A Characterization of the Mass Surveillance Potential of Road Traffic Monitors

Kirk Boyer, Hao Chen, Jingwei Chen, Jian Qiu, and Rinku Dewri*

Department of Computer Science, University of Denver, Colorado, USA
rdewri@cs.du.edu

Abstract. Modern technology allows for the detection and identification of a vehicle passing through specific locations on a road network. The most prominent technology in use leverages cameras capable of high speed image capture, back-end extraction of plate numbers, and real time membership queries in multiple databases. Various parties have a vested interest in making use of the kind of data produced by such systems, in particular to deter risky driving, analyze traffic patterns, enable unmanned toll collection, and aid law enforcement agencies. In this paper, we proceed to assess the mass surveillance potential arising from the type and frequency of data collected in these systems. We show that even when restricted to information only about the structure of a road network, one can begin to set up an effective network of traffic monitoring devices to infer the travel destinations of individuals up to a concerning level of precision. We develop a tracker placement algorithm to corroborate this claim, and provide a quantitative evaluation of the privacy risks generated by the network of trackers determined by this algorithm.

Keywords: Location privacy · Mass surveillance · ALPR

1 Introduction

A recurring theme in privacy is that the development of technologies that provide new useful forms of data often inadvertently provide access, even if indirectly, to information that is much more personal than originally intended or hoped. One such technology is automatic vehicle detection and identification, which can be used for traffic control and enforcement, monitoring accidents, toll collection, criminal pursuits, and gathering intelligence, among others. The technology is undoubtedly powerful and is beneficial for local administration and public safety services. It has seen different levels of adoption, ranging from local municipalities to large scale nationwide deployments.

As real-time image recognition technology evolves and becomes cost-effective, the installation of monitoring devices in road networks is only expected to increase. There is already a circulating argument from privacy groups that the deployment of more and more vehicle tracking devices in disparate localities could finally amalgamate into a large scale mass surveillance network. While some networks are large enough that real time tracking is surmised to be possible (e.g. the UK's National ANPR Data Center [15]), other road networks across

much of the developed world are slowly being converted to smart infrastructures aided with traffic cameras and back-end vehicle identification systems [19]. In the midst of continuous deployments, it is often difficult to comprehend when the state of traffic monitoring technology crosses the thin line between public good and personal privacy.

In this paper, we present the first known study to quantitatively assess the mass tracking potential of traffic monitoring devices. We explore how the number and geographic locations of vehicle identification cameras (or trackers in general) impact the real time tracking of a moving vehicle on a road network. Since roadways are hierarchically designed, placing trackers in specific high usage junctions can add more value from a surveillance standpoint than others. In addition, we seek to obtain a quantitative characterization of the privacy implications of such tracking, primarily as related to the locations visited by an individual.

We first develop a tracker placement algorithm that favors nodes that are central to the fastest paths joining locations on a road network graph. In this context, there are the competing interests of placing trackers at frequently passed locations, and preventing the clustering of trackers in a small region. Second, we propose four metrics to evaluate the surveillance coverage of a specific set of trackers, their effectiveness in being able to continuously track a vehicle, and consequently how much location uncertainty remains about the destinations of a travel path. An empirical evaluation of the placement algorithm with respect to these metrics in a 100 square mile area of the Denver metropolitan area of Colorado, USA suggests that, with as few as 100 trackers, one can achieve an 80% coverage, and location uncertainties close to a mile. The location privacy risks are worrisome if the number of trackers crosses into the thousands — using 1000 trackers, 97.6% of tested paths could be tracked, with most destination nodes being traceable to within 2000 feet.

The remainder of the paper is organized as follows. We present background on the tracking methodology, and insights into the design of our tracker placement algorithm in Sections 2 and 3. Section 4 presents the four evaluation metrics. We present the experimental setup in Section 5, followed by empirical results in Section 6. Section 7 discusses some related work in the domain. Finally, we conclude the paper in Section 8 with a discussion on potential refinements.

2 Road Traffic Monitoring

A road network is modeled as a directed graph $G = \langle V, E \rangle$ consisting of a vertex (node) set V and an edge set E. A node is present for each road intersection. However, nodes are typically present between two intersections as well when the path connecting them is not a straight line. This helps maintain the shape of roadways when visualized as a planar graph, and also enables accurate computation of the distance between two intersection nodes. An edge represents a road segment between two adjacent nodes. Nodes can be annotated with positioning data, and edges can be annotated with length and speed limit. Fig. 1 depicts

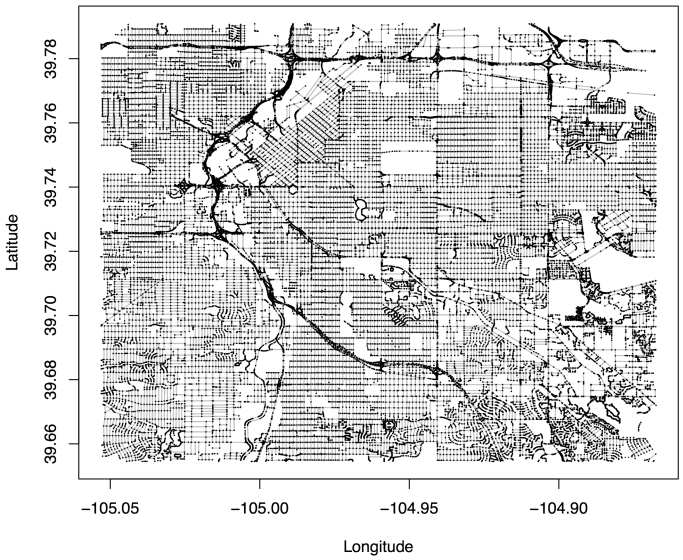


Fig. 1: An example road network graph.

a road network graph plotted based on the latitude and longitude positions of corresponding nodes.

Travel on a road network is often dictated by shortest time and fewer turns, instead of shortest length [8]. As such, a typical path between two nodes shows a hierarchy of road types, starting at local streets, joining on to higher capacity, but fewer, collector and arterial roadways, and finally to a limited number of expressways for long distance travel [1]. This pattern results in certain road segments being utilized with much higher frequency than others.

