Monitoring the San Francisco Bay Area freeway network using probe vehicles and random access radio channel

March, 1994

Jean-Paul M.G. Linnartz / Marcel Westerman / Rudi Hamerslag

TU Delft
Faculty of Civil Engineering
Department of Infrastructure
Transportation Section
Monitoring the San Francisco Bay Area Freeway Network Using Probe Vehicles and Random Access Radio Channel

Jean-Paul M.G. Linnartz
Dept. of Electrical Engineering and Computer Sciences
University of California at Berkeley, CA 94720

Marcel Westerman
Dept of Transportation Planning and Highway Engineering
Delft University of Technology, The Netherlands

Rudi Hamerslag
Dept of Transportation Planning and Highway Engineering
Delft University of Technology, The Netherlands

This work was performed as part of the California PATH Program of the University of California, in cooperation with the State of California Business, Transportation, and Housing Agency, Department of Transportation; and the United States Department of Transportation, Federal Highway Administration.

The contents of this report reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official view or policies of the State of California. This report does not constitute a standard, specification, or regulation.

Draft Path Report UCB-ITS-PWP-94-

February 1994

ISSN 1055-1417
1. **Report number**  
VK 3105.301

2. **Title report**  
Monitoring the San Francisco Bay Area Freeway Network Using Probe Vehicles and Random Access Radio Channel

3. **Authors**  
Jean-Paul M.G. Linnartz  
Marcel Westerman  
Rudi Hamerslag

4. **Institute**  
Delft University of Technology  
University of California at Berkeley

5. **Principal(s)**  
California PATH Program (IVHS)

6. **ISSN-number**  
LVV-report 0920-0592

7. **Subject**  
Transportation Modelling

8. **Research project**  
Datacommunication networks

9. **Category report**  
Research reports

10. **Date**  
March 1994

11. **Summary**

The lack of proficient real-time traffic monitoring systems is one of the major bottlenecks of Advanced Traveller Information Systems (ATIS) and Traffic Management Systems (ATMS). In this report we describe a method of collecting real-time traffic data from probe vehicles automatically sending traffic reports to one or more base stations, connected to a traffic center by a wired communications network. Analyzing and computing road traffic and message traffic flows in the San Francisco Bay Area, we study several multi-disciplinary aspects of this data collection technique, such as the relation between vehicle traffic and message traffic, the influence of road traffic congestion on communication performance, the reliability of road traffic estimates on radio network throughput and the location of base stations. The results presented in this report reveal that random access (ALOHA) transmission of traffic messages is a (spectrum) efficient, inexpensive and flexible method for collecting road traffic data and that this approach can provide reliable traffic monitoring. Not only highly accurate real time link travel times can be estimated, but also Automatic Incident Detection (AID) can performed.

12. **Remarks**

This work was performed as part of the California PATH-program and submitted as a draft PATH-report.

13. **Pages**  
58

14. **Price**  
dfl. 55,- incl.
Abstract - The lack of proficient real-time traffic monitoring systems is one of the major bottlenecks of Advanced Traveller Information Systems (ATIS) and Traffic Management Systems (ATMS). In this report we describe a method of collecting real-time traffic data from probe vehicles automatically sending traffic reports to one or more base stations, connected to a traffic center by a wired communications network. Analyzing and computing road traffic and message traffic flows in the San Francisco Bay Area, we study several multi-disciplinary aspects of this data collection technique, such as the relation between vehicle traffic and message traffic, the influence of road traffic congestion on communication performance, the reliability of road traffic estimates on radio network throughput and the location of base stations. The results presented in this report reveal that random access (ALOHA) transmission of traffic messages is a (spectrum) efficient, inexpensive and flexible method for collecting road traffic data and that this approach can provide reliable traffic monitoring. Not only highly accurate real time link travel times can be estimated, but also Automatic Incident Detection (AID) can performed.

Note: With this report, a floppy disk is available containing a simulation/animation and a numerical analysis of the performance of an ALOHA radio network for collecting road traffic data from probe vehicles in the San Francisco Bay Area.
1 Introduction

For Advanced Traffic Management Systems and Advanced Traveler Information Systems (ATMIS) accurate and reliable information is needed to observe and to control the traffic flow. In particular information about real-time travel times, speeds, real time origins and destinations and disturbances of the traffic flows are of great importance. In the past, elementary traffic data were collected by a fixed data collection system along the road infrastructure, mostly using inductive loop detectors. These fixed detectors typically provide instantaneous speeds and occupancy at the location of the detector. The desired traffic information subsequently has to be deduced from these collected elementary data. For deducing real-time travel times and real-time origins and destinations this requires complex and time demanding analyses (Westerman, 1994), while disturbances in the traffic flows due to incidents can only be detected when they have propagated to the location of a fixed detector. Although these methods may be suitable for several mid to long term and semi-dynamic traffic management decisions, the complexity, the analyses-time and the propagation delay are prohibitive for more advanced traffic control measures, such as ATMS and ATIS (Westerman and Hamerslag, 1993; OECD, 1992). This is especially the case for a statewide or nationwide traffic monitoring network, with large mutual distances between the locations of the fixed detectors. Novel, advanced traffic control systems need more reliable, more accurate and more detailed real-time traffic data, both on network level as well as on specific traffic bottlenecks.

To meet the input data requirements of ATM/IS, new techniques have to be developed for data collection and processing. In this report, we study a data collection technique where the source of real-time road network data are the vehicles themselves, functioning as moving traffic 'probes' (Ran, 1993). By means of on-board electronics, each probe vehicle keeps track of its own geographic position using for instance dead-reckoning based on wheel sensors, compass and a digitized road map and (differential) GPS satellite positioning. At random, but nonetheless more or less regular time intervals each probe vehicle transmits its travel experiences, i.e., its average or instantaneous speed, and travel time, on a radio channel, such that is may be received by a nearby base station. All base stations in a specific area are connected to a Traffic Control Center (TCC) by a wired communication network. At the TCC, the received probe vehicle data is processed into road traffic information and travel advisories.

One of the major bottlenecks of a wireless data collection technique is the data communication link from probe vehicle to base station. Several communication protocols and transmission media can be used for data transfer from many mobile transmitters (the roving probe vehicles) to one receiver (the nearby listening base station), each having its own advantages and disadvantages. In this report, we will analyze the probe vehicle concept, focussing on communication limitations. We address a radio architecture that is simple, inexpensive and spectrum-efficient for the probe vehicle transmissions. We evaluate this concept in a realistic traffic situation.
The organization of the report is as follows: In section 2, the system concept is described. A road-traffic flow calculation for the freeway network of the San Francisco Bay Area is described in section 3. A detailed analysis of the throughput of the proposed radio network takes place in section 4. This section specifies the relation between the spatial message throughput in the base stations and the spatial distribution of the traffic messages transmitted by the probe vehicles. In section 5, the models for road network and the radio network analysis are combined to perform a case study of road traffic and message traffic in the San Francisco Bay Area. This results in a specification of the requirements for monitoring the complete freeway network in the SF Bay Area using probe vehicles and random access radio channel. Sections 3 and 4 contain a theoretical analysis, while sections 2 and 5 are more descriptive and relatively self contained. Section 6 concludes this report.
2 Probe Vehicle System Concept

The application of telematics in traffic and transport has already resulted in the deployment of a number of (Advanced) Traveler Information and Traffic Management Systems (ATIS / ATMS). Examples are the U.S. Transport Advisory Radio (HAR and AHAR) and systems where subcarrier voice or data messages are added to the audio and video signals on FM and TV transmitters. The European Radio Data System/Traffic Message Channel (RDS/TMC) system, the German Autofahrer Rundfunk Information (ARI) and the British CARFAX all use FM subcarrier transmission for disseminating data to travellers. ATIS services like in-vehicle electronic mapping, real-time travel information and dynamic route guidance are under development. ATIS may help motorists in finding optimal routes to their destinations, so they lead to individual user optima. By means of several ATMS applications - such as ramp coordinated metering, variable message signing and automatic incident detection - it is tried to reach for an optimal overall system performance (system optimum). It has been observed in the past that the user optimum may differ from the system optimum.

The real-time traffic data required to generate ATIS service messages and deploy ATMS applications can be gathered from police and local authorities, sensors (Levy, 1992), weather stations and air surveillance. Few systems are yet operational that gather real-time road traffic data automatically. As travel times between two points are more reliable than measurements of the speed of vehicles at one particular point along the road, (probe) vehicles participating in the road traffic and automatically reporting the (link) time needed for travelling between two intersections are a useful source of road traffic data.

Figure 1a depicts a generic lay-out of a system in which traffic data is collected by probe vehicles and several other monitoring sources (road sensors, police, authorities, air surveillance, etc.) and distributed to a TCC. The TCC processes this data and takes measurements (e.g. ramp metering), supplies information (e.g. variable message signing) or distributes information directly to the car drivers (e.g. travel and traffic information, dynamic route guidance). Here we focus on the communication link between probe vehicle and the TCC. We do not consider strategies for processing probe data (centralized, decentralized, user optimum, system optimum, etc.).

In a practical system, it is desirable to use the communication for other services also. A detailed description of a possible solution for an 'integrated services' network using a single (30 kHz radio) channel is to be documented in a MOU 107 working paper (Linnartz, 1994). This design distinguishes downlink messages from TCC to the (probe) vehicles and uplink messages to the vehicles in the reverse direction. Downlink transmissions can be subdivided into datacasting from the TCC to all vehicles or groups of vehicles and messages from the TCC to one or more specific (probe) vehicles. Uplink messages from vehicles - via base stations - to the TCC and can be subdivided into emergency messages from a probe vehicle in danger, random-access messages (e.g. information queries, Automatic Vehicle Location/Identification), and probe data (unprocessed road traffic data). The future IVHS communication architecture may even serve a wider variety of applications and services. Each of the distinguished subclasses has its distinct traffic requirements and may therefore require a particular communication solution. In this paper, we limit ourselves to the uplink
probe transmissions. This concept thus separates uplink (probe) transmissions and downlink dissemination of traffic information into different communication channels. At the physical level, the separation is done in the time domain, but the medium-access and network (routing) protocols also differ per traffic category. We would like to emphasize that for the case considered here, there is no need for full-duplex communication. Uplink data collection from probes and downlink datacasting may be implemented without any interaction of the telecommunications traffic streams. Despite the fact that only a fraction of all available time slots is available from probes, the analysis of the system is similar to the situation where one channel is used exclusively for probes.

![Figure 1a Lay-out of generic probe vehicle system concept](image)

2.1 Related Research Activities

Prototypes of systems using 'intelligent' vehicles using distinct communication media to collect road traffic data have already been developed (OECD, 1992).

The EURO-SCOUT system (EURO-SCOUT, 1991) for instance uses infrared spectrum for the line-of-sight communication link between the EURO-SCOUT vehicles and road-side beacons. These beacons are connected to a TCC by a wired infrastructure. Several pilot projects (LISB, 1991; OECD, 1992) showed promising results. However, a major disadvantage of this system concept is the large number of beacons that are needed for a sufficiently high coverage of an urban area. Another disadvantage is that disturbances in the traffic flow can only be detected when the disturbances (or the vehicles that have experienced the disturbances) have propagated to the vicinity of a beacon. For Automatic Incident Detection this may take too long.
Another probe vehicle system concept under development is SOCRATES (Catling and Op de Beek, 1991). Although significant portions of development of this information systems is being design to operate independent of choice of the communication link, it is likely to use a modified version of the pan-European digital cellular mobile network GSM for the communication link between the SOCRATES-vehicles and the base stations. An advantage of this approach is that little extra dedicated communication infrastructure is needed. Circuit-switched (telephony) networks are fundamentally inefficient for short data messages, as generated by probes. In order to provide SOCRATES with more suitable features and a more efficient datatransmission capability, GSM’s data services have to be extended. This extension has been called a general packet radio service (GPRS: Cattling, 1994).

