2010 18th Annual IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems

A Case for Robust Semi-Experiments

Himanshu Gupta[‡,⋆]
[‡]IBM Research Laboratory
New Delhi, India
higupta8@in.ibm.com

Vinay J. Ribeiro
[⋆]Indian Institute of Technology
New Delhi, India
vinay@cse.iitd.ac.in

Anirban Mahanti
NICTA
Alexandria, NSW, Australia
anirban.mahanti@nicta.com.au

***Abstract*—In this paper, we demonstrate that anomalies in Internet traces can have a significant impact on semi-experiments that are designed to determine the causes of scaling behavior of traffic. A semi-experiment involves artificially modifying a specific aspect of a trace and studying the resulting change in scaling behavior. We demonstrate using MAWI traces that semi-experiments performed without addressing the presence of anomalies give insights that contradict widely accepted theories regarding Internet traffic scaling behavior. For example, a direct semi-experimental analysis seems to suggest that removing large flows does not result in the removal of LRD behavior and that the scaling behavior of MAWI traces is the same before and after the removal of the large flows. This observation hence challenges the well-known hypothesis that the heavy-tailed distribution of flow sizes is the primary factor causing correlation at large time-scales. To mitigate the impact of anomalies, we couple the semi-experiments with a recently proposed sketch-based procedure for robust estimation of scaling behavior. We term these "robust semi-experiments". Our analysis shows that using a robust estimation procedure enables a meaningful semi-experimental analysis and that the conclusions drawn from the robust semi-experiments agree with well-established theories regarding Internet traffic scaling behavior.**

I. Introduction

Over the past two decades, research has revealed that several aspects that pertain to the scaling of traffic are invariant over time. For example, it is known that Internet traffic displays two scaling regimes with a transition occurring in the 100ms to 1s time range [2], [8]. Prior research has shown that when Internet traffic is aggregated to large-time scales ($\geq$1s), traffic appears quite bursty and is well-modeled using *long-range dependent* (LRD) processes [12], [13]. The scaling parameter (H) at large time-scales is typically in the (0.8,1) range which is indicative of highly correlated packet arrivals [2], [12], [13]. Prior work has also shown that the scaling parameter (h) at small time-scales (1-100ms) is typically in the (0.55,0.75) range which is indicative of tiny to moderately correlated packet arrivals [6], [19].

Recently, Borgnat et al. [2] showed that the presence of anomalies can interfere with the identification of scaling behavior in traffic. They found that LRD was not observed using standard estimation techniques in many anomalous MAWI traces [1], thus giving the (incorrect) impression that LRD is not a traffic invariant. They next developed a *sketch-based procedure* to mitigate the effect of anomalies. After mitigating the effect of anomalies in the traces with the help of the proposed *sketch-based procedure*, LRD was again observed using the same estimation techniques. Our recent work [6] further re-emphasized the importance of this sketch-based procedure for robust estimation of scaling behavior by carrying out a longitudinal study of small-time scaling behavior of MAWI traces.

In the past, *semi-experiments* have often been used to determine the causes of scaling behavior of traffic. In a semi-experiment one artificially modifies a specific aspect of a trace and studies the resulting change in scaling behavior. The term was coined by Hohn et al. [8] and is an extension of the idea of block-wise shuffling introduced by Erramilli et al. [5]. Among the numerous insights derived from semi-experiments [3], [6]–[12], [15]–[17], [19], we focus on the following four in this paper:

- **P1:** The process of flow arrivals is not responsible for the biscaling structure observed at the IP level. Further, for modeling purposes, the flow arrivals can be assumed to be Poisson distributed.
- **P2:** The scaling at small time-scales has its origin in the packet patterns within flows; the LRD structure at large time-scales is not influenced by them.
- **P3:** The LRD at large time-scales is driven by the heavy tailed nature of flow durations (or sizes).
- **P4:** Dense flows are the correlation causing factors at small time-scales.

Note that all the semi-experiments conducted in the past gave scant attention to the possible presence of anomalies in traces.

In this paper, using MAWI traces, we demonstrate that anomalies in traffic traces can have a significant impact on inferences drawn from semi-experiments. In fact, semi-experiments for all the above mentioned characteristics (P1-P4) conducted without taking special care of anomalies, lead to inferences that challenge the conventional wisdom regarding Internet traffic scaling behavior and characteristics. To mitigate the impact of anomalies on semi-experiments, we design *robust semi-experiments* by combining the sketch-based procedure proposed by Borgnat et al. [2] with semi-experiments.

Our analysis shows that a semi-experiment runs the risk of drawing two types of incorrect inferences due to the presence of anomalies. We classify such errors as Type A and Type B errors. Type A errors refer to the cases where anomalies interfere with scaling behavior in a different manner before and after the modification of data. In such cases, any comparison of scaling behaviors, before and after the data modification, is meaningless and any insight derived from such comparisons

1526-7539/10 $26.00 © 2010 IEEE
DOI 10.1109/MASCOTS.2010.31

is void. Type B errors refer to the cases where anomalies interfere in a similar manner before and after the modification of data. In such cases any insights derived by comparing scaling behaviors, before and after the modification, may be correct. One, however, definitely gets an incorrect picture of the scaling behavior of the data being analyzed. A robust semi-experiment gets rid of both these types of errors.

The rest of this paper is organized as follows. Section II presents the background information relevant to the paper. It summarizes the wavelet method for estimating the scaling parameter as well as the sketch-based procedure for the robust estimation of scaling parameter. Section III summarizes the traces used in this paper, their anomalous nature and the methodology we use to study the traces. Sections IV, V, VI and VII carry out a semi-experimental and a robust semi-experimental analysis of properties P1, P2, P3 and P4, respectively and compare the obtained results. The paper then concludes with a summary of the key results.

II. Background

In this section we present the technical background that is necessary to understand the work presented in this paper. More details can be found in [2], [6], [12], [18], [19].

A. Long Range Dependence

Let $X(t)$ be a stochastic process defined for $t = (0, 1, 2, ...)$ with mean μ and variance σ^2. $X(t)$ may represent traffic volume measured in packets, bytes or sessions at time instance t. For $X(t)$, the autocorrelation function $r(k)$ is defined as follows:

$$r(k) = E[(X(t) - \mu)(X(t + k) - \mu)]. \tag{1}$$

Process $X(t)$ is called long range dependent (LRD) if its autocorrelation function is non-summable, i.e.,$\sum_{k=-\infty}^{\infty} r(k) = \infty$.

