

Gregor Nachbagauer

Verkehrsanalyse mittels Mobilfunkdaten

Masterarbeit



Institut für Semantische Datenanalyse / Knowledge Discovery
Laboratory for Brain-Computer Interfaces
Technische Universität Graz
Krenngasse 37, A - 8010 Graz
Vorstand: Assoc. Prof. Dipl.-Ing. Dr.techn. Gernot Müller-Putz

Betreuer: Ass.Prof. Dipl.-Ing. Dr.techn. Reinhold Scherer
Dipl.-Ing. Dipl.-Ing. Thomas Reiter

Begutachter: Ass.Prof. Dipl.-Ing. Dr.techn. Reinhold Scherer

Graz, 2. Oktober 2012

Resultierende Arbeiten

G. Nachbagauer, P. Schosteritsch, T. Reiter, R. Scherer, M. Cik and
M. Fellendorf. Traffic Analysis using Cellular Network Data.
Technische Universität Graz. Paper ITS Kongress 2012

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Ort, Datum

Unterschrift

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

date

signature

Verkehrsanalyse mittels Mobilfunkdaten

Kurzfassung Die Verbreitung von Mobilfunkgeräten hat in den letzten Jahren enorm zugenommen. Da diese Geräte als mobile Sensoren angesehen werden können, von denen die ungefähre Position bekannt sein muss, drängt es sich natürlich auf die daraus resultierenden ortsbezogenen Informationen zu nutzen. Vor allem für verkehrsplanerische Tätigkeiten, wo die derzeit zur Verfügung stehenden Methoden nur sehr wenige Daten liefern, welche die Grundlage für die derzeit eingesetzten Verkehrsmodelle bilden, werden neue Informationsquellen gesucht um diese Datengrundlage weiter verbessern zu können. Diese Arbeit liefert einen kurzen Überblick über bereits durchgeführte Projekte, welche sich mit Mobilfunkdaten in der Verkehrsplanung beschäftigt haben und stellt selbst Anwendungen vor, in welchen diese Daten als Grundlage genutzt werden können. Diese Anwendungen beschäftigen sich mit dem Einsatz von Mobilfunkdaten im öffentlichen Verkehr im Speziellen für den Personentransport im Eisenbahnwesen, zur Ermittlung bevorzugter Reisstrecken, zur Bestimmung eventuell auftretender Störungen des Verkehrsflusses mittels aus den Mobilfunkdaten errechneter Geschwindigkeiten, und zur Bestimmung der Auslastung einer Straße. Sämtliche dieser Aufgabenstellungen wurden mit Daten aus bestehenden Informationsquellen, welche derzeit bei der Verkehrsplanung herangezogen werden, verglichen um die erzielten Ergebnisse validieren zu können. Je nach geforderter Problemstellung zeigte es sich, dass es durchaus sinnvoll ist Mobilfunkdaten unterstützend einzusetzen. Vor allem bei den Anwendungen Personentransport im Eisenbahnwesen und bei der Ermittlung der bevorzugten Reisstrecken wurden für den betrachteten Zeitraum und die jeweiligen Strecken Ergebnisse erzielt, welche mit denen von bestehenden Quellen in einigen Punkten übereinstimmen.

Schlagworte: Mobilfunkdaten, Verkehrsanalyse, Öffentlicher Verkehr, Geschwindigkeit, Verkehrszählmesstellen.

Traffic Analysis using Mobile phone data

Abstract The number of mobile devices increased rapidly over the last years. These devices can be seen as mobile stations for which a rough estimation of their positions has to be known to the mobile network operator. It is obvious to use these location based information for other applications. In particular these data could be useful for the assistance in traffic development. Methods used in traffic development today are very expensive in time and therefore it is not possible to get continuous results which can be used of existing traffic models. For this reason there is a demand to find other resources which provide such information without these expenses. This thesis provides a short overview of projects which make use of mobile phone network data to support traffic planning. Moreover it presents applications for which it could be useful to work with this type of data. These applications use mobile phone network data to count passengers of public transports on railways, they determine the route usage of people and they detect disruptions of traffic routes under the heading of speed and they identify the capacity utilisation of a street. All of these applications have been validated with data from existing resources which are currently in use to support traffic planning. Depending on the problem the results serve a very useful purpose to use mobile phone data to support current methods. Especially the results of the passengers count of public transports on railways and the route usage of people come close to actual reached results.

Keywords: Mobile phone data, Traffic analysis, Public transport, Speed, Traffic measurement points.

Inhaltsverzeichnis

1 Einführung	8
1.1 Motivation und Zielsetzungen	8
1.2 Aufbau und Kapitelübersicht	9
2 Grundlagen	10
2.1 Mobilfunknetz	10
2.1.1 Global System for Mobile Communications (GSM)	10
2.1.2 General Packet Radio Service (GPRS)	13
2.1.3 Universal Mobile Telecommunication System (UMTS)	14
2.1.4 Zusammenfassung Mobilfunknetz	16
2.2 Mobilfunkdaten	16
3 Stand der Technik	19
4 Systemarchitektur	22
4.1 Hardware	22
4.2 Software	22
5 Methoden	22
5.1 Hashing	23
5.2 Multithreading	25
5.3 Distanzberechnung	25
6 Umsetzung	26
6.1 Einlesen und Filtern der Mobilfunkdaten	26
6.2 Abspeicherung der eingelesenen und gefilterten Daten	29
6.3 Erstellen der Anwendungen	32
6.3.1 Train Validation	33
6.3.2 Route Usage	35
6.3.3 Traffic Measurement Count	37
6.3.4 Speed Control	38
7 Ergebnisse, Validierung und Auswertung	39
7.1 Train Validation	39
7.2 Route Usage	42
7.3 Traffic Measurement Count	45
7.4 Speed Control	48
8 Diskussion	51
9 Ausblick	53

Abkürzungsverzeichnis

AuC	Authentication Center
BSS	Base Station Subsystem
BTS	Base Transceiver Station
EIR	Equipment Identity Register
FCD	Floating Car Date
GGSN	Gateway GPRS Support Node
GIP	Graphenintegrationsplattform
GSM	Global System for Mobile Communications
GSN	GPRS Support Node
GPRS	General Packet Radio Services
GPS	Global Positioning System
HLR	Home Location Register
IMEI	International Mobile Equipment Identity
IMSI	International Mobile Subscriber Identity
ISDN	Integrated Service Digital Network
LA	Location Area
LAC	Location Area Code
LAI	Location Area Identity
ME	Mobile Equipment
MS	Mobile Station
MSC	Mobile Switching Center
ORF	Österreichischer Rundfunk
ÖBB	Österreichische Bundesbahnen
PCU	Packet Control Unit
PSTN	Public Switched Telephone Network
RA	Routing Area
RNS	Radio Network Controller
SGSN	Serving GPRS Support Node
SIM	Subscriber Identity Module
UE	User Equipment
UMTS	Universal Mobile Telecommunication System
USIM	UMTS Subscriber Identity Module
UTRAN	UMTS Terrestrial Radio Access Network
VLR	Visitor Location Register

Abbildungsverzeichnis

1	Phasen der Umsetzung	9
2	GSM Architektur mit deren Komponenten und den Schnittstellen [34] . . .	11
3	Basisstation mit drei Sektoren [34]	12
4	GPRS Architektur [34]	14
5	Aufbau UTRAN mit den verwendeten Schnittstellen [14]	15
6	iTraffic System [6]	16
7	GPS Koordinaten (Blau..Antenne, Rot..empirisch ermittelte Koordinaten)	19
8	Visualisierung der unterschiedlichen Stärke des Verkehrsaufkommen an- hand einer Matrix	20
9	Java Code Hashtable	23
10	Java Code Hashmap	24
11	Java Code HashSet	24
12	Java Code MultiValueMap	24
13	Wegpunkte vom Benutzer mit eingezeichnetem Radius	27
14	Koordinaten Filter	29
15	ID mit Koordinaten an welchen diese Events ausgelöst hat (0-1 Uhr) . . .	31
16	Koordinate mit IDs welche dort eine Event ausgelöst haben (0-1 Uhr) . .	32
17	Grafische Darstellung des Ablaufes zur Auffindung der IDs	36
18	Distanz vs. Geschwindigkeit	38
19	Gefundene IDs von Graz nach Salzburg	40
20	ID über die Strecke Graz-Salzburg	41
21	Prozentuelle Aufteilung der gefundenen IDs über 14 Tage	42
22	Zuordnung der IDs zu den jeweiligen Zügen	43
23	Sektoren [Links über S6 (Semmering), Rechts über A2]	44
24	Vergleich Verkehrsmessstelle mit Ergebnissen aus der Anwendung	46
25	Ergebniss bei ungünstiger Lage der Verkehrsmessstelle	47
26	Box Plot für einen Tag Richtung Kufstein	49
27	Verhältnismäßiger Geschwindigkeitsanstieg	50

1 Einführung

1.1 Motivation und Zielsetzungen

Die Bereitstellung von Verkehrsdaten ist ein wesentlicher Bestandteil bei der Planung, dem Entwurf und dem Betrieb von Verkehrsanlagen. Bei der Gestaltung von modernen Verkehrssystemen geht es vor allem darum möglichst viele Informationsquellen zusammenzuführen, um daraus Modelle zu generieren, welche zur Beantwortung aktueller Fragestellungen im Bereich Mobilität herangezogen werden können. AnachB [7] stellt so ein multimodales Informationssystem dar, wobei in diesem Projekt folgenden Daten herangezogen werden:

- Taxi (Floating Car Data, 2500 Taxis)
- Detektoren, Sensoren (Verkehrsmessungen)
- Verkehrsmeldungen (ASFINAG [10], ORF [33], Polizei)
- Öffentlicher Verkehr (Wiener Linien [36], ÖBB [26], Park and Ride)
- Graphenintegrationsplattform GIP [17]

In diesem Projekt geht es darum Verkehrsinformationen dynamisch, intermodal, regional, kostenlos und unter Berücksichtigung von umwelt- und verkehrspolitischen Grundsätzen aufzubereiten, und daraus Verkehrsservices zu generieren. Die meisten der bestehenden Bezugsquellen können aufgrund ihrer Eigenschaften nur eine begrenzte Anzahl von Informationen liefern. Mobilfunkdaten wären ein möglicher weiterer Ansatz, weitere Ausgangsdaten für ein solches System bereitzustellen um diese weiter verbessern zu können. Sie zeichnen sich vor allem dadurch aus, dass sie eine sehr große Anzahl von Daten zur Verfügung stellen, welche für die Positionsbestimmung eingesetzt werden können. Weiters entstehen bei der Verwendung von Mobilfunkdaten keinerlei zusätzliche infrastrukturelle Kosten, da die bereits bestehende Infrastruktur des jeweiligen Mobilfunkbetreibers genutzt werden kann und jedes Mobilfunkgerät als mobiler Detektor angesehen werden kann. Diese Arbeit soll einen Überblick über die möglichen Anwendungsgebiete von Mobilfunkdaten in Bezug auf die Verkehrsplanung schaffen. Dazu wird anhand praktischer Beispiele evaluiert, ob es überhaupt möglich ist aus diesen Daten Kenntnisse für die Planung, dem Entwurf und dem Betrieb von Verkehrsanlagen zu extrahieren. Der erste Punkt beschäftigt sich mit der Analyse der Mobilfunkdaten in Bezug auf den öffentlichen Verkehr im Speziellen für den Personentransport im Eisenbahnwesen. Die zweite Anwendung verfolgt das Ziel eine mögliche Korrelation zwischen einzelnen Verkehrsmessstellen und den Mobilfunkdaten zu finden. Dabei wird ein bestimmter Streckenabschnitt betrachtet und für unterschiedliche Tage und Tageszeiten ermittelt, wie viele Verkehrsteilnehmer mit Hilfe der Mobilfunkdaten sowie mittels der Verkehrsmessstellen detektiert werden können. Weiters wurde eine Anwendung implementiert, welche es ermöglicht die Anzahl von Personen einer bestimmten Strecke zu ermitteln. Dadurch können diese Strecken miteinander verglichen werden und ein Faktor berechnet werden, welcher eine Aussage darüber trifft, welche bevorzugt von Personen verwendet wird.

Das letzte praktische Beispiel befasst sich damit die Geschwindigkeit einzelner Einträge zu bestimmen um auf bestimmten Straßenabschnitten den Verkehrsfluss abbilden zu können. Des Weiteren werden die Probleme, welche beim Umgang mit Daten aus einem Mobilfunknetz entstehen, erläutert und mögliche zukünftige Ansätze für die Verwendung von Mobilfunkdaten in der Verkehrsplanung diskutiert.

1.2 Aufbau und Kapitelübersicht

Der Aufbau der Arbeit sieht folgendermaßen aus, dass zuerst die Grundlagen, welche die Basis für die Arbeit darstellen, vorgestellt werden. Ebenfalls wird hierbei der Datenstream, auf welchen die Mobilfunkdaten beruhen, und die Informationen, welche darin enthalten sind, beschrieben. Nach diesem Kapitel wird ein Überblick über den derzeitigen Stand der Technik der Verkehrsanalyse mittels Mobilfunkdaten gegeben. Im nächsten Schritt wird dann ein kurzer Blick auf die Systemarchitektur gemacht. Hierbei wird auf die benutzte Hardware und Software eingegangen. In diesem Kapitel werden auch die verwendeten Softwarebibliotheken erwähnt und erklärt, warum diese überhaupt zum Einsatz kommen. Im darauffolgenden Kapitel werden die verwendeten Methoden, welche für die Umsetzung der Anwendungen benötigt werden beschrieben. Auf Grund der sehr großen Datenmengen ist ein Schwerpunkt der Arbeit darauf ausgerichtet, die erzielte Laufzeit der Implementierungen so gering wie möglich zu halten. Dazu werden in diesem Kapitel die dabei zum Einsatz kommenden Hilfsmittel und der Grund für die Verwendung dieser beschrieben. Weiters erläutert dieser Teil der Arbeit auch Berechnungen, welche von mehreren Anwendungen durchgeführt werden. Danach folgt die eigentliche Umsetzung des Projektes. Darin werden alle Prozessschritte (siehe Abbildung 1), die im Rahmen des Projektes für die Erfüllung der Zielsetzung notwendig waren, betrachtet. Ebenfalls



Abbildung 1: Phasen der Umsetzung

wird in diesem Kapitel ein Blick auf den Aufbau und die Funktionsweise der implementierten Anwendungen geworfen um zu zeigen, wie die Ergebnisse eigentlich zustande gekommen sind. Nach diesem Punkt kommt die eigentliche Auswertung der Ergebnisse und deren Validierung. Hierbei werden die erzielten Ergebnisse, welche mit Hilfe der vorher implementierten Software erreicht wurden, beschrieben und mittels derzeitig zur Verfügung stehenden Quellen validiert. Zum Schluss erfolgt noch eine Diskussion dieser Ergebnisse und es wird darin auch auf die Probleme, welche im Rahmen dieses Projektes aufgetaucht sind, eingegangen. Weiters wird noch ein Ausblick über die in Zukunft noch durchzuführenden Arbeiten geliefert, um die erreichten Ergebnisse noch verbessern zu können.

2 Grundlagen

In diesem Kapitel werden die Grundlagen, welche für den Aufbau dieser Arbeit von Bedeutung sind, erläutert. Dazu zählen zum einen die Systemarchitektur des Mobilfunknetzes und zum anderen der Aufbau der verwendeten Mobilfunkdaten und deren Inhalt.

2.1 Mobilfunknetz

Bei der Entwicklung vom Mobilfunknetz gibt es drei wesentliche technische Meilensteine:

- GSM - Global System for Mobile Communications
- GPRS - General Packet Radio Services
- UMTS - Universal Mobile Telecommunication System

In den nächsten Zeilen wird ein Überblick über diese Systeme folgen, wobei nicht alle Details sondern nur diejenigen Funktionen näher beschrieben werden, welche für diese Arbeit von Bedeutung sind.

2.1.1 Global System for Mobile Communications (GSM)

GSM [34][14][1] gehört zur zweiten Mobilfunkgeneration (2G) und bildet den ersten digitalen Mobilfunkstandard, welcher 1994 in Österreich eingeführt wurde. Es kann im Wesentlichen in drei Untersysteme unterteilt werden. Die Mobile Station, welche vom Teilnehmer benutzt wird. Das Base Station Subsystem, welches für die Funkverbindung verantwortlich ist und dem Netzwerk Subsystem, worin die Verbindung zu anderen Teilnehmern sowie das Mobilitätsmanagement durchgeführt wird. Abbildung 2 veranschaulicht die Systemarchitektur eines GSM Netzes und die Schnittstellen über welche die einzelnen Untersysteme miteinander kommunizieren.

Mobile Station

Die Mobile Station (MS) besteht aus dem Mobile Equipment (ME) und einer Subscriber Identity Module (SIM). Auf dieser Smartcard stehen Netzbetreiber-spezifische Daten, welche es dem Benutzer erlauben den Mobilfunkdienst überhaupt zu nutzen. Eine dieser Daten ist die sogenannte International Mobile Subscriber Identity (IMSI) mit welcher der Benutzer vom System eindeutig identifiziert werden kann. Beim ME ist dies mit Hilfe der International Mobile Equipment Identity (IMEI) möglich. Die Kommunikation zwischen den Mobile Stations und der Base Transceiver Station erfolgt über das Um-Interface [2][3].

Base Station Subsystem

Das Base Station System (BSS) besteht aus 2 Teilen, dem Base Station Controller (BSC) und der Base Transceiver Station (BTS). Die Kommunikation zwischen diesen beiden Komponenten erfolgt über das standardisierte Abis-Interface [5]. Die BTS ist für

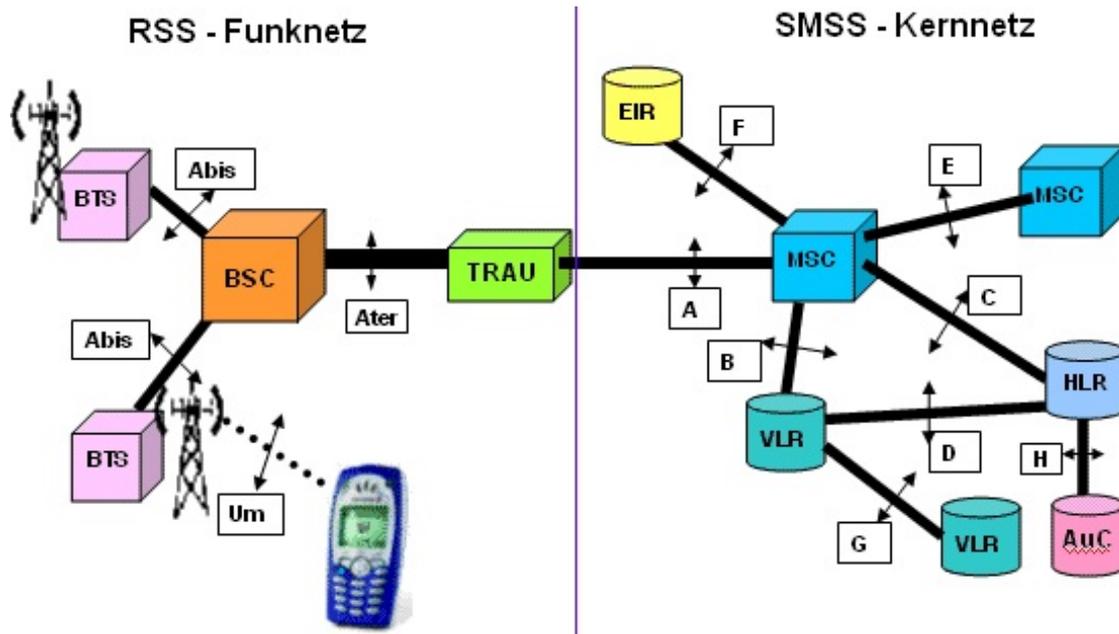


Abbildung 2: GSM Architektur mit deren Komponenten und den Schnittstellen [34]

die Kommunikation innerhalb einer Zelle (in der Praxis meistens 3 siehe Abbildung 3) zuständig und besteht aus sämtlichen Funkkomponenten, welche für eine Übertragung darin benötigt werden. Der BSC kontrolliert und steuert die zu ihm gehörigen BTSs. Reserviert die Kommunikationsfrequenz und steuert den Handover zwischen den BTSs wenn sich eine MS von einer Zelle in die nächste bewegt. Er ist die Verbindung zwischen den Mobile Stations und dem Mobile Switching Center (MSC). Die Anbindung des BSS zum MSC erfolgt über das A-Interface [4].

