Estimating Membership in a Multicast Session

Sara Alouf*
Vrije Universiteit
Department of Mathematics
De Boelelaan 1081a
1081 HV Amsterdam, The Netherlands
alouf@cs.vu.nl

Eitan Altman
INRIA
2004 Route des Lucioles
B.P. 93
06902 Sophia Antipolis, France
altman@sophia.inria.fr

Chadi Barakat
INRIA
2004 Route des Lucioles
B.P. 93
06902 Sophia Antipolis, France
cbarakat@sophia.inria.fr

Philippe Nain
INRIA
2004 Route des Lucioles
B.P. 93
06902 Sophia Antipolis, France
nain@sophia.inria.fr

ABSTRACT

We propose two novel on-line estimation algorithms to determine the size of a dynamic multicast group. We first use a Wiener filter to derive an optimal estimator for the membership size of the session in case the join process is Poisson and the lifetime of participants is distributed exponentially. We next develop the best first-order linear filter from which we derive an estimator that holds for any lifetime distribution. We apply this approach to the case where the lifetime distribution is hyperexponential. Both estimators hold under any traffic regime. Applying both estimators on real traces corresponding to video sessions, we find that both schemes behave well, one of which performs slightly better than the other in some cases. We further provide guidelines on how to tune the parameters involved in both schemes in order to achieve high quality estimation while simultaneously avoiding feedback implosion.

Categories and Subject Descriptors

G.3 [**Probability and Statistics**]: Queueing theory, Stochastic processes; H.4.3 [**Information Systems Applications**]: Communications Applications—*Computer conferencing, teleconferencing, and videoconferencing*

General Terms

Algorithms, Performance, Measurement

*This work was done when this author was at INRIA Sophia Antipolis, France.

Keywords

On-line estimation, multicast applications, membership size, $M/G/\infty$ queue, Wiener filter, first-order linear filter

1. INTRODUCTION

Since its introduction, IP multicast has seen slow deployment in the Internet. As stated in [7], the service model and architecture do not efficiently provide or address many features required for a robust implementation of multicast. However, the fact remains that IP multicast is very appealing in offering scalable point-to-multipoint delivery specially in satellite communications. Current research efforts tend to propose alternatives to IP multicast like the so-called "application layer multicast" [5, 10, 15, 20], the idea being to deploy multicast at the application layer. Also, new models to support multicast communications in a more effective way have been proposed, such as the EXPRESS multicast [14]. The latter is an extension to IP multicast that provides explicit support for large-scale multicast applications such as real-time stock quote dissemination, live sports video feeds or Internet radio and TV. EXPRESS provides as well a *best-effort* count of the number of subscribers.

This paper is motivated by the conviction that large-scale multicast applications will be widely deployed in the future as soon as the capability becomes available. We believe that membership estimates will be an essential component of this widespread deployment as they can be very useful for scalable multicast. The membership of a session can be used for feedback suppression as it is the case in current protocols such as RTP [22] and SRM [9]. In order to regulate the amount of session/control messages sent by receivers – the idea being not to exceed 5% of overall session bandwidth – these protocols use delay timers that are tuned based on membership estimates.

The membership of a multicast session can be used for charging the sources in large-scale applications. ISPs traditionally charge their customers on an input-rate basis. An alternative pricing scheme would be to charge sources based on their audience size which is more profitable in the case of millions of subscribers.

Estimating the size of a multicast session can be quite useful to many applications. Bolot, Turletti and Wakeman [4] use membership estimation to further estimate the proportion of congested receivers as needed in their videoconference system IVS. Future Internet radios and TVs will need to characterize their audience preferences and to follow the fluctuations of the audience size. Dutta, Schulzrinne and Yemini proposed an architecture for Internet radio and TV called MarconiNet [8] that relies on RTCP [22, 21]. Even though RTCP provides an easy mechanism for collecting statistics on the size of the audience, it does not scale well to large multicast session [8]. In such applications, sampling-based techniques are more appropriate.

There has been a significant research effort in devising sampling-based schemes for the estimation of the membership in multicast sessions [4, 18, 19, 11, 17, 3]. The feedback algorithms presented in [4, 18, 19, 11, 17] are all *at-least-one* scenarios in the sense that the membership estimation is based on at least one acknowledgement (ACK) coming from the receivers. In these probabilistic schemes, the receivers send ACKs to the source in reply to a specific request, either with a certain probability as in [4] or after some random time like in [18, 19, 17]. Except for [18] all these schemes assume that the size of the group remains fixed over time. In a recent work [3], we propose a dynamic scheme that tracks the variations of the membership in an optimal way. The estimation algorithm used is quite simple: the source requests from its receivers to send ACKs with probability p every S seconds; it collects the amount of ACKs received at each observation step and filters out these measurements to estimate the membership. To derive the optimal estimator we rely on a diffusion approximation for the heavy-traffic regime. Under the assumptions of Poisson join times and exponentially distributed connection times, the diffusion approximation yields linear dynamics which enables the design of the optimal filter using Kalman filter theory.

In this paper we propose two novel algorithms for estimating the membership based on the polling scheme presented in [3]. Our purpose is to develop an estimator under more general assumptions than the ones used in [3]. Our first approach is based on a Wiener filter, which provides the optimal dynamic estimator among all linear estimators. The dynamics is not required to be linear as in the case of the Kalman filter, which allows us to relax the heavy-traffic assumption made in [3]. Yet, in order to obtain explicit expressions for the parameters of the Wiener filter, we still have to assume that participants join the session according to a Poisson process, and that the time during which they stay in the multicast session, hereafter referred to as *on-time*, has an exponential distribution. Under these assumptions we design the optimal linear estimation scheme that turns out to require a filter of order one. Motivated by this structure, we then design an efficient estimation scheme for generally distributed on-times. To that end, we identify the optimal filter among all linear filters of order one. We illustrate this approach in the case of hyperexponentially distributed on-times. Alike the estimator developed with the Wiener filter, the latter estimator is valid under any traffic regime. Both estimators are then tested on real traces. Despite the fact that these traces violate the assumptions under which the estimators have been derived, very good performance are observed.

Remark 1.1. *The material presented in this paper does not require the specific use of IP multicast, nor any other multicast protocol. The solutions proposed hereafter are meant to be deployed at the application layer, and only require a multicast delivery of the requests for ACKs. This delivery can be achieved either by IP multicast, or by an application-layer multicast, or even by a* new *multicast technology. It is however assumed that the IP address of the source is available to all receivers in the session.*

The paper is organized as follows: the mathematical model of the membership is introduced in Section 2. The theory of Wiener filters is briefly presented in Section 3 and its application to the $M/M/\infty$ model comes in Section 3.1. The optimal first-order linear filter is developed in Section 4. Section 5 proposes some guidelines on how to choose parameters p and S. The robustness of both estimators is addressed through validations on real traces in Section 6. Finally, open issues are discussed in Section 7 and concluding remarks are given in Section 8.

2. MULTICAST GROUP MODELED AS AN $M/G/\infty$ QUEUE

We consider a multicast group that participants join and leave at random times. Let T_i and $T_i + D_i$ be the join time and leave time, respectively, of the i-th participant. In the following, D_i denotes the on-time of the i-th participant. Let $N(t)$ be the number of participants in the multicast group at time t or, equivalently, the size of the multicast session at time t. Without loss of generality we assume that $N(0) = 0$. Then,

$$N(t) = \sum_{i \geq 1} \mathbf{1}(T_i \leq t < T_i + D_i) \qquad (1)$$

where $\mathbf{1}(E)$ equals 1 if the event E occurs and 0 otherwise.

We shall assume that the join times form a homogeneous Poisson process (with constant intensity $0 < \lambda = 1/\mathbf{E}[T_{i+1} - T_i]$) and that the on-times form a renewal sequence of random variables (rvs) with common probability distribution $\Psi(x) = P(D_i < x)$, $0 < \mathbf{E}[D_i] < \infty$, further independent of the join times. In the following D will denote a generic rv with probability distribution $\Psi(x)$.

In the queueing terminology, $\{N(t), t \geq 0\}$ represents the occupation process (number of busy servers) in an $M/G/\infty$ queue [16].

At times $t = nS$, $n = 0, 1, \ldots$, with $S > 0$ a constant, each participant to the multicast session sends an ACK to the source with probability $0 < p < 1$ and does not send any feedback information to the source with probability $1 - p$. We assume that ACKs cannot be lost. However, this assumption can be relaxed if one knows the loss probability as it is possible to incorporate it in our feedback mechanism.