B. Scaling Analysis

The analysis of the scaling behavior of a process helps to characterize the way the process will aggregate over different time intervals. We say a process X scales with a scaling parameter h if

$$Var(X^{(m)}) \approx m^{2h-2} Var(X) \tag{2}$$

where $X^{(m)}$ is the aggregated process of X and is defined as

$$X^{(m)}(k) = \frac{1}{m} \sum_{i=km-m+1}^{i=km} X(i); k = 1, 2, ...\infty. \tag{3}$$

The scaling parameter h measures the strength of the correlations present in the data. A value of 0.5 represents the absence of correlations. For example, a Poisson process displays a scaling parameter value of 0.5. The larger the value of h, the stronger the correlations. We interpret scaling parameter values lying within the ranges [0.5, 0.6), [0.6, 0.7), [0.7, 0.8) and [0.8,1] as tiny, small, moderate and large correlations, respectively. Internet traffic usually displays large correlations at large time-scales and the corresponding scaling parameter hence lies within the range (0.8,1) [2]. We use a wavelet-based method [18] to carry out the scaling analysis. We refer the reader to Veitch et al. [18] for a detailed description of this method. A concise summary of this method follows.

C. Wavelet Estimator of Scaling Parameters

We illustrate how this estimator works using *Haar wavelets*. Consider a reference time-scale T_0 and let $T_j = 2^j T_0$ for $j = 1, 2, ...$ be increasingly coarse time-scales. To form the process at scale j, we partition the trace into consecutive and non-overlapping time intervals of size T_j and count the numbers of bytes (or packets) in these intervals. If t_i^j is the i^{th} time interval at scale $j > 0$ then t_i^j consists of the intervals t_{2i}^{j-1} and t_{2i+1}^{j-1}. Let X_i^j be the amount of traffic in t_i^j, with $X_i^j = X_{2i}^{j-1} + X_{2i+1}^{j-1}$. The *Haar wavelet coefficients* d_i^j at scale j are defined as:

$$d_i^j = 2^{-j/2}(X_{2i}^{j-1} - X_{2i+1}^{j-1}); i = 1, ..., N_j \tag{4}$$

where N_j is the number of wavelet coefficients at scale j. The *energy* ξ_j at scale j is defined as:

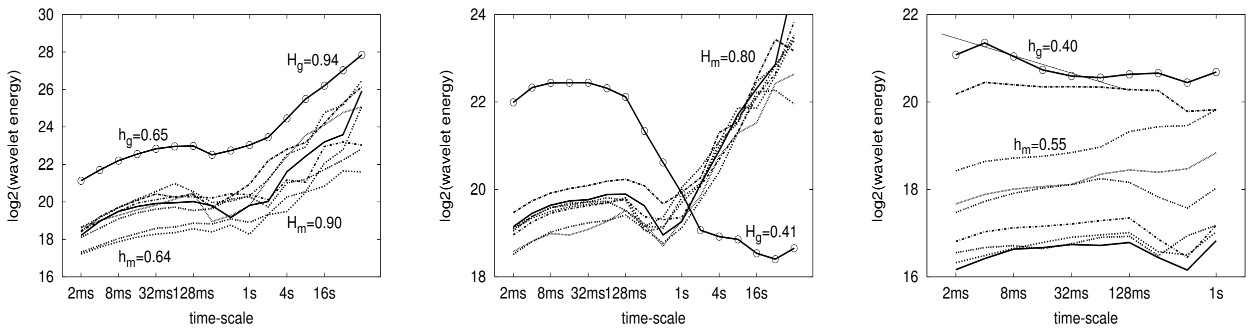
$$\xi_j = E[(d_i^j)^2] \approx \frac{\sum_i (d_i^j)^2}{N_j}. \tag{5}$$

Plotting the logarithm of energy ξ_j as a function of time-scale j gives us a *logscale diagram* (LD). The magnitude of ξ_j increases with the variability of the traffic process X^{j-1} at scale $j - 1$. The variation of energy ξ_j with j captures the scaling behavior of the process. The slope of this energy plot α estimates the *scaling parameter* h through α=2h-1. If the slope of the energy plot is roughly constant over a range of time-scales j to $j + k$, we say that the traffic process exhibits *local scaling* in the time-scales T_j to T_{j+k}.

D. Robust Estimation Technique

Borgnat et al. [2] recently showed that anomalies (e.g., attacks, congestions etc.), if present in traffic traces, can radically affect the observed scaling behavior. By looking at such a trace, one can draw misleading inferences regarding the scaling behavior of Internet traffic. For example, one may incorrectly attribute changes in scaling behavior to certain network mechanisms rather than to the network anomalies. Hence, it is important to mitigate the effect of anomalies so that the impact of network mechanisms on scaling can be precisely studied. This is especially important if our goal is to disentangle smooth long-term evolution features from day-to-day fluctuations. Borgnat et al. [2] developed a sketch-based procedure for mitigating the impact of network anomalies when studying the scaling of network traffic. This robust estimation technique is discussed next.

Let f denote a hash table of size M. We split the original collection of packets into M sub-collections, each of them consisting of all packets with identical sketch output $f(A)$ where the hashing key A is chosen as one of the packet attributes (e.g., Source IP, Destination IP). This amounts to performing random projections, preserving flow structures, since packets belonging to a given flow are assigned to the



(a) Anomaly Free Trace (b) Anomalous trace 1 (c) Anomalous trace 2

Fig. 1. Sketch based Robust Estimation

same sub-collection. A flow is defined as a set of packets with unique 5-tuple (Protocol, Source IP, Destination IP, Source Port, Destination Port) and interarrival times less than 64 seconds. Each sub-collection is aggregated at different time-scales and the scaling parameter computed. We obtain a robust estimate of the scaling parameter by computing the median over the values of scaling parameter estimated from individual sketch outputs [2]. We denote these median estimates of the scaling parameter at large and small time-scales by H_m and h_m, respectively. We use H_g and h_g to represent the scaling parameters obtained at large and small time-scales from the entire trace and term these the *global* estimates.

Statistically, one can obtain a robust estimate of a parameter by averaging estimates from independent copies of equivalent data. Finding equivalent Internet data traces is, however, non-trivial. Random projection using sketches is one crude way to obtain independent copies of equivalent traces. In the presence of anomalies, sketching the original packet stream reduces their impact, possibly restricting them to only some of the sub-collections. The correlation structure of the traffic in the sub-collections containing anomalies will likely differ from normal traffic as well as traffic in other sub-collections. The median of scaling parameter estimates over independent sketches most likely belongs to a sub-collection without anomalies and hence eliminates outliers caused due to anomalies [2].

We next present some representative examples used in prior studies to motivate the importance of robust estimation. Figure 1(a) shows the LD (upper circled plot) for an anomaly-free MAWI trace. This trace was collected on July 11, 2005 and the LD is drawn for direction UStoJp (cf. Section III-B). The anomaly-free nature of this trace was established through careful manual inspection and application of an anomaly identification algorithm [2]. The series byte-count every millisecond is used to construct the LD. The scaling parameter is estimated to be 0.94 and 0.65 at large and small time-scales, respectively. This trace hence clearly exhibits strong correlations at large time-scales and small correlations at small time-scales.

Next, we estimate the scaling parameter using the sketch-based robust estimation method [2]. We hash the trace into 8 parts using destination IP as the hash-key, and estimate the scaling parameter for each subtrace. Then the median of these 8 estimates of scaling parameter give us a robust estimate. Figure 1(a) plots all the subtrace LDs as well. All sketched subtrace LDs are found to be approximately parallel to the original LD. This hence indicates a similar scaling behavior to the original trace, both at small and large time-scales.

