Mobile Computing Hagenberg 20140614 1210455006-A Sommersemester 2014

Prof. (FH) DI Dr. Gerald Ostermayer

Road Traffic Modeling Based on Mobile Network Operator Date

Christoph Kieslich

17.03.2014

Contents

**Chapter 1**

# Kurzfassung

An dieser Stelle steht eine Zusammenfassung der Arbeit,

**Chapter 2**

# Abstract

[Sorry. Ignored \begin{english} ... \end{english}]

blah balh

**Chapter 3**

# Introduction

## 1 Problem Description

Traffic analysis is an import part in road network and mobile network simulations. These simulators require moving subscribers in order to investigate interesting effects in their networks. Moving subscribers can be generated from mobility models which try to describe the mobility of individuals.

This work focuses on estimating trajectories for mobility simulations by using mobile subscription data. A trajectory describes the path of a moving object through space as a function of time. Network operators need moving subscribers to evaluate and analyze an existing or a virtual network. Replay scenarios can be created to investigate the current or a new network with a scenario where errors accrued. The trajectory generation process involves several tasks. First useful events have to be filtered in the network. The users path has to be estimated, which involves the start and end position as well as the handover position.

Another part of this thesis is to estimate the coverage area for each cell sites. The coverage area is a crucial aspect in trajectory estimation. Trajectories need the location of a subscriber as a function of time. However GSM doesn’t expose an accurate position of the subscriber. Therefore a good representation of the coverage area allows to narrow the users position in the network. The problem is to find an approach which creates a good estimation of the coverage area for each cell site.

## 2 Motivation

In the last years mobile network simulations have to adopt to the mobility of the subscribers. Subscribers are not stationary and therefore behavior models have to be defined. These models shall represent the entire subscriber database.

Since the beginning of mobile network simulations random walk and manhattan grid approaches were used to enable mobility. The problem is that random walk and manhattan grid rely on statistical data. The statistical input used is derived from household surveys. This surveys however only depict a special moment in time. To evaluate or gather more moments numerous surveys have to be done.

Our approach is different, instead of surveys we are using mobile subscriber data. More precise call data records. The motivation behind using mobile subscriber data is that this data represents all users of the mobile network operator. We want to use this data to generate anonymous driving trajectories. The use of call data records allows the generation of trajectories for each day and time of the year. The generated trajectories adopt to daily as well seasonal changes.

## 3 Challenges

The first challenge we have to face is that we don’t have a model of the coverage area for all cell sites. Therefor we need a good representation to estimate a coarse position for each subscriber. In order to generate a trajectory the users start and end position are crucial. At first we only have the serving cell and a representation of its serving area. Therefore the position of the user within the cell has to be estimated. For this purpose we are using a combination of population and land use information data.

## 4 Goals

The main goal is to generate trajectories for each mobile subscriber. However due to the design of GSM we can only generate trajectories for mobile users which are in an active call. Additionally we are investigating approaches to annotate the driving trajectories with timing information. In the end we are also examining methods to improve the representation of the serving area with public available data (digital elevation model, land use clutter information,...).

## 5 Structure

At first we are discussing state of the art approaches to estimate driving trajectories and mobility in mobile network simulations. Followed by an overview of fundamental concepts and techniques used in this thesis. The third part explains our concept and approach which will be implemented in the succeeding chapter. At last we will present and discuss our results and give a summary of the work.

**Chapter 4**

# Related Work

The following chapter presents state-of-the-art approaches and projects related to user mobility and traffic analyis. These projects are used to derive the mobility of mobile subscribers by investigating events in mobile operator networks. They differ from classical behavior analysis in terms of penetration and accuracy. To derive the behavior of an entire population surveys and origin-destination matrices have been used in traffic modeling since the early 1960s.

Starting in the late 1990s floating phone data (FPD) has gained interested in traffic estimation and congestion detection. FPD is used due to its high penetration rate, in Austria the market penetration was $159\%$[[1]](#footnote-1) in 2013. There is a large volume of published studies describing the role of FDP for traffic analysis [Yim2001, Qiu2007, Caceres2008].

The approaches which will be presented here have a higher penetration rate due to the fact that they investigate a whole network whereas surveys can only cover parts of the entire network. Another advantage of these approaches is the adaption to seasonal changes. Surveys only depict the behavior at a certain point in time. However both road and mobile network traffic modeling depend on changes in time. By exploring events in mobile operator networks the behavior can be analyzed for every point in time.

## 1 Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from Israel

To measure the traffic speed and travel time in Israel Bar-Gera [Bar2007] used a proprietary system by Estimotion Ltd. The system works with handover updates to derive the traveled route and traffic speeds. A sequence of locations derived from the handover footprint is matched to a road segments which appears to be the most likely on the road network. An example of a handover sequence and its footprints on the road network is shown in Fig.  2.1. Despite the fact that this is a proprietary system it gives a good overview of the capabilities of floating phone data. The work by Bar-Gera shows that travel time estimation with phone data can be a replacement for loop magnetic detectors. He investigated a road segment from January to March 2005 and found that the average absolute relative difference between the two systems is $10.7\%$. However the whole system has its limitations due to the noise generated by floating phone data. They could reduce the noise by combining travel speeds from subscribers which traveled on the same road segment.

Figure 1: Handover sequences and footprints generated by a moving vehicle [Bar2007]

## 2 Generating Trajectories from Mobile Phone Data

To derive trajectories form mobile subscribers Schlaich et al. [Schlaich2010a] used location area sequences. Location area update events are issued whenever a mobile subscriber enters a new location area. These events are issued in both connected and idle states therefore trajectories can be generate even when no call is ongoing. Due to the small number of location areas in the research area, each location area was represented by a unique character. This allowed storing a location area sequence as a string. The next step was to generate routes in the research area and store the location area characters for each area they traversed. To estimate a trajectory for a mobile subscriber the location area sequence for the subscriber was compared to route sequences which have been generated before. Fig. 2.2 depicts the comparison of a mobile subscriber sequence and route sequences with high similarity.

