[1] P. Abry, R. Baraniuk, P. Flandrin, R. Riedi, and D. Veitch. Multiscale nature of network traffic. *Signal Processing Magazine, IEEE*, 19(3):28-46, May 2002. [ bib | DOI ]

The complexity and richness of telecommunications traffic is such that one may despair to find any regularity or explanatory principles. Nonetheless, the discovery of scaling behavior in teletraffic has provided hope that parsimonious models can be found. The statistics of scaling behavior present many challenges, especially in nonstationary environments. In this article, we overview the state of the art in this area, focusing on the capabilities of the wavelet transform as a key tool for unraveling the mysteries of traffic statistics and dynamics

Keywords: multiscale network traffic;nonstationary environments;parsimonious models;scaling behavior statistics;telecommunications traffic;teletraffic;traffic dynamics;traffic statistics;wavelet transform;statistical analysis;telecommunication networks;telecommunication traffic;wavelet transforms;

[2] G. Bissias, M. Liberatore, D. Jensen, and B. Levine. Privacy vulnerabilities in encrypted HTTP streams. In *Privacy Enhancing Technologies (PET)*, volume 3856 of *Lecture Notes in Computer Science*, pages 1-11. 2006. [ bib | DOI ]

Encrypting traffic does not prevent an attacker from performing some types of traffic analysis. We present a straightforward traffic analysis attack against encrypted HTTP streams that is surprisingly effective in identifying the source of the traffic. An attacker starts by creating a profile of the statistical characteristics of web requests from interesting sites, including distributions of packet sizes and inter-arrival times. Later, candidate encrypted streams are compared against these profiles. In our evaluations using real traffic, we find that many web sites are subject to this attack. With a training period of 24 hours and a 1 hour delay afterwards, the attack achieves only 23% accuracy. However, an attacker can easily pre-determine which of trained sites are easily identifiable. Accordingly, against 25 such sites, the attack achieves 40% accuracy; with three guesses, the attack achieves 100% accuracy for our data. Longer delays after training decrease accuracy, but not substantially. We also propose some countermeasures and improvements to our current method. Previous work analyzed SSL traffic to a proxy, taking advantage of a known flaw in SSL that reveals the length of each web object. In contrast, we exploit the statistical characteristics of web streams that are encrypted as a single flow, which is the case with WEP/WPA, IPsec, and SSH tunnels.

[3] M. S. Borella. Source models of network game traffic. *Computer Communications*, 23(4):403-410, 2000. [ bib | DOI ]

We study the traffic generated by sessions of a popular multi-player network game. Our analysis indicates that empirical game traffic can be characterized well by certain analytical models. While clients and servers as well as hosts with different configurations, produce different models, all models from the game are well modeled by the same families of distributions. We find that some data sets are best modeled with split distributions; that is, one portion of the data is well modeled with one particular distribution, and the rest of the data with another. We describe how our models can be simulated and discuss how host processing speed influences packet interarrival processes, which in turn influence playability. The latter is a clear indication that end user quality of service is more than just a network issue-host characteristics must be considered as well. These are empirical results that have been rarely incorporated into theoretical analyses of network traffic. As Internet gaming becomes more popular, we expect that our models will be useful for testing hardware and protocols that support gaming.

Keywords: Analytical models, Multi-player network game, Internet gaming

[4] k. claffy, D. Andersen, and P. Hick. The CAIDA Anonymized Internet Traces - Equinix, Chicago, 17 Feb 2011. Data set, 2011. [ bib | http ]

[5] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661-703, 2009. [ bib | DOI | www: ]

Power-law distributions occur in many situations of scientific interest and have significant consequences for our understanding of natural and man-made phenomena. Unfortunately, the empirical detection and characterization of power laws is made difficult by the large fluctuations that occur in the tail of the distribution. In particular, standard methods such as least-squares fitting are known to produce systematically biased estimates of parameters for power-law distributions and should not be used in most circumstances. Here we describe statistical techniques for making accurate parameter estimates for power-law data, based on maximum likelihood methods and the Kolmogorov-Smirnov statistic. We also show how to tell whether the data follow a power-law distribution at all, defining quantitative measures that indicate when the power law is a reasonable fit to the data and when it is not. We demonstrate these methods by applying them to twenty-four real-world data sets from a range of different disciplines. Each of the data sets has been conjectured previously to follow a power-law distribution. In some cases we find these conjectures to be consistent with the data while in others the power law is ruled out.

