Flooding Attacks Detection in Backbone Traffic Using Power Divergence

Ali Makke, Osman Salem, Mohamad Assaad, Hassine Moungla, Ahmed Mehaoua

Flooding Attacks Detection in Backbone Traffic Using Power Divergence

Ali Makke
University of Paris Descartes
LIPADE
45, rue des Saints-Pères
75006 Paris

Osman Salem
University of Paris Descartes
LIPADE
45, rue des Saints-Pères
75006 Paris

Mohamad Assaad
SUPELEC
Department of Telecommunications
91190 Gif-sur-Yvette

Hassine Moungla
University of Paris Descartes
LIPADE
45, rue des Saints-Pères
75006 Paris

Ahmed Mehaoua
University of Paris Descartes
LIPADE, France
POSTECH University
Division of ITCE, Korea

ABSTRACT

Flooding attacks detection in traffic of backbone networks requires generally the analysis of a huge amount of data with high accuracy and low complexity. In this paper, we propose a new scheme to detect flooding attacks in high speed networks. The proposed mechanism is based on the application of Power Divergence measures over Sketch data structure. Sketch is used for random aggregation of traffic, and Power Divergence is applied to detect deviations between current and established probability distributions of network traffic. We focus on tuning the parameter of Power Divergence to optimize the performance. We evaluate our approach using real Internet traffic traces, obtained from MAWI trans-Pacific wide transit link between USA and Japan. Our results show that the proposed approach outperforms existing solutions in terms of detection accuracy and false alarm ratio.

Categories and Subject Descriptors

C.2.3 [**Network Operation**]: Network Monitoring

General Terms

Algorithms, Design, Experimentation, Measurement, Performance, Security

Keywords

Network Anomaly Detection, DDoS, Power Divergence, Hellinger Distance, Sketch

PM2HW2N'12, October 21–22, 2012, Paphos, Cyprus.

1. INTRODUCTION

Security threats for computer network have increased significantly, which include viruses, exploitation of software vulnerabilities, worm-based attacks, Distributed Denial of Service (DDoS), etc. DDoS through TCP SYN flooding is able to make silent any web site, especially with the use of BOTNETs (roBOT NETworks) containing large number of slave machines (zombies).

Many web sites suffered from SYN flooding attack that aims to exhaust server resources and to deny access for legitimate users. Recently, PayPal has been driven offline after dropping WiKiLeaks donations (operation PayBack in 2010). The list of victim web servers is very long (Twitter, CNN, Yahoo, Amazon, Ebay, DoubleClick, etc.). Therefore, flooding attacks need to be detected in order to cope with ongoing anomaly as soon as possible.

With the distributed nature of DDoS attacks, and its impact on the performance of the routers, the detection and reaction mechanisms must be pushed to the core network (Backbone), or near the sources of attack. However, with the growing complexity in analyzing huge amount of data traffic in backbone network, the analysis of each traffic flows is unscalable and computationally expensive.

Many change point detection algorithms have been proposed, and applied to the time series resulted from the aggregation of whole network traffic in one flow. These methods are based on the identification of the change point where heavy deviation occurs in the resulted time series. However, as these methods aggregate the whole traffic in one time series, low intensity attacks is buried by the variation of background traffic in high speed networks, and may pass undetected. Moreover, the time series derived from IP traffic are subject to many variations that are irrelevant to anomaly. The time series are non-stationary and tend to change over time, leading to a lot of false alarms. Reducing the false alarms and increasing the detection accuracy in such methods are a challenging problem.

In this paper, we propose a new framework for the detection of flooding attacks. The proposed approach is intended to detect low intensity attacks in backbone traffic. It is based on Power Divergence (PD [4]) and Sketch data structure [6].

Sketch is used for dimensionality reduction and to derive time series for randomly aggregated traffic, and PD measures the difference between 2 set of probability values, and detects the deviation between these sets in normal and under anomaly conditions. The PD is a general form of divergence measures, where Kullback-Leibler (KL [3]), Hellinger Distance (HD [17]), Chi-square (χ^2 [16]), etc. can be derived from PD by changing its parameter (β). We emphasize in our analysis the optimal parameter of Power Divergence.