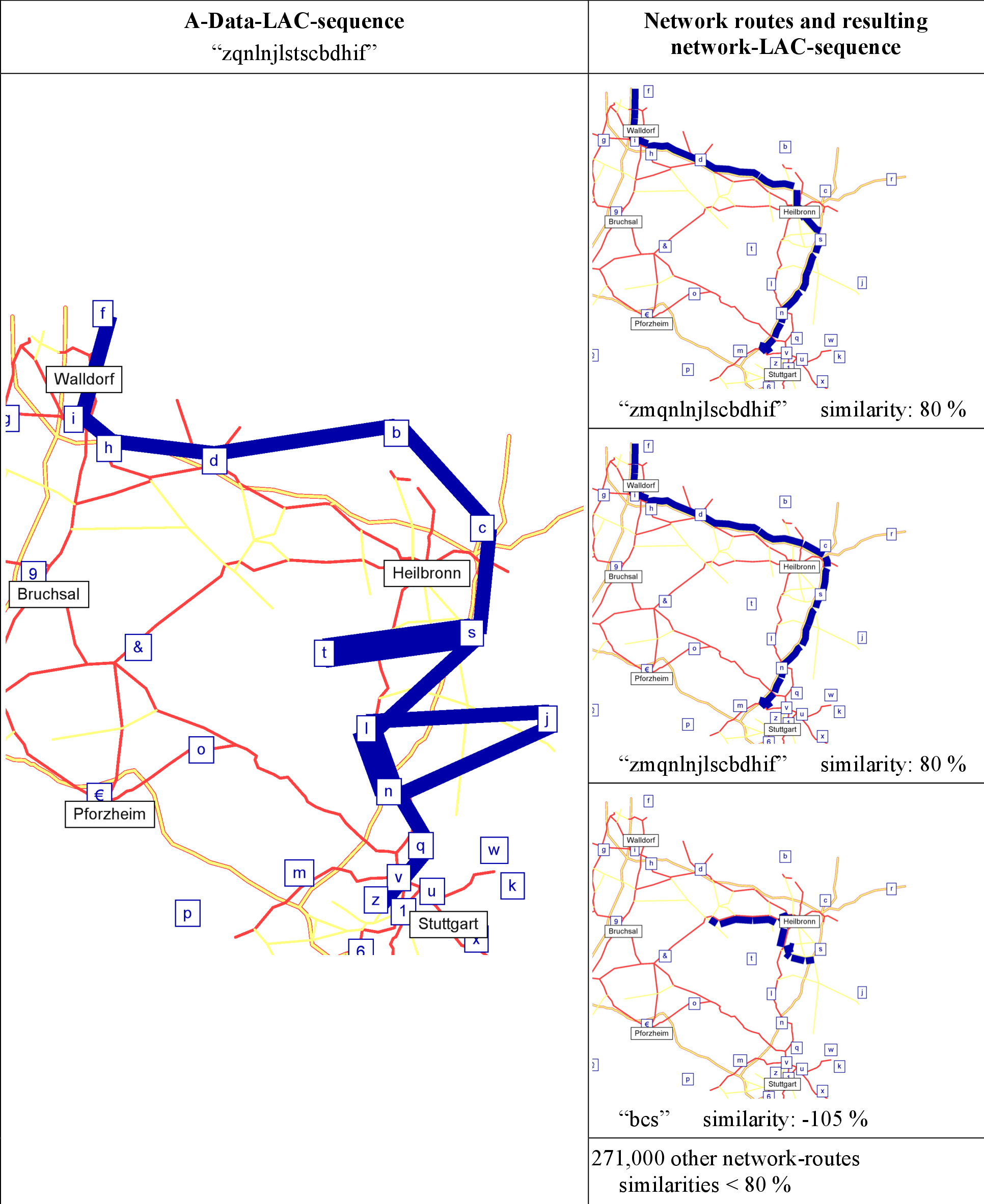


Figure 2: Results of comparison between sequences with high similarity by Schlaich et al. [Schlaich2010a]

The results showed that this technique works well for longer trips around 20 kilometers due to the fact that trips can only be estimated if the number of different location areas is greater than 3. Schlaich et al. also stated that the presented approach only generates trajectories for sim cards and not for vehicles. A vehicle can have none, one or more sim cards therefore a vehicle can generate $0..n$ trajectories where $n$ is the number of sim cards within the vehicle.

Schlaich [Schlaich2010] later used these technique to observe the route choice behavior in the Stuttgart, Germany area. It has shown that drivers react to traffic news broadcasts and variable message signs (VMS). However the acceptance of a route displayed on a VMS was only between $3$ and $17\%$.

## 3 Route Choice Estimation Based on Cellular Signaling Data

To overcome the use of origin-destination matrices Tettamanti et al. [Tettamanti2012] used a simulation framework and a route generator to generate trajectories for mobile subscribers. Instead of using location area updates as Schlaich et al. did they used handover updates. Handover updates allow to generate trajectories not only for the higher road network but also for the minor one. Their approach is based on cell area estimation as each handover update reveals the current cell area. Whenever a mobile phone reaches the boundary of the currently connected cell or when another cell has a higher reception then a handover is made to a cell with a better reception. Tettamanti et al. used Voronoi partitioning to calculate the coverage area for each cell site to estimate a coarse user position.

The main limitation when using handover updates is that handover updates are only propagated when the mobile phone is in connected mode (during a call).

The start and end of the trajectory was set by the centroid of the cell where the call originated and terminated. To derive routes a traffic modeling simulation framework *VISSIM* was used. VISSIM was used to generate routes between the start and the end of the call. For each of the generated routes the euclidean distance between each of the cell sites and the generated route was calculated. To compare different routes equation 2.1 was used as a metric. For each route $j$ the squared sum of all minimum distances $d\_{i,j}$ between the route and the cell site was calculated. The route with the minimum sum was used as a trajectory for the subscriber. In Fig. 2.3 an example for four different routes which were generated by VISSIM can be seen. Tab. 2.1 depicts the results of the above stated equations (see Equation. 2.1). In this example route 4 (see Fig. 2.3(d)) had the smallest squared sum of all four routes.

\[

\label{eq:sumsquare}

D\_j=\sum\_{i=1}^{m} d\_{i,j}^{2}

\]

(1)

\[

\label{eq:minsum}

min(D\_j), j = 1,2,\ldots,n

\]

(2)

Figure 3: Comparison of four different generated routes by Tettamanti et al. [Tettamanti2012]

Table 1: Squared deviations for each route as observed by Tettamanti et al. [Tettamanti2012]

The research by Tettamanti et al. showed that trajectories can be generated for mobile subscribers by analyzing their handover events. This can be done without the need to use special equipment and therefore be used to investigate the travel behavior for many subscribers.

