OPEN ACCESS

2nd International Electronic Conference on Entropy and Its Applications
15-30 November 2015

www.sciforum.net/conference/ecea-2

Conference Proceedings Paper – Entropy

On Entropy in Network Traffic Anomaly Detection

Jayro Santiago-Paz * and Deni Torres-Roman

CINVESTAV, Campus Guadalajara, Av. del Bosque 1145, Col. El Bajio, Zapopan, Mexico.

* Author to whom correspondence should be addressed; E-Mail: jsantiago@gdl.cinvestav.mx; Tel.: +52(33)-3777-3600, +52(33)-3777-3609.

Published: 13 November 2015

Abstract: Different systems, e.g., the anomaly-based network intrusion detection system (A-NIDS), have been continuously developed in order to ensure integrity, availability, and confidentiality of networks. In this paper, we present a structured and comprehensive overview of the research into entropy-based A-NIDS with the intention of providing researchers a quick introduction to essential aspects of this topic. The main components of the general architecture of A-NIDS based on Entropy are discussed. The achieved high detection rates prove the effective use of entropy. Finally, some open issues in entropy-based network traffic anomaly detection are also highlighted.

Keywords: Network traffic anomaly detection; entropy; network security; A-NIDS, Mutual information, KL divergence

PACS classifications: one, two, three

---

1. Introduction

Given a traffic network and its set of the selected traffic features $X = \{X_1, X_2, \ldots, X_p\}$, and N time instances of X, the normal and abnormal behaviors of the instances can be studied. The space of all instances of X builds the feature space which can be mapped to another space by employing a function such as entropy. In the literature, Shannon and generalized Rényi and Tsallis entropy estimators, as well as probability estimators (Balanced [1], Balanced II [2]), are used.

Chandola et al. (2009) [3] states that the term anomaly-based intrusion detection in networks refers to the problem of finding exceptional patterns in network traffic that do not conform to the expected normal behavior. This concept is accepted by the Internet Community.

A generic architecture of entropy-based A-NIDS is presented in figure 1, see [4]. It usually consists of two stages: the training and the testing stage. In the training stage, using a database of "normal" or free-anomaly network traffic, feature extraction, windowing, and entropy calculation modules, a "normal" profile is found. In the testing stage, using feature extraction, windowing, and entropy calculation modules, anomalies in the current network traffic are detected and classified.

Figure 1. General architecture of entropy-based A-NIDS.

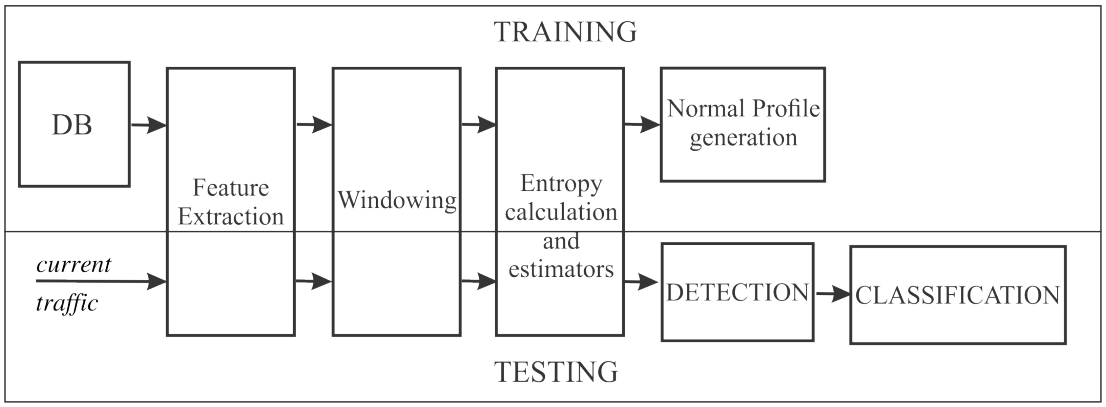

Table 1 presents the notation used in this paper.

Table 1. Notations.