The rest of this paper is organized as follows. Related work is provided in section 2. Section 3 describes briefly the Sketch data structure. The proposed approach is explained in section 4. In section 5, we present our experimental results. Finally, in section 6 we present the conclusion and our future work.

2. RELATED WORK

Several approaches have been proposed for network anomaly detection, and they are based on different techniques, such as Haar-wavelet analysis [9], entropy based method [7], CUmulative SUM (CUSUM) algorithm [15, 18], adaptive threshold [2], Exponentially Weighted Moving Average (EWMA) [20], Holt-Winters seasonal forecasting [13], data reduction techniques with Sketch [6], SNMP MIB statistical data analysis [21], Principal Component Analysis (PCA) [8], etc.

Malicious activity usually provokes an abrupt change in the statistical values of the parameters describing the traffic, such as the NetScan produced by worms outbreak, that sends a large number of SYN from the same source IP, to scan the network before propagation phase. In [19], the CUSUM algorithm is used to detect SYN flooding over one time series resulted from aggregating whole traffic in one flow. In [15] a comparison between CUSUM and adaptive threshold for the detection of SYN flooding is presented. Both algorithms have been applied over one time series (whole traffic) for scalability issues and real time processing. Low intensity attacks get smoothed with the normal variations of background traffic in backbone network and pass undetected, i.e. flooding attack with intensity 10^6 packets/s does not produce a noticeable deviation when the total number of normal packets is greater than 10^9. Furthermore, these methods use static threshold for detecting anomalies, which is not adequate with traffic variations, and may induce false alarm and miss detection. In this paper, we want to overcome these problems (aggregation in one time series and static threshold) through the use of Sketch and dynamic threshold.

The Principal Component Analysis (PCA [8]) have been proposed and used for dimensionality reduction and anomaly detection. PCA detects anomalies using the evaluation of flows correlations over single link, instead of single time series of whole traffic from many links. However, PCA is very sensitive to the number of dominant PCs. Authors in [5] proved that PCA is unable to detect high intensity attacks without the good configuration. In [12], the authors showed that methods for tuning PCA are not adequate and starting with a new data set, adjusting parameters is unexpectedly difficult.

Sketch data structure uses the random aggregation for more grained detection than aggregating whole traffic in one time series. It has been used to summarize monitored traffic in a fixed memory, and to provide scalable input for time series analysis. We will exploit the Sketch data structure to derive probability distributions.

Authors in [1] experiment the histogram-based detector in order to detect the anomaly behaviors and changes in traffic distributions. They apply Kullback-Leibler divergence between the current and previous measurement distributions. Authors in [17] apply Hellinger Distance (HD) on Sketch data structure, in order to detect divergence between current and previous distributions of the number of SIP INVITE request. In fact, HD must be near zero when probability distributions are similar, and it increases up to one whenever the distributions diverges (e.g. under Invite flooding attacks). In addition, they used the dynamic threshold proposed in [14] during their experimental analysis.

In this paper, we provide a more grained analysis then aggregating the whole traffic in one time series. We develop a general framework that increases the detection accuracy and reduces the false alarm by integrating the Power Divergence over Sketch technique. We show also that KL, HD & χ^2 measures used in [1, 17] are special cases of our framework. We will show how PD outperforms these measures by changing its parameter (β).

3. K-ARY SKETCH

Sketch generates fixed-number of time series [6] regardless the number of exiting flows. It allows more grained analysis than aggregating whole traffic in one flow. It is based on random aggregation of traffic in d different hash tables. Let $X = x_1, x_2, \ldots, x_n$ denotes the set of input stream, where each item $x_i = (\kappa_i, \nu_i)$ is identified by a key $\kappa_i \in U$, drawn from a fixed universe of items U. $\nu_i \in \mathbb{R}$ is the value associated with each key. In our model, we use $\kappa_i = DIP$ and $\nu_i = 1$ for each received SYN by the associated Destination IP (DIP).

