Privacy in Overlay-based Smart Traffic Systems

Martin Florian, Ingmar Baumgart
Institute of Telematics, Karlsruhe Institute of Technology (KIT)
76131 Karlsruhe, Germany
Email: {florian,baumgart}@kit.edu

Abstract—For smart traffic applications like dynamic route planning, communication between traffic participants is of high importance. Traditional communication architectures for smart traffic are centralized, which leads to major privacy concerns since every service provider gains a global view on the mobility behavior of all participating nodes. Recent publications on decentralized alternatives often claim to remedy privacy issues by getting rid of the centralized entity. In this paper, we test this assumption thoroughly, evaluating the privacy-aspects of overlay-based geocast systems in comparison to centralized approaches. To this means, we define an attacker model and describe two different attacks on privacy. Through simulation we show that without additional protection mechanisms, the difficulty for placing surveillance on individual nodes is low. Based on the results, we discuss possible improvements and alternative communication approaches.

I. INTRODUCTION

The increasing availability of Internet access in vehicles offers a variety of new opportunities to assist road users. Examples for such smart traffic applications, which leverage communication capabilities, are dynamic route planning or the localization and reservation of charging stations for electric vehicles. Current solutions mainly follow a centralized, server-based approach for communication, which has several inherent drawbacks:

- Lack of privacy, as all communication and service provision is handled by the service provider. Even given approaches for increasing the privacy in centralized setups, providers might not be motivated to implement them.
- Questionable scalability: With over 50 million vehicles in Germany alone, centralized solutions might quickly lead to performance bottlenecks.
- Inhibited innovation and service quality, as successful services are unlikely to share their user base with competitors. In this way, service models that depend on a large user base are difficult to deploy and cannot develop their full potential.

Decentralized overlay networks providing geocast services [6], [5], [10] are a recent development with a lot of potential for resolving these drawbacks. Roughly, the idea is the creation of a logical overlay network on top of a cellular communication network based on the Internet Protocol (IP). In this overlay network nodes propagate their location to other participating nodes and use this information for choosing overlay neighbors and forwarding messages. With OverDrive [6], this approach was specifically adapted to smart traffic scenarios, introducing mechanisms for dealing with high degrees of node mobility. The evaluation of OverDrive showed [6] the capability to address the scalability and innovation issues of traditional centralized systems.

However, the important question whether a decentralized approach is superior to a centralized approach in terms of privacy has been neglected by previous work. It is obvious that service providers in centralized approaches maintain a global view on all participating nodes. This places them in a position for easy surveillance and profile building. Overlays, on the other hand, are expected to have an inherent advantage to centralized systems in this respect, as there is no single entity maintaining a global view of all participants. In this paper, we test this hypothesis of inherent superiority thoroughly, evaluating the privacy-aspects of overlay-based geocast systems in comparison to centralized approaches.

Specifically, we make the following contributions:

1) An attacker model for privacy attacks on geocast overlays for smart traffic applications, including specific attacks for establishing a global view on all nodes in the overlay as well as identifying and tracking individual nodes.

2) A simulative evaluation of the level of privacy achievable with OverDrive in combination with a simple pseudonymization scheme.

3) A discussion of the results, drawing conclusions about the privacy aspects of overlay-based smart traffic systems as well as sketching possible approaches for improvements.

II. OVERLAY-BASED GEOCAST

A. Functionality

In the following, we will focus on OverDrive [6] as a typical representative for geocast overlays. The main service provided by OverDrive is the delivery of messages to nodes in a given geographic region. Technically, OverDrive is based around two concepts:

- An overlay neighborhood structure based on a partitioning of geographic space into concentric rings, as well as mechanisms for maintaining this structure.
- A greedy routing mechanism for forwarding messages to nodes in a desired geographic area. Messages are forwarded using connections from the aforementioned overlay neighborhood structure.

