# The 11th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2020) 

# Traffic data collection using active mobile and stationary devices 

Johan Holmgren ${ }^{\text {a }}$, Henrik Fredriksson ${ }^{\text {b,* }}$, Mattias Dahl ${ }^{\text {b }}$<br>${ }^{a}$ Department of Computer Science and Media Technology, Malmö University, Malmö, 20506, Sweden<br>${ }^{b}$ Department of Mathematics and Natural Sciences, Blekinge Institute of Technology, Karlskrona, 37179 Sweden


#### Abstract

In this paper, we study the complementary characteristics of stationary and mobile devices for traffic data collection. Since stationary devices continuously collect traffic data at fixed locations in a network, they can give insight of the traffic at particular locations over a longer period of time. Mobile devices have wider range and are able to collect traffic data over a larger geographic region. Thus, we argue that both types of technology should be considered to obtain high-quality information about vehicle movements. We present a traffic simulation model, which we use to study the share of successfully identified vehicles when considering both stationary and mobile technologies with varying identification rate. The results of our study, where we focus on freight transport in southern Sweden, confirms that it is possible to identify the majority of vehicles, even when the identification rate is low, and that the share of identified vehicles can be increased by using both stationary and mobile measurement devices.


© 2020 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0) Peer-review under responsibility of the Conference Program Chairs.

Keywords: Traffic data collection; stationary devices; mobile devices; traffic simulation

## 1. Introduction

Up-to-date and accurate data about vehicle movements and traffic flows are important in order to support the operation of traffic authorities [1]. Even though large amounts of resources are spent collecting and analyzing traffic data, there is typically a knowledge gap on how individual vehicles travel over larger areas. For example, there is often a need to gain deeper knowledge on the movements of trucks carrying dangerous goods, and to track vehicles traveling on networks where road user charging is applied [2,3]. The knowledge gap is to a large extent a consequence of the inability to identify vehicles of the dominating types of technologies that are currently in use for collecting traffic data.

In the current paper, we study how active stationary and mobile measurement devices, can be combined to analyze the movement of vehicles in a road transport network. By active measurement devices, we mean devices that are able

[^0]to detect and record some unique vehicle identity, for instance license plate number. Stationary and mobile devices have complementary characteristics, and we argue that they should be used in conjunction with each for traffic data collection. In particular, we study how probe vehicles can complement already installed stationary devices to monitor the movements of trucks. To this end, we utilize a traffic simulation model that enables us to reproduce known traffic flows on individual roads, based on random walks [4]. The simulation model is used to study freight transport in a test network in southern Sweden to evaluate the idea of using active stationary and mobile devices for data collection. The proposed model allows us, in a cost-effective way, to investigate different scenarios with varying identification rate to gain insight into how stationary and mobile devices complements each other.

The remainder of the paper is organized as follows. In Section 2, we discuss the key properties of stationary and mobile devices for traffic data collection. Section 3 describes the mathematical model of our traffic simulation model, which we used in order to evaluate our idea on traffic data collection in the simulation study presented in Section 4. Finally, we provide some concluding remarks in Section 5.

## 2. Traffic data collection using active stationary and mobile measurement devices

The basic functionality of both active stationary and active mobile devices for traffic data collection is that they should be able to detect vehicles, identify them, and communicate the collected data to the responsible authority [5]. In this section, we discuss some important properties of stationary and mobile traffic measurement technologies for the collection of detailed traffic data. We argue that the two properties of geographic coverage, which is higher for probe vehicles and the possibility to successfully detect and identify vehicles, which we argue is better for stationary devices, makes it relevant to consider both types for traffic data collection data [6].

As stationary devices are placed at specific locations in a network, they are only able to observe those vehicles that pass the device. Some types of technologies can only observe vehicles driving in one direction of bidirectional roads. This is the case for traditional cameras, which are typically only able to observe vehicles in the direction they point. Since stationary devices are permanently operating in the same location, the data they collect can give insight and data to discover patterns of the traffic flow at one particular location over time. As probe vehicles drive around in the network, they can observe vehicles wherever they happen to drive [7]. Therefore, mobile devices have significantly larger geographic coverage than stationary devices [8]. On the other hand, probe vehicles cannot collect detailed data about traffic at a specific location over time.

