Entropy-based Internet Traffic Anomaly Detection: A Case Study

Przemysław Bereziński, Józef Pawelec, Marek Małowidzki, Rafał Piotrowski

Military Communication Institute, Zegrze, Poland

{p.berezinski, j.pawelec, m.malowidzkim r.piotrowski}@wil.waw.pl

Abstract. Recently, entropy measures have shown a significant promise in detecting diverse set of network anomalies. While many different forms of entropy exist, only a few have been studied in the context of network anomaly detection. In the paper, results of our case study on entropy-based IP traffic anomaly detection are prestented[1]. Besides the well-known Shannon approach and counter-based methods, variants of Tsallis and Renyi entropies combined with a set of feature distributions were employed to study their performance using a number of representative attack traces. Results suggest that parameterized entropies with a set of correctly selected feature distributions perform better than the traditional approach based on the Shannon entropy and counter-based methods.

Keywords: anomaly detection, entropy, netflow, network traffic measurements.

1 Introduction

As the number of network security incidents grows each year [1], network intrusion detection becomes a crucial area of research. Widely used security solutions like firewalls, antivirus software and intrusion detection systems do not provide sufficient protection anymore because they do not cope with evasion techniques and not known yet (0-day) attacks. To cover this area, anomaly detection is a possible solution. Network anomalies may potentially indicate malicious activities such as worms propagation, scans, botnets communication, Denial of Service attacks, etc. The problem of a generic anomaly detection method for network anomalies is still unsolved. Recently, entropy measures have shown a significant promise in detecting diverse set of network anomalies [2-6].

Anomaly may be defined as a deviation from a norm and something which is outside the usual range of variations. Usually, in anomaly detection, a model describing normal circumstances is prepared first, then predictions based on the model are compared with actual measurements. A comprehensive survey about anomaly detection methods has been presented by Chandola et al. in [7]. There are many problems with

[1] This work has been co-financed by the Polish National Centre of Research and Development under grant no. PBS1/A3/14/2012 (SECOR - Sensor Data Correlation Engine for Attack Detection and Support of Decision Process).

anomaly detectors which have to be addressed. The main challenges are: high false positive rates, long computation time, tuning and calibration and root-cause identification [7-8].

In our previous work [9], some generalizations of entropy were described in details and some ideas about their possible use for network anomaly detection were presented. In this paper, we make two contributions. First, we present some not commonly known theory regarding entropies used in the context of anomaly detection. Second, we present results of our case study on entropy-based IP traffic anomaly detection that involved a number of entropy variants and a set of different feature distributions.

The paper is organized as follows: First, we discuss related work and overview different form of entropies. Then, we switch to flows and traces, describing a dataset and anomalies it contains. Next, we discuss the applied methodology. In the following sections, we analyze our results. We finish the paper with conclusions and proposals for future work.

2 Related work

Entropy-based approach for network anomaly detection has been of a great interest recently. This approach relies on traffic feature distributions. Feature distributions give a different view of a network activity than traditional counter-based volume metrics (like flow, packet, byte counts), which are widely used in commercial solutions. Several traffic features, i.e., header-based (addresses, ports, flags), size-based (IP or port specific percentage of flows, packets and octets) and behavior-based (in/out connections), have been suggested in the past [2],[5]. However, it is unclear which features should be used. As an example, Nychis in [2] claims that there is a strong correlation between addresses and ports and recommends the use of size and behavior-based features. On the contrary, Tellenbach in [5] reported no correlation among header-based features. A possible explanation of these contradictory results could be different data sets or, perhaps, some change in Internet traffic characteristics. We propose another possible explanation in section 7.

Although entropy is a prominent way of capturing important characteristics of distributions in a compact form (a single number), some other summarization techniques are proposed in the literature, i.e., histograms [11] and sketches [12]. Their main problem is however the proper tuning.

According to the literature, entropy of feature distributions performs better than widely used counter-based features (like flows, packets and byte counts) [15]. Volume based detection handles large traffic changes (such as bandwidth flooding attacks) well, but a large class of anomalies does not cause detectable changes of volume. Moreover, Brauckhoff et al. in [10] prove that an entropy-based approach performs better than a volumetric one in case sampled[2] flows are used.