An overview of the functioning of OverDrive can be seen in Fig. 1. The figure depicts a possible application for the
geocast service, namely the sending of a geographic query to a road segment lying ahead of the requester. From all of its neighbors, which are chosen based on a partitioning of geographic space into concentric rings, the requester greedily chooses the one neighbor that is closest to the destination region in terms of geographic distance. The request is sent via the cellular network and standard IP to this neighbor, who then forwards it according to the same rule, sending it to the one of its overlay neighbors that is closest to the destination region. Once the message arrives at a node residing in the target area, that node might, depending on the application, decide to answer the query by directly sending a response (via IP) to the requester.

The design of the overlay neighborhood structure is critical in regard to user privacy: For maintaining the neighborhood structure nodes have to communicate their locations. Namely, node locations are communicated on the following occasions:

- Periodically or after a significant change in bearing or speed, each node sends a LocationUpdate to each of its overlay neighbors.
- In response to a FindNeighborsRequest for a specific geographic position, a node returns the set of its neighbors that are closest to that position, together with the positions of those neighbors. This is necessary for discovering and choosing suitable new neighbors.
- If a node wants to add another node to its neighborhood structure, it does so by sending it a NeighborRequest, which includes its own location data.

B. Privacy Concept

Geocast overlays are expected to have inherently better privacy characteristics than centralized systems as there is no entity maintaining a global view on all participants. Despite the lack of a global observer, however, nodes still communicate their location to other entities, namely their overlay neighbors. Thus, a basic concept for protecting node identities is still necessary. In [7], the idea of combining a geocast overlay for smart traffic applications with a basic pseudonymization scheme was proposed. In the following, we will develop this idea further.

1) Revealed Information: Due to the nature of geocast overlays, each node needs to reveal several pieces of information to its overlay neighbors: (1) its IP address for enabling communication, (2) the public portion of an asymmetric cryptographic key-pair for realizing a basic level of security and (3) its approximate location, bearing and movement speed. As an important side note and in contrast to other commonly used classes of overlay protocols, geocast overlays do not use dedicated overlay identifiers but address nodes only using geographic coordinates.

To protect privacy, the linkability of position data to real world identities needs to be avoided. None of the communicated pieces of data can easily be used to perform such an identification. However, if one of the pieces of information remains static and does not change over time, location samples can be linked, leading to an increased chance that a node can be identified. As the position of a node changes naturally in a smart traffic scenario, this is mostly relevant for the IP address and public key of a node.

2) Pseudonyms: In the context of vehicular ad-hoc networks (VANETs), pseudonyms are classically defined as anonymized public keys [4], [9]. Based on the previous analysis, we propose to include the IP address of overlay nodes into the definition of a pseudonym as well, to emphasize the need for simultaneous change. Thus, for a node $X$ at time $t$:

$$\text{pseudonym}(X, t) = \text{ipaddress}(X, t) \cup \text{pubkey}(X, t)$$

3) Pseudonym Change: In [7], the changing of pseudonyms at the beginning of trips was proposed. As shown in Section IV, this might not be sufficient to protect privacy. Thus, the mixing [4], [8] of pseudonyms in the middle of trips should be considered as well. An important aspect of pseudonym change is the change of the IP address, which is highly dependent on the cellular communication service provider that is used. With the deployment of IPv6, the flexible choice of IP addresses for nodes should be feasible in the near future.

4) Dealing with Sybil Attacks: Like any fully decentralized system, geocast overlays are potentially vulnerable to Sybil attacks [3]. While usually a security threat, Sybil attacks are also a privacy threat in the context of geocast overlays. If an adversary controls a large number of nodes, he can collect location updates from all other nodes and thus easily construct a global view on all participants. As a simple countermeasure, an independent certificate authority (CA) can be used that provides signed pseudonyms to nodes. The CA may also be required to sign pseudonyms blindly [2], i.e., without being able to link pseudonyms back to real world entities.

