Estimating File-Spread in Delay Tolerant Networks under Two-hop Routing

Arshad Ali[1], Eitan Altman[2], Tijani Chahed[1], Dieter Fiems[3], Manoj Panda[2], and Lucile Sassatelli[4]

[1] Telecom SudParis, 9 rue C. Fourier, 91011 Evry Cedex, France.
{arshad.ali, tijani.chahed}@it-sudparis.eu
[2] INRIA, 2004 Route des Lucioles, 06902 Sophia-Antipolis, France.
{eitan.altman, manoj.panda}@sophia.inria.fr
[3] SMACS Research Group, Department of Telecommunications and Information Processing (TW07), St-Pietersnieuwstraat 41, B-9000 Gent, Belgium.
dieter.fiems@ugent.be
[4] I3S, 2000 Route des Lucioles, 06903 Sophia Antipolis, France.
sassatelli@i3s.unice.fr

Abstract. We consider a Delay/Disruption Tolerant Network under two-hop routing. Our objective is to estimate and track the degree of spread of a message/file in the network. Indeed, having such real-time information is critical for online control of routing and energy expenditure. It also benefits the multicasting application. With exponential inter-meeting times of mobile nodes: (i) for the estimation problem we, obtain exact expressions for the minimum mean-squared error (MMSE) estimator, and (ii) for the tracking problem, we first derive the diffusion approximations for the system dynamics and the measurements and then apply Kalman filtering. We also apply the solutions of the estimation and filtering problems to predict the time when a certain pre-defined fraction of nodes have received a copy of the message/file. Our analytical results are corroborated with extensive simulation results.

Keywords. delay/disruption tolerant networks; two-hop routing, multi-casting, estimation and tracking, Kalman filtering, level-crossing

1 Introduction

Mobile Ad hoc Networks (MANETs) aim at allowing communication between mobile users without any infrastructure. If the spatial density of mobiles in a MANET is low, then end-to-end communication between a source and a destination is limited by the lack of connectivity; in order to exchange packets, two mobile nodes must come into the radio range of each other. Owing to the intermittent connectivity, the nodes must rely on the Store-Carry-and-Forward paradigm which inherently entails a delay of communication. Such sparse MANETs are referred to as Delay Tolerant Networks (DTNs) wherein a source has to rely on the mobility of other nodes which act as relays, and takes advantage of the transmission opportunities which occur when the mobile relays come into contact. This forwarding strategy is known as opportunistic routing.

Several alternate methods of spreading multiple copies of the same message (or packet or file) have been investigated under opportunistic routing. In *epidemic routing* [11] data packets are flooded to all nodes in the network to minimize the delay. However, often the mobile nodes in DTNs have limited energy reserves and may prefer fewer transmissions to prolong network lifetime. *Spray-and-Wait routing* [10] and *probabilistic routing* [5] are proposed to achieve trade-offs between network resource consumption and protocol performance (in particular, energy versus delay trade-off). Unlike epidemic routing, in *two-hop routing* the relays do not give copies of the message to other relays.

As multiple copies of the same packet are allowed to spread in the network, it is important, in practice, to track the number of copies so as to have an online adaptive replication policy. If there is just one source spreading packets under two-hop routing, then it can have perfect knowledge of the current number of copies. However, if there are several source nodes spreading packets, then a source cannot have perfect knowledge of the current number of copies in the network. The same occurs in the case of epidemic or Spray-and-Wait routings (multi-hop routing cases) mentioned above, where the relays are allowed to themselves relay the packets to other intermediate nodes they are carrying.

In this paper, we address the case of two-hop routing. We assume that an observer node (that can be a source, but not necessarily) moves around in the network meeting with relay nodes to count the number of copies of the message. The nodes it meets inform it of whether they are carrying a copy of the packet. The problem is to get as accurate an estimation as possible of the number of nodes with copies using the measurements of the observer. The problem of adaptively controlling the spreading process will be addressed in future works.

Our Contributions: In this paper, we solve three problems. First, given the measurement of the observer at time t, we derive the exact expressions, under exponential inter-meeting times, for the *instantaneous* linear Minimum-Mean-Squared-Error (MMSE) estimator. Second, we derive a discrete time Kalman filter based on diffusion approximations for the spreading process and the measurements. Third, we estimate/predict the time at which a certain given fraction of population has received copies of the file. All analyses are substantiated by discrete event simulations.

Related Work: Mean-field approximations have been used to estimate the mean number of infected nodes under various spreading policies [13]. Such approximations are accurate when the number of nodes is sufficiently large. Our approach based on measurements with MMSE and Kalman filtering, allows to track the discrepancy between the actual process and the mean-field approximations. Also, our Kalman filter estimation is based on a second-order approximation whereas the mean-field approximations are only first-order descriptions.

A related estimation problem in wireline networks has been considered in [1], where the number of participants to a multicast session is tracked over time thanks to measurements taken by polling the users. In [1], the authors assume an *infinite* polulation from which arrivals occur and apply the diffusion approximation of the well-known $M/M/\infty$ queueing model. We, however, consider a realistic *finite* population from which arrivals occur. In [1], the Kalman filter is developed to track the fluctuations in the *stationary regime* of the $M/M/\infty$ queue. We, however, track the fluctuations in the *transient phase*. Furthermore, in [1], the delay in measurements (i.e., of the poll messages and the returning acknowledgments) is ignored. We, however, explicitly characterize the measurement process which complicates the derivation of the measurement equation.

