Estimation of Signaling Traffic for Networks with Terminal Mobility

Masters Thesis

by

Martin Ostrowski

Supervisor

Reginald Coutts

Dept of Electrical and Electronics Engineering
University of Adelaide

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

Martin Ostrowski

Estimation of Signaling Traffic for Networks with Terminal Mobility

by

Martin Ostrowski

Abstract

The future Personal Communications Service for voice will be characterised by the use of microcells and the integration of terminal mobility with fixed Intelligent Network (IN) infrastructure. Microcells in urban areas are being more widely deployed to provide adequate capacity for the increasing numbers of pedestrian cell phone users, while the reuse of infrastructure associated with the integration of terminal mobility with the fixed Intelligent Network leads to cost savings and ease of deployment of new services such as Universal Personal Telecommunications (UPT). The integration has important consequences for the required capacity of the signaling infrastructure. Microcells carry traffic with high peakedness and create large amounts of handover traffic; consequently, the supporting fixed signaling infrastructure must be carefully dimensioned if both the terminal mobility and IN service traffic are to be carried with the required quality of service. Accurate dimensioning is dependent on estimation methods that account for all traffic sources in the network; previous estimation methods have been limited to a single source. They have also focused on busy hour planning, which is not always suitable in microcell networks, since busy hours also have a spatial distribution, and have required large amounts of empirical traffic trace data to make them useable, which is not always available. This work presents a new estimation method that addresses these limitations by representing all traffic sources as the output of a single underlying process having readily available empirical parameters. The first part of the thesis presents a terminal mobility model that improves on previous estimation methods by quantifying the relationship between functional space distribution and the resulting traffic flow. The model is then extended to estimate the IN supplementary services traffic for the same set of users, leading to a combined estimation method for both sources of traffic. The second part of the thesis describes a method of using the estimated traffic flow for QoS dimensioning of microcells, using an Erlang loss formula with modified parameters. The modified parameters are used to account for the effect of handovers and improve estimation accuracy, particularly where cell size is small and user density is high. Applications of the work presented here include short and long term cell and location area layout planning, dimensioning of signaling links for a combined terminal mobility and IN service traffic and spectrum allocation for network cells.

Table of Contents

Table of Figures

1. Introduction

1.1 Motivation and Scope of Work

The traffic patterns in the fixed phone network have been extensively studied and are well understood: network dimensioning is on the basis of call blocking probability during the peak average calling period (the busy hour). A fixed network planner is only concerned with this peak calling rate, which is for all the users within the area covered by a switch.

Dimensioning a cellular network is more complicated, since *each cell* of the network must be dimensioned for busy hour blocking probability as if it were a fixed network switch. The possibility of dropping calls as users are handed over to cells with all channels occupied further increases the difficulty of planning a cellular network. A cellular network can have hundreds of cells and each one must be dimensioned separately. To do this correctly, an understanding of the physical traffic patterns in the network is needed.

At the commencement of this work, some research had already been done on the modelling of physical traffic within cells. In particular, the traffic patterns within a highway cell had been studied. However, the tendency of reducing cell size to cater for the increasing user densities, along with a shift of importance from users in vehicles to pedestrian users of mobile phones, has changed the nature of traffic patterns within cells. Many of the results applicable to highway macrocell traffic are not appropriate in the high density pedestrian microcell environment of current and future wireless phone networks. The next generation of cellular telephony is expected to cater for areas of high pedestrian density. Many future mobile handsets will be low power and only suitable for use in microcells. In addition, the cheap, mass produced handsets that are expected to be widely sold under Personal Communications Service are likely to be suitable only for low speed pedestrian mobility. This makes the understanding and estimation of traffic patterns in pedestrian microcells important for dimensioning future networks.

A parallel trend in cellular telephony is the reuse of fixed network infrastructure, especially the fixed signaling network [1]. Although fixed and cellular networks are currently almost completely separate (being connected though a gateway) the reuse of fixed network signaling infrastructure and the sharing of network elements between the fixed and mobile networks is essential to the deployment of next generation of telecommunications services. For example, Intelligent Network services rely on the use of common databases to provide some services; those databases must therefore be shared between fixed and mobile networks.

The implication of signaling network integration for signaling traffic volumes is significant. Signaling traffic in the Signaling System 7 network is generated both through terminal mobility and the invocation of supplementary services. Consequently, in the future, it will be the combined signaling traffic of terminal mobility and IN supplementary service related signaling that will form the basis of dimensioning the signaling infrastructure.

The evolution of wireless communications towards microcells and Intelligent Networks that share signaling infrastructure has brought about a need for estimation methods that address these developments.

The aim of this research is the development of a new estimation method that

- improves the accuracy of physical traffic flow estimation in microcells
- does not depend on previous traffic traces in estimating future flows and
- accounts for both Intelligent Network based supplementary services and terminal mobility signaling traffic in a hybrid network offering both types of service over a single signaling infrastructure.

The primary contribution of the work is the development of a state transition based terminal mobility model with easily obtainable empirical parameters that is also suitable for estimating supplementary service signaling traffic.

1.1.1 Previous Work

The primary area addressed in this thesis where research results have previously been published is terminal mobility modeling. A number of relevant papers exist in this area, presenting both theoretical mobility models and empirical estimation methods. Of the theoretical models, the most widely used is the fluid flow model [2]. This model likens pedestrian movement past the cell boundary to constant rate fluid flow where the flow rate is the product of pedestrian density, pedestrian velocity and cell boundary length. An interesting property of the model is that its results are independent of the shape of the boundary, depending only its length.

The model is based on the following assumptions:

i. user velocity is constant and delta distributed,

ii. flow is evenly distributed across the boundary of the cell and is proportional to the length of the cell boundary, and

iii. flow rate is conserved (a cell is neither a source nor a sink).

These assumptions will be discussed individually.

The first assumption is on the distribution of user velocity, which has generally been assumed to be constant. In reality, the velocity distribution of pedestrians inside a cell is approximately gaussian, with the parameter σ being a function of pedestrian density [18]. However, simulations performed as part of this work (see Section 3) have shown that for small cell sizes, the constant velocity assumption is reasonable, especially at high user densities where the distribution tends to a delta. More fundamentally, the estimation of average flow does not depend on the velocity distribution, but only the average velocity. Consequently, the only circumstances where the assumption may not lead to a good approximation is when estimating flow variation with time for a number of adjacent cells.

The second assumption is only valid for large, regular cells, having many uniformly distributed access points. As cell size decreases, its shape and number of access ports become more significant: a long cell with two access points at each far end will have the same rate of boundary crossings as a short one with two access points in the same relative positions. The assumption is only valid when the number of access ports is proportional to the length of the cell boundary, and the average flow rate through each port is the same. This is more likely to be true in the case of macrocells rather that microcells.

The last assumption is the most restrictive to the model's use in traffic estimation. A study of traffic patterns in an urban space reveals that cells covering space of certain functionality act as sources or sinks of traffic at different times of the day. The most obvious example is transport facilities such as train stations, that in an urban space are a source in the morning peak hours and a sink during evening hours. Such space has a profound influence on the traffic patterns around it, and cannot be represented using the fluid flow model, which assumes that flow inside a simulation space is conserved (in reality, not only is flow not conserved, but neither is the total occupancy). The source/sink problem extends further: even areas that conserve the total occupancy in the simulation space can act as sources or sinks during the day. An example is office space that, while conserving the total occupancy of the simulation space, is a sink during the morning hours, and a source during the evening hours.

The fluid flow model has been successfully used in a number of estimation scenarios and has formed the basis for several important theoretical results. However, being a theoretical model with the coarse assumptions discussed above, it cannot be used to predict traffic patterns at the cell level in a real life situation for an entire network. It must be emphasised that the fluid flow model was never intended to be used for network dimensioning, although it has been used to study the impact of cell sizes on handover signaling volumes with reference to real network scenarios.

Another model, which is more suitable to planning at a higher level of detail is the gravity model due to Rutherford [3], which predicts traffic flow towards popular destinations. Each point in the simulation space is treated either as a source or a sink and the prediction of flow is based on all possible combinations of source/sink pairs for different trip purposes in the estimation space. The name comes from its use of distance as a factor in the flow rate between two points:

$$T_{ijp} = \frac{P_{ip} . A_{jp}}{\sum_{j} A_{jp} . F_{ijp}(d)} \qquad (1)$$

where

T_{ijp} = flow between blocks i and j for trip purpose p

P_{ip} = trips produced at block i for purpose p

A_{jp} = trips produced at block j for purpose p

$F_{ijp}(d)$= friction factor between blocks i and j

The friction factor can be the distance between blocks i and j raised to some fractional power, but Rutherford notes that it has other dependencies, including time. The calibration of the friction factors is the main topic of his paper, but the empirical sources for the trip volumes are not elaborated.

Although the model in the form presented above cannot be used to obtain the time distribution of flow, it could be extended to do so by making each of the variables dependent on time. The model would then need to be calibrated for each time interval of interest.

The drawback of this model is one which is common to almost all previous estimation methods: its strong dependence on previous traffic traces for future estimation. The friction factors in the gravity model must be individually fitted to all points in the estimation space based on previously observed flows. The model is not useable where such data is not already available.

The last model previously proposed is the steady state transition model due to Myskja and Jensen [4]. This model is similar to the fluid flow model in that it uses flow conservation. It is based on the supposition that upon entering a cell, a user will leave through one of the exit ports of that cell with fixed probability, the total probability of exit being equal to 1. The entire simulation space is represented as a non-absorbing Markov chain where each state represents a cell, and edges between states are paths between cells that are assigned a fixed transition probability. The steady state transition rates can then be obtained from the Markov rate matrix for the network. The model is simple to use and is an improvement over the fluid flow model in that it recognises discrete access ports of cells, which is a more realistic representation than a continuous flow boundary, especially for small cells. A small cell would typically have no more than about 10 ports, in which case the fluid flow model boundary representation would not hold.

Although the model is convenient and simple to use when transition probabilities are available, it has a number of limitations:

- the model does not address the source/sink problem discussed previously
- the model requires empirical traffic flow data from which the transition probabilities can be derived, which is rarely available; even when such data is available, it may still not be sufficient for the derivation of transition probabilities (what is needed is the directional distribution of traffic, which is almost never available and must be measured),
- the assignment of a probability of exit from a cell through a particular port can be cumbersome if there are many ports and,
- the transition probabilities are time-dependent, which implies that many transition matrices would be required to adequately describe an extended period of time -- the transition rates for each matrix would need to be derived from observations as functions of time

The overall limitations of the models can be summarised as follows:

1) accuracy:

the estimation methods attempt to extrapolate observed traffic patterns without considering underlying phenomena that generate and perturb traffic. Hence their accuracy is limited in areas where significant traffic flow data is not already available [5][6]. They cannot be used at all in areas where traffic flow data is not available.

2) static prediction of traffic:

none of the models are able to predict the time distribution of traffic flow for the entire day of a point in the simulation space, although some can be extended to some extent by making their parameters time-dependent. Only the peak flow for the chosen busy hour can be predicted. This is a limitation for their use in wireless networks, since the busy hour will occur at different times around places of different functionality (this will be shown later by examining real traffic traces).

3) limited scalability/extensibility:

all models are heavily tied to previously observed traffic traces to make estimates about flow in the simulation space, and do not take advantage of the spatial distribution of buildings. While the existence of a fundamental relationship between space functionality distribution and traffic flow has been noted previously, it was thought to be too complex to form the basis of a prediction model. However, this work will show that such a relationship does indeed exist, and can be expressed mathematically without unreasonable complexity. The implication of the use of this relationship is that predictions can be made about simulation space where traffic traces are not available, and predictions can be made about the resulting effects on traffic flow of changing functional space distribution. This is particularly important in developing cities whose topology is likely to change in the future.

4) inability to predict IN service traffic:

it is expected that in future network the SS7 signaling backbone will be shared by both the wireless network component of the PCS service, and the supplementary services that make use of IN network elements. While the signaling traffic due to the physical mobility of users around the simulation space is proportional to the magnitude of the physical flow across cell boundaries and can hence be predicted for the busy hour by the previous models, the signaling traffic due to the invocation of IN supplementary services is not related to the physical movement of users. While previous models can estimate the traffic generated by a user's physical movements, they cannot estimate the usage patterns of IN supplementary services.

1.1.2 Contribution of This Work

This thesis presents an estimation model that improves on the estimation resolution of previous estimation methods, and develops a method of estimating the combined signaling traffic resulting form both the physical mobility of users and the usage of the IN based supplementary services.

