Performance Modelling of Opportunistic Forwarding with Imprecise Knowledge

Chiara Boldrini, Marco Conti, Andrea Passarella

Performance Modelling of Opportunistic Forwarding with Imprecise Knowledge

Chiara Boldrini, Marco Conti, and Andrea Passarella

IIT-CNR, Pisa, Italy. Email: firstname.lastname@iit.cnr.it

***Abstract*— Mobility-assisted networking is becoming very popular as a mean of delivering messages in disconnected or very dynamic networks, such as opportunistic networks. Despite the rapid growth in the number of proposals for routing protocols that exploit the mobility of nodes, there is a lack of general theoretical frameworks to be used for studying analytically their performance under different mobility conditions (e.g., exponential or Pareto inter-meeting times). Moreover, one of the main approaches to forwarding (so-called utility-based forwarding) consists in nodes collecting statistics about their behaviour (e.g., their contact patterns), and using this information to guide the forwarding process. Thus, a general theoretical framework should also be able to model the fact that the statistics collected by nodes and used to make forwarding decisions might suffer from estimation errors. In order to fill these gaps, in this paper we propose an analytical framework for the single-copy forwarding process in a mobility-assisted network that has the following characteristics: (i) it provides a closed form solution for a large class of probability distributions representing inter-meeting times, (ii) it is able to model both randomized and utility-based forwarding protocols, and (iii) it accounts for errors in the estimations of the utility values used by utility-based schemes for making forwarding decisions. We show that the framework is quite accurate and that it can be used to identify the most effective forwarding policies depending on the amount of estimation errors in the forwarding statistics.**

I. Introduction

Mobility-assisted networking is rapidly becoming very popular as a mean of delivering messages in disconnected or very dynamic networks, as in the case of opportunistic networks [1]. Opportunistic networks are networks made up of pocket devices (smartphones, tablets, etc.) that users carry with them during their normal daily routine. In opportunistic networks, messages are delivered along a multi-hop ad hoc path across the users of the network. More specifically, according to the store-carry-and-forward paradigm, nodes store messages locally and carry them while moving around, until they find a node, to which they hand over the message, deemed more suitable to reach the intended destination.

Forwarding strategies define the policy according to which messages are handed over or not upon encounter between two nodes. Two main approaches are typically identified in the literature: randomized forwarding and utility-based forwarding. With randomized forwarding, during each pairwise encounter, a message has a fixed probability $p \in [0, 1]$ to be handed over, regardless of the forwarding capabilities of the two nodes. This is the case, for example, of the Two Hop and Direct Transmission protocols [2]. With utility-based forwarding, a message is forwarded from node i to node j with a probability p_{ij} that is equal to either 1, when the utility (or fitness) of node j as a forwarder for the message is higher than that of node i, or 0, otherwise. For example, the utility can be measured taking into account how often nodes meet with the destination [3] or how much contextual information they share [4] (e.g., same friends, same visited places, or similar movement patterns).

Utility-based forwarding critically relies on the estimation of the utility parameters. As an example, assume that the utility of a node i for a message with destination d is given by the frequency μ_{id} of encounter with d. This frequency is estimated online by node i but, due to unpredictable errors (e.g., missed contacts, few contacts, etc.), the estimated value $\hat{\mu}_{id}$ can be different from the actual value μ_{id}. Thus, subject to stochastic fluctuations, it can happen that, for two generic nodes i and j, the ordering between the estimated utility values is different from the ordering between the actual values, and this would lead to wrong forwarding decisions.

The aim of this work is twofold. First, we discuss how estimation errors can be modeled for two reference utility-based forwarding strategies. Introducing estimation errors in the framework allows us to compare various forwarding approaches not only with respect to the amount of knowledge they exploit, but also with respect to the sensitivity to the accuracy of the parameters used to represent this knowledge. For example, we obtain that the intuitive result that utility-based forwarding outperforms randomised forwarding holds true only if the estimation accuracy of utility parameters is above a certain threshold. As a second contribution, we show how estimation errors can be embedded into a general framework that allows us to model the performance of single-copy opportunistic forwarding protocols, for a large class of forwarding policies and distributions of inter-meeting times. As messages are exchanged upon pairwise encounters, the time intervals between consecutive meetings of a pair of nodes (*inter-meeting times*) play a major role in the delay experienced by messages. For this reason, they are the main building block of any analytical performance model for opportunistic networks. However, since no general agreement has been reached so far about which distribution better represents the properties of inter-meeting times, we believe that having an analytical framework that can be used with different distributions is of primary importance.

The paper is organized as follows. In Section II we survey

978-3-901882-45-6/2012 - Copyright is with IFIP

the related work in the literature, while in Section III we describe our reference network model. Then, in Section IV we discuss how estimation errors can be modeled and in Section V we introduce our general analytical framework for modeling the forwarding process in an opportunistic network. Finally, in Section VI the model evaluation is provided.

II. Related work

A common classification of forwarding protocols for mobility-assisted networks breaks down algorithms into randomized and utility-based forwarding schemes. In randomized schemes messages are handed over by a tagged node i to the first encountered node j with a fixed forwarding probability $p_{ij} \in [0, 1]$ that is not dependent on the actual ability of node j to deliver the message. For example, this probability is always equal to zero but for the source-destination node pair ($p_{sd} = 1$) in the Direct Transmission forwarding strategy [2]. Thus, in this case, the source node is only allowed to deliver the message directly to the destination. Vice versa, the forwarding probability is always equal to 1 in the Epidemic protocol [5], implying that the message is handed over to the first encountered node every time. An intermediate approach between these two extremes is that of the Two Hop forwarding scheme [2], in which the source node hands over the message to the first encountered node ($p_{sj} = 1, \forall j$), but this intermediate node is only allowed to deliver the message directly to the destination ($p_{jd} = 1, p_{ji} = 0, \forall i \neq d$). Finally, the forwarding probability can take values between 0 and 1 when using a gossiping strategy [6], whose idea is to mimic the way rumors spread in real social networks.

