MULTICAST TRAFFIC MODELLING AND LOAD ALLOCATION IN HYBRID
CELLULAR NETWORK

Mat-2.108 Independent Research Project in Applied Mathematics

Erkka Jalonen
55543H
erkka.jalonen@hut.fi
Contents

Definition of terms and abbreviations 2

1 Introduction 5
1.1 The objectives of the study 6

2 Description of the network 8

3 Description of the traffic allocation problem 10

4 Static snapshot of the network 12
4.1 Model parameters 12
4.2 Notation 13
4.3 Frequency band restriction 14

5 Probabilistic model of the network 16
5.1 Introduction 16
5.2 Traffic evaluation 16
5.3 Capacity blocking in PTP cell 18
5.4 Capacity blocking in PTM cell 19
5.5 Overall blocking probability 20

6 Traffic allocation 21
6.1 Introduction to decision making agents 21
6.2 Problem complexity 22
6.3 The concept of utility 23
6.3.1 Accessibility maximization 23
6.3.2 Bandwidth efficiency maximization 23
6.3.3 Bandwidth usage minimization 24
6.3.4 Load balancing 25
6.3.5 Integrated optimization model 25
CONTENTS

6.4 Long run optimization ............................................. 25

7 Conclusions .................................................. 27

Bibliography .................................................. 28

Appendix A .................................................. 30

Appendix B .................................................. 32
Definition of terms and abbreviations

Abbreviations

UMTS Universal Mobile Telecommunication Services
RAN Radio Access Network
GSM Global System for Mobile communications
DVB-T Digital Video Broadcasting - Terrestrial
WLAN Wireless Local Area Network
GPRS General Packet Radio System
QoS Quality of Service
RX Receiver
TX Transmitter
TRX Transmitter-Receiver, Transceiver
PTP point-to-point
PTM point-to-multipoint
Terms

Terms dedicated to describe the hybrid cellular network and the traffic allocation problem.

Service A data flow with specific bit rate, duration, subscribed terminals and providing cells.

Bearer Information transmission path carrying a service. Bearer allocates an effective bandwidth. If the bearer can simultaneously deliver the service for more than one receiver then it is called point-to-multipoint (PTM) bearer, otherwise point-to-point (PTP) bearer.

Effective bandwidth Frequency bandwidth required by a bearer. The bandwidth is a function of, e.g., the service bit rate, bearer type and signaling overhead.

Point-to-point Information transmission method where each of the bearers are dedicated to deliver a specific service to a specific user. Thus in the case of several receivers several bearers must be established.

Point-to-multipoint Information transmission method where a bearer is able to deliver a specific service to a to a specific group of users. As a tradeoff for the smaller requirement of bearers, (compared to PTP) the PTM bearer may require wider effective bandwidth to deliver the same amount of data.

Cell A fixed access point capable of establishing bearers. E.g., cell of a UMTS network. Used as a synonym for base station.

RAN Radio access network. RAN is set of cells. Cells of a RAN are non-overlapping because of the cellularity [11]. E.g., GPRS network.

Terminal Mobile user equipment with one or more RXs. E.g., a laptop equipped with both UMTS and WLAN receivers.

Hybrid cellular network A cellular network combining diverse wireless systems as one access network. A set of RANs.
Chapter 1

Introduction

Research on wireless IP data multicasting in hybrid cellular networks (e.g., European Union DRiVE project [19]) has aroused questions about the most efficient ways of data delivery in multi-radio environments. A new possibility of selecting the bearer for IP multicast [16] data delivery among different radio access networks has been enabled because, firstly, user equipment is able to receive data through more than one radio access network and, secondly, operators simultaneously utilize multiple radio access networks.

The scenario of a network consisting of diverse interconnected RANs - hybrid network - has emerged because of changes in the technology environment. Firstly the present wireless systems do not contrast with each other [11],[15] to encourage taking advance of special characteristics of different technologies simultaneously. In the new scenarios we have wireless data networks with ranges of very different categories, e.g. from 30m of WLAN to about 10km of DVB-T [14],[12], and networks with very different characteristics, e.g., broadcast in DVB-T [12]. Secondly the need of multicast and broadcast arises from the growing demand of multi-user data services, e.g. television, radio, newspapers, multiplayer games and chats [15]. The benefit of broadcast [18] and multicast [18] is that the same data can be sent to arbitrary many receivers whereas with unicast the data must be sent to every receiver one by one substantially increasing the amount of consumed resources (e.g., frequency bandwidth) [18]. The present technologies are mainly all of unicast nature and therefore no major reductions to resource consumption are possible.

Figure 1.1 illustrates an example of the issue.

The figure 1.1 illustrates a situation where there are users with data terminals, user equipment, equipped with both UMTS and DVB-T receivers. Users, denoted by squares, are willing to receive a service, let us say service A. They make a request that when service A is being delivered they would like to receive it. Because operator utilizes both UMTS cells and a DVB-T cell, he should make a decision when delivering the service to the subscribed users: Should service A be delivered using only UMTS cells, using the big DVB-T cell and
Figure 1.1: An example scenario – which cells to use for data transmission if every user equipment has both DVB-T and UMTS receivers?

some of the UMTS cells or should he send anything at all? We call this as a traffic allocation problem.

This study is motivated by the question above and the assumption that, by doing intelligent traffic allocation, the capacity of the network could be increased, or equally, the resource consumption could be decreased. Accordingly the results of the study can be utilized in the business evaluation of the resource allocation.

The phrasing of the business evaluation can be divided as follows:

- What the service level traffic allocation problem is in more detail?
- What are the ideal benefits that the solution to the problem may offer?
- What are the benefits that the solution to the problem may offer in practice?

1.1 The objectives of the study

The purpose of the study is not to carry out the business evaluation but establish framework for the business evaluation of the traffic allocation.

In the study we restrict on the study of the high-level traffic modelling and ‘last mile’ traffic allocation in general hybrid cellular network. A general hybrid cellular network is a network that consists of arbitrary diverse interconnected cellular RANs. By interconnection we mean that the provider of network is able to decide which RAN to use for the delivery
of a service. By high-level traffic we mean that we restrict on considering the services as a
bulk data streams. Note that the traffic allocation concerns only allocation in the interface
between base the stations and the terminals.

To keep the document concise and coherent enough, the scope is confined as follows:

- **The technical details and facts are left out of scope.** The reason for this is the
vast amount of technological details and the dynamic state of the current standardization
concerning the field [13]. In addition this is done to maintain adequate consistency.
The facts used to construct the mathematical models in this document are considered
as given.

- **Models constructed are adequately simplified.** The diversity of technologies
combined with the overall complexity of the system, makes it infeasible to study more
complicated system.