2.1 Traffic Monitoring

Several types of traffic cameras supplement a modern road transportation infrastructure. The most common of these is a traffic detection camera, deployed at intersections to detect the presence of traffic and accordingly activate traffic lights. Such cameras are monitored in real time by the local administration, who also decides whether the footage is recorded. The availability of cost-effective high-resolution cameras opens the possibility of performing other recognition tasks on the captured images, such as passenger face detection and vehicle identification (number plate). Another category of cameras that is deployed enforces different traffic regulations such as red lights, speed limits, restricted vehicle lanes, occupancy limits, and tollways. These cameras could be supplemented with additional hardware, such as a flash light or a radio transceiver, depending

on the task for which they are set up. Since a penalty is often levied on violators, the ability to perform vehicle identification is rudimentary in such deployments. An Automated License Plate Recognition (ALPR) system can perform such a task by automatically extracting a vehicle's plate number from a captured image, comparing it to one or more databases, and reporting or recording the results [7]. ALPR systems are known to be in use to look for stolen vehicles, or vehicles registered to persons of interest, gather intelligence on criminals, and track fleets of commercial vehicles, among others. They exist as locally managed small-scale systems, and also as nationwide deployments (e.g. UK's National ANPR Data Center). Not all ALPR camera locations are public domain knowledge due to the nature of their intended use cases. ALPR systems also have the potential to act as mass surveillance systems, although it is difficult to assess their effectiveness without knowing the camera locations.

2.2 Trackers and Tracker Activations

A primary objective of this work is to gather a preliminary assessment on the mass surveillance potential of traffic monitoring cameras. We abstract out the type of camera in use, and use the term *tracker* to mean any monitoring device that is capable of identifying a vehicle during transit. Formally, a tracker is a special node in the road network graph. A vehicle passing through a tracker node is then analogous to making an entry into a database with the vehicle's identity, the tracker's location (or identifier), and a timestamp. We refer to such an event as a *tracker activation*. Multiple activations can originate from a tracker at the same time instance, meaning that a tracker is fast enough to identify multiple high-speed vehicles passing through the node at the same time. We assume that trackers are omni-directional (works irrespective of the direction of approach towards the node). A vehicle traveling through a road network generates tracker activations as it passes through tracker nodes, effectively generating a timestamped trace of its locations on the road network.

2.3 Path Reconstruction

Given a road network $G = \langle V, E \rangle$, let the ordered set $P = \{v_1, v_2, ..., v_l\}$ represent a path of length l, where $v_i \in V$, for $i = 1...l$, and $v_i \rightarrow v_{i+1} \in E$, for $i = 1...(l-1)$. Here, v_1 is the source (start) and v_l is the destination (end) node of the path, also referred to as the boundary nodes of the path. Let $T \subseteq V$ designate a set of tracker nodes. Then, $P \cap T$ represents the trackers that are activated by path P. The ordering of elements in this intersection is done as per their ordering in P. Let $\tilde{P} = P \cap T = \{v_{t_1}, v_{t_2}, ..., v_{t_s}\}$ be the ordered set of activated trackers. v_{t_1} and v_{t_s} are the corresponding boundary nodes of this set.

An entity monitoring tracker activations will observe a time series of tracker activations for each vehicle. This series can be split into multiple subsequences by observing the time difference between two successive activations – if the time difference is much larger than the time required to travel between the two trackers, then it is reasonable to split the series at this point.

Given a set of activated trackers $\tilde{P}$ obtained after splitting as above, a monitoring entity can attempt to reconstruct P based on standard notions of optimal travel time. Let $\mathsf{op}(v, v')$ denote the optimal path from source v to destination v'. The optimal path between two nodes is often the fastest path, unless it is exceedingly longer than the shortest path. For the purpose of this study, we will mean the fastest path in all our uses of the op function. One way to reconstruct P from $\tilde{P}$ is to concatenate the optimal paths between successive trackers in $\tilde{P}$. The reconstructed path, denoted as P', is given as

$$P' = \mathsf{op}(v_{t_1}, v_{t_2}) \cup ... \cup \mathsf{op}(v_{t_{s-1}}, v_{t_s}),$$

where the union operations are order-preserving. Observe that if tracker activations are generated by vehicles following the fastest path from a source to a destination, then the reconstruction is simply $P' = \mathsf{op}(v_{t_1}, v_{t_s})$. The quality of a reconstruction is then dependent on the proximity of a vehicle's tracker activations to the source and destination nodes, which in turn is dependent on the placement of trackers in the road network.

3 Tracker Placement

A trivial method to ensure that a reconstructed path exactly matches the underlying true path is to convert every node in the network to a tracker. Thereafter, the tracker activations will indicate the exact path ($P' = \tilde{P} = P$), and no reconstruction is necessary. Another alternative would be to only convert the intersection nodes to trackers, and then calculate the optimal path between two trackers as the path containing no intermediate intersections. Clearly, both solutions are cost-prohibitive.

We explore the problem of placing trackers under the situation when the number of trackers, n_T, is pre-decided. Thereafter, given a road network graph G and a positive integer n_T, determine n_T nodes to be converted to trackers such that the accuracy of path reconstructions is maximized. Specific algorithms can be tailored to maximize specific accuracy metrics; however, our approach in this initial study is to develop a generic method free from any specific metric, and then evaluate its performance with respect to different metrics. The design of specific algorithms targeting specific objective functions is left for a future study.

3.1 Frequency Based Placement

One generic approach to place the trackers is to choose nodes in the decreasing order of their frequency of use in paths. The approach is promising due to the hierarchical nature of roadways, and the conformance of driving patterns to this hierarchy in terms of path selection [18]. The node frequencies can be obtained from structural properties of the road network, or through a sampling of traffic flow using temporary devices such as pneumatic road tubes. We consider the betweenness centrality measure of connected graphs for the former [3]. The betweenness value of a node v captures the inclusion of v in the optimal paths

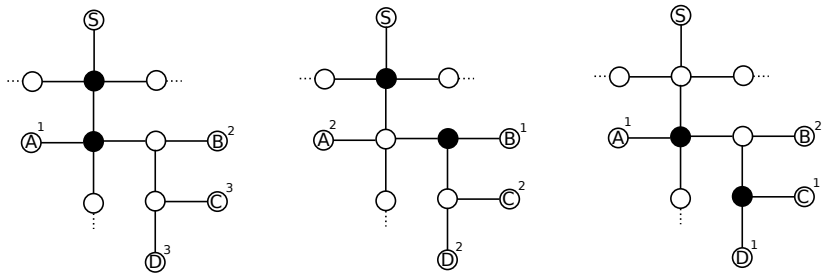


Fig. 2: Introducing separation between trackers (solid circles) can help improve reconstruction effectiveness.

between any pair of nodes. If $\mathcal{P}(v_a, v_b)$ represents all optimal paths from node v_a to v_b, then the betweenness value of a node v is given as

