Proceedings of the 43rd IEEE Conference on Decision and Control, Bahamas, December 2004

On the Statistical Distribution of Processing Times in Network Intrusion Detection

João B. D. Cabrera[a], Jaykumar Gosar[b], Wenke Lee[b] and Raman K. Mehra[a]

Scientific Systems Company, Inc.[a]
500 West Cummings Park, Suite 3000
Woburn MA 01801 USA

Georgia Institute of Technology[b]
College of Computing
801 Atlantic Drive
Atlanta, GA 30332 USA

Abstract

Intrusion Detection Systems (IDSs) are relatively complex devices that monitor information systems in search for security violations. Characterizing the service times of network IDSs is a crucial step in improving their real time performance. We analyzed about 41 million packets organized in five data sets of 10 minutes each collected at the entry point of a large production network and processed by Snort, a commonly used IDS. The processing times of the three main stages in Snort were measured. The main conclusions of our study were: (1) Rule checking accounts for about 75% of the total processing time in IDSs, with mean payload checking time being 4.5 times larger than mean header checking time. (2) The distribution of rule checking times is markedly bimodal, a direct consequence of the bimodality in packet composition in current high speed Internet traffic. (3) Header processing times have a small variance and small correlation coefficients. (4) In contrast, the distribution of payload processing times displays high variance, in a form that can be generally characterized as "slightly heavy-tailed". Explicitly, payload processing times have a Lognormal upper tail, clipped at the top 1%. This extreme 1% upper tail is better fit by an Exponential distribution. (5) Additionally, payload processing times were shown to be highly correlated, with correlation coefficients several orders of magnitude higher than the confidence bands for the standard whiteness test. The impact of these findings in the design of IDSs for real time operation in networks is discussed, and compared with existing results for processing times for Unix processes, which were shown to display pronounced heavy-tailed characteristics.

1 Introduction

Intrusion Detection Systems (IDSs) are relatively complex devices that monitor information systems in search for security violations - [5], [19]. In network-based IDSs, data packets enter the IDS and are subjected to a number of processing steps whose ultimate objective is to determine if the packet contains an intrusion or not. There are essentially three main steps in network IDSs, such as `Snort` – [6]:

- Packet decoding: Decodes the header information at the different layers and creates a data structure for the packet, which is used in the next steps.
- Preprocessing: Performs a number of preparatory steps in the packet, such as normalization, IP fragment reassembly, TCP stream reconstruction, etc.
- Rule checking: Checks if the packet contains a particular string, or a collection of strings, which are associated with an intrusion. A rule consists at a minimum of a type of packet to search (protocol type), a string of content to match and a location where that string is to be searched for – [24]. Rule checking in `Snort` has two (sub)-steps: Non-Content Matching (NCM), performed in the packets' headers and Content Matching (CM), performed in the packets' payloads.

Like any computing device operating in real time, the operational performance[1] of an IDS depends on the arrival rates of packets streaming at its input, and the service rates it provides to the packets. The two components are equally important in characterizing the performance of the IDS, and their understanding is crucial for the design of more efficient systems. The arrival rates of packets at network IDSs are the arrival rates of packets into the networking device in which it is installed, modulated by traffic shaping, if applicable. Much is known about the statistical properties of arrival rates of packets in the Internet, result of extensive research, especially in the last decade – [7] and references therein. In contrast, very little is known about the statistical properties of service times in network IDSs. The focus of the research on network IDS evaluation has been on measuring the *performance metrics* as a function of the network load, traffic characteristics (balance between protocol types, presence of fragments, etc.) and complexity of the ruleset – eg. [12], [23]. A recent study - [2] - has measured the processing times of the various components of `Snort`, but no statistical characterization was attempted. The objective was to construct synthetic workloads out of real traffic, for use in IDS benchmarking. In this paper, we study the statistical properties of

[1] By operational performance we mean the usual metrics of mean service time, percentage of dropped packets, etc.

the service times in Snort, which we believe is an essential step in designing more efficient IDSs. Service times for Unix processes were investigated in [16] and [13] leading to new strategies for load balancing and processor design. We expect a similar effect from this current study, in the design of network IDSs.