Symbol	Meaning
$D_\rho(P\|\|Q)$	Information divergence, where ρ is the order of the information divergence.
$D_1(P\|\|Q)$	Kullback-Leibler divergence
$f(x;\beta_g)$	Discriminant function where β_g are class-specific parameters
$H^R(\bullet,q)$	Rényi entropy
$H^S(\bullet)$	Shannon entropy
$H^T(\bullet,q)$	Tsallis entropy
$H(X;Y)$	Joint entropy
$H(Y\|X)$	Conditional entropy
$I(X;Y)$	Mutual information
p_i	Probability of occurrence of x_i element
$p(x,y)$	Joint probability
$p(x\|y)$	Conditional probability
q	Parameter of generalized entropies
S_X	m-dimensional space of all traffic features
S_N	p-dimensional space of free-anomaly traffic features
$W_i(L,\tau)$	i-th sliding window with L packets and τ as the overlapping parameter
$X,Y,Z \in \mathbb{R}^n$, $n=1,2,\ldots,k$	Random variables
x,y,z	Instances of random variables X,Y,Z.

Chandola et al. (2009) [3] present "a survey in anomaly detection, grouping existing techniques into different categories based on the underlying approach adopted by each technique". Bhuyan et al. (2014)

[5] provides an excellent survey in network anomaly detection describing six distinct classes of methods and systems. However, these surveys are not focused directly on entropy-based A-NIDS.

This paper presents a structured and comprehensive overview of the research into entropy-based A-NIDS with the intention of providing researchers a quick introduction to essential aspects of this topic. Using a general architecture of entropy-based A-NIDS, the different techniques proposed in the state-of-the-art of the main modules are shown. The measures of information used by researchers and the most important metrics for testing the performance of the detection and classification are presented. We also highlight some open issues in entropy-based network traffic anomaly detection.

The next sections describe the main modules of the entropy-based A-NIDS. Section 2 presents the main databases used for researchers in the field. Section 3 shows the most commonly employed features in network traffic and the windowing approach. Section 4 introduces the mathematical background, including Shannon entropy, generalized entropies, Mutual information, Kullback-Leibler divergence, conditional entropy, and the techniques to estimate their values. Section 5 presents the decision functions defined by the researchers in the detection stage. Section 6 shows the approaches employed in the classification stage and the most widely used metrics for evaluating the A-NIDS. Section 7 contains the conclusions of the paper and some important open issues of this topic.

2. Databases

Different databases have been used to evaluate the A-NIDS, and these databases are divided into two groups: synthetic and real.

The synthetic databases are generated artificially, e.g., the MIT-DARPA 1998, 1999, 2000 databases[1], which include five major categories: Denial of Service Attacks (DoS), User to Root Attacks (U2R), Remote to User Attacks (R2U) and probes.

Some real public databases are: CAIDA[2], which contains anonymized passive traffic traces from high-speed Internet backbone links, and the traffic data repository, maintained by the MAWI[3] Working Group of the WIDE Project. Other researchers have created their own databases in different universities, e.g., Carnegie Mellon University, Xi'an Jiaotong University, and Clemson University (GENI[6]), or traffic collected from backbone in SWITCH, Abilene, and Géant.

Nowadays, there is no public database large enough to exhaustively test and compare different algorithms in order to extract significant conclusions about their performances and their capabilities of classification.

3. Feature Extraction

Motoda H. and Liu H. (2002) [7] state that feature selection is a process that chooses a subset of M features from the original set of N features $M \leq N$ so that the feature space is optimally reduced according to a certain criterion [8,9]. Feature extraction is a process that extracts a set of new

---

1 http://www.ll.mit.edu/ideval/index.html
2 https://www.caida.org/data/passive/passive_2012_dataset.xml
3 http://mawi.wide.ad.jp/mawi/

features from the original features through some functional mapping [10]. Assuming that there are N features $Z_1, Z_2, \ldots, Z_N$ after feature extraction, another set of new features $X_1, X_2, \ldots, X_M (M < N)$ is obtained via the mapping functions F_i, i.e. $X_i = F_i(Z_1, Z_2, \ldots, Z_N)$.