Network Subsystem

Der Hauptteil dieses Systems bildet das Mobile Switching Center (MSC). Die Funktionsweise entspricht dem eines Public Switched Telephone Network (PSTN) oder Integrated Service Digital Network (ISDN) Switch, welche unter [18] näher definiert ist. Die Hauptfunktionen sind hierbei die Registrierung, Authentifizierung, Location Update, Handover zwischen den BSS und die Weiterleitung zu anderen Netzbetreibern (Roaming). Im Network Subsystem befinden sich weiters das Home Location Register (HLR) und das Visitor Location Register (VLR). Im HLR sind alle statischen und noch zusätzlich relevante Daten wie z.B. der Aufenthaltsort sämtlicher Teilnehmer eines Netzbetreibers enthalten. Im VLR sind für bestimmte geografische Bereiche alle aktiven MSs gespeichert. Wechselt ein Teilnehmer in diesen geografischen Bereich ist das VLR dafür verantwortlich, dass die Teilnehmerdaten aus dem HLR kopiert werden. Ebenfalls befinden sich noch das Authentication Center (AuC), welches für die Authentifizierung von SIM-Karten und der Verschlüsselung über die Funkschnittstelle verantwortlich ist, und das Equipment Identity Register (EIR), welches alle gültigen ME also die jeweiligen IMEIs für den jeweiligen

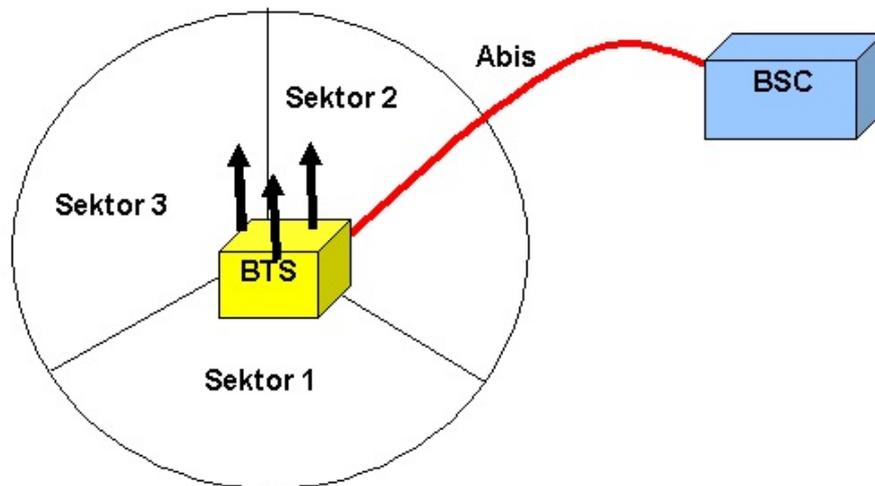


Abbildung 3: Basisstation mit drei Sektoren [34]

Netzbetreiber enthält um z.B. gestohlene MSs zu blockieren, in diesem Subsystem.

Nachfolgend werden noch Abläufe, welche für diese Arbeit von Bedeutung sind und deren Funktionsweise erklärt.

Handover

Ein Handover erfolgt prinzipiell nur dann, wenn die MS aktiv in Verwendung z.B. bei einem Telefonat ist. Dieser muss deswegen erfolgen, da sich die MS aus der Reichweite eines BTS während einer aktiven Tätigkeit bewegen kann und die Signalstärke damit zunehmende abnimmt beziehungsweise verloren geht. Damit dies nicht der Fall ist, gibt es zwei Arten von Handovers interne und externe. Ein interner Handover wird dann durchgeführt, wenn der Teilnehmer die Zelle (BTS) nicht wechselt und die Kommunikationsqualität sich so verschlechtert, dass auf einen anderen Frequenzkanal oder Zeitschlitz innerhalb der Zelle gewechselt oder wenn zwischen den BTSs eines BSC umgeschaltet werden muss. Diese Form des Handovers wird nur von den BSCs abgewickelt. Das MSC wird erst nach erfolgreicher Durchführung benachrichtigt. Von einem externen Handover spricht man dann, wenn zwischen Zellen von unterschiedlichen BSCs, welche zum gleichem MSC gehören oder zwischen Zellen unterschiedlicher MSCs umgeschaltet wird, wobei ursprüngliche MSC für die meisten anrufbezogenen Funktionen verantwortlich bleibt. Einen Handover auslösen kann sowohl die MS wie auch das MSC (für Load Balancing).

Location Update

Ein Location Update erfolgt dann, wenn eine Mobile Station, durch Empfang der Zell Location Area Identity (LAI) feststellt, dass sie sich in einer neuen Location Area (LA) befindet. Unter einer Location Area werden mehrere Zellen zusammengefasst, welche in

einem VLR verwaltet werden. Solange ein Teilnehmer innerhalb einer LA befindet, werden keine gesonderten Update Prozeduren zur Reduktion des Overheads durchgeführt. Je nach Netzbetreiber wird zusätzlich noch ein periodisches Location Update ausgeführt. Wird nun ein Update durchgeführt, wird die neue Position im VLR upgedatet und die Adresse des Eintrages im VLR im HLR abgelegt.

Einen Spezialfall bildet die erste Registrierung in einem Mobilfunknetz z.B. beim Einschalten der MS nach längerer Zeit. Dies ist darum besonders, da nach längerer Zeit die Teilnehmerdaten aus dem VLR gelöscht werden und im HLR abgelegt werden und somit keinerlei Daten im VLR existieren, die jedoch notwendig sind. Darum erfolgt hierbei kein Location Update sondern eine Location Registration mit der IMSI und der LAI der momentanen Zelle in welcher sich die MS befindet. Nach erfolgreicher Authentisierung mittels des AuC erfolgt der Eintrag ins VLR.

Aus-/Eingehender Anruf

Bei einem ausgehenden Anruf wird eine Anfrage über die am nächst liegende BTS zum Mobile Switching Center gesendet, in welchem mittels dem VLR überprüft wird, ob der Teilnehmer im Netz registriert ist oder nicht. Ist dieser registriert leitet das MSC den Anruf ans Netzwerk weiter und der Anruf wird aufgebaut.

Bei einem eingehenden Anruf wird mit Hilfe der Rufnummer der Eintrag im HLR des jeweiligen Teilnehmers identifiziert und mittels der im HLR gespeicherten VLR Adresse aus dem VLR die Location Area bestimmt in welcher sich der Teilnehmer befindet. Daraufhin wird in sämtlichen Zellen der LA ein Pagingruf mit dem im VLR abgelegten Informationen durchgeführt. Stimmen die Daten des Pagingrufes mit den Daten einer MS überein meldet sich die MS und der Anruf kann dann über die jeweilige Zelle an die MS weitergeleitet werden.

2.1.2 General Packet Radio Service (GPRS)

GPRS [34][1][20] stellt eine Erweiterung des GSM-Netzwerkes dar und ermöglichte es erstmals paketorientierte Datendienste zu nutzen. Es unterstützt mehrere Protokolle zur Datenübertragung, unter anderem IP und X.25. Der wesentliche Unterschied zum leitungsorientierten GSM-Netzwerk besteht darin das für die Dauer einer Verbindung eines Teilnehmers nicht permanent eine Übertragungsleitung zugewiesen ist. Bei GPRS teilen sich mehrerer Teilnehmer einen Übertragungskanal. Die Daten werden dabei in Pakete aufgeteilt und übertragen. Überträgt ein Teilnehmer keine Daten, kann in der Zwischenzeit ein anderer Teilnehmer seine Pakete über denselben Kanal übertragen. Im Wesentlichen wird dabei die bestehende Infrastruktur von GSM benutzt, wobei einige neue Netzwerkkomponenten zum GSM Netz hinzugefügt beziehungsweise alte Komponenten modifiziert wurden. Abbildung 4 veranschaulicht die Systemarchitektur des GPRS Netzwerkes mit den verwendeten Schnittstellen. Es wurden lediglich sogenannte GPRS Support Nodes (GSN) zum GSM Netz hinzugefügt. Dabei wird zwischen dem Gateway GPRS Support Node (GGSN), welcher für die Anbindung an externe Datenetze verantwortlich ist und dem Serving GPRS Support Node (SGSN), welcher für den

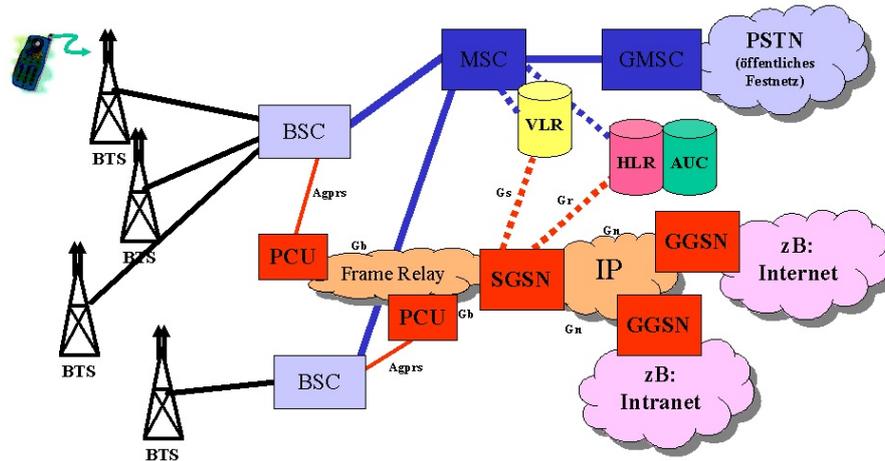


Abbildung 4: GPRS Architektur [34]

paketorientierten Dienst die Funktion des MSC übernimmt. Weiters mussten die BSCs mit der sogenannten Packet Control Unit (PCU) erweitert werden um mit dem SGSN kommunizieren zu können. Je nachdem, ob es sich um herkömmliche Sprachdaten oder um Paketdaten handelt, leitet der BSC diese entweder zum MSC oder über die PCU zum SGSN weiter. Bei den anderen Komponenten des GSM Netzes wurden lediglich Softwareupdates durchgeführt, um mit den verwendeten Funktionen von GPRS umgehen zu können. Handover beziehungsweise Location Updates wurden modifiziert, wobei die grundlegende Funktion, wie unter 2.1.1 beschrieben, erhalten geblieben ist. Unter einem Handover bei GPRS wird ein Zellupdate, welches von der MS durchgeführt wird verstanden. Anstelle der Location Area werden bei GPRS mehrere Zellen zur sogenannte Routing Area (RA) zusammengefasst.

Eine MS kann folgende drei Status, IDLE, STANDBY und ACTIVE aufweisen. Befindet sich eine MS im IDLE State, ist die GPRS Funktion für diese MS nicht im System registriert und aufgrund der daraus fehlenden Position ist es nicht möglich eine Nachricht an diese MS zu senden. Im STANDBY State ist dem SGSN lediglich die Routing Area bekannt in welcher sich die MS befindet. Mittels Paging siehe GSM Aus-/Eingehender Anruf wird danach die aktuelle Zelle in welcher sich die MS befindet ermittelt. Lediglich im ACTIVE State werden Daten zwischen der MS und dem GPRS Netzwerk übertragen. Der SGSN kennt während dieses Zustandes die aktuelle Zelle der MS.

2.1.3 Universal Mobile Telecommunication System (UMTS)

UMTS [34][14][1] gehört zur sogenannten dritten Mobilfunkgeneration (3G). Er wurde deswegen eingeführt, um zur Verfügung stehenden Frequenz-Ressourcen besser und effektiver nutzen zu können und dabei höhere Datenraten bei besserer Sicherheit zu erzielen. UMTS baut im Gegensatz zu GPRS nicht komplett auf dem GSM Netz auf. Es verfügt über einen komplett neues Funknetzsystem. Dieses System wird als UMTS Terrestrial Radio Access Network (UTRAN) bezeichnet. Die anderen Elemente (Core Network)

konnten mit kleineren Modifikationen für UMTS weiter verwendet werden. Nachfolgende Abbildung 5 veranschaulicht den Aufbau vom UTRAN mit den verwendeten Schnittstellen und deren Bezeichnung. Der Radio Network Controller (RNC) übernimmt die Funktion, welche der BSC im GSM/GPRS Netzwerk innehat. Zusätzlich übernimmt er noch einige Managementfunktionen vom Funknetzwerk, welche bei GSM/GPRS vom Core Network durchgeführt werden. NodeB bezeichnet die Funkkomponenten, welche die Zellen mit Funksignalen versorgt und ist das Äquivalent zur BTS und kann aus Kostengründen auch auf ein und denselben Masten montiert werden. Das UMTS User Equipment bezeichnet das Mobile Equipment in Verbindung mit dem UMTS Subscriber Identity Module (USIM), welche die gleiche Funktion wie das SIM bei GSM/GPRS übernimmt. Ein wesentlicher Unterschied zum GSM-Funknetz stellt die Verwendung der

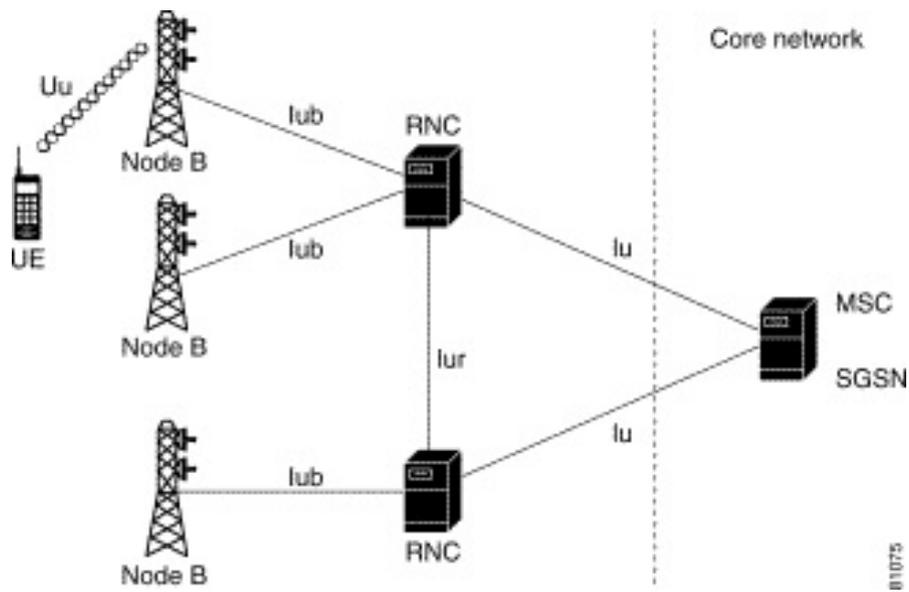


Abbildung 5: Aufbau UTRAN mit den verwendeten Schnittstellen [14]

Makrodiversität dar. Makrodiversität bedeutet, dass mehr als nur eine Funkverbindung von einem UE verwendet wird. Vereinfacht gesagt, es wird eine neue Verbindung aufgebaut bevor die alte unterbrochen wird. Es werden dabei drei Arten der Makrodiversität im UTRAN unterschieden.

- Softer Handover
Das UE befindet sich dabei im Wirkungsbereich zweier Zellen, welche zur selben NodeB gehören.
- Intra RNC Softhandover
Das UE befindet sich im Wirkungsbereich zweier Zellen unterschiedlicher NodeBs, welche jedoch am selben RNC angeschlossen sind.
- Inter RNC Softhandover
Das UE befindet sich im Wirkungsbereich zweier Zellen unterschiedlicher NodeBs, welche jedoch an unterschiedliche RNCs angeschlossen sind.

2.1.4 Zusammenfassung Mobilfunknetz

In diesem Unterkapitel werden die wichtigsten Erkenntnisse aus der Netzwerktopologie des Mobilfunknetzes, welche für den weiteren Verlauf dieser Arbeit von Bedeutung sind nochmals kurz zusammengefasst. Der erste wichtige Punkt ist, wann werden überhaupt Informationen gesendet, welche für die Positionsbestimmung eingesetzt werden können. Bei der passiven Verwendung eines mobilen Gerätes kommt es nur beim Übertritt von einer Location Area (LA) in eine Andere zu einem Event, welcher im Mobilfunknetz dementsprechend mitprotokolliert wird. Wird das mobile Gerät in irgendeiner Form aktiv verwendet kommen mehr Events zustande, da zwischen den einzelnen Mobilfunkmasten eine Prozess gestartet werden muss, damit ein Signalverlust vermieden wird. Die zweite wichtige Erkenntnis ist diese, dass anhand der Funktionsweise des Mobilfunknetzes nicht gewährleistet sein muss, dass ein Benutzer immer am vermuteten nächstgelegenen Masten angemeldet ist, sondern das dies durch eventuell auftretenden Einflussfaktoren (Geografische Begebenheiten, Netzüberlastungen, Maktrodiversität etc.) mehr oder weniger stark variieren kann.

2.2 Mobilfunkdaten

Die Mobilfunkdaten werden vom iTraffic System (Abbildung 6) der Mobilkom Austria Group zur Verfügung gestellt. Dieses System fasst mehrere Informationen beziehungs-

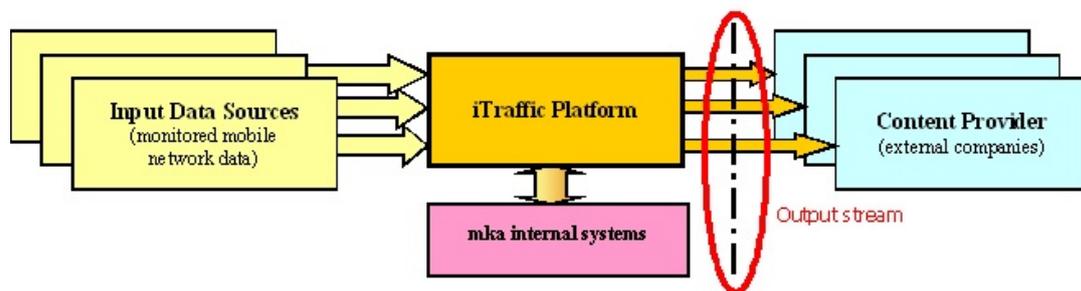


Abbildung 6: iTraffic System [6]

weise Events von unterschiedlicher Herkunft im Betreibernetz zusammen und es verfügt über eine Schnittstelle mit welcher diese Daten an Dritte weitergeleitet werden können. Die übermittelnden Daten werden danach Stundenweise abgespeichert um damit spätere Analysen durchführen zu können. Jedes dieser Events wird binär codiert und hat eine festgelegte Länge und beinhaltet mehrere Informationen.

Anonymous ID

Pseudonymisierte IMSI des Teilnehmers, welche für einen Tag unverändert bleibt. Weitere Events an einem Tag können somit ebenfalls dieser Pseudonymisierten IMSI zugeordnet werden.

Timestamp

Zeitmarke des auftretenden Events.

Event Type

Dieses Feld beinhaltet den Typ des Events. Sämtliche Events sind in Tabelle 1 ersichtlich. Die genaue Bedeutung der einzelnen Events ist unter [30] nachzulesen.

Latitude

Dieser Eintrag beschreibt die geographische Breit dargestellt in Dezimalgrad.

Longitude

Beinhaltet die geographische Länge dargestellt in Dezimalgrad.

Input Source

Liefert die jeweilige Quelle des Events. Folgende drei Quellen sind spezifiziert:

- Datacast Input Stream (2G-Daten)
- Metacast Input Stream (3G-Daten)
- GPS Input (Appello [9])

Die beiden GPS Koordinaten, Longitude und Latitude, entsprechen bei den ersten beiden Quellen nicht den tatsächlichen, sondern den durch ein empirisches Verfahren bestimmten Aufenthaltsortes des Mobilfunkteilnehmers. Dieses empirische Verfahren berücksichtigt sowohl die Ausrichtung der Antennen und die Wahrscheinlichkeit, dass sich ein mobiles Gerät an einer bestimmten Position aufhält. Im weiteren Verlauf dieser Ausarbeitung wird unter dem Begriff GPS Koordianten immer dieser, durch das empirische Verfahren ermittelter Aufenthaltsort, verstanden. Wie bereits in 2.1 erklärt sind fast immer drei Antennen an einem BTS (GSM/GPRS) oder NodeB (UMTS) montiert und deswegen ergeben sich folgenden Aufenthaltsorte welche in Abbildung 7 ersichtlich sind. Die blauen Punkte entsprechen den Mobilfunkmasten und die orangen zeigen die Aufenthaltsorte, welche durch das empirische Verfahren bestimmt wurden. Vor allem im urbanen Bereich führt diese Methode zu einer wesentlichen Abweichung zur tatsächlichen Position.

