Evaluation of an Epidemic Information Distribution Scheme in Mobile Ad-hoc Networks

Sebastian VanSyckel
University of Mannheim
Schloss, 68161 Mannheim, Germany
sebastian.vansyckel@uni-mannheim.de

Dominik Schäfer
University of Mannheim
Schloss, 68161 Mannheim, Germany
dominik.schaefer@students.uni-mannheim.de

Gregor Schiele
University of Mannheim
Schloss, 68161 Mannheim, Germany
gregor.schiele@uni-mannheim.de

Christian Becker
University of Mannheim
Schloss, 68161 Mannheim, Germany
christian.becker@uni-mannheim.de

ABSTRACT

Information dissemination in a delay-tolerant network depends on the underlying network and its characteristics, e.g. topology and density. The node mobility determines to a large degree the connectivity in the network and thus the information distribution rate. Specific routing protocols have been developed to overcome connectivity issues. However, these protocols were mainly evaluated using analytical models and random waypoint simulations. In this paper we use the real movement data collected by Nokia to evaluate a simple epidemic information distribution scheme and compare different variations of that scheme in the context of low density.

Categories and Subject Descriptors

C.2.2 [**Computer-Communication Networks**]: Network Protocols—*Routing Protocols*

General Terms

Performance, Algorithms

Keywords

Mobile Computing, Mobile Ad-hoc Networks, Model-based Routing, Nokia, Mobile Data Challenge, Lausanne Data Collection Campaign

1. INTRODUCTION

Delay-tolerant networks (DTNs) received a lot of attention from researchers in the past (see e.g. [5, 4]). They provide means to communicate without requiring an infrastructure by using ad-hoc communication between nodes. Applications of DTNs are disaster areas, local information dissemination where infrastructures may be overloaded, e.g. rock concerts, or applications that try to run cost-neutral, e.g. urban sensing, by reducing mobile telecommunication costs for users. Further, DTNs allow to reach users that do not have a mobile-broadband subscription and broadcast public service announcements to most of the population without further knowledge.

In DTNs, traditional routing protocols that first determine a route before transmitting the actual message can not succeed. Hence, specific routing protocols implementing a *store and forward* approach have been developed [7]. Using such protocols, nodes can overcome network partitions, or any other source of connectivity loss, and resend messages if they come into contact with new nodes. However, the evaluation of such protocols is mainly done either by analytical models, e.g. [8], or by simulation, e.g. [1, 13, 7], while, further, assuming a sufficient density of nodes.

In this paper we evaluate a simple epidemic model (based on [8]) with real world traces from Nokia's Lausanne Data Collection Campaign (LDCC). We specifically focus on low density situations that are known to be problematic for DTNs, and analyze different protocol extensions to overcome low density problems.

The remainder of the paper is structured as follows. We discuss relevant aspects of the Mobile Data Challenge (MDC) data set in Section 2. In Section 3, we present the system model and the DTN protocol, before we evaluate the different protocol variants in Section 4. Finally, we close the paper with a conclusion and an outlook on future work in Section 6.

2. MDC DATA SET

The raw data set of the Mobile Data Challenge is described in [9]. The data set is very comprehensive per device, with, for instance, an average of more than 12,000 Bluetooth- and 22,000 WLAN reads, respectively, per device per month. However, the number of devices is small with regard to the population (127,821) and area (41.37 km^2) of Lausanne [12], and was further split up for the different tracks of the challenge. The mapping of user id to Bluetooth MAC address is essential for our work, as we need to identify the device that receives a message. This information, however, was missing for seven devices.

In total, we use the records of 31 devices for the evaluation, which equals one device per 4123 people and 1.33 km^2, or 0.75 devices per km^2. At first sight, these numbers suggest that we evaluate information distribution in a very sparsely populated region. This comparison holds true for distribution via Bluetooth, where messages are exchanged directly between two devices participating in the same piconet. However, the amount of WLAN access points, which allow indirect message exchange, resembles a well developed infrastructure. Even though the density of participating nodes is low, the available infrastructure can partially mitigate resulting problems (see Section 4).

