

A Hough-transform-based Anomaly Detector with an Adaptive Time Interval

Romain Fontugne
The Graduate University for Advanced Studies
Tokyo, Japan

Kensuke Fukuda
National Institute of Informatics /
PRESTO JST
Tokyo, Japan

ABSTRACT

Internet traffic anomalies are a serious problem that compromise the availability of optimal network resources. Numerous anomaly detectors have recently been proposed, but maintaining their parameters optimally tuned is a difficult task that discredits their effectiveness for daily usage. This article proposes a new anomaly detection method based on pattern recognition and investigates the relationship between its parameter set and the traffic characteristics. This analysis highlights that constantly achieving a high detection rate requires continuous adjustments to the parameters according to the traffic fluctuations. Therefore, an adaptive time interval mechanism is proposed to enhance the robustness of the detection method to traffic variations. This adaptive anomaly detection method is evaluated by comparing it to three other anomaly detectors using four years of real backbone traffic. The evaluation reveals that the proposed adaptive detection method outperforms the other methods in terms of the true positive and false positive rate.

Categories and Subject Descriptors

C.2.3 [Computer-Communication Networks]: Network Operations—*Network monitoring*

General Terms

Measurement, Security, Performance

Keywords

Internet traffic, Anomaly detection, Pattern recognition

1. INTRODUCTION

The success of Internet services results in a constant network traffic growth along with an increasing number of anomalies such as remote attacks and misconfigurations. These anomalies represent a large fraction of the Internet traffic that is unwanted and penalizes legitimate users from accessing optimal network resources. Therefore, detecting and

diagnosing these threats are crucial tasks for network operators that are trying to maintain the Internet resources made available. Intensive studies have been carried out in this field, but the proposed anomaly detection methods still have important drawbacks [11, 8]. Indeed, the sensitivity of these methods to parameter tuning and traffic variations are still open issues. Therefore, in practice, selecting the optimal parameters is not intuitive as the relationship between the parameters and traffic characteristics is challenging.

Only a few works have investigated these important drawbacks currently discrediting anomaly detectors. A careful study of the approaches based on principal component analysis (PCA) was carried out by Ringberg et al. [11], and they identified four main challenges including the sensitivity to analyzed traffic and parameter tuning. In addition, an attempt to automatically tune a method based on gamma modeling and sketches was conducted by Himura et al. [8]. They designed a learning process for predicting the optimal parameters regarding the best parameters for past data. However, this method suffers from a high error rate as unexpected events do appear.

Recently, a pattern recognition based method has been proposed [6, 7]. The main idea of this detection method is to monitor the traffic in 2D pictures where anomalies appear as “lines”, which are easily identifiable using a pattern recognition technique called the Hough transform [5]. One advantage of this method is that its simple principles allow us to intuitively select a suitable parameter set. The optimal values of the parameters, however, fluctuate along with the traffic throughput variations and require continuous adjustments, making it unpractical for real usage. In order to provide a detector that is easily tunable and robust to traffic variations, this article follows a similar approach to [6], but it uses fundamentally different 2D pictures that allow for better highlighting anomalies. Moreover, the main contribution of this work is to obtain a complete understanding of the proposed method parameter set and provide a mechanism that automatically tunes it based on the traffic variations. The advantages of this adaptive method are demonstrated by comparing its results to those obtained using fixed parameter tunings and those of three other anomaly detectors using four years of real Internet traffic. The results highlight the superiority of the proposed method in terms of the true positive and false positive rates, emphasizing that automatically adjusting the parameter set in regards to the traffic fluctuations is crucial for continuously performing an accurate level of detection.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SAC'11, 21-MAR-2011, TaiChung, Taiwan

Copyright 2010 ACM 978-1-4503-0113-8/11/03 ...\$10.00.

Table 1: Different kinds of common anomalies and their particular traffic feature distributions.

Anomaly	Traffic feature distribution
Port scan	Traffic distributed in destination port space and concentrated on single destination host.
Network scan, Worm, Exploit	Traffic distributed in destination address space and concentrated on limited number of destination ports.
DDoS, Netbot, Flash crowd	Traffic distributed in source address space and concentrated on limited number of destination addresses.

