Internet Traffic Modeling

Frédéric OLIVE

790314-A153

folive@hotmail.com

March 17, 2003

Abstract

Measuring, modeling and simulating the Internet has been the focus of intensive research this past decade. Yet it has been an easy task mainly because of the enormous complexity into which the Internet has grown, but also because of the lack of tools available for measurements. However, some models and patterns have been discovered, giving a better understanding of the Internet behavior, encouraging cooperation between operators and researchers.

Introduction

During the early years of the Internet, traffic was thoroughly monitored and studied. Comprehensive research programs undertook the vast task to investigate traffic pattern and model the Internet behavior. The public and centralized structure of the original ARPANET and then of its successor NSFNET provided a single backbone with accessible measurement points, which significantly help the research community with this approach. Traffic behaviors were well identified and some traffic models roughly approximated them.

Since NSFNET disappeared in 1995, the Internet is made of the association of multiple commercial backbones. It has become an incredibly more complex system during the past decade, with the growth of the number of computers connected, the amount of information sent and the number of applications. Though the behavior of these applications have individually studied, very few research has been lead in order to study the cumulated overall behavior of them in wide-area Internet. Indeed, due to the commercial nature of the Internet, empirically grounded research in the domain has not been the foremost interest of the Internet operators. The networking industry preferred to allocate financial and engineering resources to the development of profitable advanced routing and switching systems rather than measuring tools, though they rely on the same technology. Some interesting large-scale measurement projects have nonetheless been undertaken in this competitive environment, and the research community has kept on focusing on the subject despite the lack of support and tools.

This paper, which aims at giving a state of the art in the Internet traffic measurement and modeling, is based on the studies published during the last decade on the topic. First, the measurement methodology and tools are presented in section 1. Section 2 explains the motivations behind traffic measurement and modeling, and why these are crucial issues for Internet-related research. Finally, the long section 3 describes the traffic patterns and models discovered so far.

1. Internet traffic measurements

1.1 Methodology

Traffic measurement is fundamentally based on capturing TCP/IP packets from a data flow at some key locations, such as stub networks entry points and backbone links, during a certain amount of time. Indeed packet headers contain a large amount of information which main assets are the source and destination addresses and ports, the packet size, the sequence number, and the protocol type. This data is collected and stored by coping the packet, or more likely the header only to save disk space, in a database together with a timestamp at which the packet was received. This is called a trace. Figure 1 shows an example of such traces. The analysis of the traffic characteristics during that particular period of time at that

particular location would then be made on the recorded set of data. Though, the size of each sample is limited by the capture time and the disk capacity, this technique still provides enormous quantities of data to process, thus giving acceptable statistic properties to the derived results.

TIME STAMP	PROTOCOL	SOURCE IP_ADDRESS	SRC PORT	DESTINATION IP_ADDRESS	DST PORT	IP_PKT SIZE	TCP SEQ	TCP ACK
0	IP TCP	307.246.129.64	1060	427.86.12.704	80	40	920641	412791
14966	IP TCP	561.877.104.57	7410	427.86.12.704	80	508	410104	32779
15015	IP TCP	391.82.374.90	1105	891.82.59.75	80	40	2816846	7726
22090	IP TCP	719.327.502.359	1140	526.837.913.44	80	40	1010185	14762
22126	IP TCP	582.127.755.91	1291	419.74.87.6	80	40	9557082	50482
29960	IP TCP	561.877.104.57	3741	427.86.12.704	80	40	985526	58006
29960	IP TCP	419.74.87.6	80	582.127.755.91	1291	1500	653402	57082
31724	IP TCP	419.74.87.6	80	582.127.755.91	1291	1500	654862	57082
36055	IP TCP	512.84.9.317	1125	419.74.87.628	80	311	857517	89873
36279	IP TCP	512.84.9.317	1126	419.74.87.628	80	271	857661	3293
37181	IP TCP	407.84.92.183	1207	398.54.73.39	5190	40	64202	9407
41731	IP TCP	399.81.77.33	80	342.406.374.91	1116	40	1062629	68778

Figure 1 : Example of TCP/IP packets traces. The IP addresses have been sanitized. From [2]

Optimized connection-oriented measurement tools focus on TCP *SYN* and *FIN* packets. As a matter as fact, these types of packets are exchanged between hosts during a TCP connection start-up and closure. These packets contain all the information needed to describe a particular connection in only 40 bytes. For connection related investigations, it is then useless to capture the following bigger TCP/IP data packets on the flow, thus saving a great amount of computing time and disk space.

1.2. Traffic measurement tools

Traffic capture and measurement techniques can differ greatly from one another considering their accuracy, cost, scalability, simplicity of use and reliability.

The first distinction to be made is hardware vs. software-based measurement platforms, because of the dramatic differences in data capacity and processing options. Hardware-based tools, also referred to as network traffic analyzer, are quite expensive special-purpose devices allowing fast recording of packets header using extensive storage capacities. On the contrary, software-based tools are network device which operating system has been adapted for packet capture at kernel level. These can be common PC or workstation with enhanced memory and disk space. The most common tool is this category is the well-known *tcpdump* utility that allows TCP/IP packets capture at a user level. Moreover, the usually the limited performances of software measurement device oblige to an off-line analysis of the data (i.e. after the capture is finished), while some hardware-based device allow on-line analysis and visualization options.