2.2 Radio Access

An appropriate medium for wireless probe communication is radio transmission. One solution could be to assign a dedicated radio channel, with access protocols optimized for the particular characteristics encountered in collecting road traffic data from probe vehicles. With multiple base stations in a regular cellular frequency reuse pattern, a nationwide or statewide coverage can be achieved. Polling the probe vehicles one-by-one requires substantial management efforts (transmission of synchronization signals, handovers to other base stations, updating the transmission sequence of vehicles leaving a cell, etc.) and may significantly reduce the efficiency of the network. Moreover, issues with respect to privacy are very arduous when using polling schemes.

Another possibility, which we propose here, is to not coordinate vehicle transmissions at all (ALOHA; Abramson, 1977). All vehicles in a cell transmit on the same common radio channel, in any available time slot, regardless of the possibility of transmission of travel data by other probe vehicles. The ALOHA scheme operates as follows: probes transmit traffic messages at random instants of time, accepting the risk of mutual interference between messages. If multiple messages interfere with each other, because of radio-wave propagation effects the signals are likely to be received with substantially different power. In such case the strongest signal is likely to 'capture' the receiver, while only the weaker signals are lost (Linnartz, 1993). However, these signals from remote probes may capture other base station receivers at different locations.

Mobile ALOHA radio networks with a single base station have been researched extensively in the last years. The main conclusions can be summarized by

- The maximum throughput can significantly exceed $1/e$ (36.8 %) because of receiver capture
- The mobile ALOHA channel is significantly more stable than the wired (LAN) ALOHA network
- Control of the number of admitted users (and their total average traffic load) in the system can be effective to ensure stability. This is in sharp contrast to dynamic control of the retransmission back-off time, as required in wired ALOHA networks.
• Throughput does not decrease rapidly to zero for large traffic loads
• The point of operation does not have to be at relatively low traffic loads
• Remote terminals have a lower probability of successfully transmitting their messages than nearby terminals. Remote terminals nonetheless benefit from capture as it diminishes the traffic intensity of strong interfering packets.
• Error correction redundancy does not increase throughput significantly if the channel is ‘slow fading’, i.e. if received powers are fairly constant during the packet duration
• Error detection schemes should be more effective than for wired communications over Additive White Gaussian Noise (AWGN) channels
• The throughput for spatially uniform offered message traffic is independent of packet traffic load.
• Adaptive antennas and signal processing are very effective to enhance the throughput
• Packet transmissions are preferably much shorter that the coherence time of the channel fading

Our application significantly differs from the usual ALOHA schemes where any message lost because of channel fading or interference from other transmissions is retransmitted automatically. In earlier evaluations (Linnartz, 1993), we found that the ALOHA performance can be further enhanced if transmissions also occur regardless of a frequency reuse pattern. The throughput appeared to be optimum if one assigns the same time slots on the same frequency simultaneously in all cells for probe transmissions. In such case, a probe does not have to know in which cell it is located and a base station can accept any probe report, regardless of whether it is transmitted within or outside its particular cell. Moreover, this offers ‘site diversity reception’ from probe vehicle messages as one traffic message may be received by multiple base stations (Linnartz, 1992). This scheme allows us to use the same radio channel in a large area contiguously, thus without using different frequencies in adjacent areas, as is common practice in cellular nets. Listening base stations, connected to a fixed backbone network, can be located throughout the system operational area. A network planning tool for the fixed backbone is being developed within the PATH program by Prof Walrand at U.C. Berkeley.

2.3 Receiving Base Station

A base station, as considered in our probe vehicle system concept, contains transmit and receive facilities, a (PC-style or dedicated) computer system, a GPS receiver for synchronization and connection to a high speed digital backbone. Such a base station can receive messages from probe vehicles during a certain percentage of time and be used for other IVHS services during the rest of the time. It is however possible to add additional, inexpensive base stations to achieve a higher throughput of probe vehicle messages at a particular location. This requires no further system modifications. Such an additional ‘probe data collector’ does not need transmit and synchronization facilities. As illustrated in Figure 1b, it consists of an antenna, a digital receiver, a message buffer and a simple processing unit and a telephone modem.
In the network under study, all base stations listen to the common ALOHA channel and transfer any successfully received message to the TCC. In our single-channel design, we offer about $f' = 18$ time slots per second for probe vehicles\(^1\).

In a lightly loaded radio network, a particular message may be received at multiple base stations. Hence, at the TCC duplicates have to be discarded. In a heavier loaded network, uncoordinated radio transmissions lead to message collisions which reduce the efficiency of the network, but the overhead for management is practically zero and it allows a simple system architecture and simple in-vehicle equipment.

\[^1\text{As we will compute later, the successful throughput per location per second is on the order of } S_0 \text{ (typically } S_0 \approx 0.32) \text{ times } f'. \text{ Each probe message contains at most } L = 260 \text{ bits. The throughput is thus } L S_0 f' \approx 1.5 \text{ kbit/s per base station. As some communication capacity may also be needed for remote control and telemetry of the base station, a 2400 baud telephone link may introduce some queueing delay of probe messages when they are transferred to the TCC. A telephone link with a 9600 baud modem should be amply sufficient to transfer data from the base station to TCC at negligible queueing delay.}\]
In our application, the loss of one particular travel report does not result in a retransmission: we would rather see a more recent report from another vehicle in a next time slot. This also implies that there is no need for acknowledging successfully received messages, as in conventional ALOHA radio networks. Unlike the conventional ALOHA system, where excessive retransmissions may cause system instability, our proposed system is not subject to such problem. The network performance parameter most relevant to our application is the successful throughput per road segment per time (or per minute).

2.4 Probe Vehicle Terminal

Figure 1c shows that for 'intelligent' vehicles already equipped with a navigation system, the additional equipment needed for the proposed radio system is limited to a radio modem and some data processing in the in-vehicle controller.

The radio modem contains a receiver, for receiving traffic advisories from the TCC as they are datacast to all vehicles, or data messages for interactive ATMIS services. These signals from the base station also offer slot synchronization that allows the vehicle transmitter to find the appropriate slots for transmitting probe date. Also, we will show later that in the ALOHA concept the TCC needs to specify the average intertransmission time of the probes. This requires an occasional (downlink) datacast message to all probes.
2.5 Modes of Operation

Up to now we have only addressed an entirely centralized ATMIS operation. Multiple base stations collect traffic messages from a large area. After appropriate processing at the TCC, resulting travel advisories are broadcast (for instance by HAR and subcarrier Radio Data System) and traffic messages are deployed. One of the advantages of defining a wide-area standard for probe transmissions reports is the flexibility of choosing a scope of operation for receiving and processing link traffic data from probe vehicles. We summarize a few examples to show that a migration towards a full system is possible if immediate installation of a centralized system is regarded unrealistic: once the transmission scheme is defined and experiments begin, participation may develop gradually.

**ATIS Service Providers:**
The processing of link traffic data can be performed as a public service or as a commercial service offered to subscribers. Hybrid concepts are also possible. Next to appropriate traffic control strategies, a public service may provide (system-optimum) travel advisories only for a few major highways on which the congestion is of particular concern to local authorities. Private service providers can perform a (possibly more rigorous) analysis for individual users, predicting congestion using state-of-the-art data fusion techniques and novel traffic flow models.

**Decentralized Autonomous Advisory Systems:**
A simple ATIS system may consist of a receiver, a processing unit for received traffic data, and a presentation device. Examples of presentation output devices are electronic text displays along highways, variable message signs or speech synthesizers connected to an HAR transmitter. The receiver gathers traffic data from vehicles in its vicinity. This data is used by the processing unit to evaluate the local traffic conditions and to take actions accordingly in order to reach a local optimum.

**Autonomous On-board Vehicle Navigation:**
The ALOHA concept can also work in areas where no fixed infrastructure is available. Any intelligent vehicle can receive messages offered to the ALOHA channel by other vehicles. The received signals typically contain road traffic data from an area with a range up to a few kilometers from the receiving vehicle. The received link times may be used in combination with a CD rom onboard navigation system.

In our investigations, we assume that vehicles listen to base station transmissions to obtain a time slot synchronization. If the system has to work independently of an infrastructure, some modifications are needed.

**Commercial Fleet Management and Public Transit Monitoring:**
The interest in Automatic Vehicle Location systems is growing, as it may enhance the efficiency of operation a fleet of commercial vehicles or the grade of service of a public transit system. Typical solutions proposed for the radio network is to poll all vehicles according to a regular scheme. Disadvantages of such schemes are the need for a central control, the waste of spectrum resources by sending out synchronization and sequencing messages and by polling stationary vehicles. The scheme proposed here can also be used in
a closed user group network. We believe that using the proposed single-channel approach with random transmissions is more efficient than polling vehicles in a cellular re-use pattern, less expensive to implement and more flexible to expand. However, fleet management may require a guaranteed response from particular vehicles. Some modification of the transmission scheme may be needed.
3 Road Network Analysis

In order to evaluate the proposed probe vehicle system concept in a realistic traffic situation and to study whether sufficient, adequate and useful traffic data can be collected through probes, we have estimated traffic flows on the freeways in the San Francisco Bay Area have. This chapter describes all steps in the traffic-flow calculation. The following analysis is generic, in the sense that it can be applied for other study areas and for other transmission schemes as well.

3.1 Introduction

Starting point for computing road traffic flows is a network model of transportation infrastructure (including private as well as public transport), specified by nodes (concentrations of population and employment) and connecting links (roads and rail with specific direction and capacity) to model the existing situation in the California Bay Area. An essential part of the model is the Origin-Destination (O-D) Matrix, that specifies the amount of vehicles coming from or having to the different network nodes and which may or may not be available from measurements. To determine locations and concentrations of dwelling and employment in the zones without measured OD data, we use a combined land-use and transportation model that calculates the number of departures and arrivals (trip ends) of each specified zone, assuming accessibility to be the main factor influencing land-use. At the core of this model is a multimodal static interaction model with elastic land-use constraints (Hamerslag, 1994). The model is related to the four-stage transportation model (production/attraction, distribution, mode choice, assignment) (Hamerslag, 1987), but in our case, the distribution and mode choice are calculated simultaneously. The principal difference with the existing four-stage transport model is that our model uses elastic, rather than the traditional fixed constraints. The land-use data (employment and working population) that are traditionally exogenous in a fixed constrained interaction model can also be modified endogenously.

The trip-end model computes the number of trip ends in each zone from the employment and working population or directly from the number of trips using several trip purposes. From the trip ends, the route-choice model determines the generalized time between origin-destination pairs in the car and in the public transport network. Subsequently the trip ends are distributed over the road and the public transport network along the shortest routes between every zone. This results in an origin/destination-matrix for car trips and one for public transport trips that can be refined if traffic measurements are available for calibration. In the last part of the traffic flow calculation, the OD-matrices (car as well as public transport) are assigned to the network by the assignment model which results in traffic flows on each link. So, at the end of the traffic-flow calculation, the traffic flow and mean travel time is known for each link.
A certain percentage of the assigned vehicles will function as a probe vehicle, so they generate traffic messages. As yet, we assumed that a probe vehicle transmits its location and its travel time over the last (fraction of a) road link. The ultimate contents of the traffic messages however is subject of further research. Hence, we assume that vehicles perfectly know their location, for instance through a hybrid (differential) GPS and deadreckoning positioning technique. The transmitted traffic messages can be received by base stations located in the studied area. In an ALOHA network, harmful interference (collisions) between messages transmitted in the same time slot may occur. In our investigations, message collisions are taken into account, using models for receiver capture and mobile radio-wave propagation. The final result is an overview of the number of received traffic messages and their locations, for each base station. The analysis further estimates the number of messages that are received at multiple base stations.

The (sub) models described above have been integrated into one road traffic and data communication model. The layout of this integral model has been sketched in Figure 2. A description and results of the application of each of the sub models will be given in the succeeding paragraphs of this chapter (road network analysis) and in the next chapter (radio network analysis).
road network specification

Generalized Time between OD pairs \rightarrow land-use \leftarrow Generalized Time between OD pairs

Trip ends

distribution & mode choice

car assignment

public transport assignment

road link flows
link travel times

generation of traffic messages (analytical model)

generation of traffic messages (simulation model)

calculation of spatial throughput

calculation of spatial throughput

performance of IVHS services

Figure 2 Integral model for road traffic and data communication (Adapted from Hamerslag, 1993)
3.2 Road and public transport network specification

Link networks (composed of nodes and links) are used to specify road networks and line networks (composed of nodes, links, wait functions and change-over resistances) are used to specify public transport networks. For each of these methods, the study area is divided into zones. A zone is a network node with attributes such as dwelling and employment. Each trip starts and ends in a certain zone. The nodes are connected by links representing the road infrastructure.