The improved resolution has been achieved by more accurately identifying the fundamental factors that influence the trips taken by users in an urban space and making them the basis of the estimation method. The model identifies quantitative relationships between the distribution of functional space and the resulting traffic patterns around the simulation space, and uses the time-varying relationship between source and destination functional space to map user transitions between different types of functional space onto the physical topology of the simulation space.

1.2 Intelligent Network and Signaling System 7

The Intelligent Network (IN) is a telecommunications network structure concept that was developed primarily to speed up the introduction of supplementary services such as call waiting and call forwarding. However, some IN concepts are already in use in wireless networks and the fixed IN itself is being extended to offer terminal mobility. The IN has two main components: the network elements and their functions in service execution, and a sophisticated signaling system for inter-element communication. This section firstly describes the components and structure of the IN, and then the Signaling System 7 (SS7), which is the protocol used for communications between IN elements.

1.2.1 Intelligent Network Structure

Figure 1.1 shows the elements of the Intelligent Network. Being oriented towards deploying new services, the IN is described through several services-oriented views/planes in the ITU IN recommendations (Service, Global Functional, Distributed Functional, and Physical). Since terminal mobility is not represented in any of these planes, they will not be considered in detail in this work. Only the Distributed Functional plane is relevant to signaling volume estimation, because it specifies the structure and length of messages exchanged between network elements in the execution of services. The general structure of the IN as specified in the Physical plane view is shown in Figure 1.1.

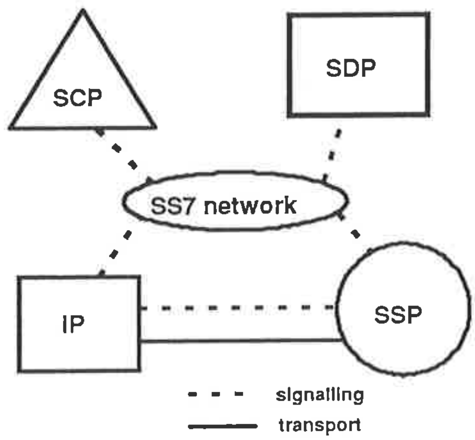


Figure 1.1 - Intelligent Network

Although a full description of the elements of the IN is beyond the scope of this document, the functions of the elements can be summarised as follows:

Service Switching Point (SSP) is responsible for basic call processing and routing of calls to other SSPs. It is the interface of the IN to the local loop and contains the bulk of the functions of a traditional telephone switch. In some IN implementations, it may also play a part in charging (metering) calls. The SSP executes a Basic Call Model (BCM), which is a finite state machine representing the state of the call. When a user makes a call, the BCM moves through states based on the actions of the user, from the "Off-Hook" state, through "Collecting Digits" to the "Null" disconnected state. Each of these states is called a Point in Call (PIC); there are 31 PICs in the IN BCM. Each PIC is potential trigger point for the invocation of a supplementary service that can be set at the beginning of the call. When a PIC armed with a trigger point is reached, the SSP suspends call processing and sends a trigger notification to the SCP, which executes appropriate service logic. When this has been done, the SCP can send a response to the SSP, which will then continue BCM call processing.

Service Control Point (SCP) implements the service logic for IN services and is responsible for the coordination of information flow between different elements of the IN. When it receives a trigger notification, it executes service logic appropriate for the current service (it may need to retrieve data from the SDP in the process) and returns the appropriate instruction to the SSP. In some cases, the SSP may be instructed to connect the user to the Intelligent Peripheral (IP) which can collect dialed digit input from the user (for example, for making a menu selection) and translate that input into an SS7 encapsulated message to the SCP.

Signal Transfer Point (STP) implements routing functions in the Intelligent Network. Messages can be routed through the SS7 network either by performing Global Title Translation (GTT) at the Signaling Connection Control Part (SCCP) level, or by Destination Point Code (DPC) at the Message Transfer Part (MTP) level of SS7 (see Section 1.2.2.1 for details). GTT is generally used for such functions as mapping toll free 800 numbers to real numbers, while DPC routing is generally used in normal transactions between SCP, SDP and SSPs that are aware of each other's network address. GTT routing procedures are not considered in this work, which assumes solely DPC routing.

Service Data Point (SDP) is a heterogeneous database that can hold user locations, service profiles and account information. Its role is in translating labels/numbers to network addresses, maintaining user profiles, including their subscription to different supplementary services, and storing call records.

Intelligent Peripheral (IP) connects to the SSP to play messages to the user and collect dialed responses from the user.

1.2.2 Signaling System 7

IN network entities communicate with each other using the SS7 protocol. The different layers of the protocol are shown in Figure 1.2. The simulation presented in following sections is based on elements of the SS7 specification, especially the Mobile Application Part (MAP). To assist in understanding the simulation, a brief description of each layer is given here. A full description of the SS7 protocol is beyond the scope of this document; for more details see references [7][8][9][10][11] and the ITU Recommendations.

The specific relevance of the different layers of SS7 to this work are as follows:

i. The Message Transfer Part (MTP) and Signaling Connection Control Part (SCCP) framing overhead is considered to obtain packet sizes for terminal mobility and supplementary services messages. These higher level messages are carried between network entities by the MTP layer.

ii. The Mobile Application Part (MAP) defines procedures and message structures for handover and location update related signaling. These procedures are explained in detail in section 1.3.2.

iii. The Intelligent Network Application Part (INAP) will coexist with MAP in the future PCS network where fixed and mobile signaling infrastructure is combined. INAP defines the signaling procedures and message structures for supplementary services. These are discussed in section 1.2.2.5 and Appendix A .

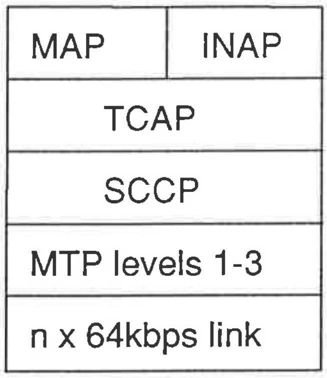


Figure 1.2 - Signaling System No 7

1.2.2.1 Message Transfer Part

MTP has three layers, corresponding roughly to the first three layers in the OSI model. MTP level 1 is responsible for framing and physical level transmission. MTP level 2 handles link level error control and sequencing of packets. MTP level 3 is responsible for the routing of packets between different network entities. For the purposes of this work, it will be assumed that routing between IN elements is transparent and the only significance of MTP is in the framing overhead it generates -- the structure of the MTP message is not important, only its length is used. Retransmission due to errors is also not considered, nor are traffic balancing procedures for distributing message volumes evenly among different routes. Such details are beyond the scope of this work. The MTP packet structure diagram in Figure 1.3 shows the number of overhead framing bits, which is the same for all MTP messages.

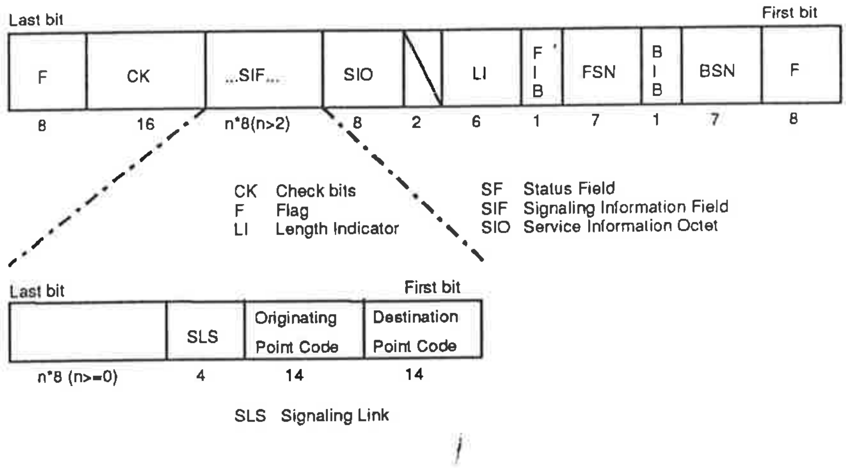


Figure 1.3 - MTP frame format

1.2.2.2 Signaling Connection Control Part

The Signaling Connection Control Part (SCCP) is defined as a layer above MTP3. It is responsible for end-to-end routing, including Global Title Translation functionality, has additional addressing capability, and also optionally provides a connection orientated service for the higher layers of SS7. SCCP defines two service classes, providing connection-oriented or connectionless service at differing degrees of reliability. As with MTP, the specific protocol procedures implemented by this layer are not important to this work -- only the framing overhead of this layer is considered, which varies according to the message contents It is assumed that a transparent message transport is provided to services above this layer.

1.2.2.3 Transaction Capabilities Part

TCAP is a layer above SCCP that is similar to the OSI (Open Systems Interconnection) Session layer. The layers beneath TCAP provide the basic functionality required for transport of queries and updates between network elements. The TCAP provides direct capabilities for full query and update transactions. Typically, a multi-part transaction will commence with a TC Begin message containing the appropriate parameters from a higher protocol (application) layer. The TC End message terminates the transaction and associated SS7 logical connection between the source and destination network entities. The basic functionality of TCAP is simple -- the complexity of this protocol layer lies in the variety of messages that can be carried by the basic TCAP mechanism. As with the lower layers, the detailed protocol procedures of TCAP such as error recovery are not used in this work, although there is some overlap with higher layer specifications, in that transactions such as location updates are represented in the higher layers as TCAP transactions where the parameters and message sequences are described by the higher layers. Like SCCP, the framing overhead of TCAP is variable and must be considered for each message individually.

1.2.2.4 Mobile Application Part

The MAP is an application level protocol that is used in mobile terminal transactions such as location updates and handovers. Channel allocation, signal strength measurement and register updates between a Home Location Register and Visitor Location register are done through the MAP.

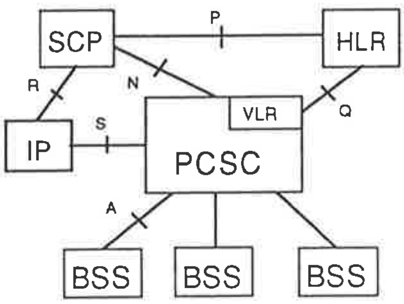
This work uses MAP specification to the extent of estimating the volume of signaling traffic per handover or location update generated in the SS7 backbone, explained in detail in Sections 1.3.2.1 and 1.3.2.2.

1.2.2.5 Intelligent Network Application Part

The INAP is used by different elements of the IN to implement supplementary services. Basic Call Model Detection Points and register queries are INAP protocol entities. The ITU INAP recommendations were used in this work to estimate the number of detection points and queries executed in the course of supplementary service usage and the number of SS7 bytes contained in each such message. For example, to obtain the signaling traffic generated as part of a IN forwarded call, the relevant INAP signaling procedures were obtained from the ITU recommendations. These procedures specify the messages exchanged between the SSP, SCP and SDPs; the length of each message can be derived from the parameters of the message and the framing overhead of the lower layers of SS7.

1.2.3 Interfaces

Figure 1.4 shows the interfaces between network elements in the IN, as defined in this work [12]. Note that the ITU Recommendations have a slightly different definition of these interfaces, as terminal mobility is not supported in the current IN model, but is the subject of standardisation work. The interfaces are implemented through nx64kbps signaling links, often dedicated channels on normal trunk groups.



BSS Base Station System
PCSC PCS Switching Center
SCP Service Control Point
HLR Home Location Register
VLR Visitor Location Register
IP Intelligent Peripheral

Figure 1.4 - Definition of interfaces

The BSS controls handovers between cells and comprises the Basestation Controller (BSC) and Basestation Transceivers (BTS). A BSC controls a single location area (see Section 1.3.2.2) containing multiple cells each with its BTS. When a user moves between cells controlled by the same BSC, no signaling traffic is generated on the A interface -- the handover is done autonomously by the BSC. This means that the location of a user is always known down to the location area level. Within a location area, the BSC locates a user through paging broadcasts.

The PCSC is equivalent to the SSP, but also implements the terminal mobility switching functions of the Mobile Switching Centre (MSC) discussed in the next section. The HLR corresponds to the SDP under IN terminology and contains user profiles and their locations. The VLR is also an SDP, acting as a cache at the PCSC for data from the HLR.