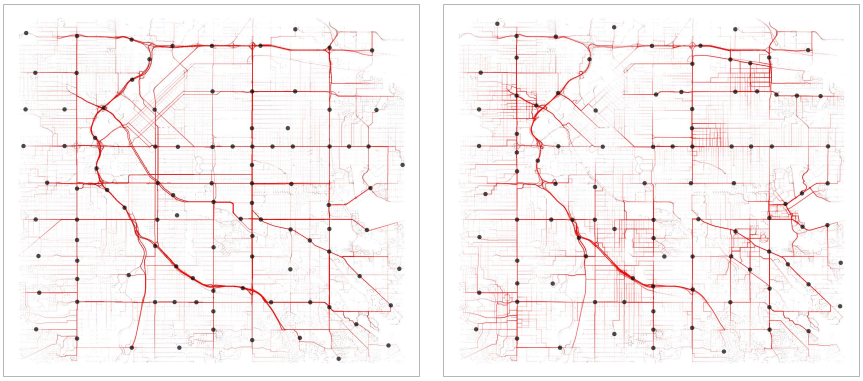
$$\beta(v) = \sum_{v \neq v_a \neq v_b} \frac{|\{P | P \in \mathcal{P}(v_a, v_b) \text{ and } v \in P\}|}{|\mathcal{P}(v_a, v_b)|}. \quad (1)$$

It can be reasonably assumed that a road network will be a connected graph. Hence, $|\mathcal{P}(v_a, v_b)| > 0$, for all v_a, v_b in the above formulation; otherwise, we ignore such pairs of nodes in the computation. Recall that we use the fastest path(s) as the optimal ones.

Node frequencies can also be obtained from available traffic count data. While betweenness accounts for paths between all pairs of nodes, and treats all such paths equally, frequency information derived from traffic counts can capture travel patterns specific to localities (e.g. a local store), central commercial locations (e.g. a business district in the city), and temporal variations throughout the day. In the following discussion, we will use the term *node frequency* to mean either betweenness or traffic count depending on the context.

3.2 Minimal Separation Between Trackers

Given the set of nodes ordered by their frequencies, the first n_T nodes can be converted to trackers so that tracker activations are triggered for a large number of paths. However, neighboring road network nodes tend to cluster together when sorted based on their frequencies. A high frequency node often has adjoining frequently used road segments. Fig. 2 illustrates the issue with a toy example. Consider the four paths originating in S and terminating in A, B, C and D. Assume that each hop (moving along an edge) takes constant time. Based on the four paths, if the two highest frequency nodes are converted to trackers (leftmost figure), then node A would be 1 hop away from a tracker activation, node B would be 2 hops away, and nodes C and D will be 3 hops away. However, the average hop count can be reduced by forcing the trackers to have some minimal separation, instead of being adjacent to each other. The center and right figures show the hop counts when using a separation of 2 hops between



(a) (b)

Fig. 3: Tracker placement using minimal separating distances and waves. Frequencies are from (a) betweenness values, and (b) traffic counts highlighting local visits.

trackers. Motivated by this observation, we choose nodes in decreasing order of their frequencies, but skip a node if a tracker already exists within some predefined distance of the node.

3.3 Waves

Minimal separating distances help disperse trackers throughout a road network. However, if the specified distance is too large, there is a possibility that the entire set of nodes is exhausted before n_T nodes are chosen as trackers. To alleviate this issue, we can restart the selection process on the remaining nodes using a smaller separating distance than in the previous iteration. Hence, our approach runs the selection process in waves (iterations) until n_T nodes have been chosen. The steps for the entire approach can be summarized as follows.

1. Let V_{sorted} be an ordered set of vertices sorted in decreasing order of their frequencies, $T = \phi$, $\boldsymbol{d} = (d_1, d_2, ..., d_w = 0)$ be a sequence of decreasing values to be used as separating distances, and dist be a distance function between nodes. Let $wave = 1$ and $V_{rejected} = \phi$.
2. Remove the first node $v \in V_{sorted}$. Add v to T if for all $v' \in T$, $\mathsf{dist}(v, v') > d_{wave}$; otherwise add v to $V_{rejeted}$ (tail end).
3. If $V_{sorted} = \phi$, increment $wave$ by 1, set $V_{sorted} = V_{rejected}$ and then $V_{rejected} = \phi$.
4. If $|T| < n_T$, repeat from step 2.

Fig. 3 illustrates the result of placing 100 trackers on a road network graph of the Denver metropolitan area in Colorado, USA. In both plots, frequently used roadways are depicted using a darker shade. For Fig. 3a, the fastest paths between all

pairs of nodes have been considered. As a result, all highways and arterial roadways have become prominently visible. For Fig. 3b, travel paths are to a nearby local business. Therefore, specific local and collector roads have gained prominence. The dispersion of trackers provided by the minimal separation method is clearly visible in both instances.

4 Evaluation Metrics

We evaluate the effectiveness of our approach by computing various metrics on the reconstructed paths corresponding to a set $\mathcal{P}_{test}$ of test paths. Given a set of tracker nodes T, we compute the reconstructed path P_i' for each test path $P_i \in \mathcal{P}_{test}$ as detailed in Section 2.3.

4.1 Reconstruction Coverage

An ill-placed set of trackers will fail to overlap with most paths. As a result, no activations will be triggered, thereby leading to no reconstruction. For mass surveillance, the more reconstructions that a set of trackers can effectuate, the better is the underlying placement. Reconstruction coverage captures this aspect as the fraction of test paths for which a reconstruction is possible.

$$Coverage = \frac{|\{P|P \in \mathcal{P}_{test} \text{ and } P \cap T \neq \phi\}|}{|\mathcal{P}_{test}|}. \tag{2}$$

Coverage does not consider the quality of a reconstruction. Even a single tracker activation is considered a successful tracking event under this metric, although no path reconstruction is possible with a single tracker activation.

4.2 Path Coverage

The path coverage metric measures the fraction of the total distance in a path $P_i \in \mathcal{P}_{test}$ that has been accurately covered by the corresponding reconstructed path P_i'. If $P_i = \{v_{i1}, v_{i2}, ..., v_{il}\}$ and $P_i' = \{v_{i1}^{'}, v_{i2}^{'}, ..., v_{il'}^{'}\}$, then

$$Path\,Coverage(P_i) = \frac{\sum_{k=1}^{l'-1} \mathsf{dist}(v_{ik}', v_{i(k+1)}')}{\sum_{k=1}^{l-1} \mathsf{dist}(v_{ik}, v_{i(k+1)})}. \tag{3}$$

This metric is appropriate when the test path P_i is also an optimal path as per the op function used during the path reconstruction. When using fastest paths, we clearly have $P_i^{'} \subseteq P_i$ and the metric signifies the fraction of travel distance for which the vehicle's location can be accurately tracked in real time.

