Modeling Spatial and Temporal Behavior of Internet Traffic Anomalies

Vidarshana Bandara, Ali Pezeshki, and Anura P. Jayasumana
Department of Electrical and Computer Engineering
Colorado State University
Fort Collins, CO 80523-1373, USA
{vwb, pezeshki, anura}@engr.colostate.edu

***Abstract*— A new approach based on graph wavelets for analyzing the spatial and temporal behavior of Internet traffic anomalies is presented. This approach is applied to Internet2 traffic measurements to evaluate the time duration and spatial spread (number of links affected) of anomalies. Based on the empirical results, a node model is proposed that captures the behavior of anomalies at individual network nodes. The model considers various aspects of anomalies, such as its origin, termination, propagation, duration and volume changes. The derivation of the model parameters requires only local node information, but the model is capable of producing network-wide anomalies whose behavior mimics network wide anomalies. Model is verified by using Internet2 traffic data. Since the proposed model can be specified using only a few parameters, it can be used in place of large anomaly traces with a great data reduction. As extensions, the model is applied over a path and an aggregated model that applies to a neighborhood in the network is also presented. A method to use the graph wavelet components found during the analysis to implement a real-time anomaly monitoring system is also discussed.**

keywords— Graph wavelets; Internet traffic anomalies; Modeling network anomalies

I. INTRODUCTION

Identifying Internet traffic anomalies, such as flash crowds and denial of service attacks, along with their spatial and temporal characteristics (e.g., life time and spatial spread) is vital for robust network design and operation. These characterizations provide critical information for designing and updating link, buffer and router capacities that are necessary for stable operation. They can also be used to identify the vulnerable regions of the network and to plan for adversarial attempts. Understanding the behavior of traffic anomalies also helps improve QoS provisioning and performance modeling. Modeling anomaly properties directly contributes for studies and also enables higher level representations. The model presented in this paper captures the behavior of anomalies at a node and then extended to aggregate a region of the network. Such aggregations capture traffic behavior between ISP level regions.

Recognizing a deviation from the usual traffic pattern is the main goal in anomaly detection and analysis. A thresholding technique in which traffic/flow rate is compared with a threshold would be an obvious choice, but due to the bursty and self-similar nature of Internet traffic [15], and daily and weekly variations, such approaches are not adequate. Frequency domain solutions in which anomalies are observed and characterized at certain frequency bands have shown some promise [2]. Wavelets decomposition and time-series analysis techniques have been used in [3] to study the statistical characteristics of network traffic anomalies.

Fourier analysis is used in [18] to find the source of an anomaly (origin-destination), as well as to characterize regular trends. A simple but effective method to detect and diagnose "black-holes," where packets are dropped in large quantities due to faults, is presented in [17]. Other anomaly detection and tracking techniques using principal component analysis [12],[18], machine learning [24], data mining [21], statistical analysis of payloads (for intrusion detection) [22], and risk modeling and Bayesian analysis (for IP fault tracking) [14],[16] have also been proposed. The reader is referred to [23] for a more comprehensive literature survey. In addition, a number of active probing techniques have been reported [1] for detecting and localizing anomalies and their spatial spread over the network, by comparing probe measurements against service level agreements. Probing techniques have also been proposed for fault diagnosis [20].

In [4], a relatively low complexity spatial analysis of network traffic is presented, which can be implemented at individual routers across the network to alleviate the need for monitoring network-wide data at a central location. Graph wavelets [7] provide a new way for spatial traffic analysis at different granularities. The idea behind graph wavelets is similar to that of standard discrete wavelet transform, where wavelet coefficients at different scales are obtained by aggregating or differencing adjacent data points in time, with appropriate weights depending on the type of wavelet. In graph wavelets, adjacent points correspond to *neighborhoods* defined on the network graph. For example a neighborhood can be a collection of nodes that are within a radius of a node of interest on the graph, or they can be a collection of links that are all connected to the same node. The size of the neighborhood in turn represents fine or coarse scales for multi-scale analysis. The former definition for neighborhoods is used in [7] for spatial traffic analysis, where the graph corresponds to the line graph of a network. The graph wavelet approach of [7] has shown great promise for discovering traffic patterns at different spatial granularities and for

Figure 1. Network nodes and links from Internet2

extracting low-dimensional representations of the traffic data. Data adaptive approaches include diffusion wavelets [6], which have been used in [5],[10] for dimension reduction for network traffic analysis.

In this paper, we propose a new approach for analyzing and modeling Internet traffic anomalies. We use graph wavelet analysis to evaluate the contribution of each link at a node, as opposed to [7] where graph wavelet analysis was done over an expanding neighborhood of the network. In [7], links at a certain depth of the neighborhood are aggregated, so the contribution of an individual link becomes less significant. Analyzing links at each node provides an analysis which does not loose granularity. Our aim is to analyze the temporal spread (lifetime) and the spatial spread (spread path and extent) of traffic anomalies across the network, and to develop simple models that capture such behavior at different spatio-temporal scales for network traffic modeling. Our main contributions are as follows:

1) We first develop a multi-scale traffic analysis framework to analyze how traffic anomalies migrate through the network. Using graph-wavelet coefficients of the traffic data we determine how long an anomaly persist in the network, which route it takes, and how deep it spreads through the network. We show that this behavior can be adequately described using well-known statistical distributions.

