Measuring the Variability of CAIDA Internet Traffic Traces

Author 1 Author 2 Author 3

***Abstract*—In this paper, we studied the variability of 17 CAIDA Internet traffic traces which were collected in 2013, 2014, 2015 and 2016. The variability of these traces was measured by using the Index of Variability. Based on the results, we outlined several important observations. In particular, the Index of Variability has the ability to reveal significant differences between traffic traces. It is dynamic and its behavior depends on several factors, such as network protocol dynamics and link speeds. In addition, traffic source link speeds have a major impact on network traffic variability (burstiness). Also, results show that there is a significant reduction in the variability for the 2015 and 2016 traces.**

I. Introduction

An extensive volume of research studies have demonstrated that Internet traffic manifests high variability, that is, it is bursty over a wide range of time scales [1]-[6]. High variability in network traffic has been shown to have a significant impact on network performance [2]. Consequently, knowledge of traffic characteristics on multiple time scales can help to improve the efficiency of traffic control mechanisms, and thus improve network performance. Particularly, the design and provision of quality-of-service-guarantees over the Internet requires the understanding of traffic characteristics, such as variability (traffic burstiness).

Most of these studies advocate that Poisson or Markovian based models are no longer appropriate for modelling the packet arrival process, since these models do not capture the traffic burstiness over a wide range of time scales. However, several other studies dispute these findings and argue that the Poisson and Markov based models are still applicable for capturing the performance relevant characteristics of the Internet core traffic [7], [8].

The study presented in this paper is motivated by the following:

- Provide evidence whether Poisson or Markovian based models are appropriate or not for modelling the packet arrival process of Internet traffic.
- Show whether Internet traffic still exhibits high variability over a wide range of time scales that are relevant to network performance.

In this paper, we present the results from measuring and analysing the variability of many CAIDA [9] Internet traffic traces. These CAIDA traces were collected in 2013, 2014, 2015 and 2016. The variability (burstiness) of these traffic traces was measured by using the Index of Variability [1]. The Index of Variability is a rigorous measure of network traffic variability that can capture the degree of burstiness of a typical network traffic process over all times scales. That is, it greatly helps in determining the complexities of traffic variability over network performance relevant time scales.

Based on the results, the following important observations were made:

- The Index of Variability has the ability to reveal significant differences between traffic traces. It is dynamic and its behavior depends on several factors, such as network protocol dynamics and link speeds.
- Source link speeds have a major impact on network traffic variability. A significant increase in source link speed can greatly reduce the burstiness of packet traffic over the network perfomance relevant time scales. As the source link speed increases, the variability is shifted to higher times scales. The 2013 and 2014 traces and subtraces exhibit significant variability over a significant range of times scales, while the ones collected in 2015 and 2016 do not. The most probable reason for this is that prior to collecting the 2015 and 2016 traces, the source link speeds must have been significantly increased.
- Link speed has a major impact on network traffic variability. A significant increase in link speed can greatly reduce the burstiness of packet traffic.
- TCP traffic traces can yield variability curves that exhibit oscillatory behavior.
- Poisson or Markovian based models can not be used to model the complexities of the 2013 and 2014 CAIDA Internet traffic traces.
- Poisson or Markovian based models can be used as analytical models for generating traffic traces similar to the 2015 and 2016 CAIDA Internet traffic traces.

The remainder of this paper is organized as follows: Section II reintroduces the definition of the *Index of Variability* and presents the significant steps for estimating $H_v(\tau)$ from data. Section III provides a short description of the CAIDA Internet traffic traces and presents the empirically obtained results. The paper concludes in Section IV.

II. Index of Variability for Packet Traffic Sequences

This section provides a brief derivation of the Index of Variability. For a detailed derivation of the Index of Variability, see [1].

A. Definition

Let $N(t)$ denote the number of events (packet arrivals) of a stationary point process in the interval $(0, t]$. For each fixed time interval $\tau > 0$, an event count sequence $Y = \{Y_n(\tau), \tau > 0, n = 1, 2, \ldots\}$ can be constructed from each point process, where $Y_n(\tau) = N[n\tau] - N[(n-1)\tau]$ denotes the number of events (packet arrivals) that have occurred during the n^{th} time interval of duration τ. Clearly, Y is also (weakly) stationary for all $\tau > 0$. Y represents a network traffic trace where $Y_n(\tau)$ denotes the number of packets observed from an arbitrary point in the network during the n^{th} time interval of duration τ. τ denotes the *time scale* of the traffic trace and represents the length (i.e., 10ms, 1s, 10s, e.t.c.) of one sample of Y.

