

Short paper: Random IDs for preserving location privacy

Stefan Schlott, Frank Kargl, Michael Weber – University of Ulm, Dept. of Media Informatics
{stefan.schlott|frank.kargl|michael.weber}@uni-ulm.de

Abstract—The privacy aspect is often neglected in many electronic systems. With the emergence of ubiquitous systems, privacy will become an even more important aspect than it has been in the past. Location privacy – the information where exactly a node is currently located – is one aspect.

This paper classifies several kinds of attacks on location privacy. It examines the results of these attacks on a scheme utilising changing identities to preserve the users' identities. Simulation results show the influence of the number of nodes and the backoff time between transmissions on the nodes' location privacy.

I. INTRODUCTION

Ubiquitous computing proposes a world filled with small, specialised computing devices. These devices provide their services to the users in an unintrusive way; ideally, the users do not even recognise them as computers [1]. They will perform regular tasks, extend the reach of the familiar computing to physical spaces, and allow interaction with the users' environment.

Beside many other challenges (e.g. low energy consumption, miniaturisation, robust networking, wireless communication), privacy is an especially important aspect. Even in one of the earliest publications on ubiquitous computing [2], Mark Weiser already recognised its importance.

Concealing one's identity forms the basis of location privacy: Otherwise, all network transmissions can be attributed to the sender, revealing his current position.

This paper focuses on location privacy and its threats. It regards the lower transmission layers; protecting ones privacy on higher layers (e.g. application layer) has to be treated in a different way. Several kinds of possible attacks are modelled; the attack most widely applicable is simulated and the effects are examined.

II. RELATED WORK

Despite the fact that there are many projects in the field of ubiquitous computing, little work has been done regarding privacy. Mark Weiser mentioned the privacy aspect in his early works on ubiquitous computing; e.g. in [3], "privacy of location" is briefly discussed.

The aspect of location privacy has received some focus. Görlach et. al. [4] classify several kinds of possible information sources and give an overview on research done in this area.

Several approaches try to employ mix networks, as described for electronic mail by Chaum et. al. in [5]. Gaia OS [6] serves as an example for this strategy: Employing the Mist

routing protocol [7], all messages are encrypted and passed along several mist routers. A so-called Lighthouse poses as the virtual communication endpoint for registered clients; traffic from and to the Lighthouse is sent through an encrypted tunnel along the mist routers.

These approaches require some kind of infrastructure - dedicated relays or cooperating nodes. In [8], Beresford et. al. examine the simple scheme of changing pseudonyms (without any supporting infrastructure). They introduced the concept of mix zones, i. e. areas without monitoring. The paper examined possible correlations of nodes entering and leaving a mix zone. They discovered that the location privacy can be surprisingly low even with relatively large mix zones.

III. SCENARIO

We assume a typical scenario where nodes are moving around freely and access certain services on nearby hosts. When no precautions are taken, service providers can easily estimate the location of accessing clients: If the client is connecting to the service directly (single hop), the service provider knows that the client is within a certain proximity. Based on the access technology, the precision of the location estimation can vary from a few meters (e. g. Bluetooth) up to several dozen meters (e. g. WaveLAN). In many cases, this is an assumption which is common sense in the context of location-based services. Further, restriction to direct communication can be a means of preserving privacy - as described with the keywords "proximity and locality" by Langheinrich et. al. in [9]. In a typical hotspot scenario (multi hop), even a remote service, located somewhere on the Internet, gets a rough idea about the position of the client if the location of the access point is known. In order to conceal their identity, nodes change their IDs (i. e. their network address) when accessing different services. By doing so, it should be very hard to create a location profile by simply relating different service accesses to each other.

Our observations focus on the single hop scenario for two reasons: First, even in a multi-hop scenario, the problems of the single hop scenario apply for the last hop. Second, even in a ubiquitous environment which usually provides some communication infrastructure, single hop communication may be the fallback if the node is within an area which is not covered by the infrastructure.

The observations of [8] were done using a real-world set of data, observed in an office building in the course of two weeks. As an example, the paper used the hallway as mix

zone; they discovered that the average number of persons, and thus the cardinality of the mix, was rather low. Combining the observation with the characteristic traffic patterns of the mix zone, a high prediction rate was possible.

The advantage of simulations is the possibility to easily change its parameters. We tried to observe both the impact of the number of nodes within the observed area and the effect of more sophisticated observers. An observer with less knowledge on the specific case should yield more general results; thus they should be treated as the minimum any attacker can achieve. The only assumption made by the observer in our simulation is the movement speed of the node - an information easily acquired (e. g. pedestrians typically move at approximately 1.0 - 1.2 m/s). More detailed information often pose a “hen-and-egg”-problem: When gathering the desired data, the observer must be able to identify and track the nodes - the exact thing location privacy tries to prevent.

IV. CLASSIFICATION OF TRACKING ATTACKS

Depending on the determination of an attacker, he may employ different techniques.

