Multiscale Energy Network Tomography

Yixiao Feng
Texas A&M University
College Station, USA
yixiao@tamu.edu

Nick Duffield, *Fellow, IEEE*
Texas A&M University
College Station, USA
duffieldng@tamu.edu

***Abstract*—Much effort and ingenuity has been applied to develop tomographic methods to derive link-level performance statistics (packet delay and loss rates) from end-to-end measurements. However, there has been recognition in recent years that network anomalies may manifest with more detailed spectral properties beyond these aggregate performance statistics. This motivates us to study whether network tomography can be used to attribute spectral properties observed in end-to-end measurements to local network regions, which would then be indicated as potential attack targets. Although spectral properties have been used to characterize network traffic and protocol and serve as features for anomaly detection, no current methods exist to localize these features to specific network links.**

In this paper we show how a tomographic analysis of end-to-end measurements can be used to estimate the wavelet energy of internal link delays. The proposed estimator, used for assessing link-level performance, is able to locate the link and scale with the highest energy that might be contributed by anomalies or congestion. We evaluate our method through world-wide realistic RTT measurements that shows a high estimation accuracy (100%) on identifying links/scales with the highest energy and small mean relative errors (<0.1).

***Index Terms*—Network tomography, Wavelet analysis, Network measurement**

I. Introduction

A. Time Series Tomography

In the canonical framework for network performance tomography we wish to recover a set of edge properties X through the linear relation $Y = AX$ that expresses measurable path properties Y that are aggregated by addition over edges, where A is the routing matrix of paths over edges. In time-series tomography the edge variables are time series $X_e = \{X_{e,t} : t \in T\}$ associated with a directed edge e and T is a temporal index set $T = \{1, \ldots, |T|\} \subset \mathbb{N}$. For each directed network path π of connected directed edges, let $Y_\pi = \{Y_{e,t} : t \in T\}$ be the time series of sums or aggregates

$$Y_{\pi,t} = \sum_{e \in E} A_{\pi,e} X_{e,t} \tag{1}$$

where $A_{\pi,e}$ is the incidence matrix of edges e over paths π, i.e., $A_{\pi,e} = 1$ if edge e occurs in path π and zero otherwise. Time-series tomography seeks to infer properties of the $\{X_e\}$ from those of the path variables $\{Y_\pi\}$. However, the linear system (1) is in general underconstrained and so does not admit a unique solution.

ISBN 978-3-903176-48-5

This paper concerns *Multiscale Energy Network Tomography*, i.e., how to infer the energy spectrum of edge time series $\{X_e\}$ at multiple time scales from pathwise time series $\{Y_e\}$ derived from end-to-end packet measurements, and how such inferences can be used to identify the edge localities of effects of anomalies according to their spectral properties.

B. Multiscale Traffic Analysis

Multiscale analysis has been used to characterize the complex nature of network traffic, examine the interaction between traffic demands and network protocols at different timescale, and compute features that support anomaly detection. Specifically, wavelet analysis provides a quantitative characterization through the set of wavelet coefficients associated with different timescales [1]. While self-similar behavior has been widely observed in network traffic traces since its original discovery in Ethernet traffic [2], wavelet analysis of WAN traffic reveals departures from self-similarity at timescales corresponding to round trip times. This is attributed to the flow control mechanism of TCP senders that is governed by acknowledgement from receivers [1]. Traffic source demands have been characterized as cascades, *i.e.,* a multifractal hierarchy of arrivals of sessions, flows within sessions, and packets within flows, each member of the cascade presenting at its own timescale [3]. The set of signal energies in different timescales can be used as feature for anomaly detection [4], [5]. Wavelet-based multifractal models have been applied to estimating effective bandwidth of network traffic [6]. These applications of wavelet-based multiscale analysis suggest that localization of observed signal energies to specific edges within a network can be valuable in identifying the origins of anomalies.

C. Spectral Properties of Anomalies

Network anomalies explore the vulnerabilities of network protocol and operations and bring down the normal traffic by taking available resources (*i.e.,* number of connections, bandwidth, etc) such as widely observed DDoS attack floods networks by an aggressive packet rate (*e.g.,* 1Tbps) [7]. Unlike the DDoS by attacking network with a high packet rate to take over network resources, low rate anomalies (*e.g.,* SlowComm [8], shrew [9], LoRDAS [10]) send very small attack traffic, which is around 10% $\sim$ 20% of the total traffic [7] based on periodic interference with TCP retransmission timers [9] and are extremely difficult to distinguish it from the normal traffic. Although the attack traffic sent by low rate anomalies

is small, it still can cause terrible damage to networks. For example, shrew attack sends burst of packets with a rate that matches the RTO value would cause network frequently timeout. Spectral analysis over signals generated by network anomalies can capture the variations over the number of packets or the transmission delay, which separates out the dominate frequencies of attack traffic from the normal traffic.

D. Problem Specification and Contribution

In this paper we address the following problems:

How to infer per link energy at multiple timescales in additive link metrics from end-to-end aggregate metrics. Figure 1 illustrates a simple two leaf tree topology where each node v_i denotes the real network device and edge $e_{i,j}$ denotes the connection between two network devices i and j. Our goal is to construct, for each timescale, an estimator $\hat{F}(X^{(i)})$ of the wavelet energy of the common path link (path intersection) $X^{(0)}$ from observed path measurements $Y^{(1)}$ and $Y^{(2)}$, i.e.,

$$\hat{F}(X^{(0)}) = f(Y^{(1)}, Y^{(2)}) \tag{2}$$

for a function f that we will determine. Furthermore, we aim to extend this result from the two-leaf tree paradigm to work with general network topologies.

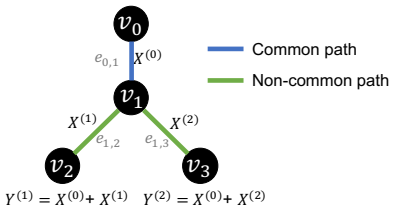


Figure 1: Two leaf tree model.

How to locate the link with the largest energy at each timescale that might indicate congestion or anomalies. The higher wavelet energy across all scales of a link indicates a larger sample variance on the collected time series. We further take the estimation results to identify the link with the largest energy and scales contributed to the most of energy, which might result from degraded link performance or anomalies. We also provide both simulation and realistic experiments on locating link with the largest wavelet energy.

Our contributions are as follows: 1) We introduced an unbiased estimator for the wavelet energy at multiple scales on the common path and non-common path using Discrete Wavelet Transform (DWT) and provided its statistical properties (§III), 2) We showed the proposed estimator can be applied on general network topologies (§IV), 3) The proposed estimator, evaluated by realistic RTT measurements gathered from the Ark platform, can accurately estimate the energy of internal links and identify those with highest energy (§V).