Among the algorithms used to reduce the number of features in network traffic anomaly detection are: PCA [11], Mutual Information and linear correlation [12], decision tree [13], and maximum entropy [14].

In network traffic, the most commonly employed features are [2,15–19]: source and destination IP addresses and source and destination port numbers. Other features extracted from headers are: protocol field, number of bytes, service, flag, and country code. Zhang et al. (2009) [20] divided the size of packets into seven types and Gu et al. (2005) [21] defined 587 packet classes based on the port number.

At flow[4] level the features selected were: flow duration, flow size distribution (FSD), and average packet size per flow. For KDD Cup 99, 41 features or a subset were employed [12,13]. On the other hand, Tellenbach et al. (2011) [22] used source port, country code and others, constructing the TES as input data.

3.1. Windowing in Network Traffic

Window-based methods group consecutive packets or flows based on a sliding window. The ith window of size L packets is represented as $W_i(L, \tau) = \{pack_k, pack_{k+1}, \ldots, pack_{k+L}\}$, with $k = iL - i\tau$, where τ is the overlapping and $\tau \in \{0, 1, \ldots, L-1\}$. When the window size is given by time, L can be different in each window. Windowing is performed in two ways: overlapping ($\tau \neq 0$) and non overlapping ($\tau = 0$) windows. The window sizes most commonly used are: 5 min [15,16,23–25], 30 min, 1 min, 100 sec, 5 sec and 0.5 sec. Some researchers use windows with a fixed length $L = 4096$ [19], 1000 [26], and 32 [4] packets. Therefore, the main objective of windowing is to reduce the data volume.

4. Entropy Concepts Used in Network Traffic Anomaly Detection

Let X be a random variable which takes values of the set $\{x_1, x_2, ..., x_M\}$, $p_i := P(X = x_i)$ the probability of occurrence of x_i, and M the cardinality of the finite set; hence, the Shannon entropy is:

$$H^S(X) = -\sum_{i=1}^{M} p_i \, log\,(p_i)\,. \tag{1}$$

Based on the Shannon entropy [27], Rényi [28] and Tsallis [29] defined generalized entropies, which are related to the q-deformed algebra. The Rényi entropy is defined as:

$$H^R(X, q) = \frac{1}{1-q} log\left(\sum_{i=1}^{M} p_i^q\right) \tag{2}$$

and the Tsallis entropy is

[4] An IP flow corresponds to an IP port-to-port traffic exchanged between two IP addresses during a period of time T.

$$H^T(X,q) = \frac{1}{q-1}\left(1 - \sum_{i=1}^{M} p_i^q\right), \tag{3}$$

when $q \to 1$ the generalized entropies are reduced to Shannon entropy. In order to compare the changes of entropy at different times, the entropy is normalized, i.e.,

$$\bar{H}(X) = \frac{H(X)}{H_{max}(X)}. \tag{4}$$

For generalized entropies, the values $q = 0.9$ and $q = 1.1$ to detect DoS and DDos atacks have been used by [30,31], respectively. The sets $q \in \{-3\} \cup \{-2, -1.75, ..., 1.75, 2\}$, $q = \{1, 2, 3, \ldots, 15\}$ and $q = \{1, 2, 3, \ldots, 10\}$ were used to detect DDos and scanning attacks, and low-rate and high-rate DDoS by [22,32,33].

4.1. Kullback-Leibler divergence

Consider two complete discrete probability distributions $P = (p_1, p_2, \ldots, p_n)$ and $Q = (q_1, q_2, \ldots, q_n)$, with $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1$, $1 \geq p_i \geq 0$, $1 \geq q_i \geq 0$, $i = 1, 2, \ldots, n$. The information divergence is a measure of the divergence between P and Q and is defined by [28]:

$$D_\rho(P||Q) = \frac{1}{\rho - 1} log\left(\sum_{i=1}^{n} p_i^\rho q_i^{1-\rho}\right), \quad \rho \geq 0, \tag{5}$$

where ρ is the order of the information divergence. Consequently, the smaller $D_\rho(P||Q)$ is, the closer the distributions P and Q are. $D_\rho(P||Q) = 0$ iff $P = Q$. When $\rho \to 1$ the Kullback-Leibler (KL) divergence [34] is obtained

$$D_1(P||Q) = \sum_{i=1}^{n}\left(p_i \, log\left(\frac{p_i}{q_i}\right)\right), \ \rho \to 1. \tag{6}$$