## 4 RoadCell – Road Traffic Estimation from Cellular Network Signaling

By using cellular network monitoring and therefore mobile subscription data Valerio et al. [Valerio2009, Valerio20092] at the FTW Forschungszentrum Telekommunikation Wien GmbH are investigating road networks in Austria. *RoadCell*[[2]](#footnote-2) is using the same mobile subscription data source as our system [RoadCell2009]. The aim of *RoadCell* is to recognize situations in the traffic flow by analyzing events in the core network of the mobile network operator. Driven by the fact that each road user is also a subscriber of a cellular network, network operator can be seen as another source to gather traffic information. Their motivation is to provide road operators with an inexpensive toolkit to observe the traffic flow without the installment of costly sensors. Thereby it is possible to not only investigate the traffic flow on the higher road network but also on minor roads where the installment of dedicated sensors cost-effective.

The idea is to observe changes in network signaling events and extract road conditions for example: drop in the handover rate;abrupt change in the LR update; (c) increase in the number of calls/SMS; change in the number of road users. Figure 2.4 shows how an accident effects the amount of Routing Area Update (RAU) and Location Area Update (LAU). An accident is indicated by sharp decrease in RAU and LAU events followed by a sharp increase in RAU an LAU events.

Figure 4: Effect of the accident on the number of combined RAU and LAU; tic=300s [Valerio20092]

Figure 2.5 depicts the system overview of the *RoadCell* project. It can be seen that despite cellular network data there are additionally data sources used. In order to map a particular event to a road segment they are using coverage predictions from the network operator. To support data processing they are using third-party sources like floating car data from taxis and public transports which are more accurate but have a lower penetration rate.

Figure 5: RoadCell System Overview [Valerio2009]

During their research found out that a combination of passive and active tracking is best to observe road conditions. Passive tracking is done without the involvement of the mobile station. This approach only utilizes the events captured in the operators core network. While passive tracking offers a high penetration and no increase in network load it is less accurate than active tracking. Active tracking requires the mobile station and core network to exchange special events which allows the network operator to get an accurate position of the mobile station. This technique is known as Time Difference of Arrival (TDOA). Valerios proposal is to use a combination of both, passive tracking is always used and once the system detects a possible accident active tracking is used for the area in which the accident occurred. This approach improves the accuracy while trying to minimize the additional network load introduced by active tracking.

**Chapter** **5**

# Fundamentals

This chapter covers fundamentals which are necessary for a better understanding of the concepts presented in this work. It gives a short introduction into mobile communications systems which our system uses. Mobile communications systems are used in our system to estimate the users coarse location. Furthermore state of the art analytical mobility models for mobile network simulations are presented. Another section covers the data (provided by A1) our system is using to generate driving trajectories. At last the process to estimate the coverage area for cell sites with Voronoi tessellation is shown.

## 1 Mobile Communication Networks - GSM

Mobile communication networks are the basis of our modern communication and interaction. However the beginning of wireless communication began in the late 19 and was shaped by the work of Hughes, Maxwell, Hertz, Tesla and Faraday.

In 1982 a new mobile communication network was developed by the Group SpÃ©cial Mobile of the CEPT (ConfÃ©rence EuropÃ©enne des Administrations des Postes et des TÃ©lÃ©communications). The main goal of GSM was to develop a Europe-wide standard for digital mobile communication. Over the past two centuries there has been a rapid increase in the use of GSM all over the world. The present section is based on the books *GSM: Switching, Services and Protocols* by JÃ¶rg EberspÃ¤cher et al. [Eberspaecher2001] and *Mobilfunknetze und ihre Protokolle 1* by Bernhard Walke [Walke2001].

Besides GSM there are other mobile communication currently in use. This systems are known as 3G (UMTS, HSPA, WCDMA, etc. ) and 4G (LET, LTE Advanced, etc. ) networks whereas GSM is a network of the second generation (2G).

### 1.1 Components - System architecture

GSM consists of a large number of components which are necessary for its operation. However we will only focus on components which are necessary to understand the functionality of our system. The overview of a GSM network with GPRS capability is visible in Fig. 3.1. The various standardized interfaces which are responsible for communication between the components can be seen as well as the different components of the *Radio Access Network* RAN and the *Core Network* CN. The RAN consists of a BTS and BSC in GSM and a Node B and a RNC in GPRS.

Figure 1: System overview of the a GPRS Network

#### 1.1.1 Mobile Station - MS

Mobile stations are the equipment used by mobile subscribers to access services provided by the network operator. The mobile station consist of two components, first the *Mobile Equipment* and second the *Subscriber Identity Module* (SIM). The SIM grants a mobile subscriber access to the network and allows to initiate or receive calls. For management purposes GSM assigns the following numbers and identities to a MS:

• **IMSI** International Mobile Station Identity

• **TMSI** Temporary Mobile Station Identity

• **MSISDN** Mobile Station International ISDN Number

• **MSRN** Mobile Station Roaming Number

#### 1.1.2 Base Transceiver Station - BTS

The BTS provides the radio channels for signaling and user traffic data with in a cell. The mobile station establishes a connection with the BTS over the air interface. The BTS consists only of a few parts, the high-frequency equipment (transceiver and receiver) and the signaling and protocol processing unit. The control remains in the BSC.

#### 1.1.3 Base Station Controller - BSC

A BSC is used to manage one or more BTS. It is directly connected to a MSC. Together with the BTS, the BSC forms the *Base Station Subsystem* (BSS). For example, the BSC executes the handover protocol and switches a MS to a new BTS.

#### 1.1.4 Mobile Switching Center - MSC

The MSC together with the databases (HLR, VLR) forms the *Mobile Switching System* (MSS). The MSC is responsible for the switching in the network. For example, the MSC performs signal routing, routing path search and service processing. Additionally the MSC has to pay attention to the allocation and administration of radio channel and the mobility of subscribers. A *Public Land Mobile Network* (PLMN) can consists of several MSCs where each is responsible for a dedicated area. Furthermore the MSC is connected to a *Gateway Mobile Switching Center* (GMSC) which forwards voice traffic between fixed and mobile networks. The GMSC enables the subscriber to call subscribers in different MSCs or PLMNs.

#### 1.1.5 Visited Location Register - VLR

The VLR stores information of all MS which are located in the serving area of the associated MSC. A VLR can be assigned to one or more MSCs. A MS can be registered in a VLR of its home PLMN or a foreign one when a roaming agreement exists.

#### 1.1.6 Home Location Register - HLR

Typically, there is one HLR per PLMN and one VLR for each MSC. The HLR stores all permanent and temporary subscriber information for all registered subscribers. Moreover it stores a coarse current location of each MS. The HLR operates as a central register needed by the MSC for routing of subscribers.

### 1.2 Radio Resource Management - RR

The GSM standard [Etsi1994] distinguishes between two modes of operation first idle and second connected mode. The radio resource management is necessary to manage the physical and logical channels.

