Automated Classification of Network Traffic Anomalies

Guilherme Fernandes and Philippe F. Owezarski

LAAS - CNRS
Université de Toulouse
7 Avenue du Colonel Roche
31077 Toulouse, France
owe@laas.fr

Abstract. Network traffic anomalies detection and characterization has been a hot topic of research for many years. Although the field is very advanced in the detection of network traffic anomalies, accurate automated classification is still a very challenging and unmet problem. This paper presents a new algorithm for automated classification of network traffic anomalies. The algorithm relies on three steps: (i) after an anomaly has been detected, identify all (or most) related packets or flow records; (ii) use these packets or flow records to derive several distinct metrics directly related to the anomaly; and (iii) classify the anomaly using these metrics in a signature-based approach. We show how this approach can act as a filter to reduce the false positive rate of detection algorithms, while providing network operators with (additional) valuable information about detected anomalies. We validate our algorithm on two different datasets: the METROSEC project database and the MAWI traffic repository.

1 Introduction

The Internet has greatly grown in complexity, changing from a single best effort service to a multi-services network that is ever more demanding of guaranteed quality of service (QoS). Network traffic anomalies can seriously impact or disrupt the normal operation of networks. It is then vital that their identification and mitigation be quickly done by network administrators. A specific type, volume anomalies, is responsible for unusual modifications on network traffic volume characteristics (normally identified on the #packets, #bytes and/or #new flows). These anomalies can be caused by a myriad of events: from physical or technical network problems (e.g. outages, routers misconfiguration), to intentionally malicious behavior (e.g. denial-of-service attacks, worms related traffic), to abrupt changes caused by legitimate traffic (e.g. flash crowds, alpha flows). This diversity coupled with the great (natural) variability of normal Internet traffic volume [16], makes the identification and mitigation of these anomalies a very challenging task.

Despite these difficulties, constant progress has been made in network traffic anomaly detection. Methods have been created to detect anomalies in single-links and network-wide data, and techniques have been used to cope with the

Y. Chen et al. (Eds.): SecureComm 2009, LNICST 19, pp. 91–100, 2009.

high dimensionality of network traffic data (e.g. sketches [13][4] and principal components [11][12]). Algorithms for network traffic anomaly detection have evolved from only being able to signal an anomaly in time (e.g. [1][17]) to providing information about the actual flows that cause the anomaly [13][4]. This information is very valuable for network administrators that need to manually verify and mitigate potential anomalies, but is still not enough. Because of the characteristics of network traffic and the frequency of anomalies, it is not feasible to manually analyze (in real-time) all anomalies detected by state-of-the-art detection algorithms. Network operators need more information than just the anomalous flows to efficiently prioritize between detected anomalies.

Although there has been some effort to characterize network traffic anomalies, automated classification has not received much attention (a notable exception is [12]). Automated classification intends to add meaningful information to the alert of a detected anomaly. Ideally, the computed information can then be used to define the type of the anomaly or to at least help characterize the underlying cause. In this paper, we present a new algorithm for automated classification of network traffic anomalies. We show how the information obtained by further analyzing the identified anomalous flows can be used in a signature-based classification module to reliably characterize different types of anomalies (e.g. DDoS, network scans, attack responses). We also show how this approach provides the flexibility needed by network operators to understand and manipulate the classification process. We do a statistical validation for the automated classification of DDoS anomalies and discuss results obtained for other type of anomalies using two different datasets: the METROSEC project (see http://www.laas.fr/METROSEC) database and the MAWI traffic repository [2].

2 Related Work

The evolution of detection algorithms (see Section 1) has been followed by several studies on the characterization of network traffic anomalies. Barford et al. [1] used a wavelet-based signal analysis on single-link volume data to characterize four classes of network anomalies: outages, flash crowds, attacks and measurement failures. Lakhina et al. used the subspace method to characterize several types of network-wide anomalies based on traffic volume metrics [11] and on traffic features [12]. Prior work has also been directed to individual types of anomalies. For example, DoS and distributed DoS (DDoS) attacks received an in-depth analysis in [15][8][14]. Jung et al. [9] studied the differences on DDoS and flash crowds behavior from a web server perspective. We thoroughly use the knowledge of such previous work to convey different attributes of the traffic anomalies that are used by our classification module to reliably label the anomaly.

