Techniques to Improve Stable Distribution Modeling of Network Traffic

Chad Bollmann*, Murali Tummala*, John C. McEachen*, James W. Scrofani*, and Mark Kragh†
*Department of Electrical and Computer Engineering
†Summer Engineering and Apprenticeship Program
Naval Postgraduate School
Monterey, California, USA
{cabollma, mtummala, mceachen, jwscrofa, mmkragh}@nps.edu

Abstract

The stable distribution has been shown to more accurately model some aspects of network traffic than alternative distributions. In this work, we quantitatively examine aspects of the modeling performance of the stable distribution as envisioned in a statistical network cyber event detection system. We examine the flexibility and robustness of the stable distribution, extending previous work by comparing the performance of the stable distribution against alternatives using three different, public network traffic data sets with a mix of traffic rates and cyber events.

After showing the stable distribution to be the overall most accurate for the examined scenarios, we use the Hellinger metric to investigate the ability of the stable distribution to reduce modeling error when using small data windows and counting periods. For the selected case and metric, the stable model is compared to a Gaussian model and is shown to produce the best overall fit as well as the best (or at worst, equivalent) fit for all counting periods. Additionally, the best stable fit occurs at a counting period that is five times shorter than the best Gaussian case. These results imply that the stable distribution can provide a more robust and accurate model than Gaussian-based alternatives in statistical network anomaly detection implementations while also facilitating faster system detection and response.

1. Introduction

Anomaly detection in computer networks has been an active field of research since the 1980s while yielding very few widely-adopted solutions [1], [2], [3]. While anonymous users can now launch denial-of-service (DoS) attacks using third parties for less than the price of a coffee, data networks continue to carry ever-increasing loads of critical information for services that require a timely and reliable connection [4].

In an attempt to address the difficulties associated with handling vast amounts of data while producing a manageable number of false detections, the majority of research efforts over the last decade appear to have focused on machine learning (ML) detection, or ML techniques in combination with other methods [5]. The focus of network anomaly detection research from the most recent survey to provide such a summary is shown in Table 1.

TABLE 1: Research efforts in network anomaly detection, 2000-2012. Adapted from [5].

Research Area	Percentage
Statistical	15
Knowledge based	3
Soft computing	18
Classification	28
Clustering	15
Combination learners	21

Our expectation is that this extended focus on ML techniques has left the statistical side of the discipline under-examined. As a result, key advances in both network anomaly detection and other, relatable disciplines have been only partially harvested. For instance, Gaussian processes are often used to model network traffic and detect anomalies, though their use can require abstractions and inaccuracies that have been known for quite some time [6], [7]. While ease-of-analysis and implementation are positive attributes of traditional approaches, our research considers alternative models with the goal of reducing modeling errors in order to improve overall detection system performance. We propose that by fully exploring and integrating non-Gaussian estimation and detection techniques and by leveraging the commoditization of computing power, we can develop real-time realizations of significantly more accurate statistical network anomaly detectors.

One objective of this paper is to convey our intended overall methodology for implementing a statistical anomaly detector that uses non-Gaussian assumptions in both the estimation and detection phases. The second objective is to examine the impacts of some of the detector design choices (*e.g.*, *a priori* distribution model, data aggregation period, and adaptive data collection techniques).

We believe that using the stable distribution in our system will permit adapting the detection process to better reflect fundamental characteristics of network traffic [7],

URI: http://hdl.handle.net/10125/50578
ISBN: 978-0-9981331-1-9

[8]. To support this, we will quantitatively show that the best distribution for our chosen traffic feature is the stable distribution and that its performance can be optimized to improve fit, which should minimize error thus produce a more accurate anomaly detection system.

We include initial results that demonstrate the speed, flexibility, and accuracy improvements enabled by the use of the stable distribution. Our goal is not to re-prove conclusions in [8], [9] but rather to quantitatively extend their work to publicly-available, reproducible datasets of real, physical network traffic across a range of cyber scenarios (*e.g.,* benign traffic, SYN flood and low-volume DoS attacks, and multi-method cyber attack). Additionally, we show using an analyzed scenario that for the maximum-accuracy case, the stable distribution permits the use of smaller data windows and counting periods than the Gaussian distribution *and* improves the overall fit to the data. Our results should be extensible to rapid detection schemes in other disciplines examining appropriately-distributed data.

This paper is organized as follows. Section 2 provides background on our anomaly detection model and approach for reducing errors in the detection process. Section 3 discuses the stable distribution and its application to our problem. Results are provided in Section 4, and conclusions are discussed in Section 5.

2. Reducing Avoidable Errors

With a goal of reducing false positive rates closer to a usable threshold [10], we examined the statistical anomaly detection process with a focus on identifying processes that introduce error through assumptions or implementation decisions that do not respect fundamental truths of the underlying problem. While these design trade-offs are inherent to most practical implementations of theoretical concepts, our goal was to identify the likely major sources of these *avoidable errors* and then assess whether they could be minimized through recent advances, or by adapting approaches from other disciplines. This approach required two distinct steps: First understanding the sources of avoidable errors, and then identifying the steps we would address to reduce these errors and improve the process. Identifying the sources of error required a review of the existing literature and development of a theoretical model for an anomaly detector.