4.3 Boundary Activation Distance

We often associate a higher privacy value to the locations and neighborhoods that we visit, instead of the path we take to arrive at such destinations. While high frequency nodes operating as trackers can provide good reconstruction

and path coverage, the associated privacy risks can be minimal if a path's source and destination are difficult to infer from a reconstruction. As such the boundary activation distance metric quantifies the average distance between the two corresponding boundary nodes in the real and the reconstructed paths. If $P_i = \{v_{i1,}v_{i2}, ..., v_{il}\}$ and $P_i' = \{v_{i1,}^{'}v_{i2}^{'}, ..., v_{il'}^{'}\}$, then

$$Boundary\ Activation(P_i) = \frac{1}{2}\big(\mathsf{dist}(v_{i1}, v_{i1}') + \mathsf{dist}(v_{il}, v_{il'}')\big). \tag{4}$$

Tracker activations happening closer to the source and destination nodes of a path will result in a smaller boundary activation distance, thereby providing smaller areas of uncertainties on the source and destination of a path.

4.4 Tracker Period

The real time tracking potential of a set of trackers can be characterized in a manner similar to that in the boundary activation distance metric. Instead of considering the location uncertainty only at the boundary nodes, the tracker period metric considers the uncertainty along the entire path as the average distance traveled between two successive tracker activations. If $P_i \cap T = \{v_{it_1}, v_{it_2}, ..., v_{it_s}\}$ are the activated tracker nodes, then

$$Tracker\ Period(P_i) = \frac{1}{s-1}\sum_{k=1}^{s-1} \mathsf{op}_{\mathsf{len}}(v_{it_k}, v_{it_{k+1}}), \tag{5}$$

where $\mathsf{op}_{\mathsf{len}}(v_a, v_b)$ is the length (travel distance) of the optimal path from node v_a to v_b. A well dispersed set of trackers will generate a consistent tracker period for a majority of the test paths.

Average values for path coverage, boundary activation distance and tracker period can be computed over the test paths in $\mathcal{P}_{test}$. In addition, summary statistics such as median and quartiles are useful in observing the variation in the metrics' values over different paths.

5 Experimental Setup

We perform the empirical evaluation of our approach on a road network graph spanning an approximately 100 square mile area of Denver, Colorado, USA (Fig. 1). The graph spans between latitudes 39.654518^oN and 39.790931^oN, and longitudes 105.053195^oW and 104.867402^oW. It consists of 40,253 vertices and 83,599 directed edges. Each node is labeled with its latitude and longitude coordinates. Each edge is labeled with the geodesic distance between the two nodes, a road segment type value, and the speed limit on the road segment. We compute the time required to travel a road segment (edge) by dividing the distance by the speed limit. Distances between nodes (the dist function) are computed using the Vincenty inverse formula for ellipsoid, which is available as the gdist function in the lmap R package. The implementation is done as a single threaded R application, running on a laptop with a 3.1 GHz Intel Core i7 processor, 16 GB

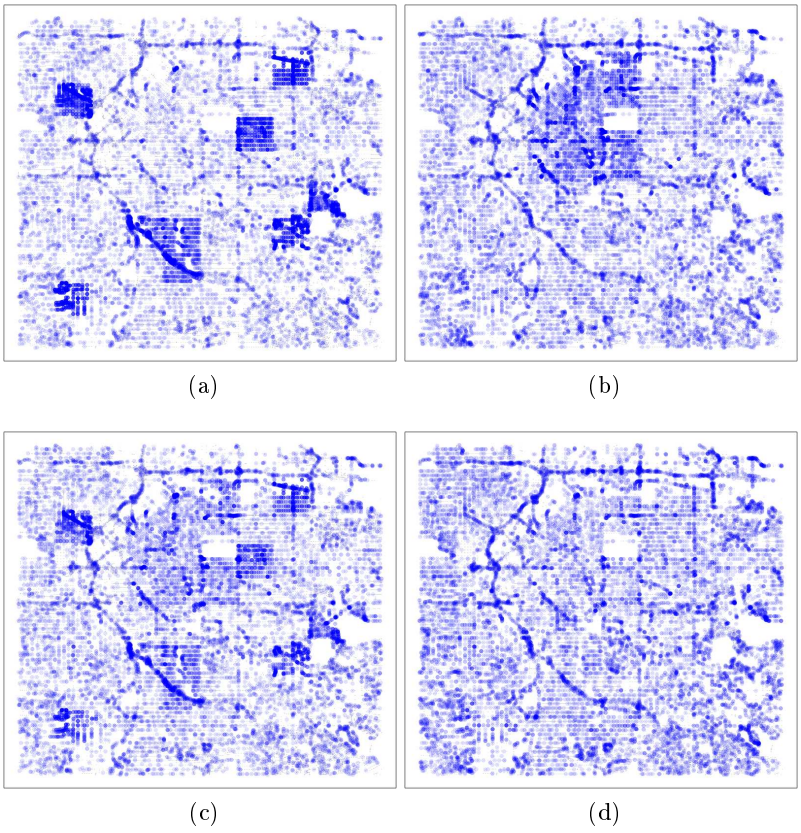
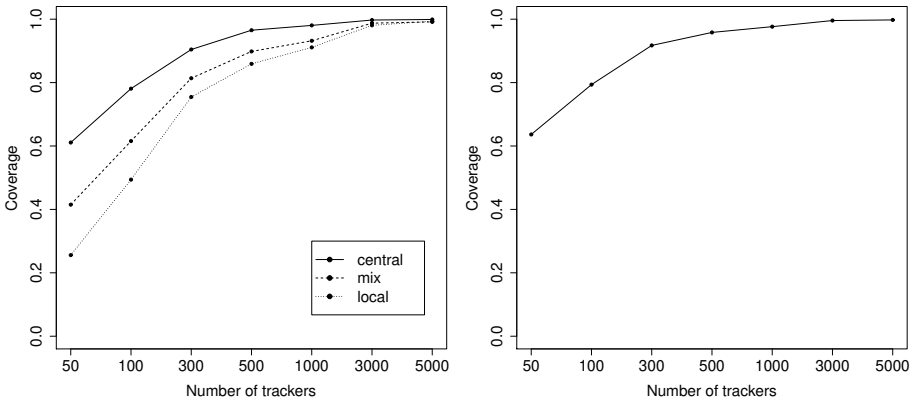


Fig. 4: Source and destination nodes of 10,000 test paths in four different generation models — (a) local (b) central (c) mix (d) free. See text for model descriptions.

memory, and OSX 10.10.5. Graph operations, such as finding fastest paths and betweenness centrality values, are performed using the igraph 1.2.1 R package.