2) We then develop a model that captures the input-output relationship between anomaly traffic at a node. We extend this model to capture the overall (composite) input-output relation of the anomaly traffic for a set of nodes in the network. We also propose a real-time distributed detection system for traffic anomalies.

In Section II, the traffic measurements and data preprocessing steps are discussed. Section III presents the proposed graph-wavelet based technique for tracing anomalies. Results explaining the spatial and temporal behavior of anomalies are presented in Section IV. A model for anomaly generation and propagation at nodes is presented in Section V. Anomaly characteristics observed on Internet2 are presented in Section VI in terms of model parameters. Application of the model is discussed in Section VII, and conclusions are drawn in Section VIII.

II. Dataset and Preprocessing

We use traffic volume measurements from Internet2 network is shown in Fig. 1, starting from Oct. 16th, 2005 [11]. The data were collected at 11 nodes probing with 5 minute intervals of 28 inbound and outbound links. However the techniques applied are very general and thus can be used to analyze measurements from any network.

Internet traffic is best characterized as bursty and self-similar [15]. Nonetheless, daily and weekly slow-varying patterns are present in traffic measures [2] as evident in Fig 2. Before applying a threshold to detect anomalies, data has to be preprocessed to remove these trends as otherwise such trends could mask an anomaly. For example the measure of anomaly (A) in Fig 2 is less than most of the peaks in the trace. Therefore the threshold is applied after de-trending the trace.

As probing is performed at 5 minute intervals, 2016 samples per link are collected over a complete week at every node. The common trend of traffic is identified by recognizing the prominent frequency components, by performing 2048 point FFT. In the Internet2 dataset, 20 frequency components sufficiently captured the common trends in a week. De-trending is done by zero-forcing these frequency components.

A simple thresholding is then performed to identify the large deviations (three times the standard deviation) back in the time domain. We consider these large deviations as traffic anomalies and we are interested to understand the spatial and temporal behavior of them. Fig. 3 shows an example of the preprocessed data: Large deviations in the raw data (blue) are preserved in both the de-trended data (cyan) and the thresholded (red) data. But the traffic trends present in the observed data are significantly suppressed in the de-trended data. Thresholding has revealed the anomalies on the dataset.

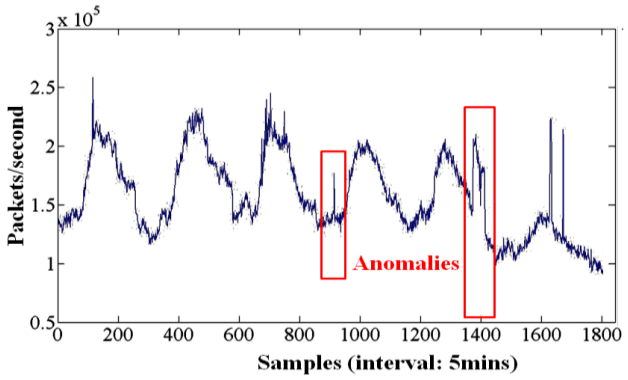

Figure 2. A sample weekly traffic trace

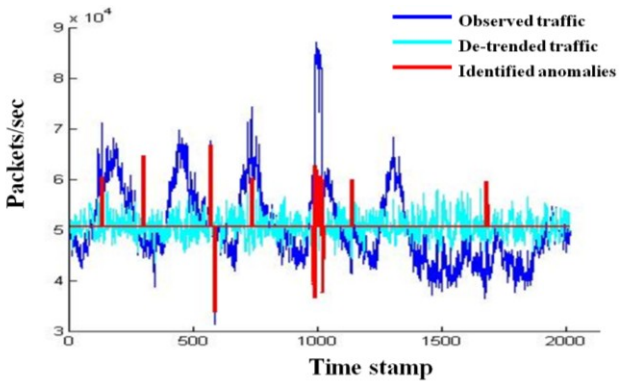

Figure 3. Detected anomalies *(marked in red)*

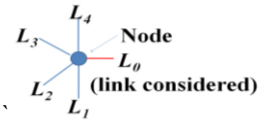

Figure 4. Multi-scale analysis of links around a node

III. GRAPH WAVELETS FOR TRACING ANOMALIES

Our aim is to analyze the spatial and temporal behavior of the anomalies across the network. Specifically, we wish to analyze the temporal spread (lifetime) and the spatial spread (spread path and extent) of the anomalies to understand how an anomaly affects the networks operation.

A. Graph Wavelets

To explain the basic idea behind graph-based wavelets, let us first recall the simplest form of Discrete Wavelet Transform (DWT), i.e., the Haar wavelet [8]. The Haar multi-scale representation of the data is obtained by forming differences between aggregated versions of the data at different scales. For instance, the Haar wavelet coefficients at the first scale are simply the differences between adjacent data points. The coefficients at higher scales are obtained by computing differences between aggregated data points in neighboring dyadic intervals.

In this paper, we apply Graph-based wavelets at each node over the links. This is different from the approach in [7], where analysis was done over the radius of neighborhood. Analysis on each node captures all the activities at each of them. Further, the wavelets coefficients reveal certain physical properties which depend of the degree of the node, as shown in Table I. These properties are explained later in Section V.