The expected number of events that have occurred during the interval $(0, t]$ is always: $E[N(t)] = \frac{t}{E[X]} = \lambda t$ where $E[X]$ is the expected interarrival time and λ is the mean event (packet) arrival rate. The index of dispersion for counts (IDC) is defined as [10]: $IDC(t) \equiv \frac{Var[N(t)]}{E[N(t)]} = \frac{Var[N(t)]}{\lambda t}$. Note that for a Poisson process, $IDC(t) = 1\ \forall t$. Since the point process is stationary, IDC has the same value over any interval of length t; thus, t can be viewed as the time scale τ of the traffic process Y. Henceforth t will be used to denote generality and τ to denote time scales, i.e., the time length of each sample of the packet-count sequence Y.

A distinctive attribute of IDC is that it is mathematically equivalent to the Aggregated Variance [13] method for estimating the Hurst parameter H of a self-similar process. For a self-similar process, plotting $\log(IDC(m\tau))$ against $\log(m)$ results in an asymptotic straight line with slope $2H - 1$. When Y is a long-range dependent (LRD) process, the slowly decaying variance property of LRD processes [2] with parameter $0 < \beta < 1$ is equivalent to an IDC curve[1] with an asymptotic straight line with slope $1 - \beta$, implying $0 < slope < 1$. When the IDC curve converges to an asymptotic straight line with $slope = 0$ for some $\tau < \infty$, then Y is considered to be a short-range dependent (SRD) process. Based on the above property of IDC, the following measure of variability is defined as follows [1]:

Definition 1: For a general stationary traffic process Y whose $IDC(\tau)$ is continuous and differentiable over $(0, \infty)$, we refer to:

$$H_v(\tau) \equiv \frac{\frac{d(\log(IDC(\tau)))}{d(\log(\tau))} + 1}{2} \qquad (1)$$

as the *Index of Variability* of Y for the time scale τ, where $\frac{d(\log(IDC(\tau)))}{d(\log(\tau))}$ is the local slope of the IDC curve at each τ when plotted in log-log coordinates.

Note that the Index of Variability is defined such that in order for an asymptotically or second-order self-similar process $H_v(\tau) \rightarrow H \in (0.5, 1)$ as $\tau \rightarrow \infty$. In case that the process is exactly self-similar, then $H_v(\tau) = H \in (0.5, 1)$ for all $\tau > 0$. That is, if $\log(IDC(\tau))$ is linear with respect to $\log(\tau)$), then $H_v(\tau)$ reduces to H. The Index of Variability can be viewed as the Hurst parameter defined at each time scale τ. In general, the process Y exhibits significant variability for those time scales τ such that $0.5 < H_v(\tau) < 1$. When $\frac{d(\log(IDC(\tau)))}{d(\log(\tau))} \rightarrow 1$, then $H_v(\tau) \rightarrow 1$, implying very high variability or burstiness. Plotting $H_v(\tau)$ versus τ would depict the fluctuation of traffic variability over different time scales for a particular traffic trace. τ.

Expanding the local slope of the IDC curve at each time scale, the following can be obtained:

$$\frac{d(\log(IDC(\tau)))}{d(\log(\tau))} = \frac{\tau}{Var[N(\tau)]} \frac{d(Var[N(\tau)])}{d\tau} - 1. \qquad (2)$$

Using the above in (1), a more convenient form of the Index of Variability is obtained:

$$H_v(\tau) = 0.5\tau \left(\frac{\frac{dVar[N(\tau)]}{d\tau}}{Var[N(\tau)]} \right) = \frac{1}{2} \left\{ 1 + \tau \left(\frac{\frac{d(IDC(\tau))}{d\tau}}{IDC(\tau)} \right) \right\} \qquad (3)$$

In case that Y is Poisson, then $\frac{d(IDC(\tau))}{d\tau} = 0$ for all τ and hence $H_v(\tau) = 0.5$ for all τ.