2 Network Model and Objectives

We consider a Delay/Disruption Tolerant Network (DTN) consisting of S_0 sources and N_0 relay nodes. We focus on the tracking of one given file (or message), generated by these S_0 sources; tracking of other files generated by the same or other sources follows the same lines. The inter-meeting times of any specific (say, the i-th) source and any specific (say, the j-th) relay node are independent and exponentially distributed random variables with parameter β. At time zero, the S_0 sources start spreading a file adopting two-hop routing. Each time a source meets with a relay, the relay gets a copy of the file. Recall that, in two-hop routing the relays do not give copies of the file to other relays.

An observer H monitors the system. The observer may be one of the sources, but not necessarily so. The inter-meeting times of the observer with any specific (say, the k-th) relay are independent and exponentially distributed random variables with parameter μ. At each contact with a relay, the observer gets to know if the relay has or does not have a copy of the file. The observer simply counts the number of contacts it has had where the relay it met had a copy of the file.

Let $X(t)$ denote the number of relays that have a copy of the file at time t. *Note that* $X(t)$ *does not include the sources.* Let $Y(t)$ denote the number of copies that the observer has counted up to time t. Henceforth, we shall refer to $\{X(t), t \geq 0\}$ as "the process" and to $\{Y(t), t \geq 0\}$ as "the observation" or "the measurement". *We assume that* $X(0) = 0$ *and* $Y(0) = 0$.

Our objectives in this paper are to solve the following problems:

- P1 Estimate the (value of the) process at time t, $X(t)$, given the observation at time t, $Y(t)$.
- P2 Estimate the process at time t, $X(t)$, given the history of observation, $\{Y(u), u \in U, U \subseteq [0, t]\}$.
- P3 Estimate the time at which the process crosses a certain level X_L.

Problem P3 is motivated by *multicast* where one would be interested to know the time when a certain number, X_L, of nodes have received a copy of the file. Problems P1 and P2, as we shall see, can be seen as intermediate steps for solving Problem P3. But, they are also important problems in their own rights.

Our approach is to use *linear* estimators that are simple to implement and useful in practice. For solving Problem P1, we use the linear Minimum-Mean-Squared-Error (MMSE) estimator, and for solving Problem P2, we use the

Kalman filter. The Kalman filter is known to be optimal in several important ways [8], [9]. We solve Problem P3 using the solutions of Problems P1 and P2.

Problem P1: Consider two correlated random variables X and Y, with their mean vector and covariance matrix given by $\binom{m_x}{m_y}$ and $\begin{pmatrix} V_{xx} & V_{xy} \\ V_{yx} & V_{yy} \end{pmatrix}$, respectively. We solve the optimal estimation problem P1 by applying Proposition 1.

Proposition 1. *The linear estimator of X given Y which minimizes the expected square estimation error is given by*

$$E[X|Y] = m_x + V_{xy}V_{yy}^{-1}(Y - m_y). \qquad \square$$

We derive the required means and covariances in Section 3.2.

Problem P2: Making simplifying assumptions, we will approximate the process $\{X(t), t \geq 0\}$ by a diffusion process. Sampling the approximate process at regular *monitoring intervals* of duration T, we shall obtain a discrete time *linear* stochastic difference equation for the process $\{X(t), t \geq 0\}$. We will also derive a discrete time linear stochastic equation relating the measurements to the process. The linearity of both the system dynamics and the measurement equations will allow us to apply the Kalman filter to use the previous estimation in order to update the current estimation optimally. This is dealt with in Section 3.3.

Problem P3: A first-order solution to the level-crossing problem is obtained by using the solution of Problem P1. A more accurate second-order solution is obtained by using the solution of Problem P2. Those two expressions will be inverted numerically and used to compare the accuracy of the MMSE and Kalman estimators in estimating the level crossing times in Section 4.

3 Dynamics of the File Spread and Observation

In this section, guided by Proposition 1, we first derive the quantities m_x, m_y, V_{xx}, and V_{xy} as functions of time. Then, we derive diffusion approximations for the process and the observation, and derive the corresponding discrete time linear stochastic equations by sampling at regular intervals.

3.1 Characterization of the Process and Observation

Let $\xi_i(t)$ denote the indicator variable that takes the value 1 if relay i, $i = 1, \ldots, N_0$, has a copy of the file at time t, and 0 otherwise. Then, we have

$$X(t) = \sum_{i=1}^{N_0} \xi_i(t). \tag{1}$$

Let T_λ^i denote the time at which relay i receives a copy of the file. Note that T_λ^i is exponentially distributed with parameter $\lambda = S_0\beta$. Then, the probability $p(t)$ that a relay has a copy of the file at time t is given by

$$p(t) = P(T_\lambda^i \leq t) = 1 - \exp(-\lambda t). \tag{2}$$

By independence of source-relay meeting events, we conclude that $X(t)$ *has a Binomial distribution with parameters* N_0 *and* $p(t)$*, i.e.,*

$$P(X(t) = k) = \binom{N_0}{k} p(t)^k (1 - p(t))^{N_0 - k}. \tag{3}$$

Given the process $\{X(t), t \geq 0\}$, the count of the observer, $Y(t)$, has a (non-homogeneous) Poisson distribution with parameter

$$\theta_y(t) = \mu \int_0^t X(u) du. \tag{4}$$

We emphasize that, $\forall t \geq 0$, $\theta_y(t)$ is a random variable, since $X(t)$ is stochastic.

