Enhancing Security and Privacy in Traffic-Monitoring Systems

This architecture separates data from identities by splitting communication from data analysis. Data suppression techniques can help prevent data mining algorithms from reconstructing private information from anonymous database samples.

Baik Hoh, Marco Gruteser, and Hui Xiong
Rutgers University

Ansaf Alrabady
General Motors

Intelligent transportation systems increasingly depend on *probe vehicles* to monitor traffic: they can automatically report position, travel time, traffic incidents, and road surface problems to a telematics service provider. This kind of traffic-monitoring system could provide good coverage and timely information on many more roadways than is possible with a fixed infrastructure such as cameras and loop detectors. This approach also promises significant reductions in infrastructure cost because the system can exploit the sensing, computing, and communications devices already installed in many modern vehicles.

Although these applications can improve travelers' safety, optimize traffic, and provide a new revenue source for car manufacturers, they also raise questions about privacy. Because users' vehicles provide data samples that include their current positions and user identities to location-monitoring services, the drivers can be tracked. Unfortunately, anonymous data collection doesn't solve this privacy problem. First, it conflicts with security—in particular, data integrity—which requires user identification. Second, even if location samples are anonymous, users can be reidentified through data mining techniques (so-called *inference attacks*).

The architecture we propose meets these privacy and data integrity requirements. It addresses privacy by separating the communication and authentication tasks (which rely on pseudonyms or identities) from data analysis and sanitization (which require access to detailed position information). Because these functions share only well-defined messages, only anonymous position information is available at the traffic-monitoring service.

We conducted a case study to evaluate how vulnerable this information is to inference attacks, and found that clustering techniques can automatically identify many vehicles' likely home locations in a typical suburban scenario. This is grounds for concern, because attackers could link home locations to household names using geocoded address databases to identify drivers. However, the results also show that data suppression techniques such as reducing sampling frequency effectively lower such reidentification risks.

Traffic-monitoring with probe vehicles

Our traffic-monitoring application estimates travel time for different routes using real-time traffic flow information. It derives that information from probe vehicle speed on different road segments. Probe vehicles monitor the road environment through in-vehicle sensors.[1] Xiaowen Dai and her colleagues have determined that we can derive useful traffic flow information if 5 percent of vehicles act as probes.[2] This approach promises reduced infrastructure installation and maintenance costs while extending sensing coverage to less-traveled roadways.

Published by the IEEE CS and IEEE ComSoc ■ 1536-1268/06/$20.00 © 2006 IEEE

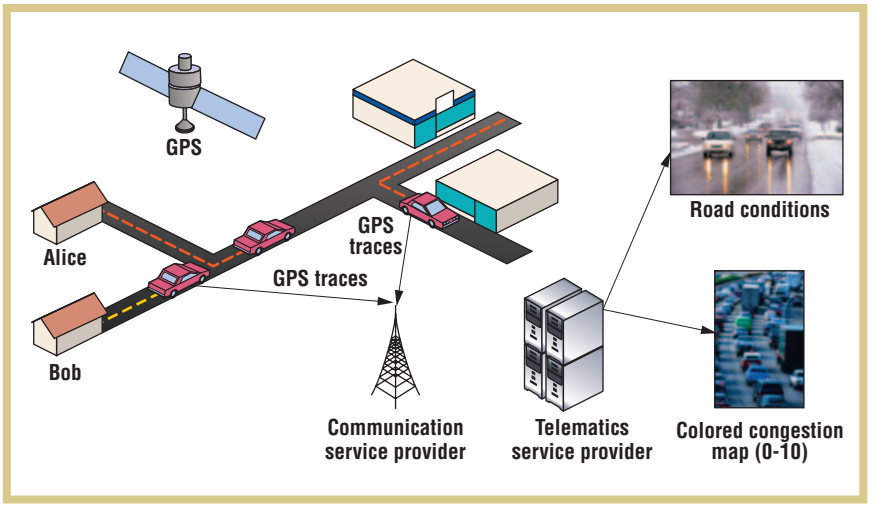


Figure 1. traffic-monitoring architecture comprising three entities: probe vehicles, a communication service provider, and a telematics service provider.

Figure 1 shows a typical traffic-monitoring architecture comprising three entities: probe vehicles, a communication service provider (CSP), and a telematics service provider (TSP)—perhaps a subsidiary of a vehicle manufacturer. Telematics is the use of GPS technology integrated with computers and mobile communication technology in vehicles. The TSP connects vehicles to a main server, perhaps through base stations leased from CSPs. Probe vehicles carry GPS receivers and communication infrastructure such as cellular links to periodically report data records (latitude, longitude, time, and speed parameters) to the traffic information system. From this information, the system can estimate current mean vehicle speed and then feed it into navigation systems or use it to build a real-time congestion map (for example, by calculating a congestion index). It can also use vehicle speed to estimate traffic density and volume using Greenshields' equation.[3] The system can then broadcast estimated traffic information to subscribers through a Web interface, where drivers can access it through their navigation systems or from home or office computers.