Entropy-based methods use the Shannon entropy [2],[15], the Titchener entropy (T-Entropy) [6], and the parameterized Renyi [3] or Tsallis [4-5] entropy. Most authors agree that there are some limitations of entropy-based detection, especially when it

[2] Many routers form flow statistics from a sampled stream of packets in order to limit consumption of resources for measurement operations.

comes to detecting small or slow attacks. This is especially true for the Shannon entropy, which has a limited descriptive power. According to the literature, the range of detectable anomalies for parameterized entropies is wider [5].

3 Entropy

In this section we present some not commonly known theory regarding entropies used in the context of anomaly detection.

Definition of entropy as a measure of disorder comes from thermodynamics and was proposed in the early 1850s by Rudolf Clausius. The statistical definition of entropy as a measure of uncertainty was developed by Ludwig Boltzmann in the 1870s. In 1948, Claude Shannon adopted entropy to information theory. We will start our quick survey with the Shannon's variant, as it is probably the most popular and commonly used entropy.

For a probability distribution $p(X = x_i)$ of a discrete random variable X , the Shannon entropy is defined as:

$$H_S(X) = \sum_{i=1}^{n} p(x_i) log_a \frac{1}{p(x_i)} \qquad (1)$$

X is the feature that can take values $\{x_1...x_n\}$ and $p(x_i)$ is the probability mass function of outcome x_i. Depending on the base of the logarithm, different units are used: *bits* (a=2), nats (a=e) or hurtleys (a=10). For the purpose of anomaly detection, sampled probabilities estimated from a number of occurrences of x_i in a time window t are typically used. The value of entropy depends on randomness (it attains maximum when probability $p(x_i)$ for every x_i is equal) but also on the value of n. In order to measure randomness only, some normalized forms can be employed. For example, an entropy value can be divided by n or by maximum entropy defined as $log_a(n)$.

Sometimes not only the degree of uncertainty is important but also the extent of changes between assumed and observed distributions, respectively denoted as q and p. A relative entropy, also known as the Kullback-Leibler divergence, may be employed to measure the size of change:

$$D_{KL}(p||q) = \sum_{i=1}^{n} p(i) log_a \frac{p(i)}{q(i)} \qquad (2)$$

The Shannon entropy assumes a tradeoff between contributions from the main mass of the distribution and the tail. To control this tradeoff, two parameterized Shannon entropy generalizations were proposed, respectively, by Renyi (1970s) [16] and Tsallis (late 1980s) [17]. If the parameter denoted as α has a positive value, it exposes the main mass (the concentration of events that occur often), if the value is negative – it refers to the tail (the dispersion caused by seldom events). Both parameterized entropies derive from the Kolmogorov-Nagumo generalization of an average:

$$\langle X \rangle_{\varphi} = \varphi^{-1}(\textstyle\sum_{i=1}^{n} p(x_i)\, \varphi(x_i)) \qquad (3),$$

where φ is a function which satisfies the postulate of additivity and φ^{-1} is the inverse function. Renyi proposed the following function φ:

$$\varphi(x_i) = 2^{(1-\alpha)x_i} \tag{4}$$

After transformations, Renyi may be given in the following form:

$$H_{R\alpha}(X) = \frac{1}{1-\alpha} \, log_a \left(\sum_{i=1}^{n} p(x_i)^{\alpha} \right) \tag{5}$$

Tsallis extended the Renyi entropy with the following function φ:

$$\varphi(x_i) = \frac{2^{(1-\alpha)x_i} - 1}{1-\alpha} \tag{6}$$

After transformations, the Tsallis entropy will be given by:

$$H_{T\alpha}(X) = \frac{1}{1-\alpha} \left(\sum_{i=1}^{n} p(x_i)^{\alpha} - 1 \right) \tag{7}$$

The normalized form of the Tsallis entropy [18] is typically defined as:

$$H_{TN\alpha}(X) = \frac{1}{1-\alpha} \left(1 - \frac{1}{\sum_{i=1}^{n} p(x_i)^{\alpha}} \right) \tag{8}$$

For both the Tsallis and Renyi entropies, parameter α exposes concentration for $\alpha > 1$ and dispersion for $\alpha < 1$. For $\alpha \to 1$, both converge to the Shannon entropy.