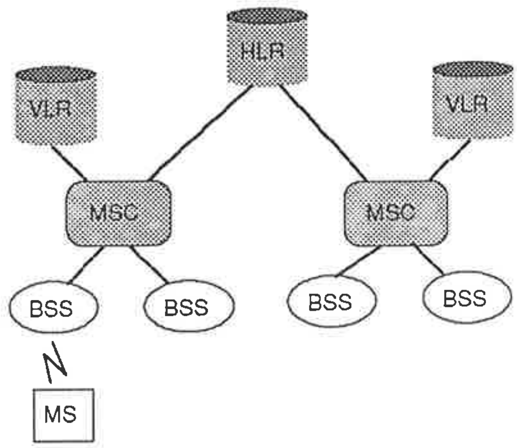
1.3 Terminal Mobility

1.3.1 Description of the Current Generation Cellular Network

The elements of the current generation cellular network are shown in MS Mobile Station

Figure 1.5, and are a subset of the combined Intelligent Network with terminal mobility shown in Figure 1.4. A user changing cells can cause signaling traffic either during the course of a conversation or when the phone is in the standby mode. The former requires a *handover* procedure, which can either be an intra location-area handover that does not require the involvement of any network entities apart form the controlling BSC, or can be inter location area, which requires communication between adjacent BSC over the SS7 network and intervention from the MSC, to update location registers. The latter only requires action when the cell change is across location areas. In this case, the location registers at the MSC must be updated. It is possible also to generate signaling traffic at the SS7 level by turning on a previously powered down handset. Although this action will generate signaling information at the MSC level, it is not analysed in this work.

The Home and Visitor Location Registers (HLR/VLR) contain databases corresponding to the SDP in the IN that hold the location of a user down to the location area level and other information such as account details. The HLR and VLR are updated during a location update or handover between location areas.



MS Mobile Station

Figure 1.5 - Cellular Network Elements

The handover and location update processes are significantly more complex than is described here. Additional complexity arises from channel allocation, and signal strength measurements. Most of these concerns are independent of the SS7 signaling traffic, and do not need to be considered in traffic analysis.

1.3.2 Characterisation of Terminal Mobility

1.3.2.1 Handover at Location Area Level

Figure 1.6 shows the information flow for a handover at the location area level. Handovers within a single location area controlled by a single BSC are executed autonomously and do not generate signaling traffic at the MSC level (the A interface). However, when an active mobile handset leaves a location area, an inter-BSC handover must be performed.

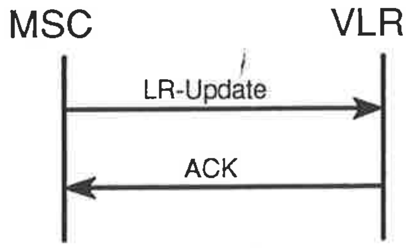


Figure 1.6 - Handover between location areas

1.3.2.2 Location Updates

Location updates are used to minimise the amount of paging required to locate a Mobile Station (MS). The location of a mobile is always known at the location area level. This is achieved through location update messages between the mobile and MSC when the location area boundary is crossed. Figure 1.7 shows the exchange of messages which results from a location area change between MSCs. This particular figure refers specifically to the GSM standard, but the procedure is identical in other systems and generates the same amount of signaling traffic at the MSC level.

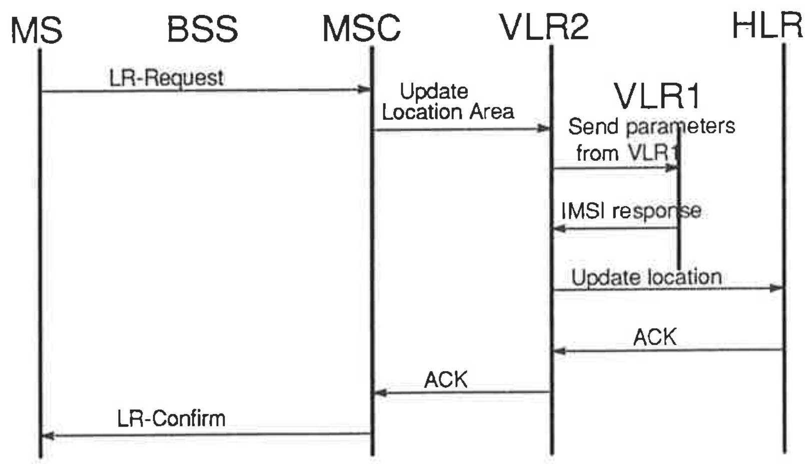


Figure 1.7 - Location Area Update

2. Traffic Modelling

This section develops the estimation method that is the main contribution of this thesis. The work is presented as follows:

- Section 2.1.1 describes how physical flow distribution is derived from functional space distributions,
- Sections 2.1.1.1 and 2.1.1.2 explain the relationship between physical flow and handover messages and offered calling traffic respectively,
- Section 2.1.2 connects the flow of users between functional space types with IN service usage and resulting signaling traffic, and
- Section 2.3 develops a queuing model for the microcell that can be used in estimating blocking probabilities in each cell in the network. The queuing model was developed because the traditional Erlang formula used in estimation blocking probability in fixed network is inaccurate for cellular networks, especially where cell size is small.

To verify the validity of the model, a simulation was done on a space for which flow distributions were available. The results of this simulation are presented in Section 3.

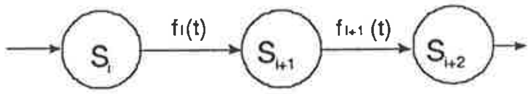
2.1 Markov Model for Intelligent Network with Terminal Mobility

Planning for the busy hour in the traditional sense is not appropriate in cellular systems, since events in the cellular network have a strong spatial as well as temporal locality. The 'busy hour' is at different times for different places. An additional complication inherent to networks where common infrastructure is used to support all forms of mobility is the dependency between the state of the user and their calling behaviour. The total signaling traffic a user generates is the superposition of the handover/location update traffic which results from their physical movements, and the MAP and TCAP SS7 signaling which results from their calling behaviour.

In the course of the day, a user is likely to change state several times, depending on their activities. For a typical city space, many users will undergo similar changes in state, as their activities are similar in nature. It has been found, for instance, that different office buildings in a city exhibit consistent traffic flow distributions over time (signatures) [13]. Such effects are the result of large groups of people engaging in similar activity - in other words, undergoing changes of state at the same time, for the same set of states.

A general representation of any user's activity during the day can be described by a continuous time absorbing Markov chain, where a state in the chain describes the user's current state and the transition probabilities are time distributed. The set of all states is called **S**, while the set of all time distributed transition probability functions is called **F**. The variables s_i and $f_i(t)$ represent elements from **S** and **F** respectively. Any trip can be described by specifying two sequences $\{s_n\}$ and $\{f_n(t)\}$:

$$\begin{aligned} \{s_n\} &= \{s_1, s_2, s_3, \ldots s_N\} \\ \{f_n(t)\} &= \{f_1(t), f_2(t), f_3(t), \ldots f_N(t)\} \\ &\text{where } s_i \in \mathbf{S} \quad f_i(t) \in \mathbf{F} \end{aligned} \qquad (2)$$



The $f_i(t)$ are derived from empirical data. For most simulation space, N is small if we consider an exact division of space into functional groups. The main sources of traffic are from office, transport, shopping, restaurant and recreational space. Each of these types of space exhibits a consistent traffic signature [22].

Particular instances of space exist which are not easily characterised, such as university grounds, or building complexes with mixed functionality. Because the space can contain more than one functionality, users can change state within the space without moving outside the space.

To overcome the problem, the space can be treated as a unique state, with its own flow signature and transition rates. Internal transitions within the space are ignored, and the average transition rate is assigned to the entire space. Such transition rates must be obtained empirically.

The sets **S** and **F** then no longer consist of only states with fundamental functionality such as office space, but also contain compound states and corresponding time distributed transition probabilities.

Transition groups of the form $S_i \rightarrow \{S_j \rightarrow S_{j+1} \rightarrow \ldots S_{j+n}\} \rightarrow S_i$ have a corresponding $f_i(t)$ with symmetrical forward and reverse components and are called *return* trips. All other trips are called *random* and do not follow a predefined path around the simulation space. There is a direct analogy between a user's state and the type of space that they occupy, because the user's trip objective will be the utilisation of a particular type of space. Therefore, the same notation S_i is used to represent both a type of space and a user's state.

2.1.1 Derivation of Spatial User Movement

To derive flow distribution, we divide the simulation space into a matrix of cells **C** of dimensions MxN. The traffic flow through any cell C_i in the matrix will be distributed with time, and will be the superposition of the flow though the cell from every other cell in the simulation space. Consequently, we must calculate the flow emanating from every cell in the matrix, and take the fraction of flow which is in the direction of C_i.

The polar distribution of flow from any cell will be a function of the density distribution of a particular type of destination space around that cell, and the distance of that space from the cell. The volume of traffic will be a function of the rate of transitions towards that type of space at any time.

Assuming that the flow signatures $f_i(t)$ are consistent for different physical instances of S_i, and the number of users in state S_i in a cell is proportional to the density of space S_i in that cell, we can determine the absolute flow rate from any cell by dividing the total number of users in the entire simulation space by the relative density of the particular space type in that cell.

The pedestrian traffic originating from a cell in the simulation space is a product of the flow rate for a space type s at time t, and the relative density of that space type in that cell. Superimposing traffic for each type of space in that cell gives the total traffic from that cell. If the simulation space is represented by a matrix on MxN cells, then, for a cell at *(x,y)*

$$total\_traffic\_generated(x,y,t) = \sum_{s}^{S}\sum_{d}^{D} g(s,d,t).\frac{\rho(s,x,y)}{\sum_{x=0}^{M}\sum_{y=0}^{N}\rho(s,x,y)} \qquad (3)$$

where $\mathbf{S} = \{\text{office, transport, shopping, food, recreation}\}$

$\mathbf{D} = \{\text{forward, return, random}\}$

$g(s,d,t) =$ traffic emanating from space type s in direction d at time t.

$\rho(s,x,y) =$ density of space type s in cell (x,y).

The direction of this traffic will be distributed proportionally to the polar density distribution of destination space, weighted with distance:

$$\delta(s,x,y,\theta) = \sum_{\substack{\forall(i,j)\in C \\ (i,j)\neq(x,y)}}^{C} \rho(s,i,j).exp\left[\ a(|i-x|+|j-y|)\ \right] \qquad (4)$$

$$(i,j)\in C \ \ iff \ \ \arctan\left(\frac{j-y}{i-x}\right) \in [\theta_1,\theta_2]$$

$\rho(s,x,y) =$ *density for space type s*

$\theta 1, \theta 2$ *are angles from* (i,j) *subtending* (x,y)

The effect of this is that, for example, most lunchtime shopping trips from an office will be towards a major shopping centre if one is nearby, but the proportion of this traffic will decrease and shift towards smaller facilities nearby if the distance increases. Similarly, commuters are likely to choose the railway station nearest their office for the forward and return trips (the maximum likely walking distance to the office from the station is called its sphere of influence). While walking distance probability distribution data is available for different types of trips, we use the exponential distribution as a reasonable approximation [14].

What remains is to determine the proportion of total traffic emanating from a cell that is destined towards each type of destination space. For example, the majority of traffic from office space during lunchtime will be distributed between food, shopping and recreational facilities, and only a negligible percentage will be towards transport. This attraction factor is a consequence of movement through state space, and is related to the varying proportions of the state transitions described previously, at different times of the day.

The pedestrian flow through a cell in the simulation space is then a superposition of the flow through that cell from every other cell in the simulation space. Some traffic from other cells will stop at destinations in the direction of the cell before that cell is reached. Therefore, the proportion of traffic passing through that cell is 1 - (ratio of traffic in direction of cell terminating before the cell and the total traffic in that direction), written as p/P_{tot}.

The equation joining all of the concepts discussed above can be written as

$$flow(x,y,t) = \sum_{\substack{(i,j)\in M \\ (i,j)\neq(x,y)}}^{M} \sum_{s_1}^{S}\sum_{s_2}^{S} \frac{p}{P_{tot}} \int_{\theta_2}^{\theta_1} \delta(s_2,x,y,\theta).d\theta.\rho_{s_1}(i,j).a(s_1,s_2).g_{s_i}(t).e_{s_i}(x,y) \qquad (5)$$

$$\begin{aligned} where \quad & \delta(s_2,x,y,\theta) = polar\ density\ of\ space\ s_2 \\ & \qquad at\ (x,y)\ in\ direction\ \theta \\ & \rho_{s_1}(i,j) = relative\ density\ of\ space\ s_1 \\ & a(s_1,s_2) = fraction\ of\ traffic\ from\ s_1\ towards\ s_2 \\ & g_{s_i}(t) = flow\ rate\ from\ space\ s_1\ at\ time\ t \end{aligned}$$

The function $e(x,y)$ requires further explanation.