Security and privacy challenges

The primary security and privacy challenges that traffic-monitoring applications face are to ensure the integrity of data samples containing speed and position information and to maintain privacy for the drivers who supply the samples. (For more information, see the "Related Work in Security and Privacy" sidebar.)

Data integrity

The integrity of the computed congestion index relies on genuine speed and position data from the probe vehicles. Malfunctioning probes and malicious parties who modify sensor readings can affect data integrity. Although malicious attacks on traffic monitoring might sound far-fetched, they appear quite plausible if you consider the gray-market devices people now buy to reduce travel time (such as infrared transmitters to change traffic lights). These new devices might manipulate the congestion index to divert traffic away from a road to reduce a particular driver's travel time or toward a particular roadway to increase revenue at a particular store. Other service providers might also try to dilute the information quality of a competing traffic-monitoring service.

Various entities can compromise data integrity:

- *Compromised vehicles*. Drivers or third parties could modify the hardware or software to report incorrect vehicle positions or speed readings. (Such modifications have occurred in European trucks' tachographs, which are supposed to record vehicles' driving times and speed to let authorities check adherence to mandatory driver rest periods.[4]
- *Impostor devices*. A device could spoof other authorized devices. This compromise is of particular concern in the form of a Sybil attack,[5] in which a device illegitimately claims multiple entities. Traffic-monitoring accuracy will degrade more if many vehicles simultaneously report incorrect information.
- *Network intermediaries*. The transmission of vehicle data over wireless and wired communication links enables intermediate network entities to modify reports.

Privacy

Proactively addressing privacy concerns in the architecture increases the potential for users to adopt the traffic-monitoring service and reduces the risk of public data-handling mishaps. Location information collected by probe vehicles raises privacy concerns because it's often precise enough to pinpoint the buildings that drivers visited, at least in suburban areas where each building has its own parking lot. Reconstructing an individual's route could provide a detailed movement profile that allows sensitive inferences. For example, recurring visits to a medical clinic could indicate illness; visits to activist organizations could hint at political opinions. While everyone's location traces deserve protection, those of political leaders, celebrities, or business leaders would likely undergo particular scrutiny. For example, frequent meetings between chief executives might indicate a pending merger or acquisition—highly desirable information for competitors and stock market speculators.

Related Work in Security and Privacy

To allow authentication of messages from vehicles while maintaining a degree of anonymity in vehicular networks, Maxim Raya and Jean-Pierre Hubaux propose frequently changing the signing keys.[1] To implement this, you could preload each vehicle with numerous anonymous public–private key pairs. Using asymmetric keys can eliminate the key agreement step in this solution. But storing keys in vehicles might allow a Sybil attack unless trusted computing hardware protects the keys. To implement this, Bryan Parno and Adrian Perrig propose installing reanonymizers (hardware that issues a fresh ID in response to valid, temporary certificates) in stoplights or tollbooths at regular intervals to refresh vehicles' anonymous keys.[2] These solutions are primarily intended for vehicle-to-vehicle communications. Our proposed architecture for vehicle-to-infrastructure communication uses fewer keys and enables easier key revocation when vehicles' keys have been compromised.

Researchers have also addressed the question of balancing security and privacy in application areas such as electronic cash and electronic voting. Pioneered by David Chaum and Eugène Van Heyst,[3] many proposals rely on group signatures that can verify a message sender's group membership while maintaining the sender's anonymity. Dan Boneh and his colleagues reduced the size of group signatures, potentially enabling their use in challenging wireless network environments.[4] Although group signatures provide a possible alternative solution, they also require a trustworthy third party to enable key revocation (to determine the original signer) and lead to large message overhead of about 200 bytes per signature. We've opted for less complex cryptographic primitives.

REFERENCES

1. M. Raya and J.-P. Hubaux, "The Security of Vehicular Ad Hoc Networks," *Proc. 3rd ACM Workshop Security of Ad Hoc and Sensor Networks* (SASN 05), ACM Press, 2005, pp. 11–21.
2. B. Parno and A. Perrig, "Challenges in Securing Vehicular Networks," *Proc. 4th Workshop Hot Topics in Networks* (HotNets-IV), ACM SIGCOMM, 2005; www.acm.org/sigs/sigcomm/HotNets-IV/program.html.
3. D. Chaum and E. Van Heyst, "Group Signatures," *Advances in Cryptology: EUROCRYPT 91*, LNCS 547, Springer, 1991, pp. 257–265.
4. D. Boneh, X. Boyen, and H. Shacham, "Short Group Signatures," *24th Ann. Int'l Cryptology Conf.* (CRYPTO 04), LNCS 3152, Springer, 2004, pp. 41–55.