Figure 1(b) depicts the LD for a trace collected during a congestion period (June 03, 2003; direction UStoJp). The LD for this trace shows that LRD no longer exists at coarse time-scales. However the distribution of flow packet counts is still heavy-tailed and hence raises doubts regarding the apparent disappearance of LRD [2]. Next we estimate the LRD parameter H using the sketch-based robust estimator. Figure 1(b) also shows LDs for all sketched sub-traces. We find that each sub-trace now shows significant variability resulting in LRD, although the original trace had none. The median value of LRD parameter is 0.8 while the global value is 0.41. This indicates that the trace is characterized by a clear LRD and that the trace showing absence of LRD is an artifact of network anomalies [2].

Figure 1(c) shows the LD (upper circled plot) for another anomalous trace (Oct 11, 2005; JptoUS). This trace is known to contain a low-intensity long-lasting spoofed flooding anomaly [4]. The anomaly consists of source IP addresses being spoofed (source IP is identical to destination IP) and destination port being 0 (which is not normally used) [4]. The global scaling estimate h_g is 0.40 which indicates the presence of a small negative correlation. This trace hence displays a small-time scaling parameter value which deviates from the usual range of (0.5,0.75) [19].

We next estimate the robust (median) value of the scaling parameter. Figure 1(c) also shows the LDs for all the sketched subtraces. Except for one LD, possibly containing an anomaly, all other LDs of subtraces now display tiny positive correlations ($h \approx 0.55$). Note that the scaling behavior of these sketches is markedly different from the scaling behavior shown by the original trace. As a result, the median estimate h_m over 8 sketches, is found to be 0.55 while the global estimate h_g was 0.40. This shows that h_g being 0.40 is an artifact of network anomalies; otherwise all sketched LDs should have had a similar value of scaling parameter [6].

From these two examples it is clear that network anomalies can interfere with the correct identification of scaling behavior and the sketch-based robust estimation procedure is able to mitigate their impact. It should be noted here that an anomaly does not necessarily imply the presence of illegitimate traffic. For example, congestion normally occurs on network links and as shown above (see Figure 1(b)) and can affect the scaling behavior. Hence if one observes an unusual value of scaling parameter, one should double-check that this deviation from normal scaling behavior is not due to any unusual event, e.g., the presence of congestion or illegitimate traffic such as a DDoS attack. Taking a median over the estimates obtained from individual sketch outputs either mitigates or subdues the effect of anomalies present in the data. This robust estimation procedure hence helps in ascertaining whether any observed change in the scaling behavior of Internet traffic is due to a change in any fundamental property of Internet traffic or not.

III. Dataset and Methodology

A. MAWI Dataset

We use publicly available traces from the MAWI repository [1]. This repository provides traces collected from a trans-Pacific backbone. A 15-minute long trace is made public for download every day. We use the traces captured at collection points B and F. The link corresponding to B (100 Mbps) was replaced by F (100 Mbps) in July 2006, and was subsequently upgraded to 150 Mbps in June 2007. Individual directions are referred to as UStoJp and JptoUS.

Overall, the traces exhibit substantial variability during the time period we consider. For example, a significant increase in throughput is observed in 2009 vis-a-vis 2001. Several long lasting congestion periods are also observed (e.g., UStoJp:2003/04 to 2004/10, UStoJp:2005/09 to 2006/06, JptoUS:2005/09 to 2006/06) [2]. Strong fluctuations in packet number are observed on UStoJp from 2004/07 to 2005/04 due to massive activities of the Sasser worm. Many anomalies such as ping floods and SYN scans have also been observed [2], [4]. We refer the reader to [2], [4] for a detailed statistical characterization of traces available from the MAWI repository.

B. Methodology

To showcase the importance of the robust semi-experimental methodology we carry out our analysis on multiple MAWI traces spanning eight years (2001-2009). We select the traces collected on 1^{st} and 15^{th} of every month (if available) from January 2001 to December 2008. This gives us a set of 180 traces spanning eight years. Overall, this suite of 180, 15-minute, packet traces is uniformly spread across eight years, and hence likely captures various variations and anomalies one expects to see in Internet traffic traces. Each trace is partitioned into two subtraces, one for traffic flowing in each direction. As the traffic is asymmetric, i.e. many flows can be observed only in one direction, we analyze each subtrace separately. The objective of using multiple traces spanning eight years is to show that a robust semi-experimental analysis is able to disentangle actual long term evolutions from time-localized events such as illegitimate traffic, congestions etc. whereas a (non-robust) semi-experimental analysis fails to do so. A robust analysis hence increases the confidence in the results obtained.

The rest of this paper follows the following conventions: (a) For all the experiments, the scaling analysis is carried out by constructing the series of byte-count every millisecond and the robust analysis is carried out by sketching every trace into 8 parts by using destination IP as the hashing key. (b) We give a name to every semi-experimental manipulation of the packet arrival process and we add these names as super-scripts to all scaling parameter estimates. For example, H_m^{A-Perm} refers to the median estimate of the scaling parameter after the data manipulation A-Perm. (c) All the longitudinal analysis plots for semi-experimental methodology are titled "SE" while all the plots for robust semi-experimental methodology are titled "Robust SE". Labels for x and y axis for all such plots are to be read as "year" and "scaling parameter", respectively. Captions for these plots indicate the name of semi-experimental manipulation, whether the scaling analysis is carried out for small or large time-scales and the directional information of MAWI traces (JptoUS or UStoJp). (c) Subscripts g and m denote global and median values, respectively. (d) The phrase "correct result" is meant to imply that the result is consistent with widely accepted theories regarding Internet traffic scaling behavior.

IV. Scaling Behavior is Independent of Flow Arrival Process

This section revisits the semi-experimental analysis of property P1 i.e., the process of flow arrivals is not responsible for the biscaling structure at the IP level. This section highlights the differences between the results obtained from semi-experimental analysis vis-a-vis robust semi-experimental analysis. The robust analysis shows that this property is found to be invariant across the entire MAWI trace-suite.

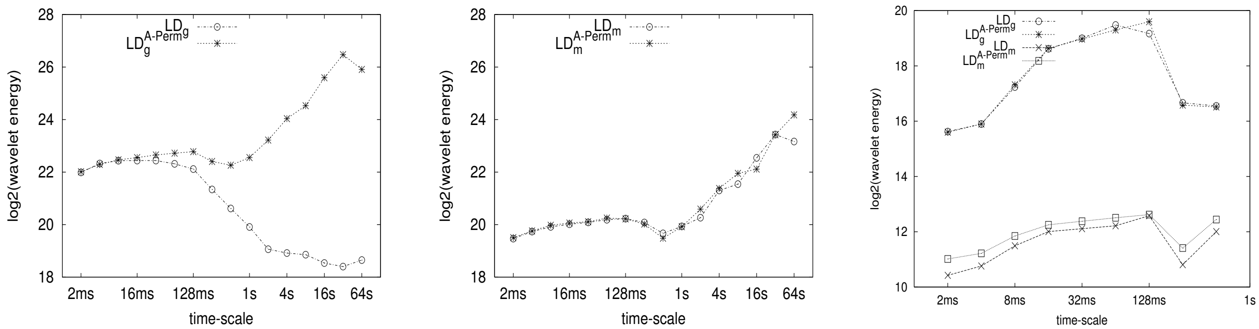
Hohn et al. [8] first established this property and we make use of the same semi-experimental analysis that they employed. Specifically, we modify the packet arrival process in the following two ways.