Other approaches include the tying of node instances to tamper-proof boxes containing signed certificates for joining the overlay. In this way an attacker needs a separate piece of hardware for each attacker node. However, this approach requires additional hardware for regular users as well, i.e., existing devices like smartphones cannot be used. Yet another approach for combating Sybil attacks is the integration of cryptographic proof of work schemes like cryptographic puzzles into overlay maintenance operations.

III. ATTACKER MODEL

In the following, we present an attacker model for evaluating the privacy characteristics of smart traffic systems based on geocast overlays. Security-related attacks will not be considered in this paper.
A. Assumptions

Our attacker model is based on following assumptions:

- The attacker is able to control several attacker nodes in the geocast network. However, Sybil attacks are not possible and the maximum number of attacker nodes is limited (see Section II-B).
- Attacker nodes are able to lie about their position. So far, no mechanisms for plausibility checks have been proposed for geocast overlays. Thus, an attacker node might claim any position, bearing and movement speed it desires.
- Attacker nodes may ignore protocol limits on the maximum number of overlay neighbors. More overlay neighbors imply higher communication costs, but the attacker might be prepared to pay such a price for gaining a more complete view on the network.

B. Attacker Goal

The goal of the attacker in our model is to trace the movement of a real-world entity that is part of a geocast overlay. This involves two steps: First, the victim’s identity must be linked to a pseudonym in the geocast overlay. Afterwards, the attacker needs to collect location updates from that node, thus keeping the victim under surveillance.

C. Establishment of a Global View

In [11], Wernke et al. give an excellent overview on privacy attacks relevant to centralized location-based systems (LBS). Since the study is based on centralized LBS, the approaches are based on a global view on all participating nodes. As no entity in a system based on geocast overlays has such a global view, we will not focus on these attacks here. Instead, we will focus on the difficulty for an attacker to establish such a view, i.e., the difficulty for an attacker to establish the same attack preconditions available to a malicious service provider in a centralized system. Given a global view, the standard analysis of these attacks applies.

We propose the following attack to establish a global view in a geocast overlay:

1) The attacker controls multiple nodes that are either regular traffic participants (moving like regular nodes) or stationary nodes scattered in the surveillance area.
2) The attacker nodes attempt to become overlay neighbors with as many regular overlay nodes as they can.
3) The attacker nodes then forward all LocationUpdate messages they receive to the attacker who combines them into a global view.

D. Surveillance of an Individual Target

Instead of building a global view on all overlay participants and then attempting standard attacks on this view, the attacker might also attempt to identify and track an individual victim with only a limited view on the network, by carefully choosing the positions of attacker nodes. In the following, we will assume that the attacker has context knowledge about his victim. Specifically, we assume that he knows the geographic position from which the victim will start its trip. Depending on the scenario, such information can be obtained easily. For example, the home address of the victim might be used.

Given such context knowledge, the attacker faces two challenges: (1) mapping the victim to an overlay node and (2) continuing to track the location of that node. For the first step, following two properties of geocast overlays can be leveraged:

Property 1: For any given pair of overlay nodes, the probability of them being overlay neighbors is inversely proportional to the geographical distance between them.

Property 2: If a node discovers another node for the first time, the probability that this node has just joined the overlay or changed its pseudonym is inversely proportional to the geographical distance between the two nodes.

Note that Property 2 is a direct implication of Property 1. With these properties of geocast overlays, the attacker can attempt to map a victim to an overlay node by placing attacker nodes around the start point of the victim (e.g., by instructing them to lie about their positions). The attacker nodes can then report all new nodes they discover to the attacker. Whenever a new node \( X \) is discovered in the vicinity of the victim start position, following reasoning applies:

- \( X \) is close to the victim start point \( \rightarrow \) it is also close to the attacker nodes.
- It is close to the attacker nodes and seen by them for the first time \( \rightarrow \) it is likely to have just joined the overlay.
- It is close to the victim start point and has likely just joined the overlay \( \rightarrow \) it likely belongs to the victim.

