DDoS Detection with Information Theory Metrics and Netflows
*A Real Case**

Domenico Vitali[1], Antonio Villani[2], Angelo Spognardi[1], Roberto Battistoni[1] and Luigi V. Mancini[1]
[1]*Dipartimento di Informatica, "Sapienza" University of Rome, Via Salaria 113, 00198, Rome, Italy*
[2]*Dipartimento di Matematica, University of Roma Tre, Largo S. L. Murialdo, 00146, Rome, Italy*

Keywords: DDoS, Attack Detection, Information Divergence, Relative Entropy, Autonomous System, Internet Security.

Abstract: Denial-of-Service (DoS) and Distributed Denial-of-Service (DDoS) constitute one of the main issues for critical Internet services. The widespread availability and simplicity of automated stressing tools has also promoted the voluntary participation to extensive attacks against known websites. Today the most effective (D)DoS detection schemes are based on information theory metrics, but their effectiveness is often evaluated with synthetic network traffic. In this work we present a comparison of the main metrics proposed in the literature carried on a huge dataset formed by real netflows. This comparison considers the ability of each metric to detect (D)DoS attacks at an early stage, in order to launch effective and timely countermeasures. The evaluation is based on a large dataset, collected from an Italian transit tier II Autonomous System (AS) located in Rome. This AS network is connected to all the three main network infrastructures present in Italy (Commercial, Research and Public Administration networks), and to several international providers (even for Internet transit purposes). Many attempted attacks to Italian critical IT infrastructures can be observed inside the network traffic of this AS. Several publicly declared attacks have been traced and many other malicious activities have been found by ex-post analysis.

1 INTRODUCTION

One of the most critical aspect of (D)DoS (Denial of Service and Distributed DoS) attacks is their artlessness and simplicity. While the synchronization of attacking entities is still performed using botnets of unaware compromised hosts, nowadays, simple word-of-mouth ways are used to coordinate volunteers attackers (e.g. chat/twitter/irc channels). Recently, one of the most used tool to perform DDoS is "LOIC" (Low Orbit Ion Cannon), a software originally designed to test the robustness of services and able to quickly flood with connections a target IP. LOIC and similar tools make these activities exploitable by all Internet users: political *hacktivists*, individuals or interested groups keep increasing their use to express disagreement against private companies or public entities. Many examples can be found in the past years (Cisco Systems, 2010). For example, in September 2010, a DDoS attack named *Operation Payback* was launched against the Motion Picture Association of America's (MPAA) web-page. Similarly, strong emphasis was given to the series of DDoS attacks against several companies which resulted in a cut off for WikiLeaks.org or to the Playstation's online store as a form of revenge towards Sony's lawsuit against the PS3 hacker George Hotz. More recently, DDoS attacks have been reported to the Italian Government and the Vatican State web sites and many other international institutions. In general, every Internet Critical Infrastructure or any sensitive economic service can be considered a possible target.

The effects of (D)DoS attacks can be serious: in the best cases, the network services hosted by the target Autonomous System become unavailable as long as the attack activity persists; in the worst cases, the session between the target AS and its ISP breaks out, making a black hole where the packets are all dropped, eventually causing a chain reaction that amplifies the attack and spreads its effect on other ASes. (D)DoS attacks are considered really challenging and have generated a large amount of research activity. Several works in the last decade try to survey metrics, strategies and tools to protect network services and to reduce the impact of such malicious activities. At

*This paper has been supported by the Prevention, Preparedness and Consequence Management of Terrorism and other Security-related Risks Programme European Commission - Directorate-General Home Affairs, under the ExTraBIRE project, HOME/2009/CIPS/AG/C2-065

Vitali D., Villani A., Spognardi A., Battistoni R. and V. Mancini L..
DDoS Detection with Information Theory Metrics and Netflows - A Real Case.
DOI: 10.5220/0004064501720181
In *Proceedings of the International Conference on Security and Cryptography* (SECRYPT-2012), pages 172-181
ISBN: 978-989-8565-24-2

the same time, new attack flavors (ip-spoofing, low-rate attacks, botnet and others) keep raising the level of challenge. Finally, privacy concerns and the lack of secure techniques to make data anonymous, keep researchers unable to freely share their own traffic datasets and network dumps, slowing and hindering the research on this topic.

This paper focuses on DoS and Distributed DoS attacks that consume the bandwidth resources of a whole AS. In the DDoS taxonomy defined by Mirkovic et al. (Mirkovic and Reiher, 2004), such kind of attack has code VT-4, since it generates an extremely large number of network flows, saturating all the router resources (CPU or ram or bandwidth capacity) of the AS (we defer the formal definition of *flow* to Section 3.2). Increasing router resources is typically helpless against bandwidth saturation attacks, mainly for *stub* ASes: they usually purchase the minimal required bandwidth, but suffer (D)DoS attacks from intermediate ASes, with much higher traffic capacity. A more effective solution is to block the malicious traffic in advance, in the upper ASes, before it could reach the target AS. The most used approach to distinguish malicious packets among the aggregated traffic at AS level is the adoption of information theory metrics, since they are able to make traffic anomalies to "emerge" from the whole traffic flows. As it will be clear in the following, the effectiveness of proposed metrics is evaluated using synthetic traffic, where attack patterns are artificially injected. In this paper, instead, we compared the metrics directly on the aggregate traffic with genuine attacks at the AS level, using a real network environment. In fact, we collected the traffic flowing through an important Italian Tier II AS, as shown in Figure 1, that plays the role of transit for some stub ASes and shares connections within other ISPs in a IXP (Internet eXchange Point). In order to protect customers privacy, we avoid any references about real IP addresses, identities or related contents in this paper. In our dataset we recorded meaningful high resources network events and several attacks that we used to evaluate and estimate the effectiveness of information theory metrics for (D)DoS attack detection, just using NetFlow data.