Referring to Fig. 4, suppose an anomaly is detected in L_0 . We wish to analyze how incoming traffic from other links contribute to the anomaly effect in L_0, or how the anomaly in L_0 affects the outgoing traffic in other links connected to the node. We take a multi-scale approach, that is, we study the anomaly propagation at various scales, with 1 corresponding to the finest scale. At scale 1, we look at weighted differences between the traffic data at L_0 and the traffic data at each of the other links one by one. These differences are viewed as graph-wavelet coefficients at scale 1. In the example depicted in Fig. 4, there are four scale-1 coefficient time series, indexed from 1 to 4 depending on their location with respect to L_0 in a clockwise fashion. The scale-1 analysis compares the link under consideration with the other links individually and captures the difference (actually similarity) between the links.If the difference between L_0 and another link is small (magnitude-wise), we conclude that the anomaly must have

TABLE I. PROPERTIES REVEALED BY GRAPH WAVELET COEFFICIENTS

Node degree	*Coefficient*		
	Scale-1	***Scale-2***	***Scale-3***
2	Originating/Propagating anomalies	N/A	N/A
3	Originating anomalies	Propagating anomalies	N/A
4	Originating anomalies	Propagating anomalies	Remaining contribution

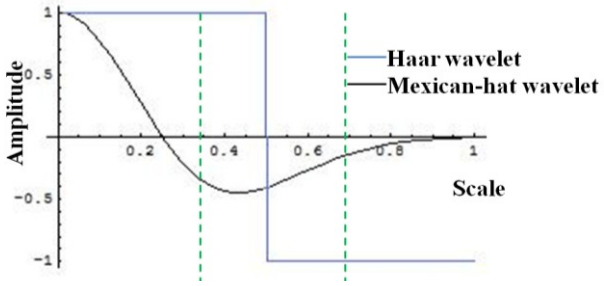

Figure 5. Partitioning of the wavelet function for weight calculation

propagated from there. On the other hand, if the difference is large then anomaly must have propagated from other links.

The scale 2 analysis is coarser, and does not identify how every individual link is similar or dissimilar to L_0. Rather it explains how the aggregated traffic at a pair of nodes contributes to or is affected by the anomaly at L_0. Letting L_i denote the traffic in link i, at scale 2 we form differences (possibly with weights) of the form L_0-(L_i+L_j), for all $i,j=1,2,\ldots,n-1$ and $i\neq j$. Here n is the total number of nodes.

At scale 3, the aggregated traffic at a collections of three links is compared with the traffic in L_0, that is we look at link differences of the form L_0-$(L_i+L_j+L_k)$, where (i,j,k) enumerate all distinct link triples. Higher scale analyses, up to $n-1$, are defined in a similar fashion. Under the Haar wavelet an N-scale analysis will have the form:

$$\hat{L}_N = L_0 - \sum_{\forall\ sizeN\ link\ sets} L_i \quad (1)$$

B. Coefficient Assignment

The values of the weights are determined by the choice of the wavelet function. To determine the weights we first partition the wavelet function in time into n intervals of equal duration and then calculate the area under each interval. An example of this partitioning is shown in Fig. 5 for the Haar and Mexican Hat wavelets and the corresponding weight values for scales 1 and 2 are given in Table II.

C. Tracing Anomalies

The next phase is to traverse the links that had a contribution to the anomaly. Figure 6 presents the pseudocode for the traversal. Two types of traversals are involved depending on the direction of the traffic analyzed. Forward tracking (by invoking *trace(forward)*) is performed by iteratively comparing an inbound link with the outbound links connected to the same node.

Similarly backward tracking (by invoking *trace(backward)*) is performed by iteratively comparing an outbound link with inbound links connected to the node. Tracing terminates when no links show significant contribution to the anomaly. This approach selects links contributing to the anomaly explicitly, and generates a map of the spread of the considered anomaly over the entire network. The only information passed from one iteration at a node to the next iteration at the other node is the duration information (time window) of the anomaly. The time window is moved and stretched/shrunk to find the best correlating time window to pick the anomaly. The same anomaly may be marked at different routers at different

TABLE II. WEIGHTS FOR SCALE 2 ANALYSIS FOR THE HAAR AND THE MEXICAN-HAT WAVELETS

	Haar wavelet	*Mexican-hat Wavelet*
scale-0	+1	0.548
scale-1	0	-0.472
scale-2	-1	-0.075

timestamps, due to propagation delays time synchronization tolerances, and duration variation due to traffic shaping. Matching the best correlating time window overcomes these issues.

IV. RESULTS

We use the anomaly detection and traversal methods presented above to investigate properties of anomalies. The revealed properties are characterized by fitting to appropriate statistical distributions.

A. Anomaly Detection and Tracing

Fig. 7 shows a sample trace of a detected anomaly (marked in red) in the link from Denver to Kansas City. Haar wavelet-based tracing indicates that the anomaly originated in the Sunny Valley to Denver link (by backward tracking), and terminated after the Kansas City to Indianapolis link (by forward tracking). For the studies presented some 4313 anomalies observed in a period of 50 weeks starting from Oct. 16th, 2005 are used.