The identification rate is the ability of the device to successfully detect and identify a vehicle. For example, for camera-based stationary devices, a high identification rate requires that there are no obstacles between the camera and the vehicle to be observed, and it needs to be sufficiently close to the vehicles to be able to detect them. For probe vehicles, the identification rate is to a large extent influenced by the characteristics of the road, including the number of lanes and the traffic density. Since probe vehicles drive around in the network, the possibility to control the external conditions is lower than for stationary devices. Probe vehicles need to be able to identify vehicles under conditions that are often less favorable than the conditions under which stationary devices need to operate, such as disadvantageous weather conditions [9]. Stationary devices, on the other hand, can be placed above the traffic, so that they are able to detect vehicles even if the traffic density is high.

Stationary devices are expensive since they need to be placed along the roads, which means that they may require both building permits, and operating personnel [10]. We would like to emphasize on the possibility of using already existing devices for data collection. For example, speed cameras could serve multiple purposes, by installing software for data collection. On the other hand, we argue that mobile devices are more economical. For example, cell phones are typically equipped with technologies suitable for vehicle detection, such as camera, Global Positioning System, and Bluetooth $[10,11]$. This enables privately owned vehicles to serve as probe vehicles; however, the willingness of private actors to use their vehicles as probe vehicles depends to a large extent on their incentives. A possible incentive is the will to contribute with data that can be used to analyze traffic, and in the longer perspective, contribute to the development of the transport network. Based on this discussion, we argue that the cost per device for mobile devices in general will be lower than the cost for stationary devices.

The number of devices to use for traffic data collection depends on the budgetary constraints of building and maintaining stationary devices, as well as the incentives to use privately owned vehicles as probe vehicles. Since we expect the cost for stationary devices to be relatively high, we believe that it will not be possible to build many new
stationary devices. However, as previously discussed, utilizing the possible synergies between different ways of using technologies currently in use, may reduce the need for new stationary devices. The number of mobile devices mainly depends on the incentives of drivers to let their vehicles act as probe vehicles.

## 3. The simulation model

Consider a transport network $(N, A)$, where $N$ is the set of nodes and $A$ is the set of links. For each link $a_{k}$, we let $\tau_{k}$ be the travel time of the link, and $\varphi_{k}(t)$ be the flow of vehicles on that link at time $t$. The set of outgoing and incoming links to a node $n$ is denoted by $O_{n}$ and $I_{n}$ respectively. In the model, we utilize hourly flows over a 24 -hour period; hence, we use periodic step functions to describe $\varphi_{k}(t)$. The functions $\varphi_{k}(t)$ are constant over a one-hour interval and have an overall period of 24 hours. Since the functions $\varphi_{k}(t)$ are constant during one-hour intervals, we may without loss of generality assume that $t$ is a discrete integer variable. We assume that the time from $t_{0}$ until the next time a vehicle enters $a_{k}$ is a random variable $X_{k}$ with the cumulative distribution function $F_{X_{k}}\left(t \mid t_{0}\right)=\int_{t_{0}}^{t} f_{X_{k}}(\xi) d \xi$, which yields the distribution $F_{k}\left(t \mid t_{0}\right)=1-\exp \left(-\int_{t_{0}}^{t} \varphi_{k}(x) d x\right)$. We assume that the number of vehicles passing a node during a time interval of length $T$ is Poisson distributed with expected value $\lambda=T \sum_{o_{k}} \varphi_{k}$ [12]. The expected number of departures from node $n$ during hour $t$ is $\delta_{n}(t)=\sum_{I_{k}} \varphi_{k}(t)$, and the time between departures is exponentially distributed.

A reasonable assumption, as well as a requirement for the model to work correctly, is that the incoming flow to a node is the same as the outgoing flow, albeit with some time delays based on $\tau_{k}$. In particular, for all $n \in N$, and all $t \in \mathbb{R}$ it is required that

$$
\begin{equation*}
\sum_{a_{k} \in I_{n}} \varphi_{k}\left(t-\tau_{k}\right)=\sum_{a_{k} \in O_{n}} \varphi_{k}(t) . \tag{1}
\end{equation*}
$$

