Detecting Anomalies in Network Traffic Using the Method of Remaining Elements

P. Velarde-Alvarado, C. Vargas-Rosales, *Senior Member, IEEE*, D. Torres-Roman, and A. Martinez-Herrera

***Abstract*—Attacks, such as port scans, DDoS and worms, threaten the functionality and reliability of IP networks. Early and accurate detection is crucial to mitigate their impact. We use the Method of Remaining Elements (MRE) to detect anomalies based on the characterization of traffic features through a proportional uncertainty measure. MRE has the functionality and performance to detect abnormal behavior and serve as the foundation for next generation network intrusion detection systems.**

***Index Terms*—Anomaly detection, traffic anomalies, entropy based intrusion detection.**

I. Introduction

TRADITIONAL security measures, like firewalls or anti-virus solutions, are not sufficient for the variety and sophistication of attacks. Early detection of potential attacks is crucial to mitigate their impact. Intrusion Detection Systems (IDS) based on entropy, [1], can be effective because malicious activities change the network traffic nature. Anomalies are characterized by unusual and significant changes in patterns of network activities disrupting behavior of traffic features. The entropy of *intrinsic features* extracted from packet headers e.g., source IP, destination IP, source port and destination port numbers, [2], does not provide sufficient sensitivity to detect some attacks that are short-term or distributed over time.

We propose the use of *proportional uncertainty* (PU) to determine the remaining values of sequences of those intrinsic features, since it provides better sensitivity to define the cutoff between remnants and significant elements than that of the *relative uncertainty* (RU) in [3]. Our results indicate that by adjusting time-slot duration and cutoff threshold β in the remaining calculations, anomalies are exposed with relatively high levels of remnant elements with respect to typical behavior.

II. Measures of Entropy

The application of Shannon's entropy, [4], in anomaly detection has disadvantages since short-term or attacks distributed over time are not clearly detected because uncertainty is either negligible or distributed as well, then, we propose the use of a modified version of the low-bias balanced estimator in [5]. For a discrete dataset X, which can take a finite number, M, of possible values $X = \{x_1, \ldots, x_M\}$; the balanced estimator for a data set of size N, is defined as

$$\widehat{H}_S^{bal}(X) = \frac{1}{N+2}\sum_{k=1}^{M}\left[(n_k+1)\sum_{j=n_k+2}^{N+2}\frac{1}{j}\right], \quad (1)$$

with n_k number of counts value x_k appears in the set. The second summation in (1) is a partial harmonic series using the Euler-Mascheroni constant, [6], $\gamma = 0.5772156649\cdots$, with asymptotic expansion of the n-th harmonic number as $\mathrm{H}_n \sim \log(n) + \gamma + (1/2)n^{-1} - (1/12)n^{-2} + (1/120)n^{-4} - (1/250)n^{-6} + \cdots$, which gives the following

$$\begin{aligned}\sum_{j=n_k+2}^{N+2}\tfrac{1}{j} &= \sum_{j=1}^{N+2}\tfrac{1}{j} - \sum_{j=1}^{n_k+1}\tfrac{1}{j}\\ &= \mathrm{H}_{N+2} - \mathrm{H}_{n_k+1}\\ &= \log\left(\tfrac{N+2}{n_k+1}\right) + \rho_{N+2} - \rho_{n_k+1},\end{aligned} \quad (2)$$

where $(\rho_{N+2} - \rho_{n_k+1})$ approaches zero when N and n_k increase indefinitely. This simplification gives us a more efficient computational formula expressed as

$$\widehat{H}^{bal-II}(X) = \frac{1}{N+2}\sum_{k=1}^{M}\left[(n_k+1)\log\left(\frac{N+2}{n_k+1}\right)\right]. \quad (3)$$

Maximum uncertainty in (3) occurs when $n_k = 1$ for $k = 1, 2, \ldots, M$, which results in $N = M$, hence from (3), we get

$$\widehat{H}_{MAX}^{bal-II}(X) = \frac{2M}{M+2}\log\left(\frac{M+2}{2}\right), \quad (4)$$

which exceeds the upper bound $\log(M)$ in Shannon's entropy. Equation (3) is our *measure of entropy*, [7], and has a significant effect on data sets for which M is closed to N, which is related to anomalous activities, i.e., when high diversity in IP addresses or port numbers occurs, (e.g., port and IP scanning or DoS attacks). Using (3) to calculate the PU in MRE, we achieve a superior and controlled anomaly exposure than that using Shannon's entropy estimator as in [3], and in [8].

III. The Method of Remaining Elements

A. Proportional Uncertainty

PU provides an index of uncertainty with respect to the maximum Shannon's entropy, i.e., the ratio of (3) to $\log(M)$. Considering (4), we can see that such ratio will be bounded above, thus for data set X and taking the limit as the alphabet size increases, we define PU as,

$$PU(X) = \frac{\widehat{H}^{bal-II}(X)}{\log(M)} \leq \lim_{M\to\infty}\frac{\frac{2M}{M+2}\log\left(\frac{M+2}{2}\right)}{\log(M)} = 2, \quad (5)$$

Manuscript received March 22, 2009. The associate editor coordinating the review of this letter and approving it for publication was G. Lazarou.

This work was partially sponsored by SEP-CONACyT project 61183 and CONACyT project 67360Y.

C. Vargas-Rosales and A. Martinez-Herrera are with the Center for Electronics and Telecommunications, ITESM-Campus Monterrey, Monterrey, N.L., 64849, Mexico (e-mail: cvargas@itesm.mx).

P. Velarde-Alvarado is with Autonomous University of Nayarit (e-mail: pvelarde@nayar.uan.mx).

D. Torres-Roman is with CINVESTAV Guadalajara, Mexico (e-mail: dtorres@gdl.cinvestav.mx).

Digital Object Identifier 10.1109/LCOMM.2009.090689

ISSN: 1089-7798