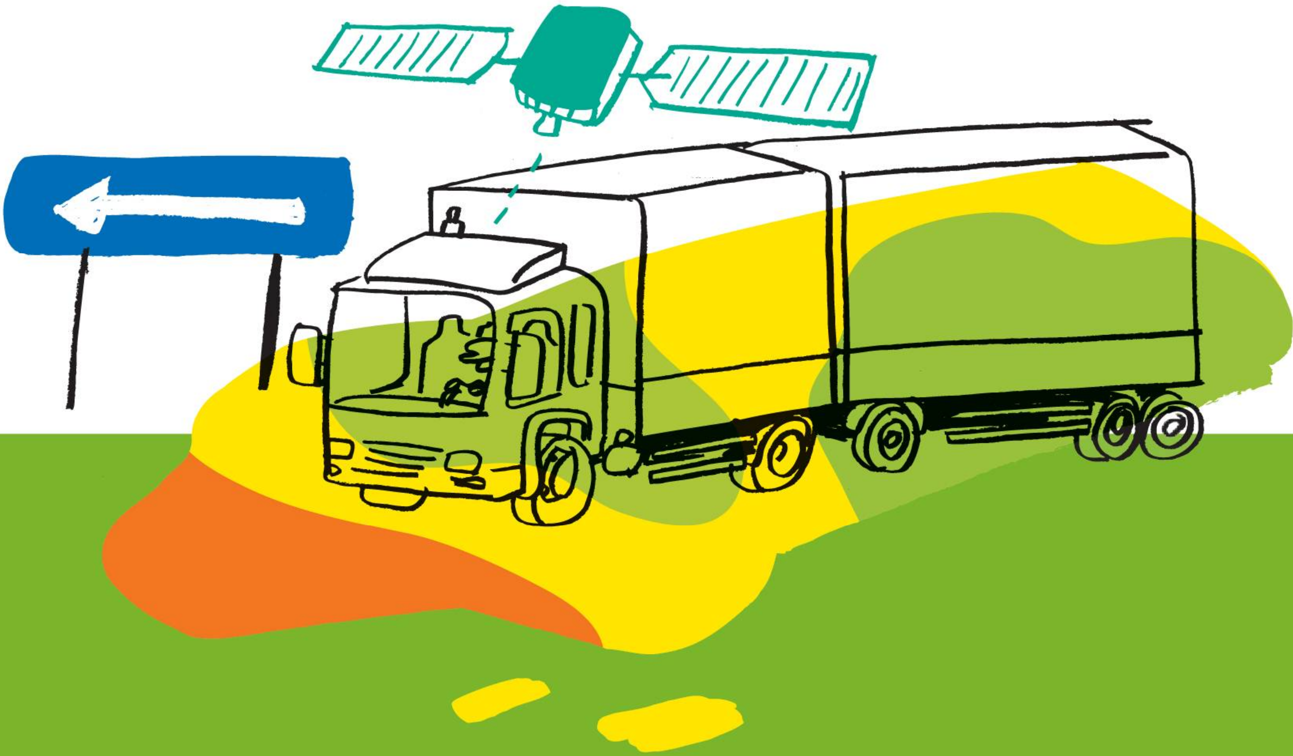

Dimensioning study for Road User Charging

Charlott Eliasson, Blekinge Institute of Technology
Markus Fiedler, Blekinge Institute of Technology

ARENA REPORT 2008:6

2008-02-29

The ARENA project

ARENA is a national project that aims to build competence for a future introduction of a road user charging system for Heavy Goods Vehicles (HGVs) in Sweden. The project has been developed in accordance with EU Directives and the Swedish public authority plans to introduce a kilometre tax for HGVs. ARENA started in 2006 and is financed by the Swedish Road Administration and the Swedish Governmental Agency for Innovation Systems. NetPort.Karlshamn is the project coordinator.

The approach of ARENA is to take a wide view and not only focus on technology. Innovation potential, consequences and possibilities related to an implementation of road user charging is also important as well as respecting that different stakeholders have different needs and requirements. This requires interaction between relevant stakeholders at an early stage. The role of the ARENA project includes the following elements:

- acting as broker both between groups of stakeholders who normally do not meet and between competitors within the same group
- develop and support knowledge both within the project but also as a coordinator between other projects

A concept for a kilometre tax system in Sweden is developed with a functional approach, which does not prescribe any technical solutions. The concept is generic rather than specific, in the sense that it should be possible to implement the result in several ways. Hence, we are trying to define the system independently from its final technical design. The motivation for this is that the time horizon for realisation is far ahead, maybe 3-6 years, and we can expect considerably changes in technical preconditions over this period. The concept includes a number of characteristics that differs from existing systems, which will reduce cost, promote innovative solutions and enable European interoperability.

The work of ARENA will continue in ARENA 2.0, where the concept will be further developed in close cooperation with the industry and relevant authorities and administrations. A full-scale demonstration will be developed for the ITS World Congress in Stockholm 2009.

Swedish Road Administration

The Swedish Road Administration (SRA) is the national authority assigned the overall responsibility for the entire road transport system in Sweden. SRAs task is to co-operate with others to develop an efficient road transport network in the direction stipulated by the Swedish Government and Parliament. SRA has been commissioned to create a safe, environmentally sound and gender-equal road transport system that contributed to regional development and offers individuals and the business community easy accessibility and high transport quality.

VINNOVA

VINNOVA (Swedish Governmental Agency for Innovation Systems) is a State authority that aims to promote growth and prosperity throughout Sweden. VINNOVAs particular area of responsibility comprises innovations linked to research and development. The tasks are to fund the needs-driven research required by a competitive business and industrial sector, and to strengthen the networks that are such a necessary part of this work.

Abstract

Road User Charging (RUC) solutions based on the concept of thin clients rely upon timely delivery of Heavy Goods Vehicles (HGV) positions towards a RUC system provider, which then performs the map matching and the calculation of the road user charges. In order to avoid data loss and late delivery, a proper dimensioning of the system is required, which is addressed by this report. We investigate the feasibility of legacy mobile and wireless systems (in particular GPRS, UMTS and WLAN) for both streaming and bulk transfer of positioning data, and we determine sensible combinations of sampling and reporting intervals in terms of efficient and economic reporting. Thereby, the chosen approach of required capacities allows for interpreting resource needs for different access technologies. We also highlight a couple of traps regarding the dimensioning, in particular of sending patterns that might cause overload at the server side of the system. It becomes obvious that mobile networks can well cope with the extra load caused by reporting HGVs, while bulk transfer in wireless hot-spots can be used as a complementary solution.

Sammanfattning

Vägavgiftslösningar baserade på konceptet med tunna klienter är beroende av tidskorrekta leveranser av positioner för lastbilar till leverantören av vägavgiftssystemet, som sedan utför kartjämförelsen samt uträkningen av vägavgifter. För att undvika dataförluster och sena leveranser måste en ordentlig dimensionering av systemet genomföras, vilket tas upp i denna rapport. Vi undersöker möjligheterna att använda befintliga mobila och trådlösa system (framför allt GPRS, UMTS och WLAN) till både streaming och bulköverföring av positioneringsdata och vi fastställer förnuftiga kombinationer av samplings- och rapporteringsintervall med tanke på effektiv och ekonomisk rapportering. Det valda tillvägagångssättet med erforderlig kapacitet öppnar upp för tolkningar av resursbehoven för de olika kommunikationsteknologierna. Vi framhåller även ett par fällor angående dimensioneringen, framför allt gällande de mönster enligt vilka data skickas, som kan förorsaka överbelastning på serversidan i systemet. Det blir tydligt att mobila nätverk på ett bra sätt kan hantera den belastningen som tillkommer vid rapportering av lastbilar, medan bulköverföring via trådlösa hot-spots kan användas som en komplimenterande lösning.

Table of Contents

1 Introduction

The ARENA project aims at the design of a Swedish Road User Charging (RUC) system. In particular, the Swedish system shall cover the whole national road network and allow for a diversification of road charges amongst other depending on the particular road an HGV uses at a particular date. This implies the need for discovering the positions of a HGV and matching them towards maps and pricing schemes. Such matching can take place on the On-Board Unit (OBU) in case map and pricing information are available; in this case, we speak about a *thick client*. An alternative solution, the *thin client*, relies upon delivery of positions towards a central server operated by a RUC provider that performs the map matching and the calculation of the road user charges. More background information and discussion of these approaches are found in [1] and [2], respectively.