This project focuses on the California (San Francisco - Oakland - San Jose) Bay Area and the road network considered has been limited to the major roads (mostly highways). First, the road infrastructure has been defined according to location, direction and capacity of highways existing in the Bay Area.

![Figure 3 Existing San Francisco Bay Area Road Network](Source: (Rand Mc.Nally, 1993))
Figure 4 Abstracted road network

Figure 5 Abstracted public transport network
The road map in Figure 3 shows the existing roads, while the sketch in Figure 4 illustrates the abstracted road network. The number and the location of the zones has subsequently been taken subordinate to the road infrastructure. This has led to 40 zones. For a realistic traffic flow calculation, a number of public transport lines has been added (see Figure 5). For trips by public transport, the speed outside the public transport network has been taken high (15 km/h), assuming that the car will be used in pre- and post-transport. In Appendix A, the attributes of the abstracted public transport network, such as capacities and free flow speeds are given.

### 3.3 Land-use model

#### 3.3.1 Theoretical background

The theoretical foundation for the model of locations and concentrations of dwelling and employment (Hamerslag, van Berkum and Replogle, 1994) is a micro-economic theory under money and time constraints. This theory says that each person maximizes the difference between utility of where he/she lives and works and the sacrifice (generalized cost, a weighted sum of travel time, distance and other costs) of the commute. The utility of working in $i$ is denoted as $u_i$, and that of living in $j$ is $u_j$. Employment and living involve costs $z_i$ and $z_j$, respectively. We denote the sacrifice (generalized cost) to make trips from zone $i$ to zone $j$ using mode $m$ as $f(z_{ijm})$. So, we write

\[ x_{ijm} = u_i - z_i + u_j - z_j - f(z_{ijm}) + e_{ijm} \]  

where we use:

- $x_{ijm}$: consumer surplus for working in $i$ and residing in $j$ by using mode $m$
- $u_i$: utility to work in $i$
- $u_j$: utility to reside in $j$
- $z_i$: generalized cost work in $i$
- $z_j$: generalized cost to live in $j$
- $z_{ijm}$: generalized cost for trip from $i$ to $j$ by mode $m$
- $e_{ijm}$: error term

When all variables $e_{ijm}$ are 0 (zero) this leads to to users optimum approach calculated by the All-or-Nothing traffic assignment method (see § 3.6). When the variables $e_{ijm}$ are Normal distributed this leads to the probit model and when all variables $e_{ijm}$ are independently Weibull distributed, the probability ($t_{ijm}$) of a person to work $i$ and reside in $j$ using mode of transport $m$ can be calculated from the logit model (Hamerslag, 1977; Ortúzar and Willumsen, 1990)

\[ t_{ijm} = \frac{\exp(x_{ijm})}{\Sigma_{ijm} \exp(x_{ijm})} \]  

The mathematical expectation $T_{ijm}$ of the number of work-home relations in OD-pair $ij$ is proportional to the production of zone $i$ (represented by $Q_i$), the attraction of zone $j$ (represented by $X_j$) and to the probability to work in zone $i$ and to reside in zone $j$, with
\[ T_{jm} = Q_i \cdot X_j \exp(x_{ijm}) / \Sigma_{ij} \exp(x_{ijm}) \]
\[ = \Omega \cdot Q_i \cdot X_j \exp(x_{ijm}) \]  \hfill (3)

The normalizing constant \( \Omega \) ensures that \( \Sigma_{jm} T_{jm} \) equals the total working population.

Substitution of (1) in (3) gives

\[ T_{jm} = \Omega \cdot Q_i \cdot X_j \exp\{u_i - z_i^{m} + u_j - z_j^{m} - f(z_{ijm})\} \]
\[ = \Omega \cdot \exp(-z_i^{m}) \cdot Q_i \cdot \exp(u_i) \cdot \exp(-z_j^{m}) \cdot X_j \cdot \exp(u_j) \cdot \exp\{f(z_{ijm})\} \]  \hfill (4)

This relation can be simplified by defining

\[ F_{ijm} \triangleq \exp\{-f(z_{ijm})\} \]  \hfill (5)

\[ l_i \triangleq \exp(-z_i^{m}) \quad \text{and} \quad n_j \triangleq \exp(-z_j^{m}) \]  \hfill (6)

\[ E_i \triangleq Q_i \cdot \exp(u_i) \quad \text{and} \quad W_j \triangleq X_j \cdot \exp(u_j) \]  \hfill (7)

where

- \( F_{ijm} \) deterrence function, indicating the influence of the resistance for traveling between zone \( i \) and zone \( j \) with mode of transport \( m \) on the probability of making that trip
- \( E_i, W_j \) polarities, indicating the utilized production of zone \( i \) and attraction of zone \( j \)
- \( l_i, n_j \) balancing factors, indicating the cost to work in zone \( i \) and to reside in zone \( j \)

Thus, polarities depend on the size of the zones (production and attraction) and also on the utility of working or living in a certain zone. So the main difference between polarities and balancing factors is that polarities involve the utilities \( (u_i^{p} \text{ and } u_j^{d}) \) while the balancing factors involve the positive or negative cost \( (z_i^{m} \text{ and } z_j^{m}) \). The influence of accessibility is specified in the balancing factors. Substitution of the distribution, equilibrium factors and polarities in (4) now gives the interaction model (Hamerslag, 1977; Ortúzar and Willumsen, 1990).

Using these definition we initially compute \( T_{jm} \), i.e., the number of work-home relations between \( i \) and \( j \) with mode of transport \( m \)

\[ T_{jm} = \Omega \cdot l_i \cdot n_j \cdot E_i \cdot W_j \cdot F_{ijm} \]  \hfill (8)

The number of work-home relations \( T_j \) between \( i \) and \( j \) is found by adding \( T_{jm} \) over all nodes \( m \), so

\[ T_j = \Sigma_m T_{jm} = \Omega \cdot l_i \cdot n_j \cdot E_i \cdot W_j \cdot F_{ijm} \]  \hfill (9)

with

\[ F_{ij} = \Sigma_m F_{ijm} \]  \hfill (10)
The demand of employment in zone $i$, $T_{i+}$, is the sum of the corresponding $i$-th row of the OD-matrix with elements $T_{ij}$. The demand of residing in zone $j$, $T_{+j}$, is the sum of the corresponding $j$-th column. That is,

$$ T_{i+} = \sum_j T_{ij} \quad (11) $$
$$ T_{+j} = \sum_i T_{ij} \quad (12) $$

where

$T_{ij}$: the number of work-home relations between $i$ and $j$

$T_{ijm}$: the number of work-home relations between $i$ and $j$ with mode $m$

$E_i$: potential employment in zone $i$ if the influence of accessibility is ignored

$W_j$: potential working population in $j$ if the influence of accessibility is ignored

$l_i, n_j$: balancing or equilibrium factors of origin $i$ and destination $j$

$F_{ijm}$: the deterrence function for relation between $i$ and $j$ with mode $m$

$F_{ij}$: the deterrence function for relation between $i$ and $j$

$\Omega$: a constant

$g, h$: parameters

$T_{i+}$: realized employment in zone $i$

$T_{+j}$: realized working population in $j$

**Deterrence function**

The deterrence function (5) specifies the influence of the generalized times $z_{ijm}$ on the probability of a trip. Next to the generalized times, in the deterrence function we have to take into account car ownership.

We write the simultaneous deterrence function $F_{ij}$ for $ij$, as

$$ F_{ij} = d_{car} \sum_m F_{ijm} + (1 - d_{car})F'_{ijm} \quad (13) $$

where $F_{ijm}$ and $F'_{ijm}$ denote the deterrence function for relation $ij$ and mode $m$ when a car is available (CA) or car is not available (NCA), respectively. These are obtained from

$$ F_{ijm} = a_m \exp(b_m \ln^2(z_{ijm} + 1)), $$

and

$$ F'_{ijm} = a'_m \exp(b'_m \ln^2(z_{ijm} + 1)) $$

where

$a_m, b_m$: parameters for mode $m$ (CA)

$a'_m, b'_m$: parameters for mode $m$ (NCA)

$d_{car}$: car density (cars/person)

$z_{ijm}$: generalized time in OD-pair $ij$ by mode $m$

Here, it should be noted that car density significantly influences $F_{ij}$. Changes in land use can better be explained by this simultaneous model than with the traditional sequential four-stage model (production/attraction, distribution, mode choice and assignment).
Estimations of employment and residency

$E_j$ and $W_j$ are the potential work- and residences if the influence of accessibility is omitted. Notice that $E_j$ and $W_j$ can not be surveyed and are not equal to row and column totals. They are proportional to the available surface (production and attraction) and the utility to reside and work in a certain zone and are determined as follows.

Working and residing involve costs influenced by the accessibility $z_i^g$ and $z_j^h$. If people prefer to work in a certain zone that has good accessibility, there will be competition to get jobs and they will have to accept somewhat lower income. Their employers will have profit which stimulates job supply. If people prefer to reside in a certain zone with a good accessibility there will be competition for homes. This leads to an increase of the cost-of-living. So the profit for the landlords and developers will be higher, resulting in increased supply. As for the demand function we introduce a supply function with an exponential form. The profit is proportional to the extra cost.

Following tradition, we use this simple mathematical relation

$$
T_{i+} = \exp(z_i^g). E_i \\
T_{i+} = \exp(z_j^h). W_j
$$

although we realize that realistic relations may be more complicated. Substitution of equation (5) yields

$$
T_{i+} = l_i^g E_i \\
T_{i+} = n_j^h W_j
$$

Elasticities of constraints

$T_{i+}$ en $T_{i+}$ are the realized employment and residences of the working population. Using (7), (8) and (12) gives

$$
T_{i+} = l_i^g E_i = \sum_j \Omega_j l_j n_j E_i W_j F_g
$$

so that

$$
l_i = \frac{1}{\sum_j \Omega_j n_j W_j F_g}^{1/(l+g)}
$$

Analogously

$$
n_j = \frac{1}{\sum_i \Omega_i l_i E_i F_g}^{1/(l+h)}
$$

Now $1/(1 + g)$ and $1/(1 + h)$ are defined as the elasticities of the constraints.
3.3.2 Specification of parameters

**Generation and distribution**
The parameters in the generation and distribution model are determined using data from the National Traffic Survey in the Netherlands (CBS, 1992) and can be regarded as being representative for the American situation as well. The parameters \( a_m \) and \( b_m \) in the lognormal multimodal deterrence function are estimated with the weighted Poisson estimator (Hamerslag and Immers, 1988). Results of this estimation are in Table 1.

