The Role of the Weibull Distribution in Internet Traffic Modeling

Muhammad Asad Arfeen, K. Pawlikowski, D. McNickle, A. Willig
University of Canterbury
Christchurch, New Zealand.
muhammad.arfeen@pg.canterbury.ac.nz,{krys.pawlikowski,don.mcnickle,andreas.willig}@canterbury.ac.nz

***Abstract*— This paper highlights the important role played by the two parameter Weibull distribution in Internet traffic modeling. Internet traffic structurally consists of sessions, flows and packets; and traverses through different tiers of service providers during its end-to-end journey. Observation of invariant heavy tails in access traffic patterns of individual users has motivated us to investigate traffic transformation/aggregation as it traverses from access to core network. We found that the flexible nature of the Weibull distribution can capture this transformation at inter-arrival level. We also present and justify our hypothesis that given a suitable scale parameter specific to a certain access media or tier, the Weibull shape parameter can be used to zoom in from session to flow and to the packet level inter-arrivals.**

I. Introduction

Identifying a parsimonious structural model for Internet traffic is still an open research problem. The idea behind the *principle of parsimony* is to find invariants and explain their behaviour in as simple and economical way as possible. Dynamics of Internet traffic have made it difficult to formulate a general model which can capture its behaviour and its transformation both structurally (sessions, flows and packets) and physically (aggregation or converged stochastic behaviour at different levels of ISP tiers). Different levels of (finite) superposition or aggregation, media (wired or wireless) provisioning, congestion and ever-increasing scale of applications are some of the reasons which contribute to this difficulty. A general model may not be very accurate but it must have some analytical intuition so that it can be applicable in a variety of conditions. Although data fitting is important part of modeling, it should not be equated with modeling. A general representative model should be based on invariant attributes which can be related mathematically and should work at all the levels or tiers of Internet hierarchy.

There exist several models (packet based) of Internet traffic that assume that it is always Long Range Dependent and ignore the parts of network where the traffic does not exhibit this property. Long Range Dependence (LRD) represents a correlation persistence phenomenon at different (higher) time scales, that can lead to higher packet loss rate than in non-LRD cases. In literature, most of the articles investigating LRD and its causes in Internet are based on data analysis of the traffic traces of university campuses or ISP (Internet Service Provider) access networks. In the Internet, LRD is present up to certain levels or tiers of ISPs, beyond which the traffic cannot be characterized as LRD [1]. Therefore, any attempt to dimension or provision all parts of networks based on such "LRD-only" based characterization of traffic may result in inappropriate provisioning. It should be noted that the theory of LRD captures traffic burstiness and quantifies it using the Hurst parameter ($1/2 < H \leq 1$, with $H = 1$ for the most bursty traffic) but it does not capture timings of burstiness events. There is a need of a simple model which can relate traffic-burstiness transformation at various tiers; and, also can relate the concept of sessions, flows and packets. The latter has been a challenging task as session identification goes with application layer traffic identification which is still an ongoing research area.

This paper presents an important twofold role played by the two parameter Weibull distribution in Internet traffic modeling: First we show that the Weibull distribution captures the transformation of inter-arrival process (packets,flows and sessions) as traffic moves from access to core network; and we further show that given a suitable scale parameter (media or tier specific) the Weibull shape parameter can be used to zoom in from session to flow and to packet level inter-arrivals. We also identify the relevant count models from literature which can be applied to the Weibull inter-arrival data.

II. Related Work

Due to its flexibility in modeling body and tail parts of empirical data, the Weibull distribution has been extensively used in Reliability and Survivability analysis. In Internet traffic analysis, a paper by Feldmann [2] was the first work to report the role of Weibull distribution. In that study, 10 different traffic traces (Ethernet segments of access networks) were analyzed and the Weibull distribution (with varying shape parameters less than 1) was found to best fit the inter-arrival process of TCP connections. It has been reported that 50% of the individual user's HTTP request inter-arrivals have Weibull shape parameters less than 0.65 which implies the presence of heavy tails in individual user's HTTP request inter-arrival process. Also, the much better fit (than HTTP request inter-arrival process of individual users) of Weibull distribution to the combined or aggregated HTTP request inter-arrivals of all users was noted there. Thus, modeling the combined HTTP request inter-arrivals of all users with Weibull distribution (shape parameter < 1) preserves the inherent burstiness of the HTTP request arrival count process in access network's aggregated traffic. UDP flow or packet inter-arrival processes were not discussed in that study.

The role of flows and packets in Internet traffic modeling has been studied in [3]. The authors analyzed the wavelet energy spectrum of flow arrival processes and packet arrival processes and concluded that the flow arrival processes do not have significant impact on the second order properties of the overall packet arrival processes. This suggested that flow and super-flow structures (sessions) can be ignored for packet arrival process modeling. Six years later the same authors reviewed their results on the basis of more extensive analysis in [4]. They asserted that flow and super-flow structures should be taken as main structural entities in modeling real Internet packet level traffic. They developed a simple Cluster Point Process (CPP) based "explorative toy model" in which the number of flows in a session is heavy tailed.

The impact of user application mix on scaling over time scales between seconds and minutes has been studied in [5]. It was found that there exists statistical dependence within user sessions which creates scaling at time scales of seconds to minutes (which are of course pertinent to users). It has been emphasized that session arrival processes contribute more in the observed scaling than the flow arrival processes.