The thin client solution is dependent on timely delivery of positions from the OBU to the server. Loss of this positioning data should be avoided, as it constitutes the basis for road user charges. As up to one hundred thousand HGVs might provide their positioning data from their On-Board Unit (OBU) towards the provider in more or less regular intervals, potentially using the same access network resources as mobile users of voice and mobile Internet service or as wireless users in (public) hotspots, such a solution needs to be robust and cost-effective. Thus, a proper dimensioning of such a system is mandatory.

On this background, this report studies dimensioning issues for RUC and aims to provide dimensioning guidelines for a Swedish RUC system. Based amongst others on assumptions made in the initial ARENA concept proposal [1] and on subsequent discussions, this report provides quantitative results for the dimensioning of the Swedish RUC system. The investigations and results presented in this report are expected to have an impact on the design of the system as such and on communication protocol between OBU and RUC provider.

The report is organised as follows. In Chapter 2, the different technologies and protocols of interest for the RUC system are described, and the concept of required capacity is introduced. With these technologies and protocols in mind, two main scenarios regarding the delivery of positioning data (streaming and bulk transfer) are presented and discussed in Chapter 3. Furthermore, the alternatives of how to report to the provider are listed in this chapter. Based on these two chapters, Chapter 4 focuses on specific technical challenges for the dimensioning of a RUC solution. It covers aspects such as HGV population; data formats; communication protocol particularities; and capacity, coverage and security issues. Chapter 5 covers the key parameters that are dealt with in this dimensioning study. In the end of this chapter, two studies of worst case scenarios in HGV traffic reveal the order of magnitude of worst case HGV density in the Swedish road network, which is an important prerequisite for the dimensioning. Chapter 6 presents case studies based on the two main scenarios from Chapter 3, and shows the estimated results in terms of required capacities as functions of the key parameters given in Chapter 5. The chapter ends with some design and dimensioning recommendations. Chapter 7 ends the report with a summary and an outlook of potential future work. The Appendix discusses calculations of positioning message lengths for both scenarios.

2 Technical Background

This part of the report aims at providing the technical background to this dimensioning study regarding the reporting between HGV and central server. We will shortly address wireless networks, communication protocols and data formats of particular interest, followed by a discussion on overload conditions in networks and the related concept of required capacity.

2.1 Wireless networks

As we target moving vehicles, wireless networks are mandatory. Here, a HGV connects on link level to a base station (BS) or an access point (AP). An important distinction has to be made between *mobile* and (ordinary) *wireless* networks. In the latter case, the wireless link replaces a cable. When the HGV moves away from the AP it is connected to, the risk for breaking the connection grows with the distance. When the HGV finally leaves the coverage area of one AP, it is not automatically connected to a new AP which is found closer by. What makes a network mobile is the handover facility between different BS. In this case, the moving HGV will be automatically connected to different BS along its way, which the end user perceives as ongoing connectivity. Still, as of today, handovers take only place within the same family of networks and within the domain of one operator. The emerging concept of *seamless communications* [3] will overcome this kind of limitations, allowing for switching between networks.

A general problem for any kind of wireless network is the dependability of the achievable data rate on the radio conditions. A bad signal-to-noise ratio (SNR) entails increased bit errors, which can be mitigated partly by lowering the speed. However, if the SNR becomes too bad, a loss of connectivity might be perceived.

2.1.1 GPRS

In the late 90's the 2nd generation (2G) mobile system, the GSM (Global System for Mobile communication) network, was extended with packet data capabilities, GPRS (General Packet Radio Service). GSM builds upon a combination of Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA). Each channel is split up into eight time slots. A voice call consumes one time slot while the call is going on. When sending data with GPRS, the given slot is only taken during that specific time. This enables many users to share the same transmission channel.

Transfer speeds of GPRS depend on class and Coding Scheme (CS) [4]. Near a Base Transceiver Station (BTS), fast coding schemes can be employed, but these are not very robust with regards to transmission problems. Far away from a BTS, robust and slower coding schemes are used. The faster coding schemes provide data rates up to 21.4 kbps per time slot (CS4), while the slowest, however most robust one (CS1) offers merely 9.05 kbps. The combination of n timeslots yields n times that capacity. Table 2.1 provides an overview.

When leaving one cell and entering another, while using GPRS for data traffic, the protocol will store and forward the traffic to the cell phone. There can be problems during this handover, since packets might get lost. An upper-layer protocol (e.g. TCP, see next session) may react to these losses with throttling of the transmission speed [4].

Table 2.1. Nominal throughput for GPRS at link level [5]

Coding Scheme # Slots	CS1 [kbps]	CS2 [kbps]	CS3 [kbps]	CS4 [kbps]
1	9.05	13.40	15.60	21.40
2	18.10	26.80	31.20	42.80
3	27.15	40.20	46.80	64.20
4	36.20	53.60	62.40	85.60
5	45.25	67.00	78.00	107.00
6	54.30	80.40	93.60	128.40
7	63.35	93.80	109.20	149.80
8	72.40	107.20	124.80	171.20

2.1.2 UMTS

The UMTS (Universal Mobile Telecommunications System) network is one of the third-generation (3G) technologies for cell phone communication, which are designed to succeed the GSM standard. UMTS is also called 3GSM to indicate its combination of 3G technology and the GSM standard. In theory, UMTS in its most common form supports up to 2 Mbps. Typically, codes supporting up to 384 kbps in downlink and 64 kbps in uplink direction are assigned to users. This capacity allocation is rather stable. If no specific code can be assigned, there is still the common transport channel available, offering rather little capacity in competition between different users. Handovers in UMTS are handled by the RRM (Radio Resource Management) protocol. With HSPA (High-Speed Packet Access), the capacity of UMTS will be extended. It is expected to support a transfer rate up to 14.4 Mbps in the downlink (HSDPA) [6], however at the expense of dependability of good radio conditions.

2.1.3 WLAN/Wi-Fi

During the recent years, Wireless Local Area Networks (WLAN) emerged as a means to offer wireless on-demand connectivity to end-users based on the IEEE 802.11 standards. WLAN are used interchangeably. Wi-Fi is today a blanket term for wireless technologies or IEEE 802.11 standards and the Wi-Fi branding is a guarantee that the product work together with other Wi-Fi products and technologies.

Different 802.11 technologies work on different frequency bands and have different ranges and transfer speeds. Similar to all is that the mobile devices connect to an access point, which may then be connected to a wired LAN and/or the Internet. The transfer speeds for WiFi is higher than the mobile technologies such as GPRS and UMTS, currently ranging from 1 Mbps (802.11b, critical radio conditions) to 54 Mbps (802.11g). Due to the collision-avoidance functionality, the user-perceived speeds are significantly lower; [7] reports a maximum of 6.5 Mbps for a single user for 802.11b and 28 Mbps for 802.11g. The total capacity of the wireless link sinks with the number of participants [8].

2.1.4 WiMAX

An upcoming technology is WiMAX, which is based on the IEEE 802.16 standards. It is developed to handle wireless connection over long distances, so called WMAN (Wireless Metropolitan Area Network). WiMAX is currently mainly used as wireless backhaul technology and might even be used to connect end-user locations (e.g. private homes) in dedicated areas. However, in contrast to WLAN/Wi-Fi, WiMAX equipment is hardly available as end-user devices; rather little is known regarding real-life performance results. For these reasons, WiMAX is not considered to be a serious alternative for position reporting at this time.

2.2 Communication protocols

2.2.1 Network layer: IP – the common basis

IP (Internet Protocol) runs over almost all physical media today and enables it to be viewed as an IP network. This packet delivery protocol is unreliable, best-effort and connectionless. It also provides fragmentation and reassembly, which means dividing and putting together datagrams when transmitting and receiving them respectively.

The IP-addresses enables routing of packets to their destinations. The addresses are of fixed length, which is 4 octets for IPv4 and 16 octets for IPv6.

2.2.2 Transport layer: TCP, UDP and others

TCP (Transmission Control Protocol) is working on a higher layer than IP and provides the connection-orientation and end-to-end reliability that IP does not have. When transferring data, TCP also provides flow control and multiplexing. The receiver, upon receiving a packet, sends an ACK that indicates with a sequence number what packet was received and the number of bytes that can, without causing problems, be received next time.

UDP (User Data Protocol) is working in the same level as TCP, but UDP basically works as an connectionless application interface to IP, with none of the reliability or flow control that TCP has. UDP only has multiplexing and demultiplexing for sending and receiving and uses ports to direct the datagrams. When sending data, there are certain limits on the network and in the transport protocols considering size. If a datagram exceeds this size, it will be segmented into two or more datagrams. This means that all the segmented packets will have to be received and reassembled before the whole datagram is received. If one of the segments does not reach the destination, when using UDP, the datagram is useless and will have to be resent.