2. ABNORMAL DISTRIBUTION OF TRAFFIC FEATURES

Recent works have identified anomalous traffic as alterations in the distributions of the traffic features [10, 6, 2, 13, 4]. For example, Table 1 lists several kinds of anomalies commonly identified in Internet traffic. Each kind of anomaly inherently affects the distribution of two traffic features. Similarly, in this article an anomaly refers to a set of flows altering the distribution of at least one of the four following traffic features: the source IP address, destination IP address, source port, and destination port. However, the proposed approach for observing these alterations in the traffic feature distributions is substantially different from that in other works. Previously, anomalies have been mainly detected by identifying the outliers in the aggregated traffic using different formalisms — e.g., signals [10], histograms [4, 2], or matrices [12] — whereas, the proposed method identifies particular patterns in pictures. The analyzed pictures are two-dimensional scatter plots, where each axis represents a traffic feature, each plot stands for traffic flows, and the particular traffic feature distributions of the anomalies are easily identifiable as lines.

Figure 1 shows two examples of the pictures analyzed in this article. Figure 1a displays traffic with regards to its destination port and destination address. This graphical representation of the traffic makes it easy to discriminate the port scan, network scan, worm, and exploit from the benign traffic as they appear as lines in the picture (Fig. 1a). Figure 1b, however, displays the traffic in regards to its destination and source addresses, and permits other kinds of anomaly to be observed. For instance, distributed denial of service (DDoS), flash crowd and botnet activities appear as horizontal lines in this scatter plot.

The three main advantages of this approach over the previous works are [11]: (1) the anomalous flows are inherently pinpointed in the scatter plots whereas the anomalous flows detected in a signal are difficult to identify. (2) The proposed approach is able to monitor the pattern of a large-scale anomaly whereas the methods detecting anomalous traffic as outliers fail if a majority of the traffic is contaminated. (3) In regards to the traffic features monitored by the pictures and the direction of the identified line, one can easily deduce the kind of observed anomaly.

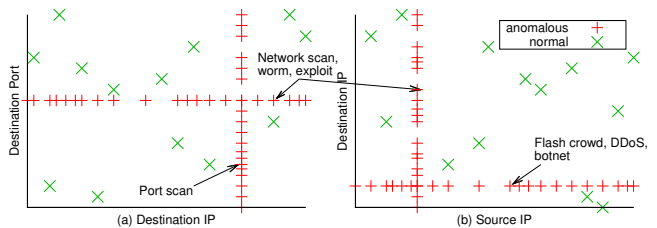


Figure 1: Example of two pictures highlighting anomalous traffic as lines.

3. ANOMALY DETECTION METHOD

The anomaly detection method proposed in this article consists of five main steps: (1) The traffic of the current time interval is mapped onto five different pictures. (2) The Hough transform is computed on each picture to uncover the plot distributions. (3) Abnormal plot distributions are detected in the Hough spaces. (4) Traffic information corresponding to the anomalous plots are retrieved and reported. (5) The time interval is shifted and step 1 is repeated.

3.1 Pictures computation

The proposed approach takes advantage of several kinds of pictures to monitor the different aspects of the traffic and highlight the different kinds of anomalies. The analyzed pictures are 2-D scatter plots designed from four traffic features: {source IP address, destination IP address, source port, destination port}. For the remainder of this paper the term *traffic features* will refer to only these four traffic features. The five picture categories correspond to all the possible pairs of traffic features containing IP address. Namely, the x and y axis of the picture, respectively, correspond to the following pairs of features:

- Source IP address, destination IP address
- Source IP address, source port
- Source IP address, destination port
- Destination IP address, source port
- Destination IP address, destination port

A flow in the analyzed pictures is represented by a plot that is located using the two following mechanisms. (1) The port space is shrunk to the size of the pictures: Lets assume a 1000-pixel picture ($ySize = 1000$) that has a y axis standing for the source port, then a http flow, i.e., $SrcPort = 80$, is plotted at $y = \lfloor SrcPort * ySize / 2^{16} \rfloor = \lfloor 80 * 1000 / 65535 \rfloor = 1$, and each pixel of the picture represents approximately $\lceil 65535 / 1000 \rceil = 65$ distinct port numbers. (2) The IP address space is at first hashed by ignoring the first h bits of the addresses and then shrunk to the size of the picture. For example, supposing $h = 16$ and a 1000 pixel wide picture ($xSize = 1000$) with an x axis as the source IP, then a flow from the source IP 192.168.10.10 is plotted at $x = \lfloor (SrcIP \bmod 2^{32-h}) * xSize / 2^{32-h} \rfloor = \lfloor (192.168.10.10 \bmod 2^{16}) * 1000 / 2^{16} \rfloor = \lfloor (0.0.10.10) * 1000 / 2^{16} \rfloor = 39$. Notice that this article only deals with square pictures, meaning that the $xSize = ySize$.