II. Related work

Much work in network tomography has focused on overcoming the limitations of underconstrained linear systems such as (1) that arises from network measurement. Several previous works have focused specifically on the problem of inferring properties of unobserved network time series $\{X_e\}$. [11] considered the related problem of time-varying traffic-matrix tomography, introducing power-law constraints between signal mean and variance in order to render the model identifiable from observations of these variables. While a modified EM approach [12] was used to find maximum likelihood solution, its computational complexity was prohibitive for real networks. Performance tomography from multicast probing is another approach to rendering an edge metric model identifiable by introducing constraints between path measurements [13]–[16]. This approach has been extended to estimate the frequencies of temporal subsequences of internal link packet loss [17] and latency [18]. Sparsity conditions have been applied as a constraint through a regularization in order to select "simpler" explanation of observed path time series [19], [20]. A computationally simpler approach has been proposed to infer link summary statistics in a parameter-free way, in particular, the variance of internal link metric time series based on the observation that for additive link metrics that are mutually independent [21]. Coupled with empirical models relating the relative ordering of metric variances and averages on path, this approach has been used to identify worst performing links and reduce dimension of the linear system (1) to the point that a restricted form can be solved for these links [22]. The conditions under which an unknown topology can be identified from covariance metrics has recently been established in [23].

Wavelet analysis has been widely used in capturing network abnormal behavior caused by anomalies (*e.g.,* [4], [5], [24]–[30]). Feldmann [3] used a wavelet-based detection mechanism on finding network anomalies and also illustrated the relationship between the wavelet energy and the RTT. Hussain [31] proposed a frame work using spectral properties to classifying DoS attacks. The low rate DoS attacks, on the other hand, send small burst of traffic to occupy network resources and lower the rate of normal TCP traffic based on periodic interference with TCP retransmission timers [9]. Many approaches (*e.g.,* [32]–[34]) on extracting features of low rate anomalies are focusing on exploring the spectral properties of traffic signals. While those works have focused on identifying spectral properties from end-to-end measurements [7], as far as we are aware these have not been explored by tomography.

III. Multiscale binary network tomography

In this section, we shall describe the unbiased estimator for inferring the common path multiscale statistics based on the network tomography of end-to-end measurements using DWT and illustrate its statistical properties.

A. Wavelet Decomposition of Signals

Let $Z = \{Z_t : t \in [T]\}$ be a signal with $T = 2^M$ components. The wavelet decomposition of Z involves writing it as linear combination of functions from an orthonormal basis as $Z_t = \sum_m \sum_n \tilde{Z}_n^m \phi_{n,t}^m$ where m and n label scale and translation parameters respectively. The wavelet coefficients are the scalar products with the corresponding basis vectors,

namely $\tilde{Z}_n^m = \sum_{t\in[T]} Z_t \phi_{n,t}^m$. Let ϕ^m be a $T/2^m \times T$ orthonormal DWT matrix (*i.e.,* constructed by shifting the Daubechies wavelet filter over time) with components $\phi_{n,t}^m$ such that wavelet coefficients $\tilde{Z}$ on scale m of time series Z can be obtained by $\phi^m Z$. DWT usually requires a dyadic length of sample size T (*i.e.,* $T = 2^M, M \in \mathbb{Z}^+$) but signal extrapolation methods (*e.g.,* zero-padding, reflect-padding, *etc.*) can be used to construct dyadic length.

B. Unbiased Estimation of Internal Links Multiscale Statistics

We start by analyzing the simple network model as shown in figure 1. The common path refers to the shared path for the two leaf nodes (*i.e.,* $X^{(0)}$) and the non-common paths refers to $X^{(1)}$ and $X^{(2)}$. Let $X^{(0)}, X^{(1)}, X^{(2)}$ be mutually independent signals on $[T]$ which are stationary in the sense that for any $S \subset [T]$ and $s \in [T]$ for which the translation $s + S \subset [T]$, $\{X_t : t \in S\}$ has the same distribution as $\{X_{s+t} : t \in S\}$. Set $Y^{(i)} = X^{(0)} + X^{(i)}$ for $i = 1, 2$ represents the aggregated path statistics. Much of our analysis will rest on the properties of the product of the wavelet transformation matrix ϕ with its adjoint. We abstract this as a T-dimensional square matrix B for which we will henceforth assume the constant signal $\mathbf{1}$ is in its null space, and our analysis will then apply to wavelet bases for which this property holds.

Define a quadratic form F of a signal Z on $\mathbb{R}^T$ by $F(Z) = Z^T BZ$, and define the estimator on the common path $X^{(0)}$

$$\hat{F}(X^{(0)}) = \frac{1}{2}\Big(F(Y^{(1)} + Y^{(2)}) - F(Y^{(1)}) - F(Y^{(2)})\Big) \quad (3)$$

Define the polynomial function G over two stationary signals $Z^{(1)}$ and $Z^{(2)}$ on $[T]$ as

$$G(Z^{(1)}, Z^{(2)}) = \sum_{t,t',s,s'\in[T]} \mathbb{E}[Z_t^{(1)} Z_s^{(1)}] B_{t,t'} B_{s,s'} \mathbb{E}[Z_{t'}^{(2)} Z_{s'}^{(2)}] \quad (4)$$

Let Δ be the bias of the estimator and

$$\Delta = \hat{F}(X^{(0)}) - F(X^{(0)}) = \sum_{i,j\in\{0,1,2\}} X^{(i)} \cdot BX^{(j)} (i \neq j) \quad (5)$$

Theorem 1. *(i)* $\mathbb{E}[BX^{(i)}] = 0$ *(stationary signal)*
(ii) $\mathbb{E}[\Delta] = 0$ *and hence* $\mathbb{E}[\hat{F}(X^{(0)})] = \mathbb{E}[F(X^{(0)})]$
(iii) $\mathrm{Var}(\Delta) = \sum_{i,j\in\{0,1,2\}} G(X^{(i)}, X^{(j)}) (i \neq j)$ [1]

Theorem 1 describes the unbiased estimator in (3) for the common path statistics based on observing aggregated path statistics Y by assuming mutual independence and stationarity of edge metrics. The bias Δ in Theorem 1(ii) is essentially the wavelet product of two time series associated with different edges such that $\Delta = X^{(1)} \cdot BX^{(2)} + X^{(2)} \cdot BX^{(0)} + X^{(0)} \cdot BX^{(1)} = 0$. Similarly, the estimator on the non-common path will simply be the difference between the observed path energy and common path energy. Define the energy estimator for the non-common paths $X^{(u)}$ (*i.e.,* $u \in \{1, 2\}$)

$$\hat{F}(X^{(u)}) = F(Y^{(u)}) - \hat{F}(X^{(0)}) \quad (6)$$

[1](iii) The covariances amongst distinct terms in (5) are zero (*e.g.,* $\mathrm{Cov}(X^{(1)} \cdot BX^{(2)}, X^{(2)} \cdot BX^{(0)}) = E[X^{(1)}] \cdot B\,\mathbb{E}[(X^{(2)})^T X^{(2)}] \cdot B\,\mathbb{E}[X^{(0)}] - E[X^{(1)}] \cdot B\mathbb{E}[X^{(2)}]E[X^{(2)}] \cdot B\mathbb{E}[X^{(0)}] = 0$ by (i)).

Theorem 2. *$\hat{F}(X^{(u)})$ is unbiased and the variance of the bias over non-common paths estimation equals to* $\mathrm{Var}(\Delta)$.

The wavelet energy on non-common path $\hat{F}(X^{(u)})$ for $u = 1, 2$ can also be estimated by equation (6) with bias $\Delta_u = \hat{F}(X^{(u)}) - F(X^{(u)}) = \Delta - 2X^{(0)}BX^{(u)}$. The non-common path estimation is unbiased as the result of wavelet products of two independent stationary signals approach to zero and $\mathrm{Var}(\Delta_u) = \mathrm{Var}(\Delta)$ (*i.e.,* Theorem 1).