The ACK interval S between two consecutive polling instants has to be larger than the largest round-trip time between a receiver and the source, so that all ACKs produced in a round reach the source before the (automatic) start of the next round. Note that in practice the source will have to regularly multicast the pair (p, S) to ensure that each participant will know these values. Throughout the paper, p and S are held fixed (see Section 5 for possible extensions).

Let Y_n be the number of ACKs received by the source at time nS. Based on the knowledge of $Y_1, \ldots, Y_n$, our objective is to find an optimal estimator (in a sense to be defined below) $\hat{N}_n$ for $N_n := N(nS)$, the size of the multicast group at time nS. In filtering parlance, Y_n is an input signal and we want to generate another signal $\hat{N}_n$ that is as close as possible to an unknown signal N_n (typically by minimizing the mean square error).

For later use we briefly review some results on the $M/G/\infty$ queue. In steady-state, the number N of busy servers is a Poisson random variable with parameter $\rho := \lambda E[D]$, namely, $P[N = j] = \rho^j \exp(-\rho)/j!$. In particular, both the mean and the variance of the number of busy servers are equal to ρ. The autocovariance function of the stationary version of the process $\{N(t), t \geq 0\}$, also denoted by $\{N(t), t \geq 0\}$, is given by [6, Eqn (5.39)]

$$\mathrm{Cov}(N(t), N(t+h)) = \lambda \int_{|h|}^{\infty} P(D > u)\, du. \qquad (2)$$

In the following we will denote by $\mathrm{Cov}_X(\cdot)$ the autocovariance function of any second-order discrete-time stationary process $\{X_n\}_n$. With this notation and the definition of the process $\{N_n\}_n$, we see from (2) that

$$\mathrm{Cov}_N(k) = \rho\, \gamma^{|k|}, \quad k = 0, \pm 1, \ldots, \qquad (3)$$

with $\gamma := \exp(-\mu S)$, when the on-times $\{D_i\}_i$ are exponentially distributed with mean $1/\mu$.

Throughout the paper, we will assume that

$$\sum_{k \geq 0} \mathrm{Cov}_N(k) < \infty. \qquad (4)$$

In other words, we will exclude the situation where the on-times are heavy-tailed (e.g. Pareto distribution).

3. WIENER FILTER

Our objective is to transform a signal Y_n (noisy observation) into another signal $\hat{N}_n$ (estimator) that is the closest to an unknown signal N_n. By closest we mean that the mean error is zero (i.e. $\mathbf{E}[\hat{N}_n] = \mathbf{E}[N_n]$) and that the mean square error is minimized.

Such a transformation can be achieved by the Wiener filter that identifies the optimal linear filter [12] (i.e. the filter that is optimal among all linear filters). This approach gives the transfer function of the linear filter, which can be transformed back to the time domain to obtain the impulse response of the filter. From the impulse response of the filter, the expression of $\hat{N}_n$ as a function of Y_n and, possibly, of $\hat{N}_{n-1}, \hat{N}_{n-2}, \ldots$, can be found. We will detail this procedure below.

Since a filter that minimizes the mean square error when the underlying processes are centered also minimizes the mean square error when the same processes are non-centered, we will derive the Wiener filter for the centered (stationary) versions of processes $\{N_n\}_n$, $\{\hat{N}_n\}_n$ and $\{Y_n\}_n$, denoted by $\{\nu_n\}_n$, $\{\hat{\nu}_n\}_n$ and $\{y_n\}_n$, respectively. We have observed in the previous section that $\mathbf{E}[N_n] = \rho$. On the other hand

$$\mathbf{E}[Y_n] = \mathbf{E}[\mathbf{E}[Y_n \mid N_n]] = \mathbf{E}[p\, N_n] = p\rho. \qquad (5)$$

Therefore $\nu_n = N_n - \rho$, $\hat{\nu}_n = \hat{N}_n - \rho$ and $y_n = Y_n - p\rho$.

Throughout the paper, z is a complex number such that $|z| = 1$. Introduce

$$S_y(z) = \sum_{k=-\infty}^{\infty} \mathrm{Cov}_y(k) z^{-k}$$

the z-transform of the autocovariance function (also called the power spectrum) of $\{y_n\}_n$.

Let $\mathrm{Cov}_{\nu y}(k) = \mathbf{E}[\nu_{n-k}\, y_n]$ be the cross-correlation function of processes $\{\nu_n\}_n$ and $\{y_n\}_n$. We also introduce[1]

$$S_{\nu y}(z) = \sum_{k=-\infty}^{\infty} \mathrm{Cov}_{\nu y}(k) z^{-k}$$

the z-transform of $\mathrm{Cov}_{\nu y}(k)$. We can express $\mathrm{Cov}_y(k)$ and $\mathrm{Cov}_{\nu y}(k)$ in terms of $\mathrm{Cov}_\nu(k)$ as follows

$$\mathrm{Cov}_y(k) = p^2 \mathrm{Cov}_\nu(k) + \mathbf{1}(k=0)\rho p(1-p) \qquad (6)$$
$$\mathrm{Cov}_{\nu y}(k) = p\, \mathrm{Cov}_\nu(k) \qquad (7)$$

where we have used the identity $\mathrm{Cov}_\nu(k) = \mathrm{Cov}_N(k)$. We are now in position to derive the Wiener filter. First, we write $S_y(z)$ as

$$S_y(z) = \sigma G(z) G(z^{-1}) \qquad (8)$$

where σ is a constant. This operation is called the canonical factorization of the power spectrum of $\{y_n\}_n$. The function $G(z)$ is the part of $S_y(z)$ that has all its zeros and poles lying in the unit disk. The function $1/G(z)$ is the transfer function of the whitening filter: it transforms $\{y_n\}_n$ into a white noise process with variance σ.

Next, we form the ratio $S_{\nu y}(z)/G(z^{-1})$. This ratio is interpreted as the transfer function of a linear filter. The impulse response of this filter has values at the left and at the right of the time origin which means that this filter is non-causal. In order to have a causal filter, we should isolate the part of the impulse response having values at the right of the time origin solely, and compute its transfer function. This can be done by expanding $S_{\nu y}(z)/G(z^{-1})$ into fractions and by considering only the fractions with zeros and poles in the unit disk. In other words, we transform the non-causal filter into a causal one. We denote the transfer function of the causal version of the filter by

$$H(z) = \left[\frac{S_{\nu y}(z)}{G(z^{-1})}\right]_+.$$

The transfer function of the optimal filter is given by [12]

$$H_o(z) = \frac{H(z)}{\sigma G(z)}.$$

It remains to invert this transfer function back into the time domain to find the desired recurrence between $\hat{\nu}_n$ and y_n and, subsequently, between the non-centered variables $\hat{N}_n$ and Y_n. This procedure is illustrated in Section 3.1 for the case that the underlying model is the $M/M/\infty$ queue.

3.1 Application to the $M/M/\infty$ model

In light of the results reported in Section 3, all what we have to do is to find expressions for $S_y(z)$ and $S_{\nu y}(z)$. This can easily be done when the underlying model is the $M/M/\infty$ queueing model, as shown below.

[1] Observe from (6) and (7) that both $S_y(z)$ and $S_{\nu y}(z)$ are well-defined for $|z| = 1$ under the assumption (4).