Keywords: power-law distributions; Pareto; Zipf; maximum likelihood; heavy-tailed distributions; likelihood ratio test; model selection

[6] A. Clauset, M. Young, and K. S. Gleditsch. On the frequency of severe terrorist events. *Journal of Conflict Resolution*, 51(1):58-87, 2007. [ bib | DOI ]

In the spirit of Lewis Richardson's original study of the statistics of deadly conflicts, we study the frequency and severity of terrorist attacks worldwide since 1968. We show that these events are uniformly characterized by the phenomenon of “scale invariance,” that is, the frequency scales as an inverse power of the severity, $P(x)$ $Ax^{-\alpha}$. We find that this property is a robust feature of terrorism, persisting when we control for economic development of the target country, the type of weapon used, and even for short time scales. Further, we show that the center of the distribution oscillates slightly with a period of roughly $\tau$˜13 years, that there exist significant temporal correlations in the frequency of severe events, and that current models of event incidence cannot account for these variations or the scale invariance property of global terrorism. Finally, we describe a simple toy model for the generation of these statistics and briefly discuss its implications.

Keywords: terrorism; frequency-severity statistics; scale invariance; Richardson's law

[7] R. Clegg, C. Di Cairano-Gilfedder, and S. Zhou. A critical look at power law modelling of the internet. *Computer Communications*, 33(3):259-268, 2010. [ bib ]

This paper takes a critical look at the usefulness of power law models of the Internet. The twin focuses of the paper are Internet traffic and topology generation. The aim of the paper is twofold. Firstly it summarises the state of the art in power law modelling particularly giving attention to existing open research questions. Secondly it provides insight into the failings of such models and where progress needs to be made for power law research to feed through to actual improvements in network performance.

[8] R. Clegg, R. Landa, and M. Rio. Criticisms of modelling packet traffic using long-range dependence (extended version). *Journal of Computer and System Sciences*, 77(5):861-868, Sept. 2011. [ bib | DOI ]

This paper criticises the notion that long-range dependence is an important contributor to the queuing behaviour of real Internet traffic. The idea is questioned in two different ways. Firstly, a class of models used to simulate Internet traffic is shown to have important theoretical flaws. It is shown that this behaviour is inconsistent with the behaviour of real traffic traces. Secondly, the notion that long-range correlations significantly affects the queuing performance of traffic is investigated by destroying those correlations in real traffic traces (by reordering). It is shown that the longer ranges of correlations are not important to mean queue length except in one case with an extremely high load.

Keywords: Long-range dependence, Queueing, Traffic

[9] M. Crovella and A. Bestavros. Self-similarity in World Wide Web traffic: evidence and possible causes. *IEEE/ACM Transactions on Networking*, 5(6):835-846, Dec. 1997. [ bib | DOI ]

The notion of self-similarity has been shown to apply to wide-area and local-area network traffic. We show evidence that the subset of network traffic that is due to World Wide Web (WWW) transfers can

show characteristics that are consistent with self-similarity, and we present a hypothesized explanation for that self-similarity. Using a set of traces of actual user executions of NCSA Mosaic, we examine the dependence structure of WWW traffic. First, we show evidence that WWW traffic exhibits behavior that is consistent with self-similar traffic models. Then we show that the self-similarity in such traffic can be explained based on the underlying distributions of WWW document sizes, the effects of caching and user preference in file transfer, the effect of user ldquo;think time rdquo;, and the superimposition of many such transfers in a local-area network. To do this, we rely on empirically measured distributions both from client traces and from data independently collected at WWW servers

