Dynamic Modeling of Internet Traffic for Intrusion Detection

Khushboo Shah

Nevis Networks

Mountain View, CA 94043

kshah@nevisnetworks.com

Edmond Jonckheere

Univeristy of Southern California

Dept. of Electrical Eng.

Los Angeles, CA-90089

jonckhee@usc.edu

Stephan Bohacek

Univeristy of Delaware

Dept. of Electrical and Computer Eng.

Newark, DE-19711

bohacek@eecis.udel.edu

Abstract

Computer network traffic is analyzed via mutual information techniques, implemented using linear and nonlinear canonical correlation analyses, with the specific objective of detecting UDP flooding attacks. NS simulation of HTTP, FTP and CBR traffic shows that flooding attacks are accompanied by a change of mutual information, either at the link being flooded or at another upstream or downstream link. This observation appears topology independent, as the technique is demonstrated on the so-called parking-lot topology, random 50-node topology, and 100-node transit-stub topology. This technique is also employ to detect UDP flooding with low false alarm rate on a backbone link. These results indicate that a change in mutual information provides a useful detection criterion when no other signature of the attack is available.

I. INTRODUCTION

Attacks on the network have become commonplace and with them intrusion detection systems (IDS's), firewalls, virus scanning, and the like have become parts of an ever growing arsenal of defense tools [MVS01], [Ken00]. If some knowledge of the nature of the attack is available, it would be easily recognizable by pattern recognition techniques. Hence, signature-based IDS is perhaps the popular IDS technique [Pax99], [Roe99]. However, when a new attack strikes, no such signature is available, in which case the only hope is through anomaly detection [SJJ02], meaning detection of some deviation of the overall system behavior from what is considered normal. Anomaly detection can be host based or network based. Host based anomaly detection is at the end user level, while network based detection is at the level of network data. The present paper is relevant to the latter, in the sense that it detects intrusion by analysis of the signals at some link.

Within network-based anomaly detection, most techniques are *count based* where the rate of occurrence (i.e., the number of events in a time period) or the absolute value of some count are monitored. A sufficiently large deviation of the count from its nominal value is assumed to signify an attack. Change-point detection schemes such as cumsum [BN93] or exponentially weighted moving average may be used to detect when the deviation of the count occurs [WZS04]. For example, TCP-SYN attacks are detected by monitoring the arrival rate of TCP-SYN packets or the number of half-open connections (e.g., [SP04]). Email worms can be detected by monitoring the number of emails sent from a mail server and by examining the number of
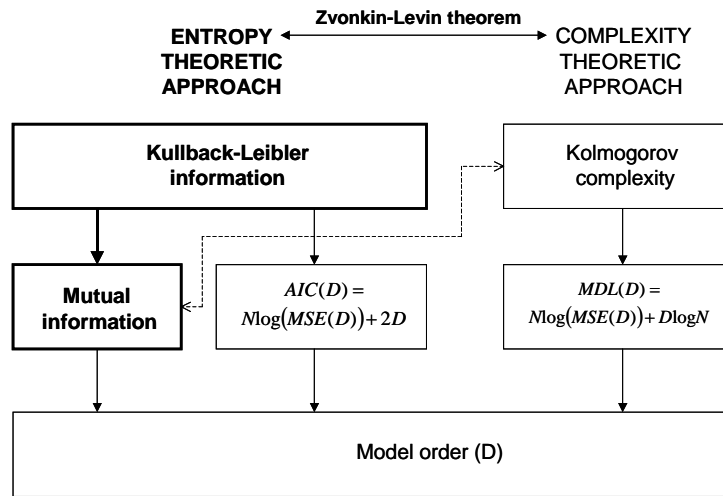
Fig. 1.   The various approaches to detect a change in the signal structure. The path taken here is boldfaced. In the Akaike Information Criterion (AIC) and the Minimum Description Length (MDL), the model order D is chosen so as to minimize AIC or MDL, respectively, where MSE denotes the mean square error and N the number of sample sets.

emails sent to certain classes of destinations [WBMW04]. The rate of DNS lookups [WKO05b] and ARP requests [WKO05a] are used to detect various types of worms. The arrival rate of certain sized UDP packets can be used to detect worms such as Code Red (e.g. [MPR02]).

The paper presents an alternative to count-based anomaly detection. More specifically, we investigate intrusion detection that is based on a possibly subtle change relevant to the *dynamical structure* of the signal. Arguably that single parameter that best encodes this dynamical structure is the order of the model of the observed time series. As already noted in [SBJ03], this model order can be obtained by either the Akaike Information Criterion (AIC) or the Minimum Description Length (MDL) criterion. The former is a Kullback-Leibler based criterion, while the latter is a Kolmogorov complexity based criterion [WD99]. A third avenue of approach utilizes the Kullback-Leibler information in a different way to produce the Akaike mutual information (MI) between past and future of the time series; model order selection is then viewed as a compromise between simplicity of the model and its ability to carry most of the mutual information; this is computationally implemented in stochastic balancing (see [**?**] and the references cited therein). The interrelation among these three approaches is depicted in Figure 1. The left hand side of the diagram refers to properties of the *statistics*, whereas the right hand side refers to properties of *sequences*. The deeper connection between the two approaches is formulated by the Zvonkin-Levin theorem [ZL70b], [SE03, Theorem 1],[Man77, p. 227]: For a stationary ergodic source emitting symbols $y(k)$ over a finite alphabet, $\lim_{n \to \infty} \frac{K((y(1),...,y(n))}{n} = \lim_{n \to \infty} \frac{H((y(1),...,y(n))}{n}$, where $K(y(1),...,y(n))$ of the complexity of the *sequence* $y(1),...,y(n)$ and $H(y(1),...,y(n))$ is the entropy of the *probability distribution* of $y(1),...,y(n)$. The other connection between complexity and mutual information, marked as a dotted line in Figure 1, is more specific to intrusion detection and expanded upon in the next section.

The specific path taken in this paper is the extreme left of the diagram of Figure 1, except that we stop short of detecting a change in model order, but rather endeavor to detect a change in mutual information.

A. Mutual Information versus Kolmogorov Complexity

Since the MI and Kolmogorov complexity both endeavor to find model order, the two approaches ought to be somehow related. To understand the similarities/discrepancies, some more formal concepts are already in order.

The mutual information between the past y_- and the future y_+ is the amount by which the Shannon entropy of the future decreases when we are given the past, that is, $H(y_+) - H(y_+|y_-)$. Practically, the past/future MI is related to the (properly weighted) mean square error between the data and the optimal predictor model. In the Gaussian case, the modeling is traditionally done by the classical innovations representation [Aka75], while, in the non-Gaussian case, the modeling could be done by such well known statistical modeling techniques as the Alternating Conditional Expectation (ACE) [BF85].

The Kolmogorov complexity $K(y)$ of a *string* y is the length of the shortest composite string $< T : u >$ such that if the string u is the input tape to the Turing machine T it produces y on the output tape and then stops [Sip97, Def. 6.20], [WD99], [ZL70a].

Information-based and complexity-based intrusion detections can be related by the sometimes loosely stated fact that high complexity means low information. Precisely, Kolmogorov proved that the most complex binary sequences are those that approach random coin tosses [Man77, p. 227], [Sip97, p. 218], which have vanishing mutual information. To generalize the latter to arbitrary shift dynamics $T : \Omega \to \Omega$ with invariant measure μ, it is convenient to use Markov partitioning $\Omega = \cup_i A_i$, so as to reduce the problem to symbolic dynamics. However, even after this conversion, the connection between complexity and mutual information does not appear to hold without the crucial ϕ-mixing condition, that is, $\left| \frac{\mu(A_i \cap T^{-k-n}A_j)}{\mu(A_i)\mu(A_j)} - 1 \right| \leq$ $\phi(k)$ for some decaying function $\phi(k)$, and uniformly for all n. For example, consider the automorphism of the torus $y(k + 1) = Fy(k) \mod 1 =: Ty(k)$, where $F \in \mathbb{Z}^{2 \times 2}$ and $\det F = \pm 1$, in which case the entropy is relative to the Lebesgue measure μ [NS89]. This entropy is well known to be $h(y_+) = \log(|\lambda(F)|_{\max})$ and the decay rate of the correlation is given by $\phi(k) \sim |\lambda(F)|_{\max}^{-k}$ [BSTV97]. By the Zvonkin-Levin theorem, the Kolmogorov complexity rate is given by $\log(|\lambda(F)|_{\max})$. As the complexity increases, the correlation decreases faster, hence so does $\sum_{i,j} \mu(A_i \cap T^{-k}A_j) \log \frac{\mu(A_i \cap T^{-k}A_j)}{\mu(A_i)\mu(T^{-k}A_j)}$ as $k \to \infty$, and from there on it can be shown [?] that the past/future mutual information $\sum_{i_k,j_l} \mu\left((\cap_{l \geq 0} T^l A_{j_l}) \cap (\cap_{k \geq 0} T^{-k} A_{i_k})\right) \log \frac{\mu((\cap_{l \geq 0} T^l A_{j_l}) \cap (\cap_{k \geq 0} T^{-k} A_{i_k}))}{\mu(\cap_{l \geq 0} T^l A_{j_l})\mu(\cap_{k \geq 0} T^{-k} A_{i_k})}$ decreases.