- **No simulation results and no business case data are presented.** The focus of
this document is the system modelling and problem solving.
Chapter 2

Description of the network

The general model of the hybrid cellular [11] network is constructed considering the characteristics of present technologies capable to deliver IP data (UMTS [7], DVB-T [12], WLAN [14] and GPRS [15]).

The hybrid cellular network studied could be characterized on the high level as follows:

- The cells are the access points of the system. They are categorized as follows: Cells capable establishing PTP bearers (e.g. GPRS), cells capable establishing PTM bearers (e.g. DVB-T) and cells with capability of establishing both PTP and PTM bearers (e.g. UMTS). Because the standardization concerning the cell with both PTP and PTM capability is still on-going [13] they are left beyond the scope. One cell can send multiple services simultaneously as far as they do not exceed its capacity [15].

- Terminals are the units demanding services in the system. Terminals are equipped with one or more RXs (e.g. both UMTS RX and WLAN RX). One terminal can receive multiple services simultaneously from multiple different sources. The terminal itself has no capacity restrictions.

- There are several different multicast services offered in the network. They could be divided to two distinct categories: Streaming services [16] are long duration fixed speed data flows that terminals can join to and resign from. Scheduled file downloads are delivered to predefined user group at predefined moments [19]. In the study scheduled file downloads are considered as streaming services with a definite duration.

- The study focuses on data delivery. No assumptions about return channels are made. The data is delivered so that a cell establishes a bearer that is used to deliver the service. After the delivery the bearer is de-established. This is how it works in the case
of a PTP bearer. With a PTM bearer the same bearer is able to deliver the service to all users covered.

- Each cell has a coverage area, specific for every service. Coverage areas of the cells of different RANs can intersect but the cellularity prevents two coverage areas of same RAN to overlap [11]. The coverage area is determined by the QoS requirements of the service delivered. In the study it is sufficient to know whether or not terminal can receive service using a specific RAN and if positive, through which cell the data will be received.

- Terminal, that both is in the coverage area of a cell determined by some service and has RX capable of receiving data from the RAN which the cell belongs in, can receive the service through the cell, assuming that the cell is the provider of the service.

- Service transmissions can be either on-air or off-air. Service being on air stands for the fact that the service is transmitted at the moment.

- Terminals are able to move, resign from a service and subscribe to a service. One terminal can receive multiple services simultaneously. Subscribing to a stream is done when the service is on-air and to scheduled file download beforehand when the service is off-air.

- A cell is able deliver a service only if it is subscribed as service provider for the service. Cell can as well unsubscribe.
Chapter 3

Description of the traffic allocation problem

The traffic allocation (or resource allocation) problem, in hybrid network described in more detail in chapter 4, is described as follows:

- A decision-making entity, referred as RAN selection system, is an independent agent doing the resource allocation decisions. It is assumed that the RAN selection system gives immediate response to a decision inquiry.

- There are only streaming services offered in the network (as stated in 2, file downloads could be represented as streaming services). Only downstream (from cell to terminal) data services are considered. Services are routed undivided.

- RAN selection system gains access to the full information about the networks (e.g. knows in the coverage area of which cell every terminal is located).

- As a response to the traffic allocation inquiry a set of terminal & network pairs is returned. The response explicitly identifies the routings [16]. In practice the system updates routing entries [16] and therefore rerouting concretely takes place only if new routing entries differ from previous.

- Traffic allocation inquiries are service specific. Firstly it means that, when the inquiry is made, the RAN selection system reroutes the service in whole and therefore may modify the routing of all the terminals receiving the service. Secondly it means that the routing of other services is not affected at all. Note that the routing is modified if and only if new routing entries differ from previous.

- RAN selection system is inquired to route a service in following three cases. Terminal moves so that it either enters or leaves the coverage area of a cell offering the service.
CHAPTER 3. DESCRIPTION OF THE TRAFFIC ALLOCATION PROBLEM

The terminals subscribe to or resign from the active service. The service is activated.
If the service is deactivated, no intelligent routing is needed.

Definitions above contain an artificial restriction - the routing of a service does not affect
the routing of other services. This was considered as feasible restriction to decrease the
complexity of the system and only has affect on the formulas presented in section 6.3.
Chapter 4

Static snapshot of the network

4.1 Model parameters

The following information is considered essential to characterize the environment of hybrid cellular network studied. The discard process leading to the choice of the parameters is not covered.

The hybrid cellular network in whole consists of several RANs, cells, terminals and services. The parameters required of every element of the network to construct a model of the network are described in the following tables. RAN parameters are described in table 4.1, cell parameters in table 4.2, terminal parameters in table 4.3 and service parameters in table 4.4.

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>N</td>
<td>Network identifier.</td>
</tr>
<tr>
<td>mode</td>
<td>0, 1</td>
<td>Cells of the network are capable establishing bearers of following type: 0 = PTP, 1 = PTM.</td>
</tr>
</tbody>
</table>

Table 4.1: RAN specific parameters
Table 4.2: Cell specific parameters

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>N</td>
<td>Cell identifier.</td>
</tr>
<tr>
<td>type</td>
<td>N</td>
<td>RAN type of the cell. Corresponds to the id of the RAN.</td>
</tr>
<tr>
<td>freq</td>
<td>$\mathbb{R}_+$</td>
<td>Available frequency band.</td>
</tr>
<tr>
<td>Eff</td>
<td>$\mathbb{R}_+$</td>
<td>Nominal bandwidth efficiency (4.3).</td>
</tr>
<tr>
<td>Ovrlh</td>
<td>$[0, 1]$</td>
<td>Proportional overhead. Overhead stand for the excess data that must be sent with the original data (frames etc.).</td>
</tr>
</tbody>
</table>

Table 4.3: Terminal specific parameters

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>N</td>
<td>Terminal identifier.</td>
</tr>
<tr>
<td>rxs</td>
<td>$\mathcal{P}(\mathbb{N})$</td>
<td>RXs that the terminal utilizes. Correspond to the ids of the RAN.</td>
</tr>
<tr>
<td>cs</td>
<td>$\mathcal{P}(\mathbb{N} \times \mathbb{N})$</td>
<td>Cells that can be used to receive the service at the moment (coverage area). Correspond to the cell id and service id pairs.</td>
</tr>
</tbody>
</table>

($\mathcal{P}(X)$ stands for the power set of $X$.)

