A Weibull Mixture Model to characterise end-to-end Internet delay at coarse time-scales

José-Alberto Hernández and Iain W. Phillips

July 5, 2005

Abstract: Traces collected at monitored points around the Internet contain representative performance information about the paths their probes traverse. Basic measurement attributes, such as delay and loss, are easy to collect and provide a means to both build and validate empirical performance models. However, the task of analysis and extracting performance conclusions from measurements remains challenging.

Ideally, performance modelling aims to find a set of self-contained parameters to describe, summarise, profile and easy display network performance status at a time. This can result in the provision of meaningful information to address applications in fault and performance management, hence providing input to network provisioning, traffic engineering and performance prediction.

In this work we present the Weibull Mixture Model, a method to characterise end-to-end network delay measurements within a few simple, accurate, representative and handleable parameters using a finite combination of Weibull distributions, with all the aforementioned benefits. The model parameters are related to meaningful delay characteristics, such as average peak and tail behaviour in a daily profile, and can be optimally found using an iterative algorithm known as Expectation Maximisation. Studies on such parameter evolution can reflect current workload status and all possible network events impacting packet dynamics, with further applications in network management.

In addition, a self-sufficient procedure to implement the Weibull Mixture Model is presented, along with a set of matching examples to real GPS synchronised measurements taken across the Internet, donated by RIPE NCC.

Keywords: Performance Modelling, Weibull Mixture Model, Expectation Maximisation, Management Applications, RIPE NCC.

1

1 Introduction

Both users and operators agree on the need to keep a network operating under its maximum achievable performance; that is, to offer the best possible quality of service to the customers, yet keeping the minimum cost to providers. However, before optimising network performance, it is necessary to describe it.

The IP Performance Measurement working group of the IETF has recognised packet delay as one of the basic metrics to characterise network performance, since it provides representative information of the state of health of a given path in the Internet. However, the delay dynamics are specified in a percentile-basis which, for certain applications, does not permit extracting useful information, and does not give a parsimonious point of view of network delay. It is rather convenient to find a model which matches the delay histogram at all percentiles and that is easily and accurately be described by a few set of parameters. However, previous work has acknowledged the difficulty of this task and revealed it is in fact challenging, if not impossible [1].

For instance, some studies have tried to match delay histograms using different probability distributions, which are similar in shape and properties to the empirical delay distribution observed from measurements. These include the use of the Shifted Gamma distribution [2], the Weibull distribution [3] and a truncated version of the Gaussian distribution [4]. However, most of these works have reported inaccurate results, due to the variability and complexity of the problem under study.

Nevertheless, such works do not mention the use of multiple probability distributions. This work presents a new model to characterise end-to-end packet delay using a finite mixture of Weibull distributions, that overcomes the limitations of single distribution approaches.

The motivation for using a mixture of Weibull distributions are manifold:

- The Weibull distribution belongs to the so-called group of heavy-tailed distributions, often found in the modelling of other networking and computer events [5].

- Previous work have acknowledged the suitability of using the Weibull distribution in modelling packet delay [3].

- The Weibull distribution approximates the queue length distribution of a router fed by fractional Brownian motion [6]. Fractional Brownian motion is a statistical model widely used in the modelling of self-similar and long-range dependent processes, such as network traffic [7, 8, 9, 10].

- Packet delay has been shown to be accurately characterised by a single Weibull

distribution in a single-hop scenario [3] and has also been analytically shown to describe the dynamics of end-to-end delays [11].

- The Weibull distribution is completely determined by two parameters only, one regards to most likely delay, and the other is concerned with the heaviness of the distribution tail. Thereby, the examination of the two parameters can easily provide insight in the fundamental characteristics of the stochastic delay experienced by packets in a given network path: average and variability.

This work introduces the Weibull Mixture Model, a method to characterise end-to-end delay using a finite combination of Weibull distributions, using the Expectation Maximisation algorithm [12]. The EM algorithm defines a two-step procedure to find the maximum likelihood estimate of an underlying probability distribution from a given sample in an iterative way. EM has traditionally been applied to mixtures of Gaussian distributions [13], but in this work it shall be used for a mixture of Weibull distributions.

The remainder of this work is organised as follows: In section 2, we introduce the EM algorithm and develop the equations defined by EM for a mixture of Weibull distributions. Section 3 briefly reviews the methodology to validate the model against measurements. Section 4 constitutes the experiments carried out with a set real network delay measurements, gently donated by the "Test Traffic Measurement" project at the RIPE NCC institution[1], and potential applications of this work. Finally, section 5 addresses conclusions and discussion.

2 The Weibull mixture model

2.1 Introduction

The Weibull mixture model assumes packet delays are distributed as a finite combination of Weibull distributions. The Weibull probability density function (PDF) is given by:

$$p(x|r,s) = \frac{sx^{s-1}}{r^s} \exp(-(x/r)^s), \quad x \geq 0, \quad r,s > 0 \tag{1}$$

where r and s refer to the scale and shape parameters respectively, which determine its structure and statistics 1. Varying the values of these two parameters highly impacts the appearance of the Weibull distribution, as depicted in figure 1. As shown, parameter r

---

[1]The RIPE NCC website: http://www.ripe.net/ttm

is closely related to the distribution peak (fig. 1, top) and parameter s is concerned with the tail behaviour (fig. 1, bottom).

[Figure 1 about here.]

Additionally, the two parameters determine the value of first- and second-order statistics, given in table 1.

[Table 1 about here.]

Notice that when $s \gg 1$, the equation $\frac{s+1}{s} \approx 1$, hence mode and mean are close to the value of parameter r. Thus, parameter r represents the values where the data are mainly located (the distribution peak). On the other hand, the parameter s is concerned with the rate at which the exponential portion of the Weibull distribution, that is, $\exp(-(x/r)^s)$, decays.

