Sketch based Anomaly Detection, Identification and Performance Evaluation

Patrice Abry, Pierre Borgnat, Guillaume Dewaele,
patrice.abry@ens-lyon.fr, pierre.borgnat@ens-lyon.fr, guillaume.dewaele@ens-lyon.fr,
Physics Dept., CNRS UMR 5672, ENS Lyon, France

Abstract

An anomaly detection procedure is defined and its statistical performance are carefully quantified. It is based on a non Gaussian modeling of the marginal distributions of random projections (sketches) of traffic aggregated jointly at different levels (multiresolution). To evaluate false negative vs. false positive in a controlled, reproducible and documented framework, we apply the detection procedure to traffic time-series from our self-made anomaly database. It is obtained by performing DDoS-type attacks, using real-world attack tools, over a real operational network. Also, we illustrate that combining sketches enables us to identify the target IP destination address and faulty packets hence opening the track to attack mitigation.

1 Motivations and Contributions

Motivations. Internet is becoming the major and universal communication network. It allows an increasing number of different activities of continuously evolving natures, ranging from on-line games, video-on-demand or IP voice telephony to web browsing or file exchanges. This implies accommodating the circulation of data with very different natures (texts, images, sounds,...), with wide variations in their characteristics (file sizes,...) and with demanding and potentially antagonist constraints and requirements (real-time, acceptable delays or loss-rates, security levels, confidentiality levels, tarifications...). These many causes of diversity in nature result in Internet Traffic flows that naturally exhibit huge fluctuations around putative average behaviors. On top of this inherent large variability are potentially superimposed non stationarities such as seasonal effects (day, nights, weekends,...) or such as harder to describe and burstier in nature events (flash crowds,...). Accompanying its increase of popularity and its being massively and diversely used, Internet becomes a place where vulnerability and security constitute central issues. Indeed, it has been observed that Internet is subject to a continuously increasing and diversifying variety of threads and attacks. Defense against attacks implies to detect their occurrence, to classify their nature or impact and to identify their origins to activate appropriate reaction mechanisms. Attack detection and mitigation is hence a major research area in Internet traffic analysis. Whatever the signal processing techniques chosen to perform anomaly detection (see [1, 5, 6, 12] for reviews and descriptions of the most prominent ones), a detection procedure always relies on the following key ingredients. First, traffic is characterized using a well-chosen (statistical) description model. This means not only to determine the average behavior of the (parameters of the) models under regular (attack-less) conditions but also to evaluate accurately the characteristics of their inevitably large variabilities and complex statistical properties. Second, a (statistical) distance between the parameters of the models estimated under attacks and those characterizing the normal conditions are computed. Third, it is necessary to decide when this distance is *large*, hence yielding a detection. This is where the fluctuations under normal conditions needs to be accurately analyzed. Then, the statistical performance of the procedure are to be validated via the production of detection rate versus false alarm rates curves. To finish with, detection procedures can be complemented with techniques enabling the identification of the origins of the attack.

Contributions. We propose here a detection procedure based on the use of a multiresolution and non Gaussian statistical modeling [10] of the marginals distributions of random projections (sketches) [4, 7, 9] aggregated at different resolution levels (cf. Section 3). Detections are obtained by the use of distances, computed from these statistics (cf. Section 4). The use of a combined collection of sketches enables us to identify the faulty packets. This inversion or identification procedure is detailed in Section 6.

In the present work, we intend to detect low volume attacks. We are not interested in situations where, because of the distributed nature of nowadays attacks, the traffic volume increase they yield close to the target is large. Indeed, for such situations, detection is easy (and does not require advanced signal processing analyses) but takes places too late: Impacts on network performance, Quality of Service degradation and waste of resources are already high. Instead, we consider detection at early stages of the attacks and close to the sources when the traffic volume increase and the net-

#id	Tool	#bots	#Total(Attack) pkt/s	Intensity	throughput
1	UDP	2	1098(74)	7%	200kb/s
2	UDP	4	3640(148)	4%	388kb/s
3	UDP	4	2190(148)	7%	388kb/s
4	TCP-SYN	2	4258(520)	12%	166kb/s
5	ICMP	4	1099(92)	8%	288kb/s
6	ICMP	4	1820(179)	10%	388kb/s
7	Mixed	4	2781(760)	27%	250kb/s
8	Smurf	4	2501(95)	4%	250kb/s

