

# Mixture Models of Endhost Network Traffic

John Mark Agosta  
Toyota ITC

Jaideep Chandrashekar  
Technicolor Research

Mark Crovella  
Boston University

Nina Taft  
Technicolor Labs

Daniel Ting  
Facebook

**Abstract**—We model a little studied type of traffic, namely the network traffic generated from endhosts. We introduce a parsimonious model of the marginal distribution for connection arrivals consisting of mixture models with both heavy and light-tailed component distributions. Our methodology assumes that the underlying user data can be fitted to one of several models, and we apply Bayesian model selection criterion to choose the preferred combination of components. Our experiments show that a simple Pareto-exponential mixture model is preferred over more complex alternatives, for a wide range of users. This model has the desirable property of modeling the entire distribution, effectively clustering the traffic into the heavy-tailed as well as the non-heavy-tailed components. Also this method quantifies the wide diversity in the observed endhost traffic.

## I. INTRODUCTION

In the last decade or so there has been a tremendous amount of research done in the area of Internet traffic modeling— [4], [14], [19] (to name just a few). This research has predominantly focused on traffic from inside a network. The reason endhost traffic models are so scarce (with some exceptions [10]) is that it is difficult to obtain the raw measurements needed, which requires installing a collection tool directly on each user’s machine, and getting the express consent of users. Nevertheless it is important to obtain a deeper understanding of end user traffic because IT management is driving computing towards self-diagnosis for trouble shooting and user controlled performance tuning.

We obtained end user data via a measurement tool that resides on laptops and thus moves with users and continues to observe network traffic as the user switches between different networks and different environments (e.g., work and home), for a population of 270 enterprise users over five weeks (§III).

Starting with this rich dataset, we focus on modeling end host activity, in particular the rate of flow initiations. A necessary first task is to estimate reliably the probability distribution of flow rates. Modeling heavy-tailed data is a notoriously fraught problem, often approached by just estimating the exponent, known as the scaling parameter of the distribution tail. Commonly used methods for estimating the scaling parameter include the *Hill estimator*, which is tricky since it relies on estimating a cut-off below which the central part of the distribution is disregarded [15]. In [3], the authors highlight the lack of care pervasive in the literature on estimating power laws. In light of this concern, we demonstrate an efficient estimator that uses the entire data set by means of *mixture models*. Hence, our first contribution is a new method (§IV-A) to estimate heavy-tailed scaling parameters.

Since we fit the entire sample, we need a method to choose models, the commonly applied one being goodness-of-fit. The limitation of this approach and their associated  $P$ -values is that they are meant to rule out hypotheses. This is certainly useful for steering data collection, but they do not provide an acceptance criterion. In our situation, with effectively an endless stream of data as a source, any reasonable model will eventually be rejected. Model selection methods give us a quantitative criterion that lets us explore a wider class of models than has hitherto been considered. Thus we do not presuppose a single parametric distribution model; instead we start with a class of *nested* mixture models (i.e. a family of models where one is a subset of another) and use *Bayes Factors* approximated in large samples by *Bayesian Information Criteria* (BIC) to select the best model for a user’s data. Since it is a requirement for Bayes Factors comparisons to compare both models on the full sample data, as a side benefit we produce models that comprise the complete distribution.

Thus our second contribution is a richer, non-parametric class of models for traffic modeling (§V). The strength of this approach is that, over a population of users, the choice of which model is best can be explained statistically.

Our third contribution lies in applying this tool to an extensive, diverse set of endhosts’ traffic data (§VI). We began by observing that distributions of users’ flow arrival counts are monotonically declining from a mode at zero. Preliminary analysis eliminated conventional component distributions (such as, e.g., Gaussian or Poisson), to concentrate on mixtures of heavier tailed exponential and Pareto distributions. Since mixtures of exponentials constitute a very flexible framework, restricting to these two distribution classes is a good approximation to the properties of the samples. We find that the majority of the endhost population can be described by a two component model whose diversity is expressed by the revealed distribution over users of estimated parameters.