Table 4.4: Service specific parameters

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>N</td>
<td>Service identifier.</td>
</tr>
<tr>
<td>br</td>
<td>$\mathbb{Z}_+$</td>
<td>Service bitrate</td>
</tr>
<tr>
<td>dt</td>
<td>$\mathbb{R}_+$</td>
<td>Service duration.</td>
</tr>
<tr>
<td>pr</td>
<td>$\mathcal{P}(\text{Cells})$</td>
<td>Cells that are providing the service.</td>
</tr>
<tr>
<td>sb</td>
<td>$\mathcal{P}(\text{Terms})$</td>
<td>Terminals that are subscribed to the service.</td>
</tr>
</tbody>
</table>

4.2 Notation

If we define $\mathcal{C}$ as the index set of all cells, $\mathcal{T}$ as the index set of all terminals and $\mathcal{S}$ as the index set of all services in the hybrid network we can introduce following definitions.

(4.2.1) $c_i, i \in \mathcal{C}$ : A cell with index $i$.

(4.2.2) $t_i, i \in \mathcal{T}$ : A Terminal with index $i$.

(4.2.3) $s_i, i \in \mathcal{S}$ : A service with index $i$.

To characterize and create the compulsory relations between index sets presented above we define as follows:

(4.2.4) $\mathcal{C}_{ptp} \subseteq \mathcal{C}$ : The cells capable establishing PTP bearers.
(4.2.5) \( \mathcal{C}_{ptm} = \mathcal{C} \setminus \mathcal{C}_{pt} \) : The cells capable of establishing PTM bears.

(4.2.6) \( \mathcal{C}_T^{(i)} \subseteq \mathcal{C}, i \in \mathcal{T} \) : The cells of the RANs the terminal \( t_i \) has RXs for.
Note: \( \mathcal{C}_T^{(i)} \) contains all the cells that \( t_i \) could receive the service through.

(4.2.7) \( \mathcal{C}_S^{(i \times j)} \subseteq \mathcal{C}, i \in \mathcal{S}, j \in \mathcal{T} \) : The cells that cover \( t_j \) with \( s_i \).
Note: This is independent of the set (4.2.6).

(4.2.8) \( \mathcal{C}_S^{(i)} \subseteq \mathcal{C}, i \in \mathcal{S} \) : The cells subscribed to provide \( s_i \).
Note: This is independent of both (4.2.6) and (4.2.7).

(4.2.9) \( \mathcal{T}_S^{(i)} \subseteq \mathcal{T}, i \in \mathcal{S} \) : The terminals subscribed to \( s_i \).
Note: This is independent of both (4.2.6) and (4.2.7).

(4.2.10) \( \mathcal{S}_{act} \subseteq \mathcal{S} \) : The on-air services.

(4.2.11) \( \mathcal{T}_{S_{act}}^{(i)} \subseteq \mathcal{T}_S^{(i)}, i \in \mathcal{S} \) : The terminals receiving the \( s_i \).
Note: \( i \notin \mathcal{S}_{act} \Rightarrow \mathcal{T}_{S_{act}}^{(i)} = \emptyset \) and \( (i \in \mathcal{S}_{act} \land j \in \mathcal{T}_S^{(i)}) \Rightarrow j \in \mathcal{T}_{S_{act}}^{(i)} \).

(4.2.12) \( \mathcal{T}_{C \times \mathcal{S}_{act}}^{(i \times j)} \subseteq \mathcal{T}_S^{(i)}, i \in \mathcal{C}_S^{(j)}, j \in \mathcal{S} \) : The terminals receiving the \( s_j \) through \( c_i \). Note: \( \bigcup_{i \in \mathcal{C}} \mathcal{T}_{C \times \mathcal{S}_{act}}^{(i \times j)} = \mathcal{T}_{S_{act}}^{(j)} \).

### 4.3 Frequency band restriction

We can simplify the difference between the general PTP bearer and PTM bearer by defining real bandwidth efficiency, meaning the total bit rate of data sent per total frequency bandwidth consumed ratio, \( e_{c \times s}^{(i \times j)} \) of the \( c_i \) for the delivery of \( s_j \)[5]. The idea is that with PTP bearers the consumed frequency band is multiplied with the amount of receivers.

\[
(4.3.1) e_{c \times s}^{(i \times j)} = \begin{cases} 
\frac{E_{c}^{(i)} \cdot (1 - O_{c}^{(i)})}{|E_{c}^{(i)} - (1 - O_{c}^{(i)})|} & i \in \mathcal{C}_{ptm} \land j \in \mathcal{S}_{act} \land \mathcal{T}_{C \times \mathcal{S}_{act}}^{(i \times k)} \neq \emptyset \\
\frac{E_{c}^{(i)} \cdot (1 - O_{c}^{(i)})}{|E_{c}^{(i)} - (1 - O_{c}^{(i)})|} & i \in \mathcal{C}_{ptp} \land j \in \mathcal{S}_{act} \land \mathcal{T}_{C \times \mathcal{S}_{act}}^{(i \times k)} \neq \emptyset \\
\text{undefined} & j \notin \mathcal{S}_{act} \land \mathcal{T}_{C \times \mathcal{S}_{act}}^{(i \times k)} = \emptyset 
\end{cases}
\]

By \( |X| \) we mean the cardinality of a finite set \( X \) and by \( E_{c}^{(i)} \) the nominal bandwidth efficiency of cell \( c_i \) and by \( O_{c}^{(i)} \) the proportional overhead in a cell \( c_i \). The nominal bandwidth efficiency describes the bit rate per frequency band ratio and proportional overhead the proportion of excess data needed for e.g. signaling [16].
CHAPTER 4. STATIC SNAPSHOT OF THE NETWORK

In practice the real bandwidth efficiency $e^{(i \times j)}_{CS}$ can be simply considered as the ratio between the service bit rate $r_{s}^{(j)}$ and frequency band allocated by the service in a cell $f_{CS}^{(i \times j)}$ [5].

(4.3.2) $e^{(i \times j)}_{CS} = \frac{r_{s}^{(j)}}{f_{CS}^{(i \times j)}}, s_{j} \in S_{act}$

Because the frequency band of a cell is limited by using equations (4.3.1) and (4.3.2) the total frequency band consumed in a cell must be less than its capacity constraint

(4.3.3) $\forall i \in C : \sum_{j \in \{k \in S_{act} \mid f_{CS}^{(i \times j)} \neq \emptyset \}} (r_{s}^{(j)} | e^{(i \times j)}_{CS} |) \leq f_{c}^{(i)}$, where $f_{c}^{(i)}$ is the total frequency band available in cell $c_{i}$. 
Chapter 5

Probabilistic model of the network

We are compelled to assume a static routing policy in the network [28]. The more complicated decision-making is used during the traffic allocation the more complex the stochastic model would be. This assumption distorts the scenario, if a dynamic routing policy is used. The restriction is done to simplify the problem so that the blocking probabilities, probabilities that a when a service is delivered a subscribed user misses out the service [23], could be analytically and explicitly described.

