The impact of location privacy on opportunistic networks

Iain Parris and Tristan Henderson
School of Computer Science
University of St Andrews
St Andrews, Fife, KY16 9SX, UK
{isp3,tnhh}@st-andrews.ac.uk

Abstract

Opportunistic networking involves forwarding messages between proximate users, who may or may not know one another. This assumes that users are willing to forward messages to each other. This assumption may not hold if users are concerned about using the opportunistic network service. One such concern may be due to privacy; for instance, users’ locations may be leaked.

A privacy-concerned user may therefore disable their mobile device’s opportunistic-networking features at various times, to preserve their privacy. This paper studies the impact of location privacy concerns on the performance of an opportunistic network. Using data from a real-world location-aware user study to develop a privacy model, we conduct trace-based simulations of various opportunistic routing protocols with two real-world traces. We find that users’ location privacy preferences may potentially reduce the delivery performance of an opportunistic network to zero.

I. INTRODUCTION

An opportunistic network leverages the various wireless devices that humans now carry, in conjunction with human contact patterns, to enable a new communication paradigm. As people come into contact with each other, their devices can communicate wirelessly to send and receive data [22].

Although the opportunistic network may facilitate communication, those users who participate may have privacy concerns — for example, regarding the confidentiality of their transmitted messages, or the potential leakage of their private information [21].

In this paper, our focus is on location privacy. A user who is concerned about privacy may choose to disable their device’s opportunistic networking features at various times, thus preserving privacy by rendering the user invisible to the network, depending on whether they feel that participating in the opportunistic network is acceptable. How might doing so affect the performance of such an opportunistic network? Our goal is to explore this question.

In this paper, we:

• present a dataset-independent empirical model of users’ location privacy preferences, based on a real-world user study of 80 participants.
• demonstrate methods for applying this privacy model, in different modes, to datasets containing human mobility patterns.
• evaluate the impact on opportunistic network routing performance on application of the privacy model.

Our contributions are to provide an empirical model of users’ location privacy preferences, and, to our knowledge, the first application of such a model to opportunistic network routing.

This paper is structured as follows. Next, we discuss related work, and introduce our empirically-determined model of location privacy preferences in Section III. We discuss the application of this privacy model to opportunistic routing protocols in Section IV. Section V evaluates the performance impact using two real-world traces, and finally in Section VI we conclude and discuss ongoing work.
II. RELATED WORK

Our work is at the overlap of two research areas: privacy in opportunistic networking, and location privacy in general. Both of these areas have been the focus of much research in recent years.

A. Privacy threats in an opportunistic network

For a participant in an opportunistic network, there are many possible privacy threats. For example:

- Disclosure of message content, enabling it to be read by unintended parties. This threat may be mitigated with encryption, if the sender and destination are able to agree on encryption keys [23].
- Messages may be traced as they progress through the opportunistic network, to infer communication patterns [18].
- Social network information may be used to inform routing decisions [14], [8], [4]. This social network information may be leaked via the routing scheme [21].
- Locations of the participants may be inferred from the messages which their mobile devices carry — whether in absolute terms (“Alice is at the supermarket”), or relative terms (“Alice and Bob were in the same location this afternoon”) [21].

Some of these privacy concerns may be mitigated by technical measures, such as encryption. Even though a public key infrastructure may not be well-suited to opportunistic communications [10], decentralised mechanisms such as identity-based cryptography [23] — where the identity of each node acts as a key — may allow cryptographic solutions to some of the privacy risks. For example, if nodes may communicate securely, then reading message content, or tracing message progression through the network, becomes much more challenging.

But even identity-based cryptography requires a global trusted third party to vouch for new nodes entering the network (by generating the necessary private keys). This may be infeasible in certain types of opportunistic network, since access to the wider Internet may be impossible before encountering a new node. In such scenarios, it may be possible to employ simpler decentralised techniques, such as obfuscating information at the routing protocol or application layer [21], or utilising trusted social contacts [10].

B. Location privacy

Location privacy has been studied in various contexts, such as sensor networks [15]; pervasive computing applications [3], [27]; and indeed opportunistic networks [19]. The focus, however, has been on investigating the trade-offs involved in the protocols which work to preserve privacy, rather than considering how users’ own privacy-preserving behaviours may affect network performance, as is our focus here.

In the context of publishing sensed location information, various obfuscation techniques have been proposed [1], [17], [28]. User studies have also been performed to determine how users respond to various types of obfuscation [5]. But the focus has been on anonymity, rather than performance of a distributed system.

Finally, we note that existing location-sharing applications have received recent publicity regarding inherent privacy threats. For example, the website “Please Rob Me”\(^1\) gathered publicly-available information from Twitter and FourSquare in order to infer whether or not a person was at their home address [12] — thus raising awareness of a privacy threat of which many users were presumably unaware.