Having acquired a likely victim node, the next step for the attacker is to continue tracking it. For this, we introduce an approach specific to geocast overlays - the follower attack. This attack is again based on Property 1. Given a node which the attacker wants to track, he instructs one of his attacker nodes to fake its location, speed and bearing in such a way that it always remains in the close vicinity of the target node. By doing so, the likelihood of the attacker node to remain a neighbor to the victim node can be increased to de facto certainty, allowing to keep the victim under continuous surveillance.

IV. Evaluation

For evaluating the privacy characteristics of geocast overlay systems, several simulation studies with the geocast overlay OverDrive were performed using the overlay simulation framework OverSim [1]. In the following, we first introduce the setup of these studies and present results. The implications of the results are then thoroughly discussed in Section V.

A. Common Setup

For all simulations, the highway network of the German state Baden-Württemberg was used as an underlying road network. Populations of vehicles acting as OverDrive nodes were simulated, using a shortest path movement model and overlay parameters as described in [6]. Unless otherwise noted, a regular nodes population of 10000 nodes was used. In addition to these regular nodes, an additional number of
attacker-controlled nodes was introduced into the overlay, depending on the evaluated scenario. No application apart from the overlay component was running on regular nodes. Thus, no application-specific location leaking was evaluated and attackers could only exploit the properties of the OverDrive protocol. On each attacker node, an attacker application was running that realizes the coordination of attacker nodes and the collective gathering of information. For each parameter setting, four simulation runs with different seeds were performed.

Based on this common setup and the attacker model from Section III, two specific attacks were evaluated: the establishment of a global view through the collection of data from multiple attacker nodes and the surveillance of an individual target using context knowledge and strategic attacker node placement.

B. Establishment of a Global View

In order to establish a global view on the overlay network, attacker nodes attempt to become overlay neighbors to as many regular nodes as possible, thus learning their geographic positions. The positions are then sent to a centralized attacker observer who combines the input of all attacker nodes into one global view of the network. The completeness of this view is verified every minute. Precision is ignored for the purposes of this study, i.e., the surveillance coverage is calculated only based on the number of known node positions and is not influenced by the deviation between known and real positions.

Several variations of attacker node properties were evaluated. Concerning mobility, attacker nodes were either stationary, with positions uniformly distributed in the underlying road network, or mobile, using the same movement model as regular nodes. The latter models the case that the attacker node is also a regular traffic participant. Concerning the maximum supported number of neighbors, configurations without a limit and configurations with a limit of 40 neighbors per ring (320 in total) were evaluated. This corresponds to two times the limit used at regular nodes. A limit might be relevant for attacker nodes if communication bandwidth is expensive or if the maintenance of a large number of neighbors is unfeasible for other reasons, e.g., due to the use of cryptopuzzles in maintenance operations.

Results: Fig. 2 shows, for different number of attackers, the measured surveillance coverage, i.e., the average percent of all non-attacker nodes whose locations are known to the attacker at each measurement interval. The plots demonstrate that as little as 10 randomly distributed, stationary attacker nodes are sufficient for achieving a global view covering the positions of more than 90% of the 10000 regular nodes, in a road network of around 5300 km length and 35750 square km spread. With 50 attacker nodes this even increases to above 98% coverage. However, a significant decrease in surveillance performance can be observed when the number of neighbors per attacker node is limited. Here, 10 stationary attacker nodes achieve surveillance coverages of only below 20%. The surveillance performance with mobile attacker nodes differs slightly, but stays in the same order of magnitude and follows the same tendency - a significant decrease in surveillance performance when the maximum number of neighbors per node is limited.

Fig. 3 depicts the distance traveled by a node until it was discovered, which is an important metric for assessing the difficulty for breaking pseudonyms based on the collected data. Note that these values are only recorded if a node was discovered at all and need to be seen in the context of the results from Fig. 2. As the start positions and times of trips are a likely anchor point for context-based attacks on pseudonymization schemes, a spatio-temporal distortion at the beginning of trips can greatly improve the difficulty for such attacks. In both scenarios (stationary and mobile attackers) the distance traveled until discovery degrades sharply with increasing numbers of attacker nodes. Unlike general surveillance coverage however, a minimum of around 50 stationary attacker nodes without a limit on the maximum number of neighbors is necessary for maintaining a satisfactory amount of precision, e.g., of only up to 2 km displacement from the start point on average.