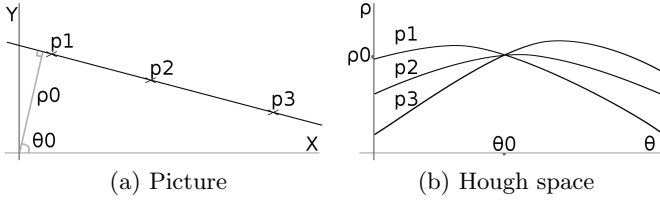


Figure 2: Principles of Hough transform.

3.2 Hough transform

A well-known image processing technique called the Hough transform [5, 7] helps us in extracting the relevant information from computed pictures. The Hough transform is commonly used to detect the parametric structures (e.g., line, circle, or ellipse) in pictures and has the advantage of being robust to noise and able to detect incomplete shapes.

The basic usage of the Hough transform allows for the identification of lines in a picture. It consists of a voting procedure, where each plot of the picture votes for the lines it belongs to. Formally, each plot in the picture $p = (x_p, y_p)$ votes for all the θ and ρ that satisfy $\rho = x_p \cdot \cos(\theta) + y_p \cdot \sin(\theta)$ (line equation in polar coordinates). All the votes are stored in a two-dimensional array, called the Hough space, in which one dimension stands for θ and one for ρ . Figure 2 depicts an example of the Hough transform. The analyzed picture (Fig.2a) contains three plots, and the votes for each plot are represented by a curve in the Hough space (Fig.2b). The maximum number of votes in the Hough space is obviously at the intersection of the three curves $I = (\theta_0, \rho_0)$, identifying the line passing through the 3 plots, $\rho_0 = x \cdot \cos(\theta_0) + y \cdot \sin(\theta_0)$.

In order to find the local maxima in the Hough space, thus the prominent lines in the picture, a robust peak detection based on the standard deviation σ of the Hough space is implemented. Therefore, all flows corresponding to the elements of the Hough space that are higher than 3σ are reported as anomalous.

3.3 Complexity

The computational complexity of the proposed method is mainly one of the Hough transforms that is linear to the number of plots in picture. In a worst case scenario, each plot represents a single flow so the number of plots in the pictures is equal to the total number of flows N . Let $f = 5$ be the number of picture categories, t the traffic duration divided by the time interval, and $n_{i,j,k}$ the number of plots in the picture k of category i at the time interval j . The cost of the proposed algorithm in the worst case is linear to N :

$$\sum_{i=1}^f \sum_{j=1}^t O(n_{i,j}) = \sum_{i=1}^5 O(N) = O(N)$$

In our experiments, the proposed method takes about one minute to analyze a 15-minute traffic trace from the MAWI archive.

4. DATA AND PROCESSING

All the experiments conducted in this work are based on the traffic traces publicly available in the MAWI archive

Table 2: Heuristics deduced from main anomalies previously reported [1, 6] and manual inspection of data-set considered in this article.

Category	Label	Details
Attack	Sasser	Traffic on ports 1023/tcp, 5554/tcp or 9898/tcp
Attack	RPC	Traffic on port 135/tcp
Attack	Ping	High ICMP traffic
Attack	Other attacks	Traffic with more than 50% of SYN, RST or FIN flag. And http, ftp, ssh, or dns traffic with more than 30% of flag SYN
Attack	NetBIOS	Traffic on ports 137/udp or 139/tcp
Special	Http	Traffic on ports 80/tcp and 8080/tcp with less than 30% of SYN flag
Special	dns, ftp, ssh	Traffic on ports 20/tcp, 21/tcp, 22/tcp or 53/tcp&udp with less than 30% of SYN flag
Special	Unknown	Traffic that does not match other heuristics

[3]. This database provides daily backbone traffic traces that contain 15-minutes of traffic taken from a trans-Pacific link between Japan and the U.S. This article particularly focuses on two data sets from the MAWI Samplepoint-B; (1) the first week of August 2004 was particularly affected by the Sasser worm [1, 6] and provides valuable support for illustrating the benefits of the proposed method. (2) All the traffic recorded from 2003 to 2006 allowed us to evaluate the global performance of the proposed method by comparing its results to the ones of other anomaly detectors.