III. EMPIRICAL MODEL OF PRIVACY

To investigate whether privacy concerns may have an impact on opportunistic network routing performance, it is necessary to have a model of such privacy concerns.

\(^1\)http://pleaserobme.com/
TABLE I
LOCATION-SHARING BEHAVIOUR ON FACEBOOK, BY PARTICIPANT CATEGORY. 80 PARTICIPANTS CARRIED A LOCATION-SENSING MOBILE PHONE FOR ONE WEEK AND WERE ASKED WHETHER THEY WOULD SHARE THEIR LOCATION AT VARIOUS TIMES AND PLACES. ROWS MAY NOT ADD UP TO PRECISELY 100% DUE TO ROUNDING.

<table>
<thead>
<tr>
<th>Category</th>
<th>Proportion in category</th>
<th>Location sharing choice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Nobody</td>
</tr>
<tr>
<td>Open</td>
<td>19% (15/80)</td>
<td>7.8%</td>
</tr>
<tr>
<td>Social</td>
<td>49% (39/80)</td>
<td>7.6%</td>
</tr>
<tr>
<td>Closed</td>
<td>23% (18/80)</td>
<td>70.1%</td>
</tr>
<tr>
<td>Variable</td>
<td>10% (8/80)</td>
<td>32.7%</td>
</tr>
</tbody>
</table>

A. User study

Developing a model requires data on users’ privacy behaviour in opportunistic networks, but collecting such data is not straightforward. To collect high-quality data, it may be required to build, deploy, and measure user behaviour in a real, large-scale opportunistic network. But this may be time-consuming and impractical — and moreover, privacy behaviour in such an experimental network may not reflect actual behaviour, since users may be unfamiliar with these new technologies and so act in different ways [20]. Thus, to develop our model, we instead measured privacy behaviour by performing a smaller-scale user study which investigated the location-sharing privacy preferences of 80 users of the popular online social network Facebook.²

The purpose of this study was to determine how widely participants would accept their current locations being broadcast to their friends online. Would a participant be happy to share some locations to the whole world, while others to select friends or to nobody at all? Are some participants more inclined to share their locations than other participants? Can we quantify location-sharing behaviours?

Participants in the experiment carried a location-sensing mobile phone for one week of their day-to-day lives. Due to resource constraints — we had 20 mobile phones available, but 80 participants — we conducted the experiment in four one-week runs, each with 20 participants. Two runs were conducted in a small UK town, St Andrews; the other two runs were in a large UK city, London. Participants were undergraduate students, who were not studying in the Computer Science department (so that they would not be known by us), and who claimed to use Facebook daily. Further experimental details can be found in [2].³

Each participant was prompted up to 20 times per day to choose how widely their current location could be published on Facebook — to everyone, to some or all of their Facebook social contacts (“friends”), or to nobody at all.

By analogy to existing location-sharing applications, and the publicity surrounding information leakage (see Section II-B), we believe that privacy choices for location-sharing behaviour when broadcasting locations via Facebook will not be dissimilar to those for an opportunistic network participant. One privacy risk associated with participation in an opportunistic network is the loss of location privacy. Should privacy threats due to information leakage in a real opportunistic network deployment receive similar publicity to similar threats in extant systems, then we believe that these location-sharing choices will converge to those for location broadcast, as we have measured in this user study.

B. Analysis

As is common in other privacy models [25], we segment the participants into categories according to their privacy behaviour, i.e., their responses to the prompted questions (see Table I). We define four categories:

²http://www.facebook.com/
³This paper describes two of the four runs — the two St Andrews runs — since it was published part-way through running the experiment.
• *Open*: Participants usually shared their location publicly with everyone, in over 50% of responses.
• *Social*: Participants usually shared their location with some or all of their Facebook friends, in over 50% of responses.
• *Closed*: Participants usually did not share their location to anybody at all, in over 50% of responses.
• *Variable*: Participants did not have consistent location-sharing behaviour. They would sometimes share with nobody, with friends, and with everyone.

For each of these four categories, we take the mean of the users’ location sharing choice proportions (*nobody*, *friends* or *everyone*) by user, in order to obtain Table I.

By simulating users and their sharing choices according to these statistics (Table I), we create a privacy model for users’ location sharing preferences. This privacy model is dataset-independent, and so may be applied to a variety of datasets for opportunistic network routing simulations.

**IV. ROUTING PROTOCOLS**

We now describe how our privacy model can be applied to opportunistic network routing simulations.