B. Anomaly Characterization

Next we characterize the anomalies observed over the entire network, in terms of the distribution of their time duration, and the distribution of their spatial spread. Since anomalies are rare occurrences, data has to be collected over a long period of time to calculate reasonably accurate statistics. A statistically significant sample set large enough to evaluate properties of anomalies is generated by averaging traffic over 10 weeks.

```
trace (direction)  ## direction = enum{forward, backward}
    Initialize Link_set ← {}
    If direction==forward
        Link_set ← {outgoing links}
    Else
        Link_set ← {incoming links}
    Call select(link set)
    If none
        exit
    Else
        For each link in the chosen links
            call trace(direction)

select (link_set)
    initialize window ← window of anomaly on current link
    ## window = struct{start_time, end_time}
    Initialize chosen_links ← {}
    for each link in link_set
        move start_time to start of anomaly on the link
        move end_time to end of anomaly on the link
        if start (end_time - time_end) > 0
            chosen_links ← chosen_links ∪ link
    return chosen_links
```

Figure 6. Pseudocodes for anomaly tracing

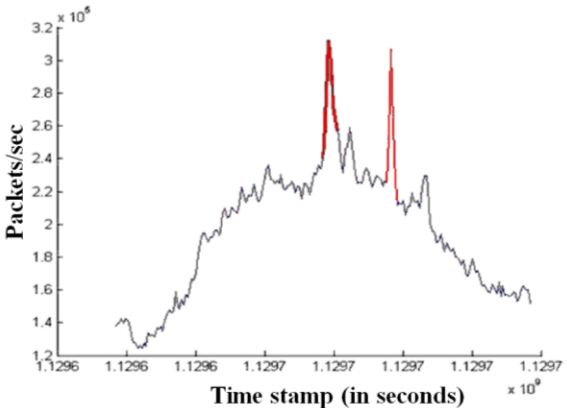

Figure 7. Traffic from Denver to Kansas City link *(the spikes marked in red are anomalies detecteds)*

Figure 8 shows the average time duration characteristics of anomalies over the entire network. These results show the distribution of the time duration of anomalies in five 10-week periods. Time distribution shows that they all follow the same distribution, as well show a thorough compliance to a geometric distribution with parameter 0.5 in timestamps. This means that about 50% of anomalies die within 5 minutes. About 25% of the anomalies last beyond 5 minutes but disappear within 10 minutes, etc.

The spatial spread is assessed in terms of number of links involved in the anomaly, and depth to which the anomaly descends. The spatial spread distribution explains the probability an anomaly would prevail in a particular number of links down the network. This distribution heavily depends on the structure of the network. Though the probability distribution for any 10 weeks period is found to be nearly the same, finding a standard distribution that will fit well (as with the case of duration analysis) is not possible.

The "depth" an anomaly propagates is less network-structure-dependent than the set of links affected by the anomaly. Therefore depth is a more reasonable measure for spatial analysis. Here, the number of levels an anomaly spread is counted, instead of the number of links. We found the distributions to be well converged, and to be close to a Log-Pearson type III distribution. The Kolmogorov-Smirnov (KS) test [9] verified this claim. The resulting distribution is shown in Fig. 9.

The KS test [9] was performed on all the 10-week datasets to assure our claims. For most of the observation sets, KS-test

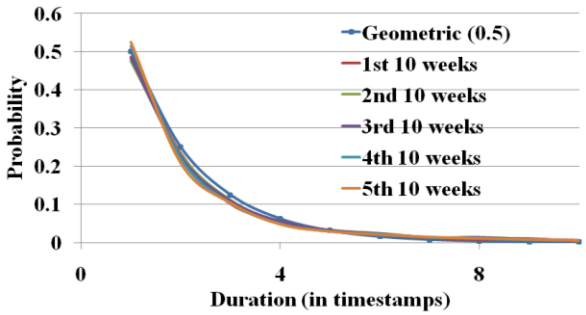

Figure 8. Probability distribution of average time duration of anomalies

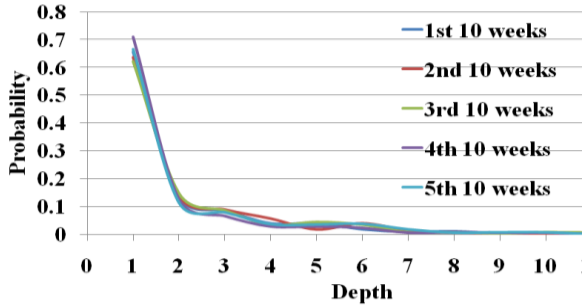

Figure 9. Probability distribution of spatial spread *(in terms of depth of the anomaly)*

confirmed similarity to the claimed distributions with a P-value close to 1.

V. ANOMALY MODEL

Being able to regenerate anomalies similar to those in practical networks is essential for accurate network modeling, evaluation and forecasting. Based on the properties of anomalies observed, a model to describe the origination, propagation and termination characteristics of anomalies is proposed next. The model is intended to capture the statistical properties of anomalies. Distinct from other Internet traffic models, it identifies and characterizes different properties of anomalies. Such a model is useful for applications such as Internet traffic simulators - to generate anomalies having realistic statistical properties. The model is validated by comparing the statistics of anomalies generated by the model and actual anomalies observed in the network. Statistics of the anomalies generated by the model showed a maximum Kullback-Leibler (KL) divergence of 8% from the statistical properties of detected anomalies. KL divergence provides a distance between two probability densities compared. A low distance value claims the probability densities are similar. Accuracy of the model further verified when applied over a path as shown later in Section VII.