2 Statistical Distributions of Processing Times

2.1 Data collection and general characteristics

Five data sets of about 10 minutes corresponding to 5-12 million packets each were collected using the standard tool tcpdump at the main entry point to the network serving the College of Computing at the Georgia Institute of Technology. The packet streams were then inputed into Snort version 2.0.5, with its standard ruleset of 1458 rules. Especially designed instrumentation tools recorded the following variables for each packet:

- P: Packet size (payload size and header size are recorded separately).
- T_g: Time spent in the grinder (packet decoding).
- T_p: Time spent in preprocessing.
- T_{d_1}: Time spent in Non-Content Matching (NCM) detection.
- T_{d_2}: Time spent in Content Matching (CM) detection.
- T_d: Time spent in detection – NCM and CM[2].
- Protocol type – TCP(http,telnet,$\cdots$), ICMP, UDP.
- Alert status – Alert type, if an alert is issued.

The data sets were collected during week days. Some relevant statistics are presented in Table 1. Sets 1 and 2 were collected on the same day. To eliminate outliers caused by measurement errors in data collection, we have repeated the same experiment twice for data set 1, and compared the resulting data records. Entries that were substantially different were deleted, and the same threshold used for deletion in the other data sets. As an example, in data set 1 we have recorded 73 packets with T_{d_2} above 1,000 μs. However, the corresponding records for the repeated experiment were substantially lower, indicating that these measurements were outliers. Note that μ_g, μ_p and μ_d are quite similar for the five data sets. Moreover, T_d represents about 75% of the overall processing time of each packet, which agrees with the study in [2]. Since T_d considerably dominates the overall service time, we focus our study on its characterization.

2.2 Bimodality and dependence on packet composition

Figure 1-(a) depicts the histogram of $\log_{10}(T_d + 1)$ for data set 1, which clearly indicates bimodality on the distribution of T_d. Figure 1-(b) depicts the histogram of P, which also displays bimodality. Similar results were verified for the other four data sets in reference to bimodality in T_d and P. To investigate the possible relationship between bimodality in P with bimodality in T_d, we plotted a scatter plot of $\log_{10}(T_d+1)$ vs. P in Figure 1-(c). It clearly shows a monotonic relationship, with large packets producing large processing times. In quantitative terms, the correlation coefficient between $\log_{10}(T_d + 1)$ and P is 0.766.

Set No.	Rate (Mbps)	Alerts (Total)	μ_g (μs)	μ_p (μs)	μ_d (μs)
1	58.6	8257	1.70	2.81	14.7
2	54.8	7446	1.64	2.74	13.6
3	101.1	7294	1.68	2.98	12.7
4	133.5	14167	1.70	3.32	13.1
5	118.8	14020	1.67	3.35	13.5

Table 1: Relevant statistics for the collected data sets. μ_g, μ_p and μ_d respectively denote the means of T_g, T_p and T_d.

Bimodality in P is a well known phenomenon in Internet traffic, as noted for example in [1] in which 127 million packets at NASA Ames Internet Exchange (AIX) were analyzed. Figure 1-(d) presents the Cumulative Distribution Function of P for data set 1, which displays great similarity with the same plot for the NASA Ames data set in [1]. it is clear from figure 1-(d) that between 30% to 40% of the packets have size close to 60 bytes, while the large remaining packets have sizes above 1,200 bytes. As noted in [1] this phenomenon can be explained by the current nature of Internet traffic. Approximately 85% of current traffic is TCP, and a large proportion is generated by bulk transfer applications such as HTTP or FTP. Consequently, the majority of the packets seen are one of two sizes: short, typically 60 or 66 byte packets which carry TCP acknowledgements but no payload and payload carrying packets with a maximum of 1500 byte packets. Given this observation, we define two types of packets for our study. Header-Only (HO) packets, containing only TCP/IP headers and no payload, and Header-and-Payload (HP) packets, containing header and payload. Clearly, all packets fall in either one of these two classes. Tables 2 and 3 present relevant statistics for the five data sets, when split into HO and HP (sub)-sets. Here, set 1HO represents the HO packets in data set 1, set 1HP represents the HP packets in data set 1, and so on. Below are some comments regarding the results in tables 2 and 3.