**A. Node categorisation**

At the start of each simulation run, each node (i.e., simulated participant) is randomly assigned to one of the categories (*open*, *social*, *closed*, *variable*), according to the proportional size of the category, for the duration of the run.

**B. Routing protocols**

We investigate two routing protocols, each with three modes of behaviour: a non-privacy-aware mode for ground truth, and two privacy-aware modes. The default (non-privacy-aware) versions of the protocols are:

• *Epidemic routing (Epid)*: Messages are flooded through the network, with copies forwarded during every encounter [24].
• *Simple social network routing (SNR)*: Each message is forwarded between members of the original sender’s social graph neighbours (*friends*). So each message contains a copy of the original sender’s friends, and is forwarded during encounters between these friends of the original sender.

**C. Privacy modes**

For each of the two routing protocols, we simulate three modes of privacy behaviour. While it is possible to think of many more behaviours, we believe that three modes is sufficient for investigating the impact of privacy, and moreover in previous work we have demonstrated that this constrained number of privacy choices is a usable compromise for privacy policies for ubiquitous computing environments [16]. Our chosen three modes are:

• *Default (D)*: Privacy preferences are ignored. We simulate this behaviour for ground truth.
• *Friendly (F)*: Nodes are modelled as being willing to share with their social network friends. If the overall privacy choice is *everyone* then the nodes behave as in the default case; if *nobody* then messages are not exchanged; if *friends* then as the default case only if the two nodes involved in this encounter are friends (otherwise messages are not exchanged).
• *PubPriv (PP)*: Nodes are modelled as either being fully public (no privacy concerns), or fully private (any privacy concerns result in disregarding the encounter) — with nothing in-between. If the overall privacy behaviour during an encounter is *everyone*, then messages are exchanged as in the default case. Otherwise (i.e., if the overall privacy behaviour is *friends* or *nobody*), messages are not exchanged.

During each encounter between a pair of nodes, each of the two nodes randomly picks a privacy behaviour of {*nobody*, *friends*, *everyone*}, weighted according to the location sharing proportions.
associated with that node’s category. Messages are then exchanged depending on the chosen privacy behaviours for that encounter. The overriding choice is the more restrictive of the two nodes’ privacy behaviours. For example, if one node picks nobody and the other picks everyone, then the overall choice is the more restrictive nobody.

V. EVALUATION AND RESULTS

We now evaluate the routing protocols, to determine the performance impact of the three modes of privacy behaviour.

A. Datasets

We perform trace-driven simulation using a custom Python opportunistic-network simulator, for two real-world datasets containing encounters and social networks.

- **LocShare**: A dataset derived from the location-sharing privacy user study described above. In the user study, participants interacted with a custom-built Facebook application called LocShare (named for location share). We derive encounters based on proximity (within 10m) of the participants in the study, inferred using the locations sensed by the mobile phones which they carried. The social network information is Facebook friendships.

- **Reality Mining**: The well-known Reality Mining dataset collected at MIT [9]. 97 university members (students and staff) carried mobile phones during their daily lives for an academic year. These phones recorded the results of periodic Bluetooth scans. We define Bluetooth encounters between participant devices in our simulations as opportunities for message exchange in an opportunistic network. As in [21], we extract social network information from the mobile phone address books.

We thus perform trace-driven simulation using the encounter traces — i.e., the times at which pairs of nodes encountered one another — obtained from each dataset.

B. Simulation parameters

We use the following simulation parameters:

- 100 runs per data point.
- 100 messages per run.
- Unicast messages, from the sender to one of the sender’s social network neighbours (friends). Note that although messages are unicast (destined for one particular recipient), the message may follow multiple paths through the network in order to reach that recipient.
- TTL of one day.
- One week per simulation.
- Infinite buffers; infinitely-fast transmission.

Following [21], messages which arrive in zero-time (i.e., from a direct link between the original sender and final recipient) are excluded from the analysis because, when sender and recipient are in proximity, a file-transfer application would be able to exploit more efficient forms of communication than an opportunistic network. By excluding these transfers, we are able to focus on the performance of the network in non-trivial opportunistic scenarios.

C. Results

Figures 1–4 show the performance of each of the routing protocols, as measured by two metrics [14]:

4For LocShare, there are four one-week parts to the dataset; we therefore simulate 25 runs with each of the four one-week parts to make up the 100 runs for each data point. For Reality Mining, we pick a random one-week interval for each of the 100 runs — but, following [21], we select only one-week intervals where there are sufficient numbers of nodes present for non-trivial routing to be possible.

5We are investigating the performance impact of privacy preferences, so we do not wish to set arbitrary constraints on buffer size or transmission rate, since these may confound the results.
• Delivery ratio: proportion of delivered messages, out of the total number of unique messages generated.