Secondly, a crucial difference separates passive from active measurements. While passive measurement techniques just record packet as they come along on the wire, active measurement techniques actually inject packets in the network and measure responses to these packets or changes in the network characteristics due to the injection. Most of the measurement are passive, or non-intrusive, but some interesting results can be achieved with active measurement tools among which *traceroute* is the best example. However active measurements can induce a part of uncertainties in the results retrieved since the injected packets can wind up triggering unexpected behavior of the network or drastically change its characteristics, thus biasing the measures.

1.3. Measured data

Measurements produce traces from which information about traffic flows characteristics is extracted. Valuable hints on the network behavior and usage are revealed the study of two traffic variables : packets and connections. Packets have a size and an inter-arrival times that are directly available in traces that log

captured packet size and arrival timestamps. Connection variables describe the characteristics of a specific TCP connection, the application transfer requests and the environment of the network at the time of the request. These characteristics derive from the TCP three-way handshake that initiate every TCP connection (Figure 2). Indeed, connection variables are the server-side round trip time, measured by the time difference between the client SYN and the server SYN/ACK, the client-side round-trip time, measured by the time difference between the server SYN/ACK and the client final ACK, the overall size of the data transferred measured from sequence numbers of the connection and individual packet size, and the connection inter-arrival time.

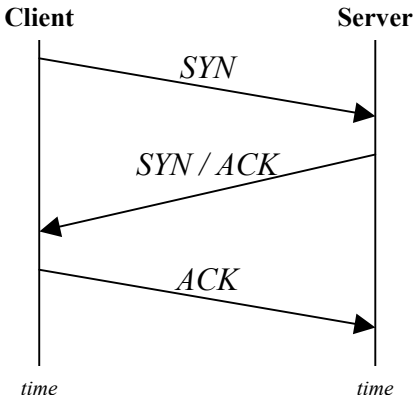

Figure 2 : TCP three-way handshake

The TCP connection flow database provides aggregates that are important to study. For example, packets sizes together with their arrival time-stamps represents an aggregated packet process : the size and the arrival times of all packets. This is meaningful since the devices at both ends of the wire have to handle the packets in time order, and their performances are directly related to the size of the packets and their arrival time. This provides a relevant framework for queuing behavior.

In the same way, connection flow data storage is important since it allows to study sub-aggregate traffic, namely packets size-time aggregate in a subset of flows. This could therefore be useful to derive flow characteristics on an application basis, by studying HTTP, FTP, SMTP etc… flows. Moreover, variable such as round-trip times and overall bytes transferred during the connection give reasonable hints with regards of congestion, while connection inter-arrival time reflects the actual behavior of users, since connections are started because of a user application request.

Finally, another important sub-aggregate process is byte counts, which represent the overall number of bytes received during equal fixed-length time intervals. Most of the Internet traffic studies have relied on bytes counts so far, since it offers an easier and concise way to trace bytes flows without the burden of storing each individual packet. However, since bytes counts are an aggregation of individual measurements, they do not always reflect all the complexity of traffic statistics, such as burstiness.

2. Goals of traffic measurements

2.1. An engineering tool

As discussed in section 1, network measurements can provide valuable information about the network behavior and usage. As a matter of fact, network measurements were first designed in order to solve engineering problems or enhance commercial services. Computer networks are really fragile : a single malfunctioning machine can crash the whole system. By screening the traffic, it is possible to identify such a malfunctioning machine and fix or isolate it. This has direct applications in network administration. Moreover, service providers or simple users want to determine how well the network is working. Measurement analysis allow to characterize network utilization and performances bottlenecks. Finally, in a more research oriented context, protocol debugging is a preliminary step to test a protocol on a restricted network before its actual world-wide deployment, and workload characterization can lead to better queuing and buffering strategies.

2.2. Towards traffic models and simulations

Measurements are needed for the crucial exploration of network traffic characteristics. Some interesting applications are applied by the engineering community in order to improve network services and devices. However, the research community focuses on the networks themselves, as an entire subject of study. Hence, in order to investigate the complexity of wide-area Internet, they thoroughly collect and analyze data regarding performances and traffic characteristics.

The first step in understanding wide-area traffic is to collect and process traffic data in order to establish statistics on a protocol and an application basis and seek for patterns. These statistics and patterns can reveal some properties of the networks or its users, but their main purpose is to build traffic models. Models intend to describe by means of statistical distributions the traffic variables. The most important variables with that regard are mainly those described in section 1.3.

Empirical models for the traffic variables distributions at an Internet site can be derived using the measured variables distributions collected at that site. Though empirical models are relatively easy to express and give reliable hints about the traffic at a specific location, analytic traffic models are preferable. Analytical models describe mathematically the distributions of traffic variable according to some adjustable set of parameters. Thus, they are likely to be applied in different locations by adapting the models parameters to the properties of the new site, and give a much better understanding of the network characteristics. However, analytical models are not easy to convey and can be subject to questioning on whether they properly describe the diverse phenomena found in wide-area traffic as well as empirical models do, and whether they faithfully capture and represent the essential characteristics of the traffic.