C. Wavelet Energy Estimation

The aforementioned estimator is constructed by the abstract matrix B. We now show how the estimation of wavelet energy is related to the estimator in (3). The energy of a signal Z at scale m is the summation of square of wavelet coefficients $\tilde{Z}_n^m$ that

$$\sigma_m^2(Z) = \sum_n (\tilde{Z}_n^m)^2 = (\tilde{Z}^m)^T \tilde{Z}^m \quad (7)$$

This is in general a *stochastic* quantity. Denote

$$B_{t,t'}^m = \sum_n \phi_{n,t}^m \phi_{n,t'}^m \quad (8)$$

which constructed by the orthonormal DWT matrix of scale m. Then we have $\sigma_m^2(Z) = F_m(Z)$ where F_m is the quadratic form $F_m(Z) = Z^T B^m Z$. Defining $\hat{F}_m$ analogously with the estimator (3) then we have the following

Theorem 3. *For each scale m, $\hat{F}_m(X^{(i)})$ is equal to $\sigma_m^2(X^{(i)})$ for $i = 0, 1, 2$ in expectation.* [2]

We can further bound the variance of the difference $\Delta_m = \hat{F}_m(X^{(0)}) - \sigma_m^2(X^{(0)})$ in terms of the energies of the underlying link by the following Theorem.

Theorem 4. $\mathrm{Var}(\Delta_m) \leq \sum_{i,j\in\{0,1,2\}} \mathbb{E}[\sigma_m^2(X^{(i)})\sigma_m^2(X^{(j)})]$ *$(i \neq j)$. (Upper bound for* $\mathrm{Var}(\Delta_m)$*)* [3]

Thus $\sum_{i,j\in\{0,1,2\}} \sigma_m^2(X^{(i)})\sigma_m^2(X^{(j)})$ $(i \neq j)$ is an upper bound for $\mathrm{Var}(\Delta_m)$ in expectation.

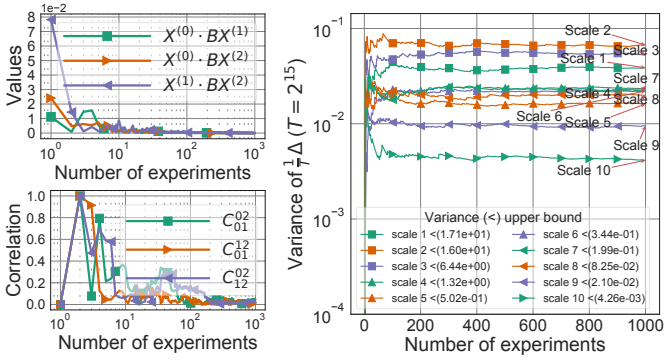
D. Statistical Properties Illustration

1) Binary Network Configuration: We configure a model based on the two-leaf tree topology in figure 1. Let ΔT denote as the time duration for a signal X with $T = 2^M$ $(M \in \mathbb{Z}^+)$ components and let Δf be the frequency increment between adjacent components. Each time series X is constructed by applying Inverse Fourier Transform over a specific power spectral density (PSD) function S_X with random phases θ, such that $X = IFT(\sqrt{S_X \cdot \Delta f} \cdot e^{i\theta}) \cdot T$. Let $X^{(0)}$ be the time series generated by $S_a = (1/(0.1 + (\frac{f}{10})^2))^2$ and $X^{(1)}$, $X^{(2)}$ be the time series generated by $S_b = ((f/10)^2 \cdot (1 + (\frac{f}{10})^2))^2$, which have large energy within low frequency component (S_a) and high frequency component (S_b) respectively. We can also choose any other PSD functions, and our goals is to show how our method recover the original frequency

[2]By equation (3), $\hat{F}_m(Z) = F_m(Z) = \sigma_m^2(Z)$.

[3]$(Cauchy - Schwarz)$, $G_m(Z^{(1)}, Z^{(2)}) = \mathbb{E}[((Z^{(1),m})^T \tilde{Z}^{(2),m})^2] \leq \mathbb{E}[(Z^{(1),m})^T \tilde{Z}^{(1),m}\ (Z^{(2),m})^T \tilde{Z}^{(2),m}] = \mathbb{E}[\sigma_m^2(Z^{(1)})\sigma_m^2(Z^{(2)})]$

information on each path. The length T of time series is set to be 2^{10} and we repeated 1000 experiments. For all experiments, we used *Haar* wavelet as our mother wavelet for simplicity on computing difference of adjacent local average to perform wavelet analysis and all other Daubechies wavelets with orthonormal basis can also be used on constructing matrix B in (8).



(a) Convergence (b) $\mathrm{Var}(\Delta)$ of the common path.

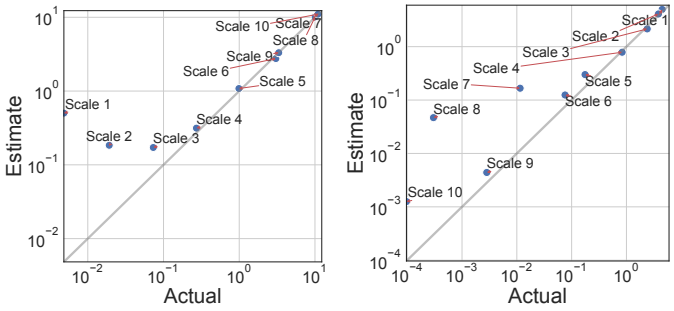
Figure 2: Binary network experiments.

2) Unbiasedness and Variance: Figure 2a illustrates unbiasedness (top figure) and correlation (bottom figure) amongst distinct terms in equation (5) (C_{ij}^{ab} in the figure represents the correlation $\rho(X^{(i)} \cdot BX^{(j)}, X^{(a)} \cdot BX^{(b)})$). The bias is computed by the average of wavelet products $X^{(0)} \cdot BX^{(1)}$ (absolute values normalized by the average of $||X^{(i)}|| * ||X^{(j)}||$). We observe how the bias approaches zero as the number of experiments increases for the normalized wavelet product, which is what we expected from Theorem 1 (ii). The empirical correlation also approaches zero as the number of experiments increase, and hence illustrates the Theorem 1(iii). The results of Figure 2 are for the finest scale $m = 1$, which also bounds the behavior for larger scales whose estimates are linear convex combinations of those from $m = 1$.

Figure 2b shows variance of bias across all scales (equation (5)). As the time series $X^{(i)}$ ($i \in \{0, 1, 2\}$) are generated according to a certain PSD function, the variance of wavelet coefficients on a certain scale m is close to constant. Legend of figure 2b shows the upper bound (*i.e.,* $\sigma_m^2(X^{(0)})\sigma_m^2(X^{(1)}) + \sigma_m^2(X^{(0)})\sigma_m^2(X^{(2)}) + \sigma_m^2(X^{(1)})\sigma_m^2(X^{(2)})$) on each scale and the variance across all scales is smaller than the upper bounds as we proved in Theorem 4.

3) Estimation Accuracy: For the configured binary network, the aggregated path statistics $Y^{(j)}$ are constructed by time series $X^{(i)}$ such that $Y^{(j)} = X^{(0)} + X^{(j)}$ ($j \in 1, 2$).