#### 1.2.1 Idle Mode

In idle mode there is no dedicated channel assigned to a MS, however the MS listens on signaling channels (BCCH and CCCH). The higher layers are only informed when a MS reaches the boundaries of a location area.

#### 1.2.2 Connected Mode

In this mode there are two deticated channels assigned to the MS, the first one is SACCH and the second one is either FACCH or SDCCH. The radio resource management provides the following service to a connected MS (a full list can be found in ETSI GSM 04.08 [Etsi1994]):

• transfer of messages on any data link layer

• establishment/release of multiframe mode on SDCCH, FACCH or SACCH.

• automatic cell reselection and handover to maintain the RR-connection

### 1.3 Handover

A handover is the transfer of an established connection to a new BTS. A handover is necessary for various reasons. First of all, a handover decision is made by the GSM network and not by the MS. As mentioned before the handover protocol is implemented in the BSS. The BSS decides to initiate a handover based on BSS criteria (channel quality, received signal level, distance between MS and BTS) and network criteria (e.g. traffic load of the network).

Handovers are only performed when the MS is in connected mode. If the MS is in idle mode and reaches the boundaries of the cell a location area update is performed if the new cell is in a different location area.

The GSM standard doesn’t define an algorithm for handover decision therefore network suppliers are responsible to implement them. A basic handover algorithm is specified in appendix A of the ETSI GMS 05.08 [Etsi19942].

##### Intra-Cell-Handover

The channel within the cell will be changed, e.g. when the channel is noisy. A change can either be done to a new frequency or to a new time-slot with the same (old) frequency.

##### Inter-Cell/Intra-BSC-Handover

The channel will be changed between two cells within the same BSC.

##### Intra-Cell/Intra-MSC-Handover

The connection will be changed between two cells in different BSCs but both are managed by the same MSC.

##### Inter-MSC-Handover

The connection will be changed between two cells which are managed by different MSCs.

The first two handovers can be carried out by the BSC if it supports handover. If this is not the case then also Intra-Cell-Handover have to be carried out by the MSC.

##### Ping-pong Handover

A ping-pong handover is an undesirable effect where a handover is made to a neighbor cell and after short time back to the original cell [Junius1995].

### 1.4 Location Area Updates - LAU

Location area updates are used to ease the localization of mobile subscribers. Therefore a MS must initiate a location area update once it leaves the location area. The HLR and VLR both store the current location area. A MS constantly measures the received signal strengths of all surrounding BTSs and reports it to the BSS to the one with the highest signal strength. If it connects to a BTS outside of the current location area then the MS initiates a location area update and tells the CN its new location area. There are two cases which can distinguished:

• change within the same VLR

• change to a new VLR

#### 1.4.1 Location Area - LA

A location area defines the area within the mobile network in which the mobile subscriber is located. Location areas are used to decrease the signaling necessary to locate a mobile subscriber within the network. Without location areas the PLMN has to initiate a paging request to all BTSs. With location areas only the BTS within the location area have to carry out the paging request.

Figure 2: Example for three location areas

One or more BTS with the same MSC can be part of one location area, in most cases they are part of the same BSC. Fig. 3.2 shows that a location area can consists of several BTS. Each location area is uniquely defined by the *Location Area Identity* (LAI). The LAI (see Tab. 3.1) consists of the Mobile Country Code (e.g. Austria 232) which identifies the country, the Mobile Network Code (e.g. A1 1) which identifies the PLMN and the Location Area which identifies the location area within the PLMN. Together with the Cell-ID (CID) the LAI forms the Global Cell-ID (GCID). The size of a location area depends on:

• cell density

• voice and data traffic

• network operator configuration

Table 1: Parts of the Location Area Identity

#### 1.4.2 Updates

For its operation the MSC (HLR, VLR) needs to know where a mobile subscriber is located. Therefore location area updates are required. There are two kinds of location area updates first normal updates and second periodic updates. Normal location area updates (NLAU) happen either when the MS is switched on or off during an IMSI attach/detach or when a MS connects to a new BTS in a different location area. Periodic location area updates (PLAU) are defined by the network operator. Each MS has a PLAU timer for this purpose, whenever the timer expires the MS performs a PLAU.

### 1.5 Antennas

Antennas enable wireless communication between a receiver and a transmitter. In GSM they are the direct air interface between the MS and the Radio Access Network (RAN). Antennas can be designed to transmit or receive signals equally in all directions (omni-directional antennas) or transmit and receive signals in a specific direction (directional or high gain antennas). In GSM networks a combination of both is used. Omni-directional are mainly used in rural areas to cover a large area whereas directional antennas are used to cover a smaller area with a higher traffic load.

#### 1.5.1 Sector

Sector antennas are a special use case of directional antennas where several directional antennas are combined and each covering a particular sector such as a $120\,^{\circ}$ horizontal pattern. In GSM the most common installations of sector antennas are $60\,^{\circ}$, $90\,^{\circ}$ and $120\,^{\circ}$. Based on the traffic load of each area smaller sectors can be combined with larger ones on the same tower (e.g. two $60\,^{\circ}$ antennas with two $120\,^{\circ}$ antennas).

#### 1.5.2 Antenna size

The antenna size or more general its coverage is defined by the antenna gain, the antenna characteristics and the transmit power. In mobile communication networks there are four different types of cells in use which are distinguished by their coverage.

##### Macrocell

Provides the primary radio coverage for a mobile network. Macrocell antennas are mounted on ground-based towers, buildings or other existing infrastructure to have a clear view over the coverage area.

##### Microcell

A microcell offers additional capacity within the coverage of a macrocell. Typically, microcell antennas are mounted at street level. Microcells have lower output power than macrocell and therefore a decreased coverage (e.g. 300 to 1000 meters).

##### Picocell

To provide coverage inside buildings or to a high numbers of users. The coverage of a picocell is 200 meters or less.

##### Femtocell

A low-powered, coverage is on the order of 10 meters, base station designed for use in home or small business [Zhang2011].

## 2 Mobile Network Mobility Simulations

Mobile networks permits its subscribers to move freely within the coverage area. To evaluate and ensure the mobility, mobile network operators need to understand the users mobility. Users mobility can be described by a mobility model. A mobility model describes the users behavior and activity using simulation and analytic models. Simulation models are based on realistic mobility scenarios whereas simplified assumptions about the users movement behavior are the foundation of analytic models.

Two major analytic models which are used in mobile network simulation are the *Random Walk* [Akyildiz2000, Bettstetter2001, Bettstetter2002] and the *Manhattan Grid* [Markoulidakis1997].