For any large space such as a railway station, a limited number of access points to the space will exist. If the separation of these access points is significantly larger than the simulation cell size, concentrated flow around the access points will be observed. The function $e(x,y)$ is defined as

$$e(x,y) = \begin{cases} n_c/n_a & (x,y)\ is\ an\ access\ cell \\ 0 & otherwise \end{cases} \qquad (6)$$

$$\begin{aligned} where\ \ & n_c = number\ of\ cells\ in\ group \\ & n_a = number\ of\ access\ cells\ in\ cell\ group \end{aligned}$$

The effect of discrete access points is only observed if the simulation cell size if sufficiently small.

The interpretation of the output from the model is important. Traffic between origin and destination will not be in a straight line, but will flow along the street grid [15]. To consider this effect, we superimpose a matrix of streets and their relative capacities over the simulation space matrix (streets are generally dimensioned for the peak flow rate they carry, so the relative capacity of every street can be said to be the same for calculation purposes).

The flow of traffic between two cells is therefore the fraction of the total traffic in the direction of the destination cell which uses a street passing through the cell whose flow signature is being calculated.

2.1.1.1 Handover Traffic

The signaling traffic resulting from terminal mobility is the product of mobile terminal penetration, the flow rate through the location area and the number of SS7 bytes per handover message.

$$L_A(t) = M.flow(A,t).n_{SS7} \qquad (7)$$

$$where \qquad A \subset \{(x,y) | x \in [0,N_1], y \in [0,N_2]\}$$

A represents the group of simulation cells comprising the location area. The flow rate to and from a location area is largely dependent on the shape and size of the location area, and the distribution of traffic through the day. The number of SS7 bytes generated per handover or location update is given in Appendix A.

2.1.1.2 Offered Call Traffic

The offered call load in any cell is proportional to the number of users occupying that cell at any time. This number consists of users who are stationary inside the cell and those users that are passing through the cell. The first component can be found by dividing the total number of users in a particular state by the relative density of the space corresponding to that state at a
The second component can be found by integrating the total flow
average sojourn time through the cell. For small cell sizes, such as
this approach is sufficiently accurate. For much larger cell sizes, t
must be considered. In that case, the steady-state cell occupancy can b

$$f'(t) = \int_{V_{min}}^{V_{max}} \frac{f(v(s),t).L}{v(s)}.ds$$

where $v(s)$ is the velocity distribution of users through

The equation essentially partitions the traffic into groups that share a consta
V_{min} to V_{max} depends on the velocity distribution being used. Typical veloc
narrow range, with $f(x, t)$~0 for $x<V_{min}$ and $x>V_{max}$.

A difficulty arises in differentiating those users that would use the cellular microcell network to make a call from those that would either use a fixed phone or an indoor wireless network. Therefore, we assume that all stationary users would not use the cellular network. The validity of this assumption depends on the structure of the wireless network(s) in the simulation space, and the functionality of wireless PCS phones. As PCS networks are deployed, this assumption can be refined.

2.1.2 Signaling Traffic Due to IN Services

To evaluate the signaling load from personal mobility and supplementary services, we must first characterise the calling behaviour of a user in a given state. This is done by specifying a set of parameters: the calling rate, and the rate of usage of each of the supplementary services we wish to consider. Each supplementary service has a trigger profile which is loaded into the switch during the call set up phase, when the user service profile is retrieved from the HLR. The number of trigger points in this profile can be used as an estimate of the signaling traffic between the SSP and SCP, over the N interface [16][17]. Where multiple services share a single trigger point, only one query is launched. The total number of trigger points per call can be found by ORing the trigger points for each service in the service set. In addition to the calling behaviour of a user in a given state, invocations of services can be associated with transitions between states. For example, call screening is likely to be invoked when users leave the office to go to lunch, and removed on their return.

The volume of SCP queries is the superposition of those queries which result from calling behaviour of a user in a given state and those which occur due to a transition between states. In addition to signaling directly related to physical movement, represented by handovers and location updates, traffic can be generated by both the existence of users in a certain state, and the transitions of users between states. This is particularly applicable in the IN case, where the invocation of services is related to the state of the user rather that their physical location. Users in an office space use different services and generate different traffic to those who are shopping, or going home. Using the simulation model, it is possible to calculate the total rates of transition between all pairs of states as well as the number of users in a given state at any time.

Estimations can be made about the traffic generated by transitions. For example, call screening is likely to be invoked when users leave the office to go to lunch, and removed on their return. This kind of behaviour has important implications for the signaling load in the network, as will be shown in the next section.

The signaling traffic observed at the SCP/SSP level is a superposition of the traffic generated by users existing in a certain state at time t, and the traffic generated by users moving between states at time t:

$$\begin{aligned} SCP\ queries(t) = & \sum_{r}^{R}\sum_{s}^{S} state\_occupancy(s,t) \,.\, static\_generation\_rate(r,s,t) \\ & \qquad .\, signalling\_traffic\_volume(r) \\ & + \sum_{r}^{R}\sum_{s1}^{S}\sum_{s2}^{S} transition\_rate(s1,s2,t) \,.\, transition\_generation\_rate(r,s1,s2) \,. \\ & \qquad .\, signalling\_traffic\_volume(r) \end{aligned} \qquad (9)$$

where $\quad$ *R is the set of services*

The static and transition generation rate functions are fractions of the total users that use the service r while in state s at time t. The *state_occupancy* function is obtained by integrating the *transition_rate* function, which is itself calculated by summing over the entire simulation space the flow of users between space pairs $s1$ and $s2$. These calculations are done numerically.

What is especially useful is that if a new service can be characterised in terms of usage probability in different types of states and transitions, as is the case for many of the services described in the current IN recommendations, then the impact of the new service on the signaling traffic at the network level can be evaluated, for an arbitrary level of service penetration.

2.2 Short Term Terminal Mobility Phenomena

So far, we have only discussed methods of predicting time averaged traffic flow. However, in any simulation cell, the average flow rate will vary significantly over short periods of time, say one minute. If we observe the flow rate through a cell in this time frame, the difference between minimum and maximum flow can typically be a factor of two (Figure 2.1).

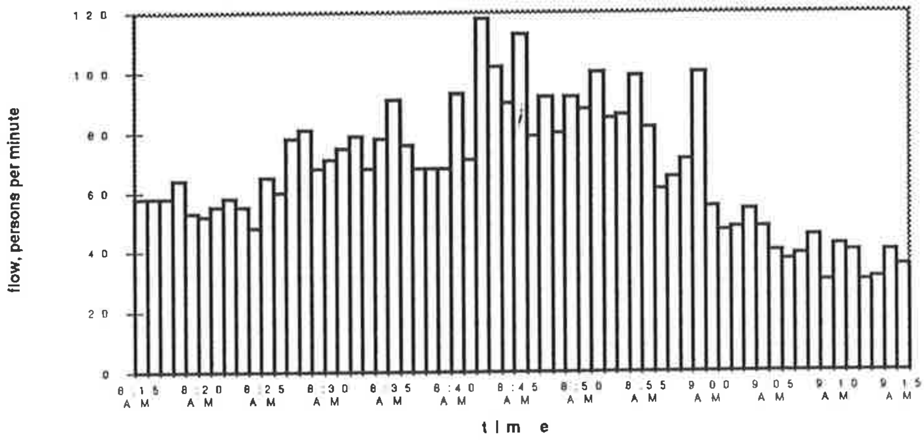


Figure 2.1 - Minute by minute variation in pedestrian flow

Two phenomena dominate pedestrian flow at this level:

a) Platooning

This effect is the result of pedestrians walking in groups and interfering with other pedestrians. A single group will impede pedestrians following it, and a platoon will form.

The severity of platooning is dependent on the density of pedestrian traffic and the walking speed, or more simply the average flow rate (Figure 2.2).

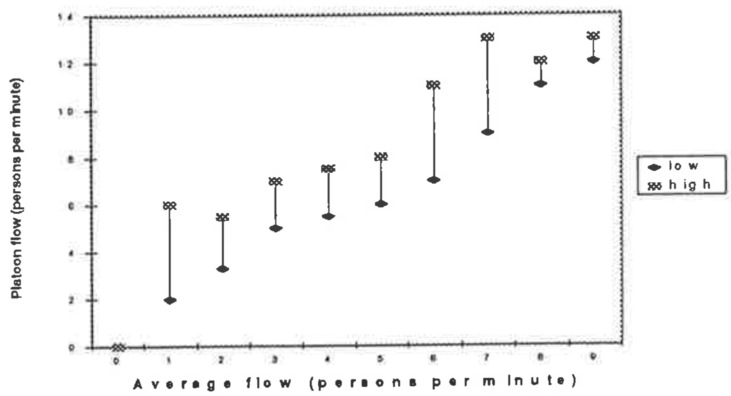


Figure 2.2 - Comparison of average and platoon flow rate

Since a critical parameter for dimensioning a cellular network is the peak flow rate, we represent platooning as the maximum expected flow rate in a short timeframe compared to the average flow rate.

Table 2.1 shows values of peak flow rate compared to average flow rate. The flow is per metre of sidewalk. These factors can be used in estimating peak flow, once the average flow is known.

Table 2.1 - Peak flow factors

Flow rate (persons/min/m)	2	4	6	8	10
Platoon factor	2.00	1.85	1.70	1.55	1.40

The short term effects of platooning do not alter call blocking probability, which should be calculated from the average flow rate. However, call dropping probability is affected by platooning, and the adjusted number should be used in this case. From a user's perspective, call dropping probability is a more critical parameter in perceived service quality than call blocking, because of the inconvenience associated with dropped calls.

Network dimensioning can account for the greater importance of low call dropping probability by assigning a threshold value for the maximum number of occupied channels which is less than the total available number of channels. The unused channels are reserved for calls which are handed over from adjacent cells. The number of reserved channels is selected to give the desired call dropping probability.

b) Traffic Signals

When a red traffic signal prevents movement, pedestrians will accumulate on either side of the traffic signal at a constant rate. The time during which pedestrians accumulate varies from about 20 to 50 seconds. The former value is usual for streets of smaller capacity, while the latter is used for larger avenues.

A simulation cell containing a traffic signal will exhibit a peakedness in the traffic apart from normal platooning effects. Although traffic signals have no effect on the average flow rate across the boundary of the cell, they have a significant impact on the maximum cell occupancy, which in turn affects the peak offered call traffic.

The peak occupancy in a cell with a traffic signal with shortest green time t_g can be expressed as

$$P = \frac{l_{c1}T_s}{2} + \frac{l_{c2}t_g}{2} \qquad (10)$$

where l_{c1} = *flow rate to cell from smaller street*

l_{c2} = *flow rate to cell from larger street*

T_s = *average cell sojourn time*

t_g = *shortest signal green time*

2.2.1.1 Speed / Density Relationship

The density of traffic flow affects not only the average flow velocity, but also the range of flow velocity. Depending on density, the amount of interference between pedestrians varies, and so the flow can vary from free to impeded to congested. The distribution of walking speed versus flow density reflects this as can be seen in Figure 2.3, a real-life measurement.

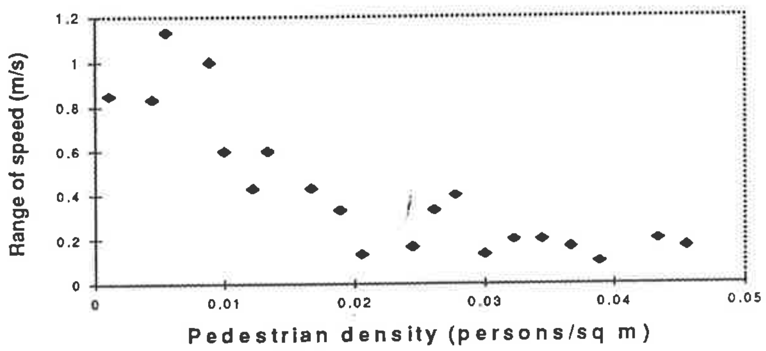


Figure 2.3 - Range of speeds at different densities

For the purposes of simulation, the relationship between speed and density plays an important role in estimating the maximum average flow rate at any time. For a given flow rate along a footpath through a cell, we calculate the resulting density and hence speed. In this way, we can determine a) the amount of platooning in a cell and b) the average cell occupancy.

A number of empirical relationships exist between speed and density. A complete treatment of these can be found in [18].