Various entities can compromise privacy:

- *Eavesdroppers*. Unauthorized third parties could monitor network transmissions for vehicle position readings and unique identifiers that would let them track vehicles. In particular, third parties could monitor wireless transmissions around particularly sensitive locations to record which vehicles recurrently visit this area. Network identifiers, such as the international mobile subscriber identifiers (IMSI) in the GSM (Global System for Mobile Communications) cell phone system, help identify recurring visits.
- *Spyware*. People with access to the on-board vehicle system could install software that directly reports vehicle positions to unauthorized network servers.
- *Insiders*. Privacy breaches through insiders are particularly insidious at the traffic-monitoring server (we'll describe this in detail later), which receives and stores reports from large numbers of different vehicles. Although access control mechanisms provide some protection, several individuals, such as system administrators, typically have root access to the system. (On 8 April 2006, *Information Week* posted a chronology of data breaches reported since the ChoicePoint incident. ChoicePoint, which maintains and sells background files on every American adult by selecting from public and private records, reported on 15 February 2005 that their 145,000 customers were at risk for identity theft. Most incidents were due to current, authorized employees in the victim companies.)

There's tension between integrity and privacy requirements. A true privacy compromise requires not only determining the places visited but also identifying the vehicle and driver. Thus, privacy can be significantly enhanced if the vehicles anonymously report position and speed information. The system can better maintain integrity, however, if the vehicles identify themselves and their identities can be authenticated. Strong authentication combined with an authority that issues and registers vehicle identities can prevent a Sybil attack, perhaps the main concern with regard to data integrity.

Architecture for anonymous data collection

To resolve the tension between data integrity and privacy, the architecture assigns the authentication and filtering functions and the actual data analysis to separate entities. One entity knows the vehicle's identity but can't access precise position and speed information; the other entity knows position and speed but not identity. The architecture also relies on encryption to prevent eavesdropping, tamper-proof hardware to reduce the risk of node compromise and spyware installation, and data sanitization to further strengthen data integrity.

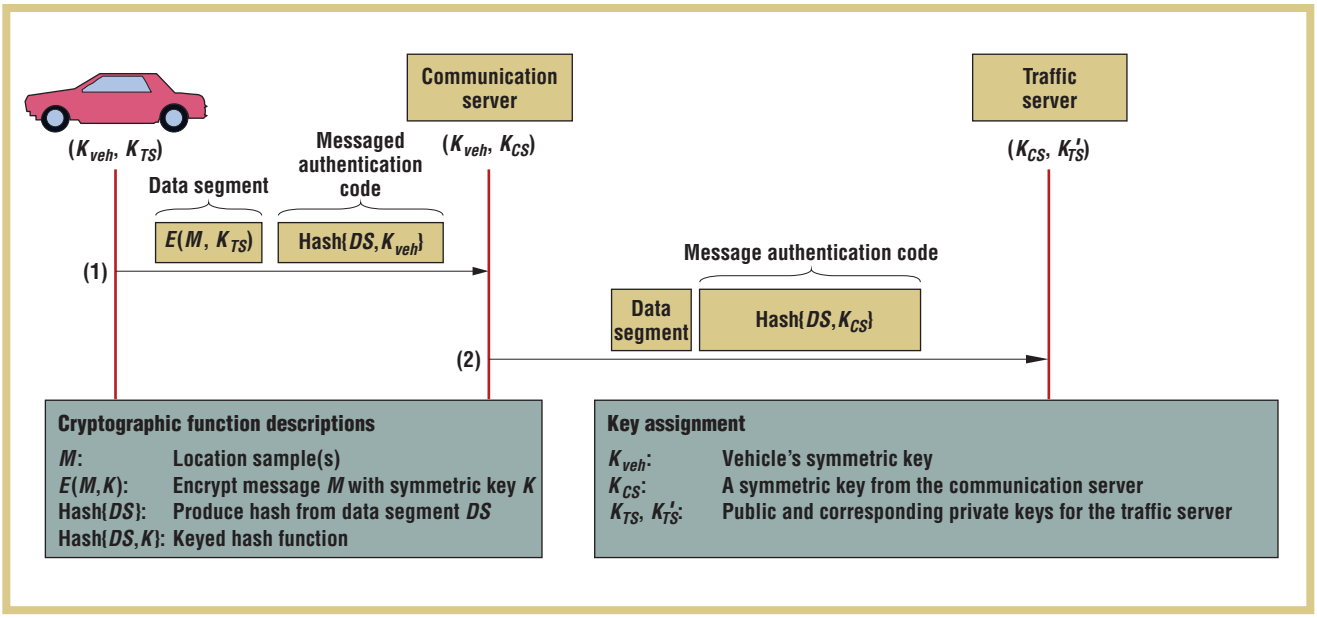


Figure 2. Traffic-monitoring architecture to ensure data integrity and anonymous data collection.