- In general terms, HO packets and HP packets constitute two very diverse populations. HP packets account for about 65% of the observed packets.
- T_d for HP packets are typically 5.5 times larger than T_d for HO packets. This is explained by the fact that HO packets are not subjected to CM processing.
- T_g for HP packets tend to be larger, a result of the packet size. T_p for HO packets are larger, an item whose explanation is under investigation.
- The bulk of the alerts – 60%-80% across the five datasets – occurs in HP packets, but the alert rate *per packet* is roughly 0.15% for both HP and HO packets, with wide variations across the five data sets.
- The results in table 3 are particularly significant. μ_{d_1} and σ_{d_1} vary very little across the five data sets, HO

[2]Clearly, $T_d = T_{d_1} + T_{d_2}$.

Set No.	Packets (Total)	Alerts (Total)	μ_g (μs)	μ_p (μs)	μ_d (μs)
1HO	1.78e6	3355	1.48	3.66	3.27
1HP	3.29e6	4902	1.82	2.35	19.5
2HO	1.64e6	3416	1.38	3.55	3.25
2HP	3.67e6	4030	1.77	2.31	17.9
3HO	3.01e6	2888	1.46	3.97	3.21
3HP	5.57e6	4406	1.80	2.45	16.5
4HO	4.33e6	2607	1.46	4.82	3.21
4HP	7.48e6	11560	1.84	2.45	17.4
5HO	3.86e6	4109	1.45	4.64	3.26
5HP	6.81e6	9911	1.79	2.62	17.9

Table 2: Relevant statistics for the data sets partitioned into HO and HP (sub)-sets.

Set No.	μ_{d_1} (μs)	σ_{d_1}	γ_{d_1}	μ_{d_2} (μs)	σ_{d_2}	γ_{d_2}
1HO	3.27	0.55	0.016	—	—	—
1HP	3.32	0.58	0.030	16.1	29.1	0.44
2HO	3.25	0.52	0.004	—	—	—
2HP	3.27	0.55	0.026	14.6	21.3	0.45
3HO	3.21	0.48	0.003	—	—	—
3HP	3.23	0.50	0.012	13.2	15.2	0.22
4HO	3.21	0.51	0.008	—	—	—
4HP	3.19	0.54	0.025	14.1	20.6	0.21
5HO	3.26	0.52	0.013	—	—	—
5HP	3.22	0.56	0.032	14.6	19.6	0.21

Table 3: Means (μ_{d_i}), standard deviations (σ_{d_i}), and correlation coefficients at lag 1 (γ_{d_i}) for NCM times and CM times for the HO and HP (sub)-sets.

and HP alike. This is not a surprise, as the headers of HP packets do not differ from the headers of HO packets. It is also very significant that γ_{d_1} is very small, indicating that T_{d_1} can be modeled as lightly correlated noise.

- Moreover, μ_{d_2} is about 4.5 times larger than μ_{d_1}, with much larger coefficients of variation: $\frac{\sigma_2}{\mu_{d_2}} \approx 1.5$, against $\frac{\sigma_1}{\mu_{d_1}} \approx 0.15$. This implies that T_{d_2} is by far the most important component of the detection stage, with T_{d_1} being lightly correlated noise with a small mean.

In summary, T_{d_2} is the main component determining the service times of network IDSs. For this reason, we will devote the remaining of the paper to its analysis.

2.3 Fitting a probability distribution to CM service times – Exploratory analysis

It is well known that the overall performance of a queuing system changes substantially if any of the processing steps have service times with a heavy-tailed nature (eg. [13], [16] and references therein). It is also well known that the presence of correlations in service times (and arrival times) affect the operation of queuing systems, introducing effects which cannot be accounted by modeling arrivals service times using independent processes [17] [22]. The results in table 3 suggest that both phenomena (heavy-tails and correlation) are present in T_{d_2}. Regarding heavy tails, coefficients of variation larger than one point to distributions with higher variance than the exponential. Regarding correlations, γ_{d_2} in the range 0.2 to 0.5 indicate that the correlations among service rates of HP packets cannot be ignored. To see this, recall that given a time series of $N \to \infty$ samples from a white noise process, the probability that the magnitude of the correlation coefficient at any lag will exceed $\frac{1.96}{\sqrt{N}}$ is 5% [3]. With $N \approx 4e6$, we have $\frac{1.96}{\sqrt{N}} \approx 0.001$, showing the extreme significance of the correlations in this case. In this paper we concentrate on the characterization of the *marginal* distribution of T_{d_2}, attempting to detect heavy-tail characteristics. Correlation effects will be investigated in detail in our future work.