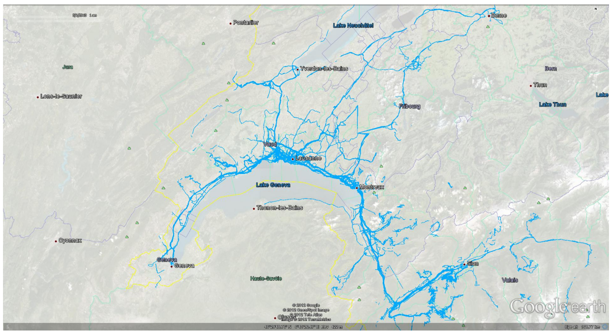

Figure 1: GPS tracks from January 1^{st} through April 30^{th}, 2010.[1]

Naturally, the devices are also not only active in Lausanne, but rather all across Switzerland. Figure 1 shows the GPS tracks that were recorded by all 38 devices of the challenge's Open Track data set in the time from January 1^{st} to April 30^{th}, 2010. This has positive effects on the reach of information, as messages are not spatially limited to the regions in which they are initially created. However, we must also consider these movement patterns as a delaying factor when we discuss the evaluation results.

We discuss the specific implications of the data set in more detail in their respective context during the evaluation in Section 4.

3. SYSTEM MODEL

In this section, we describe the system model and diffusion algorithm applied to the movement traces during the evaluation, before we present the results in the following section.

[1]Created with Google Earth.

At first, we assume the following *simple diffusion algorithm*. Messages are exchanged whenever device A discovers device B via Bluetooth, and vice versa, or devices A and B record the same WLAN access point at most 60 seconds apart – the frequency in which the devices scan for access points. We assume all nodes to be interested in every message. If a device has already received a message, it simply discards it. For both transmission options, we assume that the two devices share enough time in the MANET to exchange all messages.

However, as mentioned above, the density of nodes is very low. As a result, the devices seldomly form MANETs. To overcome this, we introduce a buffer mechanism in two additional settings, namely the *local buffer* mode as an extension of the simple algorithm, and the *global buffer* mode as an optimal benchmark. That is, we buffer the messages either locally at the access points – thus virtually expanding the MANETs in the time dimension – or in a global buffer, i.e. a centralized service in the cloud, which results in each device receiving each message as soon as it connects to any access point.

Within each of the three settings, we further distinguish between an *open* and a *closed* mode. The former uses only not secured WLAN access points, whereas the latter includes all access points. Finally, we generate several messages per day at random times and randomly assign them to a device.

4. EVALUATION RESULTS

In this section, we present and discuss the evaluation results of information dissemination using real traces. First, we examine the distribution of messages using the *simple diffusion algorithm*, reason about the critical factors, and discuss possible improvements. Afterwards, we examine the information distribution using *local buffering* of messages at access points and evaluate this approach against the benchmark *global buffering*.

Figure 2 shows the average distribution time of messages in regard to the percentage of receivers, as well as the respective minimal and maximal values, for the *simple diffusion algorithm* in MANETs of direct Bluetooth connections and shared connections with (1) all *open*, i.e. not secured access points, and (2) all access points, respectively.

Consistent with previous research based on analytical models and simulations, the information distribution rate increases in a logarithmic fashion. However, as assumed for low-density DTNs, the performance is not satisfying. Using only *open* access points (1), a message distribution of 50% takes, on average, 20 days, with a minimal of 11, a maximal of 29, and a standard deviation of 3.85. The performance of this setting drastically deteriorates further shortly after 70% of the receivers are reached, indicating that a share of the receivers are not very active participants in the network around Lausanne. Using all access points (2), the information distribution progress is very similar. The *closed* setting finally outperforms the *open* setting after about 70%. However, up to that point, both curves show the same characteristics, e.g. turning points and gradients. This suggests, that (i) the number of access points – approximately one fifth of all access points are open – is not as important as one