Another form used in the context of anomaly detection is the Titchener entropy (T-entropy) [19]. T-entropy is the gradient of linearized form of a string complexity measure called T-complexity. String complexity is a minimum number of steps required to construct a given string. As we mentioned earlier, in typical entropy-based detection, frequencies for values of discreet random variables are used to estimate probabilities. The probabilities must not depend on the occurrence of previous values. In a complexity-based approach, values are concatenated into a string in a sequence (where order matters). The string is then compressed with some algorithm and the output length is used as an estimate for the complexity; finally, the complexity becomes an estimate for entropy. More details about T-entropy is presented in our previous paper [9] and in Einman's dissertation [6].

After a short review on theory, we now switch to networks. In the following two sections, we will discuss a flow-based network analysis and a dataset we have used in our work.

4 Flow-based analysis

There are two approaches to a network traffic analysis, namely packet-based and flow-based. A flow-based approach is becoming more and more popular since it is more scalable in the context of network speed. The concept of network flows was introduced by Cisco and currently is standardized by the Internet Engineering Task Force (IETF). According to the IETF IPFIX working group [21], "A flow is defined as a set of IP packets passing an observation point in the network during a certain time interval. All packets belonging to a particular flow have a set of common properties." In the simplest form, these properties are source and destination addresses and ports. In this work we focus on flow-based network anomaly detection. Flows may be classified on the basis of several different schemes, i.e., size (elephant and mice), duration (tortoise and dragonfly) and burstiness (alpha and beta traffic) [20]. For example, an elephant is any flow with rate exceeding 1% of the link utilization. Duration of a dragonfly is less than 2 sec, while a tortoise lasts longer than 15 min. According to [20], about 45% of Internet flows are dragonflies and less than 2% are tortoises. Taxonomy of network anomalies as discussed in [14] and [15] is presented in Table 1. The table lists examples of both legitimate and malicious network activity.

Table 1 Network anomaly types according to [14] and [15]

Anomaly Type	Description
α Flows	Unusually large point to point byte transfer
DoS, DDoS, DRDoS	Single-source or distributed (also reflected) Denial of Service Attack which may be volumetric, protocol or application based
Flash Crowd	Burst of traffic to a single destination, from a "typical" distribution of sources
Network/Port Scan	Probes to many destination addresses/ports with a small set of source ports/addresses
Ingress-Shift	Traffic shift from one ingress point to another
Outage	Decrease in traffic due to equipment failures or maintenance
Point to Multi-point	Traffic from single source to many destinations, *e.g.* ,content distribution
Worms	Code propagation by exploiting vulnerable services (special case of a network scan)

5 Dataset

The data we used in our case study have been prepared in the following way: First, we captured (hopefully legitimate) traffic from a medium size corporation network using span ports and open source software - *softflowd* and *nfsen*. Then, we mixed this traffic with a subset of the labeled dataset contributed by Sperrotto et al. [13]. This set is based on data collected from a real honeypot which was running for 6 days. The honeypot featured HTTP, SSH and FTP services. The authors gathered about 14 million malicious[3] flows. From the dataset we extracted flows “responsible” for anomalies listed in Table 2.

Table 2 Selected network anomalies

Attack	Relation address /port	Size	Duration	Flows	Packets	Bytes
SSH BrutteForce (A)	1-1/n-1	S	1 h	750	20K	2,5M
WEB Scan (B)	1-1/n-1	M	14 s	660	7K	0,6M
SSH NetworkScan$_1$ (C)	1-n/n-1	M	2,5 min	15K	30K	1,7M
SSH NetworkScan$_2$ (D)	1-n/n-1	L	7 min	23K	300K	34M

Anomaly A represents a one to one brake-in to the SSH service with a dictionary based attack on username and password. This anomaly is relatively slow and small. Anomaly B represents a typical activity of a web scanner. The volume for this anomaly is a bit higher than for anomaly A but the duration is short. Anomalies C and D are examples of network scans. They are characteristic for network-wide worm propagation via service vulnerabilities. The volume for this anomalies is significantly larger.

We placed our mixed legitimate and anomalous (honeypot) traffic in a relational database. We decided to employ bidirectional flows compliant with RFC 5103 as, according to some works - e.g. [2], unidirectional flows may entail biased results in anomaly detection. This assumption required us to perform some conversion from an unidirectional to a bidirectional form.

6 Methodology

Below, we discuss data processing and selected flavors of entropies and feature distributions.