B. Fundamental Concepts

A key assumption of the techniques investigated here is that some network attacks change the structure of the traffic. In an effort to understand self-similarity, several aspects of the structure of network traffic have been extensively investigated. It has been widely reported that various aspects of the network and traffic impact the structure. For example, the autocorrelation, more specifically, the rate of decay of the autocorrelation, has been widely used to study traffic [Sta98]. This rate of decay is related to the Hurst parameter and is known to be related to the application layer parameters such as file size distribution [CB96]. In [FGW98], a wavelet-based analysis of traffic revealed a cascade structure that is dependent on transport and application protocols as well as user behavior such as mouse clicks and session duration. While much of this previous work focuses on long time scales, in [LB03], the short-time scale behavior of the "packet pattern" was studied and it was found that this pattern depends on certain network parameters such as loss rate. Here, the mutual information is used, but instead of examining the

variation over different time scales to understand self-similarity or scaling, the temporal variation is used to understand the type of traffic, specifically, to determine whether an attack is occurring.

The premise of the information theoretic approach to intrusion detection is that any kind of intrusion would disturb the dynamical structure, and hence the information structure, which the signal inherits from the interaction of TCP with the malicious flow. For example, in case of constant bit-rate (CBR) UDP flooding, packet arrival rates may become more stable than those that occur under typical TCP file transfers. In this case, the signal becomes more deterministic, hence more predictable; that is, CBR flood results in the past packet arrival rate holding more information about the future packet arrival rate. Next to CBR flooding, there are other attacks that would rather decrease the information, making the signal less predictable. It appears therefore that the traffic has to be monitored for a change in information, which should trigger the alarm. On the other hand, while flooding-based attacks may impact the mutual information, traffic anomalies that do not impact the dynamic structure would not cause a change in the mutual information. Other techniques are required to detect such attacks.

From a broader perspective, since as shown in the preceding section, the connection between rate of decay of correlation and mutual information does not appear to hold without a stronger version of mixing, it is believed that mutual information adds, next to rate of decay of correlation, a new dimension to traffic analysis.

C. Practical Implementation

Numerically, the mutual information between the past and the future of the traffic signal, or any process for that matter, is computed via Canonical Correlation Analysis (CCA) between the past and the future of the process [Aka75],[JH85]. In case of a Gaussian process, the linear CCA is adequate in the sense that the mutual information can easily be computed from the linear Canonical Correlation Coefficients (CCC's). If the traffic signal is non-Gaussian, the linear CCC's underestimate the mutual information. However, after a nonlinear pre-processing, the resulting nonlinear CCC's would yield an estimate that approaches the mutual information as closely as possible, depending on the amount of nonlinear processing that is consistent with online intrusion detection.

Several signals (e.g., link utilization, packet arrival, and queue length) are candidates for mutual information analysis by canonical correlation. However, our experiments have shown that the change in mutual information concurrent with an attack is more sizable if the average utilization over a sample period is analyzed. Since the number of arrivals during a sample period and the average utilization during a sample period differ only by a scaling factor, the mutual information of the utilization is the same as the mutual information of the number of packet arrivals.

In Section IV, three topologies are analyzed here: *parking lot topology, random 50-node topology, and 100-node transit-stub topology*. We do not consider a widely used single-bottleneck dumbbell topology in this paper, as it was shown in [SBJ03] that intrusion detection on the *dumbbell topology* is straightforward. The random 50-node and the 100-node transit-stub topologies are generated by Georgia Tech's topology generator (Gt-Itm). We use the network simulator (ns) [htta] to integrate these topologies and to generate traffic. For each topology, our study is 2-fold: linear versus nonlinear canonical correlation analysis, for varying sampling periods (ranging from 0.1 to 20 sec.). Furthermore, in Section VI, this mutual information-based detection scheme is applied to backbone network traces.

While the simulation and experiment results are promising in that they indicate that the traffic anomalies result in a significant change in the mutual information, the results should not be taken as definitive proof of the deployability of mutual information-based detection mechanisms. Rather, the intent of this paper is to illustrate the potential utility of signal processing techniques such as mutual information for the detection of network traffic anomalies. A comprehensive examination of the performance in terms of false positives and false negatives over the very wide range of types of traffic found in the Internet is currently under investigation.

D. Outline

An outline of the paper follows. Section II gives a brief overview of the related work in this area. Section III deals with the linear and nonlinear canonical correlation analyses, the mutual information, and the resulting models. Section IV presents the simulation setup. In Section V the simulation results are analyzed.

II. RELATED WORK

Today, there are generally two types of intrusion detection systems (IDS): misuse detection and anomaly detection. Misuse detection techniques attempt to model attacks on a system as specific patterns, then systematically scan the system for occurrences of these patterns [Roe99], [Pax99]. Anomaly detection approaches attempt to detect intrusions by noting significant departures from normal behavior [SBJ03], [JV91], [Den87], [GWC98], [KRL97][LB98]. Our approach falls under network-based anomaly detection as we detect intrusion by analysis of traffic signals.

Many techniques have been proposed for anomaly detection. Several of them analyze different data streams, such as data mining for network traffic [LS99], statistical analysis for audit record [JV91], sequence analysis for operating system calls [For96], information retrieval [AK98] and inductive learning [TCL99]. Statistical methods have also been developed for network anomaly detection [LTG+92], [SJJ02]. Change point detection technique has been used for detection of various flooding attacks [BKRT01], [WZS02].

Signal processing techniques, the focus of our work, have been used previously to analyze malicious network traffic and to detect ongoing attacks. In [AAB01], the authors have used wavelet coefficients across resolution levels to locate smooth and abrupt changes in variance and frequency in the given time series. [TJ03] has proposed a statistical signal processing technique based on abrupt change detection. [BKPA02] has used flow-level information to identify frequency characteristics of anomalous network traffic. [HHP03] and [CKT02] have developed spectral analysis based approach to detect DoS attack. Further, wavelets and other signal processing techniques have been extensively used to analyze both wired and wireless network traffic [ZRMD03], [PCJ+02]. Perhaps the most relevant approach along the lines of our work is Kolmogorov complexity approach to intrusion detection described in [EBH01]. The fundamental difference between our work and this work is highlighted in the introduction.

III. CANONICAL CORRELATION ANALYSIS

Here $\{y(k) \in [-b, +b] : k = ..., -1, 0, +1, ...\}$ is the centered link utilization signal (i.e., the total number of bytes that arrived during the sample period divided by the maximum possible number of bytes that could arrive during the sample period). y_k

is bounded by the bandwidth and is viewed as weakly stationary process with finite covariance $E(y(i)y(j)) = \Lambda_{i-j}$ defined over the probability space $(\Omega, \mathcal{A}, \mu)$. As such, there is no need to take infinite variance processes (for example, α-stable, H-self-similar processes [ST94]) into consideration. The past and the future of the process are defined, respectively, as

$$
\begin{aligned}
y_-[L] &= (y(k), y(k-1), ..., y(k-L+1))^T, \\
y_+[L] &= (y(k+1), ..., y(k+L))^T
\end{aligned}
$$

where L is the "lag." We will drop the notation $[L]$ whenever the size of the past or the future is irrelevant. The mutual information between the past and the future [JW92a],[Wu92],[Kul68],[Aka75] is the amount of information we acquire about the future when we are given the past. Since, technically, the entropy of a continuous-valued process does not exist, the mutual information is most easily defined in terms of past-measurable partitions A and future-measurable partitions B of the sample space Ω,

$$
\begin{aligned}
I(y_-, y_+) &= \sup_{A,B} (H(A) - H(A|B)) \\
&= \sup_{A,B} \sum_i \sum_j \log \frac{\mu(A_i \cap B_j)}{\mu(A_i)\mu(B_j)} \mu(A_i \cap B_j) \\
&= \int \int \log \frac{p(y_-, y_+)}{p(y_-)p(y_+)} p(y_-, y_+) \, dy_- \, dy_+
\end{aligned}
$$

In the above, $H(A)$ is the entropy of the partitioning A and $H(A|B)$ is the conditional entropy of the partitioning A given the partitioning B. The last equality in the above is valid only under absolute continuity conditions, in which case $p(y_-, y_+)$ is the Radon-Nikodym derivative $\frac{\mu(dy_-, dy_+)}{dy_- dy_+}$ and $p(y_-), p(y_+)$ are the marginal densities. As such, $I(y_-, y_+)$ is the Kullback-Leibler "distance" between $p(y_-, y_+)$ and $p(y_-)p(y_+)$. In this setup, it could be argued that, because $y(k)$ is a packet count under bandwidth limitation, it takes only finitely many values, so that the mutual information can still be defined as $I(y_-, y_+) = H(y_+) - H(y_+|y_-)$, where $H(y_+)$ is the entropy of the future and $H(y_+|y_-)$ is the conditional entropy of the future given the past.