Contributions. Starting from the huge dataset of **real** network traffic, in this paper we provide sundry contributions:

- we validate the theoretical research results, applying the most used information theory metrics;
- we propose the use of the above metrics on lightweight dump dataset, requiring no heavy computation or I/O efforts. Indeed, our dataset only contains compact netflow records;
- we report the ability to perform (D)DoS detection by the analysis of "aggregated" network traffic data. We claim and prove that analyzed metrics can be effectively used to detect such attacks in real-time on upstream provider side. This would prevent and mitigate attacks that focus on bandwidth saturation, since network operators could use a single network monitoring point (like a border router);
- we compare the different metrics and evaluate their effectiveness against several uncommon network activities (like (D)DoS attacks and nightly scheduled maintenance jobs), in terms of anomaly detection and robustness.

Organization of this Paper. Section 2 introduces the main representative works on (D)DoS detection; Section 3 describes our network environment and surveys on some meaningful recorded malicious events. Section 4 quickly introduces information theory metrics used for attack detection, while Section 5 presents our results. Section 6 concludes our work and draws some directions for future research on this topic.

2 RELATED WORKS

Detection and mitigation of (D)DoS attacks is still an open challenge (Di Pietro and Mancini, 2008; Curtmola et al., 2005). A systematic analysis of DDoS attacks is presented by (Mirkovic and Reiher, 2004), where the authors define a complete taxonomy of attacks, proposing different criteria such as Exploited Weaknesses, Degree of Automation, Exploited Weakness to Deny Service, Source Address Validity, Possibility of Characterization, Dynamics, Persistence of Agent Set, Victim Type or Impact on the Victim. During the exposition of our results, we usually refer to cited taxonomy. Many results like (Feinstein and Schnackenberg, 2003) and (Oshima et al., 2010) (just to cite a few) agree upon the use of Entropy and Relative Entropy (*information divergence*) as effective metrics for anomaly detection (a formal introduction to those metrics will follow in Section 4). In (Feinstein and Schnackenberg, 2003), the authors analyze several genuine network traces, using blocks of 1000 consecutive packets to compute entropy and frequency-sorted distribution of selected packet attributes. Since the network traces are not known to contain malicious activities, the authors overlay synthetic DDoS attacks at various degree of concentrations. An attack alarm is raised if the computed entropy value overcomes a threshold, pre-determined from empirical analysis.

(Oshima et al., 2010) show that the use of entropy with fixed-dimension block of packets can be very time consuming even for small organizations and imposes a CPU-burning process for big bandwidth network edge. Moreover, they introduce a dynamic threshold evaluation in order to mitigate entropy fluctuations, based on its standard deviation. Others works like (No and Ra, 2009) and (Sardana et al., 2008) improve attack detection using other information theory metrics, like cumulative entropy. Other entropy based metrics are based on the concepts of *information divergence*, such as Rényi (Li et al., 2009b) and Kullback-Leibler divergence (Li et al., 2009a). Their main advantages are the ability to improve the anomaly detection, providing at the same time earlier responses and low false positive rate (Xiang et al., 2011). The previous entropy based approaches, indeed, experience many false positives in case of fluctuations of traffic pattern. We address this and several other aspects in Section 5.

Common issues of related works is the consistency and the nature of used dataset. Almost all the papers refer to datasets that are historically consolidated (like the DARPA dataset) or that have been collected from restricted and unrepresentative traffic ((Feinstein and Schnackenberg, 2003), (Lawniczak et al., 2009)). DARPA dataset was created by the Information Systems Technology Group (IST) of MIT Lincoln Laboratory, under Defense Advanced Research Projects Agency (DARPA) and Air Force Research Laboratory (AFRL/SNHS) sponsorship. The purpose of this dataset is to collect and distribute the first standard corpora for evaluation of computer network intrusion detection systems (IDS (Di Pietro and Mancini, 2008; Di Pietro et al., 2010)). Although its nature, MIT datasets have been used by researchers to evaluate malicious activities, like DoS or DDoS in (Li et al., 2009a), (Oshima et al., 2010), (No and Ra, 2009) and (Sardana et al., 2008). As stated in (Hugh, 2000), the methodology used to generate the data by MIT Lincoln Laboratory and the nature of data itself are not appropriate for simulating different, non academic, network environments. So the experimental results of many works suffer of the above limitation. In general, all the proposed works use synthetic traffic in combination with attack-free network traces, like (Li et al., 2009b) and (Xiang et al., 2011). Such methodology can limit the validity of the research results and lack of generality. In fact, malicious activities artificially injected can miss relevant features that real activities exhibit: this lacking can produce unrealistic behavior and, in the worst case, unreal or unexpected results. The work in (Xiang et al., 2011) is an example of such limitation: in order to enrich simulations dataset, authors generate synthetic traffic and attack traffic using, respectively, Gaussian and Poisson distributions.

Huge and genuine datasets are needed today also to analyze the emergent types of coordinated and DDoS attacks, made by volunteer users, as opposite to the early crackers with bad intention. A recent analysis of high tech criminal threats to national critical infrastructures (Choo, 2010) introduced the concept of "hacktivism", to emphasize the new user role. The authors reported some real cases where citizens were involved to disrupt national infrastructures, "carrying out politically-motivated hacking and bringing down Government agencies' website". The recent Operation Payback is actually a proof of this statement. Again, in the case of historically consolidated datasets (Oshima et al., 2010), it is easy to notice that they are too old to represent recent (D)DoS attack under the hacktivists or zombies.