But models are not yet the final step in wide-area traffic studies. Traffic models provide crucial statistical results for the sake of simulation. While measurements and models explore the real characteristics of the Internet, simulation aim at exploring constructed and abstracted model of the world. Simulation is the foremost tool used by the research community to address the understanding of the Internet. It allows the exploration of complicated scenarios that would be difficult or impossible to analyze in the real world, playing a crucial role in helping the researchers to investigate the Internet traffic dynamics.

Simulations are fundamental because they reveal with a greater understanding the basic processes that rules the Internet, by analyzing a controlled model of the Internet on a restricted environment. Such knowledge could not be derived from the single analysis of traffic measurement due to the enormous number parameters such an analysis would have to cope with. However, the models used in simulation influence deeply the results, and there is always a risk of using a model simplified to the point where key aspects of the Internet behavior have been lost. Thus, modeling the dreadful complexity of the Internet properly becomes an even more dangerous and important task.

2.2. Difficulties in modeling the Internet

Though simulations appear to be the only reasonable way to characterize behavior of the traffic on the Internet and the possible effects of changes in its operation, the vital underlying models used are far from being easy to express.

First, the Internet is nothing but an assembly of diverse networking technologies and administrative domains unified thanks to the IP protocol. The price of uniform connectivity is the complex network behavior reality, that makes reliable and universal models so difficult to shape. Worse, heterogeneity exists at every scale. The link topology is heterogeneous; there is no typical link on the Internet: various competing entities use various changing technologies. Even at a protocol level, things are not as simple as they seem. For example, more than 400 different implementations of TCP have been identified on the Internet, all of them having distinct behaviors, especially in their own single way to realize congestion control. How to model traffic if does not even have unified rules ?

A second obstacle is the size of the Internet, which brings two additional problems. First, the range of the heterogeneity mentioned above is large indeed, and keeps complicating the picture as the Internet grows. The other problem is scalability. As the Internet gets bigger, many network mechanisms and protocols, which were design and tested on small network, tend to behave in an atypical fashion. Moreover, a large scale network will routinely encounter what was described as rare events during the initial deployment of its mechanisms and protocols and which subsequent effects are not known. Still following the same idea of a critical network size, even if a small number a connected computers are malfunctioning or behave strangely, this still might include thousands of them, significantly affecting the behavior of the overall system.

A third key property of the Internet is its drastic evolution in time. Thus, even if the perfect model could be found, it might only be accurate for a short period of time. The Internet has exponentially grown with incredible rates during the last decade in computer connected, number of connections, byte sent etc… indicating that rapid changes occur along many dimensions : today's Internet is not tomorrow's. This type of system is then resilient to models. Moreover, most of the changes are unpredictable due to the large numbers of organization, companies, users and policies involved. What would be the effect of providers changing their pricing policies ? What if traditional FIFO scheduling in routers is widely abandon for a fair and equal resource sharing scheduling scheme ? What would be the result of the hypothetic success of 3G mobile devices on the traffic and the infrastructure ? What will the next "killer application" after Napster ? It is nearly impossible now to answer these questions and all the more difficult to include such uncertainties into traffic models.

For all these reasons, modeling and simulating the Internet traffic is extremely painful. Models and simulations have to be as simple as possible without losing the complexity of the reality. They moreover have to give acceptable results over a large spectrum of scenarios. Faced with this nearly impossible task and a constantly changing world, researchers focused on discovering invariant, i.e. some characteristics, trends or patterns of the traffic that has been empirically shown to hold in a wide range of environment, for some significant period of time. Models would then be build on top of these invariants and verified by simulations.

3. Traffic models and patterns

The search for invariants in traffic statistics or long-life traffic patterns is crucial to express parsimonious models, which describes models that are simple enough, i.e. with a few parameters, to be applied across a wide range of conditions and environments. Parsimony is needed for an analytical model to probe successful. Since few parameters are required to describe analytically an invariant, it seems to be reasonable use them in a modeling perspective. Before investigating such empirical invariants, it is wise reminding few classical distributions and their properties often used or mentioned in the literature.

3.1. About distributions

- Poisson processes are extensively used in teletraffic theory, which originally encompassed models for the PSTN, because they nicely describe calls arrival processes in the network. The number of calls X_k received by the network in the time interval Δt follows the distribution with parameter λ :

$$P(X_k = n) = \frac{(\lambda \Delta t)^n}{n!} e^{-\lambda \Delta t}$$

  The main property of the Poisson distribution is the memoriless property, which implies exponentially distributed calls inter-arrival times.