A. Linear Canonical Correlation

The linear Canonical Correlation Analysis (CCA) is a second moment technique for computing the mutual information under the standard Gaussian assumption. Since the process $y(k)$ is bounded, the Gauss property is only an approximation of the true distribution.

Factor the covariances of the past and the future as

$$
E(y_-(k)y_-^T(k)) = L_- L_-^T, \quad E(y_+(k)y_+^T(k)) = L_+ L_+^T
$$

and then construct the canonical correlation matrix Γ along with its Singular Value Decomposition (SVD),

$$
\Gamma(y_-, y_+) := L_-^{-1} E\left(y_-(k)y_+^T(k)\right) L_+^{-T} = U^T \Sigma V
$$

where U, V are orthogonal matrices and

$$\Sigma = \begin{pmatrix} \sigma_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_L \end{pmatrix}, \quad 1 \geqslant \sigma_1 \geqslant \ldots \geqslant \sigma_L \geqslant 0$$

The σ's are called Canonical Correlation Coefficients (CCC's). Since they are all bounded by 1, it follows that, even as $L \to \infty$, the canonical correlation operator is bounded as $\|\Gamma\| \leq 1$, where $\|\cdot\|$ denotes the spectral norm. If the process is Gaussian, it is well known that

$$-\frac{1}{2} \log \det \left(I - \Gamma^T \left(y_-, y_+ \right) \Gamma \left(y_-, y_+ \right) \right) = I(y_-, y_+) \tag{1}$$

The fact that Γ is a bounded operator does not imply that $I(y_-, y_+)$ is bounded as $L \to \infty$, because $\sigma_i \leq 1$ does not imply that $-\frac{1}{2} \log \Pi_{i=1}^{\infty} \left(1 - \sigma_i^2 \right)$ exists. We will come back to this point at the end of the next subsection.

In general, with a noisy, finite length L data record, the sequence of CCC's still shows a fairly clear cutoff. Practically, in all cases, a break point $\sigma_D << \sigma_{D+1}$ is identified and a reduced model of order D is obtained after resetting the $L - D$ tail coefficients to 0. The latter is formalized in stochastic balancing and Hankel norm reduction [?], [?], [?].

A FEW NUMERICAL REMARKS: It is customary to define $L_\pm$ to be lower triangular (Cholesky factorization), although $L_\pm$ could be defined upper triangular ("anti-Cholesky" factorization), in which case Γ is near-Hankel and in fact, for $L = \infty$, it will be the Hankel operator associated with the phase of the spectral factor of y [JH85], [?], [?]. The particular way the factorization is done does not affect the CCC's. $E(y_\pm(k)y_\pm^T(k))$ might be marginally positive definite, resulting in problems with the Cholesky factorization; there is thus a need to monitor the condition number of $E(y_\pm(k)y_\pm^T(k))$. If the covariance matrix is poorly conditioned, a generalized eigenvalue approach to compute the σ's should be used.

B. Nonlinear Canonical Correlation

If the process y is not Gaussian, Eq. (1) is no longer valid. This motivates the nonlinear canonical correlation [Lar91],[JW92a],[Wu92] as a modified technique to reach the mutual information in the non-Gaussian setup. Precisely,

Theorem 1: Let $\{y(k) \in [-b, +b] : k = ..., -1, 0, +1, ...\}$ be a bounded valued weakly stationary process defined over the probability space $(\Omega, \mathcal{A}, \mu)$. Let $I(y_-, y_+)$ be the mutual information between the past and the future and let $\Gamma(\cdot, \cdot)$ denote the canonical correlation. Then

$$\sup_{f,g} \left(-\frac{1}{2} \log \det \left(I - \Gamma^T \left(f\left(y_- \right), g\left(y_+ \right) \right) \Gamma \left(f\left(y_- \right), g\left(y_+ \right) \right) \right) \right) \leq I(y_-, y_+)$$

where $f, g : [-b, +b]^L \longrightarrow \mathbb{R}^L$ are functions such that $f \circ y_-, g \circ y_+ \in L^2(\Omega, \mathcal{A}, \mu)$, $E(f(y_-)) = E(g(y_+)) = 0$, and for convenience normalized as $E\left(f^T(y_-)f(y_-) \right) = 1$, $E\left(g^T(y_+)g(y_+) \right) = 1$. Furthermore, equality is achieved iff $f\left(y_- \right), g\left(y_+ \right)$ can be made jointly Gaussian, in which case the joint past/future process is called *diagonally equivalent to Gaussian*.

Proof: See [JW92a],[Wu92] ∎

To motivate the left-hand side optimization in a practical estimation setup, consider a linear regression of $g(y_+)$ on $f(y_-)$. It is easily found that

$$\min_A E(g(y_+) - Af(y_-))^T (L_+ L_+^T)^{-1} (g(y_+) - Af(y_-))$$
$$= L - Trace\left(\Gamma^T\left(f(y_-), g(y_+)\right)\Gamma\left(f(y_-), g(y_+)\right)\right)$$

Clearly, the best choice of f, g is the one that maximizes $Trace\left(\Gamma^T\left(f(y_-), g(y_+)\right)\Gamma\left(f(y_-), g(y_+)\right)\right)$ and it is readily seen that this is achieved for the same distortion functions f, g. This latter technique calls for the maximization of the trace of $\Gamma^T\left(f(y_-), g(y_+)\right)\Gamma\left(f(y_-), g(y_+)\right)$, as was done in the approach of Larimore and Baillieul [Lar91], rather than the maximization of the mutual information, as done by Jonckheere and Wu[JW92a],[Wu92]. Not surprisingly, by how much $Trace\left(\Gamma^T\left(f(y_-), g(y_+)\right)\Gamma\left(f(y_-), g(y_+)\right)\right)$ can be increased by means of nonlinear distortion should be bounded by the mutual information; in fact, the following is true:

Theorem 2: Under the same assumnptions as Theorem 1,

$$\max_{f,g} Trace\left(\Gamma^T\left(f(y_-), g(y_+)\right)\Gamma\left(f(y_-), g(y_+)\right)\right) \leq 2I(y_-, y_+)$$

and furthermore equality holds if and only if the processes y_-, y_+ are independent.

Proof: See [JW92a],[Wu92] ∎

Using the above, it follows that

$$\text{MSE} := \lim_{L\to\infty} \frac{1}{L}\left(L - \sup_{f,g}\text{Trace}(\Gamma^T(f(y_-), g(y_+))\Gamma(f(y_-), g(y_+)))\right)$$
$$\geq 1 - 2\lim_{L\to\infty} \frac{I(y_-, y_+)}{L}$$

We define $\iota(y_-, y_+) := \lim_{L\to\infty}\frac{I(y_-, y_+)}{L}$ to be the mutual information *rate*. In case $\iota < \frac{1}{2}$, the mutual information rate is too weak and will result in a nonvanishing MSE. It can be shown that, if the system is ϕ-mixing, the mutual information rate vanishes [**?**], so that $MSE \geq 1$.