One recent study that used NetFlow technology to perform DDoS detection is (Sekar and Merwe, 2006), that proposes a multi-layer approach that combines several steps on sampled netflows. In this work, again, synthetic attacks were introduced in the real traffic gathered from a tier I ISP, in order to simulate DDoS attack. Our work, instead, is based on completely genuine traffic with known and real traffic anomalies that we were able to analyze with our metrics.

3 NETWORK ENVIRONMENT

In this section we depict the network used as case study and provide a contextualization of (D)DoS attacks. In detail, we present the network architecture, the netflow collector's position and several technical and statistical information related to the observed attacks.

3.1 Monitored Network

We remind that with the respect of the Non-Disclosure-Agreement of the *ExTrABIRE* project, no detailed information about AS (such as AS name or number) nor ISP interconnections will be provided to preserve AS and host privacy.

The monitored network is schematically summarized by AS1 in Figure 1. It is a *multihomed* Autonomous System, namely an AS that maintains several connections to more than one other ASes. AS1 provides several services, hosting web and mail servers publicly reachable and sundry x-DSL connections. AS1 uses a connection toward a Tier2 AS

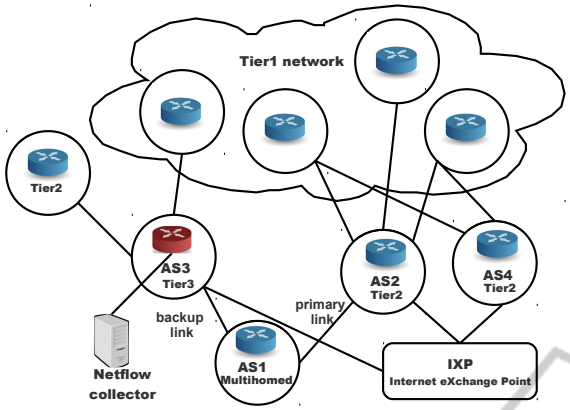

Figure 1: Network architecture.

(AS2) as upstream provider and has a secondary link (backup) with a Tier3 AS (AS3). AS2 has three high speed optical fiber links with three Tier1 AS, while AS3 receives connectivity toward Internet from a Tier2 and a Tier1 AS. Both AS2 and AS3 exchange their traffic inside an IXP (Internet eXchange Point), with several other national and international ISPs.

Since our AS is composed by heterogeneous networks and services, we state that it can be considered as a good testing case for our research activity; furthermore, it is general enough to represent many other real contexts. To show the dimension of our AS, Table 1 reports the average of the exchanged traffic. Considering that it manages thousands of unique IP addresses and that the amount of packets is around 30K every second, the monitored AS can be considered of medium size for an European member state. Despite previous studies which

Table 1: Network traffic characterization (IN/OUT).

Time of the day	Flows/s	Packets/s	Mbit/s
00:00AM-11:00AM	377/312	8.8K/6.4K	37/44
11:00AM-06:00PM	1.3K/930	21K/13K	113/54
06:00PM-11:59PM	764/575	14K/ 8K	80/27

used synthetic datasets, we use genuine traffic data and a real network setting to evaluate the considered metrics. Moreover, many works about anomaly detection over synthetic datasets report conflicting results. For instance, (Li et al., 2009a) obtained the best results with the Rényi metric while others, (Xiang et al., 2011), proposed the use of the Kullback-Leibler metric. For this reason, we choose to avoid direct comparisons with results of any synthetic dataset, and we focused only on our collected material.

3.2 Netflows Dataset

NetFlow is a Cisco^{TM} technology used for monitoring

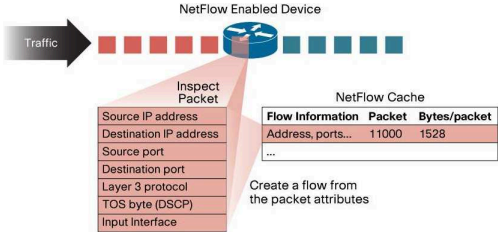

Figure 2: How netflow creates flow aggregates. This image is taken from Introduction to Cisco^{TM} IOS NetFlow.

IP traffic (Cisco Systems, 2004). Despite the classical packet collector (packet dump), NetFlow collects data in Layers 2-4 and determines applications by port numbers, aggregating the information.
NetFlow efficiently monitors a network, enabling services like traffic accounting, usage-based network billing, network planning, as well as Denial of Services monitoring. Netflow records are extremely compacts and representative, avoiding to maintain the packet's payload and making analysis and computation lighter. While several kind of attacks crafted in the traffic payload are able to circumvent the detection filter, traffic anomalies are still effectively observable. NetFlow is therefore recognized as a network monitoring tool: many research papers as well as professional software use this tool as source of data to query network status or get back data log.

The typical configuration to leverage the NetFlow protocol is made by a router with netflow capabilities and a probe (*netflow collector*) able to store received data (see Figure 1). Netflow records are sent as a *UDP* stream of bytes. A netflow-enabled router creates one record with only selected fields from the TCP headers of *each* transiting connection (Figure 2): a single netflow record is a *unidirectional* sequence of packets all sharing the 7 values source and destination IP addresses, source and destination ports (for UDP or TCP, 0 for other protocols), IP protocol, Ingress interface (SNMP ifIndex) and IP Type of Service. Other valuable information associated to the flow, like timestamp, duration, number of packets and transmitted bytes are also recorded. Then, we can consider a single flow as a record that represents the data exchanged between two hosts only in one direction, since it aggregates all the IP packets that composed a single communication session. Indeed, a single TCP connection is represented by two distinct flows in opposite directions, despite the number of IP packets or the number of exchanged bytes.