• Delivery delay: time taken for a message to first reach its destination.

Figure 1 shows that routing performance is significantly reduced for the LocShare dataset when taking into account privacy concerns in the Friendly and PubPriv privacy modes, as compared to the baseline Default mode. For epidemic routing, the median delivery ratio\(^6\) falls from 34% to 5% when using the Friendly privacy mode; the situation is compounded if we assume that users are even more private (PubPriv mode), where delivery falls further to 1%. SNR shows a similar trend: a fall in median delivery ratio from 13% (Default) to 2% (Friendly), or to zero (PubPriv). Delivery delay is, however, not significantly affected for those messages which arrive, as shown in Figure 2: there is wide variation in the delay, and each mode’s boxes overlap.

Figure 3 shows that a similar trend holds for the Reality Mining dataset. For epidemic routing, the median delivery ratio falls from 28% (Default) to 8% (Friendly), or to 3% (PubPriv). For SNR, the fall is from 13% (Default) to 6% (Friendly), or again to zero (PubPriv). Like for the LocShare dataset, Figure 4 shows that the delivery delay is not significantly affected for those messages which arrive.

\(^6\)Note that the low delivery ratios, in absolute terms, are typical of opportunistic routing simulations: the datasets used, while collected from real-world users, are from experimental settings that are not as highly-connected or dense as we would expect in a real large-scale deployment.
D. Discussion

Our results suggest that users’ privacy concerns may lead to dramatically lower routing performance for opportunistic networks. What are the implications for the designers of future systems?

If opportunistic network applications or routing protocols leak information unnecessarily, then users may become less willing to participate in the network. For example, perhaps the users would act closer to the PubPriv mode, with the associated very low delivery performance, rather than the Friendly or Default modes. It would be paramount for protocol designers to minimise the amount of unwanted private information leaked, in order to allay the privacy concerns of most users, and thus indirectly improve the performance of the network.

We also note that this may be even more pressing a concern than these results show. The privacy model described in Section III was derived from a user study involving heavy Facebook users. Perhaps such users are less privacy-concerned than the average opportunistic network user would be. Alternatively, the opposite may be true: perhaps such users are more privacy-concerned than the average person — say from having gained direct experience of privacy breaches through their Facebook experiences. Or external factors, such as changing privacy social norms or incorporating suitable incentives to encourage participation in the opportunistic network, may offset the privacy concerns and thus make the issue less pressing. We highlight this as an area for future research.
VI. CONCLUSIONS AND FUTURE WORK

In this paper, we have introduced an empirically-determined dataset-independent model for users’ location privacy concerns. We demonstrated through simulation with two real-world datasets that these privacy concerns may significantly impact opportunistic network routing performance, as measured by the metric of message delivery ratio — potentially reducing the delivery ratio to zero. Message delivery delay is, however, not significantly affected for those messages which do arrive. Our results raise a number of open questions for future work.

We plan to investigate more sophisticated privacy models. For instance, is there a correlation between privacy preferences and the location of an encounter? We have also assumed (Section III-A) that the privacy behaviour of heavy Facebook users corresponds to that of (potentially-pseudonymous) opportunistic network users. Future work needs to be conducted to test whether this assumption holds.

Future work might also investigate more complex scenarios — for example, the performance of a privacy-preserving protocol, which may alleviate users’ location privacy concerns (and hence reduce the impact on performance from user behaviour) by preserving privacy, but potentially at a performance cost from the protocol itself.

A potential limitation of our privacy model is that we asked users for their preferences to disclose their exact location — to the accuracy of a GPS sensor. Should location information be leaked via participation in opportunistic networks, then perhaps only coarser locations may be discoverable by other
users. Since coarse locations introduce a degree of obfuscation, thus implicitly providing users with increased privacy [5], we plan to investigate how user preferences may vary depending on the granularity of their shared location, within the context of an opportunistic networking application.

If users’ location privacy concerns can have such an impact on opportunistic network performance, might other privacy concerns (for example, about some malicious node gathering and broadcasting information, or the possibility of the revelation of users’ friends lists [21]) also have a performance impact? To find out, we need reliable data on other types of privacy concerns in these networks.

More data on encounters would also be useful: existing datasets used to evaluate opportunistic network protocols are relatively small-scale compared to the population of a town or city; perhaps evaluation against a larger-scale (and hence perhaps more-highly-connected) dataset would yield new insights. We are searching for such datasets.

Finally, we also note that some individuals may wish to heavily participate in the network — either altruistically [26], [11], or due to another incentive, such as reputation, payment or barter [6], [13], [7]. Perhaps this desire to participate may outweigh their usual location privacy concerns? Future work may investigate this trade-off.

REFERENCES