Performance of Epidemic Content Distribution in Opportunistic Networks

Ljubica Pajevic, Gunnar Karlsson, Ólafur Helgason
KTH Royal Institute of Technology
Linnaeus ACCESS Center
Stockholm, Sweden
{ljubica, gk, olafurr}@kth.se

ABSTRACT

Epidemic spreading is a common approach to mitigate frequent link disruptions and to support content-centric information dissemination in opportunistic networks. Stochastic models, often used to characterize epidemic processes, introduce assumptions which, on one hand, make them analytically tractable, while on the other, ignore attested characteristics of human mobility. We investigate the fitness of a simple stochastic model for content dissemination by comparison with experimental results obtained from real-life mobility traces. We examine four mobility datasets and consider content delivery delay as a performance metric. Our finding is that a homogeneous model is unable to capture the performance of content dissemination with respect to content delivery delays.

Keywords

Epidemic modeling, content distribution, opportunistic networks, ad hoc networks

1. INTRODUCTION

In opportunistic networks, highly dynamic network topology and intermittent connectivity make routing a challenge. A large body of algorithms for opportunistic networking employ epidemic spreading principle. Adopting the principles of epidemic modeling from the field of mathematical biology, stochastic modeling has become a method commonly used in networking. However, for the sake of analytical tractability, models often assume identical mobility and contact patterns for all nodes in the network. Recently, studies which consider heterogeneous networks have been emerging; yet, it is debatable whether there are savings in complexity and increased understanding from using the models compared to simulations.

In this poster, we present an empirical study of epidemic content spreading by using real-world mobility traces. Then, we consider an analytic model proposed in [3], and investigate if a homogeneous model can be utilized to evaluate the performance of opportunistic networks when the assumptions on network homogeneity are relaxed.

2. OPPORTUNISTIC CONTENT DISTRIBUTION: MODEL AND EVALUATION

The application scenario we consider is that of disseminating information by utilizing opportunistic contacts, based on user interest. Sharing local news, traffic and tourist information in public areas, public announcements at massive events, or mobile advertisements are common examples where this distribution scheme can be used. We are interested in evaluating the performance of spreading in terms of delivery delays.

Epidemic model

We consider a homogeneous set of N mobile nodes, moving in a confined area and exchanging information through intermittent contacts. Initially, a single node carries a piece of information (a content item) and $N-1$ nodes are interested in obtaining the contents. We assume that the mobility of nodes is such that inter-contact times between any two nodes can be modelled by identical, independent and exponentially distributed random variables with rate λ. This allows analysis of the epidemic process with a stochastic model based on a continuous-time Markov chain, described in [3].

To characterize the performance, we look at two metrics: the time until the information has reached all the N nodes in the network, denoted by *overall delivery time* T_{odt}, and the *individual delivery time* T_{idt} , defined as the time until an arbitrary node has obtained the contents. Their expected values, $E[T_{odt}]$ and $E[T_{idt}]$ can be found from expressions given in Tab. 1.

Table 1: Expected delivery times

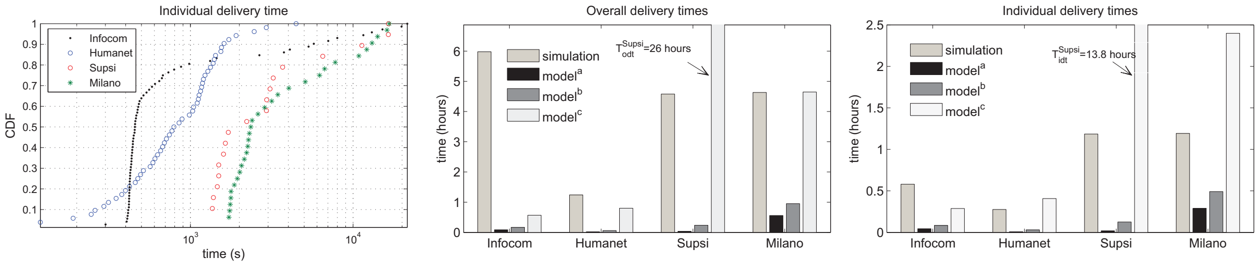
Overall delivery time $E[T_{odt}]$	$\frac{2}{N\lambda}H_{N-1}$
Individual delivery time $E[T_{idt}]$	$\frac{1}{\lambda(N-1)}H_{N-1}$

$H_n = \sum_{i=1}^{n} 1/i$ is the n-th harmonic number

Mobility Traces

To cover various scenarios, we use four experimental datasets, diverse in time granularity, number of participants in the experiments and in the experiment duration. The datasets report direct pair-wise contacts between users moving in relatively restricted areas: a conference venue, a university, or a company building. We briefly describe the contexts where the traces were collected and the acquisition methods used, and our methodology of pre-processing the traces.

