A Framework for Routing Performance Analysis in Delay Tolerant Networks with Application to Non Cooperative Networks

Giovanni Resta, and Paolo Santi, *Member, IEEE,*

Abstract—In this paper, we present a framework for analyzing routing performance in delay tolerant networks. Differently from previous work, our framework is aimed at characterizing the exact *distribution* of relevant performance metrics, which is a substantial improvement over existing studies characterizing either the expected value of the metric, or an asymptotic approximation of the actual distribution. In particular, the considered performance metrics are *packet delivery delay*, and *communication cost*, expressed as number of copies of a packet circulating in the network at the time of delivery. Our proposed framework is based on a characterization of the routing process as a stochastic coloring process, and can be applied to model performance of most stateless delay tolerant routing protocols such as epidemic, two-hops, spray and wait, etc. After introducing the framework, we present examples of its application to derive the packet delivery delay and communication cost distribution of two such protocols, namely epidemic and two-hops routing.
Characterizing packet delivery delay and communication cost distribution is important to investigate fundamental properties of delay tolerant networks. As an example, we show how packet delivery delay distribution can be used to estimate how epidemic routing performance changes in presence of different degrees of node cooperation within the network. More specifically, we consider fully cooperative, non-cooperative, and probabilistic cooperative scenarios, and derive nearly exact expressions of the packet delivery rate under these scenarios based on our proposed framework. The comparison of the obtained packet delivery rate estimation in the various cooperation scenarios suggests that even a modest level of node cooperation (probabilistic cooperation with a low probability of cooperation) is sufficient to achieve 2–fold performance improvement with respect to the most pessimistic scenario in which all potential forwarders drop packets.

Index Terms—Delay-tolerant networks; non-cooperative networks; delay-tolerant routing; packet delivery delay distribution; communication cost distribution.

I. INTRODUCTION

The delay tolerant network (DTN) paradigm [10] has attracted increased attention in the research and industrial community in recent years. Differently from other types of wireless multi-hop networks, DTNs are characterized by a very sparse node population, and by the lack of full network connectivity at virtually every time. Given these features, eventual packet delivery to the destination can be achieved only through node

mobility, which is indeed the main communication means in the network.

Since node speed is several orders of magnitude lower than speed of radio signal propagation in the air, delays expected in DTNs are orders magnitude higher than those typically experienced in fully connected wireless multihop networks. This means that applications designed for DTNs, as the name DTN suggests, must be able to tolerate large delays in information dissemination and end-to-end packet delivery.

From the above description, it is clear that a prominent parameter for evaluating network-level protocol performance in DTNs is packet delivery delay. In particular, an accurate characterization of the expected packet delivery delay, together with an estimation of the communication cost (number of copies of a packet circulating in the network), is fundamental to optimally tune the delivery delay/communication overhead tradeoff, which is at the very heart of delay-tolerant networking.

The research community has devoted substantial efforts to accurately estimating expected performance metrics in DTNs (see, e.g., [2], [3], [9], [15], [16], [19]). In some cases, though, knowing the expected value of a performance metric is not sufficient to fully understand DTN behavior in different conditions, and a characterization of the *distribution* of the metric of interest is needed. Consider for instance a situation in which the network designer is interested in comparing the expected packet delivery rate (PDR) of different DTN routing protocols, where a packet is considered as successfully delivered if and only if it is received by the destination within a certain time TTL (Time To Live) since its generation at the source. Let D be the random variable (r.v.) corresponding to the time elapsing between the instant at which the packet is generated at the source, and the instant at which a copy of the packet is first received at the destination. Knowing the expected packet delivery delay $E[D]$ gives the designer only a very inaccurate estimation of the PDR, since only 0/1 information can be deduced by the knowledge of the sole expected packet delivery delay, such as assuming PDR = 1 if $E[D] \leq \epsilon \cdot TLL$, for some $0 < \epsilon \leq 1$, and PDR = 0 otherwise. Much more accurate PDR estimation could be given if the distribution of random variable D is known, since in this case the designer could estimate the expected PDR as $P(D \leq TTL)$, i.e., using the cumulative density function of r.v. D.

In this paper, we introduce a theoretical framework for analyzing routing performance in delay tolerant networks

which is aimed at characterizing the *exact distribution* of the fundamental performance metrics described above, namely *packet delivery delay* and *communication cost*. This is in sharp contrast with existing work, which either considers only expected value of the metrics of interest ([2], [3], [9], [15], [16]), or provides an *asymptotic* approximation of their probability distribution [19].

Our proposed framework is based on a characterization of the routing process as a stochastic coloring process, and can be applied to model performance of most stateless delay tolerant routing protocols such as epidemic [18], two-hops [8], and spray and wait [15]. After introducing the framework, we present two examples of its application to compute (or accurately approximate) the packet delivery delay and communication cost distribution of epidemic and two-hops routing. The accuracy of our analytical estimation is validated through simulation of random-waypoint mobile networks, whose results show that:
– in case of epidemic routing, the presented analytical estimation is nearly exact, and much more accurate than the asymptotic approximation provided in [19] for networks of moderate size;
– in case of two-hops routing, our analytical approximation of the packet delivery delay and communication cost distribution is very accurate as long as the maximum number of packet copies circulating in the network is relatively small as compared to the network size.

We then show how the proposed framework can be used to faithfully compare performance of the two considered routing protocols, in terms of both expected PDR and average communication cost.

As a further example of application of our framework, we show how packet delivery delay distribution can be used to estimate how the performance of epidemic routing changes in presence of different degrees of node cooperation within the network. We stress that node cooperation is fundamental to ensure acceptable performance in DTNs: in fact, differently from more traditional (fully connected) types of wireless multi-hop networks, nodes are typically requested not only to act as packet forwarders, but also to store in their own buffer other nodes' packets for a very long time interval (store-and-forward communication). Thus, both energy and memory resources, which are very limited in a typical mobile node, have to be sacrificed for the other nodes' good.

In the last part of this paper, we consider different cooperation scenarios, encompassing no cooperation, full cooperation, and probabilistic cooperation, and derive nearly exact expressions of the expected PDR under these scenarios based on our proposed framework. The comparison of the obtained PDR estimations gives useful insights on the behavior of epidemic routing in presence of reduced cooperation levels within the network, the most notable of which is that even a modest level of node cooperation (probabilistic cooperation with a low cooperation probability) is sufficient to achieve up to 2-fold performance improvement over the most pessimistic scenario in which all forwarders drop packets. To the best of our knowledge, ours are the first analytical results characterizing epidemic routing performance in presence of reduced levels of cooperation within the network.

Summarizing, the novel contributions of this paper are:
i) the first theoretical framework for characterizing the *exact* packet delivery delay and communication cost *distribution* of a wide class of DTN routing protocols;
ii) the exact characterization of the packet delivery delay and communication cost distribution with epidemic routing;
iii) a very accurate approximation of the packet delivery delay and communication cost distribution with two-hops routing;
iv) the first theoretical characterization of epidemic routing performance in presence of different degrees of node cooperation within the network.

The derived bounds are shown to be significantly more accurate than the asymptotic bounds reported in [19].

The rest of this paper is organized as follows. In Section II, we survey related work. In Section III, we present our theoretical framework for estimating routing protocol performance in DTNs. In Section IV, we introduce some preliminary definitions and results which will be used in the remainder of the paper. In Section V, we present the first example of application of our theoretical framework, which is used to derive the exact packet delivery delay and communication distribution of epidemic routing. In Section VI, we present the second example of application of our theoretical framework, presenting an accurate approximation of the packet delivery delay and communication cost distribution of two-hops routing. In Section VII, we compare and assess through simulation expected performance of epidemic and two-hops routing in terms of PDR and communication cost, exploiting the theoretical results derived in sections V and VI. In Section VIII, we present an example of how the packet delivery delay distribution derived in Section V can be used to study the effects of different degrees of node cooperation on epidemic routing performance. Finally, Section IX concludes.

II. RELATED WORK

Several papers have presented analytical investigations of DTN routing protocol performance. However, most of these papers [3], [9], [15], [16] considered only derivation of the expected value of the metric of interest (typically, packet delivery delay) in their analysis. To the best of our knowledge, only a few papers [2], [19] attempted at characterizing a performance metric's distribution. However, the authors of [2] consider a network model in which the TTL is associated with each single packet replica, and thus their results cannot be used to provide bounds on the delivery time since initial packet generation. On the other hand, the results of [19] are based on ordinary differential equations and are asymptotic in nature, i.e., they become increasingly more accurate as the number of nodes in the network grows larger. As we will show later in this paper, the approach of [19], when applied to networks of reasonable size, is quite inaccurate. Most importantly, the bounds presented in [19], when applied to packet delivery delay, considerably *underestimate* the value of the metric, hence they cannot be used to provide minimal performance guarantees in DTNs.

Relevant to the last part of this paper are recent works investigating DTN performance in presence of reduced levels

of node cooperation within the network. In [6], the authors assume selfish node behavior, and present a mechanism to discourage selfish node behavior during packet exchange based on the principles of barter. In [13], the authors consider three well-known routing protocols for DTNs, and evaluate their performance (in terms of expected delivery delay and communication overhead) through simulation under different levels of node cooperation. The main finding of their study is that two-hops routing in general appears more resilient to less cooperative node behavior amongst the considered routing protocols. The above works, though, are based only on simulations, and do not provide any theoretical insight on routing protocol performance in presence of different degrees of node cooperation within the network.