1) **A**-Perm: The flow arrival process is permuted around the original arrival points and packet structure within each flow is preserved.
2) **A**-Pord: The original flow order is retained. However, the arrival times are re-positioned according to a Poisson process with the same rate.

We study and compare the scaling behaviors of the original and modified process, with and without robust estimation. An identical scaling behavior before and after the manipulation A-Perm implies that the actual order in which the flows arrive does not impact the scaling behavior. An identical behavior before and after the manipulation A-Pord implies that for modeling purposes flow arrivals can be assumed to be Poisson distributed.

A. Representative Examples

Figures 2(a) and 2(b) show a representative example (June 3^{rd}, 2003; UStoJp) motivating the robust semi-experimental



(a) Global LDs: Ex1

(b) Median LDs: Ex1

(c) Global and Median LDs: Ex2

Fig. 2. Semi-experimental vs. Robust Semi-experimental methodology: Representative Examples

methodology. This trace is the same as was used in Figure 1(b). Figure 2(a) shows global LDs drawn from traces before and after the manipulation A-Perm. These LDs show markedly different behaviors. The LD before the manipulation A-Perm (titled LD_g) suggests that LRD has disappeared while a clear LRD can be seen in the LD after the manipulation A-Perm (titled LD_g^{A-Perm}). The LD before the manipulation A-Perm (LD_g) is found to be flat at small time-scales, thereby indicating the presence of uncorrelated small-time scaling behavior. However, the LD after the manipulation A-Perm (LD_g^{A-Perm}) indicates the presence of tiny correlations. A comparison of these two LDs hence suggests that scaling behavior is affected by the flow arrival process. However, as suggested in Section II-D, the scaling behavior can be affected by the presence of anomalies and hence, we need to carry out a robust analysis.

Figure 2(b) shows median LDs (titled LD_m and LD_m^{A-Perm}) drawn from traces, before and after the manipulation A-Perm. Both median LDs show an identical scaling behavior, both at small and large time-scales. LRD can be clearly seen both before and after the manipulation A-Perm. A median based robust analysis hence suggests that scaling behavior is not impacted by flow arrival process. This trace, hence, is an example of a trace displaying type-A errors (cf. Section I) where an anomaly interferes in a different manner before and after the manipulation of the data.

Figure 2(c) plots LDs for a trace (15^{th} July, 2003, JptoUS) showing a type-B error (cf. Section I). LDs are shown only for small time-scales. The upper two plots are global LDs, before and after the manipulation A-Perm, while lower two plots are median LDs. The small-time scaling behavior is identical, before and after the manipulation A-Perm, for both global and median LDs. While global values of small-time scaling parameter (h_g, h_g^{A-Perm}) are found to be 0.84, median values (h_m, h_m^{A-Perm}) are 0.64. Further analysis reveals that only one of the sketch outputs contains a scaling parameter value greater than 0.8, both for LD_m and LD_m^{A-Perm}, thereby indicating that the anomaly is part of this sub-trace. In the absence of a robust estimation procedure, we would have correctly suggested the independence of scaling behavior from the flow arrival process. We would have, however, wrongly predicted the nature of small-time scaling behavior for this trace (strong short-range correlations instead of weak ones). These two examples hence clearly bring out the importance of robust semi-experimental methodology. Next we carry out this analysis for all 180 traces in the MAWI trace-suite.

B. Longitudinal Analysis of MAWI traces for the manipulation A-Perm

Figure 3 compares the results of semi-experimental and robust semi-experimental methodology on small-time scaling behavior for the direction JptoUS. Figure 3(a) displays global values of small-time scaling parameter, before and after the manipulation A-Perm while Figure 3(b) displays median values. The global values h_g and h_g^{A-Perm} do not match for the period 2004 to 2007 thereby suggesting that the flow arrival process affects the small-time scaling behavior of these traces. However the median values h_m and h_m^{A-Perm} are found to be almost identical. The 2004-2007 traces are hence examples of traces showing type A errors. Coupling a semi-experiment with a robust estimation procedure removes the impact of anomalies and thereby enables a meaningful comparison of scaling behavior before and after the manipulation.

Secondly as discussed in Gupta et al. [6], many traces for the period 2004-07 display a scaling parameter h_g less than 0.5, often close to 0.4, which indicates the presence of small negative correlations. Moreover few traces display a scaling parameter h_g close to 0.8 which indicates the presence of large correlations. These observations are in contrast with our current understanding [19] that Internet traffic displays tiny to small correlations at small time-scales with scaling parameter value lying in the range (0.5,0.75). Median values of small-time scaling parameter, however, are found to be in the usual range of (0.5-0.75) for all the traces, thereby clearly indicating the effect of anomalies.

Figure 3 shows that few traces contain large small-time correlations (h_g close to 0.8) and the global values of small-time scaling parameter are found to be identical before and after the manipulation A-Perm. These traces are examples of traces showing type B errors where the anomalies interfere in the same way before and after the manipulation. An identical small-time scaling parameter here correctly suggests that the small-time scaling behavior is independent of flow arrival process, however the value of scaling parameter is overestimated

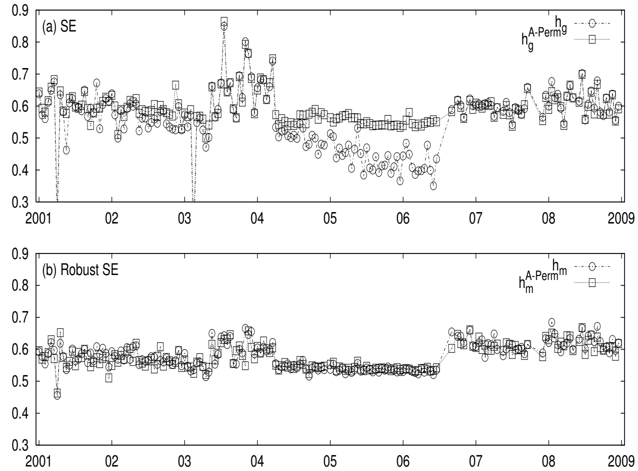


Fig. 3. A-Perm, Small Time-Scales, JptoUS

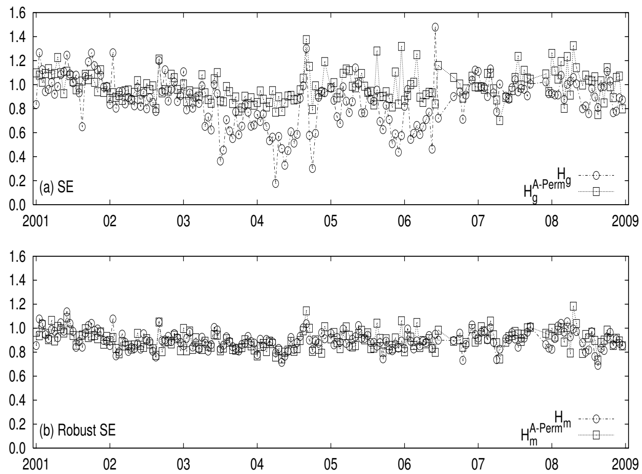


Fig. 4. A-Perm, Large Time-Scales, UStoJp

thereby giving an incorrect inference of small-time scaling behavior. Median values of small-time scaling parameter are found to be in the usual range of (0.5-0.75) as well as identical before and after the manipulation A-Perm. These observations show that a consistent picture across the entire trace-suite is obtained only by the robust semi-experimental methodology.