In [6], it has was concluded that the Scaling phenomenon observed at small time scales (sub-seconds) is related to the arrival process of packets with in flows. They highlighted the need for simple renewal process modeling as compared to complex processes such as multifractals to model traffic data.

One of the first studies on the traffic arriving at access networks (wireless hotspot, DSL and Ethernet) was published in [7] and by Crovella et al. in [8]. Both studies confirmed the dominant role of the heavy tailed Pareto distributions in modeling individual end user flows and sessions. The difference in both studies is that [7] does not use finite mixture (convex combination) models of distributions but instead we use a theorem defining simultaneous multiplexing of heavy tailed streams (in a finite context, of course, to be more realistic). In this paper, we extend the results of [7] on the basis of more recent traffic trace analysis and explain the data transformation analytically. It turns out that the two parameter Weibull distribution is very useful in parsimonious modeling of Internet traffic data (sessions, flows and packets) at different tiers of ISP; and, it can also be used to zoom in from session to flow and to packet level characteristics at a certain tier with scale parameter tier-specific.

III. THEORY

Here we briefly describe the terminology and superposition theorems used in this paper.

A. Self-Similarity & Long Range Dependence

Scale invariance (or Scaling) is an important phenomenon which makes a process self-similar. To formulate it, we first define the aggregated process or aggregated time series $X^{(m)}(i)$ of a stochastic process $X(t)$ at aggregation level m,

$$X^{(m)}(i) = \frac{1}{m} \sum_{t=m(i-1)+1}^{mi} X(t) \tag{1}$$

where m is the size of non-overlapping blocks of the time series where averaging is performed and i is the block index. The process $X(t)$ will be self-similar if it satisfies following distributional equality:

$$X \stackrel{d}{=} m^{1-H} X^{(m)} \tag{2}$$

where H is the index of self-similarity (scaling exponent), also known as Hurst parameter which is used to measure strength of correlation in a time series or process. The process $X(t)$ will be second order self-similar or Long Range Dependent if $m^{1-H}X^{(m)}$ has the same variance and auto-correlation as X for all values of aggregation level m. In other words, asymptotic second order self-similar processes (i.e $m \rightarrow \infty$ in Equation 2) are also called Long Range Dependent processes. As both Long Range Dependence and Short Range Dependence co-exist in Internet traffic, it is crucial to focus on the tail behavior of traffic parameters (flow durations, sizes, inter-arrivals etc.) which controls the transition between Long Range Dependence and Short Range Dependence. A both qualitative and quantitative method to assess LRD and its onset is the Log-scale Diagram plot (called as LD plot) which has been described in [9].

B. Heavy Tails and Long Range Dependence

Let $X_1, X_2, ..., X_n, ...$ be a sequence of n i.i.d observations with the cumulative distribution function F which will be heavy-tailed if it satisfies:

$$1 - F(x) = L(x)x^{-\alpha}, \qquad \textit{as } x \rightarrow \infty \tag{3}$$

for some $\alpha > 0$, where $L(x) > 0$ is a slowly varying function at infinity. The tail exponent or index, $\alpha > 0$, controls the rate of decay of F which is slower than exponential decay if F is heavy-tailed. Hence the tail index characterizes the tail behavior and contains useful information about extremes of the distribution. Observation of the change in value of tail index also signifies the change in physical process generating data. As discussed in [10], Hurst parameter (H) and tail index (α) are mutually related as:

$$H = \frac{3 - \alpha}{2} \tag{4}$$

1) Pareto Distribution: The Pareto distribution has the density function defined as:

$$f(x) = \begin{cases} \alpha \frac{x_m^\alpha}{x^{\alpha+1}} & x \geq x_m \\ 0 & x < x_m \end{cases}$$

where x_m is the scale parameter (smallest value) and $\alpha > 0$ is the tail index which controls the rate of decay of the Pareto cumulative distribution function. The Pareto distribution has infinite mean and infinite variance for the tail index range

$0 < \alpha \leq 1$; and, has finite mean and infinite variance for the tail index range $1 < \alpha \leq 2$; and, variance, skewness, ex-kurtosis and other higher moments start becoming finite as the value of tail index increases in integer steps ($\alpha > 2, \alpha > 3$,$\alpha > 4$...). This makes Pareto distribution quite versatile in modeling. The use of infinite variance distribution in modeling has been supported in [11] as : *"some statisticians argue that infinite variance is an inherently slippery property-how can it ever be verified? But then, independence can never be proven in the physical world, either, and few have difficulty accepting its use in modeling"*.

2) Weibull Distribution: The Weibull distribution has the density function defined as:

$$f(x) = \begin{cases} \frac{c}{b}\left(\frac{x}{b}\right)^{c-1} e^{(-x/b)^c} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

where c is shape parameter and b is scale parameter. The Weibull distribution is quite flexible and can take on different forms based on the value of the shape parameter as shown in Figure 1. Exponential and Rayleigh distributions are special cases of Weibull distribution. The Weibull distribution is heavy tailed for shape parameter $c < 1$. Maximum Likelihood Estimators of the shape parameter c and scale parameter b can be obtained as the solution of the following simultaneous equations:

$$c = \frac{n}{\frac{1}{b}\sum_{i=1}^{n} x_i^c \log x_i - \sum_{i=1}^{n} \log x_i} \qquad (5)$$

and,

$$b = \left[\frac{1}{n}\sum_{i=1}^{n} x_i^c\right]^{1/c} \qquad (6)$$

See [12] for parameter estimation methods for Weibull distribution. The Weibull distribution with shape parameter $c < 1$ is heavy tailed (less heavy tailed than Pareto distribution) but has all the moments finite. This makes it a suitable candidate for convergence modeling of both waist and tail parts in a heavy tailed multiplexing environment.