Figure 2 illustrates the entities and cryptographic schemes involved in transmitting a data sample from a vehicle. We distinguish the communication server (CS) from the traffic server (TS). The CS, which a CSP provides, maintains network connections and authenticates users but doesn't access location and speed data. The TS receives anonymous data from the CS, decrypts and sanitizes it, and computes the real-time congestion maps. In a real implementation, a cellular-phone service provider could provide the CS, and a TSP could provide the traffic server. The two parties would likely enter a contractual relationship as business partners that would prohibit the exchange of any privacy-sensitive information beyond that specified in this architecture. To further increase user confidence, an independent agency could audit the information exchange between these parties.

One pair of keys enables encryption between the TS and vehicles. Every vehicle knows the TS's public key K_{TS} and uses it to encrypt a location sample. We refer to this encrypted message as a *data segment*. Since the TS can decrypt the DS only with its private key, this layer of encryption protects location privacy against eavesdroppers.

The CS shares a separate symmetric key K_{veh} with each vehicle and knows the network identifiers (such as IMSI in GSM networks). Using this key, the CS can authenticate incoming data samples and ensure that authorized probe vehicles are transmitting them. If they're valid, the CS then removes all network identifiers and the message authentication code from the vehicle and attaches its own MAC using a third key K_{CS} established between the TS and the CS.

Key distribution and storage

The proposed architecture requires storing K_{veh} in vehicles. If an intruder can easily extract secret vehicle keys from multiple cars, the intruder could insert large numbers of incorrect data samples into the traffic-monitoring system. Thus the key should be stored in tamper-proof hardware. The TS's public key K_{TS}, on the other hand, need not be stored in tamper-proof hardware as long as users (or vehicles) can verify its integrity and authenticity.

Vehicles' keys initially can be embedded by the manufacturer and updated during regular government vehicle inspections or regular maintenance. This lets the manufacturer or inspection agency replace keys if they've been compromised. If more frequent key updates are necessary, we can extend the architecture to allow over-the-air provisioning of new keys.

A sanitizer for traffic-monitoring systems

The cryptographic authentication mechanisms can address Sybil attacks (provided that the keys are hard to generate) and message modifications by network intermediaries, but they can't prevent incorrect reports from compromised vehicles. So, the TS should sanitize received data.

Techniques for sanity checking include outlier detection, consistency checking, and rule-based classification.[6,7] We can leverage these techniques to build a sanitizer component for traffic-monitoring systems. For example, the sanitizer could test data integrity by comparing an anonymous vehicle's claimed speed on a specific road segment at a specific time with

- statistics that other vehicles report in the same situation,
- statistics collected one month earlier in the same situation, or
- adjacent location sample data reported by the same vehicle.

For example, if a malicious vehicle sends a fake message reporting low speed (a severe traffic jam) but the sanitizer finds that most probe vehicles on the same road segment at a similar time report high speed, the system can easily detect this as an unreliable message.

We can extend the system to actively blacklist vehicles that submit apparently incorrect data. Because only the CS maintains identities, the TS must return the message with the incorrect data to the CS. The CS in turn looks up this message's originator (this requires buffering messages for a certain time window) and drops all further messages from this vehicle until the CS can establish its integrity through other means.

Discussion

As we've already said, this architecture can provide privacy guarantees against basic eavesdropping and insider attacks through encryption and the separation of identity and position information.

More sophisticated intrusions at the CSP and the TSP are also possible.

CS integrity

The proposed architecture assumes that the CS is trustworthy with respect to data integrity. The architecture provides no cryptographic protection against the CS spoofing, replaying, or dropping messages. To relax this trustworthiness assumption, the sanitizer could easily filter replayed messages at the TS, because no two messages should contain the identical GPS time stamp and position. We could also add a basic degree of protection against spoofed messages through an additional symmetric key, K_{INT}, that all vehicles and the TS share. Vehicles can use this key to generate a MAC for each location update message that the TS can verify without being able to identify the vehicle. However, vehicles and the TS would need to update this key regularly because a key shared by many vehicles is difficult to keep secret. Identifying dropped messages proves most difficult.

Because the TSP database mixes anonymous location samples with all vehicles, private information is hard to extract. Nevertheless, breaches can still occur.

For a more comprehensive solution, the TSP should continuously monitor the traffic data's quality by cross-checking with other data sources and monitoring consumer complaints. Monitoring should let the TSP identify if the CS has inserted a continuous bias in the data. It might also make an additional authentication key (K_{CS}) unnecessary.

In this architecture, we've deliberately emphasized privacy protection, because privacy leaks are often more difficult to identify than integrity problems. Because the CSP and TSP will enter a mutually beneficial contractual relationship, both parties will want to maintain data integrity and monitor the possibility of insider attacks. Individual drivers, however, have fewer resources to verify that their private data hasn't been compromised.