- Casual listeners are observers with no different techniques than any other common node. Such nodes are equipped with simple radial antennas, thus they cannot determine the direction of an incoming transmission.
- Trackers are able to determine the position of a single node. This can be done by using several directional antennas, triangulating the exact position. Optionally, a tracker may employ an additional radial antenna, giving information on the network traffic within its proximity.
- An all-seeing observer has the bird’s eye view of all transmissions. That means he knows the positions of all nodes currently sending data.

We assume that the communication done by the nodes has some session characteristics. A single request for information and the corresponding response, transmitted using e. g. TCP, would form such a session (though sessions may be longer, of course). During such a session, a node’s address is immutable, otherwise the session would break.

Nodes are assumed to change their identification after every session. That means that they can be tracked for the time of an ongoing session. After changing the address, a casual listener will be unable to distinguish a newly appeared node from the sender of the last observed transmission - but the other two attacker models can. As a consequence, a node wanting to avoid to be tracked has to finish all sessions. Observers are left with the last position of the node. In order to re-identify a node when it starts transmitting again, the observers can exploit the following information:

- Movement speed: By approximating the maximum distance a node can cover within the elapsed time since the end of the previous transmission, an observer can reduce the area where the node is now to be expected.
- Movement patterns: The environment may force the node to move along certain paths. For example, some obstacles

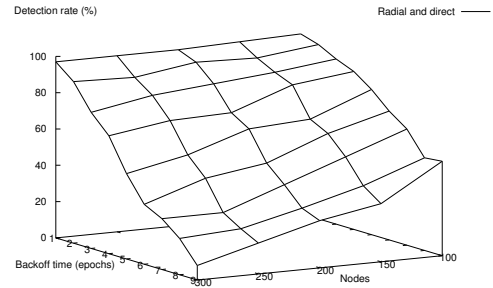


Fig. 1. Random direction, radial and directional antennas

like shelves in a shop may block some possible routes. Even if there are no obstacles, the environment may imply that some movement directions are more likely than other (as exploited in [8] and [10]).

- Traffic patterns: Even though the communication is encrypted, the size, the amount, and the frequency of transmitted packets may reveal some information helping to identify a node.

V. SIMULATION

As mentioned above, our simulations tried to regard the problem in a very general way. In our simulation, the observer only exploited the movement speed of the nodes. The other possibilities of extracting data described in the previous chapter were left aside since they depend heavily on the precise scenario. Two factors were the main subject of our simulations: The dependency of privacy on the number of nodes and on the backoff time.

The simulations were done for a 100x100 meter environment. The nodes moved at 1 m/s. Each parameter set ran for the total time of 10000 transmissions per node. All nodes followed a certain traffic pattern, consisting of a transmission period (during which the node could be tracked) and a backoff time (during which the node was invisible for observers). In the traffic pattern used for the simulations, both transmission time and backoff time were static (but unknown to the observer); in the beginning of the simulation, the startup of the nodes was delayed by a random time. The simulation used clocked time, i. e. at each clock tick, the nodes advanced according to the node speed, and the attacker made his observations.

As described in the previous chapter, a casual observer with no further exploitable information has no chance to distinguish the nodes. A tracker may follow the movements of a node until it finishes its transmission. After that, the area of potential locations of the node becomes a circle around the last known position with an increasing radius. This circular area forms a temporary mix zone for the node; if a new transmission starts within this area, the observer can only guess whether this is the tracked node. In our simulation, where we assumed that the observer has no further information, we observed the quality of the assumption that the next transmission detected within the search area originates from the tracked node.

In the simulation, the observer makes his guess and is afterwards told which node would have been correct. In reality,

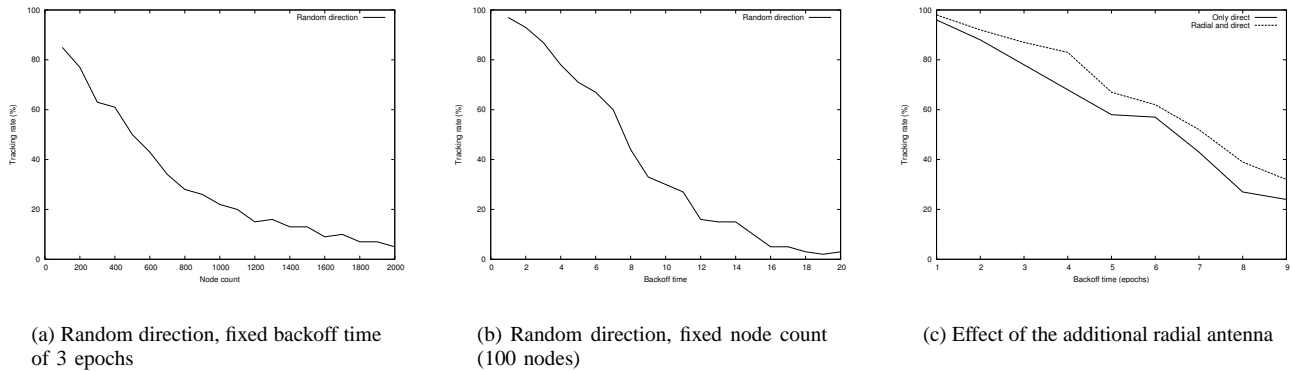


Fig. 2. Simulation results

an observer would not get that information, leaving him with chained probabilities when trying to track a node over several consecutive backoff periods.