Event Type
GB.ATTACH
GB.CELL.UPDATE
GB.OTHERS
GB.ROUTING_AREA.UPDATE_RA.UPDATING
IU.CS.PAGING
IU.PS.PAGING
IU.OTHER.TLLI
IU.OTHER.TMSI
IUPS.ATTACH
IUPS.DETACH
IUPS.ROUTING_AREA.UPDATE_RA.UPDATING
IUPS.ROUTING_AREA.UPDATE.COMBINED_RA.LA.UPDATING
IUPS.ROUTING_AREA.UPDATE.COMBINED_RA.LA.UPDATING WITH IMSI.ATTACH
IUPS.ROUTING_AREA.UPDATE.PERIODIC.UPDATING
IUCS.DETACH
IUCS.SMS.MOBILE.ORIGINATED
IUCS.SMS.MOBILE.TERMINATED
IUCS.SETUP
IUCS.CONNECT.ACKNOWLEDGE
IUCS.DISCONNECT
IUCS.LOCATION.UPDATE.NORMAL
IUCS.LOCATION.UPDATE.PERIODIC
IUCS.LOCATION.UPDATE.IMSI.ATTACH
A.DISCONNECT
A.OTHER
DETACH
SHORT.MESSAGE.SERVICE
MOBILE.TERMINATING.CALL
CM.REESTABLISHMENT
LOCATION.UPDATE
HANDOVER.CELL.UPDATE
SUPPLEMENTARY.SERVICE
MOBILE.ORIGINATED.CALL
EMERGENCY.CALL
SETUP
CONNECT.ACKNOWLEDGE
CLOSURE
GB.PAGING.PS
GPS.NOTIFY

Tabelle 1: Event Typen [6]



Abbildung 7: GPS Koordinaten (Blau..Antenne, Rot..empirisch ermittelte Koordinaten)

Ist ein Event von der Quelle GPS (Appello) so entsprechen die GPS Koordinaten der tatsächlichen Position des Mobilfunkteilnehmers. Diese Anzahl ist jedoch nur in sehr geringem Maße verfügbar. Der genaue prozentuelle Anteil kann aus [30] entnommen werden

3 Stand der Technik

Derzeit gibt es nur eine beschränkte Anzahl an Informationsquellen, welche Daten zur Verfügung stellen die echten Tatsachen (Zählungen etc.) entsprechen um draus Verkehrsmodelle zu generieren. Die ursprünglichsten dabei eingesetzten Methoden sind zum einen Befragungen, welche mittels eines Wegetagebuches durchgeführt werden und die Verwendung von Detektoren. Bei den Befragungen werden Informationen (Reiseziel und Zweck, Soziodemografischen Daten usw.) von einzelnen Personen erfasst und diese verarbeitet. Aus den erhobenen Daten werden dann Quell-Ziel Matrizen gebildet. Diese Matrizen bilden das Verkehrsaufkommen zwischen den einzelnen Einträgen (Bezirke etc.) ab. Abbildung 8 zeigt eine grafische Darstellung von einer dieser Matrizen aus welcher die verschiedenen Stärken des Verkehrsaufkommen zwischen unterschiedlichen geografischen Bereichen ersichtlich ist. Bei der zweiten Methode werden stationär angebrachte Detektoren (Induktionsschleifen, Verkehrskameras etc.) zur Bestimmung von Verkehrscharakteristiken herangezogen. Mit diesem Verfahren ist es jedoch nicht möglich Fahrtenverläufe von einzelnen Verkehrsteilnehmern zu bestimmen, da diese sich meist in einem örtlich begrenzten Umfeld befinden und somit nur eine beschränkte Repräsentativität auf die Grundgesamtheit (alle Menschen in Österreich) liefern. Bei der Verwendung von Verkehrskameras ist es im Prinzip möglich durch die Identifikation mittels des KFZ Kenn-

3 Stand der Technik

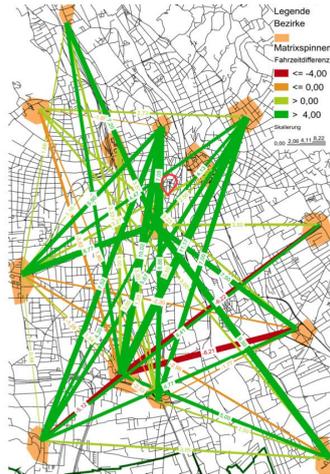


Abbildung 8: Visualisierung der unterschiedlichen Stärke des Verkehrsaufkommen anhand einer Matrix

zeichens, einzelne Verkehrsteilnehmer auch auf längere Strecken zu detektieren und somit Fahrtenverläufe rekonstruieren zu können. Die Probleme dabei sind jedoch die Kosten der benötigten Infrastruktur und der Datenschutz, da keine personenbezogenen Daten über einen längeren Zeitraum für solche Zwecke gespeichert werden dürfen. Das Datenschutzproblem kann jedoch durch Anonymisierung beziehungsweise Pseudonymisierung behoben werden wo hingegen der Kostenfaktor durch die damit verbundene Infrastruktur weiter ansteigt. Einen Überblick über diese Methoden liefert [23]. Weiters werden Verkehrsdaten von bereits bestehenden Systemen, wie etwa von [10] und auch von [33] unterstützend bei der Generierung von Verkehrsmodellen eingesetzt.

Seit mehreren Jahren wird auch der Einsatz von Mobilfunkdaten im Verkehrswesen erforscht. Eine Vielzahl der Publikationen bedient sich speziell ausgerüsteter Geräte beziehungsweise sie nutzen die integrierte GPS Funktion von Mobilfunkgeräten um ortsbezogene Daten zu erhalten. Es werden dabei nicht die Daten aus dem Mobilfunknetz, sondern eigens genierte Daten, für die Analysen herangezogen. Die meisten dieser Auswertungen bedienen sich sogenannter Floating Car Daten (FCD). Dazu werden ein oder mehrere Beobachtungsobjekte, welche in den Verkehrsfluss integriert werden, eingesetzt, um Verkehrsdaten zu liefern. Aus diesen gesammelten Daten ist es dann möglich Verkehrsmodelle für einen bestimmten Bereich zu generieren. [22] liefert einen Überblick über die Projekte, welche anhand von FCD durchgeführt wurden. Aufgrund der hohen Strukturkosten, eignen sich solche Systeme nur bedingt für den dauerhaften Einsatz bei der Erstellung von Verkehrsanalysen beziehungsweise Verkehrsmodellen.

Der zweite Ansatz ist die Daten, die im Mobilfunknetz selbst generiert werden als Grundlage zu verwenden. Auch auf diesem Sektor gibt es schon einige Veröffentlichungen, die nur mit diesen Daten arbeiten. Da es nur möglich ist aktive Benutzer eines Mobilfunkgerätes genauer zu detektieren kann hierbei wieder eine Unterscheidung getroffen werden, ob sämtliche Daten oder nur die Daten der aktiven Benutzer für die Analysen herangezo-

gen werden. Werden nur die aktiven Benutzer herangezogen, führt dies dazu, dass sich die Anzahl der zur Verfügung stehenden Daten stark reduziert wird. Das Paper [29] zeigt, dass es bei der Verwendung von Location Area Codes (LAC) und den dazugehörigen Updates möglich ist, Trajektorien zu bilden, welche danach einer Straße zugeordnet werden können. Es wurden dabei geeignete Filter implementiert, welche zunächst auf die Daten angewendet wurden, um sämtliche IDs des betrachteten geografischen Gebietes zu erhalten, welche sich von einer Location Area in eine andere bewegt haben, um somit den Fahrtenverlauf nachbilden zu können. Es zeigte sich, dass es durchaus möglich ist mit Hilfe von Location Area Updates, IDs einer Straße zuzuordnen zu können, wobei IDs, welche sich nur in einer Location Area aufgehalten haben aufgrund der verwendeten Methodik nicht detektiert werden konnten.

Ein weiteres Projekt wurde in Italien für die Stadt Rom durchgeführt.[13] Dabei wurde ein online Stream, welcher von der Telecom Italia zur Verfügung gestellt wurde, benutzt, um Verkehrsanalysen durchzuführen. Es erfolgte dabei eine räumliche Unterteilung in mehrere Quadrate, welche eine Größe von 250x250m im städtischen Bereich und 500x500m im außerstädtischen Bereich aufweisen. Die Größe ergibt sich aus der Dichte der jeweiligen Funkmasten in einem geografischen Bereich. Für jedes dieser Quadrate wurden dann mehrere Parameter, wie zum Beispiel die Anzahl der User, welche sich darin bewegen, ermittelt. Die dabei zur Verfügung stehenden Mobilfunkdaten beinhalten jedoch wesentlich mehr Informationen wie zum Beispiel die Signalstärke, welche zur Positionsbestimmung eingesetzt werden konnten. Somit war es ihnen möglich verschiedensten Anwendungen zu erstellen, welche anhand der bei dieser Arbeit zur Verfügung stehenden Daten gar nicht möglich wären.

[11] liefert einen guten Überblick über die bis 2010 durchgeführten Projekte, welche sich Mobilfunkdaten zu nutze gemacht haben. Ebenfalls wird darin eine Anwendung für den Großraum Amsterdam vorgestellt. Im wesentlichen wurden dabei alle Events, welche an einer bestimmten Position aufgetreten sind über einen längeren Zeitraum erfasst und danach verglichen. Es zeigt sich, dass sich je nach Position und Zeitpunkt (Wochenende, Wochentag etc.) gewisse Parameter wie zum Beispiel die Gesprächsintensität verändern. In dieser Arbeit werden nicht nur Events von mobilen Geräten, welche aktiv verwendet werden, für die Analysen herangezogen sondern es wird der Ansatz verwendet, dass sämtliche Events gleichermaßen berücksichtigt werden. Dies hat vor allem den Grund, dass vorerst eine Aussage darüber getroffen werden soll, ob die für diese Arbeit zur Verfügung stehenden Daten generell für verkehrsplanerische Tätigkeiten herangezogen werden können oder nicht. Für spätere speziellere Aufgabenstellung wäre es sicher von Vorteil nur aktive Geräte zu verwenden, da es mit Hilfe der dadurch reduzierten Anzahl von Informationen trotzdem noch möglich sein sollte Analysen durchzuführen zu können (siehe [22]).

4 Systemarchitektur

4.1 Hardware

Die ganze Implementierung wurde so ausgelegt, dass keine besonderen Hardwareanforderungen notwendig sind. Es ist jedoch von Vorteil, wenn ein möglichst großer Hauptspeicher zur Verfügung steht, da alle Objekte, welche gespeichert werden, in diesen geschrieben werden. Die Anwendungen wurden auf folgenden System getestet:

- Prozessor: Intel(R) Core(TM) i7-2630QM CPU @ 2.00GHz
- Hauptspeicher: 8GB DDR3
- Festplatten: Toshiba 650GB mit 7200RPM

4.2 Software

Die zu Implementierenden Anwendungen wurden vollständig die Java entwickelt. Als Entwicklungsumgebung wurde Eclipse SDK Version 3.7.0 verwendet. Um die Anwendung ausführen zu können ist lediglich eine Installation von Java erforderlich. Mit folgender Java Version wurde die Anwendung getestet:

- Version 6 Update 29

Die grafische Repräsentation der Landkarte wurde mit Hilfe des JXMapKit durchgeführt. JXMapKit ist ein Projekt von Swinglabs.org[32] und verfügt standardmäßig bereits über sehr viele nützliche Methoden um Anwendungen im Geoinformationsbereich leichter realisieren zu können. Das Kartenmaterial wird dabei von OpenStreetMaps[27] zur Verfügung gestellt, da für die Benützung keinerlei Lizenzen notwendig sind. Die digitalen Karten werden dabei kachelbasiert angeboten und über einen definierten Webservice abgerufen. Dazu mussten folgende Libraries in das Projekt mit eingebunden werden:

- swingx-1.6.jar
- swingx-ws-1.0.jar

Zur Erstellung der Diagramme wurde JFreeChart[15] verwendet, welches nur noch als Library in die Anwendung miteingebunden werden musste. Mit Hilfe dieser Library ist es möglich eine Vielzahl von Diagrammtypen zu erstellen und diese je nach Anwendung zu verändern.

5 Methoden

Dieser Abschnitt soll einen Überblick über die in der Arbeit verwendeten Methoden beziehungsweise deren Funktionsweise liefern. Dies Methoden wurden vor allem deswegen verwendet, da diese speziell für den Umgang mit großen Datenmengen (ca. 50GB pro Tag), so wie sie in dieser Arbeit für sämtliche Aufgabenstellungen vorhanden sind, konzipiert sind.

5.1 Hashing

Dieses Verfahren wird dazu verwendet, um Datenelemente in großen Datenmengen zu finden. Es wird vor allem darum eingesetzt, da mit Hilfe dieser Methode üblicherweise das Einfügen beziehungsweise das Entfernen in konstanter Zeit $O(1)$ ermöglicht wird. Dabei wird zu jedem Datenobjekt mittels einer Hashfunktion ein Hashwert berechnet, welcher danach als Index in einer Tabelle verwendet wird. Zur Berechnung dieses Hashwertes wird zumindest ein Schlüssel (String, Integer etc.) benötigt. Im Idealfall verfügt jedes dieser Datenobjekte über einen eigenen Hashwert. Ist dies nicht der Fall, so muss dies extra berücksichtigt werden um das gewünschte Ergebnis zu erhalten, da jeder Schlüssel einzigartig sein muss. Hashing wurde in den entwickelten Anwendungen deswegen eingesetzt, da es häufig notwendig ist aus einer Vielzahl von Daten möglichst effizient ein doppeltes Auftreten ein und der selben ID erkennen zu können. Ist dies der Fall so ist es damit möglich so ein Mehrfachvorkommen einfach zu ignorieren, dass heißt sämtliche andere Events einer ID außer acht zu lassen oder diese zu berücksichtigen und sie dementsprechend zu verarbeiten, dass sie zu einem späteren Zeitpunkt zur Verfügung stehen. Es sollte dabei jedoch berücksichtigt werden, dass möglichst sparsam mit den Ressourcen umgegangen wird und nur die nötigsten Mehrfachvorkommen berücksichtigt werden. In Java gibt es folgenden Datenstrukturen welche Hashing verwenden:

- Hashtable
- HashMap
- HashSet

Hashtable

Stellen die grundlegende Speicherstruktur dar, um zu einem jeweiligen Schlüssel den dazugehörigen Wert zu speichern. Sowohl der Schlüssel wie auch der Wert müssen ungleich null sein, da es ansonsten zu einer NullPointerException kommt. Weiters sind Hashtables synchronisiert. Das bedeutet, dass zu ein und derselben Zeit nur ein Thread auf diese Zugriff hat.

```
Hashtable<Integer,String>; cityTable = new Hashtable<Integer,String>();
cityTable.put(1, "Graz");
cityTable.put(2, "Wien");
cityTable.put(3, null); /* NullPointerException at runtime*/
```

Abbildung 9: Java Code Hashtable

HashMap

Erlauben ebenfalls Schlüssel-Wertpaare, wobei sowohl der Schlüssel wie auch der Wert null sein können. Ein weiterer Unterschied besteht auch noch darin, dass Hashmaps nicht synchronisiert sind, was zu einer besseren Performance beiträgt.

```
HashMap<Integer,String> productMap = new HashMap<Integer,String>();  
productMap.put(1, "Auto");  
productMap.put(2, "Motorrad");  
productMap.put(3, null);
```

Abbildung 10: Java Code HashMap

HashSet

Hashsets werden vor allem dafür eingesetzt, um eine Liste mit einzigartigen Einträgen zu erhalten. Es erfolgt dabei keine Zuordnung zu einem Schlüssel und somit kann jeder Wert nur einmalig vorkommen. Mehrere Anforderungen erfordern es jedoch, dass zu

```
HashSet<String> stateSet = new HashSet<String>();  
stateSet.add ("CA");  
stateSet.add ("WI");  
stateSet.add ("NY");
```

Abbildung 11: Java Code HashSet

einem Schlüssel mehrere Werte gespeichert werden können. Um dies zu gewährleisten wird die von der Apache Commons Collections 3.2 API[8] zur Verfügung gestellte MultiValueMap benutzt. Dabei werden zu einem Schlüssel mehrere Werte gespeichert. Diese Werte werden in Form einer Liste zu dem jeweiligen Schlüssel gespeichert.

```
MultiValueMap flatSet = new MultiValueMap();  
flatSet.put("Kueche","Herd");  
flatSet.put("Kueche","Kuehlschrank");  
flatSet.put("Bad","Spiegel");  
flatSet.put("Bad","Waschmaschine");
```

Abbildung 12: Java Code MultiValueMap

5.2 Multithreading

Darunter wird das gleichzeitige Abarbeiten mehrerer Threads innerhalb eines einzelnen Prozesses verstanden. Es kann sowohl nur auf einem wie auch auf mehrere Prozessorkerne durchgeführt werden. Der Unterschied dabei besteht darin, dass es bei einem Prozessorkern nur zu einer quasi gleichzeitigen Abarbeitung kommt wo hingegen bei mehreren Kernen es zu einer echt-parallelen Ausführung kommt. Bei einem Prozessorkern führt die Erstellung von zu vielen Threads zu einer wesentlichen Reduktion der Gesamtleistung. Aus diesem Grund sollten die Anzahl mehrere Threads möglichst gering gehalten werden. Jedoch auch bei Prozessoren mit mehreren Kernen kann eine zu große Anzahl von Threads, welche gleichzeitig abgearbeitet werden sollen zu einem zu großen Overhead führen.

5.3 Distanzberechnung

Zur Berechnung der Distanz zwischen zwei GPS Koordinaten gibt es mehrere Möglichkeiten. Sämtliche dieser Berechnungsvarianten wurden aus [16] entnommen. Die einfachste Methode beruht auf den Satz des Pythagoras, da für kurze Entfernungen die Längen- und Breitenkreise gerade Linien darstellen und parallel beziehungsweise rechtwinklig zueinander verlaufen.

$$\begin{aligned}
 dis &= \sqrt{dx * dx + dy * dy} \\
 dis_x &= 71.5 * (lon1 - lon2) \\
 dis_y &= 111.3 * (lat1 - lat2) \\
 lat1, lat2, lon1, lon2 &\text{ in Grad}
 \end{aligned} \tag{1}$$

Die Konstante 111.3 in Formel 1 entspricht dabei dem Abstand zwischen zwei Breitenkreisen in km und 71.5 ist der durchschnittliche Abstand zwischen zwei Längenkreisen in den Breiten in denen sich Mitteleuropa befindet.

Das Problem dabei liegt in diesem Abstand zwischen zwei Längenkreisen. Dieser variiert in der Abhängigkeit der geografischen Breite. Aus diesem Grund gibt es eine verbesserte Formel zur Entfernungsberechnung die diesen Umstand berücksichtigt.

$$\begin{aligned}
 dis &= \sqrt{dx * dx + dy * dy} \\
 dis_x &= 111.3 * \cos((lat1 + lat2)/2 * 0.01745) * (lon1 - lon2) \\
 dis_y &= 111.3 * (lat1 - lat2) \\
 lat1, lat2, lon1, lon2 &\text{ in Grad}
 \end{aligned} \tag{2}$$

In dieser Arbeit wird jedoch die Entfernungsberechnung durchgeführt, welche auch die Kugeloberfläche der Erde berücksichtigt. Der Grund für die Entscheidung diese Methode zu wählen liegt einfach darin, dass damit ausgeschlossen ist, dass es bei größeren Distanzen zu Abweichungen kommen kann, da hierbei die Längen- und Breitenkreise kein rechtwinkliges Gitternetz mehr bilden.