### Table 1 Specification of parameters \( a_m \) and \( b_m \)

<table>
<thead>
<tr>
<th>Category</th>
<th>( a_m )</th>
<th>( b_m )</th>
</tr>
</thead>
<tbody>
<tr>
<td>car drivers</td>
<td>30</td>
<td>-0.42</td>
</tr>
<tr>
<td>car passengers (CA/CNA)</td>
<td>8</td>
<td>-0.40</td>
</tr>
<tr>
<td>PT passengers (CA)</td>
<td>25</td>
<td>-0.42</td>
</tr>
<tr>
<td>PT passengers (NCA)</td>
<td>115</td>
<td>-0.42</td>
</tr>
</tbody>
</table>

*The equilibrium factors*
To determine \( E_i \) and \( W_j \) we start with an initial estimate for \( F_\theta \), \( T_{i+} \) and \( T_{s+j} \). With a traditional Gauss-Seidel procedure (Furness iteration) (Hamerslag, 1977) \( \Omega \) and the products \( r_i \) and \( s_j \) can be calculated, where

\[
 r_i = d \ l_i \ E_i \quad \text{and} \quad s_j = d \ n_j \ W_j \quad (19)
\]

The values \( l_i \), \( n_j \), \( E_i \) and \( W_j \) are not known separately. However

\[
 T_{i+} = l_i^* \ E_i = \sum_j \ \Omega \ l_i \ n_j \ E_i \ W_j \ F_\theta \quad (14)
\]

so

\[
 l_i \ E_i = (l_i^* \ E_i) / (\sum_j \ \Omega \ n_j \ W_j \ F_\theta) \quad (20)
\]

Substituting \( s_j \), we get, after some algebraic operations,

\[
 l_i = \left\{ 1 / (\sum_j \ \Omega \ s_j \ F_\theta) \right\}^{1/(1+\theta)} \quad (21)
\]

and with (19)

\[
 E_i = r_i / l_i \quad (22)
\]

Analogously, we find

\[
 n_j = \left\{ 1 / (\sum_i \ \Omega \ r_i \ F_\theta) \right\}^{1/(1+h)} \quad (23)
\]

\[
 W_j = s_j / n_j \quad (24)
\]

So \( l_i \), \( n_j \), \( E_i \) and \( W_j \) can be determined separately.

According to (Metropolitan Transportation Commission, 1993) the total number of commuters in the San Francisco Bay Area is approximately 1,000,000. The initial values for \( T_{i+} \) and \( T_{s+j} \) were found by distributing the 1,000,000 peak hour trips homogenously over the 40 zones defined in the network specification. So

\[
 E_i = W_j = 25,000, \ for \ all \ i, j = 1, 2, \ldots, 40
\]
Elasticities of constraints

$1/(1+g)$ and $1/(1+h)$ are defined as the elasticities of the constraints. Notice that when $1/(1+g) = 1$ and $1/(1+h) = 1$ the model becomes the traditional double fixed constrained interaction-model (Hamerslag, 1977; Orthúzar and Willumsen, 1990). When $1/(1+g) = 0$ and $1/(1+h) = 0$ then $l_i = 1$ and $n_j = 1$ yields a direct gravity type model. Other values may also be possible.

In (Hamerslag, 1977) a sensitivity test has been carried out which showed that the values $1/(1+g) = -2$ and $1/(1+h) = -0.2$ showed reasonable results in determining locations of dwelling and employment. These values where also used in the computations for the San Francisco Bay Area.

3.3.3 Results

The concentrations of dwelling and employment determined with the land-use model with 72 nodes and elastic constraints have been compared with evening commuter data that has been surveyed at certain points in the Bay Area (Metropolitan Transportation Commission, 1993). Figure 6 and Table 2 show both the surveyed and the estimated data.

![Surveyed versus estimated trip ends](image)

*Figure 6 Surveyed versus estimated trip ends*
Table 2 Surveyed and estimated trip ends

<table>
<thead>
<tr>
<th>Aggregated zone</th>
<th>Surveyed commuters</th>
<th>Estimated employment T,</th>
<th>Surveyed / estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>San Francisco</td>
<td>555,447</td>
<td>514,078</td>
<td>1.08</td>
</tr>
<tr>
<td>Oakland</td>
<td>174,116</td>
<td>315,350</td>
<td>0.56</td>
</tr>
<tr>
<td>Redwood City</td>
<td>36,862</td>
<td>39,117</td>
<td>0.94</td>
</tr>
<tr>
<td>San Jose</td>
<td>319,397</td>
<td>234,039</td>
<td>1.36</td>
</tr>
<tr>
<td>Walnut Creek</td>
<td>47,065</td>
<td>55,843</td>
<td>0.84</td>
</tr>
<tr>
<td>Concord</td>
<td>55,514</td>
<td>37,535</td>
<td>1.48</td>
</tr>
<tr>
<td>San Rafael</td>
<td>35,919</td>
<td>28,358</td>
<td>1.27</td>
</tr>
<tr>
<td>Total</td>
<td>1,224,320</td>
<td>1,224,320</td>
<td>1.00</td>
</tr>
</tbody>
</table>

The comparison of the estimated employment and working population of aggregated zones of the specified road network with surveyed commuter data provides an acceptable validation of the model outcomes. The similarities between the data from both sources as shown in Figure 6 are sufficiently high to serve as input data in order to perform a realistic traffic flow calculation for the San Francisco Bay area.

3.4 Trip end model

From the determined concentrations of dwelling and employment the trip end model calculates the number of tripends from the employment and working population. The number of departures are calculated irrespective of the destination and the number of arrivals are calculated irrespective of the origin. The total adult population is calculated by

\[ \text{Population}_j = E_j / pw \] (25)

where
- \( \text{Population}_j \) total adult population in zone \( j \)
- \( E_j \) estimated employment in zone \( j \)
- \( pw \) ratio working population/adults (\( pw = .60 \))

Subsequently the number of trips by car and public transport is determined as follows:

\[ \text{trips} = \{ \alpha \cdot d_{car} + (\beta \cdot d_{car} + \tau) \cdot (1 - d_{car}) \} \cdot \text{phf} / pw \] (26)

where
- \( \alpha, \beta, \tau \) parameters (3.01, .14 and .24 respectively)
- \( d_{car} \) car density in cars/adult (\( d_{car} = .80 \))
- \( \text{phf} \) peak hour fraction (\( \text{phf} = .10 \))
The total number of trips is splitted out over the trip-purposes home-work, work-home, home-home and work-work

\[ D_{i,p} = \text{trips} \cdot T_{i+} \cdot f_p \]
\[ A_{j,p} = \text{trips} \cdot T_{+j} \cdot f_p \]

where
- \( p \) trip purpose home-work, work-home, home-home, work-work
- \( f_p \) fractions per trip purpose
- \( O_{i,p} \) departures from zone \( i \) for trip purpose \( p \)
- \( D_{i,p} \) arrivals in zone \( i \) for trip purpose \( p \)

The Figure below shows the lay-out of an origin-destination matrix. The right column and the bottom row depict the trip ends. Appendix B encloses the computed trip ends.

![Origin-Destination Matrix Example](image)

**Figure 7 Example of an origin-destination matrix with known (estimated) trip ends**

### 3.5 Distribution model

In the distribution model the origin-destination matrix with car trips and the origin-destination matrix with public transport trips are calculated by connecting the departures and arrivals determined in the trip end model. The same formulas as in the trip end model are used. However now the constraints are fixed, so the elasticities

\[ 1/(1 + g) = 1 \]
\[ 1/(1 + h) = 1. \]

After these calculation the cells of the origin-destination matrices for car and for public transport are filled completely. An example of the lay-out of a filled origin-destination matrix is given in Figure 8. The computed values of the cells are enclosed in appendix C.
Validation of the results of the OD-matrix estimation could have been possible as a twenty-four hours origin-destination matrix of a part of the studied network was available (MTC, 1993). However, the error introduced by disaggregation of this matrix into a peak hour matrix would be too high to allow meaningful comparison (Hamerslag and Immers, 1988). Moreover, our estimation method for the complete traffic flow calculation (including a OD estimation method) is also feasible under conditions when no detailed OD-data is available which is the case in most situations. The estimated OD-matrix has been produced to perform a traffic flow calculation in order to obtain a realistic peak hour traffic situation. This should be interpreted as a mean peak hour traffic situation and will be used to study the behaviour of the proposed probe vehicle data collection technique under realistic road traffic conditions. As it only serves as an evaluation method the demands with respect to its accuracy are not critical.

3.6 Traffic assignment model

We will now assign the end-to-end trips specified in the origin-destination (OD) matrix to the abstracted road network according to shortest routes. Where the OD-matrix only denotes the starting and ending point of each trip, the traffic assignment model calculates the route of these trips.

The ‘All-or-Nothing’ (AON)-assignment is the most simple traffic assignment model and it constitutes the basis for all other traffic assignment models. The shortest path between each OD pair is calculated only once and all trips between this OD-pair follow this particular route while all other routes remain unused. Especially in congested networks this leads to unrealistic results as alterations in link travel times effected by occurring congestion are not taken into account.

The ‘equilibrium assignment’ estimates shortest routes and traffic flows recursively, allowing for traffic reacting on congestion computed in previous iterations, which gives more realistic results. It can be shown that these iterations converge to a system equilibrium (Hamerslag, 1977). However, complete knowledge about the current and future travel times of all links
is assumed for all traffic participants, while in conventional road traffic without mature ATMIS this knowledge is incomplete.

In the ‘stochastic equilibrium assignment’ also alternative routes, which may be slightly longer than the shortest one, will be chosen as they are subjectively the shortest for certain travelers. During each iteration step, randomness is introduced to take into consideration this uncertainty.

In our investigations, we used the ‘stochastic equilibrium assignment’ as this traffic assignment method gives the most realistic results. A more detailed and more complete overview of traffic assignment models can be found in (Hamerslag, 1977; Ortúzar and Willumsen, 1990; May, 1990).

The amount of road traffic assigned to a certain road link $a$ depends on the travel resistance $Z_a$ of that road link (expressed as travel time), which directly evolves from the actual traffic load $q_a$ on that link and leads to an iterative computation of the ultimate link traffic flows as $q$ is a function of $Z$ which, in turn, is a function of $q$. Each iteration step $i$ of the applied stochastic equilibrium assignment consists of the following phases:

1. Determine the operational travel time $Z_a[i]$ on each link by randomizing the travel time $Z_a[i-1]$ obtained in the previous iteration step

   \[ Z_a[i] = \begin{cases} 
   Z_a[i-1] + c X_{a,i} / Z_a[i-1] & \text{for } i=2,3,\ldots \\
   Z_{a,0} & \text{for } i=1 
   \end{cases} \tag{28} \]

   where index $a$ denotes the road link, $c$ is a coefficient with a value between 0.2 and 0.8 (default 0.5) indicating the level of uncertainty and $X_{a,i}$ is a random variable from a normal probability distribution. In the initial iteration step ($i = 1$), the travel time equals the free flow travel time $Z_{a,0}$, i.e. the travel time if $q_a = 0$.

2. The randomized link travel resistances $Z_a[i]$ are used to calculate the shortest route paths for each of the origin-destination pairs specified in the OD-matrix. Each iteration step $i$ thus gives a different shortest route tree $B_i$ which is kept for further calculation.

3. The vehicle traffic specified by the OD-matrix is assigned to the shortest-route tree according to the All-or-Nothing assignment. This results in an estimate of the traffic flow on each link $a$, denoted as $q_o^+[i]$.

4. To take into account that in general travelers do not have perfect and complete knowledge of all link resistances only a portion $\alpha$ of all traffic will choose this (objectively) fastest route, while portion $1 - \alpha$ is assumed to take the previously obtained fastest route which is the fastest from a subjective point of view for these travelers.
So, in the $i$-th iteration step the traffic flow on road link $a$ is given by the linear combination

$$q_a[i] = \begin{cases} 
(1-\alpha)q_a[i-1] + \alpha q_a'[i] & \text{for } i=2,3,... \\
q_a'[1] & \text{for } i=1 
\end{cases}$$

(30)

5. In the last phase of each iteration step the travel time $Z_a$ on freeway link $a$ of the road network is estimated from the actual traffic load on that link using the BPR-function (Hamerslag, 1977)

$$Z_a[i] = Z_{a,0} \left( 1 + k \left( \frac{q_a[i]}{c_a} \right)^b \right)$$

(31)

where $q_a$ and $c_a$ are the actual and the steady flow ($k=1$) or practical ($k=0.15$) capacity of link $a$, respectively, expressed in vehicles per hour and $Z_{a,0}$ is the link travel time in the unloaded (free-flow) situation. Typically $b$ is in the range of 2 to 6. The maximum traffic load $c_a$ is determined by the network specifications.

This iteration scheme, involving a tree-search algorithm, a traffic assignment algorithm and the defined time-delay and uncertainty-relations, is repeatedly performed with level of uncertainty $\alpha = 0.2$ and $b = 4.5$ and $k = 1$ until the stop criterium (20 iterations) has been reached. The results provide an estimate of the link traffic flows $q_a = q_a[20]$.