Alongside randomized strategies there are utility-based schemes. In short, with utility-based schemes a message is handed over from node i to node j upon encounter only if the utility that node j brings to the message is higher than that brought by node i. The utility of a node as forwarder measures what are the chances that it improves, or speeds up, the delivery towards the destination. Utility-based schemes differentiate in how they define the utility of a relay. Network-level information such as time since the last encounter (Spray&Focus [7]) or frequency of encounters (PROPHET [3]) can be used. The intentional exploitation of the social component in user mobility is allowed by social-aware utility-based protocols, in which either information on the social graph describing relationships between users (BUBBLE [8], SimBet [9]) or so-called contextual information (HiBOp [4], SocialCast [10]) is exploited in order to quantify the ability of nodes to deliver messages. All the above schemes require the online estimation of the parameters (e.g., frequency of encounters) used to compute the forwarding utility. When these estimations are flawed, clearly the protocols do not behave as expected and this might lead to wrong forwarding decisions.

From the modeling standpoint, we are not aware of existing contributions that consider the effects of imprecise estimations on the performance of forwarding protocols for opportunistic networks. Among the existing models that assume accurate estimations, the vast majority rely on the i.i.d. [11], [12] or i.n.i.d. [13]–[15] exponential inter-meeting times assumption. To the best of our knowledge, the only contributions that consider inter-meeting times following a distribution different from the exponential, and their effects on the expected delay of messages, are [16] and [17]. However, in [16] only i.i.d. Pareto inter-meeting times have been studied, while [17] was exclusively focused on deriving conditions for the convergence of the expected delay. This brief literature overview highlights the need for new analytical models able to compute the expected delay, or other fundamental forwarding metrics, when utility statistics are imprecise and inter-meeting times are general i.n.i.d. random variables.

M_{ij}	inter-meeting time for the i, j node pair
R_{ij}	residual inter-meeting time for the i, j node pair
M_i	inter-any-meeting time for node i
R_i	residual inter-any-meeting time for node i
μ_{ij}	contact rate for the i, j node pair
μ_i	contact rate for inter-any-meeting time M_i
$\hat{\mu}_{ij}$	contact rate for the i, j node pair resulting from an online estimation process
p_{ij}	transition probabilities of the forwarding Markov process
p_{ij}^{φ}	probability that node i hands over the message to node j upon encounter when forwarding policy φ is in use
T_i^{exit}	time before node i hands over the message to any other node or, equivalently, time before the forwarding Markov process exits from state i
D_i^d	delay of a message generated by node i and addressed to node d
$\mathcal{P}_i$	set comprising all nodes that can be encountered by node i

TABLE I
NOTATION

III. Network model

In this paper, we consider a network of N mobile nodes. We assume the following delivery process: upon encounter, two nodes can exchange messages, which (analogously to the bundle in DTN terminology) are atomic units that cannot be fragmented. The actual messages that are exchanged depend on the forwarding policy used. In order to isolate and highlight the effect of node mobility from other effects, we make the following assumptions. First, we assume that messages can be exchanged only at the beginning of a contact and that the transmission of the relayed messages can be always completed within the duration of a contact. Second, we assume that nodes have infinite buffer space. All the above assumptions are common in the literature on opportunistic networks modelling and ensure that the resulting analytical model is tractable.

Under these assumptions, the delay of messages in an opportunistic network only depends on the mobility of nodes and on the forwarding strategy used to route messages. As far as mobility is concerned, throughout the paper we will rely on the concept of inter(-any)-meeting time and residual inter(-any)-meeting time, whose definitions are provided below. The *inter-meeting time* M_{ij} between node i and node j is defined as the time between two consecutive meetings between the same pair of nodes. The inter-any-meeting time M_i for node i is defined as the time between two consecutive meetings between node i and *any* other node j. By definition, the inverse of the expectation of inter(-any)-meeting times gives the inter(-any)-meeting rates. The inter(-any)-meeting rate is denoted as μ_{ij} (μ_i) in the following. For the sake of simplicity, we assume that inter(-any)-meeting rates do not vary with time. As we assume that inter-meeting times for each fixed node pair i, j are independent and identically distributed, the meeting process between node i and node j can be modelled as a renewal process.

The concept of residual time comes into the picture because, in general, the message generation process and the meeting process are asynchronous. This means that the time at which a message is generated by a generic node i can be considered as a random point in time with respect to the evolution of the contact process between i and any other node. Thus, starting from its generation, this message has to wait for at least a residual inter-meeting time before being handed over to another node. Assuming that node i and node j are not in contact at a generic time t_r, the residual inter-meeting time $R_{ij}(t)$ between them is defined as the time interval between t_r and the first time node i and node j come into contact again. From this, the definition of the residual inter-any-meeting times follows straightforwardly.

The notation that we use throughout the paper is summarized in Table I.

A. Forwarding policies

In this work, for the utility-based schemes, we focus on two reference strategies, namely Direct Acquaintance (DA) and Social Forwarding (SF), which exemplify key aspects of utility-based forwarding protocols available in the literature. In fact, each utility-based scheme defines a criterion for classifying how good a given node is as relay for a specific destination. Based on this criterion, utility-based schemes derive what we call *fitness*, i.e., a measure of how fit the node is as relay. Let us denote with $f^{\varphi}_{i,d}$ the fitness, measured according to forwarding strategy φ, of node i as relay for messages with destination d. Upon encounter with node j, node i will hand over the message to j only if $f^{\varphi}_{j,d}$ is greater than $f^{\varphi}_{i,d}$. The algorithm for computing the fitness of a generic node i as relay can range from very simple to quite complex. Its specific definition is out of the scope of the paper, since our main goal is to define a general framework and to provide some significant examples of application. For this reason, we have identified two utility-based strategies that abstract the main features of the proposals available in the literature in terms of the extent of the information exploited. The simplified utility-based policies that we use are defined below. The advantage of the Social Forwarding strategy with respect to Direct Acquaintance is that the former is able to capture also the component of the fitness associated with the transitivity of encounters.

Definition 1 (Direct Acquaintance): The source and each intermediate relay hand over the message to the first encountered node having a higher fitness, where the fitness $f^{DA}_{i,d}$ of a generic node i for a message with destination d is defined as the estimated frequency $\hat{\mu}_{id}$ of a direct meeting with the destination d ($f^{DA}_{i,d} = \hat{\mu}_{i,d}, \forall i \neq d$).