Previous work has proposed ways to (automatically) convey more information about network traffic (e.g. by creating and labeling clusters [12][5]) and to provide prioritization (e.g. by using heuristics such as unexpectedness [5]). Specific to network traffic anomalies, the unsupervised approach of [12] creates

clusters based on how anomalies are represented in the entropy space of their traffic features (i.e. IP addresses and ports). Although all anomalies that belong to a specific cluster share a given characteristic, this approach is clearly not enough to uniquely classify an anomaly (as shown by their results). Closer to our work, Kim et al. [10] study how different types of DoS attacks and port scans behave, creating rules to detect and classify them based either on flow header information or on statistical analysis of the flow traffic. Our algorithm aims at general automated classification of network traffic anomalies which are *just being detected*.

3 Anomaly Classification

Our algorithm defines three steps for anomaly classification: (i) after an anomaly has been detected, identify all (or most) related packets or flow; (ii) use these packets or flow records to derive several distinct metrics directly related to the anomaly; and (iii) classify the anomaly using these metrics with a signature-based approach. These steps are based on the fact that much information is needed to reliably classify different types of anomalies and even to distinguish between subtypes, like the many types of DoS attacks. Since current detection algorithms are based on few parameters (i.e. traffic volume metrics or traffic features like IP address and ports), steps are necessary to obtain more information about the anomaly. Naturally, the best source of information are records on the packets or flows that actually cause the anomaly. From now on, we will refer only to packets traces, but similar results can be obtained using flow records.

To test our classification algorithm, we use a variation of the simple traffic volume anomaly detection algorithm presented on [6]. The detection algorithm can be explained as follows. Given a trace of duration T and a time-scale granularity of Δ (i.e. 30s throughout this paper), divide the trace in N slots where $N \in [1, T/\Delta]$. For each slot i obtain the data time series X of each traffic volume metric $\in \{$#packets, #bytes, #syn$\}$. Obtain the *absolute deltoids* [3] P of X and calculate their standard deviation σ_p. For any p_i over the threshold $K * \sigma_p$, mark its slot as anomalous. Using the deltoids of the data time series is important to consider the variation over the amplitude of the curve instead of the variation of network traffic, as the latter is insignificant due to its natural high variability. Our choice of metrics is based on [11] (with #syn instead of #new flows), but the algorithm permits the use of any other data time series.

Detection of low intensity anomalies is important especially for DDoS anomalies [16] and for anomalies in highly aggregated traffic. To detect low intensity anomalies, we apply the detection algorithm to different aggregation levels at the same time. Aggregation is done based on destination IP address and a bit mask modifier for each packet. In this paper we use the following prefix sizes as aggregation levels /0 (i.e. whole traffic), /8, /16 and /24. As with any other detection algorithm, this increase in sensitivity generates a higher rate of false positives (i.e. normal traffic variations are considered anomalous). With the multi-level feature, the algorithm presented above is particularly sensitive to infrequent

communications where only a few packets are seen for a given network/mask aggregation. Although this would generally make the algorithm unusable, we show how the classification process can be used as a filter to greatly reduce the number of false positives. The simplicity of the detection algorithm makes the next step (i.e. identification of corresponding packets and derivation of metrics) a straightforward task and permitted us to concentrate on the characterization of the anomalies.

3.1 Gathering Information

With the characterization of network traffic anomalies done in previous work [1][11][12], we see that different types of anomalies can affect volume metrics and traffic features, such as IP addresses and ports, in the same manner. This clearly shows that we cannot do reliable classification based only on these metrics, and further information needs to be identified. We then introduce the notion of anomaly *attributes*. An attribute is a feature that helps to characterize a specific anomaly (see Table 1). The classification module uses signatures based on attributes derived directly from the packets that compose the anomaly.