Despite its benefits, the use of a single Weibull distribution does not provide enough flexibility to accurately match real end-to-end delay histograms. For this reason, the use of a finite weighted M-sized combination of Weibull probability distributions shall be considered, namely:

$$p(x|q,r,s) = \sum_{j=1}^{M} q_j p(x|r_j,s_j) \tag{2}$$

where q_j represents the weight of the j-th Weibull component, which characterised by its own scale and shape parameters (r_j, s_j). Obviously, $\sum_j q_j = 1$.

Concluding, the model is totally characterised by the $3 \times M$ set of parameters, Θ:

$$\Theta = [q_1, r_1, s_1, \ldots, q_M, r_M, s_M] \tag{3}$$

representing weight, q_j, rough average r_j and tail s_j behaviour of all of its components. The problem now moves to obtaining an estimate of the optimal model parameters $\hat{\Theta}^*$ for a given set of delay measurements $\mathbf{x}$ collected by the network administrator. In other words:

Problem statement: Given a set of N measurements $\mathbf{x} = [x_1, \ldots, x_N]^T$ that constitutes the delay experienced by packet probes for a certain link, the modelling problem involves finding the optimal set of $3 \times M$ parameters, $\hat{\Theta}^*$, that best fits the measurement histogram.

This implies maximising the log-likelihood function, that is:

$$\log L(\Theta|\mathbf{x}) = \log \prod_{i=1}^{N} \sum_{j=1}^{M} q_j p(x_i|r_j, s_j) = \sum_{i=1}^{N} \log \Big(\sum_{j=1}^{M} q_j p(x_i|r_j, s_j) \Big) \qquad (4)$$

which encompasses computing the sum of a logarithm of sums, and often very hard to obtain analytical expressions by traditional Maximum Likelihood procedures.

To accomplish a tractable solution, a data augmentation step is necessary [14].

2.2 Introducing hidden labels

Let us introduce a second random variable, z, which represents the probability that a data sample comes from a single component of the mixture-model. That is, $p(z_i = j|x_i, \Theta)$ represents the probability of the measurement value x_i to be drawn from the j-th Weibull distribution for a particular model Θ. For this reason, the set of augmented data $\{z_i\}_{i=1}^{N}$ is often referred as to the "labels".

It is straightforward that, for a given value x_i, if its original component $z_i = j$ is known, then $p(x_i|z_i = j, \Theta) = p(x_i|r_j, s_j)$. Also, the component weights can be reformulated as $p(z = j|\Theta) = q_j$.

In fact, this notation agrees with the total probability theorem that establishes:

$$p(x|\Theta) = \sum_{j} p(x, z = j|\Theta) = \sum_{j} p(x|z = j, \Theta)p(z = j|\Theta) = \sum_{j=1}^{M} q_j p(x|r_j, s_j) \qquad (5)$$

The new complete log-likelihood function becomes:

$$\log L(\Theta|\mathbf{x}, \mathbf{z}) = \log \prod_{i=1}^{N} q_{z_i} p(xi|r_{z_i}, s_{z_i}) = \sum_{i=1}^{N} \log q_{z_i} p(xi|r_{z_i}, s_{z_i}) \qquad (6)$$

It is worth remarking that by introducing the auxiliary labels z, we have transformed a sum of logarithms of a sum (see eq. 4) into a more convenient sum of logarithms (see eq. 6). However, the second formulation has the missing information z handicap, which needs be estimated separately. To proceed, the Expectation Maximisation shall be used [14, 13].

2.3 The Expectation Maximisation algorithm

The EM algorithm is a method for finding the maximum-likelihood estimate of an underlying distribution from a given sample in an iterative way. It is especially useful

when part of the information is hidden or have missing values making the task of estimation particularly difficult. Such missing information regards to the labels $\{z_i\}_{i=1}^N$ for this particular problem.

Essentially, the EM algorithm defines two steps repeated iteratively until convergence: (1) an expectation step (E-step), where the expectancy of the complete log-likelihood function (eq.) is calculated with respect to the observed information $\mathbf{x}$ and the current state of the model parameters estimate $\Theta^{(t)}$; and (2) a maximisation step (M-step) where the previous amount is maximised with respect to the parameters Θ, see:

$$
\begin{aligned}
\text{E-step:} \quad Q(\Theta, \Theta^{(t)}) &= E[\log L(\Theta|\mathbf{x}, \mathbf{z})|\mathbf{x}, \Theta^{(t)}] & (7)\\
\text{M-step:} \quad \Theta^{(t+1)} &= \arg\max_{\Theta} Q(\Theta, \Theta^{(t)}) & (8)
\end{aligned}
$$

After a certain number of iterations (from 10 to 100 depending on the particular problem), the algorithm converges to a local maximum in the likelihood function, giving the optimal set of parameters, Θ_{ML}.