Figure 3 gives estimated wavelet energy across multiple scales compared with the actual wavelet energy on the common path (*i.e.,* e_0). The estimated values (*i.e.,* y-axis) are very close to the actual values as those points align with the diagonal line (*i.e.,* gray line) well. The theoretical variance σ^2 of the signal $X^{(0)}$ with spectrum S_X is decomposed into



(a) Common path estimates (b) non-common path estimates

Figure 3: Wavelet energy estimation

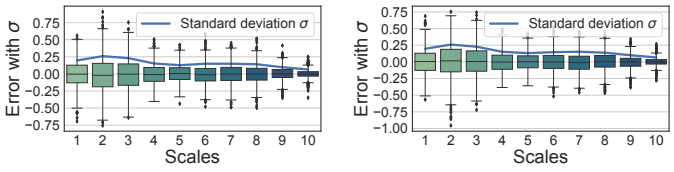
different scales and we have

$$\sigma^2 = \int S_X df = \frac{1}{T} \sum_m ||\tilde{X}_m^{(0)}||^2 \tag{9}$$

where $||\tilde{X}_m^{(0)}||^2$ is the wavelet energy at scale m of $X^{(0)}$. Specifically, the wavelet energy σ_m^2 at scale m is essentially summarizing information in the corresponding spectrum S_X such that

$$\sigma_m^2 \approx 2 \int_{\frac{1}{4m}}^{\frac{1}{2m}} S(f) df \tag{10}$$

[35]. For the common path, as we expected, center scales (*e.g.,* scale 7-10) have more wavelet energy compared to other scales according to the assigned PSD function S_a. Similarly, for the non-common path, large wavelet energy is observed over small scales (*e.g.,* 1-5). Figure 4a shows errors with standard deviation and we can see those error are very small, which also characterizes the variance of Δ in equation (5) at each scale.



(a) Common path (b) non-common path

Figure 4: Errors across all scales

IV. Generalizations

The foregoing work in section III has focused on the canonical two-leaf tree. In this section we outline network generalizations and discuss the effects of nonstationary signals as abnormal traffic would generally break stationary assumption.

A. Network Generalization

We assume a directed path $P_{uv} \subset E$ is specified from each vertex u to v in V such that for each $u \in V$ the edges $\cup_{v \in V \setminus \{u\}} P_{uv}$ form a tree, and likewise $\cup_{u \in V \setminus \{v\}} P_{vu}$.

The key technical result enabling our approach is the following lemma, which says that F is additive in expectation on mutually independent signals.

Lemma 5. *Let X and X' be mutually independent and stationary signals on T. $\mathbb{E}[F(X+X')] = \mathbb{E}[F(X)] + \mathbb{E}[F(X')]$.*

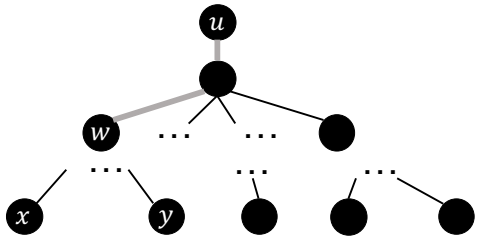


Figure 5: General topology. Grey lines indicate the common path of leaf nodes x and y

Now consider a vertex u and the source tree $T = (V_T, E_T)$ formed by the paths $\{P_{uv} : v \in R\}$ for some $R \subset V$. (Receiver trees can be treated in the same manner.) Without loss of generality we assume that R comprises leaf nodes of this tree; if not we can partition T into trees that have this property. For any interior node $w \in V_T$ let $R_w \subset R$ be the set of leaf nodes descended from w. To any leaf pair $\{x, y\} \subset R_w$ we associated a binary logical tree with edges formed by the subpaths $u \to w$, $w \to x$ and $w \to y$ as illustrated in Figure 5. Thus based the composite signals $Y^{(x)} = \sum_{e \in P_{ux} X_e}$ we form for each scale m the estimator

$$\hat{F}_m^{xy}(Y^{(w)}) = \frac{1}{2}\left(F_m(Y^{(x)} + Y^{(y)}) - F_m(Y^{(x)}) - F_m(Y^{(y)})\right) \tag{11}$$

of $\sigma_m^2(Y^{(w)})$, which is unbiased according to Theorem 1. Convex combinations of estimates from distinct pairs are also unbiased and are expected to reduce variance, for example the average $\hat{F}_m^{\text{avg}}(Y^{(w)}) = (|R_w|(|R_w|-1))^{-1} \sum_{x \neq y \in R_W} F_m^{xy}(Y^{(w)})$; see also [21].

Due to Lemma 5, for any two internal vertices w, w' in the original tree T, we can form an unbiased estimator of the energy $\sigma_m^2(Y^{(w,w')})$ for the total signal $Y^{(w,w')} - \sum_{e \in P_{ww'}} X_e$ associated with the path from w to w' by $\hat{F}_m^{\text{avg}}(Y^{(w)}) - \hat{F}_m^{\text{avg}}(Y^{(w')})$ where we assume w' is closer to the root u of T than w.

Finally, a procedure for fusing a set of source and receiver trees derived from end-to-end measurements on any increasing path additive metric has recently been established in [23], including non-symmetric routing. Applied to the present case, this would enable establishing the virtual network topology that expresses the common contributions of network edges to end-to-end signal energy at any energy scale.

B. Nonstationary Time Series

By Theorem 1, $\mathbb{E}[BX] = 0$ for a stationary signal X, which is invalid for nonstationary signals. We hereby use the figure 1 to discuss when at least one nonstationry signal found in $X^{(i)} (i \in \{0, 1, 2\})$ and they are mutually independent.

Lemma 6. *If no more than one signal is nonstationary, then $\mathbb{E}[\hat{F}(X^{(0)})] = \mathbb{E}[F(X^{(0)})]$.* [4]

[4] $\mathbb{E}[X^{(i)} B X^{(j)}] = 0$, thus $\mathbb{E}[\hat{F}(X^{(0)})] = \mathbb{E}[F(X^{(0)})]$.

Theorems of the unbiased estimator properties derivated in the section III remain valid even if one of the internal link statistics of the two leaf tree model is nonstationary. This can be helpful when identifying the degraded link results from anomalies, hardware failures, *etc.* . For nonstationary signals observed in at least two of the links, we need more constrains to achieve unbiasedness. Let $d^{(i)}$ be the backward differences of time series $X^{(i)}$ such that $d^{(i)}$th order backward difference of $X^{(i)}$ are second order stationary time series with zero mean. Let L denote the even length of Daubechies wavelet filter.

Lemma 7. $\mathbb{E}[\hat{F}(X^{(0)})] = \mathbb{E}[F(X^{(0)})]$ *($L \geq 2 \cdot max\{d^{(i)}\}$).* [5]

Lemma 7 provides a constrain on nonstationary time series to have an unbiased estimator $\hat{F}(X^{(0)})$. With $L \geq 2 * max\{d^{(i)}\}$, we make sure that wavelet coefficients for each collected time series $X^{(i)}$ generated by wavelet filter with length L are second-order stationary time series with zero mean and hence the estimator $\hat{F}(X^{(0)})$ is unbiased.