- Distributions with an heavy-tail property are also common in traffic models. A distribution is heavy-tailed, if its complementary probability distribution exhibits an hyperbolic behavior (or a linear behavior on log-log scale) for large x-values :

$$1 - F(x) = P[X > x] \sim k\, x^{-\beta} \text{ as } x \to \infty,\ 0 < \beta < 2 \text{ and } k > 0$$

- The Pareto distribution is the simplest heavy-tailed distribution since it is hyperbolic over its entire range :

$$F(x) = P[X \le x] = 1 - \left(\frac{k}{x}\right)^{\alpha},\ \alpha, k > 0,\ x \ge k$$

- A zero-mean stochastic process $X = (X_k : k \ge 1)$ is self-similar or fractal with scaling factor $H \in [0.5, 1]$, if $\forall m \ge 1$, $X^{(m)} = m^{H-1} X$, where $X^{(m)}(k) = m^{-1}\left(X_{(m-1)k+1} + \cdots + X_{km}\right)$. Simply put, the distribution of the aggregated series is the same as that of the original, except for a range change in scale.

- A process with long-range dependence has an autocorrelation function $r(k) \sim k^{-\beta}$ as $k \to \infty$, where $0<\beta<1$. Thus, the autocorrelation function follows a power-law, as compared to exponential decays exhibited in traditional traffic distributions, such as the Poisson distribution. Hence, heavy-tailed distributions, which also have power-law decays exhibits long-range dependencies. Furthermore, it is proven that self-similar processes have the long-range dependence property.

- Inter-arrivals times can be approximated by Weibull distributions, where the parameters depend on the connection rate ρ. With $\lambda(\rho)$ being the shape parameter, $\alpha(\rho)$ the scale parameter and υ the inter-arrival variable, then :

$$\left(\frac{\upsilon}{\alpha(\rho)}\right)^{\lambda(\rho)} = u$$

where u is a unit exponential. The main characteristics of this distribution is that it tends toward an exponential distribution as λ tends toward 1.

3.2. Daily and weekly traffic patterns

Network activities follows human activity daily and weekly pattern, beginning around 8-9 AM and rising until lunch time around 11-12 AM, slightly decreasing until 1 PM where it picks back up again until 5-8 PM and then declining as the business day ends. The network activity renews in the early evenings hours peaking at 10-12 PM as users now use the network from their home. Figure 3 describe such a behavior, measured on a major U.S. East Cost backbone trunk ([4]), and also reveal the presence of a diminished traffic pattern on week-ends and holidays.

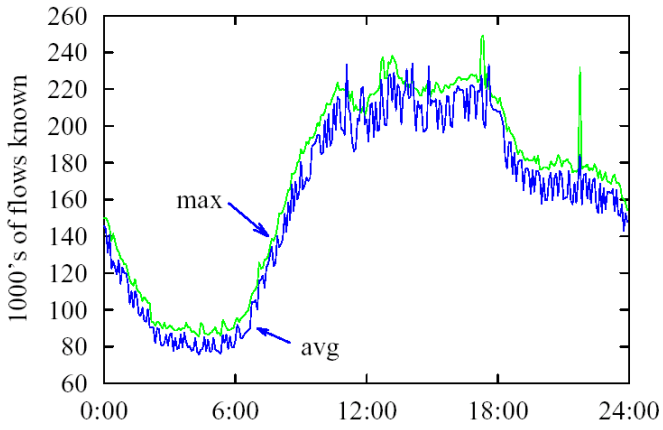

Figure 3a : Flow volume for 1 day [4]

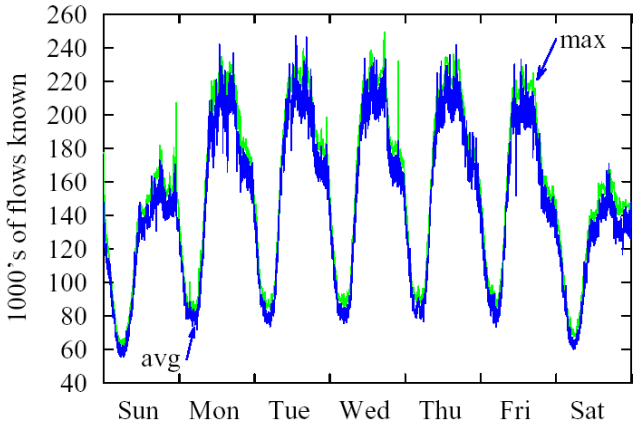

Figure 3b : Flow volume for 7 days [4]

However, significant variations in this pattern exist between different protocols, for example between human and non-human initiated protocols such as NNTP. There even exist different pattern for the same protocol : HTTP patterns are affected by switching from work-related to leisure-related Web-surfing. Though, for a particular flow of traffic, diurnal and weekly patterns are considered as an invariant, which is particularly useful to model and generate traffic plausible variations in simulations.

3.3. IP traffic pattern

In the same way, IP traffic composition in wide-area Internet has been studied and described for a long time. Though the overall amount of traffic has been growing exponentially through the last decade, the repartition of the TCP/IP suite various protocols in the traffic still remains rather constant. This allows to create simple models for traffic generation. [6] presents such results on a protocol and application basis (Tables 1 and 2, Figures 4 and 5).