2.1. Anomaly Detection Model

The steps of basic statistical anomaly detection are straightforward. A simple implementation first obtains and preprocesses its data, network traffic in our case. The detector then compares the measured data to the normal (*i.e.*, benign conditions) estimate, computing the deviation. If the deviation exceeds an allowed threshold, the detector declares an anomaly. Algorithms designed to improve detection accuracy introduce complexity to this design, through steps such as updating estimates of normal, adjusting the threshold, or considering more than one feature of the traffic (*e.g.*, packets, bytes, network addresses, etc.).

To assist in understanding design choices and identify candidate processes for reducing errors, we developed a detailed model of our intended detection process. Shown in Fig. 1, our model collects network traffic for a given period of time, the processing *window*, ϖ. To eliminate unnecessary data in the window, the system filters the collected data for specific features (*e.g.*, destination IP address or packets per unit time). A separate process then measures some aspect of these features over a uniform counting period, the sub-window (Δ_{sw}). Each measurement yields an individual vector element, p_i. To produce the detector input, the process then concatenates M instances of these measurements p_i to produce the detector input, the *feature vector* $\overrightarrow{p}$.

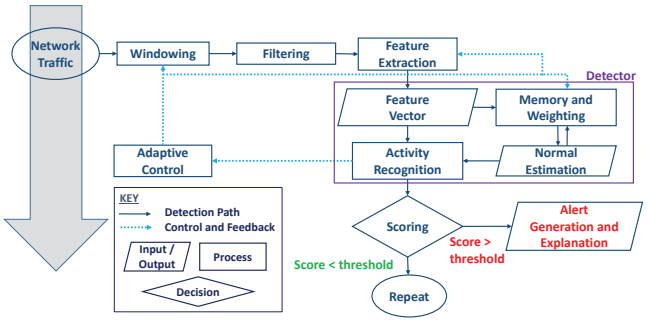

Figure 1: Detection process model

Next the detector compares $\overrightarrow{p}$ to an estimated vector of the feature's normal, or expected, values. This estimate of normal is produced though a combination of memory, weighting (or decay), and updating processes. If the estimate differs substantially from the current measurement, based on a distance metric and threshold, an alert is generated through a scoring and reporting process. Finally, an adaptive control mechanism can update the threshold, window length, monitored features, and memory processes, depending on the complexity of the implemented model.

Mathematically, the number of M elements p_i for $i \in [1, M]$ are related to window and sub-window sizes by

$$M = \varpi / \Delta_{sw}, \tag{1}$$

where

$$\varpi = [t_{stop} - t_{start}] \tag{2}$$

and t_{start} and t_{stop} are the start and stop times of the data processing window, respectively.

As we discuss in Section 4.3, the size of the window and the sub-window can significantly affect the fit of the model and distribution of features, and thus could influence the performance of the entire detection system.

2.2. Framework for Reducing Errors

Our review of the literature combined with our process model in Fig. 1 suggested modifying three system processes to improve overall system accuracy. First, the traffic model should reflect the fundamental nature of the monitored feature. Gaussian distributions are frequently assumed, but

recent work has shown that alternatives, such as the stable distribution, more accurately reflect many distributions of network traffic features [8], [11].

Similarly, our second modification is that the detection algorithm ("Activity Recognition" in fig. 1) should reflect the nature of the inputs, both the signal and the noise. It is a long-identified principle of the radar and underwater acoustic disciplines that detectors implementing tests based on non-Gaussian assumptions significantly improve detection accuracy in non-Gaussian noise conditions [12], [13]. This concept has been validated in additional areas including the impulsive noise environment that characterizes wireless communications [14].

Our third process for improvement and the focus of the work in this paper is data windowing. It is our prediction that, when more flexible and accurate non-Gaussian distributions are used to model the data, we can further improve overall system accuracy through tuning the amount of data aggregation, or window size, and the detector threshold.

This three-process approach is depicted in Fig. 2, which presents our framework for improving accuracy through reducing accepted errors in design and implementation. In this framework, the cascading nature of avoidable errors is evident and suggests that small improvements in the initial stages of the process could yield large improvements in results.

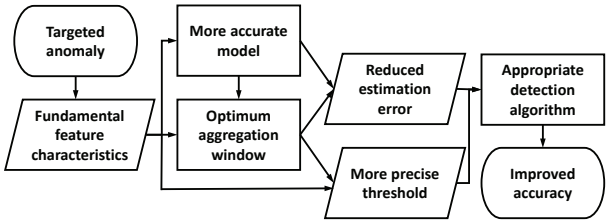

Figure 2: Chosen framework for reducing avoidable errors

2.3. Background on Windowing

As examining the windowing process is the principal contribution of this paper, some additional context for our research may be useful. As a consequence of the Central Limit Theorem, the distribution of finite-variance data at sufficiently large aggregations should tend towards Gaussian. However, the distributions of certain aspects of network traffic have been shown to be fundamentally heavy-tailed (*i.e.*, not of finite variance) [7], [9].