We analyzed our dataset stored in a relational database. Stored procedures were implemented to capture different feature distributions. The anomalies search area was not limited by any filter (per direction, protocol, etc.). We analyzed flows within fixed (5 min) time windows (with no sliding). Next, the Tsallis or Renyi entropies for positive and negative α values were calculated for distributions listed in Table 3. These distributions are commonly employed in entropy-based analysis except for flows duration which is our proposal.

[3] All flows from honeypots are malicious by its nature, so there is a need to mix them with legitimate traffic for the purpose of anomaly detection testing. With such custom-made datasets benchmarking is hampered.

Table 3 Selected traffic feature distributions

Feature	Probability mass function
src(dst)address(port)	$\frac{\text{number of } x_i \text{ as src(dst) address(port)}}{\text{total number of src(dst) addresses(ports)}}$
flows duration	$\frac{\text{number of flows with } x_i \text{ as duration}}{\text{total number of flows}}$
packets, bytes	$\frac{\text{number of pkts(bytes) with } x_i \text{ as src(dst) address(port)}}{\text{total number of pkts(bytes)}}$
in(out)-degree	$\frac{\text{number of adresses with } x_i \text{ as in(out) degree}}{\text{total number of addresses}}$

The following set of entropies presented in Section 3 was selected:

Table 4 Selected forms of parameterized entropy

Denotation	Formula	Comments
Tsallis$_1$	$H = \frac{1 - \sum_i p_i^\alpha}{1 - \alpha}$	general form
Tsallis$_2$	$H = \frac{1 - \sum_i p_i^\alpha}{(1 - \alpha)\sum_i p_i^\alpha}$	normalization
Renyi	$H = \frac{log_2 \sum_i p_i^\alpha}{1 - \alpha}$	general form
Shannon	$H = \frac{log_2 \sum_i p_i^\alpha}{1 - \alpha} \;, \alpha \to 1$	obtained from Renyi ($\alpha \to 1$)

This selection was based on some promising results with flow-based network anomaly detection as reported by [3],[5]. We believe that T-entropy (which was not selected) is more appropriate for packet-based detection, where the order of packets, e.g., requests and responses from servers, really matters.

During the training phase, a profile was built using time-period specific min and max entropy values computed for every <feature, α> pair. During the detection phase, the observed entropy H_α was compared with the min and max values stored in the profile, according to the following rule:

$$result_\alpha(x_i) = \frac{H_\alpha(x_i) - (min_\alpha - k * min_\alpha)}{(max_\alpha + k * max_\alpha) - (min_\alpha - k * min_\alpha)} \qquad (9)$$

$$k - threshold\ margin, k \in \langle 0, 0.3 \rangle$$

According to this rule, threshold exceeding is indicated as abnormal dispersion for values less than zero or abnormal concentration for values higher than one. Abnormal dispersion or concentration for different feature distributions is characteristic for anomalies. For example, during a port scan, a high dispersion in port numbers and high concentration in addresses is observable. Detection is based on the relative value of entropy with respect to the distance between min and max. Coefficient *k* in the formula (9) determines a margin for min and max boundaries and may be used for a tuning purposes. A high value of k, e.g. $k = 0.3$, limits the number of false positives while a low value ($k \to 0$) increases detection rate.

For comparison, we also employed a traditional counter-based approach for flow, packet and byte counts. We assumed ideal conditions to measure detection rate for anomalies. We used the same part of legitimate traffic during training and in the detection phase so no false positives could be observed. To measure the false positives rate, we cross-checked legitimate traffic using two halves of the profile.

Finally, linear and rank correlation of entropy timeseries for different α values and different feature distributions was performed in order to define a proper range of α values, and to verify the legitimacy to use a broad spectrum of features. The results are presented in the next section.

7 Results

We obtained the best results for detection of high dispersion and concentration with the Tsallis_1 and Renyi entropies. Comparing the Tsallis_1 and Renyi, we observed higher values of threshold excess (more significant peaks of entropy values) with Tsallis_1, although such sensitivity was also visible in the form of sharp false positives with legitimate traffic cross check. The Renyi entropy was not so sensitive to anomalies (the excess of threshold was smaller than Tsallis_1) but was a bit less vulnerable to false positives. With a traditional counter-based approach anomalies A and B were undetectable by a flow, packet and byte counts, while anomalies C and D were detectable only by flow count. The results for Tsallis_2 were ambiguous (slightly exceeding threshold). The Shannon entropy detection failed for anomalies A and B. All results are presented in Table 5. The markings +, +/-, – denote, respectively, successful, indecisive and unsuccessful detection.