4.2. Mutual information

The conditional entropy of a variable Y given X, with alphabet $\mathfrak{X}$ and $\mathfrak{Y}$, respectively, is defined as:

$$H(Y|X) = -\sum_{x\in\mathfrak{X}} p(x) \sum_{y\in\mathfrak{Y}} p(y|x) \, log\left(p(y|x)\right) \tag{7}$$

$$= -\sum_{x\in\mathfrak{X}} \sum_{y\in\mathfrak{Y}} p(x,y) \, log\left(p(y|x)\right). \tag{8}$$

The mutual information (MI) [35] between two random variables X and Y is a measure of the amount of knowledge of Y supplied by X or vice versa. If X and Y are independent, then their mutual information is zero. The MI of two random variables X and Y is defined as:

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X;Y) \tag{9}$$

where $H(\bullet)$ is entropy, $H(X|Y)$ and $H(Y|X)$ are conditional entropies, $H(X;Y)$ is the joint entropy of X and Y, defined as

$$H(X;Y) = -\sum_{x\in\mathfrak{X}}\sum_{y\in\mathfrak{Y}} p(x,y)\, log\,(p(x,y)) \tag{10}$$

where $p(x,y)$ is the joint probability mass function.

The MI equation can be written as:

$$I(X;Y) = \sum_{x\in\mathfrak{X}}\sum_{y\in\mathfrak{Y}} p(x,y)\, log\left(\frac{p(x,y)}{p(x)p(y)}\right) \tag{11}$$

where $p(x)$ and $p(y)$ are marginal probability mass functions of X and Y, respectively. In order to estimate the MI between X, Y, it is necessary to estimate $p(x,y)$.

4.3. Entropy calculation

As the full probability distribution is generally not known or not completely known, different probability estimators are used, e.g., relative frequency, Balanced, and Balanced II, and consequently, a "true" probability distribution is built. The entropy is calculated using these estimators; the more accurate the estimators, the better the entropy estimates.

Rahmani et al. (2009) [36] noted that time series of IP-flow number and aggregate traffic size are strongly statistically dependent, and when an attack occurs, it causes a rupture in the time series of joint entropy values. In order to calculate the joint entropy $H(X;Y)$ they employed $p(x,y)$ of the time series X and Y using either the Gamma density probability function (when the number of connections was small) or the central limit theorem (when the number of connections was large enough). Liu et al. (2010) [18] calculated the conditional entropy $H(Y|X)$ where Y and X are two of the most widely used traffic variables: source and destination Ip addresses.

Amiri et al. (2011) [12] used an estimator of MI developed by Kraskov [37], which employs entropy estimates from k-nearest neighbors distances. Velarde-Alvarado et al. (2009) [2] estimated entropy values using the balanced estimator II as a probability estimator.

5. Anomaly detection

An anomaly in network traffic is a data pattern that does not conform to those representing a normal traffic behavior. Therefore, anomaly detection is a broad field, where numerous anomaly detection methods are used for different applications.

Assuming that 1) $X \in \mathbb{R}^p$ is a p-dimensional real-valued random variable with a domain $S_X \subset \mathbb{R}^p$ representing traffic features, 2) x_i are instances of X, i.e. $x_i \in S_X$, and 3) data patterns of normal behavior are represented by the subspace $S_N \subset S_X$, anomaly detection determines whether an instance x_i belongs to S_N or not.

The space S_X can be partitioned or divided into classes with the help of decision functions, allowing further classification.

5.1. Specific Decision Functions

The KL divergence $D_1(P||Q)$ is used as decision functions in [20,21]. In addition, [20] classified the abnormal situations into different classes. In [38], both KL divergence and Maximum entropy are used. Yan et al. (2008) [25] used $D_{0.5}(P||Q)$.