III. THE ROUTING ANALYSIS FRAMEWORK

Consider a delay-tolerant network composed of M nodes, and denote by $\mathcal{M}$ the set of nodes. Among the M nodes, a node s is chosen uniformly at random as the packet source, and another node $d \neq s$ is chosen uniformly at random as the packet destination. The goal of this section is to present a framework for analyzing the performance of DTN routing protocols used to route the packet from s to d. More specifically, we are interested in characterizing the *probability distribution* of the *packet delivery delay* and *communication cost*, which are defined as follows. The packet delivery delay is defined as the time elapsing between the instant at which a packet is generated at s, and the instant at which a copy of the packet is first received by d. The communication cost is defined as the number of nodes holding at least one copy of the packet present in the network at the instant at which a copy of the packet is first received by d.

The proposed framework is based upon the following assumptions:

1) *low load*: network traffic is low, so that buffer capacity on the nodes is not an issue (i.e., it can be considered as virtually infinite).

2) *transmission range and encounter*: two nodes can communicate iff they are within distance r, where r is the transmission range; this is equivalent to assuming isotropic, deterministic propagation of the radio signal with distance. We say that an *encounter* between two nodes occur when they come into each other transmission range.

3) *no contention*: any communicating pair of nodes does not interfere with any other pair communicating at the same time. This assumption is justified by the very low node density in a typical DTN scenario, and by the relatively low network load scenario considered in this paper.

4) *fast transmissions*: relative speed between arbitrary node pairs is very low compared to transmission time; in other words, we assume that the duration of node encounters is always sufficient for the two nodes to exchange the content of their buffers.

The process of packet dissemination within the network, by means of which a packet is eventually delivered to destination, is governed by a routing protocol. The routing protocol defines the rules according to which new copies of the packet are
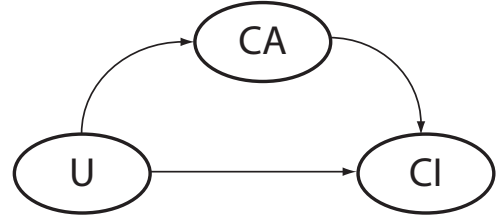

Fig. 1. Transition diagram of the coloring process.

generated, and buffered packets are exchanged between nodes upon encounters. The presented framework can be applied to any *stateless*, *deterministic* routing protocol for DTNs. By *stateless*, we mean that decisions about whether a copy of the packet should be delivered to an encountered node do not depend on a notion of state (such as, e.g., history of past encounters, information related to social relationships between nodes, etc.). By *deterministic*, we mean that routing decisions are not influenced by random choices.

The packet dissemination process is modeled as a *coloring process*. Nodes in $\mathcal{M} - \{s, d\}$ can be in one of the following states:

– *uncolored* (U): nodes in this state have not yet received a copy of the packet;

– *colored active* (CA): nodes in this state have at least two copies of the packet; these nodes are allowed to deliver one or more copies of the packet to *any* uncolored node;

– *colored inactive* (CI): nodes in this state have one copy of the packet; these nodes are allowed to delivery the only copy of the packet they have only to the final destination.

The source node can only be in either CA or CI state, while the destination can only be in either U or CI state.

Possible transitions between states for non-source nodes are depicted in Figure 1: a node starts from the U state, and eventually transitions to either the CA or CI state depending on the specific routing protocol. A node in CA state may eventually transition to CI state, while the opposite transition is not possible. In the following, we say that a node is *colored* if it is in either CA or CI state, while it is *uncolored* otherwise. It is evident from the transition diagram reported in Figure 1 that the number of colored nodes within the network cannot decrease over time.

Let L be the maximum number of copies of a packet (excluding possible copy delivered to the destination) allowed to circulate in the network by the routing protocol at hand. The coloring process starts at time $t_0 = 0$, when the source node is in state CA, and all other nodes are in state U (see Figure 2). The relevant events in the considered stochastic process are "coloring events", representing the situation in which a node in CA state encounters a node in U state, or a node in CA/CI state encounters the destination. The i-th coloring event is denoted E_i, and occurs at time T_i. Upon occurrence of E_i (i.e., at time $t > T_i$), the number of colored nodes in the network becomes $i + 1$. Thus, a simple relationship is established between the number of coloring events, and the number of nodes holding at least one copy of the packet in the network. As we shall see, this relationship can be used to derive the communication cost distribution by simply estimating the probability that the destination is the i-th colored node, with $1 \leq i \leq L$.
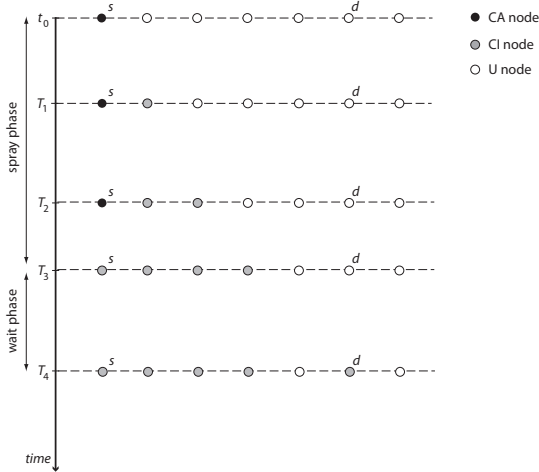
Fig. 2. The stochastic coloring process. In the example, $L = 4$. Initially, only the source node is colored (and active). At time T_1, the first node is colored and transitions to state CI. At time T_3, four copies of the packet are circulating in the network: the spray phase ends, and the wait phase starts. At time T_4, the destination node is colored and the coloring process ends.

The coloring process ends when the destination node becomes colored. It is easy to see that the number of colored nodes (including the source) at any point in time *before* the destination is colored is at most L. Thus, the maximum possible number of coloring events occurring in the process is L, accounting for up to $L-1$ coloring events between a node in CA and a node in U state, and coloring of the destination.

The coloring process can be thought as divided into two phases (see Figure 2), called the *spray* and the *wait* phase [15]: in the spray phase, some of the colored nodes are in CA state, and they can deliver extra copies of the packet to any uncolored node; in the wait phase, all colored nodes are in CI state, and the only possible coloring event occurs when the destination first encounters a colored node and becomes colored. Given the upper bound L on the number of packet copies circulating in the network, it is easy to see that, unless the destination is colored earlier, the spray phase ends at time T_{L-1}, i.e., when the $(L-1)$-th node is colored.

Let E_d denote the event "destination is colored", which occurs at time T_d. Our goal in the following is to characterize the probability distribution of random variable T_d representing the time at which the destination is colored. More specifically, we want to compute

$$P(T_d < t)$$

for any given $t > 0$.

We first observe that, given the above description of the coloring process, we have $T_d = T_i$ for some $i = 1, \ldots, L$, with $T_d = T_i$ indicating that the destination is the i-th colored node in the process. Denote by $E_{d,i}$ the event "destination is the i-th colored node". It is immediate to see that the $E_{d,i}$s are mutually disjoint events, and the probability of interest can be computed as follows:

$$P(T_d < t) = \sum_{i=1}^{L} P(T_i < t | E_{d,i}) \cdot P(E_{d,i}) .$$

We have then reduced the problem of computing the distribution of T_d to the two following subproblems: 1) com-

puting the distribution of r.v. T_i, conditioned on $E_{d,i}$; and 2) computing the probability of event $E_{d,i}$. Note that computing probabilities $P(E_{d,i})$ is exactly what is needed to derive the communication cost distribution, which is then a byproduct of 2).

In order to compute the distribution of T_i, we observe that T_i can be expressed as a sum of random variables $\tau_1, \ldots, \tau_i$, where r.v. $\tau_1 = T_1 - t_0$, and $\tau_j = T_j - T_{j-1}$ for $1 < j \leq i$. Thus, we can write:

$$P(T_i < t | E_{d,i}) = P(\sum_{j=1}^{i} \tau_j < t | E_{d,i}) .$$

Clearly, the distribution of τ_j random variables depends on the features of the mobility pattern, and more specifically on the *expected meeting time* between nodes. The expected meeting time of a mobility model is defined as the time elapsed between a random time instant (after node spatial distribution has reached the stationary state) and the first encounter between a pair of nodes moving according to the model [17]. In the following, to simplify derivations, we make the assumption that nodes move according to an arbitrary mobility model with *exponentially distributed* meeting time with rate $\frac{1}{emt}$ between arbitrary node pairs, where emt is the expected meeting time between arbitrary node pairs. This is a quite standard assumption in the analysis of DTN performance [15], [16], [19], which has been formally proved to hold for some mobility models (e.g., random walks [1]), and validated through simulation-based analysis for common mobility models such as random waypoint, random direction, and so on [17]. Indeed, recent analyses of real-world DTN traces [12] have shown that meeting time follows a mixed power law and exponential (for the tail) distribution. Extending our analysis to mixed power law/exponential meeting time distributions is beyond the scope of this paper, and is left for future work.

IV. PRELIMINARIES

In the rest of this paper, we will present examples showing how the analytical framework described in the previous section can be used to derive or to accurately approximate the packet delivery delay and communication cost distribution of popular DTN routing protocols. Before presenting the technical derivations, we need some preliminary definitions and basic results.

The following notation will be used throughout the paper. We use notation $f(\lambda, x)$ and $F(\lambda, x)$ to denote the pdf and cdf of an exponential random variable of parameter λ, namely $f(\lambda, x) = \lambda e^{-\lambda x}$ and $F(\lambda, x) = 1 - e^{-\lambda x}$.

