

Performance of Opportunistic Content Distribution under Different Levels of Cooperation

Ólafur Ragnar Helgason*, Franck Legendre†, Vincent Lenders‡, Martin May§ and Gunnar Karlsson*

*ACCESS Linnaeus Center and School of Electrical Engineering

KTH, Royal Institute of Technology

100 44 Stockholm, Sweden

Email: {olafurr,gk}@ee.kth.se

†Communication Systems Group

ETH Zürich, 8092 Zürich, Switzerland

Email: legendre@tik.ee.ethz.ch

‡Armasuisse

3000 Bern, Switzerland

Email: vincent.lenders@armasuisse.ch

§Technicolor

Issy-les-Moulineaux, France

Email: martin.may@technicolor.fr

Abstract—In this paper, we study the performance of opportunistic content distribution in a delay-tolerant network (DTN). Our primary interest is the benefit of node cooperation on the content dissemination and we investigate thoroughly how different levels of cooperation affect the content delivery delay. The evaluation is based on two complementary approaches, analytical stochastic models (Markov chains) and simulations that use real-world contact traces with high time resolution. These two independent approaches show agreeing results on the qualitative behavior of the system. Our major finding is that node collaboration – even if limited – dramatically improves performance compared to non-cooperation. In particular, we find that limiting the number of times a node shares a content item gives only a slightly worse performance than under unlimited cooperation. Besides, we find that a critical number of nodes is necessary to disseminate content efficiently. Eventually, based on the analytic model we also find that the details of the mobility show little effect on the delivery time when the nodes cooperate.

I. INTRODUCTION

In recent years, there have been proposed systems for opportunistic content distribution in the wireless ad-hoc domain. Examples are those targeted towards people in metropolitan areas [1], [2], [3], [4], vehicular environments [5], [6], and wildlife sensor networks [7], [8]. Mobile nodes collaborate in distributing content by sharing and downloading content in a peer-to-peer fashion over the radio contacts that are established when two nodes are within each others communication range. Still there is much work to be done in understanding thoroughly the behaviour and performance of these systems and how they are affected by individual parameters such as mobility, node resources, radio propagation, cooperation degree, etc.

Stochastic modelling is probably the most commonly used method to analytically evaluate the performance of mobile wireless networking systems of the nature we consider in

this work. Sometimes, these theoretical models can, under simplifying assumptions, give important insight about fundamental issues such as system capacity or scaling properties. In most cases however, the analytical models quickly become intractable when the simplifying assumptions are lifted. Therefore, simulations are commonly used to evaluate scenarios that are analytically intractable and to assess the effect of simplifying assumptions made in the analytic models. However, even with the advanced simulators used to date, capturing all the necessary parameters that affect the performance of mobile wireless systems is non-trivial. Evaluating efficiency in real-life is therefore still very important, even though experiments are exposed to many parameters which cannot be controlled (radio propagation, interference, mobility, repeatability etc.). Finally, relating and comparing results from theory, simulations and experiments is a significant challenge.

In this paper we evaluate the performance of opportunistic content distribution from one node to all (broadcast dissemination) using one-hop unicast communications and considering different cooperation strategies. We use two performance evaluation approaches i.e., a stochastic Markovian model on the one hand and simulations based on experimental traces on the other hand. The analytical model and the trace-driven simulations are complementary methods to study the opportunistic content distribution and the goal is to obtain qualitative results on the system behaviour and performance rather than giving an exact quantitative comparison. Specifically, we study how the content distribution is affected by node cooperation. We consider different levels of cooperation where the willingness of the nodes to share content with peers differs, i.e., (i) no, (ii) unlimited, and (iii) limited sharing. With the two latter models, *unlimited* and *limited* sharing, we expect all nodes to be involved in the content distribution. Yet, since devices, such as mobile phones, small audio/video players and PDA's, are

small and resource limited we presume that the capability of the nodes to cooperate will be restricted. In particular, limited battery lifetime will discourage cooperative transmissions of data.

In the field of delay- and disruption tolerant networks (DTN's) there are some works [9], [10] that relate theory and experimental results, similarly to what we do here. However, they are mostly focused on the delivery of a message to a single given destination while we consider the delivery of a message to a group of users. Also, the measurement granularity (i.e. time interval between individual measurement probes) of our experimental traces is significantly finer than that of the traces from many previous experiments which allows us to capture also the shorter contacts in our measurements.

The contributions of this paper are:

- Markovian models of opportunistic content distribution for three different cooperation strategies.
- Simulation results based on real-life mobility traces with fine granularity (high time resolution).
- Comparison of the analytic and simulation results.

The main findings are:

- Markov models are well suited to capture qualitatively the scaling properties of the system.
- The performance is significantly improved when nodes cooperate and it only degrades slightly when the content sharing is restricted.
- The performance is not very sensitive to the particular details of the mobility, such as the inter-contact time distribution, in the case when nodes cooperate.
- For a good content spreading performance a critical population of nodes is required. After reaching this critical size, a further increase in population will not significantly improve the delivery time.

The remainder of this paper is organized as follows. Section II defines the network model, the different cooperation strategies, and the performance metrics. These strategies are formalized as Markov chains in Section III with analytical and simulation results. Section IV presents an empirical evaluation using real-world traces. Section V summarizes related work and we conclude our work with future directions in Section VI.

II. WIRELESS CONTENT DISTRIBUTION MODEL

We consider a scenario in which N nodes are moving in an area A . The scenario is that of a delay-tolerant network (DTN) with a relatively sparse node density and intermittent connectivity where a connected network cannot be assumed. Hence, we model the network as a series of pair-wise contacts. The *contact duration* is defined as the time between the moment two nodes move into wireless range until they move apart. The *inter-contact time* is defined as the elapsed time between two consecutive contacts of the same pair of nodes.