A Sketch S is a 2D array of $d \times w$ cell (as shown in Fig. 1), contains d hash arrays. The arrival of a flow with key κ_i increments its associated counter in the j^{th} hash table by ν_i ($S_{j,h_j(\kappa_i)} += \nu_i$). The update procedure is realized by d different hash functions, chosen from the set of 2-universal hash functions $h_j(\kappa_i) = \{((a_j\kappa_i + b_j) \bmod P_U) \bmod w\}$, to uniformly distribute κ_i over hash tables and to reduce collisions. The parameter P_U is a prime number larger than the maximum number in the universe. a_j and b_j are random integers smaller than P_U, with $a_j \neq 0$. Using d hash functions, the probability that two keys are aggregated in the same bucket over the d hash tables is $(1/w)^d$.

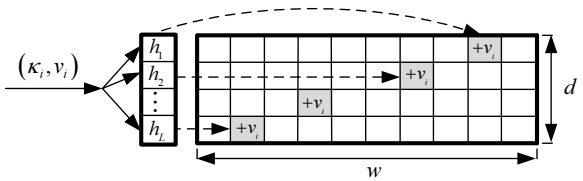

Figure 1: Sketch data structure

Sketch randomly aggregates multiple IP addresses in the same bucket, if the value resulted from hashing the addresses is the same ($h(DIP_1) = h(DIP_2) = j$). The counter in each bucket is used to derive a probability as the ratio of counter

value to the sum of whole counters in one hash table :

$$p_{ij} = \frac{S[i][j]}{\sum_{j=1}^{w} S[i][j]} \quad (1)$$

4. PROPOSED APPROACH

The approach used in this paper to detect flooding attacks is based on Power Divergence. In fact, the idea is to estimate the subjective prior distribution of the traffic and to use it as a baseline probability. Let $P_i = [p_{i,1}, ..., p_{i,w}]$ denotes the probability distribution derived from the i^{th} hash table, in the discrete time interval T. In presence of attacks, the probability distribution changes. One can use this change to detect anomalies. However, with the normal traffic variations, this probability distribution changes also even in the absence of anomaly. This is called false alarms. The objective then is to find a method that detects the attacks and remove the false alarms.

This motivates the need for a quantitative measure of information or more generally a decision theoretic measure of divergence between the basic probability P_i and some other distribution Q_i. Power Divergences are generalizations of this decision measure and are associated with strictly convex functions. The Power Divergence has been first defined in [4] and equivalent variants (up to a scale factor β) of this divergence are discussed in [11]. The divergence measure is therefore the decision measure that generalizes KL, HD and χ^2 divergence to a broad class of divergence of order β. In fact, the PD is a measure of distance between two probability set of order β given as follows:

$$PD_i = PD(P_i||Q_i) = \frac{\sum_{j=1}^{w} p_{ij}(p_{ij}/q_{ij})^{\beta-1} - 1}{\beta(\beta-1)} \quad (2)$$

where P_i is the probability distribution in the current time interval, and Q_i is the probability distribution in previous interval. $PD_i = 0$ iff P_i and Q_i are identical ($p_{ij} = q_{ij}$), and $PD_i > 0$ when $P_i \neq Q_i$. PD_i must be near zero under normal traffic, with a large deviation (one spike) when distributions change occurs. d measures $(PD_1, PD_2, \ldots, PD_d)$ are resulted from the d hash tables in Eq. 2. Anomaly induces spike in PD_i, and when more than L values of PD_i exceed a dynamic threshold, an alarm is raised.

The Power Divergence presents some interesting special cases when changing its parameter β, as shown in table 1. Obviously, Power Divergence outperforms then KL, χ^2 and HD measures as they are special cases. In fact, by changing the values of β, one can optimize the detection of attacks compared to the KL, χ^2 and HD measures. By experiments, we will show numerically that for different values of β, the detection efficiency changes. The optimal value of β can then be obtained through our experiments.

β	Divergence measure
-1	$\frac{1}{2} \times \chi^2(Q_i\|\|P_i)$
0	$KL(Q_i\|\|P_i)$
0.5	$4 \times HD(P_i\|\|Q_i)$
1	$KL(P_i\|\|Q_i)$
2	$\frac{1}{2} \times \chi^2(P_i\|\|Q_i)$

Table 1: Special cases of Power Divergence

P_i and Q_i in Eq. 1 are derived from the i^{th} hash table in Sketch data structure, in two consecutive discrete intervals. Firstly, the shared counters of the sketch are continuously updated from ongoing traffic during a time interval T. At the end of each interval, the probability $p_{i,j}$ is calculated, and PD is applied to detect deviations.