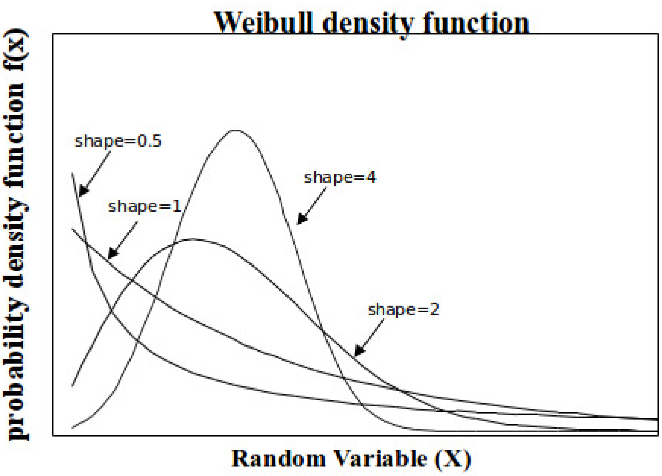

Figure 1: Weibull Distribution probability density function.

C. Superposition theorem : Non-Heavy Tailed Inter-arrivals

Cox's superposition theorem (1954) [13] can be used to describe traffic at high multiplexing levels. He considered the superposition of n independent renewal sequences and proved that as $n \to \infty$, the count process in the superposed sequence tends to Poisson and corresponding inter-arrival distribution tends to exponential. Cinlar (1968) generalized the same result to m-dimensional point process. For finite n, Cinlar and Dudley (1972) showed that the rate of convergence in distribution to Poisson is n^{-1} and if component streams are heterogeneous (i.e with different intensities) then this rate of convergence in distribution is of the order of $\sum_{i=1}^{n}(\frac{\lambda_i}{\lambda})^2$, where λ_i is the intensity of the ith component process of n processes and $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$. In [14], an extended Palm's theorem (proposition 11.2.VI) is presented as: A process obtained by superposition of n independent replicates of a stationary point process, and then by dilating the time scale by a factor k, will converge weakly in distribution to a Poisson process. In [15], it has been argued (empirically) that this applies to a point process even if the inter-arrivals have a heavy tailed distribution. But this applies to very high levels of superposition or multiplexing (i.e. in core or backbone links).

D. Superposition theorem : Heavy Tailed Inter-arrivals

It is analytically difficult to generalize the Poisson approximation result if the renewal sequences taking part in superposition process are finite and have heavy-tailed inter-arrival times. In such a case the superposed process may not even be renewal. Mitov (2006) [16] presented an approximation in the case where component streams have heavy-tailed inter-arrival time distribution with infinite mean and infinite variance, i.e for the tail index range $0 < \alpha \leq 1$. Mitov showed that this superposition converges asymptotically to inter-arrival process with Weibull distribution, with shape (c) and scale (b) parameters given by:

$$c \to 1 - \alpha \qquad (7)$$

$$b \to C \sin \frac{\pi\alpha}{\pi(1-\alpha)} \qquad (8)$$

provided that $nt^{1-\alpha} \to C$, where $0 < C < \infty$.

Unlike [10], [17]–[19], the unique feature of Mitov's approximation is that here $t \to \infty$ and $n \to \infty$ simultaneously. To the best of our knowledge no approximation has been proposed if individual streams or components have finite mean but infinite variance, i.e for the tail index range $1 < \alpha \leq 2$. In Section VI, we provide experimental evidence that, in such a case, the convergence of inter-arrival distribution in superposed stream also appears to be Weibull (shape < 1).

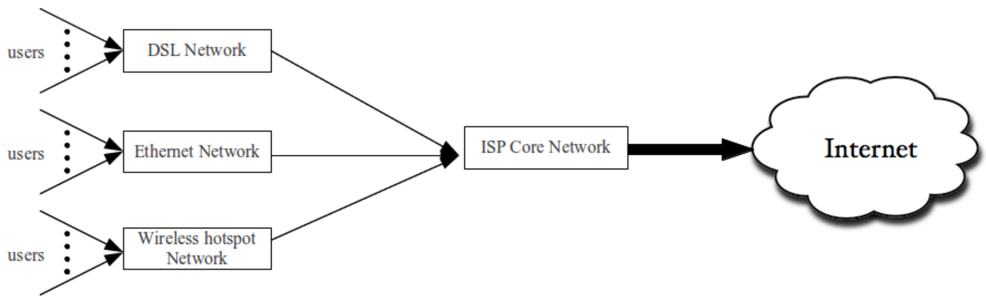

Figure 2: Traffic Capture at DSL, Ethernet, Wireless hotspot network and at their multiplexing point (ISP Core network) of a New Zealand ISP.