Consider n i.i.d. continuous random variables $X_1, \ldots, X_n$, and let $\overline{X}_1, \ldots, \overline{X}_n$ be a realization of the n random variables. We now order the values of the realization in increasing order, starting from the smallest, and denote with $\overline{X}_{(1)}, \ldots, \overline{X}_{(n)}$ the ordered values. Each of the $\overline{X}_{(i)}$, for $i = 1, \ldots, n$, can be considered as a realization of a random variable $X_{(i)}$, which is known as the *i-th order statistic* of random variables $X_1, \ldots, X_n$ (note that $X_{(1)} = \min\{X_1, \ldots, X_n\}$ and $X_{(n)} = \max\{X_1, \ldots, X_n\}$). Denoting by $\psi(x)$ and $\Psi(x)$

the pdf and cdf, respectively, of each of the X_i, the pdf of the i-th order statistic of random variables $X_1, \ldots, X_n$ is [7]:

$$Ord(n, i, x) = \frac{n!}{(i-1)!(n-i)!} \Psi(x)^{i-1} (1 - \Psi(x))^{n-i} \psi(x) \ .$$

In the following, we denote by $Ord(n, i, \lambda, x)$ the pdf of the i-th order statistic of a set of n i.i.d. exponential random variables of parameter λ.

Fact 1: The first order statistic (minimum) of a set of n i.i.d. exponential random variables of parameter λ is an exponentially distributed random variable of parameter $n\lambda$. Formally,

$$Ord(n, 1, \lambda, x) = n\lambda e^{-n\lambda x} \ .$$

The following lemma from [4] will be useful in the following:

Lemma 1: Let $(X_i)_{i=1\ldots n}$, $n \geq 2$, be independent exponential random variables with pairwise distinct respective parameters λ_i. Then the pdf of their sum is

$$f_{X_1+X_2+\cdots+X_n}(x) = \left[\prod_{i=1}^{n} \lambda_i \right] \sum_{j=1}^{n} \frac{e^{-\lambda_j x}}{\prod_{\substack{k \neq j \\ k=1}}^{n} (\lambda_k - \lambda_j)}, \qquad x > 0.$$

The specific DTN routing protocols considered in the applications of the proposed analytical framework are the following:

i) Epidemic [18]: the maximum number of allowed packet copies circulating in the network is set to $L = M$. Initially, only the source is in CA state. Each time a CA node encounters an uncolored node, it generates *two* new copies of the packet, and deliver them to the uncolored node. This mechanism ensures that all colored nodes (except the destination) are in CA state, thus maximizing the speed of packet spreading within the network. Under our working assumption of unlimited buffer capacity, epidemic routing is optimal in terms of packet delivery delay, at the expense of a significant communication cost.

ii) Two-hops [8]: the maximum number of allowed packet copies circulating in the network is set to $L \leq M$. The source node initially generates L copies of the packet, and delivers a single copy of the packet to the first $L - 1$ encountered uncolored nodes. This mechanism ensures that only the source node can be in CA state, while all other colored nodes are in CI state and can deliver the packet only to the destination. This routing algorithm is equivalent to Spray and Wait routing with source spraying [15].

V. ANALYSIS OF EPIDEMIC ROUTING

In this section, we provide a first example of usage of our framework to derive the probability distribution of the packet delivery delay and communication cost with epidemic routing.

Theorem 1: The cumulative density function of random variable T_d representing delivery time with epidemic routing is defined as

$$P(T_d < t) = \frac{1}{M-1} \sum_{i=1}^{M-1} P(Col_{i,t}) \ ,$$
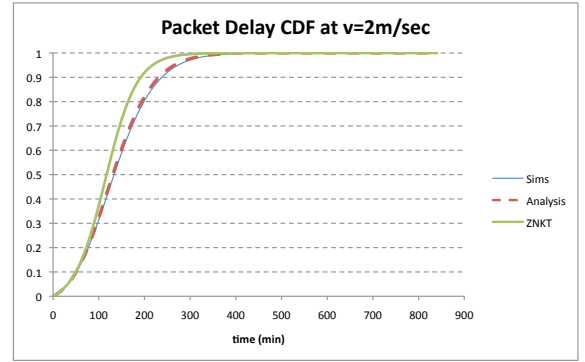

Fig. 3. Experimental and analytical packet delivery delay cdf with $v = 2m/sec$.
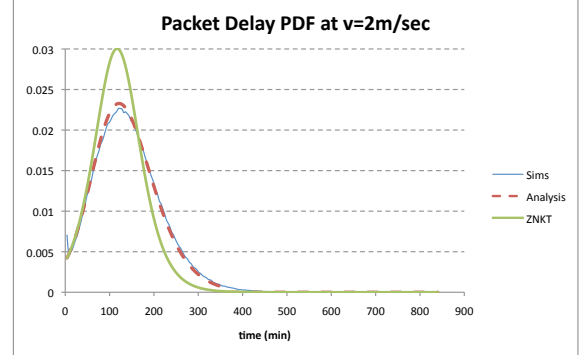

Fig. 4. Experimental and analytical packet delivery delay pdf with $v = 2m/sec$.

where $P(Col_{1,t})$ is defined in equation (3) of the proof, the $P(Col_{i,t})$s are defined in equation (4) of the proof, and the pdf of r.v. T_{i-1} needed to compute the $P(Col_{i,t})$s is obtained according to Lemma 1, where the rates of the involved exponential r.v.s are defined as in formula (5) in the proof.

Proof: The proofs of all results presented in this paper are reported in the supplemental material. ∎

Corollary 1: Let C be the random variable representing the number of nodes holding at least one copy of the packet in the network at delivery time. With epidemic routing, we have

$$P(C = i) = \begin{cases} \frac{1}{M-1} & \text{for } 1 \leq i \leq M-1 \\ 0 & \text{otherwise} \end{cases} \ .$$

To validate our analysis, we have performed a set of simulations using a self-developed simulator, in which $M = 30$ nodes are initially distributed uniformly at random in a square area of $10km$ side. Nodes have a transmission range of $250m$, and move according to the random waypoint (RWP) mobility model [11] with no pause time and fixed speed v. Since the stationary node spatial distribution generated by the RWP model is not uniform [5], we let the nodes initially move for a large number of steps (1000) in order to reach the stationary distribution. After stabilization, a randomly selected node generates a packet directed towards a randomly selected destination, and the corresponding packet delivery delay is recorded when the packet first reaches the destination. Similarly, we record the number of packet copies in the network at delivery time. We performed a large set of such experiments (more than 10^5 for each parameter setting), and grouped stored packet delivery delays in bins of $4min$ duration. The experimental cdf and pdf of packet delivery
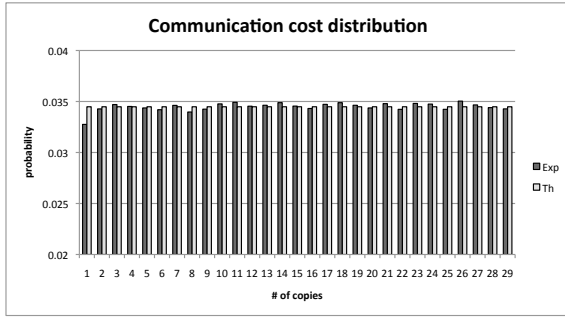
Fig. 5. Experimental and analytical communication cost distribution with $v = 2m/sec$ and Epidemic routing.

delay is then derived by computing the number of elements in each bin. Note that, since the communication cost distribution is discrete, no binning is needed and we simply keep trace of the number of packet copies recorded in each experiment to estimate the probability distribution.

We have considered three settings of parameter v, ranging from pedestrian ($v = 2m/sec$) to vehicular ($v = 15m/sec$) mobility. Figures 3 and 4 reports the derived packet delivery delay experimental distributions when $v = 2m/sec$ (very similar results have been obtained also with other settings of v). The figures also report the analytical packet delivery delay distribution computed (using Mathematica^{TM}) according to the analysis reported in the supplemental material, where ϵ is set to 0.1 in the definition of the λ_j terms. The values of the emt for the random waypoint mobility model has been computed according to the formula reported in [17], which we report for completeness:

$$emt = \frac{N}{3.5 \cdot v \cdot r} \ ,$$

where N is the area of the mobility region, r is the node transmission range, and v is the node speed.

For the sake of comparison, we report also the asymptotical packet delivery delay distribution derived in [19] (ZNKT curve). As seen from the figures, differently from the derivations of [19], the analysis reported in this paper can be used to faithfully reproduce the packet delivery delay distribution.

The distribution of communication cost obtained with experiments when $v = 2m/sec$ is reported in Figure 5. As seen from the histograms, our theoretical characterization matches experimental results very well.

VI. ANALYSIS OF TWO-HOPS ROUTING

In this section, we show how the framework presented in Section III can be used to approximate the packet delivery delay distribution of two-hops routing. The derivation in case of two-hops routing is much more involved than in case of epidemic routing, since the process of spreading the packet in the network is asymmetric. More specifically, during the spraying phase the coloring process is asymmetric, since only the source is in CA state, while the other colored nodes are in CI state and can color only the destination. In the wait phase, the coloring process becomes symmetric, since all colored nodes (including the source, which now holds only the last copy of the packet) are in CI state. As we shall see, the difficulty in deriving the exact packet delivery delay and
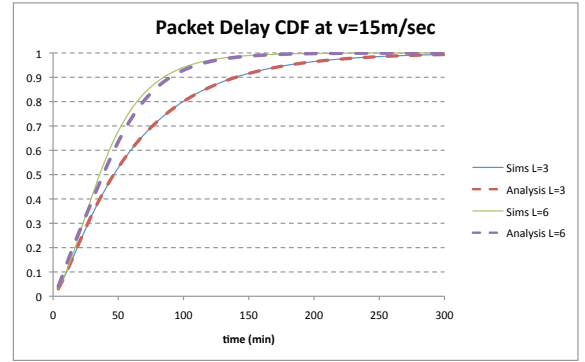

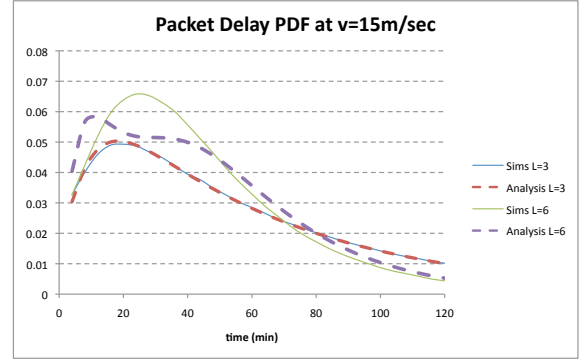Fig. 6. Experimental and analytical packet delivery delay cdf with $v = 15m/sec$ and different values of L.