There are several concepts and definitions of tailweight in the statistical literature – [14]. Here we present two definitions which have been extensively used in network traffic modeling – [21].

Definition 1 (Heavy-Tailedness in terms of asymptotics) The distribution for a random variable X is called heavy-tailed if

$$\mathbf{P}(X \geq x) \sim cx^{-\beta}, \text{ as } x \to \infty, \ \beta \geq 0.$$

By this it is meant that for some constants $\beta \geq 0$ and c, we have

$$\lim_{x \to \infty} \frac{\mathbf{P}(X \geq x)}{(cx^{-\beta})} = 1 \quad [17], [21] \square$$

The second definition is based on the concept of Conditional Mean Exceedance defined below.

Definition 2 (Conditional Mean Exceedance) The Conditional Mean Exceedance (*CME*) for a random variable X is defined as

$$CME(x) = \mathbf{E}(X - x | X \geq x),$$

where $\mathbf{E}(\cdot)$ denotes expected value – [14], Definition 3.4 $\square$

If x describes the waiting time of a certain process, $CME(x)$ has a simple interpretation. It is how much more time the process is expected to last, given that it has already lasted for x time units.

Definition 3 (Heavy-Tailedness in terms of *CME*) A distribution is called heavy-tailed if $CME(x)$ is an increasing function of x and light-tailed if $CME(x)$ is a decreasing function of x. If $CME(x)$ is constant, then the Cumulative Distribution Function (CDF) of X must be the Exponential distribution function – [14], Definition 3.5 $\square$

We now define three distributions which are widely used in traffic modeling: the Pareto distribution, the Exponential distribution and the Lognormal distribution.

Definition 4 (Pareto distribution – [11]) The Pareto distribution with location parameter a and shape parameter b has CDF $F(x)$ given by:

$$F(x) = \mathbf{P}(X \leq x) = 1 - \left(\frac{a}{x}\right)^b, \ a, b > 0, \ x \geq a \qquad (1)$$

with corresponding Probability Density Function (PDF) $f(x) = ba^b x^{-b-1}$ $\square$

Definition 5 (Exponential distribution – [11]) The Exponential distribution with location parameter α and shape parameter β has CDF $F(x)$ given by:

$$F(x) = \mathbf{P}(X \leq x) = 1 - \exp\left(-\frac{x-\alpha}{\beta}\right), \beta > 0,\ x \geq \alpha$$

with corresponding PDF $f(x) = \frac{1}{\beta}\exp\left(-\frac{x-\alpha}{\beta}\right)$ □

Definition 6 (Lognormal distribution – [11]) The Lognormal distribution with location parameter θ, shape parameter σ and scale parameter m has PDF $f(x)$ given by:

$$f(x) = \frac{1}{(x-\theta)\sigma\sqrt{2\pi}} \times \exp\left(-\frac{\left[\log\left(\frac{x-\theta}{m}\right)\right]^2}{2\sigma^2}\right)$$
$$m,\ \sigma > 0,\ x \geq \theta$$

The CDF does not have a closed form □

The Pareto distribution is heavy-tailed according to both Definition 1 and Definition 3[3]. The Exponential distribution is not heavy-tailed according to both definitions. The Lognormal distribution in not heavy-tailed according to Definition 1 [21], but does have a monotonically increasing *CME*, being therefore heavy-tailed according to Definition 3.

To determine if a heavy-tailed distribution (Pareto, Lognormal or others) is a good fit for a sample is subject of active research – eg. [7], [8], [9], [10], and references therein. A thorough description of the various approaches and difficulties is provided in [10]. We proceed on an exploratory manner, analyzing properties of the sample related with Definitions 1 and 3.