Figure 4 compares the results for semi-experimental and robust semi-experimental methodologies on large-time scaling behavior for direction UStoJp. Figure 4(a) displays global values of the LRD parameters H_g and H_g^{A-Perm}, before and after the manipulation A-Perm, while Figure 4(b) displays median values. Many traces for period 2003-2007 show a large-time scaling parameter H_g less than 0.6 which indicates the absence of LRD behavior. Borgnat et al. [2] discuss that this is an artifact of network anomalies. A robust analysis shows that these do possess LRD and the median values of scaling parameter H_m are more than 0.8. For these traces, a semi-experimental analysis hence compares scaling parameter values which are not true indicators of large-time scaling behavior of Internet traffic. These traces hence show a different value of global LRD parameter H, before and after the manipulation A-Perm. This hence indicates that flow arrival process affects LRD behavior. However median values of LRD parameter, H_m and H_m^{A-Perm}, are found to be almost identical across the entire 8 years. An absence of robust estimation procedure hence will wrongly suggest that the property P1 is not an invariant. An analysis of the small-time scaling and LRD behavior for directions UStoJp and JptoUS, respectively, shows similar observations. Plots for these two cases are not shown for the lack of space.

C. Longitudinal Analysis of MAWI traces for the manipulation A-Pord

We next analyze the behavior of MAWI traces for the manipulation A-Pord. Hohn et al. [8] showed that Internet traces display identical scaling behavior, before and after the manipulation A-Pord. A-Pord keeps the original order of flows but randomizes flow arrivals according to a Poisson distribution of the same rate as the original arrival process. An identical scaling behavior of Internet traffic, before and after A-Pord, forms the basis of cluster process models where flow arrivals are modeled as Poisson and packets for each flow are modeled using a Gamma distribution [9].

Figure 5 shows the comparison of the semi-experimental and robust semi-experimental methodologies. Here we use scatter plots to show the comparison. Figures 5(a) and 5(b) compare small-time scaling behavior for direction JptoUS. Figures 5(c) and 5(d) compare large-time scaling behavior for direction UStoJp. A comparison of global and median values of scaling parameter before and after the manipulation A-Pord, again highlights the importance of robust analysis.

Figure 5(a) shows that many data points stray from the 45° line. This indicates that for many traces the global values of small-time scaling parameter, h_g and h_g^{A-Pord}, do not match. This hence suggests that property P1 is not an invariant. However robust semi-experiments present a different picture. All the values of h_m and h_m^{A-Pord} across 8 years are found to be nicely distributed along the 45° line (Figure 5(b)) thereby clearly indicating that the scaling behavior is not altered after the manipulation A-Pord and property P1 is indeed an invariant. Moreover all the median values of small-time scaling parameter, before and after the manipulation A-Pord, are found to lie within the usual range of (0.5-0.75) consistent with prior studies [19].

Similar trends are observed in Figures 5(c) and 5(d). Figure 5(c) shows a large dispersion. Many data points deviate from the 45 degree line and hence indicate a significant difference in the values of LRD parameter H_g and H_g^{A-Pord}, before and after the manipulation A-Pord. Global estimates of the LRD parameters, H_g and H_g^{A-Pord}, span a wide range of values. For many traces the value of H_g is less than 0.6 thereby indicating the absence of LRD [2]. However median values of the LRD parameter, H_m and H_m^{A-Pord}, are found

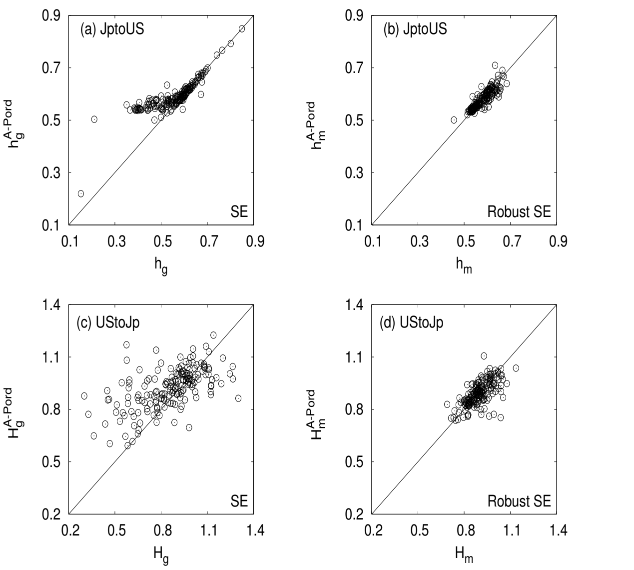


Fig. 5. Semi-experimental vs. Robust Semi-experimental methodology: A-Pord

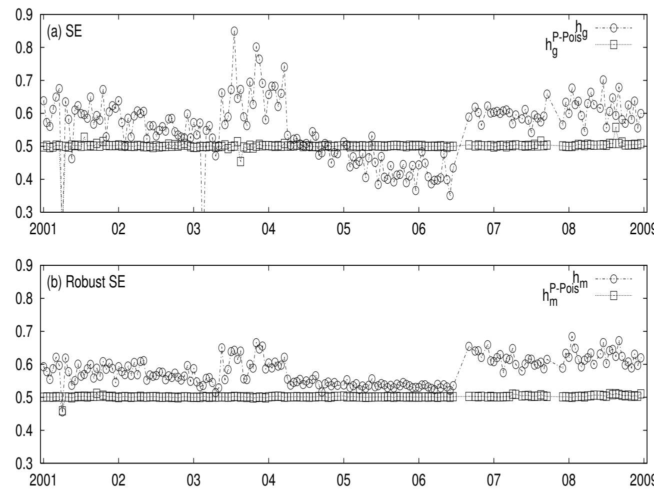


Fig. 6. P-Pois, Small Time-Scales, JptoUS

to be nicely clustered along the 45° line, which clearly shows that the large dispersion in Figure 5(c) is due to the presence of anomalies. The scaling behavior indeed remains identical, before and after the manipulation A-Pord, which in turn shows that property P1 is an invariant.