SCTP (Stream Control Transmission Protocol) [9] is a recent addition to the transport layer protocols. It can be described as a lightweight TCP, as it is considered to have similar properties as TCP. It was designed, from the beginning, to deliver PSTN signaling messages over IP, but it also supports many other types of applications. Until now, SCTP is hardly used in the end-user context; it might become a serious alternative to TCP and UDP in the future.

2.2.3 SSL/TLS

SSL (Secure Sockets Layer) is a predecessor to *TLS* (Transport Layer Security), or TLS is a version of SSL that got its name when IETF adopted and changed it. They both provide security services for various higher layer protocols using the TCP/IP, using cryptography, such as providing a cryptographically protected channel for HTTP-requests. SSL/TLS uses certificates that were primarily intended to support electronic commerce, where the users want assurance that they are communicating with legitimate merchants. One of the purposes of SSL/TLS is to authenticate the server and possibly also the client. The latter is known as mutual authentication. The three basic phases of SSL/TLS are as follows:

- Peer negotiation for algorithm support;
- Key exchange and authentication;

- Symmetric cipher encryption and message authentication.

Out of the major implementations of SSL/TLS, only OpenSSL implements compression. Reference [10] states: "One advantage of TLS is that it is application protocol independent. Higher level protocols can layer on top of the TLS protocol transparently. The TLS standard, however, does not specify how protocols add security with TLS; the decisions on how to initiate TLS handshaking and how to interpret the authentication certificates exchanged are left to the judgment of the designers and implementors of protocols that run on top of TLS."

2.2.4 HTTP(S)

HTTP (HyperText Transfer Protocol) has initially been used to transfer data on the WWW (World Wide Web), but emerged as a universal protocol for data exchange (between so-called web services), for instance by exchanging XML documents, cf. Section 2.3.

HTTPS, indicating Secure HTTP, uses SSL to connect the client to the server for secure communication [11].

2.3 Data formats

As high-level format for data exchange and machine/business-to-machine/business communication, in particular in the context of the Web, *XML* (eXtensible Markup Language) [12] is considered as today's standard format. The price for its flexibility regarding user-definable data and corresponding human readability consists in rather large amounts of structuring information using so-called tags signalling the start and the end of a particular amount of data[1]. The potentially heavy overhead for structuring information entails that the data to be sent would increase in size, requiring more resources and yielding higher cost.

Binary data is a low level data format, which is typically used by machines in constrained environments. Compact binary data means that there would be less data to be sent and therefore the cost for sending it would also be lower as compared to a high-level format. However, the data of this format would not be readable at all without having the coding scheme readily available, since it shows no characters but 0 and 1 (or a range between 0 and E in hexadecimal representation).

The ASCII data format is found in-between the high and low level of formats and could be a good compromise for the two former formats. It would not be readable for anyone but persons knowing the way in which the data is organized. It shows characters that can be separated in order to retrieve the information needed.

2.4 Overload and bottlenecks

Overload implies that the capacity of a resource is insufficient as compared to the current demand. The notion is typically associated with the term *bottleneck*, which denotes a temporary or permanent shortage in capacity along a communication path [13]. We distinguish the following cases:

- *Temporary overload*, when the demand exceeds the capacity for a limited amount of time. Typically, this leads to queues, involving delays and perhaps even data loss if there are no (more) waiting facilities available.
- *Permanent overload*, when the demand permanently exceeds the capacity. This inevitably leads to data loss, or – if feedback towards the application is provided – to an ever-growing backlog of information to be transferred.

[1] For instance, <mydata>123</mydata> implies 3 B (15 %) for the data and 17 B (85 %) for the tags

Obviously, overload has a time dimension and may show up differently on different time scales. Consider an overload that lasts for some seconds. Typically, delays in the order of several seconds could be observed, and thus, the one-second time scale is affected. However, when describing that process on a minute time scale, that particular queue might come and go within the same minute and thus be invisible on the one-minute time scale.

2.5 Required capacity

The required capacity assures that the amount of workload that arrives within a time interval ΔT is processed completely within the same time interval. Consequently, it is a function of this time interval, i.e. $C(\Delta T)$. If resources at least equal to the required capacity were available at *all* time intervals, the consequences of overload would be limited to one time interval ΔT at a time.

As an example, assume a stream of constantly 100 requests per second. If the server offered handling those C(1 s) = 100 requests per second, there was never any queue. Assume now that the stream of requests toggles its intensity between 80 and 120 requests per second. The latter case means overload on the 1 s-scale: the queue would grow by 20 requests during one second. During the next second, it would sink by 20 requests during one second as only 80 requests were present in view of a capacity of 100 requests. In order to avoid any queue on the one-second scale, the required capacity was C(1 s) = 120 requests per second. However, the capacity on the one minute scale still amounts C(1 min) = 100 requests per second (or 6000 requests per minute), as at the end of that minute, the queue size has not changed. Obviously, dimensioning for longer time scales helps to save capacity, exploiting the fact that longer time horizons usually lead to smoother data delivery.

3 Scenarios

This section discusses different ways according to which positioning data can be sent from the HGV towards the server. In the end, it shortly reviews some reporting alternatives.

3.1 Streaming scenario

Streaming in the ICT context means sending data on a regular basis. Typical examples include voice, video and gaming traffic. Positioning data taken up more or less regularly also qualifies for this type of data transport. Important advantages of streaming are as follows:

- A responsive control loop is enabled, as the limited dead time allows for a reasonably short reaction time (e.g. on missing or corrupt data);
- Successfully streamed data has left the sender and does not pile up there.

However, there are also some challenges associated with streaming:

- A HGV on the move basically needs sufficiently well-working mobile connectivity, which might not be available everywhere along the route the HGV has to take;
- Streaming needs consolation of data loss and a specific handling of outages and (possibly subsequent) overload situations, as data is piling up at the sender if no connectivity is available;
- Scalability issues have to be taken into account, as many simultaneously reporting units entail large data rates.

A typical choice for streaming is UDP, which due to its simplicity and lack of error and flow control allows for quick data delivery. The protocol is stateless and does not take any responsibility for data delivery. Consequently, a tracking of success/failure has to be performed by the application. Thus, UDP-based streaming has to be implemented as

1. polling: the server asks for data, and the client answers;
2. ACK-ing/positive acknowledgements: the client sends, and the server acknowledges successful reception;
3. NAK-ing/negative acknowledgements: the client sends, and the server polls only if the data is not received at a specific point in time.

Option 3 involves the least amount of traffic, but most effort at the server that has to keep a time-based state for each reporting unit. Typically, raw data (i.e. without or with hardly any structuring information) is transferred via UDP.

Authentication and encryption are also performed on message level, i.e. application level. On the wireless link, we rely on the link-layer security mechanisms. The link between the GGSN of the mobile network and the server is sheltered by one VPN tunnel shared by all connections [14].

3.2 Bulk transfer scenario

In this scenario, positioning data is accumulated and sent as a bulk, e.g. when good data connectivity is available. The following advantages apply for bulk transfer:

- It is not as dependent on continuous coverage and connectivity as streaming;
- A connection needs to exist only during the time of delivery.

Specific challenges of bulk transfer are as follows:

- Once connected, quite large amounts of data might need to be transferred;

- Large times during which an OBU is disconnected imply large dead-times in payment control loops;
- Scalability on the server side can be an issue, if many simultaneous connections are established and kept open [15][16][17].

On transport level, the protocol TCP is used, trying for of complete delivery of the data at the expense of throughput reductions and delays in case of network problems[2]. TCP can be augmented by TLS/SSL, providing security on top of TCP at the expense of additional overhead. A TLS/TCP socket can be fed by raw data (in case of absence of structuring information), or by structured XML documents via HTTPS.

3.3 Coexistence of both scenarios

Looking at the advantages and challenges discussed in Sections 3.1 and 3.2, it becomes obvious that streaming and bulk transfer complement each other.

Streaming might need bulk transfer in order to cope with coverage problems. While a HGV is out of coverage or facing overload, it accumulates positioning data. As soon as it approaches an area of good coverage, as e.g. provided by UMTS or a WLAN hotspot, it may upload the accumulated data as one batch.

On the other hand, streaming might be a meaningful complement to bulk transfer in order not to pile up data unnecessarily, thus shortening bulk transfer times and providing positioning information more quickly.