We consider four probabilistic node selection models, and generate 10,000 test paths from each. For test path generation, we divide the graph area into a 10×10 grid of cells, and assign a probability of selection to each cell. Nodes are chosen by first choosing a cell as per the assigned probabilities, and then randomly picking a node within the chosen cell. In the *free* model, equal probabilities are assigned to each cell, and source/destination pairs are chosen accordingly. For the *central* model, specific cells in the Denver business district are assigned a significantly higher probability of selection. Once a source (or destination) node is chosen based on these probabilities, the other node is chosen similar to as in the free model. This results in most paths being directed towards, or away from, a central area in the city. In the *local* model, few cells containing



(a) Using training frequency counts (b) Using betweenness on free model

Fig. 5: Coverage evaluation.

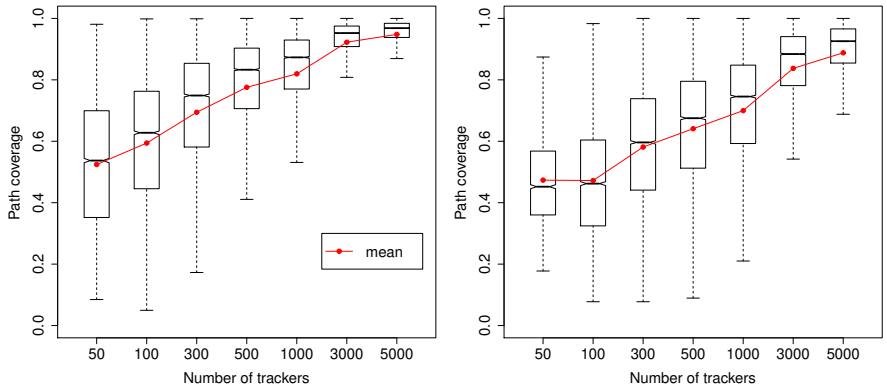
locally prominent shopping stores are first identified. A source node is chosen as in the free model. The destination node is then chosen from the closest locally prominent cell. Finally, the *mix* model contains paths of types generated in both the central and the local models. The sampled source and destination nodes of the paths in the four models are depicted in Fig. 4. The fastest path between a chosen source and destination pair is taken as the respective test path. The generated paths are usually the same as obtained from road navigation services.

We implemented our approach to place trackers based on both the betweenness centrality measure, as well as traffic counts. In the absence of real traffic count data, we obtained the counts for the latter from a set of training paths generated in a manner similar to that of test paths. We generated separate sets of 10,000 training paths from each of the local, central and mix models, and computed the node frequencies in each case from the corresponding training paths set. Unless stated otherwise, the free model test paths are used to evaluate the trackers placed based on betweenness values, while the local, central and mix model test paths are used to evaluate trackers placed based on frequency counts computed from the respective training paths.

We present results for tracker counts (n_T) of 50, 100, 300, 500, 1000, 3000, and 5000, using a separation distance vector of $\boldsymbol{d} = (1, 0.5, 0.25, 0.1, 0)$ mile for all but the case of $n_T = 50$. For $n_T = 50$, we start with a separation distance of 2 miles, followed by the listed values. Where appropriate, metrics are computed only on the test paths for which a reconstruction is possible.

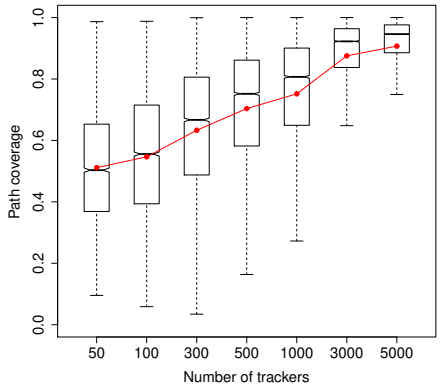
6 Results

In the following, we first report results on the mass surveillance potential of the tracker placement approach, and later discuss some parameter influence.

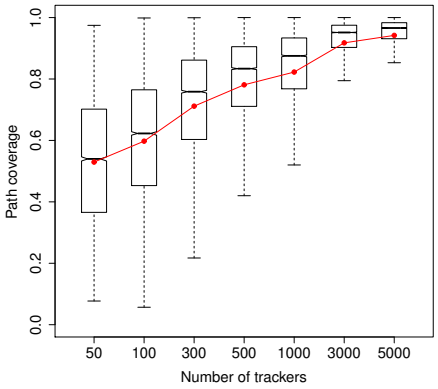


(a) Central model

(b) Local model



(c) Mix model



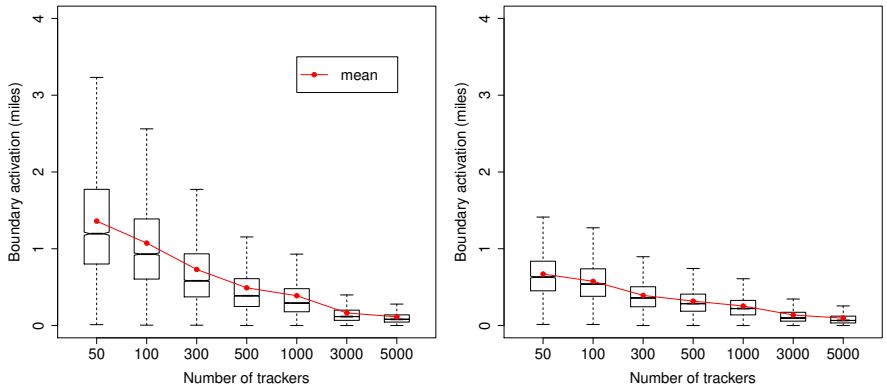
(d) Free model

Fig. 6: Path coverage evaluation.

6.1 Mass Tracking

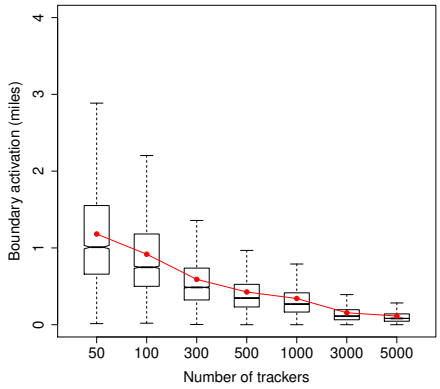
Reconstruction coverage. Fig. 5 shows the reconstruction coverage of the placement algorithm in the four models. Except in the local model, some form of reconstruction becomes possible for 80% or more of the test paths with 300 or more trackers. The coverage is also reasonably good in the mix and free models for as few as 100 trackers. The local model requires more trackers to achieve coverage similar to the central or mix models. This is expected since traffic is more likely to be restricted to localities, and hence trackers have mostly local utility. A tracker has better overall utility when it can be on the path of local as well as centrally directed traffic, as in the central and mix models.