Accordingly, it is our belief that in some detection implementations, depending on the amount of aggregation (*i.e.*, window and sub-window size), processes might display Gaussian characteristics at larger data windows and non-Gaussian characteristics at smaller aggregations. Indeed, we show in Section 4.3 that while counts of packet per sub-window are Gaussian-distributed for large windows, for small windows (and sub-windows) these feature vectors instead possess a stable distribution.

Thus, in order to determine the appropriateness and accuracy of stable models, we must examine smaller window lengths. Previous work in the literature has focused on choosing window sizes that are small enough to minimize accuracy impacts from other properties of network traffic, such as non-stationarity and self-similarity [6], [11]. These concerns may dominate when using large aggregation windows that are most appropriate for a Gaussian distribution; for instance, [11] used 30-minute windows as the basis for avoiding non-stationarity effects.

However, the flexibility of the stable distribution permits significant fit improvements over the Gaussian case when using small windows, encouraging the introduction of new window size criteria and enabling the adaptation of the data to the model. As such, we propose a different optimization criterion that can shrink the window even further, and in some cases to seconds: That of maximizing the fit of the distribution to the data. The shorter windows that result thus remove error through providing a more accurate estimation of normal while offering the potential to improve the response time of the detection system.

3. The Stable Distribution and Anomaly Detection

We will now provide some brief (due to time constraints and extensive documentation in existing literature) background into the mathematical foundations and practical applications of stable distributions. We refer the interested reader to the cited references for more comprehensive background, including [8], [11], [15], and [16].

3.1. Mathematical Characterization

The stable distribution is characterized by four parameters representing the size of the tail (α), asymmetry or skewness (β), dispersion (γ), and location (μ). These four parameters provide tremendous flexibility, allowing symmetric as well as wholly-skewed forms, but also complicate fitting the distribution to data. The stable parameters are constrained per Table 2.

TABLE 2: Stable distribution parameters and constraints

Parameter	Property	Range
α	Tail size	$(0, 2]$
β	Asymmetry	$[-1, 1]$
γ	Spread	$[0, \infty)$
μ	Location	$\Re$

The computational costs of applying the stable distribution are further increased by its lack of a closed-form solution. It must be defined using its characteristic function which has two cases depending on the value of α,

$$E\left[e^{i\theta Z}\right] = e^{-\gamma^{\alpha}|\theta|^{\alpha}\left[1-i\beta\tan(\frac{\pi\alpha}{2})\mathbf{sign}(\theta)\right]+i\mu\theta} \quad (3)$$

for the case of $\alpha \neq 1$, and

$$E\left[e^{i\theta Z}\right] = e^{-\gamma|\theta|\left[1+i\beta\frac{2}{\pi}(ln|\theta|)\,\mathbf{sign}(\theta)\right]+i\mu\theta} \quad (4)$$

for $\alpha = 1$ [15].

The stable distribution is more flexible than other distributions partially because of its four degrees of freedom (*i.e.*, parameters α, β, γ, and μ). This allows for the implementation of very expressive models to more accurately reflect both small and large changes in the underlying time series. This modeling quality has undoubtedly contributed to the stable distribution's adoption in numerous applications including financial forecasting, sound propagation in water, radar, geology, and astronomy, among many others [16], [17].

The stable distribution possesses special cases with closed-form solutions and heavy tails, including the Gaussian (where $\alpha = 2$), Cauchy ($\alpha = 1$), and Lévy ($\alpha = 0.5$) distributions. The Gaussian is pertinent as the limiting case for large window and sub-window sizes, as shown in Section 4.3. Finally, to model highly-impulsive data series where $\alpha < 1$, the stable distribution can be restricted to $(-\infty, \mu]$ or $[\mu, \infty)$ through constraining $\beta = \{-1, 1\}$ and μ appropriately [16].

3.2. Application to Anomaly Detection

The stable distribution has repeatedly been found to improve signal modeling and detection in the presence of impulsive noise (or high-variance outliers) [13], [17], [18]. The existence of high variance outliers (demonstrated as heavy tails) in the distributions of many aspects of network traffic has been previously demonstrated [7]. These long tails, as well as the characterization of network traffic as the result of a combination of the many random processes of transmitting, transporting, and routing data from source to destination, warrant the application of the stable distribution to the networking field as well [18].

However, there are limited examples of this application in networking literature; most recently, stable distributions appear to have been used by Simmross-Wattenberg *et al.* as traffic models in 2008 [9] and as part of a network anomaly detector in 2011 [11]. Our review of the literature identified no subsequent work using stable distributions in network anomaly detection, though other disciplines such as radar and image processing continue to apply the stable distribution to detection problems for the reasons discussed previously [19], [20].