5.1 Introduction

We assume that the system we are modelling is of the scale that there is no significant terminal movement out from and into the system. In addition we consider that the service assortment and the cells with their configurations remains fixed. This means that the variables (4.2.1) – (4.2.3), (4.2.4) – (4.2.6) and (4.2.8) are considered remaining constant within the time frame examined. So the variables are (4.2.7), (4.2.9), (4.2.10), (4.2.11) and (4.2.12). Because the mobile terminals moving uniformly inside the system both cause and decrease traffic, we can also consider (4.2.7) as constant. Further, a simple traffic model combines both (4.2.9) and (4.2.10). The traffic model along with the blocking probabilities stipulate the values of (4.2.11) and (4.2.12).

5.2 Traffic evaluation

For every service $s_i$ we collect the terminals, that have the same tuple of cells that they can receive service $s_i$ through, to disjoint subsets $T_{D_{s, i}}^{k(x)} \subseteq T_i, i \in \mathcal{S}, k \in D_{s}^{(i)} \subset \mathbb{Z}_+ \cup \{0\}$ (where $D_{s}^{(i)}$ denotes index set). To be able to receive service $s_i$ through cell $c_j$ the terminal should be in the cells coverage area where the service QoS is considered adequate, the terminal has
to be equipped with the specific RX and the cell $c_j$ has to be the service $s_i$ provider. This characterization can be formalized like a equivalence classes,

\begin{equation}
\forall j, l \in T: \quad j, l \in T_{D_{k \times S}}^{(k \times i)} \text{ for common } k \in D_{S_{i}}^{(i)} \iff 
C_{S_{i}}^{(i)} \cap C_{T_{D_{k \times S}}^{(k \times i)}}^{(j \times i)} \cap C_{T_{D_{k \times S}}^{(k \times i)}}^{(l \times i)} = C_{S_{i}}^{(i)} \cap C_{T_{D_{k \times S}}^{(k \times i)}}^{(j \times i)} \cap C_{T_{D_{k \times S}}^{(k \times i)}}^{(l \times i)}.
\end{equation}

This means that two terminals belong to the same subset if and only if they have the exactly same cells to choose the delivery route from. Furthermore we define that $T_{D_{k \times S}}^{(0 \times i)}$ is the set of terminals that are not able to receive $s_i$ at all. Also it is valid that $\forall i \in S: \quad \bigcup_{k \in D_{d}} T_{D_{k \times S}}^{(k \times i)} = T$.

We assume that the terminals subscribe to the services according to a Poisson process with constant intensity $\lambda$. Actually this holds true only with infinite terminal population but offers a good approximation of environments with a large number of users. Further, we assume that each terminal chooses the service $s_i$ independently of others and from the same preference distribution, $\alpha_i$, $(0 \leq \alpha_i \leq 1, \sum_{i \in S} \alpha_i = 1)$ being the probability that $s_i$ is chosen (e.g. Zipf distribution [30]). Now we can approximate that each of the terminal subsets $T_{D_{k \times S}}^{(k \times i)}$ offer arrival intensity for $s_i$

\begin{equation}
\lambda_{D_{k \times S}}^{(k \times i \times \eta)} = \frac{\alpha_i \lambda |T_{D_{k \times S}}^{(k \times i)}|}{|T|}.
\end{equation}

The equation above follows from the general properties of Poisson process [23].

If it is assumed that there is a fixed policy $\eta$ that defines the routing possibilities $p_{D_{k \times S \times C}}^{(k \times i \times \eta)}$ where $c_j$ is one of the cells that $s_i$ can be delivered to terminals $T_{D_{k \times S}}^{(k \times i)}$ through --- we also define that trivially $p_{D_{k \times S \times C}}^{(k \times i \times \eta)}(\eta) = 0$ if $s_i$ cannot be delivered through $c_j$ to $T_{D_{k \times S}}^{(k \times i)}$. Furthermore we define $p_{D_{k \times S \times C}}^{(k \times i \times \eta)}(\eta)$ as the possibility that the policy prevents terminals $T_{D_{k \times S}}^{(k \times i)}$ altogether from receiving the service $s_i$ (intentional blocking). Now it holds true that $\forall i \in S$,

\[
\sum_{\eta \in D_{d}^{(i)}} p_{D_{k \times S \times C}}^{(k \times i \times \eta)} = 1.
\]

This means that for every cell $c_j$ the offered arrival intensity for $s_i$ is

\begin{equation}
\lambda_{j \times S}^{(j \times i)} = \sum_{k \in D_{k \times S}} \left( \lambda_{D_{k \times S}}^{(k \times i \times \eta)} \cdot p_{D_{k \times S \times C}}^{(k \times i \times \eta)}(\eta) \right).
\end{equation}

We assume that the users’ holding times for $s_i$ are arbitrary distributed with mean $1/\mu_{s_{i}}^{(i)}$. Note that $\mu_{s_{i}}^{(i)} \leq t_{s_{i}}^{(i)}$. The offered traffic intensity $a = \lambda \mu_{s_{i}}^{-1}$ for $s_i$ becomes,

\begin{equation}
a_{j \times S}^{(j \times i)} = \sum_{k \in D_{k \times S}} \left( \lambda_{D_{k \times S}}^{(k \times i \times \eta)} \cdot p_{D_{k \times S \times C}}^{(k \times i \times \eta)}(\eta) \cdot \mu_{s_{i}}^{(i)} \right)
\end{equation}

We now briefly deduce the blocking probabilities for both point-to-point and point-to-
multipoint cells. We define:

(5.2.5) \( \beta_{s \times c}^{(i \times j)} \), time blocking of \( s_i \) in \( c_j \). The proportion of time that enough capacity units are occupied in \( c_j \) to block the further delivery of \( s_i \) [26].

(5.2.6) \( B_{s \times c}^{(i \times j)} \), service blocking of \( s_i \) in \( c_j \). The probability that the attempt to establish a bearer to deliver \( s_i \) is blocked [26].

(5.2.7) \( b_{s \times c}^{(i \times j)} \), call blocking of \( s_i \) in \( c_j \). The probability that the terminal’s attempt to receive the service \( s_i \) is blocked [26].

In this study we are mainly focusing on call blocking \( b_{s \times c}^{(i \times j)} \).

### 5.3 Capacity blocking in PTP cell

The service level of a point-to-point data network can be analyzed as multi bit rate network, where the service mixture consists of streams with various effective bandwidths. The following formulas are deduced in [24]. Note also that because the blocking of only one cell is considered, excess subscribers are not used.