We study the performance of the content distribution process by investigating the time that it takes for content available at one node at time $t = 0$ to spread to all other $N - 1$ nodes. We

assume that it is possible to exchange an entire content item of a node over a single contact using unicast communication at the link layer.¹ In practice, this implies that the contact duration or the wireless link throughput is high compared to the content size.

We are interested in how the content dissemination process is affected by the degree at which the nodes cooperate to share content. In particular, we identify three strategies to spread content with different degrees of node cooperation.

No cooperation: Nodes do not cooperate in spreading the content. A node receives content only when it enters in contact with the initial content provider (that published the content at $t = 0$) which shares it with all nodes.

Unlimited cooperation: Nodes cooperate without any restrictions. That is, every time a mobile node is in contact with any other node that is interested in content, it shares that content with the requesting node.

Limited cooperation: Nodes cooperate, but restrict the number of times they share a given content item. The motivation for this type of limitation is typically an incentive to save a node's own resources (i.e., battery, CPU, link load). We denote by k the number of times a node will provide each content item to other nodes. For example, when k is equal to one, each nodes will provide the content over a maximum of one contact. For $k = 2$, all nodes will share the content with a maximum of two other nodes.

For all the cooperation strategies, we look at the performance of the content dissemination time. To characterize this process, we look at two metrics: the *overall* and the *individual* delivery times. The *overall delivery time*, denoted by T_{odt} , is the time until content has spread to all the N nodes. Hence, $T_{odt} = \max(\{T_i\}_{i=1,\dots,N})$ where T_i is the time at which node i obtains the content. The *individual delivery time*, denoted by T_{idt} , is the time that it takes for an arbitrary node to obtain the content. When all the nodes are identical and if the mobility processes of the nodes are identical and independent then the sequence of random variables $\{T_1, T_2, \dots, T_N\}$ are IID and they have the same distribution as T_{idt} . From a performance viewpoint, T_{odt} is a measure of the performance of the system as a whole while T_{idt} is a measure of the system performance as seen from an arbitrary node.

Borrowing terms from epidemic modelling [11], the nodes in our model are classified as *susceptible*, *infected* or *recovered*. In epidemic modelling a susceptible individual will become infected by the epidemic disease when it meets an infected individual. After some time an infected individual may recover from the disease and become immune and thus no longer infect other susceptible nodes. Analogously in our model, whenever a node that has some content (infected) makes a contact with as susceptible node, the latter node obtains the content and becomes infected. A recovered node is a node that carries the content but does not share it, as

¹Broadcast communications at the link layer are unreliable and difficult to implement so we focus on the one-hop unicast case.

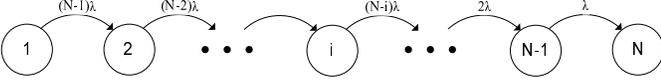


Fig. 1. Markov chain for *no* cooperation.

in the *no*- or limited cooperation strategies. We will use this terminology throughout this paper.

In the following sections, we explore the performance of this content distribution model considering different assumptions for the node population (homogeneous vs. non-homogeneous), the inter-contact time determined by the mobility process (Markovian, non-Markovian), and the area A (bounded vs. unbounded).

III. ANALYSIS BASED ON STOCHASTIC MODELS

In this section, we describe a stochastic model, based on Continuous-time Markov Chains (CTMC), for opportunistic content distribution. We are interested in capturing how the degree in which nodes cooperate affects the performance of the spreading and how the spreading performance scales as the number of nodes grows.

We consider a homogeneous set of nodes moving in bounded area. We denote by the random variable $\tau_{i,j}(n)$ the n^{th} inter-contact time between nodes i and j and assume that the sequence of inter-contact times $\{\tau_{i,j}(n)\}_{i,j,n}$ are identical, independent and exponentially distributed random variables with rate λ . This allows us to model the three cooperation strategies introduced in the previous section using Continuous-Time Markov Chains. Later in this section we relax the Markovian assumption and compare the analytical model with simulations with non-exponentially distributed inter-contact times.

A. No Cooperation

$N - 1$ susceptible nodes are interested in the particular content item held by a single infected provider. Since the nodes are not cooperating for spreading the content the only way to obtain it is by infection from the content provider which is willing to share content with all nodes it meets without limitation.

We denote by the random variable $X(t)$ the number of infected nodes at time t . The process $\{X(t); t \geq 0\}$ is a pure birth process with rate $\lambda_i = (N - i)\lambda$ for all states $i = 1, \dots, N-1$ as shown in Fig.1. The CTMC for this process is absorbing with $\{i = N\}$ as the absorbing state; all other states are transient.

In a CTMC, the state sojourn time is exponentially distributed with rate equal to the sum of the rates going out of the state. Therefore the mean time to absorption from the initial state $\{i = 1\}$ is equal to the overall delivery time T_{odt} , and its mean is calculated as the sum of the mean time spent in each of the transients states. We therefore have that

$$E[T_{odt}] = \sum_{i=1}^{N-1} \frac{1}{\lambda_i} = \frac{1}{\lambda} \sum_{i=1}^{N-1} \frac{1}{(N-i)} = \frac{1}{\lambda} \sum_{j=1}^{N-1} \frac{1}{j} = \frac{1}{\lambda} H_{N-1} \quad (1)$$

where $H_n = \sum_{i=1}^n \frac{1}{i}$ is the n -th harmonic number. It is well known that the harmonic series does not converge when $n \rightarrow \infty$. An asymptotic expansion for the harmonic numbers is $H_n = \gamma + \ln(n) + \mathcal{O}(\frac{1}{n})$ where γ is Euler's constant. Thus we have

$$E[T_{odt}] = \frac{1}{\lambda} \left(\gamma + \ln(N-1) + \mathcal{O}\left(\frac{1}{N-1}\right) \right) \quad (2)$$

from which we deduce that $E[T_{odt}] \in \mathcal{O}(\ln(N))$.