Definition 2 (Social Forwarding): Messages are delivered through a path with positive gradient of fitness, where the fitness $f^{SF}_{i,d}$ of node i for a message addressed to node d is computed as the weighted sum of the fitness $f^{DA}_{i,d}$ for a direct acquaintance and the fitness $f^{I}_{i,d}$ for an indirect meeting ($f^{SF}_{i,d} = \xi f^{DA}_{i,d} + (1-\xi) f^{I}_{i,d}, \quad \text{where } 0 < \xi < 1$). Component $f^{I}_{i,d}$ is a measure of the probability of being indirectly connected to the destination or, in other words, of the likelihood of being connected to nodes that have high delivery probability for destination d. In the general case, it can be recursively defined as the average of the fitness of the encountered nodes, which implies $f^{I}_{i,d} = \frac{1}{|\mathcal{P}_i|} \sum_{j \in \mathcal{P}_i} \gamma f^{DA}_{j,d} + (1-\gamma) f^{I}_{j,d}$, where $\gamma \in [0, 1]$ prioritizes either direct acquaintance or the indirect fitness and $\mathcal{P}_i$ denotes the set of nodes that can be encountered by node i.

Parameter γ is a weight that can be tuned in order to prioritize what neighbour j directly sees ($\gamma \to 1$, in this case) or what the neighbours of j see ($\gamma \to 0$, in this case). Parameter γ can be in general different from ξ in order to weight differently the fitness values associated directly with node i itself and those related to its neighbours. For the sake of simplicity, in the following we assume $\gamma = 1$.

IV. Modelling Estimation Errors

The fitness values discussed above are estimated by nodes exploiting some control information that they exchange or infer upon encounter. However, these estimations might be affected by the length of the neighbor discovery interval and by the properties of the different technologies with which user devices communicate with each other (Bluetooth, WiFi, etc.). Thus, some encounters may be missed, some others erroneously detected, repeated short contacts might be considered a single long contact, while an actual long contact may be split into smaller ones. Estimations might also suffer from memory limitations that force to store no more than n bites of data. Thus, overall, it is unlikely that the estimated fitness value will match exactly the actual value.

In this work, we take into account estimation errors by representing them as random errors. Thus, for a forwarding policy φ, considering our forwarding fitness as a function of a set of estimated parameters $\hat{\pi}_1, ..., \hat{\pi}_m$, i.e., $f^{\varphi}_{i,d} = g_{\varphi}(\hat{\pi}_1, ..., \hat{\pi}_m)$, random errors can be accounted for by considering each $\hat{\pi}_z$ as drawn from a Normal distribution with mean π_z, i.e, the actual value of the parameter, and variance σ^2_z, for each z.

In the case of the utility-based strategies defined in Section III-A, the forwarding fitness is a function of the estimated meeting rates between nodes. Thus, measurement errors on such rates can be taken into account by modelling them as $\hat{\mu}_{ij} \sim \mathcal{N}(\mu_{ij}, \sigma^2_{ij})$. Let us assume that node i meets another node j and that node i has some outstanding messages. Node i has thus to make forwarding decisions about whether or not to hand them over to j. We call pairwise forwarding probability p^{φ}_{ij} the probability that, upon encounter, node i hands over to j messages with destination d. When estimation errors are present, the estimated inter-meeting rates $\hat{\mu}_{id}$ will be in general different from the exact μ_{id}. Thus, even when $\mu_{id} > \mu_{jd}$, there is a chance that node i hands over these messages to node j. Theorems 1 and 2 give the probability of this event under the Direct Acquaintance and Social Forwarding policy.

Theorem 1 (p_{ij}^{DA} for Direct Acquaintance): When the destination node is d and estimation errors are modelled as random errors, the pairwise forwarding probability p_{ij}^{DA} is given by the following:

$$p_{ij}^{DA} = \frac{1}{2}\left[1 + \text{Erf}\left(\frac{\mu_{jd} - \mu_{id}}{\sqrt{2(\sigma_{id}^2 + \sigma_{jd}^2)}}\right)\right] \tag{1}$$

Proof: Let us consider a node i that is deciding whether to forward to node j a message addressed to node d, and is using forwarding strategy φ. Under the assumption that $\hat{\mu}_{ij} \sim \mathcal{N}(\mu_{ij}, \sigma_{ij}^2)$, according to Definition 1, $f_{i,d}^{DA} \sim N(\mu_{id}, \sigma_{id}^2)$, and $f_{j,d}^{DA} \sim N(\mu_{jd}, \sigma_{jd}^2)$. The probability that node i hands over the message to node j is equal to $P(f_{i,d}^{DA} < f_{j,d}^{DA}) = P(f_{i,d}^{DA} - f_{j,d}^{DA} < 0)$. The distribution of the difference $f_{i,d}^{DA} - f_{j,d}^{DA}$ of two independent Normal random variables is again a Normal random variable, with mean $\mu_{id} - \mu_{jd}$ and variance $\sigma_{id}^2 + \sigma_{jd}^2$. Then, Equation 1 results from evaluating the CDF of the difference $f_{i,d}^{DA} - f_{j,d}^{DA}$ in zero. ■

Now we compute the pairwise forwarding probability p_{ij}^{φ} under the Social Forwarding scheme.

Theorem 2 (p_{ij}^{SF} for Social Forwarding): The forwarding probability p_{ij}^{SF} when the destination is d is given by $p_{ij}^{SF} = \frac{1}{2}\left[1 + \text{Erf}\left(\frac{\mu_{j,d}^{SF} - \mu_{i,d}^{SF}}{\sqrt{2*[(\sigma_{i,d}^{SF})^2 + (\sigma_{j,d}^{SF})^2]}}\right)\right]$. Expectation $\mu_{i,d}^{SF}$ and variance $\sigma_{i,d}^{SF}$ for node i with respect to node j are equal to $\mu_{id}^{SF} = \xi\mu_{id} + \frac{(1-\xi)\sum_{z\in\mathcal{P}_i^j}\mu_{zd}}{|\mathcal{P}_i|}$ and $(\sigma_{id}^{SF})^2 = \xi^2\sigma_{id}^2 + \frac{(1-\xi)^2\sum_{z\in\mathcal{P}_i^j}\sigma_{zd}^2}{|\mathcal{P}_i|^2}$, where $\mathcal{P}_i$ denotes the set of nodes encountered by i, and ξ is a configurable weight. An analogous expression holds for node j with respect to node i.