2.3 Microcell Call Process Characterisation

It has been shown that call arrivals and service time distributions are different for the microcell than for the fixed telephony case. The calling population inside a microcell can change rapidly with time, giving rise to additional call arrivals and departures. These consist of calls that are handed over to the cell while they are in progress, and those calls that originate in the cell and are handed over to another cell while the call is still in progress.

We consider a cell where pedestrians enter one side of the cell, travel through the cell in a fixed amount of time, and leave the other end of the cell. Note that this has the same traffic properties as cells with bidirectional traffic and cells with more than one direction of motion, provided that all paths through the cell have the same sojourn times. The fundamental assumption is that cell sojourn time is constant. This approximation is valid when pedestrian traffic is congested, with significant interactions between pedestrians. This is the case in high density urban areas during periods of high flow rate. In the case where paths through the cell of significantly different length exist, the analysis can be easily extended by partitioning the traffic into separate streams with different cell sojourn times.

The following definitions are used:

t_h = call holding time, exponentially distributed with mean T_h
T_s = cell sojourn time
t_s = remaining cell sojourn time after call origination inside cell
λ_c = call arrival rate for calls originated inside cell
λ_p = arrival rate of active mobile terminals into cell

The offered traffic in the cell comprises:

a) calls that are handed over from other cells

This can be further divided into calls that terminate in the cell and those that continue on to the next cell. The fraction of calls that enter the cell and continue to the next cell is given by the expression

$$P(t_h > T_s) = e^{-T_s/T_h} \qquad (11)$$

It is obvious that where the sojourn time is not much greater than the mean holding time, the fraction of calls which continue through to the next cell is not negligible and must be considered. This is the case in the microcell with radius less than ~400m , and any sized cell where significant paths exist for which the sojourn time is not much greater than the mean call holding time.

b) calls that are initiated in the cell

These calls can also be subdivided into those that complete inside the cell and those that do not. Those calls that complete inside the cell have exponentially distributed holding times. The distribution of call holding times for those calls which continue to the next cell is assumed to be uniform, that is a call is equally likely to start at any point in the cell. In this case, the call holding time is equal to the remaining sojourn time with distribution

$$p_s(t_s)=\begin{cases} 1/T_s & 0<t_s<T_s \\ 0 & otherwise \end{cases} \qquad (12)$$

Hence for a given remaining sojourn time t_s, uniformly distributed, the probability of a handover is

$$P(t_h > t_s) = \int_0^{T_s} \frac{e^{-t/T_h}}{T_s}.dt = \frac{T_h}{T_s}\left(1-e^{-T_s/T_h}\right) \qquad (13)$$

Figure 2.4 shows values for the fraction of calls which are handed over to the next cell, for different values of sojourn time.

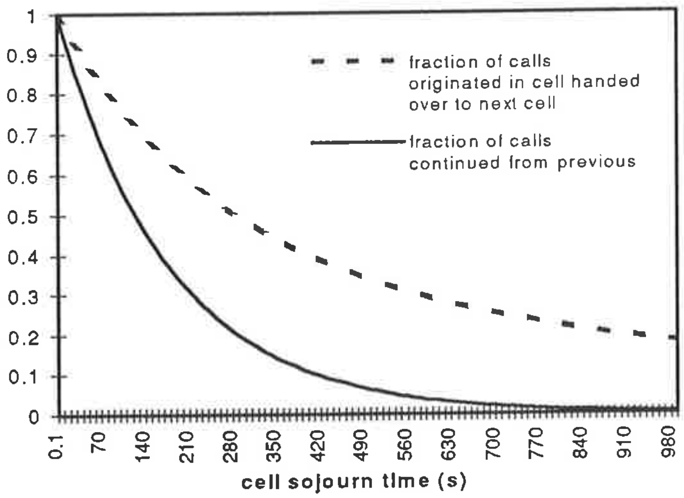


Figure 2.4 - Fraction of calls with handovers

2.3.1 *Queuing Model for Cell*

Using the above results, a queuing model can be developed for the microcell. We take advantage of the assumption that all arrivals in the system are Poisson.

The components of the total call arrival rate into the cell are

λ_{oc} = arrival rate of continuing calls to cell that are subsequently handed over

λ_{nc} = arrival rate of calls originated in cell that are subsequently handed over

λ_{oe} = arrival rate of continuing calls to cell that terminate in cell

λ_{ne} = arrival rate of calls originated in cell that terminate in cell

Using the results from the previous section

$$
\begin{aligned}
\lambda_{oc} &= \lambda_c \lambda_p T_h e^{-T_s/T_h} \\
\lambda_{nc} &= \frac{\lambda_c T_h}{T_s}\left(1 - e^{-T_s/T_h}\right) \\
\lambda_{oe} &= \lambda_c \lambda_p T_h \left(1 - e^{-T_s/T_h}\right) \\
\lambda_{ne} &= \lambda_c \left[1 - \frac{T_h}{T_s}\left(1 - e^{-T_s/T_h}\right)\right]
\end{aligned}
\qquad (14)
$$

These equations are useful in calculating the proportions of handed-over calls compared to calls originated in the cell when provisioning for spare channels to achieve the desired call dropping rate. Calls that are handed over are usually treated with a higher priority than originating calls, because the failure to assign a channel for such a call means that a call in progress is dropped, rather than a new call being blocked.

Since all arrivals are Poisson, we represent the cell as a M|G|c|c queue, combining the handed over arrivals and new calls originated in the cell. The handed over calls have a total arrival rate of $\lambda_1 = \lambda_{oc} + \lambda_{oe}$ and the calls originated in the cell have an arrival rate of $\lambda_2 = \lambda_{nc} + \lambda_{ne} = \lambda_c$.

The expected channel occupancy times are derived using results from [19] and [20]. For a new call originating in the cell, the holding time is

$$s_1 = \min\,(t_h,\, t_s)$$

where t_s has the distribution $p_s(t)$, defined previously.

The expected channel occupancy time of a new call is

$$E[s_1] = T_h - \frac{2T_h^{\ 2}}{T_s}\left[1 - p_s^*(T_h)\right] = T_h - \frac{2T_h^{\ 2}}{T_s}\left[1 - \frac{T_h}{T_s}\left(1 - e^{-T_s/T_h}\right)\right] \qquad (15)$$

where $p_s^*(T_h) = \mathcal{L}\{p_s(T_h)\} = \frac{T_h}{T_s}\left(1 - e^{-T_s/T_h}\right)$

For a call handed over from another cell, the channel occupancy time is given by

$$s_2 = min\,(\,t_h, T_s\,)$$

with expected value

$$E[s_2] = T_h\left(1 - p_h^*(T_h)\right) = T_h\left(1 - e^{-T_s/T_h}\right) \qquad (16)$$

where $p_h^*(T_h) = \mathcal{L}\{p_h(T_h)\} = \mathcal{L}\{\delta(T_h - T_s)\} = e^{-T_s/T_h}$

For the M|G|c|c queue, the probability distribution of occupied channels is the same as for the M|M|c|c queue, and is given by:

$$p_v(n) = \frac{\rho^n / n!}{\sum_{k=0}^{N} \rho^k / k!} \qquad (17)$$

where

$$\rho = \lambda_1 E[s_1] + \lambda_2 E[s_2]$$

Only the quantities λ_c, λ_p and T_h are needed in the above equation; all other quantities are derived from these. λ_c and T_h are independent parameters which are an input to the simulation model presented previously, whereas λp is the estimated flow into the cell as given by equation (4), an output from the simulation model.

3. Results

To verify the analytical model, a 2km by 2km block of downtown Manhattan was used in a simulation. This space was chosen because of the availability of flow data for comparison with simulation outputs and has a regular street grid, which simplifies simulation.

This section describes the simulation performed to verify the model presented in Section 2. Section 3.1 describes:

- the simulation program structure
- simulation outputs
- the functional space distribution in the simulation space
- existing flow patterns for particular space types
- state transitions for trips taken within the simulation space
- mobility parameters for the simulation space, including service penetration levels.

These sections follow quite closely the derivations in Section 2 and describe how empirical data is used in the equations derived in Sections 2.1.1 and 2.1.2.

Section 3.2 presents the flow output of the simulation model, giving both the space and time flow distributions in the space considered, as derived in Section 2.1.1. Section 3.3 gives the signaling traffic levels for the simulation space, as derived in Section 2.1.2. Section 3.4 discusses the results of the model in comparison with previous models.

3.1 Simulation Description

3.1.1 Simulation Program

The simulation program consisted of approximately 5000 lines of C++ code, initially developed under MS Windows and later ported to UNIX.

The program input consisted of ASCII text files containing information on

- the density distributions of space of different functionality, derived from functional maps of the urban space
- traffic flow signatures from spaces of particular functionality
- the street grid and cell layout of the simulation space

An intermediate processing module generated polar density distribution information from the functional density data, as described in Section 2.1.1.

The flow processing component of the program produced the flow versus time vectors for each simulation cell, calculated as the sum of flows from all other cells in the simulation space.

The output formatting component converted the simulation output binary data files into either spatial heat maps or time distributed point flow signatures, suitable for import into the Matlab graphing module for display.

The code was ported to UNIX as it was very memory and CPU intensive. The simulation time was approximately 4 hours on a Sun SPARCstation 10 for the given urban area, and intermediate data files for passing information between program components were typically between 5 and 10 Mb in size.

The model yielded the flow distribution for the space for the time between 7:30am and 9:00pm. The outputs can be summarised as follows:

i. the spatial distribution of flows at different times was obtained by calculating the flows through each cell in the network --this is presented in the form of heat maps which show the number of pedestrians passing through a cell per 15 minutes (this is a typical sampling period for real-life pedestrian flow measurement [22]) at different times of the day,

ii. the simulation output of the time distribution of flow at different points was found and compared with previously measured flow data,

iii. the signaling traffic resulting from terminal mobility (handovers and location updates) together with the traffic generated by supplementary services was calculated.

The other contributions of this work -- the estimation of blocking probability for microcells with modified parameters for the Erlang loss formula, and the use of short term mobility phenomena in the estimation of peak flow were not tested and are presented as a theoretical results.

3. 1.2 Description of Simulation Space

The estimation model requires the density distribution of functional space; these were obtained from functional maps of the city and are shown in Figures 3.1 to 3.5. For clarity, the physical features of the city such as streets are not shown. The labeling on the axes is arbitrary -- the density maps show an aerial view of the section of the city being considered.

Five types of space were identified in the simulation: office, transport, shopping, restaurant and recreational. These were the only types of space with significant density in the downtown area. The distributions were obtained by digitising a functional distribution map of downtown Manhattan.

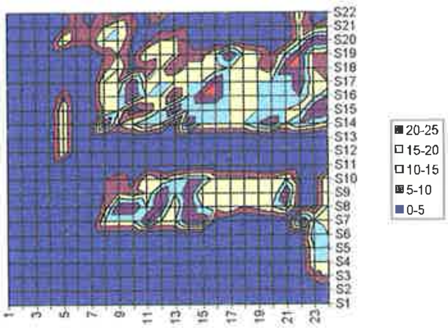


Figure 3.1 - Office space

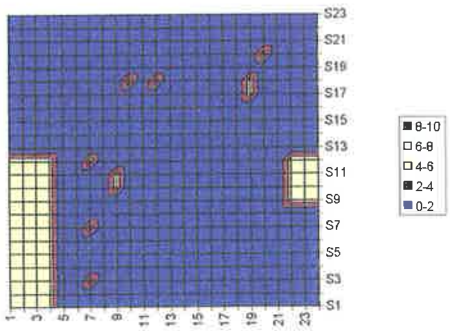


Figure 3.4 - Recreational space

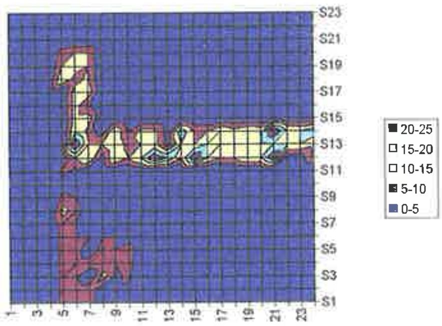


Figure 3.2 - Shopping space

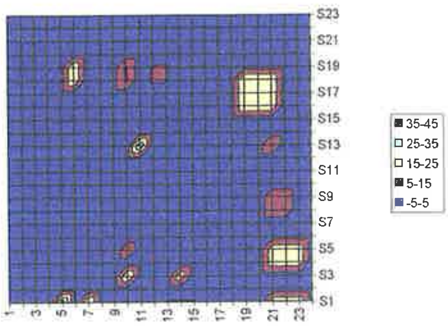


Figure 3.5 - Transport space

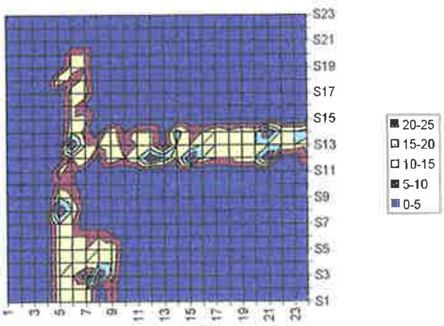


Figure 3.3 - Restaurant space

3.1.3 *Flow signatures for particular space types*

The following recurring flow patterns (called *signatures* in civil engineering) were obtained by sampling and averaging real-life flows around instances of space with a particular functionality in the Manhattan urban space [22]. The forward traffic is that which emanates from the space that will eventually return to the space, the reverse traffic being the corresponding returning flow. The random component emanates from the space without returning. This component of traffic requires that flow be conserved by an equivalent inflow from surrounding cells.

