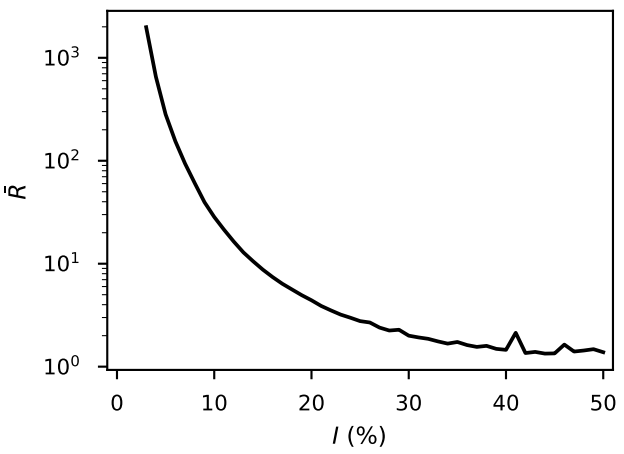


Fig. 4. Plot of the performance comparison measure, $\bar{R}$, versus attack intensity, I (%). See the text for a description of the measure $\bar{R}$.

AUC of 0.7. Similarly, $N_{\text{test}} \approx 400$ is the minimum required value for the AB detector to achieve the same AUC of 0.7. Denote by $R(\text{AUC}, I)$ the ratio of minimum required N_{test} values to achieve a desired AUC. Continuing with the example, it follows that $R(0.7, 10\%) = 400$; i.e., for attacks of intensity $I = 10\%$, the AB detector requires a test subsequence length 400 times that of the LRT detector to achieve the same AUC of 0.7. Finally, denote by $\bar{R}(I)$ the average of $R(\text{AUC}, I)$ over all AUC values, which has been plotted in the figure. The curve in Fig. 4 demonstrates two key points. First, for small intensities, the AB detector requires a significantly larger N_{test} than the LRT detector. Second, as the intensity increases, $\bar{R}(I)$ decays slowly; even at $I = 50\%$, the AB detector requires on average 1.5 times the N_{test} of the LRT detector.

V. Conclusion

The performance loss due to the lack of an attack model in the anomaly-based approach to DoS detection was quantified. Real background network traffic from the MAWI archive was superposed with synthetic DoS attacks. Detection performance was evaluated in terms of the AUC, with varying attack intensity and test subsequence length. It was observed that the performance loss is most notable for attacks of small intensities ($I < 5\%$). The performance loss diminishes as I increases, however at a slow rate; at a high attack intensity ($I = 50\%$), the AB detector still requires a test subsequence of length 1.5 times that of the LRT detector in order to match its performance on average.

Three future directions are proposed. 1) Although independence has been assumed here, it has long been known that network traffic exhibits long-range dependence. Furthermore, while the gamma distribution accommodates the skewness in the data, it does not address the inherent heavy-tailedness of network traffic. Research into improved modelling of network traffic is ongoing and these models can be incorporated into the present framework for improved detection accuracy. 2) The simplistic attack model can also be extended to better reflect the effects of the attack on the underlying network protocols. For instance, the present framework provides a platform to utilize previous work on the spectral characterization of DoS attacks [12] for enhanced detection. The framework can also easily be extended to attacks other than the DoS. 3) In this study, the attack model was either completely known or completely unknown. It is perhaps unrealistic to know the attack model completely before the attack occurs. Therefore, the effects of an attack model mismatch could be studied. This would allow to identify whether, realistically, it may be more effective not to assume an attack model at all; i.e., to take the anomaly-based approach, rather than to assume a mismatched attack model.

Acknowledgements

The authors gratefully acknowledge the interest of Dr Joey Jansen van Vuuren, Manager of the Cybersecurity Centre of Innovation, CSIR, South Africa, in this project and her continuous encouragement. The lead author (MWL) is indebted to ARMSCOR for being awarded the prestigious Ledger Bursary Grant. The authors also acknowledge the award of a joint UCT/CSIR Seed Grant, which enabled the initiation of the project.

References

[1] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: Methods, systems and tools," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 303–336, 2014.

[2] C. Gates and C. Taylor, "Challenging the anomaly detection paradigm: A provocative discussion," in *Proc. 2006 New Security Paradigms Workshop*, Sep., pp. 21–29.

[3] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in *Proc. 2010 IEEE Symp. Security and Privacy*, May, pp. 305–316.

[4] A. G. Tartakovsky, B. L. Rozovskii, R. B. Blaek, and H. Kim, "Detection of intrusions in information systems by sequential change-point methods," *Statistical Methodology*, vol. 3, no. 3, pp. 252 – 293, 2006.

[5] K. Cho, K. Mitsuya, and A. Kato, "Traffic data repository at the WIDE project," in *Proc. USENIX Annu. Technical Conf.*, Jun. 2000. [Online]. Available: http://mawi.wide.ad.jp/mawi/

[6] A. Scherrer, N. Larrieu, P. Owezarski, P. Borgnat, and P. Abry, "Non-gaussian and long memory statistical characterizations for internet traffic with anomalies," *IEEE Trans. Depend. Sec. Comput.*, vol. 4, no. 1, pp. 56–70, Jan. 2007.

[7] F. Simmross-Wattenberg, J. I. Asensio-Perez, P. C. de-la Higuera, M. Martin-Fernandez, I. A. Dimitriadis, and C. Alberola-Lopez, "Anomaly detection in network traffic based on statistical inference and α-stable modeling," *IEEE Trans. Depend. Sec. Comput.*, vol. 8, no. 4, pp. 494–509, Jul. 2011.

[8] G. Dewaele, K. Fukuda, P. Borgnat, P. Abry, and K. Cho, "Extracting hidden anomalies using sketch and non gaussian multiresolution statistical detection procedures," in *Proc. 2007 Workshop Large Scale Attack Defense*, Aug. 2007, pp. 145–152.

[9] P. Borgnat, G. Dewaele, K. Fukuda, P. Abry, and K. Cho, "Seven years and one day: Sketching the evolution of internet traffic," in *IEEE INFOCOM*, Apr. 2009, pp. 711–719.

[10] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1998.

[11] C. M. Bishop, "Novelty detection and neural network validation," in *Proc. IEEE Vision, Image and Signal Process.*, Aug. 1994, pp. 217–222.

[12] A. Hussain, J. Heidemann, and C. Papadopoulos, "A framework for classifying denial of service attacks," in *Proc. 2003 Conf. Appl., Technol., Architectures, and Protocols for Comput. Commun.*, Aug. 2003, pp. 99–110.