Proof: First, we derive the fitness $f_{i,d}^{SF}$ for the Social Forwarding policy. From Definition 2 we know that fitness $f_{i,d}^{SF}$ has two components, namely $f_{i,d}^{DA}$ and $f_{i,d}^{I}$. As explained in the proof of Theorem 1, $f_{i,d}^{DA} \sim N(\mu_{id}, \sigma_{id}^2)$. As for $f_{i,d}^{I}$, it being defined as the arithmetic mean of $f_{z,d}^{DA}$ for all nodes z encountered by node i, again we have that $f_{i,d}^{I}$ follows a normal distribution. In fact, the arithmetic mean is just the sum of the $f_{z,d}^{DA}$ values divided by a constant (the number of peers in $\mathcal{P}_i$). From standard probability theory we know that the sum $X = \sum_i X_i$, where X_i is a normal random variable with expectation μ_i and variance σ_i^2 is again a normal random variable with expectation $\sum_i \mu_i$ and variance $\sum_i \sigma_i^2$. In addition, for a normal random variable $\mathcal{N}(\mu, \sigma^2)$ the following holds true: $a\mathcal{N}(\mu, \sigma^2) = \mathcal{N}(a\mu, a^2\sigma^2)$. Thus, for the arithmetic mean of $|\mathcal{P}_i|$ normal random variables we have:

$$f_{i,d}^{I} = \mathcal{N}\left(\frac{\sum_{z\in\mathcal{P}_i}\mu_{zd}}{|\mathcal{P}_i|}, \frac{\sum_{z\in\mathcal{P}_i}\sigma_{zd}^2}{|\mathcal{P}_i|^2}\right). \tag{2}$$

In fitness $f_{i,d}^{SF}$, both $f_{i,d}^{DA}$ and $f_{i,d}^{I}$ are multiplied by a constant, ξ and $(1-\xi)$ respectively, and then added together. Recursively applying basic properties of normal random variables, we obtain the following:

$$f_{i,d}^{SF} = \mathcal{N}\left(\mu_{id}^{SF}, (\sigma_{id}^{SF})^2\right) \tag{3}$$

where $\mu_{id}^{SF} = \xi\mu_{id} + \frac{(1-\xi)\sum_{z\in\mathcal{P}_i}\mu_{zd}}{|\mathcal{P}_i|}$, $(\sigma_{id}^{SF})^2 = \xi^2\sigma_{id}^2 + \frac{(1-\xi)^2\sum_{z\in\mathcal{P}_i}\sigma_{zd}^2}{|\mathcal{P}_i|^2}$, and $\mathcal{P}_i$ is defined as the set of peers encountered by node i. Please note that our simplifying assumption of parameter $\gamma = 1$ in Definition 2 does not affect our results. In fact, $\gamma < 1$ only involves additional weighted sums of Normal random variables.

In this second part of the proof we use the above results in order to derive the pairwise forwarding probability p_{ij}^{SF}. To this aim, let us consider a node i that is deciding whether to forward to node j a message addressed to node d. The probability that node i hands over the message to node j is equal to $P(f_{i,d}^{SF} < f_{j,d}^{SF}) = P(f_{i,d}^{SF} - f_{j,d}^{SF} < 0)$. Differently from Theorem 1, here $f_{i,d}^{SF}$ and $f_{j,d}^{SF}$ may be correlated. In fact, peers that both i and j meet contribute to the indirect fitness values $f_{i,d}^{I}$ and $f_{j,d}^{I}$. Since it is not trivial to evaluate the impact of correlation, in this work, as a first approximation, we decided to neglect it. Thus, Equation 3 simply follows after applying the properties of the difference between two independent Normal random variables. ■

Now that we have defined p_{ij}^{DA} and p_{ij}^{SF} for our utility-based policies in the case of errors in the estimated fitness, we use them in the analytical model that we discuss below.

V. The Framework

We use a semi-Markov process with N states to model the opportunistic forwarding process. A semi-Markov process is one that changes state in accordance with a Markov chain (called *embedded* or *jump* chain) but where transitions between states can take a random amount of time with an arbitrary distribution. As such, it is fully described by the transition matrix associated with its embedded chain and by $T_i^{exit}, \forall i = 0, \cdots, N$, where T_i^{exit} denotes the distribution of the time that the semi-Markov process spends in state i before making a transition.

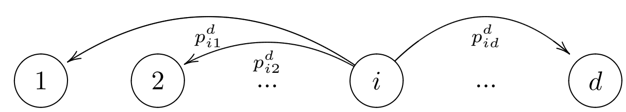


Fig. 1. Fragment of the embedded markov chain (valid for all $i \neq d$)

We express our semi-Markov process associated with the single-copy message forwarding process in terms of the embedded Markov chain in Figure 1. Assuming that node i is currently holding a message whose destination[1] is d, the probability p_{ij}^d that node i will delegate the forwarding of the message to another node j is a function of both the likelihood of meeting node j and the probability that node i will hand

[1]The chain is different for different destinations, because the useful relays are generally not the same. However, for the sake of readability, in the following we drop superscript d.

over the message to node j according to the forwarding policy in use. The state associated with the destination node d is absorbing, because in state d the forwarding process is completed.