We begin by defining, using (4.3.1) and (4.3.2), the possible states \( \Omega \) of a cell \( c_j \) with capacity \( f_i^{(j)} \) as follows:

(5.3.1) \( \Omega = \{ \mathbf{n} \in \mathbb{N}^K \mid 0 \leq \mathbf{w}^T \cdot \mathbf{n} \leq f_i^{(j)} \} \)

In the formula above \( K \) is the amount of services delivered through \( c_j \) and \( \mathbf{n} \in \mathbb{N}^K \) is a column vector with \( n_k = |T_{c_j}^{(j \times k)}| \) where \( k \) are the services the cell \( c_j \) is providing. Column vector \( \mathbf{w} \in \mathbb{R}^K \) contains the corresponding effective frequency bandwidths \( w_k = s_x^{(k)} / (E^{(j)}_c) \cdot (1 - O_{c}^{(j)}) \geq 0. \)

Blocking states \( B^i \) of the service \( s_i \) in a cell are defined correspondingly

(5.3.2) \( B^i = \{ \mathbf{n} \in \mathbb{N}^K \mid \mathbf{n} + \mathbf{e}_i \notin \Omega , k_l = s_i \} \),

where \( \mathbf{e}_i \) is a canonical unit vector with \( e_{ik} = \delta_{ik} \).

The state probabilities \( \pi(n) = P(N = n) \) of the cell can now be expressed as

(5.3.3) \( \pi(n) = \frac{\prod_{k=1}^{K} \left( \frac{s_{j \times k}}{n_k} \right)^{n_k}}{\sum_{\mathbf{m} \in \Omega} \prod_{k=1}^{K} \left( \frac{s_{j \times k}}{m_k} \right)^{m_k}} \).

And therefore time blocking probabilities \( b_{s \times c}^{(i \times j)} \) of service \( s_i \) in a point-to-point cell \( c_i \) can be expressed as follows:

(5.3.4) \( b_{s \times c}^{(i \times j)} = P(n \in B^i) = \sum_{\mathbf{m} \in B^i} \pi(m) = \frac{\sum_{\mathbf{m} \in B^i} \prod_{k=1}^{K} \left( \frac{s_{j \times k}}{m_k} \right)^{m_k}}{\sum_{\mathbf{m} \in \Omega} \prod_{k=1}^{K} \left( \frac{s_{j \times k}}{m_k} \right)^{m_k}} \).
CHAPTER 5. PROBABILISTIC MODEL OF THE NETWORK

In point-to-point network it hold true that (for service $s_i$ and cell $c_j$) service blocking probability $b^{(i\times j)}_{s\times c}$ is equal to call blocking probability $b^{(i\times j)}_{s\times c}$. Because the arrival process is Poisson, $b^{(i\times j)}_{s\times c} = B^{(i\times j)}_{s\times c} = \beta^{(i\times j)}_{s\times c}$ holds true.

The blocking probabilities are complex to calculate using the formula 5.3.4, which is why several alternative methods for calculating the blocking probabilities are developed. Some of them are exact (e.g. The Kaufman - Roberts recursion [24] and Convolution method [24]) and some approximate (e.g. establishing estimates by using Monte - Carlo simulation [21]).

5.4 Capacity blocking in PTM cell

The following formulas are deduced in [30],[26].

The probability $b^{(i\times j)}_{s\times c}$ that terminal’s attempt to subscribe to service $s_i$ fails is,

$$ b^{(i\times j)}_{s\times c} = \frac{B^{(i\times j)}_{s\times c}}{1 - B^{(i\times j)}_{s\times c} (e^{\lambda_{s\times c}} - 1) + 1} $$

where $B^{(i\times j)}_{s\times c}$ is the probability that the attempt to turn service $s_i$ on in the cell $c_j$ fails. Naturally $b^{(i\times j)}_{s\times c} \leq B^{(i\times j)}_{s\times c}$, because in point-to-multipoint cell, if service $s_i$ is already on the air, a new terminal doesn’t require any additional capacity.

Unlike with the point-to-point blocking model, we define the maximum capacity in bps $w^{(j)}_c$ for cell $c_j$ (using (4.3.1) and (4.3.2)) as $w^{(j)}_c = \left[ f^{(j)}_c \cdot E^{(j)}_c \cdot (1 - O^{(j)}_c) \right] \geq 0$ and let the effective bandwidths for service $s_i$ be $r^{(i)}_s$. Thus:

$$ B^{(i\times j)}_{s\times c} = \frac{\sum_{k=0}^{w^{(j)}_c} \pi^{(j\times s\times f)}_{c\times s\times f} \left( \binom{k+j}{j} \right)}{\sum_{k=0}^{w^{(j)}_c} \pi^{(j\times s\times f)}_{c\times s\times f}} $$

Where $\pi^{(j\times s\times f)}_{c\times s\times f}$ is the probability that $k$ capacity units are occupied in cell $c_j$ with service $s_i$ removed. The removal of service $s_i$ is dictated by the theory of Engset’s system [24] - as if arriving service activation were an outsider observer. This implies that $B^{(i\times j)}_{s\times c} \leq \beta^{(i\times j)}_{s\times c}$, where time blocking probability $\beta^{(i\times j)}_{s\times c}$ is determined from the formula similar to (5.4.2) without service $s_i$ removed. (The formulas are deduced e.g. in [26].)

The occupancy probabilities of service $s_i$ can be identified from the probability generating function:

$$(5.4.3) \quad \sum_{k=0}^{\infty} (\pi^{(j\times s\times f)}_{c\times s\times f} \cdot z^k) = \prod_{t\in\mathcal{S}\setminus\{i\}} (e^{-a^{(j\times s\times f)}_{c\times s\times f}} + (1 - e^{-a^{(j\times s\times f)}_{c\times s\times f}}) \cdot z^{a^{(j\times s\times f)}_{c\times s\times f}})$$

As with the point-to-point cells several methods for efficient calculation of the blocking probabilities are developed. For further information see e.g. [30].
5.5 Overall blocking probability

After knowing the blocking probabilities of individual cells, we can define the blocking probability of the overall network - in other words the probability that the first service reception of an arbitrary terminal is blocked. To calculate the overall blocking probability the theory of embedded Markov chains is used [25],[23].