## II. RELATED WORK

Heavy tailed statistics have been documented in numerous phenomena in network traffic; in the popularity of web pages [2], in traffic demands [8]; in network topology [15], in TCP inter-arrival times [7], in wireless LAN traffic [16], and many others. The seminal work by Leland et al. [14] studied LAN traffic and convincingly demonstrated that actual network traffic is self-similar or long-range dependent in nature (i.e., bursty over a wide range of time scales). Our work differs by revealing the distribution of models over users rather than

aggregating all users. Secondly, we observe the power law nature of traffic in the first-order statistics of traffic rates, rather than in the second-order autocorrelation properties. For a more detailed comparison to prior art, we refer the reader to a longer version of the paper at [1].

The idea of using mixture models for Internet traffic has been proposed in other contexts before [9]. That work proposes using hyperexponential models as a tractable way to approximate a heavy-tailed distribution. Our work instead does not assume the presence of a heavy tail, and instead uses mixture models to extend the range of models to consider.

### III. DATASET DESCRIPTION

The dataset consists of traces collected at 270 enterprise end-hosts (90% laptops) over a period of approximately 5 weeks. Each host was associated with a unique user for the entire trace collection period, and ran a corporate standard build of Windows XP that included a number of enterprise IT applications.

Packet-level traces were collected *on* the end-hosts, providing a longitudinal view of the traffic even as they moved in and out of the network and across interfaces (wired and wireless). The trace logging software included a wrapper around WinDump to log only packet headers. It also tracked changes in IP address or interface, restarting the trace collection as required. The logged data was uploaded opportunistically a few times a day to a central server (the logging was paused during the upload). Overall, we obtained over 400 Gb of packet traces, which were then converted into flows using BRO [18].

The starting time of each flow generates a point process in continuous time that we bin over non-overlapping, fixed size time-windows to create a time series for each user. Each user trace was binned for 8 different window sizes, starting at 4 seconds, and increasing in multiples of 2, up to 512 seconds. Each bin contains a count of the new flow arrivals. The *flow count* events within each time-window or *bin* are the random variables modeled in this work. In our datasets the median sample size was 9771 intervals, and the maximum was 264,000. Zeros could occur in bins because the host was turned off (or asleep), or else the host was disconnected from the network during that bin. We filter out all such bins and in the resulting data, we see zeros only because there were no flows originated in that bin (and the machine was turned on). That being said, we model the flow events when the counts are nonzero since our goal is to characterize the distribution of active traffic.

### IV. METHODOLOGY

#### A. Mixture Models with Heavy Tails

A probability *mixture model* is a convex combination of probability densities. A mixture model can be thought of as a hierarchical model where the mixing weights determine the probability of each of the component models, which in turn generate the sample. Since all components share the same support, any sample point could in principle have been generated by any component, but with possibly vanishingly

small probability. Such models are familiar in the Statistics literature, [6] [11] and have become a mainstay in the machine learning community [12].

A finite mixture model’s probability density is defined by  $k$  component densities,  $f_i(x)$ , and mixture fractions  $m_i$ , with parameters  $\mathbf{m}, \theta$  as given by:

$$f(x | \mathbf{m}, \theta) = \sum_{i=1}^k m_i f_i(x | \theta_i), \quad (1)$$

$$\text{s.t. } \sum_{i=1}^k m_i = 1, m_i > 0.$$

where the  $\theta_i$  are the component parameters, and  $\mathbf{m} = m_1 \dots m_k$ .

We consider the following nested family of models: a Pareto only model labeled (P), a mixture of one exponential and one Pareto (EP), and a mixture of two exponentials and one Pareto (EEP). The “pure power-law” model we fit is

$$f(x | \alpha) = Cx^{-\alpha} = \frac{1}{\zeta(\alpha, x_{min})} x^{-\alpha}, x \in \mathcal{N} \quad (\text{P})$$

$$\zeta(\alpha, x_{min}) = \sum_{n=0}^{\infty} (n + x_{min})^{-\alpha}. \quad (2)$$

where  $x$  takes on positive integer values, for which we use the discrete version of the Pareto density (referred to also as the *Zeta* distribution).