When L values of PD_i are larger than dynamically updated threshold, we raise an alarm. However, PD induces only two spikes (at the start and at the end of attack). As we want to continuously raise alarms for whole duration of the attack, the prior distribution Q_i will stop sliding by keeping its value until the end of the attack. However, with the variations of normal traffic, and the similarity of DDoS attacks with flash crowd, we suppose that flooding attacks will span for many intervals to overload a server, in contrast to flash crowd and normal variation. Thus reduce the false alarms. Therefore, we will trigger an alarm only if the deviation lasts more than η intervals.

Detection threshold

Let $PD_{i,kT}$ represent the time series of resulted Power Divergences from different time intervals. To detect deviations in $PD_{i,kT}$, we derive a subsequent time series $\overline{PD}_{i,kT}$ that contains the values of $PD_{i,kT}$ smaller than dynamic threshold, i.e. without spikes values. As the empirical rule states that 95% of data fall within 2 standard deviations of mean, $\overline{PD}_{i,kT}$ contains only values that satisfy:

$$\overline{PD}_{i,kT} < \mu_{i,(k-1)T} + 2.\sigma_{i,(k-1)T} \quad (3)$$

Where $\mu_{i,kT}$ & $\sigma_{i,kT}$ are the mean and the standard deviation of $\overline{PD}_{i,kT}$ respectively. $\mu_{i,kT}$ and $\sigma_{i,kT}$ are updated dynamically using the EWMA (Exponentially Weighted Moving Average):

$$\mu_{i,kT} = \alpha\mu_{i,(k-1)T} + (1-\alpha)\overline{PD}_{i,(k-1)T} \quad (4)$$

$$\sigma^2_{i,kT} = \alpha\sigma^2_{i,(k-1)T} + (1-\alpha)(\overline{PD}_{i,kT} - \mu_{i,kT})^2 \quad (5)$$

The threshold is updated dynamically by adjusting the value of $\mu_{i,kT}$ and $\sigma_{i,kT}$ as shown in eqs. 4 & 5. $PD_{i,kT}$ falls down the threshold $(\mu_{i,(k-1)T} + 2.\sigma_{i,(k-1)T})$ under normal condition, and exceeds the dynamic threshold under flooding attack. The decision function for alarms is given in Eq. 6. To reduce false alarm resulted from normal traffic variations, $PD_{i,kT}$ must exceed the dynamic threshold for η consecutive intervals before raising an alarm.

$$d(A_i) = \begin{cases} 1 & if\ PD_{i,kT} \geq \mu_{i,(k-1)T} + 2.\sigma_{i,(k-1)T} \\ & and\ \eta \geq 3 \\ 0 & Otherwise \end{cases} \quad (6)$$

5. EXPERIMENTAL RESULTS

In this section, we present the performance analysis results for the Power Divergence over Sketch, for SYN flooding attacks detection. Afterward, we conduct analysis to study the impact of parameters on true positive and false alarm ratio. Since web servers use TCP protocol, SYN flooding is the most commonly used attack. Therefore, we focus on presenting the experiment results for SYN flooding detection (due to space limitations), but the proposed approach had applied for the detection of any type of flooding (with TCP, UDP, ICMP, etc.).