The susceptible nodes can only obtain content from the single provider. Since all the nodes are identical and the inter-contact times are IID, the mean individual delivery time is equal to the the mean of the inter-contact time, or

$$E[T_{idt}] = \frac{1}{\lambda} \quad (3)$$

B. Unlimited Cooperation

With unlimited cooperation, an infected node will remain infected forever and is therefore always willing to share content. As before, we assume that there are initially $N - 1$ susceptible nodes in the area. We denote by $X(t)$ the number of infected nodes at time t and thus $N - X(t)$ is the number of susceptible nodes for $t > 0$.

The process $\{X(t); t \geq 0\}$ is a pure-birth Markov process with positive transition rates

From	To	Rate	(4)
i	$i + 1$	$i(N - i)\lambda$	

The Markov chain in (4) is a variation of the *simple stochastic Markovian epidemic* [11] and its transition diagram is shown in Fig. 2. Similarly to the previous case, $\{i = N\}$ is the only absorbing state and we are interested again in the mean overall delivery time, $E[T_{odt}]$, which equals the mean time to absorption from state $\{i = 1\}$. We are also interested in the mean of the individual delivery time $E[T_{idt}]$.

Let $R_{i,i+1}$ denotes the time that it takes for the process, starting from state i , to reach state $i + 1$, $i \geq 1$. $R_{i,i+1}$ is exponential with rate $(N - i)\lambda$ and thus

$$E[R_{i,i+1}] = \frac{1}{\lambda_i} = \frac{1}{i(N-i)\lambda}$$

The time it takes all the nodes to obtain the content is $T_{odt} = R_{1,N}$ and its expected value is $E[R_{1,N}] = E[R_{1,2}] + E[R_{2,3}] + \dots + E[R_{N-1,N}]$. We thus have

$$E[T_{odt}] = E[R_{1,N}] = \sum_{i=1}^{N-1} E[R_{i,i+1}] = \frac{1}{\lambda} \sum_{i=1}^{N-1} \frac{1}{i(N-i)} \quad (5)$$

and since

$$\sum_{i=1}^{N-1} \frac{1}{i(N-i)} = \frac{1}{N} \left(\sum_{i=1}^{N-1} \frac{1}{i} + \sum_{i=1}^{N-1} \frac{1}{N-i} \right) = \frac{2}{N} H_{N-1} \quad (6)$$

we have the following asymptotic expansion for $E[T_{odt}]$

$$E[T_{odt}] = \frac{2}{\lambda N} \left(\gamma + \ln(N-1) + \mathcal{O}\left(\frac{1}{N-1}\right) \right) \quad (7)$$

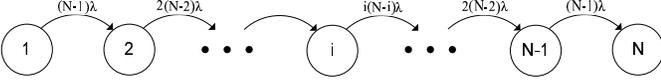


Fig. 2. Markov chain for *unlimited* cooperation.

where γ is Euler's constant. Since $\lim_{x \rightarrow \infty} \frac{k_1 \ln(x-1)}{x} = 0$ and $\lim_{x \rightarrow \infty} \frac{k_2}{x} = 0$, where k_1 and k_2 are constants, we have that $E[T_{odt}] \in \mathcal{O}(1)$.

Thus, when the nodes utilize the peer contacts to cooperatively share content the mean overall delivery time is bounded by a constant and does not grow to infinity when $N \rightarrow \infty$ as is the case when there is no peer cooperation (2).

To obtain $E[T_{idt}]$, we denote by the random variable $T_{k,N-1}$ the time until k out of the $N-1$ susceptible nodes have become infected. We also introduce the event $B = \{A \text{ given susceptible node is the } k\text{-th to become infected}\}$. Then we have that

$$\begin{aligned} E[T_{idt}] &= \sum_{k=1}^{N-1} E[T_{k,N-1}] P\{B\} \\ &= \sum_{k=1}^{N-1} E[T_{k,N-1}] \frac{1}{N-1} \end{aligned} \quad (8)$$

where $P\{B\} = \frac{1}{N-1}$ is uniform since all nodes are identical and inter-contact times are IID. $E[T_{k,N-1}]$ is the mean time that it takes the Markov chain (4) to reach state $\{i = k+1\}$ and therefore we have

$$E[T_{k,N-1}] = \frac{1}{\lambda} \sum_{i=1}^k \frac{1}{i(N-i)} \quad (9)$$

and thus

$$E[T_{idt}] = \frac{1}{\lambda(N-1)} \sum_{k=1}^{N-1} \sum_{i=1}^k \frac{1}{i(N-i)} \quad (10)$$

By writing out terms in the double, it is easily shown that

$$\begin{aligned} \sum_{k=1}^{N-1} \sum_{i=1}^k \frac{1}{i(N-i)} &= \\ &= 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{N-2} + \frac{1}{N-1} \\ &= H_{N-1} \end{aligned}$$

The mean individual delivery time is thus

$$E[T_{idt}] = \frac{1}{\lambda(N-1)} H_{N-1} \quad (11)$$

from which we can deduce that $E[T_{idt}] \in \mathcal{O}(1)$.