Due to the lack of ground truth data for backbone traffic, the evaluation of the proposed detector relies on heuristics that is fundamentally independent from the principle of the proposed method (Table 2). Indeed, these heuristics is based on well-known port numbers and abnormal usages of TCP flags [1, 6], whereas the proposed method uses only the port numbers as indexes and does not rely on the application information related to them nor the TCP flags. Heuristics classifies traffic into two categories, *attack* and *special*, and helps in quantifying the effectiveness of the detection method.

An anomaly detector is expected to report more traffic classified as attacks than those labeled special. Thus, the *accuracy* of a detector is defined as the ratio of the alarms classified as attacks by the heuristics listed in Table 2.

5. PARAMETER TUNING AND DRAWBACKS

5.1 Experimental parameter tuning

The following experiments aim at finding the optimal parameter tuning of the proposed method using one week of traffic affected by the Sasser worm (Section 4). Furthermore, these experiments uncover the correlation between the two main parameters, i.e., the size of picture and the time interval, and show that the performances of the proposed method are not affected by any variance in the h value as long as the number of possible indexes is higher than the picture size, $2^{32-h} > xSize$.

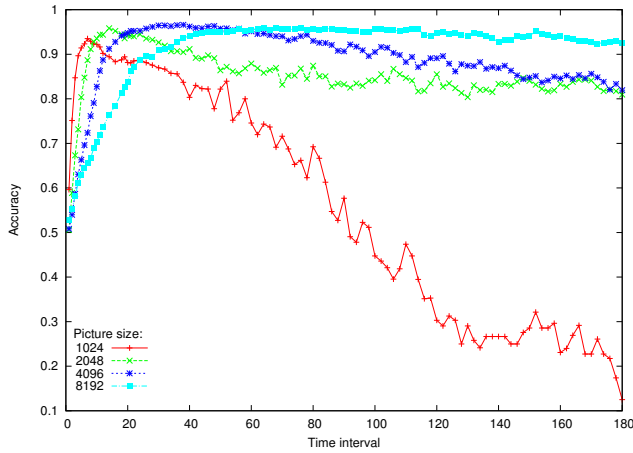
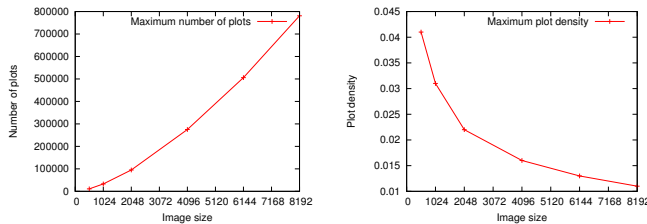


Figure 3: Accuracy of proposed method using four different picture sizes.



(a) Maximum acceptable number of plots (b) Maximum acceptable plot density

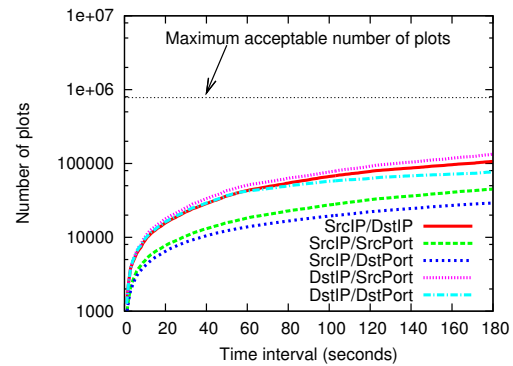
Figure 4: Evaluation of maximum acceptable number of plots to perform the Hough transform. The plot density is the maximum acceptable number of plots over the picture area.

Figure 3 depicts the average accuracy of the detection method using numerous parameter values. It highlights that the proposed method is able to achieve an accuracy that is higher than 0.9 for any time interval $> 4s$ and a suitable picture size. Furthermore, Fig. 3 indicates that the optimal picture size is proportional to the size of the time interval. For instance, if the time interval is less than 8s the best performance is obtained with a picture size set to 1024, whereas the time interval ranges (9, 16) are suitable for a picture size equal to 2048, and so forth. Intuitively, a larger time interval involves a greater number of plots in the pictures; thus, to avoid meaningless saturated pictures, the optimal size of a picture increases along with the size of the time interval.