Invoking the finite variance property, we construct Hilbert space bases for the subspaces of $L^2(\Omega, \mathcal{A}, \mu)$ of past, future measurable functions $\Omega \longrightarrow \mathbb{R}$. The distortion functions f, g will be expressed as linear combinations of those basis functions, leading to yet another computational implementation of the nonlinear CCA in addition to the sequential selection of Larimore and Ballieul [Lar91] and the integral equation approach of [JW92a], [Wu92]. In case of finite lag L, since $y(k)$ is defined over a compact set $[-b, +b]$, by a well known theorem any function of y_-, y_+ can be *uniformly* approximated by polynomials; hence we choose polynomials $p_j(y_-), q_j(y_+); j = 1, 2, ...$ such that $E_- p_j = E_+ q_j = 0$, and forming bases of the Lebesgue spaces of zero-mean past-measurable, future measurable functions, respectively. Since

$$f_i(y_-) = l.i.m._{N\to\infty}\sum_{j=1}^{N}\phi_{ij}p_j(y_-), \quad g_i(y_+) = l.i.m._{N\to\infty}\sum_{j=1}^{N}\gamma_{ij}q_j(y_-)$$

for least squares fitting coefficients ϕ_{ij}, γ_{ij}, the nonlinear CCA therefore reduces to

$$\sup_{\phi,\gamma} \left(-\frac{1}{2} \log \det \left(I - \Gamma \left(\phi p \left(y_- \right), \gamma q \left(y_+ \right) \right) \Gamma^T \left(\phi p \left(y_- \right), \gamma q \left(y_+ \right) \right) \right) \right)$$

where ϕ, γ are the arrays made up with the coefficients ϕ_{ij}, γ_{ij}. The solution ϕ, γ is far from unique even under the normalization condition on f, g, because there is still the freedom to pre-multiply ϕ, γ by orthogonal transformations. If $L < \infty$, the above supremum is nontrivial and is easily accomplished via linear CCA of $p \left(y_- \right), q \left(y_+ \right)$, that is, via SVD of $\Gamma \left(p \left(y_- \right), q \left(y_+ \right) \right)$. Specifically, do the factorizations

$$E \left(p \left(y_- \right) p \left(y_- \right)^T \right) = L_- L_-^T, \quad E \left(q(y_+) q(y_+)^T \right) = L_+ L_+^T$$

along with the SVD

$$\Gamma \left(p \left(y_- \right), q \left(y_+ \right) \right) = \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}^T \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}, \quad I \geq \Sigma_1 \gg \Sigma_2 \geq 0$$

The coefficients of the optimal distortion functions are given by

$$\phi = U_1 L_-^{-1}, \quad \gamma = V_1 L_+^{-1}$$

Even when $L < \infty$, the Hilbert space basis will still be infinite-dimensional, so that the arrays ϕ, γ will be "fat." In this case, we have

$$\sup_{\phi,\gamma} \left(-\frac{1}{2} \log \det \left(I - \Gamma \left(\phi p \left(y_- \right), \gamma q \left(y_+ \right) \right) \Gamma^T \left(\phi p \left(y_- \right), \gamma q \left(y_+ \right) \right) \right) \right) \leq -\frac{1}{2} \log \det \left(I - \Gamma \left(p \left(y_- \right), q \left(y_+ \right) \right) \Gamma^T \left(p \left(y_- \right), q \left(y_+ \right) \right) \right)$$

In other words, the CCA of the Hilbert space basis (the right-hand side) provides a bound on what the nonlinear CCA can achieve (the left-hand side).

A feature that is already present in the linear CCA of traffic signals, but that becomes much more pronounced in the nonlinear CCA, is that the head of the CCC sequence, $\sigma_1, \sigma_2, \sigma_3, ..., \sigma_D$, remains close to one before dropping abruptly near zero. This phenomenon is, to our knowledge, unique to traffic signals and points to some deterministic features in the dynamics.

NUMERICAL REMARK: Practically, p, q are chosen as simple monomials or Chebyshev polynomials in the components of the past, future. It is important to scale the large powers appearing in $p(y_-), q(y_+)$, for otherwise the high power terms become dominant over the low power terms.

IV. SIMULATION SETUP

We used the Network Simulator (ns) developed by LBNL to set up our simulation environment [httb]. Ns is a discrete event simulator widely accepted for networking research. It provides a substantial support for simulation of TCP, routing, and multicast protocols over wired and wireless (local and satellite) networks. Moreover, ns generates Constant Bit Rate (CBR), TELNET, FTP, HTTP, etc. traffic. The simulator also has a small collection of mathematical functions that can be used to implement exponential, uniform, Pareto, etc. random variables. We used this capability to setup the network environment that

TABLE I

CBR TRAFFIC PARAMETERS FOR PARKING LOT TOPOLOGY

Trial	CBR Traffic	
	packet size	interval (sec)
1	250	0.06
2	300	0.07
3	350	0.08
4	400	0.09
5	450	0.1

synthesized HTTP and CBR traffic.

A dynamical model for normal TCP traffic was synthesized from the signals obtained by sending HTTP traffic from the sources to the destinations at random times. For HTTP traffic, the file size distribution was modeled as a general ON/OFF behavior with a combination of heavy-tailed and light tailed sojourn times, while the inter-page time and the inter-object per page time distributions were set to be exponential. The page size was set to be constant and the object per page size to be Pareto to replicate today's network bursty traffic [PE95], [LTWW94]. In summary, HTTP traffic can be parametrized by the following parameters in ns : number of sessions, inter-session time, session size, inter page time, page size, inter object time, average object size, and shape parameter of object size (exponent (α) in Pareto distribution).

In addition to this background (HTTP) traffic, a large number of small size CBR packets were sent over some UDP connections from model the attack scenario [CER]. CBR traffic can be parameterized by packet size and interval.

We ran several trials to cover a wide range of parameters for each topological setting. Each run was executed for 30000 simulated seconds, logging the traffic at the 0.01 second granularity.

V. RESULTS AND INTERPRETATION

In this section, we show how the mutual information changes under CBR attack. Three topologies are considered; parking lot topology, 50-node random topology, and 100-node transit-stub topology. For parking lot topology, we carried out two experiments. The first experiment gives an idea of how the mutual information is affected under the attack, while the second experiment shows how the attack can be detected at a link different than the attacked link. In a more complicated setting, we consider 50-node random topology. Moreover, to see if the mutual information is a useful tool in detection of infrastructure attacks, such as flooding a bottleneck link, we use 100-node transit-stub topology.

A. Parking Lot Topology

Figure 2 shows the "Parking Lot" topology. The nodes S_i (i = 8,10,12) are sources and the nodes D_i (i = 9,11,13) are destinations. The sources send traffic to their downstream destinations. In addition to this background (HTTP) traffic, a large number of CBR packets are sent over several UDP connections from source nodes to the victim node to model the attack scenario [CER]. Specifically, source nodes 8 and 10 each send 15 CBR flows to the victim node 4. The intensity of CBR and HTTP traffic is varied in each trial. Here, we show the results for 5 trials. The parameters of CBR and HTTP traffic for each trial are shown in Tables I and II. Here the link speed is 10 Mbps and the latency of the each link is 20 ms.
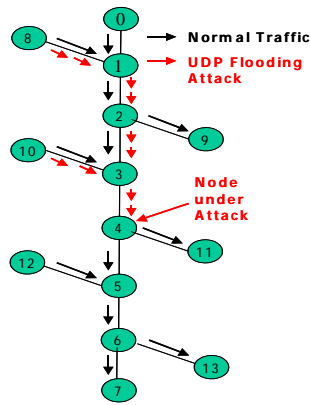
Fig. 2. Parking lot topology. Normal traffic is a HTTP traffic, while UDP packet storm attack is simulated by sending CBR traffic downstream from the sources 8 and 10 to the vicitm 4.