In Santiago-Paz et al. (2015) [4], a decision function is based on the Mahalanobis distance [39] $\mathbf{d}_M^2(\mathbf{x}_i)$, and a second decision function is given by $f(\mathbf{x}_i) = \sum_i^N \alpha_i k(\mathbf{x}_i, \mathbf{x}) - b$ for One Class-Support Vector Machine (OC-SVM), where $k(\mathbf{x}_i, \mathbf{x})$ is a kernel. Huang et al. (2006) [40] computed the Rényi entropy ($q = 3$) of the Coiflet and Daubechies wavelets.

In Velarde-Alvarado et al. (2009) [2], used the proportional uncertainty (PU) and the method of remaining elements (MRE) to detect anomalies. Tellenbach et al. (2011) [22] used Kalman filter, PCA, and KLE as anomaly detection methods. Ma et al. (2014) [41] established a function decision based on the entropy of the source IP address $\hat{H}_s$ and the entropy of the destination IP address $\hat{H}_d$. In [42,43], a function decision based on entropy and a range of values was used to detect anomalies.

The use of the entropy allows the A-NIDS to achieve high detection rates, see table 2. In addition, new measures based on entropy should be studied and used as a basis for other decision functions.

Table 2. Results of network traffic anomaly detection using entropy.

Author	Information metric	Database	Anomaly	TNR [%]
Gu et al. (2005)	KL divergence	–	Portscan	91.0
Zhang et al. (2009)	KL divergence	MIT-DARPA	DoS	87.10
			Probe	68.18
			R2L	79.49
			U2R	60.87
Liu et al.(2010)	Conditional entropy	CAIDA	DDoS	93.0
Ferreira et al. (2011)	Shannon, Rényi, Tsallis	KDD Cup 99	DoS	93.18
			Probe	79.20
			R2L	97.76
			U2R	95.05
Amiri et al. (2011)	Mutual Information	KDD Cup 99	DoS	90.02
			Probe	99.97
			R2L	99.98
			U2R	95.0
Santiago-Paz et al.(2015)	Shannon, Rényi, Tsallis	LAN, MIT-DARPA subset	Worms, DoS, Portscan	99.83

6. Classification

Gupta et al. (2014) [44] state that given: 1) a training data set of the form $\{(x_i, \mathbf{y}_i)\}$, where $x_i \in S_X$ is a feature vector or data pattern and $\mathbf{y}_i \subset \{1, \ldots, G\}$ is the subset of the G class labels that are known to be correct labels for x_i, 2) a discriminant function $f(x; \beta_g)$ with class-specific parameters β_g for each class with $g = 1, \ldots, G$; then class discriminant functions are used to classify an instance x as the class label that solves $arg\, max_g f(x; \beta_g)$.

Lakhina et al. (2005) [16] apply two clustering algorithms: k-means and hierarchical agglomeration, using a vector $\tilde{\mathbf{h}} = [\tilde{\mathbf{H}}(\text{srcIP}), \tilde{\mathbf{H}}(\text{dstIP}), \tilde{\mathbf{H}}(\text{srcPort}), \tilde{\mathbf{H}}(\text{dstPort})]$. Xu et al., (2005) [23] define three "free" feature dimensions and introduce an *"Entropy-based Significant Cluster Extraction Algorithm"* for clustering.

Lima et al. (2011) [13] use the WEKA[5] Simple K-Means algorithm , which employs Euclidean distance as a measure to compute distances between instances and clusters. Support Vector Machine is applied by Tellenbach et al. (2011) [22] to classify the anomalies. Yao et al. (2012) [45] use the Random Forests Test.

Santiago-Paz et al. (2014) [19] present the *Entropy and Mahalanobis Distance (EMD) based Algorithm* to define elliptical regions in the feature space. In [4], OC-SVM and k-temporal nearest neighbors are used to improve accuracy in classification.