A. Proposed Model to Emulate Anomaly Behavior at a Node

The proposed model captures the behavior of anomalies at a network node. Figure 10 shows the basic structure of the proposed model for a node having three links. Traffic flows marked are only for the link-i. Each link connects the node at an interface block, represented with the detailed structure shown in Fig. 11. To characterize the core, a splitting ratio is used where a fraction α_{ij} (α_{ik}), of anomalies received at interface-*i* propagate to interface-*j* (*k*).

Parameters of the model are derived using the characteristics observed in the network. The core is specified using a splitting ratio, which is the fraction of anomalies that would take each path. The interface block is specified using two probability values and three probability distributions as presented in Table III, which also indicates the values for an example node based on measurements.P_{abs} is the probability a received anomaly would be absorbed, i.e., moved out of the network. For certain anomalies a responding anomaly can be observed on the same link in the reverse direction. These can

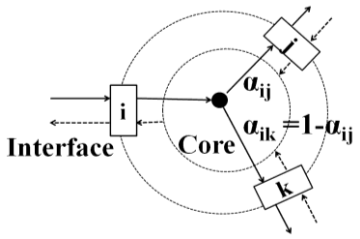

Figure 10. Basic anomaly model for a network node

be interpreted as acknowledgements for the anomalies received. The probability of having a responding anomaly is identified with P_{res}. The inter-anomaly gap is found to have an exponential distribution (noted with Exp(mean) and valued in hours) is marked as pdf_{gen}. When anomalies propagate and get responded their volume is adjusted. These factor by which the volume is adjusted is explained using a normal distribution (noted with Norm(mean, standard deviation)) and identified with pdf_{vol} and pdf_{res_vol} respectively.

B. Utilizing the Model

The proposed model is capable of regenerating anomalies having similar statistical properties as observed in the actual network. The model covers the three types of outgoing anomalies:

1) An originating anomaly
2) An anomaly as a response to a received anomaly
3) As propagation of a received anomaly

The inter-arrival time of originating anomalies and the initial anomaly volume showed a wide range of behaviors on different links. Yet in general, it could be observed that an originating anomaly would have about 3 million to 18 million packets distributed logarithmically, and the inter-arrival time between originating anomalies have a Poisson distribution. When inter-arrival times of anomalies of each link were fitted with an exponential distribution they only had standard error of about 5%. An example case is shown in Fig. 12, for anomalies observed in the Chicago to New York link. The histogram of inter-arrival times of anomalies is shown in bars in Fig 12. The histogram can be closely approximated to an exponential curve as shown. It was also noted that the distributions used to characterize the volume adjustment when

TABLE III. PARAMETERS IN THE INTERFACE BLOCK

Parameter	Definition	Sample values (for Kansas City node's Denver interface)
P_{abs}	Probability of absorbing a received anomaly completely	68%
pdf_{gen}	Distribution of inter anomaly gap (in hours)	Exp(45.33)
P_{res}	Probability of responding to a received anomaly	3%
pdf_{vol}	Distribution of volume asjustment	Norm(1,0.188)
pdf_{res_vol}	Distribution of volume change of the responding anomaly	Norm(1,0.305)

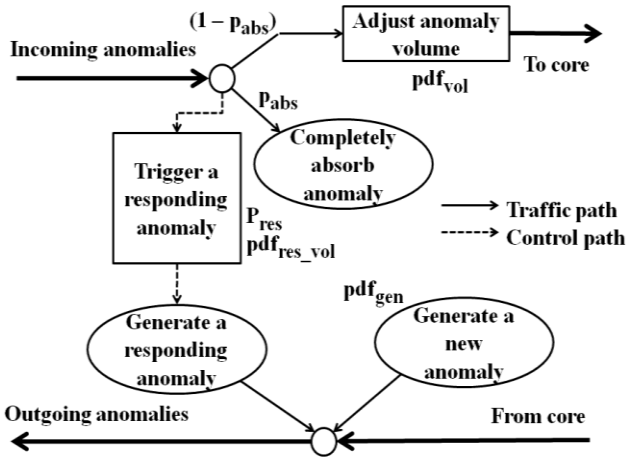

Figure 11. Structure of the interface model

an anomaly propagate, and when a responding anomaly is generated could be approximated using a normal distribution. A standard error of about 10% was seen at worse case.

Once the model parameters (shown in Fig 10 and 11) are identified for a network, anomalies having similar statistical properties can be generated. The complete process to regenerate anomalies having similar statistical behaviors to the observed is summarized using a pseudocode in Fig. 13. Originating anomalies are generated with an inter-anomaly gap of pdf_{gen}. A receiving node generates a responding anomaly in the reverse link with probability P_{res}, with the volume related to the incoming anomaly by a multiplicative factor randomly distributed according to pdf_{res_vol}. A receiving node absorbs the received anomaly with P_{abs}, or propagates it after changing the volume by a multiplicative factor given by pdf_{vol} and forwards to the core. If the node has more than one out-going link, the anomaly will choose a link with the links splitting ratio. The outgoing interface spread the anomaly volume over a period given by geometric(0.5) distribution.

VI. Anomaly Parameters of Internet2

The model captures the statistical properties of anomalies observed at a node. Thus now it becomes possible to evaluate or classify the anomalies, and also regenerate anomalies having the same statistical behavior as observed at each node.