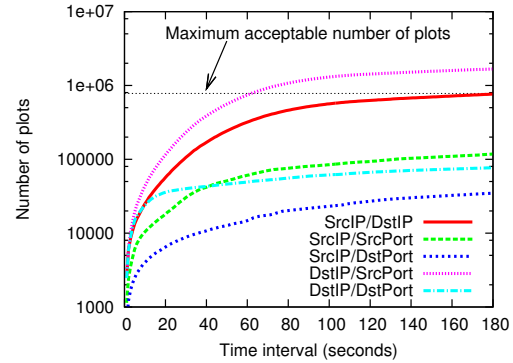
Although the specific values given here are suitable for the analyzed traffic, different values might be more effective for traffic having different properties. Obviously, traffic with the same properties but a higher throughput displays more plots in the pictures, and thus in this case, smaller time intervals are required to maintain an acceptable number of plots in the pictures.

5.2 Evaluation of optimal parameter

The time interval is the parameter that controls the amount of traffic displayed in the pictures. Thus, as the proposed method inherently translates the traffic flows to the plots in the pictures, the time interval allows us to select the quan-



(a) Traffic with few Sasser activities



(b) Traffic with many Sasser activities

Figure 5: Plot growth for different picture categories ($xSize = 8192$).

tity of plots appearing in the pictures. The challenge in setting the time interval is the trade-off between displaying enough plots to have relevant pictures and limiting the surrounding noise representing the legitimate traffic and hiding anomalies.

The sensitivity of the implemented Hough transform to the number of plots in the pictures is analyzed using synthetic pictures that have a random line and various amounts of uniformly distributed noise. The algorithm was performed 100 times on different pictures with the same level of noise. If the 100 tests are successful then the noise is increased and the algorithm is again performed. The highest noise level for which all 100 executions of the algorithm succeed defines the maximum acceptable number of plots in a picture. This experiment was conducted using six different picture sizes, as indicated in Fig. 4a. As expected, the maximum acceptable number of plots in the pictures increases with the picture size. Figure 4a shows that the maximum acceptable number of plots for picture sizes of 1024, 2048, 4096, and 8192 are respectively 33000, 95000, 275000, and 781000. Figure 4b shows that this increase is not linear to the area of the picture and the common upper bound for all the considered picture sizes is approximately 1% of the picture area.

5.3 Dispersion of plots in pictures

The previous section provided an insight on how to select the suitable time interval for a particular picture, but the proposed method analyzes five different pictures at the same time. A crucial task is to understand the divergence

between the different kinds of pictures. Since the five picture categories monitor distinct feature spaces, plots corresponding to the same traffic are differently dispersed in all the pictures. Therefore, the traffic is usually depicted by using a different number of plots for two pictures from different categories. For example, Fig. 5a shows the number of plots for the five kinds of pictures for several time interval sizes. This figure highlights that the number of plots appearing in each picture category increases at different rates. A slow increase in the number of plots means that many flows share the same instance in the monitored feature spaces, whereas a rapid growth highlights the flows spreading into the observed feature spaces. The rate of increase of the plots for each picture category is strongly related to the throughput and the dispersion of the traffic in the feature space.

Since anomalies alter the traffic feature distribution, they also significantly affect the increase in the number of plots. Figure 5b is a typical example where the increase in plots for certain picture categories is rapidly increasing due to anomalous traffic. Indeed, the traffic analyzed in Fig. 5b contains an outbreak of the Sasser worm highlighting a considerable increase in the number of plots for two picture categories monitoring the destination address. This observation is in accord with the behavior of the Sasser worm manually observed in the traffic trace, that is, the worm tries to infect numerous remote hosts to spread throughout the network.

Despite their differences, the two traffic analyzed in Fig. 5 are taken from the same traffic trace (Fig. 5b representing the first three minutes of the traffic trace, whereas Fig. 5a is the traffic recorded three minutes later), illustrating two drawbacks of the proposed method. (1) For the same traffic, the number of plots in all the picture categories is significantly different. Thus, the suitable time interval for a picture from a certain category does not necessarily suit the pictures from the other categories. (2) The increase in plots for a certain picture category sharply varies especially when anomalous traffic appears. Thus, the suitable time interval for a single picture category fluctuates over time.

6. ADAPTIVE TIME INTERVAL

Here, an improved version of the anomaly detection method is proposed to overcome the drawback identified in the previous section. This new version assigns different time intervals to all the picture categories and adapts these time intervals to the traffic variation. Therefore, the value of the time intervals is no longer a fixed value taken as an input, but it is automatically computed by taking into account the throughput and the traffic distribution in the traffic feature spaces.