B. Measuring the Index of Variability from Traffic Traces

The estimation of the Index of Variability from traffic traces requires the computation of the first derivative of the variance curve (i.e., $Var[N(\tau)]$) from discrete samples ($Var[N(\tau_i)]$, $i = 1, \cdots, n$). To accomplish this, an analytic function that best fits the discrete variance data must first be estimated. This in turn requires the use of an interpolation scheme. There exist many such as scheme, which are either based on polynomial-based interpolation methods, or cubic and smoothing splines [11]-[12].

Since we use the sample variances as the estimates of $Var[N(\tau_i)]$, $i = 1, \cdots, n$, these estimates of the variances are considered to be noisy samples. The smoothing spline interpolation methods are known to have optimal properties for estimating continuous functions and their derivatives from a finite number of noisy samples [11], [12]. Note that non-smoothing interpolation methods such as cubic spline produce estimated curves that pass through all the sample points. Therefore, in case of noisy data, non-smoothing interpolation methods yield rough curves, and therefore erroneously high first derivatives.

1) Smoothing Spline Interpolation: For a given data series (x_i, y_i), $i = 1, 2, \cdots, n$, the smooth function $f(x)$ is the solution of the minimization problem

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \xi \int_{x_1}^{x_n} (f^{(k)})^2 \, du, \qquad (4)$$

where ξ is the smoothing parameter and $f^{(k)}$ is the k^{th} derivative of f. For $k = 2$, f is just a cubic smoothing spline.

The first term in (4) is the residual sum of squares, an indicator of the goodness-of-fit of the spline curve to the data. In other words, it measures the degree of fidelity of the smoothing spline function to the data. The second term measures the roughness of the resulting smoothing spline curve. The roughness of a function can be characterized by its curvature. For example, if a function is a straight line, then

[1] In log-log coordinates.

its 2^{nd} derivative (and therefore, roughness) is zero. That is, the second term is a penalty term that measures how close the function is to a straight line.

The smoothing parameter ξ plays an important role. It weights two aspects: smoothness and fit. Large values of ξ give a smoother curve, while small values of ξ result in a closer fit.

2) Steps for Estimating $H_v(\tau)$ from Traffic Traces: In this subsection we briefly describe a practical method for estimating the Index of Variability from traffic traces. This method was first introduced in [1]. Let assume that a traffic trace is the realization of a second-order ergodic point process whose variance curve is continuous and differentiable. $H_v(\tau)$ can be estimated as follows:

(I) Using the *Aggregated Variance* method [13] estimate the variance-time sequence: $\widehat{Var}[N(\tau_i)],\ i = 1, \cdots, n$.

(II) Using an appropriate smoothing spline implementation estimate the smoothing spline $\widetilde{Var}[N(\tau)]$ from $\widehat{Var}[N(\tau_i)],\ i = 1, \cdots, n$.

(III) Using (3) estimate the Index of Variability $\widehat{H_v}(\tau)$.

To estimate the smoothing splines, we used the *MATLAB Spline Toolbox*. Its smoothing spline implementation is based on the Reinsch's approach [11], [12]. This algorithm computes the optimal smoothing parameter ξ such that the penalized residual sum of squares is less than a tolerance value $\epsilon > 0$. In all cases, we used the default value of k (= 2) and $\epsilon = 0.0001$. Note that prior to estimating the Index of Variability for a particular traffic trace, we did not perform any test of differentiability. Rather, we made the restrictive assumption that their variance curves are differentiable. The accuracy and robustness of this procedure was validated in [1] by estimating and matching the Index of Variability curves from test data.

III. CAIDA Internet Traffic Datasets

In this study, using the steps outlined in the previous section, we estimated the Index of Variability curves ($H_v(\tau)$) for 17 empirically collected CAIDA Internet traffic traces [9]. We processed 4 traces from the 2013 *Anonymized Internet Traces* (AIT) dataset, 7 from the 2014 AIT dataset, 4 from the 2015 AIT dataset, and 2 traces from the 2016 AIT dataset. All these traces were collected using the *equinix-chicago* passive monitor [15]. In addition, all these traces were collected when the backbone ISP link was 10GigE.

Each trace is 1-hour long containing the following: Detailed traces files in a compressed pcap format and packet times files that contain nanosecond-precision timestamps. The timestamps in each time file line up exactly with the packets in the corresponding pcap file. Note that the timestamps in each pcap file are truncated to a microsecond precision.