The left-hand side of the equation is the flow going into node $n$ at time $t$, and the right-hand side is the flow going out from node $n$ at time $t$. From a logistic point of view, it is difficult to measure the traffic flow of a transportation network over multiple points simultaneously over a 24 -hour period. This may cause deviations in the flow data such that nodes have larger inflow than outflow or vice-versa over a 24 -hour period. The data ${ }^{1}$ used in our study did not satisfy equation (1) and needed calibration. Calibration of the data to satisfy equation (1) for each one-hour interval and node, can be done by introducing loops (links of the form $a=\left(n_{i}, n_{i}\right)$ ) at each node to act as "parking" [4]. When a vehicle reaches a node, it will leave the system with a particular probability and be "parked" for a certain amount of time. In the model, we assign all loops a travel time $\tau_{k}=1$ (the same length as the constant interval of the step function). In this way, we can always find flows $\varphi_{k}(t)$ on the loops such that the system is balanced.

In our study, we assume that the measurement devices are able to record and communicate the vehicles' identity (e.g., license plate number), position, travel direction, and the time for the observation. Since we expect that the possibility for probe vehicles to successfully detect and identify other vehicles is considerably lower than for stationary devices, there is probably a need for a rather high number of probe vehicles. We use the country of Sweden, which has around 684,000 registered trucks ${ }^{2}$, as an example. If we assume that only $10 \%$ of the trucks operate as probe vehicles, there would still be as many as 68400 mobile devices in the network. If we also consider that private actors, as well as other vehicle types than trucks could operate as probe vehicles, there could be a quite large number of mobile devices even though a very low percentage of the vehicles act as probe vehicles.

## 4. Simulation study

In this section, we describe our simulation study which illustrates how stationary and mobile devices can be used for traffic data collection. The main idea is to repeatedly simulate the movements of trucks for a selection of routes in our

[^1]traffic network, and log each observation made by a device. Our performance measure is the share of the simulated traffic that is successfully observed and identified by either a stationary or mobile device, for each of the 24 hours of the day with varying identification rate.

### 4.1. Scenario description

We consider a road traffic network in a geographic region in southern Sweden (see Fig. 1). In the study, we let speed cameras act as stationary devices, and freight trucks are used as probe vehicles. We focus on freight transport since trucks may have fewer concerns regarding privacy issues and may therefore be more prone to operate as probe vehicles. The main purpose of our simulation study is to evaluate and illustrate our idea of using stationary and mobile devices for collecting detailed (individual level) data about the movements of trucks. By studying the traffic flow map in Fig. 1 and the underlying truck flow data (average daily flows), we identified the road transport network that we considered relevant to consider in our simulation study. The network consists of 33 nodes and 102 uni-directed links.


Fig. 1: The traffic network used in our simulation study on top of the traffic flow map that we used to identify an appropriate traffic network. The traffic flow map was provided by the Swedish Transport Administration.

The nodes and links were selected to capture the major traffic flows in the considered region. The nodes are selected in locations where routes are expected to intersect, and where it is obvious that large traffic flows may split in the network. This is the reason that some nodes are located very close to each other. For each link, we extracted the corresponding distribution of the traffic volumes for each of the 24 hours of a day and identified the speed cameras located along the roads.

According to the best of our knowledge, speed cameras in Sweden are only used for identifying vehicles violating the speed limits. As discussed in Section 2, we assume that the functionality of speed cameras can be extended to also operate as devices for traffic data collection. The density of speed cameras in the whole of Sweden is smaller than in the network used in our studies. The reasons for this is that the considered region is rather densely populated with a high amount of traffic (there are large parts of Sweden, in particular in the north, with lower traffic volume and fewer speed cameras) and we mainly included roads where there are high traffic volumes, which are also those roads that typically contain speed cameras. In total, we included 121 speed cameras in the network. However, the cameras are not evenly distributed in the network as the number of cameras on the links varies between zero and eleven (see Table 1). There

Table 1: Links (node pairs), travel times, link length, and number of speed cameras along the links. Each pair of nodes represents two links in opposite directions. For instance, the node pair 3,32 represents the two links $(3,32)$ and $(32,3)$, where one speed camera is located along the link $(3,32)$ and two speed cameras along the opposite direction $(32,3)$, denoted $1 / 2$ in the "Cameras" column.