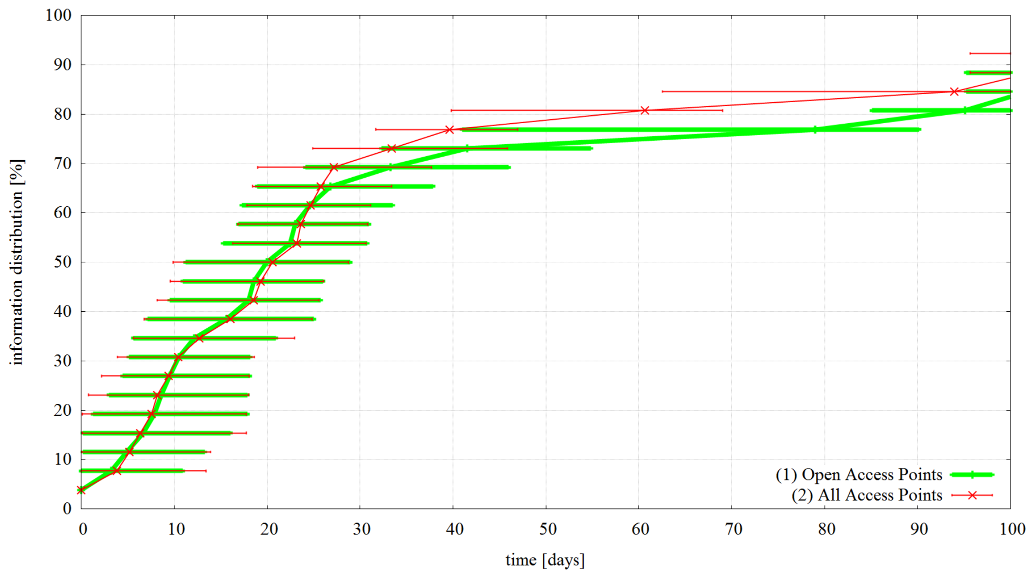

Figure 2: Information distribution using Bluetooth connections and shared access point connections.

naively assumes, and (ii) the randomized message creation time and initial delivery are negligible. We found, however, that the reason behind this strong similarity is mainly due to the consistency in everyday movement patterns, which manifests itself in the sequence in which messages traverse the network of nodes. In numbers, 12.16% of messages follow the same sequence of 24 nodes, 22.97% the same of 16 nodes, 45.95% the same of 13 nodes, 50% the same of seven nodes, and 63.51% the same of four nodes. As soon as a message reaches the starting point of one of these sequences, e.g. two students meeting in class every day, it most likely traverses through the rest of the network together with other recent messages. As a result, both the number of access points as well as the randomness become less important.

Nevertheless, the simulation shows that the *simple diffusion algorithm* is not applicable for DTNs with low density. This is due to the fact that the users are rarely in close enough proximity to each other in order to form MANETs. Consequently, we should be able to improve the performance, if we are able to expand either the size or the duration of the networks. Larger networks exist in Lausanne, e.g. at the university, but most access points likely operate independently, e.g. in downtown cafes. Hence, as described in Section 3, we propose to expand the time dimension by adding a *local message buffer* to the access points of the infrastructure. That is, the individual access points maintain and further propagate the set of messages that have been transferred to them by devices crossing through their range. As a benchmark, we additionally simulate information distribution using a *global buffer*, i.e. a buffer in the cloud, with the effect that all access points share the same set of messages.

Figure 3 depicts the average distribution times with (1,2) *local buffering* and (3,4) *global buffering* – the benchmark – of messages, using (1,3) only *open* access points and (2,4) all access points, respectively. All four distribution rate curves show the characteristic logarithmic growth.

As expected, the *global buffering* mode outperforms the *local buffering* mode. Still, the improvement over the *simple algorithm* is significant. Comparing the results of *local buffering* to those of the simple approach, the messages are distributed roughly four times as fast, e.g. reaching 50% in five instead of 20 days, and 70% in nine instead of 34 days. For several scenarios, e.g. public service announcements, these rates are acceptable, despite the very low density. Further, within both buffering approaches, the difference between the *open* and *closed* modes is more significant than with the simple approach, i.e. the performance improves with the number of access points. Hence, with buffering, independent movement patterns become more important again. For example, students that do not sit in the same class can still exchange messages if their ways around campus, temporally independent, partially overlay. Finally, all four rates experience a significant drop-off in growth past a certain level, most notably in (4) past the 80% mark. This, again, shows that the remaining 20% of devices are seldomly active.

5. RELATED WORK

The epidemic routing approach we used for our evaluation is flooding-based, i.e. it distributes messages throughout the network using replication and transmission to new contacts. In practice, such approaches are not very efficient [6]. Hence, several protocols have been developed that aim at reducing the network load by deciding whether forwarding a given message to a given node is beneficial. Examples are model-based routing [2], PRoPHET [10], and SimBet [3]. However, they are usually evaluated using simulation instead of real traces. In [11], a number of protocols have been evaluated using contact traces derived from real network logs, but low-density DTNs were not considered.

6. CONCLUSION AND FUTURE WORK

In this paper, we evaluated a simple information distribution scheme based on epidemic routing using real movement traces. We found that in low-density DTNs, the consistency