Path coverage. The path coverage statistics on the fraction of covered test paths (reconstruction possible) for each model are depicted in Fig. 6. We report summary statistics using a boxplot, which shows the minimum, first quartile,

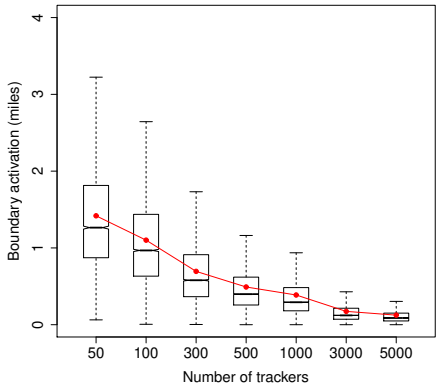


(a) Central model

(b) Local model



(c) Mix model

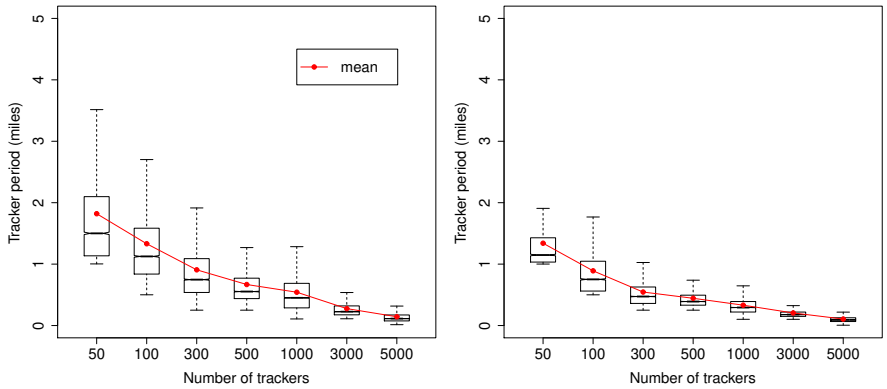


(d) Free model

Fig. 7: Boundary activation distance evaluation.

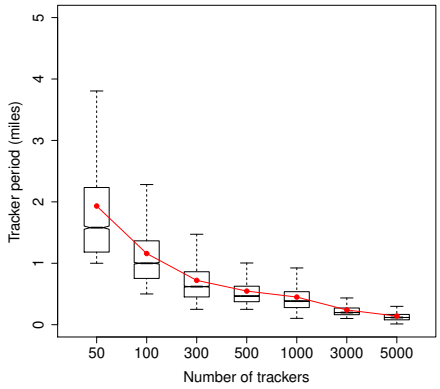
median, third quartile, and maximum values. In general, the results for the local model lie somewhere between that of the central and mix models, with the free model statistics being similar or better than those in the mix model. An average path coverage of 60-70% is achieved with 300 trackers, with 75% of the test paths being reconstructed with at least a 50% path coverage (the local model is slightly lower). While the minimum and maximum values fluctuate, the quartile boundaries show a general increasing trend with increasing number of trackers.

Boundary activation distance. From the standpoint of source and destination privacy, the boundary activation distance is a critical measure. Path source and destination is within an average of less than a mile from the closest tracker activations when using 300 trackers. The uncertainty reduces to half a mile or less with 500 trackers. The interquartile range in most cases is itself half a mile or less, indicating that the metric's value even with the observed varia-

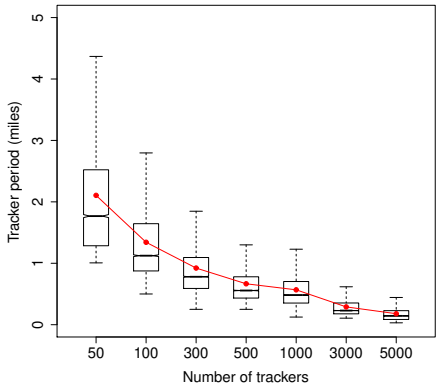


(a) Central model

(b) Local model



(c) Mix model



(d) Free model

Fig. 8: Tracker period evaluation.

tions is concernedly low. It can be seen that the metric has much lower values in the local model compared to the other models. This is because trackers in this model locally serve a smaller targeted region, and the source nodes are often in the close vicinity.

Tracker period. The uncertainty in continuous tracking is depicted in terms of the tracker period metric in Fig. 8. Although the values are slightly larger than the boundary activation distances, the trends are similar in nature. Effectively, tracker activations happen every mile or less (on an average) with 300 or more trackers. The quartile values are indicative of a good dispersion of trackers across the entire area.

In summary, the evaluation indicates that trackers placed based on node frequencies have a significant mass tracking potential with as few as 100 or 300 trackers. A vehicle can also be localized to a small region at all times, which can

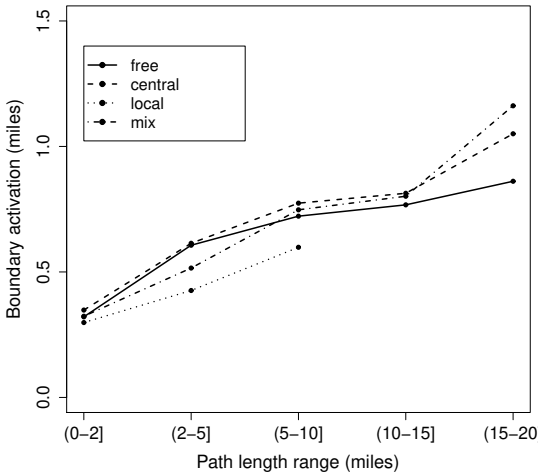


Fig. 9: Path length versus boundary activation distance when using 300 trackers.

have adverse implications in terms of location privacy. The potential is amplified almost to the extent of targeted tracking when deploying more than 1000 trackers is no longer cost-prohibitive. Comparatively, using the betweenness centrality measure (as in the free model evaluation) provides similar effectiveness as when using traffic counts. The advantage is that betweenness is a structural property of the graph, and does not require sample data of traffic counts. Hence, this approach is more generic and attractive. Of course, to be robust, the betweenness approach should perform reasonably well even when test paths follow specific distributions, as in the central, mix, or local models (Section 6.3).