Expanding E-step and taking separate derivatives with respect to the different parameters (M-step) lead to the following formulation:

$$
q_j^{(t+1)} = \frac{1}{N} \sum_{i=1}^{N} p(z_i = j | x_i, \Theta^{(t)}) \tag{9}
$$

$$
r_j^{(t+1)} = \Big[\frac{\sum_{i=1}^{N} x^{s_j^{(t)}} p(z_i = j | x_i, \Theta^{(t)})}{\sum_{i=1}^{N} p(z_i = j | x_i, \Theta^{(t)})}\Big]^{1/s_j^{(t)}} \tag{10}
$$

$$
s_j^{(t+1)} = \frac{\sum_{i=1}^{N} p(z_i = j | x_i, \Theta^{(t)})}{\sum_{i=1}^{N} \left(\left(\frac{x_i}{r_j^{(t)}}\right)^{s_j^{(t)}} - 1\right) \log\left(\frac{x_i}{r_j^{(t)}}\right) p(z_i = j | x_i, \Theta^{(t)})} \tag{11}
$$

To complete the iterative circle, the probability $p(z_i = j | x_i, \Theta)$ can be calculated using the Bayes' rule as:

$$
\begin{aligned}
p(z_i = j | x_i, \Theta^{(t)}) &= \frac{p(x_i, z_i = j | \Theta^{(t)})}{p(x_i | \Theta^{(t)})} = \\
&= \frac{p(x_i | z_i = j, \Theta^{(t)}) p(z_i = j | \Theta^{(t)})}{\sum_{k=1}^{M} p(x_i | z_i = j, \Theta^{(t)}) p(z_i = j | \Theta^{(t)})} = \\
&= \frac{q_j p(x_i | r_j, s_j)}{\sum_k q_k p(x_i | r_k, s_k)}
\end{aligned}
\tag{12}
$$

Table 2 summarises the procedure in a batch way.

[Table 2 about here.]

3 Model validation

The hypothesis of an underlying probability distribution or combination of probability distributions on a measured data must be adequately checked, tested and validated. To do so, we have used both visual matching techniques and quantitative techniques to measure the matching accuracy of the model.

Visual techniques whether a model "looks" suitable to a given data, and can help assessing if a given model should be accepted or rejected in a first instance. These include comparing the histogram against the Weibull mixture model, the same plot in log-log scale and a quantile-quantile plot [15].

In addition to qualitative techniques, we have used the λ^2 test [16] to give insight in the accuracy of the Weibull mixture model. Essentially, the λ^2 test is derived from the χ^2-test, but slightly modified to overcome its limitations.

Given a histogram obtained from data set $\mathbf{x}$, the χ^2 methods computes a discrepancy quantity, named D^2, which measures, for every histogram bin, the difference between the number of observations falling in its range, and the expected number of observations that should fall in according to the proposed model. That is, for the k-th bin, the discrepancy should be:

$$
D_k^2 = \frac{(o_k - e_k)^2}{e_k}
\tag{13}
$$

where o_k refers to the true number of observations falling in the k-th histogram bin, and e_k regards the expected number of observations according to the model. That is, for an N-sized histogram, such number of expected observations is $e_k = N \int_{k-\text{th bin}} p(x|\Theta) dx$.

The χ^2-metric comprises the sum of discrepancies for all the histogram bins, that

7

is, $D^2 = \sum_k D_k^2$. Then, if the Weibull mixture model is correct, D^2 is χ^2 distributed with $K - 3M - 1$ degrees of freedom, with M being the number of Weibull components in the mixture. Hence, for a given significance level α (typically $\alpha = 0.05$), the model is accepted if:

$$D^2 < \chi_{\alpha^2, K-3M-1} \tag{14}$$

However, the χ^2 test can output very different results depending on the histogram's bin-size chosen. To overcome this limitation, previous work [16, 17] suggests the use of a related discrepancy measure, named λ^2, which aims to correct the χ^2 value to achieve bin-size insensitiveness. The expression for λ^2 is:

$$\lambda^2 = \frac{\chi^2 - \xi - 3M}{N - 1} \tag{15}$$

where ξ is computed as $\xi = \sum_k \frac{o_k - e_k}{e_k}$. In other words, ξ is the same equation as χ^2 except of the numerator is not squared.

The λ^2-test also provides means to compare discrepancies of models. That is, for a given data sample $\mathbf{x}$, a model Θ_1 better fits it than a model Θ_2 if $\hat{\lambda}^2_{\Theta_1} < \hat{\lambda}^2_{\Theta_2}$. This criteria shall be used to compare whether an M_1-sized model is more accurate than a M_2-sized one.

In addition to this, the model shall also be tested on its capability to not only match histograms at every point, but also to accurately estimate first- and second-order moments, sucha as mean and variance. Data-based estimates are:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{16}$$

$$\sigma_x^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2 \tag{17}$$

The estimates of the two above, based on the model parameters are given by:

$$\bar{x} = \sum_j q_j \bar{x}_j = \sum_j q_j r_j \Gamma(\frac{s_j + 1}{s_j}) \tag{18}$$

$$\sigma_x^2 = \sum_j q_j (\sigma_j^2 + \bar{x}_j^2) - \bar{x} = \sum_j q_j r_j \Gamma(\frac{s+2}{s}) - \left(\sum_j q_j r_j \Gamma(\frac{s_j + 2}{s_j}) \right)^2 \tag{19}$$

4 Experiments, results and applications

4.1 Measurement testbed

The Weibull Mixture Model has been applied to more than 700000 one-way delay measurements, collected from around 35 monitored points spread around the world, donated by the RIPE NCC institution.

The measuring strategy followed involved collecting around 3000 GPS-synchronised (few hundred nanoseconds accuracy) one-way delays uniformly spread on a 24-hour period. This implies a resolution of:

$$\text{Resolution} = \frac{3000 \text{ readings}}{24 \times 60 \text{ mins}} \approx 2 \text{ measurements/min}$$

In the following experiments, we have subtracted the minimum value from the the delay measurement set, thus giving only the stochastic queing portion of it. However, we have kept the original scale considering both elements to show the effect of the queueing delay over the total delay.

In what follows, section 4.2 shows the model matching in a single example with further conclusions of some aspects of the algorithm, such as optimal mixture-size and time of convergence. Section 4.3 explores the results obtained through the whole measurement set of experiments. Finally, section 4.4 shows how the Weibull Mixture Model can be applied to detecting routing step changes.