Location privacy at the CSP

The proposed architecture provides location privacy to drivers with respect to the CSP, because only the TSP knows the secret key to decrypt the GPS samples. Although the CSP could probably use wireless network localization methods to obtain the mobile node's position, these methods would be significantly less accurate.

For example, cell phone localization techniques in the US were designed to Federal Communications Commission specifications. The E911 Phase II mandate states that a system should be able to locate 67 percent of calls within 100 meters and 95 percent of calls within 300 meters. So, we can expect commonly used technologies such as Uplink Time Difference of Arrival to provide an order-of-magnitude less accuracy than in-vehicle GPS, which typically achieves better than 10-meter accuracy. Assisted GPS technology, which relies on GPS chips in cell phone handsets, might be more precise. A-GPS could be easily disabled, however, for in-vehicle deployment.

Location privacy at the TSP

Because the TSP database mixes anonymous location samples from all vehicles, private information is hard to extract. If multiple vehicles cross paths, discerning which sample belongs to which vehicle is difficult. Nevertheless, breaches can still occur—intruders can access decrypted location samples from the TSP's database.

Here are two risk scenarios arising from data mining techniques in which privacy might be compromised even if an anonymous data collection architecture is deployed:

- *Home identification*. An intruder might identify a home's location from probe vehicle drivers as a first step toward identifying a particular driver.

- *Target tracking*. An intruder might reconstruct paths from anonymous traces and use them to link the driver to sensitive places that he or she visited.

Although these techniques are most useful in conjunction—a privacy compromise requires both identifying the driver and acquiring sensitive information about the individual—we concentrate here on home identification.

Home identification. Clustering can be an effective tool for home identification.[8] Clustering analysis provides insight into data by dividing objects into groups (clusters) such that the objects in a cluster are more similar to each other than to the objects in other clusters.

Consider the case in which an authorized, legitimate, but malicious employee at the TSP accesses GPS trace data collected as defined in our proposed architecture. Since location samples are anonymous, at most, the adversary can obtain a collection of GPS location samples without user identity. In addition, because of measurement inaccuracies and the possibility of using different parking spots, the exact endpoints of a GPS trace might differ by hundreds of feet, even though a vehicle visits the same place.

However, clustering techniques can smooth out such noisy GPS traces and allow automatic identification of repeatedly visited places. Specifically, clustering algorithms can automatically group a set of location samples that likely belong to the same destination: anonymized location samples with low-to-zero speed might be candidates for endpoints, and the centroid of this cluster of endpoints provides a good estimate of the destination. We can improve the estimate's accuracy by knowing road topology as provided by digital road maps such as those from GoogleMap or MapQuest. In our clustering practice, we've developed a set of heuristic rules to filter out irrelevant location samples. For instance, we can differentiate stationary from moving GPS location samples by looking at GPS speed information. Also, we can use time information to distinguish home locations from other kinds of destinations. If the marked time is from 4 p.m. to midnight and we detect no subsequent moving GPS location samples before the next morning, the destination is more likely to be a home than a workplace.

Indeed, place identification is a general technique to extract potentially sensitive information about a driver's habits and interests.[9] This information is also directly related to the ability to identify a driver from anonymous traces. Generally, the more information intruders know about a data subject (workplace, home location, gym visits, favorite restaurant, and so on), the more likely they can identify that driver. We believe that home identification provides the highest risk, because there's usually a one-to-one mapping between a typical suburban home and a household, and home owners and occupants are public knowledge through telephone white pages or real estate records.

Target tracking. Adversaries can use target tracking to reconstruct paths from anonymous samples or segments,[10,11] especially once they've identified a home location. Privacy risks go beyond knowing a home location once they've linked potentially sensitive information or places to this home. Target tracking lets an adversary follow the traces reported by a vehicle to other locations, thereby linking information about other places to the driver's identity. However, these techniques don't work well in urban areas because buildings, bridges, and tunnels often block GPS signals; they're more effective in suburban areas, which contain less dense GPS traces.

Target tracking lets an adversary follow the traces reported by a vehicle to other locations, thereby linking information about other places to the driver's identity.