Despite a lack of follow-on work in the networking and cyber-security fields, we believe that the flexibility of the stable distribution, combined with its improved ability to accurately model network traffic, fully justifies its adoption in network anomaly detection. This flexibility, including available constraints, allows the stable distribution to represent fundamental characteristics of network traffic without approximation or abstraction. For instance, network traffic features are usually one-sided, asymmetric, and wholly positive, characteristics which are native to only a few other distributions [9]. Also, as we will see in Section 4.2, some distributions previously used in literature (*e.g.*, Weibull, gamma) cannot model processes with zero counts, unlike the stable distribution [6].

The computational cost of utilizing stable distributions is a likely contributor to its limited adoption to date, even given the advantages discussed above [16]. However, computing power continues to increase while costs fall, facilitating practical implementation of real-time network cyber security approaches using the stable distribution. Further, other disciplines, particularly the financial field, have long accepted the costs of the stable distribution in order to harvest its benefits of robustness and accuracy. These benefits, as applied to network traffic modeling, are demonstrated in the next section.

4. Results

Our preliminary applications of stable distributions to network traffic quantitatively confirm the conclusions in [9]. The flexibility and robustness of the stable distribution enable models that more accurately depict network traffic than other alternatives, including Gaussian, Poisson, and gamma distributions, across a range of windows and traffic scenarios.

4.1. Examined Datasets and Applied Metrics

Previous studies have quantitatively demonstrated the improved modeling accuracy of stable distributions at both higher (approx. 30 Mbps) and lower (approx. 0.37 Mbps) link volumes under benign conditions [11]. We were interested in the *flexibility* of the stable distribution as characterized by its performance across a range of traffic scenarios, including sparse links, as well as during transitions in traffic volume due to background network processes, and during attacks.

The ISCX dataset [21] was chosen to investigate stable distribution accuracy in many different scenarios and at low link volumes; average link volume was 74 kbps for the 14 June trace. This sparse link required data windows on the order of minutes or hours, and large sub-windows, while our research objective is ultimately data windows of less than a minute and sub-windows of milliseconds. However, the ISCX dataset is publicly-available, contains both high and low-volume DoS attacks, and is well-documented [22]. Also, the variety of scenarios and relatively low traffic volume helped to demonstrate the flexibility of the stable distribution. For the ISCX scenarios, we examined only the portion of the network traffic destined for the web server (IP address 192.168.5.122) in two traces from different days, 14 and 15 June.

To validate the effectiveness of the stable distribution at higher link volumes (16 Mbps), we used a dataset available from the 2010 Mid-Atlantic Collegiate Cyber Defense Competition (MACCDC) [23], again filtered for a single IP address (41.204.84.14). The role of this host was not identified in the dataset documentation, but we judged that

filtering for the host was an acceptable abstraction to support our research goal of identifying various DoS attacks on web servers.

We then expanded our high link speed analysis using data from the Measurement and Analysis on the Wide Internet (MAWI) archive. These datasets are 15-minute truncated captures of anonymized real-world traffic transiting a trans-Pacific backbone link [24]. The selected MAWI trace [25] allowed us to extend our analysis to backbone speeds (approximately 750 Mbps) and seemed most appropriate for investigating the affects of sub-window size due to its real-life traffic mix.

Figure 3 summarizes the results of analyzing these 4 different network traffic traces under 15 different scenarios, for various traffic conditions and window sizes. The feature collected in all scenarios was packets per sub-window, since this is a non-spoofable characteristic of DoS attacks as well as many other cybersecurity anomalies.

Feature vectors from these different scenarios were processed using native functionality of MATLAB [26] to fit various distributions to each filtered trace, and the log likelihood (LL) results from the maximum likelihood fit were recorded and used to assess the fit of each distribution to the to the associated data set. For each scenario, the results were then normalized by the LL value of the best-performing (*i.e.*, closest fit) distribution. For each distribution, the performance across each scenario was then averaged and is shown by the "Avg. Normalized LL" in the bottom row of the figure. Not all scenarios are shown in Fig. 3 due to space constraints; the displayed scenarios are discussed in further detail in Section 4.2.

Scenario	Trace	Scenario	Subwindow [sec]	Window [min]	Normalized log-likelihood of given fit: Stable	Weibull	Exponential	Gaussian
1	ISCX 15th	Benign Low Vol	60	83	**1.00**	1.07	1.14	1.45
2	ISCX 15th	Benign Mid Vol	60	83	1.01	**1.00**	1.02	1.03
3	ISCX 15th	Attack Transition	60	83	**1.00**	1.03	2.10	1.03
4	ISCX 15th	Flood Attack	60	61	**1.00**	**1.00**	3.12	1.01
8	ISCX 15th	Scenario 1	10	25	**1.00**	1.01	1.17	1.67
9	ISCX 15th	Scenario 1 (5 min)	10	5	1.04	**1.00**	**1.00**	1.08
10	MACCDC	30 Second Trace	1	0.5	**1.00**		1.09	1.05
11	MACCDC	30 Second Trace	0.1	0.5	**1.00**		1.82	2.40
12	MACCDC	Full Trace	1	5.6	**1.00**		1.08	1.00
13	MACCDC	Full Trace	0.1	5.6	**1.00**		1.55	1.98
14	MAWI	Benign 20150925	0.002	0.07	**1.00**		1.22	1.01
15	MAWI	Attack 20150925	0.01	0.10	**1.00**		1.27	1.01
Avg. Normalized LL for ALL Scenarios (some not shown)					**1.00**	**1.02**	**1.38**	**1.30**

Figure 3: Normalized, best fit assessment for distributions in various traffic scenarios. Values closer to one are better. LL results of gamma and Poisson distributions were omitted for display purposes; the Weibull distribution performed better than the gamma in nearly all scenarios.