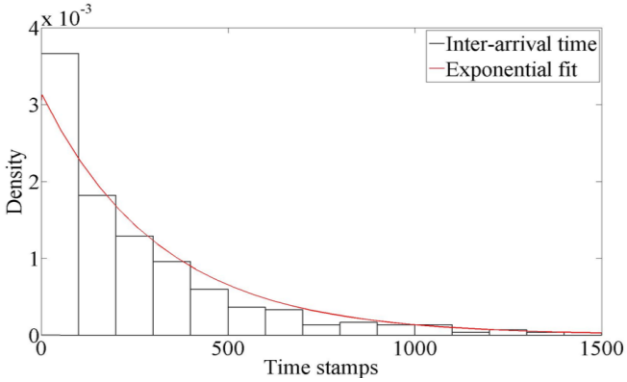

Figure 12. Inter-arrival time distribution on the Chicago to New York link Structure of the interface model

Anomaly receive process
Call responding anomaly generation process
With P_{abs} absorb anomaly
If the anomaly not absorbed
Adjust volume according to pdf_{vol}
Pass to core

Responding anomaly generation process
With $(1\text{-}P_{res})$ discard the responding anomaly
If not discarded
Construct an anomaly according to pdf_{res_vol}
Force an outgoing anomaly

Anomaly passing to core process
According to splitting ratio choose the outgoing link
Pass the anomaly

Anomaly generation process
According to pdf_{gen} or by a trigger from responding anomaly generation
Spread the anomaly over time according to geometric(0.5)
Transmit anomaly

Figure 13. Pseudocode to impletement the model

Tables IV and V presents model parameter values for selected nodes of Internet2 network. Table V lists the splitting ratios. A propagating anomaly would pick an outgoing link based on the splitting ratio.

The link having a higher ratio is referred as the "dominant link" and the other as the "other link." As many phenomena of interest such as buffer overflow, packet loss, etc. occur due to traffic anomalies, such characterization is useful for network design and resource provisioning. Notably, only a few parameters (measured locally) are required to capture the statistical properties of anomalies over the entire network. As shown Section VII, these parameters can be extended over a path or a region. An extension to the model over a path and region is discussed in Section VII.

TABLE IV. Parameters in the Interface Block for Selected Nodes

Node	Link	P_{abs} (%)	Standard deviation of normal pdf_{vol}	P_{res} (%)	Standard deviation of normal pdf_{res_vol}	Mean of exponential pdf_{gen} (hours)
KSCY	Dnvr	39.3	2.17	23.5	0.926	45.33
	Ipls	31.8	0.617	14.4	1.27	42.28
	Hstn	77.1	57.8	29.4	2.60	33.86
HSTN	Kscy	73.0	31.9	31.6	2.50	37.96
	Losa	37.8	16.4	33.2	48.3	42.95
	atla	48.8	1.48	28.2	2.10	28.57
DNVR	Sttl	65.0	12.9	22.5	2.07	41.96
	Snva	50.8	6.68	31.0	2.68	24.11
	Kscy	23.0	1.72	15.9	10.3	38.38
IPLS	Kscy	30.1	0.599	18.0	1.83	41.79
	Chin	24.0	5.97	19.1	14.4	44.22
	Atla	73.5	12.5	26.1	1.02	37.11
STTL	Snva	93.1	201.0	13.1	114.0	30.13
	Dnvr	92.5	3.19	19.3	2.31	42.52
LOSA	Snva	86.9	5.93	20.1	2.52	21.80
	Hstn	81.9	2.76	30.8	5.24	25.88
WASH	Atla	76.5	3.60	31.7	3.25	29.73
	Nycm	82.6	1.55	9.76	3.38	49.90

TABLE V. SPLITTING RATIOS FOR SELECTED NODES

Link	Preferred link	Splitting ratio (%)	Other link
atla-hstn	hstn-losa	95.0	hstn-kscy
atla-ipls	ipls-chin	70.5	ipls-kscy
chin-ipls	ipls-kscy	96.1	ipls-atla
dnvr-kscy	kscy-ipls	88.5	kscy-hstn
dnvr-snva	snva-losa	97.0	snva-sttl

TABLE VI. COMPARISON BETWEEN P_WS

Propagation	Actual P_W(%)	Model P_W(%)	Error
losa-snva-dnvr-kscy	68.1	70.4	2.3
nycm-wash-atla-hstn	95.0	88.9	6.1
sttl-snva-losa-hstn	98.6	98.6	0.0
chin-nycm-wash-atla	92.5	93.0	0.5
ipls-atla-hstn-losa	98.7	99.1	0.4

VII. APPLICATION OF THE MODEL

A. Anomaly Propagation Characteristics

The model proposed above captures propagation of anomalies over the network. The probability that an anomaly would survive over a path depends on P_{abs}s and $\alpha_{i,j}s$ of intermediate nodes and given by:

$$P_W = \prod_{i \in W} \left((1 - P_{abs,i}) \alpha_{c,i} \right) \quad (2)$$

where, W={links of the path}, $\alpha_{c,i}$: splitting ratio for the chosen outgoing link from link-i. A comparison between the P_W's derived from the model and the actual is shown in Table VI. It compares the probability an anomaly would survive over path provided by the model and actually observed. The error column states the difference between the probability values. The "Actual P_W" is the value observed in the network. The "Model P_W" is the value calculated using model parameters using equation (2). The error column is the difference between the two probability values.