6.1. The Classifier Metrics

Given a classifier and an instance, there are four possible outcomes:[6] TN, FP, FN, and TP. With these entries, the following statistics are computed [46]: Accuracy (AC) is the proportion of the total number of predictions that were correct: $AC = \frac{TN+TP}{TN+FP+FN+TP}$; True Positive Rate (TPR) is the proportion of positive cases that were correctly identified: $TPR = \frac{TP}{FN+TP}$; True Negative Rate (TNR) is the proportion of negative cases that were classified correctly: $TNR = \frac{TN}{TN+FP}$; False Negative Rate (FNR) is the proportion of positive cases that were incorrectly classified as negative: $FNR = \frac{FN}{FN+TP}$; and F-measure is a measure of a test's accuracy: $F\text{-measure} = \frac{2*TPR*AC}{TPR+AC}$. In addition, Receiver Operating Characteristic[7] (ROC) graphs illustrate the performance of a classifier.

Although, there are several types of distances, classifiers based on new closeness and farness measures of data patterns and pattern clusters should be studied.

7. Conclusions

This paper provides a structured and comprehensive overview of the state-of-the-art in entropy-based A-NIDS. Using a general architecture of Entropy-based A-NIDS, the different techniques proposed in the state-of-the-art of the main modules are shown. The measures of information used by researchers and the most commonly employed metrics for testing the performance of the detection and classification are presented. The achieved high detection rates prove the effective use of entropy.

Open Issues

---

[5] http://www.cs.waikato.ac.nz/ml/weka/

[6] TN is the number of correct predictions that an instance is negative, FP is the number of incorrect predictions that an instance is positive, FN is the number of incorrect predictions that an instance is negative, and TP is the number of correct predictions that an instance is positive.

[7] ROC graphs are two-dimensional graphs in which an (FP rate, TP rate) pair corresponding to a single point in Receiver Operating Characteristic space.

Nowadays, there is no public database large enough to exhaustively test and compare different algorithms in order to extract significant conclusions about their performances and their capabilities of classification. Therefore, the construction of a common database with real "normal" and anomalous traffic for the evaluation of A-NIDS is needed.

The value of the q parameter for generalized entropies is found experimentally; its correct choice for the best anomaly detection is an open research problem.

For different networks, the larger the slot size, the more different the entropy behaviors. In the near future, this behavior including more and recent traces in order to determine whether the learned model from a certain network can be used in a different network should be addressed.

Another open issue is related to the adequate window size for reducing the data volume, ensuring good entropy estimates and early detection of anomalies.

The set of decision functions and classifiers with new closeness and farness entropy-based measures should be enhanced.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Bonachela, J.A.; Hinrichsen, H.; Munoz, M.A. Entropy estimates of small data sets. *Journal of Physics A: Mathematical and Theoretical* **2008**, *41*, 202001.
2. Velarde-Alvarado, P.; Vargas-Rosales, C.; Torres-Roman, D.; Martinez-Herrera, A. Detecting anomalies in network traffic using the method of remaining elements. *Communications Letters, IEEE* **2009**, *13*, 462–464.
3. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly Detection: A Survey. *ACM Comput. Surv.* **2009**, *41*, 15:1–15:58.
4. Santiago-Paz, J.; Torres-Roman, D.; Figueroa-Ypiña, A.; Argaez-Xool, J. Using Generalized Entropies and OC-SVM with Mahalanobis Kernel for Detection and Classification of Anomalies in Network Traffic. *Entropy* **2015**, *17*, 6239.
5. Bhuyan, M.H.; Bhattacharyya, D.K.; Kalita, J.K. Network anomaly detection: methods, systems and tools. *Communications Surveys & Tutorials, IEEE* **2014**, *16*, 303–336.
6. Berman, M.; Chase, J.S.; Landweber, L.; Nakao, A.; Ott, M.; Raychaudhuri, D.; Ricci, R.; Seskar, I. GENI: A federated testbed for innovative network experiments. *Computer Networks* **2014**, *61*, 5 – 23. Special issue on Future Internet Testbeds âĂŞ Part I.
7. Motoda, H.; Liu, H. Feature selection, extraction and construction. *Communication of IICM (Institute of Information and Computing Machinery, Taiwan) Vol* **2002**, *5*, 67–72.
8. Blum, A.L.; Langley, P. Selection of relevant features and examples in machine learning. *Artificial intelligence* **1997**, *97*, 245–271.
9. Dash, M.; Liu, H. Feature selection for classification. *Intelligent data analysis* **1997**, *1*, 131–156.
10. Wyse, N.; Dubes, R.; Jain, A.K. A critical evaluation of intrinsic dimensionality algorithms. *Pattern recognition in practice* **1980**, pp. 415–425.