The proposed improvement consists of controlling the amount of monitored traffic based on the quantity of plots in the picture instead of the time interval. The Hough transform is performed only if a certain number of plots p are displayed in the picture (regardless of the time interval corresponding to the traffic mapped into the picture), and other pictures keep monitoring the traffic until they display a sufficient number of plots, p . Therefore, all the pictures stand for different time intervals and the Hough transform is performed at different instants of time for each picture. The first two steps of the algorithm proposed in Section 3 are replaced by: (1) Map traffic to pictures until a picture displays p plots. (2) Compute the Hough transform for pictures with p plots. In addition, the time interval parameter is replaced by p ,

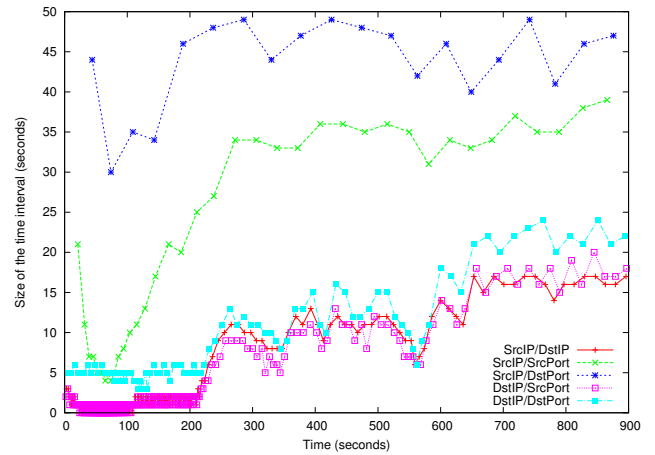


Figure 6: Evolution of time interval corresponding to pictures computed during 15 minutes of traffic.

which is the number of plots required to perform the Hough transform. The value of p is directly deduced from the picture size to assure the success of the Hough transform. The upper bound for p is 1% of the picture area (Section 5.2), and the lower values help in quickly reporting the anomalies since the Hough transform is performed earlier. However, too small p values result in irrelevant pictures as the sample traffic displayed in pictures is insignificant. In the following experiments, p is arbitrarily set to 0.5% of the picture area, $p = 0.05 \cdot xSize^2$. Hereafter, this new version of the detection method is referred to as the adaptive method.

6.1 Performance improvement

The benefit of the adaptive method is evaluated by using one week of traffic (Section 4). For clarity reasons and because all the traffic traces reach a similar conclusion, the following focuses only on the first day of the analyzed traffic.

6.1.1 Robustness to traffic variation

Figure 6 displays the time intervals corresponding to all the pictures computed during the analysis of the 15 minutes of traffic. The first four minutes of this traffic are significantly affected by the Sasser worm resulting in a higher throughput and an increase in the number of destination addresses. Nevertheless, the method successfully handled the traffic variation, that is, the time intervals represented by the pictures monitoring the destination address remain from 1 to 5 seconds during the Sasser outbreak (Fig. 6). However, the same quantities range from 14 to 25 seconds during the last four minutes of traffic, where the traffic is much less polluted by the Sasser worm. This example illustrates the benefit of the adaptive method since selecting a fixed value for the time interval of the basic method is challenging.

6.1.2 Accuracy gain

The only parameter of the adaptive method is the picture size, and by setting it to three different values, namely 1024, 2048, and 4096, the same high accuracy score is observed, 0.99, 0.98, and 0.99, respectively. However, the number of reported alarms decreases as the picture size increases, which is 373, 173, and 117 events respectively. Thus, for the fol-

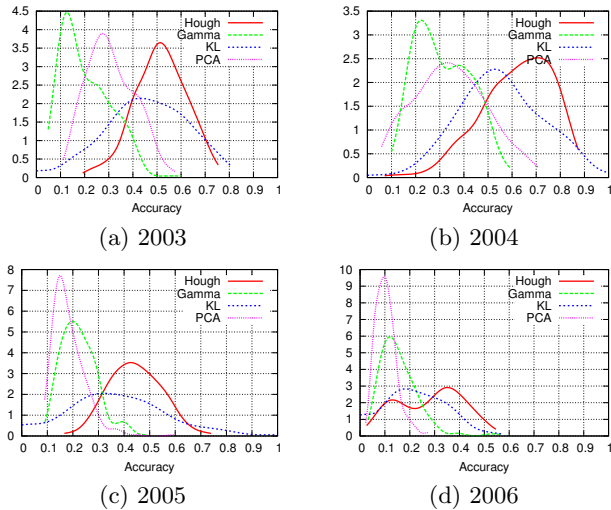


Figure 7: PDF of accuracy of four detectors for four years of MAWI traffic.

lowing experiments the picture size is set to 1024 in order to report as much anomalous traffic as possible.