A netflow-enabled router sends to the probe a single flow as soon as the relative connection expires. This can happen when 1) when TCP connection reaches the end of the byte stream (FIN flag or RST flag) are set; 2) when a flow is idle for a specific

timeout; 3) if a connection exceeds long live terms (30 minutes by default).

The use of NetFlow technology has several advantages with respect to the raw packet sniffing, since using just few information it is able to give a lightweight picture of monitored network. Several researchers proposed use netflow collectors as IDSs (Intrusion Detection Systems), traffic classifiers as well as a specific security tools ((Chan et al., 2008), (Dübendorfer et al., 2005)). To have a flavor of the advantages of a netflow dataset in terms of dimension and required effort, we can consider this simple statistic of our data set: a 2 GBytes of full netflow entries contains 110 millions of flows, 2 billions of packets and about $1,5$ TByte of exchanged data, corresponding to the data gathered in one single day. Until today, our dataset consists in a collection of 12 months long netflow records (about 900 GBytes) and keeps growing.

We note that, despite the advantage of low computational requirements to process an extensive amount of data, netflows inevitably sacrifice many valuable information related to traffic payload: carried attacks to single host services, virus or malware specific signatures, particular malformed packets and so on are dropped and cannot be recovered from netflows. Clearly comparison between detection effectiveness of NetFlow and Deep Packet Inspection is a very interesting problem that deserves a full study all its own. In our particular case, due to some privacy issues, we could not access to the whole payload to perform a full packet inspection. In fact, we monitor several governmental networks that exchange sensitive data and any access to their packet payloads is restricted. Furthermore, the probe used in our project is not enough powerful to handle the huge volume of traffic generated by the monitored networks. The above points prevented us to perform a full traffic inspection and comparison with the collected netflows across the 12 months of monitoring. Our work is based on real events and we cannot be aware of them in advance, then it is impossible to recover full network traffic information of the attack through an a posteriori analysis.

Comparing a synthetic dataset with our real dataset in order to have an estimation of false positive or false negative events raises many difficulties. First of all, we should build a representative and huge synthetic dataset similar to ours, but that is out of the scope of this work. Moreover, there would not be any guarantee that the synthetic dataset has the same traffic behavior than our veritable dataset. Finally, we probably would still remain without any real knowledge of the actual attacks our dataset contains.

4 ENTROPY AND RELATIVE ENTROPY METRICS

The use of entropy analysis aims to capture fine-grained patterns in traffic distributions that simple volume based metrics cannot identify. Interestingly, information theory based metrics enable sophisticated anomaly detections directly with the whole traffic that are difficult to provide with simpler metrics, like aggregated traffic workload, number of packets or single host traffic. As it will be described in the next sections, the events detected by combinatorial metrics are not really predominant when observed with the traditional ways: within the aggregate traffic of our ISP (order of 1Gbit/s), a (D)DOS attack against a single VLAN is not noticeable, since the Mbits needed to perform a (D)DOS attack are well-hidden in the aggregate traffic, and would not determine any apparent anomaly. On the other hand, the Kullback-Leibler divergence is effectively able to notice the anomaly and to raise an alarm. To provide the same level of accuracy, any traditional metric should be continuously evaluated on every possible target, in order to detect the anomalies. Combinatorial metrics, instead, are able to detect the anomalies within the whole traffic; moreover, those are less affected by the fluctuations of traffic workload or any any other quantitative measure.

The first metric we studied is the *simple entropy*, that captures the degree of dispersal/concentration of a distribution. Then, we also considered two relative entropy measures, namely the *Kullback Leibler* divergence (Li et al., 2009a) and *Rényi* divergence (Li et al., 2009b).

The concept of *Entropy* was introduced by Shannon in (Shannon, 1948). The classic definition says that entropy is a measure of the *uncertainty associated with a random variable*. The entropy $H(X)$ of a discrete random variable X is defined as:

$$H(X) = -\sum_i p_i \log_2 p_i \qquad (1)$$

where $p_i = P[X = i]$ is the probability that X assumes the value i.

Relative entropy (also known as *information divergence*) is a non-symmetric measure of the similarity between two probability distributions P and Q and quantifies the distance between two statistical objects. The Kullback-Leibler divergence equation (Li et al., 2009a) we used is:

$$D_{KL}(P||Q) = \sum_i P(i) log \frac{P(i)}{Q(i)} \qquad (2)$$

A low D_{KL} value means a high similarity in the two probability distributions, on the other hand, high di-

vergence values correspond to low similarity. Since it is not symmetric, the divergence measure can not be strictly considered a *metric* (i.e. $D_{KL}(P||Q) \neq D_{KL}(Q||P)$).

The Rényi divergence generalizes the Kullback-Leibler divergence, providing a family of metrics based on a parameter α. Formally, we use (Li et al., 2009b):

$$D_\alpha(P||Q) = \frac{1}{1-\alpha} \log \sum_i \frac{p_i^\alpha}{q_i^{\alpha-1}} \quad (3)$$

Notice that $D_{a\to 1}(P||Q) = D_{KL}(P||Q)$. Intuitively, Rényi divergence with high values of α takes in higher account the more likely events, while with low values of α it considers more equally all the events, regardless of their likelihood.

4.1 Metrics Implementation

In this section we describe how we exploit netflow data to implement the mentioned metrics. As stated by (Nychis et al., 2008), port and address IP distributions are highly correlated in network traffic. For this reason we only considered source and destination IP.

