On the Inefficient Use of Entropy for Anomaly Detection

Mobin Javed[1], Ayesha Binte Ashfaq[1], M. Zubair Shafiq[2], Syed Ali Khayam[1]

[1] National University of Sciences & Technology, Islamabad, Pakistan
[2] nexGIN RC, FAST National University, Islamabad, Pakistan
{mobin.javed,ayesha.ashfaq,ali.khayam}@seecs.edu.pk, zubair.shafiq@nexginrc.org

Abstract. Entropy-based measures have been widely deployed in anomaly detection systems (ADSes) to quantify behavioral patterns [1]. The entropy measure has shown significant promise in detecting diverse set of anomalies present in networks and end-hosts. We argue that the full potential of entropy-based anomaly detection is currently not being exploited because of its inefficient use. In support of this argument, we highlight three important shortcomings of existing entropy-based ADSes. We then propose efficient entropy usage – supported by preliminary evaluations – to mitigate these shortcomings.

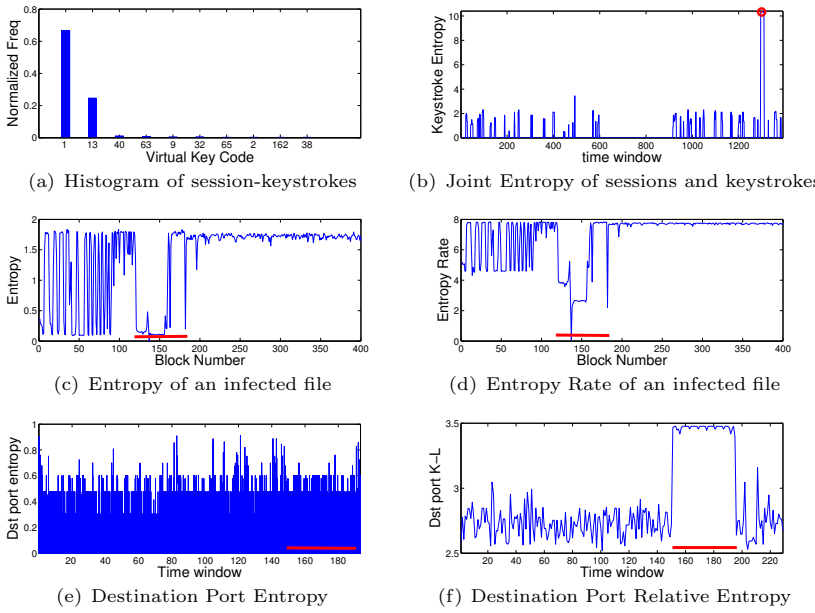
1 Entropy Limitations and Countermeasures

1.1 Feature correlation should be retained

Current ADSes perform entropy analysis on *marginal distributions* of features. In general, significant correlation exists across traffic and/or host features which is not being leveraged by these ADSes. As a proof-of-concept example, we propose to detect malicious network sessions by noting that the histogram of keystrokes which are used to initiate network sessions is skewed [see Fig. 1(a)] and perturbation in this metric can easily reveal the presence of an anomaly; network traffic and keystroke data were collected before and after infecting a human-operated computer with the low-rate `Rbot-AQJ` worm. While analyzing the entropies of the marginal keystroke distribution and/or the marginal session distribution is clearly not useful, Fig. 1(b) shows that quantifying these features using joint (session-keystroke) entropy can easily detect anomalous activity.

1.2 Spatial/temporal correlation should be retained

Another limitation of the entropy measure is its inability to take spatial/temporal correlation of benign patterns into account. Such correlations can prove useful in the detection of subtle anomalies. For instance, Fig. 1(c) shows the block-wise (block size = 1KB) entropy of a `PDF` file which is infected by an embedded executable malware. It is evident that entropy is unable to provide clear perturbations required for detection. On the other hand, entropy rate [Fig. 1(d)], which models and accounts for the spatial/temporal correlation, provides very clear perturbations at the infected file blocks; entropy rate quantifies the average entropy of conditional distributions [2].



(a) Histogram of session-keystrokes (b) Joint Entropy of sessions and keystrokes

(c) Entropy of an infected file (d) Entropy Rate of an infected file

(e) Destination Port Entropy (f) Destination Port Relative Entropy

Fig. 1. Examples to support the limitations of the current use of entropy.

1.3 Randomness quantification is not enough

Entropy cannot distinguish between differing distributions with the same amount of uncertainty; e.g., entropy of the normalized distribution of a source producing 90 packets on 80 and 10 packets on port 21 is the same as a source producing 900 packets on port 6666 and 100 packets on port 6667. Thus anomalies which do not perturb randomness go undetected. Fig. 1(e) shows a case where the Blaster worm cannot be detected in the destination port entropy time-series; endpoint traffic dataset from [3] is used for this experiment. This limitation arises due to the fact that entropy does not take the individual port numbers into account. It is, therefore, important to perform a symbol-by-symbol comparison between benign and observed distributions. This can be achieved by computing the *relative entropy* of the distributions. Fig. 1(f) shows that K-L divergence time series of destination port is perturbed due to the presence of `Blaster` worm. We argue that relative entropy is the main reason for the Maximum Entropy Detector's [4] superior accuracy [3].

References

1. A Bibliography of Entropy-based Anomaly Detectors, `http://www.wisnet.seecs.edu.pk/Bibliographies/Bibliography_Entropy_Based_ADSes.pdf`.
2. Shafiq, M.Z., Khayam, S.A., Farooq, M.:Embedded Malware Detection using Markov n-grams.In: DIMVA (2008)
3. Ashfaq, A.B., Robert, M.J., Mumtaz, A., Ali, M.Q., Sajjad, A., Khayam, S.A.:A Comparative Analysis of Anomaly Detectors under Portscan Attacks.In:RAID (2008)
4. Gu, Y., McCullum, A., Towsley, D.: Detecting anomalies in network traffic using maximum entropy estimation. In: ACM/Usenix IMC (2005)