Darum wird bei dieser Berechnung zur Methode der sphärischen Trigonometrie gegriffen um die gewünschte Distanz zu erhalten.

$$dis = 6378.137 * \arccos(\sin(lat1) * \sin(lat2) + \cos(lat1) * \cos(lat2) * \cos(lon2 - lon1)) \quad (3)$$

lat1, lat2, lon1, lon2 in Rad

Bei allen vorgestellten Methoden muss darauf Rücksicht genommen werden, dass bei den meisten Programmiersprachen der Winkel in Radiant angegeben werden muss. Die Umrechnung von Grad auf Rad erfolgt durch die folgende Formel.

$$1^\circ = \pi/180rad \approx 0.01745 \quad (4)$$

6 Umsetzung

Grundsätzlich lässt sich die Umsetzung in drei große Arbeitspakete unterteilen, welche folgendermaßen aussehen:

1. Einlesen und Filtern der Mobilfunkdaten
2. Abspeicherung der eingelesenen und gefilterten Daten
3. Erstellen der Anwendungen

6.1 Einlesen und Filtern der Mobilfunkdaten

In der ersten Phase geht es vor allem darum, dass die Daten, welche vom iTraffic System zur Verfügung gestellt werden eingelesen und gefiltert werden. Beim Einlesen der Daten kommt es aufgrund der großen Datenmenge, welche sich bei ca. 50GB pro Tag befindet, vor allem darauf an, ein Verfahren zum Einlesen zu wählen, welches möglichst effizient arbeitet und die Daten so aufzubereiten, dass sie für möglichst viele Anwendungen herangezogen werden können, ohne die gesamten Daten für einen bestimmten Zeitraum neuerlich einlesen zu müssen. Aus diesem Grund kann der Benutzer der Anwendung eine Strecke seiner Wahl mittels von ihm gesetzten Wegpunkten markieren, und die Daten werden danach so eingelesen, dass zum einen GPS Koordinaten mit den aufgetretenen Events einer ID und zum anderen sämtliche IDs mit den jeweiligen GPS Koordinaten gespeichert werden. Diese Wegpunkte können vom Benutzer auf der Landkarte mittels der Maus eingezeichnet werden. Je nach Anzahl dieser definierten Wegpunkte kann eine unterschiedliche Genauigkeit bei der Zuordnung zu einer Strecke erzielt werden.

Bei den IDs mit den jeweiligen GPS Koordinaten wird nur der erste aufgetretenen Event an dieser Koordinate gespeichert. Das hat jedoch diesen Nachteil, dass somit für eine Strecke jede ID nur in eine Richtung unterwegs sein kann, da sie an einer GPS Koordinate zu einem späteren Zeitpunkt nicht mehr berücksichtigt wird. Da diese Form der Speicherstruktur aber derzeit nur für die Geschwindigkeitsberechnung benötigt wird, wirkt sich dieser Nachteil nicht dramatisch auf das gewünschte Ergebnis aus, da die

Anzahl der detektierten Geschwindigkeiten trotzdem ausreichend ist. Jeder dieser vom Benutzer gesetzten Wegpunkte verfügt über einen definierten Radius, in welchen sich die GPS Koordinaten der Events befinden können. Weiters werden diese Wegpunkte durchnummeriert. Dies hat diesen Grund, da damit die GPS Koordinaten einem Wegpunkt zugeordnet werden können, was wiederum später bei der Richtungsbestimmung benötigt wird. Ebenfalls können mehrere Wegpunkte zu einem Set zusammengefügt werden, damit ein geografischer Bereich besser betrachtet werden kann. Die Anzahl der Wegpunkte und der Radius von diesen Wegpunkten kann dabei vom Benutzer frei gewählt werden. Abbildung 13 veranschaulicht beispielhaft wie der Benutzer eine beliebige Strecke markieren kann. Somit ist es möglich je nach gewünschter späterer Verwendung einen un-

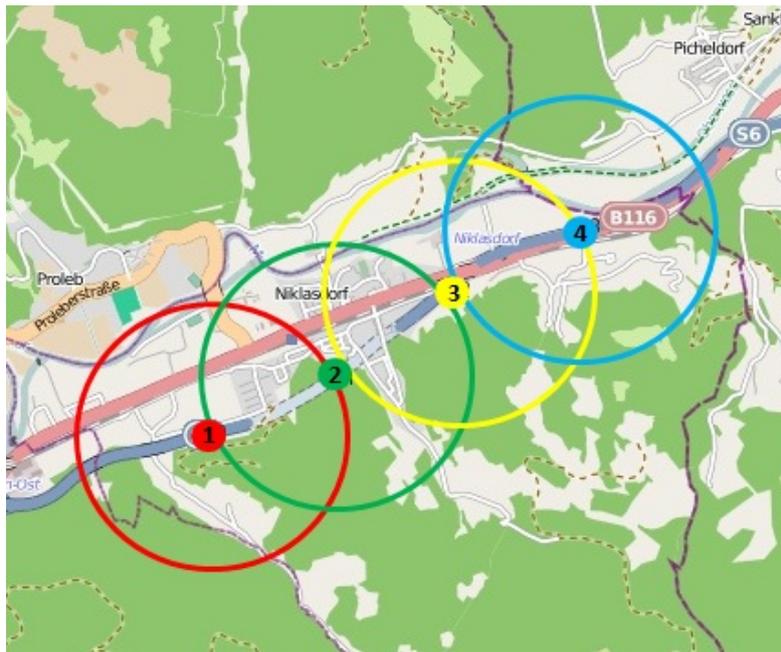


Abbildung 13: Wegpunkte vom Benutzer mit eingezeichnetem Radius

terschiedlichen Genauigkeitsgrad erzielen zu können. Folgende Tabelle 2 veranschaulicht diese beiden Speicherstrukturen. Mit Hilfe dieser Daten ist es nun möglich unterschied-

IDs	GPS Koordinaten
GPS Koordinate A, Zeit	ID A, Zeit, Eventtyp etc.
GPS Koordinate B, Zeit	ID B, Zeit, Eventtyp etc.
GPS Koordinate C, Zeit	ID C, Zeit, Eventtyp etc.
usw.	usw.

Tabelle 2: Speicherstruktur

lichste Analysen zu einer Strecke durchzuführen.

Um ein geeignetes Verfahren zum Einlesen ermitteln zu können wurden mehrere Methoden zum Einlesen getestet um danach die unterschiedlichen Geschwindigkeiten beim Einlesen miteinander verglichen. Es stellte sich dabei heraus, dass es keinen signifikanten Unterschied ausmacht, welche Methode gewählt wird, solange folgenden Dinge berücksichtigt werden:

- Einlesen eines Byteblockes und nicht jedes Byte einzeln einlesen (z.B.: 128 KByte)
- Verwendung von non-thread-safe Klassen (z.B.: FileChannel)
- Vermeidung von Kopiervorgängen zwischen der Java Virtual Machine und dem Betriebssystem, durch den Einsatz von Memory Mapping oder direkt allokierten Buffern

Die ermittelten Ergebnisse decken sich auch mit Ergebnissen von anderen Referenzen siehe [25][31]. Die Anwendungen dieser Arbeit verwenden zum Einlesen einen FileChannel mit einem direkt allokierten Buffer von 128 KByte. Das heißt, es werden immer 128KByte eingelesen und diese danach dementsprechend weiterverarbeitet.

Um die Dauer des Einlesens noch zu verbessern wurde Multithreading eingesetzt. Dabei wurde der Prozess des Einlesens auf zwei Threads aufgeteilt. Die Aufgabe des ersten Threads bestand dabei darin, die Events einzulesen und die einzelnen Inhalte dieses Events (Latitude, Longitude etc.) zu dekodieren und diesen für den zweiten Thread zur Verfügung zu stellen. Dieser wiederum hatte die Aufgabe die jeweiligen Events zu filtern. Hierbei geht es darum, nur diejenigen Events zu bekommen, welche sich auf irgendeiner GPS Koordinate auf der markierten Strecke des Benutzers befinden. Um dabei nicht sämtliche GPS Koordinaten mit allen Wegpunkten vergleichen zu müssen, wurde eine eigene Filtermethode implementiert.

Filter

Zuerst wurden sämtliche vom Benutzer gesetzten Wegpunkte, zum einen anhand ihrer geografischen Breite und zum anderen anhand ihrer geografischen Länge sortiert. Dies wurde mittels einem Treeset[19] und eigens generierten Vergleichklassen realisiert. Danach wurden die zwei benachbarten Einträge des einzufügenden Events mittels den dafür vorgesehen Operationen des Treesets ermittelt. Dieser Vorgang wird zum einen für das Treeset, welches nach der geografischen Breite und zum anderen für das Treeset, welches nach geografischer Länge sortiert ist durchgeführt. Als Ergebnis bleiben vier mögliche Wegpunkte über, zu denen der Event zugeordnet werden kann. Diese vier Wegpunkte beschreiben die jenigen Punkte, welche in alle vier Himmelsrichtungen die jeweils kürzeste Distanz aufweisen. Abbildung 14 veranschaulicht dieses vorgehen. Der Event kann nur zu einem dieser vier gefundenen Wegpunkte gehören. Alle Anderen können vernachlässigt werden. Somit müssen lediglich diese vier Wegpunkt herangezogen werden. Ist die Distanz zu einem dieser vier Wegpunkte kleiner als der vom Benutzer definierte Radius, handelt es sich um ein Event, welcher der markierten Strecke zugeordnet werden kann. Ist die Distanz zu mehreren Wegpunkten unter dem vordefinierten Radius, dann wird der Event demjenigen Wegpunkt zugeordnet, zu welchem dieser die geringste Distanz

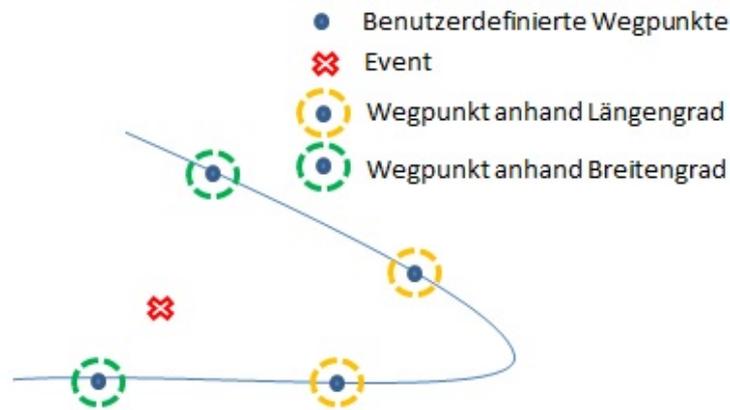


Abbildung 14: Koordinaten Filter

aufweist. Es zeigte sich, dass es mit Hilfe der Aufteilung auf zwei Threads und dem selbst erstellten Filter zu einer Reduktion der Laufzeit um bis zu ca. 40% kommt.

Der nächste Punkt beschäftigte sich damit herauszufinden, ob die Laufzeit noch weiter verbessert werden kann, wenn mehrere Files (also mehrere Stunden) gleichzeitig eingelesen und gefilter werden. Als Indikatoren wurden dafür zum einen die Einlesedauer und zum anderen die CPU Auslastung verwendet. Es stellte sich heraus, dass das gleichzeitige Einlesen von drei Files zur kürzesten Zeit beim Einlesen führt. Aus diesem Grund werden jeweils drei Files (Stunde) gleichzeitig eingelesen und durch einen eigenen Thread gefiltert. Diese Anzahl der Files, welche gleichzeitig eingelesen werden können variiert jedoch in der Abhängigkeit von der jeweils eingesetzten Systemarchitektur. Aus diesem Grund ist es jederzeit möglich diese Anzahl zu reduzieren oder zu erhöhen. Mit Hilfe dieser Faktoren ist es möglich eine Strecke für einen Tag innerhalb von ca. einer Stunde einzulesen und zu filtern.

6.2 Abspeicherung der eingelesenen und gefilterten Daten

Die eingelesenen und gefilterten Daten werden so abgespeichert, dass sie für spätere Analysen durch die erstellten Anwendungen zur Verfügung stehen. Die hat vor allem den Grund, dass das Einlesen von einer Strecke nur einmal zu erfolgen hat, und die daraus ermittelten Daten auch für andere Anwendungen noch verwendet werden können. Dazu werden die gewonnen Daten in Files geschrieben, welche danach in eine eigens erstellte Ordnerhierarchie abgespeichert werden. Diese Hierarchie sieht folgendermaßen aus.

- Anwendungsname
- Streckennamen
- Tag

- Typ
- Stunde
- Files

Der Anwendungsname ergibt aus der Anwendung, für welche die Daten primär eingelesen wurden. Natürlich ist es auch möglich, dass Daten, welche für eine Anwendung eingelesen wurden auch für eine Andere herangezogen werden können. Der Streckenname wird vom Benutzer zu Beginn des Einlesevorganges eingetragen. Damit ist es möglich, dass jederzeit weitere Tage einer Strecke zugeordnet werden können. Dazu muss der Benutzer lediglich einen Streckennamen eintragen, welcher bereits vorhanden ist. Der Typ gibt an wie die Daten abgespeichert werden. Wie bereits vorher erwähnt werden zum einen die GPS Koordinaten mit allen IDs, die bei diesen Koordinaten detektiert wurden, und zum anderen alle IDs mit den GPS Koordinaten, welche von dieser ID passiert wurden, gespeichert. Dies wird vor allem deswegen durchgeführt, da es beim Einlesen zu keiner wesentlich schlechteren Programmlaufzeit kommt, da sämtliche Objekte sowieso im Hauptspeicher angelegt sind und die Daten danach für die Anwendungen nicht wieder umgeschichtet werden müssen. Aus demselben Grund werden auch gleich für jede ID Geschwindigkeitsmatrizen und Richtungsmatrizen gebildet und ebenfalls in das dazugehörige File gespeichert. Zur Bildung der Geschwindigkeitsmatrizen werden sämtliche GPS Koordinaten einer ID betrachtet und mit Hilfe des jeweiligen Zeitstempels t und der Distanz s (Formel 3) zwischen den GPS Koordinaten die Geschwindigkeit v mit $v=s/t$ berechnet. Somit wurde für alle IDs einer Strecke sämtliche Geschwindigkeiten dieser ID berechnet und gespeichert. Die Richtungsmatrix wird in Abhängigkeit von den Zeitstempeln und der Angehörigkeit der GPS Koordinate zu einem vom User gesetzten Wegpunkt bestimmt. Es wird dabei eine Matrix aufgebaut, welche die Werte null, eins und eine Konstante beinhalten kann. Der Wert null steht dafür, dass sich die ID in Richtung des vom Benutzer zuerst gesetzten Wegpunktes bewegt. Der Wert eins steht für das Gegenteil, also eine ID bewegt sich in Richtung des vom Benutzer zuletzt gesetzten Wegpunktes. Findet die Bewegung zwischen den GPS Koordinaten auf einem Wegpunkt statt wird dies mittels der Konstante *99999999.99* gekennzeichnet.

Weiters werden noch sämtliche vom Benutzer gesetzten Wegpunkte in das jeweilige File geschrieben. Somit ist es zum einen möglich jederzeit die von Benutzer selektierte Strecke wieder zu verwenden um zum Beispiel einen neuen Tag einlesen zu können und zum anderen wird dabei beim Typ GPS Koordinaten mit IDs gleich die jeweilige Zugehörigkeit zu einem Wegpunkt mitprotokolliert.

Diese gespeicherten Files werden danach von den jeweiligen Anwendungen eingelesen und deren Daten für die weiteren Analysen verwendet. Nachfolgenden Abbildungen 15 und 16 veranschaulichen exemplarisch die so abgespeicherten Dateien.

Abbildung 15 zeigt die Speicherstruktur einer ID X und die Koordinaten an denen diese ID X innerhalb einer Stunde einen Event ausgelöst hat. Als „Control Points“ werden dabei die Wegpunkte, welche von Benutzer auf der Karte bevor dem Einlesen eingezeichnet wurden, bezeichnet. Jeder dieser Punkte verfügt wie bereits erwähnt über einen bestimmten Radius, welcher in km angegeben ist und über eine SetID zu welchem Set

ID X

Graz-Wien_A2

Control Points:

```
FilterGPS [lat=47.07058949794368, lon=15.435104370117188, rad=5.0, setID=0]
FilterGPS [lat=47.275735100729406, lon=16.019439697265625, rad=2.0, setID=1]
FilterGPS [lat=47.274337475394645, lon=16.046905517578125, rad=2.0, setID=1]
.....
FilterGPS [lat=47.56193242677904, lon=16.121749877929688, rad=2.0, setID=1]
FilterGPS [lat=47.73262723170871, lon=16.158485412597656, rad=1.0, setID=2]
.....
FilterGPS [lat=47.89263630713117, lon=16.206893920898438, rad=1.0, setID=2]
```

Mobile Radio Stations:

```
0 MobRadioStations_NoEntities [latitude=47.46338653564453,
longitude=16.0958194732666, timestamp=07.05.2012 00:51:33, userWayPoint=13]
1 MobRadioStations_NoEntities [latitude=47.52595901489258,
longitude=16.113311767578125, timestamp=07.05.2012 00:56:21, userWayPoint=17]
```

Distance Matrix in Meter

```
DisMat - 0.0000,7075.4310,
DisMat - 7075.4310,0.0000,
```

Time Difference Matrix in Seconds

```
TimeDiffMat - 0.0000,288.0000,
TimeDiffMat - 288.0000,0.0000,
```

Speed Matrix in km/h

```
Speed - 0.0000,88.4429,
Speed - 88.4429,0.0000,
```

Direction Matrix

```
DirMat - 99999999.9900,1.0000,
DirMat - 1.0000,99999999.9900,
```

Abbildung 15: ID mit Koordinaten an welchen diese Events ausgelöst hat (0-1 Uhr)

6 Umsetzung

dieser zugeordnet wurde. Die jeweiligen Koordinaten an denen ein Event aufgetreten ist, wurden als „Mobile Radion Stations“ bezeichnet. Es werden dabei zu jeder dieser Einträge die Zeit und der Wegpunkt, zu welchem er zugeordnet wurde gespeichert. Am Ende jeder Datei stehen die für eine ID ermittelten Distanz-, Zeit-, Geschwindigkeits- und Richtungsmatrizen.

Abbildung 16 veranschaulicht wie IDs zu einer dieser empirisch ermittelten Koordinaten

Koordinate X

Graz-Wien_A2

Control Points:

```
0 FilterGPS [lat=47.07058949794368, lon=15.435104370117188, rad=5.0, setID=0]
1 FilterGPS [lat=47.275735100729406, lon=16.019439697265625, rad=2.0, setID=1]
2 FilterGPS [lat=47.274337475394645, lon=16.046905517578125, rad=2.0, setID=1]
.....
19 FilterGPS [lat=47.56193242677904, lon=16.121749877929688, rad=2.0, setID=1]
20 FilterGPS [lat=47.73262723170871, lon=16.158485412597656, rad=1.0, setID=2]
.....
35 FilterGPS [lat=47.85832433461554, lon=16.18560791015625, rad=1.0, setID=2] X
.....
39 FilterGPS [lat=47.89263630713117, lon=16.206893920898438, rad=1.0, setID=2]
```

Entities:

```
0 Entity [id=EA29EE2E46CD4725FF6264D19721EF5436A47283A96B2F5E43660B6E6025327D,
timeStampList=07.05.2012 00:22:59,07.05.2012 00:23:00]
1 Entity [id=2A3F6729EA897B2A7CD885B7DC05F292F7FBD3406695C11B9380388E4910422A,
timeStampList=07.05.2012 00:49:55,07.05.2012 00:49:54,07.05.2012 00:01:18]
.....
9 Entity [id=C9D6A1DE64F8B055E919974D2D654DDDF7FBD3406695C11B9380388E4910422A,
timeStampList=07.05.2012 00:55:17,07.05.2012 00:55:19,07.05.2012 00:56:01,
07.05.2012 00:56:00,07.05.2012 00:02:56,07.05.2012 00:56:43,07.05.2012 00:56:45,
07.05.2012 00:29:44,07.05.2012 00:58:12,07.05.2012 00:58:10,07.05.2012 00:57:26,
07.05.2012 00:58:55,07.05.2012 00:57:28,07.05.2012 00:58:53,07.05.2012 00:29:43]
.....
14 Entity [id=105860975E2D147E4AB68788F14D34A0F7FBD3406695C11B9380388E4910422A,
timeStampList=07.05.2012 00:10:36,07.05.2012 00:10:33,07.05.2012 00:10:37]
```

Abbildung 16: Koordinate mit IDs welche dort eine Event ausgelöst haben (0-1 Uhr)

für eine Stunde gespeichert werden. Jede dieser IDs verfügt dabei über zumindest einen Zeitstempel. Das „X“ bei den Kontrollpunkten markiert dabei zu welchem vom Benutzer gesetzten Wegpunkt die Koordinate zugeordnet wurde.