Using the network protocol analyzer *tshark* [16], we also generated *ipv4*, *ipv6*, *tcp*, *udp*, and *http* subtraces from six of the traces. We did that to better understand the impact of these protocols on the variabilty of the aggregated traffic trace. Table I lists the number of packes collected for each of the subtraces, including the percentage to the total number of packets for each aggregated trace. The trace label indicates the date that it was collected and the direction of the packet flow (i.e., dirA indicates the Chicago to Seattle flow and dirB indicates the Seattle to Chicago flow). We also listed the pacjet, bits and flow rates for each of the traces, as well as the link load. With the exception of one trace (20140619-dirB, load: 62.7%), the link was underutilized during the collection of the rest of the traces. More detailed information about these traces can be found in [15], [9].

A. Empirical Results

Figure 1 depicts the estimated Index of Variability curves for all the 17 CAIDA aggregated traces. As shown, 9 of the 2013 and 2014 traces exhibit moderate to high variability (burstiness) over all relevant time scales. Unlike the Hurst parameter, the Index of Variability clearly captures the fluctuating burtiness of each traffic trace at each time scale. Only two of the 2014 traces have very low variability for most time scales, however, it slowly increases at high time scales. Clearly, the 2015 and 2016 packet traces do not exhibit any variability over the times scales under study. That is, $H_v(\tau) \approx 0.5$ for all relavent time scales.

Figure 2 compares the variability of the aggregated traces to the variability of its *ipv4*, *ipv6*, *tcp*, *udp*, and *http* subtraces for the following traffic traces: 20130529-dirA, 20130529-dirB, 20130620-dirA, 20130620-dirB, 20160218-dirA, and 20160218-dirB. The results shown in this figure are surprising, and nowhere else such results were previously reported.

20130529 Traces: The *ipv4* subtraces basically have the same variability curves as with their corresponding aggregated ones. This is expected since the aggregated traces consist mostly of IPv4 packets (99.98% and 99.96%). For the dirA trace, 88% of the packets were generated by the TCP transport protocol. Hence, the variability curve of the *tcp* subtrace is almost the same as with the varability curve of the aggreageted. However, for the dirB trace, the packets generated by TCP is about 68%, and therefore, the variability curve of the *tcp* subtrace is different than the one of the aggregated. The variability curves of the *http* subtraces are almost the same as with the ones of the *tcp* subtraces, since HTTP operates over TCP and, as seen in Table I, the majority of TCP packets were generated by the HTTP application.

Both *ipv6* subtraces depict high fluctuation in their variability curves with an increasing trend. Since the number of IPv6 packets is only a very small fraction of the total number of packets of the aggregated trace, the high variability of the *ipv6* subtraces do not seem to have any impact on the variability of the aggregated traces.

The *udp* subtraces are characterized by almost monotically increasing variability curves. This is due to the nature of the UDP protocol, a bursty packet (datagram) transmission protocol. Comparing the *tcp* varibility curves with the curves of the *udp* subtraces, it becomes evident that the oscillatory behavior of the *tcp* curves is due to the dynamics of TCP.

20130620 Traces: For both directions, the variability curves of the agreggated traces follow closely to the ones of the *tcp* and *http* subtraces, for the same reason as mentioned above.

TABLE I

THE CAIDA INTERNET TRAFFIC TRACES THAT WE ANALYSED, INCLUDING THE NUMBER OF IPV4 AND IPV6 PACKETS INCLUDED IN THESE TRACES. ALSO, THE NUMBER OF TCP, UDP AND HTTP PACKETS IS SHOWN ONLY FOR THE 2013 AND 2016 TRACES. IN ADDITION, THIS TABLE SHOWS SOME IMPARTANT TRAFFIC STATISTICS FOR EACH TRACE.