3.2 Derivation of the Means and (Co)variances

Lemma 1. *(i) The process* $X(t)$ *has mean* $m_x(t)$ *and variance* $V_{xx}(t)$ *given by:*

$$m_x(t) = N_0 p(t) \quad , \quad V_{xx}(t) = N_0 p(t)(1 - p(t)) \; .$$

(ii) The process $Y(t)$ *has mean* $m_y(t)$ *and variance* $V_{yy}(t)$ *given by* $m_y(t) = \mu m_x(t) E[T_x(t)$ *and:*

$$V_{yy}(t) = m_y(t) + m_x(t)\mu^2 \left(E[T_x^2(t)] - (E[T_x(t)])^2 \right) + \mu^2 E[T_x^2(t)] V_{xx}(t) \; ,$$

where $T_x(t)$ *is a random process with:*

$$E[T_x(t)] = \frac{t}{1 - \exp(-\lambda t)} - \frac{1}{\lambda} \quad , \quad E[T_x^2(t)] = \frac{\exp(\lambda t)}{\lambda^3} \left(\lambda^2 t^2 - 2\lambda t + 2 \right) - \frac{2}{\lambda^3} \; .$$

(iii) The cross-correlation between $X(t)$ *and* $Y(t)$ *is given by:* $V_{yx}(t) = m_y(t)(1 - p(t))$.

Proof: From Equation (3), we have: $m_x(t) = E[X(t)] = N_0 p(t)$ and $V_{xx}(t) = var(X(t)) = N_0 p(t)(1 - p(t))$. Next, we compute the distribution of $\theta_y(t)$.

$$\theta_y(t) = \sum_{i=1}^{N_0} \mu \int_0^t \xi_i(u) du = \sum_{i=1}^{N_0} \mu \max(t - T_\lambda^i, 0) = \sum_{i=1}^{X(t)} \mu (t - T_\lambda^i) = \sum_{i=1}^{X(t)} \mu T_x^i(t),$$

where $T_x^i(t)$ are i.i.d. random variables distributed like the truncated random variable $T_x(t)$ (truncated at t) with the following distribution:

$$P(T_x(t) > a) = P(t - T_\lambda^i > a | T_\lambda^i \leq t) = \frac{1 - \exp(-\lambda(t - a))}{1 - \exp(-\lambda t)} \quad \text{for } 0 \leq a \leq t. \tag{5}$$

whereby $E[T_x(t)]$ and $E[T_x^2(t)]$ given above.

$$m_y(t) = E[Y(t)] = E_X[E_Y[Y(t)|X(t)]] = E[\theta_y(t)] = \mu m_x(t) E[T_x(t)] \tag{6}$$

$$
\begin{aligned}
V_{yy}(t) = var(Y(t)) &= E_X[var_Y(Y(t)|X(t))] + var_X(E_Y[Y(t)|X(t)]) \\
&= E[\theta_y(t)] + var(\theta_y(t)), \qquad (7)
\end{aligned}
$$

since the variance of a Poisson random variable is equal to its mean. As before, $E[\theta_y(t)] = m_y(t)$ and $var(\theta_y(t))$ is obtained as follows:

$$
\begin{aligned}
var(\theta_y(t)) &= E[X(t)]var(\mu T_x(t)) + E[\mu^2 T_x(t)^2]var(X(t)) \\
&= m_x(t)\mu^2\left(E[T_x^2(t)] - (E[T_x(t)])^2\right) + \mu^2 E[T_x^2(t)]V_{xx}(t) \qquad (8)
\end{aligned}
$$

$$
\begin{aligned}
V_{yx}(t) = V_{xy}(t) &= E[X(t)Y(t)] - E[X(t)]E[Y(t)] \\
&= E_X[E_Y[X(t)Y(t)|X(t)]] - m_x(t)m_y(t) \\
&= E[X(t)\theta_y(t)] - m_x(t)m_y(t) = E[X^2(t)]\mu E[T_x(t)] - m_x(t)m_y(t) \\
&= \left(V_{xx}(t) + (m_x(t))^2\right)\mu E[T_x(t)] - m_x(t)m_y(t) = m_y(t)(1-p(t)) \qquad (9)
\end{aligned}
$$

■

3.3 Fluid and Diffusion Approximations

The process $\{X(t), t \geq 0\}$ can be viewed either as a *state-dependent queue* [7] or as a *density-dependent Markov process* [4]. We obtain the fluid and difusion approximations for the process $\{X(t), t \geq 0\}$ by viewing it as a single-server Markovian queue with state-dependent arrival rates, zero service rate and infinite buffer, and then applying the framework of [7]. A brief informal background on fluid and difusion approximations has been provided in the Appendix.

Consider the sequence $M_X^{(n)}/M_X^{(n)}/1/\infty/n$, $n = 1, 2, \ldots$, of state-dependent Markovian queueing systems, where index n denotes the size of the population from which the arrivals are drawn and $X^{(n)}(t)$ denotes the queue length at time t of the n-th system. The analogy with our DTN is as follows. The quantities n and $X^{(n)}(t)$ of the queueing system correspond to the quantities N_0 and $X(t)$, respectively, in our DTN. In analogy with our DTN, we let the arrival and departure rates for the n-th queueing system at state $X^{(n)}$ to be

$$
\lambda^{(n)}(X^{(n)}) = \lambda(n - X^{(n)}), \quad \text{and} \quad \mu^{(n)}(X^{(n)}) = 0, \qquad (10)
$$

respectively. First, we obtain the fluid limits of the process and measurement.