Once the forwarding Markov process is completely defined in terms of transition probabilities and exit times, we can exploit well known algorithms for Markov chain transient analysis in order to compute significant properties of the forwarding process. For example, the expected delay $E[D_s^d]$ from node s to node d can be computed (Equation 4) as the expected hitting time on state d starting from the source node s. In Equation 4, T_i^{exit} denotes the time before the message leaves node i and p_{ij} the probability that the message is handed over to node j by node i. Please note that p_{ij} is different from p_{ij}^{φ}, since the latter is conditioned on the fact that the meeting is with node j, while the former accounts also for the different meeting probabilities with the peers.

$$\begin{cases} E[D_i^d] = 0 & i = d \\ E[D_i^d] = E[T_i^{exit}] + \sum_{j \neq d} p_{ij} E[D_j^d] & \forall i \neq d. \end{cases} \tag{4}$$

The key step in order to solve the system in Equation 4 is to derive $E[T_i^{exit}]$ and p_{ij}. Unfortunately, computing $E[T_i^{exit}]$ can be prohibitive when inter-meeting times follow distributions different from the simple exponential. Intuitively, the problem is that T_i^{exit} can be obtained as the minimum of the time T_{ij}^{exit} that takes to node i to forward the message to each potential next hop j. Then, each T_{ij}^{exit} is the sum of the residual inter-meeting time before node i meets node j and the inter-meeting time between node i and j taken a geometrically distributed number of times (because, at each encounter, node i can hand over the message to node j with a certain probability). In the end, we get that T_i^{exit} is given by the minimum of a weighted sum of random variables, which rarely has a closed form (not even for the expectation). For example, it has no closed form for the Pareto and Pareto with exponential cut-off inter-meeting times, cases that have been often found in real mobility traces. Thus, in this work we propose an approximate model that relies on the concept of inter-any and residual inter-any meeting times. The advantage of this model is that we get rid of the minimum by modeling the time before the next encounter with *any* other node, and assuming that the probability that exactly node j is the next encounter is equivalent to the long run proportion between the rate of encounters between node i and node j and the rate of encounter of i with any node. The disadvantage is that we intentionally neglect the impact of the memory of probability distributions. When there is indeed no memory, as in the case of exponential distribution, the proposed framework is exact. When there is memory, the framework is approximated. The error introduced by the approximation is evaluated in Section VI in the case of Pareto distributed inter-meeting times.

In the following we provide a general formulation for both $E[T_i^{exit}]$ and p_{ij} in terms of the inter-any-meeting time M_i and the residual inter-any-meeting time R_i (proofs can be found in the appendix). This general formulation will be specialized later in the paper based on the distribution of inter-any-meeting times considered, showing that the proposed approximate model can be conveniently used also with distributions difficult to deal with, like the Pareto and Pareto with exponential cut-off.

Theorem 3 (Expected exit time T_i^{exit}): The expectation of the exit time T_i^{exit}, i.e., the time required for the chain to exit from state i when the forwarding policy φ is used, is given by the following:

$$E[T_i^{exit}] = E[R_i] + \left(\frac{1 - p_i^{\varphi}}{p_i^{\varphi}}\right) E[M_i], \tag{5}$$

where p_i^{φ} is equal to $\sum_{j \in \mathcal{P}_i} p_{ij}^{\varphi} * \frac{\mu_{ij}}{\mu_i}$, the forwarding probability p_{ij}^{φ} can be computed as described in Section IV, and μ_{ij} and μ_i are the meeting rates between node i and node j, and between node i and any other node, respectively.

Theorem 4 (Transition Probability p_{ij}): The transition probability p_{ij} is given by $p_{ij} = \frac{p_{ij}^{\varphi} \mu_{ij}}{\sum_z p_{iz}^{\varphi} \mu_{iz}}$, where p_{ij}^{φ} can be computed as described in Section IV, and μ_{ij} (μ_{iz}) is the inter-meeting rate between node i and node j (z).

Please note that under this simplified model, transition probabilities are not dependent on the specific distribution of inter-(any)-meeting times but only on their expectations. Instead, as highlighted by Theorem 3, T_i^{exit} depends on the distribution of inter-any-meeting times, which in turn characterizes the distribution of residuals R_i. Below we derive the closed form solution of $E[T_i^{exit}]$ for three reference probability distributions commonly used in the literature of opportunistic networks. Please note that the Pareto and Pareto with exponential cut-off case could not have been solved without using inter-any-meeting times.

A. The Exponential Case

In this section we apply Theorems 3 and 4 to the case of exponentially distributed inter-any-meeting times, i.e., $M_i \sim Exp(\lambda_i)$.

Lemma 1: When the inter-any-meeting time M_i follows an exponential distribution with rate λ_i for all i and the rate of pairwise intermeeting times is μ_{ij}, the expectation of T_i^{exit}, i.e., the time before the semi-Markov process exits state i, is given by the following:

$$E[T_i^{exit}] = \frac{1}{\sum_{j \in \mathcal{P}_i} \mu_{ij} p_{ij}^{\varphi}}, \tag{6}$$

where p_{ij}^{φ} can be computed as described in Section IV.

Proof: Equation 6 follows from the application of Theorem 3. $E[M_i]$ corresponds to the expectation of $M_i \sim Exp(\lambda_i)$, which is equal to $\frac{1}{\lambda_i}$. By the memoryless property of the exponential distribution, the residual R_i of exponentially distributed inter-meeting times features an exponential distribution with the same rate. Thus, its expectation $E[R_i]$ is simply $\frac{1}{\lambda_i}$. Then, noting that $\mu_i = \lambda_i$ in the case of an exponential distribution, Equation 6 follows after simple substitutions. ■

It is easy to show that the results presented in this section are exact when the exponential distribution for the inter-any-meeting time is the result of pairwise inter-meeting times being exponentially distributed.

Corollary 1: When inter-meeting times M_{ij} between any generic node pair i, j follows an exponential distribution with rate λ_{ij}, Lemma 1 and Theorem 4 involve no approximations.