Mit Hilfe dieser Dateien ist es möglich mehrere Analysen einer Strecke durchzuführen, ohne dass diese neuerliche eingelesen werden muss. Sämtliche erstellte Anwendungen greifen auf diese Daten zurück um daraus die gewünschten Informationen zu erhalten.

6.3 Erstellen der Anwendungen

Das letzte Arbeitspaket beschäftigte sich mit der Erstellung von Anwendungen. Der Schwerpunkt dabei liegt auf der Analyse von Langstrecken, da hierbei eine wesentlich bessere Zuordnung zu einer Straße möglich ist. Im urbanen Raum ist dies durch die ungenaue Positionsbestimmung mittels des empirischen Modells nicht gewährleistet. Aus

diesem Grund wurden folgende vier Anwendungen realisiert.

- Der erste Punkt beschäftigt sich mit der Analyse der Mobilfunkdaten in Bezug auf den öffentlichen Verkehr im speziellen für den Personentransport im Eisenbahnwesen. Es geht vor allem darum, eine Aussage zu treffen, wie viele Personen von A nach B fahren. (Train Validation)
- Bei der dieser Anwendung geht es darum, eine Aussage zu treffen, welche Strecke von Personen bevorzugt ausgewählt wird, um von einem Punkt zu einem anderen Punkt zu gelangen. (Route Usage)
- Die dritte Anwendung beschäftigt sich damit, eine Korrelation zwischen einzelnen Verkehrsmessstellen und den Mobilfunkdaten zu finden. Dabei wird ein bestimmter Streckenabschnitt betrachtet und für unterschiedliche Tage und Tageszeiten ermittelt, wie viele Verkehrsteilnehmer mit Hilfe der Mobilfunkdaten detektiert werden können. Diese Ergebnisse werden danach mit den Werten von Verkehrsmessstellen verglichen, um einen möglichen Korrelationsfaktor bestimmen zu können. (Traffic Measurement Count)
- Die nächste Implementierung ermittelt die Geschwindigkeit einzelner Einträge um auf bestimmten Straßenabschnitten den Fluss des Verkehrs abbilden zu können. Dieses Verfahren kann als Basis für die Erkennung für zeitlich-räumliche Störungen herangezogen werden. (Speed Control)

Sämtliche dieser Anwendungen werden mit Messungen, Zählungen und bestehenden Verkehrsmodellen verglichen, ob die mittels der Anwendungen erhaltenen Ergebnisse mit den realen Ergebnissen oder mit den Ergebnissen aus den unterschiedlichen Modellen ganz, teilweise oder überhaupt nicht übereinstimmen.

Die ersten drei oben genannten Anwendungen (Train Validation, Traffic Measurement Count und Route Usage) verwenden ähnliche Abläufe. Aus diesem Grund wird die erste (Train Validation) detaillierter vorgestellt und bei den anderen nur die Unterschiede, welche zu dieser genauer vorgestellten Anwendung bestehen, beschrieben.

6.3.1 Train Validation

Bei dieser Anwendung wird zuerst mittels eines Dialoges, die gewünschte bereits eingeleseene Bahnstrecke und der jeweilig gewünschte Zeitrahmen ausgewählt. Dieser Zeitrahmen kann sich von nur einer Stunde bis hin zu einem ganzen Tag erstrecken. Als Typ werden dabei die GPS Koordinaten mit den IDs verwendet, da bei dieser Anwendung die jeweiligen IDs zu den gewissen Bahnhöfen benötigt werden. Die Implementierung ist so ausgelegt, dass derzeit nur ein Zwischenbahnhof berücksichtigt wird. Jeder der Bahnhöfe entspricht einem vom Benutzer gesetzten Wegpunkt. Wobei der erstgesetzte Wegpunkt als Startbahnhof und der letztgesetzte Wegpunkt als Zielbahnhof gesehen wird. Dies ist nur eine formale Definition, da dies auch umgekehrt möglich wäre, da zu diesem Zeitpunkt noch nicht bekannt ist, welcher dieser beiden tatsächlich der Start- oder Zielbahnhof ist. Jeder dieser Wegpunkte verfügt über einen beliebig definierbaren

Radius. Damit ist es möglich ein Ergebnis zu erhalten, welches wirklich nur diejenigen Personen liefert, welche sich im näheren Umkreis eines ausgewählten Bahnhofes aufgehalten haben.

Zuerst werden die ausgewählten Stunden eingelesen und so gespeichert, dass zu jeder Stunde sämtliche IDs der Bahnhöfe in folgender Form $Stunde = 0 \{Bahnhof = 0 \{idX\{ZeitmarkeX, ZeitmarkeY \dots\}, idY, idZ \dots\}, \dots Bahnhof = 2 \{idC, idD, idX \dots\}\}$ vorliegen. Um dieses Ergebnis zu erhalten müssen sämtliche GPS Koordinaten welche für eine Stunde gefunden wurden den jeweiligen Bahnhöfen zugeordnet werden. Dazu werden die einzelnen Files, von denen jeweils eines einer GPS Koordinate entspricht, ebenfalls Zeile für Zeile eingelesen. Wird eine Zeile gefunden, welche mit X endet wird aus dieser Zeile die jeweilige Nummer des Wegpunktes extrahiert. Danach werden sämtliche IDs mit allen an dieser GPS Koordinate detektierten Zeitmarken zu diesem Bahnhof hinzugefügt. Da es nicht notwendig ist sämtliche Zeitmarken einer ID zu erfassen, werden nur die Zeitmarken, welche einen bestimmten definierten zeitlichen Abstand (z.B.: 10 Minuten) zueinander aufweisen mit berücksichtigt. Dies hat diesen Grund, da es für das Ergebnis keinen Einfluss hat, wenn eine ID an einer GPS Koordinaten innerhalb dieses Abstandes mehrmals detektiert wurde, da diese in diesem Zeitraum sowieso nur in eine Richtung unterwegs sein kann. Dieser zeitliche Abstand sollte je nach Strecke so angepasst werden, dass möglichst keine Einträge einer ID welche diese Strecke öfters verwendet hat verloren gehen.

Diese Vorgehensweise wird nur für sämtlich ausgewählte Stunden wiederholt, bis der gewünschte Zeitrahmen vollständig ist und sämtliche Stunden in der gewünschten Form vorliegen. Nach diesem Einlesen und zuordnen der IDs, werden danach die IDs herausgefiltert, welche in dem eingelesenen Zeitrahmen an sämtlichen Bahnhöfen detektiert wurden. Von diesen IDs werden danach die jeweiligen Zeitmarken von allen Stunden zu dieser ID sortiert zusammengefügt, sodass von jeder ID zu allen Bahnhöfen sämtliche Zeitmarken zur Verfügung stehen. $ID = X \{Bahnhof = 0 \{ZeitmarkeA, ZeitmarkeB \dots\} \dots Bahnhof 2 \{ZeitmarkeX, ZeitmarkeY \dots\}\}$ wobei ZeitmarkeA_X vor Zeitmarke B_Y ist. Aus diesem Ergebnis kann das zeitliche Verhalten einer ID auf dieser Strecke nachvollzogen werden. Dazu werden zuerst die ersten Zeitmarken vom Startbahnhof und vom Zielbahnhof miteinander verglichen. Daraus kann zuerst einmal eine Richtung bestimmt werden. Gibt es zwischen diesen beiden Zeitmarken auch noch eine Zeitmarke beim Zwischenbahnhof so werden zu dieser ID diese Zeitmarken hinzugefügt. Verfügt eine ID am Startbahnhof über mehrere Zeitmarken, welche vor der Endzeit liegen, so wird die jeweils späteste als Startzeit definiert. Das gleiche gilt für die Zeitmarken am Zwischenbahnhof. $ID = X \{Startzeit, Zwischenzeit, Endzeit\}$

Danach werden beim Startbahnhof alle Zeiten, welche vor der Endzeit am Zielbahnhof liegen gelöscht. Ebenfalls werden alle Zeiten beim Zwischenbahnhof, welche sich zwischen der Startzeit und Endzeit befinden gelöscht. Die Endzeit wird nicht gelöscht, da diese beim nächsten Durchlauf als Startzeit fungiert und damit wiederum zu einem Eintrag unter den oben genannten Bedingungen führen kann. Nach diesen Löschvorgängen beginnt das ganze wieder bei dem Vergleich der ersten Zeitmarken vom Startbahnhof und vom Zielbahnhof. Dies wird solange wiederholt, bis die Liste der Zeitmarken bei irgendeinem Bahnhof leer ist. Dieser Vorgang wird dann für sämtliche IDs wiederholt sodass

zu jeder Richtung sämtliche IDs mit den Zeitmarken gefunden wurden. Sollte eine ID mehrmals in die gleiche Richtung gefahren sein wird einfach ein Eintrag zur jeweiligen ID hinzugefügt. $ID = X \{Starzeit1, Zwischenzeit1, Endzeit1, \dots, StartzeitN, ZwischenzeitN, EndzeitN\}$

Abbildung 17 veranschaulicht in grafischer Form das Vorgehen, um die jeweiligen Personen zu finden, welche sich von einem Startbahnhof über einen Zwischenbahnhof zu einem Endbahnhof bewegt haben.

Das Ergebnis wird danach noch mittels JFreeChart visualisiert. Dazu wird ein Gantt Diagramm verwendet. Aus Gründen der Übersichtlichkeit wird für jede Richtung ein eigenes Diagramm erstellt. Normalerweise wird diese Form der Diagramme im Projektmanagement eingesetzt um einzelne Aktivitäten zeitlich darstellen zu können. Dabei entspricht der Startzeitpunkt im Diagramm der Startzeit am Bahnhof und der Endzeitpunkt der Endzeit am Bahnhof. Weiters kann bei diesem Diagrammtyp eingetragen werden, wie weit eine Aktivität bereits abgeschlossen ist. Diese Funktion wird dazu verwendet, um jeweilige Reisezeit von Startbahnhof bis zum Zwischenbahnhof darstellen zu können. Dazu werden zwei Unteraktivitäten erstellt, wobei für die erste Unteraktivität der Startzeitpunkt der Zeit am Startbahnhof entspricht und die Endzeit der Zeit am Zwischenbahnhof. Für die zweite Unteraktivität entspricht der Startzeitpunkt der Zeit am Zwischenbahnhof und die Endzeit der Zeit am Endbahnhof. Danach wird die erste Unteraktivität als abgeschlossen eingetragen und eingefärbt. Ziel dabei sollte es sein eventuell auftretende Verzögerung einer Teilstrecke zuordnen zu können.

6.3.2 Route Usage

Die nächste Anwendung stellt eine Erweiterung der ersten Anwendung dar. Es sollte hierbei möglich sein, den Unterschied zwischen der Benützung zweier Strecken, welche den gleichen Anfangs- beziehungsweise Endpunkt aufweisen, zu veranschaulichen. Es werden dabei anstelle nur eines Wegpunktes (Bahnhof) mehrere Wegpunkte, welche beim Einlesen der Strecke als Set definiert wurden, für die Auswertung verwendet. Diese Implementierung ermöglicht eine flexible Selektierung der Strecke in ein Anfangs-, Zwischen- und Endset. Anfang und Ende kann dabei wieder als formale Definition gesehen werden da dies für jede ID variiert. Die Funktionsweise unterscheidet sich im Wesentlichen nur daran, dass ein größerer geografischer Bereich für jeden der drei möglichen Wegpunkte (Set von Wegpunkten) vom Benutzer definiert werden kann. Für diese drei Wegpunkte werden danach wieder die gleichen Speicherstrukturen erstellt, damit am Schluss zu jeder ID wieder die Zeitmarken für die unterschiedlichen Sets aufscheinen. Daraus wird danach die Richtung bestimmt, in welche sich die jeweilige ID bewegt hat. Für die Visualisierung der Ergebnisse ist ein Gantt Diagramm nicht geeignet, da einfach zu viele IDs über einen Tag gesehen die markierten Strecken durchfahren und somit die Übersichtlichkeit nicht mehr gewährleistet ist. Aus diesem Grund wurde zur grafischen Darstellung ein Balkendiagramm für jede der beiden Richtungen, welche jeweils die Startzeiten und die Ankunftszeiten stundenweise aufsummieren, erstellt. Dazu werden im Gegensatz zur ersten Anwendung die verschiedenen Zeiten nicht zu einer ID hinzugefügt sondern es wird zur jeweiligen Stunde und Richtung in die sich die ID bewegt ein Eintrag hinzuaddiert.

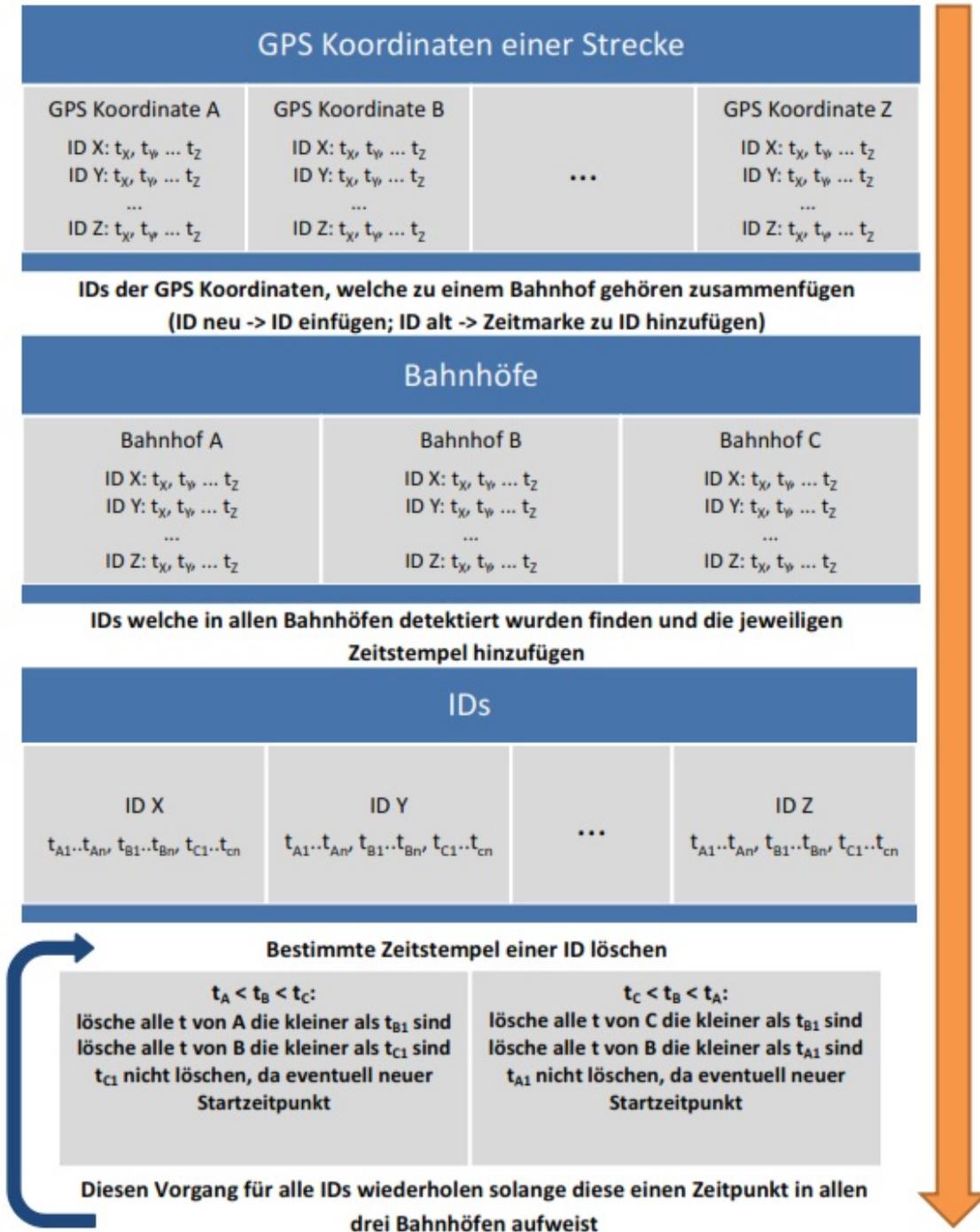


Abbildung 17: Grafische Darstellung des Ablaufes zur Auffindung der IDs

Dies hat zwar folgenden Nachteil, dass dabei die Zuordnung zu einer ID verloren geht, aber das ganze zur Veranschaulichung der unterschiedlichen Streckenauslastung ausreicht. Diese Summen der einzelnen Stunden unterschiedlicher Strecken können danach miteinander verglichen werden und ein Verhältniss daraus berechnet werden. Somit wird ein Ergebnis erzielt, welches zeigt wieviele Personen die jeweilige Strecke im Vergleich zu einer Anderen benützen.