Lemma 2. *(i) Consider the rescaling $x^{(n)}(t) := X^{(n)}(t)/n$. The limit of the sequence $\{x^{(n)}(t), t \geq 0\}$, $n = 1, 2, \ldots$, as $n \uparrow \infty$, is given by*

$$
x(t) = 1 - \exp(-\lambda t). \qquad (11)
$$

(ii) The fluid limit $\{y(t), t \geq 0\}$ associated with the sequence $\{Y^{(n)}(t)/n, t \geq 0\}$, $n = 1, 2, \ldots$, is given by

$$
y(t) = \mu \int_0^t x(u)du. \qquad (12)
$$

Proof: Proof of (i): Applying Theorem 4.1 of [7] (or, Theorem 3.1 of [4]), the fluid limit $\{x(t), t \geq 0\}$ is given by the unique solution to the Ordinary Differential Equation (ODE) $\frac{dx(t)}{dt} = \lambda(1 - x(t))$, with initial condition $x(0) = 0$, where $\lambda = S_0\beta$. Whereby the result.
Proof of (ii): Consider the sequence of processes $\{Y^{(n)}(t), t \geq 0\}$, $n = 1, 2, \dots,$ where, for each n, $\{Y^{(n)}(t), t \geq 0\}$ is a *doubly stochastic* Poisson process [2] with (stochastic) intensity function $\mu X^{(n)}(t)$, i.e., we have

$$Y^{(n)}(t) = \mathcal{P}\left(\mu \int_0^t X^{(n)}(u)du\right), \tag{13}$$

where $\{\mathcal{P}(t), t \geq 0\}$ denotes a Poisson process of unit intensity. Consider the rescaling $y^{(n)}(t) = Y^{(n)}(t)/n$ and the mappings

$$\phi_{1,n}(t) = \frac{\mathcal{P}(nt) - nt}{n}, \quad \phi_{2,n}(t) = \frac{\mu}{n}\int_0^t X^{(n)}(u)du = \mu \int_0^t x^{(n)}(u)du.$$

It is easy to see that $y^{(n)}(t) = (\phi_{1,n} \circ \phi_{2,n})(t) + \phi_{2,n}(t)$ where $(f \circ g)(x)$ denotes $f(g(x))$. Note that, as $n \uparrow \infty$, we have $\phi_{1,n}(t) \to 0$ and $\phi_{2,n}(t) \to \mu \int_0^t x(u)du$, almost surely. Applying the Continuous Mapping Theorem (CMT) (see Theorem 13.2.1 of [12]), we obtain the fluid limit $\{y(t), t \geq 0\}$. ■
Next, we obtain the diffusion limits of the process and measurement.

Theorem 1. *(i) Consider the rescaling $v_x^{(n)}(t) = \sqrt{n}(x^{(n)}(t) - x(t))$. The diffusion limit $\{v_x(t), t \geq 0\}$, i.e., the limit of the sequence $\{v_x^{(n)}(t), t \geq 0\}$, $n = 1, 2, \dots,$ as $n \uparrow \infty$, is given by*

$$v_x(t) = \sqrt{\lambda}\int_0^t e^{-\lambda(t-u/2)}dB_1(u) = e^{-\lambda(t-s)}v_x(s) + \sqrt{\lambda}\int_s^t e^{-\lambda(t-u/2)}dB_1(u)\ . \tag{14}$$

(ii) Consider the rescaling $v_y^{(n)}(t) = \sqrt{n}(y^{(n)}(t) - y(t))$. The diffusion limit $\{v_y(t), t \geq 0\}$ is given by

$$v_y(t) = \int_0^t \sqrt{\mu x(u)}dB_2(u) + \frac{\mu}{\sqrt{\lambda}}\int_0^t e^{-\lambda u/2}dB_1(u) - \frac{\mu}{\lambda}v_x(t), \tag{15}$$

where $B_1(t)$ and $B_2(t)$ are independent standard Brownian motions.

Proof: Proof of (i): Applying Theorem 4.2 of [7], the diffusion limit $\{v_x(t), t \geq 0\}$ associated with the sequence $\{X^{(n)}(t), t \geq 0\}$, $n = 1, 2, \dots,$ is given by the unique (strong) solution to the *linear* Stochastic Differential Equation (SDE)

$$dv_x(t) = -\lambda v_x(t)dt + \sqrt{\lambda(1 - x(t))}dB_1(t), \tag{16}$$

with initial condition $v_x(0) \sim \mathcal{N}(0, 0)$, where $B_1(t)$ denotes a *standard Brownian motion*. Solving (16) (see page 354 of [3]), we obtain the result for all $0 \leq t < \infty$.
Proof of (ii): Defining the mapping $\phi_{3,n}(t) = \frac{\mathcal{P}(nt)-nt}{\sqrt{n}}$, it is easy to see that

$$v_y^{(n)}(t) = (\phi_{3,n} \circ \phi_{2,n})(t) + \mu\int_0^t v_x^{(n)}(u)du.$$

Noting that the diffusion limit associated with $\phi_{3,n}(t)$ is a standard Brownian motion $B_2(t)$ (which is independent of $B_1(t)$ on which $v_x(t)$ depends), and applying CMT (see Theorem 13.2.1 of [12]), we obtain the diffusion limit $\{v_y(t), t \geq 0\}$ associated with the sequence $\{Y^{(n)}(t), t \geq 0\}$, $n = 1, 2, \ldots$, as

$$v_y(t) = B_2\left(\mu \int_0^t x(u)du\right) + \mu \int_0^t v_x(u)du.$$

Whereby the result. ■

3.4 The Kalman Filter

Defining $v_{x,k} := v_x(kT)$, where $T(> 0)$ is some periodic interval at which we want to track the process $X(t)$, we obtain from (14) the *system dynamic equation* as:

$$v_{x,k+1} = \alpha v_{x,k} + w_k, \quad k = 0, 1, 2, \ldots, \tag{17}$$

where $\alpha = e^{-\lambda T}$, and

$$w_k = \sqrt{\lambda} \int_{kT}^{(k+1)T} e^{-\lambda((k+1)T - u/2)} dB_1(u).$$

Defining $v_{y,k} := v_y(kT)$, we obtain from (15) the *measurement equation* as:

$$v_{y,k} = \gamma v_{x,k} + z_k, \quad k = 0, 1, 2, \ldots, \tag{18}$$

where $\gamma = -\dfrac{\mu}{\lambda}$, and $z_k = r_k + s_k$, where

$$r_k = \frac{\mu}{\sqrt{\lambda}} \int_0^{kT} e^{-\lambda u/2} dB_1(u), \quad \text{and} \quad s_k = \int_0^{kT} \sqrt{\mu x(u)} dB_2(u).$$