The comparison between the two versions of the method emphasizes the better false positive and true positive rates of the adaptive method. Namely, it identifies 369 source addresses infected by Sasser (i.e. 86% of the Sasser traffic manually identified). However, the basic method, with identical parameters but a fixed time interval of 10 seconds, identifies only 258 source addresses related to Sasser (i.e. 60% of the Sasser traffic manually identified). The basic version of the method is able to identify the same amount of Sasser traffic only if the time interval is set to one second, however, in this case 229 http traffics were also reported and a manual inspection revealed that they are benign traffic regarded as false positive alarms.

7. EVALUATION

The adaptive detection method is evaluated by analyzing four years of MAWI traffic (i.e. 2003, 2004, 2005, and 2006) and comparing its results to the outputs of three other anomaly detectors based on different theoretical backgrounds, which are: (1) the well-known PCA-based detector [10] (in this work the implementation of this detector relies on sketches to analyze traffic taken from a single link [9]), (2) the detection method based on multi-scale gamma modeling and sketches [4], and (3) the detector based on the Kullback-Leibler (KL) divergence and association rule mining [2]. The picture size parameter of the adaptive method is set to 1024, and the Gamma-based method is optimally tuned using the method proposed by Himura et al. [8], whereas, the parameters of the PCA and the KL methods are set with fixed and arbitrary values that are globally suitable for the analyzed MAWI traffic.

The four detectors aim at finding any kinds of traffic anomaly by inspecting only IP header. However, they aggregate traffic using different formalisms, i.e., the proposed method monitor the traffic using pictures whereas the PCA-based one analyzes time series and the gamma and KL detectors take advantage of histograms.

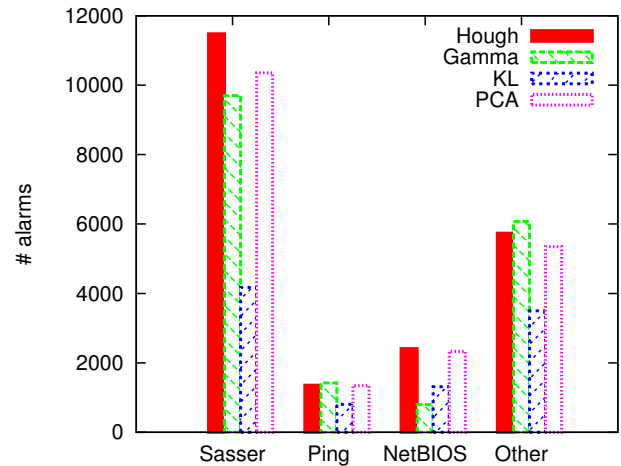


Figure 8: Breakdown of alarms reported by four detectors and classified as attacks during four years of MAWI traffic.

Figure 7 shows the accuracy achieved by the four detectors for each year of analyzed traffic. The average accuracy of the proposed method is higher than the one of the three other detectors during the four years of MAWI traffic. Among the three other detection methods the KL-based one is the best detector in terms of accuracy, moreover, it occasionally outperforms the method proposed in this article (Fig. 7b and Fig. 7c).

The circumstances in which the KL-based detector remarkably outperforms the other detectors were thoroughly inspected and this highlighted the fact that this detector reports a high ratio of attacks but out of only a small number of alarms. Consequently, the KL-based detector achieves a high attack ratio along with a high false negative rate (i.e. missed anomalies). Figure 8 shows the quantity of attacks reported by each detector classified with the labels from Table 2 (RPC is omitted as only 11 alarms of this kind were identified in the four years of traffic) and emphasizes the large amount of anomalies missed by the KL-based one.