The detection algorithm that we use in this work makes it straightforward to get these packets. A detected anomaly is identified by its slot, network address and mask. We also know exactly why it was considered anomalous (i.e. the deltoid for one or more of the volume metrics was above the threshold). Using this information we then proceed to read all the packets in the corresponding slot that are destined to that network, so that we can find the responsible destination hosts (i.e. IP address/32). Our idea of responsible destinations is similar to the notion of dominant IP address range and/or port of [11]. In our algorithm, the set of responsible destinations is composed of all the destination hosts that appear in any of the possible combinations of minimum sets that would bring the anomaly's corresponding deltoid below a fraction of the original threshold. After identifying these hosts, we follow an equivalent approach to determine the responsible sources, ports and protocols. This notion could also be applied to any other traffic feature. Potentially, finding the packets (or flows) that compose an anomaly can be done with any detection algorithm that identifies the starting time and anomalous flows of the anomalies (e.g. [13][4]).

During the anomaly detection and responsible flows identification phases we compute the attributes shown in Table 1. Attributes *found* and *impactlevel* are specific to the detection algorithm we use in this work, but similar attributes should be available for other detection algorithms. The rest of the attributes are derived while identifying the responsible flows. This list is by no means absolute and can be extended. These attributes were the ones we identified as useful during this work and are justified in Section 3.2.

3.2 Classification

General Idea. The main objective of our algorithm is to automatically label network traffic anomalies while they are being detected. The vast number of

Table 1. Attributes derived from a given anomaly. *p*, *b* and *s* are for packets, bytes and syn respectively.

Attribute	Description
found{p,b,s}	If metric was anomalous, value of P, zero otherwise.
impactlevel{p,b,s}	# of anomalous parent aggregation levels due to this anomaly.
#respdest	Number of responsible destinations.
#rsrc/#rdst	Ratio of responsible sources to responsible destinations.
avg#rdstports	Average number of responsible destination/source ports.
avg#rsrcports	Average number of responsible source ports.
#rpkt/#rdstport	Ratio of number of packets to responsible destination ports.
#rpkt/#rsrc	Average number of packets of responsible sources.
bpprop	Average packet size (only packets of the anomaly).
spprop	Ratio of number of syn to number of packets of the anomaly.
samesrcpred	If a specific responsible source appears for the majority of dests.
samesrcportpred	If the majority of responsible sources use the same source ports.
oneportpred	If only one destination port dominated.
invprotopred	If packets using invalid protocol numbers or types dominated.
invalidpred	If the anomaly was mainly consisted of (other) invalid packets.
landpred	If most packets had the same source and destination IPs.
echopred	If most packets were of type ICMP Echo Request/Reply.
icmppred	If most packets were ICMP of any other type.
rstpred	If most packets were TCP with RST flag set.

different types of anomalies [11] and the variations of individual types make it necessary to create very specialized signatures to achieve low misclassification rates. To this extent, we define three types of signatures: (i) universal, (ii) strong and (iii) local. Universal signatures are rules that should never misclassify an anomaly independently of network characteristics. Strong signatures are expected to have low misclassification rates but usually rely on some kind of threshold (and thresholds are difficult to set). Local signatures are defined by network administrators specifically to their domain. Note that they can choose how to best label these anomalies and change thresholds to suit their needs.

We will now discuss the anomalies that we have studied and show some examples of how the attributes we have identified can be used to create strong or even universal signatures for them. The idea is to give the reader a better understanding of how automated classification can be done using these attributes and to show the expressiveness of our algorithm. New attributes and rules can certainly be identified by expert network administrators.

DoS Characterization. Denial-of-service (DoS) attacks are malicious attempts to negate access to network resources [15]. Distributed denial-of-service (DDoS) attacks are (flooding) DoS attacks which use multiple sources to cause much more damage while being hardly detectable. These attacks are extremely common [15][8] and can greatly reduce the QoS of a network even when it has enough resources to cope with the attack [16]. DDoS anomalies may greatly affect the time series of #packets, #flows or both [11][1], and the distributions of destination and source

Table 2. Examples of strong signatures used in this work. *gr* stands for the time series granularity and *sspp* is an abbreviation for the attribute *samesrcportspred*.

Id	Anomaly Type	Signature
1	ICMP Echo DDoS	#respdest == 1 and echopred and (#rpkt/#rdstport > 30*gr or #rsrc/#rdest > 15)
2	TCP SYN DDoS	#respdest == 1 and founds and spprop > 0.9 and oneportpred and #rpkt/#rdstport > 10*gr
3	Network Scan	#respdest > 200 and samesrcpred
4	SYN Port Scan	#respdest == 1 and #rsrc/#rdest == 1 and spprop > 0.8 and avg#rdstports > 5
5	Attack Response	#respdest == 1 and (rstpred or icmppred) and foundp > 20*gr and (not (impactlevelp == 3)) and (#rsrc/#rdest == 1 or sspp)

addresses and ports [12]. However, these characteristics are shared with other types of anomalies, and more detailed information is needed to create robust signatures for their automated classification.