| Link (node pair) | Travel time (min) | Length (km) | Cameras <br> (amount) | Link (node pair) | Travel time (min) | Length (km) | Cameras <br> (amount) | Link (node pair) | Travel time (min) | Length (km) | Cameras <br> (amount) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1,13 | 11 | 20.0 |  | 7,12 | 35 | 39.8 |  | 16,20 | 13 | 22.0 |  |
| 1,14 | 47 | 66.3 | 2/2 | 7,20 | 15 | 16.0 | 2/2 | 16,22 | 23 | 31.6 | 3/4 |
| 2,17 | 35 | 62.6 | 1/0 | 8,10 | 59 | 80.2 | 11/11 | 17, 18 | 6 | 9.9 |  |
| 2,29 | 2 | 4.2 | 0/1 | 8,27 | 71 | 101.0 | 2/2 | 18,23 | 7 | 11.8 |  |
| 2,31 | 11 | 20.5 |  | 9,21 | 18 | 28.7 | 0/1 | 18,25 | 19 | 29.9 |  |
| 3,4 | 6 | 9.7 |  | 9,24 | 24 | 34.8 |  | 19, 30 | 22 | 32.3 |  |
| 3,12 | 26 | 33.7 | 0/1 | 10,21 | 17 | 25.0 | $1 / 0$ | 21, 24 | 25 | 31.4 | 0/1 |
| 3,32 | 20 | 26.7 | 1/2 | 11, 12 | 3 | 4.3 |  | 22, 28 | 20 | 25.3 | 2/1 |
| 4,5 | 17 | 25.9 | 1/1 | 11,29 | 7 | 10.9 |  | 23, 25 | 16 | 18.1 | 4/3 |
| 4,7 | 23 | 31.0 | 3/3 | 11,33 | 7 | 8.8 | 1/1 | 23,28 | 24 | 36.1 |  |
| 4,19 | 13 | 18.8 |  | 13,31 | 16 | 29.1 |  | 24, 27 | 37 | 48.9 | 3/3 |
| 5,6 | 30 | 41.0 | 3/2 | 13,32 | 31 | 39.6 |  | 25, 28 | 47 | 45.6 |  |
| 5,9 | 38 | 55.3 | 2/2 | 14,26 | 38 | 58.3 | 1/1 | 26,27 | 5 | 7.0 |  |
| 5,19 | 27 | 33.3 |  | 14,32 | 38 | 71.2 |  | 29,33 | 5 | 9.0 |  |
| 6,7 | 8 | 10.8 | 2/2 | 15,16 | 2 | 3.1 |  | 26,30 | 38 | 54.2 | 6/6 |
| 6,20 | 8 | 12.0 |  | 15,17 | 4 | 6.4 |  | 31,33 | 12 | 14.3 | 2/3 |
| 6,22 | 21 | 27.0 | 2/2 | 15,23 | 17 | 20.7 | 4/5 | 32, 33 | 12 | 21.7 |  |

are 45 links with one or more cameras ( $44.1 \%$ of the links) and 57 links without cameras ( $55.9 \%$ of the links). For the links equipped with at least one speed camera, the mean and median numbers of cameras are 2.69 and 2 respectively.

As a route consists of a sequence of links, we argue that it is necessary to consider the number and characteristics of links to define the main characteristics of a route. Since the network contains 33 nodes and 102 links, the network contains 102 routes of length one, 326 routes (i.e., valid link sequences) of length two, and 1050 routes of length three. Our route selection process aimed to identify a "minimum" set of routes to be included in our simulation study. We claim that a successful identification of a vehicle depends primarily on the number of devices it is expected to encounter on the link, which is: 1) the number of speed cameras located along the link and 2) the number of expected probe vehicle meetings (i.e., probe vehicles that are simultaneously driving on the opposite link). To be able to define representative routes to include in our analysis, we considered routes of length one, two and three, which we categorized based on partitions of the links according to the two above mentioned criteria.