Sampling frequency and home identification: a case study

This case study analyzes the effectiveness of home identification techniques on the TSP's data sets. Our objective is to shed light on the most serious privacy question raised in the discussion of our architecture. Is anonymous data collection enough to protect user privacy? If not, what sampling frequency provides enough data without unduly raising the privacy risk? Intuitively, the privacy risk decreases when the system operates with lower sampling frequency (the frequency with which probe vehicles provide position updates). So, a judicious choice of the sampling frequency is critical in our proposed architecture. Operating at reduced sampling frequency is a basic data suppression technique that you can derive from known concepts such as mix zones[12] or cloaking techniques.[13] In this case study, we consider how effectively this basic approach reduces the home identification risk. Of course, the challenge in devising a suppression technique is to improve privacy while not unnecessarily reducing service quality. Other researchers have looked at the effect on

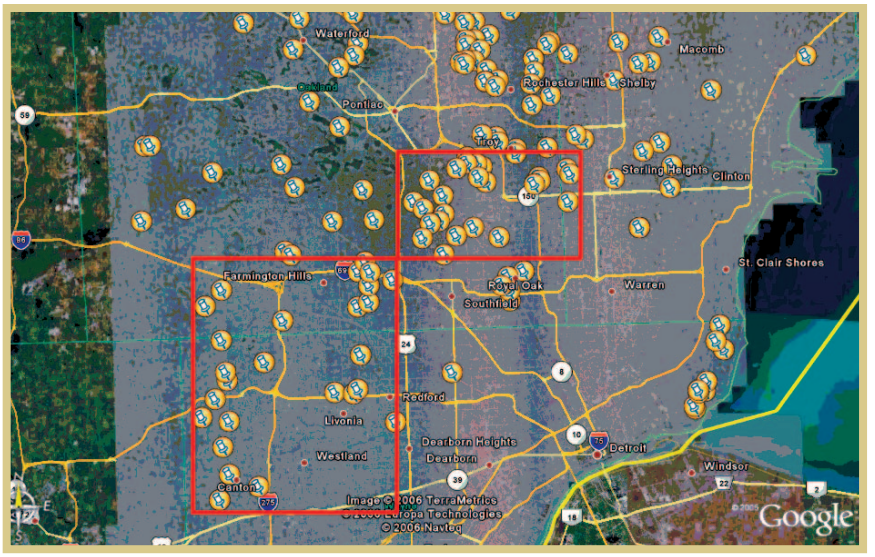


Figure 3. Plausible home locations in two target regions (the red rectangles), identified through manual inspection. The study considered 65 homes in a 25 × 25 km area.

service quality,[2] so our study concentrates on privacy aspects.

Our study uses a data set containing GPS traces from vehicles driving in the larger Detroit, Michigan, area. For privacy reasons, we had no specific information about the vehicles or drivers except that the drivers were volunteers. Each GPS sample comprises vehicle ID, time stamp, longitude, latitude, speed, and heading information. Each vehicle records a GPS sample every minute while its ignition is switched on, for a period of one week. This means that the traces contain both spatial and temporal gaps. No data is provided while the vehicle is parked with its ignition switched off. In addition, data was unavailable when the GPS receiver was acquiring or had lost a satellite fix (for example, because of obstruction from high-rise buildings).

Clustering-based home identification algorithm

For our home identification algorithm, we use a *k*-means clustering algorithm on anonymous location samples to identify frequently visited places. We then refine the resulting clusters using several heuristics. First, a set of anonymous location samples near a home likely have low to zero speed. Second, vehicles are often parked overnight at homes. Specifically, the key steps of the algorithm are the following:

1. Drop location samples that are too high-speed (> 1 meter/second) from the set of all vehicles (the remaining samples contain the candidate trip endpoints).
2. Select a target region of interest to improve computational efficiency, and drop samples outside this region.
3. Apply the *k*-means pairwise clustering algorithm to samples in the target region and store the returned cluster centroids.
4. Filter the candidate home locations out of all centroids using heuristic A, arrival time, and heuristic B, zoning information.

Step 3 repeats to calculate the centroids of clusters until it finally groups all location samples into the optimum number of clusters. The *k*-means pairwise clustering algorithm in step 3 doesn't have a priori knowledge of the optimum number of clusters at the initial run. So, it uses all locations obtained after step 2 as initial clusters and keeps merging close ones into fewer clusters at each run. The merging process stops when every centroid has all its elements within a certain distance (D_{th}) on the average. D_{th} should be chosen according to different home densities. If the home density is too dense, keep D_{th} small enough to differentiate the locations of other vehicles living near each other. In our simulations, we use a value of 100 m for this threshold, which we derived from the region's actual home density.

Filtering with heuristic A eliminates all centroids that don't have any evening visits. We define an evening visit as a location sample arriving between 4 p.m. and midnight. Filtering with heuristic B eliminates all centroids outside residential areas. In our experiments, we've eliminated centroids outside residential areas by manually inspecting satellite imagery (using Google Earth). You could automate this process by obtaining geographic-information-system data sets with city zoning information.