TABLE II

HTTP TRAFFIC PARAMETERS FOR PARKING LOT TOPOLOGY

Trial	HTTP traffic							
	number of sessions	inter-session time (sec)	session size	inter-page time (sec)	page size	inter object time (sec)	average object size	object size shape parameter
1	2500	2.5	1000	75	5	0.05	60	1.1
2	3000	3	1200	90	6	0.06	72	1.2
3	3500	3.5	1400	105	7	0.07	84	1.3
4	4000	4	1600	120	8	0.08	96	1.4
5	4500	4.5	1800	135	9	0.09	108	1.5
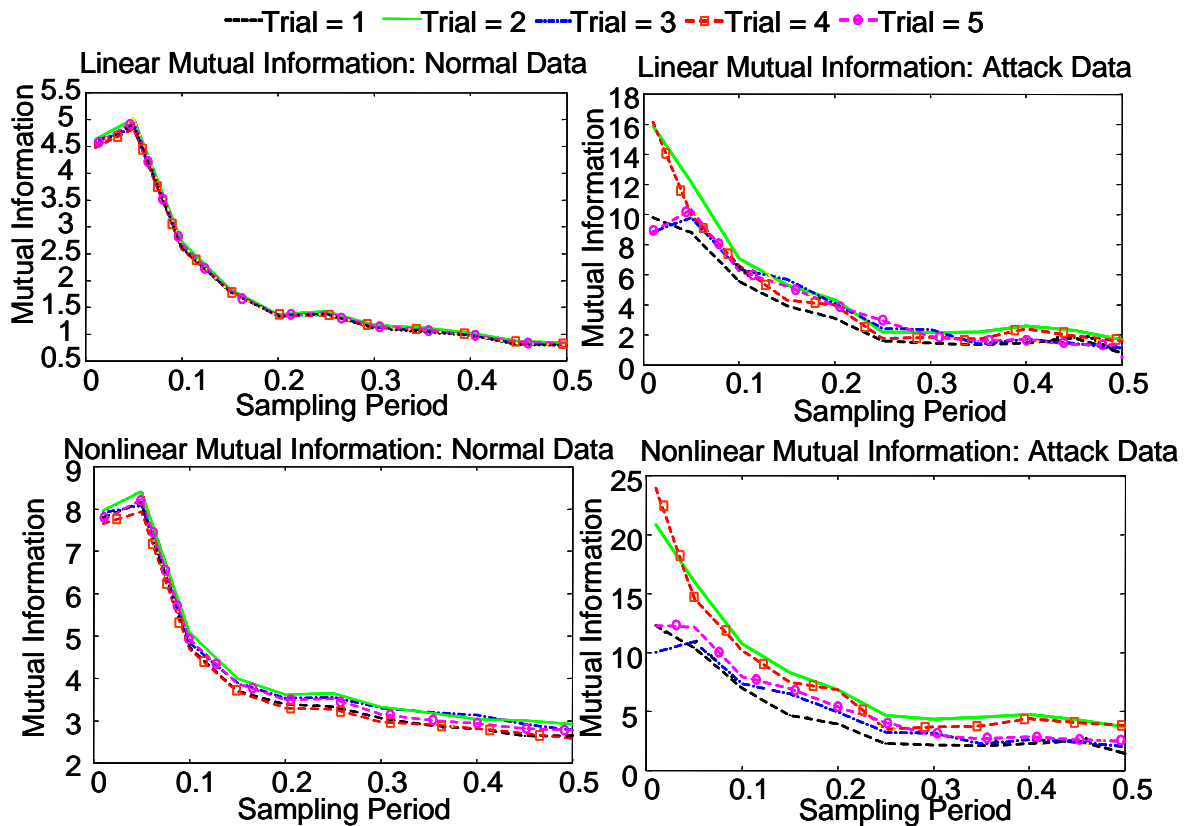
Fig. 3. Mutual information vs sample period for Parking lot topology. The upper frames show the linear mutual information while the lower frames show nonlinear mutual information. The left-hand side plots are for normal traffic while the right-hand side plots are for attack traffic.

1) Experiment 1: HTTP Traffic under CBR Attack; Monitored Link same as Flooded link; Linear versus Nonlinear Analysis:
In this experiment, the impact of intensity of traffic on the ability to detect an attack is explored. Here, the background traffic is
HTTP and the attack traffic is CBR. Intensity of HTTP traffic can be varied by changing such parameters as number of sessions,
number of pages, number of objects, etc. in `ns`(Table II). The intensity of CBR traffic is also varied (Table I). The link under
attack is 3-4 and the monitored link for the detection is also 3-4 in Figure 2. The upper frames of Figure 3 show the linear
mutual information for different sample intervals for normal and attack traffic. Note that the mutual information is derived from
the average link utilization over the sample period (i.e., the number of byte that arrived during the sample period divided by
the maximum possible number of bytes that could arrive during the sample period). Note that the mutual information for the
normal traffic remains the same for different trials. The justification of the latter is that the mutual information is unchanged
under scaling; it only depends on the dynamics, which in this case remains that of HTTP traffic. From trial 1 to trial 5,
the intensity of HTTP traffic increases while the intensity of CBR traffic decreases. As the relative intensity of CBR traffic
increases, the traffic becomes more predictable. This can be seen as the increase in the mutual information in the attack traffic.
Observe that for the trial 1, the increase in the mutual information under attack is small; the justification is the small amount
of CBR traffic. Another experiment was performed in which the intensity of CBR traffic was kept constant. This experiment
also showed a clear increase in mutual information under significant amount of CBR traffic.

The lower frames of Figure 3 shows the nonlinear mutual information for normal and attack traffic. Observe that for normal
traffic the nonlinear mutual information is higher than the linear mutual information. Since TCP has complicated dynamics,
higher correlation and hence higher mutual information is achieved by nonlinear distortion of the past and the future. This
also holds true for the attack traffic. However, for this set-up, the relative increase in linear and non-linear mutual information
remains almost the same.

2) Experiment 2: Monitored Link Downstream of the Flooded Link: In this experiment, the flooded link is still 3-4, but the
link utilization is monitored along link 4-5. The simulation set-up is the same as Experiment 1. The linear mutual information
is computed for the link utilization 4-5. Figure 4 shows significant increase in the linear mutual information for the attack
traffic as compared to the normal traffic. In conclusion, the mutual information can pick up the difference in the statistical
structure of the signal, even when the signal is not recorded on the flooded link. This differs from count-based schemes that
typically focus on observing the attack directly.

B. Random 50 Node Topology

In the more complicated "50-node" random topology (Fig. 5) generated by Georgia Tech's topology generator (Gt-Itm), 20
nodes are set as the sources and 20 nodes are set as the destinations. The maximum link speed is 1.5 Mbps while the minimum
link speed is 10 Mbps. The propagation delay varies between 20 to 120 ms. HTTP requests are sent at random times from
random clients to random servers. All the sources send 5 CBR flows to the target node 14 during the attack. The CBR and
HTTP traffic parameters for various trials for this set-up are listed in Tables III and IV. Each trial was executed for 30000
simulated seconds, logging the traffic at 0.01 second granularity. The monitored link is 14-30.

Figure 6 shows the linear and nonlinear mutual information for the monitored link. The results are consistent with the
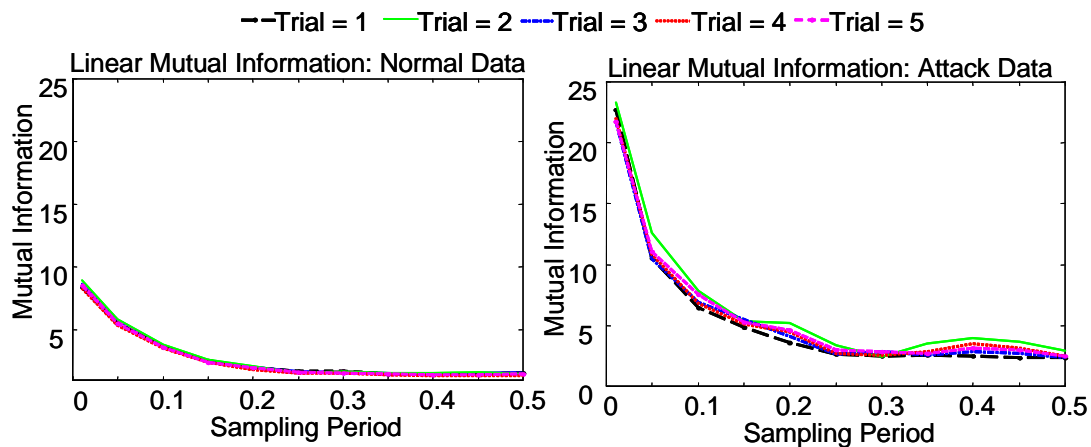
Fig. 4. Linear mutual information vs sample period for Parking lot topology. The flooded link is 3-4 while the monitored link is 4-5. Observe the difference between the mutual information.
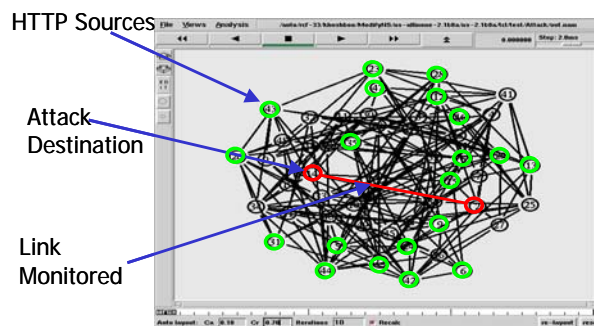

Fig. 5. 50-node random topology. The target node 14 and the monitored link is 14-30

results obtained for the parking lot topology, meaning that the mutual information increases in case of an attack. Furthermore, the increase in the mutual information under attack is much more sizable for this topology as compared with the elementary baseline topology.