Defining, $n_k := n_{1,k} + n_{2,k}$, where

$$n_{1,k} = \frac{\mu}{\sqrt{\lambda}} \int_{kT}^{(k+1)T} e^{-\lambda u/2} dB_1(u), \quad \text{and} \quad n_{2,k} = \int_{kT}^{(k+1)T} \sqrt{\mu x(u)} dB_2(u),$$

we obtain, $r_{k+1} = r_k + n_{1,k}$, $s_{k+1} = s_k + n_{2,k}$ and $z_{k+1} = z_k + n_k$.

Notice that, the process noise w is white, but the measurement noise z is *colored*. We *whiten* the measurement noise by defining $v'_{y,k} := v_{y,k+1} - v_{y,k}$, and derive the *modified measurement equation* as:

$$\begin{aligned} v'_{y,k} &= v_{y,k+1} - v_{y,k} = \gamma v_{x,k+1} + z_{k+1} - \gamma v_{x,k} - z_k \quad (19) \\ &= \gamma(\alpha v_{x,k} + w_k) + z_{k+1} - \gamma v_{x,k} - z_k = \gamma' v_{x,k} + z'_k, \quad (20) \end{aligned}$$

where $\gamma' = \gamma(\alpha - 1)$ and $z'_k = \gamma w_k + n_k$. Notice that the modified measurement noise z' is white. The modified measurement noise z' and the original (unmodified) measurement noise z are both correlated with the process noise w.

However, the modified measurement noise at (the discrete) time k, z'_k, is uncorrelated with the process noise up to time $k-1$, $\{w_j\}$, $j = 0, 1, \dots, k-1$. Thus, $M_k := E[w_{k-1} z'_k] = 0$, and we can apply a standard Kalman filter (see page 187 of [9]) with the system dynamics and (modified) measurement equations

$$v_{x,k} = \alpha v_{x,k-1} + w_{k-1} \tag{21}$$

$$v'_{y,k} = \gamma' v_{x,k} + z'_k, \tag{22}$$

where $\{w_k, k = 0, 1, \dots\}$ and $\{z'_k, k = 0, 1, \dots\}$ are white noise sequences with

$$w_k \sim \mathcal{N}(0, Q_k), \quad Q_k := E[w_k^2], \quad \text{and} \quad z'_k \sim \mathcal{N}(0, R_k), \quad R_k := E[(z'_k)^2].$$

It can be shown that

$$E[w_k^2] = (1-\alpha)\alpha^{k+1}, \text{ and } E[(z'_k)^2] = \gamma^2 E[w_k^2] + 2\gamma E[w_k n_k] + E[n_k^2], \text{ where}$$

$$E[w_k n_k] = \mu T \alpha^{k+1} \quad \text{and} \quad E[n_k^2] = \mu T - \gamma\,(1-\gamma)\,(1-\alpha)\alpha^k.$$

Let $\hat{v}^-_{x,k}$ and $\hat{v}^+_{x,k}$ denote the estimates for $v_{x,k}$ before and after taking into account the measurement, respectively, at time k. Let P_k^- and P_k^+ denote the covariances of the corresponding estimation errors. Let $\hat{v}'_{y,k} = \hat{v}_{y,k+1} - \hat{v}_{y,k}$, where $\hat{v}_{y,k} = \sqrt{N_0}((\hat{Y}(kT)/N_0) - y(kT))$, and $\hat{Y}(kT)$ and $y(kT)$ denote the actual measurement (i.e., observer count) and the value of $y(t)$, respectively, at time $t = kT$. Starting with $\hat{v}^+_{x,0} = 0$ and $P_0^+ = 0$, we apply the following Kalman filter equations (see Equations 5.17-5.19 of [9]) repeatedly at all time k:

$$\hat{v}^-_{x,k} = \alpha \hat{v}^+_{x,k-1} \tag{23}$$

$$P_k^- = \alpha^2 P_{k-1}^+ + Q_{k-1} \tag{24}$$

$$P_k^+ = \left((P_k^-)^{-1} + (\gamma')^2/R_k\right)^{-1} \tag{25}$$

$$K_k = \gamma' P_k^+ / R_k \tag{26}$$

$$\hat{v}^+_{x,k} = \hat{v}^-_{x,k} + K_k(\hat{v}'_{y,k} - \gamma' \hat{v}^-_{x,k}) \tag{27}$$

where K_k denotes the Kalman filter gain at time k. We obtain the estimates for the process as $\hat{X}(kT) = N_0 x(kT) + \sqrt{N_0}\hat{v}^+_{x,k}$, where $\sqrt{N_0}\hat{v}^+_{x,k}$ provides an estimate of the fluctuation of the process about its mean, at time $t = kT$.