IV. Problem Statement and Challenges

We address two research problems regarding parsimonious structural modeling of Internet traffic. Firstly, how to define a simple transformation process which can model the point process behaviour of traffic (session, flows, packet) as it traverses through access tiers to core tiers of Internet hierarchy. Secondly, how the Sessions, Flows and Packet level inter-arrivals can be related and modeled in a simple and unified framework.

Having seen extensive use of Weibull distribution and its flexibility in modeling body and heavy/non-heavy tails, we hypothesize that the Weibull distribution can model traffic transformation as it traverses from access to core tiers; and, the Weibull shape parameter ($0 \rightarrow 1$) can be used to zoom in from session to flow and to packet level statistics. Before presenting our methodology and analysis in the next section, we point to some of the challenges in the analysis of sessions and flows:

- The term Session has been perceived differently among researchers. To some it refers to the whole duration of one login of user to the Internet. For others, it refers to single web page browsing duration. For some others, Sessions are separated by long breaks in individual user's activity.
- Unlike layer 4 flows, Sessions cannot be identified by a 5 tuple (source IP address and port number, destination IP address and port number, transport protocol value). The only exception being the Session having one flow only.
- It is difficult to map sessions quantitatively to flows due to an ever growing application mix. The number of flows in Session are reported to be heavy tailed in [20].
- Application layer traffic classification can definitely help in characterizing user sessions but this is still an unsolved problem.
- Grouping flows belonging to a session is also not easy as the traffic traces contain both Peer to Peer (P2P) and Skype traffic, and for such traffic, the 5-tuple changes (these applications do port hopping in the same flow) even for the same flow and thus a single flow appears as multiple and totally different flows.

In Section VII, we describe how we attempted to overcome the above challenges.

V. Description of data set

The traffic traces of an unnamed New Zealand ISP were captured in 2012 by the WAND Research Group at the University of Waikato. The trace files can be obtained from [21]. The ISP provides wireless hotspot, DSL and Ethernet connectivity in an urban environment. The 24 hour traffic has been captured at Wireless hotspot, DSL and Ethernet access networks on 20th January 2012. All traffic of these three different access networks traverses through the ISP core network and has also been captured there. This enables us to examine multiplexed or aggregated traffic in the core network as well. Figure 2 shows the block diagram of network under consideration. We use 1 hour traffic segments for analysis.

VI. Access to Core traffic transformation

Having observed heavy tails in access patterns (Session and Flow Inter-arrivals) of individual users, here we extend results of our previous study [7] and assess both cases i.e $0 < \alpha \leq 1$ (infinite mean, infinite variance) and $1 < \alpha \leq 2$ (infinite mean, finite variance). We emulate the point process nature of traffic arriving at access network by finite superposition of heavy tailed streams using the D-ITG traffic generator [22] on a 1Gbps Local Area Network. As reported in [7], the tail of flow inter-arrivals lies in the whole range of $0 < \alpha \leq 2$, so we select α values from this range. The different values of heavy tail index ranging from $\alpha = 0.3$ to $\alpha = 1.7$ are used to emulate traffic arriving from different access network subscribers (wireless hotspot, dsl and ethernet users) at the corresponding access network. The point process nature of the superposition experiment makes the results applicable to any heavy tailed component point process be it session, flow or packet arrival process.

A. *Case:* $\mathbf{0 < \alpha \leq 1}$

In Figure 3, the superposition of streams with heavy tailed inter-arrivals (infinite mean, infinite variance) have been shown. The dotted line is the limit line from by Mitov's theorem (see Equation 7). The Weibull Shape parameter estimation method has been described in the Section III-B.2. In the case of α =0.3 and 0.5, the experimental convergence to the expected Weibull Shape parameter value (defined by the Mitov's theorem) is observed even for finite number