C. 100 Node Transit Stub

CERT has noted that DoS attacks on links and routers are increasing [Cen]. A coordinated attack can be planted by many end hosts that all send packets that will eventually traverse the same link thereby hogging all link bandwidth. In the present experiment, we explore the possibility of detecting such an attack. A 100-node transit-stub topology is generated by Georgia Tech's topology generator (Gt-Itm). As shown in Figure 7, there is only one HTTP server and 20 HTTP clients. There are 13

TABLE III

CBR TRAFFIC PARAMETERS FOR RANDOM 50-NODE AND 100 NODE TRANSIT-STUB TOPOLOGIES

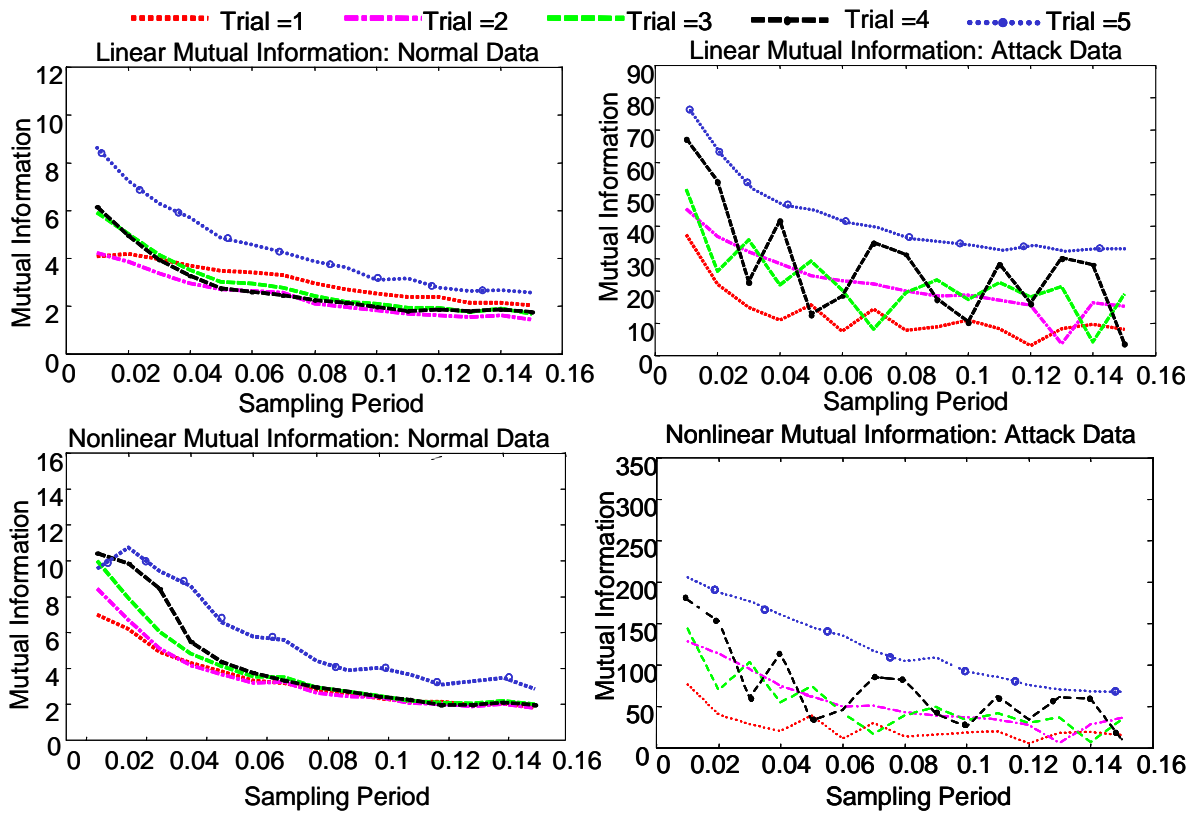| Trial	CBR Traffic	
	packet size	interval (sec)
1	25	0.11
2	50	0.12
3	75	0.13
4	100	0.14
5	125	0.15

Fig. 6. 50-node random topology. The upper frames show the linear mutual information while the lower frames show nonlinear mutual information. The left-hand side plots are for normal traffic while the right-hand side plots are for attack traffic.

TABLE IV

HTTP TRAFFIC PARAMETERS FOR RANDOM 50-NODE AND 100 NODE TRANSIT-STUB TOPOLOGIES

Trial	HTTP traffic							
	number of sessions	inter-session time (sec)	session size	inter-page time (sec)	page size	inter object time (sec)	average object size	object size shape parameter
1	400	1	200	15	1	0.01	12	1.1
2	800	2	400	30	2	0.02	24	1.2
3	1200	3	600	45	3	0.03	36	1.3
4	1600	4	800	60	4	0.04	48	1.4
5	2000	5	1000	75	5	0.05	60	1.5

attack sources and 13 attack destinations. Each attack source sends 20 CBR flows to every attack destination. All the attack sources use bottleneck link 2-0 to send traffic. The focus here is the HTTP client that uses the link 0-2 to send HTTP requests and the link 2-0 to receive the HTTP server response. We ran 5 different trials by varying CBR and HTTP traffic parameters (see Tables III and IV). Each trial was executed for 30000 simulated seconds, logging the traffic at 0.01 second granularity. The monitored link is 2-0.

Figure 8 shows the time-series of link utilization of various links. The left-most frame in Figure 8 shows the link utilization for the upstream server link, the center frame shows the link utilization for the bottleneck link and the right-most frame shows the link utilization for the upstream client link. It can be seen that, during the attack, the client of interest has zero link utilization, meaning the client completely stops getting HTTP data packets since almost all the bandwidth of the link 2-0 is used by the attack traffic. On the other hand, there is no visible difference in the link utilization of upstream server link nor
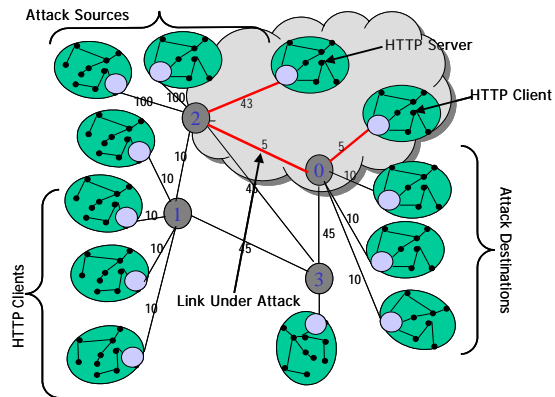
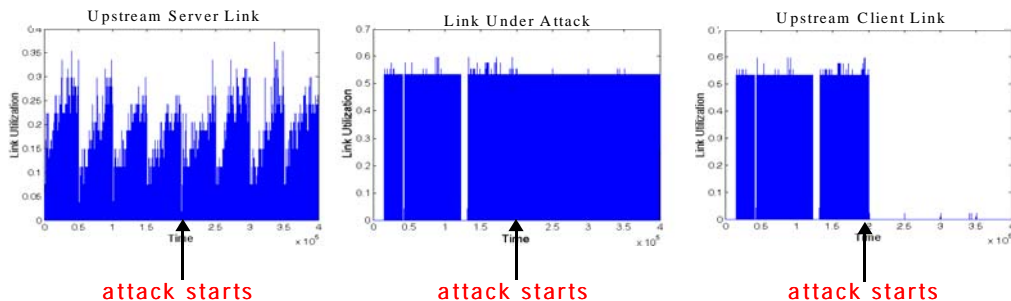Fig. 7.   100-node transit-stub topology. The link under attack is 0-2.

Fig. 8.   Link utilization vs time.

in the link utilization of the bottleneck link after the attack.

To detect this attack, we use the nonlinear mutual information computed for the link utilization observed on the bottleneck link 2-0. Figure 9 shows the mutual information plots for this experiment for different trials. It can be seen that there is a significant change in the mutual information, even though the attack cannot be seen by visual inspection of the link utilization plots. It is important to note that since the link utilization remains constant during the attack, count-based methods that simply consider the amplitude of the link utilization during a sample period are unable to detect the attack.

VI. EXPERIMENTAL STUDY

To further investigate mutual information-based detection schemes, traces from a backbone link were used. Specifically, we examine packet traces captured on SONET OC-48 links by CAIDA monitors. The link runs from San Jose, CA to Seattle, WA and belongs to US tier 1 backbone Internet Service Provider (ISP). The traces were collected by Linux-based monitor with Dag 4.11 network cards and packet capture software originally developed at the University of Waikato and currently produced by Endace. The data was collected over a 1 hour period on August 8, 2002. During this time, the average link utilization was 14.7%. The packet trace captured a UDP flooding attack. The detection of this attack is used as a test case for examining the performance of mutual information-based detection.