Keywords: LAN traffic;NCSA Mosaic;WAN traffic;WWW document size distribution;WWW servers;WWW traffic;World Wide Web traffic;caching;client traces;dependence structure;empirically measured distributions;file transfer;local-area network;self-similar traffic models;self-similarity;statistical tests;user preference;user think time;wide-area network;Internet;local area networks;performance evaluation;statistical analysis;telecommunication traffic;wide area networks;

[10] R. B. D'Agostino and M. A. Stephens, editors. *Goodness-of-fit techniques*. Marcel Dekker, Inc., 1986. [ bib ]

[11] S. Dharmapurikar and V. Paxson. Robust TCP stream reassembly in the presence of adversaries. In *Proceedings of the 14th USENIX Security Symposium*, pages 5-5, 2005. [ bib ]

There is a growing interest in designing high-speed network devices to perform packet processing at semantic levels above the network layer. Some examples are layer-7 switches, content inspection and transformation systems, and network intrusion detection/prevention systems. Such systems must maintain per-flow state in order to correctly perform their higher-level processing. A basic operation inherent to per-flow state management for a transport protocol such as TCP is the task of reassembling any out-of-sequence packets delivered by an underlying unreliable network protocol such as IP. This seemingly prosaic task of reassembling the byte stream becomes an order of magnitude more difficultto soundly execute when conducted in the presence of an adversary whose goal is to either subvert the higher-level analysis or impede the operation of legitimate traffic sharing the same network path.

We present a design of a hardware-based high-speed TCP reassembly mechanism that is robust against attacks. It is intended to serve as a module used to construct a variety of network analysis systems, especially intrusion prevention systems. Using trace-driven analysis of out-of-sequence packets, we first characterize the dynamics of benign TCP trafficand show how we can leverage the results to design a reassembly mechanism that is efficient when dealing with non-attack traffic. We then refine the mechanism to keep the system effective in the presence of adversaries. We show that although the damage caused by an adversary cannot be completely eliminated, it is possible to mitigate the damage to a great extent by careful design and resource allocation. Finally, we quantify the trade-off between resource availability and damage from an adversary in terms of Zombie equations that specify, for a given configuration of our system, the number of compromised machines an attacker must have under their control in order to exceed a specified notion of “acceptablecollateral damage.”

[12] M. D. H. M. Moghaddam, B. Li and I. Goldberg. SkypeMorph: Protocol obfuscation for Tor bridges. Technical Report CACR 2012-08, Centre for Applied Cryptographic Research (CACR), University of Waterloo, Canada, 2012. http://cacr.uwaterloo.ca/techreports/2012/cacr2012-08.pdf. [ bib ]

[13] D. Herrmann, R. Wendolsky, and H. Federrath. Website fingerprinting: attacking popular privacy enhancing technologies with the multinomial naïve-bayes classifier. In *Proc. of Workshop on Cloud Computing Security (CCSW)*, pages 31-42, 2009. [ bib | DOI ]

Privacy enhancing technologies like OpenSSL, OpenVPN or Tor establish an encrypted tunnel that enables users to hide content and addresses of requested websites from external observers This protection is endangered by local traffic analysis attacks that allow an external, passive attacker between the PET system and the user to uncover the identity of the requested sites. However, existing proposals for such attacks are not practicable yet.

We present a novel method that applies common text mining techniques to the normalised frequency

distribution of observable IP packet sizes. Our classifier correctly identifies up to 97% of requests on a sample of 775 sites and over 300,000 real-world traffic dumps recorded over a two-month period. It outperforms previously known methods like Jaccard's classifier and Naïve Bayes that neglect packet frequencies altogether or rely on absolute frequency values, respectively. Our method is system-agnostic: it can be used against any PET without alteration. Closed-world results indicate that many popular single-hop and even multi-hop systems like Tor and JonDonym are vulnerable against this general fingerprinting attack. Furthermore, we discuss important real-world issues, namely false alarms and the influence of the browser cache on accuracy.