Since the speed cameras are unevenly distributed (see Table 1), we partitioned the links into two groups, links without any speed camera, and links with at least one speed camera. The two link groups were further partitioned into subgroups, based on the number of expected meetings. In our calculations of the expected number of meetings, we assumed that the number of departures on a link is uniformly distributed over each of the 24 hours of the day. This is a simplification we made to simplify the evaluation. Here, it should be emphasized that this assumption was made only during the process of categorizing the routes, and not in the simulation environment where we assumed the number of departures is Poisson distributed. For each of the considered links, the amount of traffic varies over the day; in particular, the traffic volumes during nights are always lower than the traffic in the day. The peak hours vary for the links, but in general, the hour distributions of the traffic flows are not essentially too different for the different links. Still, we argue that we get a valid indication of how the number of meetings is expected to vary for the different links even if we do not consider the variations over the day. We let $a_{k}$ and $a_{l}$ be opposite links, that is, $a_{k}=\left(n_{i}, n_{j}\right)$ and $a_{l}=\left(n_{j}, n_{i}\right)$ for some nodes $n_{i}$ and $n_{j}$. When a truck enters $a_{l}$, the expected number of trucks that is at the same time driving on $a_{k}$ is $\frac{\tau_{k}}{24} \int_{0}^{24} \varphi_{k}(t) d t$. The expected number of trucks to enter $a_{k}$ while the truck traverses $a_{l}$ is $\frac{\tau_{k}}{24} \int_{0}^{24} \varphi_{k}(t) d t$. In total, the expected number of trucks that the truck traversing $a_{l}$ will meet is $2 \times \frac{\tau_{k}}{24} \int_{0}^{24} \varphi_{k}(t) d t$. It should also be mentioned that the expected traveling speed of all links in the considered scenario is identical (approximately $68 \mathrm{~km} / \mathrm{h}$ ). In addition, we further assumed that the length and estimated daily truck traffic volume of any link is identical to the length and truck traffic volume of its opposite link, i.e., $\int_{0}^{24} \varphi_{k}(t) d t=\int_{0}^{24} \varphi_{l}(t) d t$.

The number of expected vehicles that a truck traveling on a particular link is expected to meet, varies from 0.054 to 222 , where the mean value is 25.5 , and the median in 15.7. In addition, the cumulative distribution over the measure (see Fig. 2) clearly shows that the smaller values of expected meetings dominate. For example, $98 \%$ of the values are below $200,96.1 \%$ of the values are below 100 , and $90.2 \%$ of the values are below 50 . It can be also seen in the histogram over the expected number of meetings (see Fig. 2) that the smaller values are dominating. In Fig. 2, we also show the cumulative distribution over the measure; however, with logarithmic $x$-axis, which except for the extremely
large and small values shows a rather linear behavior. Based on the number of expected meetings, the routes of length one was partitioned by the value ranges [ $0.0539,8.54]$, $[8.90,20.9]$, and $[23.9,222]$ into six route categories (each value range gave two route categories, with or without speed camera). The links that are a part of routes of length two or three, was partitioned by the value ranges [0.0539, 15.7] ( 50 links) and [17.1, 222] ( 52 links) into four route categories. From the link partitions, we considered 56 different route categories. From each one of the 56 route categories, we randomly selected one route to be included in our output analysis.


Fig. 2: Empirical cumulative distribution (left), histogram (middle), and logarithmic empirical cumulative distribution (right) over the number of expected meetings.

### 4.2. Results and analysis

The simulation, using 13200 trucks, was run for a 11-day period (with a 24 hour warm-up period), where each successful identification was logged. We examined the share of the vehicle fleet that was identified at least once on the considered routes. The results of our simulation study, i.e., the share of successful identifications, are presented in Table 2 and Fig. 3, where we considered varying identification rate, and assumed identical identification rate for all stationary and mobile devices.

Table 2: The average share of traffic identified at least once on all routes with varying identification rate.