V. Archipelago RTT measurements

In this section we consider an extension of our approach to tomography from realistic round-trip time (RTT) measurements taken by the globally distributed Ark measurement platform [36] without the assumptions on the independent and stationary of time series. We first described how to obtain the useful RTT data sets from Ark measurements, and then illustrate how the unbiased estimation of end-to-end measurements can be further applied on RTT measurements to locate the link with the largest energy (possibly resulted from heavy daily traffic or degradation) in Section V-B, V-C.

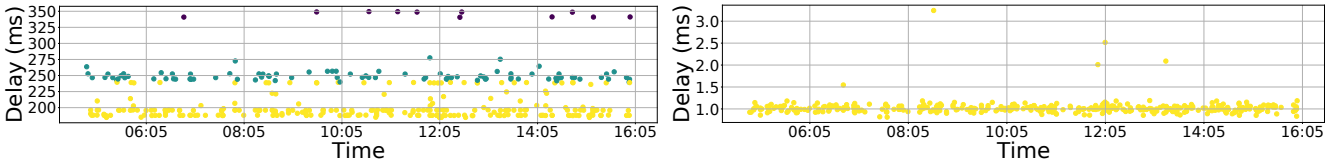
A. Use of the Ark Platform

1) Ark Platform Capabilities: The Ark platform is deployed across 115 cities in 47 countries and it is equipped with several monitoring functions such as topology discovery [37] and congestion detection [38], [39]. For the present paper, we are concerned with Ark's traceroute measurements, launched from monitor nodes to randomly selected addresses within each of the routed /24 networks, of which there are over 10 million. Each monitor node uses Scamper [40] to collect RTT measurements to the destination host and each intermediate hop RTT via ICMP responses at an aggregate rate of 100pps.

2) RTT Data for Subpaths: In this study we use the Ark platform to provide RTT time series for subpaths forming tree subtopologies rooted at a monitoring node. Specifically, given a monitoring node v_0, we form a raw time series of RTT measurements between v_0 and any other node v_i traversed by probe packets from v_0, comprising the RTT values associated all ICMP responses from v_i back to v_0.

3) Derived Time Series and Data Quality: Ark reports probe dispatch times at a 1 second granularity, thus we derive a time series for study by averaging all raw RTT values for measurement from v_0 to a given node v_i associated with each 1 second bin. We wish to provide time series of RRT

[5] If $L \geq 2 \cdot d^{(i)}$, then wavelet coefficients generated by the wavelet filter with length L is a stationary process with zero mean [35].



(a) Link $v_5 - v_7$

(b) Link $v_0 - v_2$

Figure 6: RTT collection over of internal links

measurements for a set of nodes $\{v_i\}$, but we are in practice constrained in the choice of location of v_i relative to v_0 in terms of hop count, since with a greater hop separation, fewer destinations will be reached from v_0 through v_i and hence the sparser the time series. For this reason, we kept the hop count separations as small as possible. After time alignment of the collected time series over derived topology, the length of measurements T is truncated to satisfy the dyadic length $T = 2^m$ ($m \in \mathbb{Z}^+$) for computing wavelet coefficients.

B. Example of A Derived General Topology from Ark

We extracted a topology from Ark measurements at monitor node acc-gh (`196.49.14.12`) as shown in figure 7 with three leaf nodes v_3, v_6, v_7. The monitor node continuously collects ICMP responses from each of the network device, which forms a RTT time series for each path originated from the monitor node. In the network tomography context, we assume that the only observed RTT measurements are from the monitor node to each leaf node (*i.e.,* v_0 to v_j ($j \in \{3, 6, 7\}$)), and our goal is to estimate the wavelet energy of internal links or paths based on those observed path measurements.

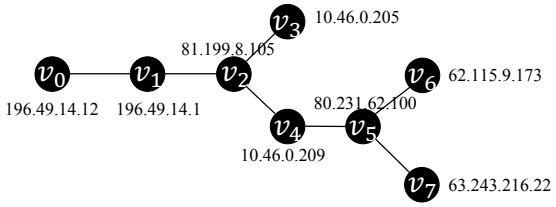


Figure 7: Subtopology from monitor node of acc-gh

The two-leaf tree topology rooted at the monitor node with common path of $v_0 - v_2$ can be formed by leaf node pairs of (v_3, v_6) or (v_3, v_7). Similarly, another two-leaf tree topology with common path of $v_0 - v_5$ can be constructed by (v_6, v_7). The energy estimation of the common path can be performed by selecting any two leaf nodes shared the common path as in equation (11), hence we will use leaf node pairs of (v_3, v_6), (v_6, v_7) for the two common paths estimation. The non-common paths energy estimation follows equation (6). Additionally, the internal path (v_2-v_5) can be estimated by the difference between the wavelet energy of the two common paths.

The evaluation is taken from *2019-01-01* and the results are shown in the figure 8. The wavelet energy is normalized by $1/T$ and RTT measurements unit is *millisecond*. We compared the estimated energy with the actual energy computed from the RTT time series associated with different internal links or paths. Different lines in the figure 8 show the actual energy with error bars indicated the estimation errors across all scales for internal links or paths. The largest energy is observed on link $v_5 - v_7$ with scale 1 contributed the most, which is around 4×10^2. Both the common paths $v_0 - v_2$ and $v_0 - v_5$ have small wavelet energy (<2) across all scales with scale 1 being the scale with the largest energy. The estimation of the common path $v_2 - v_5$, which is the difference between the two common paths estimates, are close to the estimates of $v_0 - v_5$ because we have very small energy on $v_0 - v_2$ (*i.e.,* bottom line of the figure 8). The non-common paths (except $v_5 - v_7$) estimation also have small wavelet energy across all scales with the largest energy at scale 1. Over all the estimation for the internal links based on the aggregated path measurements $v_0 - \{v_3, v_6, v_7\}$, our estimator is able to accurately locate the link with the largest energy (*i.e.,* $v_5 - v_7$).

The time series of RTT measurements of the highest-energy link ($v_5 - v_7$) and the smallest-energy link ($v_0 - v_2$) are shown in the figure 6. The highest-energy link has maximum $150ms$ deviated from the normal delay around $200ms$, while the smallest-energy link has delays varied within $3ms$. The amount of wavelet energy increases as the frequency and magnitude of large link delay increase, and our estimator accurately inferred the wavelet energy of each internal link and identified the link(s) that has the largest energy.