Keywords: forensics, low-latency anonymity, text mining, traffic analysis

[14] T. R. Kevin P. Dyer, Scott E. Coull and T. Shrimpton. Peek-a-boo, I still see you: Why efficient traffic analysis countermeasures fail. In *Proc. of IEEE Symposium on Security and Privacy*, May 2012. [ bib ]

[15] W. Lian, F. Monrose, and J. McHugh. Traffic classification using visual motifs: an empirical evaluation. In *Proc. of Symposium on Visualization for Cyber Security (VizSec)*, pages 70-78, 2010. [ bib | DOI ]

In this paper, we explore the effectiveness of using graphical methods for isolating the differences between common application protocols-both in their transient and steady-state behavior. Specifically, we take advantage of the observation that many Internet application protocols proscribe a very specific series of client/server interactions that are clearly visible in the sizes and timing of packets produced at the network layer and below. We show how so-called “visual motifs” built on these features can be used to assist a human operator to recognize application protocols in unidentified traffic. From a practical point of view, visual traffic classification can be used, for example, for anomaly detection to verify that all traffic to a web server on TCP port 80 does indeed exhibit the characteristic behavior patterns of HTTP, or for misuse detection to find unauthorized servers or to identify traffic generated by prohibited applications. We present our technique for building a classifier based on the notion of visual motifs and report on our experience using this technique to automatically classify on-the-wire behavioral patterns from network flow data collected from a campus network. Specifically, we analyze over 1 billion flows corresponding to over 5 million sessions on nearly 200 distinct ports and show that our approach achieves high recall and precision.

Keywords: evaluation, information visualization, network management, security, traffic analysis, traffic visualization

[16] M. Liberatore and B. N. Levine. Inferring the source of encrypted HTTP connections. In *Proceedings of ACM Conference on Computer and Communications Security (CCS)*, pages 255-263, 2006. [ bib | DOI ]

We examine the effectiveness of two traffic analysis techniques for identifying encrypted HTTP streams. The techniques are based upon classification algorithms, identifying encrypted traffic on the basis of similarities to features in a library of known profiles. We show that these profiles need not be collected immediately before the encrypted stream; these methods can be used to identify traffic observed both well before and well after the library is created. We give evidence that these techniques will exhibit the scalability necessary to be effective on the Internet. We examine several methods of actively countering the techniques, and we find that such countermeasures are effective, but at a significant increase in the size of the traffic stream. Our claims are substantiated by experiments and simulation on over 400,000 traffic streams we collected from 2,000 distinct web sites during a two month period.

Keywords: low-latency anonymity, network forensics, traffic analysis

[17] R. A. Lockhart and M. A. Stephens. Estimation and tests of fit for the three-parameter Weibull distribution. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(3):491-500, 1994. [ bib | www: ]

Estimation techniques are given for the three-parameter Weibull distribution, with all parameters unknown. Tables are given for the empirical distribution function statistics W<sup>2</sup>,

U<sup>2</sup> and A<sup>2</sup>, for testing for the distribution.

[18] A. Moore, D. Zuev, and M. L. Crogan. Discriminators for use in flow-based classification. Technical Report RR-05-13, Dept. of Computer Science, Queen Mary University of London, Aug. 2005. [ bib ]

[19] A. Panchenko, L. Niessen, A. Zinnen, and T. Engel. Website fingerprinting in onion routing based anonymization networks. In *Proc. of Workshop on Privacy in the Electronic Society (WPES)*, pages 103-114, 2011. [ bib | DOI ]

> Low-latency anonymization networks such as Tor and JAP claim to hide the recipient and the content of communications from a *local observer*, i.e., an entity that can eavesdrop the traffic between the user and the first anonymization node. Especially users in totalitarian regimes strongly depend on such networks to freely communicate. For these people, anonymity is particularly important and an analysis of the anonymization methods against various attacks is necessary to ensure adequate protection. In this paper we show that anonymity in Tor and JAP is not as strong as expected so far and cannot resist website fingerprinting attacks under certain circumstances. We first define features for *website fingerprinting* solely based on volume, time, and direction of the traffic. As a result, the subsequent classification becomes much easier. We apply support vector machines with the introduced features. We are able to improve recognition results of existing works on a given state-of-the-art dataset in Tor from 3% to 55% and in JAP from 20% to 80%. The datasets assume a closed-world with 775 websites only. In a next step, we transfer our findings to a more complex and realistic open-world scenario, i.e., recognition of several websites in a set of thousands of random unknown websites. To the best of our knowledge, this work is the first successful attack in the open-world scenario. We achieve a surprisingly high true positive rate of up to 73% for a false positive rate of 0.05%. Finally, we show preliminary results of a proof-of-concept implementation that applies camouflage as a countermeasure to hamper the fingerprinting attack. For JAP, the detection rate decreases from 80% to 4% and for Tor it drops from 55% to about 3%.
>
> Keywords: anonymous communication, pattern recognition, privacy, traffic analysis, website fingerprinting