	Bytes %	Packets %	Flow %
TCP	95	90	75
UDP	5	10	20
ICMP	<1	<1	5
Others (IPv6, IP-in-IP)	<1	<1	<1

Table 1 : IP protocols traffic composition [4].

	Bytes %	Packets %	Flow %
www	75	70	75
DNS	1	3	18
SMTP	5	5	2
FTP data	5	3	<1
NNTP	2	<1	<1
Telnet	<1	<1	<1

Table 2 : IP applications traffic composition [4].

Furthermore, Figures 4 and 5 show that the traffic composition does not significantly change throughout the day even if the traffic volume varies. The dominant protocol is by far TCP, followed by UDP, while the main applications are the Web and DNS. It is then easy to deduce that web-surfing is the main activity for users. Traffic models and simulations should focus first on Web related applications and protocol, since they account for nearly 75 % of the traffic.

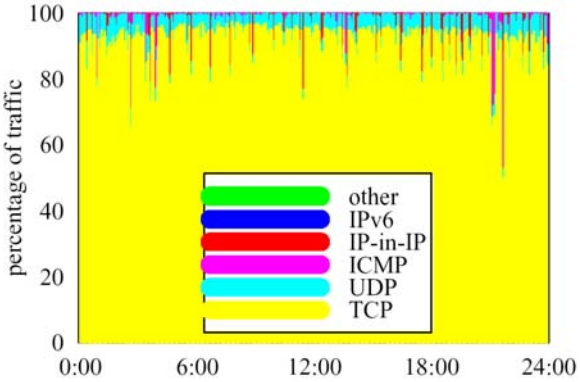

Figure 4 : Daily IP protocols traffic composition [4].

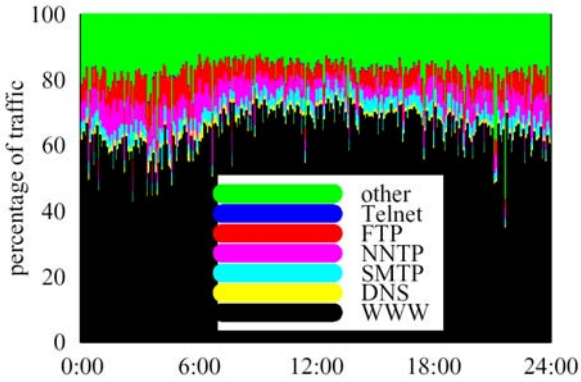

Figure 5 : Daily IP application traffic composition [4].

3.4. Poisson session inter-arrival times

An other user-related pattern describes requests or sessions inter-arrival times on servers. The study presented in [5] reveals that that the model suggested by Paxson and Floyd in [3] applies. The postulate was that request inter-arrival times for services such as FTP or HTTP could be well-described using Poisson processes, provided that the rate of the Poisson process is allowed to vary on an hourly basis. Figure 6 shows a comparison of the actual connection inter-arrival time for a one-hour interval at a popular Web server (for 1166 traces and a mean time of 3.09 s) with an exponential distribution with the same mean value. The empirical measure seems to follow quite closely the analytical distribution for short time scales, which validates partially the suggested model. The heavier tail presented by the empirical distribution might be the consequence of automatic machine initiated requests. It is furthermore worth noticing that the multiple connections that comprise a session do not follow a Poisson model, mainly because connections within a session are correlated.

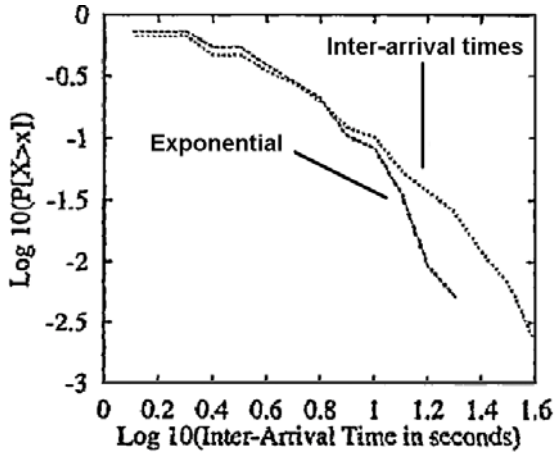

Figure 6 : Comparison of request inter-arrival time with the exponential distribution [5]

3.5. Log-normal connections sizes and durations

Though connections can not be modeled as Poisson processes, their size and duration follow log-normal distributions, as shown on Figure 7. The body of the empirical distribution matches almost perfectly the analytical distribution. Moreover, the distribution also provides another invariant which has to deal with the empirical distribution heavy-tailed property with $\beta \approx 1.3$.