11. Kanda, Y.; Fontugne, R.; Fukuda, K.; Sugawara, T. ADMIRE: Anomaly detection method using entropy-based PCA with three-step sketches. *Computer Communications* **2013**, *36*, 575–588.
12. Amiri, F.; Yousefi, M.R.; Lucas, C.; Shakery, A.; Yazdani, N. Mutual information-based feature selection for intrusion detection systems. *Journal of Network and Computer Applications* **2011**, *34*, 1184–1199.
13. Lima, C.F.L.; Assis, F.M.; De Souza, C.P. A comparative study of use of Shannon, Rényi and Tsallis entropy for attribute selecting in network intrusion detection. Measurements and Networking Proceedings (M&N), 2011 IEEE International Workshop on. IEEE, 2011, pp. 77–82.
14. Ashfaq, A.B.; Rizvi, S.; Javed, M.; Khayam, S.A.; Ali, M.Q.; Al-Shaer, E. Information theoretic feature space slicing for statistical anomaly detection. *Journal of Network and Computer Applications* **2014**, *41*, 473–487.
15. Wagner, A.; Plattner, B. Entropy Based Worm and Anomaly Detection in Fast IP Networks. In Proceedings of 14th IEEE WET ICE / STCA security workshop. IEEE, 2005, pp. 172–177.
16. Lakhina, A.; Crovella, M.; Diot, C. Mining Anomalies Using Traffic Feature Distributions. *SIGCOMM Comput. Commun. Rev.* **2005**, *35*, 217–228.
17. Chang, S.; Qiu, X.; Gao, Z.; Qi, F.; Liu, K. A flow-based anomaly detection method using entropy and multiple traffic features. Broadband Network and Multimedia Technology (IC-BNMT), 2010 3rd IEEE International Conference on. IEEE, 2010, pp. 223–227.
18. Liu, Y.; Yin, J.; Cheng, J.; Zhang, B. Detecting DDoS attacks using conditional entropy. Computer Application and System Modeling (ICCASM), 2010 International Conference on. IEEE, 2010, Vol. 13, pp. V13–278.
19. Santiago-Paz, J.; Torres-Roman, D. Characterization of worm attacks using entropy, Mahalanobis distance and K-nearest neighbors. Electronics, Communications and Computers (CONIELECOMP), 2014 International Conference on, 2014, pp. 200–205.
20. Zhang, Y.l.; Han, Z.g.; Ren, J.x. A network anomaly detection method based on relative entropy theory. Electronic Commerce and Security, 2009. ISECS'09. Second International Symposium on. IEEE, 2009, Vol. 1, pp. 231–235.
21. Gu, Y.; McCallum, A.; Towsley, D. Detecting anomalies in network traffic using maximum entropy estimation. Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement. USENIX Association, 2005, pp. 32–32.
22. Tellenbach, B.; Burkhart, M.; Schatzmann, D.; Gugelmann, D.; Sornette, D. Accurate Network Anomaly Classification with Generalized Entropy Metrics. *Comput. Netw.* **2011**, *55*, 3485–3502.
23. Xu, K.; Zhang, Z.L.; Bhattacharyya, S. Profiling Internet Backbone Traffic: Behavior Models and Applications. *SIGCOMM Comput. Commun. Rev.* **2005**, *35*, 169–180.
24. Nychis, G.; Sekar, V.; Andersen, D.G.; Kim, H.; Zhang, H. An Empirical Evaluation of Entropy-based Traffic Anomaly Detection. Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement; ACM: New York, NY, USA, 2008; IMC '08, pp. 151–156.
25. Yan, R.; Zheng, Q.; Peng, W. Multi-scale entropy and renyi cross entropy based traffic anomaly detection. Communication Systems, 2008. ICCS 2008. 11th IEEE Singapore International Conference on. IEEE, 2008, pp. 554–558.