The PCA and Gamma-based detectors, however, report the same quantity of attacks as the proposed method along with numerous alarms classified as special (Fig. 7). Although the proposed method is more sensitive to Sasser and attacks towards NetBIOS services, the Gamma-based method detected slightly more ping flood traffic (66 alarms) and traffic labeled as other attacks (337 alarms) for the four years of analyzed traffic. Nevertheless, the PCA and Gamma-based detectors were considerably worse than the adaptive method in terms of accuracy, and this drawback is due to the quantity of traffic classified as special that was reported by these two detectors (i.e. high false positive rate).

The advantage of the adaptive method is to consistently adapt its time interval over the four years of analyzed traffic, and therefore, it constantly detects a large quantity of anomalous traffic while the number of reported benign traffic is low.

8. CONCLUSIONS

This article proposed a new anomaly detection method that takes advantage of image processing techniques to iden-

tify the flows with abnormal traffic feature distributions. Crucial challenges rarely addressed in the appropriate literature were uncovered by investigating the major drawbacks of this method; the sensitivity of anomaly detectors to traffic variations and the role of the time scale in anomaly detection. Addressing these two issues resulted in a significant improvement for the proposed detection method that overcomes any adverse conditions as it analyzes traffic within a time interval that is automatically adapted to the traffic throughput and the distribution of traffic features.

The evaluation of this adaptive method using real Internet traffic highlighted its ability to maintain a high detection rate while the traffic was significantly altered by anomalies. Therefore, these experiments indicated that the adaptive time interval enabled 26% more worm traffic to be detected, and decreased the false positive rate. The adaptive detection method proposed in this paper is also validated by comparing it with three other detection methods and using four years of real backbone traffic. The results highlighted that the proposed adaptive method allows for the detection of almost all the anomalies reported by the other detectors while it achieves the lowest false positive rate.

Acknowledgments

The authors are thankful to Yosuke Himura, Yoshiki Kanda, Patrice Abry, Pierre Borgnat, and Guillaume Dewaele for their valuable comments.

References

- [1] P. Borgnat, G. Dewaele, K. Fukuda, P. Abry, and K. Cho. Seven years and one day: Sketching the evolution of internet traffic. *INFOCOM '09*, pages 711–719, 2009.
- [2] D. Brauckhoff, X. Dimitropoulos, A. Wagner, and K. Salamatian. Anomaly extraction in backbone networks using association rules. *IMC '09*, pages 28–34, 2009.
- [3] K. Cho, K. Mitsuya, and A. Kato. Traffic data repository at the WIDE project. In *USENIX 2000 Annual Technical Conference: FREENIX Track*, pages 263–270, 2000.
- [4] G. Dewaele, K. Fukuda, P. Borgnat, P. Abry, and K. Cho. Extracting hidden anomalies using sketch and non gaussian multiresolution statistical detection procedures. *SIGCOMM LSAD '07*, pages 145–152, 2007.
- [5] R. O. Duda and P. E. Hart. Use of the hough transformation to detect lines and curves in pictures. *Commun. ACM*, 15(1):11–15, 1972.
- [6] R. Fontugne, Y. Himura, and K. Fukuda. Evaluation of anomaly detection method based on pattern recognition. *IEICE Trans. on Commun.*, E93-B(2):328–335, February 2010.
- [7] K. Fukuda and R. Fontugne. Estimating speed of scanning activities with a hough transform. *ICC '10*, page 5, 2010.
- [8] Y. Himura, K. Fukuda, K. Cho, and H. Esaki. An automatic and dynamic parameter tuning of a statistics-based anomaly detection algorithm. *ICC '09*, page 6, 2009.
- [9] Y. Kanda, K. Fukuda, and T. Sugawara. An evaluation of anomaly detection based on sketch and pca. *GLOBECOM '10*, 2010.
- [10] A. Lakhina, M. Crovella, and C. Diot. Mining anomalies using traffic feature distributions. *SIGCOMM '05*, pages 217–228, 2005.
- [11] H. Ringberg, A. Soule, J. Rexford, and C. Diot. Sensitivity of pca for traffic anomaly detection. *SIGMETRICS Perform. Eval. Rev.*, 35(1):109–120, 2007.
- [12] A. Soule, H. Ringberg, F. Silveira, and C. Diot. Challenging the supremacy of traffic matrices in anomaly detection. *IMC '07*, pages 105–110, 2007.
- [13] K. Xu, Z.-L. Zhang, and S. Bhattacharyya. Internet traffic behavior profiling for network security monitoring. *IEEE/ACM Trans. Netw.*, 16(6):1241–1252, 2008.