As an anomaly propagates, its volume changes according to pdf_{vol} of each node in the path. The effective pdf of volume adjustment over a path is the product of individual pdfs over each link. As these pdfs are independent, the effective pdf can be expressed as follows:

$$pdf_{vol,W} = M^{-1}\{\prod_{i \in W} M\{pdf_{vol,i}\}\} \quad (3)$$

where W={links of the path}, M is the Mellin transform [19]. When the above statistics were derived for a number paths we observe a standard error of about 7% between the estimated and actual values.

B. Model for Node Aggregates

Further attempts were made to devise a higher level node model, which can capture a region, as in Fig. 14. Having each node characterized with a very few parameters, eases forming higher level nodes.

Aggregating nodes hides links, and a region has multiple internal paths. Entry points to the region will become interfaces to the aggregated node. The probability of absorption (P_{abs}) for an interface in the aggregated node is given by:

$$P_{abs} = 1 - \sum_{w \in T} P_w \quad (4)$$

where, T is the set of all internal paths starting from the interested interfaces of the aggregated node. The splitting ratio of an anomaly arriving at interface-i, choosing interface-j (α_{ij}) depends on all the possible internal paths exist between the two interfaces, and the splittings encountered internally,and is given by:

$$\alpha_{ij} = \sum_{t \in T_{ij}} \prod_{k \in t} \alpha_k \quad (5)$$

where, T_{ij} is the set of all the paths between interface-i to interface-j, α_k: splitting ratio of choosing link-k.

When Equations (4) and (5) are applied for the portion of network indicated in Fig. 14(a), to reach Kansas City from Los Angeles, there are two internal paths in the aggregated node. The first path has a probability of 29.75% to reach the interface to Kansas City. The path-2 (thru Seattle) has a P_W of 0.16%. Thus the cumulative probability to reach at the interface to Kansas City from Los Angeles interface is 29.91%. Therefore according to the model P_{abs} for the interface towards Los Angeles is 70.1%. The actual value was found to be 62.4%.

C. A Real-time Distributed Monitoring System

Although the analysis presented here is aimed at deriving a statistical model, the analysis can easily be extended to a real-time distributed system for monitoring anomalies. The wavelet coefficients presented in Section III(A) are computed using traffic around a node. Thus each router is capable of constructing a data-structure with its wavelet coefficients.

MIB (Management Information Base) on each router will perform the Fourier based detection on each of its links, and the wavelet analysis to produce the data-structure with the coefficients shown in Table I. By exchanging the data-structures with neighbors a description of anomalies in the local network can be constructed. The data-structure of a node contains the anomaly information on all its links. When a node is aware of it neighbors' data-structures, the node is aware of anomalies in a local network up to two links deep.

When an anomaly propagates towards a certain node, the observing node could forward all known coefficient data-

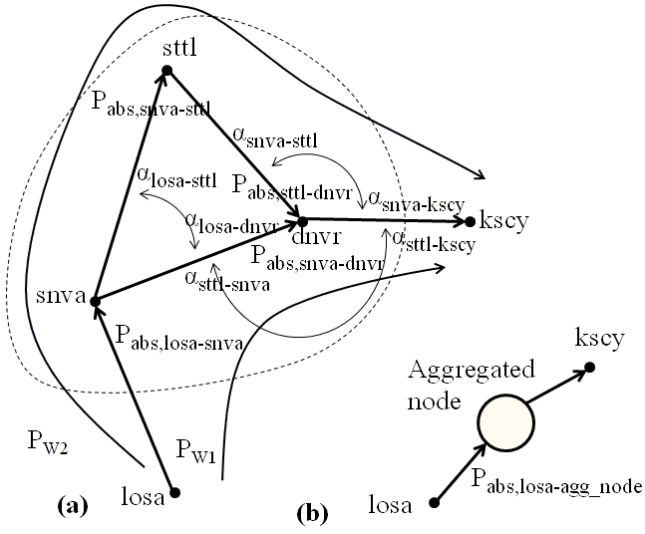

Figure 14. Portion of the network made into a higher level node

```
Throughput sampler
    Set Timer to sampling time
    At expiry
        Sample all links
        Call anomaly detector
        Call form wavelet coefficients
        Call exchange coefficients
Anomaly detector
    Apply Fourier transform to sample record
    Threshold and mark anomalies
Form wavelet coefficients
    Perform link comparison to form wavelet coefficients
Exchange coefficients
    Send coefficients to all neighbors
    Receive coefficients from all neighbors
    Construct message with own coefficients and neighbor coefficients
    Forward the message in the anomaly direction
```

Figure 15. Pseudocode for real-time distributed monitoring system

structures to the same node. Thus all nodes experiencing an anomaly have the capacity to develop the complete spatio-temporal map of the anomaly. Figure 15 shows the pseudocode for this scheme. Such a scheme enables any node observing an anomaly to derive properties of the anomaly. If the anomaly behavior of the network has been modeled, then these properties will enable predicting and decision making.

VIII. CONCLUSIONS

A method to characterize and model Internet traffic anomalies was proposed. To detect anomalies a simple Fourier-based method was employed. A new approach based on graph wavelets was developed to analyze the spatial and temporal behavior of anomalies. Measurement data for Internet2 was used to evaluate model parameters and to validate the region model. In particular, we studied how an anomaly propagates from one link to other links connected to the same node in the network (or vice versa). We demonstrated that the time duration and spatial spread of anomalies, observed in the Internet2 data used in this paper, can be adequately captured by standard statistical distributions. A node model capable of capturing the input-output relation for the anomaly traffic in a node was developed. This model was then extended to a composite input-output model, capturing the anomaly propagation over a path or region.