Proof: When inter-meeting times are exponential, the meeting process is a Poisson process. Thus, the inter-any-meeting time process can be seen as the superposition of a set of Poisson process, each with inter-meeting time M_{ij}, for all $j \in \mathcal{P}_i$. It is a well known result that the superposition of Poisson processes generates another Poisson process whose inter-arrival times are exponentially distributed with a rate that is the sum of the rates of each individual Poisson process ($\lambda_i = \sum_{j \in \mathcal{P}_i} \lambda_{ij}$) [18]. From the memoryless property of the exponential distribution, it follows that the residual of exponential inter-any-meeting times features an exponential distribution with the same rate $\sum_{j \in \mathcal{P}_i} \lambda_{ij}$. Since we are able to go from pairwise inter-meeting times to inter-any-meeting times without any approximation (thanks to the memoryless property of the exponential distribution), the proposed analytical framework based on inter-any-meeting times is exact. ■

B. The Power Law Case

In this section we assume that the inter-any-meeting time M_i follows a Pareto distribution with shape α_i and scale b_i for all i. The corresponding CCDF is $F_{M_i}(t) = \left(\frac{b_i}{b_i+t}\right)^{\alpha_i}$. The residual inter-any-meeting time in this case is again Pareto distributed, with rate $\alpha_i - 1$ [19]. Under these assumptions the following lemma holds.

Lemma 2: When the inter-any-meeting time M_i follows a Pareto distribution with shape α_i and scale b_i for each i, and the rate of pairwise intermeeting times is μ_{ij}, the expected exit time from state i is given by the following:

$$E[T_i^{exit}] = b_i \left(\frac{1}{2 - 3\alpha_i + \alpha_i^2} + \frac{1}{\sum_{j \in \mathcal{P}_i} p_{ij}^{\varphi} \mu_{ij}} \right), \qquad (7)$$

where p_{ij}^{φ} can be computed as described in Section IV.

Proof: In order to apply Theorem 3, we need compute the expectation of the inter-any-meeting time and of the residual inter-any-meeting time. From standard probability theory, we obtain $E[M_i] = \frac{b_i}{\alpha_i - 1}$ and $E[R_i] = \frac{b_i}{\alpha_i - 2}$. Then, Equation 7 follows after simple substitutions. ■

C. The Power Law with Exponential Cut-Off Case

In this section we assume that the inter-any-meeting time M_i follows a power law distribution with exponential cut-off described by shape α_i, scale b_i and rate λ_i. The corresponding CCDF is $F_{M_i}(t) = \frac{\Gamma(-\alpha_i, \lambda_i t)}{\Gamma(-\alpha_i, \lambda_i b_i)}$.

The expectation of the exit time T_i^{exit} is then provided in the following lemma.

Lemma 3: The expected exit time from state i is given by the following:

$$\begin{aligned} E[T_i^{exit}] &= \frac{1}{2}\left(\frac{1-\alpha_i}{\lambda_i} + \frac{c_i^{1-\alpha_i} e^{-c_i}}{\lambda_i \Gamma(1-\alpha_i, c_i)} + \right. \\ &\left. - \frac{2\Gamma(1-\alpha_i, c_i)}{\lambda_i \Gamma(-\alpha_i, c_i)} + \frac{2}{\sum_{j \in \mathcal{P}_i} p_{ij}^{\varphi} \mu_{ij}} \right) \qquad (8) \end{aligned}$$

where c_i is equal to $\lambda_i b_i$.

Proof: As in the power law case, we need to compute the expectation of M_i and R_i. The residual inter-any-meeting time can be derived as described in [19]. From standard probability theory, we obtain the following:

$$E[M_i] = \frac{\Gamma(1-\alpha_i, \lambda_i b_i)}{\lambda_i \Gamma(-\alpha_i, \lambda_i b_i)}$$

$$E[R_i] = \frac{1 - \alpha_i + \frac{e^{-\lambda_i b_i}(\lambda_i b_i)^{1-\alpha_i}}{\Gamma(1-\alpha_i, \lambda_i b_i)}}{2\lambda_i}.$$

Then Equation 8 follows after simple substition from Theorem 3. ■

VI. Model Evaluation

In this section we (i) study how the effectiveness of forwarding protocols changes when the estimation errors increase, and (ii) evaluate the approximation introduced by the proposed model based on inter-any-meeting times when memoryful probability distributions are considered. The performance of the utility-based forwarding schemes defined in Section III-A are compared against those of the Direct Transmission (DT), Always Forward (AF), and Two Hop (2H) schemes, which are common baseline randomized reference protocols. The AF schemes forces nodes to hand over the message to the first node encountered. Please note that the Always Forward strategy is the single-copy counterpart of Epidemic routing [5]. Epidemic routing is known to be optimal under ideal conditions, because it exploits all possible paths towards the destination. However, AF cannot be considered optimal (and indeed it is far from being optimal in our results), since a single copy cannot exploit all possible paths. For the description of DT and 2H, please refer to Section II.

The scenario we consider comprises 15 nodes, which are divided into three communities, $C1$, $C2$, and $C3$. We assume that nodes belonging to the same community all meet with each other. Nodes that belong to different communities do not meet, unless they are *travellers*. A traveller is a node that is in touch with more than one community. We define two traveller nodes, one that connects communities $C1$ and $C2$, and the other one that connects communities $C1$ and $C3$. The network of users is connected: there is a path connecting any two nodes of the network. However, not all forwarding strategies might be able to find them. As far as node mobility is concerned, in order for the results to be comparable, we require that the expectation of the inter-meeting times for the same node pair i, j is the same despite of the distribution being considered. In the following we use the exponential and Pareto distribution, thus we impose $\frac{1}{\lambda_{ij}} = \frac{\alpha_{ij} b_{ij}}{\alpha_{ij} - 1}$, where λ_{ij} is the rate of the exponential distribution, while α_{ij} and b_{ij} are the exponent and scale of the Pareto distribution. For the sake of example, we set $\lambda_{ij} = 1s^{-1}$ and $\alpha_{ij} = 5.5, b_{ij} = 4.5s$ for nodes that are confined within a single community. Instead, we model the fact that travelers divide their time between different communities by halving their rate of encounter ($\lambda_{ij} = 0.5$ and $\alpha_{ij} = 3.25, b_{ij} = 4.5s$, where at least either i or j is a traveller). For simulations, we run $10000s$ of simulated time, and results are shown with 99% confidence intervals.