Because real home addresses were unavailable (we omitted driver identities in the data set for privacy reasons), we manually inspected the unmodified week-long traces overlayed on satellite images to identify plausible home locations as an experiment baseline. To make the evaluation feasible, we analyzed a subset of the region covered by the 239 traces in the data set (each trace corresponds to one driver). The subset contained the two residential regions (together a 25 × 25 km area) marked with rectangles in figure 3. We found 65 plausible homes through manual inspection. We then compared the automated algorithms to the results of the manual inspection. For the algorithm evaluation, we considered a home correctly identified if the algorithm and manual inspection gave the same answer. The results indicate whether the inspection task could be automated for mass surveillance purposes.

Because no real ground truth was available, the experiment doesn't definitely

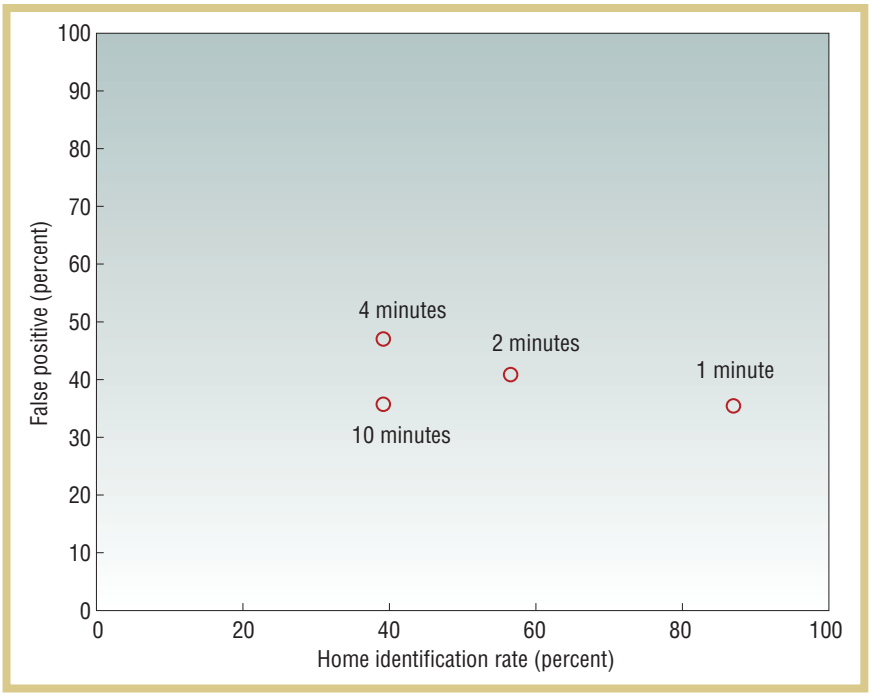


Figure 4. Results of the home identification algorithm using four different sampling intervals.

answer whether we identified the drivers' real home locations. The 65 reference locations we chose manually, however, each contained a single home that stood out as a likely home location—the drivers visited this location much more frequently at night than others. So, we believe manual inspection provided a reasonable approximation of real home positions.

To examine the effectiveness of reducing sampling frequency, we measured the home identification rate (how many homes out of 65 we correctly detected) and the false positive (how many of the estimated home locations were incorrect) at sample intervals. False positives can be caused by many vehicles waiting at traffic lights or the cluster centroid shifting to a neighbor's house because of inaccurate location reports. In addition to the standard 1-minute sample interval (which produces one location trace per minute), we considered 2-, 4-, and 10-minute intervals.

Results

Figure 4 shows that at the standard rate, the home identification algorithm correctly located about 85 percent of the homes, albeit also returning a large number of false positives. Reducing the sampling frequency decreases the home identification rate to 40 percent for the 4-minute interval, with similar false-positive rates. Although there's no clear linear trend—for example, home identification with 10-minute intervals performed better than with 4-minute intervals because it happened to generate fewer candidate centroids after clustering—the results indicate that data suppression algorithms can reduce the home identification risk and thereby increase privacy. Also, although the home identification intrusion technique we evaluated suffered from many false positives, it can be at least effective for automated prefiltering, followed by manual inspection to remove false positives.

The degree of privacy protection this architecture provides depends on judiciously choosing the frequency with which probes send in their position updates. Sampling frequencies higher than one sample per minute, as frequently considered for traffic-monitoring applications,[1,2,14] allow data mining techniques to reidentify many of the probe vehicles. To provide a high degree of privacy protection, traffic-monitoring systems should operate at sample frequencies of at least several minutes or employ more sophisticated data suppression mechanisms that can optimize both privacy and data quality. P

ACKNOWLEDGMENTS

This work was supported in part by the US National Science Foundation under grant CN5-0524475.