C. Limited Cooperation

Here we restrict the uploading such that nodes will share each content item only k times to save resources. Once a node has reached this limit, it recovers and stops sharing. Let $X(t)$ be the number of infected and recovered nodes at time t (note the slightly different meaning of $X(t)$ here

compared to previous cases), and $Y_m(t)$ the number of infected nodes who have to share the content m times before becoming recovered, where $m = 1, \dots, k$. Thus right after a susceptible node obtains the content it will become infected and since it has never shared its content Y_k will increase by one. Also, after an infected node that has shared its content n times ($n = 0 \dots k-1$) meets a susceptible node, Y_{k-n} will decrease by one and Y_{k-n-1} will increase by one.

When $k = 1$ each node being infected will then only share the content once (except for the last infected node). Hence, only a single node will be infected at each time until the content is fully spread. In fact, when a susceptible node meets an infected node, they will become infected and recovered respectively and thus there is always just one infected node at a time in the system. Therefore this stochastic process is the same as the one-dimensional Markov chain $\{X(t); t \geq 0\}$ for the *No* cooperation strategy in Fig.1. It should also be clear that when $k \rightarrow \infty$ a node will share a content whenever it is asked for and then this scenario is the same as the previous *Unlimited* strategy (4).

For a general $k > 1$ the Markov process for the system will consist of $k+1$ random variables since we have to count the number of nodes for each of the sequence of variables $\{Y_m(t); t \geq 0\}_{m=1, \dots, k}$ as well as the total number of nodes that have been infected. Here we give the model for the case $k = 2$. Initial conditions are $X(0) = Y_2(0) = 1$, $Y_1(0) = 0$. The positive state transition rates for the process $\{(X(t), Y_1(t), Y_2(t)); t \geq 0\}$ are

From	To	Rate
(i, j, k)	$(i+1, j+1, k)$	$k(N-i)\lambda$
	$(i+1, j-1, k+1)$	$j(N-i)\lambda$

(12)

This 3-dimensional Markov chain does not easily lend itself to closed-form solutions as the previous models. Therefore we resort to numerical analysis and simulations in the following subsection.

D. Numerical results

All the CTMCs in the previous subsections reach an absorbing state when the number of infected plus possibly recovered nodes reaches N , i.e. when all nodes have obtained the content. We obtain closed-form expressions for the overall delivery time and individual delivery time for the first two cooperation strategies and we resort to numerical calculation for the limited cooperation strategy in (12). Our evaluation uses a mean inter-contact time of 697s which is in accordance with what is obtained in our experimental traces as described in Section IV.

In Fig.3(a), we plot the mean overall delivery time $E[T_{odt}]$ as a function of the node population N . We compare the cases for no cooperation, unlimited cooperation, and limited cooperation with $k = 2$. As pointed out before, a consequence of the IID assumption is that the limited cooperation case with $k = 1$ is equal to the no cooperation case and therefore we do not plot that explicitly. We clearly see that when nodes cooperate, the delivery time is significantly reduced. From

the figure we can observe the asymptotic behaviour of the no cooperation and unlimited cooperation models that was derived in (2) and (7). The results for the limited-cooperation case are particularly interesting. If each node only shares the content twice the overall delivery time is only slightly longer than what it is in the unlimited cooperation case. Moreover, it seems to have the same asymptotic behaviour as the unlimited cooperation case although we have not proved this analytically. In the limited cooperation model of (12), the state space grows quite fast as the number of nodes increases. The size of the transition-rate matrix \mathbf{Q} of the Markov Chain grows as $O(N^3)$ and therefore we can only solve (12) exactly for small systems (a maximum of $N = 36$ in our evaluations). However, even with systems of this size, the performance differences between the individual cooperation schemes are very evident. For larger values of N ($N > 36$), we resort to simulation of (12). For each value of N , a mean value of T_{odt} is obtained from 1000 simulation runs (95% confidence intervals are omitted from figure since they cannot be distinguished from the mean value). For a discussion on how to numerically find the mean-time to absorption in a CTMC we refer to [12].

Fig.3(b) shows the mean individual delivery time $E[T_{idt}]$ for the different cooperation models. Data for the limited cooperation case is obtained from simulations and, as for the overall delivery time, it is interesting to see that individual delivery time is only slightly longer for the limited cooperation than what it is for the unlimited case. It is also interesting to see that initially, when N goes from 1 to approximately 20, there is a sharp decrease in $E[T_{idt}]$ which then levels out as N increases. This suggests that for obtaining good performance, an initial mass of nodes is needed and a further increase beyond that will only slightly improve performance.

In Fig.3(c), we study the effect of the Markovian assumption in our analytic models. It has recently been proven that the inter-contact time between two independent mobile nodes decays at least exponentially fast for commonly used mobility models, such as the Random Waypoint and Random Direction mobility models, as long as the area boundary is finite [13]. However, some experiments indicate that the inter-contact time distribution has a heavy tail and obeys a power-law [9] while other results are more optimistic. In particular, the Log-normal distribution has been shown to be a good match for three commonly used references sets of traces [14] and the Weibull distribution is exhibited in our own (see Section IV) as well as other traces [15]. Fig.3(c) therefore compares results from the Markovian analytic model with results obtained by simulating the no cooperation and unlimited cooperation models with a Log-Normal and Weibull inter-contact time distribution, both having the same mean value of $697s$ as with the exponential distribution. The Log-Normal distribution has the same variance ($0.49 \times 10^6 s^2$) as the exponential distribution while the Weibull distribution has a higher variance ($10^6 s^2$) which matches our experimental results. Simulation results from 1000 runs are plotted with 95% confidence intervals for the no cooperation case but these are omitted for the unlimited cooperation case since they can barely be distinguished from

the mean value in the plot. Direct numerical comparison shows some quantitative difference. In the no cooperation scenario, the mean overall delivery time is longest for the Weibull distribution while it is shortest for the exponential case. For the unlimited cooperation case this dominance is not the same, in this case the Weibull distribution has the shortest mean overall delivery time while the Log-Normal distribution results in the slowest. Despite these numerical differences we conclude that, from a qualitative viewpoint, the assumption of exponentially distributed inter-contact times in our Markov models captures the behavior of the system, in particular when comparing the difference between cooperation types.