3.4 Reporting alternatives

No matter whether streaming of bulk transfer is applied, we can distinguish between different reporting alternatives, involving different kinds of reporting objects:

- **Time-based reporting:** positions are determined at regular (possibly adaptive) instances of time;
- **Link-based reporting:** when a new link, e.g. a motorway, is entered, the corresponding positioning and/or link information is conveyed;
- **Distance-based reporting:** having traveled a certain distance puts forward the need for position reporting;
- **Payment reporting:** as soon as a certain amount of RUC is accumulated, a payment report might be necessary;
- **Final reporting:** when a HGV is about to leave the country, it might need to dump any accumulated data.

The way we deal with these reporting alternatives in a generic way in this study will be discussed further in Section 5.1.

[2] In case of very bad network conditions, a TCP connection can starve and be ended by a time-out before all data is delivered.

4 Dimensioning Challenges

The correct dimensioning of a RUC solution is a prerequisite for satisfactory behaviour. Amongst others, dimensioning has to address the frequency and volume of positioning data, the HGV density in network cells, the needs of communication protocols, overheads of security solutions, etc. This section discusses particular challenges encountered in the context of RUC systems.

4.1 HGV positioning data

HGV's should report how they travel by (i) positions, (ii) links or (iii) payments. In any of the cases, positioning data needs to be collected, since payment and link switching data can only be derived from the positioning data, e.g. from comparisons with on-board maps and payment information.

The positioning data might need to be collected with a short interval to make sure that the route of the HGV is calculated correctly and that no road switching is missed. This will potentially produce a lot of data, which may urge the use of lightweight protocols for sending the (raw) data, cf. Section 4.3. The big amount of data also stresses the point of calibrating the sampling interval to a level where it is as large (i.e. as long between samples) as possible, without any risk of loosing route information, to reduce the amount of data that need to be sent.

All data that is collected by a HGV, or calculated from the collected data, should be sent to the central server. If the amount of data to be streamed is limited and spread over time, sending data by streaming will not likely overload the mobile network. Though, with streaming, the network can be overloaded if there are more HGVs in a specific area than there is capacity for in the specific network. Sending data by bulk transfer (i.e. sending a large amount of data at the same time) can, like streaming, be done only if there is network coverage providing sufficient capacity, Since HGVs stop at different times and places, the risk that they will overload the mobile network is limited to extreme situations. Such situations might arise as many HGVs are stopping for the night and sending all data collected during that day, or a ferry is leaving from Sweden and all HGVs dump their remaining data before leaving. Another situation entailing risk of overload for bulk transfer is when the HGV has collected too much data and the capacity of the specific mobile network is not enough, i.e. after travelling through a large area of non-coverage.

4.2 Positioning data format

The main challenge when choosing the data format is to consider machine readability and compactness, the latter in order to reduce the amount of data to be transmitted and thus cost.

The format of the data depends on presentation level and target for the data. Data can be presented on a high level, i.e. it is readily readable by a person. This simplifies both understanding of and eavesdropping on the information to be transferred. The other option is low-level data, which is compact, readable by machines and hardly to be interpreted without additional formatting information. These features make it the format of choice for position reporting.

There is a possibility to compress the data. Thereby, the cost of transmitting (and storing) data is also reduced. The achievable level of compression depends strongly on the compression algorithm and the amount of redundancy contained in the data to be compressed; a reduction by 50 % is not uncommon, and in some cases, reductions of up to 80 % were observed.

4.3 Supported number of HGVs in an access network cell

The mobility pattern of HGVs determines their concentration in certain geographical areas, covered by cells belonging to mobile networks or wireless access points. For each considered network, this raises a couple of issues.

Since the GSM network has been extended to handle GPRS, both voice and data communication compete for the same resources (channels and timeslots therein). In many cases, revenue-maximising GSM communication comes at first hand. However, given a typical dimensioning of a GSM network for a blocking probability of 2 %, the average number of time slots occupied by voice is three out of seven slots. On one hand, this provides room for GPRS traffic – in fact, some slots might be reserved for GPRS in order to avoid starvation – but it also means that a carrier could easily get all remaining slots filled if the numbers of phone calls and reporting HGVs in the range of the cell grow beyond feasibility at the same time.

UMTS has similar problems as GSM/GPRS has with available time slots. UMTS is able to allocate code spaces to users beyond the common transport channel, providing them with dedicated capacity (e.g. 64 or 16 kbps). If too many HGVs are present in a certain cell, the latter might run out of codes and/or capacity on the common transport channel, with very low user-perceived throughput as a consequence.

WLAN has been designed as a wireless commodity, offering simple connection to hot-spots. A typical WLAN problem is that such hotspots are frequently over-populated with users, resulting in high collision rates and unsustainably low user-perceived throughputs.

Obviously, the natural mobility of HGVs implies the risk of overload, cf. Section 2.4, the consequence of which is inadequate resource allocation to some or all HGVs reporting via a certain operator in a certain area. The affected HGVs will not be able to transfer their data. To them, capacity starvation appears as loss of connectivity. Due to the dynamic nature of the problem, overload problems can quickly appear/disappear. Therefore, precautions have to be taken to detect and mitigate overload conditions. This calls for a suitable monitoring solution.

4.4 Required capacity

Overload and its consequences are avoided if the required capacity does not exceed the offered capacity. Expressing the resource needs of HGVs in terms of required capacity and comparing corresponding values with offerings by different wireless and mobile networks helps to gain an impression to which extent the reporting is sustainable.

A particular challenge consists in the fact that the reporting happens in upstream direction, which was not foreseen for transporting large amounts of data. For instance GPRS typically provides max. 2 timeslots upstreams as compared to max. 6 timeslots downstreams (ratio 1:3) In UMTS, the user-perceived capacity in downlink direction can reach 384 kbps as compared to 64 kbps in the uplink (ratio 1:6). The reason for designing the system with asymmetric capacity is that most user-oriented applications download data rather than they upload data. E.g. downloading a webpage and sending ACKs in response typically generates 40 B upstream for each 3 kB downstream (ratio 1:75). The latter example proposes that the capacity competition in the upstream between RUC reporting and other applications using the same cell seems to be limited. Still, the fact that the upstream capacity is quite limited in mobile scenarios still poses a challenge for dimensioning the HGV reporting.

Another dimensioning issue is the timescale for which the required capacity is determined. Assume the HGV reporting system is said to be able to manage 108000 requests per hour. On average, this means 1800 requests per minute and 30 requests per second, the latter of which is quite a decent number. If, however, the whole workload of 108000 requests would arrive during one minute, we would encounter 1800 requests per second on average, which is quite a heavy load. A delivery of the whole workload (108000 requests) during one second will definitely crash the system. For this reason, it is important to identify the timescale for which the dimensioning shall be valid, and to be specific about the expected request patterns.

The example above also shows the potential of capacity savings by controlling the request patterns. Uncoordinated requests imply the need for over dimensioning, still without guarantee for overload avoidance.

4.5 Coverage

Since the mobile network coverage is variable, especially on the countryside, it is inevitable for a HGV to end up in an area with no coverage, sooner or later. The HGV will then not be able to send any data and will have to save all collected data, until it is once again in an area with coverage. This will lead to “crowded areas” around the areas with no coverage, where many HGVs have a lot of data to send to the server. This can cause overload in the used network in the specific cell or area.

4.6 Communication protocols

For sending data over a mobile link, both UDP (used for streaming) and TCP (used for file transfer) have disadvantages. UDP can be risky, because it has no recovery mechanism for handling lost data. When using UDP, the latter has to be performed by the application. TCP can also be risky in a highly lossy environment, like a mobile link, where the recovery mechanism that handles lost packets can result in user-perceived throughput being lowered beyond feasibility because of frequent TCP timeouts and retransmissions. The big advantage of both UDP and TCP is that they can be considered as lightweight protocols concerning overhead.

When streaming, UDP is preferred in order to keep the flow of data and the simplicity of it. UDP also helps keeping the throughput, since it does not react to packet losses. Of course the loss of data has to be corrected by the application, which entails extra traffic, but the sending protocol is still keeping the throughput high. A general problem for streaming is segmentation of the data messages. If one out of n UDP messages is lost, the remaining n–1 messages also become worthless. Therefore, messages should be short enough (i.e. should be sent often enough) to avoid segmentation. Streaming with TCP would not be feasible, since connections need to be established and torn down for each small sending. Keeping connections open all the time can entail scalability problems at the server (as outlined below). There is also an overhanging risk of throughput limitations due to TCP’s reaction to network problems, such as loss of data. The use of XML for streaming is not feasible either because of the enormous amounts of overhead due to the given structure of XML documents. In this case, most space in the XML data (and thus most traffic) would be structuring information such as tags.

For bulk transfer, TCP is preferred. Here it is important that all data arrives, but not necessarily in a continuous fashion as a data flow. When transferring a larger amount of data it is more likely that a packet will be lost at some point of each transfer, which is taken care of by TCP. In this case application does not have to worry about error correcting when sending the data, as is the case with UDP; however, there is the risk of capacity starvation in case of bad connectivity.