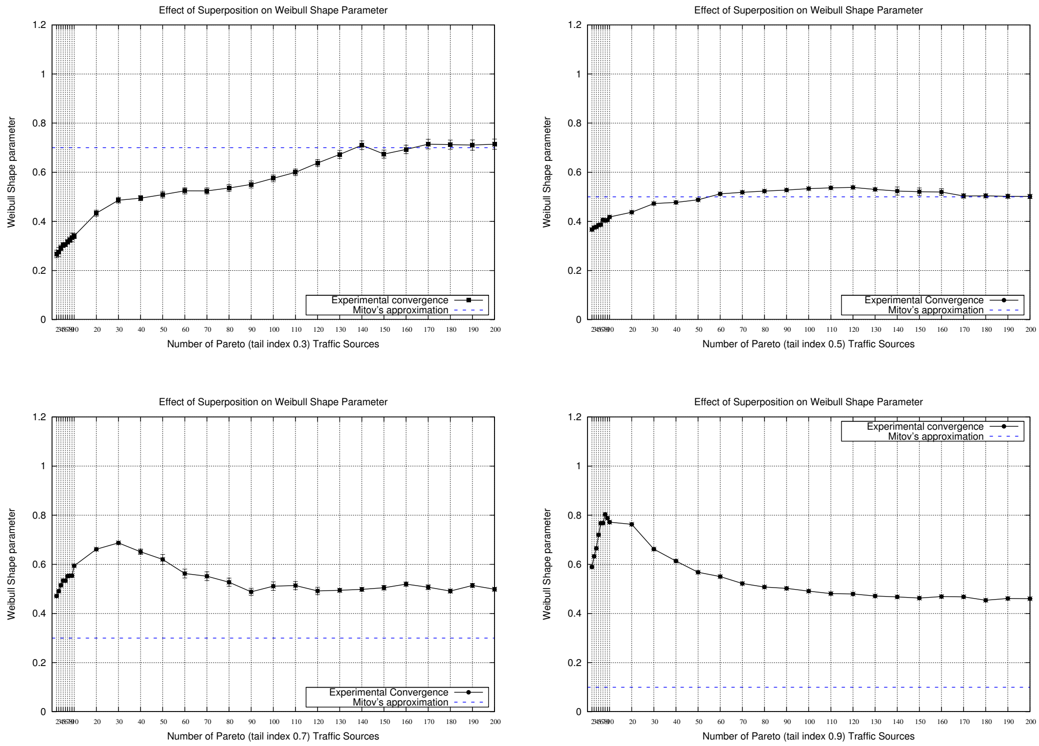

Figure 3: Case $0 < \alpha \leq 1$: Effect of the superposition of streams with Pareto distributed heavy tailed inter-arrivals on Weibull shape parameter. Emulation of traffic arriving at access network.

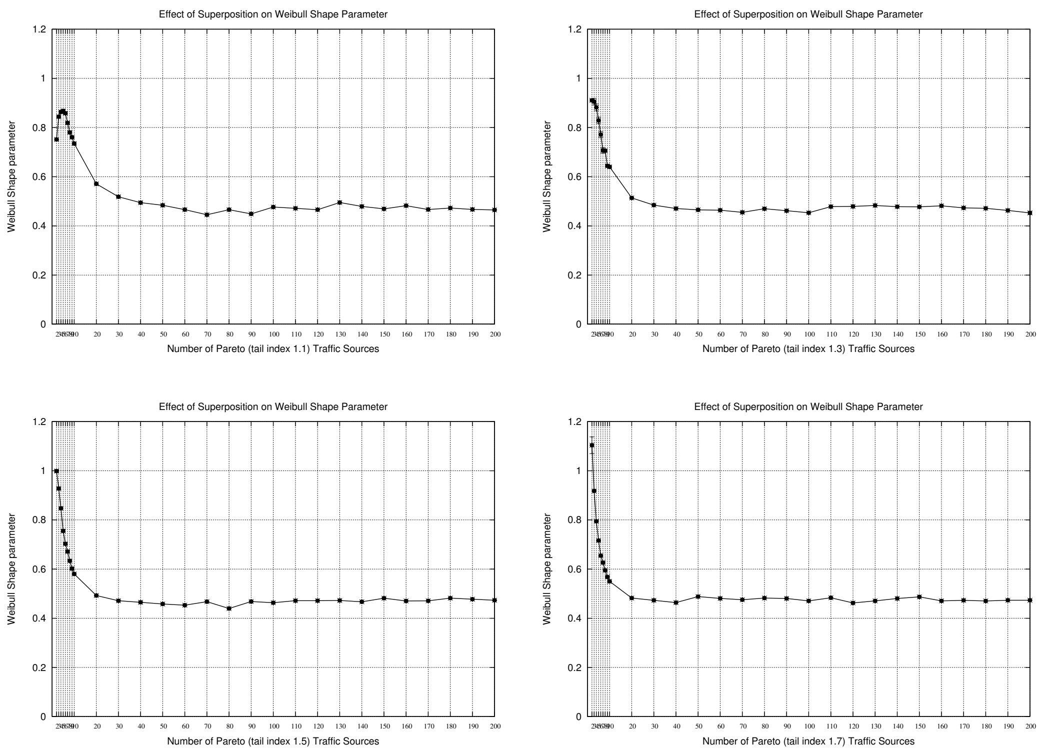

Figure 4: Case $1 < \alpha \leq 2$: Effect of the superposition of streams with Pareto distributed heavy tailed inter-arrivals on Weibull shape parameter. Emulation of traffic arriving at access network.

of sources. For the case α =0.7 and 0.9, the convergence is achieved only in infinite case as limit. In this case, the experimental level of convergence significantly differs from theoretical convergence sense but do becomes parallel to theoretical convergence (asymptotic convergence). Oscillations in final convergence are observed because scale parameter is sinusoidal function of α (See equation 8). A small initial peak is observed which disappears as the value of α increases beyond 1. We believe that this initial anomaly corresponds to distributional transformation phase in which Pareto behaviour changes towards heavy tailed Weibull behaviour. It is noteworthy that for such a superposition there exits no analytical counting process approximation known till now.

B. Case: $1 < \alpha \leq 2$

This range is well investigated in self-similarity literature for heavy tailed ON/OFF traffic models where ON and OFF periods are Pareto distributed with the heavy tail index range $1 < \alpha \leq 2$. Accordingly, the accumulated count process converges to Fractional Brownian Motion with Hurst parameter $H = (3 - \alpha)/2$ (Taqqu's Theorem [10]). The onset of LRD or Scaling according to $H = (3 - \alpha)/2$ is not observed for finite number of sources and time. The applicability of this relation has been investigated in [23] and it has been concluded that "the relation $H = (3 - \alpha)/2$ as predicted by Taqqu's theorem is mostly beyond practical quantitative observability in the current Internet". In [18], the result by Taqqu has been extended for pure renewal processes with heavy tailed inter-arrival time distribution ($1 < \alpha \leq 2$) and a similar FBM approximation has been found, though with different scaling and centering, as:

$$A(t) \approx \frac{mt}{\mu} + \sqrt{\sigma_\beta^2 m} B_H(t) \tag{9}$$

where A_t counts the accumulated number of packets generated by m independent heavy tailed streams ($1 < \alpha \leq 2$) in time t, μ is the mean and σ_β is sample variance. $B_H(t)$ denotes Fractional Brownian Motion (FBM) with index H, $1/2 < H \leq 1$.