REFERENCES

1. T. Ishizaka, A. Fukuda, and S. Narupiti, "Evaluation of Probe Vehicle System by Using Micro Simulation Model and Cost Analysis," *J. Eastern Asia Soc. Transportation Studies*, vol. 6, 2005, pp. 2502–2514.
2. X. Dai, M. Ferman, and R. Roesser, "A Simulation Evaluation of a Real-Time Traffic Information System Using Probe Vehicles," *Proc. 2003 IEEE Intelligent Transportation Systems*, vol. 1, IEEE Press, 2003, pp. 475–480.
3. B. Greenshields, "A Study of Traffic Capacity," *Highway Research Board Proc.*, vol. 14, 1935, pp. 448–477.
4. R. Anderson, *Security Engineering: A Guide to Building Dependable Distributed Systems*, John Wiley & Sons, 2001, pp. 222-225.
5. J.R. Douceur, "The Sybil Attack," *Proc. 1st Int'l Workshop Peer-to-Peer Systems* (IPTPS 02), 2002; www.cs.rice.edu/Conferences/IPTPS02/101.pdf.
6. N. Adam, V. Janeja, and V. Atluri, "Neighborhood Based Detection of Anomalies in High Dimensional Spatio-Temporal Sensor Datasets," *Proc. 2004 ACM Symp. Applied Computing*, ACM Press, 2004, pp. 576–583.
7. D. Beneventano et al., "Consistency Checking in Complex Object Database Schemata with Integrity Constraints," *Proc. IEEE Trans. Knowledge and Data Eng.*, vol. 10, no. 4, 1998, pp. 576–598.

8. A.K. Jain and R.C. Dubes, *Algorithms for Clustering Data*, Prentice Hall, 1988.

9. D. Ashbrook and T. Starner, "Using GPS to Learn Significant Locations and Predict Movement across Multiple Users," *Personal Ubiquitous Computing*, vol. 7, no. 5, 2003, pp. 275–286.

10. D. Reid, "An Algorithm for Tracking Multiple Targets," *IEEE Trans. Automatic Control*, vol. 24, no. 6, 1979, pp. 843–854.

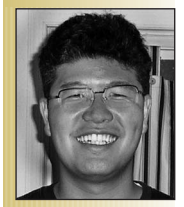
11. M. Gruteser and B. Hoh, "On the Anonymity of Periodic Location Samples," *Proc. 2nd Int'l Conf. Security in Pervasive Computing*, LNCS 3674, Springer, 2005, pp. 179–192.

12. A. Beresford and F. Stajano, "Mix Zones: User Privacy in Location-Aware Services," *IEEE Workshop Pervasive Computing and Comm. Security* (PerSec 04), 2004; www.cl.cam.ac.uk/~fms27/papers/2004-BeresfordSta-mix.pdf.

13. M. Gruteser and X. Liu, "Protecting Privacy in Continuous Location-Tracking Applications," *IEEE Security and Privacy*, vol. 2, no. 2, 2004, pp. 28–34.

14. R. Cayford and T. Johnson, "Operational Parameters Affecting Use of Anonymous Cell Phone Tracking for Generating Traffic Information," *Proc. Inst. of Transportation Studies 82nd TRB Ann. Meeting*, vol. 1, no. 3, 2003.

the AUTHORS



Baik Hoh is a PhD student in the Electrical and Computer Engineering Department and WINLAB (Wireless Information Network Laboratory) at Rutgers University. His research includes privacy-enhancing technologies and mobile worm and virus quarantine. He received his MS in electrical and computer engineering from the Korea Advanced Institute of Science and Technology. He's a student member of the IEEE Computer Society and the ACM. Contact him at WINLAB, Rutgers Univ., 671 Rte. 1 S., North Brunswick, NJ 08902; baikhoh@winlab.rutgers.edu.

Marco Gruteser is an assistant professor at WINLAB, Rutgers University. His research interests include location-aware networking and building reliable, secure, and privacy-aware communication systems for vehicular networks. He received his PhD in computer science from the University of Colorado at Boulder. He is a member of the IEEE Computer Society and the ACM. Contact him at WINLAB, Rutgers Univ., 671 Rte. 1 S., North Brunswick, NJ 08902; gruteser@winlab.rutgers.edu.

Hui Xiong is an assistant professor of computer information systems in the Management Science and Information Systems Department at Rutgers University. His research interests include data mining, spatial databases, statistical computing, and geographic information systems. He received his PhD in computer science from the University of Minnesota. He is coeditor of *Clustering and Information Retrieval* and coeditor in chief of the *Encyclopedia of Geographical Information Science*. He is a member of the IEEE Computer Society and the ACM. Contact him at Rutgers Univ., Ackerson Hall, 200K, 180 University Ave., Newark, NJ 07102; hui@rbs.rutgers.edu.

Ansaf Alrabady is a senior research engineer at General Motors. His main research is in communication security and embedded system development for automotive applications. He received his PhD in computer engineering from Wayne State University. In 2001, he received the Automotive Hall of Fame Young Leadership and Excellence award for his contributions to the automotive industry. Contact him at General Motors, 30500 Mound Rd., Warren, MI 48090-9055; ansaf.alrabady@gm.com.