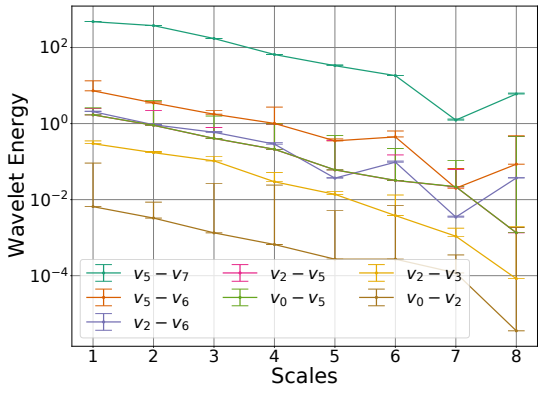


Figure 8: Estimation over the general topology

C. Estimator Performance over the Ark Collections

1) Dataset: We extracted subtopologies from 9 different monitor nodes in different locations across the world and constructed a virtual two-leaf tree topology for each of them with

Table I: Jan. 2019 probing stats. Probings from monitor node v_0 to 2 leaf nodes with average #. valid probings T per cycle.

v_0	Leaf node 1	Leaf node 2	T	#. hops
acc-gh	10.46.0.205	10.46.0.209	15,769	3, 3
cjj-kr	134.75.23.1	134.75.23.9	18,162	2, 2
dtw2-us	12.123.159.50	12.123.159.54	3,346	6, 6
eug-us	10.252.9.13	10.252.10.13	9,157	3, 3
yhu-ca	62.115.134.52	62.115.137.142	7,636	4, 4
pbh2-bt	103.80.109.65	103.80.109.1	2,021	2, 2
sao-br	154.54.11.1	63.223.54.30	391	5, 5
tij-mx	38.140.128.49	69.174.12.97	13,312	3, 3
wbu-us	38.140.128.49	69.174.12.97	15,746	2, 2

selected leaf nodes. For a topology with multiple destinations (*i.e.,* the number of leaf nodes ≥ 3), two or more two-leaf trees (the smallest unit) can be constructed by selecting any two destinations (Section IV). For each selected monitor node, we evaluated 1 month RTT measurements (≥ 1 probing cycle per day) with statistics shown in table I where the number of hops shows how many hops between the monitor node and each leaf node. The estimation for each monitor node is conducted for every probing cycle during this month, and the input data (*i.e.,* known in advance) for the estimator is the end-to-end delay and the corresponding topology.

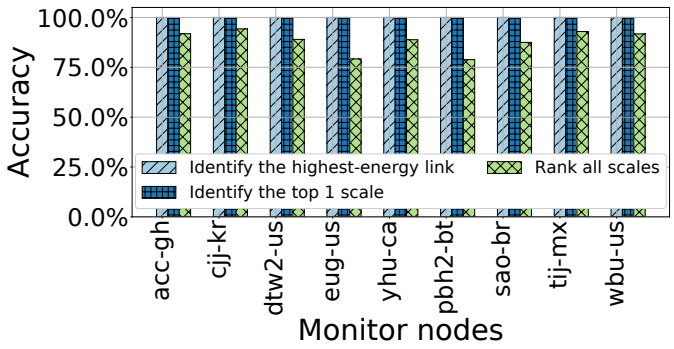


Figure 9: Accuracy

2) Accuracy: Figure 9 shows the accuracy on identifying the highest-energy link and ranking its scales according to their wavelet energy at each monitor node. The accuracy is calculated by the average accuracy over all probing cycles. For each probing cycle, the accuracy of identifying the highest-energy link is either 100% or 0% and the accuracy of ranking scales is $1 - \frac{\#.false}{\#.scales}$. Each of the first bar of different monitor nodes (shown in the x-axis) in the figure 9 shows the accuracy of identifying the highest-energy link and our estimator achieved 100% accuracy over all monitor nodes. Each second bar shows the accuracy of finding the scale with the largest energy (Top 1 scale) and we also achieved 100% accuracy. We observed small errors on ranking the scales with smaller energy as shown in each third bar. The proposed estimator can precisely identify the highest-energy link and scale with small errors over scales that have very small wavelet energy (*e.g.,* $< 10^{-3}$).

3) Mean relative error: Figure 10 shows the mean relative error (MRE) of estimated the wavelet energy across all scales on the link with the largest total wavelet energy, which is the sum of wavelet energy across all scales. The MRE is $\frac{1}{C}\sum_c \frac{1}{M_c}\sum_m \frac{|\hat{F}_m^c(X) - F_m^c(X)|}{F_m^c(X)}$, where $\hat{F}_m^c(X)$ represents the estimates in cth probing cycle and M_c, C are the largest number of scales observed for each probing cycle and the number of probing cycles taken in this month respectively.

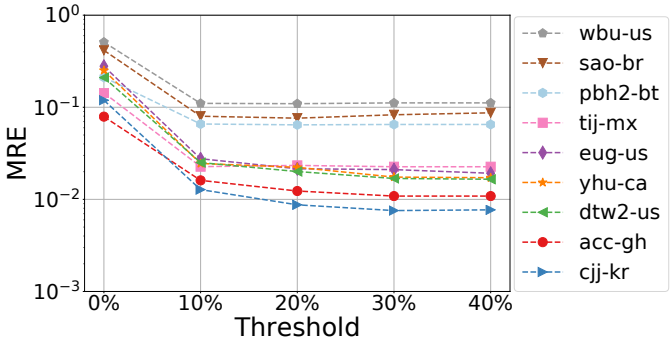
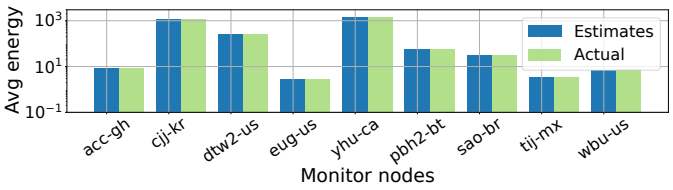


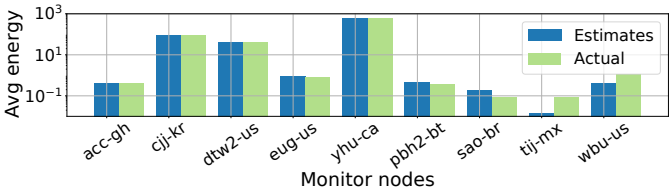
Figure 10: MRE of selected monitor nodes

We applied a filter over the energy at each scale and only calculated MRE of scales with energy larger than a certain percentage of total estimated energy as shown in the x-axis of figure 10. For example, threshold = 10% is to compute MRE of those scales with energy larger than 10% of the total estimated energy of all scales. We stopped at 40% of the total energy as only 1 scale left for all monitor nodes. Large MRE is observed when threshold = 0%, which contains those scales with small energy (*e.g.,* $< 1 \times 10^{-3}$) and is more sensitive to the amount of energy difference between estimates and actual values. On the other hand, as we increase the threshold ($\geq$ 10% of total energy) to eliminate small wavelet energy at some scales, the MREs are small (≤ 0.1) over the selected monitor nodes.

4) Average wavelet energy.: Figure 11 shows the average energy per probing cycle on the link with the highest energy and all other links for both estimates (green) and actual (blue) and Figure 12 shows relative errors. The average energy is calculated by the total normalized wavelet energy ($\frac{1}{T}\hat{F}(X)$) of all probing cycles divided by the total number of probing cycles such that $\frac{1}{C}\sum_c \frac{1}{T}\hat{F}(X)$ where T is length of time series captured at each probing cycle. As we can see from the figure 11 (a) for the link with the highest energy, all estimates are accurate for every monitor node with small relative errors. We aggregate the energy over all other links except the link with the highest energy, and the results are also accurate as shown in figure 11 (b). Large relative errors (~ 1) are observed in cities such as *sao-br* and *tij-mx* but they have small absolute errors as indicated by the bars. Those high wavelet energies (*e.g.,* > 10) estimated over certain links of each monitor node are extracted features for the link(s) experiencing high frequency or magnitude of link delay, which is similar to the illustration in V-B. By using the inference method, we are able to efficiently and accurately locate those links through end-to-end measurements.