Network flows are aggregated into time blocks of a fixed size (1 minute by default). Let f^t be the number of flows that cross the monitored network in a time block. Let f_i^t be the number of flows that have IP_i as source (or destination) address. For each time block t, the entropy is evaluated by the following formula:

$$H(X) = -\sum_{\forall distinct\ IP\ i} \frac{f_i^t}{f^t} \log_2 \frac{f_i^t}{f^t} \quad (4)$$

Concerning the relative entropy metrics (Kullback-Leibler and Rényi), we associate p_i to the packet distribution over a time block t and q_i to the packet distribution of the previous time block $t-1$:

$$p_i = \frac{f_i^t}{f^t}, q_i = \frac{f_i^{t-1}}{f^{t-1}}$$

We compute Kullback-Leibler divergence as follows:

$$D_{KL}(t||t+1) = \sum_i \frac{f_i^t}{f^t} \log \frac{\frac{f_i^t}{f^t}}{\frac{f_i^{t-1}}{f^{t-1}}} = \sum_i \frac{f_i^t}{f^t} \log \frac{f_i^t f^{t-1}}{f^t f_i^{t-1}}$$

and Rényi divergence as:

$$D_\alpha(P||Q) = \frac{1}{1-\alpha} \log \left(\sum_i \frac{(\frac{f_i^t}{f^t})^\alpha}{(\frac{f_i^{t-1}}{f^{t-1}})^{\alpha-1}} \right)$$

Notice that we only considered the entries that appear in both t and $t-1$ time blocks, since the relative entropy imposes $Q(i) > 0$ for each $P(i) > 0$. Another key aspect is the choice of the parameter α in the Rényi divergence. According to the results of (Li et al., 2009b), in our experiments we set the value of α to 5. The time block dimension affects the relationship between detection *reactivity* and detection *sensibility*, directly influencing the results. With empirical analysis we find that 1 minute is a good compromise between reactivity and sensibility. For a matter of space, we do not discuss here the results obtained using different time block dimensions.

5 ATTACK AND ANOMALY ANALYSIS

In this section we report the comparison of the three metrics presented in the above section, applied to several anomalies collected in our data set. Remember that our netflow dataset refers to the period between September 2010 and August 2011, that has been the scenario for several (D)DoS episodes, in Italy and abroad. In order to make a complete and fair comparison between Entropy, Kullback-Leibler and Rényi metrics with the previous research results, we evaluated the above metrics considering separately the destination and the source *IP* distributions.

Our experiments considered the whole dataset of netflows and are based on the evaluation of the three metrics for all the 12 months. [2]. However, since we did not have a complete knowledge of the attacks present during the whole period, we only considered as attacks the official report of the AS administrators. We believe that the detailed analysis of these events and the comparison of the different metrics is the most relevant contribution of our paper. As depicted in the next figures, the reported anomalies correspond to metric fluctuations that produced peaks in their values. Once such high peaks were identified, we conducted a deeper inspection in order to capture the motivations behind the anomaly. This kind of analysis produced several insights about the behaviors and limitations of the metrics.

In addition to the reported anomalies, we chose to analyze another kind of network activity, namely the abnormal traffic generated by scheduled and automated administration activities (e.g. scheduled backups or maintenance procedures). Since those activities can make sensible service outage to users, they are programmed in the period that spans from 12:00am to 8:00am, when regular traffic is low and the connection flows reach their minimum. By observing the relative netflows is possible to identify

[2]Please refer to the website www.extrabire.eu for the complete result sets

sudden and relatively short mutations of the traffic pattern, resulting in a deep alteration of the metrics.

5.1 Sample Events

We report the results of four sample events (E_1, E_2,E_3, E_4): Table 2 summarizes how the implemented metrics (Entropy H, Kullback Leibler KL and Rényi R) reacted during these events.
The three metrics are evaluated both on source and destination *IP* with the exception of the Rényi divergence which is evaluated only on destination *IP*. We chose to not report the Rényi on source address due to its extremely fuzzy behavior. With the shortened form H_s, KL_s and H_d, KL_d, R_d we refer to entropy, Kullback Leibler and Rényi respectively evaluated on source and destination *IP* distributions.

For each event (the rows of table 2) we report the behavior of all metrics (the columns of table 2) with the following notation: if we observe an abrupt variation to a higher value we say that the metric *Increases*; otherwise if we observe an abrupt variation to a smaller value we say that the metric *Decreases*. Finally we indicate the absence of observable variation with *Unvaried*.

Table 2: Events and metrics reactions.

	E_1 (DoS)	E_2 (DoS)	E_3 (DDoS)	E_4 (Routines)
KL_d	*Increases*	*Increases*	*Increases*	*Unvaried*
H_d	*Decreases*	*Decreases*	*Decreases*	*Increases*
R_d	*Unvaried*	*Unvaried*	*Unvaried*	*Unvaried*
KL_s	*Unvaried*	*Increases*	*Increases*	*Unvaried*
H_s	*Decreases*	*Decreases*	*Increases*	*Decreases*

All the comparison charts we show in this section are composed by all the metric results. Actually, in order to avoid metrics overlapping, we plot Entropy and Kullback-Leibler results with the source IP mirrored with respect to the x axis.