Fig. 7. Experimental and analytical packet delivery delay pdf with $v = 15m/sec$ and different values of L.

communication cost distribution with two-hops routing lies in the asymmetry of coloring during the spraying phase.

The derivation of the (approximate) packet delivery delay and communication cost distribution with two-hops routing are reported in the supplemental material.

In order to estimate the impact of the several approximations made in our analysis on the accuracy of the analytical estimation of the packet delivery delay and communication cost cdf, we have done extensive simulations, using the same setting as in the experiments with epidemic routing. Figures 6 and 7 report the results of the simulation experiments with velocity $v = 15m/sec$, along with our analytical estimation, for two values of L ($L = 3$ and $L = 6$). The results reported in the figure confirms that our analysis is indeed very accurate as long as the $\frac{L}{M}$ ratio is low (≤ 0.1), i.e., for most cases of practical relevance. As the $\frac{L}{M}$ ratio increases, the approximation reported in equation (12) – see supplemental material – becomes looser, yielding to a less accurate estimation of the packet delivery delay distribution. This effect is evident in Figure 7, where the analytical curve for $L = 6$ has a quite different shape from the experimental distribution. Similar results have been obtained with lower values of velocity.

The experimental communication cost distribution is reported in Figure 8 for the case of $L = 6$, which is the most critical for packet delivery delay distribution. As seen from the histogram, the approximations made in computing the communication cost distribution – mainly due to assumed independence of the coloring events at intermediate hops (see supplemental material) – result in some minor difference between the predicted communication cost distribution and the one derived from experiments.
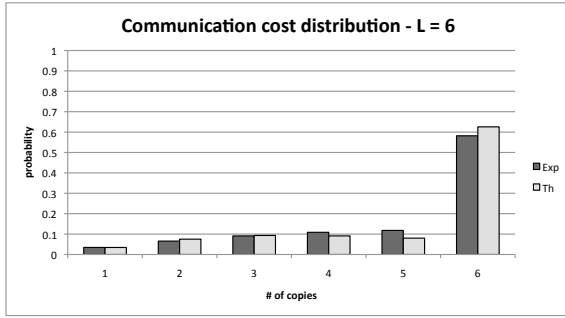
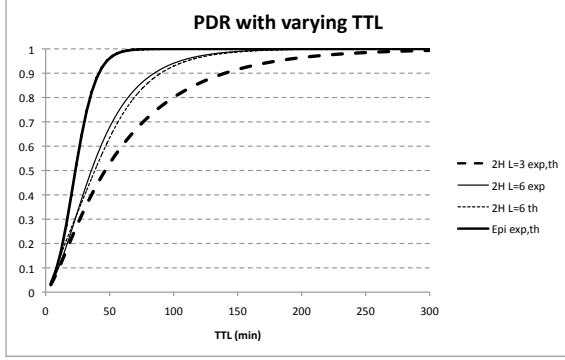Fig. 8. Experimental and analytical communication cost distribution with $v = 2m/sec$ and two-hops routing.

Fig. 9. Experimental and analytical estimation of packet delivery rate with epidemic and two-hops routing and varying TTL values. Node speed is set to $15m/sec$.

protocol	exp	thr
epidemic	15.04	15
two-hops $L = 3$	2.86	2.85
two-hops $L = 4$	3.67	3.65
two-hops $L = 6$	4.96	4.98
binary SW $L = 4$	3.73	–

TABLE I
EXPECTED COMMUNICATION COST OF THE VARIOUS ROUTING
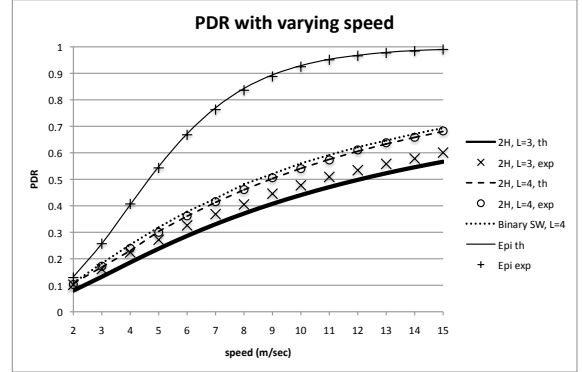PROTOCOLS WITH $TTL = 60min$ AND $v = 15m/sec$.

Fig. 10. Experimental and analytical estimation of packet delivery rate with epidemic and two-hops routing and varying node speed. The TTL is set to $3600sec$.

VII. COMPARISON OF EPIDEMIC AND TWO-HOPS ROUTING

In this section, we show how our proposed theoretical framework can be used to faithfully predict and compare performance of epidemic and two-hops routing. More specifically, we will compare performance of the two protocols with respect to expected packet delivery rate and expected communication cost.

We first notice that, once a notion of TimeToLive (TTL) of a packet is defined, the expected packet delivery rate of a certain routing protocol can be easily derived through our framework by computing the probability that random variable T_d is at most TTL, i.e., by using the cdf of random variable T_d as computed in sections V and VI. Similarly, the expected communication cost can be computed in a straightforward manner using the communication cost distributions derived in sections V and VI. Figure 9 reports the expected packet delivery rate as a function of TTL for epidemic routing, and two-hops routing with $L = 3$ and 6. Node velocity is set to $v = 15m/sec$. As seen from the figure, epidemic routing is more effective than two-hops routing in terms of expected PDR: for instance, when $TTL = 60min$, epidemic routing achieves an expected PDR of 98.9%, compared to 60.2% achieved by two-hops routing with $L = 3$ and to 76.8% achieved by two-hops routing with $L = 6$. However, the price to pay for this packet delivery performance increase is communication cost, which is 15.04 for epidemic routing, 2.86 for two-hops routing with $L = 3$, and 4.96 for two-hops routing with $L = 6$ (see Table I).

We now consider routing protocol performance for a fixed $TTL = 60min$ and varying node speed. To make a more through comparison, we consider an additional routing proto-

col, namely the binary SW protocol defined in [15] as follows:

 – the source node initially holds $L = 2^h$ copies of the message, for some integer $h > 0$; when a node holding $K \leq L$ copies of the message encounters a new node, it delivers to the new node $\frac{K}{2}$ copies of the message, and keeps the remaining $\frac{K}{2}$ with it. When a node remains with a single copy of the message, it can deliver the message only to the destination node.

Binary SW routing has been shown in [15] to provide a very good tradeoff between packet delivery delay and communication cost. Due to the difficulties in theoretically analyzing its performance [14], binary SW routing protocol is evaluated only by means of simulation.

The expected packet delivery rate with epidemic, two-hops, and binary SW routing for $TTL = 60min$ and different speed values is reported in Figure 10. Parameter L in two-hops routing is set to 3 and 4, while binary SW is evaluated only with $L = 4$ (recall that L must be a power of 2 in Binary SW). The figure reports both the plots obtained through simulation, and those predicted using our theoretical framework (only for epidemic and two-hops routing). As seen from the plots, our theoretical framework can be used to faithfully predict routing performance, especially in case of epidemic routing. As for the comparison of the three routing protocols, as expected epidemic routing considerably outperforms two-hops and binary SW routing in terms of expected PDR, with PDR almost doubled for intermediate speed ranges. However, the price to pay is a much higher communication cost, as reported in Table I, which refers to the case of $v = 15m/sec$. As seen from the table, the expected communication cost with epidemic routing is more than five times larger than in case of two-hops routing with $L = 3$, and about four times larger than in the case of two-hops and binary SW routing with $L = 4$. The values reported in the table also show that our

theoretical framework can be used to provide a very accurate estimate of average communication cost with both routing protocols. Finally, we observe that binary SW routing provides a slight improvement over two-hops routing with the same value of $L = 4$ in terms of expected PDR, with a comparable communication cost, thus confirming its (slight) superiority over two-hops routing in the considered setting.

VIII. APPLICATION TO NON-COOPERATIVE NETWORKS

The material in this section, reported in the supplemental material, shows how the packet delivery delay distribution derived in Section V can be used to evaluate the effects of different degrees of node cooperation on the performance of epidemic routing.

IX. CONCLUSIONS AND FUTURE WORK

In this paper, we have introduced a framework for deriving the exact packet delivery delay and communication cost distribution for a large class of delay tolerant network routing protocols. We have also presented examples of application of our framework to specific routing protocols, namely epidemic and two-hops routing. Furthermore, we have shown how accurate characterization of the packet delivery delay distribution can be used to investigate epidemic routing performance in presence of different degrees of node cooperation within the network.