The mutual information of the time series of the average link utilization over a 62 msec sample intervals was computed. Based on the first 1000 samples, the nominal mutual information was determined. We denote this nominal value as $\bar{I}$, whereas the mutual information found after processing a new observation is denoted I_k. We take the lag to be 30 and compute the
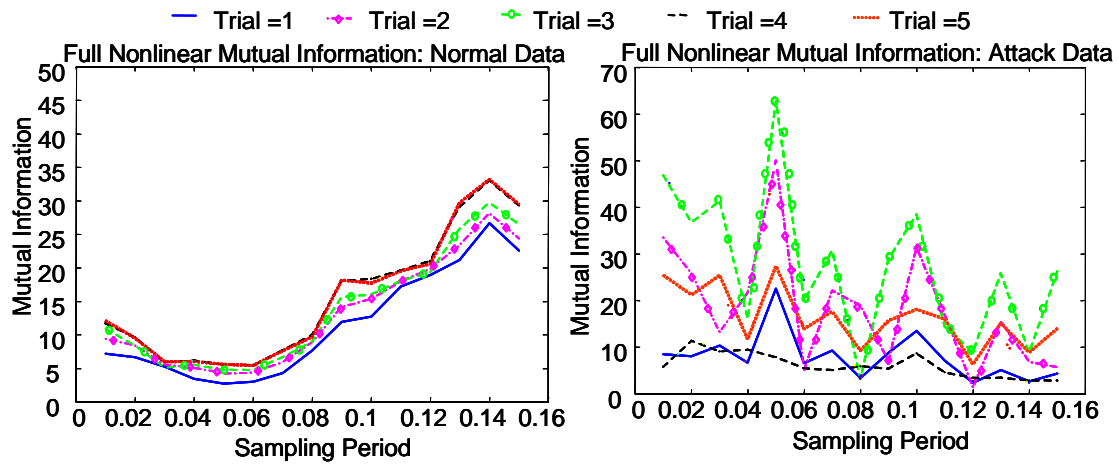
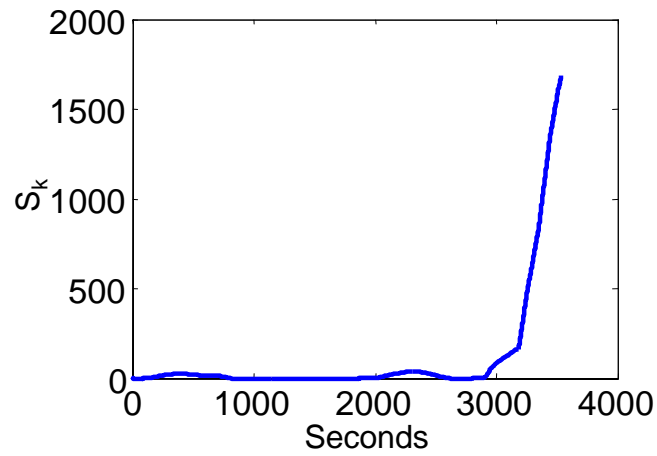Fig. 9.   100-node transit-stub topology. The plot shows the nonlinear mutual information.

Fig. 10.   Time series of S_k, the cumsum of the mutual information. This data is from a back-bone link. The steep increase at around the 10000th sample is due to a UDP flooding attack.
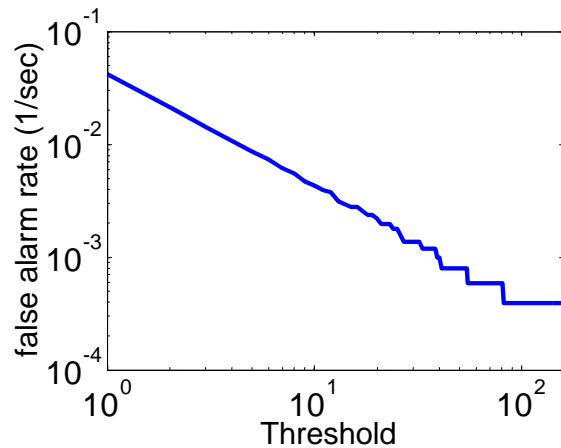

Fig. 11.   False alram rate vs. Threshold. No false alarms occured for Threshold above 160. However, the attack was detected if Threshold is below 1600.

mutual information based on a window of 1000 observations. We employ the cumsum [BN93] technique to distinguish normal mutual information from abnormal mutual information. Specifically, an attack is declared when $S_k >$`Threshold`, where $S_{k+1} = \max\left(0, S_k + I_k - \bar{I}\right)$, with $S_0 = 0$. Figure 10 shows the time series of S before and during the UDP flooding attack. The start of the attack can easily be observed by the sharp rise in S toward the end of the trial.

Clearly, the performance of the detection scheme is related to the value of `Threshold`. Figure 11 shows the relationship between the false alarm rate and `Threshold`. A false alarm is said to occur if S_k exceeds `Threshold`. After a false alarm, S is reset to 0 and the time series is continued to be processed. As expected, as `Threshold` grows, the false alarm rate decreases. No false alarms occurred for threshold above 160, hence no points are included for `Threshold`> 160. However, as long as `Threshold` is below 1600, the attack is detected.

VII. CONCLUDING REMARKS

The investigations reported here have demonstrated that some specific attack scenarios, while perhaps not visible by naked eye observation of traffic plots, nevertheless create dynamical shift substantial enough for the mutual information to be affected in a sizable manner. It appears that the signal to be monitored is the link utilization at some link in the vicinity of the target of the attack. Results have shown that mutual information is especially useful in detecting flooding attacks such as CBR attacks. Other attacks, like SYN, which disrupts the normal sequencing of control and data packets, would require a distinction between control and data packets, which is left for further research. From a broader perspective, it appears that TCP traffic has a mutual information signature distinct from that of non-TCP traffic, so that any deviation, malicious or not, from TCP would be detectable. While the utility of mutual information has been demonstrated through simulations and experiments, further work is required to determine the performance under the wide range of traffic scenarios found in real networks.

REFERENCES

[BN93]   M. BASSEVILLE AND I. NIKIFOROV. "Detection of abrupt changes: theory and application". Prentice Hall, Englewood Cliffs, NJ (1993).

[MPR02]   JELENA MIRKOVIC, GREGORY PRIER, AND PETER L. REIHER. Attacking DDoS at the source. In "Proceedings of the 10th IEEE International Conference on Network Protocols (ICNP)", pages 312–321. IEEE (2002).

[SP04]   V. SIRIS AND F. PAPAGALOU. Application of anomaly detection algorithms for detecting SYN flooding attacks. In "Proceedings of IEEE Global Telecommunications Conference (Globecom 2004)", volume 4, pages 2050–2054. IEEE (2004).

[WBMW04]   CYNTHIA WONG, STAN BIELSKI, JONATHAN M. MCCUNE, AND CHENXI WANG. A study of massmailing worms. In "WORMŠ04". ACM (2004).

[WKO05a]   DAVID WHYTE, EVANGELOS KRANAKIS, AND P.C. VAN OORSCHOT. ARP-based detection of scanning worms within an enterprise network. In "Proceedings of the Annual Computer Security Applications Conference (ACSAC 2005)" (2005).

[WKO05b]   DAVID WHYTE, EVANGELOS KRANAKIS, AND P.C. VAN OORSCHOT. DNS-based detection of scanning worms in an enterprise network. In "Proceedings of the 12th Annual Network and Distributed System Security Symposium (NDSS 2005)" (2005).

[WZS04]   HAINING WANG, DANLU ZHANG, AND KANG G. SHIN. Change-point monitoring for the detection of DoS attacks. *IEEE Transactions on Dependable and Secure Computing* **1**(4), 193–208 (2004).

[AAB01]   V. ALARCON-AQUINO AND J. A. BARRIA. Anomaly detection in communication networks using wavelets. *IEEE-Proceedings-Communications* **148(6)**, 355–62 (2001).

[AK98]   R. ANDERSON AND A. KHATTAK. The use of information retrieval techniques for intrusion detection. *First International Workshop on the Recent Advances in Intrusion Detection (RAID)* (1998).

[Aka75]   H. AKAIKE. Markovian representation of stochastic processes by canonical variables. *SIAM J. Control* **13**, 162–173 (1975).

[BF85]   L. BREIMAN AND J. H. FRIEDMAN. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association* **80**, 580–619 (1985).

[BKPA02] PAUL BARFORD, JEFFERY KLINE, DAVID PLONKA, AND RON AMOS. A signal analysis of network traffic anomalies. *ACM SIGCOMM Internet Measurement Workshop* (2002).