E_1 - DoS Attack. This episode has been classified as a DoS attack. In fact, we made a brief statistic to show which IP generated the highest number of flows directed to our network and we found that there was a single *IP* playing a primary role during the attack against one single host server. There were also few other *IP* addresses participating to the attack, with a smaller contribution. All the metrics correctly detected the malicious activity, as shown by the fluctuations and the spikes of Figure 3. Indeed, Rényi distribution shows the lower peak. The DoS nature of the attack is well described by the downfall of both entropy lines (in the lower part of the plot) around

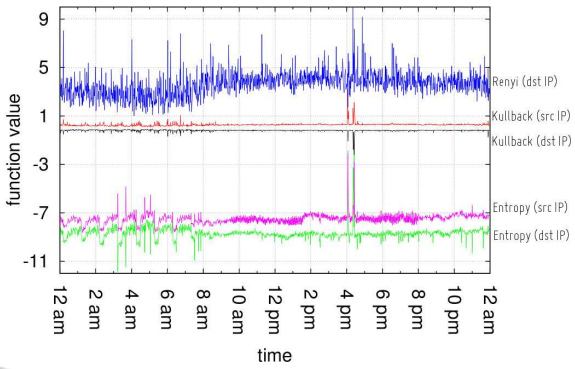

Figure 3: E_1 — metrics comparison.

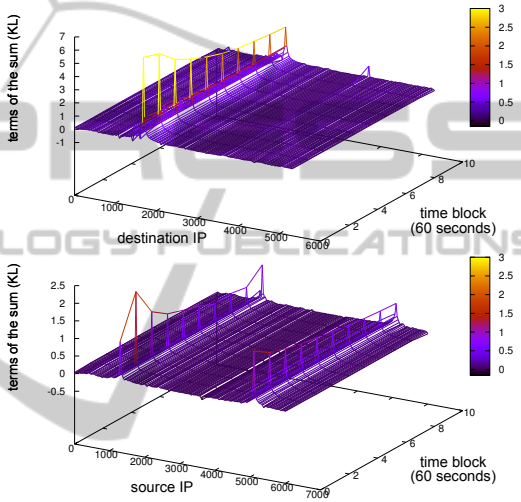

Figure 4: E_1 — Kullback-Leibler details on destination (upper plot) and source (lower plot) IP address.

4:00pm. This behavior expresses that a small number of source addresses generates the largest amount of connections towards a small set of destinations, namely the typical scenario of a DoS attack. At the same time, KL on destination *IP* grows significantly. To have a deeper insight of KL behaviors, we graphically report the contribution of every destination *IP* to the final KL_s and KL_d values. In Upper Figure 4 we report the first ten time-blocks since the beginning of the malicious activity. It is evident how the final value of KL_d is obtained by the contribution of one main component (the victim host 2000), while the contribution of the other hosts is negligible. On the other hand, since during a DoS attack, only few sources generate the largest part of the traffic, each attacking host addresses many flows towards the victim host. Sampling the network traffic we will see that the attacking *IPs* are the most frequent among the source addresses. This anomaly is perfectly captured by the peaks of lower Figure 4, that corresponds to the main contributors to the KL_s value.

In the same plot of Figure 3 is possible to observe the anomaly that we introduced as administration activity (maintenance jobs) and that we labeled with E_4: before *7:00AM* indeed both source and destination entropy metrics rise and fall continuously, since they generate maximum (respectively minimum) traffic compared to high (respectively low) traffic. More details will be provided in Section 5.1.

E_2 - DoS Attack. In this episode, the main contribution to the attack came from a single *IP* and consists in more than 50% of all the flows towards one single victim host; moreover, the five most active *IP*s have generated the 93.8% of the whole traffic.
In this event the victim host does not involve a large portion of network flows, while several other services of the networks (web, mail, DNS servers etc.) generated the larger amount of flows. Nevertheless, the traffic diversity expressed when the attack occurred has been well detected by both *KL* measures. This aspect represents a scalability factor of this measure and suggests that the attack is detectable among the whole aggregated traffic: that is, the attack emerges from the traffic thanks to its informational fingerprint. As entropy line shows (Figure 5), the attack starts soon after *1:00PM*. The entropy on both attributes decreases,

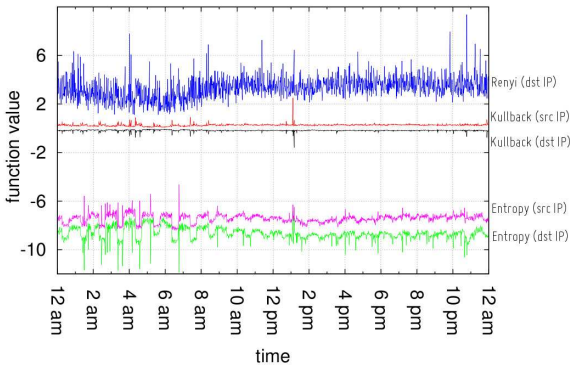

Figure 5: E_2 — metrics evaluations.

representing a non-uniform distribution of destination IP as well as source IP fields. The Rényi distribution reveals a small peak, but this is hidden by the fuzzy behavior it exhibits.
Even in this case it is possible to observe the perturbations due to the maintenance jobs: in the case of the entropy, the peaks are higher than the ones relative to the attack E_2, generating some false positive (as it will be clear in the following). Rényi divergence also suffers the same issue. In this particular DoS attack, the intensity of malicious traffic is significantly lower than E_1, making the detection more difficult. Indeed, the entropy peaks associated to the attack are not really evident since they are lower than the false positive

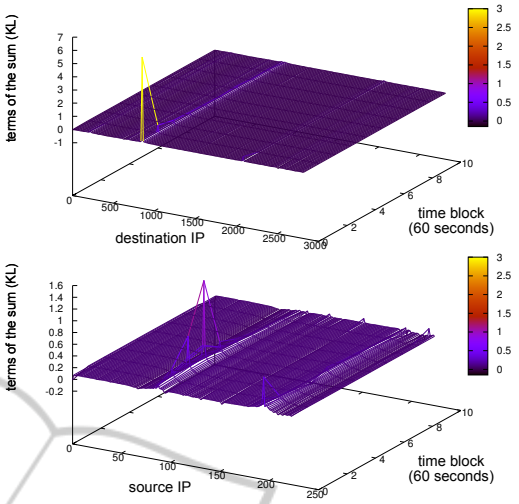

Figure 6: Kullback Leibler details on destination (upper plot) and source (lower plot) IP address.

of the early morning; nevertheless, the *KL* is still able to detect the anomaly. Again, the deeper representations of the *KL* contributors at the time of the attack (Figure 6) show how this metric correctly reveals the attack and characterizes it as a DoS.