We use the real internet MAWI [10] trans-Pacific traces from 15/04/2010 12h00 to 18h15 as few hours in the life of

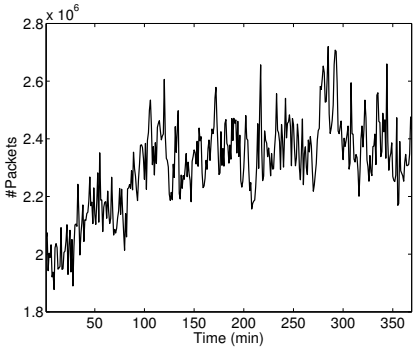

Figure 2: Total number of packets under normal condition

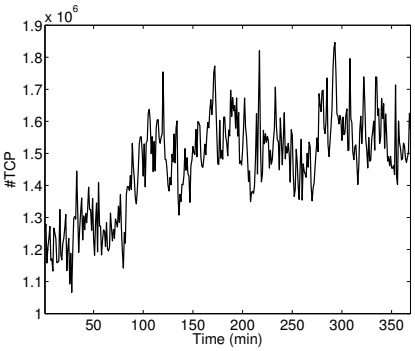

Figure 3: Total number of TCP segments under normal condition

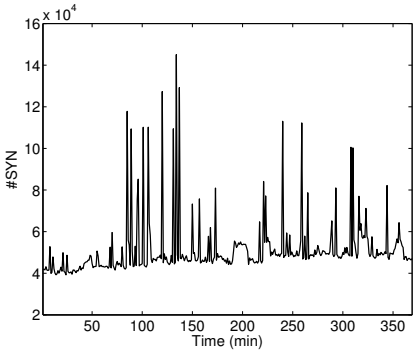

Figure 4: Total number of SYN segments under normal condition

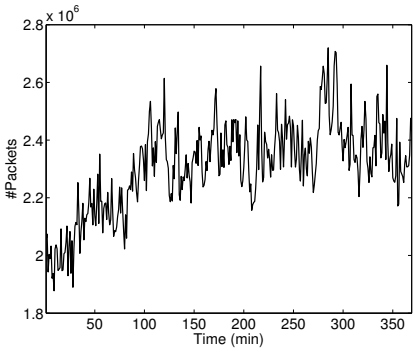

Figure 5: Total number of packets under DDoS attacks

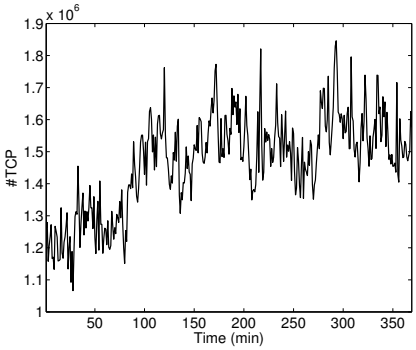

Figure 6: Total number of TCP segments under DDoS attacks

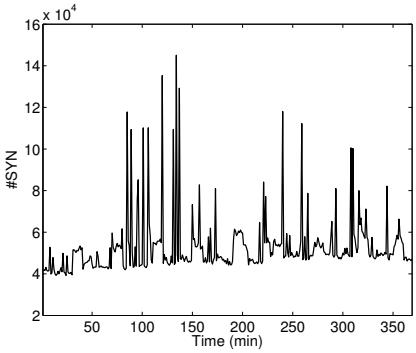

Figure 7: Total number of SYN segments under DDoS attacks

the internet, to test the efficiency of these used algorithm. We have analyzed these traces of wide area network, using Sketch with $\kappa_i = DIP$, and $v_i = 1$ for SYN request only. The used parameters in our implementation are: $w = 1024$, $d = 5$, $\eta = 3$, $\alpha = 0.2$, $L = 3$. For the sake of clarity, we present the results of Power Divergence over the first line of Sketch.

The effect of Sketch parameters (width and depth) on the detection accuracy had analyzed in [6]. Authors provide a detailed analysis of the impact of the number of hash function (d) and the Sketch width (w) on the detection accuracy. They found that Sketch provides better accuracy when increasing d & w, but at the cost of increasing the required memory and the computational complexity. A trade-off between accuracy and complexity is required, where $d = 5$ and $w = 1024$ were chosen as the lowest values to achieve this trade-off in our experiments. Also, the time interval T is subject to reduce the detection delay and the computational complexity. For a value of $T = 1sec$ the computational complexity increases, and for $T = 5min$ the detection delay increases. Therefore, we chose $T = 1min$ as tradeoff.