The exponential–Pareto model is defined as

$$f(x | \mathbf{m}, \lambda_1, \alpha) = m_1 \lambda_1 e^{-\lambda_1 x} + (1 - m_1) Cx^{-\alpha}. \quad (\text{EP})$$

The mixture variable adds another degree of freedom, revealing the relative contributions of the components.

The two exponential–Pareto mixture density model is:

$$f(x | \mathbf{m}, \lambda_1, \lambda_2, \alpha) = m_1 \lambda_1 e^{-\lambda_1 x} \quad (\text{EEP})$$

$$+ m_2 \lambda_2 e^{-\lambda_2 x}$$

$$+ (1 - m_1 - m_2) Cx^{-\alpha}.$$

We were motivated originally to consider this set of models because visual exploration of the data showed traffic flow distributions with a mode left-most, then a monotone decrease, with a linear segment on a semi-log plot in the dense part of the distribution, followed by a long, heavy tail.

The intent behind using a family of models is to capture the diversity across users. The EEP model is capable of fitting any combination of the 3 component distributions, although in practice we almost always see a heavy-tailed component. In terms of degrees of freedom, these are very parsimonious models; the EP model has 3 parameters, and the EEP has only 5.

#### B. Estimating Model Parameters

Desirable properties of maximum likelihood estimation (MLE) recommend its use to estimate model parameters. Besides being asymptotically efficient, if the model does contain the true data generating distribution and is differentiable in

quadratic mean (DQM) [20], the MLE converges to the true parameters at a rate  $O(1/\sqrt{n})$ . If not, the MLE still converges to the best approximation to the true distribution within the model’s constraints at the same rate.

Instead of a conventional Expectation-Maximization (EM) methods, we solved for the MLE as constrained optimization problem using an interior point method to enforce the constraints on the model parameters. We found EM converged slowly, probably due to the common mode of the components. The method uses a concave barrier function that decreases steeply to  $-\infty$  at the boundary of the constraint set, preventing estimates from violating constraints and making it amenable to unconstrained solution methods. The weight is reduced on each iteration until the barrier becomes negligible. These unconstrained problems are solved using the `optim()` function in the statistical package **R**, which implements a Quasi-Newton optimization method. To exclude bad solutions, we also added constraints  $\alpha < 4$  and  $\lambda < 3.5$  so that the component parameters do not grow too large. Since the mixture model typically contains local optima, we performed the optimization multiple times with random initializations to find the global maximum.

### C. Model Selection

Given multiple probability models for the same sample, *model selection* uses a comparative metric called a *Bayes Factor (BF)* as a means of comparing which model is more probable. In practice the estimated mixing weights will find the correct model: Components with insubstantial weights can be ignored, leaving only desired components. Model selection, in addition reveals the strength of the comparison, and can be applied generally, not only to mixture models. Our explanation of model selection borrows extensively from Kass and Raftery [13]. This can be understood using the odds ratio form of Bayes rule, where the posterior odds—the ratios of posteriors—between two models, is expressed as the product of the *BF* and the prior odds. So, for example to compare the model  $\mathcal{M}_P$  to the proposed model  $\mathcal{M}_{EP}$ , the posterior odds will be

$$\frac{P(\mathcal{M}_{EP} | \mathcal{D})}{P(\mathcal{M}_P | \mathcal{D})} = \frac{P(\mathcal{D} | \mathcal{M}_{EP})}{P(\mathcal{D} | \mathcal{M}_P)} \frac{P(\mathcal{M}_{EP})}{P(\mathcal{M}_P)} \quad (3)$$

where the middle term in this equation, the Bayes factor, *BF*, is defined as the ratio of marginal likelihoods:

$$BF_{EP,P} = \frac{P(\mathcal{D} | \mathcal{M}_{EP})}{P(\mathcal{D} | \mathcal{M}_P)} \quad (4)$$