This observation may be enhanced by the tracker by using an additional radial antenna so he can listen to network traffic outside the temporary mix zone. By observing that traffic, he can distinguish between new transmissions inside the mix zone and nodes with ongoing transmissions which accidentally enter the observed area.

An all-seeing observer might further enhance the method by verifying if any other node currently in the backoff phase could reach the observed temporary mix zone. This calculation can be done based on the last known position and on the assumed movement speed. But this enhancement would imply that no new nodes enter the observed area, which is unrealistic in most situations. Due to that reason, we left aside that case for our simulations.

VI. RESULTS

Figure 1 shows the simulation results for a tracking observer with an additional radial antenna. The nodes move according to the random direction model, i. e. every step is made in a direction randomly chosen. If a node would hit the wall, the node is reflected. As expected, both the number of nodes and the backoff time increase the chance that another node enters the temporary mix zone. The backoff time has obviously similar influence on the result as the number of nodes. Two further simulation runs, comparing scenarios with fixed node count and fixed backoff time, confirm this observation (fig. 2(a) and fig. 2(b)).

We also examined the gain for the tracking observer using an additional radial antenna. The simulation results are shown in fig. 2(c). Surprisingly, the benefit of the additional information is far less than it might have been expected.

Different movement models will result in different detection ratios. The detection rate of the random waypoint variants is slightly worse than the random direction simulation due to border-/center effects; in random waypoint simulations, the node density at the borders of the simulation area is lower than in the middle parts. Thus, tracking nodes in the middle of the area observed becomes harder.

VII. SUMMARY AND OUTLOOK

The paper pointed out several kinds of observers and multiple means of gathering data. Simulation runs tried to determine the effectiveness of several of these scenarios. As expected, the simple scheme of changing IDs is far from perfect. But given a decent number of nodes, an adequate amount of location privacy can be achieved by choosing an appropriate backoff time. When regarding tracking over several consecutive backoff periods, a feasible amount of privacy can be achieved.

As future work, it should be elaborated how the movement behaviour can be used for better predictions within the temporal mix zone; some may improve tracking (like in the examples of [8]), others might prove more challenging for an observer. Further, the other possible data sources allow clustering the observed nodes according to their behaviour, thus reducing the anonymity set for the node to be tracked.

REFERENCES

- [1] D. A. Norman, *The Invisible Computer*. Cambridge, Massachusetts; London, England: The MIT Press, 1998.
- [2] M. Weiser, "The Computer for the Twenty-First Century," *Scientific American*, pp. 94–110, September 1991.
- [3] —, "Some computer science issues in ubiquitous computing," *Commun. ACM*, vol. 36, no. 7, pp. 74–84, 1993.
- [4] A. Görlach, A. Heinemann, and W. W. Terpstra, *Privacy, Security and Trust within the Context of Pervasive Computing*, ser. The Kluwer International Series in Engineering and Computer Science. Springer, 2004, vol. 780, ch. Survey on Location Privacy in Pervasive Computing.
- [5] D. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," Thesis (M.S. in Computer Science), University of California, Berkeley, Berkeley, CA, USA, June 1979.
- [6] R. H. Campbell, J. Al-Muhtadi, P. Naldurg, G. Sampemane, and M. D. Mickunas, "Towards security and privacy for pervasive computing," in *ISSS*, ser. Lecture Notes in Computer Science, vol. 2609. Springer, 2002, pp. 1–15.
- [7] J. Al-Muhtadi, R. Campbell, A. Kapadia, M. D. Mickunas, and S. Yi, "Routing through the mist: Privacy preserving communication in ubiquitous computing environments," in *ICDCS '02: Proceedings of the 22nd International Conference on Distributed Computing Systems (ICDCS'02)*. IEEE Computer Society, 2002, p. 74.
- [8] A. R. Beresford and F. Stajano, "Location privacy in pervasive computing," *IEEE Pervasive Computing*, vol. 2, no. 1, pp. 46–55, 2003.
- [9] M. Langheinrich, "Privacy by design - principles of privacy-aware ubiquitous systems," in *UbiComp*, ser. Lecture Notes in Computer Science, vol. 2201. Springer, 2001, pp. 273–291.
- [10] A. Beresford and F. Stajano, "Mix zones: User privacy in location-aware services," in *PerCom Workshops*. IEEE Computer Society, 2004, pp. 127–131.