Let us first determine $S_y(z)$. By using (6) and (3) together with the property that $\mathrm{Cov}_N(k) = \mathrm{Cov}_\nu(k)$, we find

$$\mathrm{Cov}_y(k) = \begin{cases} p^2 \rho \gamma^{|k|}, & \text{for } k \neq 0 \\ p\rho, & \text{for } k = 0. \end{cases}$$

Since $\gamma < 1$ and $|z| = 1$, the z-transform of $\mathrm{Cov}_y(k)$ is

$$S_y(z) = \frac{p\rho\left[\gamma(p-1)z^2 + [1+\gamma^2(1-2p)]z + \gamma(p-1)\right]}{z(1-\gamma z)(1-\gamma z^{-1})}.$$

The second-order polynomial in the variable z in the numerator has two positive real roots given by r and $1/r$, with

$$r = \frac{1+\gamma^2(1-2p) - \sqrt{(1-\gamma^2)[1-\gamma^2(1-2p)^2]}}{2\gamma(1-p)}.$$

Note that $r < 1$. Hence

$$\begin{aligned} S_y(z) &= \frac{\gamma\rho p(1-p)}{r}\left[\frac{(1-rz)(1-rz^{-1})}{(1-\gamma z)(1-\gamma z^{-1})}\right] \\ &= \sigma\, G(z)\, G(z^{-1}) \end{aligned}$$

with

$$\sigma := \frac{\gamma\rho p(1-p)}{r} \quad \text{and} \quad G(z) := \frac{1-rz^{-1}}{1-\gamma z^{-1}}.$$

We now compute $S_{\nu y}(z)$. From (7) and (3) we find

$$\mathrm{Cov}_{\nu y}(k) = p\rho\gamma^{|k|}$$

so that

$$S_{\nu y}(z) = \frac{p\rho(1-\gamma^2)}{(1-\gamma z)(1-\gamma z^{-1})}.$$

The transfer function $H(z)$ is given by

$$H(z) = \left[\frac{S_{\nu y}(z)}{G(z^{-1})}\right]_+ = \frac{\rho p(1-\gamma^2)}{(1-\gamma r)(1-\gamma z^{-1})}$$

and the transfer function $H_o(z)$ of the optimal filter takes here the simple form

$$H_o(z) = \frac{\rho p(1-\gamma^2)}{\sigma(1-\gamma r)(1-rz^{-1})} = \frac{B}{1-Az^{-1}}$$

where

$$A = r, \qquad B = \frac{\rho p(1-\gamma^2)}{\sigma(1-\gamma r)} = \frac{r(1-\gamma^2)}{\gamma(1-p)(1-\gamma r)}.$$

The impulse response of this linear filter is given by the first-order recurrence relation [12]

$$\hat{\nu}_n = A\hat{\nu}_{n-1} + By_n$$

with $\hat{\nu}_n$ the estimator of ν_n. We now return to the original processes $\{N_n\}_n$ and $\{Y_n\}_n$, to finally obtain the optimal linear filter:

$$\hat{N}_n = A\hat{N}_{n-1} + BY_n + \rho(1-A-pB). \tag{9}$$

It is interesting to compare this filter with the Kalman filter derived in [3][2]. They appear to be the same! This result is somehow expected, since both the Kalman filter and the Wiener filter are optimal (among the class of linear filters) in the sense that they minimize the mean square error. The key point is that the Kalman filter used in [3] was derived under a heavy traffic assumption, while the Wiener filter computed in the present paper holds for any value of the model parameters λ and μ. This partly explains why the estimator in [3] behaves well under light or moderate traffic as experimentally observed in that paper.

[2]Recall that a Kalman filter is the optimal filter under the condition of linear dynamics and observation, which does not hold in our case. However, the dynamics does converge to a linear diffusion as the traffic load tends to infinity, allowing us in [3] to obtain a Kalman filter which is optimal for the asymptotic heavy traffic regime.

We conclude this section by computing the mean square error $\epsilon_{min} := \mathbf{E}[(N_n - \hat{N}_n)^2]$ of our estimator. It is known that [12]

$$\epsilon_{min} = \sum_{k=1}^{M} \mathrm{Res}\,[F(z), z_k]$$

with

$$F(z) := \frac{1}{z}\left[S_\nu(z) - H_o(z)S_{\nu y}\left(z^{-1}\right)\right]$$

where $z_1, \ldots, z_M$ are the poles (if any) of the function $F(z)$ lying in the unit disk. The notation $\mathrm{Res}\,[F(z), z_k]$ stands for the residue of $F(z)$ at point $z = z_k$, namely, the coefficient of $1/(z - z_k)$ in the Laurent series expansion of $F(z)$ in the vicinity of z_k.

Specializing $F(z)$ to the values of $S_\nu(z)$, $S_{\nu y}(z)$ and $H_o(z)$ found earlier, yields

$$F(z) = \frac{\rho(1-\gamma^2)((1-Bp)z - A)}{(1-\gamma z)(z-\gamma)(z-A)}.$$

This function has two poles inside the unit circle which are located at $z = A$ and $z = \gamma$; the residues of $F(z)$ at these poles are given by $-\rho pAB(1-\gamma^2)/((1-\gamma A)(A-\gamma))$ and $\rho(1+pB\gamma/(A-\gamma))$, respectively. Summing up these residues gives

$$\epsilon_{min} = \rho\left(1 - \frac{Bp}{1-\gamma A}\right).$$

By using the expressions of A and B, we finally obtain

$$\epsilon_{min} = \rho\frac{-(1-\gamma^2) + \sqrt{(1-\gamma^2)(1-\gamma^2(1-2p)^2)}}{2\gamma^2 p}. \tag{10}$$

This expression for ϵ_{min} can be used to tune the parameters p and γ or equivalently S (see Section 5).

4. OPTIMAL FIRST-ORDER LINEAR FILTER

The theory reported in Section 3 applies to any on-time distribution $\Psi(x)$ (with the exception of heavy-tailed distributions). However, it is not easy to identify the function $G(z)$ that appears in the canonical factorization of the spectrum $S_y(z)$ (see (8)) and thereby the optimal filter, except when the on-times are exponentially distributed rvs (see Section 3.1).

In this section we will determine the first-order linear filter that minimizes the mean square error. Observe that, unlike the Wiener filter, the proposed approach will not return the optimal filter among all linear filters but simply the optimal linear filter among all first-order linear filters. We will illustrate this approach at the end of this section in the case where $\Psi(x)$ is an hyperexponential distribution.

Recall the definition of the centered processes $\{\nu_n\}_n$, $\{\hat{\nu}_n\}_n$ and $\{y_n\}_n$ made at the beginning of Section 3.

The methodology is simple: we want to find constants $A \in (0,1)$ and B such that $\epsilon := \mathbf{E}[(\nu_n - \hat{\nu}_n)^2]$ is minimized when the process $\{\hat{\nu}_n\}_n$ satisfies the following first-order recurrence relation

$$\hat{\nu}_n = A\hat{\nu}_{n-1} + By_n. \tag{11}$$

In steady-state this implies that

$$\hat{\nu}_n = B\sum_{k=0}^{\infty} A^k y_{n-k}. \tag{12}$$

The mean square error ϵ is equal to

$$\epsilon = \mathbf{E}\left[\hat{\nu}_n^2\right] + \mathbf{E}\left[\nu_n^2\right] - 2\mathbf{E}\left[\hat{\nu}_n\nu_n\right].$$

We have $\mathbf{E}[\nu_n^2] = \mathbf{E}[(N_n - \rho)^2] = \rho$ (see Section 2). From (12) and (7) we find

$$\mathbf{E}\left[\hat{\nu}_n\nu_n\right] = pB\sum_{k=0}^{\infty} A^k \mathrm{Cov}_\nu(k) = pBg(A)$$

where

$$g(z) := \sum_{k=0}^{\infty} z^k \mathrm{Cov}_\nu(k). \tag{13}$$

The power series $g(z)$ converges for $|z| < 1$ (Hint: $k \to \mathrm{Cov}_\nu(k)$ is nonincreasing) and is therefore differentiable for $|z| < 1$. We will denote by $g'(z)$ its derivative.

It remains to express $\mathbf{E}[\hat{\nu}_n^2]$ in terms of the parameters A and B. Squaring both sides of (11) and then taking the expectation, yields

$$\mathbf{E}\left[\hat{\nu}_n^2\right] = \left(\frac{B}{1-A^2}\right)\left(2A\mathbf{E}[\hat{\nu}_{n-1}y_n] + B\mathbf{E}\left[y_n^2\right]\right).$$

With the identities $\mathbf{E}[y_n^2] = \mathrm{Cov}_y(0) = \rho p$ (see (6)) and $\mathbf{E}[\hat{\nu}_{n-1}y_n] = Bp^2\left(g(A) - \rho\right)/A$ (Hint: use (12), (7) and $\mathrm{Cov}_\nu(0) = \rho$), we obtain

$$\mathbf{E}\left[\hat{\nu}_n^2\right] = \left(\frac{pB^2}{1-A^2}\right)\left(2pg(A) + \rho(1-2p)\right).$$

Finally, the mean square error is

$$\epsilon = \rho - 2pBg(A) + \left(\frac{pB^2}{1-A^2}\right)\left(2pg(A) + \rho(1-2p)\right). \tag{14}$$

In order to minimize ϵ, $A \in (0,1)$ and B must be the solution of the following system of equations:

$$\left\{\begin{array}{rcl} \dfrac{\partial\epsilon}{\partial A} & = & \dfrac{2pB}{1-A^2}\left(AB\left[\dfrac{2pg(A)+\rho(1-2p)}{1-A^2}\right]\right. \\ & & \left. +g'(A)(pB - (1-A^2))\right) = 0 \\ \dfrac{\partial\epsilon}{\partial B} & = & 2p\left(B\left[\dfrac{2pg(A)+\rho(1-2p)}{1-A^2}\right] - g(A)\right) = 0. \end{array}\right.$$

The 2nd equation gives

$$B = \frac{g(A)(1-A^2)}{2pg(A) + \rho(1-2p)}. \tag{15}$$

Substituting this value of B into the 1st equation shows that A must satisfy

$$\begin{aligned} &Ag(A)(2pg(A) + \rho(1-2p)) \\ &\quad -g'(A)(1-A^2)(pg(A) + \rho(1-2p)) = 0. \end{aligned} \tag{16}$$

If this equation has a unique solution $A \in (0,1)$, then substituting this value of A into (15) will give the optimal pair (A,B).