The larger *BF*, the greater the weight of evidence for the EP model. As for the prior term, an unprejudiced rule implies equal model priors, in which case the Bayes Factor and the posterior odds-ratio are equal. This criterion is similar to the maximum likelihood ratio, but rather than taking the probability at the maximum, one integrates over the range of model parameters  $\theta$ , resulting in a correction for the degrees of freedom of the models. Adding more parameters to a model and thus increasing its degrees of freedom can

Odds	$\log_{10}(BF)$	$\log(BF)$	Strength of comparison
20:1	1.3	3	“substantial”
100:1	2	4.5	“strong”
1000:1	3	7	“decisive”

TABLE I: Interpretation of Bayes Factor strengths

only increase the likelihood at the maximum but does not necessarily improve marginal likelihood. This criterion trades off simplicity with accuracy—a built-in “Occam’s Razor.”

### D. Interpreting The Weight of Evidence

Interpreting the magnitude of a *BF* is commonly done by considering the ratio as an odds ratio, e.g., odds of 20 to 1 in favor of the model in the numerator corresponds to a  $BF = 20$ , or, using natural logs, to  $\log BF \simeq 3$ . Table I shows a standard convention [13] that we adopt for interpreting the strength of Bayes Factors, with their suggested labels. For comparison between E and EP models, we give precedence to the conventional model, and hence require an log odds-ratio significantly greater than zero—we use 10—which is well into the “decisive” range, corresponding to an odds ration of greater than 20,000. If the EP model is selected, then we compute  $\log BF_{EEP,EP}$ . Again, if this factor is above 10, then EEP is selected, otherwise the final choice is EP. Of course, the test is symmetric and the ratio may be expressed either way. A negative  $\log BF_{EP,P}$  would be evidence *against* the EP model, in favor of P.

### E. Approximation by BIC

In practice the integral implied by  $P(\mathcal{M} | \mathcal{D}, \theta)$  requires a prior over the  $\theta$ . Recall these parameters are bounded, so the integral is equivalent to setting a proper, uniform prior over their range. Since with such large samples likelihood values are strongly peaked around their maximum and numerical integration works poorly, we use a common approximation to *BF*.

With large samples, *BF* is approximated by the Bayes Information Criterion (BIC). BIC is often presented as a correction to maximum log likelihood to account for the degrees of freedom of a model. BIC is defined as

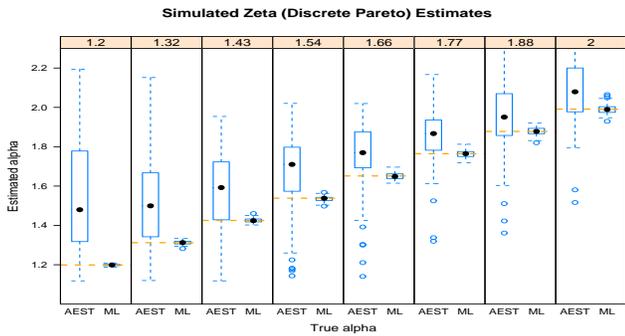
$$BIC = \log P(\mathcal{D} | \mathcal{M}, \hat{\theta}) - \log(N) \cdot d/2 \quad (5)$$

where  $N$  is the sample size and  $d$  is the numbers of parameters in the model. In our experimental work we computed both Laplace approximations and BIC corrections and found to our satisfaction that they agreed with each other to within a fraction of a percent on the dataset.

With the BIC approximation, the log Bayes Factor becomes  $\log BF_{EP,P} = BIC_{EP} - BIC_P$

## V. VALIDATION

We validated our model-fitting and selection method by first showing that the estimates produced are accurate, in comparison to a widely used power-law tail fitting procedure. We used synthetic data from an EP mixture that is of a



**Fig. 1:** With synthetic Pareto-tailed data over  $1 < \alpha < 2$  an EP mixture model estimator performs accurately, and with less variance, than the `aest` method.

type covered by both procedures, where the true value of the parameters of the generating data is known.