Figure 2-(a) presents the CCDF (Complementary CDF) plots for T_{d_2} in the five data sets. CCDF plots are obtained by graphing $\log_{10}[1-F(x)]$ vs. $\log_{10}(x)$. By taking logarithms in equation (1) we are left with

$$\log_{10}(1-F(x)) = -b\log_{10}(x) + b\log_{10}(a)$$

So, if a Pareto distribution with shape parameter b is a good fit for a sample $x_1, x_2, \cdots, x_N$, then CCDF plots should approximately fall on a straight line, with slope $-b$. Visual inspection suggests that linear fitting for CCDF can be attempted in the five data sets for processing times in the range 20 μs up to about 200 μs. For T_{d_2} larger than 200 μs, linear fit is no longer possible. Next, we consider the *CME* plots, presented in figure 2-(b). $CME(x)$ increases monotonically in the range 20 μs up to about 100 μs for all data sets, indicating that a Pareto or a Lognormal distribution may provide adequate fit. For $T_{d_2} > 100$ μs, the behavior of $CME(x)$ is hard to characterize. Given these two items (CCDF plots and *CME* plots), it is natural to define three distinct regions for the distribution of T_{d_2}. The ***bulk*** of the distribution, corresponding to values less or equal to 20 μs, the ***Clipped Upper Tail*** (CUT), corresponding to the region $T_{d_2} \in (20\ \mu\text{s}, 100\ \mu\text{s}]$, and the ***Extreme Upper Tail*** (EUT), where $T_{d_2} > 100$ μs. The first two lines of table 4 display the percentage of samples falling in the CUT and EUT regions for each of the five data sets. The general conclusion is that CUT roughly corresponds to the top 5% of the distribution, censoring the top 1%, which constitutes EUT. To test for heavy-tailedness in CUT and EUT, we compare the fits of Pareto, Exponential and Lognormal distributions to the samples in each of these regions. For CUT, the location parameter is equal to 20, and for EUT the location parameter is equal to 100. Maximum Likelihood Estimators (MLEs) were used to estimate the shape and scale parameters for the three distributions. Given a, the MLE for b in Pareto(a,b) is given by $\hat{b} = \frac{N}{\sum_{i=1}^{N}\log(\frac{x_i}{a})}$, the MLE for β in Exponential(a,β) is the sample mean of the a-shifted sample, i.e. $\{x_i - a\}$. Finally, MLEs for m and σ in Lognormal(a,m,σ) follow from the property that the logarithm of the $a-$shifted sample is normally distributed with mean m and standard deviation σ. To compare the samples with the distributions we utilized the χ^2 measure of discrepancy, defined below:

Definition 7 (χ^2 measure of discrepancy – [18], [20]) Suppose we have observed N samples of a random variable Y which we want to model using another model distribution Z. We partition Z into k bins. Each bin has a probability p_i associated with it, which is the proportion of the distribution Z falling into the ith bin. Let Y_i be the number of observations that actually fell into the ith bin. The χ^2 measure of discrepancy is defined as $\frac{X^2}{N}$, where X^2 is given by:

$$X^2 = \sum_{i=1}^{k}\frac{(Y_i - Np_i)^2}{Np_i}$$ □

The χ^2 measure of discrepancy is naturally suggested by the classical χ^2 test of hypothesis. The difference here is that one wants to compare the fit of several distributions to a sample. The last two lines of Table 4 show the results. As the χ^2 measure is sensitive to bin size [18], we repeated the test for two bin sizes, for both CUT and EUT. The results remain the same for both bin sizes. Lognormal fit is the best for CUT in all data sets, except one. Exponential fit is the best for EUT in all data sets. These results can be confirmed in figure 2-(c) which shows the CDFs for data set 1 in both the CUT and EUT regions, and the corresponding fits.

Size (%)	1HP	2HP	3HP	4HP	5HP
CUT	5.65	5.59	3.03	3.80	4.75
EUT	1.16	0.46	0.38	0.82	0.81
CUT (w=1, 10)	Log	Exp	Log	Log	Log
EUT (w=10, 100)	Exp	Exp	Exp	Exp	Exp

Table 4: Size of the CUT and EUT regions, and best fits for different bin sizes in the CUT and EUT regions using the chi-square metric.