Trace	**Number of IPv4 Pkts (Fraction)**	**Number of IPv6 Pkts (Fraction)**	**Number of TCP Pkts (Fraction)**	**Number of UDP Pkts (Fraction)**	**Number of HTTP Pkts (Fraction)**	**Pkts/s**	**Bits/s**	**Load**	**Flows/s**
20130529-dirA	1107530928 **(99.98%)**	201139 **(0.0182%)**	979703011 **(88.44%)**	122417898 **(11.05%)**	804682651 **(72.64%)**	297.78k	1.62G	16.3%	8.50k
20130529-dirB	1284234871 **(99.96%)**	456221 **(0.0355%)**	873723503 **(68.01%)**	374057272 **(29.11%)**	695070788 **(54.10%)**	345.35k	2.11G	21.2%	16.40k
20130620-dirA	1179974998 **(99.98%)**	196779 **(0.0167%)**	1063609052 **(90.12%)**	111689230 **(9.46%)**	929718974 **(78.78%)**	317.25k	2.21G	22.2%	7.41k
20130620-dirB	1465553045 **(99.98%)**	299262 **(0.0200%)**	1184436853 **(80.80%)**	263167922 **(17.95%)**	976258761 **(66.60%)**	394.05k	2.79G	28.0%	17.16k
20140320-dirA	1116009669 (99.59%)	4553693 (0.4064%)	- -	- -	- -	301.23k	1.46G	14.6%	8.24k
20140320-dirB	2100185181 (99.96%)	745912 (0.0355%)	- -	- -	- -	564.77k	4.32G	43.4%	18.41k
20140619-dirA	967713972 (99.98%)	200208 (0.0207%)	- -	- -	- -	260.19k	1.62G	16.3%	9.23k
20140619-dirB	2822246142 (99.95%)	1446259 (0.0512%)	- -	- -	- -	759.06k	6.25G	62.7%	28.26k
20140918-dirA	1898079340 (99.84%)	3063416 (0.1611%)	- -	- -	- -	511.06k	3.93G	39.5%	11.01k
20140918-dirB	2026574556 (99.96%)	907872 (0.0448%)	- -	- -	- -	545.02k	4.06G	40.7%	27.29k
20141218-dirA	1586482873 (99.92%)	1330366 (0.0838%)	- -	- -	- -	426.83k	2.82G	28.3%	12.11k
20150219-dirA	1245244103 (99.91%)	1155036 (0.0927%)	- -	- -	- -	328.43k	2.11G	21.2%	8.95k
20150219-dirB	2318481104 (99.65%)	8100606 (0.3482%)	- -	- -	- -	613.07k	4.36G	43.8%	24.95k
20150917-dirA	1109704671 (95.00%)	58427102 (5.0000%)	- -	- -	- -	313.59k	1.78G	17.9%	7.51k
20150917-dirB	1487443672 (98.86%)	17092717 (1.1361%)	- -	- -	- -	403.90k	2.93G	29.4%	9.18k
20160218-dirA	1916892228 **(96.25%)**	74647492 **(3.7482%)**	1775824548 **(89.17%)**	198136019 **(9.95%)**	1487647284 **(74.80%)**	524.97k	3.09G	31.0%	12.55k
20160218-dirB	1653211782 **(94.83%)**	90123691 **(5.1696%)**	1388687750 **(79.66%)**	336057375 **(19.27%)**	1155301323 **(66.27%)**	459.97k	3.25G	32.6%	10.10k

Again, these curves exhibit an oscillatory behavior due to the dynamics of TCP. For the *udp* subtraces, the variability curves have a similar trend as with the curves of the 20130529 traces. Note that, for the dirB traces, the UDP packets make up about 18% of the total traffic trace. For the dirA traces, were not enought IPv6 packets toward to the end of the traces, so we were unable to accurately compute the *ipv6* subtrace variability curve of the same timescale range. Notable is the difference between the *ipv6* variability curves between this dirB subtrace and 20130529 subtraces. It is clear so far that the behavior of the variability curve of a traffic trace, and hence its burstiness, is dynamic.

20160218 Traces: Suprisingly , all the subtraces of these traces on both directions exhibit no variability over the times scales under study ($H_v(\tau) \approx 0.5$ for all relavent time scales). These results, and the ones from Figure 1, lead to the following question: why the 2013 and 2014 traces and subtraces exhibit significant variability and the 2015 and 2016 traces (and their subtraces) do not? What changed between the year 2014 and 2015? Analysing all the information about the traces, we could not find a justifable reason. By further investigation, we found the most probable cause in [17]: the source link speeds have signficantly increased. The study in [17] presents results that shows when the source link speed increases, the range of time scales (lower part of times scales) that the variability is either very low or absent increases towards to higher time scales. Due to this, we strongly believe that the high variability of the 2015 and 2016 traces has shifted to much higher time scales, depending on how much the source link speeds have increased.