There are three distinct types of blocking that can appear. One of them is blocking that appears when a terminal cannot receive the service at all because it either is not equipped with a particular RX or there are no cells in range sending the service (see section 4.2: terminal $t_j$ cannot receive service $s_i$ at all if and only if $j \in \mathcal{T}_{D \times S}^{(k \times i \times 0)}$). The blocking for those terminals is inevitable. Out of range the blocking probability can be defined as

$$(5.5.1) \quad P(\text{out of range blocking}) = \frac{1}{|\mathcal{D}|} \sum_{i \in \mathcal{S}} \alpha_i |\mathcal{T}_{D \times S}^{(0 \times i)}|.$$ 

Another blocking type is blocking due to the routing policy. This means that there is a possibility that the routing policy forbids the delivery of the service (see section 4.2: terminal $t_j$ can experience policy blocking of service $s_i$ if and only if $j \in \mathcal{T}_{D \times S}^{(k \times i \times 0)}$). The blocking caused by the policy $\eta$ used can be defined as:

$$(5.5.2) \quad P(\text{policy blocking}) = \frac{1}{|\mathcal{D}|} \sum_{i \in \mathcal{S}} \alpha_i \sum_{k \in \mathcal{D}_{\eta}^{(i)}} |\mathcal{T}_{D \times S}^{(k \times i \times 0)}| P_{D \times S \times C}^{(k \times i \times 0)}.$$ 

The third blocking type is the blocking due to the limited bandwidth capacity of the cells. This can be defined as:

$$(5.5.3) \quad P(\text{capacity blocking}) = \frac{1}{|\mathcal{D}|} \sum_{i \in \mathcal{S}} \alpha_i \sum_{j \in \mathcal{C}^{(i)}} \sum_{k \in \mathcal{D}_{\eta}^{(i)}} |\mathcal{T}_{D \times S}^{(k \times i \times j)}| P_{D \times S \times C}^{(k \times i \times j)}.$$ 

The three blocking situations are mutually exclusive because the blocked requests in (5.5.1) are not included in (5.5.2) and the blocked requests in both (5.5.1) and (5.5.2) are not included in (5.5.3). Therefore the overall blocking probability, regardless of the blocking type, is obtained by adding up the blocking probabilities (5.5.1), (5.5.2) and (5.5.3). The overall call blocking $b_{\text{sum}}$ is therefore:

$$(5.5.4) \quad b_{\text{sum}} = \frac{1}{|\mathcal{D}|} \sum_{i \in \mathcal{S}} \alpha_i \left( |\mathcal{T}_{D \times S}^{(0 \times i)}| + \sum_{k \in \mathcal{D}_{\eta}^{(i)}} |\mathcal{T}_{D \times S}^{(k \times i \times 0)}| P_{D \times S \times C}^{(k \times i \times 0)} + \sum_{j \in \mathcal{C}^{(i)}} \sum_{k \in \mathcal{D}_{\eta}^{(i)}} |\mathcal{T}_{D \times S}^{(k \times i \times j)}| P_{D \times S \times C}^{(k \times i \times j)} \right).$$

The analytical calculation of blocking probabilities is feasible only with simple cases where the policy $\eta$ is, e.g., either a static random policy or a static pattern allocation policy [28] (above $\eta$ was considered a static random policy). When the policy $\eta$ used in a network is dynamic and makes non-trivial choices, it will become infeasible to calculate the exact blocking probabilities. Methods used to calculate approximate values for blocking probabilities are described in section 6.4.
Chapter 6

Traffic allocation

Although in the stochastic model of the previous chapter we defined routing possibilities \( p_{D \times S \times D}^{(k \times s \times d)}(\eta) \) to be fixed and dictated by the policy \( \eta \), policies like \( \eta \) are merely one possible approach to the solving of the traffic allocation problem.

6.1 Introduction to decision making agents

The traffic allocation, in practice RAN selection, is assumed to be done in the backbone network. Therefore the decision-making entity has been conceived to be an ideal rational agent. The terminology used in this section is similar with the terminology used in [4] and [28].

It is considered that the agent gains access to the complete information about the environment it operates in. Although this may not be the real situation (e.g. architecture of UMTS network [8]), it is a sensible assumption because we are interested in ideal outcome of the traffic allocation. The environment is also considered deterministic, non-episodic, dynamic and discrete [4].

In practice the agent operates by updating the routing table [2] entries so that it won’t act by giving immediate responses to routing inquiries. Because the system is dynamic, the routings suggested by the selection system may be forbidden (e.g. architecture of the UMTS network [8]), because of e.g. the capacity restrictions. These kind of collisions are managed by restarting the decision process with a more strictly confined state space.

There are numerous alternative methods of decision-making that could be used to solve the traffic allocation problem described. Agents could be characterized as follows:

- Autonomous/Nonautonomous/Learning agents

Agent lacks autonomy if its’ decision bases completely on built-in knowledge. E.g. agents using fixed routing policies fall into this category.
• Stochastic/Deterministic If the routing contains contingency, the agent is called stochastic. E.g. load balancing by routing services to networks by probabilities [28].

• Static/Dynamic

Static agents operate under time and state independent characteristics of the system, whereas dynamic agents operate under time dependent information. E.g. load balancing by using predetermined routing patterns is considered static.

• Reflex-/Goal-based-/Utility-based agent

Reflex agents make decision by using condition-action rules, goal-based agents make decision to actions in order to achieve goals and utility-based agents are able to operate with multiple goals [4]. Because the two last-mentioned agents base their decision on autonomous reasoning, they also need information about the environment and the impacts of their decisions [4].

By the rule of thumb the more complex and intelligent the agent is, the more effective it can be [4]. The selection of the optimal agent is beyond the scope of this study. Instead an ideal selection system is considered to determine the maximum profit possible to achieve by using utility based intelligent routing.

6.2 Problem complexity

If we consider an agent that routes the service by minimizing (or maximizing) a given objective function, we end up having a general combinatorial optimization problem. Despite the restriction that routing of a service will not change the routing of other services we still have an enormous amount of combinations to plough through.

Because the cells of the same RAN cannot intersect, there can be at most $n_{RAN} \geq n_{RX}$ alternative cells for every a terminal ($n_{RAN}$ is the number of different RANs in the system and $n_{RX}$ is the number of different RXs in the terminal). This implies that the computational worst case complexity of the cell selection problem with $n_{TERM}$ terminals ordering the service is $O(n_{RAN}^{n_{TERM}})$ because for every terminal subscribed to the service we need to select the RAN used. The problem is therefore NP complete [1].

A combinatorial optimization problem can be solved by using an algorithm that bases on methods described in [1] and [3]. A backtracking brute force method that ploughs through all the combinations is described in Appendix A and a backtracking branch&bound method is described in Appendix B. Heuristic methods, such as genetic algorithms and simulated annealing [1], require validation and are therefore out of scope of this study.

In respect of further algorithm development, first an accurate and simple algorithm implementation is needed because verified solutions are needed to validate algorithms that either
are so complex that there is a considerable chance of implementation defect or are of heuristic nature and therefore give only approximate solutions.

6.3 The concept of utility

There are no certain objectives that the RAN selection should be based on. Therefore some aspects were identified and studied: Economic aspects, throughput maximization aspects, resource consumption aspects and aspects concerning technical details.