While the framework and specific results presented in this paper advance state-of-the-art in delay tolerant routing protocol performance evaluation, we admit that our results can be considered only as a first step towards gaining a better understanding of DTN routing protocol performance in general, and in non-cooperative (or partially cooperative) scenarios in particular. Several avenues for further research are open, such as deriving an analytical characterization of packet delivery delay and communication cost distribution with binary spray and wait routing [15], evaluating the effect of limited buffer size when network traffic is medium/high, including more sophisticated non-cooperative node behaviors in the analysis, and so on. We believe the tools and techniques presented in this paper will be helpful for those purposes.

REFERENCES

[1] D. Aldous, J. Fill, "Reversible Markov Chains and Random Walks on Graphs", *monograph in preparation*, available at http://www.stat.berkeley.edu/~aldous/RWG/
[2] A. Al Hanbali, A. A. Kherani, P. Nain, "Simple Models for the Performance Evaluation of a Class of Two-Hop Relay Protocols", *IFIP Networking*, pp. 191–202, 2007.
[3] A. Al Hanbali, P. Nain, E. Altman, "Performance of Ad Hoc Networks with two-hop relay routing and limited packet lifetime", *Performance Evaluations*, Vol. 65, n. 1-2, pp. 463–483, 2008.
[4] M. Balazs, *Sum of Independent Exponential Random Variables with Different Parameters*, http://www.math.bme.hu/~balazs/sumexp.html.
[5] C. Bettstetter, G. Resta, P. Santi, "The Node Distribution of the Random Waypoint Mobility Model for Wireless Ad Hoc Networks", *IEEE Trans. on Mobile Computing*, Vol. 2, n.3, pp. 257–269, July-Sept. 2003.
[6] L. Buttyan, L. Dora, M. Felegyhazi, I. Vajda, "Barter-based Cooperation in Delay-Tolerant Personal Wireless Networks", . *Proc. IEEE WoWMoM*, pp. 1–6, 2007.
[7] H.A. David, H.N. Nagaraja, *Order Statistics*, John Wiley and Sons, 2003.
[8] M. Grossglauser, D.N.C. Tse, "Mobility Increases the Capacity of Ad-Hoc Wireless Networks", *Proc. IEEE Infocom*, pp. 1360–1369, 2001.
[9] R. Groenevelt, P. Nain, G. Koole, "The Message Delay in Mobile Ad Hoc Networks", *Performance Evaluation*, vol. 62, n. 1–4, pp. 210–228, 2005.
[10] S. Jain, K. Fall, R. Patra, "Routing in Delay Tolerant Networking", Proc. ACM Sigcomm, pp. 145–158, 2004.
[11] D.B. Johnson, D.A. Maltz, "Dynamic Source Routing in Ad Hoc Wireless Networks", *Mobile Computing*, Kluwer Academic Publishers, pp. 153–181, 1996.
[12] T. Karagiannis, J.-Y. Le Boudec, M. Vojnovic, "Power Law and Exponential Decay of Inter Contact Times Between Mobile Devices", *Proc. ACM Mobicom*, pp. 183–194, 2007.
[13] A. Panagakis, A. Vaios, I. Stavrakakis, "On the Effects of Cooperation in DTNs", *Proc. IEEE COMSWARE*, pp. 1–6, 2007.
[14] G. Resta, P. Santi, "The Effects of Node Cooperation Level on Routing Performance in Delay Tolerant Networks", *Proc. IEEE Secon*, pp 413–421, 2009.
[15] T. Spyropoulos, K. Psounis, C.S. Raghavendra, "Efficient Routing in Intermittently Connected Mobile Networks: The Multicopy Case", *IEEE Trans. on Networking*, Vol. 16, n. 1, pp. 77–90, 2008.
[16] T. Spyropoulos, K. Psounis, C.S. Raghavendra, "Efficient Routing in Intermittently Connected Mobile Networks: The Single-copy Case", *IEEE Trans. on Networking*, Vol. 16, n. 1, pp. 63–76, 2008.
[17] T. Spyropoulos, K. Psounis, C.S. Raghavendra, "Performance Analysis of Mobility-assisted Routing", *Proc. ACM MobiHoc*, pp. 49–60, 2006.
[18] A. Vahdat, D. Becker, "Epidemic Routing for Partially Connected Ad Hoc Networks", *Tech. Rep. CS-200006*, Duke University, April 2000.
[19] X. Zhang, G. Neglia, J. Kurose, D. Towsley, "Performance Modeling of Epidemic Routing", *Computer Networks*, Vol. 51, pp. 2867–2891, 2007.

G. Resta received the MS degree in computer science from the University of Pisa, Italy, in 1988. In 1996, he became a researcher at the Istituto di Matematica Computazionale of the Italian National Research Council (CNR), Pisa. He is now a senior researcher at the Istituto di Informatica e Telematica (CNR) in Pisa. His research interests include computational complexity (especially in relation to linear algebra problems), parallel and distributed computing, and the study of structural properties of wireless ad hoc networks.

P. Santi received the Laura Degree and Ph.D. degree in computer science from the University of Pisa in 1994 and 2000, respectively. He has been researcher at the Istituto di Informatica e Telematica del CNR in Pisa, Italy, since 2001. During his career, he visited Georgia Institute of Technology in 2001, and Carnegie Mellon University in 2003. His research interests include fault-tolerant computing in multi-processor systems (during PhD studies), and, more recently, the investigation of fundamental properties of wireless multihop networks such as connectivity, lifetime, capacity, mobility modeling, and cooperation issues. He has contributed more than 60 papers and a book in the field of wireless ad hoc and sensor networking, he is Associate Editor of IEEE Trans. on Mobile Computing and IEEE Trans. on Parallel and Distributed Systems, he has been General Co-Chair of ACM VANET 2007 and 2008, and he is involved in the organizational and technical program committee of several conferences in the field. He is a member of IEEE Computer Society and a senior member of ACM and SIGMOBILE.

Supplemental Material for Paper "A Framework for Routing Performance Analysis in Delay Tolerant Networks with Application to Non Cooperative Networks"

Giovanni Resta, and Paolo Santi, *Member, IEEE,*

I. PROOF OF THEOREM 1

We recall that our framework dictates that, in order to evaluate the packet delivery delay distribution, the following need to be derived:

$$P(T_d < t) = \sum_{i=1}^{L} P(\sum_{j=1}^{i} \tau_j < t | E_{d,i}) \cdot P(E_{d,i}) . \quad (1)$$

We starting observing that the maximum number of packet copies (excluding possible copy delivered to destination) allowed to circulate in the network under epidemic routing is $M-1$, implying that the first summation in the above formula is up to the term $M-1$. Furthermore, since with epidemic routing every colored node is in CA state and the destination is chosen uniformly at random amongst the $M-1$ non-source nodes, we have $P(E_{d,i}) = \frac{1}{M-1}$ for each i, and (1) can be rewritten as follows:

$$P(T_d < t) = \frac{1}{M-1} \sum_{i=1}^{M-1} P(\sum_{j=1}^{i} \tau_j < t | E_{d,i}) . \quad (2)$$

Let us now consider the terms $Prob(\sum_{j=1}^{i} \tau_j < t | E_{d,i}) = P(Col_{i,t})$. We treat separately the case $i = 1$ and $i \geq 2$. If $i = 1$, the destination node is the first node encountered by the source, and $P(Col_{1,t})$ equals the probability that the two nodes meet within time t. Given that the meeting time between an arbitrary pair of nodes is exponentially distributed with rate $\frac{1}{emt}$, and observing that conditioning on $E_{d,1}$ implies that the distribution of the meeting time between the source and the destination (which is the first node met by the source amongst the $M-1$ nodes) is exponential with rate $(M-1)/emt$ (Fact **??**), we have:

$$P(Col_{1,t}) = F\left(\frac{M-1}{emt}, t\right) . \quad (3)$$

Let us now consider the case $i \geq 2$. We recall that, starting from T_{i-1} at which the $(i-1)$-th node is colored, there are i colored nodes in the network. Let $MT_{d,i} = \tau_{i-1} | E_{d,i}$ denote the meeting time between the destination and an arbitrary

colored node, conditioned on the event that destination is the i-th colored node. Given the conditioning event, the destination is the *first* node among the remaining $(M-i)$ nodes meeting one of the i nodes with a copy of the packet. Hence, $MT_{d,i}$ can be expressed as the minimum of a set of i exponential random variables with the same rate $\frac{(M-i)}{emt}$, whose distribution is exponential with rate $\frac{i(M-i)}{emt}$. We can then write, for $i \geq 2$:

$$
\begin{aligned}
P(Col_{i,t}) &= \int_0^t P(Col_{i,t} | T_{i-1} = x) f_{T_{i-1}}(x) dx = \\
&= \int_0^t F\left(\frac{i(M-i)}{emt}, t-x\right) f_{T_{i-1}}(x) dx, (4)
\end{aligned}
$$

where $f_{T_{i-1}}$ is the pdf of random variable T_{i-1}. In order to compute $f_{T_{i-1}}$, we observe that $T_{i-1} = \sum_{j=1}^{i-1} \tau_j$, where τ_j is the random variable corresponding to the meeting time between any of the j colored nodes and one of the remaining $(M - j)$ nodes. Similarly to above, we have that each of the τ_js is an exponential random variable with rate $\frac{j(M-j)}{emt}$, and that T_{i-1} is a sum of exponentially distributed random variables with different rates, where the rate λ_j of the j-th variable is $\frac{j(M-j)}{emt}$. We can then derive the distribution of T_{i-1} by Lemma 1 which, however, requires the parameters of the exponential random variables to be distinct. Unfortunately, this is not the case with the λ_js, since the rate of τ_j equals the rate of τ_{M-j}, for $j = 1, \ldots, \lfloor M/2 \rfloor$. To circumvent this problem, we artificially decrease the rate λ_j of the τ_js random variables with $j > \lfloor M/2 \rfloor$ by multiplying them by $(1 - \epsilon)$, for a small enough $\epsilon > 0$[1].