IV. Conclusion

In this paper, we analysed the variability (burstiness) of 17 CAIDA Internet traffic traces collected in 2013, 2014, 2015 and 2016. The variability of these traces was obtained by estimating the Index of Variability curves ($H_v(\tau)$), a measure of network traffic variability that has the ability to discern qualitative differences between various traces. The following summarizes the key observations:

- The Index of Variability clearly can capture the fluctuating variability of each traffic trace over at each time scale.
- The Index of Variability has the ability to reveal significant differences between traffic traces. It is dynamic and

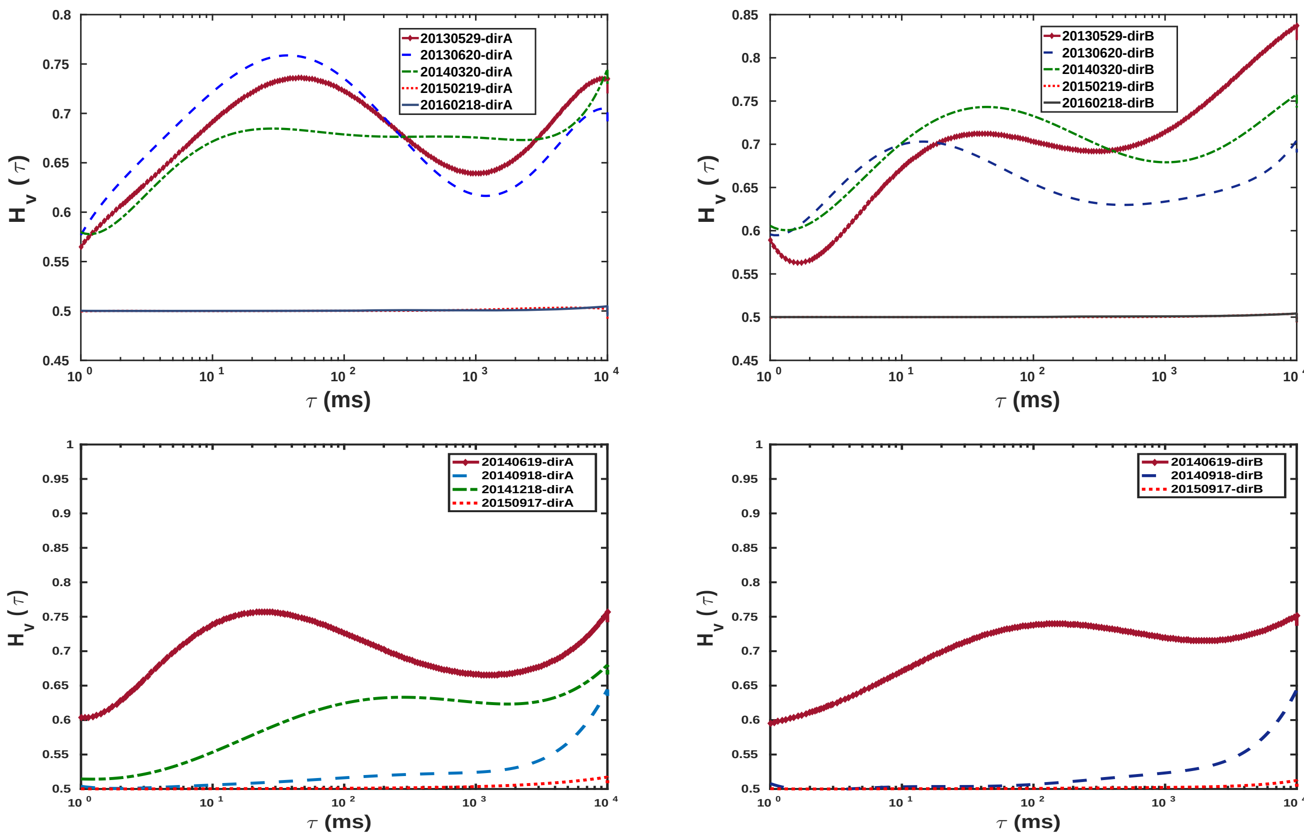

Fig. 1. Each plot shows the estimated Index of Variability (IV) Curves for the traces indicated by the plot legends. The IV curves reveal that the 2013 and 2014 traffic traces exhibit burstiness, however the 2015 and 2016 traces do not.

its behavior depends on several factors, such as network protocol dynamics and link speeds.