It is shown in Appendix A that (16) has always a unique solution in $[0,1)$ (in particular) if $g'(x) > 0$ in $[0,1)$. This condition will hold as long as $P(D > S) > 0$. In practice, one can always select S such that this condition is true.

The reader can check that the filter defined in (11) with the optimal pair (A,B) is the same as the Wiener filter found in Section 3.1 when the on-times are exponentially distributed.

We now illustrate the approach developed in this section by considering the situation where on-times have a hyper-exponential distribution. More precisely, we assume that

$$\Psi(x) = 1 - \sum_{l=1}^{L} p_l e^{-\mu_l x} \tag{17}$$

with $0 < p_l < 1$, $l = 1,2,\ldots,L$, and $\sum_{l=1}^{L} p_l = 1$. In this setting the underlying queueing model can be seen as L independent $M/M/\infty$ queues in parallel. The arrival rate to queue l is $p_l\lambda$ and the service rate is μ_l. Define $\gamma_l := \exp(-\mu_l S)$, $\rho_l := p_l\lambda/\mu_l$ so that $\rho = \sum_{l=1}^{L}\rho_l$. The autocovariance function of the process $\{\nu_n\}_n$ is equal to

$$\mathrm{Cov}_\nu(k) = \left\{\begin{array}{ll} \rho & \text{for } k = 0 \\ \sum_{l=1}^{L}\rho_l\gamma_l^{|k|} & \text{for } k \neq 0 \end{array}\right.$$

so that

$$g(A) = \sum_{l=1}^{L}\frac{\rho_l}{1 - A\gamma_l}.$$

Numerical example[3]: $L = 2$, $p = 0.0106$ and $S = 2.5s$. Also

$$\begin{array}{rclrclrcl} 1/\mu_1 & = & 3897s & \rho_1 & = & 19.5 & \gamma_1 & = & 0.999359 \\ 1/\mu_2 & = & 480061s & \rho_2 & = & 75.1 & \gamma_2 & = & 0.999995 \\ 1/\mu & = & 18316s & \rho & = & 94.7 & & & \end{array}$$

The optimal first-order filter is

$$\hat{N}_n = 0.99879456\,\hat{N}_{n-1} + 0.10720289\,Y_n + 0.006540864.$$

For comparison the Wiener filter found in Section 3.1 (for exponential on-times) is

$$\hat{N}_n = 0.99828589\,\hat{N}_{n-1} + 0.14885344\,Y_n + 0.012900081.$$

5. GUIDELINES ON CHOOSING ACKS PARAMETERS

A "good" pair (p,S) should (i) limit the feedback implosion while at the same time (ii) achieve a good quality of the estimator. Of course (i) and (ii) are antinomic and therefore a trade-off must be found. This trade-off will be formalized as follows: we want to select a pair (p,S) so that the mean number of ACKs generated every S seconds (see (5)) and the relative error of the variance of the estimator (denoted as η) are bounded from above by given constants, namely

$$\left\{\begin{array}{l} \mathbf{E}[Y_n] = p\rho \leq \alpha \\ \eta = \dfrac{Var(N_n) - Var(\hat{N}_n)}{Var(N_n)} \leq \beta. \end{array}\right. \tag{18}$$

[3]The values of the parameters come from the trace called $video_1$ investigated in Section 6.

When $\hat{N}_n$ is optimal then $Var(N_n) - Var(\hat{N}_n) = \mathbf{E}[(N_n - \hat{N}_n)^2]$ and η becomes the "normalized mean square error" [13, p. 202]. Optimality was shown for the $M/M/\infty$ queue, therefore

$$\eta = \frac{\epsilon_{min}}{\rho}$$

with ϵ_{min} given in (10).

For given constants α and β, it is easy to solve the constrained optimization problem defined in (18), provided that η is known.

For the $M/M/\infty$ model, where ϵ_{min} is given in (10), we find that $p = \alpha/\rho$ and that S, or equivalently γ, is the unique positive solution of the equation $\epsilon_{min} = \rho\beta$.

The problem now is to choose constants α and β so that conditions (i) and (ii) are satisfied. We have found that α in the range $[0.5, 1]$ and $\beta \leq 0.15$ give satisfactory results.

We conclude this section with general remarks on how to adapt the parameters p and S to important variations in the membership. The estimation schemes in Sections 3.1 and 4 have been obtained under the assumption that parameters p and S are fixed. However, the filters constructed in Sections 3.1 and 4 can still be used if p and/or S change over time, provided that these modifications do not prevent the system to be most of the time in steady-state. In that setting, a new filter will have to be recomputed after each modification (e.g. each time the number of ACKs collected by the source in a given period is significantly different from $p\rho$, the current expected number of ACKs).

6. VALIDATION WITH REAL TRACES

In this section we apply the estimators developed in Sections 3.1 and 4 to four traces corresponding to video multicast sessions. Two types of estimators will be used: the estimator – denoted as $\hat{N}_n^E$ – found in (9) when the the dynamics of population is modeled as an $M/M/\infty$ queue; the estimator – denoted as $\hat{N}_n^{H_2}$ – derived in Section 4 in case the join times are Poisson and the on-times have a 2-stage hyperexponential distribution ($M/H_2/\infty$ model).

The objective is twofold: we want to investigate the quality of both estimators when compared to real life conditions, and we want to identify the best one.

We have collected four MBone traces – denoted $video_i, i = 1, \ldots, 4$ – between August 2001 and September 2001 using the *MListen* tool [1]. Each trace corresponds to a long-lived video session (see duration of each session in Table 1, where the superscript "d" stands for "days"). We have run both algorithms (estimators) on each trace.

For each trace we have identified the parameters of the $M/M/\infty$ model (parameters λ and μ, or equivalently parameters ρ and μ) and of the $M/H_2/\infty$ model (parameters ρ, μ_1, μ_2, p_1 and $p_2 = 1 - p_1$ – see definitions in Section 4). The values of these parameters are reported in columns 3–8 in Table 1. Details on how these values have been obtained are given in Appendix B.

Parameters p and S have been chosen by following the guidelines presented in Section 5, namely $\alpha \in \{0.5, 1\}$ and $\beta \in \{0.1, 0.15\}$. Values of these parameters are listed in columns 9–10 in Table 1. The performance of estimators $\hat{N}_n^E$ and $\hat{N}_n^{H_2}$ are reported in Tables 2 and 3.

Table 2 reports several order statistics (columns 3–7) and the sample mean of the relative error $\frac{|N_n - \hat{N}_n|}{N_n}$ (column 2), where $\hat{N}_n$ is either $\hat{N}_n^E$ or $\hat{N}_n^{H_2}$. All results are expressed in percentages. The first observation is that both estimators perform reasonably well. The sample mean of the relative error is always less than 6.82% and is as low as 3.79%; when averaged over all experiments, this sample mean is less than 4.5% for both $\hat{N}_n^E$ and $\hat{N}_n^{H_2}$ (see last two rows). The last column gives the 95th percentile and reads as follows: the relative error achieved on trace $video_3$ by $\hat{N}_n^E$ (resp. $\hat{N}_n^{H_2}$) is 95% of the time less than 11.00% (resp. 12.56%). The second observation is that no scheme is uniformly better than the other over an entire session but their sample means are very close to each other (see column 2). For instance, $\hat{N}_n^E$ performs better than $\hat{N}_n^{H_2}$ regarding the 90th and the 95th percentiles whereas the result is reversed regarding the 25th percentile. It looks like the relative error on $\hat{N}_n^{H_2}$ is empirically more dispersed around its mean than is the relative error on $\hat{N}_n^E$, and has a longer tail. Across all sessions (see last two rows), 75% of the time $\hat{N}_n^{H_2}$ performs better than $\hat{N}_n^E$. This improvement does not come for free, since it requires the identification of 4 parameters (ρ, μ_1, μ_2 and p_1) instead of 2 (ρ and μ) for $\hat{N}_n^E$.