Table 1. Anomaly Database: Characteristics of the Attacks. **All attacks are obtained using Tfn2k, except the one on first line (Trin00). All illustrations reported here are obtained from the trace of last line (#id 8).**

work degradation impacts caused by the illegitimate traffic remain invisible and negligible. This can be seen as the case where a router collects the packets generated by of a small number of bots. Because we want to be able to assess the statistical performance of the proposed detection procedure in a comparable and reproducible manner, we have chosen an approach that consists of a trade-off: We work with real traffic collected on an operational network (RENATER) and real DDoS type attacks performed with real-world attack tool [3]. However, they are performed by ourselves in a controlled manner (cf. Section 2) hence enabling false positive *vs.* false negative evaluation (cf. Section 5).

2 Anomaly Database

Attacks were performed using Trin00 and Tfn2k, two well-known real world tools commonly involved in computer network attacks. While Trin00 daemons create only UDP packet flooding attacks, Tfn2k consists of a more versatile tool enabling various types of DDoS attacks: UDP flooding, TCP/SYN flooding, ICMP/Echo flooding, combination of these three methods, SMURF (sending Echo packets to a broadcast adress, replacing the IP of the sender with the one of the victim) and Targa3 (malformatted packets). For reasons made clear above, attacks are kept very low in volume and have no significant impact on network performance (and hence cannot be subsumed to global mean or variance elementary changes). Attacks were conducted on the real operational RENATER network (the French high speed network used by research and education institutions), and involved attacking machines hosted in four different sites in France, while the attacked machine consists of a single computer. Attacks are conducted during afternoon hours, and are mixed with real, legitimate, internet traffic. Attacks typically last 10 minutes and traffic is collected before and after, during an hour. These campaigns of experiments, whose descriptions are summarized in Table 1, provide us with a labeled database of real traffic traces containing low intensity bandwidth controlled and documented anomalies. Fig. 2, top left, presents one such trace.

3 Modeling

Multiresolution-Gamma Modeling. Aggregated traffic X_Δ is not in general Gaussian and its marginal probability density function (PDF) $f_\Delta(x)$ is likely to vary both with the aggregation level Δ and the degree of traffic multiplexing. In [10], we proposed to describe $f_\Delta(x)$ by means of Gamma distributions, $f_\Delta(x) = x^{\alpha-1}e^{-x/\beta}/(\Gamma(\alpha)\beta^\alpha)$. A Gamma random variable is characterized by its scale and shape parameters $\beta > 0$, and $\alpha > 0$. By modeling jointly a collection of aggregation levels Δs, the Multiresolution-Gamma Modeling proposed here catches not only the marginal distributions of the aggregated traffic but also its time correlation structure. Indeed, X_Δ being an uncorrelated time-series would yield $\alpha_\Delta = \alpha_0\Delta$ and $\beta_\Delta = \beta_0$. Therefore, any departure from these elementary behaviors with respect to Δ constitutes a signature of the time correlations in X_Δ (cf. [10] for details).

Random Projections or Sketches. Inspired from [4, 7, 9], we split traces by means of random projections, or sketches, obtained as the outputs of H different M-output hash tables. Hash functions are constructed using the fast-tabulation method of [11]. Here, Hash functions are applied to the IP destination ($IPdst$): $m_i = h_n(IPdst_i) \in \{1, ..., M\}$, where $\{i, t_i\}$ refers to the i-th packet of the trace and to its arrival time. This can straightforwardly be extended by hashing the remainder of the 5-tuple of the IP header. The m-th outputs of the n-th sketch are aggregated yielding the $X_\Delta^{n,m}(t)$ time series. Fig. 2 presents examples of sketches (left column) and shows that the Multiresolution-Gamma Modeling detailed above equally relevantly describes the entire trace X_Δ and the sketch outputs $X_\Delta^{n,m}(t)$, both for regular traffic and traffic with anomalies. Unlike for the entire trace, the anomaly is clearly apparent, here in output $m = 20$, validating the intuition that sketching highlights the anomaly and hence increases the Signal to Noise Ratio for the detection procedure.