- Source link speeds have a major impact on network traffic variability. A significant increase in source link speed can greatly reduce the burstiness of packet traffic over the network perfomance relevant time scales. As the source link speed increases, the variability is shifted to higher times scales.
- TCP traffic traces can yield variability curves that exhibit oscillatory behavior. It was found in [17] that this oscillatory and periodic behavior of the variability curve occurs when the maximum TCP window size is smaller than the delay-bandwidth product.
- The 2013 and 2014 traces and subtraces exhibit significant variability over a significant range of times scales, while the ones collected in 2015 and 2016 do not. The most probable reason for this is that prior to collecting the 2015 and 2016 traces, the source link speeds must have been significantly increased.

Measuring the variability of empirical CAIDA Internet traffic traces using the Index of Variability enabled us to draw the following observation regarding traffic modeling:

- Poisson or Markovian based models can not be used to model the complexities of the 2013 and 2014 CAIDA Internet traffic traces.
- The results indicate that Poisson or Markovian based models can be used as analytical models for generating traffic traces similar to the 2015 and 2016 CAIDA Internet traffic traces. However, it is important to note that these legacy models can only be used to approximate the behavior of network traffic under certain network scenarios in which source link speeds are extremely high.

References

[1] G. Y. Lazarou, J. Baca, V. S. Frost and J. B. Evans, "Describing Network Traffic Using the Index of Variability," *IEEE/ACM Trans. on Networking,* vol. 17, no. 5, pp. 1672–1683, Oct. 2009.

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

[3] G. Terdik and T. Gyires, "Levy Flights and Fractal Modeling of Internet Traffic," *IEEE/ACM Trans. on Networking,* vol. 17, no. 1, pp. 120–129, Feb. 2009.

[4] H. J. Fowler and W. Leland, "Local area network traffic characteristics, with implication for network congestion management," *IEEE JSAC,* vol. 9, no. 7, pp. 1139–1149, Sept. 1991.

[5] V. Petkov, R. Rajagopal and K. Obraczka, "Characterizing Per-Application Network Traffic Using Entropy," *ACM Transactions on Modeling and Computer Simulation,* vol. 23, no. 2, article 14, May 2013.

[6] P. R. M. Inacio, M. M. Freire, M. Pereira and P. P Monteiro, "Fast Synthesis of Persistent Fractional Brownian Motion." *ACM Transactions on Modeling and Computer Simulation,* vol. 22, no. 2, article 11, March 2012.