4.2 Visual example of model fit

End-to-end delays typically show high variability patterns with bell-shaped histograms, such as the one shown in figure 2 (top and middle-left). The use of a single Weibull distribution gives optimal estimates of $\hat{r} = 0.7418$ and $\hat{s} = 1.4887$. According to table 1, the model's mean is located at $\bar{x} = \hat{r}\Gamma(\frac{\hat{s}+1}{\hat{s}})^{1/\hat{s}} = 0.6703$. This implies that the majority of the packets will experience 0.67ms of queueing delay which, in addition to the 12.16 ms of transmission delay, gives a total average delay of 12.83 ms. The estimated variance is $\sigma^2 = 0.2097$.

[Figure 2 about here.]

The plots at the bottom show the accuracy of the matching via the log-log and QQ plots. It can be clearly seen from them that the matching accuracy is relatively poor, especially for fitting the tail. If two Weibull components are considered, the matching

9

accuracy significantly improves, as shown in fig. 3. The EM algorithm gives model parameter estimates of $q_1 = 0.4363$, $r_1 = 0.4503$ and $s_1 = 2.2824$ for the first component and $q_2 = 0.5637$, $r_2 = 0.9787$ and $s_2 = 1.7037$ for the second Weibull. This yields a mean value of $\bar{x} = 0.6661$ and variance of $\sigma^2 = 0.2277$, which are slightly different than the 1-Weibull model estimates. Again, the QQ-plot reveals that the matching is still not perfect.

[Figure 3 about here.]

Figures 4–7 show the accuracy achieved when considering 3, 5, 7 and 10 Weibull components respectively. Seemingly, adding more components provides a gradually better matching accuracy, at the price of increasing the model complexity though.

[Figure 4 about here.]

[Figure 5 about here.]

[Figure 6 about here.]

[Figure 7 about here.]

Two questions arise from these results. Firstly, how many Weibull distributions are required to obtain a good fit? And secondly, how many algorithm iterations are necessary to achieve convergence? Figure 8 addresses these questions. Fig. 8-bottom shows the evolution of the discrepancy $\hat{\lambda}^2$ with the number of iterations, for various values of M. As shown in fig. 8-top, the use of a single Weibull offers poor performance compared to the results achieved with $M = 2$. Indeed, the use of two components or more results more appropriate (fig. 8-bottom) with $M = 5$ being the best possible choice. The same plots reflects an interesting fact: the use of too many components degrades the matching discrepancy. This effect has previously been analysed and treated [18].

[Figure 8 about here.]

Additionally, the same figure 8 shows how to choose a suitable number of algorithm iterations. In practice, when considering two or more Weibull components, at least 20 iterations are necessary to achieve convergence. Generally, the more components considered, the more iterations needed.

More detailed results are given in table 3. In quantitative terms and after convergence, it is concluded that the use of a single Weibull distribution to fit delay histograms does not capture the delay dynamics of this particular end-to-end link. However, the

10

Weibull Mixture Model with at least two components significantly improves the matching accuracy. Moreover, a Weibull Mixture Model considering three components and beyond only achieves a slightly better match. Nevertheless, the use of more than five components performs even worse results in terms of λ^2.

[Table 3 about here.]

With regard to first- and second-order moment estimates, table 4 shows the results for various sizes of the Weibull Mixture Model according to equations 19, and a comparison to data estimates as given by equation 17. According to this, the mean and variance for $M = 2$ and $M = 3$ provides the closest results to the sample mean and variance.

[Table 4 about here.]

In conclusion, these results show that the use of a *M*-sized mixture of Weibull distributions ($M > 1$) provides a good delay histogram fit. The experimental results are consistent with this hypothesis, not only at every specific quantile of the probability density function, but also in terms of estimating general sample properties such as the mean and the variance.

The next section attempts to show how many of the remaining 244 experiments have shown similar good fitting results.

4.3 Full measurement testings

Figure 9 shows histograms of the λ^2 discrepancy values obtained for the 245 experiments for mixture sizes of 1, 2, 3, 5 and 7.

[Figure 9 about here.]

Table 5 shows bounds on the discrepancy achieved by the 10 to 100 percentages of the experiments to which different *M*-sized models have been applied.

[Table 5 about here.]

As shown, if using $M = 5$ Weibull components, 50% of the experiments achieve discrepancy results of $\lambda^2 = 0.0289$. This value is typical of a model fit as accurate as the one shown in fig. 3. Furthermore, up to 90% of the cases exhibit $\lambda^2 = 0.1045$, which corresponds to a similar accuracy than the shown in fig. 3.

It is worth remarking that the use of $M > 5$ components may degrade performance results, whereas the use of $M < 5$ might lead to poorly accurate fits. $M = 5$ stands as

11

the optimal choice in terms of accuracy. However, the complexity of handling $3 \times M$ parameters may suggest more practical $M = 2$ and $M = 3$ sized models, at the expense of losing some accuracy.

4.4   On detecting step changes

Step changes in delay profiles are important events as they clearly cause a significant degradation in network performance. They are often attached to changes in routing configuration for various reasons: link failure, power shutdowns and routing exchange protocols misbehaving [19]. The ability to early detect and locate such incidents is key for a quick recovery.

The Weibull Mixture Model provides enough flexibility to detect such situations. Figure 10 shows the matching how two Weibull distributions combined together accuracy captures the bimodal histogram shown in this step change model. A close examination of the model parameters (see table 6) shall reveal two Weibull components, each located at a different histogram peak. This information is very valuable to network operators since they can easily detect step change events just by monitoring the evolution of the model parameters over time.

[Figure 10 about here.]

[Table 6 about here.]