[20] V. Paxson. End-to-end routing behavior in the Internet. *SIGCOMM Comput. Commun. Rev.*, 26:25-38, August 1996. [ bib | DOI ]

> The large-scale behavior of routing in the Internet has gone virtually without any formal study, the exception being Chinoy's analysis of the dynamics of Internet routing information [Ch93]. We report on an analysis of 40,000 end-to-end route measurements conducted using repeated “traceroutes” between 37 Internet sites. We analyze the routing behavior for pathological conditions, routing stability, and routing symmetry. For pathologies, we characterize the prevalence of routing loops, erroneous routing, infrastructure failures, and temporary outages. We find that the likelihood of encountering a major routing pathology more than doubled between the end of 1994 and the end of 1995, rising from 1.5% to 3.4%. For routing stability, we define two separate types of stability, “prevalence” meaning the overall likelihood that a particular route is encountered, and “persistence,” the likelihood that a route remains unchanged over a long period of time. We find that Internet paths are heavily dominated by a single prevalent route, but that the time periods over which routes persist show wide variation, ranging from seconds up to days. About 2/3's of the Internet paths had routes persisting for either days or weeks. For routing symmetry, we look at the likelihood that a path through the Internet visits at least one different city in the two directions. At the end of 1995, this was the case half the time, and at least one different autonomous system was visited 30% of the time.

[21] J.-F. Raymond. Traffic analysis: Protocols, attacks, design issues, and open problems. In *Designing Privacy Enhancing Technologies*, volume 2009 of *Lecture Notes in Computer Science*, pages 10-29. 2001. [ bib | DOI ]

> We present the trafic analysis problem and expose the most important protocols, attacks and design issues. Afterwards, we propose directions for further research. As we are mostly interested in efficient and practical Internet based protocols, most of the emphasis is placed on mix based constructions. The presentation is informal in that no complex definitions and proofs are presented, the aim being more to give a thorough introduction than to present deep new insights.

[22] M. Sanli, E. G. Schmidt, and H. C. Güran. FPGEN: A fast, scalable and programmable traffic generator for the performance evaluation of high-speed computer networks. *Performance Evaluation*, 68(12):1276-1290, Dec. 2011. [ bib | DOI ]

Testing today's high-speed network equipment requires the generation of network traffic which is similar to the real Internet traffic at Gbps line rates. There are many software-based traffic generators which can generate packets according to different stochastic distributions. However, they are not suitable for high-speed hardware test platforms. This paper describes FPGEN (Fast Packet GENerator), a programmable random traffic generator which is entirely implemented on FPGA (Field Programmable Gate Array). FPGEN can generate variable packet sizes and traffic with Poisson and Markov-modulated on-off statistics at OC-48 rate per interface. Our work that is presented in this paper includes the theoretical design of FPGEN, the hardware design of the FPGA-based traffic generator board (printed circuit board design and construction) and the implementation of FPGEN on FPGA. Our experimental study demonstrates that FPGEN can achieve both the desired rate and statistical properties for the generated traffic.