The LL value results from use of the maximum likelihood algorithm to estimate the unknown parameter θ of a distribution that is most likely to produce a known data series with samples X_i. The likelihood function of the unknown parameter, $\varphi(\theta)$, can be defined in terms of the product of the probability of observing each X_i for a given value of θ such that

$$\varphi(\theta) = f(X_1|\theta) \times ... \times f(X_n|\theta) = \prod_{i=1}^{M} f(X_i|\theta) \quad (5)$$

for integer $i \in [1, M]$ [27]. By taking the natural log of (5) and finding the value of θ that maximizes this likelihood, $\hat{\theta}$, the log-likelihood is calculated as [27]

$$\ln \varphi(\hat{\theta}) = \sum_{i=1}^{M} \ln f(X_i|\hat{\theta}) \quad (6)$$

LL is frequently used to determine which distribution provides the *most accurate* fit [27]. However, the nature of the LL prevents it from being a preferred metric for comparing the fit of a single distribution across data sets of different sizes, because the LL can be seen in (6) to be the result of a sum, and as such is dependent on the size of the dataset used to produce the estimation.

To assess fit of the stable and Gaussian distributions as a function of sub-window size, we utilized the average Hellinger metric [28], defined as

$$S(\overrightarrow{p}, f_Z(z)) = \frac{1}{N}\sqrt{2\sum_{j=1}^{N}\left(\sqrt{h_j} - \sqrt{f_Z(j)}\right)^2} \quad (7)$$

where $f_Z(z)$ is obtained from the results of the maximum likelihood fit to the histogram of $\overrightarrow{p}$. This histogram has N bins, each with mean location j and bin value h_j for $j \in [1, N]$. The metric in (7) is averaged by integer $N \in (0, \infty)$ to enable comparing fits across data series of different sizes.

4.2. Accuracy and Robustness

The stable distribution provides significant accuracy improvements over the selected alternative distributions, as shown by the Average Normalized LL in Fig. 3. Scenarios were chosen to examine to assess the flexibility of the stable distribution under various traffic conditions: benign, low volume; benign, medium volume, bursty conditions characteristic of background network processes; part-benign and part-attack; high-volume DoS; and low-volume DoS.

Additionally, Scenarios 8 and 9 examined the boundaries of the stable distribution's performance using part of the same data as Scenario 1 but with significantly smaller windows and sub-windows. Scenarios 10-12 used the MACCDC trace to continue exploring the effects of sub-window and window sizing at higher data rates and the suitability of the stable distribution to model abnormal traffic conditions (in this case, heavy cyber attack with limited benign background). Scenarios 14 and 15 confirmed the previous results using the most realistic, highest-speed trace (MAWI).

Regarding the examined distributions in Fig. 3, we apply the exponential distribution in all scenarios as a sort of limit marker for sparse link conditions, because it demonstrates the best fit under the lowest-traffic scenarios (*e.g.*, Scenario 9), where there are a large number of sub-windows with low or zero packet counts. The Weibull and gamma distributions were not assessed for all scenarios. These distributions, special cases of the exponential distribution, are constrained $\in (0, \infty)$, and thus can only be used as models in low-traffic scenarios if sub-windows with a packet count of zero

are excluded during the curve fit. These distributions were initially included to assess potential alternatives, but because sub-windows with zero packet counts are valid outcomes on sparse links, we do not consider these distributions viable alternatives and discontinued their analysis after Scenario 10. Finally, the Poisson distribution (results not shown) was only assessed for a few scenarios before analysis was discontinued due to its significantly lower performance than all other distributions, as previously demonstrated in the literature [9].

Some interesting conclusions can be drawn from Fig. 3. First, as demonstrated in other disciplines, the heavy-tail of the stable distribution makes it more robust to outliers than other distributions. Scenario 1, while low volume on average (mean of 39 packets per minute), included 4 outlier packet storms greater than 1500 packets per minute. These outliers explain the stable distribution's improved relative results as compared to Scenario 2, which had no significant outliers. This robustness continued during actual attacks and transition periods, when attacks began, such as Scenarios 3, 4, and 7.

The flexibility of the stable distribution is demonstrated by Scenarios 9 - 11, where we attempted to assess the limits of the stable distribution by both shrinking the window size and the aggregation period. This produced a large number of sub-windows with zero packet counts and an extremely left-skewed packet count distribution; in Scenario 9 the exponential distribution was the most accurate. The stable distribution's performance did not suffer extremely under these conditions, however, still producing a better fit than the Gaussian. This suggests that Scenario 9 parameters provide a sort of lower bound regarding the smallest achievable window and sub-window sizes for the trace's 16 Mbps link rate.