3 Impact on IDS design and Conclusions

The main conclusions from the previous section are: (1) Processing times in `Snort` are dominated by the content matching stage, denoted by T_{d_2}; (2) in contrast to general Unix processes, which display strong heavy-tailedness [16], T_{d_2} has a Lognormal upper tail[4], clipped at the top 1%. This extreme 1% upper tail does not have heavy-tail characteristics.

[3] We note that for a Pareto distribution with $b > 1$, $CME(x) = \frac{x}{b-1}$.

[4] We note that the shape parameter for Unix processes in [16] falls in the range $1.05 < b < 1.25$. The shape parameter for Pareto fit of T_{d_2} in

Evidence for heavy tails in service times of Unix processes suggested the use of preemptive scheduling disciplines for such systems [13], [16]. The results in this paper indicate that service times in network IDSs are more regular than general processes in a distributed computing environment. Hence, preemptive disciplines are not necessary.

The long term objective of our research is to develop queuing models for network IDSs, to serve as the basis for load management and reconfiguration schemes [4], [15]. The development of appropriate models for the service times of IDSs is obviously an important element in this effort. The results in this paper represent the first step in the development of such models, which was the characterization of the marginal distribution of the service times. We stress however that the T_{d_2} time series is highly correlated. Queuing models derived on basis of the marginal sample distributions tend to underestimate the various metrics of operational performance, such as packet loss and mean service times [22]. Work is currently in progress to develop a model of service times incorporating the correlation effects. Such model could be used as part of a monitoring scheme, in which the model parameters are updated on line and used to compute on line estimates for the metrics of operational performance to be used in reconfiguration and load management. Work is in progress as well to develop such monitoring schemes.

Acknowledgement The research reported here was performed under contract DAAD17-03-C-0108 with the U.S. Army Research Laboratory. We are grateful to our Program Manager, Anthony Pressley, for his encouragement.

References

[1] Packet Length Distributions, 2002. Available at http://www.caida.org/analysis/AIX/plen_hist/ on August 2004.

[2] S. Antonatos, K. G. Anagnostakis, and E. P. Markatos. Generating Realistic Workloads for Network Intrusion Detection Systems. In *Proceedings of the 4th International Workshop on Software and Performance*, 2004.

[3] M. S. Bartlett. On the Theoretical Specification of Sampling Properties of Autocorrelated Time Series. *Journal of the Royal Statistical Society*, 27, 1946.

[4] J. B. D. Cabrera, W. Lee, R. K. Prasanth, L. Lewis, and R. K. Mehra. Optimization and Control Problems in Real Time Intrusion Detection. In *Proceedings of the 41st IEEE Conference on Decision and Control*, pages 1408–1413, Las Vegas, NV, December 2002.

[5] J. B. D. Cabrera and R. K. Mehra. Control and Estimation Methods in Information Assurance - A Tutorial in Intrusion Detection Systems. In *Proceedings of the 41st IEEE Conference on Decision and Control*, pages 1402–1407, December 2002.

[6] B. Caswell. *Snort 2.0 Intrusion Detection*. Syngress Publishing, Inc., 2003.

[7] W. S. Cleveland and D. X. Sun. Internet Traffic Data. In A. Raftery, M. A. Tanner, and M. T. Wells, editors, *Statistics in the 21st Century*, pages 214–228. Chapman & Hall/CRC, 2002.

[8] M. E. Crovella and M. S. Taqqu. Estimating the Heavy Tail Index from Scaling Properties. *Methodology and Computing in Applied Probability*, 1, 1999.

the CUT region for the five data sets studied in this paper fall in the range $1.8 < b < 2.0$.

[9] A. B. Downey. Evidence for long-tailed distributions in the Internet. In *Proceedings of the ACM SIGCOMM Internet Measurement Workshop*, November 2001.

[10] A. B. Downey. Lognormal and Pareto Distributions in the Internet, 2003. Submitted to *Computer Communications*, available at http://allendowney.com/research/longtail/ on August 2004.