The above relation suffers from the same practical observability issue in real Internet as described in [23]; and, also due to non-Gaussian marginal distribution of access patterns of individual users.

Figure 4 shows the superposition of streams with heavy tailed inter-arrivals (finite mean,infinite variance). In literature, there do not exists any analytical result accounting for the convergence of inter-arrivals in this case. Our empirical results show that in such a case, a fast convergence to the Weibull Shape parameter (in the range 0.45 to 0.5) is achieved even in the finite source sense for all values of heavy tail index in the range $1 < \alpha \leq 2$. Finite mean, indeed, does help in the fast convergence in this case.

It should be noted that further superposition of streams with the Weibull distributed inter-arrivals (shape parameter<1) maintains the shape parameter in superposed stream even in the finite source context. This is in accordance with Mitov's approximation as described previously.

C. DISCUSSION

In sections VI-A and VI-B, the access network traffic has been emulated by the superposition of heavy tailed streams ($0 < \alpha \leq 2$) and results are shown in Figures 3 and 4. These results are applicable to any point process, be it a process describing session, flow or packet level inter-arrivals. Apart from Mitov's theorem, the Weibull Quantile-Quantile plots also gives good fit (see [7]). It should be noted that in some studies, like [15], [24], the packet inter-arrivals in tier-1 backbone traffic have been reported to be exponential or nearly uncorrelated. This is, of course, due to high multiplexing and presence of less-heavy tails in individual user's packet level traffic. According to [15], the exponential convergence result of Palm's theorem (see Section III-C) can ultimately be applied in high multiplexing environment even if individual packet level streams' inter-arrivals are heavy tailed.

The convergence of inter-arrivals (of sessions,flows and packets) to Weibull distribution has encouraged us to mention a recently introduced count model corresponding to Weibull distributed inter-arrivals, proposed by McShane (2008) [25]. This count model offers the same conceptual elegance and usefulness as the Exponential-Poisson connection. McShane's model is applicable for all possible values of shape parameters and can model counts resulting from heavy tailed and non-heavy tailed inter-arrivals in access and core networks respectively. The Weibull count model is given as:

$$P(N(t) = n) = \sum_{j=n}^{\infty} \frac{(-1)^{j+n}(\lambda t^c)^j \alpha_j^n}{\Gamma(cj + 1)}$$

where $n = 0, 1, 2, \ldots$, c is Weibull shape parameter, λ is the rate (variable), $\alpha_j^0 = \Gamma(cj + 1)/\Gamma(j + 1)$ for $j = 0, 1, 2, \ldots$, and $\alpha_j^{n+1} = \sum_{m=n}^{j-1} \alpha_m^n \Gamma(cj - cm + 1)/\Gamma(j - m + 1)$, for $n = 0, 1, 2, \ldots$ and for $j = n + 1, n + 2, n + 3, \ldots$.

The performance comparison of Weibull count model versus other count models can be found in [25]. Moreover, this count model supports both over-dispersed and under-dispersed data, therefore, it can be used to model more dynamic traffic situations in access and core networks.

VII. SESSIONS TO FLOWS TO PACKETS

The concept of session is user specific, therefore we assess traffic originating from users in access networks and use a simple method to identify sessions from the trace files. First we have identified all unique source IP addresses and then separated their traffic. On the basis of TCP SYN segments and unique 5-tuple for UDP segments, we identify new flows and record their time stamps. In such a way the new flows belonging to all users have been identified and stored separately, giving us flow inter-arrival data for every user. On the basis of empirical observations for most of the users, we have assumed that a general user Sessions are spread across time scales from minimum of 10 second to several minutes. We