| Time period | Identification rate |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
| 0-1 | 0.243 | 0.384 | 0.491 | 0.570 | 0.634 | 0.675 | 0.713 | 0.737 | 0.757 | 0.777 |
| 1-2 | 0.259 | 0.400 | 0.500 | 0.586 | 0.646 | 0.696 | 0.729 | 0.758 | 0.781 | 0.798 |
| 2-3 | 0.288 | 0.446 | 0.549 | 0.625 | 0.688 | 0.726 | 0.755 | 0.778 | 0.799 | 0.813 |
| 3-4 | 0.350 | 0.524 | 0.638 | 0.712 | 0.760 | 0.796 | 0.833 | 0.865 | 0.881 | 0.898 |
| 4-5 | 0.482 | 0.692 | 0.803 | 0.862 | 0.906 | 0.931 | 0.945 | 0.956 | 0.966 | 0.973 |
| 5-6 | 0.720 | 0.874 | 0.928 | 0.951 | 0.969 | 0.983 | 0.990 | 0.993 | 0.994 | 0.996 |
| 6-7 | 0.706 | 0.885 | 0.934 | 0.957 | 0.975 | 0.978 | 0.983 | 0.988 | 0.992 | 0.997 |
| 7-8 | 0.729 | 0.889 | 0.936 | 0.963 | 0.981 | 0.987 | 0.992 | 0.997 | 0.997 | 0.999 |
| 8-9 | 0.721 | 0.882 | 0.944 | 0.965 | 0.979 | 0.987 | 0.990 | 0.994 | 0.996 | 0.999 |
| 9-10 | 0.728 | 0.901 | 0.944 | 0.965 | 0.979 | 0.984 | 0.990 | 0.994 | 0.998 | 0.999 |
| 10-11 | 0.738 | 0.879 | 0.935 | 0.958 | 0.973 | 0.983 | 0.990 | 0.995 | 0.996 | 0.999 |
| 11-12 | 0.744 | 0.889 | 0.945 | 0.968 | 0.977 | 0.987 | 0.991 | 0.994 | 0.996 | 0.999 |
| 12-13 | 0.723 | 0.864 | 0.925 | 0.951 | 0.969 | 0.978 | 0.983 | 0.989 | 0.991 | 0.992 |
| 13-14 | 0.732 | 0.872 | 0.931 | 0.961 | 0.971 | 0.980 | 0.986 | 0.990 | 0.994 | 0.997 |
| 14-15 | 0.731 | 0.889 | 0.946 | 0.965 | 0.975 | 0.982 | 0.988 | 0.989 | 0.993 | 0.998 |
| 15-16 | 0.700 | 0.873 | 0.934 | 0.960 | 0.976 | 0.981 | 0.984 | 0.987 | 0.990 | 0.993 |
| 16-17 | 0.639 | 0.806 | 0.883 | 0.926 | 0.947 | 0.965 | 0.974 | 0.979 | 0.985 | 0.990 |
| 17-18 | 0.577 | 0.757 | 0.838 | 0.886 | 0.911 | 0.937 | 0.949 | 0.958 | 0.967 | 0.972 |
| 18-19 | 0.506 | 0.692 | 0.794 | 0.856 | 0.899 | 0.926 | 0.943 | 0.955 | 0.963 | 0.968 |
| 19-20 | 0.464 | 0.652 | 0.757 | 0.822 | 0.857 | 0.885 | 0.908 | 0.922 | 0.937 | 0.944 |
| 20-21 | 0.406 | 0.582 | 0.687 | 0.766 | 0.817 | 0.854 | 0.875 | 0.898 | 0.911 | 0.928 |
| 21-22 | 0.356 | 0.541 | 0.655 | 0.727 | 0.785 | 0.827 | 0.855 | 0.880 | 0.900 | 0.918 |
| 22-23 | 0.302 | 0.477 | 0.597 | 0.666 | 0.721 | 0.758 | 0.795 | 0.825 | 0.844 | 0.855 |
| 23-24 | 0.259 | 0.401 | 0.505 | 0.583 | 0.643 | 0.682 | 0.709 | 0.728 | 0.743 | 0.755 |

For probe vehicles, we also include the share of the vehicles acting as probe vehicles in the identification rate. If we let the share of vehicles that act as probe vehicles be $x$, and the share of encountered vehicles that a mobile device is


Fig. 3: Share of vehicles that were successfully identified at least once per hour for the routes, where the identification rate are varied in the same way for stationary and mobile devices. The share of successfully identified traffic is the average taken over all simulated routes in the network for each hour and each identification rate.
expected to successfully identify be $y$, then the probability of a successful identification for mobile devices, according to our definition, is $x \times y$. We argue that $x \times y$ should be proportional to the total number of identifications made (i.e., no additional information is gained by knowing both $x$ and $y$ ). The results confirm that it is possible to obtain a high share of successfully identified vehicles, even for a low percentage of the transportation fleet acting as probe vehicles, and for low identification rate. In the diagrams in Fig. 4 and Fig. 5, we illustrate how the share of successfully identified vehicles is expected to vary over a 24 -hour period for varying identification rate for one link without speed cameras, and one link with two speed cameras, both with expected numbers of meetings in the range [0.0539, 8.54]. Fig. 4a clearly shows that it is challenging to successfully identify vehicles during the night when the number of expected meetings is low, and there is no stationary device along the link. Fig. 5a indicates that for the links with a low number of expected meetings, the share of successfully identified traffic can be significantly increased by using (especially during nighttime), in this case, two stationary devices. On links with a reasonably high number of expected meetings, the share of identified vehicles is expected to be rather high, even for lower shares of vehicles acting as probe vehicles or lower identification rate.