We analyze these traces using the proposed approach with κ_i=SIP (Source IP) and $v_i = 1$ for SYN segment. We find many scanning anomalies (SSH Scan, RPC & Netbios Scan, etc.). Afterward, we apply our approach with κ_i=DIP (Destination IP), and we don't find any existing SYN flooding attacks in these raw traces. Therefore, we inject 9 real DDoS attacks with different intensity to simulate distributed SYN flooding attacks. These attacks are inserted each 30 minutes, and the duration of each one is 10 minutes. The intensity of these attacks begins with a value of 10000 and decreases until 2000 SYN/min.

Fig. 2 and Fig. 5 show the variation of the total number of packets before and after the SYN flooding attacks. By comparing these variations, no noticeable deviations between both figures, and inserted attacks don't induce heavy deviations in the time series of the total number of packets. Indeed, Fig. 3 and Fig. 6 show the variation of the number of TCP segments before and after the SYN flooding attacks. We can notice that the shape of traffic variations in both figures is similar. This can be explained by the fact that the intensity of SYN flooding is relatively small when compared to the intensity of the total number of TCP segments. In such cases, the detection of these low intensity attacks is very challenging.

Fig. 4 and Fig. 7 show the variation of SYN before and after SYN flooding attacks. In fact, we can notice the large variations in the total number of SYN (Fig. 4) before attacks injection. Therefore, the aggregation of whole traffic in one time series produces a lot of false alarms for these heavy deviations. On the other hand, the injected SYN flooding attacks produce small variations in the number of SYN (lower than existing normal variations) as shown in Fig. 7. We may not notice the difference between Fig. 4 & Fig. 7 without a deep inspection, given the low intensity of flooding attacks and the large variations in the number of SYN.

We conducted many experiments by varying the values of β in order to find the optimal one. We present the variation

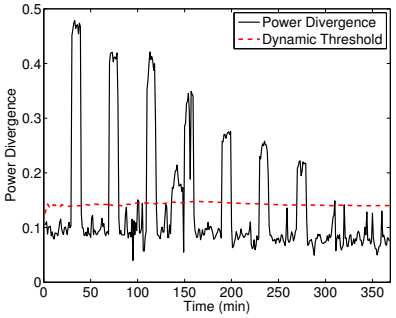

Figure 8: PD for $\beta = 0.5$ (Hellinger Distance)

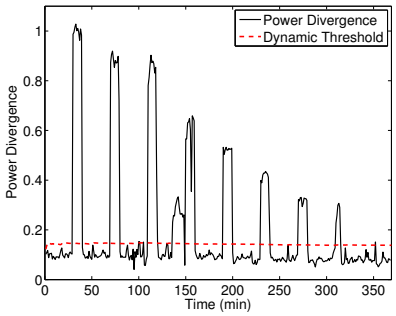

Figure 9: PD for $\beta = 1$ (Kullback-Leibler)

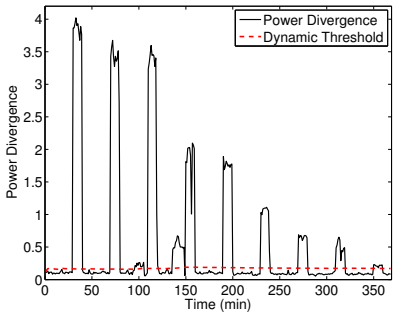

Figure 10: PD for $\beta = 1.5$ (Unknown Divergence)

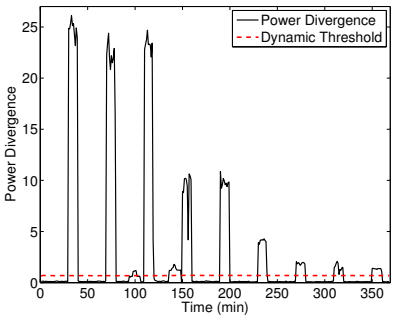

Figure 11: PD for $\beta = 2$ (χ^2 divergence)

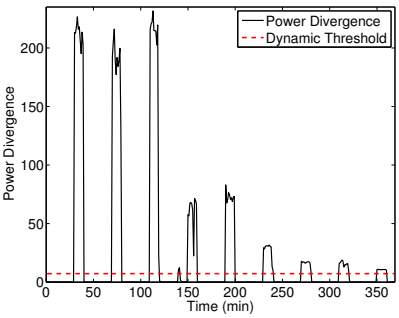

Figure 12: PD for $\beta = 2.5$ (Unknown Divergence)