Hence, the rates of the τ_j variables are defined as:

$$\lambda_j = \begin{cases} \frac{j(M-j)}{emt} & \text{if } j \leq \lfloor \frac{M}{2} \rfloor \\ (1 - \epsilon) \cdot \frac{j(M-j)}{emt} & \text{if } \lfloor \frac{M}{2} \rfloor < j \leq M - 2 \end{cases} , \quad (5)$$

and Lemma 1 can be applied to compute $f_{T_{i-1}}(t)$.

II. ANALYSIS OF TWO-HOPS ROUTING

According to our framework, the packet delivery delay distribution can be computed as follows:

$$P(T_d < t) = \sum_{i=1}^{L} P(\sum_{j=1}^{i} \tau_j < t | E_{d,i}) \cdot P(E_{d,i}) . \quad (6)$$

---

[1]Our choice of decreasing, instead of increasing, the actual rate of the MT_js goes in the direction of providing an *upper bound* to the actual packet delivery delay, which is more useful than a lower bound for, e.g., QoS estimation.

Note that the outer summation ends at L, the maximum number of copies allowed to circulate in the network according to the two-hops routing protocol.

We start providing an exact characterization of the first and last term in the summation of (6), which are easy to derive.

The first term in the summation corresponds to the situation in which the destination is the first colored node, and the spraying phase does not even start. We are then in the same conditions as in the case of epidemic routing, and we can write:

$$P(\tau_1 < t|E_{d,1}) \cdot P(E_{d,1}) = F\left(\frac{M-1}{emt}, t\right) \cdot \frac{1}{(M-1)} \ . \ (7)$$

The case $i = L$ is also easy to handle. When $i = L$, the spraying process is finished (i.e., $(L-1)$ nodes have been colored by the source), and coloring of the destination occurs during the wait phase. Note that in the wait phase the coloring process is symmetric, hence any of the L nodes currently holding a copy of the packet can color the destination. Denote by $S_1, \ldots, S_{M-1}$ the random variables corresponding to the first time source node S meets node i after packet generation, for $i = 1, \ldots, M-1$ (see Figure 1). It is easy to see that the starting time of the wait phase is random variable $S_{(L-1)}$, i.e., the $(L-1)-th$ order statistic of random variables $S_1, \ldots, S_{M-1}$, whose pdf is $Ord(M-1, L-1, 1/emt, x)$. Furthermore, conditioned on $S_{(L-1)} = x$, the probability of delivering the packet to destination within time t is given by the cdf at time $(TTL - x)$ of an exponential random variable with rate L/emt, which represents the time at which the first amongst the L colored nodes meets the destination. We can then write:

$$P(\sum_{j=1}^{L} \tau_j < t|E_{d,L}) =$$

$$= \int_0^t F\left(\frac{L}{emt}, t-x\right) Ord(M-1, L-1, 1/emt, x)dx \ . \ (8)$$

The value of $P(E_{d,L})$ is computed as follows:

$$P(E_{d,L}) = 1 - \sum_{i=1}^{L-1} P(E_{d,i}) \ , \tag{9}$$

where the $P(E_{d,i})$s when $i = 2, \ldots, L-1$ are computed below.

Let us now consider the case $1 < i < L$. We first observe that event $E_{d,i}$ can be divided into mutually disjoint events $E_{d,i}^S$ and $E_{d,i}^{\overline{S}}$, corresponding to the situation in which the destination is colored by the source s or by a non-source node. Conditioned on $E_{d,i}^S$, the probability of delivering the packet to the destination within time t can be computed observing that the destination is the i-th node encountered by s, and that the random time of this encounter corresponds to the i-th order statistic of random variables $S_1, \ldots, S_{M-1}$. We can then write:

$$P(\sum_{j=1}^{i} \tau_j < t|E_{d,i}^S) = \int_0^t Ord(M-1, i, 1/emt, x) \ dx \ . \ (10)$$

We now compute $P(E_{d,i}^{\overline{S}})$, from which $P(E_{d,i}^S)$ can be trivially derived. We need to introduce some further notation
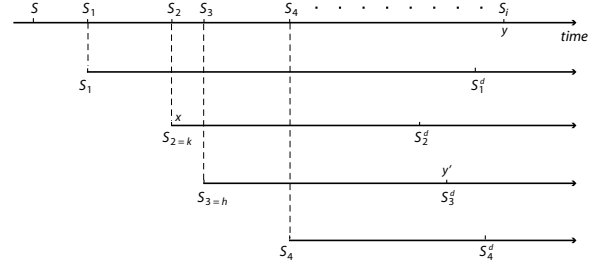

Fig. 1. The destination coloring stochastic process during the spraying phase.

(see Figure 1). We use random variables S_h^d to denote the time at which node h (which has been colored at time S_h) first meets the destination. Note that, conditioned on a specific value of $S_h = x$, $S_h^d - t$ has exponential distribution of rate $1/emt$, for $h = 1, \ldots, i-1$. In order to compute $P(E_{d,i}^{\overline{S}})$, we further subdivide $E_{d,i}^{\overline{S}}$ into disjoint events $E_{d,i,k}^{\overline{S}}$, $k = 1, \ldots, i-1$, where $E_{d,i,k}^{\overline{S}}$ is the event "the destination is colored by the k-th non-source node"[2]. Note that the above occurs if and only if $S_k^d < S_h^d$ for each $h = 1, \ldots, i-1$ (with $h \neq k$) *and* $S_k^d < S_i$ (see Figure 1).

Let us first compute the probability that $S_k^d < S_h^d$, for arbitrary distinct k, h, with $1 < k, h < i$. Conditioned on $S_h^D = y'$ and $S_k = x$, the pdf of event $S_k^d < S_h^d$ equals the probability that random variable $\hat{S}_k = S_k^d - S_k$, which is exponential of parameter $1/emt$, is below $(y' - x)$ (see Figure 1). We further observe that the pdf of variable S_k is $Ord(M-1, k, 1/emt, x)$. We can then write:

$$P(S_k^d < S_h^d) = \int_0^\infty G(y') f_{S_h^d}(y')dy' \ , \tag{11}$$

where $f_{S_h^d}(y')$ is the pdf of random variable S_h^d and

$$G(y') = \int_0^{y'} F\left(\frac{1}{emt}, y'-x\right) Ord(M-1, k, \frac{1}{emt}, x)dx \ .$$

In order to derive $f_{S_h^d}(y')$, we observe that, by setting $\hat{S}_h = S_h^d - S_h$, we can write $S_h^d = S_h + \hat{S}_h$. Random variable S_h is the h-th order statistic of random variables $S_1, \ldots, S_{M-1}$, and $\hat{S}_h$ is an exponential random variable of parameter $1/emt$. Hence, $f_{S_h^d}(x)$ is the pdf of a random variable which is the sum of two independent random variables with different (and nontrivial) distributions, which is not easy to derive. To circumvent this problem, we approximate the pdf of S_h with that of S_1 (i.e., of the first order statistic), which is exponential with rate $(M-1)/emt$. Now, S_h^d can be considered as the sum of two independent exponential random variables with different rates, whose distribution is given by Lemma 1. Summarizing, we have

$$f_{S_h^d}(y') \approx f_{S_1 + \hat{S}_h}(y') \ , \tag{12}$$

where $f_{S_1 + \hat{S}_h}(y')$ is defined as in Lemma 1.

Note that the probability mass of variable S_h is skewed to the right in the time axis with respect to that of variable S_1, with an increasing skewness for increasing h. Hence, our approximation of $f_{S_h^d}(y')$ becomes less and less accurate as

---

[2]Implicit here is the assumption that non-source nodes are ordered according to their meeting time with the source.

h increases. However, the simulation results presented in the following show that the above approximation is very accurate as long as the ratio L/M is around 0.1 or below. Deriving a better approximation of $f_{S_h^d}(y')$ is left for future work.

We now compute the probability that $S_k^d < S_i$, for any $1 < k < i$. Similarly to above, we observe that, conditioned on $S_i = y$ and $S_k = x$, the pdf of event $S_k^d < S_i$ equals the probability that $\hat{S}_k = S_k^d - S_k$ is below $(y - x)$ (see Figure 1). Furthermore, we observe that the pdf of S_i is $Ord(M - 1, i, 1/emt, y)$, and that of S_k is $Ord(M - 1, k, 1/emt, x)$. We can thus write:

$$P(S_k^d < S_i) = \int_0^\infty H(y) Ord(M - 1, i, \frac{1}{emt}, y)\, dy\,, \quad (13)$$

where

$$H(y) = \int_0^y F\left(\frac{1}{emt}, y - x\right) Ord(M - 1, k, \frac{1}{emt}, x)\, dx\,.$$

We can now approximate $P(E_{d,i}^{\overline{S}})$ as follows:

$$P(E_{d,i}^{\overline{S}}) \approx \sum_{k=1}^{i-1} \left(P(S_k^d < S_i) \prod_{\substack{h=1 \\ h \neq k}}^{i-1} P(S_k^d < S_h^d) \right)\,. \quad (14)$$

Note that in the above equation we assume that pairs of events $(S_k^d < S_h^d, S_k^d < S_{h'}^d)$, with $h \neq h'$, are mutually independent, which is not true in general.