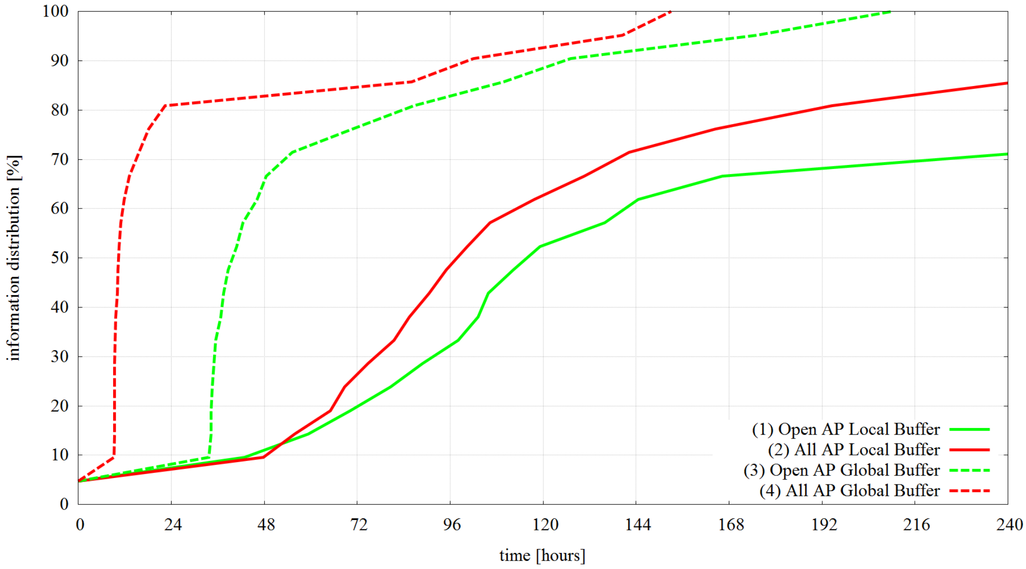

Figure 3: Information distribution with local buffering at access points vs. global buffering in the cloud.

in everyday user movement has a higher impact on the information distribution rate than the number of access points in the infrastructure. Further, we showed that message buffering at access points significantly improves the performance in low-density DTNs, such that epidemic routing becomes applicable again for non time-critical applications.

In future work, we would like to evaluate the same information dissemination scheme using the traces of all participants of the LDCC, increasing the density of the DTN to approximately 4.8 devices per km^2. In that, we would further examine the significance of consistent movement patterns and research the implications of these patterns for routing approaches based on world models and history of encounters.

7. REFERENCES

[1] C. Becker, M. Bauer, and J. Hähner. Usenet on the fly - supporting the locality of information in spontaneous networking environments. In *Proc. of the Workshop on Ad hoc Communications and Collaboration in Ubiquitous Computing Environments*, 2002.

[2] C. Becker and G. Schiele. New mechanisms for routing in ad hoc networks. In *Proc. of the Fourth Plenary Cabernet Workshop*, 2001.

[3] E. M. Daly and M. Haahr. Social network analysis for routing in disconnected delay-tolerant manets. In *Proc. of the 8th ACM International Symposium on Mobile Ad-hoc Networking and Computing (MobiHoc '07)*, 2007.

[4] K. Fall. A delay-tolerant network architecture for challenged internets. In *Proc. of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '03)*, 2003.

[5] S. Farrell, V. Cahill, D. Geraghty, I. Humphreys, and P. McDonald. When tcp breaks: Delay- and disruption-tolerant networking. *Internet Computing, IEEE*, 10(4):72–78, 2006.

[6] K. Harras, K. Almeroth, and E. Belding-Royer. Delay tolerant mobile networks (dtmns): Controlled flooding in sparse mobile networks. In *NETWORKING 2005. Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications Systems*, pages 41–91. Springer, 2005.

[7] S. Jain, K. Fall, and R. Patra. Routing in a delay tolerant network. In *Proc. of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '04)*, 2004.

[8] A. Khelil, C. Becker, J. Tian, and K. Rothermel. An epidemic model for information diffusion in manets. In *Proc. of the 5th ACM International Workshop on Modeling Analysis and Simulation of Wireless and Mobile Systems (MSWiM '02)*, 2002.

[9] J. K. Laurila, D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, T.-M.-T. Do, O. Dousse, J. Eberle, and M. Miettinen. The mobile data challenge: Big data for mobile computing research. In *Proc. of the Workshop Mobile Data Challenge by Nokia, in conjunction with The Tenth International Conference on Pervasive Computing (PERVASIVE 2012)*, 2012.

[10] A. Lindgren, A. Doria, and O. Schelén. Probabilistic routing in intermittently connected networks. In *Service Assurance with Partial and Intermittent Resources*, pages 239–254. Springer, 2004.

[11] L. Song and D. F. Kotz. Evaluating opportunistic routing protocols with large realistic contact traces. In *Proc. Second ACM Workshop on Challenged Networks (CHANTS '07)*, 2007.

[12] Swiss Federal Statistical Office. Population and households statistics (statpop), 2010.

[13] J. Tian, J. Hähner, C. Becker, I. Stepanov, and K. Rothermel. Graph-based mobility model for mobile ad hoc network simulation. In *Proc. of the ACM/IEEE 35th Annual Simulation Symposium (ANSS '02)*, 2002.