### 2.1 Random walk

The mobility of entities in nature is unpredictable, therefore a variety of models exists in literature. These models try to describe the mobility for different scenarios. The random walk model is a memory-less system and therefore maintains no record of previous locations and speed. In each iteration dices a new direction and speed within a predefined range [Camp2002]. An example for a generated random walk is depicted in Fig. 3.3. The speed is limited to the interval $0..5$ and angle is limited to $-15..15^\circ$.

Figure 3: Example for a random walk

### 2.2 Manhattan Grid

In a Manhattan grid scenario streets are aligned in a square grid. It is named after the district Manhattan in the City of New York. Fig. 3.4 shows a map of Manhattan and its road network. It can be seen that roads consists of straight lines and each junction forms a right angle.

The Manhattan Grid mobility model can work on any road network. At each junction the logic is executed which decides if the current street shall be used or a different one. The logic can be modeled as a random number generator with a defined probability density function. For example the probability to drive straight can be $60\%$, turn right $30\%$ and turn left $10\%$. The Manhattan Grid mobility model is an extension to the random walk model where participants are not allowed to move freely but rather move on a defined road network.

Figure 4: Manhattan Road Network

## 3 A1 Data

We are using mobile subscription data from A1. A1 is the largest network operator in Austria with $5.3$ million subscribers. The used data was recorded on the network of A1 between Monday, 22 November 2010 and Sunday, 28 November 2010. The captured interfaces, events and the data structure of the system will be depict in the succeeding subsections.

Fig. 3.5 shows the distribution of events for Monday, 22 November 2010. It can be seen from this that the distribution for handover events is not the same as for mobile originated calls. A strong increase of handover events (see Fig. 3.5a) takes place at 17 o’clock. This is due to the end of work of subscribers. People are leaving from work and start calling there friends while on there way home. On the other hand a strong decrease of call activity (see Fig. 3.5b) takes place between 12 and 13 o’clock where people are having lunch.

[Sorry. Ignored \begin{subfigure} ... \end{subfigure}]

 [Sorry. Ignored \begin{subfigure} ... \end{subfigure}]

Figure 5: Normalized Distribution of Handover Events (a) and Mobile Originated Calls on Monday, 22 November 2010

### 3.1 System Architecture

To capture useful events, monitoring units need to be installed in the Core Network of A1. Fig 3.6 illustrates the system architecture used by A1 to capture events. The architecture is almost the same as for a typical GSM system, the only differences are the monitoring units responsible for capturing and forwarding events. In order to intercept the network, the monitoring units are attached to the main interfaces (A, IuCS, Gb/IuPS, IuPS) of the Core Network. The processing unit is responsible for aggregating events and ensuring anonymity.

Figure 6: System Architecture of the A1 Monitoring System

#### 3.1.1 Interfaces

The monitoring units capture both *Circuit Switched* (CS) and *Packed Switched* (PS) data. Therefore it intercepts the communication of CS and PS interfaces. The CS interfaces are A and IuCS and they belong to the MSC and enable voice calls and sms. The PS interfaces are the Gb/IuPS and the IuPS. Both are connected to the SGSN and enable subscribers to use packed orientated services (e.g. Internet).

### 3.2 Data Structure

Each event captured by the monitoring unit is forwarded to the processing unit for further aggregation. An event is encoded into a binary format and then forwarded to a consumer. Each event has a fixed length (88 bytes) and its structure is shown in Fig. 3.8. Not all fields were used by our system and therefore we only describe the ones we used.

#### 3.2.1 Anonymous ID

This 32 bytes field contains a unique anonymous identifier of the mobile terminal. It changes every day at midnight, i.e. a device can be followed for maximum 24 hours time span before the assigned anonymous ID changes.

#### 3.2.2 Timestamp

The timestamp represents the exact time of occurrence of the event (i.e., the instant when the event has been captured by the monitoring unit).

#### 3.2.3 Latitude and Longitude

These two fields represent the position of the BTS in which the event occurred. They are both encoded as decimal numbers representing the WGS84 coordinates in degrees. The decimal number is encoded according to IEEE 754 floating-point double format[IEEE754].

#### 3.2.4 Event type

Every event includes a field (i.e., Event type) that indicates which type of event has been detected. All possible signaling events are described in Table 3.2.

#### 3.2.5 Angle

This field defines the installment of a sector antenna. It is the angle at which the antenna is mounted on the mast. Fig. 3.7 illustrates the mounting of a directional antenna on a mast.

Figure 7: Example for a antenna installation

[Sorry. Ignored \begin{bytefield} ... \end{bytefield}]

Figure 8: Bitstream structure of the A1 Interface

#### 3.2.6 Encoding example

The following example illustrates the decoding of the input bit stream.

E4A2E4E263A2B54214A2B28E09AE02C36089219123696993EF9E1DF9D8647F18

496DC7340006E61B4030747CF7F849A84048143577861E6800000000000005780200

00000000000023000000000000000000000000000000

- Anonymous ID

Encoded Byte Array = 0xE4A2E4E263A2B54214A2B28E09­AE02C360­89219123696993EF9E1DF9D8647F18

- Timestamp: Wed Jan 14 12:06:28 CET 2009, in microseconds: 1231931188452123

Encoded Byte Array = 0x496DC7340006E61B

- Latitude: 16,45503187

Encoded Byte Array = 0x4030747CF7F849A8

- Longitude: 48,15788168

Encoded Byte Array = 0x4048143577861E68

- Radius: 1400

Encoded Byte Array = 0x0000000000000578

- Input source: Metawin

Encoded Byte Array = 0x02

- Reserved: 0

Encoded Byte Array = 0x000000

- Reserved: 0

Encoded Byte Array = 0x00000000

- Event type: Emergency Call

Encoded Byte Array = 0x23

- Reserved: 0

Encoded Byte Array = 0x00

- Reserved: 0

Encoded Byte Array = 0x0000

- Angle: 0

Encoded Byte Array = 0x0000000000000000

- Speed: 0

Encoded Byte Array = 0x0000000

### 3.3 Events

In this subsection we describe how everyday users activities are visible in the A1 data stream. The type of events detected by the monitoring units from each terminal depends on the type of terminal. While it is not feasible to cover all events (complete list of events is shown in Tab. 3.2), we only describe the ones (mobility related) used in our system.

#### 3.3.1 Calls

When a terminal establishes a call with another subscriber a Mobile Originated Call event is created. The receiving terminal will create a Mobile Terminated Call event. Whenever a call is ended (e.g. one subscriber hangs up)a A Disconnect event is created. However in our research of the data stream, there was no evidence of a A Disconnect event.