IV. ANALYSIS WITH REAL-WORLD MOBILITY TRACES

In this section, we analyze empirically the different strategies of cooperation using real-world mobility traces. With this evaluation, we lift some of the simplifying, yet strong, assumptions presented in Section III: The traces we use in this section have been obtained with a setting where the mobile nodes and users are not bound to a fixed area. Furthermore, our previous model assumed an homogeneous node population (i.e., i.i.d.) with Markovian inter-meeting times (except for the simulations), whereas the trace population here is non-homogeneous and the empirical inter-contact time distribution is determined by the mobility of the users.

A. Experimental setup

The traces (contact patterns) are collected in scenario with volunteering test users carrying one of 20 identical HP iPAQ hx2400 pocket PCs running Microsoft Windows Mobile Edition 2003. The iPAQs are connected using their integrated IEEE 802.11b WLAN module (a Samsung SWL-2750C chipset) operating in ad-hoc mode. The devices are broadcasting two beacons per second with a random offset to avoid synchronized access to the wireless medium. These beacons allow us to determine which devices were in contact over the time of the experiment and to later reconstruct the relative mobility pattern of the users (i.e., contact and inter-contact patterns). The experiment involved $N = 20$ users during 5 working days. These users are staff members of a research lab, all working on the same floor. However, since the users may leave the floor during the experiment, the area A should be considered as unbounded. In analyzing the empirical inter-contact time distribution of our traces, we found that it is best fitted with a Weibull distribution with shape $\alpha = 0.71$ and characteristic life $\beta = 558$. For more details about the traces, we refer to [16].

We simulate the different content distribution strategies by replaying the traces. For the traces, we consider only the period between 10am and 5pm in order to have enough mobility (a significant amount of people were not in the office before 10am or after 5pm). In our experiment, the content infected node is fixed for the whole five days and emits new content every hour. Hence, our simulated setting consists in total of 30 disseminated pieces of content (5 days \times 6 pieces of content).

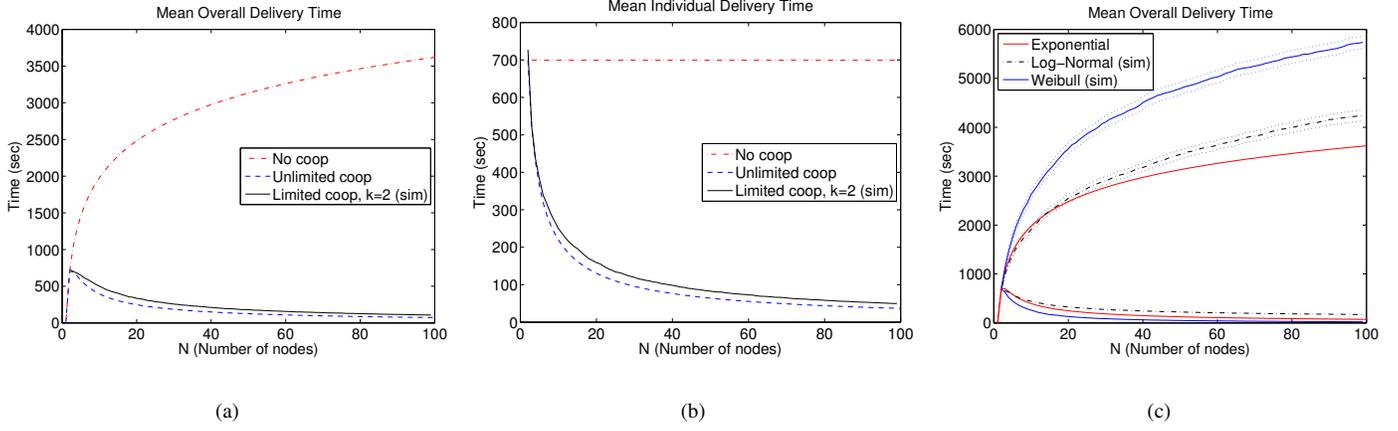


Fig. 3. Overall and individual delivery time for the different cooperation strategies. 3(a): Overall delivery time for no, unlimited and limited cooperation. 3(b): Individual delivery time for no, unlimited and limited cooperation. 3(c): Overall delivery time for three different inter-contact time distributions for no cooperation (top three curves) and unlimited cooperation (lower three curves).