We can now compute $P(E_{d,i}^S)$ by observing that the probability that the i-th node met by the source is the destination is $1/(M - (i - 1))$, implying:

$$P(E_{d,i}^S) = \left(1 - P(E_{d,i}^{\overline{S}})\right) \cdot \frac{1}{M - (i - 1)}\,. \quad (15)$$

In order to complete the derivation, we are left to compute the probability $P(\sum_{j=1}^i \tau_j < t | E_{d,i,k}^{\overline{S}})$, i.e., the probability that the packet is delivered to the destination within time t conditioned on the event that the destination is colored by the k-th non-source node, for some $1 < k < i$. Since under such conditioning the destination is colored at time S_k^d (see Figure 1), we have:

$$P(\sum_{j=1}^i \tau_j < t | E_{d,i,k}^{\overline{S}}) =$$

$$= \int_0^t F(\frac{M - i}{emt}, TTL - x) Ord(M - 1, k, \frac{1}{emt}, x)\, dx\,. \quad (16)$$

In the above equation, we have used the fact that, conditioned on event $E_{d,i,k}^{\overline{S}}$, the pdf of random variable S_k^d is closely approximated by the pdf of an exponential random variable of parameter $(M - i)/emt$, corresponding to the first encounter (which, conditioned on $E_{d,i,k}^{\overline{S}}$, we know is with the k-th relay node) between the destination and one of the i nodes (including the source) currently holding a copy of the packet.

Summarizing, we can compute any of the terms $P(\sum_{j=1}^i \tau_j < t | E_{d,i}) \cdot P(E_{d,i})$ with $1 < i < L$ in the summation of (6) as follows:

$$P(\sum_{j=1}^i \tau_j < t | E_{d,i}) \cdot P(E_{d,i}) =$$

$$= P(\sum_{j=1}^i \tau_j < t | E_{d,i}^S) P(E_{d,i}^S) +$$

$$+ \sum_{k=1}^{i-1} P(\sum_{j=1}^i \tau_j < t | E_{d,i,k}^{\overline{S}}) P(E_{d,i,k}^{\overline{S}})\,,$$

where $P(E_{d,i,k}^{\overline{S}})$ is defined as follows:

$$P(E_{d,i,k}^{\overline{S}}) = P(S_k^d < S_i) \cdot \left(\prod_{\substack{h=1 \\ h \neq k}}^{i-1} P(S_k^d < S_h^d) \right)\,.$$

The communication cost distribution can be readily obtained from the above derivations, and is defined as follows:

Proposition 1: Let C be the random variable representing the number of nodes holding at least one copy of the packet in the network at delivery time. With two-hops routing, we have

$$P(C = i) = \begin{cases} \frac{1}{M-1} & \text{if } i = 1 \\ \left(1 - P(E_{d,i}^{\overline{S}})\right) \cdot \frac{1}{M-(i-1)} & \text{if } 1 < i < L \\ 1 - \sum_{j=1}^{L-1} P(C = j) & \text{if } i = L \\ 0 & \text{otherwise} \end{cases}\,,$$

where $P(E_{d,i}^{\overline{S}})$ is defined in equation (14).

III. Application to non-cooperative networks

We consider the following cooperation scenarios:
a) Coop: in this scenario, nodes behave according to the routing protocol specifications (*full cooperation*).
b) Def: in this scenario, nodes discard packets received by other nodes (unless the node itself is the destination), and correctly send own packets (*no cooperation*).
c) Rand: in this scenario, nodes forward a packet received by other nodes with fixed probability p.
d) TfT: in this scenario, nodes forward a packet received by other nodes with probability p, where p depends on network conditions; for definiteness, in the following we assume that p equals the observed PDR for own packets.

Routing performance will be evaluated in terms of expected PDR, where a packet is considered to be correctly received by the destination if and only if the packet is delivered within time TTL (time to live) since its generation, where TTL is a tunable parameter. Observe that, by the Law of Large Numbers, evaluating the expected PDR is equivalent to computing the probability of the event "a packet generated at an arbitrary node u and destined to node v is delivered to v within time TTL". Hence, the expected PDR can be readily obtained by computing $P(T_d < TTL)$, i.e. the value of the packet delivery delay cdf at point TTL.

A. Full cooperation scenario

Given the above discussion, the expected PDR of epidemic routing in the full cooperation scenario can be simply obtained by computing the packet delivery delay cdf derived in Section V at $t = TTL$.

To validate the accuracy of our analysis, we have performed a set of simulations. The simulation setting is similar to the one
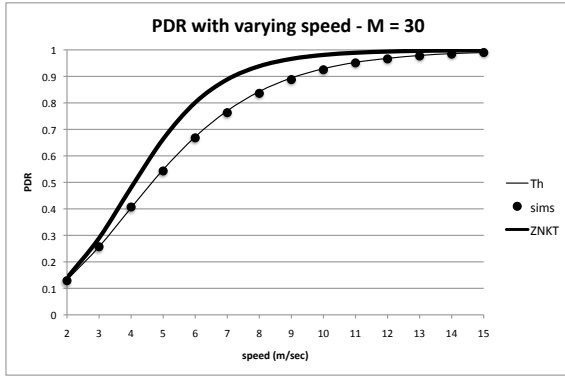
Fig. 2. Probability of correctly delivering a packet within time $TTL = 3600s$ with epidemic routing for RWP mobile networks with $M = 30$ nodes and increasing node velocity. Fully cooperative scenario.

used in Section V, except that the outcome of an experiment is a binary value 0/1 computed depending on whether the packet is delivered to the destination within time $TTL = 3600s$ (value 1) or not (value 0). We performed a large set of such experiments (more than 20000 for each parameter setting), and experimentally estimated $P_{FC}(T_d < TTL)$ (where FC stands for Full Cooperation) as the number of successful experiments (i.e., experiments returning value 1) over the total number of experiments.

The results of this experimental estimation of $P_{FC}(T_d < TTL)$ when $M = 30$ and nodes' speed v is varied between 2 and $15m/sec$ are reported in Figure 2. The figure reports also the analytical estimation computed according to the analysis reported in Section V, as well as the asymptotical bound for the PDR derived in [1] (ZNKT curve). As seen from the figure, our analytical PDR estimation is very accurate for all the range of speed considered, while the bound of [1] is accurate only when the experimental PDR is either very low or very high.

B. No cooperation scenario

In case all nodes are defective, the protocol used to route packets between nodes is irrelevant, since the only possible way to deliver a packet to destination is through direct transmission between source and destination. Under our working assumption of exponentially distributed meeting times, the probability that source and destination come into each other transmission range within time TTL is given by:

$$P_{NC}(T_d < TTL) = \left(1 - e^{-\frac{TTL}{emt}}\right) \, ,$$

where NC stands for No Cooperation.

The above characterization has been validated through simulation, which has shown an almost perfect overlapping between simulation results and analytical estimate.

C. Rand cooperation scenario

We recall that in case of *Rand* cooperation, a node which is not the final destination forwards the packet with a given fixed probability p. The relevant cases are when $0 < p < 1$, since otherwise we are either in the no cooperation ($p = 0$) or full cooperation ($p = 1$) scenario.

The node coloring process under epidemic routing and *Rand* cooperation is more complex than in the full cooperation

scenario, due to one additional source of randomness in packet propagation, namely, the probability p of forwarding the packet when a node in CA state meets an uncolored one. Since node coloring is no longer deterministic, not only the expected meeting time, but also the expected inter-meeting time[3] of the underlying mobility model comes into play. To better understand this point, let us analyze the process of coloring the first non-source node. When the source meets the first (uncolored) node – which occurs according to an exponential random variable with rate $(M - 1)/emt$ –, the new node is colored with probability $p < 1$. So, with probability p, the rate of the first coloring is the same as in the fully cooperative scenario. However, with probability $(1 - p)$, the new node (call it u) remains uncolored, and the rate of first coloring is determined not only by the rate of encounter of the source with the second uncolored node – which is a random variable with distribution equal to the pdf of the second order statistics of random variables S_is defined in the previous section –, but also by the rate of re-encounter of the source with node u. More specifically, denoting by $S_{(i)}$ the i-th order statistic of the S_is, and by S_u the time between the first and second encounter between the source and node u, we have that the second encounter between the source and an uncolored node is given by $\min\{S_{(2)}, S_{(1)} + S_u\}$. At the time of the second encounter, we again have a node coloring with probability p, and the above reasoning is re-iterated.

As seen from the above description, the characterization of relevant stochastic properties of epidemic routing under *Rand* cooperation is very complex, as it involves mixing estimation of expected meeting and inter-meeting times between different groups of nodes. For this reason, we consider a related stochastic coloring process, in which potential forwarding nodes are a-priori divided into cooperative and non-cooperative nodes. More specifically, we assume that a *fixed* fraction p of the $M - 2$ potential forwarders are cooperative (i.e., forward a packet with probability 1), while the remaining $(1-p)(M-2)$ nodes are non-cooperative (i.e, forward a packet with probability 0). The analysis of this second coloring process is straightforward, since it is equivalent to performing epidemic routing in a fully cooperative network with $p(M-2)+1$ – the $p(M-2)$ cooperative relays plus the destination – (instead of $(M-1)$) nodes. Thus, the expected PDR of epidemic routing under probabilistic cooperation with parameter $0 < p < 1$ can be closely approximated by the packet delivery delay cdf at $t = TTL$ in a network composed of $p(M-2)+1$ nodes.