[11] M. Evans, N. Hastings, and B. Peacock. *Statistical Distributions*. John Wiley and Sons, Inc., New York, Third edition, 2000.

[12] M. Hall and K. Wiley. Capacity verification for high speed network intrusion detection systems. In *Proceedings of the 5th International Symposium on Recent Advances in Intrusion Detection*, October 2002.

[13] M. Harchol-Balter and A. B. Downey. Exploiting process lifetime distributions for dynamic load balancing. *ACM Transactions on Computer Systems*, 15(3):253–285, 1997.

[14] T. P. Hettmansperger and M. A. Keenan. Tailweight, Statistical Inference and Families of Distributions - A Brief Survey. In G. P. Patil et al., editors, *Statistical Distributions in Scientific Work*. D. Reidel Publishing Company, Dordrecht, Holland, 1980.

[15] W. Lee, J. B. D. Cabrera, A. Thomas, N. Balwalli, S. Saluja, and Y. Zhang. Performance Adaptation in Real-Time Intrusion Detection Systems. In *Proceedings of the 5th International Symposium on Recent Advances in Intrusion Detection*, October 2002.

[16] W. E. Leland and T. J. Ott. Load-balancing heuristics and process behavior. In *Proceedings of ACM SIGMETRICS*, pages 54–69, May 1986.

[17] W. E. Leland, M. S. Taqqu, W. W. Willinger, and D. V. Wilson. On the Self-Similar Nature of Ethernet Traffic (Extended Version). *IEEE/ACM Transactions on Networking*, 2:1–15, 1994.

[18] D. S. Moore. Measures of lack of fit from tests of chi-squared type. *Journal of Statistical Planning and Inference*, 10:151–166, 1984.

[19] S. Northcutt. *Network Intrusion Detection - An Analyst's Handbook*. New Riders Publishing, 1999.

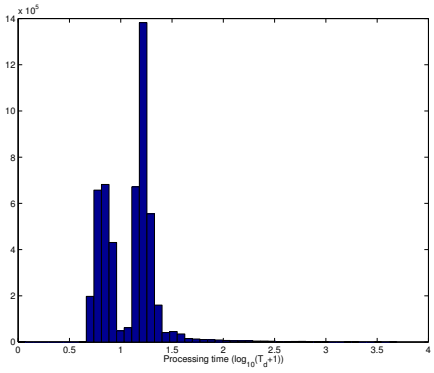
[20] V. Paxson. Empirically-derived analytic models of wide-area TCP connections. *IEEE/ACM Transactions on Networking*, 2(4):316–336, August 1994.

[21] V. Paxson and S. Floyd. Wide-area Traffic: The Failure of Poisson Modeling. *IEEE/ACM Transactions on Networking*, 3(3):226–244, June 1995.

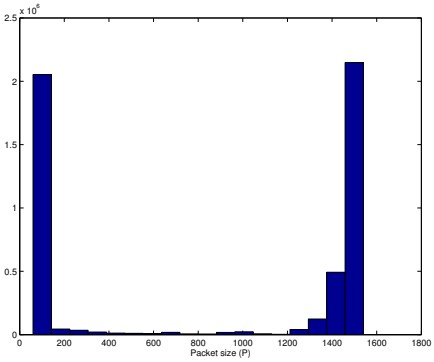
[22] E. A. Peköz and M. Lapré. Inequalities for Queues with a Learning Server. *Queueing Systems*, 37(4):337–347, March 2001.

[23] L. Schaelicke, T. Slabach, B. Moore, and C. Freeland. Characterizing the performance of intrusion detection sensors. In *Proceedings of the 6th International Symposium on Recent Advances in Intrusion Detection*, September 2003.

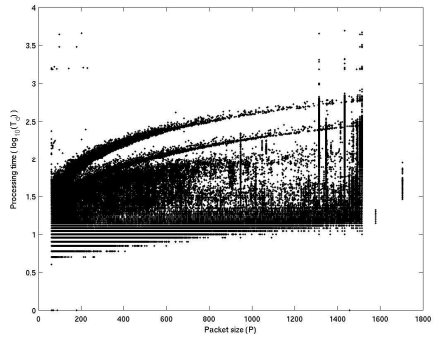
[24] N. Tuck, T. Sherwood, B. Calder, and G. Varghese. Deterministic Memory-Efficient String Matching Algorithms for Intrusion Detection. In *Proceedings of IEEE INFOCOM*, Hong Kong, March 2004.