Table 5 The effectiveness of selected entropies and volume counters

Attack	**Tsallis_1**	**Tsallis_2**	**Renyi**	**Shannon**	**Flows**	**Packets**	**Bytes**
SSH BrutteForce (A)	+	+/-	+	-	-	-	-
WEB Scan (B)	+	+	+	-	-	-	-
SSH NetworkScan_1 (C)	+	+/-	+	+	+	-	-
SSH NetworkScan_2 (D)	+	+/-	+	+	+	-	-

We noticed that with an entropy-based approach some feature distributions work better than others. The best results were obtained by using addresses, ports and flows duration distributions, although we think that the set of proper features is specific for a particular anomaly– thus, a number of different, uncorrelated features (see the second part of the section) should be employed. Abnormally high dispersion in destination addresses distribution for anomalies C and D exposed by negative values of alpha parameters, is depicted in Fig. 1. We can see time t on x axis (5 minute time windows), result r (where value above one means abnormal dispersion and below zero means abnormal concentration - formula (9)) on y axis and α values on z axis. Anomaly durations are marked on the time axis.

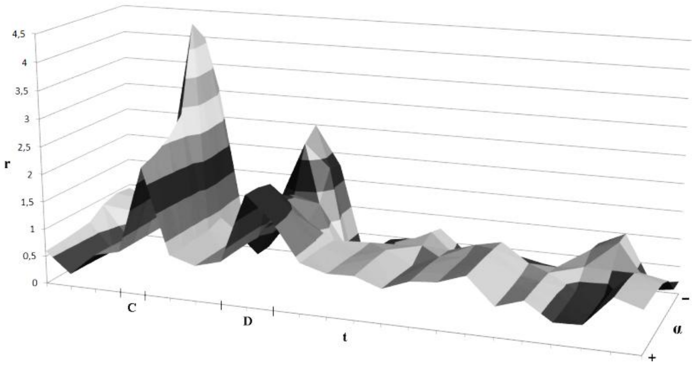

Fig. 1. Abnormally high dispersion in destination addresses for anomalies C and D (Renyi$_1$)

Abnormal concentration of flows duration for anomaly B - which is typical for anomalies with fixed data stream - is depicted in Fig. 2.

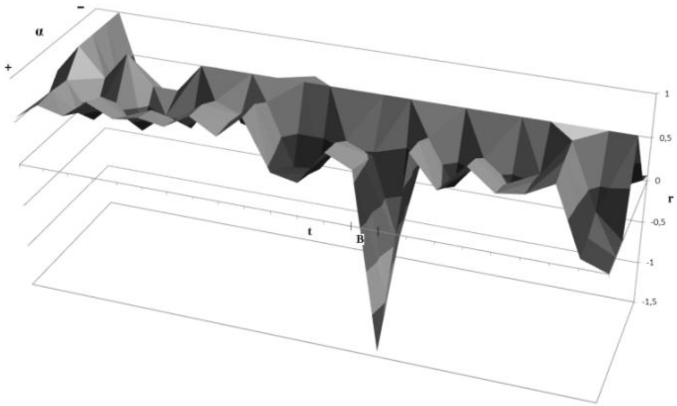

Fig. 2. Abnormally high concentration in flows duration for anomaly B (Tsallis$_1$)

Fig. 3 shows unsuccessful detection (no excess of threshold) of anomaly with common approach based on flow, packet and byte counters.

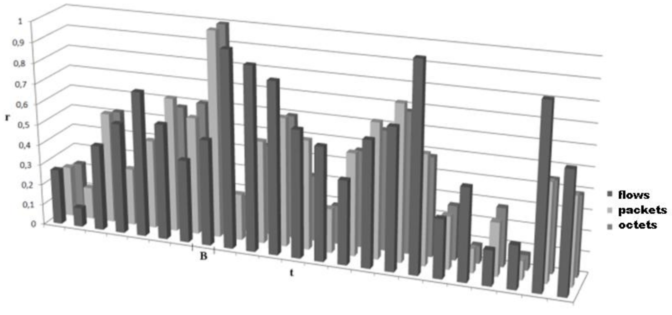

Fig. 3. Unsuccessful detection of anomaly B with a counter-based approach

In the remainder of this section, we analyze correlations for various α values and for various feature distributions. This is important as strong correlation suggests that some results are closely related to each other and thus it may be sufficient to restrict the scope of analysis without impairing its validity.