Universal signatures for DDoS anomalies can be defined by analyzing the types of DDoS attacks that use packets which do not comply with the used protocol specification. For example, many attacks have been seen in the wild to use either minimum size IP packets (i.e. 40 bytes) [8], an invalid protocol (e.g. IP protocol 0 or 255 [15][8]), or using land packets for flooding (i.e. packets with the same source and destination IP) [4]. A simple and direct rule would be *if invalidpred or invprotopred or landpred then label as DoS* (see Table 1 for a description of the attributes used). Note that all the identification information (e.g. source(s) and destination IP and port, protocol, etc.) is given as part of the alert.

Creating universal signatures for DDoS anomalies generated by attacks that use compliant packets is very difficult. For this type of attacks we try to develop strong signatures using a rich variety of attributes. Table 2 shows some of the signatures used in this work. For example, the second signature of Table 2 classifies TCP SYN attacks destined to a specific service (*oneportpred*) with an average of 10 or more packets per second (*#rpkt/#rdstport*). It uses *founds* and *spprop* to verify that most of the packets that generate the anomaly have (only) the TCP SYN flag set.

Other Anomalies. We will now quickly go over the other type of anomalies and the most interesting attributes we have identified for each one. *Network scans* [14] are probing attempts to identify the availability of a specific service on many different machines. Network scans can be reliably characterized by a single source communicating with many destinations (i.e. attributes *#respdest* and *samesrcpred*). Stronger signatures can also use *bpprop*, *foundsyn*, *spprop*, *oneportpred* and *#rpkt/#rdstport* to improve accuracy and maybe lower the threshold for *#respdest*. *Port scans* are similar but concentrate on one destination to discover which services the host is running. They should create very little traffic but may have a noticeable impact on #syn. They are characterized by one source, one destination and multiple ports with few packets being used. Signature 4 of Table 2 shows an example for classifying TCP SYN port scans.

Flash crowds (FC) can be defined as a sudden surge of legitimate client requests for a resource. The distributed nature of FCs makes it difficult to distinguish them from DDoS attacks [9]. Attributes include *#rsrc/#rdst*, *oneportpred*, *foundsyn*, *foundpkts* and *#rpkt/#rsrc*, while also taking into consideration that they should only be detectable in higher granularities (i.e. > 5min). *Alpha flows* are unusual high-rate byte transfers from a single source to a single destination, having a strong impact in #bytes and #packets [11]. They also tend to use much bigger packets than DoS attacks. Normally, port information is used to identify known operations that create alpha flows (e.g. scheduled backups). Attributes include *impactlevelbytes*, *impactlevelpkts*, *#respdest*, *#rsrc/#rdst*, *bpprop* and *foundsyn*, and actual ports might be defined.

Finally, *attack response* anomalies are generated by victims of attacks (e.g. DDoS or scans). These response packets are normally either TCP packets with RST ACK, RST or SYN ACK flags set, or ICMP control packets [15]. The line between attack responses and low intensity DDoS anomalies is very thin, especially as these packets are known to be used in DDoS reflector attacks [8]. Signature 5 of Table 2 shows a unified signature for detecting responses to flooding attacks and to scanning attempts.

Local Signatures. The flexibility of being able to understand, add and modify the way that anomalies are classified is a key feature for the applicability of automated network traffic anomaly detection and classification on real networks. Network operators may modify (or disable) strong signatures (i.e. by changing thresholds and/or labels), and also develop local (i.e. domain specific) signatures. For example, instead of trying to separate attack responses from DDoS attacks that use TCP RST packets, a signature might be defined as *if #respdest == 1 and rstpred and impactlevelp > 2 then label as StrongRSTAnomaly*. The flexibility provided by this approach can also be used to reduce false positives of detection algorithms. The rationale is that a wide range of signatures can be defined to potentially cover most of the true anomalies and a default label — applied to any anomalies that did not match one of these signatures — could then be discarded by network operators. This reduces the detection rate of true anomalies but trades the false positive rate of the detection algorithm for the misclassifications of the signatures defined.