4.3. Effect of Window and Sub-window Sizes

Two other important conclusions can be drawn from Figs. 3, and can be conveyed with the help of Fig. 4 and 5, which compare packet count histograms for different sub-window sizes. These figures demonstrate the importance of the stable distribution in network traffic modeling when utilizing small windows. Small windows are desirable because they speed detection system response time; faster data aggregation and comparison leads to faster detection of cyber security events and response by network defenders. Comparing Scenarios 12 and 10 (or 13 and 11) for stable fit *relative* to the Gaussian case show that the gain from using the stable distribution improves as window size shrinks from minutes to seconds. The stable fit converges to Gaussian in the nearly 6-minute window and 1 second sub-window case in Fig. 4, but when a smaller window is used (Fig. 5), the better fit becomes decidedly non-Gaussian due to the heavy left tail.

A similar effect for sub-window can be visually observed in Fig. 4. The symmetric and near-Gaussian data distribution of the 1 second sub-window case in becomes asymmetric and heavy-tailed when 100 millisecond sub-windows

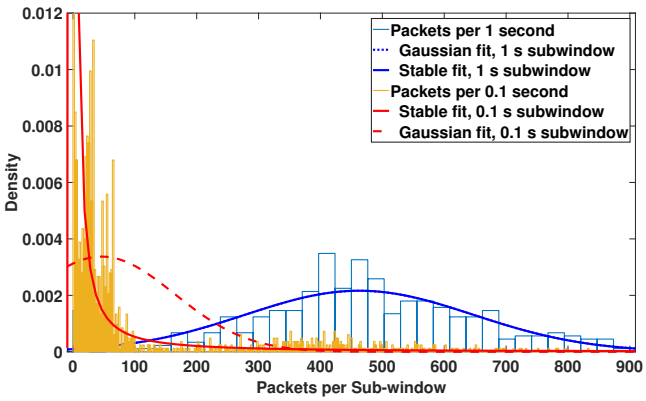

Figure 4: Measured and estimated distributions of packet count of the MACCDC trace over 336 seconds (Scenarios 12 and 13). Note that for the 1 second sub-window case, the Gaussian fit plots on top of the stable fit. The origin bin for the 0.1 second case is truncated for display purposes, and has a value of 0.237.

are used. The impacts of both sub-window and window qualitatively show that the stable distribution provides a better model at smaller aggregation periods. This result is intuitively satisfying; aggregating less data should move us farther from the limiting results of the central limit theorem that lead to a Gaussian result. The implication of this result is that for anomaly detection systems which use shorter windows to lower response times, the stable distribution should be used to model traffic.

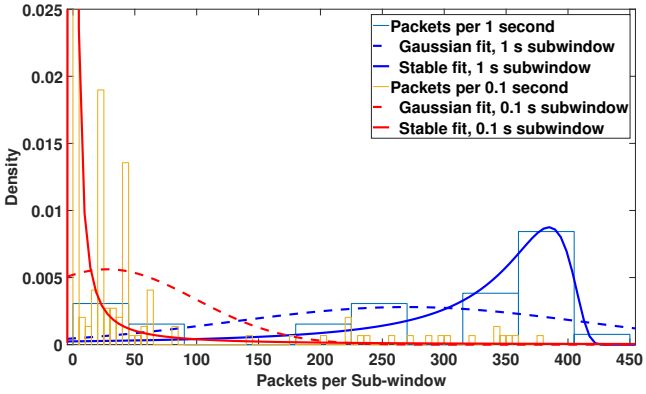

Figure 5: Measured and estimated distributions of packet count of the MACCDC trace over 30 seconds (Scenarios 10 and 11). The origin bin for the 0.1 second case is truncated for display purposes, and has a value of 0.131.

4.4. Towards Adaptive Sub-window Sizes

The visual changes in fit of the stable distribution for different sub-windows imply that it should be possible to select a sub-window size that minimizes stable fit residuals (*i.e.*, fit error) for a given link traffic rate. Quantitatively, the relative LLs of Scenarios 11 and 12 begin to confirm this;

the Gaussian distribution's fit grows relatively worse worse as the sub-window length is shortened from 1 sec to 0.1 sec.

To examine this idea, we analyzed a portion of the MAWI trace by determining the average Hellinger distance between the data and best-fit stable and Gaussian distributions using six sub-window sizes and a fixed window size of four seconds. Table 3 and Fig. 6 show the results of this analysis. We find that for a constant window size, the average divergence (*i.e.*, error of fit) of both distributions changes with sub-window size and that both cases possess a global minimum (within the sub-window sizes examined). The global minimum demonstrates that it should be possible to optimize fit (and thus detection system accuracy) by adapting sub-window size when using the stable distribution.

TABLE 3: Average Hellinger distance of stable and Gaussian fits for constant window size. The 0.5 ms case is not shown for display purposes; both values in this case are 2.46e-2.