Table 2: Mean and percentiles of $|N_n - \hat{N}_n|/N_n$

Trace		Mean	25	50	75	90	95
$video_1$	$\hat{N}_n^E$	6.82	1.09	2.42	5.25	11.5	19.4
	$\hat{N}_n^{H_2}$	6.12	1.08	2.55	6.31	13.5	20.6
$video_2$	$\hat{N}_n^E$	4.19	1.41	3.08	5.43	8.66	11.9
	$\hat{N}_n^{H_2}$	4.12	0.98	2.14	4.41	8.78	12.6
$video_3$	$\hat{N}_n^E$	4.20	1.55	3.26	5.71	8.71	11.0
	$\hat{N}_n^{H_2}$	3.98	1.07	2.36	4.83	9.35	12.6
$video_4$	$\hat{N}_n^E$	3.79	1.23	2.57	4.51	7.50	11.0
	$\hat{N}_n^{H_2}$	4.06	1.02	2.21	4.39	8.98	14.7
overall	$\hat{N}_n^E$	4.44	1.33	2.88	5.22	8.60	12.0
	$\hat{N}_n^{H_2}$	4.34	1.02	2.26	4.73	9.61	14.2

Table 3: Empirical mean and variance of $N_n - \hat{N}_n$

Trace		Mean	Variance	ϵ_{min}, ϵ	η
$video_1$	$\hat{N}_n^E$	-0.112	12.664	13.942	0.147
	$\hat{N}_n^{H_2}$	-0.047	12.851	12.120	
$video_2$	$\hat{N}_n^E$	0.006	0.495	1.407	0.099
	$\hat{N}_n^{H_2}$	0.019	0.785	0.396	
$video_3$	$\hat{N}_n^E$	0.037	0.207	0.737	0.091
	$\hat{N}_n^{H_2}$	0.019	0.229	0.208	
$video_4$	$\hat{N}_n^E$	0.052	0.911	1.566	0.087
	$\hat{N}_n^{H_2}$	0.065	1.423	0.676	

Table 3 reports the sample mean and the sample variance of the error $N_n - \hat{N}_n$. In the 4th column, we list the theoretical variance. It is given by ϵ_{min} for $\hat{N}_n^E$ (see (10)) and by ϵ for $\hat{N}_n^{H_2}$ (see (14)). The expected average $\mathbf{E}[N_n - \hat{N}_n]$ is zero in both approaches. Both estimators $\hat{N}_n^E$ and $\hat{N}_n^{H_2}$ have almost no bias (see column 2), and their empirical variances closely match the theoretical ones given by ϵ_{min} and ϵ, respectively. It is of interest to point out that for the 4 traces studied, ϵ, the theoretical mean square error provided by $\hat{N}_n^{H_2}$, is smaller than ϵ_{min}, the theoretical mean square error provided by $\hat{N}_n^E$ (however, this result is reversed if we consider the empirical mean square errors). Thus, $\hat{N}_n^{H_2}$ is

Table 1: Parameter identification

Trace	Session lifetime	ρ	$1/\mu$	$1/\mu_1$	$1/\mu_2$	p_1	p_2	p	S	α	β
$video_1$	$3^d\,13^h\,33^m\,20^s$	94.7	18316	3897	480061	0.97	0.03	0.011	2.5	1.0	0.15
$video_2$	$11^d\;\;1^h\,46^m\;\;8^s$	14.1	16476	1	226498	0.93	0.07	0.034	3.2	0.5	0.1
$video_3$	$50^d\,22^h\,13^m\,20^s$	8.1	66823	1	900854	0.93	0.07	0.062	20.0	0.5	0.1
$video_4$	$29^d\,16^h\,43^m\,13^s$	17.9	83390	1	473268	0.82	0.18	0.028	10.0	0.5	0.1

Table 4: Distributions that best fitted into the inter-arrivals and on-times sequences.

Trace	Best fit for inter-arrivals sequence	Best fit for on-times sequence
$video_1$	Lognormal with $\mu = 3.38$, $d = 1.49$	Weibull with shape 0.35, scale 3700
$video_2$	Lognormal with $\mu = 5.20$, $d = 1.68$	Weibull with shape 0.26, scale 1400
$video_3$	Weibull with shape 0.65, scale 3500	Lognormal with $\mu = 5.08$, $d = 3.32$
$video_4$	Weibull with shape 0.55, scale 2700	Weibull with shape 0.18, scale 4000

more efficient[4] than $\hat{N}_n^E$ (again, $\hat{N}_n^E$ is empirically more efficient than $\hat{N}_n^{H_2}$). The last column provides the relative error on $Var(\hat{N}_n^E)$, called η $(= \epsilon_{min}/\rho)$ in Section 5.

In Fig. 1 (resp. 2, 3 and 4) we plot the variations of membership for session $video_1$ (resp. $video_2$, $video_3$ and $video_4$), together with the estimates returned by $\hat{N}_n^E$ and $\hat{N}_n^{H_2}$. Among all 4 sessions, session $video_1$ presents the highest variations in N_n. Fig. 1(a) (resp. 2(a), 3(a) and 4(a)) displays three curves: the membership of the video session, the estimation returned by $\hat{N}_n^E$, labeled "Exponential", and the estimation returned by $\hat{N}_n^{H_2}$, labeled "Hyperexponential". It is clearly visible, especially at the left-hand side of graph 1(a), that $\hat{N}_n^E$ tracks better the session dynamics than $\hat{N}_n^{H_2}$. Both estimators $\hat{N}_n^E$ and $\hat{N}_n^{H_2}$ have been derived under some specific and restrictive assumptions: Poisson join times for both of them, exponential (resp. 2-stage hyperexponential) on-times for the first (resp. second) one. It is interesting to know whether or not these assumptions were violated in each session $video_i$, $i = 1, \ldots, 4$. We have therefore carried out a statistical analysis of each trace in order to determine the nature of their join time process and of their on-time sequence.

As shown in Table 4 and Figs. 1, 2, 3 and 4, parts (b) and (c), neither is the join time process Poisson nor are the on-times exponentially distributed (or hyperexponentially distributed) for any of the traces. The inter-join times and the on-times appear to follow subexponential distributions (Weibull and Lognormal distributions), a situation quite different from the assumptions under which the estimators have been obtained. Despite these significant differences, the estimators behave well and therefore show a good robustness to assumption violations.

In summary, both estimators perform very well when applied to real traces and are robust to significant deviations from their (theoretical) domain of validity. Estimator $\hat{N}_n^{H_2}$ returns the best global performance for the relative error criterion, but does not track high fluctuations as well as $\hat{N}_n^E$. Overall, we have found that $\hat{N}_n^E$ is a good estimator, both in terms of its performance and its usability since it only requires the knowledge of two parameters: ρ and μ.

[4]An estimator is said to be more efficient if it has a smaller variance.

7. OPEN ISSUES

The main pending issue concerns the knowledge of parameters ρ and μ (or equivalently any couple of parameters among ρ, λ and μ, since $\rho = \lambda/\mu$ in steady-state). When these parameters are not known, the source should estimate them. Again, the source could estimate any two parameters among ρ, λ and μ and infer the third one.

One possible way of estimating λ is to let a newly arrived receiver send a "heartbeat" to the source with a certain (constant) probability q (q should be small enough to avoid overwhelming the source with heartbeats). The source would then use the arrival time t_m of the mth heartbeat to estimate λ. The maximum likelihood estimator is $\hat{\lambda} = m/(qt_m)$. This estimator is unbiased and consistent by the strong law of large numbers ($\lim_{m\to\infty} t_m/m = 1/(q\lambda)$).

In a similar way, the source can estimate μ if receivers probabilistically send a "goodbye" message reporting their lifetime when they leave the session. Let $\tau_{m'}$ be the lifetime indicated in the m'th goodbye message received at the source, then the maximum likelihood estimator of μ is simply $\hat{\mu} = m'/(\sum_{i=1}^{m'} \tau_{m'})$. The estimator $\hat{\mu}$ is unbiased and consistent.

A natural estimator for ρ is $\hat{\rho} = \mathbf{E}[\hat{N}_n]$. As long as there is no estimation of both ρ and μ, it is not possible to compute the filter coefficient A and B. Then only a naive estimator for N_n can be used, defined as the ratio of the number of ACKs received Y_n over the ACK probability p (this estimator does not perform well as shown in [2]). Observe that $\mathbf{E}[Y_n/p] = \rho$.