E_3 - DDoS Attack. In this case we describe a Distributed DoS attack, characterized by a large number of attack sources. In this event the most active host generates only the 0.5% of the traffic flows. This kind of attack is really different from E_1 of Section 5.1, where the most active *IP* addresses more than half of total flows. Figure 7 reports the metric behavior. As in E_1, Rényi distribution seems to generate several peaks associated to a non-attack instances. The most significant example can be found around *6:00PM*. The attack started soon after *10:00PM*: both entropy metrics reveal the event and catch its DDoS nature. The abrupt growth of source *IP* entropy line suggests that there was a great amount of diversity in this field. The peak of destination *IP* entropy rep-

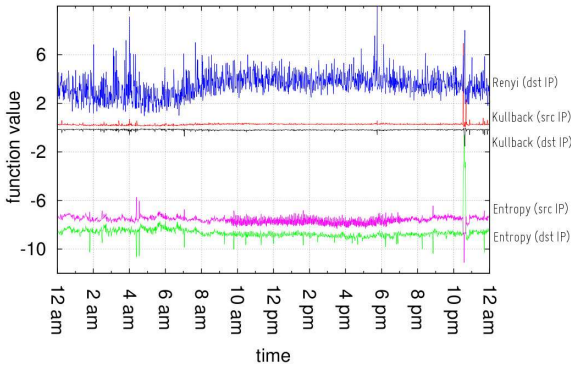

Figure 7: E_3 — metrics evaluations.

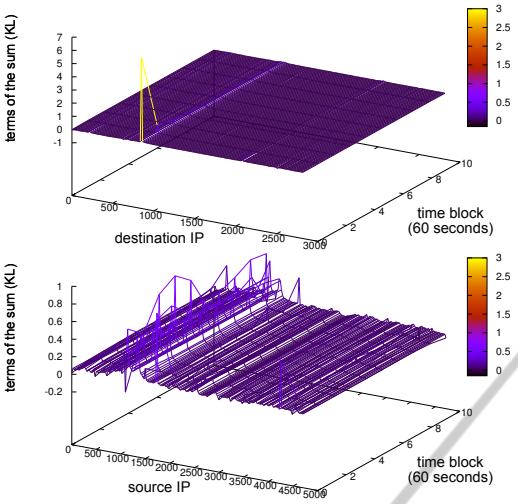

Figure 8: Kullback Leibler details on destination (upper plot) and source (lower plot) IP address.

resents that there is an anomalous variation in the connected endpoints. Attack dynamic is represented in upper Figure 8, where the roughness and quickness of the malicious event causes a jump of the *KL* value. The presence of several new entities drawn by the DDoS attack induces a continuous variation in the source *IP* distribution (see lower Figure 8) and, then, causes the *KL* to fluctuate constantly. As opposite to previous cases, the plot shows that the variation of the *KL* metric is caused by multiple components, that contribute to its final value.

E_4 - Maintenance Jobs. In order to explore how entropy metrics are prone to false positive (see introduction of Section 5.1), we perform a deep analysis of maintenance job events. These events are common to all networks and consist in backup activities scheduled during the early hours of each days, aimed to reduce host workload and service degradation. Likewise other cases, we plot in Figure 9 a graph showing how each *IP* contributes to the final *KL* value. The component's order of magnitude is clearly smaller than the other *KL* detailed graphs. *KL* values, as well as the values of entropy metrics, are sensible to traffic variation. Since the entropy metrics sense destinations (respectively sources) *IP* address distribution diversity, they notice a lacks of regularity in the traffic flows and increase their values. On the opposite, *KL* values warns distributions divergence, but the low level of traffic activity attenuates the final result, keeping the value of the metric under below suspicious value.

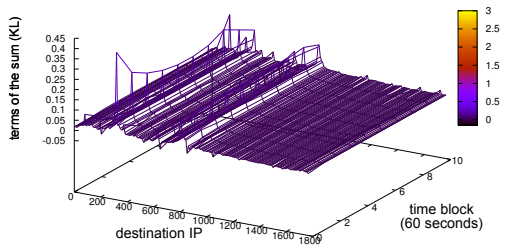

Figure 9: E_4 — metrics evaluations.

5.2 Metrics Comparison

Our experiments show that all the aforementioned metrics are able to detect the traffic alterations but some of them are prone to a high false positive rate. In particular we observed that the Rényi is the more unstable metric: it exhibits many spikes during the whole analysis, making very unsuitable its use for (D)DoS anomaly detection with netflows. Similarly, the Entropy metric shows many fluctuations, making difficult to find a feature related to (D)DoS attacks. The Kullback-Leibler (KL), instead, appear to have the more stable trend, showing evident spikes only during the attack events. According to other studies like (Xiang et al., 2011), our analysis also suggests that the *KL* is the most suitable information theory metric to detect (D)DoS.