• MOBILE TERMINATED CALL (0x1D): terminal receives a call

• MOBILE ORIGINATED CALL (0x22): terminal originates a call

• A DISCONNECT (0x18): The call is closed

#### 3.3.2 Location Area Update

Devices are most of the time in "idle" mode where they are switched on but are not involved in any voice/data communication. In this state the device must still be reachable by the network. Thus, when a terminal changes to a cell in a different location area, it sends a Location Area Update message to the core network. This message appears in the data stream as:

• LOCATION UPDATE (0x1F), in case of 2G terminal

• IUCS LOCATION UPDATE NORMAL (0x15), in case of 3G terminal

If a devices does not change its Location Area for a specific time and the PLAU timer expires it send some kind of keep alive message to the core network.

• IUPS RA PERIODIC UPDATE (0x0C) and IUCS LOC UPD PERIODIC (0x17) in case of 3G terminal (for PS and CS domain respectively)

• LOCATION UPDATE (0x1F) in case of 2G terminals

Fig. 3.9 illustrates where a location area update (red cell) happens when a subscriber traverses different location areas on its way.

Figure 9: Cells and location areas traversed by a subscriber

#### 3.3.3 Handover Cell Update

When a call is established the terminal is said to be in active "connected" state (see 3.1.2). In this state, the network is informed about each change of cell, regardless of the LA to which the cell belongs. A limitation is that the A1 data stream is not able to capture cell changes for 3G terminals (these messages do not reach the core network, but terminate in the radio access network, i.e. they cannot be monitored by the systems). However cell changes for 2G terminals appear as:

• GB CELL CHANGE (0x02), in case of an ongoing data connection,

• HANDOVER CELL UPDATE (0x20), in case of an ongoing call.

## 4 OpenCoverageMap

OpenCoverageMap was a project and a master thesis done by a colleague of mine at the University of Applied Science Upper Austria. Dieter Schlosser [Schlosser2012] goal was to create an open network coverage map for mobile networks. Instead of relying on measurements and coverage reports handed out by network operators he was using smartphones which were doing the measurements. This approach, where users a carrying out work, is know as crowd-sourcing [Surowiecki2004].

### 4.1 Purpose

In our project we are using OpenCoverageMap data to evaluate our approaches. In order to measure the network, OpenCoverageMap captures events similar to the events captured by A1. The OpenCoverageMap application which is installed on the smartphone periodically records the position (using GPS), the phone state as well as the currently connected cell site. Tab. 3.3 shows an example of data which was recorded by the OpenCoverageMap application.

Table 2: Data recorded by OpenCoverageMap for the user 57

The GPS information allows us to evaluate a route estimated by our system with the actual traveled route (see  5.5.2 for more information).

### 4.2 Data Structure

OpenCoverageMap does not capture events, therefore our systems needs to covert the OpenCoverageMap data stream to the same format as the A1 data steam. Call establishment and termination events are generate by comparing the timestamps of each record. OpenCoverageMap logs the state of the MS every second. By iterating over the records and comparing the current state with the previous one it is possible to extract call establishment and termination events. To generate handover events, an iteration over all records is done, each record compares the cell id and lac with the previous one. If both are similar no event is generated, however if they are different a handover event is generated. After the conversion is done the OpenCoverageMap event stream looks the same as the A1 event stream.

## 5 Voronoi Tessellation

We want to generate trajectories for each mobile subscriber and therefore we need the subscribers coarse location. For this purpose we first need an estimation of the coverage area of all cell sites. When a mobile subscriber is connected to a cell site we know that his current position must be withing the coverage area of the cell site. A fast approximation of the coverage area can be done with Voronoi diagrams. As mentioned before Tettamanti2012 et al. [Tettamanti2012] used Voronoi tessellation to estimate the coverage area.

Voronoi diagrams are widely used in computer science, e.g. pattern matching, space division, cluster analysis, collision detection etc. as mentioned by Aurenhammer [Aurenhammer1991].

Figure 10: Voronoi diagram of 16 random points in a 2D space

### 5.1 Point Tessellation

The following assumptions are based on *Voronoi diagrams–a survey of a fundamental geometric data structure* by Aurenhammer [Aurenhammer1991] *Computational Geometry: Algorithms and Applications* by Makr de Berg [Berg2000].

Let $P=\{p\_1,p\_2,...,p\_n\}$ be a set of $n$ distinct points, in our case the location of each cell site. The Voronoi diagram of $P$ is a subdivision of the plane into $n$ cells. The property that each Voronoi cell must fulfill is that a point $q$ lies in the cell respective to a site $p\_i$ only if $dist\left(q,p\_i\right) < dist\left(q,p\_j\right)$ for each $p\_j \in P$ where $i \neq j$. Each cell defines the coverage area for the n-th cell site. Moreover each cell is a (possibly unbounded) open convex polygon. Fig. 3.10 illustrates a Voronoi diagram for 16 random points in a two dimensional space. It can be seen that the outer cells are not closed and therefore shape an open convex polygon. In Section 3.5.2 we show a method to produce closed polygons. \[

dist(p,q)=\sqrt{\left(p\_x-q\_x\right)^2+\left(p\_y-q\_y\right)^2}

\]

(1)

### 5.2 Boundaries

Voronoi diagrams are unbounded which means that the coordinates of vertices can be infinite. Unbounded polygons are undesired in topology computation. If we would use unbounded polygons for coverage estimation the area for this cell sites would be infinite. A simple approach to eliminate this effect is to clip the Voronoi diagram with a rectangular bounding box. However this approach is only satisfying if the sites boundaries shape a rectangle. A better approach is to compute the convex hull over all sites and clip it with the Voronoi diagram.

### 5.3 Cell Tower Segmentation

We mentioned before that a GSM network consists not only of omni directional antennas but also of sectored antennas. The coverage area of each sector is defined by the angle at which the antenna is mounted on the tower. If we would only consider the location of each cell sites, sectors with different angles would have the same Voronoi polygon and therefore the same coverage area. In order to get a better representation of the coverage area for each sector we moved the location of each sector based on its angle $\phi$. The definition of the movement can be found in Equation 3.2. The constant factor means a maximum movement of 3 meters in either direction. \[

\label{eq:move}

x=x+\frac{cos(\phi)}{50000},y=y+\frac{cos(\phi)}{50000}

\]

(2)

**Chapter 6**

# Trajectory Generation Pipeline

In this chapter,

Figure 1: Developed trajectory generation pipeline. The trapezoids represent the single steps in the pipeline, the rounded rectangles illustrate the outputs of the particular steps

## 1 Input Data

This section focuses on input data which is consumed by the system. The gathering of input data is the first step in the developed pipeline. Generally input data can be distinguished between four data types, which are used by the system: geographical data, socio-statistical data, mobile subscription data (A1 and OCM), road network data. The different data types are described in the succeeding sections.