6.2 Impact of Path Length

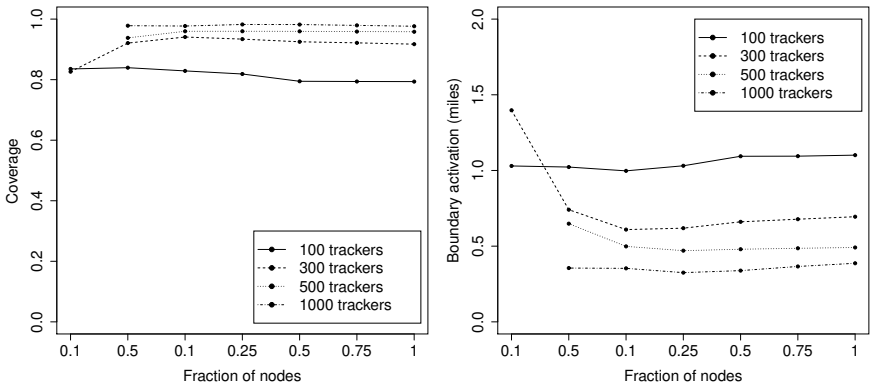
The test paths used in the evaluations range from 0-10 miles in the local model, and 0-20 miles in the other three models. It is imperative to ask if the boundary activation distance statistics have been skewed due to significantly differing performance in long and short paths. Fig. 9 depicts the mean boundary activation distance in subsets of test paths grouped by path length. While we observe a tendency for the boundary activation distance to slowly grow with longer paths, the absolute values themselves are within half a mile of the collective mean. We do not consider this variation to be significant enough that it can alleviate any pertinent privacy concerns when traveling longer distances.

6.3 Impact of path distributions

Table 1 shows the coverage and boundary activation distance means/variances resulting from evaluations using the set of test paths from each of the four models when 300 trackers are placed based solely on the betweenness centrality

Table 1: Performance of the betweenness approach with 300 trackers on different model specific test paths.

	Boundary activation distance		Coverage
	mean (miles)	variance	
free	0.69	0.21	0.92
central	0.72	0.22	0.91
local	0.50	0.08	0.55
mix	0.64	0.18	0.72



(a) Coverage evaluation.

(b) Boundary activation distance evaluation.

Fig. 10: Impact of considering a reduced set of nodes for tracker placement.

measures. We have already observed that the approach provides more than 90% coverage and less than a mile of boundary activation distance on test paths generated uniformly at random (free model). The results here indicate that the betweenness approach performs at par with an approach executed with model specific traffic counts data. Comparing the results in Table 1 and those in Figs. 5 and 7 (300 trackers), we can see that the metric values are almost identical, except for a lower coverage in case of the local model's test paths.

6.4 Candidate Nodes for Tracker Placement

The betweenness centrality values (and frequency of node usage in general) of a road network follow a heavy tailed distribution. For example, in the Denver area graph we use here, 75% of the sum of the betweenness values comes from that of only 12.5% of the most frequently used nodes in the network. Since our approach enforces minimal separation between trackers, it is possible that, within a wave, nodes that are insignificant in terms of their centrality in fastest paths gets chosen as trackers. It therefore begs the question of whether tracking performance

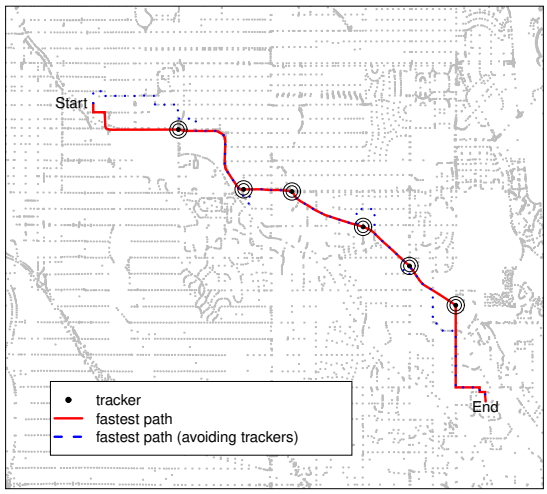


Fig. 11: A tracking avoidance example.

Table 2: Impact of avoiding trackers on travel time and connectivity.

(a) Mean percentage increase in travel time

	Number of trackers						
	50	100	300	500	1000	3000	5000
free	12.9	20.5	32.8	47.7	60.3	100.1	108.9
central	11.6	19.4	30.7	45.6	53.8	103.9	50.2
local	17.2	26.3	39.0	62.2	99.7	200.4	61.9
mix	12.1	19.3	32.9	50.3	62.4	97.2	45.9

(b) Percentage of test paths with no avoidance path

	Number of trackers						
	50	100	300	500	1000	3000	5000
free	2.4	4.5	9.1	18.8	28.0	94.4	99.7
central	2.0	3.8	7.5	15.9	26.9	96.1	99.8
local	3.3	4.1	12.7	16.4	30.0	85.9	99.1
mix	1.4	3.4	10.0	17.9	27.1	91.1	98.9

can be improved by restricting the choice of trackers to few most frequently used nodes only. Fig. 10 depicts the coverage and mean boundary activation distance on the free model test paths when trackers are chosen from the top 1%, 5%, 10%, 25%, 50%, 75% and 100% nodes sorted by their betweenness values. Since the total number of nodes is 40,253, results are unavailable when restricting to the top 1% of nodes while requiring 500 or 1000 trackers. We observe minor, but insignificant, improvements when restricting to the top 10-25% of the nodes. Clearly, restricting to too few nodes will nullify the advantages of the minimal separation approach owing to the iterative reduction of the separation distance.

6.5 Tracking Avoidance

As the final experiment, we explored how the travel time will be affected if tracker locations are known, and an attempt is made to travel along the fastest path that avoids the tracker nodes. As shown in Fig. 11, an avoidance path could simply

include rerouting in the vicinity of the tracker nodes, or separate (potentially slower) subpaths altogether. Depending on the source and destination of a path, it is also possible that an avoidance path is impossible since a node critical to the connectivity between the boundary nodes has been marked as a tracker. Table 2 shows the average percentage increase in travel time of the various test paths in the four models (trackers placed by the default approaches). We observe an average of ≈30% increase in travel time with 300 trackers in place. In addition, ≈10% of the test paths have no avoidance paths. The travel time may not be too bad, but the path itself can be inconvenient on a regular basis! The values become significantly worse for 1000 or more trackers.

7 Related Work

The issue of privacy in automated traffic enforcements has received significant attention from academicians and civil liberty groups alike. Blumberg et al. highlighted that automated vehicle identification systems should be designed to cater no more information than what is necessary to enforce traffic laws [2]. To such an end, they proposed a camera-free protocol for traffic monitoring where the identity of a vehicle can be learned only if it violated a traffic law. The Australian Privacy Foundation argued that automated vehicle identification can be inaccurate, and thereby result in unreasonable, embarrassing, or even dangerous, conclusions for law-abiding citizens [4]. The effect of such social implications can be much greater than the possible deterrence that the technology can have on criminals. Besides, the ability to link vehicle sightings to travel patterns can expose individuals to malicious uses such as theft and discrimination, in addition to creating the avenue for personal habits to be scrutinized in a legal proceeding [5]. The absence of national standards and policies for transportation data storage and access can make it a challenging task to assure privacy and accountability to the traveling public. A key challenge that has been identified in this context is to succinctly define terms such as "casual observation" and "targeted surveillance" [9]. Irrespective of the numerous privacy preservation proposals offered, adoption has been minimal, and the threat continues to grow [6,17].