We can also observe the difficulty to define a threshold value that could be used to determine if a spike corresponds to an attack or not. Threshold selection is easier with simple metrics (like packet number or packet size), but it seems much harder with statistical metrics (Chang et al., 2006; Sardana et al., 2008). Both the Entropy and Rényi metrics show very unstable values and, in coincidence with some known attacks (like E_2 and E_3), exhibit lower values than the ones obtained during the regular traffic. In our analysis *KL* metric assume a value greater than 1 during the known attacks, suggesting a possible *empirical* threshold. However, the evaluation of *KL* threshold is a subtle argument and is out of the scope of this work.

6 CONCLUSIONS AND FUTURE WORKS

In this paper we reported our study on real and huge netflow data set, to efficiently detect (Distributed) Denial of Services attacks at Autonomous Systems level. We were able to compare and evaluate the main information theory metrics proposed in the literature, bringing several insights. We show that malicious activities can be detected in aggregated traffic, making the attacks to emerge from the whole set of aggregated

netflows. We observed that the Kullback-Leibler metric seems to be the best suited to analyze huge amount of traffic, since it has been able to detect DoS and DDoS activity, maintaining a low level of false positives.

An interesting challenge is the formal definition of a threshold value, whose correctness distinguish legitimate and malicious activities. In the future we plan to release an obfuscated version of our dataset providing the community with a common ground, where the proposed solutions can be fairly compared.

REFERENCES

Chan, Y.-T. F., Shoniregun, C. A., and Akmayeva, G. A. (2008). A netflow based internet-worm detecting system in large network. In Pichappan, P. and Abraham, A., editors, *ICDIM*, pages 581–586. IEEE.

Chang, C. I., Du, Y., Wang, J., Guo, S. M., and Thouin, P. D. (2006). Survey and comparative analysis of entropy and relative entropy thresholding techniques. *Vision, Image and Signal Processing, IEE Proceedings -*, 153(6):837–850.

Choo, K.-K. R. (2010). High tech criminal threats to the national information infrastructure. *Inf. Secur. Tech. Rep.*, 15:104–111.

Cisco Systems (2004). Cisco Systems NetFlow Services Export Version 9. rfc3954.

Cisco Systems (2010). Cisco 2010 Annual Security Report, Highlighting global security threats and trends. http://www.cisco.com/en/US/prod/vpndevc/annual_security_report.html.

Curtmola, R., Sorbo, A. D., Ateniese, G., and Del, A. (2005). On the performance and analysis of dns security extensions. In *in Proceedings of CANS*, pages 288–303. SpringerVerlag.

Di Pietro, R. and Mancini, L. V. (2008). *Intrusion Detection Systems*. Springer-Verlag.

Di Pietro, R., Oligeri, G., Soriente, C., and Tsudik, G. (2010). Intrusion-Resilience in Mobile Unattended WSNs. In *INFOCOM*, pages 2303–2311. IEEE.

Dübendorfer, T., Wagner, A., and Plattner, B. (2005). A framework for real-time worm attack detection and backbone monitoring. In *IWCIP 2005*.

Feinstein, L. and Schnackenberg, D. (2003). Statistical approaches to DDOS attack detection and response. In *In Proceedings of the DARPA Information Survivability Conference and Exposition*, pages 303–314.

Hugh, J. M. (2000). Testing intrusion detection systems: a critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by lincoln laboratory. *ACM Trans. Inf. Syst. Secur.*, 3:262–294.

Lawniczak, A. T., Di Stefano, B. N., and Wu, H. (2009). Detection & study of DDoS attacks via entropy in data network models. CISDA'09, pages 59–66, Piscataway, NJ, USA. IEEE Press.

Li, K., Zhou, W., and Yu, S. (2009a). Effective metric for detecting distributed denial-of-service attacks based on information divergence. *IET Communications*, 3(12):1851–1860.

Li, K., Zhou, W., Yu, S., and Dai, B. (2009b). Effective DDOS attacks detection using generalized entropy metric. ICA3PP '09, pages 266–280, Berlin, Heidelberg. Springer-Verlag.

Mirkovic, J. and Reiher, P. (2004). A taxonomy of DDOS attack and DDOS defense mechanisms. *SIGCOMM Comput. Commun. Rev.*, 34:39–53.

No, G. and Ra, I. (2009). An efficient and reliable DDOS attack detection using a fast entropy computation method. ISCIT'09, pages 1223–1228, Piscataway, NJ, USA. IEEE Press.

Nychis, G., Sekar, V., Andersen, D. G., Kim, H., and Zhang, H. (2008). An empirical evaluation of entropy-based traffic anomaly detection. IMC '08, pages 151–156, New York, NY, USA. ACM.

Oshima, S., Nakashima, T., and Sueyoshi, T. (2010). DDoS detection technique using statistical analysis to generate quick response time. BWCCA '10, pages 672–677, Washington, DC, USA. IEEE Computer Society.

Sardana, A., Joshi, R., and Kim, T.-h. (2008). Deciding optimal entropic thresholds to calibrate the detection mechanism for variable rate DDOS attacks in ISP domain. In *ISA*, pages 270–275, Washington, DC, USA. IEEE Computer Society.

Sekar, V. and Merwe, J. V. D. (2006). Lads: Large-scale automated ddos detection system. In *In Proc. of USENIX ATC*, pages 171–184.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27:379–423.

Xiang, Y., Li, K., and Zhou, W. (2011). Low-rate DDOS attacks detection and traceback by using new information metrics. In *Information Forensics and Security, IEEE Transactions*, volume 99. IEEE Press.