One might want to treat the estimation of N_n, μ and ρ as a joint parameter estimation problem. Unfortunately, the problem becomes much more complicated, and we are not able at the moment to say anything about its resolution. Alternatively, one might want to replace ρ by $\hat{\rho} = \mathbf{E}[\hat{N}_n]$ in (9) and investigate on the performance of the resulting algorithm. It is possible to go even farther by replacing ρ by Y_n/p, instead of using $\mathbf{E}[Y_n/p]$. This will add more burstiness to the dynamics of $\hat{N}_n$ and (9) becomes

$$\hat{N}_n = A\hat{N}_{n-1} + (1 - A)\frac{Y_n}{p}.$$

The latter equation is nothing but an exponential weighted moving average (EWMA) estimator for the membership process N_n. We expect this EWMA estimator to perform better than $\hat{N}_n$, as given in (9), over highly dynamic sessions, as its auto-regressive equation has been derived by replacing the constant term ρ by the bursty term Y_n/p. It is also expected that this EWMA estimator will react faster to important and sudden changes in the membership process. However, its performance over real sessions still needs to be investigated more carefully. Observe that the use of this EWMA estimator relies on the prior knowledge of $\gamma = \exp(-\mu S)$ solely, and therefore either "goodbye" or "hello" messages are still required to compute $\hat{\mu}$.

The first option is to directly estimate μ as indicated above, but the first estimate will be delayed until the reception of the first goodbye message, which might necessitates the departure of several customers. The second option is to estimate both ρ and λ as indicated earlier in the section and to estimate μ as $\hat{\lambda}/\hat{\rho}$, which is not expected to perform as well as the first option, but will return a first estimate much faster. At the expense of a larger warm-up period, we believe that the first option is preferable. Another advantage of this option concerns the possible modification of q' by the source, in order to control the volume of goodbye messages sent, which clearly cannot be done with the hello messages.

8. CONCLUSION

The major contribution of this work is the design of two novel estimators for evaluating the membership in multicast sessions. We have designed estimators capable of efficiently tracking the dynamics of multicast sessions while simultaneously avoiding feedback implosion. In contrast to the estimator proposed in [3] which was designed under heavy traffic assumptions, our schemes do not place such restrictions.

Relying on the Wiener filter theory, we have computed the optimal linear estimator for session membership when the underlying model is an $M/M/\infty$ queue. The optimality refers to the unbiasedness of the estimator and to the fact that the mean square error is minimized. We have also developed the optimal first-order linear filter in case the on-time distribution is arbitrary and have derived the associated estimator in case the on-times have a 2-stage hyperexponential distribution.

Both obtained estimators have been validated on real traces. Their performance have been shown to be excellent, one of them showing a good ability to adapt to highly dynamic multicast sessions.

9. ACKNOWLEDGMENTS

The authors wish to thank Prof. Patrick Thiran for helpful suggestions.

APPENDIX

A. EXISTENCE AND UNIQUENESS OF THE SOLUTION

LEMMA A.1. *Define*

$$\begin{aligned} f(x) \quad &:= \quad (2pg(x) + \rho(1-2p))xg(x) \\ &\quad\quad -(pg(x) + \rho(1-2p))(1-x^2)g'(x), \end{aligned}$$

where $g(x)$ is given in (13).

If $g'(x) > 0$ for $x \in [0,1)$, then $f(x)$ has a unique zero in $[0,1)$.

Proof. Write $f(x)$ as $f(x) = f_+(x) - f_-(x)$ where

$$\begin{aligned} f_+(x) &:= [2p(g(x)-\rho)+\rho]\, xg(x) \\ f_-(x) &:= [p(g(x)-\rho)+\rho(1-p)]\,(1-x^2)g'(x). \end{aligned}$$

The derivative of $f_-(x)$ is given by $f'_-(x) = -\alpha x^2 - \beta x + \alpha$ using

$$\begin{aligned} \alpha &:= p\left(g'(x)\right)^2 + [p(g(x)-\rho)+\rho(1-p)]g''(x), \\ \beta &:= 2[p(g(x)-\rho)+\rho(1-p)]g'(x). \end{aligned}$$

Since $g'(x) > 0$ and $g(x) > \rho$ (see (13)), it is seen that $\alpha > 0$ and $\beta > 0$ which implies that $f''_-(x) = -2\alpha x - \beta < 0$ for $x \in [0,1)$. We therefore have that $f'_-(x)$ is strictly decreasing in $[0,1)$, with $f'_-(0) = \alpha > 0$ and $f'_-(1) = -\beta < 0$. Thus, the function $f'_-(x)$ has only one zero in $[0,1)$.

Therefore, and under the assumptions of the lemma, it is seen that:

(i) $f_+(x)$ is continuous and strictly increasing in $[0,1)$ with $f_+(0) = 0$;

(ii) $f_-(x)$ is continuous in $[0,1)$ and $f_-(1) = 0$. There exists $x_0 \in (0,1)$ such that $f_-(x)$ is strictly increasing in $[0,x_0)$, strictly decreasing in $(x_0,1)$ and $f'_-(x_0) = 0$.

We deduce from the above that $f(x)$ has a unique zero in $[0,1)$ if $g'(x) > 0$ in $[0,1)$. This condition will hold as long as $P(D > S) > 0$. In practice, one can always select S such that this condition is true.

B. COMPUTING PARAMETERS FROM TRACE

Each trace records $(T_i, D_i), i \geq 1$. To use the $M/M/\infty$ queue model, we identify $1/\lambda = \mathbf{E}[T_{i+1} - T_i]$ and $1/\mu = \mathbf{E}[D]$, and deduce $\rho = \lambda/\mu$. To use the $M/H_2/\infty$ queue model, λ is computed as before. Identifying μ_1, μ_2, p_1 and p_2 requires the knowledge of the first three moments of D (recall that $p_2 = 1 - p_1$). For a 2-stage hyperexponential distribution, the kth moment is given by

$$\mathbf{E}[D^k] = \sum_{l=1}^{2} \frac{p_l k!}{(\mu_l)^k} = k!\left(\frac{p_1}{(\mu_1)^k} + \frac{p_2}{(\mu_2)^k}\right), \text{ for } k \geq 1.$$

The parameters μ_1, μ_2, p_1 and p_2 are then solution to the following system of four equations, where σ_l stands for $1/\mu_l$ with $l = 1,2$.

$$p_1 + p_2 = 1 \tag{19}$$
$$p_1\sigma_1 + p_2\sigma_2 = \mathbf{E}[D] \tag{20}$$
$$p_1\sigma_1^2 + p_2\sigma_2^2 = \mathbf{E}[D^2]/2 \tag{21}$$
$$p_1\sigma_1^3 + p_2\sigma_2^3 = \mathbf{E}[D^3]/6. \tag{22}$$

Equations (19) and (20) readily give

$$p_1 = \frac{\sigma_2 - \mathbf{E}[D]}{\sigma_2 - \sigma_1}, \qquad p_2 = \frac{\mathbf{E}[D] - \sigma_1}{\sigma_2 - \sigma_1}. \tag{23}$$

Substituting Equations (23) for p_1 and p_2 into (21) yields

$$\mathbf{E}[D^2]/2 = \mathbf{E}[D](\sigma_2 + \sigma_1) - \sigma_1\sigma_2. \tag{24}$$

Substituting them into (22) yields

$$\mathbf{E}[D^3]/6 = \mathbf{E}[D]((\sigma_2+\sigma_1)^2 - \sigma_1\sigma_2) - \sigma_1\sigma_2(\sigma_2+\sigma_1). \tag{25}$$

Introduce now S_σ and P_σ as the sum and the product of σ_1 and σ_2, respectively. Equations (24) and (25) become

$$\begin{aligned}\mathbf{E}[D^2]/2 &= \mathbf{E}[D]S_\sigma - P_\sigma \\ \mathbf{E}[D^3]/6 &= \mathbf{E}[D](S_\sigma^2 - P_\sigma) - P_\sigma S_\sigma \\ &= (\mathbf{E}[D^2]/2)\, S_\sigma - \mathbf{E}[D]P_\sigma\end{aligned}$$

where the latter identity is obtained when using the first one. It then follows that

$$S_\sigma = \frac{3\mathbf{E}[D]\mathbf{E}[D^2] - \mathbf{E}[D^3]}{3(2\mathbf{E}[D]^2 - \mathbf{E}[D^2])}, \tag{26}$$

$$P_\sigma = \frac{3\mathbf{E}[D^2]^2 - 2\mathbf{E}[D]\mathbf{E}[D^3]}{6(2\mathbf{E}[D]^2 - \mathbf{E}[D^2])}, \tag{27}$$

and $\sigma_1 = 1/\mu_1$ and $\sigma_2 = 1/\mu_2$ are the (positive) solutions of $x^2 - S_\sigma\, x + P_\sigma = 0$. Namely,

$$\sigma_{1,2} = 1/2 \times \left(S_\sigma \pm \sqrt{(S_\sigma)^2 - 4P_\sigma}\right).$$

We now can compute p_1 and p_2 as given in Equations (23). It is then possible to calculate $\rho_l = p_l\lambda/\mu_l$ and $\gamma_l = \exp(-\mu_l S)$ for $l = 1, 2$. Last $\rho = \rho_1 + \rho_2$.