The tail-fitting method used for comparison is a widely used tool for estimating the  $\alpha$  parameter of  $\alpha$ -stable distributions, based on a scaling property of sums of heavy-tailed random variables [17]. We used a publicly available implementation, called `aest` [5]. We see in Fig. 1 that the range of  $\hat{\alpha}$ 's in the columns subtitled “ML” for the mixture model estimates, is within a few percent of the true value, unlike the `aest`  $\hat{\alpha}$ 's that have high variance and bias. See [1] for details.

Next we validate that model selection by pair-wise comparison of BIC scores does indeed select the right model. Since the EEP model subsumes the other two, the model with more parameters will always fit better, so the model choice is driven by the penalty due to the BIC penalty term.

The test data consisted of pseudo-random samples with known parameters  $\hat{\alpha}, \hat{m}_{\{1,2\}}, \hat{\lambda}_{\{1,2\}}$ , generated from each of the three models, P, EP and EEP. We ran 100 test cases over a range of sample sizes from 500 to 20,000 points, in the style of an empirical “design of experiments” to find what sample sizes were necessary to show adequate model selection results. We ran 3 pair-wise comparisons: EP vs. P on EP data, EEP vs EP on EP data and EEP vs EP on EEP data.

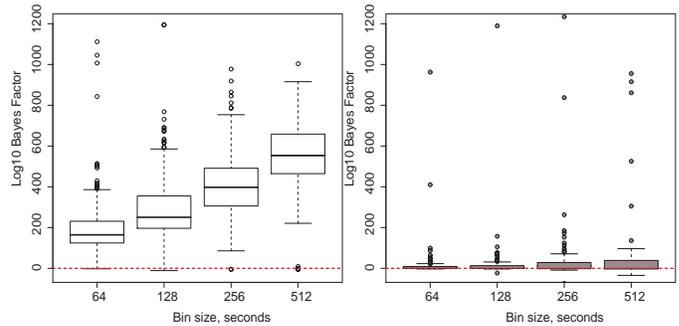
In Table II, we summarize the ability of our model selection method to distinguish the 3 hypotheses. For each test, we state the number of samples and the Bayes Factor level so achieved, using the conventions *substantial*, *strong*, or *decisive* in Table I. For the first two tests we list sample sizes for two levels.

The more complicated the model comparison, the larger the sample required for the same strength of differentiation. In short, the EP model can be selected “substantially” with traces of no less than about 1000 samples. The EEP model requires about 10 times the sample to be selected at the same level. This is reason to believe that requiring samples on the order of a few thousand (or at most 10,000) is a fairly light requirement compared to the typical size of our sample traces.

For sake of comparison, the Results reveal that the actual Bayes Factors computed on the data have values ranging in the hundreds, with sample sizes in the thousands and tens of thousands—clearly at the “decisive” level, and orders of magnitude larger than seen in these validation tests!

Truth	Model Choice:	Min Number Samples	$\log_{10}$ BF strength
EP	EP vs. P	1000 5000	substantial decisive
EP	EEP vs. EP	1000 10,000	substantial strong
EEP	EEP vs. EP	9000	substantial

**TABLE II:** Sample sizes and the strength of comparison they achieve with simulated data, for different model comparisons.



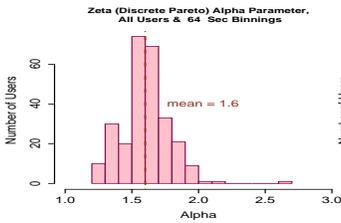
**Fig. 2:** Boxplot of BIC for P vs. EP models.