[BKRT01] RUDOLF B. BLAZEK, HONGJOONG KIM, BORIS ROZOVSKII, AND ALEXANDER TARTAKOVSKY. A novel approach to detection of denial-of-service attacks via adaptive sequential and batch sequential change-point detection methods. *IEEE Systems, Man and Cybernetics Information Assurance Workshop* (2001).

[BSTV97] F. BRINI, S. SIBONI, G. TURCHETTI, AND S. VAIENTI. Decay of correlation for the automorphism of the torus $\mathbb{T}^2$. *Nonlinearity* **10**, 1257–1268 (1997).

[CB96]   M. E. CROVELLA AND A. BESTAVROS. Self-similarity in world wide web traffic: Evidence and possible causes. *IEEE/ACM Transactions on Networking* **6**, 835–846 (1996).

[Cen]   CERT-COORDINATION CENTER. Overview of attack trends. *http://www.cert.org/archive/pdf/attack$_t$rends.pdf*.

[CER]   CERT. 'CERT advisory CA-96.01: UDP port denial-of-service attack. *ftp://info.cert.org/pub/cert_advisories/ca-96.01.udp_service_denial*.

[CKT02]   CHEN-MOU CHENG, H.T. KUNG, AND KOAN-SIN TAN. Use of spectral analysis in defense against DoS attacks. *IEEE GLOBECOM* (2002).

[Den87]   D. E. DENNNIG. An intrusion detection model. *IEEE Transactions on Software Engineering* **13(2)**, 222–232 (February 1987).

[EBH01]   S. EVANS, S. F. BUSH, AND J. HERSHEY. Information assurance through kolmogorov complexity. *DARPA Information Survivability Conference and Exposition 2* (2001).

[FGW98]   ANJA FELDMANN, ANNA C. GILBERT, AND WALTER WILLINGER. Data networks as cascades: Investigating the multifractal nature of internet WAN traffic. *SIGCOMM* pages 42–55 (1998).

[For96]   S. FORREST. A sense of self for UNIX processes. *Proceedings of IEEE Symposium on Security and Privacy* pages 120–128 (May 1996).

[GWC98]   A. GHOSH, J. WANKEN, AND F. CHARRON. Detection anomalous and unknown intrusions agains programs. *Annual Computer Security Applications Conference* pages 259–267 (Decemeber 1998).

[HHP03]   A. HUSSAIN, J. HEIDEMANN, AND C. PAPADOPOULOS. A framework for classifying denial of service attacks. *ACM SIGCOMM* (2003).

[htta]   HTTP://WWW.ISI.EDU/NSNAM.

[httb]   HTTP://WWW.ISI.EDU/NSNAM.

[JH85]   E. JONCKHEERE AND J. HELTON. Power spectrum reduction by optimal hankel norm approximation of the phaseof the outer spectral factor. *IEEE Transactions on Automatic Control* **30, No. 12**, 1192–1201 (1985).

[JV91]   H. S. JAVITZ AND A. VALDES. The SRI IDES statistical anomaly detector. *IEEE Symposium on Seciryt and Privacy* (May 1991).

[JW92a]   E. JONCKHEERE AND B. F. WU. Mutual kolmogorov-sinai entropy approach to nonlinear estimation. *IEEE Conference on Decision and Control* pages 2226–2232 (1992).

[JW92b]   E. JONCKHEERE AND B. F. WU. Mutual kolmogorov-sinai entropy approach to nonlinear estimation. *IEEE Conference on Decision and Control* pages 2226–2232 (1992).

[Ken00]   S. KENT. On the trail of intrusions into information systems. *IEEE Spectrum* **37 (12)**, 52–56 (December 2000).

[KRL97]   C. KO, M. RUSCHITZKA, AND K. LEVITT. Execution monitoring of security-critical programs in distributed systems: A specification-based approach. *IEEE Symposium on Security and Privacy* pages 175–187 (May 1997).

[Kul68]   S. KULLBACK. Information theory and statistics. *Dover* (1968).

[Lar91]   W. E. LARIMORE. Identification and filtering of nonlinear systems using canonical variate analysis. *Nonlinear Modeling and Forecasting, SFI studies in the Sciences of Complexity* **12** (1991).

[LB98]   T. LANE AND C. E. BRODLEY. Temporal sequence learning and data reduction for anomaly detection. *5þ ACM conference on Computer and Communications security* pages 150–158 (1998).

[LB03]   N. X. LIU AND J. S. BARAS. On scaling property of network traffic in small scales. *submitted, Computer Networks* (2003).

[LS99]   W. LEE AND S. STOLFO. A framework for construction features and models for intrusion detection systems. *5þ ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (August 1999).

[LTG+92] T. LUNT, A. TAMARU, F. GILHAM, R. JAGANNATHAN, P. NEUMANN, H. JAVITZ, A. VALDES, AND T. GARVEY. A real-time intrusion detection expert system (IDES). *Technical report, Computer Science Laboratory, SRI international* (1992).

[LTWW94] W. LELAND, M. TAQQU, W. WILLINGER, AND D. WILSON. On the self-similar nature of ethernet traffic(extended version). *IEEE/ACM Transactions on Networking* pages 1–15 (1994).

[Man77] Y. I . MANIN. "A Course in Mathematical Logic". Springer-Verlag (1977).

[MVS01] DAVID MOORE, GEOFFREY VOELKER, AND STEFAN SAVAGE. Inferring internet denial of service activity. *USENIX Security Symposium* (2001).

[NS89] V. V. NEMYTSKII AND V. V. STEPANOV. "Qualitative Theory of Differential Equations". Dover, New York (1989).

[Pax99] V. PAXSON. Bro: A system for detecting network intruders in real-time. *IEEE Computer Networks* **31(23-24)**, 2435–2463 (1999).

[PCJ+02] CRAIG PARTRIDGE, DAVID COUSINS, ALDEN JACKSON, RAJESH KRISHNAN, TUSHAR SAXENA, AND W. TIMOTHY STRAYER. Using signal proceesing to analyze wireless data traf?c. *ACM workshop on Wireless Security* (2002).

[PE95] P. PRUTHI AND A. ERRAMILLI. Heavy-tailed ON/OFF source behavior and self-similar traffic,. *IEEE* pages 445–450 (1995).

[Roe99] M. ROESCH. Snort-lightweight intrusion detection for networks. *USENIX LISA Conference* (November 1999).

[SBJ03] K. SHAH, S. BOHACEK, AND E. JONCKHEERE. The predictability of data network traffic. *American Control Conference* (2003).

[SE03] M. SOW AND A. ELEFTHERIADIS. Complexity distorsion theory. *IEEE Transactions on Information Theory* **49**, 604–608 (2003).

[Sip97] M. SIPSER. Introduction to the theory of computation. *PWS Publishing Company* page Boston: PWS Publishing Company (1997).

[SJJ02] S. STANIFORD, J.A. HOAGLAND, AND J.M. MCALERNEY. Practical automated detection of stealthy portscans. *Journal of Computer Security* (2002).

[ST94] G. SAMORODNITSKY AND M. S. TAQQU. Stable non-gaussian random processes, stochastic models with infinite variance. *Chapman Hall* (1994).

[Sta98] WILLIAM STALLINGS. "High-Speed Networks TCP/IP and ATM Design Principles". Prentice Hall, 1st edition edition (1998).

[TCL99] H.S. TENG, K. CHEN, AND S. C-Y LU. Adaptive real-time anomaly detection using inductively generated sequential patterns. *IEEE Symposium on Security and Privacy* (1999).

[TJ03] MARINA THOTTAN AND CHUANYI JI. Anomaly detection in IP networks. *IEEE Transactions on signal processing* **51(8)**, 2191–2204 (August 2003).

[WD99] C. S. WALLACE AND D. L. DOWE. Minimum message length and Kolmogorov complexity. *The Computer Journal* **42, no. 4**, 270–283 (1999).

[Wu92] B. F. WU. Identification and control of chaotic processes-TheKolmogorov-sinai entropy approach. *ph.d. dissertation, dept. of electrical engineering–systems, university of southern california* (1992).

[WZS02] H. WANG, D. ZHANG, AND K. SHIN. Detecting SYN flooding attacks. *IEEE INFOCOM* (2002).

[ZL70a] A. ZVONKIN AND L. LEVIN. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Mathematics Surveys* **256**, 83–124 (1970).

[ZL70b] A. K. ZVONKIN AND LA LEVIN. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Mathematics Surveys* **256**, 83–124 (1970).

[ZRMD03] ZHI-LI ZHANG, VINAY RIBEIRO, SUE MOON, AND CHRISTOPHE DIOT. Small-time scaling behaviors of internet backbone traf?c: An empirical study. *IEEE INFOCOM* (2003).