For sending from a mobile device to the central server, the use of SSL has been discussed since is provides a certain amount of security. This would mean, in an every day scenario with streaming, approximately 10000 connections from mobile devices to the central server. Since SSL is not stateless, the server would be handling all connections (states) simultaneously, which would not be scalable for such a large number of connections, even if the server was to be extremely over dimensioned. In a case like this, it is always best to find a solution that is scalable from all points, and not to start with over dimensioning central parts of a system.

An evaluation of SSL [15] shows that the maximum throughput is as much as seven times lower when using SSL connections than with a normal connection. Another study [16], in the same line, shows that even when adding up to three extra CPUs the throughput will, despite a considerable increase in replies per second, still reach a point of saturation from where it will decrease in an exponential manner. Even a study of the follow-up of SSL, TLS [17], shows results of the same character. These studies underline the argumentation not to over dimension the server in the first design steps, but to find a solution that is truly scalable.

Because of the limitations outlined above, TLS/SSL does not appear to be an option for streaming. However, it could be used for bulk transfer in special scenarios. If a device for some reason have been in an area with radio shadow for some time and a lot of data needs to be sent, SSL could propose a solution to transfer all this data as a file. If designed and parameterised properly, this would not generate too many simultaneous connections for the server, since not all devices would connect simultaneously. Section 6.2 will discuss these challenges in detail.

4.7 Security

The data transported from the mobile device can be sensitive information, e.g. regarding competition between companies. Because of this, the data need to be encrypted in some way. The most vulnerable points, in a system like this, would be everything that is connected to the mobile device and also the link between the GGSN and the central server. Therefore, it is important that the mobile device can not be eavesdropped upon and that the link between the GGSN and the Central Server is secured from any kind of attacks, since it can be on a public and exposed link. In the latter case, a VPN solution between GGSN and server helps to encrypt data without running into scalability problems that would appear if the server had to maintain a security relationship to each reporting HGV.

Concerns in this area are that security must not be applied in several layers [18] and that the overhead of the security solution must stay proportional to the transferred data. Section 6.2 will illustrate the latter compromise.

5 Key dimensioning parameters

This section presents the most essential parameters for dimensioning the RUC system, both for the streaming and the bulk transfer scenario. First, different positioning alternatives are discussed, followed by a list of the key parameters relevant to positioning. Then, we turn our attention to a key problem for dimensioning, which is to correctly estimate the number of HGVs in a cell.

5.1 HGV positioning alternatives

The time instants at which the position is sampled from the GNSS can be triggered in different ways, which are denoted *positioning alternatives* and give raise to different types of positioning intervals:

1. **Fixed positioning intervals.** In order to assure a certain spatial and temporal resolution, positions are taken regularly in time. For instance, a fixed positioning interval of 4 s implies a spatial resolution of max. 100 m if the speed does not exceed 90 km/h.
2. **Adaptive positioning intervals.** If merely a certain spatial resolution is required, speed-dependent sampling can be used. Figure 5.1 illustrates the required sampling interval to achieve a resolution of 100 m.

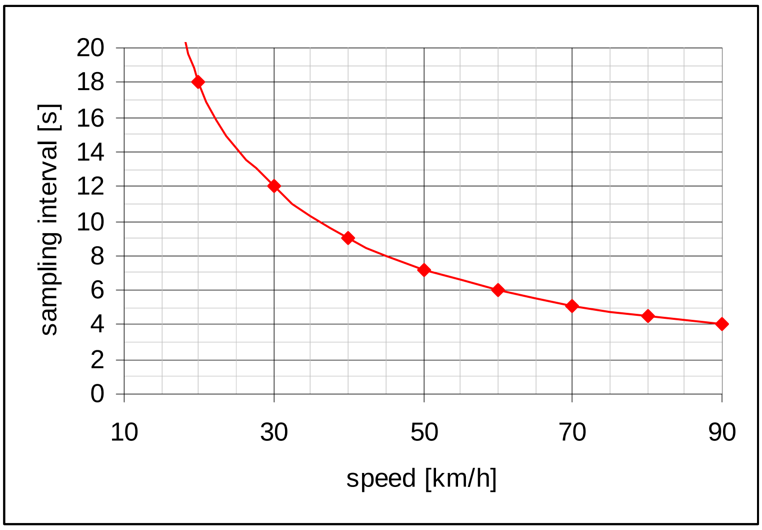

Figure 5.1. Sampling interval vs. speed for resolution of 100 m.

   No reporting takes place when the HGV is standing still.

3. **Event-controlled sampling** implies adaptive positioning intervals. For instance, an HGV might be required to report data when it
   a. enters a new link (link-based reporting);
   b. has traveled a certain distance (distance-based reporting);
   c. reaches a certain threshold of accumulated charges (payment-based reporting);
   d. is about to leave the country (final reporting).

For the dimensioning studies, we primarily focus on fixed times in order to prevent the parameter space from exploding. Results for adaptive systems can be derived from the ones for fixed systems by inter- or extrapolating. The fixed parameters to be used in the sequel may also be regarded as worst cases for the adaptive cases.[3]

Once sampled, positioning data can be sent immediately, or accumulated in order to form a positioning **report**, which is typically sent at a lower frequency than given by the sampling interval.

5.2 HGV positioning parameters

The following key parameters apply:

1. The **position sampling interval** t_S denotes the time interval between two successive readings of the GNSS position. We will consider intervals of 1 s, 4 s, 10 s, 30 s and 1 min.
2. The **position reporting interval** t_R denotes the interval between two reports of accumulated positioning data. We will consider intervals of 1 min, 5 min, 1 h, 8 h, 1 day and 3 days.
3. The **length of a positioning sample** l_{pos} is given by the data format. We will consider a common length of 10 B for longitude, latitude and time stamp. In order to save precious bytes, these values are proposed to be coded in a differential way with regards to reference values. For instance, the most northern and southern points in Sweden have latitudes of 69˚3'36" and 55˚20'13", respectively. The number of degrees alone consumes two bytes; however, the difference (14˚) only consumes one byte. As an additional way to minimize the amount of data, compression can be applied (cf. Section 6.1.7).

From these parameters and observing the worst-case assumption that any HGV samples and reports using fixed time intervals, we derive:

- The number of positioning samples per report

$$p = t_R / t_S; \qquad (5.1)$$

- The total length of the positioning report data

$$P = l_{pos}\, p. \qquad (5.2)$$

5.3 Number of HGVs in network cells

The number of HGVs in a network cell can be estimated in different ways:

1. Direct estimation from counts;
2. Indirect estimation from the cell size combined with the population density of HGVs.

Both methods suffer from incomplete information. The main problem is that information on cell boundaries is kept secret by the operators, which affects both methods: the area for counts is not known, and even if the HGV density was known, the unknown size of a cell makes it impossible to calculate the desired number.

A mobile operator has information about which cell a particular mobile unit, identified by its phone number (MSISDN)[4], is connected to. However, operators are reluctant to share such information for reasons of privacy and business.

A work-around to these problems is to take operator-unspecific information into account how mobile systems are built, which largely depends on local user densities. Cells take a

[3] For instance, an HGV that is standing still would not need to report. In the worst case scenario, however, each HGV is assumed to report at a constant rate.

[4] An unambiguous one-to-one matching of phone number and HGV has to be made sure by the RUC provider.

certain number of customers. For instance, one carrier (i.e. one frequency with seven timeslots) allows for roughly three simultaneous phone calls at a call blocking risk of 2 %. Consider a heavy-loaded road with many hundreds of cars per hour. To provide car drivers with sufficient capacity, many cells (potentially of different operators) need to exclusively cover this road, e.g. by applying special antenna lobes. Assuming such as cell structure, we are able to estimate the number of HGVs in a cell from traffic counts and a supposed length of the cell along the road. In order to get an impression on the worst-case load, it is advisable to use traffic counts during the busy time of day. From this data, we are able to obtain qualified estimations based on real-world observations which is illustrated in two examples

5.3.1 E22 in Lund

The Swedish Road Administration (Vägverket) offers traffic counts for different types of vehicles in the Skåne region with a maximal time resolution of 15 minutes, expressed in vehicles per hour [19]. We selected a route with a large amount of HGV traffic, the E22 northbound close to Lund on May 31, 2007[5]. During this day, 288 HGVs, 269 HGVs with trailer, and 248 busses were counted, giving a total of 805 HGVs per day. The busy hour was found between 22:00 and 23:00 (44 HGVs) and the busy quarter between 22:00 and 22:15 (27 HGVs). The corresponding values are compared in Table 5.1, where the factors compare the results for different averaging intervals. Obviously, the traffic intensity during the busy quarter is more than three times as high as the daily average. Traffic in the opposite direction was spread much more equally in time.