**Fig. 3:** Boxplot of BIC comparison for EP vs. EEP models.

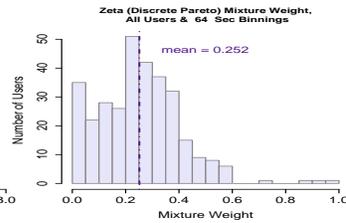
## VI. RESULTS

**Choice of Models:** We use our methodology to select the best model for each of our 270 users. In Fig. 2 we show a box-plot of the log of the Bayes Factor (or difference in BICs) of the P and EP models against bin-size on the x-axis. We can see that for nearly all users we can select the two component EP model as ‘decisive’, according to Table 2. There are a very small number of users—roughly a dozen— whose  $\log BF_{EP,P}$  was near zero, suggesting a Pareto-only model. Not only is the two component mixture model EP preferred for all the other users, but it is strongly preferred as evidenced by the high Bayes factor values. We observe a small trend here in that as the bin sizes increase, the log Bayes factor ratio gets larger, indicating that for larger bin sizes, the exponential component plays an increasingly dominant role. Next we compare the EP and EEP models. Fig. 3 plots the  $BF$  distribution for all users, for each of 4 bin sizes. Interestingly, we see that at bin sizes of 64 and 128, the Bayes factors are close to zero for the majority of the users. Since the two models are fairly indistinguishable here, we again select the model of lower complexity, namely EP for nearly all the users save a few outliers. At larger bin sizes, we do see some users for whom the EEP model is selected. Overall, our method assigns the EEP model to roughly 30% of the users and the EP model to the remaining 70%.

The percentage of users that were assigned a given model depends upon the bin size. However, we see clear trends. The fraction of users assigned a Pareto-only model was always less than 5%, the fraction assigned an EP model varied from 50-85% and the fraction assigned an EEP ranged from 15-40%. We conclude two things from this section. First, the flexibility we have built into our methodology is important and needed because the best model for one endhost is not necessarily the same for another endhost. Second, for the



**Fig. 4:** Histogram of estimated  $\alpha$  values across users.



**Fig. 5:** Histogram of estimated  $m_\alpha$  across users.

majority of the endhosts, the mixture model consisting of one exponential and one Pareto is clearly the preferred model.

**User Behavior:** As indicated in §II, there is a growing interest in understanding the range of variation of user behavior. We now look at some model details to explore the range of parameters selected across users, and the amount of mixing between the two model components. We computed an EP model for all our users, and examined the resulting  $\alpha$  and  $\lambda$  values. We first observed that there is little correlation between  $\alpha$  and  $\lambda$  values within the set of endhost EP models. This is reassuring, as it indicates that the fitting process does not introduce dependencies between the two component distributions, and that properties of one distribution do not affect the other.

In Fig. 4 we show the histograms of  $\alpha$  values over users for a bin size of 64 sec. We see that the values of  $\alpha$  range from 1.3 to 2.3 across the users; different users have very different properties in terms of the heaviness of the tail of the distribution. Roughly 1/6 of our users have  $\alpha < 1.5$  implying a fairly heavy tail, while most users have  $\alpha$  values around 1.6 or 1.7. It is interesting that we do have a small number users (4) with  $\alpha > 2$  indicating a finite second moment.

We now look more closely at the user mix of the two components of the model. A value of  $m$  close to 0 implies that the model is dominated by the exponential distribution, (when  $m = 0$  there is no Pareto component in the model). Similarly when  $m$  is close to 1 the Pareto component dominates the behavior of the model. To see the range of  $m$  values chosen across our users, we provide a histogram of this mixing factor in Fig. 5. The frequency on the y-axis denotes the number of users whose  $m$  parameter is that indicated on the x-axis. Only 3 users picked an  $m$  very close to 1, indicating that the pure Pareto model suites practically none of our users—in agreement with the Bayes Factors conclusions. Most of the users have an  $m$  parameter less than 0.4, and roughly half of our users had  $m < 0.25$  indicating the dominance of the exponential component in the model. The  $m$  values are fairly well spread across the range 0 to 0.5 (roughly). We can also interpret this range of  $m$  as a indication of user diversity, in that their mixing fractions differ substantially.

## VII. CONCLUSION

To the best of our knowledge, this is the first paper to study heavy tails of traffic from endhosts, and to study heavy-tailed network traffic using mixture models, employing model

selection. We have shown strong evidence that the rate of initiation of flows in end host traffic, over a variety of users, is almost always heavy-tailed. The scaling parameter varies widely, between 1.0 and 2.0, and on average the heavy tailed component makes up about one quarter of the traffic. We demonstrated that a model selection approach using the Bayesian Information Criteria (BIC) rather than a goodness-of-fit test, applied to a family of mixture models is both accurate and versatile. We showed that this versatility was needed to yield good models for all 270 diverse users. This underscores the value of a method that does not presuppose a single distribution model for flow traffic.