4 Detection

Intuitions. The central empirical observation is reported in Fig. 1: drastic changes in the shape (not only in the value) of $\alpha_\Delta^{n,m}$ take place when attacks occur. This is clearly seen on the sketch output containing the attack ($m = 20$ here), while no change is visible for other sketch outputs. This change in $\alpha_\Delta^{n,m}$ as a function $j = \log_2 \Delta$ betrays a sharp modification in the short-time correlations of the traffic, and not only in its mean or standard deviations. Tracking this correlation structure alteration consists of the core of the detection procedure described below.

Distances. Traces under analysis are split into adjacent non overlapping time windows of length T. Independently for each time window and each aggregation level, one computes a Mean Quadratic Distance (MQD) be-

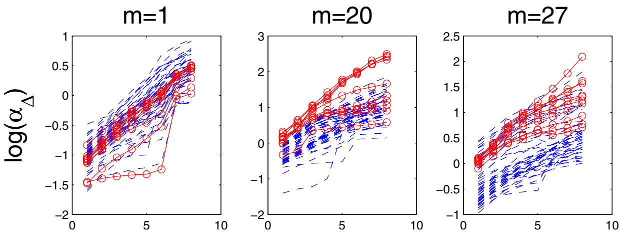

Figure 1. MULTIRESOLUTION GAMMA PARAMETERS. **Evolution of $\alpha_\Delta^{n,m}$ with respect to Δ for 3 sketch outputs. Time windows containing the attack are marked in (red, 'o'). Attack (located in $m = 20$) is clearly betrayed by a significant change in the shape of $\alpha_\Delta^{n,m}$.**

tween α_Δ and β_Δ measured on the current time window l and on a a priori chosen reference window, $D_\alpha(l) = \frac{1}{J}\sum_{j=1}^{J}\left(\hat{\alpha}_{2^j}(l) - \hat{\alpha}_{2^j}(ref)\right)^2$, the definition for D_β is identical, mutatis mutandis: Other distances can be used (see e.g., [2]). For instance, Kullback divergences (KD) between the PDFs $f_{\Delta,l}$ and $f_{\Delta,Ref}$ are defined as [2]: $K_\Delta^{(1d)}(l) = \int(f_{\Delta,l} - f_{\Delta,Ref})(\ln f_{\Delta,l} - \ln f_{\Delta,Ref})dx$. $K_\Delta^{(1d)}$s at various Δs are combined to produce multiresolution distances. Multidimensional KD between $f_{\Delta,\Delta',l}$ and $f_{\Delta,\Delta',Ref}$ can also be computed (cf. [10]).

Alarms. Finally, distances are thresholded to yield alarms.

5 Performance

Parameter Setting. Here, $T = 1$min and $J = 9$, i.e., $\Delta = \{2^1, \ldots, 2^9\} * 5$ms. We use $2 \le H \le 10$ different hash functions (conceived from known initial random seeds), with $5 \le M \le 50$. The reference is computed from the entire traffic collected before and after the self-made anomaly, and therefore assumed to be regular traffic.

Distances as a response to attacks. Fig. 2 displays distances computed for the entire trace and 3 different sketch outputs. For $m = 20$, the distances clearly yield unambiguous alarms for each 1min time-windows during the anomaly. This is relevant as output $m = 20$ contains the anomaly (this is easily checked as the target IP address is chosen a priori). Also, one notices that clear changes in mean or variance can be seen in other sketch outputs ($m = 1, 27$ shown here). Interestingly enough, the distances do not raise alarms for those cases. It clearly illustrates that to correctly distinguish anomalies from a mere increase of traffic, correlations are to be taken into account, hence the interest of the Multiresolution-Gamma modeling. Fig. 2 clearly shows that the anomaly is detected within its 1st or 2nd minute of occurrence, i.e., extremely quickly.

False Positive vs. False Negative. Statistical performance for detection procedures are usually assessed via Probability of Detection vs. Probability of False Alarm curves, such as those reported in Fig. 3. A collection of threshold values λ are used. For each λ, an alarm is triggered for the time-windows l where the distance bypasses λ. Then, our documented database enables us to count the number of correct and incorrect alarms. These results are extremely satisfactory, with false alarm rates below the percent for high detection rates for MQD distance D_α despite the fact that attacks have very low intensity and can hardly be seen. Kullback distances yield even better results (cf. Fig 3).