C. REFERENCES

[1] K. C. Almeroth and M. H. Ammar. *MListen*, 1995. http://www.cc.gatech.edu/computing/Telecomm/mbone/.

[2] S. Alouf. *Parameter Estimation and Performance Analysis of Several Networking Applications*. PhD thesis, University of Nice-Sophia Antipolis, November 2002. Available at http://www.inria.fr/mistral/personnel/Sara.Alouf/.

[3] S. Alouf, E. Altman, and P. Nain. Optimal on-line estimation of the size of a dynamic multicast group. In *Proc. of IEEE INFOCOM '02, New York, New York*, volume 2, pages 1109–1118, June 2002.

[4] J.-C. Bolot, T. Turletti, and I. Wakeman. Scalable feedback control for multicast video distribution in the Internet. In *Proc. of ACM SIGCOMM '94, London, UK*, pages 58–67, September 1994.

[5] Y. Chu, S. Rao, and H. Zhang. A case for end system multicast. In *Proc. of ACM SIGMETRICS '00, Santa Clara, California*, pages 1–12, June 2000.

[6] D. R. Cox and V. Isham. *Point Processes*. Chapman and Hall, New York, 1980.

[7] C. Diot, B. N. Levine, B. Lyles, H. Kassem, and D. Balensiefen. Deployment issues for the IP multicast service and architecture. *IEEE Network magazine, Special Issue on Multicasting*, 14(1):78–88, January/February 2000.

[8] A. Dutta, H. Schulzrinne, and Y. Yemini. MarconiNet - an architecture for Internet radio and TV networks. In *Proc. of NOSSDAV '99, Basking Ridge, New Jersey*, June 1999.

[9] S. Floyd, V. Jacobson, S. McCanne, C. Liu, and L. Zhang. A reliable multicast framework for light-weight sessions and application level framing. In *Proc. of ACM SIGCOMM '95, Cambridge, Massachusetts*, pages 342–356, August 1995.

[10] P. Francis. Yoid: Extending the internet multicast architecture. Unrefereed report, April 2000. Available at http://www.icir.org/yoid/docs/index.html.

[11] T. Friedman and D. Towsley. Multicast session membership size estimation. In *Proc. of IEEE INFOCOM '99, New York, New York*, volume 2, pages 965–972, March 1999.

[12] S. Haykin. *Modern Filters*. Macmillan, New York, 1989.

[13] S. Haykin. *Adaptive Filter Theory*. Prentice Hall, 3rd edition, 1996.

[14] H. W. Holbrook and D. R. Cheriton. IP multicast channels: EXPRESS support for large-scale single-source applications. In *Proc. of ACM SIGCOMM '99, Cambridge, Massachusetts*, pages 65–78, September 1999.

[15] J. Jannotti, D. Gifford, K. Johnson, M. Kaashoek, and J. O'Toole. Overcast: Reliable multicasting with an overlay network. In *Proc. of USENIX OSDI '00, San Diego, California*, pages 197–212, October 2000.

[16] L. Kleinrock. *Queueing Systems: Theory*, volume 1. John Wiley and Sons, 1975.

[17] C. Liu and J. Nonnenmacher. Broadcast audience estimation. In *Proc. of IEEE INFOCOM '00, Tel Aviv, Israel*, volume 2, pages 952–960, March 2000.

[18] J. Nonnenmacher. *Reliable multicast transport to large groups*. PhD thesis, Ecole Polytechnique Federale de Lausanne, Switzerland, July 1998.

[19] J. Nonnenmacher and E. Biersack. Scalable feedback for large groups. *IEEE/ACM Trans. on Networking*, 7(3):375–386, June 1999.

[20] D. Pendarakis, S. Shi, D. Verma, and M. Waldvogel. ALMI: An application level multicast infrastructure. In *Proc. of USENIX USITS '01, San Francisco, California*, pages 49–60, March 2001.

[21] J. Rosenberg and H. Schulzrinne. Timer reconsideration for enhanced RTP scalability. In *Proc. of IEEE INFOCOM '98, San Francisco, California*, volume 1, pages 233–241, March/April 1998.

[22] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. RTP: a transport protocol for real-time applications. RFC 1889, Network Working Group, January 1996.

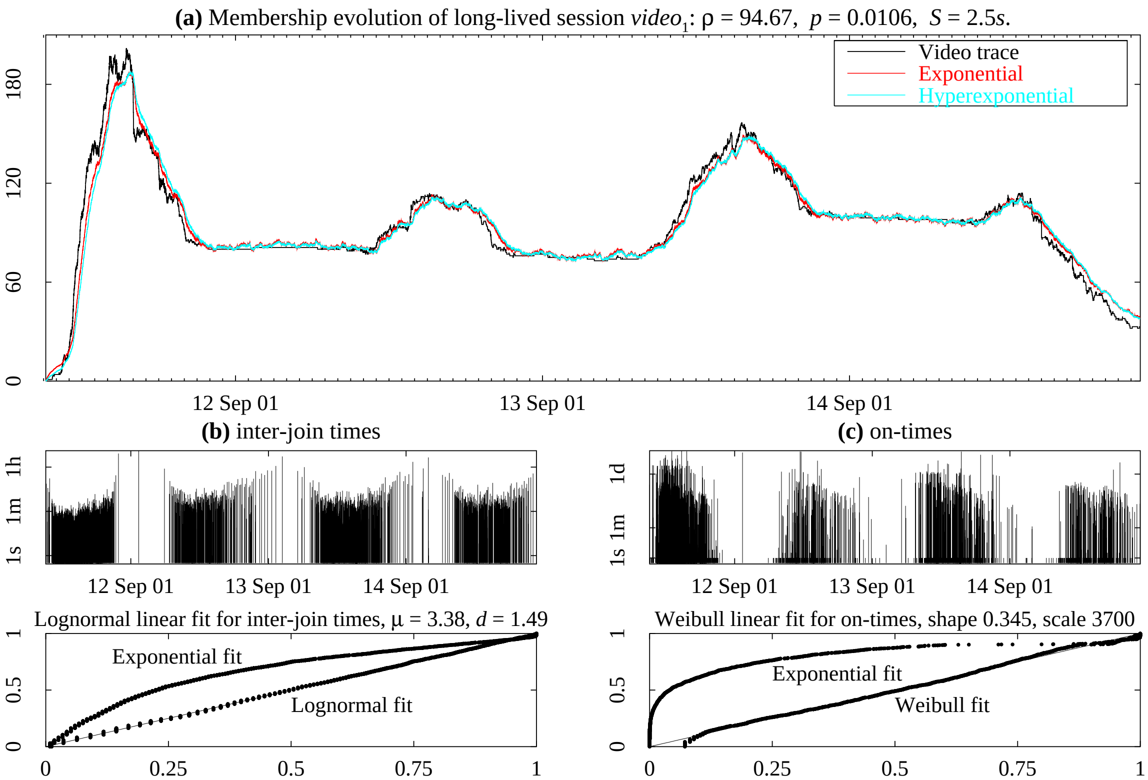

Figure 1: Membership estimation of session $video_1$ and corresponding probability plots