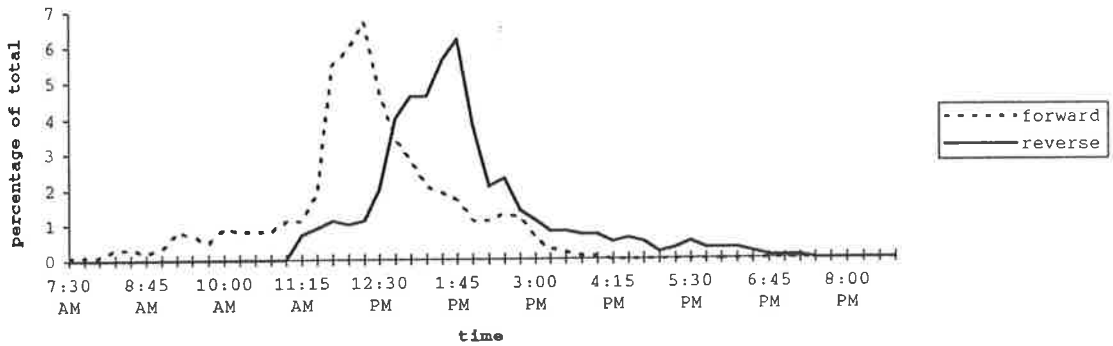


Figure 3.6 - Distribution of flow from office space

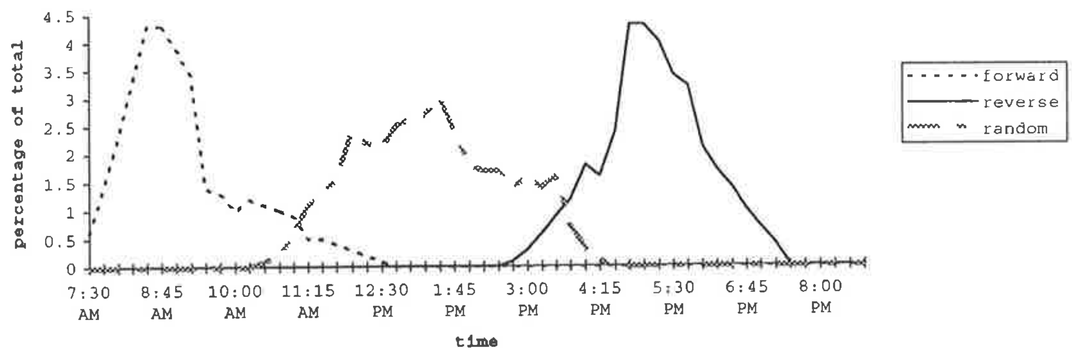


Figure 3.7 - Distribution of flow from transport space

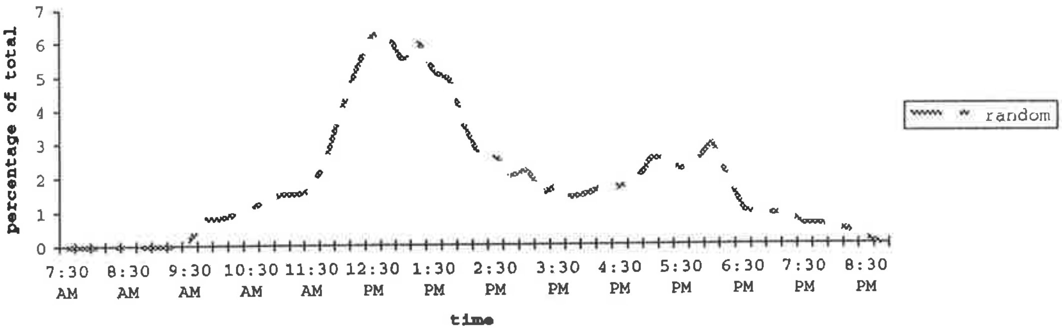


Figure 3.8 - Distribution of flow from shopping space

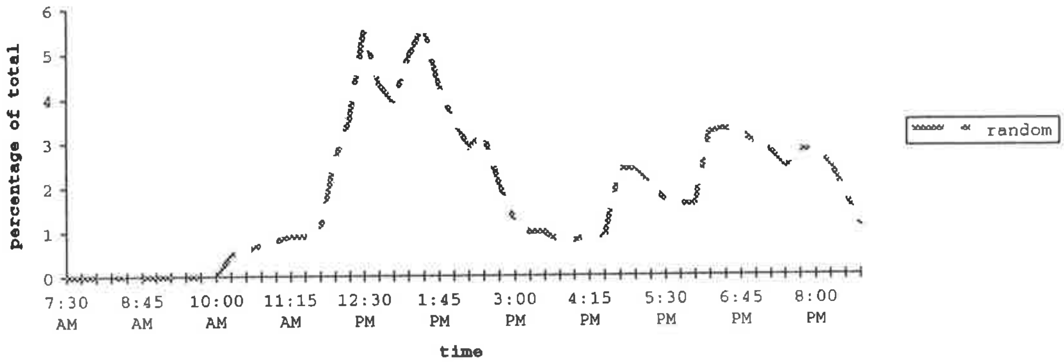


Figure 3.9 - Distribution of flow from restaurant space

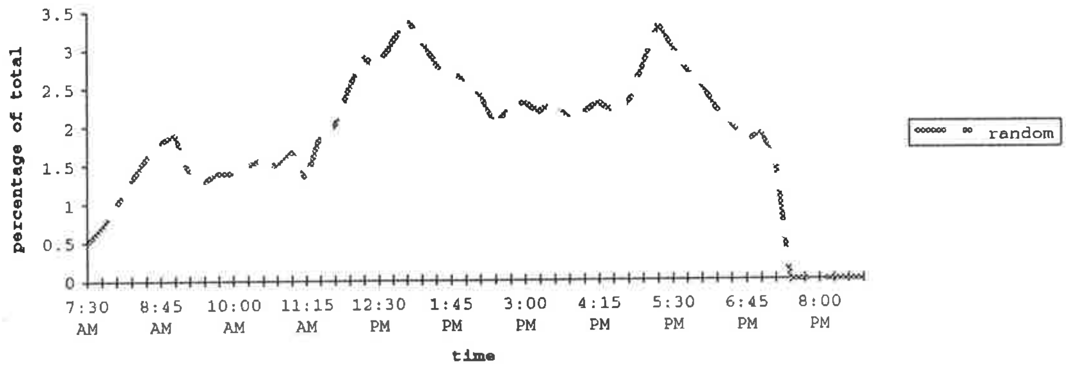


Figure 3.10 - Distribution of flow from recreational space

3.1.4 Types of trips considered

To obtain the resulting flow patterns for the simulation space given the functional space distributions and space flow signatures above, the distribution of transition rates between different space types is required.

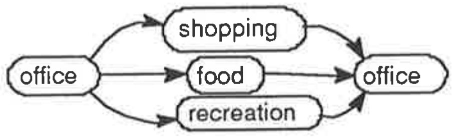
The majority of pedestrian traffic is generated by the following trips through state space:

a) Commuting

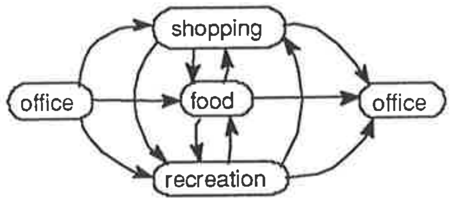
This generates *return* trips between transport facilities (eg. a train station) and office space. The forward trips peak in the morning, with the reverse trip, which occurs along the same route in the afternoon.

b) Lunchbreak

This is subdivided into two components. Firstly, trips to food and recreational facilities (eg. parks) are likely to be *return* trips (ie. returning along the same route). A percentage of trips to shopping space will also be *return* trips.



The other type of trip originating from the office during lunchtime is a *random* trip (eg. walk through park, shopping without specific purpose) for which the forward and return paths are different.

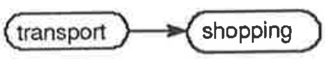


c) Shopping exchange traffic

A large percentage of *random* trips to shops will create new *random* trips towards other shops. *Random* trips will also be generated towards transport/food/recreational facilities.

d) Trips to city for shopping

These are characterised by random traffic from transport facilities to shopping space.



3.1.5 *Simulation Parameters*

We assume the following parameters for the PCS network [23][24][25]:

Table 3.1 - PCS network parameters

	Outgoing Calls		*Incoming Calls*	
	Call Rate	Holding time	Call Rate	Holding time
Office	3	180	*	180
Transport	0.5	80	*	80
Shop	1	120	*	120
Restaurant	0	0	*	80
Recreation	1	1	*	240

* Incoming call rate is dependent on time of day, not location of recipient. Assume average call rate of 2 per hr normally, and 0.5 per hr during lunch and outside business hours. Note that incoming calls which are screened still generate detection point and database query traffic.

Table 2.2 - Penetration rates of services

	Penetration rates
Mobile terminal	50%
UPT	30%
Call Hold with Voicemail	60%
Credit Calls	10% of all outcalls
Call Screening	80%

Notes

1) Mobile users do not subscribe to UPT.

2) Assume 80% of users of call screening will activate it when they leave the office to go to lunch (and deactivate it when they return) or when going home.

3) User interface aspects are not considered, and may generate additional traffic.