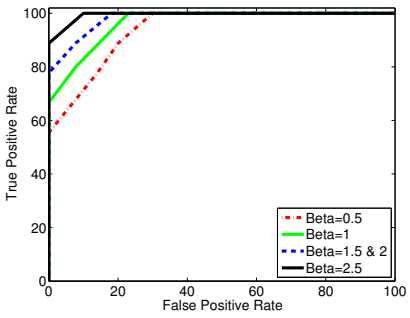

Figure 13: Receiver Operating Characteristic (variation of β)

of Power Divergence with the dynamic threshold (given in Eq. 3) on the traffic with variable intensity attacks. Fig. 8 shows the variation of PD with $\beta = 0.5$ ($PD = 4 \times HD$), Fig. 9 with $\beta = 1$ ($PD = KL$), Fig. 10 with $\beta = 1.5$, Fig. 11 with $\beta = 2$ ($PD = 1/2 \times \chi^2$), and Fig. 12 with $\beta = 2.5$. Whenever the Power Divergence exceeds the dynamic threshold, an alarm is raised. It is important to note the scale difference between these figures, where the intensity of raised pikes increases when increasing the value of β. We don't use the same scale for the sake of clarity.

To evaluate the performance of Power Divergence with different values of β, we investigate the detection rate and the false alarm rate. Receiver Operating Characteristic (ROC) is used for accuracy analysis when varying the value of the threshold. ROC curve shows the variation of the true positive (Eq. 7) in term of false alarm rate (Eq. 8):

$$DR = \frac{TP}{TP + FN} \times 100 \quad (7)$$

Where TP is the number of true positive alerts, and FN is the number of false negative. The false alarm rate is defined as the ratio of false alarms to the number of raised alarms:

$$FAR = \frac{FP}{TP + FP} \times 100 \quad (8)$$

In Fig. 13, we present the ROC curve for different values of β. We found through experiments that for $\beta = 0.5$ (PD is proportional to HD), PD achieves a DR=100% with a FAR=30%, for $\beta = 1$ (PD is equal to KL), PD achieves a DR=100% with a FAR=23%, for $\beta = 1.5$ & $\beta = 2$ (PD is proportional to χ^2), PD achieves a DR=100% with a FAR=18%, and for $\beta \geq 2.5$, PD achieves a DR=100% with a FAR=10%. It is important to note, that we conduct many experiments with different values of β (with a step 0.5 and up to a value of 100). We found through experiments, with a value of $\beta \geq 2.5$, the intensity of PD for detected attacks increases significantly and proportionally to the attack intensity, but without any change in the ROC. Therefore, PD with $\beta = 2.5$ outperforms existing measures and achieves the optimal performance.

6. CONCLUSION AND PERSPECTIVES

In this paper, we proposed a new framework based on Sketch and Power Divergence for anomaly detection over high speed links. The proposed approach evaluated on real traces with DDoS SYN flooding attacks. Our experimental results showed the capacity of PD in the detection of low intensity attacks. We presented the results of PD with different values of β, as well as the associated ROC curves when varying the threshold. We proved through experiments that PD outperforms existing measures (HD, KL & χ^2) when increasing the value of (β). We found that for $\beta = 2.5$, PD achieves the optimal performance.

In our future work, we will focus on defense mechanisms, in order to trigger automatic reaction against ongoing attacks. We also intend to provide a method for reducing the amount of monitored data on high speed networks, and to analyze the impact of sampling on the precision of PD divergence measure.

7. ACKNOWLEDGMENTS

This research was supported by Korea Science and Engi-

neering Foundation, under the World Class University (WCU) program.

8. REFERENCES

[1] D. Brauckhoff, X. Dimitropoulos, A. Wagnerand, and K. Salamatian. Anomaly Extraction in Backbone Networks Using Association Rules. In *Proceedings of the 9th ACM SIGCOMM conference on Internet Measurement Conference (IMC'09)*, pages 28–34, 2009.