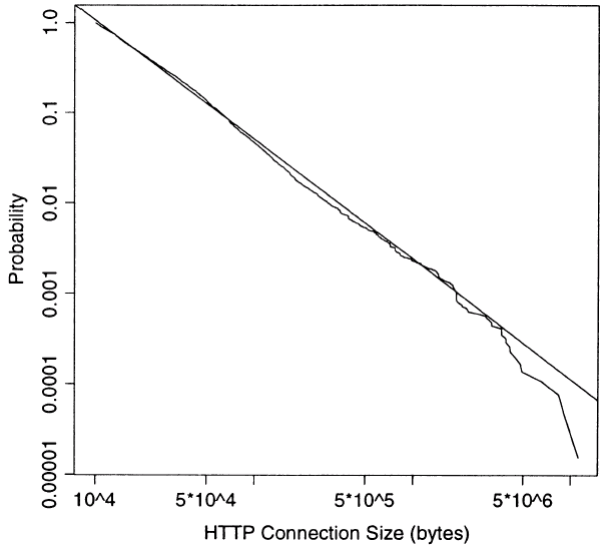

Figure 7 : Log-log complementary distribution plot of HTTP connection sizes at server X [6]. For 32,630 connections of at least 10,000 bytes.

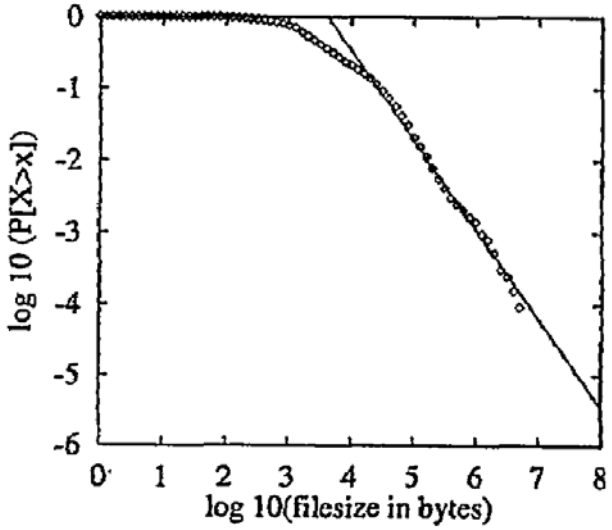

Figure 8 : Log-Log complementary distribution of file size on server X [5]. The heavy-match match a Pareto distribution with $\alpha = 1.24$.

Together with log-log distributions of traffic variables, the heavy tailed property of connection size and duration is not something completely new and atypical. They appear in a wide-range of computer related phenomena, such as CPU time consumed by processes, sizes of Unix files, Ethernet bursts etc… which are also classical invariants.

The heavy-tail distributions of connections size and duration can actually be explained by the shape of the file size distribution on Web servers (Figure 8). The heavy-tail parameters match in both connection size and file size distributions, and assuming constant or slow varying transfer byte rates, the connection duration distribution shape should also follow the file size distribution.

3.6. Long-range dependence and self-similarity

Like user-related network behavior, packets traffic has also its share of invariants. Long-term correlations in packet arrivals in aggregated traffic are well-described by self-similar processes, contrary to the traditional teletraffic theory and the common intuition that long-term correlations should rapidly die-out and that traffic observed on large time scale should be relatively smooth.

[7] provides strong empirical evidences about self-similarities and long-range dependencies in various aspects of the World Wide Web transfers. Namely, transmission times as well as periods of activity and inactivity during the transfer operation are self-similar because of the basic characteristics of information organization. Indeed, the size distribution of available files on the Web roughly follows a Pareto distribution (Figure 8). Other studies have revealed other self-similar variables in Internet traffic, such as packet and connection inter-arrival times. It seems today that Internet traffic exhibits self-similarity at many scales leading the research community to lean towards multifractal traffic modeling schemes.

Yet related to long-range persistence, the Internet nonstationarity has received much less attention, until it was clearly demonstrated at Bell Labs [8] using HTTP start times. Nonstationarity means that the profound statistical properties of the traffic can change in time. A changing number of superposed changing traffic sources has been found to be the main cause to the Internet nonstationarity. As shown on Figure 9a, as the rate connection increases at a Web server the Weibull shape parameter for the request arrival time distribution tends to 1, which means that connection inter-arrival times tend to be

exponentially distributed : the request inter-arrival time process switches from self-similar to Poisson. This result is somewhat surprising : the number of active connection has a dramatic effect on traffic characteristics.

Figure 9b reveals that the packet inter-arrival times follow the same trend as load increases, suggesting that more traffic parameters are affected by the network push back toward Poisson behavior. Indeed [8] provides strong evidence that the queuing behavior, the round-trip times and the transferred file sizes are also nonstationary. At low connection loads the traffic is self-similar creating burstiness (uncorrelated packet arrival time), but as the load increases the laws of superposition push the arrivals toward Poisson, the size toward independence and reduces the variability of the main traffic variables.