3.2 Simulation Output - Call Traffic

3.2.1 Spatial distribution of physical flow

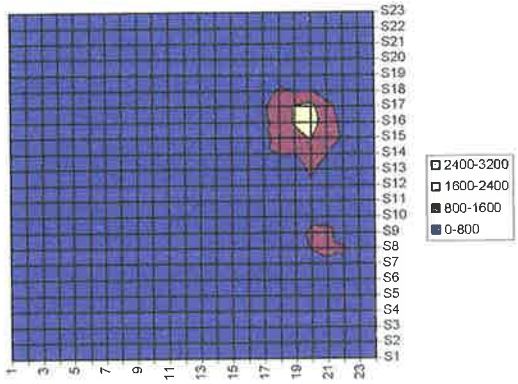


Figure 3.11 - 8:00 am

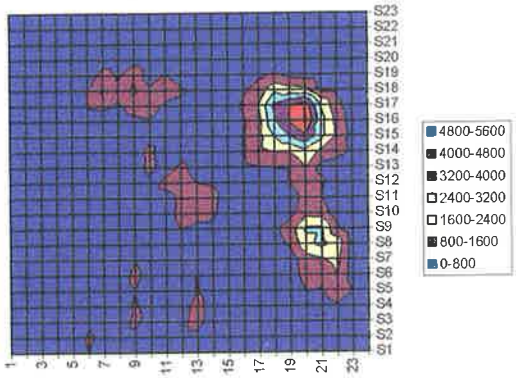


Figure 3.12 - 8:30 am

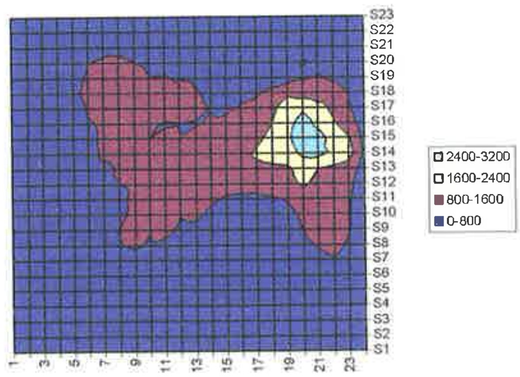


Figure 3.13 - 12:00 pm

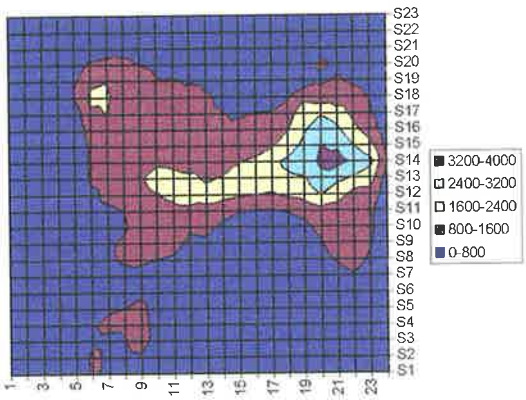


Figure 3.14 - 1:00 pm

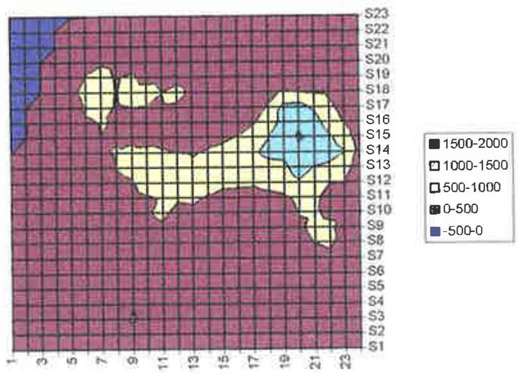


Figure 3.15 - 3:30 pm

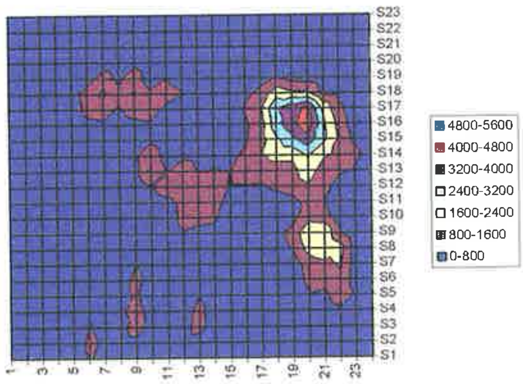


Figure 3.16 - 5:00 pm

Discussion

Figures 3.11 to 3.16 show the spatial distribution of flow at different times in the day, shown as cell boundary (grid square) crossings per 15 min period. Comparing these distributions with the functional distribution of space (Figs 3.1 to 3.5) reveals the influence of space type on the time distribution of flow. Figure 3.12 shows a concentration of flow around major transport facilities, whereas peak flow occurs between office and shop/restaurant areas around noon.

The offered calling traffic distribution is crucial in dimensioning cellular systems. Current planning methods are more concerned with providing adequate coverage than consistency of QoS, but as user densities increase, the expected calling traffic in a cell will become critical to the dimensioning of the network. The time distribution of this traffic will be important when multilayer cells and dynamic topology configuration are implemented. Knowing where more channels are required at different times assists in planning such networks.

3.2.2 Time distribution of physical flow

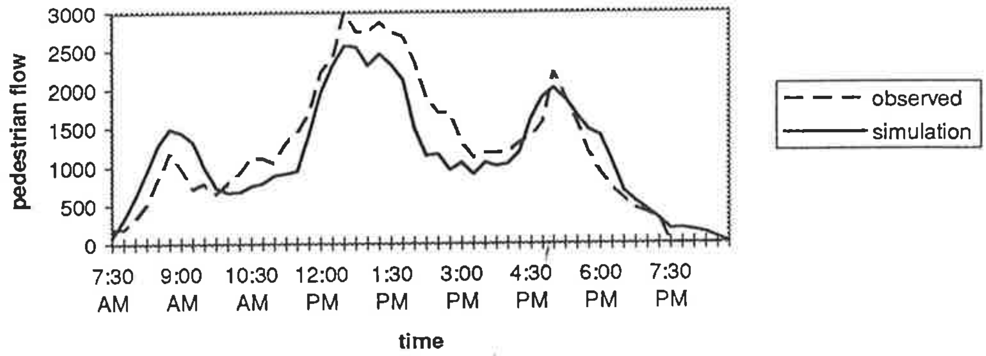


Figure 3.17- Fifth Ave. 44th to 47th sts.

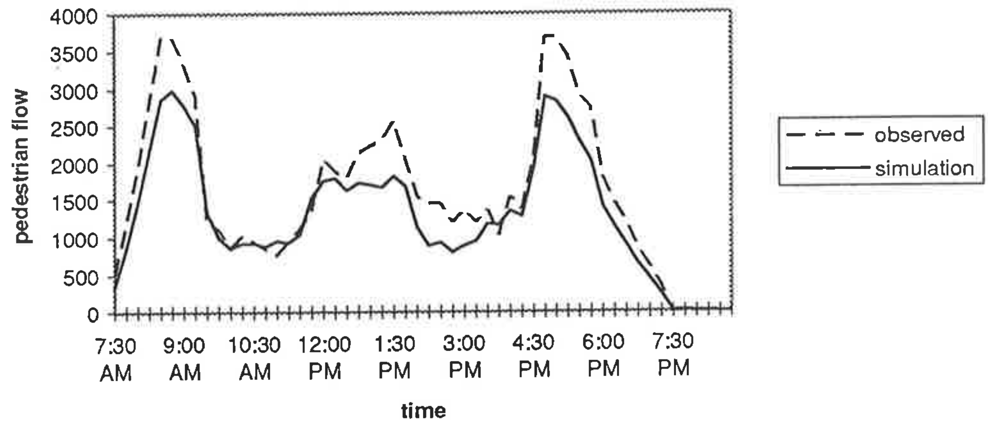


Figure 3.18 - Grand Central Station, Near Escalators

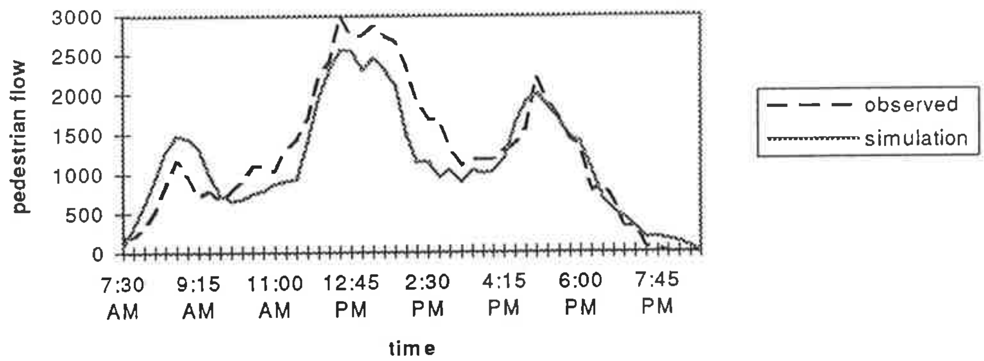


Figure 3.19 - 5th Avenue at 48th St

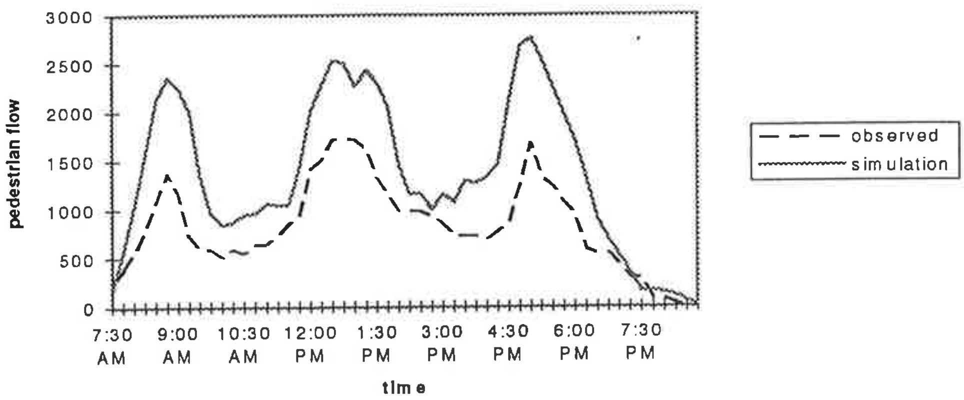


Figure 3.20 - 48th Street Average

Discussion

The flow signatures give insight into the capabilities and limitations of the model. Figures 3.17 to 3.19 represent areas that have well defined functionality and high concentrations of adjacent spaces with different functionalities. As could be expected, the simulation output is accurate in areas of high single-mode flows, as estimations of population density become more accurate, anomalies in movement due to abnormal path topology, and the influence of non-modal flows becomes less significant.

The model is less accurate at predicting exact flows where space functionality is indistinct, or abnormal path topology perturbs or impedes flow, as can be seen from Figure 3.20. Here, the irregularity of the functional distribution (which is inexact at the sampling points) together with the constraints in paths between source and destination space leads to a less accurate estimate.

The accuracy of the model becomes poor when the flow magnitudes are small, and the dominant trip mode is significantly influenced by other non-modal flows.

These observations suggest that a statistical approach, where the instantaneous flow at any point is represented by a random variable, would give more accurate results. However, sufficient data was not available to test the distributions of such random variables. Intuition would suggest that the normal distribution could be used -- the variance of the distribution would be a function of the topology of the sampling point. However, lack of empirical data makes this difficult to verify.

3.3 Simulation Output - Signaling Traffic

3.3.1 Effect of Location Area Strategy

To evaluate the volume of database queries due to terminal mobility, the space was divided into four location areas. Two possible layouts were evaluated. The first divided the space into four adjacent rectangular areas, equal in average volume of users. The second consisted of four areas, one of which covered the shopping district, and the other three around it. Although this might intuitively seem like a good idea, as it minimises the number of location changes for the high volume of traffic between the shops along the main street, it is obvious from the results (Figure 3.21) that the lunchtime and commuting traffic far outweighs the effects of the shopping exchange traffic.

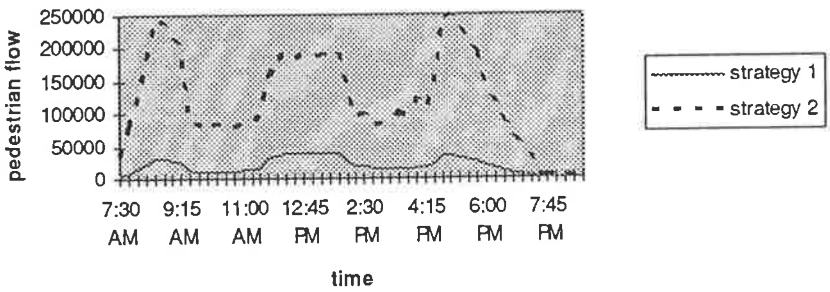


Figure 3.21 - Comparison of Two Location Update Strategies

It is difficult to evaluate such layouts using conventional traffic surveys, even if they are available. Data about the lunchtime busy hour is by itself insufficient to estimate the signalling traffic due to the physical movement of users. A time dependent map of hot spots is needed. Location area planning is one of the unique advantages of using the estimation method presented in this work. For the simulation output of signaling load due to location area changes in this simulation scenario, the first strategy was used.

3.3.2 Signaling Load On Defined Interfaces

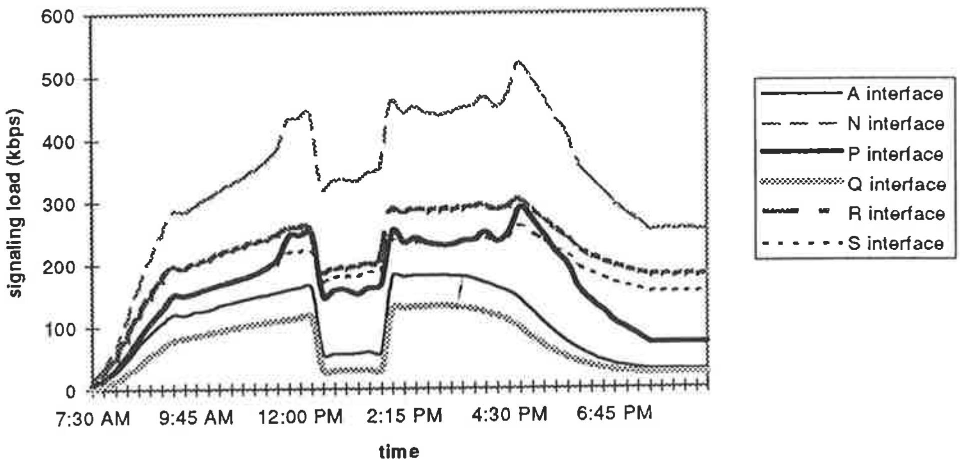


Figure 3.22 - Signaling load

The signaling load shown if figure 3.22 represents the total traffic due to supplementary services. The simulation output is generated as described in Section 2.1.2.