26. Quan, Q.; Hong-Yi, C.; Rui, Z. Entropy based method for network anomaly detection. Dependable Computing, 2009. PRDC'09. 15th IEEE Pacific Rim International Symposium on. IEEE, 2009, pp. 189–191.
27. Shannon, C.E. A Mathematical Theory of Communication. *Bell System Technical Journal* **1948**, *27*, 379–423.
28. Rényi, A. On measures of entropy and information. Fourth Berkeley symposium on mathematical statistics and probability, 1961, Vol. 1, pp. 547–561.
29. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics* **1988**, *52*, 479–487.
30. Ziviani, A.; Gomes, A.T.A.; Monsores, M.L.; Rodrigues, P.S. Network anomaly detection using nonextensive entropy. *Communications Letters, IEEE* **2007**, *11*, 1034–1036.
31. Ma, X.; Chen, Y. DDoS Detection method based on chaos analysis of network traffic entropy. *Communications Letters, IEEE* **2014**, *18*, 114–117.
32. Bhuyan, M.H.; Bhattacharyya, D.; Kalita, J. An empirical evaluation of information metrics for low-rate and high-rate {DDoS} attack detection. *Pattern Recognition Letters* **2015**, *51*, 1 – 7.
33. Xiang, Y.; Li, K.; Zhou, W. Low-rate DDoS attacks detection and traceback by using new information metrics. *Information Forensics and Security, IEEE Transactions on* **2011**, *6*, 426–437.
34. Kullback, S.; Leibler, R.A. On information and sufficiency. *The Annals of Mathematical Statistics* **1951**, pp. 79–86.
35. Cover, T.M.; Thomas, J.A. *Elements of information theory*; John Wiley & Sons, 2012.
36. Rahmani, H.; Sahli, N.; Kammoun, F. Joint entropy analysis model for DDoS attack detection. Information Assurance and Security, 2009. IAS'09. Fifth International Conference on. IEEE, 2009, Vol. 2, pp. 267–271.
37. Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Physical review E* **2004**, *69*, 066138.
38. Coluccia, A.; DâĂŹAlconzo, A.; Ricciato, F. Distribution-based anomaly detection via generalized likelihood ratio test: A general Maximum Entropy approach. *Computer Networks* **2013**, *57*, 3446–3462.
39. Mahalanobis, P.C. On the generalised distance in statistics. Proceedings National Institute of Science, India, 1936, Vol. 2, pp. 49–55.
40. Huang, C.T.; Thareja, S.; Shin, Y.J. Wavelet-based real time detection of network traffic anomalies. Securecomm and Workshops, 2006. IEEE, 2006, pp. 1–7.
41. Ma, X.; Chen, Y. DDoS Detection method based on chaos analysis of network traffic entropy. *Communications Letters, IEEE* **2014**, *18*, 114–117.
42. Bereziński, P.; Jasiul, B.; Szpyrka, M. An Entropy-Based Network Anomaly Detection Method. *Entropy* **2015**, *17*, 2367–2408.
43. Özçelik, İ.; Brooks, R.R. Deceiving entropy based DoS detection. *Computers & Security* **2015**, *48*, 234–245.
44. Gupta, M.R.; Bengio, S.; Weston, J. Training highly multiclass classifiers. *The Journal of Machine Learning Research* **2014**, *15*, 1461–1492.

45. Yao, D.; Yin, M.; Luo, J.; Zhang, S. Network Anomaly Detection Using Random Forests and Entropy of Traffic Features. Multimedia Information Networking and Security (MINES), 2012 Fourth International Conference on. IEEE, 2012, pp. 926–929.
46. Kohavi, R.; provost, F. Glossary of Terms. *Mach. Learn.* **1998**, *30*, 271–274.