For case (i), in which we study how the relative performance of forwarding protocols changes when estimation errors increase, we assume that node pairs meet according to an exponential distribution. This implies that their inter-any-meeting times are also exponentially distributed and that the framework is exact (Corollary 1). This lets us focus only on the effects of imprecise utility estimation. Figure 2 shows how the percentage of node pairs for which each policy is able to provide the lowest expected delay changes when the standard deviation of the Normal distribution used for modeling estimation errors is changed. Here the lowest expected delay is computed as the lowest expected delay among those provided by all the policies considered. This has the following two implications. First, since there might be ties (two or more strategies that all provide the minimum expected delay), the values obtained fixing a given value of standard deviation do not necessarily add up to 100%. Second, since Figure 2 provides a *relative* ranking, the values associated with randomized strategies (DT, AF, 2H) can change even if these strategies are not sensitive to variations in the estimation accuracy. However, such change is connected with utility-based strategies performing worse, not with randomized strategies performing better.

From Figure 2 the following behavior emerges. While with no errors the utility-based policies are able to provide the lowest expected delay for a large fraction of pairs (100% for SF and $\sim 75\%$ for DA), their performance rapidly drops as the chances of estimation errors increase, up to the point when the simple AF and DT overtake them. This is reasonable, as utility-based strategies rely on the accuracy of the predicted utility for making good decisions. Among the utility-based policies, the SF strategy appears to be both more effective when utility information is exact and, at the same time, more resilient to errors.

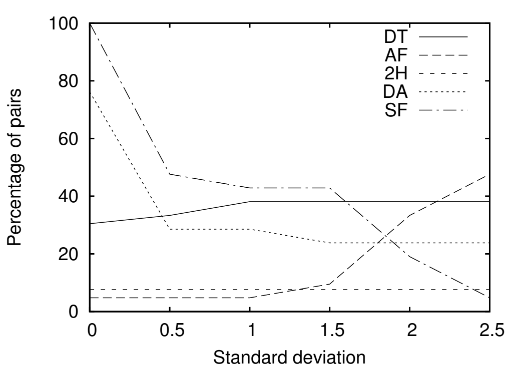
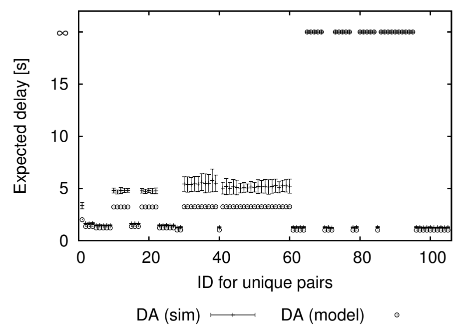
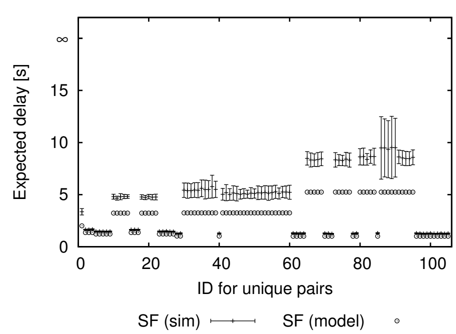


Fig. 2. Percentage of pairs for which each policy provides the lowest expected delay when the standard deviation is varied

For case (ii), in which we want to evaluate the approximation introduced by the proposed model based on inter-any-meeting times, we assume that node pairs meet according to a Pareto distribution. As discussed in [20], the inter-any-meeting times following from Pareto inter-meeting times can be *approximated* with a Pareto distribution. Thus, using this approximation, we derive that inter-any-meeting times to be used in our analytical framework, starting from the pairwise inter-meeting times that we have used for simulations. In Figure 3 we evaluate the case with no estimation error, as this allows us to focus only on the effects of the approximation based on inter-any-meeting times. We focus on the expected delay that the DA and SF policies provide per node pair. In order to identify the different distinct pairs, we assign an identifier to each of them (note that we assume $E[D_{ij}] = E[D_{ji}]$, thus pair i, j is the same as pair j, i). Figure 3(a) compares the expected delay provided by DA as obtained from simulations against analytical predictions, finding them in good agreement. Similar results hold for the SF policy (Figure 3(b)). As expected, the model introduces some approximations. After analyzing the differences in performance between the different node pairs, we found that the biggest difference corresponds to source-destination pairs for which messages travel along longer multi-hop paths. Intuitively, the longer the path, the longer the memory of distributions that our model neglects, from which the discrepancy between analytical and simulation results follows.



(a)



(b)

Fig. 3. Expected delay provided by the utility-based DA and SF forwarding protocols (no estimation error)

VII. Conclusion

In this paper we have proposed an approximated analytical model for the forwarding process in an opportunistic network.

The advantage of this model is twofold. First, it is able to account for imprecise estimations of the utility function on which utility-based forwarding schemes rely. Second, it can be easily used even when inter-meeting times are i.n.i.d. and not exponential. To the best of our knowledge, this is the first model that accommodates both these two aspects.

As a case study, we have used the proposed analytical framework in order to evaluate the relative performance of a general class of forwarding protocols when the accuracy of the utility information is varied. Our results show that the performance of utility-based forwarding schemes may degrade significantly when the utility information is not accurate and that, in this case, simple randomized strategies may even prove more effective.

ACKNOWLEDGMENT

This work was partially funded by the European Commission under the SCAMPI (FP7-FIRE 258414), RECOGNITION (FET-AWARENESS 257756), and EINS (FP7-FIRE 288021) projects.

REFERENCES

[1] M. Conti and M. Kumar, "Opportunities in opportunistic computing," *IEEE Computer*, vol. 43, no. 1, pp. 42–50, 2010.

[2] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Trans. Netw.*, vol. 10, no. 4, pp. 477–486, 2002.

[3] A. Lindgren, A. Doria, and O. Schelén, "Probabilistic routing in intermittently connected networks," *LNCS*, pp. 239–254, 2004.

[4] C. Boldrini, M. Conti, and A. Passarella, "Exploiting users' social relations to forward data in opportunistic networks: The HiBOp solution," *Pervasive and Mobile Computing*, vol. 4, no. 5, pp. 633–657, 2008.