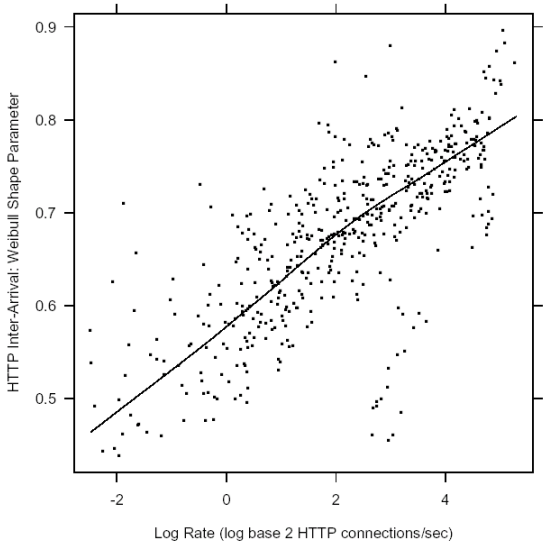

Figure 9a : Weibull shape parameter λ for HTTP request inter-arrival times against the connection rate [8]. As λ tends to 1, the arrival process tends to Poisson.

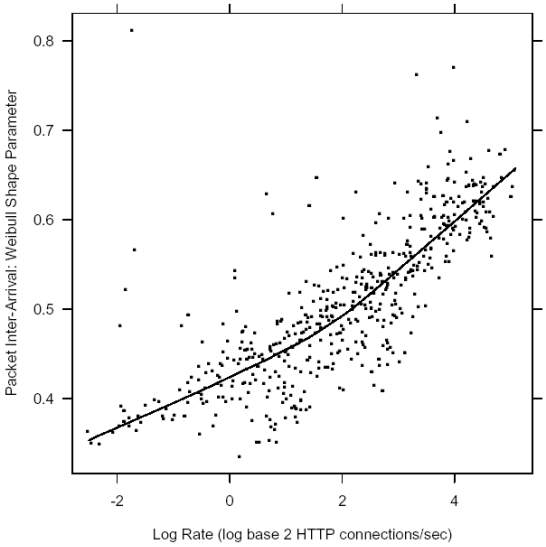

Figure 9a : Weibull shape parameter for packets inter-arrival times against the connection rate [8]. As λ tends to 1, the arrival process tends to Poisson.

3.7. Power-laws in global topology

Despite the apparent randomness and heterogeneity of the Internet topology, [9] discovers surprisingly simple power-laws relationships that accurately describes it. From the inter-domain routing BGP tables, the Internet could represented as a directed graph with routers as nodes and links as edges (Figure 10 and 11), to which principles of graph theory can be applied. The results are derived from the Internet topology in December 1998, which at that time counted 4389 nodes, 8256 edges and an average of 3.76 outdegree per node.

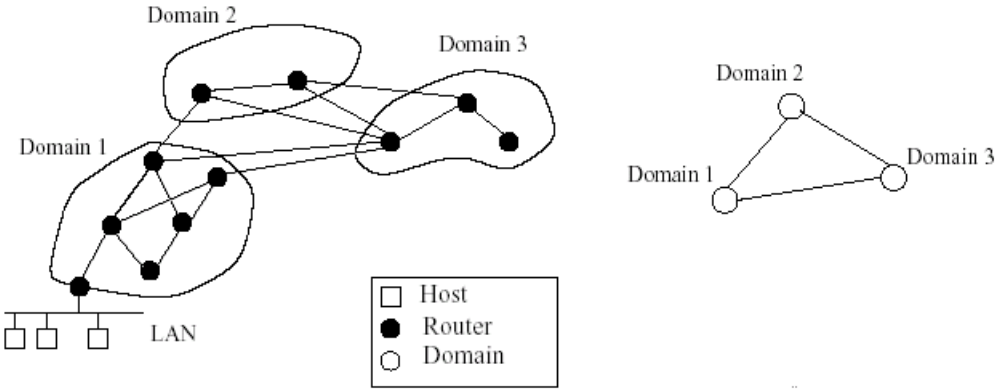

Figure 10 : Graph representation of the Internet topology [9].

The basic parameters of the study are the nodes outdegrees, d_v, i.e. the number of outward directed edges from a given node, and the ranks of the nodes, r_v, which are their index in the order of decreasing outdegrees. [9] presents number of results but here are the two most striking observations it provides.

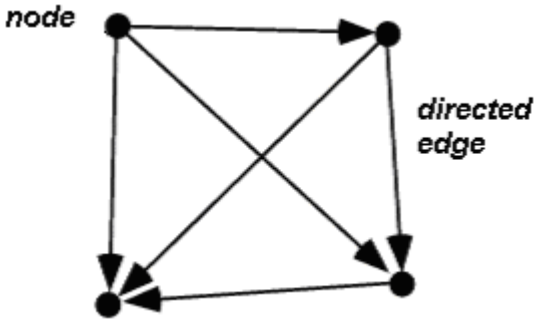

Figure 11 : Graph representation terminology

First the plots of the pair (r_v, d_v) in log-log scale are approximated by a simple linear regression (Figure 12) : the outdegree of a node is proportional to the rank of that node to the power of a constant.

Secondly, the frequency, f_d, of occurrence of an outdegree d, follows a similar law since it is proportional to the power of that outdegree, with a constant coefficient (Figure 13) .