The usage of centrality measures to capture traffic flows has been explored earlier in the context of urban transportation planning. Kazerani and Winter argued that human agents demonstrate travel behavior that cannot be captured with topological characteristics of a road network alone. As such, centrality measures based on traditional notions of shortest path are insufficient [12,13]. Further, sources and destinations of travel are not uniformly distributed, and change over time, which contradicts the premise behind measures such as betweenness. We partially observed the impact of these dynamics in our study; albeit, the advantages of pursuing non-generic measures was minimal in mass tracking, at least when working with fixed trackers only. In fact, in a large scale study involving 360,000 San Francisco Bay area users and 680,000 Boston area users spanning a three week period, nodes with the top 25% of betweenness and degree values were reported to be of topological importance from a driving standpoint [18].

Park and Yilmaz observed that node centrality in downtown versus residential locations can be quite different owing to the presence of alternative pathways in the grid-like design of the former [16]. Gao et al. also reported a similar conclusion when attempting to understand urban traffic-flow characteristics [10]. As a consequence in this work, attempts to monitor traffic inside a core urban area will require the placement of more trackers. Combining multiple centrality measures to understand traffic flow can also be useful [11].

The closest work with respect to determining ideal tracker locations on a road network is in a recent study by Ma et al. [14]. Their approach operates on a grid of blocks overlaid on the road network, and identifies prominent blocks based on factors such as number of unique vehicles crossing a block, amount of vehicle traffic, time when a vehicle is out of surveillance, and average camera-hit intervals. The placement strategy is tied to a set of already available GPS traces — a dependency that we sought to avoid in this study.

8 Conclusion and Future Work

Through this work, we have demonstrated that existing traffic monitoring capabilities hold the potential to support a high-precision mass surveillance network without the need for an excessive level of investment. Our node tracker placement algorithm can determine critical locations on a road network that are often parts of high traffic routes, and also disperse trackers over a region to enable a large coverage. Results suggest that more disclosure is needed in how and where traffic monitoring locations are chosen in an evolving infrastructure so as to ensure that evaluations of the nature performed here can be carried out to quantitatively assess potential privacy risks.

In this work, we approached tracker placement as a static problem. However, traffic patterns change at different times of the day, creating a varying distribution of node usage frequencies. Although we have demonstrated that the betweenness approach adapts well to specific distributions, it remains to be evaluated if tracking capabilities are enhanced when mobile trackers can be deployed. Mobile trackers can change locations depending on changing traffic patterns and have the potential to address low coverage areas dynamically. What mix of fixed and mobile trackers can provide an attractive solution is an interesting direction to explore.

Traffic cameras are currently deployed mostly in accident prone and high crime areas. Our placement algorithm targets good surveillance possibility. With access to a database of cameras with their deployed locations, we can assess where the current mass tracking capabilities stand in a region of interest. Such a study can hopefully provide a much needed quantitative dimension to the debate on service versus privacy in the transportation infrastructure.

References

1. American Association of State Highway and Transportation Officials: A Guide

for Achieving Flexibility in Highway Design, chap. 1.4.1 Functional Classification (2004)
2. Blumberg, A.J., Keeler, L.S., Shelat, A.: Automated traffic enforcement which respects "driver privacy". In: Proceedings of 2005 IEEE Intelligent Transportation Systems. pp. 941–946 (2005)
3. Brandes, U.: A faster algorithm for betweenness centrality. Journal of Mathematical Sociology **25**(2), 163–177 (2001)
4. Clarke, R.: The covert implementation of mass vehicle surveillance in Australia. In: Proceedings of the Fourth Workshop on the Social Implications of National Security. pp. 45–59 (2009)
5. Cottrill, C.D.: Examining privacy and surveillance in urban areas: A transportation context. In: Proceedings of 3rd Hot Topics in Privacy Enhancing Technologies. pp. 1–13 (2010)
6. Crump, C.: You are being tracked: How license plate readers are being used to record Americans' movements. Tech. rep., American Civil Liberties Union (2013)
7. Du, S., Ibrahim, M., Shehata, M., Badawy, W.: Automatic license plate recognition (ALPR): A state-of-the-art review. IEEE Transactions on Circuits and Systems for Video Technology **23**(2), 311–325 (2013)
8. Duckham, M., Kulik, L.: "Simplest" paths: Automated route selection for navigation. In: Proceedings of the International Conference on Spatial Information Theory. pp. 169–185 (2003)
9. Fries, R.N., Gahrooei, M.R., Chowdhury, M., Conway, A.J.: Meeting privacy challenges while advancing intelligent transportation systems. Transportation Research Part C: Emerging Technologies **25**, 34–45 (2012)
10. Gao, S., Wang, Y., Gao, Y., Liu, Y.: Understanding urban traffic-flow characteristics: A rethinking of betweenness centrality. Environment and Planning B: Urban Analytics and City Science **40**(1), 135–153 (2013)
11. Jayasinghe, A., Sano, K., Nishiuchi, H.: Explaining traffic flow patterns using centrality measures. International Journal for Traffic and Transport Engineering **5**(2), 134–149 (2015)
12. Kazerani, A., Winter, S.: Can betweenness centrality explain traffic flow? In: Proceedings of the 12th AGILE International Conference on Geographic Information Science. pp. 1–9 (2009)
13. Kazerani, A., Winter, S.: Modified betweenness centrality for predicting traffic flow. In: Proceedings of the 10th International Conference on GeoComputation. pp. 1–5 (2009)
14. Ma, X., He, Y., Luo, X., Li, J., Zhao, M., An, B., Guan, X.: Vehicle traffic driven camera placement for better metropolis security surveillance. IEEE Intelligent Systems p. preprint (2018)
15. Mandeville, B.: Automatic number plate recognition (ANPR) strategy 2016 - 2020. Tech. rep., National Police Chef's Council, UK (2016)
16. Park, K., Yilmaz, A.: A social network analysis approach to analyze road networks. In: Proceedings of the ASPRS Annual Conference (2010)
17. Schwartz, A.: Chicago's video surveillance cameras: A pervasive and poorly regulated threat to our privacy. Nortwestern Journal of Technology and Intellectual Property **11**(2) (2012)
18. Wang, P., Hunter, T., Bayen, A.M., Schechtner, K., González, M.C.: Understanding road usage patterns in urban areas. Nature Scientific Reports **2**(1001) (2012)
19. Wikipedia: Automatic number-plate recognition. https://en.wikipedia.org/wiki/Automatic_number-plate_recognition#Usage (August 13 2018)