D. Discussion

Since prior studies did not use robust methods, we may wonder regarding the accuracy of the conclusions drawn therein. Are the traces used in prior studies anomaly-free or are they all instances of traces showing type-B errors? In an informal contribution Ricciato et al. [14] also raise similar concerns. They term the anomalous flows as *pseudo-flows* and wonder whether the prior studies recognized the presence of pseudo-flows and how they handled it. Ricciato et al. further ask whether future work should separate the legitimate flows from pseudo-flows and analyze them independently. Our analysis throws light on some of these concerns.

Our analysis is carried out on a large number of traces spanning eight years. To the best of our knowledge, ours is the first semi-experimental study to do so. Some of these traces are heavily impacted by time-localized events like congestions, illegitimate traffic etc. When one compares the results of a semi-experimental analysis across all traces spanning multiple years (or links), one is likely to see that the conclusions derived are not identical throughout the dataset due to the impact of time/space-localized events on a subset of the traces. As argued by Borgnat et al. [2], a robust estimation procedure disentangles the long term evolutions from time-localized events, and not surprisingly, the conclusions derived by a robust semi-experimental analysis are found to be consistent across the entire data-set. The principal argument made in this paper is that consistent results across the entire MAWI data-set are obtained only by a robust semi-experimental analysis.

As prior studies carried out their analysis only on few traces spanning only 2-3 years, they might not have had the opportunity to study the impact of anomalies in as much detail as we have. Also, even if anomalies were present, they might not have been strong enough to impact the scaling behavior. Dewaele et al. [4] find that almost all MAWI traces contain some sort of anomaly. However, our study finds that anomalies impact the semi-experimental analysis of only a subset of traces. For most of the remaining cases, the semi-experimental analysis without a robust estimation procedure gives expected results.

V. Packet Patterns within Flows Impact Scaling Behavior

In this section we analyze a semi-experiment which modifies the packet structure within individual flows. Specifically, we use the manipulation P-Pois, originally introduced by Hohn et al. [8], which is defined as follows:

- P-Pois: Flow arrival times, flow duration etc. are retained in full. Within each flow, packet arrival times are replaced by a Poisson process of the same rate.

Hohn et al. [8] observed that the effect of randomizing the packet patterns within flows, i.e. manipulation P-Pois, is restricted to small time-scales only. The manipulation P-Pois results in the disappearance of any correlations at small time-scales and the LDs at small time-scales become flat. On the other hand large scale behavior remains unaffected. Based on these observations Hohn et al. [8] concluded that the scaling structure at small time-scales has its origin in the packet patterns within flows and the LRD structure at large time-scales is not influenced by the packet level structure within flows. In this section we revisit these observations in the context of robust analysis.

Figure 6 shows the small-time scaling behavior of MAWI traces, for direction JptoUS, after the manipulation P-Pois. Figure 6(a) compares the global values of small-time scaling parameter (h_g and h_g^{P-Pois}) while Figure 6(b) compares

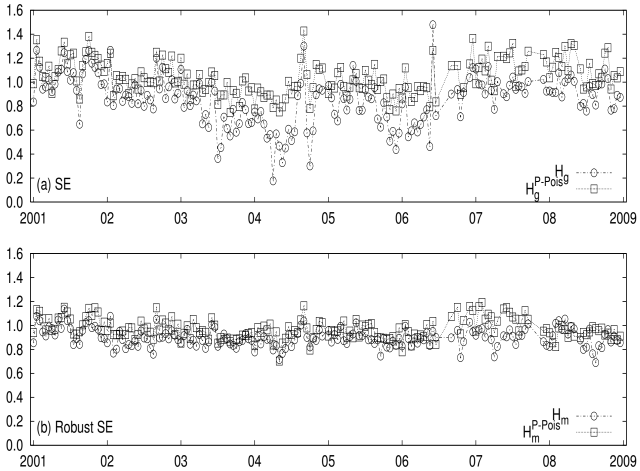


Fig. 7. P-Pois, Large Time-Scales, UStoJp

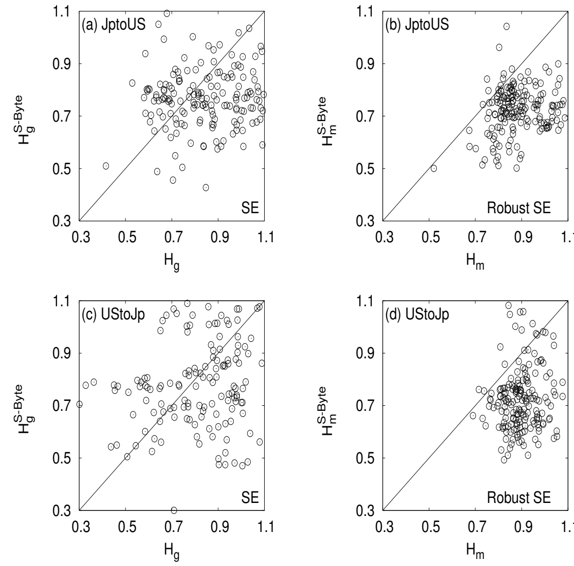


Fig. 8. Semi-experimental vs. Robust Semi-experimental methodology: S-Byte

the median values of small-time scaling parameter (h_m and h_m^{P-Pois}). Figure 6(a) shows that the manipulation P-Pois, results in the disappearance of short range correlations throughout the MAWI trace-suite. The small-time scaling parameter is consistently found to be close to 0.5 thereby validating the observations by Hohn et al. [8].

Figure 6(a) hence shows that a semi-experimental methodology correctly infers that the manipulation P-Pois results in the disappearance of short-range correlations. However the problem with this semi-experimental analysis is that it fails to capture the correct small-time scaling behavior before the manipulation P-Pois. As discussed in Section IV, many traces have negative short range correlations and few traces display large short range correlations. Figure 6(b) presents the overall consistent picture, both internally and when compared with prior studies. It shows that MAWI traces spanning 8 years consistently show tiny to moderate short range correlations with small-time scaling parameter lying within the range (0.5-0.75) and that randomizing the packet arrivals, i.e. the manipulation P-Pois, strips all the traces of any short range correlations they possess.

Figure 7 presents the effect of the manipulation P-Pois on large-time scaling behavior of MAWI traces. Figure 7(a) compares the global values of LRD parameter, H_g and H_g^{P-Pois}, obtained from the semi-experimental analysis and Figure 7(b) compares the median values, H_m and H_m^{P-Pois}, obtained from the robust semi-experiments. Once again a semi-experimental methodology raises similar issues. For many 2003-2007 traces, the LRD parameter does not match before and after the manipulation P-Pois. A deviation for these traces is observed because LRD parameter of these traces does not correctly reflect the large-time scaling behavior of the Internet traffic, as discussed in Section IV and in Borgnat et al. [2]. The LRD parameter for these traces is less than 0.6 thereby implying disappearance of LRD. A semi-experimental methodology hence incorrectly suggests that the LRD behavior of these traces does not remain the same after the manipulation P-Pois.