4 Validation

We use two datasets to validate our algorithm: the METROSEC project traces with artificially created anomalies and the MAWI traffic repository with anomalies seen on the wild. We concentrate on DDoS anomalies for their importance and multiformity. If we are able to successfully separate different DDoS anomalies from normal traffic and from other types of anomalies, it might follow that general automated classification of network traffic anomalies is possible. Note that because of space limitation, only the most significant results are presented. A full description of the validation process and results can be found in [7].

4.1 Data

The METROSEC traces consist of real traffic collected on the French operational network RENATER with simulated attacks performed using real DDoS attack tools. This dataset was created in the context of the METROSEC research project to, among other goals, study the nature and impact of anomalies on networks' QoS. This dataset has been used for validation by a number of different studies on anomaly detection (e.g. [17]). For the validation of our algorithm, we use 14 METROSEC traces containing DDoS attacks of intensities ranging from very low (i.e. 4-10% of the whole traffic) to very high (i.e. 87-92%). The attacks also vary in type (i.e. from TCP SYN flooding to Smurf attacks), number of attacking hosts (i.e. 1-4) and duration.

On the other hand, the MAWI dataset has real undocumented anomalies. It is composed of 15 minutes packets traces collected daily at 2PM from a Japanese network called WIDE since 1999 to present. These traces are provided publicly after being anonymized and stripped of their payload data (see http://mawi.wide.ad.jp/). Although these traces are undocumented, the authors of [4] started an effort to label anomalies found in this database. We randomly selected a total of 30 traces from 2001 to 2006 from which some had already been identified by [4] to contain DDoS anomalies. Using this second dataset is important to verify that our algorithm is not restricted to a single network nor to artificial attacks.

4.2 Methodology

The validation of our algorithm is divided in two parts. In the first part, a (proper) statistical validation is done using the METROSEC traces for the classification of DDoS anomalies. Different levels of sensitivity of the detection algorithm are used by varying its K parameter from 1.5 to 6. The classification signatures used are the same for all values of K, but only DDoS related signatures are considered. In the second part, the classification performance of our algorithm is tested for different types of anomalies (i.e. DDoS, port and network scan, and attack response) on both of the datasets presented in the previous section. A fixed K of 2 is used, and all the signatures are enabled (including the *same* DDoS signatures used in the first part). A granularity of 30 seconds and the levels of aggregation 0, 8, 16 and 24 are used in the detection algorithm for both parts.

4.3 Results and Discussion

The classification performance for the first part of our validation was very similiar for all values of K (i.e. the algorithm achieved a very high rate of correct classifications with a *very* small rate of misclassifications). The results obtained with K equal to 2 include 23 true positives (i.e. DDoS anomalies correctly classified), 2 false positives (i.e. non-DDoS anomalies misclassified as DDoS), 1 false negative (i.e. misclassified DDoS anomaly) and 455731 true negatives (i.e. non-DDoS anomalies classified as non-DDoS). Further analysis showed that one of

the false positives was actually a real, unexpected DDoS ICMP reflector attack, and the attack responsible for the false negative was correctly classified in a subsequent anomaly.

The results for the second part of our validation were equally promising. On the METROSEC traces, the non-DDoS signatures found a total of 16 port scans, 13 attack responses and 2471 network scans. Manual analysis showed that all port scans and 10 attack responses were true positives. We were not able to identify the nature of the other 3 attack responses. Network scans were not manually analyzed, but the signature used (see Table 2) has a very low (if not inexistent) misclassification rate. Running the algorithm on the 30 fifteen minutes MAWI traces resulted in 22 DDoS, 4429 network scan, 5233 port scan and 72 attack response anomalies in a total of 2.5 million anomalies detected. Manual analysis and cross-referencing with the results of [4] revealed 19 true positives (of which 6 had not been detected by [4]), 3 false positives that might be ICMP reflector attacks, and 9 (known) false negatives. The false negatives were mainly due to the detection algorithm used, and are not a limitation of our classification approach or of the signatures used. Preliminary analysis of the other type of anomalies showed that many of them were due to worm scannings (and responses), with Sasser and Dabber variants being particularly common.