Keywords: FPGA, High-speed traffic generator, Poisson traffic, Markov-modulated on-off traffic

[23] D. Tammaro, S. Valenti, D. Rossi, and P. Antonio. Exploiting packet-sampling measurements for traffic characterization and classification. *International Journal of Network Management*, 2012. [ bib | DOI ]

The use of packet sampling for traffic measurement has become mandatory for network operators to cope with the huge amount of data transmitted in today's networks, powered by increasingly faster transmission technologies. Therefore, many networking tasks must already deal with such reduced data, more available but less rich in information. In this work we assess the impact of packet sampling on various network monitoring-activities, with a particular focus on traffic characterization and classification. We process an extremely heterogeneous dataset composed of four packet-level traces (representative of different access technologies and operational environments) with a traffic monitor able to apply different sampling policies and rates to the traffic and extract several features both in aggregated and per-flow fashion, providing empirical evidences of the impact of packet sampling on both traffic measurement and traffic classification. First, we analyze feature distortion, quantified by means of two statistical metrics: most features appear already deteriorated under low sampling step, no matter the sampling policy, while only a few remain consistent under harsh sampling conditions, which may even cause some artifacts, undermining the correctness of measurements. Second, we evaluate the performance of traffic classification under sampling. The information content of features, even though deteriorated, still allows a good classification accuracy, provided that the classifier is trained with data obtained at the same sampling rate of the target data. The accuracy is also due to a thoughtful choice of a smart sampling policy which biases the sampling towards packets carrying the most useful information.

[24] X. Tang et al. Characterizing impulsive network traffic using truncated α-stable processes. *IEEE Communications Letters*, 13(12):980-982, Dec. 2009. [ bib | DOI ]

It has been recently recognized that aggregated traffic in a variety of networks exhibits a similar impulsiveness over a wide range of aggregation levels, but approaches a Gaussian distribution in the limit as the aggregation level grows. Although several traffic models have been proposed in the past decade, their accuracy in simultaneously characterizing the above properties still needs to be further improved. In this letter, we propose a truncated #x003B1;-stable process model which is able to capture the impulsiveness of observed network traffic as well as its tendency toward the Gaussian distribution with aggregation. An inherent physical mechanism is also proposed to give insight into the underlying meaning of the proposed model. Simulation results show that the proposed process achieves close agreement with real traffic and outperforms previous models.

Keywords: Gaussian distribution;aggregated traffic;impulsive network traffic;truncated #x003B1;-stable processes;Gaussian distribution;telecommunication traffic;

[25] G. Terdik and T. Gyires. Does the internet still demonstrate fractal nature? In *International Conference on Networks*, pages 30-34, Mar. 2009. [ bib | DOI ]

The self-similar nature of bursty Internet traffic has been investigated for the last decade. A first generation of papers, approximately from 1994 to 2004, argued that the traditionally used Poisson models oversimplified the characteristics of network traffic and were not appropriate for modeling bursty, local-area, and wide-area network traffic. Since 2004, a second generation of papers has challenged the suitability of these results in networks of the new century and has claimed that the traditional Poisson-based and other models are still more appropriate for characterizing todaypsilas Internet traffic. A possible explanation was that as the speed and amount of Internet traffic grow spectacularly, any irregularity of the network traffic, such as self-similarity, might cancel out as a consequence of high-speed optical connections, new communications protocols, and the vast number of multiplexed flows. These papers analyzed traffic traces of Internet backbone collected in 2003. In one of our previous papers we applied the theory of smoothly truncated Levy flights and the linear fractal model in examining the variability of Internet traffic from self-similar to Poisson. We demonstrated that the series of interarrival times was still close to a self-similar process, but the burstiness of the packet lengths decreased significantly compared to earlier traces. Since then, new traffic traces have been made public, including ones captured from the Internet backbone in 2008. In this paper we analyze these traffic traces and apply our new analytical methods to illustrate the tendency of Internet traffic burstiness. Ultimately, we attempt to answer the question: Does the Internet still demonstrate fractal nature?