Economic aspects are beyond the scope. This is because the experiences on existing technologies advise that there are no short-term variable costs in multicast networks. E.g. in [9] 30% of the total costs in DVB-T network during the study period were capital expenditures and 70% were operating expenditures such as personnel and maintenance costs. Because operating costs are considered fixed, there are no variable costs the algorithm could affect. Income is dependent on the user response, which requires further study.

Throughput maximization is related to the blocking probability minimization [22] and load balancing [28]. Blocking probability minimization is described as an example in section 6.4.

Resource consumption aspects are related to resources considered scarce. One of them is frequency band (spectrum) [15], which have been taken into closer examination.

Technical aspects considered are related to issues like battery power consumption and excess signaling traffic. Excess signaling may arise when the selection system reroutes a great amount services over and over again because of the noise of unceasing arrivals and departures. It is sensible to avoid excess routing but, as a tradeoff, hysteresis to the system may emerge. These concepts are out of scope of this document.

6.3.1 Accessibility maximization

With the accessibility maximization of \( s_i \) we mean the maximization of the total number of terminals receiving the service when it is on-air.

\[
(6.3.1) \quad \text{maximize } \mathcal{T}_S^{(i)} \text{ by changing sets } \mathcal{T}_{C_S}^{(j, i)}, j \in \mathcal{C}_S^{(i)} \text{ within solution space defined by (4.3.3)}. 
\]

The accessibility maximization can be done both alone and integrated to some other objective.

6.3.2 Bandwidth efficiency maximization

The traditional bandwidth efficiency is defined as the ratio between the bit rate and the bandwidth allocated per delivery [5]. Because there is no particular way to define bandwidth
CHAPTER 6. TRAFFIC ALLOCATION

efficiency for the whole network consisting of multiple cells we could use e.g. a weighted average of the bandwidth efficiency of cells to attain a value to be compared with other systems.

\[ (6.3.2) \quad \text{maximize} \quad \frac{\sum_{j \in \{k \in C_{\text{s}} : |T_{c_{\text{s}} - d_{k}}^{|i|} | \neq 0\} \{c_{c_{\text{s}}}, w_{c}^{(i)}\}} {\sum_{j \in \{k \in C_{\text{s}} : |T_{c_{\text{s}} - d_{k}}^{|i|} | \neq 0\} \{w_{c}^{(j)}\}} \] by changing sets

\[ T_{C_{\text{s}} - d_{k}}^{(jx)} ; j \in c_{S}^{(i)} \text{ within solution space defined by (4.3.3)}. \]

Above \( w_{c}^{(i)} \) stands for the weight of the cell \( c_{i} \). Good weights could be terminals under the cell, cell’s interference area in \( \text{m}^2 \) etc.

It’s easy to see that all by itself (6.3.2) would give a trivial solution where only one bearer (bearer with the best bandwidth efficiency) would be established. Therefore it should be used along with e.g. (6.3.1).

6.3.3 Bandwidth usage minimization

Frequency bands in space and time can be considered as a scarce natural resource [15], and therefore we should use them as effectively as possible. We introduce a new quantity:

\[ (6.3.3) \quad \Phi = f \cdot t \cdot a, [\Phi] = \text{Hz} \cdot \text{s} \cdot \text{m}^2. \]

\( f \), frequency band allocated

\( t \), time the frequency range is reserved

\( a \), reserved geographical area (cell’s interference area)

If we now try deliver a service consuming as little \( \Phi \) as possible, we end up minimizing:

\[ (6.3.4) \quad \text{minimize} \quad \sum_{j \in \{k \in C_{\text{s}} : |T_{c_{\text{s}} - d_{k}}^{|i|} | \neq 0\} \{b_{s}^{(i)} \cdot t_{s}^{(i)} \cdot a^{(j)}\} \] by changing sets

\[ T_{C_{\text{s}} - d_{k}}^{(jx)} ; j \in c_{S}^{(i)} \text{ within solution space defined by (4.3.3)}. \]

Above \( b_{s}^{(i)} \) stands for the bit rate of the service \( s_{i} \), \( t_{s}^{(i)} \) for the duration service \( s_{i} \) keeps the frequency range reserved and \( a^{(j)} \) for the reserved geographical area. It should be noticed that in case of a streaming service \( t_{s}^{(i)} \) depends on how long the terminal is subscribed to the service and in case of a file download \( t_{s}^{(i)} \) depends on the bearers capability to deliver data. If we assume that \( t_{s}^{(i)} \) is independent of the bearer and the terminal, we could simplify the formula as follows:

\[ (6.3.5) \quad \text{minimize} \quad \sum_{j \in \{k \in C_{\text{s}} : |T_{c_{\text{s}} - d_{k}}^{|i|} | \neq 0\} \{a_{j}^{(i)} \}} \] by changing sets

\[ T_{C_{\text{s}} - d_{k}}^{(jx)} ; j \in c_{S}^{(i)} \text{ within solution space defined by (4.3.3)}. \]
Again it should be notice that the formulas above should be used along some other objective (e.g. accessibility maximization (6.3.1)). Otherwise a trivial solution with no one receiving the service would occur.

6.3.4 Load balancing

Article [28] concerns load balancing between cellular base stations and it can be used as an introduction to the different policies.

Load balancing is about allocating the incoming traffic uniformly into the network. Several methods are described in [28]. To retain the approach where the objective is explicitly defined by a mathematical function, let us define a penalty function so that the penalty rate of change approaches negative infinity when the proportional load, denoted by \( \frac{f_c(j)}{f_c^{(j)}} \), in a cell \( c_j \) approaches one. If we let this rate of change function be \( -1/(1 - \frac{f_c(j)}{f_c^{(j)}}) \), we would be lead to maximize the overall proportional load, when optimizing the service \( s_i \) with the following formula:

\[
\text{maximize } \prod_{j \in C_{c_i}^j} \frac{f_c(j)}{f_c^{(j)}} \quad \text{within solution space defined by (4.3.3)}.
\]

Above \( f_c(j) \) denotes the sum of the effective bandwidth required by services delivered through cell \( c_j \).

Note that the derivation of the (6.3.6) using the penalty rate of change was particularly used to demonstrate the process leading to (6.3.6).

6.3.5 Integrated optimization model

The integrated model is,

\[
\text{maximize } \sum_i (\theta_i \cdot f_i),
\]

where \( f_i \)'s are the objective functions and \( \theta_i \) the weights set to functions by importance. However, in practice, the values of \( f_i \) mostly are incomparable. Some methods to solve the problem are described in [20].

Another way of integrating multiple objective functions together is to set them priorities over each other [21]. This makes some objectives more important than others.

6.4 Long run optimization

The optimization criteria described in section 6.3 based on the idea that the we wanted to achieve the optimum at a certain moment. If we abandon the objective and try to maximize
the long run expected revenues (or equally minimize the costs) instead, we end up having a continuous time Markov decision process [24], [25].