Table 5.1. Comparison of numbers of HGV in different time intervals, observed on May 31, 2007, E22 northbound at Lund South.

Time interval	Number of HGV	No. of HGV per hour	No. of HGV per quarter	Factors	
00:00 – 24:00	805	33.5	8.4	1.0	
22:00 – 23:00	44	44	11	1.3	1.0
22:00 – 22:15	27	108	27	3.2	2.5

At 90 km/h, the travel time through a cell of length 15 km amounts to 10 min. Taking the above-mentioned busy-quarter-based values into account, we arrive at an expected value of 18 northgoing HGVs in such a cell. A smaller cell size, which seems rather the case, would entail less HGVs. So the estimation of the order of magnitude of the number of HGVs is some few tens.

5.3.2 Verköleden in Karlskrona

A second observation was carried out on the sole road in Karlskrona connecting the landing stage of the ferry Gdynia (Poland) – Karlskrona with the road network such as E22 towards Malmö and Kalmar, 28 towards downtown and Vetlanda, and 122 towards Växjö. This road is most probably the one that is most busy with HGVs, namely upon arrival of a ferry. The northgoing HGV traffic was observed on May 14, 2007, from 07:30 to 08:45. The times when HGVs passed a point of reference were recorded. Based on these times, counts within intervals of variable length (1 min, 2 min, …, 30 min) were taken and evaluated. During the busy hour from 07:45 to 08:45, there was a stream of 70 HGVs with varying intensity; some HGVs were following each other almost bumper-to-bumper (i.e. at two-second distance), while there were also gaps in the stream up to several minutes, which matches the observed

[5] http://193.181.63.7/, follow link "Historik", select place (Lund Södra Nord E22 Norrgående) and save to file Lund Södra Nord E22 Norrgående, 070531 0000 - 070601 0000.xls

positive lag-1 autocorrelation of 28 % of the inter-HGV times[6]. The mean inter-HGV time was 50 s and the standard deviation 55 s. For each interval length, Figure 5.2 shows the maximal number of HGVs encountered in the set of evaluated intervals, while Figure 5.3 shows the corresponding ratio of the maximal to the average number of HGVs across these intervals.

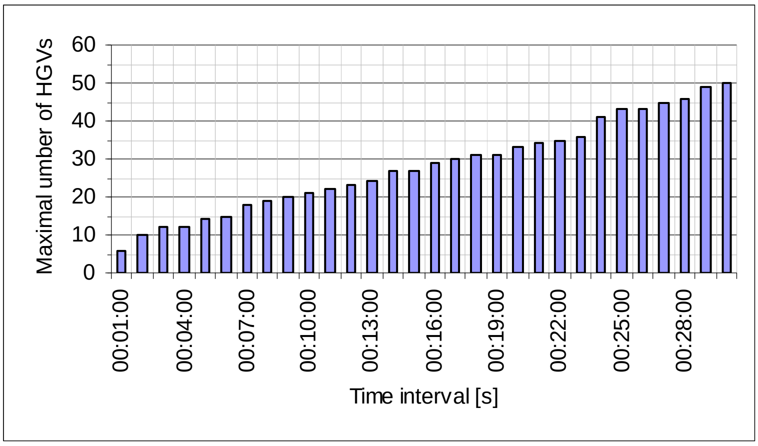

Figure 5.2. Maximal number of HGV observed during a given time interval versus the time interval.

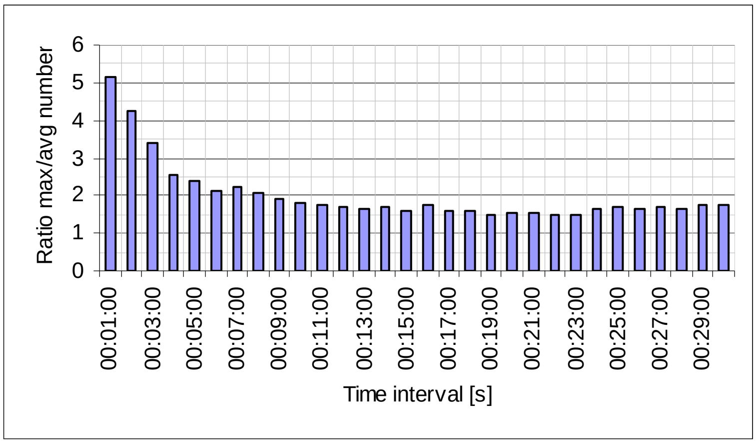

Figure 5.3. Ratio of maximal number of HGV observed during a given time interval and average number during the whole observation time versus the time interval.

The length of the road between the landing stage and the first major crossroad is about 6 km, which a HGV traveling at 70 km/h makes in about 5 min. Given the geography of Karlskrona, this could be a typical cell size. We observed a maximal number of 15 HGVs during this time, which closely resembles the Lund data. We observe a ratio of about 2.5 between the maximal number of HGVs observed during 5 minutes and the hourly average, which reduces to values in the range of 1.5 to 1.8 as the intervals get longer.

From all observations, we can see that a typical number of HGVs to be taken into account per cell is found in the order of ten to thirty. Several overlapping cells, potentially belonging to different operators, will decrease those numbers.

[6] Obviously, the traffic does not form a Poisson process (in that case, the standard deviation of the inter-HGV time would have matched the average, and the autocorrelation would have vanished). Thus, we unfortunately do not have access to a powerful dimensioning theory.

6 Case studies

This section will present and discuss numerical results for both the streaming and the bulk transfer scenario. The required capacities, given in bit per second (bps), kilobit per second (kbps) or Megabit per second (Mbps), denote long-term averages. This implies that if positioning data is sent in an uncoordinated way as outlined in Chapter 2.5, resource needs on short time scales (e.g. ΔT = 1 s) might exceed the shown values by one or several magnitudes. Thus, each case will be accompanied by some hints on how an efficient resource utilisation can be achieved.

The section will also present conclusions with regards to the design of a Swedish RUC system that can be drawn based on these results.

6.1 Streaming scenario

In the streaming scenario, outlined in Section 3.2, we will focus on the required capacity $C(N)$ as a function of the number N of HGVs in a cell. We round up the observations presented in Section 5.3 to the next order of magnitude, i.e. consider N_{max} = 100 HGVs as the maximal number of HGVs per cell. This helps us to easily scale down the results to the number N of interest.

UDP will be used as the transport layer protocol for transferring the positioning data from the HGV via PLMN, GGSN and VPN towards the server. The throughput on the VPN is then given by the sum of the throughputs of all reporting HGVs, where the number of 100000 is taken as an upper bound for the fleet of reporting HGVs in Sweden.

We limit ourselves to such combinations of sampling interval t_S and reporting interval t_R (5.1) that avoid segmentation into several UDP datagrams. If a position reporting was segmented into n datagrams and only one datagram got lost, the reconstruction of this data would need a major selective retransmission and post-processing effort; alternatively, all datagrams belonging to the damaged message are resent. From that point of view, it is much more efficient to have a one-to-one relationship between message and datagram. If a message gets too long, it is better to use a smaller reporting interval instead.

The application data consists out of p positioning samples, for which we assume a length of l_{pos} = 10 B, and an identification (ID) number of length 4 B. The calculation of the length L of a positioning message and of the maximal number of samples in order to avoid segmentation is found in Appendix A. We obtain the required capacity

$$\mathrm{r} = L \,/\, t_R \qquad (6.1)$$

as average over the reporting interval ($\Delta T = t_R$).

6.1.1 Reporting interval t_R = 1 min

In this case, all sampling intervals $t_S \in \{1\text{ s}, 4\text{ s}, 10\text{ s}, 30\text{ s}, 1\text{ min}\}$ can be applied without the risk of segmentation ($p \leq 60$). Figure 6.1 shows the required capacity as a function of the number of reporting HGVs.