### 1.1 Geographical Input Data

As mentioned before GSM doesn’t expose an accurate location of the subscriber. In GSM only the Cell-ID of the current connected cell is known. However, to calculate a route between the start and end of a call a more accurate location is needed. Because the coverage area of cell can be from as little as up to . Therefore we need to set boundaries in which subscribers can be located. An assumption is that the majority of subscribers will start or end their journey in an urban fabric whereas only a small fraction of subscribers will start or end their journey in open land. In course of this project the geographic information is derived from data from the *CORINE*[[3]](#footnote-3) project which was initiated by the European Commission in 1985. The main purpose of this project was to generate a geographic information system for the member states of the European Union. The *CORINE* land cover (CLC) project is one essential part of the *CORINE* projects which aim to develop a land cover information system for twelve member states of the European Union (Effective 2006). The *CORINE* land cover project distinguishes between five main categories and in total 44 land cover classes. Out of the five only two categories are applicable for the purpose of defining boundaries where a subscriber can be located: *artificial surfaces* and *agricultural areas*. The three remaining categories –*forest and semi-natural areas*, *wetlands* and *water bodies*– can be considered as areas where subscribers are not starting or ending their calls. The *CORINE* land cover data for Austria can be obtained from the following service[[4]](#footnote-4). The land cover data is provided in two different data types, first as a raster image with a grid size of 100 meters and second as a Shapefile..Figure 4.2 illustrates the land cover map[[5]](#footnote-5) of Austria. The red spots within the image represents urban fabric areas whereas green denotes to forest and semi-natural areas.

Figure 2: Corine land cover map of Austria

### 1.2 Socio-statistical Input Data

The second kind of input data used in the generation pipeline are population density maps. Population density maps belong to the class of socio-statistical and provide information about the distribution of population in a geographical area. Gallego et al. [Gallego2010, Gallego2011] describe several disaggregation methods which can be used to generate a population density map of the European Union. Their approach uses information from the *CORINE* land cover project to derive a population density grid with 100 meters grid spacing. The population density grid is used in the developed system to narrow the location of a user within the current connected cell. This works in conjunction with *CORINE* land cover. A more detailed description of the described process will be presented later. The population density grid can be obtained freely from the European Environment Agency (EEA) [[6]](#footnote-6). Similar to the *CORINE* land cover data the population density grid is available in two data types.

### 1.3 A1 Dataset

The mobile subscription data which is provided by A1, is used to generate trajectories for subscribers. To generate trajectories the developed system is fed with an event stream for a special user. This event stream contains all events which have been captured during one day. As we mentioned before in Section 3.3.2 the anonymous user id changes at least every 24 hours. Therefore it is only possible to extract trajectories for one day per subscriber. Each event stream can contain zero, one or calls. Every call in the event stream represents a trajectory.

### 1.4 OCM Dataset

In order to validate our approach we need not only an event stream but also the path traveled by the subscriber. Therefore, the developed system uses an event stream provided by the OpenCoverageMap project. After a conversion to the A1 event format this event stream can be provided to the developed system. In addition to the A1 event stream it contains the traveled path which was recorded via GPS on the mobile phone. The GPS tracks allows the system to validate the route finding process as well as the timing estimation. The comparison between the estimated trajectory and the actual traveled route helps to validate the routing. By calculating the speed from the GPS the system can validate how well the timing estimation has been done. More details to the validation process will be provided in Section 4.4,4.5.

### 1.5 Road Network

The last input data for the developed system is a representation of the road network. Individuals walk or drive on predefined paths: streets, paths, etc. To generate trajectories the system needs to have a representation of the road network to assign the path of subscribers on the road network. A freely available and open-source road network is provided by the OpenStreetMap[[7]](#footnote-7) project. The aim of OpenStreetMap is to provide a world wide map which can be used without any royalties. Contributers all over the world are feeding OpenStreetMap with new data and therefore OpenStreetMap is very accurate for locations where the number of contributers is high. Figure 4.3 illustrates a map[[8]](#footnote-8) of Linz, Austria generated by OpenStreetMap. It can be seen that OpenStreetMap provides a large variety of data (e.g. streets, paths, buildings, POI, etc.).

The developed system utilizes the road network from OpenStreetMap to estimate a start and end position of the trajectory as well for route generation. Additionally OpenStreetMap is used to validate the timing estimation for subscribers where no GPS path is available which is the case for A1 subscribers.

Figure 3: OpenStreetMap map of Linz, Austria

### 1.6 Voronoi Coverage

### 1.7 Coverage Planning

## 2 Call Detection

This section describes the process of filtering calls within each event stream. The two used event streams – A1 and OCM – are using different events to signalize call establishment and termination. After this process the system can generate a trajectory for each call found in the event stream.

### 2.1 A1 Calls

To signalize the establishment of a call in the A1 event stream either a *Mobile Originated Call* or a *Mobile Terminated Call* event is issued. The first event is used when a mobile station initiates a call with another subscriber. The mobile station of the other subscriber will issue the second event once the call has been established with the first subscriber.

When either of the involved subscribers terminates the call a *A Disconnect* event shall be issued by both subscribers. However, during investigation of the event stream we haven’t found any occurrence of this event. Therefore the developed system is using an approach by which it investigates events which have been issued after a call establishment. If the system recognizes either one of the following events: *SMS*,*Mobile Originated Call*,*Mobile Terminated Call*,*Location Area Update* or *IMSI detach* it generates a call termination event before.

### 2.2 OCM Calls

Rather than capturing events in the network the OpenCoverageMap project periodically stores the state of the mobile phone. The state can either be *idle* or *connected*. In order to detect a call the system has to find changes in the state. If a state change is detected the corresponding event is issued. A call establishment is detected when the state changes from *idle* to *connected*. In contrast to the call establishment the call termination is detected when the state changes from *connected* to *idle*.

The detection of call terminations is more accurate for OpenCoverageMap than for A1. This due to the fact that OpenCoverageMap periodically stores the state of the mobile phone whereas in order to detect a call termination for A1 events the system has to recognize a follow-up event.

## 3 Route calculation

The following section presents the process of calculation a route for a mobile subscriber. In general each call is used to calculate one route which is later transfered to a trajectory. The process consists of several steps which will be described in the succeeding sections.