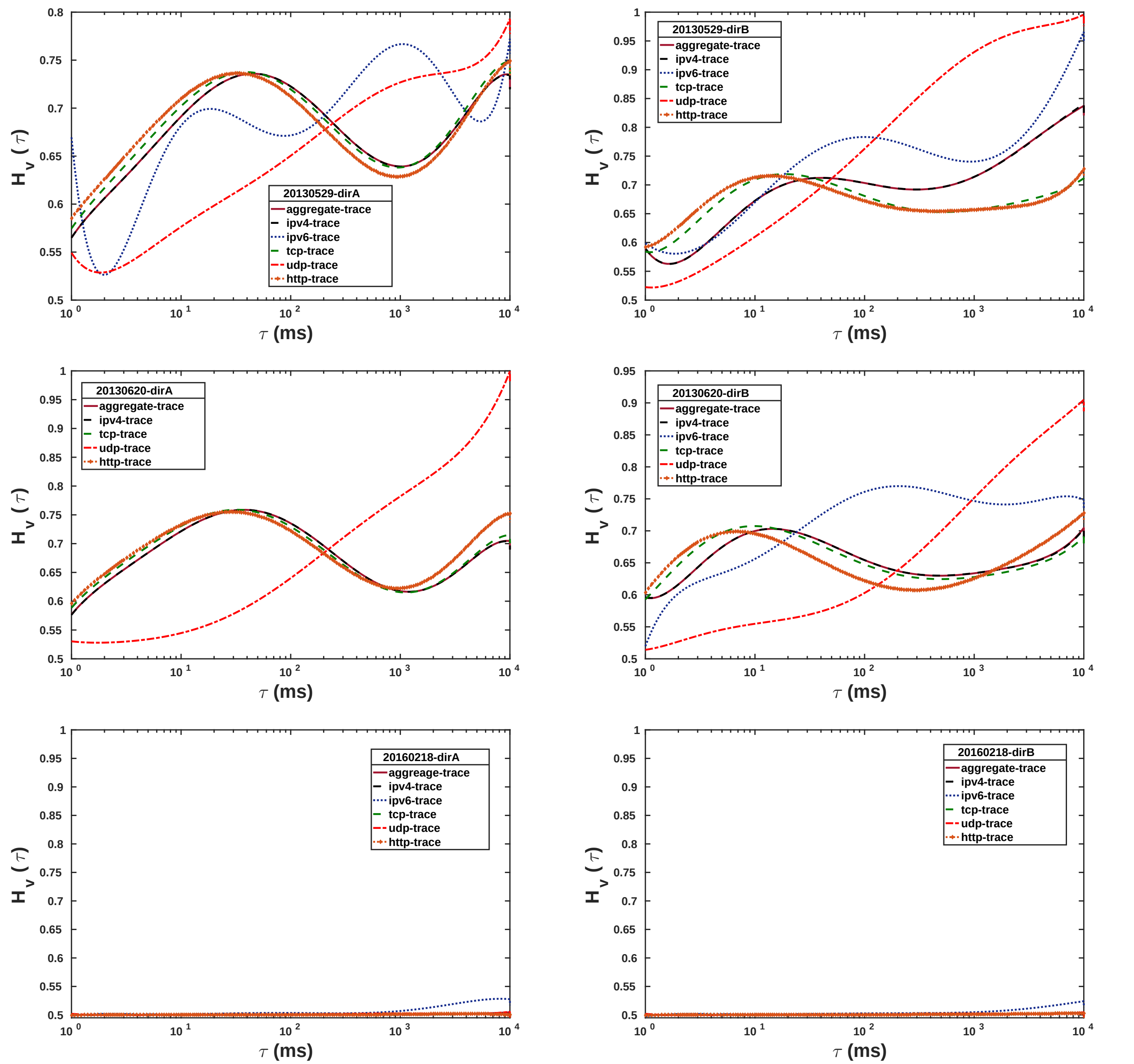

Fig. 2. Each plot shows the estimated Index of Variability (IV) Curves for the aggregate traces and *ipv4*, *ipv6*, *tcp*, *udp*, and *http* subtraces indicated by the plot legends.

[7] T. Karagiannis, M. Molle and M. Faloutsos, "Long-range dependence ten years of Internet traffic modeling," *IEEE Internet Computing,* vol. 8, no. 5, pp. 57-64, Sept.-Oct. 2004.

[8] S. L. Scott and P. Smyth, "The Markov Modulated Poisson Process and Markov Poisson Cascade with Applications to Web Traffic Modeling," *Bayesian Statistics 7*, J. M. Bernardo, M. J Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West (Eds.), *Oxford University Press*, 2003.

[9] The CAIDA Anonymized Internet Traces Datasets. [Online]. Available: https://data.caida.org/datasets/passive-2015/README-2015.txt.

[10] D. R. Cox and P. A. W. Lewis, *The Statistical Analysis of Series of Events,* London:Methuen, 1966.

[11] C. M. Reinsch, "Smoothing by spline functions," *Numerische Mathematik*, vol. 10, pp. 177–183, 1967.

[12] R. L. Eubank, *Spline Smoothing and Nonparametric Regression,* New York:Marcel Dekker, 1988.

[13] J. Beran, *Statistics for Long-Memory Processes. Monographs on Statistics and Applied Probability*, New York: Chapman and Hall, 1994.

[14] K. P. Murphy, *Machine Learning: A Probabilistic Perspective,* Cambridge, Massachusets: MIT Press, 2012.

[15] CAIDA Passive Monitor: equinix-chicago.[Online]. Available: http://www.caida.org/data/monitors/passive-equinix-chicago.xml.

[16] TShark: network protocol analyzer. [Online]. Available: https://www.wireshark.org/docs/man-pages/tshark.html.

[17] G. Y. Lazarou, "On the Variability of Internet Traffic," Ph.D. Dissertation, Electrical Engineering and Computer Science, University of Kansas, Lawrence, 2000.