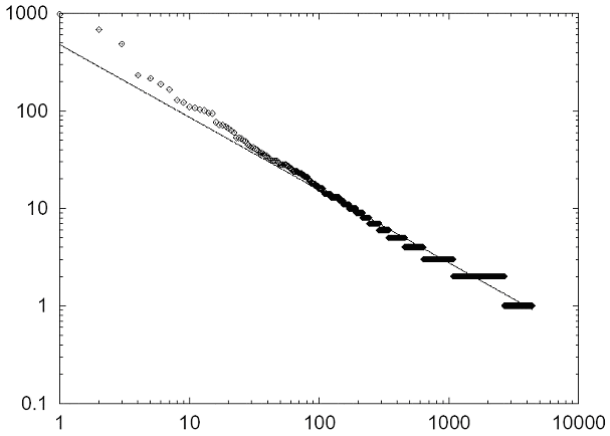

Figure 12 : Log-Log plot of the outdegree d_v versus the rank r_v in the sequence of deceasing outdegree [9].

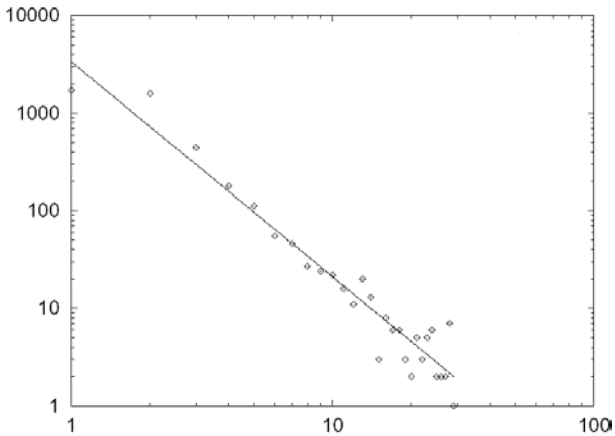

Figure 13 : Log-Log plot of the frequency f_d versus the outdegree d [9].

The Internet topology proves to have a certain level of order : nodes and edges (i.e. routers and links) distribution follows a rather simple pattern. Being able to construct a faithful but simple topological representation of Internet is clearly an important result for modeling and simulating the Internet.

Conclusion

Finally, the research community has managed what some thought impossible at first glance : finding simple models to describe the unbelievable complexity of the Internet. Though most models do not provide closed analytical expressions, with few parameters to play with, they allow to understand and faithfully reproduce some of the Internet characteristics for the sake of simulations. The search for analytical models [10] was recently boosted by the discovery of the fractal and even multifractal structure of some aspect of wide-area Internet traffic. The development of this models relies on advanced mathematical theory and could hopefully provides new hints.

Furthermore, the whole Internet community is now becoming aware that everybody has something to gain from traffic measurements : application designers have to understand the basic mechanisms of the Internet behavior to develop non-disruptive and optimized applications, ISPs need to deploy measurement tools to monitor their network and perform troubleshooting, pricing, and performance enhancements, etc … As the Internet enters another consecutive decade of sustained growth, organizations involved in its operation and those focused on its study must find a way to cooperate to provide a reliable world-wide measurement platform. Hopefully, the surprising results painfully achieved by standalone researchers have produced some interesting incentives that start to pave the way toward a global unified measurement system.

References

[1] K.C. Claffy, "*Measuring the Internet*", Internet Computing, IEEE, Vol. 9, No. 1, pp. 73-75, Jan/Feb 2000.

[2] C. Williamson, "*Internet traffic measurement*", Internet Computing, IEEE, Vol. 5, No. 6, pp. 70-74, Nov/Dec 2001.

[3] S. Floyd and V. Paxson, "*Difficulties in Simulating the Internet*", IEEE/ACM Transactions on Networking, Vol. 9, No. 4, pp. 392-403, 2001.

[4] K. Thompson, G. Miller, and R. Wilder, "*Wide-area Internet Traffic Patterns and Characteristics*", IEEE Network, Vol. 11, No. 6, pp. 10-23, Nov/Dec 1997.

[5] M. Arlitt and C. Williamson, "*Internet Web Servers: Workload Characterization and Performance Implications*", IEEE/ACM Transactions on Networking, Vol. 5, No. 5, pp. 815-826, October 1997.

[6] W. Willinger and V. Paxson, "*Where Mathematics Meets the Internet*", Notices of the American Mathematical Society, Vol. 45, No. 8, pp. 961-970, August 1998.

[7] M. Crovella and A. Bestavros, "*Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes*", IEEE/ACM Transactions on Networking, Vol. 5, No. 6, pp. 835-846, December 1997.

[8] J. Cao, W. Cleveland, D. Lin, and D. Sun, "*On the Nonstationarity of Internet Traffic*", Proceedings of ACMSIGMETRICS, Cambridge, MA, pp. 102-112, June 2001.

[9] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "*On Power-Law Relationships of the Internet Topology*", Proceedings of ACM SIGCOMM, pp. 251-262, Cambridge, MA, September 1999.

[10] V. Paxson, "*Empirically-Derived Analytic Models of Wide-Area TCP Connections*", IEEE/ACM Transactions on Networking, Vol. 2, No. 4, pp. 316-336, August 1994.