6.3.3 Traffic Measurement Count

Diese Anwendung verwendet gewisse Methoden, welche auch von den ersten beiden zur Analyse benutzt werden. Zu Beginn wählt der Benutzer ebenfalls die gewünschte Strecke und den gewünschten Zeitraum aus. Danach wird die ausgewählte Strecke auf einer Landkarte dargestellt, damit der Benutzer die Position der Verkehrsmessstelle auf der Karte hinzufügen kann. Diese Position muss lediglich zwischen zwei Wegpunkten von der eingelesenen Strecke liegen. Diese zwei Wegpunkte und deren Nummer werden danach von dem Programm ermittelt. Diese Nummern werden darum benötigt, da auf Basis von diesen die jeweiligen IDs getrennt werden. Wegpunkte, welche eine niedrigere Nummer als die kleinere Nummer der beiden ermittelten aufweisen werden zu einem Set hinzugefügt und Wegpunkte mit einer höheren Nummer als die größere ermittelte Nummer zu dem anderen Set hinzugefügt. Dadurch ergeben sich zwei Sets von Wegpunkten, welche miteinander verglichen werden können. Dazu werden wieder die jeweils die ersten Zeitmarken einer ID in den beiden Sets bestimmt um damit die Richtung bestimmen zu können. Das weitere Vorgehen zur Bestimmung von Mehrfachaufkommen einer ID ist analog zu der Methode, welche schon bei den beiden vorher beschriebenen Anwendungen verwendet wurde. Zur Darstellung der Ergebnisse wird wiederum ein Balkendiagramm verwendet. Es wird hierbei nur der Zeitpunkt, wo die ID im Endset detektiert wurde berücksichtigt. Weiters kann der Benutzer einen zeitlichen Filter benutzen, in welchem Zeitrahmen sich eine ID vom Startset ins Endset bewegt haben muss. Wird dieser nicht benutzt, werde sämtliche Vorkommen einer ID berücksichtigt. Da heißt sollte eine ID um beispielsweise sieben Uhr im Startset einen Event auslösen und dann erst wieder um siebzehn Uhr im Endset würde dies zu einem Eintrag zum siebzehn Uhr führen. Nach Vorliegen dieser Ergebnisse wurde noch ein Korrelationsfaktor zwischen den realen Werten eine Verkehrsmessstelle und den ermittelten Werten aus der Anwendung bestimmt. Der Korrelationsfaktor [21] wird folgendermaßen berechnet:

$$Kor(x, y) = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

wobei

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i \\ \bar{y} &= \frac{1}{n} \sum_{i=1}^n y_i \end{aligned} \tag{5}$$

6.3.4 Speed Control

Bei der letzten Anwendung wird ein Geschwindigkeitsprofil über eine eingelesene Strecke erstellt. Hierbei werden wiederum die gewünschte Strecke und der Zeitrahmen ausgewählt. Im Gegensatz zu den drei vorher implementierten Anwendungen wird dabei der Typ IDs mit GPS Koordinaten zur Analyse verwendet. Aufgrund der sehr großen Anzahl von IDs werden zum einlesen der Files mehrere Threads verwendet, wobei ein Thread immer für eine bestimmte Anzahl von IDs pro Stunde verantwortlich ist. Dazu wird auf die bereits beim Einlesen der Strecke erstellten Matrizen zurückgegriffen. Es werden die jeweiligen Geschwindigkeiten einer ID unter Berücksichtigung der Richtung von den jeweiligen Thread abgespeichert. Diese Daten von den einzelnen Threads werden danach zusammengefügt. Daraus resultieren zwei Datenstrukturen, welche sämtliche Geschwindigkeiten beinhalten. Diese Geschwindigkeiten können bei Bedarf noch nach deren Distanz gefiltert werden. Die Implementierung eines solchen Filters wurde deswegen durchgeführt um zu geringe Distanzen zu Vermeiden. Abbildung 18 zeigt diese Problematik, dass bei zu geringen Distanzen die detektierten Geschwindigkeiten zu sehr streuen. Zur Re-

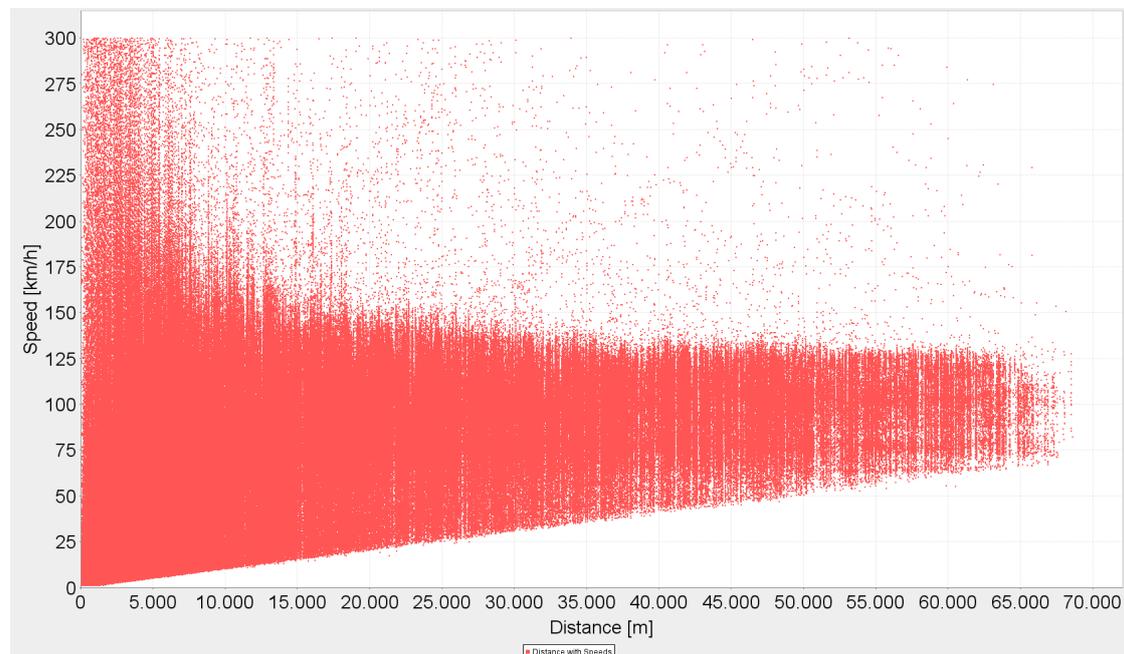


Abbildung 18: Distanz vs. Geschwindigkeit

präsentation der Ergebnisse werden die Geschwindigkeiten zu einem Stundenweise in Geschwindigkeitsbereiche unterteilt. Diese Geschwindigkeitsbereiche können individuell angepasst werden. Wird eine Geschwindigkeit einem dieser Bereiche zugeordnet, wird der Zähler des jeweiligen Bereiches um eins erhöht. Aus diesen einzelnen Zählern wird danach ein Balkendiagramm erstellt. Weiters wird ein Boxplot sämtlicher ungefilterten oder gefilterten Geschwindigkeiten über einen Tag erstellt um daraus das Geschwindig-

keitsprofil für einen Tag erkenne zu können.

Bei all diesen Abläufen kommen die in 5.1 vorgestellten Speicherstrukturen zum Einsatz, um eine möglichst geringe Programmlaufzeit erzielen zu können.

7 Ergebnisse, Validierung und Auswertung

In diesem Kapitel werden die Ergebnisse der oben vorgestellten Anwendungen aufgearbeitet und mittels Ergebnissen aus anderen Quellen validiert. Es wurde ein Zeitraum von vierzehn Tagen, bestehend aus zwei jeweils zusammenhängenden Woche analysiert. Bei der Auswahl dieser beiden Wochen ist darauf Rücksicht genommen worden, dass es sich dabei um so genannten Normwochen handelt, welche weder einen Feiertag noch Ferien beinhalten. Als erste Woche wurde die Kalenderwoche 19 vom 7.5.2012 bis zum 13.5.2012 und als zweite Woche die KW 24 vom 11.6.2012 bis zum 17.6.2012 ausgewählt. Bei einem Tag, nämlich dem 15.06.2012, sind die bezogenen Daten vom iTraffic System nicht über den ganzen Tag vorhanden. Es fehlen die Daten von vier bis acht Uhr. Weiters liefert ein weiterer Tag, nämlich der 7.5.2012, um 13 Uhr keine korrekten Ergebnisse. Dieses Faktum, führt leider dazu, dass einige Ergebnisse verzerrt dargestellt werden. Zur Hochrechnung der jeweiligen Ergebnisse wird der Marktanteil der A1 Telekom Austria mit 41 Prozent beziffert. Diese Zahl kann aus dem Bericht für das dritte Quartal 2011 der Rundfunk und Telekom Regulierungs-GmbH (RTR-GmbH)[28] entnommen werden.

7.1 Train Validation

Bei dieser Anwendung wurde die Strecke Graz-Salzburg für den oben genannten Zeitraum betrachtet. Das Ergebnis für einen Tag ist in folgender Abbildung 19 für die Strecke Graz nach Salzburg ersichtlich. Analog zu diesem Diagramm wird auch noch ein Diagramm für die entgegengesetzte Richtung, also von Salzburg nach Graz erstellt. Der blaue Balken bildet die jeweilige Fahrzeit einer ID ab. Der grüne Unterbalken beziffert die Fahrtzeit zum jeweils ausgewählten Zwischenbahnhof. Mit Hilfe dieses Diagrammes ist es möglich, die gefundenen IDs jeweils einem Zug unter zu Hilfenahme des Fahrplanes zuzuordnen, solange die Fahrtzeit nicht zu sehr von der normalen Fahrtzeit abweicht. Weicht diese zu sehr von der erwarteten Fahrtzeit ab, kann dies daran liegen, dass die Startzeit nicht der realen Abfahrtszeit entspricht. Diese Möglichkeit ergibt sich dadurch, da die Zeitmarken nur den jeweils zuletzt ausgelösten Event am Startset repräsentieren. Darum kann es durchaus passieren, dass eine ID schon vorher einmal in dem jeweiligen Set detektiert wurde und danach bei der Abfahrt keinen neuerlichen Event ausgelöst hat. Ein ähnliches Problem ergibt sich bei der Bestimmung der Zeitmarke für das Zielset. Es ist durchaus möglich, dass eine ID im Zielset erst später einen Event auslöst, als dies laut Fahrplan der Fall sein sollte. Zur genaueren Analyse dieser Problematik wäre es möglich diese IDs, deren Fahrzeit nicht der gewünschten Fahrtzeit entspricht, genauer zu betrachten. Dazu könnte einfach die ID herangezogen werden und sämtliche Events dieser ID über den gesamten Tag betrachtet werden. Daraus sollte es

7 Ergebnisse, Validierung und Auswertung

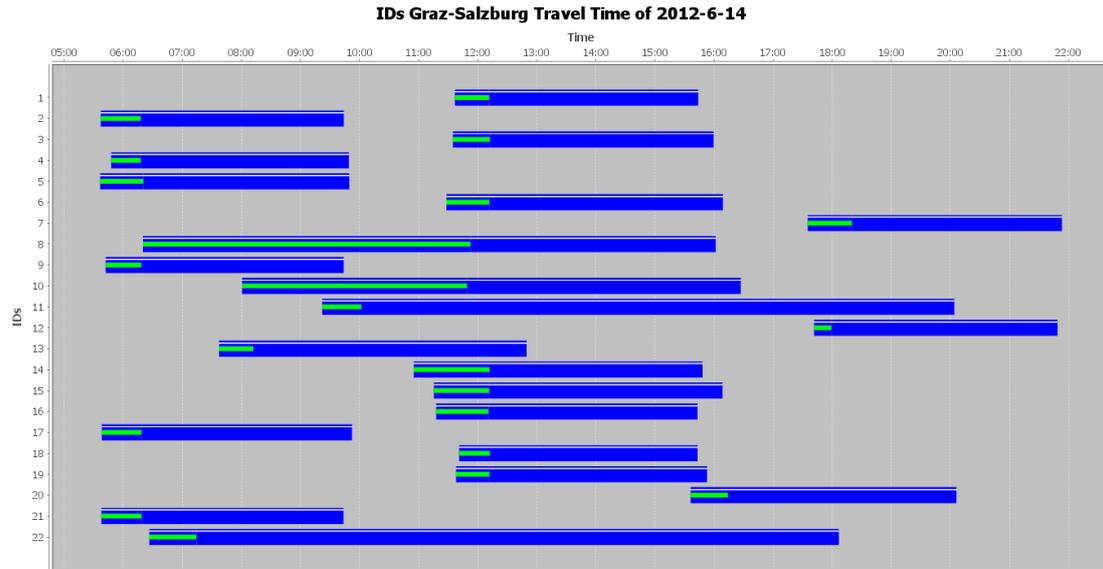


Abbildung 19: Gefundene IDs von Graz nach Salzburg

möglich sein die tatsächliche Fahrzeit näher bestimmen zu können, unter der Voraussetzung, dass es überhaupt noch weitere Events dieser ID gibt, welche zur Aufklärung dieses Problems herangezogen werden können. Abbildung 20 veranschaulicht exemplarisch so eine mögliche Visualisierung einer ID für die Strecke Graz-Salzburg. Da das primäre Ziel dieser Anwendung jedoch nicht die Zuordnung einer ID zu einem jeweiligen Zug ist, sondern lediglich die Anzahl der gefundenen IDs von Bedeutung ist, wird dieser Aspekt derzeit nicht berücksichtigt und die überdurchschnittlich lange Fahrzeit nicht weiter betrachtet.

Für den betrachteten Zeitraum ergeben sich folgende in Tabelle 3 ersichtlichen Werte für die Anzahl von IDs, welche an den jeweiligen Tagen die markierte Strecke durchfahren haben. Die Anzahl der IDs variiert dabei von Tag zu Tag, wobei ersichtlich ist, dass der Mittelwert von beiden Wochen annähernd gleich ist. Auffallend ist weiters, dass mehr IDs von Graz nach Salzburg unterwegs sind als umgekehrt. Dies ist bei jedem der vierzehn analysierten Tage der Fall. Abbildung 21 zeigt die Prozentuelle Aufteilung der Zuordnung der gefundenen IDs zu einem jeweiligen Zug. Es zeigt sich, dass es durchaus möglich ist, einen Großteil der detektierten IDs einem Zug zuzuordnen zu können. Lediglich bei 12% von der Gesamtanzahl der IDs konnte keine eindeutige Zuordnung zu einem Zug durchgeführt werden. Die nachfolgenden Abbildung 22 zeigt diese Zuordnung zu den jeweiligen Zügen, welche laut Fahrplan der ÖBB[26] auf dieser Strecke unterwegs sind. Die Validierung erfolgte auf Basis der Studie „Verkehrssimulation zwischen Graz und Salzburg“[35], welche von der Technischen Universität Graz durchgeführt wurde. Laut dieser fahren an einem durchschnittlichen Werktag 34 Fahrgäste mit dem Zug vom Hauptbahnhof Graz zum Hauptbahnhof Salzburg und 30 Fahrgäste in die Gegenrichtung. Der Wert von Salzburg nach Graz entspricht nahezu dem Wert, welcher auch von

7 Ergebnisse, Validierung und Auswertung

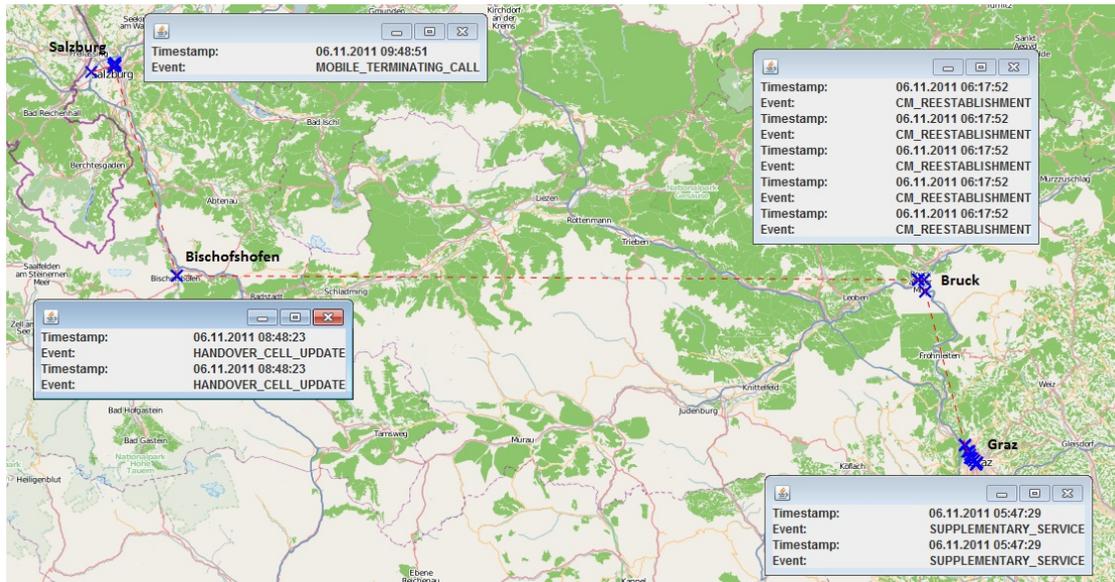


Abbildung 20: ID über die Strecke Graz-Salzburg

Tag	Richtung		Hochrechnung(41%)	
	Graz	Salzburg	Graz	Salzburg
07.05.2012	15	18	37	44
08.05.2012	7	19	18	47
09.05.2012	12	13	30	32
10.05.2012	9	18	22	44
11.05.2012	19	28	47	69
12.05.2012	13	18	32	44
13.05.2012	16	17	40	42
Mittelwert	13	19	33	46
11.06.2012	12	17	30	42
12.06.2012	7	14	18	35
13.06.2012	7	19	18	47
14.06.2012	22	24	54	59
15.06.2012	7	20	18	49
16.06.2012	8	17	20	42
17.06.2012	17	20	42	49
Mittelwert	12	19	29	47

Tabelle 3: Anzahl der gefundenen IDs für die Strecke Graz-Salzburg

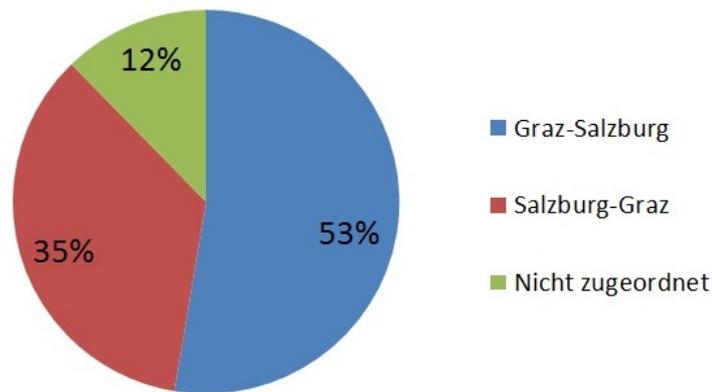


Abbildung 21: Prozentuelle Aufteilung der gefundenen IDs über 14 Tage

Verkehrssimulation erreicht wird. In die Gegenrichtung weicht das Ergebnis etwas stärker ab.

7.2 Route Usage

Hierbei wurde die Strecke von Graz nach Wien und umgekehrt für den gegebenen Zeitrahmen analysiert. Die genannte Strecke wurde aus diesem Grund gewählt, da sich dabei die Frage stellt, wie viele Personen fahren von Graz nach Wien und umgekehrt über die A2 beziehungsweise die S6 (Semmering). Um diese Frage zu klären wurde die Strecke in jeweils drei Sektoren unterteilt. Der erste Sektor bildet den Raum Graz ab. Dazu wurde ein Kreis mit einem Radius von 5 Kilometer aufgespannt. Den zweiten Sektor bilden zum einen eine Teilstrecke von der A2 und zum anderen eine Teilstrecke von der S6. Der dritte Sektor bildet die Strecke nach dem Zusammentreffen der A2 und der S6 ab. Abbildung 23 veranschaulicht diese geografische Aufteilung auf der Landkarte. In dieser Abbildung ist der Sektor, welcher den Raum Graz abbildet rot markiert. Blau stellt jeweils das Zwischenset dar und gelb den Sektor, in welchem die A2 und die S6 zusammentreffen.