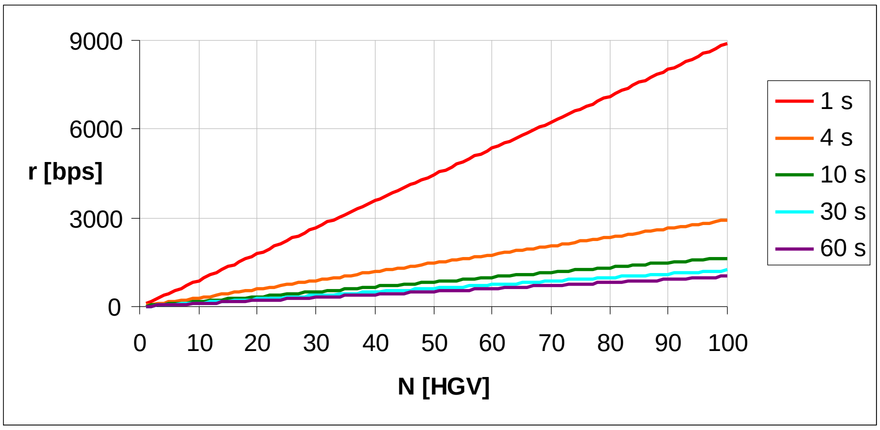

Figure 6.1. Required capacity versus the number of HGVs for a reporting interval t_R = 1 min and various sampling intervals.

Obviously, the shorter the sampling interval, the more messages are contained in a report, and the larger is the required capacity. For t_S = 1 s, we reach 9 kbps for 100 HGVs. For t_S = 4 s, the required capacity is reduced to 3 kbps for 100 HGVs, and increasing t_S implies sinking capacity requirements.

6.1.2 Reporting interval t_R = 5 min

In this case, the sampling intervals $t_S \in$ {4 s, 10 s, 30 s, 1 min} can be applied without the risk of segmentation ($p \leq 60$). In the following Figure 6.2, results for the sampling interval t_S = 1 s are indicated by a dashed line.

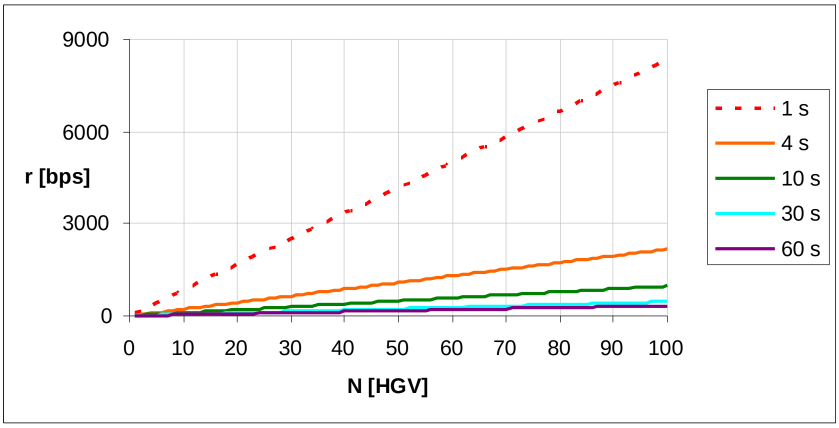

Figure 6.2. Required capacity versus the number of HGVs for a reporting interval t_R = 5 min and various sampling intervals.

Comparing Figure 6.2 with Figure 6.1, we realise that required capacities have become smaller, which is due to the comparatively higher number of positioning samples p per message, i.e. less overhead and thus a better utilisation. Now, for t_S = 4 s, 100 HGVs produce slightly more than 2 kbps.

6.1.3 Reporting interval t_R = 1 h

For such a long reporting interval, merely the sampling intervals $t_S \in$ {30 s, 1 min} avoid segmentation.

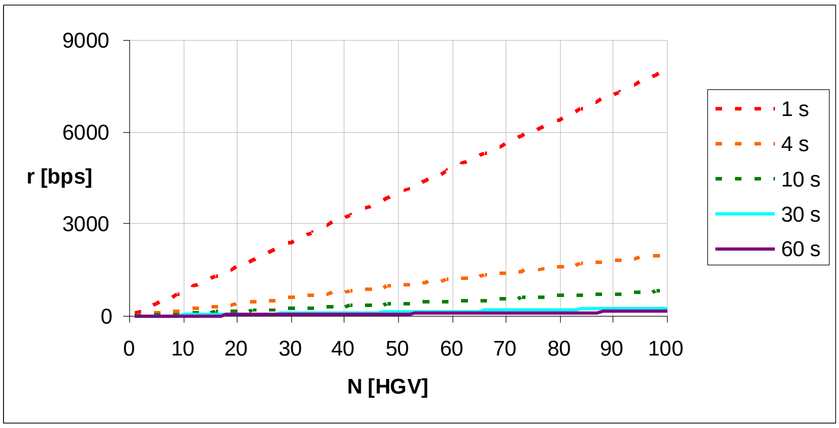

Figure 6.3. Required capacity versus the number of HGVs for a reporting interval t_R = 1 min and various sampling intervals.

As compared the results shown in Figure 6.3 to the ones shown in Figures 6.1 and 6.2, we again perceive a sinking capacity requirement. The only (in the sense of segmentation) feasible cases produce much less than 1 kbps even in the case of 100 HGVs.

6.1.4 Summary

For the cases avoiding segmentation, Table 6.1 compares the following parameters:

- *Relative overhead*

  $O = L/P - 1$ (6.2)

  with L given by Equation (A.1) and P by Equation (5.2), which illustrates the relative part of the overhead ($L - P$) as compared to the amount of positioning data P and thus provides an efficiency index of the reporting as such;
- *Data rate* per HGV r, given by Equation (6.1), which provides an idea about the average contribution to the system load;
- *Volume*

  $V = L \cdot 240\ \text{h} / t_R$ (6.3)

  per month when assuming 240 h reporting time per month, i.e. 30 days à 8 h or 20 days à 12 h; and
- Cost

  $C = V / (16\ \text{kr/MB})$ (6.4)

  per month, assuming a typical professional user price for mobile data traffic.

Table 6.1. Different performance parameters of combinations of sampling and reporting intervals avoiding segmentation.

t_S	1 s	4 s		10 s		30 s			1 min		
t_R	1 min	1min	5 min	1 min	5 min	1 min	5 min	1 h	1 min	5 min	1 h
p	60	15	75	6	30	2	10	120	1	5	60
O	11 %	47 %	8 %	107 %	21 %	360 %	72 %	5 %	660 %	148 %	11 %
r	89 bps	29 bps	22 bps	17 bps	10 bps	12 bps	5 bps	3 bps	10 bps	3 bps	1 bps
V	9.2 MB	3.0 MB	2.2 MB	1.7 MB	1.0 MB	1.3 MB	0.5 MB	0.3 MB	1.1 MB	0.4 MB	0.2 MB
C	147 kr	48 kr	36 kr	27 kr	16 kr	20 kr	8 kr	5 kr	17 kr	5 kr	2 kr

As seen from Table 6.1, longer reporting intervals (implying more positioning samples per report) lead to less overhead, less volume and less cost. Short sampling intervals (implying a high resolution in time and thus also in space) mean comparably large required capacities, volumes and cost. When looking for a high time resolution combined with moderate cost, the combination $t_S = 4$ s and $t_R = 5$ min provides a decent compromise with regards to overhead, capacity needs, volume and cost. In case of a low time resolution ($t_S = 30$ s...1 min), "streaming" data once per hour reaches a high degree of resource use at comparably very low capacity needs, volume and cost.

6.1.5 Capacity considerations

The combination of $t_S = 1$ s and $t_R = 1$ min yields the maximal required capacity. For $N = 100$ HGVs, it basically matches the maximal available capacity of a GPRS carrier when one time slot and Coding Scheme 1 is available. All other combinations need less capacity. However, due to resource sharing effects, the actually available GPRS capacity might be pretty much smaller, say some few kbps, which is basically reached when using $t_S = 4$ s. Several GPRS carriers, time slots or even networks alleviate this problem.

For 100000 HGVs, we immediately can see that no matter of the reporting or sampling interval, the throughput is upper-bounded by 10 Mbps, which can be considered as a rather modest load on a VPN link.

However, these decent capacity values merely apply if the momentary throughput is found pretty close to its average value. Consider the case of $t_S = 4$ s and $t_R = 5$ min. $N = 100$ HGVs generate 2.2 kbps during 5 min (300 s). If all of them would like to get their data through during the same second, the required speed would raise by factor 300 to 660 kbps during that second. Considering the server side (and thus 100000 HGVs), the corresponding peak capacity requirement would amount to 660 Mbps during that particular second, which would have the same effect on a server or a link as a heavy Denial-of-Service attack.

6.1.6 Recommended sending pattern

In order to avoid such traffic peakedness, it might be advisable to make different OBUs send at different times in a controlled *proactive* way. As the OBUs are connected to GNSS, a time reference is available. The placement of the OBUs and their identity numbers within the country and within the reach of certain access points is a pseudo-random process. Assume a minimal reporting interval covers j seconds. We recommend to split this interval into j slots to let OBU i send in slot (i mod j) $\in \{0, 1, \ldots, j - 1\}$. This way, we end up with a pseudo-random discrete-uniform-type distribution of time slots in the access networks and even on the link towards the server.