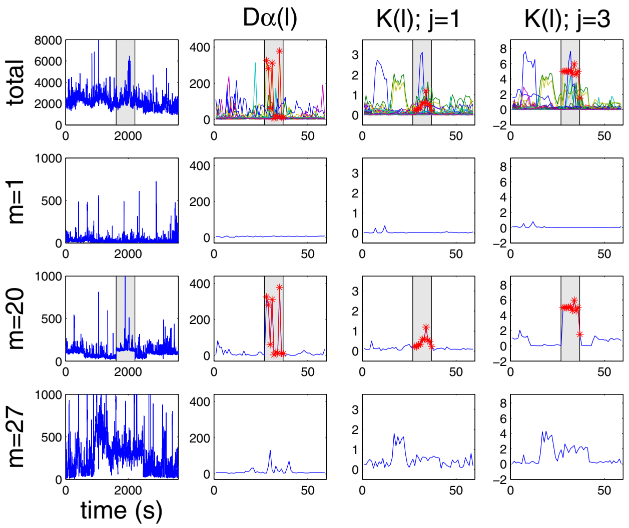

Figure 2. DISTANCES. **Trace aggregated at $\Delta = 1$s, top: total traffic, bottom: Sketch Outputs $m = 1, 20, 27$. Left to right: MQD D_α and K_Δ^{1d} distances, for $\Delta = 10$ms and $\Delta = 40$ms, as functions of time. Anomaly (here, in $m = 20$) clearly yields large values of the distances, hence detections. Note that the distances do not take large values despite the clear shift in means observed in sketches $m = 1, 27$ alarms and hence do not trigger false negatives.**

Confidence level. From the database and the use of the traces when no anomaly occurs, one can estimate the PDF of the distances under regular conditions. From this, the detection procedure can not only output a decision (alarm is triggered or not) but also its confidence level (the so-called p-value). This will be detailed elsewhere, together with further improvements obtained from consecutive threshold bypasses.

6 Target and Faulty Packet Identification

Because sketches sort the traffic, they can be involved in target and faulty packet identification. However, inverting the sketch procedure implies the use of a collection of H different hash functions, each with M outputs. Essentially, an IP destination address ($IPdst$) potentially identifies the attack target, if $\forall n = 1, \ldots, H$, $m_n^A = h_n(IPdst_i)$ where

N_{IP}	$H =$	1	2	3	4	5	6	7
83629		2054	51	2	1	1	1	1
2^{32}		107148219	2677374	67432	1719	51	1	1

Table 2. IDENTIFICATION: **Number of IP identified as the potential target, function of the numbers of sketches used and the number of IP addresses monitored (83629 in the trace).**

m_n^A stands for the output number of the n-th sketch procedure where an alarm is triggered. Obviously, the sketch procedure is prone to collisions (different $IPdst$ share the same sketch outputs), and a large enough number H of sketches is required to ensure an ambiguity-less inversion. The use of k-universal (with $k \geq 2$) hash functions plays here a key role, as it ensures that the average number of collisions (between any $IPdst$ and that of the attack target) diminishes exponentially fast with the number of sketches H [8]: $\#_C = N_{IP} M^{-2H}$, where N_{IP} denote the number of IP addresses in the analyzed traffic. Moreover with $k \geq 4$, the variance of $\#_C$ remains equally low. Obviously, to ensure a one-to-one identification, $\#_C$ needs to be kept low (below 1). In practice, in this work, $N_{IP} \simeq 2^{17} \ll 2^{32}$, with $k = 4$, and $M = 40$, $H = 4$ is large enough (cf. Table 2). The computational efficiency of the table-based hashing procedure virtually enables to apply the inversion procedure to each of the 2^{32} possible IP addresses, in a couple of minutes. However, this can be further improved by retaining only the $IPdst$s actually appearing in the analyzed traffic. A list containing all $IPdst$s currently observed is kept. This list can be flushed regularly, as long as no anomaly is detected. IP address target identifications opens the track for faulty packet identification. This is a promising research direction being currently developed and implemented.