5   Summary, discussion and further works

This work has presented the Weibull Mixture Model, a way to characterise, summarise and profile delay experienced between edges in the Internet. Such model consists of combining multiple Weibull distributions to fit the delay histogram obtained from packet probes, taken at different times of the day. An easy, comprehensive and powerful statistic technique, known as Expectation Maximisation algorithm, has been utilised to obtain the optimal model parameters and given rise to a procedure implementation that can be easily run by network administrators and managers.

The Weibull Mixture Model has been checked and tested with more than 700000 GPS-accurate measurements collected under real network activity. The proposed model gives accurate results qualitatively and quantitatively in more than 90% of the experiments considered, providing a close fit at all percentiles of the delay histogram and inferring basic first- and second-order moments from its parameters only.

12

The implications of reducing daily delay behaviour from around 3000 measurements into $3M$ parameters (often ranging 6 to 15 the most) are manyfold:

- The Weibull Mixture Model allows a certain grade of profiling. Traditionally, network managers have categorised delay behaviour in terms of percentiles, which allow a poor representation of the real histogram (just one point of it per percentile). This model rather permits a richer representation as it gives all the points of the histogram with a fewer number of parameters.

- Statistics are easy to compute. Delay mean and standard deviation and other first and second-order moments, very useful when negotiating Service Level Agreements, arise from the model parameters easily.

- A new set of metrics to help managers and users assess network performance status, i.e. traffic workload, network availability and connectivity, grade of congestion, etc. has been defined, and its applicability shown. The analysis of the model parameters can help identify and diagnosis sudden events in network activity, such as network misconfiguration, power shutdowns, denial-of-use attacks, etc. as they dramatically impact the model parameters.

- Workload profiling and performance benchmarking. Profiling link workload is key to keep track of traffic demand evolution and assess which areas require urgent capacity upgrades. The Weibull Mixture Model parameters effectively captures the delay histogram features and constitute an optimal way to profile, record and store performance status of a network.

Ongoing research shall attempt to produce an EM-based real-time algorithm able to adapt and adjust model parameters at every new measurement sample. This would allow a finer tracking of end-to-end performance status with further applications in early detection and troubleshooting of performance degradation events; real-time workload estimation and hotspot location and even permit a certain grade of automatic traffic redirecting and engineering to avoid congestion and load balancing.

Additionally, it is intended to investigate the possible implications of embedding real-time network status inference via the Weibull Mixture Model into TCP's congestion advoidance and slow-start mechanisms. This consists of optimally adjusting the congestion window of TCP with the information provided by the Weibull-mixture model on attempts to early detect and react from congestion based on delay analysis rather than packet loss. For example, a sudden decrease of parameter s is clearly an effect of increase in delay variability, which can be considered a symptom of network congestion.

References

[1] B. Y. Choi, S. Moon, Z. L. Zhang, K. Papagiannaki, C. Diot, Analysis of point-to-point packet delay in an operational network, in: Proceedings of INFOCOM, 2004.

[2] A. Mukherjee, On the dynamics and significance of low frequency components of Internet load, Internetworking: Research and Experience 5 (4) (1994) 163–205.

[3] K. Papagiannaki, S. Moon, C. Fraligh, F. Thiran, F. Tobagi, C. Diot, Analysis of measured single-hop delay from an operational backbone network, IEEE Journal on Selected Areas in Communications. Special Issue on Internet and WWW Measurement, Mapping and Modelling 21 (6).

[4] T. Elteto, S. Molnar, On the distribution of Round-Trip Times in TCP/IP networks, in: Proceedins of the 24th Conference on Local Computer Netwoks, 1999, pp. 172–181.

[5] K. Park, G. Kim, M. Crovella, On the relationship between file sizes, transport protocols, and self-similar network traffic, in: Proceedings of the Forth International Conference on Network Protocols, Columbus, OH, USA, 1996, pp. 171–180.

[6] I. Norros, On the use of fractional Brownian motion in the theory of connectionless networks, IEEE Journal on Selected Areas in Communications. 13 (6) (1995) 953–962.

[7] W. E. Leland, M. S. Taqqu, W. Willinger, D. V. Wilson, On the self-similar nature of Ethernet traffic (extended version), IEEE/ACM Transactions on Networking 2 (1) (1994) 1–15.

[8] V. Paxson, S. Floyd, Wide-area traffic: the failure of Poisson modeling, IEEE/ACM Transactions on Networking 3 (3) (1995) 226–244.

[9] R. Beran, J. Sherman, M. Taqqu, W. Willinger, Long-range dependence in variable-bit rate video traffic, IEEE/ACM Transactions on Networking 43 (2/3/4) (1995) 1566–1579.

[10] M. Crovella, A. Bestavros, Self-similarity in World Wide Web traffic: evidence and possible causes, IEEE/ACM Transactions on Networking (5) (1997) 835–846.

[11] M. Lelarge, Z. Liu, C. H. Xia, Asymptotic tail distribution of end-to-end delay in networks of queues with self-similar cross traffic, in: Proceedings of the INFO-COMM'04, 2004.

[12] A. Dempster, N. Laird, D. Rubin, Maximum Likelihood for incomplete data via de EM algorithm, Journal of the Royal Statistics Society, Series B 39 (1) (1977) 1–38.

[13] J. A. Bilmes, A gentle tutorial of the EM algorithm and its application to parameter estimation for gaussian mixture and hidden markov models, International Computer Science Institute.

[14] M. A. Tanner, Tools for statistical inference. Methods for the exploration of posterior distributions and likelihood functions, 3rd Edition, Springer series in statistics, 1996.

[15] L. C. Hamilton, Regression with graphics: A second course in applied statistics, Brooks/Cole, 1992.

[16] V. Paxson, Empirically-derived analytic models of wide-area TCP connections, IEEE/ACM Transactions on Networking 2 (4) (1994) 316–336.