[2] S. Bu, R. Wang, and H. Zhou. Anomaly Network Traffic Detection Based on Auto-Adapted Parameters Method. In *Proceedings of the 4th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM'08)*, pages 601–607, 2008.

[3] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley-Interscience, 1991.

[4] J. Havrda and F. Chavrat. Quantification method of classification processes: The concept of structural α-entropy. *Kybernetika*, 3:30–35, 1967.

[5] N. L. D. Khoa, T. Babaie, S. Chawla, and Z. Zaidi. Network Anomaly Detection Using a Commute Distance Based Approach. In *International Conference on Data Mining Workshops (ICDW'10)*, pages 943–950, 2010.

[6] B. Krishnamurthy, S. Sen, Y. Zhang, and Y. Chen. Sketch-based Change Detection: Methods, Evaluation, and Applications. In *Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement (IMC'03)*, pages 234–247, 2003.

[7] A. Lakhina, M. Crovella, and C. Diot. Mining Anomalies using Traffic Feature Distributions. *SIGCOMM Comput. Commun. Rev.*, 35(4):217–228, 2005.

[8] A. Lakhina, M. Crovella, and C. Diot. Mining anomalies using traffic feature distributions. *SIGCOMM Comput. Commun. Rev.*, 35(4):217–228, 2005.

[9] W. Lu and A. A. Ghorbani. Network Anomaly Detection Based on Wavelet Analysis. *EURASIP Journal on Advances in Signal Processing*, pages 1–16, 2009.

[10] MAWI working group traffic archive. http://mawi.wide.ad.jp/mawi/.

[11] P. N. Rathie and P. Kannappan. A Directed-Divergence Function of Type β. *Inform. Contr.*, 20:38–45, 1972.

[12] H. Ringberg, A. Soule, J. Rexford, and C. Diot. Sensitivity of PCA for Traffic Anomaly Detection. In *Proceedings of the ACM SIGMETRICS'07 Conference*, pages 109–120, 2007.

[13] R. Schweller, Z. Li, Y. Chen, Y. Gao, A.Gupta, Y. Zhang, P.A.Dinda, M.-Y. Kao, and G. Memik. Reversible Sketches: Enabling Monitoring and Analysis Over High-Speed Data Streams. *IEEE/ACM Transactions on Networking*, 15(5):1059–1072, 2007.

[14] H. Sengar, H. Wang, D. Wijesekera, and S. Jajodia. Detecting VoIP Floods Using the Hellinger Distance. *IEEE Trans. Parallel Distrib. Syst.*, 19:794–805, 2008.

[15] V. A. Siris and F. Papagalou. Application of Anomaly Detection Algorithms for Detecting SYN Flooding Attacks. In *Proceedings of IEEE GLOBECOM'04*, volume 4, pages 2050–2054, 2004.

[16] I. J. Taneja. Bounds on Triangular Discrimination, Harmonic Mean and Symmetric Chi-Square Divergences. *Journal of Concrete and Applicable Mathematics*, 4(1):91–111, 2006.

[17] J. Tang, Y. Cheng, and C. Zhou. Sketch-Based SIP Flooding Detection Using Hellinger Distance. In *Proceedings of the 28th IEEE conference on Global telecommunications (GLOBECOM'09)*, GLOBECOM'09, pages 3380–3385, 2009.

[18] H. Wang, D. Zhang, and K. G. Shin. SYN-dog: Sniffing SYN Flooding Sources. In *IEEE ICDCS'02*, pages 421–428, 2002.

[19] H. Wang, D. Zhang, and K. G. Shin. Change-Point Monitoring for the Detection of DoS Attacks. *IEEE Trans. On Dependable and Secure Computing*, 1(4):1993–2004, 2004.

[20] N. Ye, S. Vilbert, and Q. Chen. Computer Intrusion Detection Through EWMA for Autocorrelated and Uncorrelated Data. *IEEE Transactions on Reliability*, 51(1):75–82, 2003.

[21] J. Yu, H. Lee, M.-S. Kim, and D. Park. Traffic Flooding Attack Detection with SNMP MIB using SVM. *Computer Communications*, 31(17):4212–4219, 2008.