7 Conclusions and Perspectives

This work shows that the proposed detection procedure, tracking a local correlation change in traffic via a multiresolution Gamma modeling combined with a sketch procedure, is able to detect attacks even with very low intensity levels. Using our self-made documented anomaly database, we are able to assess its statistical performance in a controlled and reproducible manner. Also, the detection procedure outputs the identification of the attack target, hence enabling reaction and mitigation.

For sketch analysis, the fast-tabulation method of [11] is scalable for more loaded networks. Therefore, the full procedure, combining hash, multiresolution modeling (implemented by the wavelet pyramidal algorithm) and distances computed over sliding windows, can potentially be implemented on-line. Furthermore, automatic choices of the reference time-window (based on non-stationary signal processing methods) and of the threshold are considered, together with multi-point measurements. Such further developments are under current investigations.

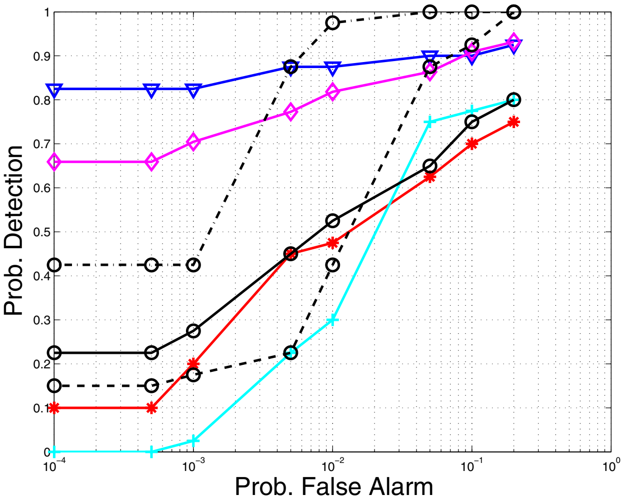

Figure 3. DETECTION VS. FALSE ALARM RATES. $T = 1\text{min}$. **For distance D_α and for attacks #id 3(7%,*), 4(12%,$\triangledown$), 6(10%,+), 7(27%,$\diamond$), 8(4%,o) in Table 1. For 8(4%,o), dashed and mixed lines correspond to K_Δ^{1d}, with $\Delta = 10\text{ms}$ and $\Delta = 40\text{ms}$, for comparison.**

References

[1] P. Barford, J. Kline, D. Plonka, and A. Ron. A signal analysis of network traffic anomalies. In *ACM/SIGCOMM IMW*, Marseille, France, Nov. 2002.

[2] M. Basseville. Distance measures for signal processing and pattern recognition. *Signal Processing*, 18:349–369, 1989.

[3] P. Borgnat et al. Détection d'attaques de déni de service par un modèle non gaussien multirésolution , In *CFIP'2006*, Nov. 2006.

[4] B. Krishnamurty, S. Sen, Y. Zhang, and Y. Chen. Sketch-based change detection: Methods, evaluation, and applications. In *ACM IMC*, Oct. 2003.

[5] A. Lakhina, M. Crovella, and C. Diot. Diagnosing network-wide traffic anomalies. In *SIGCOMM*, Aug. 2004.

[6] L. Li and G. Lee. DDoS attack detection and wavelets. In *International Conference on computer communications and networks*, Aug. 2003.

[7] X. Li et al. Detection and identification of network anomalies using sketch subspaces. In *ACM IMC*, Oct. 2006.

[8] M. Dietzfelbinger, J. Gil, Y. Matias, and N. Pippenger. Polynomial hash functions are reliable. In *Proc. 19th ICALP, LNCS 623*, pages 235–246, 1992.

[9] S. Muthukrishnan. Data streams: Algorithms and applications. In *ACM SIAM SODA*, Jan. 2003.

[10] A. Scherrer, N. Larrieu, P. Owezarski, P. Borgnat, and P. Abry. Non gaussian and long memory statistical characterisations for internet traffic with anomalies. *IEEE Trans. on Depend. and Secure Comp.*, Sept. 2006. To Appear.

[11] M. Thorup and Y. Zhang. Tabulation based 4-universal hashing with applications to second moment estimation. In *Proc. ACM-SIAM SODA*, Jan. 2004.

[12] Y. Zhang, Z. Ge, A. Greenberg, and M. Roughan. Network anomography. In *ACM IMC*, Oct. 2005.