### 3.1 Start and End Position Estimation

The first step in the route calculation process is to estimate the subscribers start and end position. As mentioned before GSM only exposes the Cell-ID which only defines a boundary in which the subscriber is located. To narrow this boundaries the system is using geographical input (*CORNINE* land cover) and socio-statistical maps (population density maps).

By using *CORNINE* land cover data the system can define areas where it is unlikely that a subscriber will start or end his journey. This step can be parameterized by giving each of the CLC classes a percentage factor. The percentage defines how likely it is that a subscriber starts or ends a call in this class.

Our second second assumption is that subscribers will more likely start or end their call in a higher populated area within the cell boundaries. More subscribers are located in higher populated areas than in less populated ones. Therefore the system is using population density maps in order to better estimate the start or end position. Figure 4.4 shows an example for a population density map for Vienna, Austria. In this example it can be seen that the density is higher in the inner districts than in the outer ones.

Figure 4: Population density map of Vienna, Austria. The population density is higher in darker (purple) than in brigher areas

#### 3.1.1 Corine Land Cover Clipping

By clipping and therefore removing unwanted areas the system can narrows the boundaries in which the subscriber is located. To clip the coverage area of a cell the system loads the *Corine* land cover map for the particular area. The next step is to clip the cell coverage area with useful *CORINE* land cover categories such as *artificial surfaces* and *agricultural areas*. This technique known as polygon clipping removes unwanted area –*forest and semi-natural areas*, *wetlands* and *water bodies* – from the cell.

#### 3.1.2 Population Density Estimation

After removing unwanted areas within the cell coverage area the next step is to estimate the subscribers position. This step is the same for start and end positions. By using the population density information of the coverage area the system dices a position based on the population density. Therefore more subscribers will be located in higher populated areas.

##### Population Density Clipping

The first step is to load the population density map for the coverage area. A random number generator with a defined probability density function will created. The probability density function is derived from the population density map. When the system loads the population density map for the coverage area it will get polygons with different population densities. More generally speaking each area with a unique population density is represented by a polygon. Afterwards the random number generator is used to dice a area of interest. By using the population density as function for the random number generator areas with a hight population density will be picked more often than less populated areas. Once the system picked a polygon based on the random number generator it clips this polygon with the coverage area. This process narrows the boundaries of the subscribers location. Following the subscribers location will be set by dicing a location within the remaining coverage area.

##### Random Point in Coverage Area

Dicing a random location into a rectangular shape is an easy task, however the shape of the remaining coverage area will unlikely be a rectangular. A simple approach is using the bounding box of the coverage area polygon and dicing coordinate pairs for the bounding box as long as one coordinate pair lays within the coverage polygon. This simple approach has one major disadvantage because it is unpredictable. There is not estimation for how long it would take that one coordinate pair is contained in the coverage polygon.

A more sophisticated approach is to triangulate the coverage polygon into triangles. Triangulation of convex monotype polygons can be done with algorithms by A. Fournier and D.Y. Montuno [Fournier1984] or Godfried Toussaint [Toussaint1984]. The developed system uses the GDAL project [GDAL] which provides an implementation of polygon triangulation for many platforms and programming languages.

## 4 Route validation

## 5 Timing estimation

**Chapter** **7**

# Implementation

## 1 System Overview

## 2 Architecture

## 3 Call detection / filtering

### 3.1 Call Length Distribution

## 4 Stat and End Position

### 4.1 Proof of Hypothesis

### 4.2 Population and Land Use Information

#### 4.2.1 Population Density

#### 4.2.2 CLC

### 4.3 Polygon Triangulation

### 4.4 Random Point Distribution

### 4.5 PL/R Interface

## 5 Routing

### 5.1 Origin Destination Routing

### 5.2 Route Validation

#### 5.2.1 Time Ratio

#### 5.2.2 Euclidean Distance

#### 5.2.3 Hausdorff metric

## 6 Event clustering

#### 6.0.4 Geometric Clustering

#### 6.0.5 Event Clustering

## 7 Time Synchronization

## 8 Coverage Analysis

### 8.1 Radio Network Planning

### 8.2 Digital Elevation Model

### 8.3 Buildings

### 8.4 Radio Signal Strength

**Chapter 8**

# Results

## 1 Routing

## 2 Timing

### 2.1 Rural

#### 2.1.1 Cluster

#### 2.1.2 Cell

#### 2.1.3 Hybrid

### 2.2 Urban

#### 2.2.1 Cluster

#### 2.2.2 Cell

#### 2.2.3 Hybrid

### 2.3 Highway

#### 2.3.1 Cluster

#### 2.3.2 Cell

#### 2.3.3 Hybrid

## 3 Handover Positions

**Chapter 9**

# Discussion

**Chapter 10**

# Summary

1. Press conference by the Forum Mobilkommunication [press conference link](http://www.fmk.at/Medien/Pressekonferenzen/FMK-Jahrespressekonferenz-2012) [↑](#footnote-ref-1)
2. Project website: <http://www.ftw.at/forschung-innovation/projekte/roadcell> last accessed March 12 2014 [↑](#footnote-ref-2)
3. CORINE: "Coordination of Information on the Environment";

   [http://www.eea.europa.](http://www.eea.europa.eu/publications/COR0-landcover)

   [eu/publications/COR0-landcover](http://www.eea.europa.eu/publications/COR0-landcover), last accessed on March 11,2014 [↑](#footnote-ref-3)
4. <http://www.data.gv.at/datensatz/?id=9246f37d-da69-4442-9504-ebd006a059bb>, last accessed on March 11,2014 [↑](#footnote-ref-4)
5. Image source: <http://www.eea.europa.eu/data-and-maps/figures/corine-land-cover-2000-by-country/clc00_at_national.eps/image_original>l, last accessed on March 11,2014. [↑](#footnote-ref-5)
6. [http://www.eea.europa.eu/data-and-maps/data/population-density-disaggregated-with-](http://www.eea.europa.eu/data-and-maps/data/population-density-disaggregated-with-corine-land-cover-2000-2)

   [corine-land-cover-2000-2](http://www.eea.europa.eu/data-and-maps/data/population-density-disaggregated-with-corine-land-cover-2000-2), last accessed on March 11,2014 [↑](#footnote-ref-6)
7. <http://www.openstreetmap.org/about>, last accessed on March 12, 2014 [↑](#footnote-ref-7)
8. Image source: <http://render.openstreetmap.org/cgi-bin/export?bbox=14.278879165649414,48.31028478774528,14.302825927734377,48.320590159843626&scale=5754&format=pdf>, last accessed on March 12, 2014 [↑](#footnote-ref-8)