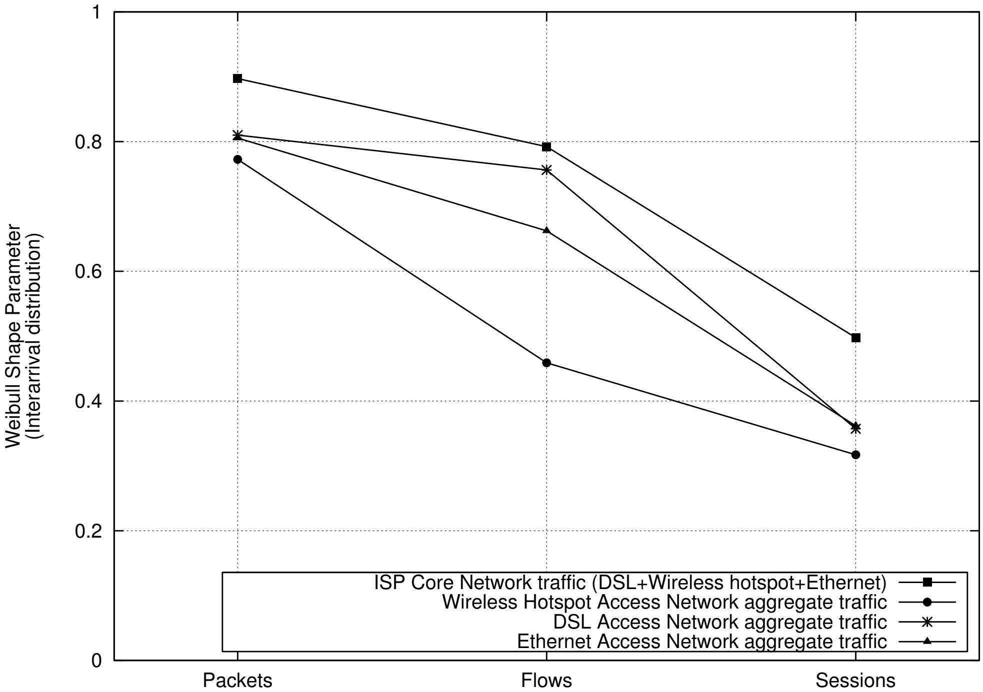

Figure 5: Weibull shape parameter vs packet-, flow-, session- inter-arrivals for different access networks and ISP core network.

check autocorrelation of flow inter-arrival data of every user and group highly correlated flows into a single session with the session time stamp being the time stamp of first flow in it. This way Session time stamps of all users have been obtained. This also gives Session Inter-arrival data of individual user. Now we combine all user session time stamps and sort them. This gives session time stamps in aggregated trace and thus converged or aggregated session inter-arrival data can be obtained.

Regarding traffic trace analysis software it will be pertinent to mention recently developed flow manager library by WAND group, LibFlowManager [26], which we have used to identify new UDP flows as well as TCP flows. This library maintains a large active flow table for both TCP and UDP flows. The UDP flows are timed out or expired if no UDP packet associated with the corresponding 5 -tuple is seen for 2 minutes; and, the TCP half open connections are expired if no TCP packet associated with it is seen for 4 minutes.

Figure 5 shows the analysis of inter-arrivals of packets, flows and sessions. It can be seen that for all the three different access networks (DSL, Ethernet and Wireless hotspot), the Weibull shape parameter is close to unity for packet inter-arrivals and it starts decreasing (becoming more heavy tailed) as we move to flow inter-arrivals and session inter-arrivals. Thus the decrease in Weibull shape parameter value from unity (approximately) to lower values models packet inter-arrivals, flow inter-arrivals and session inter-arrivals respectively. Thus the Weibull shape parameter can be used to zoom in/out between packet, flow and session inter-arrivals at a certain access or core network. Practically speaking the value of scale parameter depends on access media and tier (traffic load). In general, it is lower for core networks as compared to the scale parameter values of inter-arrivals in the attached access networks.

Figure 5 also shows that the Weibull shape parameter values of packet, flow and session inter-arrivals increases (due to high multiplexing) at the core ISP level compared to the corresponding values at the access networks. This means that the tail of these random variables becomes less heavy and thus traffic burstiness decreases (or traffic tends to Short Range Dependent) as it moves from access to core network. Differences in the access traffic patterns are also evident in Figure 5, for example, wireless hot spot traffic has least values of shape parameter for packets, flows and session inter-arrivals. This is because of higher think times and careful usage (due to cost and data rate issues) of wireless hotspot users. DSL and Ethernet users have comparatively higher Weibull shape parameter values for packets, flows and session inter-arrivals. It should be noted that such a difference may not act as a signature difference between traffic at DSL, Ethernet and Wireless hotspot networks; nevertheless, the variations in Weibull shape parameter at a certain tier may indicate some event or anomaly which is an interesting area for future research.

VIII. Conclusion

Structural modeling of Internet traffic has always been a challenging research area. We have shown the versatile role played by the simple two parameter Weibull distribution in Internet traffic structural modeling. The Weibull shape parameter can capture traffic inter-arrival (packets, flows and sessions) dynamics as it traverses from access to core networks; and, can also be used to zoom in/out between packet-, flow- and session inter-arrivals at a certain tier (access or core). Based on an extensive analysis of the traffic traces obtained from different access networks and their core network, we found that the Weibull shape parameter of packet inter-arrivals is greater than shape parameter of flow inter-arrivals and which in turn is greater than shape parameter of session inter-arrivals. We believe that such a relation can be used to quantitatively relate inter-arrival times of packets, flows and sessions. The results of this paper are useful for specifying or controlling inter-arrivals of packets, flows and application layer sessions at different tiers of various networks in simulations. This paper also opens potential avenues of future research, for e.g, time series analysis of Weibull shape parameter for event or anomaly detection at a certain tier; and, using the Weibull count data model for traffic modeling and forecasting.

IX. Acknowledgement

This research has been partially funded by Research Education and Advanced Networks New Zealand (REANNZ) under research grant E5468. The authors are thankful to Shane Alcock of WAND research group at University of Waikato for providing ISP traffic traces of DSL, Ethernet and Wireless hotspot access networks.

References

[1] J. Cao, W. S. Cleveland, D. Lin, and D. X. Sun, "Internet traffic tends to poisson and independent as the load increases," tech. rep., 2001.