Lastly, Fig. 4 shows the results from simulation runs with a fixed number of attacker nodes (20) and a variable number of regular nodes (up to 80000). The goal of these runs was to evaluate the feasibility of establishing a global view in larger networks, respectively the relationship between surveillance
success and the size of the node population kept under surveillance. It can be seen that surveillance performance degrades only little for stationary attacker nodes without a limit on the number of neighbors. For all other configurations, a significant drop in surveillance performance can be observed for growing network sizes. For mobile attacker nodes without a limit, this is due to the higher neighborhood structure instability.

C. Identification of an Individual Target

Instead of building a global view, the goal of the second attack is to identify a pseudonymized victim using context knowledge and only a minimal number of attacker nodes.

In the specific experiment that was constructed, a random location in the road network is marked as a victim start point. In addition to the large population of regular nodes starting their trips from random locations as described in [6], several victim nodes are created at the victim start point at intervals randomly distributed around 2 minutes. The attacker knows the coordinates of the victim start point. However, this is the only information he has in order to distinguish between victim nodes and regular nodes. Since the success of the attacker likely depends on the placement of the victim start points, e.g., because it influences the likelihood of regular nodes appearing in the vicinity of that point as well, twice as many simulation runs were performed per configuration in this experiment.

In order to identify victim nodes, the attacker follows these steps:

1) At the beginning of the simulation, he places (through location faking) all of his attacker nodes at random positions in a 1 km radius around the victim start point.
2) Each new neighbor of an attacker node is reported to the attacker observer.
3) If a reported node is seen for the first time by the attacker observer and if this node’s position is within a 1 km radius of the victim start point, the node is assumed to be one of the victim nodes (following the reasoning from Section III-D).

Results: Concerning the recognition of victim nodes, two metrics are especially interesting: the victim recognition rate and the victim recognition false positive rate, i.e., the amount nodes that were falsely identified as victims by the attacker. Concerning the victim recognition rate, the results in Fig. 5 demonstrate that 2 attacker nodes placed in vicinity of the victim start point are sufficient for identifying the victim with a probability of above 90% on average. Whenever an attacker marks a node as a victim, he may be wrong with a 35% chance, as can be seen in Fig. 5. The false positive rate is potentially unavoidable, due to the chance of regular, non-victim nodes starting their trips in the same area. The measured false positive rates are very broadly spread between different simulation runs. This is due to the large impact of the choice of a victim start point, e.g., in the probability of other nodes starting their trips close by. Neither the recognition success rate nor the false positive rate changes significantly with growing numbers of attacker nodes.

Assuming that the attacker follows potential victims probabilistically based on his certainty, the probability for the real victim to be followed is a combination of the recognition success rate $p_r$ and the recognition false positive rate $p_f$:

$$p_{\text{detection}} = p_r(1 - p_f)$$

Thus, with a recognition probability of 90% and a false positive rate of 35%, a victim has a 58.5% chance of being marked and followed by the attacker.

V. Discussion

The main question of this work is whether geocast overlays in general and OverDrive in detail outperform centralized approaches in terms of the difficulty for adversaries to keep individual vehicles under surveillance. In the case that the adversary in a centralized approach is the service provider himself, this is very much so. The service provider in a centralized system always has a global view on all participating nodes at no additional cost. Additionally, as an adversary he is unlikely to have an interest in enabling privacy-preserving mechanisms like the use of pseudonyms for users of his service.

An adversary that is not the service provider of a centralized system has the following options to attack privacy:

1) Obtaining position data from the service provider in the case of a centralized system.
2) Obtaining approximate position data from the victim’s cellular data provider.
3) Following the target physically.
4) Performing an attack on the geocast overlay in the case of a decentralized system.