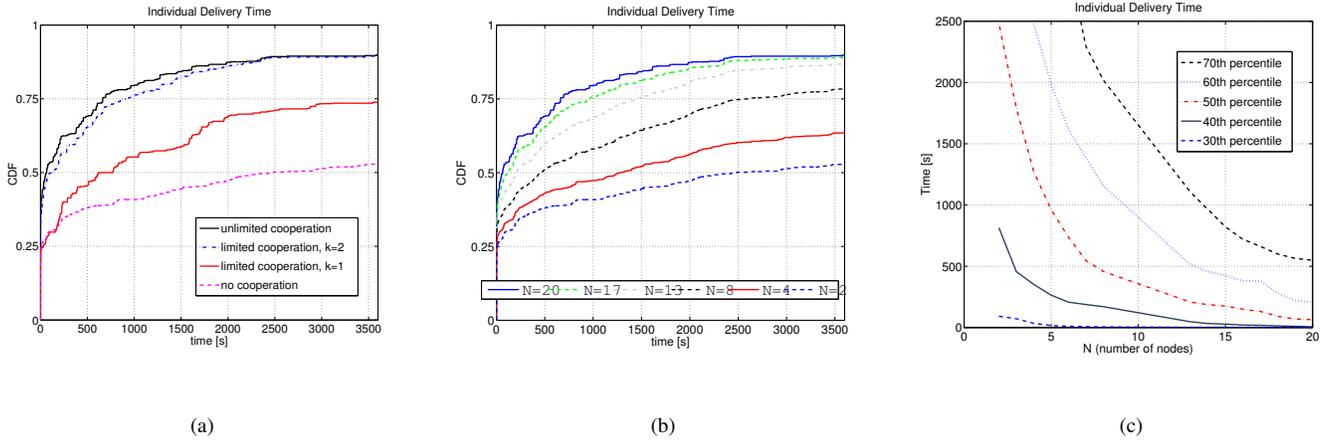


Fig. 4. Empirical individual delivery time for the trace-based scenario. 4(a): CDF for no, unlimited and limited cooperation. 4(b): CDF for unlimited cooperation varying the node population. 4(c): Individual delivery time versus N for different percentiles.

B. Impact of node cooperation

Since the area of A is unbounded in the experimental setting, some nodes might possibly never get infected when they move away. For this reason, we do not consider here the overall delivery time of the content but only compare the individual delivery times (T_{idt}) and look at the overall empirical cumulative distribution function (CDF). Figure 4(a) shows the corresponding empirical CDFs for the unlimited, limited, and no cooperation content distribution modes. We clearly see the benefit of cooperation versus no cooperation. By applying unlimited cooperation, the individual delivery time of the first 50% of deliveries is decreased from approximately 2500 seconds to 64 seconds (a factor of almost 40!). After one hour, unlimited cooperation manages to achieve a delivery that is 37.2% higher than no cooperation. Furthermore, we observe that by limiting the cooperation to two uploads ($k = 2$), the distribution process is only slightly worse compared to

unlimited cooperation. This is in agreement with our analytic results in Fig.3(a) and Fig.3(b).

An interesting phenomenon is observed in Figure 4(a) of our simulations when we compare no cooperation to limited cooperation with $k = 1$. In the no cooperation case the content emitter is fixed whereas in the limited cooperation with $k = 1$ the content is spread from one mobile node to another (although there is always just one infected node at each instance of time). We see that the individual delivery time for a given percentile is shorter in the limited cooperation with $k = 1$ than what it is in the no-cooperation case. This reveals that the nodes in our experiment are not homogeneous: some nodes are more efficient in distributing the contents than the fixed node. Hence we note the benefit of mobility in spreading content, even though the content is disseminated by one node at a time, as in the no cooperation case. This effect is not captured by our analytical models because the models for no cooperation and limited cooperation with $k = 1$ are the same.

We recall from Section III that this is a consequence of the assumption that nodes are homogeneous and that the inter-contact times of all nodes are IID.

C. Impact of node population

To assess how the node population N affects the distribution process, we performed simulations in which we omit particular nodes that were present in the traced experiment. This allows us to simulate settings with arbitrary node densities $N = 2, \dots, 20$. A reduced node population can be viewed as a scenario with less nodes or a scenario in which only part of the nodes are cooperating in the content distribution.

The CDFs for six selected values of N under the unlimited cooperation strategy are plotted in Figure 4(b). To avoid random selection of particular nodes which are not effective at distributing content, the presented CDFs are computed for all possible node selection combinations from the complete set of nodes in the traced experiment. We clearly see how the number of cooperating nodes affects the content delivery time. Figure 4(c) shows the same results under a slightly different viewpoint. The individual delivery time is plotted versus N on the horizontal axis over all nodes and all content. Since not all the nodes will eventually retrieve some content, we plot different percentiles. For example the 50th percentile represents the time for 50% of all content to be delivered in the simulations. Starting with two nodes (one infected and one susceptible node) and as N increases, we observe an initial sharp decrease of the delivery time which then flattens out as N increases. This same behavior is also seen in the analytical results in Fig.3(b). This suggests that the content distribution requires an initial critical population of nodes and that a further increase will only contribute slightly to reducing the delivery time.

The impact of the node population on the delivery time is supported by both methods: note that Fig.3(b) and Fig.4(c) show the same behaviors; Fig.3(c) and Fig.4(c) show furthermore that although the inter-contact distribution has a quantitative effect on the exact numerical values, qualitative results are the same when there is cooperation.

V. RELATED WORK

There are other works within the area of delay-tolerant networks that study real user traces and contrast experimental results with analytical models, most notably the DieselNet [10], [5] and the Huggle [17], [9] projects. These works however focus on routing protocols for unicast delivery while our work focuses on broadcasting a message in a DTN. Also, our traces capture mobility at more fine-grained precision than the other works. We use a 0.5s measurement interval while [5] and [9] use 7s and 120s intervals respectively.

Content dissemination systems, that utilize opportunistic contacts to exchange content between mobile devices carried people in metropolitan areas, have been proposed in the 7DS system [1] and the work on infostations [2]. In 7DS, nodes cooperate by sharing data, rebroadcasting, and caching popular entries similarly to our unlimited cooperation strategy. In the

mobile infostation, cooperation is limited and based on social contracts, similar to a tit-for-tat approach, which can in part be captured by our limited cooperation model with $k = 1$. The results of this paper are therefore applicable to these systems. Our results also apply to recent DTN content dissemination systems in vehicular environments [6], and wildlife sensor networks [7], [8].