[2] A. Feldmann, *Characteristics of TCP Connection Arrivals*, pp. 367–399. Self-Similar Network Traffic and Performance Evaluation, John Wiley & Sons, Inc., 2002.

[3] N. Hohn, D. Veitch, and P. Abry, "The impact of the flow arrival process in internet traffic," in *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*, vol. 6, pp. VI – 37–40 vol.6, april 2003.

[4] F. Ricciato, A. Coluccia, A. D'Alconzo, D. Veitch, P. Borgnat, and P. Abry, "On the role of flows and sessions in internet traffic modeling: An explorative toy-model," in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, pp. 1 –8, 30 2009-dec. 4 2009.

[5] S. Uhlig, O. Bonaventure, and C. Rapier, "3d-ld: a graphical wavelet-based method for analyzing scaling processes," in *Proc. of the 15th ITC Specialist Seminar*, 2002.

[6] N. Hohn, D. Veitch, and P. Abry, "Does fractal scaling at the ip level depend on tcp flow arrival processes?," in *Proc. ACM SIGCOMM Internet Measurement Workshop (IMW-2002)*, (Marseille), pp. 63–68, Nov. 6-8 2002.

[7] M. Arfeen, K. Pawlikowski, D. McNickle, and A. Willig, "Towards a combined traffic modeling framework for access and core networks," in *Telecommunication Networks and Applications Conference (ATNAC), 2012 Australasian*, pp. 1 –7, nov. 2012.

[8] J.-M. Agosta, J. Chandrasekar, M. Crovella, N. Taft, and D. Ting, "Mixture models of endhost network traffic," in *Proceedings of the Infocom 2013 Miniconference*, (Turin, Italy), 2013.

[9] P. Abry, D. Veitch, and P. Flandrin, "Long-range dependence: revisiting aggregation with wavelets," *Journal of Time Series Analysis (Bernoulli Society)*, vol. 19, pp. 253–266, May 1998.

[10] M. S. Taqqu, W. Willinger, and R. Sherman, "Proof of a fundamental result in self-similar traffic modeling," *SIGCOMM Comput. Commun. Rev.*, vol. 27, pp. 5–23, April 1997.

[11] S. Floyd and V. Paxson, "Difficulties in simulating the internet," *IEEE/ACM Transactions on Networking*, vol. 9, pp. 392–403, 2001.

[12] H. Rinne, *The Weibull Distribution A Handbook*. Chapman and Hall/CRC, 2008.

[13] D. R. Cox and W. L. Smith, "On the superposition of renewal processes," *Biometrika*, vol. 41, no. 1/2, pp. pp. 91–99, 1954.

[14] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes. Vol. II.* Probability and its Applications (New York), New York: Springer, second ed., 2008. General theory and structure.

[15] J. Cao, W. S. Cleveland, D. Lin, and D. X. Sun, "On the nonstationarity of internet traffic," in *Proceedings of the 2001 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, SIGMETRICS '01, (New York, NY, USA), pp. 102–112, ACM, 2001.

[16] K. V. Mitov and N. M. Yanev, "Superposition of renewal processes with heavy-tailed interarrival times," *Statistics &; Probability Letters*, vol. 76, no. 6, pp. 555 – 561, 2006.

[17] W. Willinger, V. Paxson, and M. S. Taqqu, *Self-similarity and heavy tails: structural modeling of network traffic*, pp. 27–53. Cambridge, MA, USA: Birkhauser Boston Inc., 1998.

[18] R. Gaigalas and I. Kaj, "Convergence of scaled renewal processes and a packet arrival model," *Bernoulli*, vol. 9, no. 4, pp. pp. 671–703, 2003.

[19] L. Lopez-Oliveros and S. I. Resnick, "On the superposition of heterogeneous traffic at large time scales," *Stochastic Systems*, vol. 1, pp. pp. 1–37, 2011.

[20] A. Bianco, G. Mardente, M. Mellia, M. Munafo, and L. Muscariello, "Web user session characterization via clustering techniques," in *Global Telecommunications Conference, 2005. GLOBECOM'05. IEEE*, vol. 2, pp. 6–pp, IEEE, 2005.

[21] WAND Network Research Group, "WITS Archive." url-http://research.wand.net.nz/wits/, 2007.

[22] A. P. Alessio Botta, Alberto Dainotti, "Multi-protocol and multi-platform traffic generation and measurement," in *INFOCOM 2007 DEMO Session*, 2007.

[23] P. Abry, P. Borgnat, F. Ricciato, A. Scherrer, and D. Veitch, "Revisiting an old friend: on the observability of the relation between long range dependence and heavy tail," *Telecommunication Systems*, vol. 43, pp. 147–165, 2010. 10.1007/s11235-009-9205-6.

[24] Z.-L. Zhang, V. Ribeiro, S. Moon, and C. Diot, "Small-time scaling behaviors of internet backbone traffic: an empirical study," in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 3, pp. 1826 – 1836 vol.3, march-3 april 2003.

[25] B. McShane, M. Adrian, E. T. Bradlow, and P. S. Fader, "Count models based on weibull interarrival times," *Journal of Business & Economic Statistics*, vol. 26, pp. 369–378, 2008.

[26] WAND Network Research Group, "Libflowmanager." URL-http://research.wand.net.nz/software/libflowmanager.php (as accessed on 1-06-2013).