Figure 9 illustrates the estimates of road traffic flows found according to this scheme. The different hatches represent $q_a / c_a$, i.e., the amount of vehicle traffic traveling on a road link in proportion to the capacity of that link and provide estimates of link travel times. A proportion of $q_a / c_a$ is 0.85 and higher indicates congested traffic. The estimated road link flows and road link travel times are enclosed in appendix D.

Detailed validation of the estimated flows with real traffic measurements did not yet take place. Rough comparison with the traffic situation on the freeways in the San Francisco Bay Area during average peak hours indicate that the estimations are realistic. For our purpose, creating a realistic traffic situation to evaluate the proposed probe vehicle concept, this verification is adequate as it pictures the "average" traffic situation in the Bay Area. We used a static traffic assignment the traffic situation is a steady state during the whole evening peak hour. Refinement of the model by using a dynamic traffic assignment algorithm would give better insight in time varying traffic fluctuations. This has not been conducted here because this study concentrates on possibilities of dynamic traffic data collection by probe vehicles rather than on processing dynamic traffic data into useful traffic information.
3.7 Message traffic model

The fleet of probe vehicles is modelled as a certain fraction \( \zeta \) \((0 < \zeta < 1)\) of the assigned vehicles, equipped with location and radio communication equipment. By means of simulations and analytical computations, the effectiveness of several probe vehicle concepts can be analyzed. For our study, the probe vehicles randomly generate traffic messages according to the slotted ALOHA transmission scheme (Walrand, 1991). These messages contain the position of the probe vehicle and its experienced link travel time. Due to the characteristics of this transmission scheme mutual collisions between messages transmitted from different probes will occur. To study the effect of these collisions on the spatial distribution of the traffic messages that will be received by a listening base station (the spatial throughput) we developed two computer programs, namely a simulation and a numerical evaluation of an analytical model. The latter will be called the 'analytical' approach in the following sections.
In the simulation model, probe vehicles are simulated sequentially as they are driving along the calculated shortest routes from their specific origin to their specific destination experiencing different travel times on different road links. Each probe transmits on average once every $T$ seconds, choosing a random time slot. For every transmitted message we store its time slot and after all probes have reached their destinations, the database with transmitted messages is sorted on time slot. Subsequently for each transmitted message, its received power and the joint local-mean interference power from other messages in the same time slot is obtained from Egli’s path loss model (Egli, 1957; see chapter 4). The local-mean signal-to-interference-plus-noise ratio then gives the probability of successful reception. A random experiment decides whether the message is received successfully at a base station. This is repeated for all base stations.

In the analytical model we focus on one particular road link $a$ of the freeway with a traffic flow of $q_a$ vehicles per hour and average travel time of $Z_a$ minutes. The expected number of vehicles present on road link $a$ is denoted as $D(a)$. Little’s Law expresses $D(a)$ as the product of the input traffic flow $q_a$ and the mean link travel travel time $Z_a$. For a penetration grade $\xi$ ($0 < \xi < 1$), the expected number of probe vehicles that are present in segment $a$ is denoted as $\lambda_a$, where

$$\lambda_a = \xi q_a Z_a$$  \hspace{1cm} (32)

Each probe transmits on average once every $T$ seconds, choosing a random time slot. Since the number of vehicles is large, the message transmission process becomes Poissonian. The mean number of messages per time slot from segment $a$ is

$$G(a) = \frac{\xi q_a Z_a}{T f_s}$$  \hspace{1cm} (33)

where the sampling rate $f_s$ denotes the number of slots preserved for probe vehicle reports per second.

According to the road traffic assignment model, the traffic density only changes at junctions, on-off ramps or intersections, so the road network computations are been performed with relatively large links. The propagation characteristics can however greatly vary over the length of such link. Therefore, the analytical evaluation of the radio network considers shorter road segments, subdividing the longer road links.

Both the simulation and the analytical model calculate the spatial throughput using the same models and assumptions to be formulated in the next chapter.
4 Radio Network Analysis

An essential element in the probe vehicle system is the wireless data link from many vehicles to one or more base stations. This is studied in this chapter.

4.1 Radio Coverage and Effect of Collisions

4.1.1 Channel and Capture Model

We address a radio channel with Rayleigh fading, caused by multipath reception (Jakes, 1974), and path loss according to Egli's semi-empirical model for UHF groundwave (Egli, 1957). We consider one particular 'test' signal, denoted by index 0, attempting to reach one of the base stations, denoted as $A, B, \ldots$. We assume that a message is received successfully if and only if the signal-to-joint-interference (C/I)-ratio of this signal at receiver $A$ is above the receiver threshold $z$. This event is denoted as $A_0$. For narrowband transmission, with incoherent Differential Quadrature Phase Modulation (DQPSK), $z$ is typically in the range $4 \ldots 10$ (6 \ldots 10 dB). Because of fading and path loss, the received powers are random variables. Assuming constant received power at least during a packet transmission time, the probability of successful reception is (Linnartz, 1993)

$$
\Pr\{p_{A,0} > zp_{A,i}\} = \int_{-\infty}^{\infty} f_{p_{A,i}}(x) \int_{-\infty}^{\infty} f_{p_{A,0}}(y) dy \; dx \tag{34}
$$

where $p_{A,i}$ is the joint interference-plus-noise power at receiver $A$ while $p_{A,0}$ is the instantaneous power of the wanted signal at $A$ and $f_p$ denotes a probability density function. In a Rayleigh-fading channel, the instantaneous power is exponentially distributed around the local-mean power (Jakes, 1974) (Linnartz, 1993). The local-mean power will be expressed later in terms of the propagation distances, antenna heights, etc.

4.1.2 Product-Form Expression for Capture Probability

Assuming the local-mean power $\bar{p}_{A,0}$ to be known in (34), one can write

$$
P\{p_{A,0} > zp_{A,i}|\bar{p}_{A,0}\} = \int e^{-\frac{zp_{A,i}}{\bar{p}_{A,0}}} dF_{p_{A,i}}(x) \Delta \chi\left\{p_{A,i}, \frac{z}{\bar{p}_{A,0}}\right\}. \tag{35}
$$

where $\chi\{f, s\} = \mathbb{E}[\exp\{-sp_i\}]$ denotes the characteristic function (cf) of the joint interference power $p$, the point $s$. For incoherent cumulation of $N$ statistically independent signals, the pdf of the joint interference power is the $N$-fold convolution of the pdf of the individual powers. Laplace Transformation then results in the multiplication of $N$ factors, each containing a characteristic function of the received power from an individual component. So,

$$
P\{p_{A,0} > zp_{A,i}|\bar{p}_{A,0}\} = \prod_{i=1}^{N} \chi\left\{p_{A,i}, \frac{z}{p_{A,0}}\right\}. \tag{36}
$$
The total number of probe vehicles in the system is denoted as $N$, with index $i$ denoting one particular probe, randomly transmitting with probability $P(i_{\text{on}})$ where $i_{\text{on}}$ denotes the event that transmitter $i$ is switched on. Since $p_{A,i} = 0$ if $i$ is OFF, the cf becomes

$$\chi \{p_{A,i} > s\} = 1 - P(i_{\text{on}}) + E[e^{-sp_{A,i}}|i_{\text{on}}]$$

(37)

We denote the conditional probability of unsuccessful reception $A_0$ given that (only) interfering terminal $i$ is active as $W_{A,i}$ ($0 \leq W \leq 1$), called the 'vulnerability weight' factor. Thus, $W_{A,i} = 1 - E[\exp\{-sp_{A,i}\}|i_{\text{on}}]$. We model the receiver noise floor as an additional interfering signal with constant, i.e., non-fading power For a finite population of $N$ terminals, the probability that a 'test' packet captures receiver $A$, given all locations of terminals ($i = 0, 1, 2, \ldots, N$) becomes

$$P(A_0 | \text{all } r,f_i) = P_{NA} \prod_{i=1}^{N} \left[ 1 - W_{A,i} P(i_{\text{on}}) \right],$$

(38)

where $P_{NA}$ denotes the probability that the test signal (at distance $r$) is received successfully at $A$ in an interference-free but noise-limited channel. Thus $P_{NA}$ describes the 'coverage' of probe vehicle signals.

It has been shown analytically (Linnartz, 1993) that for a Rayleigh fading channel without shadowing, thus with known local-mean powers, the weight factor becomes

$$W_{A,i} \triangleq P(p_{A,0} > p_{A,i} | p_{A,0}, p_{A,i}) = \frac{\overline{p}_{A,i}}{\overline{p}_{A,i} + p_{A,0}}$$

(39)

The local-mean power $\overline{p}_{A,i}$ of signal $i$ at receiver $A$ is taken according to Egli's model

$$\overline{p}_{A,i} = G_A(\phi_i) P_i h_i^2 h_A^2 r_i^{-4} \left(\frac{40 \text{ MHz}}{f_c}\right)^2$$

(40)

where $G_A(\phi)$ is the antenna gain pattern of receiver $A$ in the direction $\phi$ of the probe, $P_i$ is the (ERP) transmit power by probe vehicle $i$, $h_i$ is the antenna height at the probe, $h_A$ is the base station antenna height, $r_i$ is the propagation distance and $f_c$ is the carrier frequency. We have limited our investigations to omni-directional antennas with $G_A(\phi) = 1$ for all $\phi$. We assume that the vehicle antenna heights and effectively radiated powers are equal for all probes.

4.1.3 Signal Coverage

Signals experience a noise floor with spectral power density $F k T_0$, where $F$ is the noise factor (in absolute units; not in dB) accounting for additional man-made noise, above the thermal background noise, $k$ is Boltzmann's constant and $T_0$ is the noise temperature (290 K). Requiring that the energy per bit is a factor $z$ above the noise floor gives, in an interference-free situation,
\[ P_{NA} = P(P_{A0} > z F k T_0 r_b) = \exp \left\{ - \frac{z F k T_0 r_b}{P_{A0}} \right\} \]  

where \( r_b \) is the channel bit rate.

### 4.2 Throughput of ALOHA scheme

We average the capture probability over all possible locations of the \( n \) active interferers. The \( n \)-dimensional integral over the product can be written as the product of integrals since each (weight-) factor contains the local-mean power of only a single interfering signal. Moreover, for i.i.d locations of \( n \) interferers, the probability of the test packet capturing receiver \( A \) is

\[
P(A_0 | n, r_0, \phi_0) = P_{NA} \int \cdots \int \left[ \prod_{i=1}^{n} 1 - W_{A,i} \right] dF_{r_1 \phi_1} dF_{r_2 \phi_2} \cdots dF_{r_n \phi_n}
\]

\[ = P_{NA} \left[ 1 - \int W_{A,i} dF_{r_1 \phi_1} \right]^n \]

where we used \( P(i_{ON}) = 1 \) for interferers known to be active.

In order to study the Poisson field of interferers, we that the limit for the number of terminals \( N \to \infty \) under the condition that the total offered traffic \( N P(i_{ON}) = G = \Sigma G(a) \) is a constant determined by the total number of probes in the road network and their transmission rate. For the interfering message traffic being a spatial Poisson process, the capture probability becomes a sum of exponential functions, viz.,

\[
P(A_0 | r_0, \phi_0) = \sum_{n=0}^{\infty} \frac{G_i^n}{n!} e^{-G_i} P(A_0 | n, r_0, \phi_0)
\]

For an (practically) infinite population of probe vehicle terminals, data packets are transmitted with a spatial distribution \( G(a) \) where \( a \) denotes the freeway segment at a distance \( r_a \) from the base station; see (33). Thus

\[
P(A_0 | r_0, \phi_0) = P_{NA} \exp \left\{ - \sum_a W_{A,a} G(a) \right\}
\]

\[ = P_{NA} \exp \left\{ - \sum_a \frac{z r_a^4}{z r_0^4 + r_a^4} G(a) \right\} \]

The successful throughput, in messages per slot from segment \( a \) becomes

\[
S(a) = G(a) P(A_0 | r_0, \phi_0)
\]

\[ = \frac{\zeta q_a Z_a}{f_s T} P_{NA} \exp \left\{ - \sum_a \frac{z r_a^4}{z r_0^4 + r_a^4} \zeta q_a Z_a \right\} \]

\[ \frac{z r_a^4}{z r_0^4 + r_a^4} = f_s T \]

31
Throughput per link per second is the sum of the throughput from all its segments, $\Sigma_{\omega} f_s S(\omega)$.

### 4.3 Capture Probability with Site Diversity

We return to our expressions conditional on known locations of terminals. The probability $P(k_{ON} \mid A_0)$ that a terminal $k$ has transmitted an interfering packet, given that the wanted packet captures base station $A$, is found from Bayes rule

$$P(k_{ON} \mid A_0, \{\text{all } r, \phi\}) = \frac{P(A_0 \mid k_{ON}, \{\text{all } r, \phi\})}{P(A_0 \mid \{\text{all } r, \phi\})} P(k_{ON}), \quad (46)$$

where $P(A_0 \mid k_{ON})$ is the probability that a wanted segment is received successfully, despite the knowledge that interferer $k$ was active, but without any information on the activity of other terminals. We conditioned, for time being, on exact knowledge of terminal locations $(r, \phi)$.