Geocast overlays can be said to outperform centralized approaches in the case of an attacker that is not the service provider if the exploitation of the geocast overlay for surveillance is more difficult than any of the other available options. The listed alternatives to attacking the overlay have in common that a global view on all present users is provided. The results from Section IV show that the establishment of a global view in a geocast overlay is not entirely unfeasible. With control over 50 nodes without a limit on the number of overlay neighbors, an attacker is able to reconstruct an almost complete view over a network of 10000 nodes running the unmodified geocast overlay protocol OverDrive as described in [6]. However, this view does not feature real identities and its spatio-temporal resolution is not optimal for breaking pseudonymization schemes. Additionally, the surveillance
performance of the attacker was shown to decrease if the maximum number of neighbors per overlay node is limited. Thus, the implementation of schemes that guarantee such a limit, e.g., through the introduction of mutual checks between overlay neighbors or cryptographic puzzles per overlay operation, seems worthwhile to achieve privacy.

Another approach to attack the privacy of geocast overlay users is to use context knowledge. The results from Section IV show that on average, a context-based pseudonym breaking attack on OverDrive with 2 attacker nodes leads to a probability of correct identification of around 58%. This result is especially problematic in connection with follower attacks. After the identification of a victim, it can be followed at very low cost by faking the location, bearing or speed of an attacker node. Several countermeasures to this approach are possible. For one, plausibility checks against location faking, e.g., using short range wireless communication technologies, can be implemented. This would hamper the initial victim identification as well, as it forces attacker nodes to be physically residing in the vicinity of the victim start point. Then, a pseudonym mixing scheme for geocast overlays should be used, that enables the change of pseudonyms during trips and also mixes neighbor sets. For example, a collaborative mixing approach as described in [8] might be appropriate. With all approaches, the additional challenge exists that all privacy-enhancing measures apply to attacker nodes as well. So, for example, attacker nodes can change their pseudonyms as well, thus circumventing the effects of detection.

With this added problem of attacker node anonymity, the question arises whether a hybrid communication approach might not be more suitable for achieving the benefits of fully flat and decentralized geocast overlays while improving privacy and possibly also the bandwidth overhead for participating nodes. Given a number of publicly known stationary nodes for example, more transparency can be realized as to with whom ones location data is shared with. Such stationary geocast nodes can be set up by municipalities or similar organizations interested in efficient road traffic within a given region. The cost of such a setup and the actual benefit in terms of privacy needs to be verified in further work.

VI. RELATED WORK

Pseudonymization schemes for user privacy in vehicular ad-hoc networking (VANET) scenarios have been proposed and evaluated thoroughly - [9], [4]. However, VANETs focus on local one- and few-hop communication and not on the formation of a global overlay network.

Privacy in general has also been a major topic in the context of location based systems (LBS). [11] gives an excellent overview over different techniques and attacks. However, vehicular and traffic scenarios are not in the focus in LBS privacy research. Additionally, according to our knowledge, no decentralized, overlay-based geocast systems like OverDrive [6], GeoKad [5] or GeoKad [10] have yet been thoroughly analyzed and evaluated in terms of user privacy.

VII. CONCLUSION

The main goal of this paper was the evaluation of the privacy characteristics of smart traffic systems based on geocast overlays in comparison to centralized approaches. To this end, we developed an attacker model and proposed two specific attacks. The efficiency of these attacks on the geocast overlay OverDrive was then evaluated through simulation. Results show that the difficulty for placing surveillance on individual nodes is low. In a typical scenario, victims have a chance of around 41% to remain untracked by an attacker with context knowledge and the ability to fake node locations. The establishment of a global view on all participants was shown to be difficult if limits on the number of attacker nodes and on the maximum number of neighbors per attacker node are imposed. Open challenges include the development of specific protection mechanisms and dealing with attacker anonymity.

ACKNOWLEDGMENT

This research was supported by the German Federal Ministry of Economics and Technology as part of the iZEUS project 01ME12013. The authors are responsible for the content.

REFERENCES