- **Infocom** traces [5], obtained at Infocom 2006, report contacts between 78 experiment participants carrying iMote devices during four days. The scanning interval was 120 seconds.
- **Humanet** traces [1] describe human mobility in an office scenario, reporting proximity traces of 52 participants during one working day. The users were carrying Bluetooth customized devices, which were scanning every 5 seconds to capture direct contacts with other devices.
- **Supsi** dataset [2] comprises of contacts in a group of 39 participants, from three institutes, located in two buildings. The participants were carrying sensor nodes with



(a) (b) (c)

Figure 1: CDFs of the contents delivery times (a). Comparison of the overall (b) and individual delivery times (c).

a transmission range of 5 meters and a scanning cycle of 10 milliseconds.

- **Milano** dataset [4] was obtained by 44 participants. Contacts were logged by customized radio devices, operating with a transmission range of 10 meters and a configurable scanning interval of one second.

The duration of the experiments varies from a single day (*Humanet*) to several weeks (*Supsi*); to obtain inter-contact times, we consider only the days when a considerable number of contacts were recorded, and observe only the contact events that occurred during work hours (usually from 9:00 to 18:00).

Evaluation

We want to assess the capability of a homogeneous epidemic model to capture the process of epidemic content spreading in real-life scenarios. First, we simulate the epidemic process in four scenarios, and plot the CDFs of the contents delivery times in Fig. 1(a). For each of the traces, we choose a single day when the nodes were most active, seen as the number of contacts recorded during that day. The analytic model can be seen as a simple and efficient tool to estimate the performance; its simplicity stems from the fact that it requires only two input parameters: the number of nodes in the network and the node contact rate. In order to validate the analytic model, we compute the same metrics, overall and individual delivery times, given by the expressions in Tab. 1.

As a first evaluation, we assume that node interactions can be well described by the aggregate inter-contact time distributions, that is, the empirical distribution of inter-contact times estimated over all possible node pairs. From the aggregate distribution we find the contact rate as the reciprocal of average inter-contact time. Figs. 1 (b) and (c) depict the simulation results and the expected delivery times (denoted by **model**^{a}). Clearly, this method significantly underestimates the delivery delays, calling for more careful investigation of pair-wise node interactions. Therefore, we propose the second method to estimate the contact rate for a set of nodes. First, for every pair of nodes which reported contacts we find the average contact rate. Then, we find the empirical distributions of those contact rates and perform curve fitting. In all four cases, log-normal distribution seems to provide the best fit for the contact rate distributions. Next, we find the average contact rates from the fitted distributions and calculate the expected delivery times. Bars denoted by **model**^{b} in Fig. 1 correspond to this case. With a slight improvement from the previous method, the discrepancies between the analytic and simulation results are still significant. Acknowledging the fact that contacts for many node pairs are not observable from the traces (from 9% in the *Infocom* to almost 70% in the *Supsi* trace), we propose the third method to estimate average contact rate over all node pairs, by compensating for the missing node pairs. To all node pairs which have not recorded any contacts—the two nodes which haven't met during the experiment—we assign the same value for the average inter-contact time (for the delays in Fig. 1 that value equals the duration of the full trace), find the average contact rate over all pairs, and compute the delivery times. These results are denoted by **model**^{c}. We observe that this method gives better estimation than the previous two. However, the inconsistency (in some scenarios the method underestimates delivery delays, while in other vastly overestimates), makes it impracticable for use in general.

3. SUMMARY

We studied the performance of epidemic content distribution in opportunistic networks and empirically evaluated the content delivery delays by using four mobility datasets, chosen to represent a small system of users moving in a bounded area. We proposed three methods of treating the statistical data obtained from the traces, and showed that the homogeneous model is unable to accurately capture the epidemic process of the real-life scenarios. For our future work, we will aim at modeling epidemic spreading in heterogeneous systems by using other types of stochastic models.

4. REFERENCES

[1] J. M. Cabero, V. Molina, I. Urteaga, F. Liberal, and J. L. Martin. CRAWDAD data set Tecnalia Humanet (v. 2012-06-12), June 2012.

[2] A. Förster, K. Garg, H. A. Nguyen, and S. Giordano. On context awareness and social distance in human mobility traces. In *Proc.ACM*, MobiOpp '12, New York, NY, USA, 2012.

[3] O. Helgason, F. Legendre, V. Lenders, M. May, and G. Karlsson. Performance of opportunistic content distribution under different levels of cooperation. In *Proc. European Wireless Conference (EW)*, 2010.

[4] P. Meroni, S. Gaito, E. Pagani, and G. P. Rossi. CRAWDAD data set unimi/pmtr (v. 2008-12-01), Dec. 2008.

[5] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau. CRAWDAD trace cambridge/haggle/imote/infocom2006, May 2009.