The conditional probability that the test packet captures receiver $B$ given that it also captures receiver $A$ is found as

$$P(B_0 \mid A_0, \{\text{all } r, \phi\}) = \prod_{k=1}^{N} P(1 - W_{B,k} P(k_{ON} \mid A_0, \{\text{all } r, \phi\}))$$

$$= P_{NB} \prod_{k=1}^{N} \left(1 - \frac{1 - W_{A,k} P(k_{ON})}{1 - W_{A,k}} W_{B,k} P(k_{ON})\right), \quad (47)$$

where $P_{NB}$ is the probability that the received signal fails to exceed the noise floor at receiver $B$. The probability that a packet from terminal 0 captures at least one of the two base stations $(A_0 \cup B_0)$, given the position of all terminals $i (i = 0, 1, 2, \cdots, N)$, equals

$$P(A_0 \cup B_0 \mid \{\text{all } r, \phi\}) = P(A_0 \mid \{\text{all } r, \phi\}) + P(B_0 \mid \{\text{all } r, \phi\}) - P(A_0 \mid \{\text{all } r, \phi\}) P(B_0 \mid A_0, \{\text{all } r, \phi\}). \quad (48)$$

Inserting (38) and (47) in (48) gives

$$P(A_0 \cap B_0 \mid \{\text{all } r, \phi\}) = P_{NA} P_{NB} \prod_{i=1}^{N} (1 - W_{A,B,i} P(i_{ON})), \quad (49)$$

where we introduced the joint weight function

$$W_{A,B,i} \Delta W_{A,i} + W_{B,i} - W_{A,i} W_{B,i} \quad (50)$$

The $W_{A,B,i}$ can be interpreted as a factor weighing the disturbance caused by an interfering packet signal from position $x_i$ to a reception of a data packet by terminal $j$ at the two base stations $A$ and $B$ simultaneously. For the interfering traffic being a spatial Poisson process, the capture probability becomes a sum of exponential functions, viz.,
\[ P(A_0 \cup B_0 \mid \text{all } r, \phi) = \sum_{n=0}^{\infty} \frac{G_r^n}{n!} e^{-G_r} \cdot P(A_0 \cup B_0 \mid \text{all } r, \phi) \]

which in fact can be written as three sums, the first containing the throughput of receiver \( A \), the second representing the throughput of receiver \( B \) and the third accounting for the 'joint' throughput.

We take the limit for \( N \to \infty \), assuming fixed offered traffic, and express the capture probability terms of \( G(r, \phi) \). The result can be written in the elegantly structured expression

\[ P(A_0 \cup B_0 \mid r_0, \phi_0) = P_N A \exp\left\{ -\sum_a W_{A,i} G(a) \right\} + P_N B \exp\left\{ -\sum_a W_{B,i} G(a) \right\} - P_N A P_N B \exp\left\{ -\sum_a W_{A,B,i} G(a) \right\} \]

This equation offers a mathematical expression for the probability of capturing at least one receiver, given an arbitrary spatial distribution of the offered packet traffic, described by \( G(a) \). The expression contains three terms: the first two terms are of the form of the capture probability for the individual receivers \( A \) and \( B \) respectively; the third term compensates for successful reception at two receivers simultaneously.

### 4.4 Message Throughput per Link

For two-branch site diversity, the throughput of segment \( a_0 \) is

\[ S(a_0) = G(a_0) P(A_0 \cup B_0 \mid r_0, \phi_0) \]

\[ = \frac{\zeta q_a Z_a}{f_s T} \left[ P_N A \exp\left\{ -\sum_a W_{A,i} \frac{\zeta q_a Z_a}{f_s T} \right\} + P_N B \exp\left\{ -\sum_a W_{B,i} \frac{\zeta q_a Z_a}{f_s T} \right\} - P_N B P_N A \exp\left\{ -\sum_a W_{A,B,i} \frac{\zeta q_a Z_a}{f_s T} \right\} \right] \]

A similar expression can be derived for more cooperating base station. In our computations, we considered the nearest three base stations. In the case that also a fourth base station receives messages from a particular link, mostly the common throughput, containing only duplicates, appeared to be large leaving little additional throughput. The number of messages uniquely received at base station 4 appeared negligible. We found that the computations for three base stations appeared to be sufficiently accurate (Linnartz and Westerman, 1994).

It can be seen from the above expression that for a system of given architecture, the ratio of the penetration grade \( \zeta \) and the transmission interval \( T \) is a crucial parameter. The throughput of successful messages initially increases with increasing \( \zeta/T \) but decreases if \( \zeta/T \) becomes too large. If the penetration grade is know, the system can control \( T \) by sending broadcast messages to the probes. The optimum value of \( T \) however depends on the distance between
the receiving base station and the link of interest. This will be illustrated in later sections. For a detailed theoretical discussion of the performance of wireless ALOHA nets we refer to (Linnartz, 1993).

If a link contains several segments, the total link throughput is the sum of the throughput of all segments.
5 Case Study: Road Traffic and Message Traffic in the San Francisco Bay Area

We developed two computer tools, a simulation and an analysis, according to the specifications as described in chapter 3 and 4. These programs have been applied to the San Francisco Bay Area. In this chapter the results will be presented and discussed.

5.1 Input of System Parameters

In the first phase of the program, the link traffic flows $q_a$ and link travel times $T_a$ are computed iteratively as described in chapter 3. For these iterations, a delay and a cost function are used. The delay function estimates the link travel times from earlier computed link flows and the cost function represents the non-rational human choice caused by uncertain and incomplete information by randomizing the estimated travel times. According to the randomized travel times, car drivers choose the shortest route from origin to destination. The delay and the cost function contain several variables. These variables have default values and can be adjusted if desired. Figures 10 and 11 show the input screens that allow these variables to be changed in the computer program.

The delay function is given by:

$$T_{flow} = T_{min} \times (1 + \alpha \times (Flow/\text{Cap}) \times B)$$

The random cost function is given by:

$$T_{rand} = T_{o} + \alpha \times \text{RND} \times \text{SQR}(T_{o})$$

- $T_{flow}$ is the number of trips per link
- $\text{Cap}$ is the steady flow capacity
- $T_{min}$ is the link cost on a unloaded link
- $T_{flow}$ is the link cost on a loaded link
- $\alpha$ is a random number from a normal probability distribution
- $T_{o}$ is the input generalized travel time
- $T_{rand}$ is the randomized time

In Figure 12, the computed road link flows and link travel times are depicted using the default values of delay and cost function. In Figure 12, the thickness of the links presents the expected number of vehicles that are present on that link. The hatch of the links denotes the travel time (delay) on that link according to the legend given in the left side of Figure 12.

A certain percentage $\gamma$ of the assigned vehicles are designated as probe vehicles and regularly transmit traffic messages according to the discussed transmission scheme. The default system parameters are realistic for transmission over a UHF mobile radio channel, and can be altered if desired. Figures 13 and 14 show the input screens for specifying parameters in the path loss model and in the model for calculating the probability of successful reception. A theoretical derivation of these models is given in chapter 4.
Delay function: 
\[ T_{\text{flow}} = T_{\text{min}} \cdot (1 + 1 \cdot \frac{\text{FLOW}}{\text{Cap}})^4 \]

\[ \text{FLOW}(i) = \text{FLOW}(i') \cdot (0.2) + \text{FLOW}(i-1) \cdot (1 - 0.2) \]

**Figure 12** Computed road traffic flows during the evening rush hour on the freeway network in the San Francisco Oakland - San Jose Bay Area

**Figure 13** Input screen of the path loss model

**PATH LOSS (Model Eqn)**

\[ P_r = \left( \frac{(H_r \cdot 2 \cdot H_t \cdot 2)}{R_o \cdot 4} \right) \times G_r \times G_t \times \left( \frac{P_t}{48 \text{ MHz}} \right) \cdot f^2 \]

- \( P_r \) = received local-mean power
- \( H_r \) = base station antenna height
- \( H_t \) = vehicle antenna height
- \( R_o \) = distance between vehicle and base station
- \( G_r \) = base station antenna gain
- \( G_t \) = vehicle antenna gain
- \( P_t \) = vehicle transmit power
- \( f \) = carrier frequency

**Figure 14** Input screen of the model for calculating the probability of successful reception

\[ P(Pr > z \cdot H) = \exp \left( -z \cdot k \cdot T_o \cdot F \cdot \text{rb} / P_r \right) \]

- \( P_r \) = received local mean power
- \( z \) = receiver threshold
- \( k \) = Boltzmann\'s constant
- \( T_o \) = 290 Kelvin
- \( F \) = non-natural noise factor
- \( \text{rb} \) = bit rate
- \( 1 \) = 88E+04 bit/s

Each probe transmits, on the average, once every \( T \) seconds, choosing a random time slot. The number of time slots per second that are reserved for probe vehicle messages is denoted as \( f_r \). These three parameters \( f_r \), \( T \) and \( f_r \) are the main variables which can be adjusted to specific applications. Our default value of \( f_r = 18 \) slots per second corresponds to a design of a single-
channel IVHS communication architecture (Linnartz, 1994). This design also allows other services to be offered through 30 kHz of radio bandwidth. A higher sampling rate may claim too much of the available spectrum at the expense of other services and a lower sampling rate appears to significantly decrease the efficiency of the probe vehicle concept. This leaves the parameters $\zeta$ and $T$ to be varied in the transmission scheme, but our earlier analysis showed that in the relevant expressions, the ratio $\zeta/T$ occurs as the important parameter. If the probe penetration grade is sufficiently large, only the average number of messages per link matters. At large penetration grades, the interval between transmission can be reduced to optimize the number of transmitted messages per unit of time.

After the traffic flow has been calculated, the spatial throughput of messages from the probes is calculated. This can be done either by means of the numerical evaluation of analytical expression or by simulation. The theory behind both methods has been described in chapter 3.

5.1.1 Numerical Evaluation of the Analytical model

Figure 15 shows the screen to change the parameters $\zeta$, $T$, and $f_r$ in the 'analytical' program. The flows in this figure are the computed road traffic flows and are the same as those pictured in Figure 12. Figure 16 addresses default transmission parameters $\zeta = 0.1\%$, $T = 60$ seconds and $f_r = 18$ probe time slots per seconds. In Figure 16, the thickness of the links denotes the expected number of traffic messages $G(a)$ transmitted from that link. This amount is proportional to the amount of link vehicle traffic and the link travel time. The hatch of the links in Figure 16 represent the amount of vehicles that are present on that link and should be read according to the legend given in the left side of Figure 16 (mark the differences in interpretation of thickness and hatches between Figure 15 and Figure 16!)