[17] S. Pederson, M. Johnson, Estimating model discrepancy, Technometrics 32 (3) (1990) 305–314.

[18] R. Bellman, Adaptive control processes: A guided tour, Princeton University Press, 1961.

[19] V. Paxson, End-to-end routing behaviour in the Internet, IEEE/ACM Trans. Networking 5 (5) (1997) 601–615.
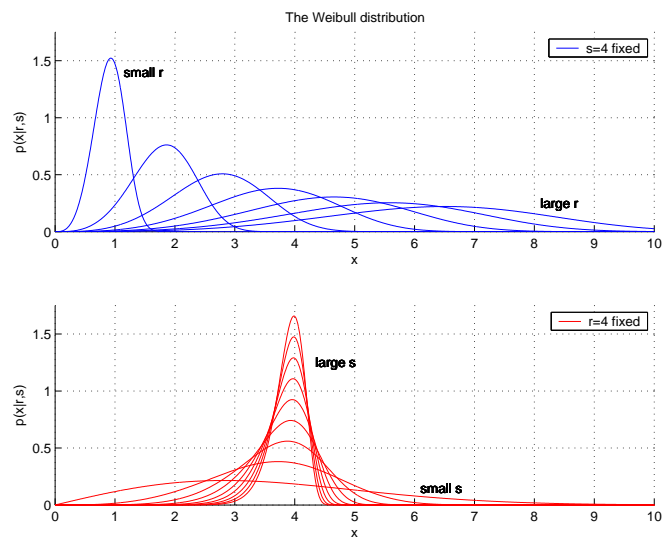
List of Figures

Figure 1: The Weibull distribution. Plots for different values of its scale (top) and shape (bottom) parameter.
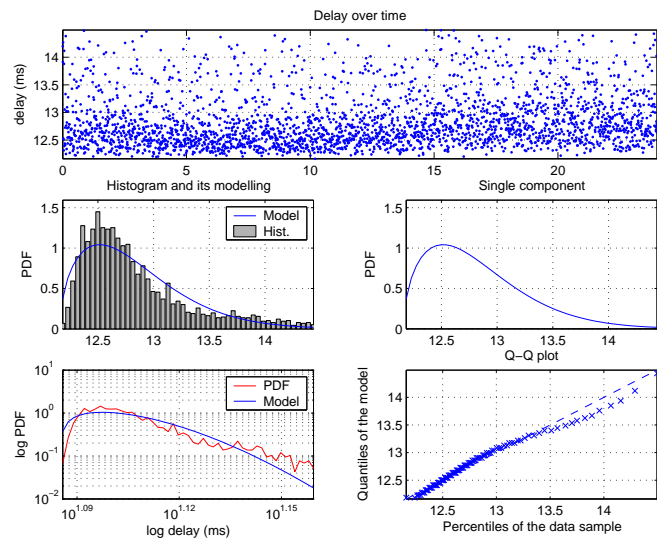
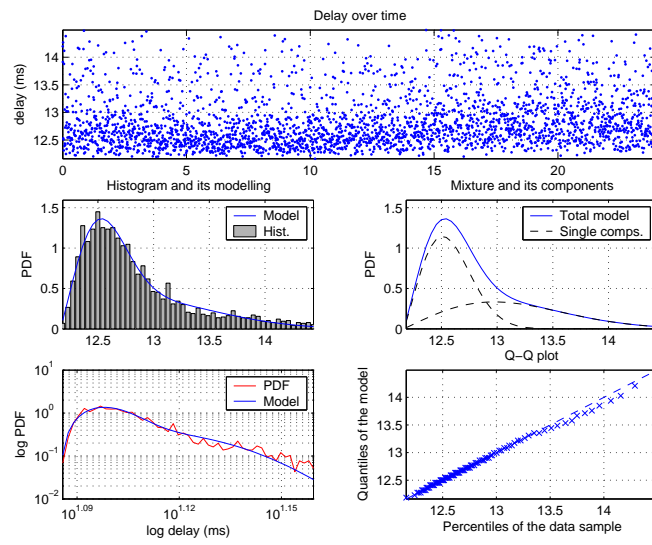Figure 2: Matching example with a single Weibull distribution.

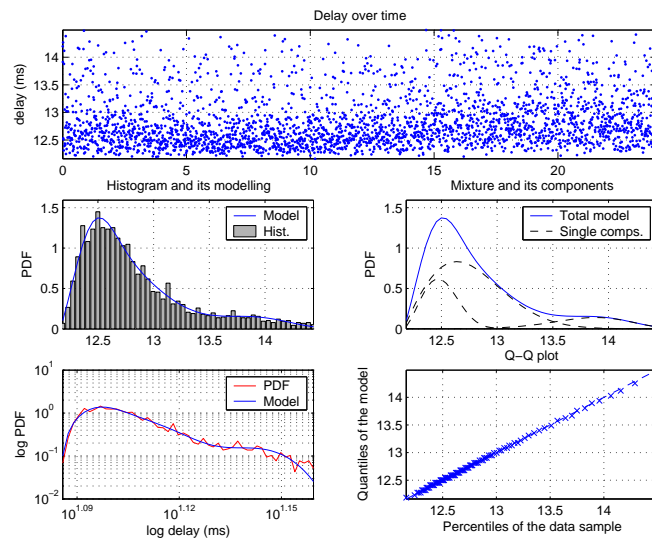Figure 3: Matching example with a 2-Weibull Mixture Model.