[5] A. Vahdat and D. Becker, "Epidemic routing for partially connected ad hoc networks," Citeseer, Tech. Rep., 2000.

[6] Z. Haas, J. Halpern, and L. Li, "Gossip-based ad hoc routing," *IEEE/ACM Trans. Netw.*, vol. 14, no. 3, pp. 479–491, 2006.

[7] T. Spyropoulos, K. Psounis, and C. Raghavendra, "Efficient routing in intermittently connected mobile networks: The single copy case," *IEEE/ACM Trans. Netw.*, vol. 16, no. 1, pp. 63–76, 2008.

[8] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: Social-based forwarding in delay tolerant networks," *IEEE Trans. Mobile Comp.*, p. 14, 2010.

[9] E. Daly and M. Haahr, "Social network analysis for information flow in disconnected Delay-Tolerant MANETs," *IEEE Trans. Mobile Comp.*, pp. 606–621, 2008.

[10] P. Costa, C. Mascolo, M. Musolesi, and G. Picco, "Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 5, pp. 748–760, 2008.

[11] Z. Haas and T. Small, "A new networking model for biological applications of ad hoc sensor networks," *IEEE/ACM Trans. Netw.*, vol. 14, no. 1, pp. 27–40, 2006.

[12] R. Groenevelt, P. Nain, and G. Koole, "The message delay in mobile ad hoc networks," *Per. Eval.*, vol. 62, no. 1-4, pp. 210–228, 2005.

[13] T. Spyropoulos, T. Turletti, and K. Obraczka, "Routing in Delay-Tolerant Networks Comprising Heterogeneous Node Populations," *IEEE Trans. Mobile Comp.*, pp. 1132–1147, 2009.

[14] C. Lee and D. Eun, "Exploiting Heterogeneity in Mobile Opportunistic Networks: An Analytic Approach," in *IEEE SECON'10*, 2010, pp. 1–9.

[15] C. Boldrini, M. Conti, and A. Passarella, "Modelling social-aware forwarding in opportunistic networks," in *Proc. of PERFORM'10*, 2010.

[16] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Trans. Mobile Comp.*, pp. 606–620, 2007.

[17] C. Boldrini, M. Conti, and A. Passarella, "Less is more: Long paths do not help the convergence of social-oblivious forwarding in opportunistic networks," in *ACM/SIGMOBILE MobiOpp'12*, 2012, pp. 1–8.

[18] D. Cox, *Renewal theory*. Methuen London, 1962.

[19] C. Boldrini, M. Conti, and A. Passarella, "From pareto inter-contact times to residuals," *IEEE Commun. Lett.*, no. 99, pp. 1–3, 2011.

[20] W. Whitt, "Approximating a point process by a renewal process, i: Two basic methods," *Operations Research*, pp. 125–147, 1982.

APPENDIX

Proof for Theorem 3: Let us assume that the Markov chain is currently in state i, or, equivalently, that the message is currently on node i. At each new encounter, node i will hand over the message with probability p_i^φ, which depends on the forwarding strategy φ in use. The fact that the message generation process is asynchronous with respect to the encounter process implies that each message has to initially wait at least for a residual inter-any-meeting time R_i^{any} before being handed over to another node. Then, upon contact, the message either leaves node i with probability p_i^φ or stays with i with probability $1 - p_i^\varphi$. If the message is not handed over at the first contact, it has to wait for the next contact or, equivalently, it has to wait for M_i before the next transmission opportunity. This process is repeated until the message is relayed. Following the above line of reasoning, we can write the overall time T_i^{exit} a message stays with node i before being handed over as follows:

$$P\left(T_i^{exit}=t\right)=p_i^\varphi P(R_i=t)+\sum_{n=2}^{\infty}\left[p_i^\varphi\left(1-p_i^\varphi\right)^{n-1}\cdot P\left(R_i+\sum_{m=1}^{n-1}M_i=t\right)\right]$$

Exploiting the linearity of the expectation, $E[T_i^{exit}]$ can be obtained as follows:

$$E[T_i^{exit}]=p_i^\varphi E[R_i]+\sum_{n=2}^{\infty}\left[p_i^\varphi\left(1-p_i^\varphi\right)^{n-1}\cdot(E[R_i]+(n-1)E[M_i])\right]$$

The series in the above equation is convergent. Then, by simple manipulation, we obtain Equation 5. ■

Proof for Theorem 4: First we consider the probability p_i^φ that node i hands over the message to any of the next encounters. p_i^φ can be computed by conditioning on the probability of meeting a specific node j and the probability that node i hands over a message to node j. Thus we have $p_i^\varphi = \sum_{j\in\mathcal{P}_i} p_{ij}^\varphi * p_{ij}^e$, where p_{ij}^φ denotes the probability that node i hands over a message to node j when they meet and p_{ij}^e gives the probability of such an event. Both the pairwise contact process and the inter-any contact process can be seen as renewal processes. According to renewal theory, on the long run $\frac{N(t)}{t} \to \lambda$, where $N(t)$ denotes the number of renewal intervals (here, contacts) and λ indicates the rate of the renewal process. If we apply this result to both the pairwise contact process and the inter-any contact process and we consider their ratio, we obtain $p_{ij}^e = \frac{\mu_{ij}}{\mu_i}$, where μ_{ij} denotes the rate of the process of encounter between node i and node j and μ_i is the rate of the inter-any contact process. The pairwise forwarding probability p_{ij}^φ can be computed as in Section IV in the case of estimation errors for μ_{ij}. Thus, we obtain $p_i^\varphi = \sum_{j\in\mathcal{P}_i} p_{ij}^\varphi * \frac{\mu_{ij}}{\mu_i}$.

Finally, in order to compute the transition probability p_{ij}, we are interested in the probability of a forwarding event involving exactly node j, since this gives the probability of the Markov chain moving from state i to state j. Again, this follows from the long run proportion between two regenerative processes, one describing jumps from state i to state j and one describing jumps from state i to any other state. Thus, we obtain that the transition probability p_{ij} is approximately equal to $\frac{p_{ij}^\varphi \mu_{ij}}{\sum_z p_{iz}^\varphi \mu_{iz}}$. ■