In the screen of Figure 16, listening base stations can be located by moving the cursor. The square indicates the current location of the cursor and the X, Y-coordinate of this location is shown in the bottom-left side of the screen. By pressing the < + > key the base station can be placed at the current cursor location.
The height of the base station (default $h = 100$ meters; Figure 17) and the antenna gain pattern should be specified next. Only the omni-directional antenna gain pattern is available in the program now, but other patterns could be included if required. The two circles around the base station indicate the area where, for an interference-free time slot (no collision), the probability of successful reception is .9 (inner circle) and .1 (outer circle). The radius of both circles and the surface enclosed by them is shown in the top-right hand corner of the screen. After the base stations have been placed the throughput for each road link at the different base stations can be calculated by pressing the $<$C$>$ key.

At this stage, the program has sufficient information to compute link throughput. In order to present the results, it initially asks criteria for the number of road traffic samples per road link per minute required to perform reliable IVHS services. The program user can specify thresholds for Automatic Incident Detection, Advanced Traffic Management/Traveler Information Systems and (traditional) Traffic Management/Traveler Information Systems.
As illustrated in Figure 19, we have provisionally set the requirements to 3 messages per link per minute for AID, 5 messages per link per 5 minutes for ATM/IS and 1 message per link per 5 minutes for (traditional) TM/IS. In this way the hatches give some indication of how well AID and ATM/IS can be performed. However, the number of probe messages per link per minute required to perform reliable ATM/IS, is to be investigated further. We speculate the number of messages required per link may depend on the expected vehicle density and velocity, rather than being a constant number as in our case.

The throughput per road link (expressed in received traffic messages per minute from that road link) is calculated by taking into account the possibility that one traffic message can be received by more than one base station (site diversity: see par. 4.4). The analysis considers only the three nearest base stations for each link, as farther base stations contribute little to the throughput.

For each road link, the distance to each of its three nearest base stations is calculated. Using the functions discussed in chapter 4, for each base station, the number of messages received from that specific link per unit of time is determined taking into account site diversity. These results are written to an output file. So, for each road link, the expected number of traffic messages received per minute (= 18 times 60 time slots) is pictured (see Figure 20).

Finally, for all road links the throughput is known. These results are pictured in Figure 21 (default values, specified numbers and locations of base stations, etc.). The thickness of the links in Figure 21 again denotes the amount of transmitted traffic messages from that link. The hatch of the links in Figure 21 represents the amount of traffic messages that were received from that link and should be read according to the legend given in the left side of the Figure.

5.1.2 Simulation

In the simulation, the parameters $\xi$, $T$ and $f_s$ can be altered in a separate menu. After having specified the probe vehicle penetration grade $\xi$ the number of probe vehicles on the road network is calculated.

The screen shows:
- the number of vehicles from origin $i$ to destination $j$
- the number of probe vehicles from origin $i$ to destination $j$
- the cumulative number of vehicles
- the cumulative number of probe vehicles.

Figure 21 Computed spatial throughput per road link expressed in received traffic messages per minute
After specifying the number of time slots per second reserved for uplink probe vehicle datatransmissions $f_s$ and the mean time interval between successive transmission $T$ the probes are sequentially simulated traveling along their (subjective) shortest routes from their origin $i$ to their destination $j$ experiencing different link travel times $Z_a$ and transmitting a traffic message every $T$ seconds. Figure 22 pictures this process.

As the probes are simulated sequentially, a database of transmitted traffic messages is built in order of probe vehicles. This database subsequently is sorted on time. Plotting the items in this time-sorted database shows the picture of the traffic flow as it is transmitted by probes. Figures 23 and 24 show two phases in this process. Each dot represents a transmitted traffic messages. The hue of the dots indicates the perceived speed according to the legend in the bottom right corner. In the screen in Figure 24 the base stations can be placed in the same way as described under the analytical model.

Simulation of spatial throughput occurs in a different way as in the analysis. In the analytical model we focused on one particular road link, in the simulation model we focus on one particular traffic message and calculate the probability that it will be received by any of the listening base stations. A random experiment is performed to determine whether the packet captures the receiver. The results of the throughput calculation for one base station are depicted in Figure 25. The large dots represent a received traffic message, the smaller dots represent a transmitted traffic message that could not capture one of the base stations. Again the hue of the dots indicates the perceived speed.
Combining the throughput per base station and accounting for site diversity gives the total amount of received traffic messages, for this specific situation (default values, specified numbers and locations of base stations, etc.), as pictured in Figure 26. As all base stations are connected to a traffic control center by a wired infrastructure Figure 26 represents the information that has come available in this traffic control center. Based on this information traffic control strategies can be deployed.

As the results of the simulation model are subject to stochastic variations from sample to sample, the analytical method has been used for further computations.

5.2 Results

The equations derived in chapter 4 with respect to the spatial message throughput showed that a very important parameter influencing the number of messages received from a road link \( a \) by a certain base station is the distance from this link \( a \) to this base station, \( r(a) \). In order to analyze the effect of the distance from a link to a base station on the throughput of that link, base stations have been placed at three different locations in the Bay Area (see Figure 27). The first location (the top one: \( A \)) represents a position with heavily occupied road links very nearby. The second location (the middle one: \( B \)) is surrounded by several road links with high and several road links with low traffic occupancies. The third and last base station (the bottom one: \( C \)) is located in a fairly quiet traffic environment.
Throughput as function of distance and traffic density (per road link)

Figure 28 Throughput in base station A per road link as function of traffic density

Figure 29 Throughput in base station B per road link as function of traffic density

Figure 30 Throughput in base station C per road link as function of traffic density
In Figures 28, 29 and 30 the quantitative relationship between the distance from a road link to a base station, \( r(a) \), the number of transmitted traffic messages per time slot from that road link, \( G(a) \), which is proportional to the link traffic density \( D(a) \), and the number of received traffic messages per time slot from that road link by the base station, \( S(a) \), has been graphed. The probe vehicle penetration rate \( \xi \) amounts \( 0.1\% \), the number of time slots preserved for probe vehicle transmissions \( f_s \) is 18, the mean time interval between transmissions \( T \) is 60 seconds and site diversity is not taken into account here.

The graphs show that the number of messages received in the base stations from a road link extensively decreases as the distance between base station and road link increases. This effect occurs in all three situations nearly to the same extent and the maximum distance from where transmitted messages are still received amounts approximately 25 kilometers. We found empirically that when configuring a traffic monitoring network using \( 0.1\% \) probe vehicles \((T = 60, f_s = 18, \text{etc.) the radius of the coverage area (cell) of the base stations should not exceed 25 kilometers. The practical radius however will be smaller depending on the requirements of the traffic data quality (e.g. number of samples per road link per minute).}

The Figures 31, 32 and 33 graphically reproduce the same situations as in the former graphs where independent base stations were located at three different positions A, B and C respectively. The fringe of the circles now indicate a cell-size of about \( 350 \text{ km}^2 \) (radius \( \approx 10 \text{ km} \)). The thickness of the road links in the Figures represent the amount of car traffic and the hatches of the links indicate the computed number of received traffic messages from the road links according to the specified AID and (A)TM/IS requirements. Also these Figures show that from a relatively large area around the base station traffic messages are received. As the distance to the base station increases the number of received messages decreases.
An increasing probe vehicle penetration rate intensifies this effect. As more probe vehicles are present on the road network that transmit traffic messages, the messages from links close to the position of the base station repress the messages transmitted from farther links. This means that an increasing probe vehicle penetration rate limits the practical cell-size of a base station even more.

This is also shown by the next three graphs (34, 35 and 36) that show the same situation as in graph 21 (relation between \( r(a) \), \( G(a) \) and \( S(a) \) for a base station placed at location A and a penetration rate \( \xi \) of .1%) but now with a probe vehicle penetration rate of 1%, 5% and 10% respectively.

In the first case, with a penetration grade of .1% (graph 21), messages could be received over ranges of 25 km. The total interfering message volume generated within a range proportional to the distance of the link under consideration appeared a relevant measure for the probability of successful reception. This observation corresponds with a vulnerability model proposed by (Abramso, 1977), assuming that a packet from distance \( r_p \) is always successful if no interference occurs with a circle of radius \( c r_p \). In Figure 21, the total offered traffic is small enough to ensure a large throughput. That is, for a penetration rate of only .1% the total number of offered messages remains relatively small, so the communication needs are well within the channel capacity. The channel capacity can be used more effectively at a large penetration rate and a proportionally larger offered traffic. This increases the throughput for nearby road links, but the increased message intensity is experienced as interference by signals from more remote probes. Graph 34 - 36 illustrate this effect for an increasing penetration rate. Note that the scales on the right Y-axes are not the same in these curves.

If the penetration rate exceeds approximately 5% (\( \xi = 0.05 \)), the channel capacity limits the throughput of the messages. Initially only the throughput from remote road links is affected. Virtually no messages are received from links at more than, say, 14 km. Throughput from nearby links, however, is large, particularly if they contain dense road traffic. As seen from Equation (45), the throughput per link per base station is of the generic form

\[
S(a) = c_1 \frac{\xi}{T} \exp\{-c_2 \frac{\xi}{T} r(a)^2 / T\}
\]

where the constants \( c_1 \) and \( c_2 \) depend on the road network lay-out its traffic load. Hence, throughput initially always increases with increasing \( \xi \) but decreases rapidly beyond a certain optimum. For a network with multiple receivers at various distances and site diversity, the relation between the capture probability and the penetration rate may be more complicated.

Simplified theoretical models with perfectly uniform offered traffic suggest that the total throughput remains constant if the offered traffic per unit area increases. However, with increasing offered traffic, only messages from very close to the receiver contribute to the throughput. More realistic models, assuming that vehicles are always at least separated some minimum distance from the receive antenna, show a vanishing throughput with increasing offered traffic. For the SF Bay Area modelled in graph 36, beyond a penetration rate of 10%, the total number of traffic messages captured by base stations decreases substantially, unless the transmission interval \( T \) is increased appropriately.
Throughput as function of distance and traffic density (per road link)

Figure 34 Throughput in base station A per road link as function of traffic density at a penetration rate of 1%

Throughput as function of distance and traffic density (per road link)

Figure 35 Throughput in base station A per road link as function of traffic density at a penetration rate of 5%

Throughput as function of distance and traffic density (per road link)

Figure 36 Throughput in base station A per road link as function of traffic density at a penetration rate of 10%
The graphs 37, 38 and 39 address our special case in detail. The number of received traffic messages from four specific road links at certain distances (approx. 3, 6, 9 and 12 kilometers) from a base station placed at positions A, B and C respectively are depicted as a function of probe vehicle penetration rate. For all three locations we see that, to a certain extent, the throughput of each road link increases at an increasing probe vehicle penetration rate. If the base station is located in an area with heavy traffic, as is the case at location A, the throughput of the remote links (12 and 9 km resp.) already decreases at a low penetration rate (1 and 2\% resp.). At a penetration rate of about 4\%, the throughput of more nearby links (6 km) decreasing also. Monitoring the road link within small radius (about 3 km) continues to improve upto a relatively large penetration rate (10 \%) and remains possible until about 50\% of all vehicles serves as a probe and transmit once every minute.

For base stations placed at position B (graph 37) and C (graph 38), we see similar effects. Many traffic messages are transmitted from the road link 6 kilometers separated from the location B. For this reason the number of traffic messages received from this link continues to exceed the number of messages received from the closest link. For a high penetration rate of about 25 \%, the total number of competing messages is too high and the performance of monitoring the remote, heavily occupied road link severely diminishes in favour of the nearest road link.

In the latter situation (base station at position C) there are relatively few messages from nearby the base station so more traffic messages from links further away are received. However, at an increasing penetration rate, traffic messages from the nearest link conflict harmfully with remote messages. These graphs show that it is difficult to make a general judgement about the effect of an increasing probe vehicle penetration rate.
Throughput per road link at several distances as function of penetration rate

Figure 37 Throughput per road link at several distances as function of penetration rate; location A

Throughput per road link at several distances as function of penetration rate

Figure 38 Throughput per road link at several distances as function of penetration rate; location B

Throughput per road link at several distances as function of penetration rate

Figure 39 Throughput per road link at several distances as function of penetration rate; location C
At any particular distance, the number of successfully received messages decreases if the offered message traffic increases without bounds. Except in the unrealistic case that transmitting probes can be arbitrarily close to the base station, the total throughput received at a base stations eventually also approaches zero. Graph 40 shows the total throughput per base station as function of the probe vehicle penetration rate.