4 Performance of Analytical Prediction and Estimation based on Measurements

In this section, we evaluate: (i) the quality of estimation provided by the MMSE estimator and the Kalman filter, and (ii) the accuracy of the predictions about the level-crossing times based on the estimation. We also comment on the prediction effectiveness of the fluid model of the process. We simulate a DTN as described in Section 2 for the following scenarios: (1) **Scenario 1:** $N_0 = 50$, $\beta = 0.02$, $T = 1.0$, $\mu = \beta$, (2) **Scenario 2:** $N_0 = 50$, $\beta = 0.02$, $T = 0.1$,

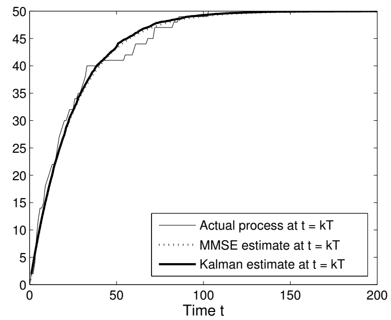

Fig. 1. Performance of MMSE estimation and Kalman filter estimation of the process for Scenario 1.

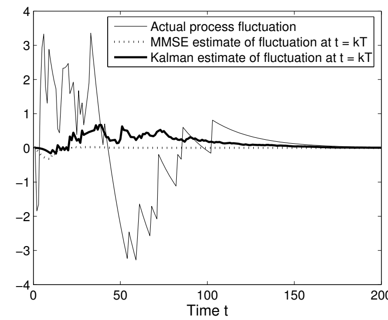

Fig. 2. MMSE estimation and Kalman filter estimation of the process fluctuations for Scenario 1.

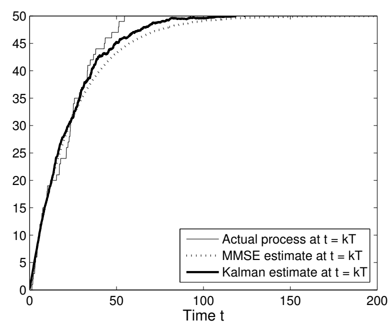

Fig. 3. Performance of MMSE estimation and Kalman filter estimation of the process for Scenario 2.

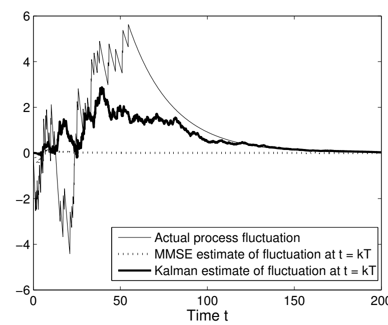

Fig. 4. MMSE estimation and Kalman filter estimation of the process fluctuations for Scenario 2.

$\mu = 10\beta$, and (3) **Scenario 3:** $N_0 = 50$, $\beta = 0.02$, $T = 0.01$, $\mu = 100\beta$, (4) **Scenario 4:** $N_0 = 200$, $\beta = 0.02$, $T = 1.0$, $\mu = \beta$, (5) **Scenario 5:** $N_0 = 1000$, $\beta = 0.02$, $T = 1.0$, $\mu = \beta$. In all scenarios, we have $S_0 = 2$.

In Figure 1 we depict the performance of the MMSE estimator and the Kalman filter for Scenario 1. We note that the estimations by both the MMSE estimator and the Kalman filter are very close to each other and indeed close to the fluid approximation $N_0x(t)$ of $X(t)$. In Figure 2, we show the estimations of the fluctuations about the fluid limit for Scenario 1, and notice that neither the MMSE estimator nor the Kalman filter is able to successfully track the fluctuations in this scenario. We suspect that the inability to track the fluctuations in Scenario 1 is primarily due to the insufficiency of measurement data.

To verify if the inability to track the fluctuations in Scenario 1 is indeed due to the insufficiency of measurement data, we examine Scenario 2 (Figures 3 and 4) and Scenario 3 (Figures 5 and 6). In Scenario 2 (resp. Scenario 3), we increase

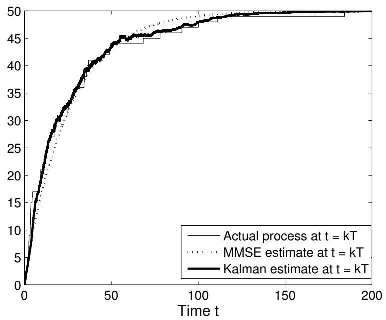

Fig. 5. Performance of MMSE estimation and Kalman filter estimation of the process for Scenario 3.

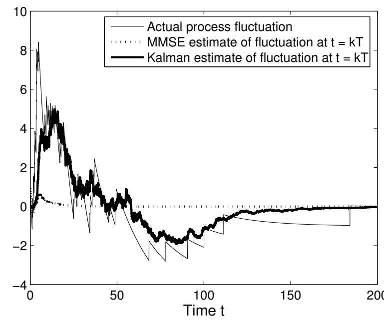

Fig. 6. MMSE estimation and Kalman filter estimation of the process fluctuations for Scenario 3.

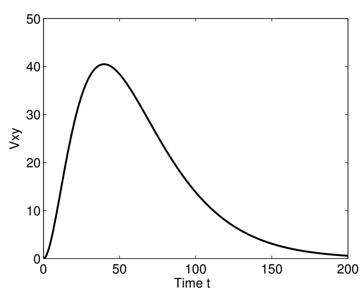

Fig. 7. $V_{xy}(t)$ as a function of time.