The volume of signaling traffic on the A and Q interfaces (as defined in Figure 1.4) is lower, due to the relatively large location area size, and small signaling unit size for a location update. IN supplementary services entail more complex signaling and hence a higher signaling load. The simulation output agrees with other estimates [26] , but also gives the time distribution of the signaling load.

3.4 Comparison with other models

Sufficient data about the simulation space was not available to make estimations using other models. Also, because of differing output from the different models, it is difficult to make a direct quantitative comparison of the accuracy of the models. Some qualitative comparisons are made below:

i. fluid flow model -- this model cannot be verified in this simulation scenario. Using the model, one might derive the average cell boundary crossing rate for some arbitrary cell layout plan. The model gives no information about the location, time or magnitude of flows, only a single number for average boundary crossing rate.

ii. gravity model -- this model makes predictions about future flows based on available flow data. Since two sets of flow data at different times were not available, this model could not be verified. It should be noted that this model also does not predict the location, time or magnitude of flows, only the change in those flows based on predicted future changes of flow tendencies between different blocks.

iii. the steady state transition model -- this model could not be verified in this simulation scenario, as it requires directional flow data. Some cell topology would already need to be in place, and the transition rates between those cells would need to be available.

It can be concluded that although each of the models above has a place in network planning, none of them are able to make predictions at the level of detail presented in this work. Conversely, the model presented in this thesis can be used to make the predictions of the other models; for example, a simulation space can be partitioned into equisized cells and the average boundary crossing rate calculated. The transition probabilities from one cell to another could also be derived form the relative flow volumes form one cell to another.

4. Conclusions

The accurate estimation of pedestrian traffic in a microcell environment is becoming increasingly important, because of the increasing proportion of pedestrian cellular phone users compared to those using cell phones in vehicles and the reduction of cell sizes in areas of high user density. To date, little work has been done in estimating traffic under these circumstances.

In addition, with the increasing tendency to reuse fixed network infrastructure in the cellular networks, a method for estimating the signaling traffic in a network with combined terminal and personal mobilities and Intelligent Network based supplementary services is needed. This work presented such a method:

1) a relationship between functional space and resulting traffic flow was established, leading to the derivation of an expression mapping user transitions between functional space onto the physical topology of the simulation space,

2) a relationship between physical flow and handover and location update traffic and offered calling load was derived,

3) an Erlang distribution with modified parameters was proposed for increased accuracy in calculating blocking probability in a pedestrian microcell, and

4) a relationship between the user movement between states and resulting IN supplementary service traffic was established.

The model was verified by comparing with the predictions of other models and comparing the model output with empirical data. The estimation method gives a higher level of spatial and time flow distribution detail than previous models and in addition provides additional information on combined signaling rates essential to the dimensioning of wireless networks. Unlike previous methods, the one presented in this work uses widely available data for simulation input. The method has potential value to wireless network planning of use of common signaling infrastructure for wireless networks and fixed Intelligent Networks.

5. Future Work

Two aspects of the model can be extended/improved:

i. a statistical approach could be taken to estimating flows to account for the decreased accuracy of the model in areas of small, mixed-mode flow. The model's output would then be a high/low estimate of the flow based on a desired confidence level. This would require the investigation of the relationship between flow density and the variance of flow magnitude. The flow magnitude would then be a distribution rather than a constant value. The gaussian distribution intuitively seems like a good choice and has the advantage of having only two parameters, mean and variance. The mean would be the average flow already predicted by the model, while the variance would be a function of average flow magnitude and the relative densities of surrounding space. The exact function would need to be calibrated against available flow data

ii. the model assumes that the flow patterns are homogeneous around the simulation space and does not account for perturbations of traffic by physical topology of the thoroughfares. Where simulation space layout is highly irregular and traffic is significantly diverted from the most direct route between source and destination, a more sophisticated technique is required to predict the resulting path distribution of the traffic.

6. Acknowledgments

I would like to thank the following people for their assistance in this research:

Prof Reginald Coutts, CTIN (main supervisor)

Prof Randy Katz, Dept Computer Science, University of California at Berkeley (remote supervisor)

Keith Sklower, Dept Computer Science, University of California at Berkeley

Dr Marek Kwiatkowski, Teletraffic Research Center, University of Adelaide

Dr Edward Chlebus, Teletraffic Research Center, University of Adelaide

Dr Ken Sarkies, Dept Electrical Engineering, University of Adelaide

Dr Sanjiv Nanda, Bellcore

John Leske, CTIN

Tony Smith, CTIN

Ted Buot, CTIN

Fujio Watanabe, CTIN and Communications Research Lab, Japan

7. References

[1] D.J. Clark and S.Cutcliffe, "Fixed Network Infrastructure for Cellular Networks", Mobile and Personal Conference, pp67-72, 1993.

[2] R. Thomas, H.Gilbert and G.Mazziotto, "Influence of the Mobile Station on the Performance of a Radio Cellular Network", Proceedings of the 3rd Nordic Seminar, Paper 9.4, Copenhagen, September, 1988.

[3] G.S. Rutherford, "Use of the Gravity Model For Pedestrian Travel Distribution", Transportation Research Record 728, HRB, National Research Council, Washington, DC, 1979,pp 53-58.

[4] A.Myskja and T.Jensen, "Performance Study for Wireless Access Networks for UPT: Multilayer Cell Structures for Future Mobile Communication", UPT 1992, Genova, pp 159-172.

[5] P.N. Seneviratne and M.Javid, "Estimating and Updating Flows of Pedestrian Facilities in the Central Business District", Transportation Research Record 1281, HRB, National Research Council, Washington, DC, pp 97-101.

[6] "City of Perth: Central Area Survey 1977", Perth City Council, Perth, Australia, 1980.

[7] K.Jabbari, "Common Channel Signaling System No. 7 for ISDN and Intelligent Networks", Proc. of the IEEE, vol. 79, no. 2, Feb 1991.

[8] D.Wilson, "Signaling System No. 7, IS-41 and Cellular Telephony Networking", Proc. of the IEEE, vol. 80, April 1992.

[9] A. Modarressi and R.Skoog, "Signaling System No. 7: A Tutorial", IEEE Communications Magazine, July 1990.

[10] G.Willmann and P.Kuhn, "Performance Modeling of Signaling System No. 7", IEEE Communications Magazine, July 1990.

[11] B.W.Unger et al., "Simulation of SS7 Common Channel Signaling", IEEE Communications Magazine, March 1994.

[12] P.Li, "Commercial Impact of Choosing a Cellular Technology", AT&T Bell Labs, Singapore.

[13] D.C. Kneebone and R.D. Munro, "Work Trips Generated By Large City Office Building", 2nd Australian Road Research Board Conference, Melbourne, 1964, pp 141-160.

[14] T.M. Ciolek, "Pedestrian Movement And Stationary Behaviour in a Public Setting", Department of Anthropology, Australian National University, Canberra, ACT, Australia, 1978.

[15] C.J.Khisty, "Non-Euclidean Metrics in Nonmotorized Transportation", Transportation Research Record 1281, HRB, National Research Council, Washington, DC, pp 102-111.

[16] ITU-T Recommendation Q.1219.

[17] ITU-T Recommendation Q.1218.

[18] S.Older, "Movement of Pedestrians on Footways in Shopping Streets", Traffic Engineering and Control, August 1968, pp 160-164.

[19] Y.Lin, L.Chang and A.Noerpel, "Modeling Hierarchical Microcell/Macrocell PCS Architecture", for submission.

[20] Y.Lin, L.Chang and A.Noerpel, "A Personal Communication Network Handoff Model and Its Applications", for submission.

[21] "Urban Space: A Framework For Analysis", Regional Planning Association.

[22] "Urban Design Manhattan", Regional Planning Association.

[23] S. Wallinder, "Implementation of UPT - Universal Personal Communication", Ericsson Review No.1, pp 40-48, 1994

[24] R.Cameron, "UPT Implemetation in a Capability Set 1 Network Environment", ICUPC '93.

[25] D. Squires, "International Case Study: Providing Effective Intelligent Network Solutions for the Rapid Future Growth of PCS", IN 95, Sydney, Australia.

[26] C.N.Lo, S. Mohan and R.S.Wolff, "An Estimate of Network Transaction Volume to Support Voice and Data PCS", 8th Int'l Teletraffic Seminar Universal Personal Communications, pp293-311, October 1992.

[27] G.Pollini, K.Meier-Hellerstern and D.Goodman, "Signaling Traffic Volume Generated by Mobile and Personal Communications", IEEE Communications, June 1995, pp 60-65.

Appendix A - SS7 Traffic Generated By Mobility and IN Services

Terminal Mobility

This signaling traffic will occur at the A-interface, if we assume that the VLR is integrated with the PCSC. The number of SS7 bytes per change of location area is shown in table A.1, taken from [27].

Table A.1 - SS7 bytes from user activity

Interface	*Orig. Call*	*Term. Call*	*Location Update*
A	120	253	119
Q	-	280	-

Mobile originated and terminated calls also cause signaling traffic. This traffic is characterised in terms of the number of SS7 bytes for an originating and terminating call. The signaling will occur at the A interface according to the values in table A.2.

Supplementary Services

Table A.2 - IN signaling bytes

	Interface			
	S	*N*	*P*	*R*
UPT outcall	1104	928	696	464
UPT registration	504	232	464	232
Call Hold	358	0	252	232
Credit Call	252	590	504	464
Call Screen	938	812	232	348

Appendix B - Papers Covered by This Thesis

"Simulation Model for Intelligent Network Based PCS", with R.Coutts, TRLabs Wireless 95, Calgary, Alberta.

"Estimation of SS7 Traffic in UMTS Network", GLOBECOM 95, Singapore.

Glossary of Terms

Call forwarding	a supplementary service that allows a user to redirect call from one
number to	another
Call holding time	the time interval between the callee going off-hook to answer the call and
the	completion of the call
Call screening	a supplementary service that allows a user to define either an inclusion
list (a	list of the only numbers that will be passed) or an exclusion list (a
list of the	numbers which will specifically be blocked)
Credit call	a call made using a credit card that is charged the that credit card account
Handover	the transfer of a user's call from a radio channel in one cell to a free radio
	channel in another cell without the interruption of the call
In-band signaling	signaling that uses the same physical channel as the associated data flow
Location area	a group of cells under the autonomous control of a BSC that does not
require	messaging to the MSC for handovers between cells within that
location area	
Location update	a message sent to the central location database whenever a user moves
from	one location area to another and the MS is either in an active
(conversation)	or standby (powered on, but not in conversation) mode
Non-modal flow	a physical traffic flow that is not directly related to the presence of
adjacent	functional space
M/G/c/c queue	a queue with Poisson arrivals, general call holding distribution, c
	available servers and calls blocked when all servers are busy
Macrocell	an outdoor cell with radius greater than 500m
Microcell	an outdoor cell radius between 100m and 500m
Mixed-mode flow	a physical traffic flow that results from the adjacency of spaces of
different	functionality

Modal flow — a physical traffic flow that is directly related to the positions and concentrations of surrounding functional space; can be either single mode or mixed mode

Out of band signaling — signaling that occurs over a channel physically separate from the associated data flow

Platooning — the effect of clustering of pedestrians due to interactions and interference between individuals and groups

Single mode flow — a physical traffic flow that is generated entirely by a space of a single functionality

Sojourn time — the time interval between a user entering a cell and subsequently leaving it

Supplementary service — a enhanced calling service offered by an operator to residential and business users in addition to the basic telephone service

Acronyms

BCM	Basic Call Model
BSC	Basestation Controller
BSS	Basestation System
BTS	Basestation Transceiver
HLR	Home Location Register
INAP	Intelligent Network Application Part
IP	Intelligent Peripheral
ITU	International Telecommunications Union
MAP	Mobile Application Part
MS	Mobile Station
MSC	Mobile Switching Centre
MTP	Message Transfer Part
OSI	Open Systems Interconnection
PCS	Personal Communications Service
PCSC	Personal Communications Switching Centre
QoS	Quality of Service
SCCP	Singling Connection Control Part
SCP	Signaling Control Point
SDP	Signaling Data Point
SS7	Signaling System 7
SSP	Signaling Switching Point
STP	Signaling Transfer Point
TCAP	Transaction Capabilities Application Part
UPT	Universal Personal Telecommunications
VLR	Visitor Location Register