(a) Link with the highest energy



(b) All other links

Figure 11: Average energy per probing cycle

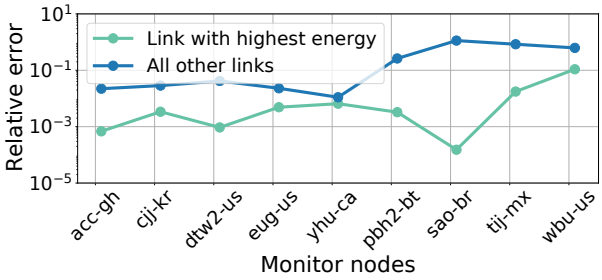
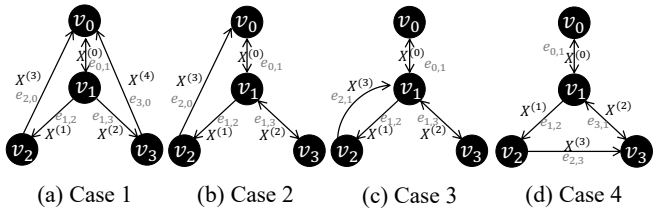


Figure 12: Relative error



(a) Case 1 (b) Case 2 (c) Case 3 (d) Case 4

Figure 13: Asymmetric scenarios.

VI. Discussion

A. Complexity and Estimation Improvement

For a two-leaf tree model, the complexity requires $\mathcal{O}(n)$ (DWT) with n being the length of time series. Assume the network is abstracted by a full binary tree and we collected n end-to-end measurements at each leaf node, then it requires $\mathcal{O}(nm)$ to estimate the energy over each internal link where m is the number of leaf nodes.

Let $Y^{(1)} = X_1^{(0)} + X^{(1)}$, $Y^{(2)} = X_2^{(0)} + X^{(2)}$ be the path aggregation of a simple two leaf tree model. The estimation errors increase if $X_1^{(0)}$ and $X_2^{(0)}$ are uncorrelated as $\mathbb{E}[\hat{F}(X^{(0)})] \approx 0 \neq \mathbb{E}[F(X^{(0)})]$. Therefore, the probing mechanism by enabling dispatch of back to back probes can further improve the estimation accuracy. Previous works used similar assumptions to justify tomography using trains unicast probe packets striped across a set of destinations in order to emulate multicast probes [41], [42].

B. Asymmetric Routing

Besides the symmetric routing in previous Ark evaluation, the estimator can also be applied to asymmetric mechanism. Here we illustrate several cases using the simple tree model in Figure 1 and assume the time series collected on each link satisfied the constrains provided in previous sections. Let $X^{(i)}$ denote the one way delay (OWD) measurements taken on each link and let $Y^{(1)}$ and $Y^{(2)}$ denote the RTT measurements from root node to the two destinations.

Case 1. In Figure 13a, we have $Y^{(1)} = X^{(0)} + X^{(1)} + X^{(3)}$ and $Y^{(2)} = X^{(0)} + X^{(2)} + X^{(4)}$, and

$$\hat{F}(X^{(0)}) = X^{(0)} \cdot BX^{(0)} + \sum_{a\in\{1,3\},b\in\{2,4\}} X^{(a)} \cdot BX^{(b)} \quad (12)$$

The RTT measurements $Y^{(j)}$ ($j \in \{1,2\}$) only contain one way measurements on the common path from v_0 to v_1. According to theorem 1, we have $\mathbb{E}[\sum_{a\in\{1,3\},b\in\{2,4\}} X^{(a)} \cdot BX^{(b)}] = 0$ and $\mathbb{E}[\hat{F}(X^{(0)})] = \mathbb{E}[X^{(0)} \cdot BX^{(0)}] = \mathbb{E}[F(X^{(0)})]$. $\hat{F}(X^{(0)})$ estimates the OWD of the common path from the RTT measurements of Y. Similarly we can obtain the estimator for other cases.

Case 2. We have $Y^{(1)} = X^{(0)} + X^{(1)} + X^{(3)}$ and $Y^{(1)} = 2X^{(0)} + 2X^{(2)}$ if we assume the OWD measurements taken for each direction on a path are the same, and our estimator based on the path observations is

$$\begin{aligned}\hat{F}(X^{(0)}) &= \frac{1}{4}\left(F(Y^{(1)} + Y^{(2)}) - F(Y^{(1)}) - F(Y^{(2)})\right) \\ &= X^{(0)} \cdot BX^{(0)} + \frac{1}{2}\sum_{a\in\{1,3\},b\in\{2\}} X^{(a)} \cdot BX^{(b)}\end{aligned} \quad (13)$$

Case 3. Figure 13c shows when the non-common path has asymmetric routing. This is similar to the symmetric routing but the difference is that the wavelet energy estimation of the non-common path is (*i.e.,* $(X^{(1)} + X^{(3)}) \cdot B(X^{(1)} + X^{(3)})$.

Case 4. We have $Y^{(1)} = 2X^{(0)} + X^{(1)} + X^{(2)} + X^{(3)}$ and $Y^{(2)} = 2X^{(0)} + 2X^{(2)}$. Therefore, our estimator on the common path can be described as

$$\hat{F}(X^{(0)}) = \frac{1}{4}\left(F(Y^{(1)} + Y^{(2)}) - F(Y^{(1)}) - 2F(Y^{(2)})\right) \quad (14)$$

VII. Conclusion

Multiscale analysis using Discrete Wavelet Transforms enabled us to characterize the complex properties of network traffic by extracting the details across frequencies and provided valuable statistics in differentiating the normal patterns and the anomalies. This paper proposed an unbiased estimator on the multiscale energy of internal links based on the tomography of end-to-end measurements to locate the network internal link with the largest energy. We first show how the estimator

of the path intersection is constructed on a canonical two leaf tree model, then we explored how the different PSD would contribute to the wavelet energy over different scales. We evaluated our approach using the realistic RTT measurements collected by CAIDA Ark project to construct the real subtopologies and infer the wavelet energy of internal path from the observed source to destination RTT measurements. We believe the established novel approach can potentially be used to provide comprehensive network internal statistics analysis for detecting network anomalies and congestion control.

REFERENCES

[1] A. Gilbert, "Multiscale analysis and data networks," *Applied and Computational Harmonic Analysis*, vol. 10, pp. 185–202, 05 2001.

[2] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking*, vol. 2, no. 1, pp. 1–15, 1994.

[3] A. Feldmann, A. C. Gilbert, and W. Willinger, "Data networks as cascades: Investigating the multifractal nature of internet wan traffic," in *Proceedings of the ACM SIGCOMM '98 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, ser. SIGCOMM '98, 1998, p. 42–55.

[4] C.-T. Huang, S. Thareja, and Y.-J. Shin, "Wavelet-based real time detection of network traffic anomalies," *2006 Securecomm and Workshops*, pp. 1–7, 2006.

[5] A. Dainotti, A. Pescapè, and G. Ventre, "A cascade architecture for dos attacks detection based on the wavelet transform," *Journal of Computer Security*, vol. 17, pp. 945–968, 2009.

[6] F. H. V. Teles and L. L. Ling, "Adaptive wavelet-based multifractal model applied to the effective bandwidth estimation of network traffic flows," *IET Communications*, vol. 3, pp. 906–919, 2009.