However as shown in Figure 7(b), robust semi-experiments get rid of these discrepancies. A robust estimation procedure estimates the values of LRD parameter in the range of (0.8-1). LRD behavior, before and after the manipulation P-Pois, is hence consistently found to be same throughout the eight years.

VI. LRD IS DRIVEN BY HEAVY TAILED FLOW SIZES

We next study the manipulation S-Byte, defined as follows:

- S-Byte: All flows above the 70% percentile in terms of byte-counts are removed from the trace.

It is a well-known fact that LRD is caused by the heavy tailed nature of flow sizes (or durations) [3], [8], [12]. This manipulation removes the large flows thereby truncating the heavy tail of distribution of flow sizes. Hohn et al. [8] showed that this manipulation results in the disappearance of LRD behavior.

Figure 8 shows the scaling behavior of MAWI traces, before and after the manipulation S-Byte. We use scatter plots to show the comparison. Figure 8(a) compares the global values of scaling parameter (H_g, H_g^{S-Byte}) while Figure 8(b) compares the median values (H_m, H_m^{S-Byte}). A comparison of Figure 8(a) and 8(b) again sheds light on the importance of a robust estimation methodology.

From Figure 8(a) we do not observe a clear reduction in the scaling parameter values after the removal of the largest 30% flows. For many traces, the scaling parameter values before and after the manipulation S-Byte, lie along the 45° line which indicates that the values H_g and H_g^{S-Byte} are almost the same for these traces. For many traces, the value of scaling parameter H_g^{S-Byte} is more than 0.8 which indicates that these traces show large correlations even after large flows have been removed. Moreover, many data points lie above 45° line which indicates that the strength of correlations at large time-scales has increased after the removal of large flows for the corresponding traces. These observations hence suggest that large flows are not the primary correlations causing factors at large time-scales. We also observe that the global scaling parameter values (H_g) for many traces are less than 0.8, thereby indicating the absence of strong correlations at large time-scales for these traces. Moreover, many traces display a global scaling parameter value less than 0.6 which suggests that LRD is not a ubiquitous phenomenon.

Figure 8(b) compares the median estimates and presents a consistent and expected picture. Most of the scaling parameter values for the original trace (H_m) are between the usual range (0.8,1). Almost all data-points lie below 45° line. This indicates that removing the largest 30% flows, results in a clear decrease of scaling parameter across the entire MAWI trace-suite. Except for few traces, the value of scaling parameter post S-Byte manipulation is less than 0.8 and often less than 0.7. This indicates that the removal of large flows has removed the strong correlations present in the original traces. This semi-experiment hence shows that LRD behavior is driven by the heavy tailed nature of flow sizes and this conclusion, across the entire MAWI trace-suite, is correctly reached by a robust semi-experimental analysis.

Figure 8(c) and 8(d) show the same analysis for direction UStoJp. Once again the global values of the scaling parameter, H_g and H_g^{S-Byte}, are close for many traces , thereby challenging the theory that large flows are the primary correlation causing factors at large time-scales. Median estimates (Figure 8(b)), however, clearly show that the removal of large flows removes the strong correlations and hence affects the LRD behavior. These observations hence clearly show that the presence of anomalies can significantly cloud the insights obtained from a semi-experimental analysis and a robust analysis hence is necessary to mitigate such effects.

VII. Dense Flows drive Small-Time Scaling Behavior

Hohn et al. [8] showed that packet patterns within individual flows influence small-time scaling behavior. Zhang et al. [19] later showed that small-time scaling behavior is driven by the packet patterns within dense flows. This section revisits the semi-experimental methodology carried out by Zhang et al. [19] to show the impact of dense flows on small-time scaling behavior of Internet traffic.

Zhang et al. [19] defined a dense flow as a flow which has at least 50% packet interarrival times less than a certain threshold T. A flow otherwise is called a sparse flow. Intuitively, a dense flow has bursts of densely clustered packet arrivals. Zhang et al. [19] next partitioned the network traces in two components, dense and sparse. The aggregate of dense and sparse flows form the dense and sparse components, respectively. The study then showed that the aggregate of sparse flows has a much smaller scaling parameter at small time-scales, $h \leq 0.6$. This hence shows that dense flows are the correlation causing factors at small time-scales. The semi-experiment of this section involves comparing the scaling behavior of original trace vis-a-vis its sparse component. Specifically, the manipulation studied in this section, *T-Sparse* is defined as follows:

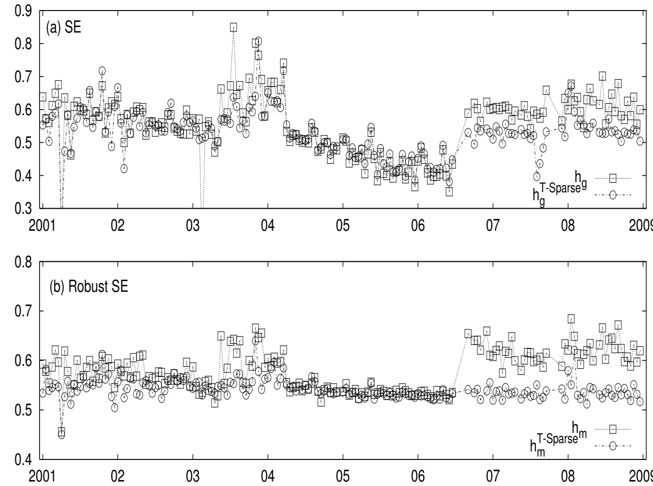


Fig. 9. T-Sparse, Small Time-Scales, JptoUS

1) T-Sparse: All flows with 50% packet interarrival times less than a threshold T are identified and removed from the original trace.

Figure 9 compares the results from the semi-experimental and robust semi-experimental methodologies for the manipulation *T-sparse*. Threshold T is taken to be 2ms. Figure 9(a) plots the values of small-time scaling parameter, before and after the manipulation T-Sparse, across the MAWI trace-suite. There are couple of problems with the results obtained by the semi-experimental methodology. First, as previously seen in Sections IV and V, the nature of small-time scaling behavior is misinterpreted for the 2004-07 traces. Hence a proper comparison of scaling parameter values, before and after the manipulation T-Sparse, cannot be carried out. Second, a close look reveals that for many traces the value of small-time scaling parameter of the sparse component is found to be more than 0.6, with few traces even showing a value more than 0.7. This indicates the presence of small to moderate correlations in the sparse component, which is counter-intuitive. As dense flows have already been removed, this begs the question: what is causing the correlations in the sparse component of these traces? The presence of correlations in the sparse component of these traces hence suggests that dense flows are not the primary correlation causing factors at small time-scales for these traces.