We have verified the accuracy of our analytical estimation of PDR with epidemic routing under *Rand* cooperation through extensive simulation. The simulation setting is the same as in the previous section, with the only difference that packets are propagated with fixed probability p upon encounters. Note that we have chosen settings for p such that $p(M - 2)$ is an integer. Figures 3 and 4 report the experimental cdf and pdf obtained with $v = 15m/sec$ (very similar results have been obtained with other values of v). The figures also report the corresponding distribution obtained using our proposed

---

[3]The expected inter-meeting time is defined as the expected time between two successive meetings between an arbitrary node pair.
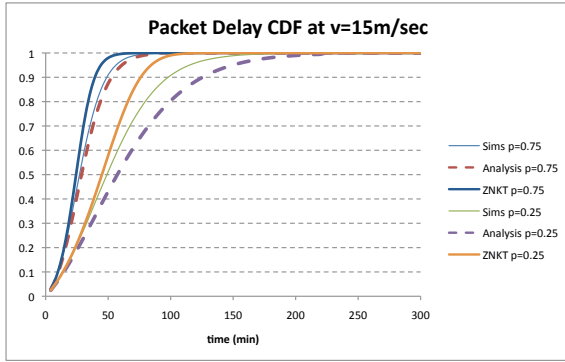
Fig. 3. Experimental and analytical packet delivery delay cdf with $v = 15m/sec$ in the *Rand* cooperation scenario with $p = 0.25$ and $p = 0.75$.
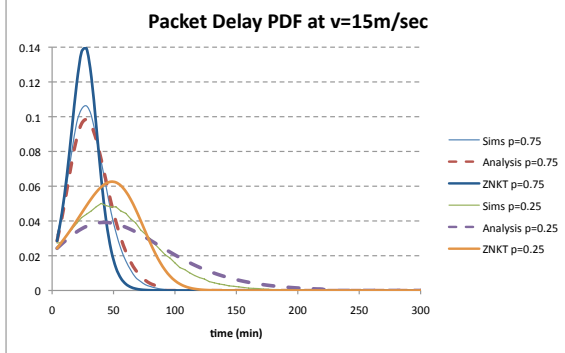

Fig. 4. Experimental and analytical packet delivery delay pdf with $v = 15m/sec$ in the *Rand* cooperation scenario with $p = 0.25$ and $p = 0.75$.

approximation, and the one reported in [1]. As seen from the figure, our proposed approximation is significantly closer to the actual distribution than that presented in [1]. Most significantly, our approximation can be used to *lower bound* the probability of packet delivery at a certain time, while that of [1] upper bounds such probability. This can be seen very well in Figure 4, which shows that while our approximated pdf skews the probability mass towards larger delivery times with respect to the actual distribution, the opposite happens with the pdf derived from [1]. We can then conclude that, differently from the estimation of [1], our bound can be used to provide *minimal performance guarantees* of epidemic routing under *Rand* cooperation.

We have also evaluated, for a fixed value of $TTL = 3600sec$, the PDR as obtained through simulations, and through the formulas presented herein and in [1]. The results are reported in Figure 5. Confirming the previous observation, our estimation turns out to be a relatively accurate *lower bound* to the actual PDR, in contrast to the ZNKT bound, which considerably *overestimates* performance. It is also worth observing that while ZNKT characterization turns out to be relatively accurate only for relatively low or high PDR values, our characterization is relatively accurate for the whole range of PDR values, and it is indeed very accurate for $p = 0.75$.

D. TfT scenario

The characterization of epidemic routing performance under *Rand* cooperation presented in the previous sub-section can be used to predict stationary operational points of a network under the *TfT* scenario. We recall that with *TfT* cooperation,
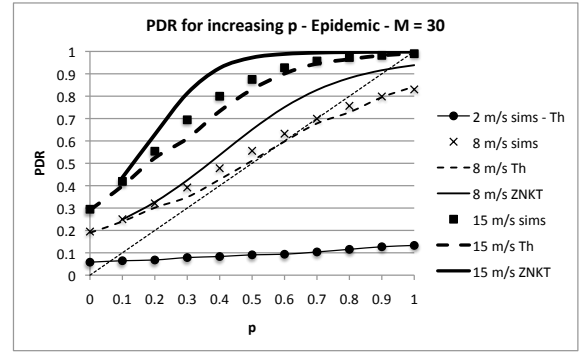

Fig. 6. Probability of correctly delivering a packet within time $TTL = 3600s$ with epidemic routing and *Rand* cooperation with various values of p. The $2m/s$ curves for both analytical bounds are overlapped with simulation results, and are not shown.

the packet forwarding probability is adaptive to network conditions: more specifically, the packet forwarding probability p_u for node u equals the long-term observed PDR of the packets sent by u.

Stationary network operational points can be estimated by characterizing the PDR under *Rand* cooperation with different values of the parameter p, and finding the value(s) $\bar{p}$ of p such that the resulting PDR equals $\bar{p}$. To see that these are stationary operational points of a network under *TfT* cooperation, assume all nodes are currently observing a PDR of $\bar{p}$; given *TfT* cooperation, all nodes will forward packets with probability $\bar{p}$, i.e., they behave like a network with *Rand* cooperation of parameter $\bar{p}$, which will keep the observed PDR (and, hence, the cooperation level) at $\bar{p}$.

The characterization of stationary operational points under *TfT* cooperation by means of simulation, and using our analytical bound and the one of [1], are reported in Figure 6. We first observe the existence of stationary operational points for all considered speed values; operational points can be very bad (below 0.1 at $2m/s$), relatively good (around 0.7 at $8m/s$), or even close to optimal performance (at $15m/s$). This indicates that the effect of *TfT* cooperation on epidemic routing performance is highly dependent on network conditions: *the better the network conditions, the lesser the effect of the reduced cooperation level*. To be more specific, when the underlying network performance is bad (lowest speed), the performance drop due to *TfT* cooperation w.r.t. fully cooperative scenario is in the order of 50% (PDR drops from 0.13 to 0.06); in the average speed case, the performance drop is in the order of 16% (PDR drops from 0.83 to 0.7); finally, in case of very good network performance (highest speed), the performance drop is negligible.

Concerning the accuracy of the analytical bounds, we observe that while the ZNKT bound can be used to accurately estimate stationary operational points only when epidemic routing performance under full cooperation is either very bad of very good, it considerably overestimates the stationary operational point for the intermediate speed value. In particular, according to ZNKT bound the stationary operational point when the speed is $8m/s$ is around 0.9, while it is around 0.7 according to the simulation results. On the other hand, our analytical bound is able to provide an accurate estimate of
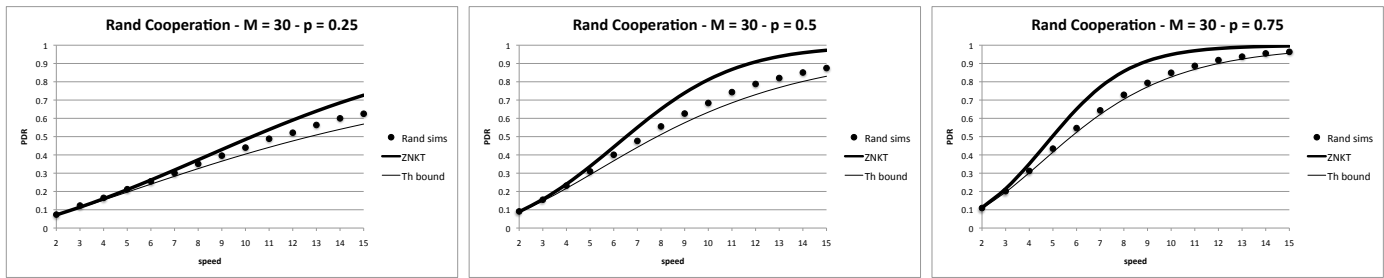
Fig. 5. Probability of correctly delivering a packet within time $TTL = 3600s$ with epidemic routing for RWP mobile networks with $M = 30$, $Rand$ cooperation, and varying node velocity for different values of p.
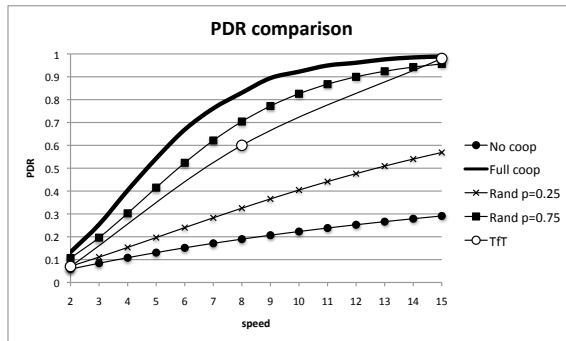

Fig. 7. Probability of correctly delivering a packet within time $TTL = 3600s$ with epidemic routing in different cooperation scenarios and varying speed.

stationary operational points for the entire range of considered speeds (the estimate is around 0.6 when speed is $8m/s$).

E. Performance comparison

In Figure 7, we report the bounds on PDR with different degrees of node cooperation computed with our formulas for varying node speed. As seen from the figure, even a modest degree of node cooperation (*Rand* cooperation with $p = 0.25$) is sufficient to substantially increase performance with respect to the most pessimistic scenario of no cooperation. In particular, PDR is increased of as much as of a factor of two for the highest speed value. It is also interesting to observe that *TfT* cooperation even more substantially increases performance over the most pessimistic cooperation scenario (up to a factor of three improvement), providing a performance similar to the one obtained with full cooperation for the higher speed values.

REFERENCES

[1] X. Zhang, G. Neglia, J. Kurose, D. Towsley, "Performance Modeling of Epidemic Routing", *Computer Networks*, Vol. 51, pp. 2867–2891, 2007.