Our analytic models are related to those in the mathematical field of epidemic modelling, which has a long history where both stochastic and deterministic models are used to study the spreading of infectious diseases [11], [18]. Epidemic modelling has received considerable attention from the networking research community, as there are many scenarios that arise which are analogous to the spreading of epidemics. Some examples are distributed computing and peer-to-peer networks [19] [20], database updates and maintenance [21], ad-hoc networks[22], multicast [23], [24] and more. In delay-tolerant networks, stochastic models, related to epidemic theory, have been used to study the performance of different unicast routing approaches [25], [8], [26]. In [27] we introduced an early version of some of the stochastic models of section III in this paper, which we have extended herein to capture other performance parameters. Also, in [27] we did not provide any comparison with experimental data. In [28] the authors develop a model based on ordinary differential equations (ODE) to study the performance of epidemic routing in DTN's. The ODE models appear as fluid limits of Markovian models under appropriate scaling as the number of nodes grow. Although stochastic models more accurately capture the system behavior, ODE fluid models can often be solved when the stochastic model is not tractable. It is our intention to study fluid models for some of the stochastic models presented in this paper.

Recently, many paradigms have been revisited with network coding approaches [29], [30], [31]. However, networking coding might in practice be costly and difficult to implement [32], and it requires adaption to provide a performance gain in the wireless case [33].

Performance evaluation for cooperative peer-to-peer systems in the wired Internet, such as Bittorent [34] or Avalanche [35], are closely related to our work, the latter having some comparable findings to ours. However, due to mobility and the nature of the wireless medium, content dissemination in a DTN setting faces dynamics (node arrival/departure, churn) on a much shorter timescale than what is generally experienced in a wired setting.

VI. SUMMARY AND FURTHER DIRECTIONS

In this work, we studied the effect of cooperation on the performance of opportunistic content dissemination networks. We considered different levels of cooperation in sharing content i.e., (i) no, (ii) unlimited, and (iii) limited cooperation. Our results are obtained from two complementary approaches, an analytic Markovian model and simulations based on real-life experimental traces.

The stochastic model assumes homogeneous nodes that have exponentially distributed inter-contact times, but we also

give simulation results for other non-exponential inter-contact time distributions. We also evaluate the different strategies in a realistic scenario by replaying fine-grained user traces (with a measurement time resolution of 0.5s), obtained from an experiment with real-users. The evaluations give results that are in agreement. The major finding is that by limiting the number of uploads, we achieve only a slightly worse performance than when nodes cooperate in an unlimited manner. This is confirmed by experimental results and by analyzing the asymptotic behavior of the stochastic models. We are not aware of any previous work that has performed a comparison between different content dissemination strategies in the particular setup we are targeting and our findings give insights on issues such as the effect of the population size, inhomogeneity of nodes and the effects of a non-Markovian mobility processes.

As part of our future work, we intend to further support our experimental evaluation by comparison with more experimental traces, collected by ourselves and others. For the stochastic models, we intend to obtain asymptotic bounds for the limited cooperation model, given by the Markov chain in (12), and to verify analytically if the unlimited and limited cooperation models are asymptotically equivalent as our experimental and simulation results indicate. Eventually, we will investigate the benefit of the limited cooperation strategy on energy savings but also in terms of fairness, sharing incentive, and storage management.

REFERENCES

- [1] M. Papadopouli and H. Schulzrinne, "Seven degrees of separation in mobile ad hoc networks," in *Proceedings of the IEEE GLOBECOM*, 2000.
- [2] W. H. Yuen, R. D. Yates, and S. C. Mau, "Exploiting data diversity and multiuser diversity in noncooperative mobile infostation networks," in *Proceedings of IEEE INFOCOM*, San Francisco, USA, March 2003.
- [3] G. Karlsson, V. Lenders, and M. May, "Delay-tolerant broadcasting," *IEEE Transactions on Broadcasting*, vol. 53, no. 2, pp. 369 – 381, Mar. 2007.
- [4] M. Motani, V. Srinivasan, and P. Nuggehalli, "Peoplenet: Engineering a wireless virtual social network," in *Proceedings of ACM MobiCom*, 2005.
- [5] J. Burgess, B. Gallagher, D. Jensen, and B. Levine, "MaxProp: Routing for Vehicle-Based Disruption-Tolerant Networks," in *Proceeding of IEEE INFOCOM*, Barcelona, Spain, April 2006.
- [6] J. Ott and D. Kutscher, "A Disconnection-Tolerant Transport for Drive-thru Internet Environments," in *Proceedings of IEEE INFOCOM*, Miami, USA, March 2005.
- [7] P. Juang, H. Oki, Y. Wang, M. Martonosi, L.-S. Peh, and D. Rubenstein, "Energy-Efficient Computing for Wildlife Tracking: Design Tradeoffs and Early Experiences with ZebraNet," in *Proceedings of the Tenth International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS-X)*, San Jose, CA, USA, October 2002.
- [8] T. Small and Z. J. Haas, "The shared wireless infostation model: a new ad hoc networking paradigm (or where there is a whale, there is a way)," in *Proc. ACM MobiHoc*, Annapolis, Maryland, USA, 2003, pp. 233–244.
- [9] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of Human Mobility on the Design of Opportunistic Forwarding Algorithms," in *Proceedings of IEEE INFOCOM*, Barcelona, Spain, April 2006.
- [10] X. Zhang, J. Kurose, B. N. Levine, D. Towsley, and H. Zhang, "Study of a Bus-Based Disruption Tolerant Network: Mobility Modeling and Impact on Routing," in *Proc. ACM Intl. Conf. on Mobile Computing and Networking (Mobicom)*, September 2007, pp. 195–206.
- [11] D. Daley and J. Gani, *Epidemic Modelling: An Introduction*. Cambridge, United Kingdom: Cambridge University Press, 1999.
- [12] P. Bremaud, *Markov chains : Gibbs fields, Monte Carlo simulation, and queues*. New York, USA: Springer-Verlag, 1998.
- [13] H. Cai and D. Y. Eun, "Crossing over the bounded domain: from exponential to power-law inter-meeting time in manet," in *Proc. MobiCom*, Montreal, Quebec, Canada, 2007, pp. 159–170.