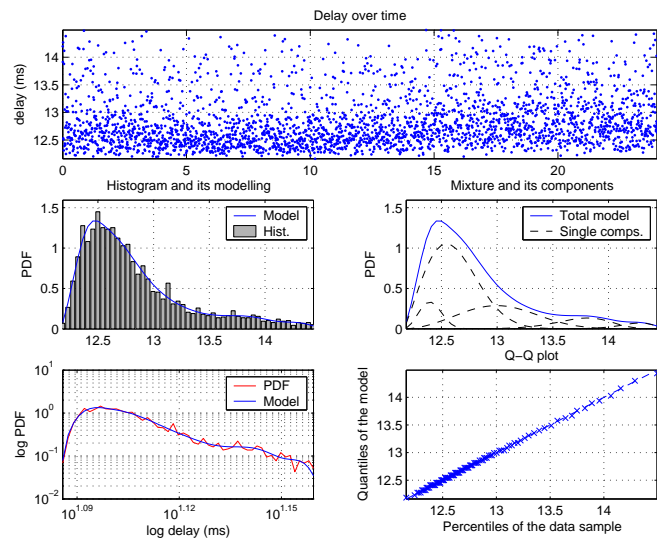Figure 4: Matching example with a 3-Weibull Mixture Model.

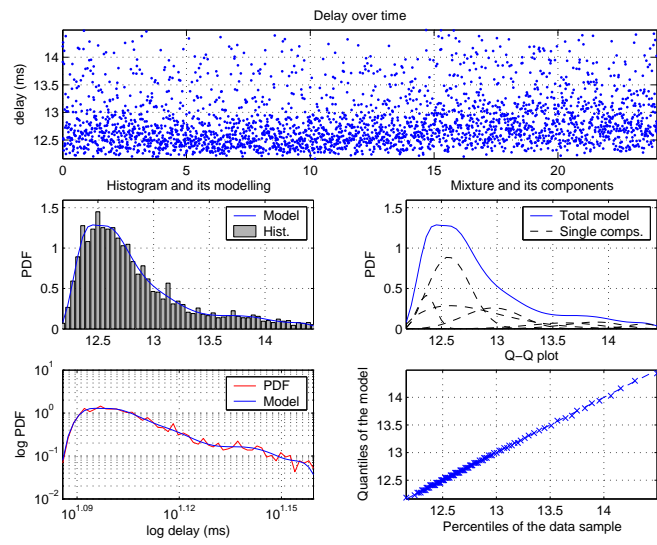Figure 5: Matching example with a 5-Weibull Mixture Model.

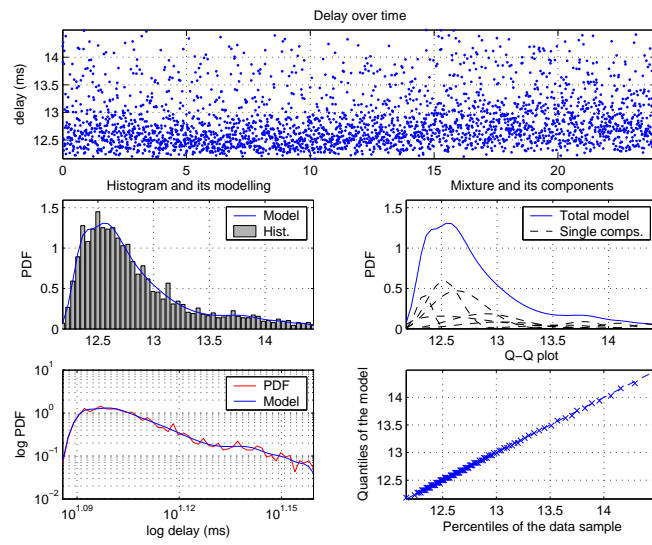Figure 6: Matching example with a 7-Weibull Mixture Model.

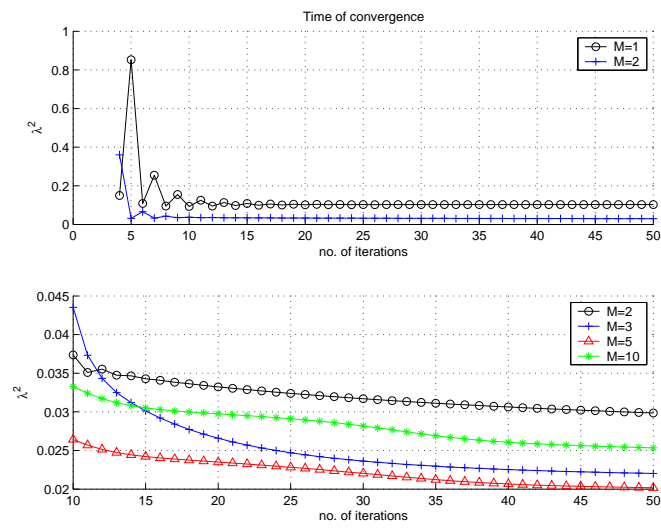Figure 7: Matching example with a 10-Weibull Mixture Model.