Diese drei Sektoren bilden die jeweiligen Sets, mit welchen danach die Auswertung durchgeführt wurde. Die Zahl der detektierten IDs wurde dabei für einen ganzen Tag zusammengefasst und mit der jeweils anderen Zahl der gefundenen IDs für die zweite Strecke verglichen. Der Faktor wurde durch eine Division dieser Werte berechnet. Dabei wurden immer die gefundenen IDs über einen ganzen Tag von der A2 durch die Anzahl der gefundenen IDs von der S6 durchdividiert. Danach wurde daraus der Mittelwert von den Wochentagen und von den Wochenenden der beiden Wochen gebildet. Das Ergebnis zeigt, dass im Mittel auf eine Person, welche über den Semmering Richtung Wien am Wochentag fährt, 1.8 und am Wochenende 1.5 Personen über die A2 fahren. Richtung Graz kommen auf 1 Person die über den Semmering fährt 2.4 Personen am Wochentag und 2 Personen am Wochenende, welche über die A2 fahren. Referenzergebnisse aus dem

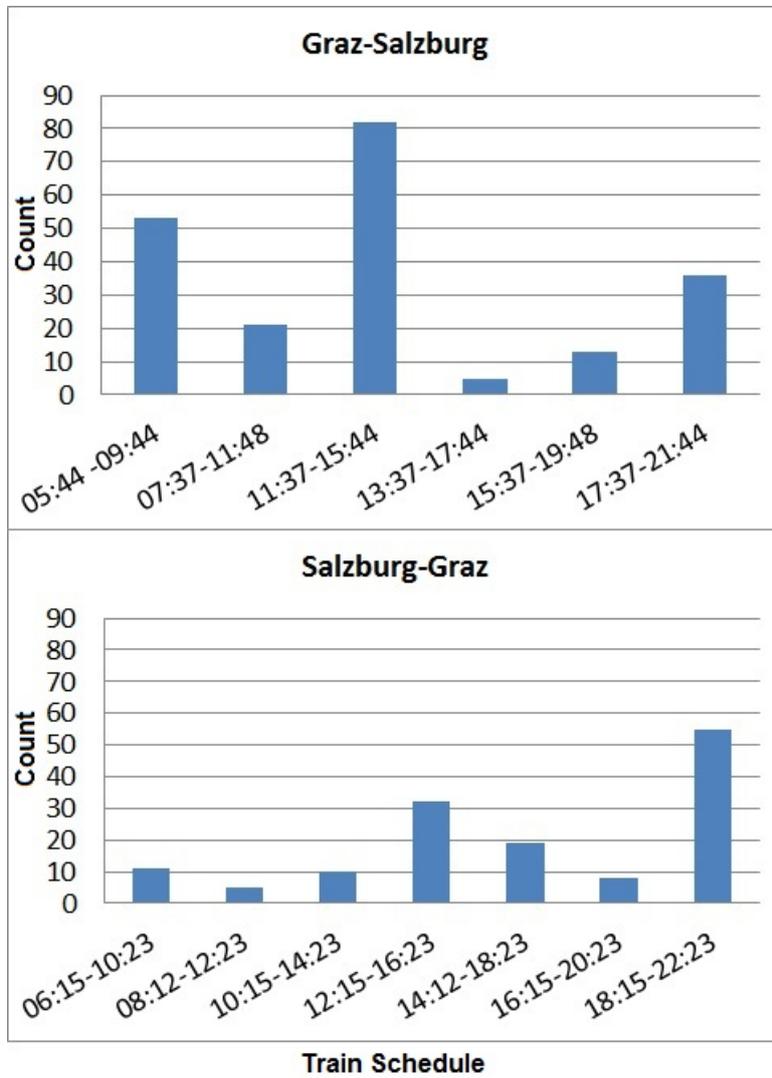


Abbildung 22: Zuordnung der IDs zu den jeweiligen Zügen

7 Ergebnisse, Validierung und Auswertung

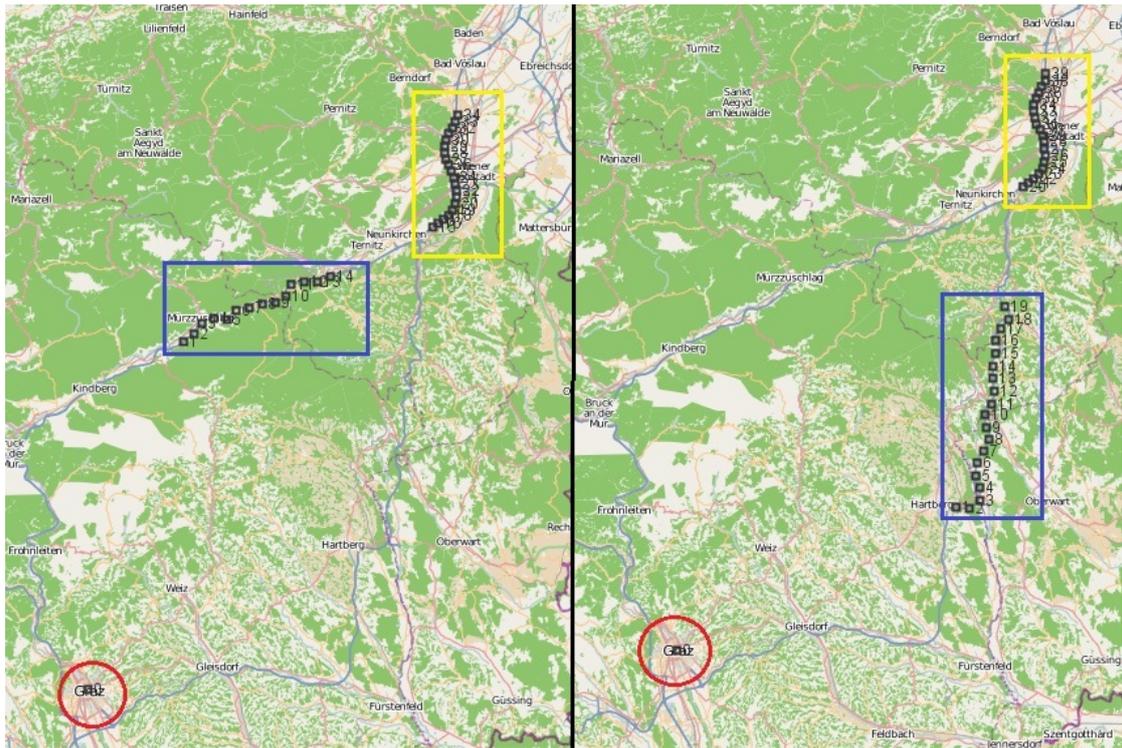


Abbildung 23: Sektoren [Links über S6 (Semmering), Rechts über A2]

Verkehrsmodell Österreich [12] zeigen, dass es an einem durchschnittlichen Werktag 30% der IV-Fahrten vom Verkehrsbezirk der Stadt Graz in den Verkehrsbezirk der Stadt Wien und zurück über den Semmering (S6 und S35) und 70% der Fahrten über den Wechsel (A2) führen. Diese Ergebnisse sind mit einer Schwankungsbreite von +/- 5% behaftet. Demzufolge kommen auf 1 Person über den Semmering 2.3 Personen, welche über die A2 fahren. Dieses Ergebnis entspricht jenem Ergebnis, welches die Anwendung von Wien nach Graz erzielt. In die Gegenrichtung ist das erzielte Ergebnis ein wenig geringer.

7.3 Traffic Measurement Count

Bei dieser Anwendung wird die Autobahn A12 von Kufstein bis Innsbruck und umgekehrt für die Analyse herangezogen. Dazu wurde zunächst diese Strecke markiert und danach über den gewünschten Zeitraum eingelesen. Danach wurden auf dieser Strecke die jeweiligen Verkehrsmessstellen eingezeichnet und die Anwendung gestartet. Es gab dabei vier Verkehrsmessstellen (Kufstein, Kundl, Schwaz, Ampass), von denen Validierungsdaten [10] zur Verfügung gestanden haben. Diese Daten wurden mit den Ergebnissen von der Anwendung verglichen. Es zeigte sich, dass es nicht möglich ist den Wert, welcher von der Verkehrsmessstelle geliefert wird mit den Mobilfunkdaten zu erreichen. Dies hat vor allem diesen Grund, dass die Zuordnung zu einer Straße mittels der Mobilfunkdaten nicht erreicht werden kann. Um diese erzielen zu können, müsste eine wesentlich bessere geografische Auflösung vorhanden sein. Dies wirkt sich vor allem auf diese Strecke aus, da parallel zu der Autobahn (A12) die Bundesstraße und auch noch die Eisenbahnlinie verläuft. Aus diesem Grund werden zu viele IDs gefunden, welche über einen gewünschten Wegpunkt gefahren sind, da sowohl die IDs die sich auf der Bundesstraße bewegen und auch die IDs, welche mittels der Bahn unterwegs sind in das Ergebnis mit einfließen. Abbildung 24 veranschaulicht diese Problematik wobei bei den Ergebnissen der Anwendung der Marktanteil nicht berücksichtigt wurde.

Ein weiteres großes Problem stellt die Lage der Verkehrsmessstelle dar. Befindet sich diese am Anfang oder auch am Ende der markierten Strecke werden nur von der Anwendung nur sehr wenige IDs gefunden, welche sich über diese Messstelle bewegt haben. Dies hat diesen Grund, da dadurch das eine Set sehr klein ist und darin nur sehr wenige Einträge gefunden werden, welche mit den Einträgen des zweiten Sets verglichen werden können. Dies wirkt sich vor allem auf die Messstelle, welche sich in der Nähe von Kufstein befindet aus, da sich diese zwischen den ersten und den zweiten Wegpunkt befindet und somit befinden sich in dem einen Set nur ein Wegpunkt zu welchen Einträge gefunden werden können. Abbildung 25 zeigt, dass für diesen Messpunkt wesentlich weniger IDs gefunden werden, als dies bei den übrigen Messstellen der Fall ist.

Trotz dieser Probleme ist nach Auswertung sämtlicher Tage ersichtlich, dass die Ergebnisse von den Verkehrsmessstellen und der Anwendung einen ähnlichen Verlauf aufweisen. Aus diesem Grund wurde ein Korrelationsfaktor siehe Formel 5 berechnet, welcher diese Gemeinsamkeit widerspiegeln soll. In nachstehende Tabelle 4 sind sämtliche Ergebnisse für die Verkehrsmessstelle Kundl für die untersuchten Tage und die unterschiedlichen Richtungen aufgelistet. Daraus ist ersichtlich, dass eine hohe Korrelation zwischen den beiden Ergebnissen in Bezug auf den Werteverlauf über einen Tag gese-

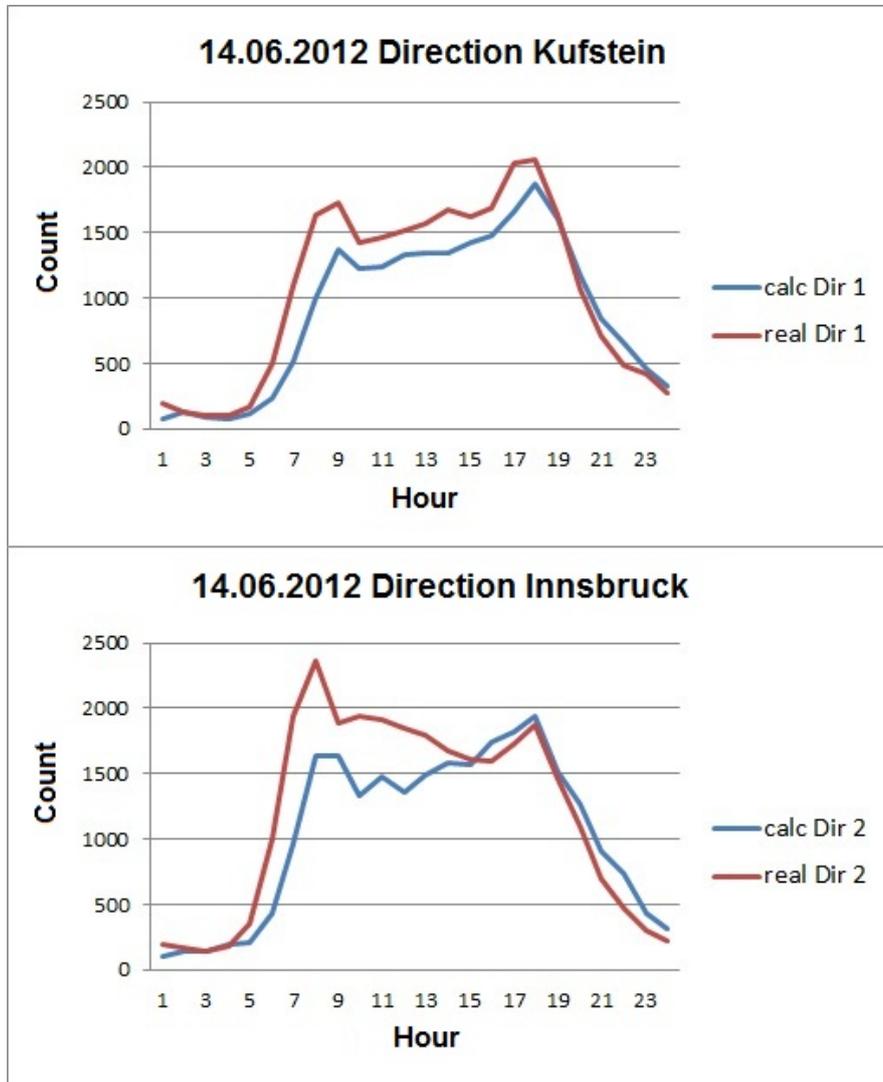


Abbildung 24: Vergleich Verkehrsmessstelle mit Ergebnissen aus der Anwendung

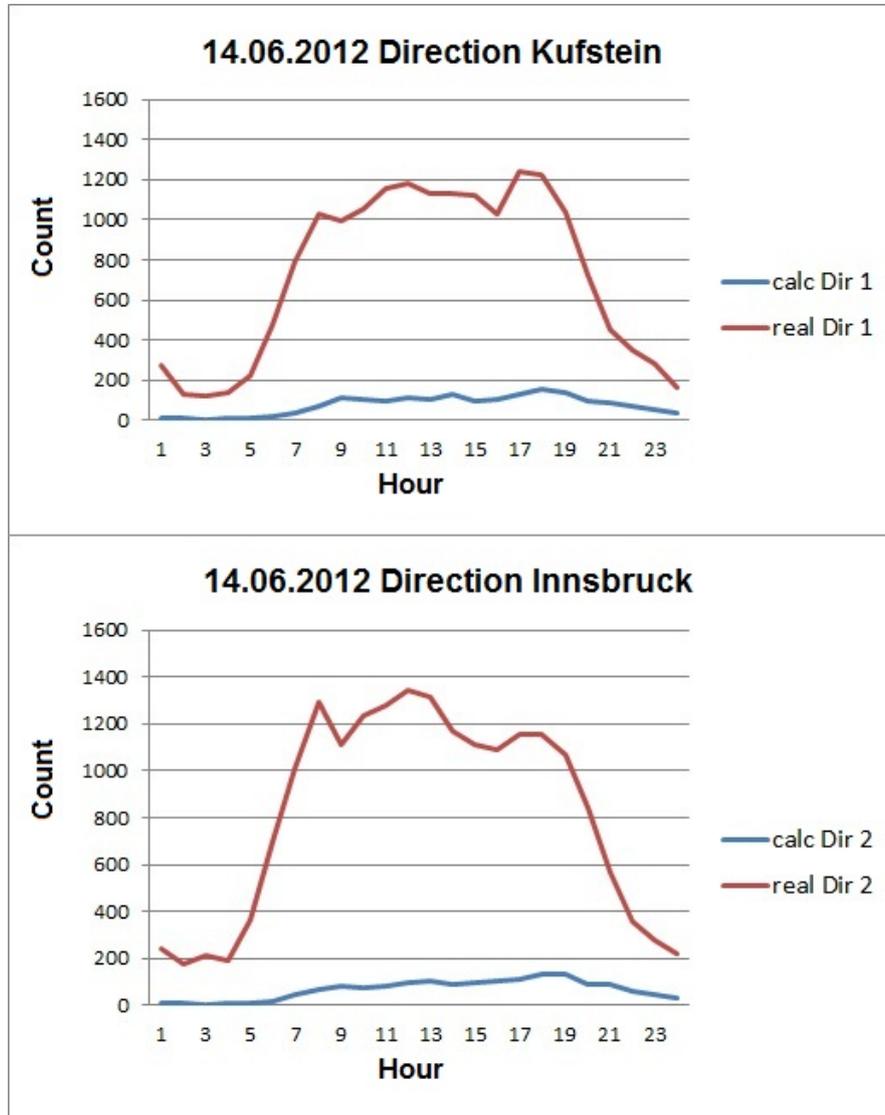


Abbildung 25: Ergebniss bei ungünstiger Lage der Verkehrsmessstelle

Tag		Korrelation in Richtung	
		Kufstein	Innsbruck
Montag	07.05.2012	0,823595723	0,843818491
Dienstag	08.05.2012	0,821209423	0,881888323
Mittwoch	09.05.2012	0,881018708	0,870771796
Donnerstag	10.05.2012	0,928905818	0,883727032
Freitag	11.05.2012	0,935621674	0,926996858
Samstag	12.05.2012	0,934482913	0,878396842
Sonntag	13.05.2012	0,736972666	0,500285489
Montag	11.06.2012	0,963239572	0,98250848
Dienstag	12.06.2012	0,845060767	0,816274622
Mittwoch	13.06.2012	0,831768623	0,883058773
Donnerstag	14.06.2012	0,94077601	0,884610267
Freitag	15.06.2012	0,793642671	0,71038714
Samstag	16.06.2012	0,928805256	0,933898155
Sonntag	17.06.2012	0,824058689	0,777543047

Tabelle 4: Korrelationsfaktor für die Messstelle Kundl über den gesamten Zeitraum

hen zu finden ist. Ausgenommen dabei sind der 15.06.2012 und die jeweiligen Sonntage. Die schlechte Korrelation für den 15.06.2012 lässt sich dadurch erklären, dass wie weiter oben bereits erwähnt nicht alle Daten für den ganzen Tag vorhanden sind. Für jede untersuchte Zählstelle weist der Sonntag die geringste Korrelation auf, da sich hierbei die jeweiligen Spitzenwerte sehr unterscheiden. Ein möglicher Grund könnte darin liegen, dass der Sonntag vermehrt für Familienausflüge genutzt wird und das Mobiltelefon nicht sehr häufig benutzt wird und somit auch die auftretenden Events weniger werden und diese erst am Abend wieder ansteigen und somit auch wieder mehr IDs detektiert werden. Um eine Aussage über die Signifikanz der Korrelation treffen zu können, wurde ein Signifikanzwert mittels Matlab berechnet [24]. Als schlechtester Wert kam dabei der Wert 0.013 (1.3%) für den 13.05.2012 in Richtung Innsbruck heraus. Dieser Wert spiegelt die Wahrscheinlichkeit wieder das eine Korrelation zufällig entstanden ist. Sämtliche andere Wahrscheinlichkeiten liegen unter 0.02%.

7.4 Speed Control

Zur Betrachtung der Geschwindigkeiten wurde wieder die gleiche Strecke von Kufstein nach Innsbruck und umgekehrt für die Auswertung herangezogen. Es wurde dabei für jede Stunde und für jede Richtung ein Balkendiagramm erstellt, welches fünf Geschwindigkeitsbereiche aufweist. Weiters wurde für jeden Tag ein Boxplot erstellt und die gesammelten Geschwindigkeiten darin eingetragen. Der Boxplot, siehe Abbildung 26, dient hauptsächlich dazu um einen Geschwindigkeitsbereich darstellen zu können, in welchen sich die unterschiedlichen IDs bewegen. Die Ergebnisse für den ausgewählten Zeitraum

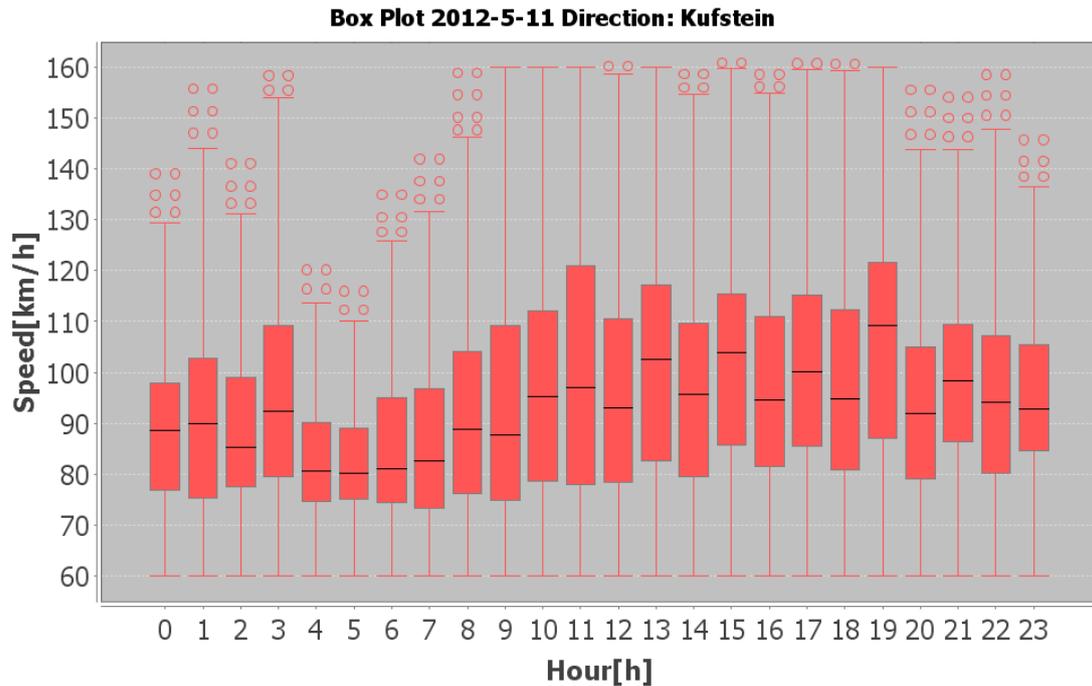


Abbildung 26: Box Plot für einen Tag Richtung Kufstein

zeigen, dass dieser Geschwindigkeitsbereich sehr variabel ist wobei sich der Median zwischen einer Geschwindigkeit von 75 - 100 km/h bewegt und sich mehr als 60% der detektierten Geschwindigkeiten in dem Bereich von 60 - 160 km/h befinden. Der Vergleich zwischen den Ergebnissen der einzelnen Geschwindigkeitsbereiche pro Stunde mit den Verkehrsmeldungen [33] für den betrachteten Zeitraum brachte kein zufriedenstellendes Ergebnis. Selbst die Meldung eines Staus führte zu keiner Veränderung beziehungsweise Verschiebung der Geschwindigkeiten in den jeweiligen geringeren Geschwindigkeitsbereichen der betroffenen Stunde.