5 Conclusions

In this paper we presented a new approach for automated classification of network traffic anomalies. We defined an initial set of anomaly attributes and characterized different types of anomalies (e.g. DDoS, network scans, etc) using them. We showed how automated classification can be done (succesfully) using these attributes within a signature-based approach and leveraging on the capability of state-of-the-art detection algorithms to identify the anomalous flows. We evaluated our work using two very different sets of packets traces with real network traffic and several anomalies. The results obtained illustrate the expressiveness of our approach to differentiate between many types of DDoS anomalies and other anomalies (including normal traffic variations), and strongly hint that general automated classification is possible. On future work we intend to explore the subtleties of other types of anomalies and to see how state-of-the-art identification algorithms can be easily integrated to our classification approach.

Acknowledgment

This work has been done in the framework of the ECODE project funded by the European commission under grant FP7-ICT-2007-2/223936.

References

1. Barford, P., Kline, J., Plonka, D., Ron, A.: A signal analysis of network traffic anomalies. In: Internet Measurment Workshop, Marseille (November 2002)
2. Cho, K., Mitsuya, K., Kato, A.: Traffic data repository at the wide project. In: USENIX ATEC, San Diego, California (2000)

3. Cormode, G., Muthukrishnan, S.: What's new: finding significant differences in network data streams. IEEE/ACM Trans. Netw. 13(6), 1219–1232 (2005)
4. Dewaele, G., Fukuda, K., Borgnat, P., Abry, P., Cho, K.: Extracting hidden anomalies using sketch and non gaussian multiresolution statistical detection procedures. In: Workshop on Large-Scale Attack Defense (LSAD), Kyoto, Japan (2007)
5. Estan, C., Savage, S., Varghese, G.: Automatically inferring patterns of resource consumption in network traffic. In: ACM SIGCOMM, Karlsruhe (2003)
6. Farraposo, S., Owezarski, P., Monteiro, E.: A multi-scale tomographic algorithm for detecting and classifying traffic anomalies. In: IEEE ICC, Glasgow (June 2007)
7. Fernandes, G., Owezarski, P.: Automated classification of network traffic anomalies. LAAS Report No 08468 (2008)
8. Hussain, A., Heidemann, J., Papadopoulos, C.: A framework for classifying denial of service attacks. In: ACM SIGCOMM, Karlsruhe (2003)
9. Jung, J., Krishnamurthy, B., Rabinovich, M.: Flash crowds and denial of service attacks: Characterization and implications for cdns and web sites. In: WWW, Honolulu, Hawaii (May 2002)
10. Kim, M.-S., Kong, H.-J., Hong, S.-C., Chung, S.-H., Hong, J.: A flow-based method for abnormal network traffic detection. In: IEEE/IFIP Network Operations and Management Symposium, Seoul (April 2004)
11. Lakhina, A., Crovella, M., Diot, C.: Characterization of network-wide anomalies in traffic flows. In: Internet Measurement Conference, Taormina, Italy (2004)
12. Lakhina, A., Crovella, M., Diot, C.: Mining anomalies using traffic feature distributions. In: ACM SIGCOMM, Philadelphia (2005)
13. Li, X., Bian, F., Crovella, M., Diot, C., Govindan, R., Iannaccone, G., Lakhina, A.: Detection and identification of network anomalies using sketch subspaces. In: Internet Measurement Conference, Rio de Janeiro, Brazil (2006)
14. Mirkovic, J., Reiher, P.: A taxonomy of ddos attack and ddos defense mechanisms. SIGCOMM Comput. Commun. Rev. 34(2), 39–53 (2004)
15. Moore, D., Voelker, G.M., Savage, S.: Inferring internet denial-of-service activity. In: USENIX SSYM, Washington, DC (2001)
16. Owezarski, P.: On the impact of dos attacks on internet traffic characteristics and qos. In: ICCCN (October 2005)
17. Scherrer, A., Larrieu, N., Owezarski, P., Borgnat, P., Abry, P.: Non-gaussian and long memory statistical characterizations for internet traffic with anomalies. IEEE Trans. Dependable Secur. Comput. 4(1), 56–70 (2007)