The results of correlation between entropy timeseries for different α values are presented in Table 6. The table shows the pairwise Tsallis$_1$ α correlation scores from range <-1..1> where scopes |1-0.9|, |0.9-0.7|, |0.7-0.5|, |0.5-0| denote, respectively, strong, medium, weak, and no correlation. The sign determines if the correlation is positive (+/no sign) or negative (-). The presented values (see Table 6) are an average from 15 different feature distributions scores. For the Renyi entropy, the results were similar.

Table 6 Results of linear and rank correlation of α

Pearson	α=-3	α=-2	α=-1	α=0	α=1	α=2	α=3
α=-3	1	0,9	0,9	0,66	0	-0,06	-0,09
α=-2	-	1	0,9	0,69	0	-0,06	-0,09
α=-1	-	-	1	0,75	0	-0,05	-0,08
α=0	-	-	-	1	0	0,18	0,12
α=1	-	-	-	-	1	0,88	0,82
α=2	-	-	-	-	-	1	0,97
α=3	-	-	-	-	-	-	1

Spearman	α=-3	α=-2	α=-1	α=0	α=1	α=2	α=3
α=-3	1	0,9	0,8	0,46	0	-0,09	-0,11
α=-2	-	1	0,9	0,57	0	-0,07	-0,1
α=-1	-	-	1	0,72	0	-0,06	-0,09
α=0	-	-	-	1	0	0,2	0,15
α=1	-	-	-	-	1	0,87	0,79
α=2	-	-	-	-	-	1	0,98
α=3	-	-	-	-	-	-	1

It should be noticed, that there is a strong positive linear (Pearson) and rank (Spearman) correlation for negative α values and strong positive correlation between α values which are higher than 1. For α = 0 there is some small positive correlation with negative values. For α = 1 (Shannon) there is a medium correlation with α = 2 and α = 3. These results suggest that it is sufficient to use α values from range <-2,2> to have different sensitivity levels of entropy. Some interesting results of pairwise correlation between the Tsallis$_1$ entropy timeseries of different feature distributions are presented in Table 7 and Table 8 (the results for the Renyi entropy were similar).

Table 7 Results of correlation of features for α=-3

Pearson	ip_src	ip_dst	port_src	port_dst	indegree	outdegree
ip_src	1	0,8	0,89	0,91	0,37	0,35
ip_dst	-	1	0,98	0,89	0,27	0,55
port_src	-	-	1	0,86	0,15	0,5
port_dst	-	-	-	1	0,41	0,53
indegree	-	-	-	-	1	0,27
outdegree	-	-	-	-	-	1

Spearman	ip_src	ip_dst	port_src	port_dst	indegree	outdegree
ip_src	1	0,9	0,85	0,87	0,47	0,69
ip_dst	-	1	0,96	0,89	0,43	0,83
port_src	-	-	1	0,83	0,3	0,69
port_dst	-	-	-	1	0,52	0,76
indegree	-	-	-	-	1	0,48
outdegree	-	-	-	-	-	1

Table 8 Results of correlation of features for α=3

Pearson	ip_src	ip_dst	port_src	port_dst	indegree	outdegree
ip_src	1	-	-0,34	-0,02	-0,07	0,44
ip_dst	-	1	-0,29	0,05	0,08	-0,28
port_src	-	-	1	-0,42	0,59	-0,04
port_dst	-	-	-	1	-0,39	0,01
indegree	-	-	-	-	1	0,03
outdegree	-	-	-	-	-	1

Spearman	ip_src	ip_dst	port_src	port_dst	indegree	outdegree
ip_src	1	0,0	-0,21	0,07	0,21	0,366
ip_dst	-	1	-0,31	0,07	0,08	-0,35
port_src	-	-	1	-0,55	0,64	0,23
port_dst	-	-	-	1	-0,53	0,12
indegree	-	-	-	-	1	0,18
outdegree	-	-	-	-	-	1

We presented results for one positive and one negative value of α because these results differ significantly. Averaging (based on results from the whole range of α values) would hide an essential property. It is noticeable that there is a strong positive correlation of addresses and ports for negative values of α but no correlation for positive values. Thus both Nychis [2] and Tellenbach [5] could have been right.