**Acknowledgements**— We wish to thank Eve Schooler and others at Intel who were closely involved in the collection of the data. Most of the work in this paper was carried out when three of the authors were employed at Intel.

## REFERENCES

- [1] AGOSTA, J. M., CHANDRASHEKAR, J., CROVELLA, M., TAFT, N., AND TING, D. Mixture models of endhost network traffic. *arXiv 1212.2744 [cs.NI]* (2012).
- [2] BRESLAU, L., CUE, P., CAO, P., FAN, L., PHILLIPS, G., AND SHENKER, S. Web caching and zipf-like distributions: Evidence and implications. In *In INFOCOM* (1999), pp. 126–134.
- [3] CLAUSET, A., SHALIZI, C. R., AND NEWMAN, M. E. J. Power-law distributions in empirical data. *SIAM Review* (2009).
- [4] CROVELLA, M. E., AND BESTAVROS, A. Self-similarity in world wide web traffic: evidence and possible causes. *IEEE/ACM Trans. on Networking* 5, 6 (1997).
- [5] CROVELLA, M. E., AND TAQQU, M. S. Estimating the heavy tail index from scaling properties. In *Methodology and Computing in Applied Probability* (1999).
- [6] EVERITT, B. S., AND HAND, D. J. *Finite Mixture Distributions*. Chapman and Hall, London, 1981.
- [7] FELDMANN, A. *Self-Similar Network Traffic and Performance Evaluation. Chapter 2*. John Wiley & Sons, New York, 2002.
- [8] FELDMANN, A., GREENBERG, A., LUND, C., REINGOLD, N., REXFORD, J., AND TRUE, F. Deriving traffic demands for operational ip networks: Methodology and experience. *IEEE/ACM Transactions on Networking* 9 (2001), 265–279.
- [9] FELDMANN, A., AND WHITT, W. Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. In *Proceedings of IEEE INFOCOM'97* (April 1997).
- [10] GIROIRE, F., CHANDRASHEKAR, J., IANNACCONE, G., PAPAGIANNAKI, K., SCHOOLER, E., AND TAFT, N. The Cubicle Vs. The Coffee Shop: Behaviora Modes in Enterprise End-Users. *PAM* (2008).
- [11] J. M. MARIN, K. M., AND ROBERT, C. Bayesian modelling and inference on mixtures of distributions. Tech. rep., CEREMADE, Universite Paris Dauphine, February 2004.
- [12] JORDAN, M. I., AND JACOBS, R. A. Hierarchical mixtures of experts and the em algorithm. *Neural Computation* 6 (1994), 181–214.
- [13] KASS, R. E., AND RAFTERY, A. E. Bayes factors. *Journal of the American Statistical Association* 90, 430 (1995), 773–795.
- [14] LELAND, W. E., TAQQ, M. S., WILLINGER, W., AND WILSON, D. V. On the self-similar nature of Ethernet traffic. In *ACM SIGCOMM*.
- [15] LI, L., ALDERSON, D., WILLINGER, W., AND DOYLE, J. C. A First-Principles Approach to Understanding the Internet’s Router-level Topology. *Proc. ACM SIGCOMM* (2004).
- [16] LUO, S., LI, J., PARK, K., AND LEVY, R. Exploiting Heavy-Tailed Statistics for Predictable QoS Routing in Ad-Hoc Wireless Networks. *IEEE Infocom* (2008).
- [17] PAPAGIANNAKI, K., TAFT, N., AND DIOT, C. Impact of flow dynamics on traffic engineering design principles. In *INFOCOM* (2004).
- [18] PAXSON, V. Bro: A system for detecting network intruders in real-time. *Computer Networks* (1999).
- [19] PAXSON, V., AND FLOYD, S. Wide-area traffic: the failure of poisson modeling. In *SIGCOMM* (1994).
- [20] VAN DER VAART, A. W. *Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press, June 2000.