Although the state space needed on the calculation of the blocking probabilities expands exponentially (section 6.2), it can be reduced as follows. When we possess the service subscription arrival intensities, as demonstrated in 5.2, we can define a static traffic allocation policy for every state. Let us denote by \( a = a_i(\eta) \) the routing decision in state \( i \) when some minimization criteria is used (e.g. one of the criteria described in section 6.3). Again let us denote by \( c_i(a) \) the cost per time unit of the decision \( a_i \) (e.g. in accessibility maximization 6.3.1 the number of users blocked). Now let us use exponential discounting \( e^{-\alpha t}, \alpha > 0 \) to increase the relative significance of the states most probable to end in. Now the expected total discounted cost incurred [25] when the initial state is \( i \) can be described as:

\[
(6.4.1) \quad V_\eta(i) = E_\eta \left[ \sum_{k=1}^{\infty} \left( e^{-\alpha(\tau_1 + \cdots + \tau_{k-1})} \int_0^{\tau_k} c(X_k, a_k)e^{-\alpha t} dt \right) | X_1 = i \right]
\]

Where \( \tau_k \) is the expected time between the \((k-1)\)th and \( k \)th transition. The \( V_\eta(i) \) is minimized by finding the optimal policy \( \eta^* \). From [25] it follows that:

\[
(6.4.2) \quad V_{\eta^*}(i) = \min_{\eta} \left\{ \sum_{j=1}^{\infty} \int_0^{\infty} \left( \int_0^t e^{-\alpha s} c_i(a) ds + V_{\eta^*}(j)e^{-\alpha t} \right) dF_{ij}(t[a]) \right\}
\]

Where \( P_{ij}(a) \) are the transition probabilities from state \( i \) to state \( j \) determined by the routing decision \( a \) and \( F_{ij}(t[a]) \) is the distribution function of the time \( t \) until the transition takes place. Both \( P_{ij} \) and \( F_{ij} \) can easily be calculated from the exponential arrival data described in chapter 5 and the corresponding costs e.g. from the values of the function described in section 6.3.

Solving the differential equation (6.4.2) concerns still a state space of exponential order. The problem comes feasible is we neglect/approximate the terms with a small significance in the equation. This limits us to the 'stochastic neighborhood' of the beginning state \( i \). This discussion is beyond the scope of the study. Methods for solving the (6.4.2) are presented in [24], [25].

Note that if we define \( c_i(a) \) to be the number of users blocked, we end up having a discounting expected blocking probability over time minimization method.
Chapter 7

Conclusions

The study presents the modelling part and problem survey for the assessment of the utility of an ideal intelligent traffic allocation system.

- Mathematical modelling (chap.4) defines the simplified system and its parameters.
- Probabilistic modelling (chap.5) defines the dynamics of the system. The model describes the dynamics of the simplified network, described in the static model, when the assumptions on traffic arrivals with exponential distributions are valid.
- Problem definition (chap.6) formalized the problem and introduced some solution methods. Utility oriented solving turned out be a combinatorial optimization problem. Long run optimization by using Markov decision processes can in practice only produce approximated solutions.
- Utility determination (chap.6) deepened the concept of the utility of a traffic allocation scheme. The determination of the utility in hybrid network was found complicated.

The study indicated that the utility assessment requires further work as follows.

Ideal traffic allocation system described in section 6.3 selects the optimum traffic allocation scheme among all possible traffic allocation schemes. Finding the optimum in both instant moment (sec.6.3) and in the long run (sec.6.4) are combinatorial problems with computational complexity of exponential order. In practice this may mean that effective heuristic or approximating methods are needed to carry out the assessment.

In section 6.3 we skimmed the surface of the study on the utility of the different routing schemes. Eventually it is about business evaluation, which implies that the utility should be an economic quantity and it should be verified with simulations constructed on the data from presumably business scenarios.
Bibliography


BIBLIOGRAPHY


Appendix A

The simplest accurate method to solve a combinatorial problem with an arbitrary objective function is to go through all the possible configurations. One implementation is presented below. The running time of the algorithm is of exponential order.

procedure brute_force( p: problem configuration )
begin
    best_solution <- temp_solution <- trivial solution
    loop_variable <- TRUE

    while loop_variable equals TRUE
    begin
        temp_solution <- get_next_solution( p, temp_solution )
        if temp_solution equals NULL then loop_variable <- FALSE
        if profit temp_solution is better than profit best_solution
        then best_solution <- temp_solution
    end
    return best_solution
end

procedure get_next_solution( p: problem configuration,
                             s: old solution )
begin
    i <- terminal count

    while i is larger than 0
    begin
        if there are no more choices for terminal i then
        begin
            s <- reset choices for terminal i
            s <- get next choice for terminal i
            i <- i--
        end
    end
if there are choices for terminal i then
begin
    s \leftarrow \text{get next choice for terminal } i$
    \text{return } s$
end
end
return NULL
Appendix B

A simple backtracking algorithm that would assuredly give the global optimal solution without checking the trivially unfeasible solutions would speed up optimization significantly. One possible algorithm is a branch & bound algorithm introduced below. The running time of the algorithm is of exponential order.

procedure branch_and_bound( p: problem configuration )
begin
    best_solution <- temp_solution <- trivial solution
    upperbound <- infinity
    loop_variable <- TRUE
    
    while loop_variable equals TRUE
    begin
        temp_solution, upperbound <- get_next_solution_bnb( p,
            temp_solution, upperbound, best_solution )
        if temp_solution equals NULL
        then loop_variable <- FALSE
        if profit temp_solution is better than profit best_solution
        then best_solution <- temp_solution
    end
    return best_solution
end

procedure get_next_solution_bnb( p: problem configuration,
    s: old solution,
    u: upperbound,
    b: best_solution )
begin
    i <- terminal count
    
    while i is larger than 0
    begin
    }
if there are no more choices for terminal $i$ then
begin
  $s \leftarrow$ reset choices for terminal $i$
  $s \leftarrow$ get next choice for terminal $i$
  $i \leftarrow i - 1$
end

if there are choices for terminal $i$ then
begin
  $s \leftarrow$ get next choice for terminal $i$
  while $j = j + 1$ is smaller or equal to terminal count
begin
  upperbound $\leftarrow$ calculate upperbound when choices
  for terminals from 0 to $j - 1$ are known
  if upperbound is smaller than the profit then
  begin
    $i \leftarrow j - 1$
    $j \leftarrow \text{infinity} // \text{way out from the loop}$
  end
  end
  if $j$ is different from infinity then return $s$
end
end
return $\text{NULL}$
end