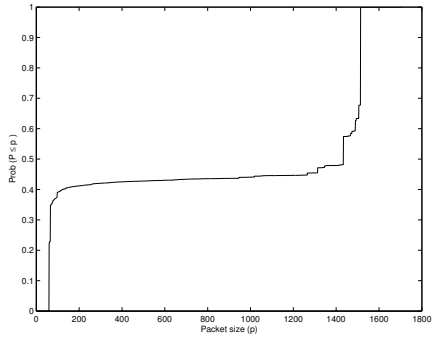
(a) Histogram of $\log_{10}(T_d + 1)$ (μs).



(b) Histogram of P (bytes).

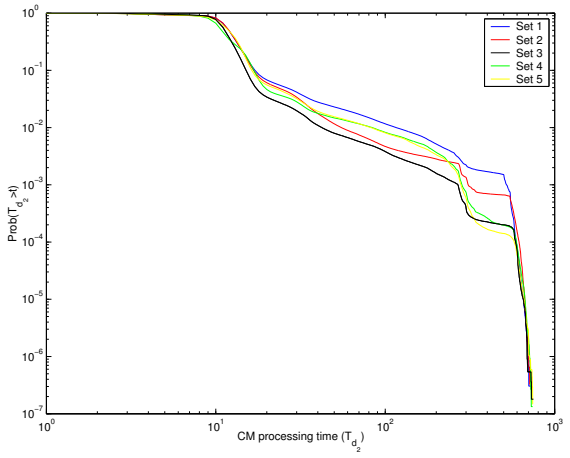


(c) $\log_{10}(T_d + 1)$ vs. P.

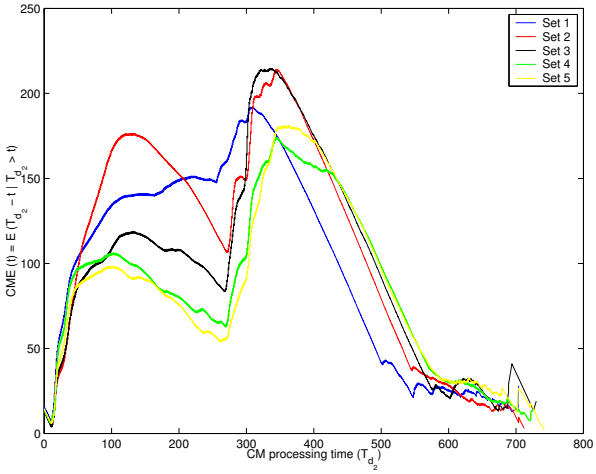


(d) CDF of P (bytes).

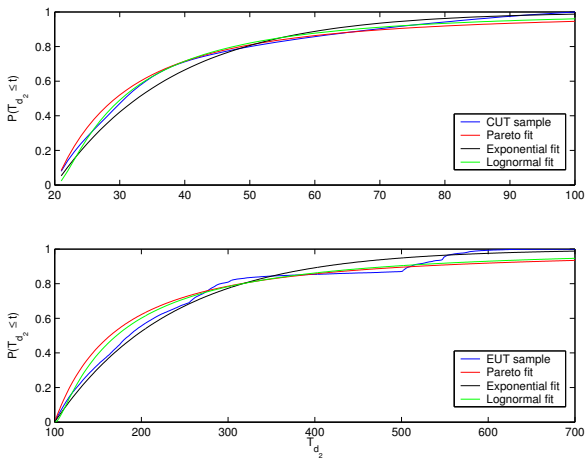
Figure 1: Bimodality of processing times and packet sizes.



(a) CCDF plots for the five data sets.



(b) *CME* plots for the five data sets.



(c) CDF plots: Pareto, Exponential and Lognormal fits in the CUT (top plot – Lognormal is best) and EUT (bottom plot – Exponential is best) regions for data set 1.

Figure 2: Investigating heavy tails in CM processing time.