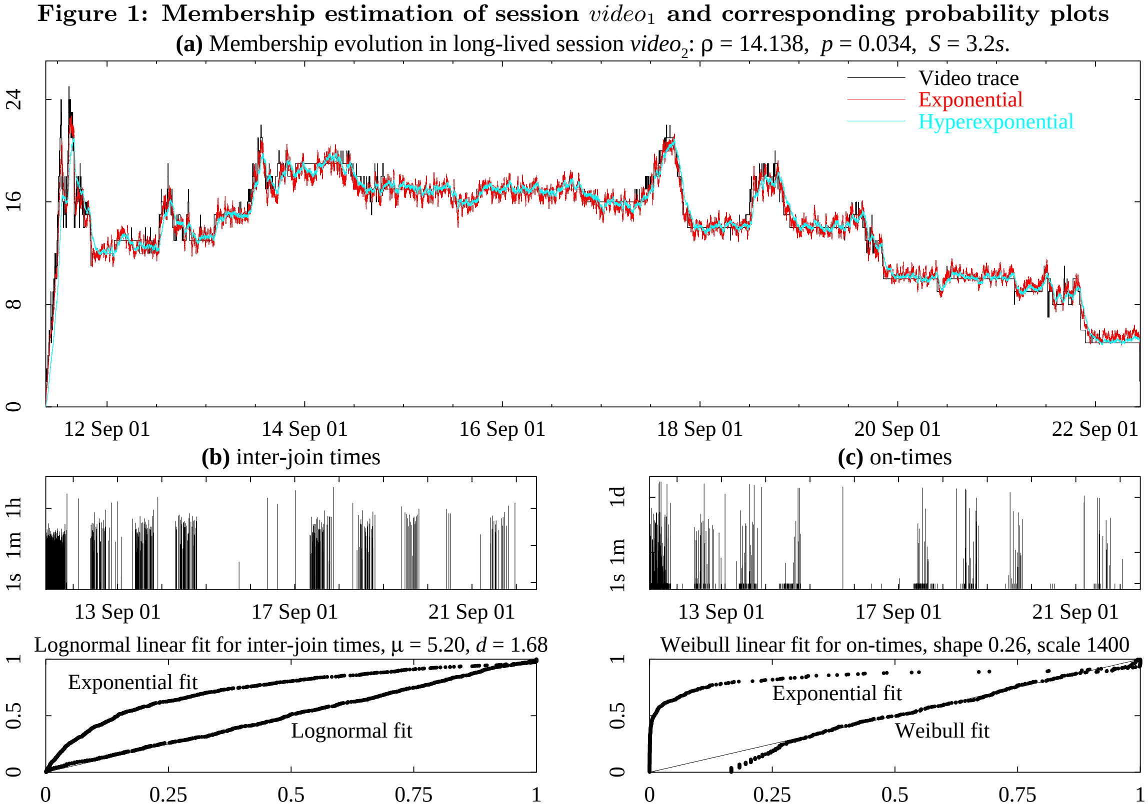

Figure 2: Membership estimation of session $video_2$ and corresponding probability plots

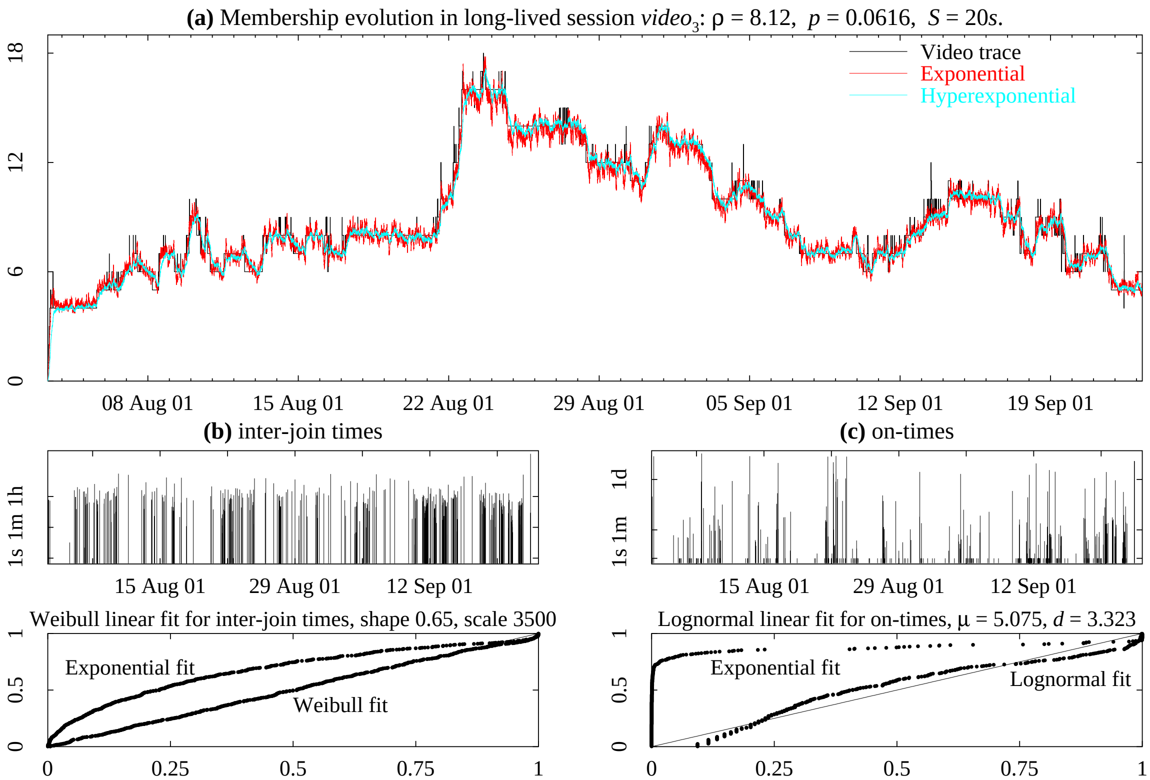

Figure 3: Membership estimation of session $video_3$ and corresponding probability plots

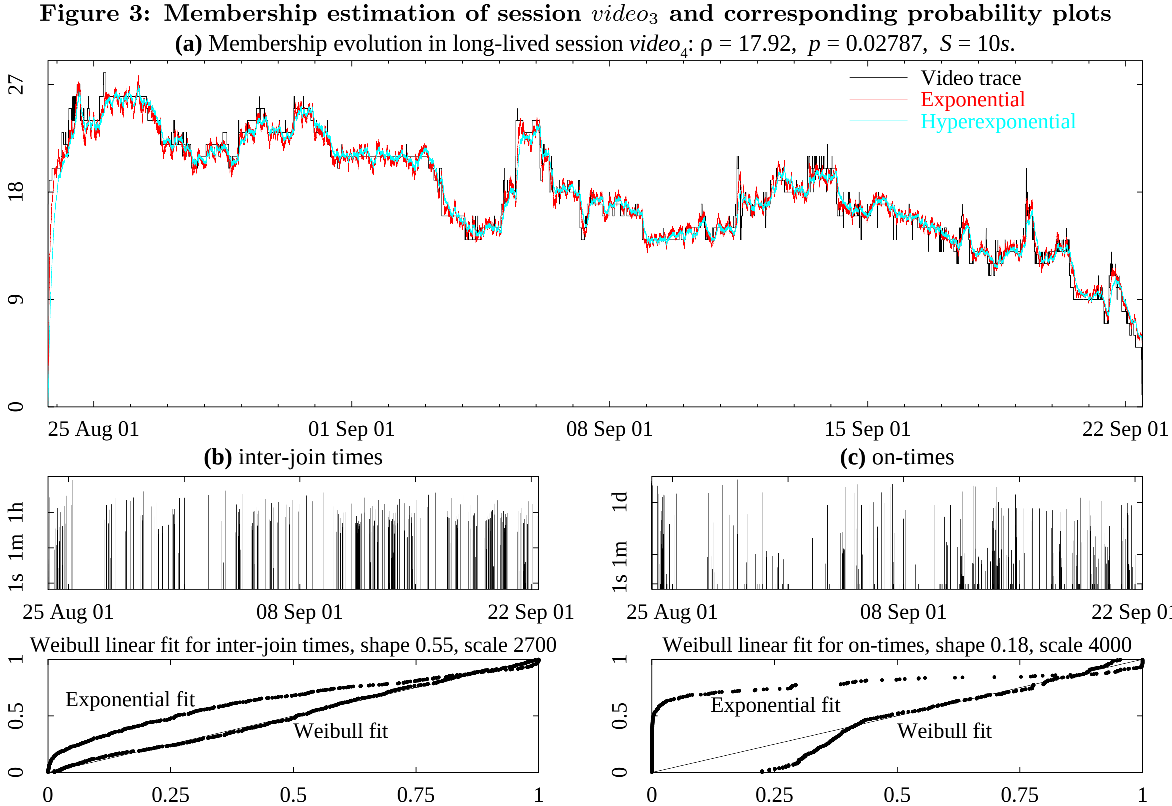

Figure 4: Membership estimation of session $video_4$ and corresponding probability plots