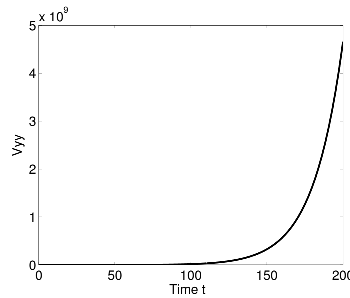

Fig. 8. $V_{yy}(t)$ as a function of time.

the rate μ at which measurements are taken by the observer by a factor 10 (resp. 100). To make better use of faster measurements and avoid smoothening of measurement data over longer time intervals, we also decrease the monitoring interval T by the same factor. We observe that the performance of the Kalman filter is much improved in Scenario 2 with faster measurements during the later phase of spreading (compare Figure 4 with Figure 2). Comparing Figure 6 with Figures 4 and 2, we observe that tracking of the fluctuations by the Kalman filter is extremely accurate in Scenario 3. This accurate tracking of the fluctuations results in extremely accurate tracking of the process itself (see Figure 5).

In Figures 3-6, we observe that the MMSE estimator fails to make use of faster measurements. In fact, it stays very close to the fluid approximation. This can be explained as follows. The MMSE estimation differs from the fluid approximation $N_0x(t)$ (which is equal to $m_x(t)$) by the term $V_{xy}(t)V_{yy}(t)^{-1}(Y(t)-m_y(t))$. From the expressions for $V_{xy}(t)$ and $V_{yy}(t)$ in Section 3.2, it can be seen that $V_{xy}(t)$ initially increases sublinearly, but then quickly decreases exponentially with t (see Figure 7), and $V_{yy}(t)$ increases superlinearly with t (see Figure 8). Thus, except for an initial phase, the effect of the measurement $(Y(t) - m_y(t))$ is

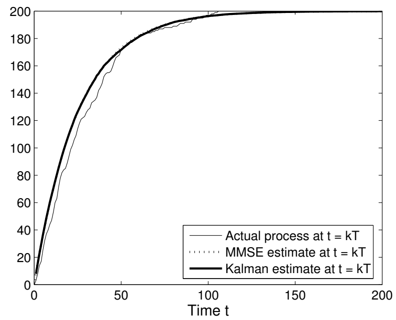

Fig. 9. Performance of MMSE estimation and Kalman filter estimation of the process with Scenario 4.

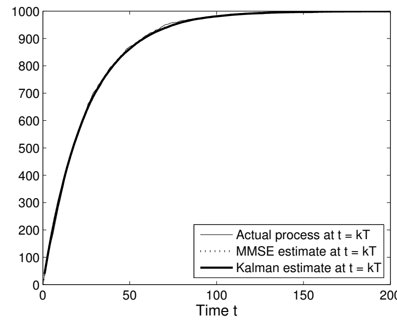

Fig. 10. Performance of MMSE estimation and Kalman filter estimation of the process with Scenario 5.

diminished by the factor $V_{xy}(t)V_{yy}(t)^{-1}$. Increasing μ by a factor K increases $V_{xy}(t)$ by a factor K, but also increases $V_{yy}(t)$ by a factor K^2. Thus, increasing μ by a factor K results in an overall attenuation of the measurement $(Y(t) - m_y(t))$ by a factor K (see Section 3.2). Furthermore, the difference between the measurement $Y(t)$ and its mean $m_y(t)$ also decreases with t. In summary, we can expect the performance of the MMSE estimator to get worse with time.

Next, we examine the situations in which the fluid approximation itself can be used as a good predictor. Suppose that we increase the area of the network by a factor K keeping the density of nodes constant. Then, we increase both S_0 and N_0 by K, but decrease the source-relay meeting rate β by K. Then, the net rate at which meetings occur in the network increases from $S_0N_0\beta = \lambda N_0$ to $KS_0N_0\beta = \lambda(KN_0)$. This scaling is equivalent to increasing only N_0 by a factor K keeping S_0 and β constant as in (10). Thus, if the area of the network is large so that N_0 is large, then the fluid model can be a good predictor. We demonstrate this by Figures 9 and 10 which correspond to Scenarios 4 and 5, respectively. Note that Scenarios 4 and 5 are derived from Scenario 1 by scaling as above with a scaling factor $K = 4$ and $K = 20$, respectively. Comparing Figures 1, 9 and 10, we observe that the process becomes smoother and closer to the fluid approximation with increase in the number of nodes N_0.

Level-Crossing Times: Next, we compare the accuracy of the MMSE and the Kalman estimators in estimating the level crossing times by computing the percentage error w.r.t. the level crossing times of the actual process and averaging over 100 runs. Fixing the threshold levels at $X_L = 0.15N_0, 0.25N_0, 0.50N_0, 0.75N_0$, and $0.90N_0$, we obtained average percentage errors for estimates of level crossing times by the MMSE and the Kalman estimators for Scenario 3. We summarize the results as follows: $X_L = 0.15N_0, e(MMSE) = 25.32\%, e(Kalman) = 23.72\%$; $X_L = 0.25N_0, e(MMSE) = 22.63\%, e(Kalman) = 16.07\%$; $X_L = 0.50N_0, e(MMSE) = 14.98\%, e(Kalman) = 8.32\%$; $X_L = 0.75N_0, e(MMSE) = 12.71\%, e(Kalman) = 7.33\%$; $X_L = 0.90N_0, e(MMSE) = 14.80\%, e(Kalman) =$

9.94%. We conducted similar experiments (not reported here due to lack of space) with different parameter settings and observed similar trends. The Kalman filter shows a slightly better performance than MMSE during the initial phase of spreading when the threshold level is small (say, $X_L \leq 0.15N_0$). However, the Kalman filter strictly outperforms the MMSE estimator for higher threshold levels (say, $X_L \geq 0.25N_0$) because it takes into account all previous sample measures.