Figure 8: Discrepancy for various mixture sizes at different iterations of the algorithm.
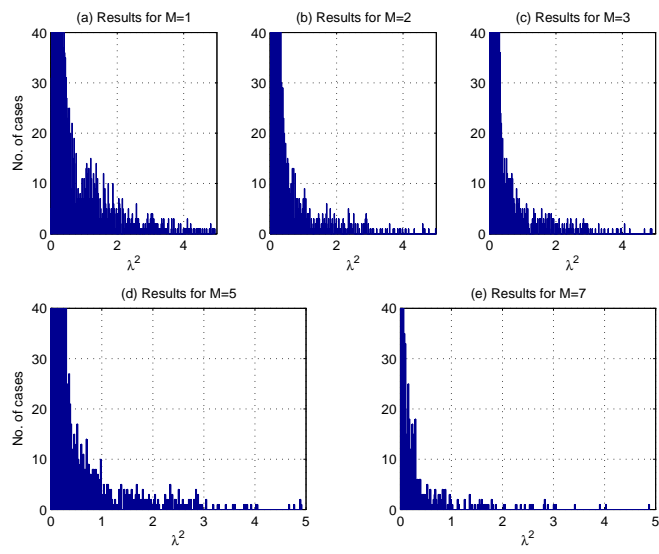
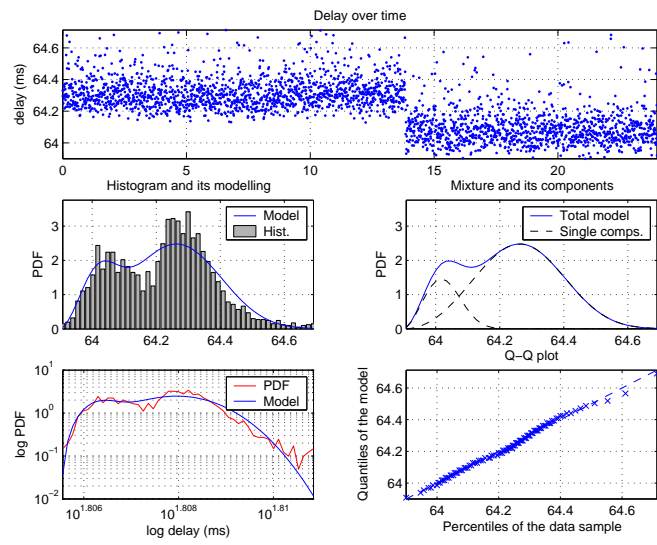Figure 9: Discrepancy results for the 245 experiments.

Figure 10: A 2-Weibull Mixture Model matching to step change event.

List of Tables

Mode:	$\tilde{x} = r(\frac{s-1}{s})^{1/s}$
Median:	$\check{x} = r(\ln 2)^{1/s}$
Mean:	$\bar{x} = r\Gamma(\frac{s+1}{s})$
Std. dev.:	$\sigma_x = r\sqrt{\Gamma(\frac{s+2}{s}) - \Gamma(\frac{s+1}{s})^2}$

Table 1: Main first and second-order statistics of the Weibull distribution

---

1. Initialisation.

Choose M as the size of the mixture

Initialise parameters $\mathbf{q}$ as a flat distribution and $\mathbf{r}$ and $\mathbf{s}$ uniformly spread over a certain interval (r_{min}, r_{max}) and (s_{min}, s_{max}).

2. REPEAT until convergence.

Compute the elements of the $N \times M$ matrix Px:

$$P_x(i,j) = q_j \frac{s_j x_i^{s_j-1}}{r_j^{s_j}} \exp\left(-\left(\tfrac{x_i}{r_j}\right)^{s_j}\right)$$

Compute the elements of the $N \times M$ matrix Pz:

$$P_z(i,j) = \frac{P_x(i,j)}{\mathbf{1}_M^T P_x(i,*)}$$

Compute parameters $\mathbf{q}$:

$$q_j = \tfrac{1}{N} P_z(*,j)^T \mathbf{1}_N$$

Compute parameters $\mathbf{r}$:

$$r_j = \left(\frac{\mathbf{x}^T P_z(*,j)}{N q_j}\right)^{1/s_j}$$

Compute parameters $\mathbf{s}$:

$$s_j = \frac{N q_j}{\left(\left(\left(\tfrac{\mathbf{x}}{r_j}\right)^{s_j} - 1\right) \log\left(\tfrac{\mathbf{x}}{r_j}\right)\right)^T P_z(*,j)}$$

Back to step 2.

Table 2: Summary of the implementation of Weibull Mixture Model.

M	λ^2
1	0.1031
2	0.0324
3	0.0247
4	0.0221
5	0.0228
6	0.0247
7	0.0263
8	0.0265
9	0.0277
10	0.0291

Table 3: λ^2 values for various mixture model-sizes.

M-size	$\bar{x}$	$\hat{\sigma}_x^2$
$M = 1$	0.6703	0.2097
$M = 2$	0.6661	0.2277
$M = 3$	0.6664	0.2270
$M = 4$	0.6663	0.2271
$M = 5$	0.6662	0.2270
$M = 6$	0.6661	0.2275
$M = 7$	0.6661	0.2272
$M = 8$	0.6661	0.2275
$M = 9$	0.6661	0.2274
$M = 10$	0.6661	0.2274
Data	0.6664	0.2271

Table 4: Mean and variance estimates for different values of M.

	$M = 1$	$M = 2$	$M = 3$	$M = 5$	$M = 7$
λ^2_{10}	0.0753	0.0330	0.0188	0.0141	0.0159
λ^2_{20}	0.1258	0.0533	0.0311	0.0234	0.0241
λ^2_{30}	0.1894	0.0709	0.0429	0.0347	0.0350
λ^2_{40}	0.2589	0.0905	0.0599	0.0491	0.0476
λ^2_{50}	0.3261	0.1181	0.0858	0.0690	0.0697
λ^2_{60}	0.4549	0.1819	0.1402	0.1067	0.1117
λ^2_{70}	0.7541	0.2610	0.2307	0.1909	0.1834
λ^2_{80}	1.3471	0.3726	0.3194	0.2912	0.2939
λ^2_{90}	2.5785	0.8759	0.8112	0.7042	0.7502
λ^2_{100}	7.3261	4.2342	2.1157	2.9203	3.1371

Table 5: Quantiles of the λ^2 results.

	$j = 1$	$j = 2$
q_j	0.1916	0.8084
r_j	0.1422	0.4108
s_j	2.5948	3.1123

Table 6: Weibull Model parameters for a step-change model.