Das große Problem bei dieser Anwendung liegt darin, dass wieder IDs in das Ergebnis miteinfließen, welche sich eigentlich gar nicht auf dieser Strecke befinden. Somit kann keine klare Aussage darüber getroffen werden, ob auf der zu analysierenden Strecke eine Störung aufgetreten ist oder nicht. Was sich jedoch zeigte ist, dass ein verhältnismäßiger Geschwindigkeitsanstieg während des Tages auftritt, welcher in Abbildung 27 ersichtlich ist. Bei einigen Tagen kam es bei manchen Stunden jedoch wieder zu einer Verringerung der Geschwindigkeit, welche jedoch keinen bekannten Störungen zugeordnet werden konnte.

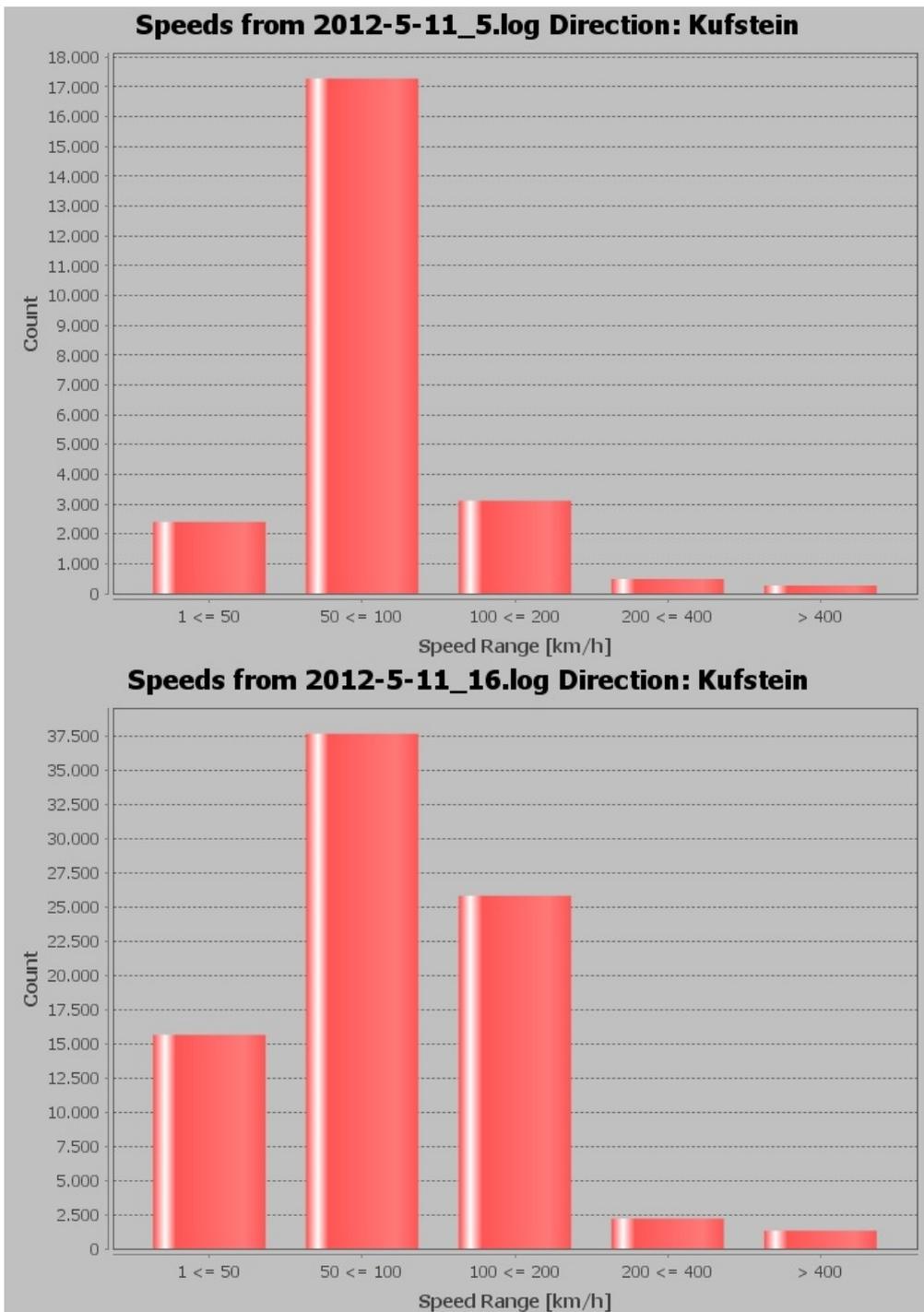


Abbildung 27: Verhältnismäßiger Geschwindigkeitsanstieg

8 Diskussion

Das große Problem bei dem Umgang mit den Daten lag nicht in den verwendeten Methoden, da sich diese meistens auf Einfügeoperation in bestehende Datenstrukturen beschränkten, sondern in den Überlegungen selbst, wie die zur Verfügung stehenden Daten eingesetzt werden konnten, um sinnvolle Ergebnisse zu erhalten. Das heißt, dass ein Großteil der Zeit damit verbracht wurde, den Inhalt der zur Verfügung stehenden Daten zu verstehen und diese richtig deuten zu können um diese möglichst zielführend einsetzen zu können. Vor allem werden durch den Aufbau der Mobilfunknetze die Anwendungsmöglichkeiten der darin generierten Daten in bestimmten Bereichen sehr eingeschränkt, da es einfach an der benötigten geografischen Genauigkeit fehlt und passive Benutzer nicht oft detektiert werden. Es stellt sich dabei heraus, dass es in den meisten Fällen am sinnvollsten war, diese Überlegungen so einfach wie möglich zu halten. Je mehr Einschränkungen (vor allem im zeitlichen Bereich) getroffen wurden, desto schlechtere Ergebnisse kamen zustande. Dies hatte auch einen weiteren Vorteil, da sich damit die Laufzeit der Implementierungen verkürzt hat, wodurch wesentlich mehr Ergebnisse in kürzerer Zeit erzielt werden konnten und können.

Die Ergebnisse dieser Arbeit zeigen, dass der Einsatz von Mobilfunkdaten bei der Verkehrsplanung durchaus Potenzial zugesprochen werden kann. Es sind jedoch einige Abstriche zu machen, was die dabei erreichbare geografische Genauigkeit betrifft. So wird es nicht möglich sein, eine genaue Zuordnung zu sämtlichen Straßen durchführen zu können. Dieses Problem tritt vor allem auf diesen Abschnitten auf, wo parallel zu einer Strecke weitere Straßen oder Bahnstrecken verlaufen. Diese Ungenauigkeit führt auch dazu, dass Analysen im urbanen Bereich mit den bei dieser Arbeit zur Verfügung stehenden Daten nicht durchgeführt werden können. Ein weiteres Problem stellt die Einschränkung dar, dass nur die Daten von einem Mobilfunkanbieter zur Verfügung gestanden haben, wodurch sich Unregelmäßigkeiten bei den Ergebnissen ergeben können. Ebenfalls kann anhand der Mobilfunkdaten nicht ausgeschlossen werden, dass mehrere Personen, welche sich in einem Fahrzeug aufhalten dementsprechende oft mitgezählt werden, oder das eine Person über mehrere SIM-Karten verfügt, wodurch es wiederum zu einer Verfälschung der Ergebnisse kommen kann. Die Anwendung, welche sich mit der Analyse der Strecke von Kufstein bis Innsbruck beschäftigt hat, veranschaulicht diese Problematik, dass es nicht möglich ist auf die realen Werte eines stationären Detektors zu gelangen. Das Ergebnis zeigt jedoch, dass es wiederum möglich ist, die jeweiligen Lastspitzen eines Tages zu bestimmen was auch aus den Ergebnissen von anderer Projekten[11][13] ersichtlich ist.

Für die Bestimmung örtlicher Störungen anhand der aus den Mobilfunkdaten generierten Geschwindigkeiten, ist der Ansatz sämtliche Geschwindigkeiten welche in einem geografischen Gebiet ermittelt werden können nicht der Richtige. Dazu wäre es sinnvoller wie in [22] beschrieben nur die aktiven Teilnehmer zur Analyse heranzuziehen. Es würde sich dadurch zwar die Anzahl der gefunden Geschwindigkeiten stark reduzieren aber es sollte trotzdem möglich sein aus diesen ermittelten Geschwindigkeiten eine Aussage darüber treffen zu können, ob irgendeine Störung des Verkehrsflusses aufgetreten ist oder nicht. Was hierbei jedoch wiederum berücksichtigt werden muss, ist die Möglichkeit, dass aus

den oben genannten Gründen, die ermittelten Geschwindigkeiten nicht auf der jeweiligen Strecke zustande gekommen sein könnten.

Die Anwendung die sich mit der Auslastung zweier Strecken, welche den selben Start-beziehungsweise Endpunkt aufweisen, beschäftigt hat zeigt das es durchaus Möglich ist diejenige Strecke herauszufinden, welche von den Straßenteilnehmern bevorzugt benutzt wird. Die erzielte Genauigkeit variiert von den jeweiligen Tagen, welche betrachtet werden und auch wie die einzelnen Sets positioniert werden. Das heißt je genauer und umfangreicher die einzelnen Sets eingezeichnet werden, desto bessere Ergebnisse können erreicht werden, da dadurch die Anzahl von Personen (IDs), welche nicht detektiert werden verringert wird. Die besten Ergebnisse dieser Arbeit zeigt die Anwendung, welche sich mit dem öffentlichen Verkehr im speziellen für den Personentransport im Eisenbahnwesen beschäftigt hat. Es ist dabei möglich einen Großteil der gefundenen Personen einem jeweiligen Zug zuzuordnen. Des Weiteren stimmte die Anzahl der aus den Mobilfunkdaten gefundenen Personen, welche sich über einen Tag gesehen von Graz nach Salzburg bewegt haben mit der Anzahl des Modells, welches derzeit zur Berechnung dieser Werte verwendet wird zumindest in eine Richtung überein. Das liegt vor allem daran, dass sehr viele Reisende, welche mit dem Zug unterwegs sind, während der Fahrt aktiv mit deren mobilen Gerät arbeiten und darum mehr Events, welche detektiert werden können, zustande kommen. Ebenfalls ist es hierbei durch den bekannten Streckenverlauf möglich die jeweiligen Personen mit sehr hoher Wahrscheinlichkeit einer Strecke zuzuordnen zu können und somit die Beeinflussung der Ergebnisse durch sich nicht auf dieser Strecke befindlicher Personen nahezu reduziert werden kann. Die verwendete Methode keinen zeitlichen Filter einzubauen, das heißt die Fahrdauer nicht einzuschränken, ermöglichte es auch unabhängig von zeitlichen Störungen Einträge einen Zug zuzuordnen.

Zu all den erreichten Ergebnissen muss gesagt werden, dass dabei lediglich ein Zeitraum von vierzehn Tagen betrachtet wurde und damit noch keine wirklich klaren Aussagen getroffen werden können. Weiters gehen bei den Ergebnissen sämtliche Personen verloren, welche sich um die Zeit bewegt haben, wo die jeweiligen pseudonymisierten IDs umgestellt werden. Da dies in dem Zeitraum zwischen 24 und 1 Uhr passiert kann dieser Zeitraum auch nicht mittels den Mobilfunkdaten nachgebildet werden beziehungsweise es können keine Einträge gefunden werden, bei denen der Startpunkt am Vortag des Endpunktes liegt. Darüber hinaus ist es mittels Mobilfunkdaten nicht möglich sämtliche Inhalte (Reisezweck, Soziodemografischen Daten etc.), welche zum Beispiel bei einer Befragung von Verkehrsteilnehmern erzielt werden können, nachzubilden. Trotz dieser Einschränkungen liefert diese Ausarbeitung Ergebnisse welche zeigen, dass es sich durchaus lohnt Mobilfunkdaten bei einigen Problemstellungen, welche bei der Verkehrsplanung auftreten einzusetzen und damit die Methoden, welche derzeit eingesetzt werden zu unterstützen. Vor allem kann mit deren Einsatz die Ergebnisse von Verkehrssystemen wie [7] verbessert werden, da mit Hilfe von Mobilfunkdaten weitere Informationen verfügbar gemacht werden, welche in diese Modelle mit einfließen können. Natürlich nur mit der Einschränkung, dass es im urbanen Bereich aufgrund der empirisch ermittelten geografischen Position zu erheblichen Ungenauigkeiten kommen kann.

9 Ausblick

Um bessere Aussagen über die jeweiligen Anwendungen treffen zu können, geht es daran, diese über einen längeren Zeitraum und über unterschiedliche Strecken Auswertungen durchführen zu lassen und die jeweiligen Methoden zu verfeinern. Vor allem sollte ein Verfahren entwickelt werden, welches es ermöglicht eine genauere Zuordnung zu einer jeweiligen Strecke realisieren zu können. Dazu müsste der Verlauf einer jeweiligen ID über einen längeren Zeitraum betrachtet werden um eine Rückschluss auf die gewählte Strecke erzielen zu können. Diese exakte Zuordnung stellt derzeit bei den bei dieser Arbeit zur Verfügung stehenden Daten die größte Herausforderung dar. Vor allem auf Kurzstrecken wird sich eine genaue Zuordnung nicht realisieren lassen, solange nicht mehr Information (Signalstärke etc.) in den bezogenen Daten zur Verfügung stehen. Weiters könnte mit Hilfe von geeigneten Methoden (z.B.: geeignete Speicherstrukturen, Sortierung anhand eines definierten Kriteriums) die Datenmenge so aufbereitet werden, dass es ermöglicht wird online Analysen durchzuführen. Diese gefilterten und aufbereiteten Daten könnten durch Abspeicherung in eine Datenbank für spätere Anwendungen zugänglich gemacht werden. Wie diese Arbeit zeigt können solange es nicht möglich ist eine genaue Zuordnung durchzuführen Analysen nur unter bestimmten Voraussetzungen, wie die häufige Verwendung eines Mobilfunkgeräts und die geeignete geografische Begebenheit einer Strecke, aussagekräftige Ergebnisse liefern, da ansonsten die verschiedensten Einflussfaktoren eine zu große Rolle spielen. Aber vor allem in der Kombination mit anderen Floating Car Data Quellen und einer jeweiligen Gewichtung der erreichten Ergebnisse sollte es möglich sein gute Modelle zu generieren, welche sich zur Unterstützung bei der Verkehrsplanung eignen. Das größte Potential für zukünftige Arbeiten liegt sicher bei der Analyse von Langstrecken, da es hierbei aufgrund der Begebenheiten des Mobilfunknetzes (Location Area Updates) möglich ist sämtliche Personen auf einer Strecke zu erfassen. Mit einigen kleinen Verbesserungen und mit dem Bezug von Mobilfunkdaten, welche nicht nur von einem Anbieter stammen, sollte es auf jedenfall möglich sein, die bei dieser Arbeit erzielten Ergebnisse noch verbessern zu können.

Literatur

- [1] 3gpp. Specification, . URL <http://www.3gpp.org/specifications>. 25.06.2012.
- [2] 3gpp. 04 Series, . URL <http://www.3gpp.org/ftp/Specs/html-info/04-series.htm>. 25.06.2012.
- [3] 3gpp. 05 Series, . URL <http://http://www.3gpp.org/ftp/Specs/html-info/05-series.htm>. 25.06.2012.
- [4] 3gpp. 48.001 BSS-MSC Interface, General aspects, . URL <http://www.3gpp.org/ftp/Specs/html-info/48001.htm>. 25.06.2012.
- [5] 3gpp. 48.052 BSC-BTS Interface principles, . URL <http://www.3gpp.org/ftp/Specs/html-info/48052.htm>. 25.06.2012.
- [6] A1. iTraffic Output Stream Specification. 2011.
- [7] AnachB Verkehrssystem. URL <http://www.anachb.at>. 25.06.2012.
- [8] Apache Commons Collections 3.2.1 API. URL <http://commons.apache.org/collections/api-release/index.html>. 25.06.2012.
- [9] Appello. URL <http://http://www.appello.se/>. 25.06.2012.
- [10] ASFINAG Verkehrsstatistik 2012. URL <http://www.asfinag.at>. 25.06.2012.
- [11] M. T. Borzacchiello. Real-time data from mobile phone networks for urban incidence and traffic management - A review of John Steenbruggen Maria Teresa Borzacchiello Peter Nijkamp Henk Scholten. 2010.
- [12] Bundesministerium für Verkehr, Innovation und Technologie, Austria. Verkehrsmo-
dell Austria.
- [13] F. Calabrese, M. Colonna, and P. Lovisolo. Real-time urban monitoring using cell phones: A case study in rome. *Intelligent*, 12(1):141–151, 2010. URL http://ieeexplore.ieee.org/xpls/abs/_all.jsp?arnumber=5594641.
- [14] Cisco. Documentation. URL http://docstore.mik.ua/univercd/cc/td/doc/product/wireless/moblwrls/cmx/mmg_sg/cmxgsm.htm. 25.06.2012.
- [15] D. Gilbert and T. Morgner. jFreeChart. URL <http://www.jfree.org/jfreechart/>. 25.06.2012.
- [16] GPS Distanzberechnung. URL <http://www.kompe.de/gps/distcalc.html>. 25.06.2012.
- [17] Graphenintegrationsplattform. URL <http://www.staedtebund.gv.at/ausschuesse/gis-koordinatoren/gip.html>. 25.06.2012.

- [18] ITU-T. Telecommunication Standardization Sector. URL <http://www.itu.int/ITU-T/index.html>. 25.06.2012.
- [19] Java TreeSet. URL <http://docs.oracle.com/javase/1.4.2/docs/api/java/util/TreeSet.html>. 25.06.2012.
- [20] C. Jian and D. J. Goodman. General Packet Radio System. Technical report, Rutgers University. 1997.
- [21] Korrelationskoeffizient. URL <http://de.wikipedia.org/wiki/Korrelationskoeffizient>. 25.06.2012.
- [22] M. Linauer. Generierung streckenbezogener Verkehrsdaten als Basis für den Einsatz in Verkehrstelematiksystemen. (1), 2005. URL <http://content.grin.com/document/v48514.pdf>.
- [23] F. Maier. Schriftenreihe Heft 13 Abschnittsweise Regressionsanalyse zur Schätzung von Verkehrskenngrößen München 2010. Technical report, Technische Universität München. 2010.
- [24] Matlab Korrelationskoeffizient und P-Wert. URL <http://www.mathworks.de/de/help/matlab/ref/corrcoef.html>. 25.06.2012.
- [25] Nadeau Software Consulting. URL http://nadeausoftware.com/articles/2008/02/java_tip_how_read_files_quickly. 25.06.2012.
- [26] OEBB. URL <http://www.oebb.at/>. 2012.06.25.
- [27] Open Street Map. URL <http://www.openstreetmap.org/>. 25.06.2012.
- [28] Rundfunk und Telekom Regulierungs-GmbH. RTR Telekom Monitor 1/2012. Technical Report September 2011, 2012.
- [29] J. Schlaich and T. Otterstatter. Mobilfunkdaten im Verkehrswesen - Teil 1: Beobachtung von Ortsveränderungen. *Traffic Engineering*, 2011.
- [30] P. Schosteritsch. Traffic analysis using mobil phone data. *Technische Universität Graz*, 2012.
- [31] Stackoverflow. URL <http://stackoverflow.com/questions/964332/java-large-files-disk-io-performance>. 25.06.2012.
- [32] SwingLab. JXMapKit. URL <http://today.java.net/pub/a/today/2007/10/30/building-maps-into-swing-app-with-jxmapviewer.html>. 25.06.2012.
- [33] TMC Meldungen Oe3. URL <http://oe3.orf.at/verkehr/stories/370103/>. 25.06.2012.
- [34] UMTS-Link. URL <http://www.umtslink.at>. 25.06.2012.

Literatur

- [35] P. Veit and M. Fellendorf. Verkehrssimulation zwischen Graz und Salzburg. *Unveröffentlichte Studie der Technische Universität Graz im Auftrag des Amtes der Steiermärkischen Landesregierung*, 2012.
- [36] Wiener Linien. URL <http://www.wienerlinien.at/eportal/>. 25.06.2012.