The Figures 41, 42, 43 and 44 graphically reproduce the same situations as in the former graphs and examines the influence of an increasing probe vehicle penetration rate.
6 Discussion

Properties of the probe vehicle system using ALOHA transmission scheme
The graphs in chapter 5 reveal that the road areas that can be monitored adequately, critically depend on the location of the receiving base stations. Moreover it depends on the spatial distribution of the offered traffic, i.e., from the lay-out of the road network, the amount of road traffic and the intertransmission time chosen by the probe vehicles. At a relatively low probe vehicle penetration grade ($\xi < \pm 1\%$) the distance between a certain road link $a$ and the receiving base station, $r(a)$, is the most important factor influencing the throughput of that road link $S(a)$, followed by the (probe) vehicle density on that road link, $D(a)$. If the probe vehicle penetration grade becomes sufficiently large ($\xi > \pm 1\%$), the size of coverage area of a base station diminishes and the number of participating probe vehicles on a road link within the 'cell' becomes less critical, particularly since the intertransmission interval can be optimized to avoid loss in throughput due to excessive message collisions. These effects are an evident consequence of the modelled (slotted) radio ALOHA system.

Properties of probe vehicle systems in general
From a traffic engineering prospective, the possibly distorting effect that road density and distance to the base station have on throughput is very relevant to develop new methods for road monitoring using probe vehicles. For traffic control, continuous traffic monitoring, reliable estimation of actual and future traffic conditions and quick detection of disturbances of the traffic flow are essential. To this end, two categories of parameters can be distinguished:
1) vehicle density (or flow), expressed in $\text{vehicles/km}$ and $\text{vehicles/hour}$, respectively, and 2) travel time (or speed), expressed in $\text{hour}$ and $\text{km/hour}$, respectively.

The parameters of the first category can only directly be determined from measurements of (nearly) all vehicles on a road link or segment\(^1\). This can be explained by optimistically assuming that the penetration grade $\xi$ of probes is exactly known. Typically, $\xi$ is in the order of 0.1 to 10 with low $\xi$ prevailing during the early introduction phases. The traffic flow is monitored using small and discrete time periods of for instance 1 or 5 minutes. Due to fluctuations in the composition of the traffic flow the actual number of probes present on a road link in a certain time period will vary to such a extend that no reliable statistical conclusions on traffic density can be drawn from received probe vehicle messages. Moreover, the capture process of the proposed (slotted) ALOHA transmission scheme introduces randomness in the received throughput as not all transmitted traffic messages will be received by a base station. Furthermore, the offered traffic $G(a)$ and car density $q_a$ are not a function of the throughput $S(a)$. Even if one could know the statistics of the interfering traffic from other road links exactly, two solutions for $q_aZ_{\text{ao}}$ exists in Equation (45).

\(^1\)The ALOHA system appears fundamentally unsuitable for monitoring category 1 parameters, because of the near-far effects in the spatial throughput. There is however a clear case where a random access radio link can estimate density or flow, if the receiver is located along a freeway, and considers only messages generated by cars on that freeway. If the penetration rate of radio transmitters is large and known, for instance if the radio system is also used for electronic toll collection, then it can accurately measure the number of cars passing the receiver.
A start has been made in developing algorithms for estimating vehicle density and flow from probe vehicle data using speed/flow diagrams (e.g. Ortúzar and Willumsen, 1990).

For obtaining the parameters of the second category the probe vehicle system concept in general (and the proposed system using ALOHA transmission scheme in particular) is preeminently suitable. Traffic data collected by traditional (fixed) detectors concerns data of a cross-section of a road link only. The second category parameters however regard data of a crosscut of a road link, while deriving crosscut data from cross-section data can introduce significant deviations and is less reliable. The results of this study showed that travel times can reliably be obtained through probe vehicles, even if their penetration rate is low.

The properties of the probe monitoring technique system are thus complementary to those of loop detectors, which are useful for category 1 observations, but less suitable for category 2 data. We surmise that the combination of probe vehicles with loop detector allows the development of novel traffic monitoring techniques that substantially outperform existing methods. However, the fundamental properties and behaviour of these new methods are not yet know and fundamental studies appear necessary. As far as we could observe, even for systems using only probes, thus without loops, it is unknown how many vehicle samples are required to reliably estimate the probability density function of the vehicle travel time (see also Boyce, Hicks and Sen, 1991) or to perform Automatic Incident Detection (AID). For AID simple algorithms exist based on observations of category 1, but not based on category 2 probe measurements. Such algorithms are yet to be researched for probe data or combined data of category 1 and 2.

Control of the transmission interval $T$

The usefulness of a traffic monitoring system using probe vehicles is among others related to the amount of received messages from the road links. This depends on the number of offered, i.e., transmitted massages per unit of time $\rho z / f T$. This is determined by the number of vehicles per road link, the penetration rate of probe vehicles, and their transmission protocol. As the offered traffic increases, the distance between link and base station becomes increasingly important. Mutual interference of messages can be controlled effectively by appropriately adjusting the transmission interval $T$. This study showed that it is likely to be necessary to dynamically adjust $T$, according to the actual radio traffic density, optimizing the throughput. In a more advanced system, one could make the transmission scheme $T$ depending on the actual road traffic conditions, examples are

- more reports from links where an incident is suspected to have happened
- more reports from links where few (probe) vehicles are present
- more reports from secondary roads
- more reports from links far away from any receiving base station
- transmission of reports only if travel time differs from daily average

Penetration rate required for reliable ATMIS

In many expressions, the parameter $\rho z / f T$ appears. This misleadingly suggests that the ATMIS performance only depends on the ratio $\rho / T$. Yet, the quality and the reliability of the information about the actual traffic conditions that can be deduced from the data in the probe vehicle samples is the important criteria. When the penetration rate is very low, this can of course not be compensated by a very short intertransmission interval. From a traffic engineering perspective it is not very useful to have vehicles transmit more than once a minute. In these
cases the TCC may often receive messages from the same probe, which may bias the sample set to the behaviour of one particular driver. More insight is needed into the effect of correlated samples per link, taken from a limited set of probes only.

Analyses (e.g. graphs 34, 35 and 36) indicate that a reasonable value of $\zeta/T$ appears to be in the range of 1 to 4, greatly depending on the transmission scheme, e.g. the values of $f_T = 18$, the throughput requirements for ATMIS services, and (as shown in graphs 37, 38 and 39) on the traffic situation in the vicinity of a base station. Without having specified proper ATMIS requirements it is difficult to make a general judgement about the minimum or optimum probe vehicle penetration rate. Assuming

- our specified default throughput thresholds for AID and (A)TMIS:
  - AID: 3 messages per link per minute,
  - ATMIS: 5 messages per link per 5 minutes and
  - traditional TMIS: 1 message per link per 5 minutes
- an $\zeta/T$ ratio of ($\zeta/T = 4$)
- an intertransmission of at least $T = 1$ minute,

a probe penetration rate of 4% ($\zeta = 4\%$) is required for sufficiently reliably monitoring the freeways by means of the proposed probe vehicle system. Cattling (1994) expects that a penetration rate of dynamic route guidance systems will be 5% by 2000. Until such penetration rate is reached dynamic route guidance may be offered, but AID may not be feasible though probe vehicle monitoring.

We used the above parameters for a tentative design of lay-out of base stations for monitoring the San Francisco Bay Area during peak hours. Figure 44 (simulation) and 45 (analysis) show that 10 base stations offer highly sufficient throughput from all road segments. In this situation high quality AID can be performed on all major road links and reliable ATMIS is possible on the complete San Francisco Bay Area freeway network.
Figure 44 System configuration for monitoring the complete freeway network of the San Francisco Bay Area during peak hours using probe vehicles and random access radio channel (simulation model)
Penetration rate 1.00%, time interval 60 seconds, 18 time slots per second

received messages per link per minute

- 3.0 AID
- 1.0 ATM/IS
- 0.2 TM/IS

Figure 45 System configuration for monitoring the complete freeway network of the San Francisco Bay Area during peak hours by probe vehicles (analytical model)
7 Conclusions

A very suitable means of real-time monitoring the traffic situation on a road network is to collect real-time travel times from roving probe vehicles. We addressed a random-access (ALOHA) transmission scheme and found it a (spectrum) efficient, inexpensive and flexible method for transferring traffic reports from probe vehicles to listening base stations. Such a probe vehicle system can be part of a larger system, such as a dynamic route guidance system. It provides adequate real-time road traffic data, which can be used by traffic management systems for real-time traffic control and by traveler information systems for supplying travel advisories.

To study the performance of the proposed probe vehicle system, we considered a realistic road traffic situation, encountered in the San Francisco Bay Area. A road traffic flow calculation has been performed. This comprehended a simultaneous land-use and transportation model to estimate origin and destination (O-D) data and a traffic assignment model to estimate traffic flows and travel times over road links. We evaluated the relationship between transmitted and received traffic messages generated by probe vehicles participating in the road traffic. In particular, we estimated the number of successfully received messages per unit of time per link, as it depends on system parameters. In particular, the probe vehicle penetration rate, the intertransmission time, the distance to the nearest base station and the number of available time slots per second appeared relevant. Although the system receives disproportionally more messages from certain links than from others, it appeared feasible to deploy a radio system collecting travel times from probe vehicles. Based on a traffic coherence time of 5 minutes, we estimated that at a probe vehicle penetration rate of 4%, an intertransmission interval between transmission of 1 minute, 18 slots per second and about 10 base stations are required to adequately monitor each road link.

The proposed system concept can efficiently gather road link travel times. Road link densities (or flows) may not be accurately obtained from the probe system, but methods are being developed to deduce these data and other techniques exist for directly collecting these data.

With an increasing penetration rate, or at very high road traffic densities, the intertransmission time needs to be modified to avoid excessive transmission conflicts limiting the throughput.
Recommendations for further study

A crucial issue in any probe vehicle system is the required number of probe samples per unit of time, needed to perform various ATMIS services. This requires specification of the service, e.g. AID or ATMIS, and the algorithms and criteria to be used for these services. The number of successfully received messages is an important performance criterion. However, it is not simple to directly relate the performance of an ATMIS service, such as automatic incident detection (AID), to the throughput of probe messages per link. As yet, the processing of collected probe vehicle data into travel advisories and traffic control measures has not been addressed here. However, our continued investigations include methods for combining probe vehicle data and induction loop data.

Moreover, further specification of the message data format is needed. For instance, one needs to specify whether link travel times or speeds are reported, and how these can be efficiently encoded. A related problem is the specification of appropriate transmission intervals, which may depend on the travel experiences by the probes.

The probe vehicle concept appears to be very suitable to collect real-time traffic data from secondary roads and roads in urban or metropolitan areas. The performance of this type of road networks was tentatively addressed in an earlier report by Linnartz and Gamba, but a further study, including a realistic road traffic model, is recommended.

We further recommend a study of the transmission performance of other communication services, such as datacasting or interactive (two-way) communication, using the same model for vehicle densities. A further comparison with polling may be of interest if the performance requirements for the radio system include guarantee reception of messages for from particular vehicles. In particular, this may be relevant if the concept is applied to fleet management.

Methods for investigating other (directional) antenna patterns have been developed theoretically, but these have not yet been implemented.

The resolution of the 'analytical' computer tool may need to be refined. At this moment, it is assumed that propagation conditions and traffic conditions are approximately constant over one entire road link, i.e., we have not split this into separate road segments.
7 References


IVHS (1992), *An overview of the IVHS program through FY 1992*, Washington, USA.


MTC (1993), Metropolitan Transportation Commission, Planning and Modeling Division, Oakland.


Appendices

A Specified Public Transport Network

B Estimated Trip Ends

C Estimated Origin-Destination Matrix

D Specified Road Network and
   Estimated Road Link Flows and Road Link Travel Times

These appendixes contain tables used in our investigations. They are not reproduced in print here, but they are available on floppy disk.