Keywords: Poisson process;bursty Internet traffic;fractal nature;truncated Levy flight;Internet;stochastic processes;telecommunication traffic;

[26] G. Terdik and T. Gyires. Lévy flights and fractal modeling of internet traffic. *IEEE/ACM Trans. Netw.*, 17:120-129, Feb. 2009. [ bib | DOI ]

The relation between burstiness and self-similarity of network traffic was identified in numerous papers in the past decade. These papers suggested that the widely used Poisson based models were not suitable for modeling bursty, local-area and wide-area network traffic. Poisson models were abandoned as unrealistic and simplistic characterizations of network traffic. Recent papers have challenged the accuracy of these results in today's networks. Authors of these papers believe that it is time to reexamine the Poisson traffic assumption. The explanation is that as the amount of Internet traffic grows dramatically, any irregularity of the network traffic, such as burstiness, might cancel out because of the huge number of different multiplexed flows. Some of these results are based on analyses of particular OC48 Internet backbone connections and other historical traffic traces. We analyzed the same traffic traces and applied new methods to characterize them in terms of packet interarrival times and packet lengths. The major contribution of the paper is the application of two new analytical methods. We apply the theory of smoothly truncated Levy flights and the linear fractal model in examining the variability of Internet traffic from self-similar to Poisson. The paper demonstrates that the series of interarrival times is still close to a self-similar process, but the burstiness of the packet lengths decreases significantly compared to earlier traces.

Keywords: Lévy flights, burstiness, fractal modelling, long-range dependence, network traffic

[27] W. Willinger, M. S. Taqqu, W. E. Leland, and D. V. Wilson. Self-similarity in high-speed packet traffic: Analysis and modeling of ethernet traffic measurements. *Statistical Science*, 10(1):67-85, Feb. 1995. [ bib | DOI ]

Traffic modeling of today's communication networks is a prime example of the role statistical inference methods for stochastic processes play in such classical areas of applied probability as queueing theory or performance analysis. In practice, however, statistics and applied probability have failed to interface. As a result, traffic modeling and performance analysis rely heavily on subjective arguments; hence, debates concerning the validity of a proposed model and its predicted performance abound. In this paper, we show how a careful statistical analysis of large sets of actual traffic measurements can reveal new features of network traffic that have gone unnoticed by the literature and, yet, seem to have serious implications for predicted network performance. We use hundreds of millions of high-quality traffic measurements from an Ethernet local area network to demonstrate that Ethernet traffic is statistically self-similar and that this property clearly distinguishes between currently used models for packet traffic and our measured data. We also indicate how such a unique data set (in terms of size and quality) (i) can be used to illustrate a number of different statistical inference methods for self-similar processes, (ii) gives rise to new and challenging problems in statistics, statistical computing and probabilistic modeling and (iii) opens up new areas of mathematical research in queueing theory and performance analysis of future high-speed networks.

[28] C. Wright, C. Connelly, T. Braje, J. Rabek, L. Rossey, and R. Cunningham. Generating client workloads and high-fidelity network traffic for controllable, repeatable experiments in computer security. In *Recent Advances in Intrusion Detection (RAID)*, volume 6307 of *Lecture Notes in Computer Science*, pages 218-237. 2010. [ bib | DOI ]

> Rigorous scientific experimentation in system and network security remains an elusive goal. Recent work has outlined three basic requirements for experiments, namely that hypotheses must be *falsifiable*, experiments must be *controllable*, and experiments must be *repeatable* and *reproducible*. Despite their simplicity, these goals are difficult to achieve, especially when dealing with client-side threats and defenses, where often user input is required as part of the experiment. In this paper, we present techniques for making experiments involving security and client-side desktop applications like web browsers, PDF readers, or host-based firewalls or intrusion detection systems more *controllable* and more easily *repeatable*. First, we present techniques for using statistical models of user behavior to drive real, binary, GUI-enabled application programs in place of a human user. Second, we present techniques based on adaptive replay of application dialog that allow us to quickly and efficiently reproduce reasonable mock-ups of remotely-hosted applications to give the illusion of Internet connectedness on an isolated testbed. We demonstrate the utility of these techniques in an example experiment comparing the system resource consumption of a Windows machine running anti-virus protection versus an unprotected system.
>
> Keywords: Network Testbeds; Assessment and Benchmarking; Traffic Generation

[29] C. V. Wright, S. E. Coull, and F. Monrose. Traffic Morphing: An efficient defense against statistical traffic analysis. In *Proc. of Network and Distributed System Security Symposium (NDSS)*, Feb. 2009. [ bib | .pdf ]

---

*This file was generated by bibtex2html 1.96.*