Fig. 4: Share of successfully identified vehicles for a link without speed cameras, and expected numbers of meetings in the range [0.0539, 8.54].

## 5. Concluding remarks

We have presented an idea of using active stationary and mobile measurement technologies for traffic data collection. Based on a discussion on the complementing characteristics of stationary and mobile measurement devices, we suggest


Fig. 5: Share of successfully identified traffic for a link with two speed cameras, and expected numbers of meetings in the range [0.0539, 8.54].
that both types should be considered for traffic data collection. To validate our idea, we studied transport in southern Sweden, using a traffic simulation model. The purpose of the study was to investigate the relevance of our idea of using both technologies for traffic data collection with varying identification rate, where speed cameras and freight trucks act as stationary and mobile devices, respectively. The results show that a reasonably high share of successfully identified vehicles can be achieved even for low identification rate. Future work includes studying the integrity aspects of the suggested idea (data encryption) and developing methods for identifying incorrect data.

## Acknowledgments

This work was partly funded by the Regional Council of Blekinge, the Swedish Agency for Innovation Systems, the Swedish Transport Administration, and the Swedish Transport Agency.

## References

[1] G. Leduc, et al., Road Traffic Data: Collection Methods and Applications, Working Papers on Energy, Transport and Climate Change 1 (55) (2008) 1-55.
[2] M. Noureddine, M. Ristic, Route planning for hazardous materials transportation: Multicriteria decision making approach, Decision Making: Applications in Management and Engineering 2 (1) (2019) 66-85.
[3] W. Y. Ochieng, M. A. Quddus, R. J. North, R. B. Noland, Technologies to measure indicators for road user charging, Proceedings of the Institution of Civil Engineers-Transport 163 (2) (2010) 63-72.
[4] M. Dahl, J. Holmgren, H. Fredriksson, E. Laksman, Significant route identification using daily 24-hour traffic flows, The $23^{\text {rd }}$ IEEE International Conference on Intelligent Transportation Systems, In press (2020).
[5] L. A. Klein, M. K. Mills, D. R. Gibson, et al., Traffic detector handbook: Volume I, Tech. rep., Turner-Fairbank Highway Research Center (2006).
[6] F. Meng, S. C. Wong, W. Wong, Y. Li, Estimation of scaling factors for traffic counts based on stationary and mobile sources of data, International Journal of Intelligent Transportation Systems Research 15 (3) (2017) 180-191.
[7] P. J. Tarnoff, D. M. Bullock, S. E. Young, J. Wasson, N. Ganig, J. R. Sturdevant, Continuing evolution of travel time data information collection and processing, Tech. rep. (2009).
[8] T. Seo, T. Kusakabe, Probe vehicle-based traffic state estimation method with spacing information and conservation law, Transportation Research Part C: Emerging Technologies 59 (2015) 391-403.
[9] E. Jenelius, H. N. Koutsopoulos, Travel time estimation for urban road networks using low frequency probe vehicle data, Transportation Research Part B: Methodological 53 (2013) 64-81.
[10] J. C. Herrera, D. B. Work, R. Herring, X. J. Ban, Q. Jacobson, A. M. Bayen, Evaluation of traffic data obtained via GPS-enabled mobile phones: The mobile century field experiment, Transportation Research Part C: Emerging Technologies 18 (4) (2010) 568-583.
[11] C. Nanthawichit, T. Nakatsuji, H. Suzuki, Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway, Transportation Research Record 1855 (1) (2003) 49-59.
[12] F. Mannering, W. Kilareski, S. Washburn, Principles of Highway Engineering and Traffic Analysis, 5th Edition, John Wiley \& Sons, 2013.


[^0]:    * Corresponding author. Tel.: +46-455-385548

    E-mail address: henrik.fredriksson@bth.se

[^1]:    ${ }^{1}$ The Swedish national road database (NVDB), provided by the Swedish Transport Administration.
    ${ }^{2}$ The Swedish Transport Agency (accessed 25 June 2020)