ACKNOWLEDGMENT

This work is supported in part by NSF grants CNS-0720889 and CCF-0916314.

REFERENCES

[1] P. Barford, N. Duffield, A. Ron, and J. Sommers, "Network Performance Anomaly Detection and Localization," *Proc. IEEE INFOCOM*, April, 2009, pp1-15.

[2] P. Barford, J. Kline, D. Plonka, and A. Ron, "A Signal Analysis of Network Traffic Anomalies." *Proc. 2nd ACM SIGCOMM Workshop on Internet Measurment*, Marseille, France, Nov. 06 - 08, 2002, pp71-82.

[3] P. Barford and D. Plonka, "Characteristics of Network Traffic Flow Anomalies," *Proc. 1st ACM SIGCOMM Workshop on Internet Measurement*, 2001, pp.69-73.

[4] P. Chhabra, C. Scott, E.D. Kolaczyk, and M. Crovella, "Distributed Spatial Anomaly Detection," *Proc. INFOCOM '08*, 13-18 April 2008, pp.1705-1713.

[5] M. Coates, Y. Pointurier, and M. Rabbat, "Compressed Network Monitoring for IP and All-Optical Networks," *Proc. 7th ACM SIGCOMM Conference on Internet measurement,* 2007, pp.241-252.

[6] R.R. Coifman and M. Maggioni, "Diffusion Wavelets," *Appl. Comp. Harm. Anal.*, 2006, vol. 21 no. 1, pp.53--94.

[7] M. Crovella and E. Kolaczyk, "Graph Wavelets for Spatial Traffic Analysis," *Proc. IEEE INFOCOM*, San Francisco, April 2003, pp1-10.

[8] I. Daubechies, "Ten Lectures on Wavelets", *SIAM*, 1992.

[9] M. H. DeGroot, Ch9 in *Probability and Statistics, 3rd ed.* Reading, MA: Addison-Wesley, 1991.

[10] J. Haupt, W.U. Bajwa, M. Rabbat, and R. Nowak, "Compressed Sensing for Networked Data," *IEEE Signal Processing Magazine - Special Issue on Compressive Sensing*, March 2008, vol. 25, no. 2, pp.92-101.

[11] http://noc.net.internet2.edu/i2network/live-network-status/historical-abilene-data.html

[12] L. Huang, X. Nguyen, M. Garofalakis, M.I. Jordan, A. Joseph, and N. Taft, "In-network PCA and Anomaly Detection," *In Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, 2007, pp.617-624.

[13] P. Huang, A. Feldmann, and W. Willinger, "A Non-intrusive, Waveletbased Approach to Detecting Network Performance Anomalies," *Proc. of ACM Internet Measurement Workshop '01*, November 2001, pp213-227.

[14] S. Kandula, D. Katabi, and J. Vasseur, "Shrink: A Tool for Failure Diagnosis in IP Networks," *Proc. of ACM SIGCOMM MineNet Workshop*, August 2005, pp172-178.

[15] T. Karagiannis, M. Molle, and M. Faloutsos, "Long-Range Dependence: Ten Years of Internet Traffic Modeling," *Internet Computing*, IEEE, Sept.-Oct. 2004, vol.8, no.5, pp.57-64.

[16] R. Kompella, J. Yates, A. Greenberg, and A. Snoeren, "IP Fault Location via Risk Modeling," *Proc. of NSDI '05*, May 2005, pp57-70.

[17] R. Kompella, J. Yates, A. Greenberg, and A. Snoeren, "Detection and Localization of Network Black Holes," *Proc. of IEEE INFOCOM '07*, May 2007, pp.2180-2188.

[18] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing Network-Wide Traffic Anomalies," *Proc. Conference on Applications, Technologies, Architectures, and Protocols For Computer Communications-SIGCOMM '04*, 2004, pp. 219-230.

[19] Z.A. Lomnicki, "On the Distribution of Products of Random Variables," *Journal of the Royal Statistical Society. Series B (Methodological)*, Blackwell Publishing for the Royal Statistical Society, vol. 29, no. 3 (1967), pp.513-524.

[20] M. Natu and A. Sethi, "Efficient Probing Techniques for Fault Diagnosis," *Proc. of the IEEE Conference on Internet Monitoring and Protection (ICIMP '07)*, July 2007, p20.

[21] B.R. Raghunath and S.N. Mahadeo, "Network Intrusion Detection System (NIDS)," *1st International Conference on Emerging Trends in Engineering and Technology, 2008 (ICETET '08)*, 16-18 July 2008, pp.1272 – 1277.

[22] S. Shanbhag and T. Wolf, "Correlation and Collaboration in Anomaly Detection," *Global Telecommunications Conference - IEEE GLOBECOM 2008*, Nov. 30 2008-Dec. 4 2008, pp.1 – 6,

[23] M. Steindera and A. Sethi, "A Survey of Fault Localization Techniques in Computer Networks," Science of Computer Programming, 2004, vol. 53, pp.165–194.

[24] M. Thottan and J. Chuanyi, "Anomaly Detection in IP Networks," *IEEE Trans.s Signal Processing*, Aug. 2003, vol. 51, no. 8, pp. 2191-2204.