Distribution	**Sub-window [ms]**				
	1	2	5	10	20
Stable	1.08e-3	7.79e-4	1.01e-3	1.34e-3	1.51e-3
Gaussian	2.47e-3	1.78e-3	1.60e-3	1.43e-3	1.50e-3

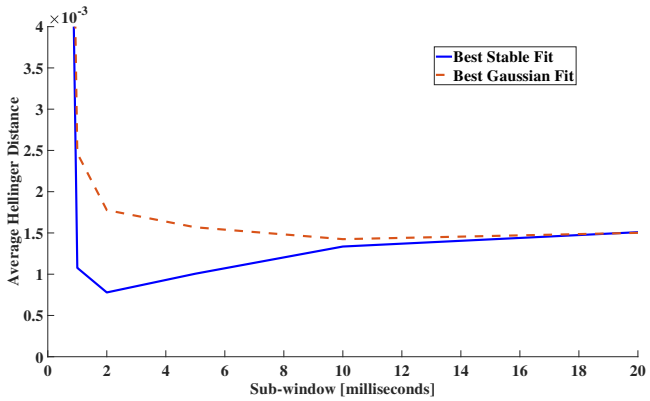

Figure 6: Average Hellinger distance of the best-fit stable and Gaussian distributions, for a fixed 4 second window with varying sub-windows. The 0.5 millisecond case was truncated for display purposes, and has a value of 0.02462 for both fitted distributions.

Three aspects of Fig. 6 warrant further discussion. Defining the smallest distance as the *best accuracy*, Fig. 6 shows that the stable distribution's accuracy in the 1, 2, and 5 ms cases are better than the Gaussian distribution's *best* case. Second, the stable distribution provides its best accuracy at a sub-window that is 5 times shorter than the Gaussian best case, which should lead to a significantly faster anomaly detection time.

Finally, in the 0.5 and 20 millisecond sub-window cases, the maximum likelihood estimate for the stable distribution converged to the Gaussian distribution, as demonstrated by their equivalent accuracy. Reinforcing this conclusion, the fitted stable distribution parameters in these two cases are the special Gaussian case, with $\alpha = 2$. Thus, the stable distribution should provide better or equivalent accuracy to the Gaussian fit in *all* cases. In all, these results demonstrate the flexibility of of the stable distribution and its ability provide accurate estimation at both large and small aggregations.

Moving forward, we continue to acquire large, real-world data sets and confirm our initial results with more extensive analysis. We are also assessing empirical and analytical solutions for optimizing window size based on native features of the link being monitored, such as background traffic rate. We expect that analyzing additional realistic datasets containing a mixture of background traffic will reinforce our conclusions and more thoroughly demonstrate the improved fit of the stable distribution at smaller aggregations.

5. Conclusion

This work has demonstrated benefits of using stable distributions to detect cyber events. For the three datasets and range of scenarios examined, the stable distribution was shown to be the most suitable distribution for modeling the selected feature of packets per sub-window across a range of different aggregation sizes and network event scenarios. Integrating the stable distribution model into our planned detection system, an item of future work, should facilitate more robust detection even in the presence of large fluctuations of link traffic volume and conditions.

As shown in Fig. 3, the stable distribution can provide better estimation of normal conditions at low link rates as well as smaller sub-window and window lengths. This will speed detection system response since the time to identify an anomaly is constrained by the window size and processing costs.

Also, regarding the selection of window and sub-window sizes, we have clearly shown through Fig. 6 that the accuracy of the estimated stable fit is significantly influenced by sub-window size. This demonstrates that fit measures can be used to optimize the selection of sub-window length when using the stable distribution. The impact of window size on distribution fit was also demonstrated qualitatively for the Gaussian case using Figs. 4 and 5; we expect that this can be quantitatively extended to the stable distribution in the future.

Harnessing these conclusions should enable the introduction of new data-adaptive network security techniques that optimize model fit against shorter system response times. We expect that combining stable methods of modeling traffic and detecting anomalies with the data-adaptive techniques demonstrated in this work will significantly improve the end-to-end accuracy of our planned detection system.

Acknowledgments

This work was supported by the Laboratory for Telecommunication Sciences. We thank the panel of anonymous reviewers for their time and valuable insights which improved the quality of this paper.

6. References

[1] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in *2010 IEEE symposium on security and privacy*. IEEE, 2010, pp. 305–316.

[2] A. Patcha and J.-M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," *Computer networks*, vol. 51, no. 12, pp. 3448–3470, 2007.

[3] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," *computers & security*, vol. 28, no. 1, pp. 18–28, 2009.

[4] J. Networks. White Paper - Defending Against Application-Layer DDoS Attacks. Accessed Nov. 16, 2016. [Online]. Available: www.juniper.net

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

[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 Transactions on Dependable and Secure Computing*, vol. 4, no. 1, pp. 56–70, 2007.

[7] W. Willinger, V. Paxson, and M. S. Taqqu, "Self-similarity and heavy tails: Structural modeling of network traffic," *A practical guide to heavy tails: statistical techniques and applications*, vol. 23, pp. 27–53, 1998.