- [14] V. Conan, J. Leguay, and T. Friedman, "Characterizing pairwise inter-contact patterns in delay tolerant networks," in *Proc ACM Autonomics*, Rome, Italy, Oct. 2007.
- [15] D. Lelescu, U. C. Kozat, R. Jain, and M. Balakrishnan, "Model T++: An Empirical Joint Space-Time Registration Model," in *Proceedings of ACM MOBIHOC*, Florence, Italy, 2006.
- [16] V. Lenders, J. Wagner, and M. May, "Analyzing the Impact of Mobility in Ad Hoc Networks," in *ACM/Sigmobile Workshop REALMAN 2006*, Florence, Italy, May 2006.
- [17] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, "Pocket Switched Networks and Human Mobility in Conference Environments," in *Proceedings of the ACM SIGCOMM 2005 Workshop on Delay-Tolerant Networking (WDTN)*, Philadelphia, Pennsylvania, USA, August 2005.
- [18] H. Andersson and T. Britton, *Stochastic Epidemic Models and Their Statistical Analysis*. New York, USA: Springer-Verlag, 1999.
- [19] P. T. Eugster, R. Guerraoui, A.-M. Kermarec, and L. Massoulie, "From epidemics to distributed computing," *IEEE Computer*, vol. 37, no. 5, pp. 60–67, 2004.
- [20] S. Voulgaris and M. van Steen, "An epidemic protocol for managing routing tables in very large peer-to-peer networks," in *Proc. of IFIP/IEEE DSOM*. Berlin, Germany: Springer-Verlag, 2003.
- [21] A. Demers, D. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. Sturgis, D. Swinehart, and D. Terry, "Epidemic algorithms for replicated database maintenance," in *Proc ACM PODC*, Vancouver, British Columbia, Canada, 1987, pp. 1–12.
- [22] A. Khelil, C. Becker, J. Tian, and K. Rothermel, "An epidemic model for information diffusion in manets," in *Proc. ACM MSWiM*, Atlanta, Georgia, USA, 2002, pp. 54–60.
- [23] A.-M. Kermarec and A. J. Ganesh, "Efficient and adaptive epidemic-style protocols for reliable and scalable multicast," *IEEE Trans. Parallel Distrib. Syst.*, vol. 17, no. 7, pp. 593–605, 2006.
- [24] K. P. Birman, M. Hayden, O. Ozkasap, Z. Xiao, M. Budiu, and Y. Minsky, "Bimodal multicast," *ACM Trans. Comput. Syst.*, vol. 17, no. 2, pp. 41–88, 1999.
- [25] R. Groenevelt, P. Nain, and G. Koole, "The message delay in mobile ad hoc networks," *Perform. Eval.*, vol. 62, no. 1–4, pp. 210–228, 2005.
- [26] T. Small and Z. J. Haas, "Resource and performance tradeoffs in delay-tolerant wireless networks," in *Proc. WDTN'05 ACM SIGCOMM workshop on Delay-tolerant networking*, Philadelphia, Pennsylvania, USA, 2005, pp. 260–267.
- [27] Ó. R. Helgason and G. Karlsson, "On the effect of cooperation in wireless content distribution," in *Proceedings of the IEEE/IFIP WONS 2008*, Garmisch-Partenkirchen, Germany, Jan. 2008.
- [28] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, "Performance modeling of epidemic routing," *Comput. Networks*, vol. 51, no. 10, pp. 2867–2891, 2007.
- [29] S. Deb, M. Médard, and C. Choute, "Algebraic gossip: a network coding approach to optimal multiple rumor mongering," *IEEE/ACM Transactions on Networking (TON)*, vol. 14, no. SI, pp. 2486–2507, 2006.
- [30] C. Fragouli, J. Widmer, and J.-Y. Le Boudec, "A network coding approach to energy efficient broadcasting: From theory to practice," in *Proc. IEEE Infocom*, Barcelona, Spain, 2006.
- [31] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "Xors in the air: practical wireless network coding," *IEEE/ACM Transactions on Networking (TON)*, vol. 16, no. 3, pp. 597–510, 2008.
- [32] M. Wang and B. Li, "How practical is network coding?" in *Proc. IEEE IWQoS*, June 2006.
- [33] P. Chaporkar and A. Proutiere, "Adaptive network coding and scheduling for maximizing throughput in wireless networks," in *Proc. ACM MobiCom*, Montreal, Quebec, Canada, Sept. 2007.
- [34] BitTorrent.org, "Protocol Specification," <http://www.bittorrent.org/protocol.html>, 2006.
- [35] C. Gkantsidis and P. R. Rodriguez, "Network coding for large scale content distribution," in *Proc. IEEE Infocom*, Miami, USA, 2005.