Figure 9(b) compares the median estimate of the small-time scaling parameter. Robust semi-experiments once again present a consistent picture. The scaling parameter h_m for 2004-07 lies in range (0.5-0.75) and the scaling parameter for sparse component $h_m^{T-Sparse}$ is consistently less than 0.6 for all the traces thereby alleviating the doubts raised by the semi-experimental methodology. The 2007-09 traces display small or moderate correlations ($0.6 < h_m < 0.7$) and hence, a

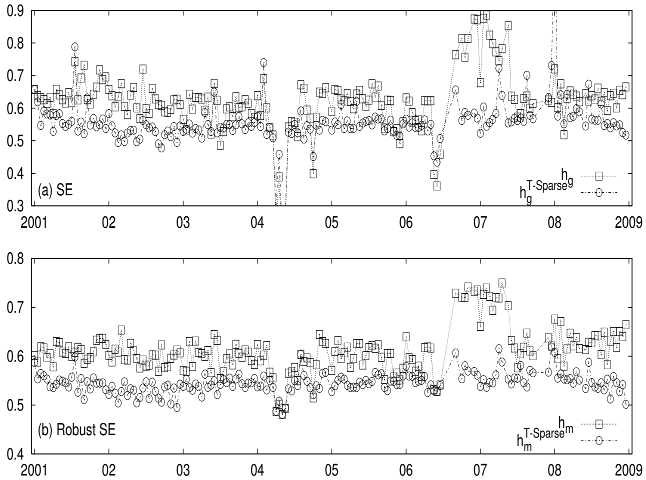


Fig. 10. T-Sparse, Small Time-Scales, UStoJp

clear difference can be seen in the scaling parameter values of original traces and their sparse component. The 2001-03 traces mostly have a scaling parameter close to 0.6 and hence only a small dip is observed in the scaling parameter values ($h_m^{T-Sparse}$) of the sparse components. The 2004-07 traces have a scaling parameter value close to 0.55 which indicates almost uncorrelated scaling and hence no difference is found in the scaling parameter values of the original and sparse components. The results obtained are consistent with the results shown in Section V where a complete randomization of the packet arrivals results in the removal of short range correlations altogether. Both the manipulations, P-Pois and T-Sparse, are different ways of removing bursts of densely clustered packets. The manipulation P-Pois, completely randomizes the packet arrival process for all the flows and hence the resulting trace is found to be uncorrelated.

Figure 10 shows the results of robust analysis for the manipulation T-sparse, for direction UStoJp. The scaling parameter values for sparse component $h_m^{T-Sparse}$ are consistently found to be less than 0.6 throughout the 8 years. As traces in direction UStoJp consistently display small to moderate short range correlations ($0.6<h_m<0.75$), a clear difference between the scaling parameter values of aggregate and sparse component (h_m and $h_m^{T-Sparse}$) is observed across the entire MAWI trace-suite.

There are also few examples of traces showing type B errors. Figure 10 shows that the global scaling parameter values (h_g) for few 2006-08 traces exceed the usual range. Scaling parameter values for these traces are found to be greater than 0.75 which represents large short range correlations. However the scaling parameter values of the sparse component $h_m^{T-Sparse}$ for these traces are found to be less than 0.6 which implies that the removal of dense flows causes a significant decrease in the value of small-time scaling parameter. This semi-experiment hence correctly validates the status of dense flows as the primary correlation causing factors at small time-scales. It, however, incorrectly infers the small-time scaling behavior of these traces.

VIII. Conclusions

This paper presented a case for coupling semi-experiments with a robust estimation procedure. We argued, using traces that span a period of eight years, that in the presence of anomalies, a semi-experimental methodology may result in incorrect inferences regarding Internet traffic scaling behavior. We revisited the semi-experimental analysis of few well-known characteristics of Internet traffic and showed that in the presence of anomalies, the inferences derived from a semi-experimental analysis challenge conventional theories regarding the scaling of Internet traffic. We demonstrated that a robust estimation procedure mitigates the impact of anomalies and hence increases the confidence in the results obtained from semi-experiments.

References

[1] MAWI working group traffic archive. http://tracer.csl.sony.co.jp/mawi.

[2] P. Borgnat, G. Dawaele, K. Fukuda, P. Abry, and K. Cho. Seven years and one day: Sketching the evolution of Internet traffic. In *Proceedings of INFOCOM*, 2009.

[3] M. E. Crovella and A. Bestavros. Self-similarity in World Wide Web traffic: evidence and possible causes. *IEEE/ACM Transaction on Networking*, 5(6), 1997.

[4] G. Dawaele, K. Fukuda, P. Borgnat, P. Abry, and K. Cho. Extracting hidden anomalies using sketch and non-gaussian multiresolution statistical detection procedure. In *Proceedings of SIGCOMM LSAD*, 2007.

[5] A. Erramilli, O. Narayan, and W. Willinger. Experimental queueing analysis with long-range dependent packet traffic. *IEEE Transactions on Networking*, 4(2), 1996.

[6] H. Gupta, V. J. Ribeiro, and A. Mahanti. A longitudinal study of small-time scaling behavior of Internet traffic. In *Proceedings of IFIP Networking*, 2010.

[7] N. Hohn, K. Papagiannaki, and D. Veitch. Capturing router congestion and delay. *IEEE Transactions on Networking*, 17(3), 2009.

[8] N. Hohn, D. Veitch, and P. Abry. Does fractal scaling at the IP level depend on TCP flow arrival processes? In *Proceedings of IMW*, 2002.

[9] N. Hohn, D. Veitch, and P. Abry. Cluster processes: A natural language for network traffic. *IEEE Transactions on Signal Processing*, 51(8), 2003.

[10] N. Hohn, D. Veitch, and T. Ye. Splitting and merging of packet traffic: Measurement and modelling. *Performance Evaluation*, 62(1-4), 2005.

[11] H. Jiang and C. Dovrolis. Source-level IP packet bursts: Causes and effects. In *Proceedings of IMC*, 2003.

[12] W. E. Leland, M. Taqqu, W. Willinger, and D. Wilson. On the self similar nature of Ethernet traffic. *IEEE/ACM Transaction on Networking*, 2(1), 1994.

[13] V. Paxson and S. Floyd. Wide area traffic: the failure of Poisson modelling. *IEEE/ACM Transaction on Networking*, 3(3), 1995.

[14] F. Ricciato. Some remarks to recent papers on traffic analysis. *Computer Commnication Review*, 36(3), 2006.

[15] J. Ridoux, A. Nucci, and D. Veitch. Seeing the difference in IP traffic: Wireless versus wireline. In *Proceedings of INFOCOM*, 2006.

[16] C. Rolland, J. Ridoux, and B. Baynat. Catching IP traffic burstiness with a lightweight generator. In *Proceedings of IFIP Networking*, 2007.

[17] C. Rolland, J. Ridoux, and B. Baynat. LiTGen, a lightweight traffic generator: application to P2P and mail wireless traffic. In *Proceedings of PAM*, 2007.

[18] D. Veitch and P. Abry. A statistical test for time constancy of scaling exponents. *IEEE Transaction on Signal Processing*, 49(10), 2001.

[19] Z. L. Zhang, V. J. Ribeiro, S. Moon, and C. Diot. Small-time scaling behaviors of Internet backbone traffic: An empirical study. In *Proceedings of INFOCOM*, 2003.