[8] G. Xiaohu, Z. Guangxi, and Z. Yaoting, "On the testing for alpha-stable distributions of network traffic," *Computer Communications*, vol. 27, no. 5, pp. 447–457, 2004.

[9] F. Simmross-Wattenberg, A. Tristan-Vega, P. Casaseca-de-la Higuera, J. I. Asensio-Perez, M. Martın-Fernandez, Y. A. Dimitriadis, and C. Alberola-Lopez, "Modelling network traffic as α-stable stochastic processes: An approach towards anomaly detection," *Proc. VII Jornadas de Ingenieria Telematica (JITEL)*, pp. 25–32, 2008.

[10] S. Axelsson, "The base-rate fallacy and the difficulty of intrusion detection," *ACM Transactions on Information and System Security (TISSEC)*, vol. 3, no. 3, pp. 186–205, 2000.

[11] F. Simmross-Wattenberg, J. I. Asensio-Perez, P. Casaseca-de-la Higuera, M. Martin-Fernandez, I. A. Dimitriadis, and C. Alberola-Lopez, "Anomaly detection in network traffic based on statistical inference and alpha-stable modeling," *IEEE Transactions on Dependable and Secure Computing*, vol. 8, no. 4, pp. 494–509, 2011.

[12] S. Banerjee and M. Agrawal, "Underwater acoustic noise with generalized gaussian statistics: Effects on error performance," in *OCEANS-Bergen, 2013 MTS/IEEE*. IEEE, 2013, pp. 1–8.

[13] G. Tsihrintzis and C. Nikias, "Evaluation of fractional, lower-order statistics-based detection algorithms on real radar sea-clutter data," *IEE Proceedings-Radar, Sonar and Navigation*, vol. 144, no. 1, pp. 29–38, 1997.

[14] H. El Ghannudi, L. Clavier, N. Azzaoui, F. Septier, and P.-A. Rolland, "α-stable interference modeling and cauchy receiver for an ir-uwb ad hoc network," *IEEE Transactions on Communications*, vol. 58, no. 6, pp. 1748–1757, 2010.

[15] G. Samoradnitsky and M. S. Taqqu, *Stable non-Gaussian random processes: stochastic models with infinite variance*. CRC press, 1994, vol. 1.

[16] J. P. Nolan, "Modeling financial data with stable distributions," *Handbook of Heavy Tailed Distributions in Finance, Handbooks in Finance: Book*, vol. 1, pp. 105–130, 2003.

[17] R. D. Pierce, "Application of the positive alpha-stable distribution," in *Higher-Order Statistics, 1997., Proceedings of the IEEE Signal Processing Workshop on*. IEEE, 1997, pp. 420–424.

[18] S. Painter, "Random fractal models of heterogeneity: The levy-stable approach," *Mathematical Geology*, vol. 27, no. 7, pp. 813–830, 1995.

[19] V. A. Aalo, K. P. Peppas, and G. Efthymoglou, "Performance of cacfar detectors in nonhomogeneous positive alpha-stable clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 3, pp. 2027–2038, 2015.

[20] H. Sadreazami, M. O. Ahmad, and M. S. Swamy, "A study of multiplicative watermark detection in the contourlet domain using alpha-stable distributions," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4348–4360, 2014.

[21] Candadian Institute for Cybersecurity. Intrusion detection evaluation dataset. Information Security Center of Excellence. Accessed 20 May 2017. [Online]. Available: http://www.unb.ca/cic/research/datasets/ids.html

[22] A. Shiravi, H. Shiravi, M. Tavallaee, and A. A. Ghorbani, "Toward developing a systematic approach to generate benchmark datasets for intrusion detection," *computers & security*, vol. 31, no. 3, pp. 357–374, 2012.

[23] 2010 Mid-Atlantic Collegiate Cyber Defense Competition. Hosted by Netresec. Accessed 10 Jun 2017. [Online]. Available: https://www.netresec.com/?page=MACCDC

[24] R. Fontugne, P. Borgnat, P. Abry, and K. Fukuda, "Mawilab: combining diverse anomaly detectors for automated anomaly labeling and performance benchmarking," in *Proceedings of the 6th International COnference*. ACM, 2010, p. 8.

[25] MAWI. Dataset for 25 September, 2015. Measurement and Analysis on the WIDE Internet. Accessed 20 Jun 2017. [Online]. Available: http://mawi.wide.ad.jp/mawi/samplepoint-F/2015/201509251400.html

[26] MATLAB, *version 9.2 (R2017a)*. Natick, Massachusetts: The MathWorks Inc., 2017.

[27] D. Panchenko. 18.443 Statistics for Applications. Massachusetts Institute of Technology: MIT OpenCourseWare. Accessed 06 Jun 2017. [Online]. Available: https://ocw.mit.edu

[28] D. J. Weller-Fahy, B. J. Borghetti, and A. A. Sodemann, "A survey of distance and similarity measures used within network intrusion anomaly detection," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 70–91, 2015.