[7] R. Mathew and V. Katkar, "Survey of low rate dos attack detection mechanisms," in *Proceedings of the International Conference & Workshop on Emerging Trends in Technology*, ser. ICWET '11, 2011, p. 955–958.

[8] E. Cambiaso, G. Papaleo, and M. Aiello, "Slowcomm: Design, development and performance evaluation of a new slow dos attack," *Journal of Information Security and Applications*, vol. 35, pp. 23–31, 08 2017.

[9] A. Kuzmanovic and E. W. Knightly, "Low-rate tcp-targeted denial of service attacks: The shrew vs. the mice and elephants," in *Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, ser. SIGCOMM '03, 2003, p. 75–86.

[10] G. Maciá-Fernández, J. E. Díaz-Verdejo, P. García-Teodoro, and F. de Toro-Negro, "Lordas: A low-rate dos attack against application servers," in *Critical Information Infrastructures Security*, J. Lopez and B. M. Hämmerli, Eds. Springer Berlin Heidelberg, 2008, pp. 197–209.

[11] B. Yu, J. Cao, D. Davis, and S. V. Wiel, "Time-varying network tomography: router link data," in *2000 IEEE International Symposium on Information Theory (Cat. No.00CH37060)*, 2000, pp. 79–.

[12] Y. Vardi, "Network tomography: Estimating source-destination traffic intensities from link data," *Journal of the American Statistical Association*, vol. 91, pp. 365–377, 03 1996.

[13] R. Caceres, N. Duffield, J. Horowitz, and D. F. Towsley, "Multicast based inference of network internal loss characteristics," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2462–2480, Nov. 1999.

[14] N. G. Duffield, J. Horowitz, F. L. Presti, and D. Towsley, "Multicast topology inference from measured end-to-end loss," *IEEE Transactions on Information Theory*, vol. 48, no. 1, pp. 26–45, Jan 2002.

[15] N. Duffield, J. Horowitz, F. Lo Presti, and D. Towsley, "Explicit loss inference in multicast tomography," *Information Theory, IEEE Transactions on*, vol. 52, pp. 3852 – 3855, 09 2006.

[16] Yolanda Tsang, M. Coates, and R. D. Nowak, "Network delay tomography," *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2125–2136, 2003.

[17] V. Arya, N. Duffield, and D. Veitch, "Multicast inference of temporal loss characteristics," *Perform. Eval.*, vol. 64, pp. 1169–1180, 10 2007.

[18] V. Arya, N. G. Duffield, and D. Veitch, "Temporal delay tomography," in *IEEE INFOCOM 2008 - The 27th Conference on Computer Communications*, 2008, pp. 276–280.

[19] N. Duffield, "Network tomography of binary network performance characteristics," *Information Theory, IEEE Transactions on*, vol. 52, pp. 5373 – 5388, 01 2007.

[20] V. Arya and D. Veitch, "Sparsity without the complexity: Loss localisation using tree measurements," *CoRR*, vol. abs/1108.1377, 2011.

[21] N. Duffield and F. Lo Presti, "Network tomography from measured end-to-end delay covariance," *Networking, IEEE/ACM Transactions on*, vol. 12, pp. 978– 992, 01 2005.

[22] H. X. Nguyen and P. Thiran, "Network loss inference with second order statistics of end-to-end flows," ser. IMC '07. New York, NY, USA: Association for Computing Machinery, 2007, p. 227–240.

[23] G. Berkolaiko, N. Duffield, M. Ettehad, and K. Manousakis, "Graph reconstruction from path correlation data," *Inverse Problems*, vol. 35, 04 2018.

[24] C. Callegari, M. Pagano, S. Giordano, and T. Pepe, "Combining wavelet analysis and information theory for network anomaly detection," in *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies*, ser. ISABEL '11, 2011.

[25] J. Gao, G. Hu, X. Yao, and R. K. C. Chang, "Anomaly detection of network traffic based on wavelet packet," in *2006 Asia-Pacific Conference on Communications*, 2006, pp. 1–5.

[26] S. S. Kim and A. Reddy, "Statistical techniques for detecting traffic anomalies through packet header data," *Networking, IEEE/ACM Transactions on*, vol. 16, pp. 562 – 575, 07 2008.

[27] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide traffic anomalies," *Computer Communication Review*, vol. 34, 10 2004.

[28] ——, "Mining anomalies using traffic feature distributions," in *Proceedings of the 2005 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, ser. SIGCOMM '05, 2005, p. 217–228.

[29] A. Dainotti, A. Pescapé, and G. Ventre, "Nis04-1: Wavelet-based detection of dos attacks," in *IEEE globecom 2006*. IEEE, 2006, pp. 1–6.

[30] W. Lu and A. Ghorbani, "Network anomaly detection based on wavelet analysis," *EURASIP J. Adv. Sig. Proc.*, vol. 2009, 01 2009.

[31] A. Hussain, J. Heidemann, and C. Papadopoulos, "A framework for classifying denial of service attacks," *Computer Communication Review*, vol. 33, 07 2003.

[32] Z. Wu, R. Hu, and M. Yue, "Flow-oriented detection of low-rate denial of service attacks[j]," *International Journal of Communication Systems*, vol. 29, 04 2014.

[33] Y. Chen, K. Hwang, and Y.-K. Kwok, "Collaborative defense against periodic shrew ddos attacks in frequency domain," *ACM Transactions on Information and System Security*, vol. 3, 01 2005.

[34] M. Yue, L. Liu, Z. Wu, and M. Wang, "Identifying ldos attack traffic based on wavelet energy spectrum and combined neural network," *International Journal of Communication Systems*, vol. 31, 10 2017.

[35] D. Percival, "On estimation of the wavelet variance," *Biometrika*, vol. 82, 01 2012.

[36] The CAIDA UCSD IPv4 Routed /24 Topology Dataset, 2019. [Online]. Available: https://www.caida.org/data/active/ipv4_routed_24_topology_dataset.xml

[37] V. Giotsas, G. Smaragdakis, B. Huffaker, M. Luckie, and k. claffy, "Mapping peering interconnections to a facility," in *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*, vol. 37, 12 2015, pp. 1–13.

[38] M. Luckie, A. Dhamdhere, D. Clark, B. Huffaker, and k. claffy, "Challenges in inferring internet interdomain congestion," in *Proceedings of the 2014 Conference on Internet Measurement Conference*, ser. IMC '14, 2014, p. 15–22.

[39] D. Clark, S. Bauer, W. Lehr, K. Claffy, A. Dhamdhere, B. Huffaker, and M. Luckie, "Measurement and analysis of internet interconnection and congestion," *SSRN Electronic Journal*, 01 2014.

[40] M. Luckie, "Scamper: A scalable and extensible packet prober for active measurement of the internet," in *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '10, 2010, p. 239–245.

[41] M. J. Coates and R. D. Nowak, "Sequential monte carlo inference of internal delays in nonstationary data networks," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 366–376, 2002.

[42] N. Duffield, F. Lo Presti, V. Paxson, and D. Towsley, "Network loss tomography using striped unicast probes," *IEEE/ACM Transactions on Networking*, vol. 14, no. 4, pp. 697–710, 2006.