8 Conclusion and future work

Concluding the results of our case study, we can observe that:

i) not normalized Tsallis and Renyi entropies performed best;

ii) the Shannon entropy and counter-based methods performed poorly; in fact, they were unable to detect neither small nor medium-size attacks;

iii) a broad spectrum of features provides a better flexibility to detect different types of anomalies;

iv) among a large set of network traffic feature distributions, addresses, ports, and flows durations proved in our study to be the best choices.

While we admit that our experiments were limited to a small number of cases, we also believe that the cases were representative. The analyzed dataset contained traces of typical network attacks. Thus, while more research work is necessary to validate the effectiveness of the methods that performed well, the poor performance of the Shannon entropy and counter-based methods allows to question whether they are the right approach to anomaly detection.

The work described in the paper will be the basis for implementation of a sensor cooperating with a multi-source event correlation engine. Thus, we are concerned with a high detection rate rather than a small false positives ratio. Future works will include classification based on results from various features and anomaly details extraction (e.g., addresses and ports of top contributors to malicious traffic). We are also considering the analysis of additional features. We hope we will be able to report valuable results.

References

1. Verizon Risk Team ‘Data Breach Investigations report’, Verizon , 2012
2. Nychis G. et al. *‘An Empirical Evaluation of Entropy-based Traffic Anomaly Detection’* in ACM SIGCOMM conference on Internet Measurement, 2008.
3. Ruoyu Y. et al. *‘Multi-scale entropy and renyi cross entropy based traffic anomaly detection’* in IEEE International Conference on Communication Systems (ICCS), 2008.
4. Ziviani A.et al. *'Network Anomaly Detection using Nonextensive Entropy'* in IEEE Communications Letters, IEEE Press, vol.11, no. 12, 2007.
5. Tellenbach B. ‘Detection, Classification and Visualization of Anomalies using Generalized Entropy Metrics’, Dis. Th., Elektro-Technische Hohschule Zurich, 2012.
6. Eimann R. *Network Event Detection with Entropy Measures,* Dis. Th., University of Auckland, 2008.
7. Chandola V. et al. ‘Anomaly detection: A survey’. ACM Comput. Surv., 41(3), 2009.
8. Brauckhoff D. *‘Network Traffic anomaly Detection and Classification’*, Dis. Th., Elektro-Technische Hohschule Zurich, 2010.
9. Pawelec J. et al. *‘Entropy Measures For Internet Traffic Anomaly Detection’* in TransComp conference on Computer Systems, Industry and Transport, 2013

10. Brauckhoff D.et al. *‘Impact of packet sampling on anomaly detection metrics*’ in ACM SIGCOMM conference on Internet Measurement, 2006.
11. Stoecklin M. et al. *‘A two layered anomaly detection technique based on multi-modal flow behavior models'* in PAM conference on Passive and active network measurement, Springer, 2008.
12. Dimitropoulos X. et al. 'The eternal sunshine of the sketch data structure', Computer Networks, vol. 52, no. 17, 2008.
13. Sperotto A. et al. *‘A Labeled Data Set For Flow-based Intrusion Detection*’, in IEEE International Workshop on IP Operations and Management (IPOM), Berlin, 2009.
14. Plonka D., Barford P. *'Network anomaly confirmation, diagnosis and remediation'* in Allerton conference on Communication, control, and computing, IEEE Press, 2009.
15. Lakhina A et al. *‘Mining anomalies using traffic feature distributions’* in ACM SIGCOMM conference on Internet Measurement, 2005.
16. Renyi A. *Probability Theory*, North-Holland, Amsterdam, 1970.
17. Tsallis C. ‘Possible Generalization of Boltzmann-Gibbs Statistics’, J. Statistical Physics, vol. 52, no. 1-2, 1988.
18. Gupta P., Kumar V. *‘General Pseudoadditivity of Kapur’s Entropy prescribed by the existence of equilibrium’* in the International Journal of Scientific & Engineering Research, vol. 1 no. 3, 2010
19. Titchener M. ‘Deterministic Complexity and Entropy’ Fundamenta Informaticae, vol. 64, no. 1-4, 2005.
20. Lan K., Heidemann J., *‘On the correlation of Internet flow characteristics’*, Technical Report ISI-TR-574, USC/Information Sciences Institute, 2003.
21. Claise B. ‘Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information', RFC 5101, 2008.