5 Conclusion

In this paper, we tackled the problem of estimating file-spread in DTNs with two-hop routing. Apart from providing solid analytical basis to our estimation framework, we also provided insightful conclusions validated with simulations. Some of the important insights are: (i) the deterministic fluid model cam indeed be a good predictor with a large number of nodes, (ii) the Kalman filter can track the spreading process quite accurately provided that measurements as well as updates are taken sufficiently fast, (iii) during the initial phase of spreading when the amount of sample measures is still low, the MMSE estimator can be used for estimating the level crossing times of sufficiently low threshold levels, and (iv) as time progresses, the MMSE estimator becomes less useful, but the Kalman filter would be available at later phases to provide accurate estimates. Applying the real-time estimations for online adaptive control of the spreading process is a topic of our ongoing research.

Appendix

In this appendix, we provide a brief informal background on fluid and diffusion limits and approximations. Please refer to [6], [7] and [12] for more details.

Intuitively speaking, the fluid approximation provides the first-order deterministic approximation to a stochastic process and represents its *average* behavior. The diffusion approximation provides the second-order approximation to a stochastic process representing its average behavior added with *random fluctuations about the average* (usually, in terms of a Brownian motion).

Consider a sequence $\{Z^{(n)}(t), t \geq 0\}$, $n = 1, 2, \ldots$, of stochastic processes. Index n represents some quantity which is *scaled up* to infinity in order to study the sequence of processes at the limit, as $n \uparrow \infty$. For queueing systems, n might represent "the number of servers" (as in infinite server approximations) or "a multiplying factor of one or more transition rates" (as in heavy-traffic approximations) or some other quantity w.r.t. which the scaling is performed.

Consider the Strong Law of Large Numbers (SLLN) type rescaling $z^{(n)}(t) := Z^{(n)}(t)/n$. Under certain conditions, as $n \uparrow \infty$, the sequence of rescaled processes $\{z^{(n)}(t), t \geq 0\}$, $n = 1, 2, \ldots$, converges almost surely to a deterministic process $\{z(t), t \geq 0\}$ (see, for example, Theorem 4.1 of [7]). Then, the limit $\{z(t), t \geq 0\}$ is called the *fluid limit* associated with the sequence $\{Z^{(n)}(t), t \geq 0\}$, $n = 1, 2, \ldots$, and the approximation

$$Z^{(n)}(t) \approx nz(t), \quad \forall t \geq 0, \tag{28}$$

is called the *fluid approximation* for the n-th system.

Consider now the Central Limit Theorem (CLT) type rescaling $v_z^{(n)}(t) = \sqrt{n}(z^{(n)}(t) - z(t))$, which amplifies the deviation of the rescaled process $\{z^{(n)}(t), t \geq 0\}$ from the fluid limit $\{z(t), t \geq 0\}$. Under certain conditions, as $n \uparrow \infty$, the sequence of rescaled processes $\{v_z^{(n)}(t), t \geq 0\}$, $n = 1, 2, \ldots$, converges weakly to a *diffusion process* (or a continuous-time Markov process with continuous sample paths) $\{v_z(t), t \geq 0\}$ (see, for example, Theorem 4.2 of [7]). Then, $v_z(t)$ is called the *diffusion limit* associated with the sequence $\{Z^{(n)}(t), t \geq 0\}$, $n = 1, 2, \ldots$, and the approximation

$$Z^{(n)}(t) \stackrel{d}{\approx} nz(t) + \sqrt{n}v_z(t), \quad \forall t \geq 0, \tag{29}$$

is called the *diffusion approximation* for the n-th system, where $\stackrel{d}{\approx}$ means "approximately distributed as". In particular, if $v_z(0)$ is a Gaussian random variable, then $\{v_z(t), t \geq 0\}$ is a Gaussian process and it is completely characterized by its mean and auto-covariance functions.

References

1. Alouf, S., Altman, E., Nain, P.: Optimal on-line estimation of the size of a dynamic multicast group. In: Infocom'02. vol. 2, pp. 1109–1118 (June 2002)
2. Bremaud, P.: Point Processes and Queues: Martingale Dynamics. Springer (1981)
3. Karatzas, I., Shreve, S.E.: Brownian Motion and Stochastic Calculus. Springer-Verlag, New York (1988)
4. Kurtz, T.G.: Solutions of Ordinary Differential Equations as Limits of Pure Jump Markov Processes. Journal of Applied Probability 7(1), 49–58 (1970)
5. Lindgren, A., Doria, A., Schelen, O.: Probabilistic Routing in Intermittently Connected Networks. In: ACM Mobicom (2003)
6. Mandelbaum, A., Massey, W., Reiman, M.I.: Strong approximations for markovian service networks. Queueing Systems 30, 149–201 (1998)
7. Mandelbaum, A., Pats, G.: State-dependent queues: approximations and applications. In: Kelly, F., Williams, R.J. (eds.) IMA Volumes in Mathematics and Its Applications, vol. 71, pp. 239–282. Springer-Verlag, Berlin (1995)
8. Poor, H.V.: An Introduction to Signal Detection and Estimation. Springer Verlag, second edition edn. (1994)
9. Simon, D.: Optimal State Estimation. John Wiley & Sons, Inc. (2006)
10. Spyropoulos, T., Psounis, K., Raghavendra, C.: Efficient Routing in Intermittently Connected Mobile Networks: The Multi-copy Case. IEEE/ACM Transaction on Networking 16(1), 63–76 (Feb 2008)
11. Vahdat, A., Becker, D.: Epidemic Routing for Partially-Connected Ad Hoc Networks. In: Duke University, Tech. Rep. CS-200006 (2000)
12. Whitt, W.: Stochastic-Process Limits. Springer (2002)
13. Zhang, X., Neglia, G., Kurose, J., Towsley, D.: Performance Modeling of Epidemic Routing. Computer Networks 51, 2867–2891 (2007)