Yet another *reactive* possibility consists in using a so-called Random Backoff Scheme, *cf.* amongst others [4], which aims at resolving overload conditions by making the unlucky station re-sending the data after a randomized (typically exponentially distributed time). The effects on the required capacity will be described in the next section. However, Random Backoff is known to have an inherent risk of instability in case of high offered loads, which is due to the reactive nature of the method.

6.1.7 Impact of compression and loss

Until now, the case study has not taken the effects of compression and data loss into account. These will now be considered in the form of correction factors to the required capacity estimations.

The reduction factor $f_{compr} \in]0, 1]$ yielded by compression is highly dependent on the properties of the data to be compressed and on the capabilities of the compression algorithm. As a simple approximation, we can assume that the factor simply scales the length of the positioning messages. Consequently, in Equation (5.2) or (A.1), l_{pos} is to be replaced by $f_{compr} \cdot l_{pos}$. Typical compression factors lie in the order of magnitude of 50 %. Table 6.2 illustrates the impact of the compression factor on the relative overhead *O* (6.2), the required capacity *r* per HGV (6.1), the volume per month *V* (6.3) and the cost per month *C* (6.4), all for t_S = 4 s and t_S = 5 min.

Table 6.2. Different performance parameters as a function of the compression factor.

f_{compr}	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2
length	10 B	9 B	75	6	30	2	10	120	1
O	8 %	11 %	11 %	12 %	16 %	18 %	21 %	33 %	47 %
r	22 bps	20 bps	18 bps	16 bps	14 bps	12 bps	10 bps	8 bps	6 bps
V	2.2 MB	2.1 MB	1.8 MB	1.6 MB	1.4 MB	1.2 MB	1.0 MB	0.8 MB	0.6 MB
C	36 kr	33 kr	29 kr	26 kr	23 kr	20 kr	16 kr	13 kr	10 kr

Basically, we observe a linear downscaling of the capacity, roughly as $f_{compr} \cdot C(N)$. However, the relative overhead increases as the compression factor increases.

Similarly, the effect of loss can be taken into account through a resend factor $f_{resend} > 1$, which scales the required capacity in a similar fashion as described above ($f_{resend} \cdot C(N)$). Compensation for 10 % loss means $f_{resend} \geq 1.1$ and appears roughly as prolonged messages. Unfortunately, resending is likely to increase the burstiness of the traffic, which imposes the need for some extra spare capacity and implies the risk of instability in case of heavy loss.

6.2 Bulk transfer scenario

When HGVs cannot stream their data, they need to save the positions to report until there is an occasion for a bulk transmission. Such an occasion might be (1) a necessary break during which a HGV stands still at a parking place with good mobile or maybe even WLAN connectivity; (2) the need to report outstanding positions when leaving the country, which in Sweden means (a) entering a ferry terminal, (b) crossing a bridge, or (c) using one of the few land customs facilities to Norway or Finland, which also can be covered by mobile or wireless networks. So the main dimensioning issue for bulk transfer addresses the capacity *C*(*N*) required by *N* HGVs to complete their transmission of positioning data within a certain *time horizon* t_H, which then has to be related to the offerings of mobile and wireless networks. In order to avoid misunderstandings, it has to be pointed out that the time horizon has nothing to do with the concept of grace time, which denotes the allowed latency of positioning information from the viewpoint of the provider. Meeting a certain time horizon is rather important for the functioning of the system, which will be illustrated in this and the next section.

Figure 6.4 illustrates the relationship between t_S, t_R and t_H. The different colours symbolize data from different reporting intervals that are to be sent in the corresponding next interval. Obviously, the *time horizon* t_H is reasonably upper-bounded by the reporting interval t_H.

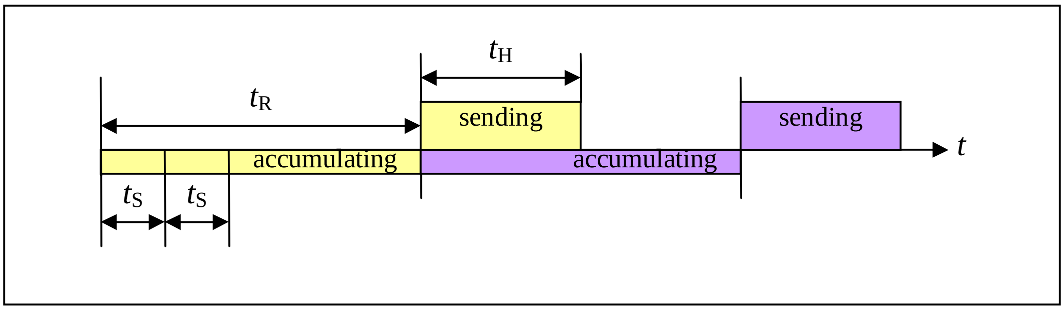

Figure 6.4. Relationship between sampling interval t_S, reporting interval t_R and time horizon t_H.

As outlined in Section 3.2, bulk transfer will use TLS/SSL and TCP. Similarly to the streaming case, the application data consists of $p = t_S / t_R$ positioning samples (5.1), for which we assume a length of l_{pos} = 10 B, plus the ID à 4 B. The calculation of the length L of the positioning message is discussed in Appendix B.

In the following, we consider the following parameters: (a) reporting interval $t_R \in$ {1 min, 5 min, 1 h, 8 h, 1 d, 3 d}; (b) sampling interval $t_S \in$ {1 s, 4 s, 10 s, 30 s, 1 min}.

6.2.1 Time horizon

On the server side, the number of simultaneous connections might be an issue, which is determined by the time horizon t_H. Assuming 100000 HGVs and a uniform distribution of bulk transfers during the day, we obtain the following figures:

- for $t_R = t_H$ = 1 min or 5 min: 100000 simultaneous connections (cf. streaming)[7];
- for t_R = 1 h and t_H = 15 min: 25000 simultaneous connections[8];
- for t_R = 8 h and t_H = 15 min: 3125 simultaneous connections;
- for t_R = 1 day and t_H = 15 min: 1042 simultaneous connections;
- for t_R = 3 days and t_H = 15 min: 347 simultaneous connections.

Only the last two values seem to be reasonable with regards to server scalability; the third case could be handled when installing a server park. In the sequel, we consider an adaptation of the time horizon such that a rather unproblematic number of simultaneous connections (here 500) is not surpassed:

- for t_R = 1 min, this means t_H = 300 ms, which is too short;
- for t_R = 5 min, this means t_H = 1.5 s, which is hardly possible in practice;
- for t_R = 1 h, this means t_H = 18 s, which is still pretty short;
- for t_R = 8 h, this means t_H = 144 s = 2 min 24 s;
- for t_R = 1 day, this means t_H = 432 s = 7 min 12 s;
- for t_R = 3 days this means t_H = 1296 s = 21 min 36 s.

As the check-in time at a ferry terminal typically spans one hour, the latter values seem to be feasible in any case. On the other hand, a land border passing might imply the need for a shorter time horizon (e.g. 1 min).

Taking the time horizon instead of the reporting interval into account, we obtain the required capacity as follows:

$$r' = L / t_H \geq r, \qquad (6.5)$$

with L calculated as described in Appendix B.

6.2.2 Bulk transfer performance and cost

The following tables illustrate performance and cost arising from different parameter choices for the bulk transfer scenario.

[7] As the time horizon matches the reporting interval, everyone has to report continuously.
[8] Here, the time horizon is only a quarter of the reporting interval, which leads to a decrease of the number of simultaneous connections by factor four.

Table 6.3. Relative overhead *o* (cf. Equation (6.2)) for different combinations of reporting intervals/time horizons and sampling intervals; results for non-practicable time horizons in brackets.

t_S \ t_R (t_H)	1 min (0.3 s)	5 min (1.5 s)	1 h 18 s	8 h 2 min 24 s	1 day 7 min 12 s	3 days 21 min 36 s
1 s	(94 %)	(22 %)	4 %	3 %	3 %	3 %
4 s	(379 %)	(75 %)	9 %	4 %	3 %	3 %
10 s	(937 %)	(187 %)	18 %	5 %	4 %	3 %
30 s	(2850 %)	(570 %)	47 %	8 %	5 %	4 %
1 min	(5740 %)	(1144 %)	94 %	14 %	6 %	4 %

Table 6.3 reveals that the overhead rises as reporting intervals decrease (which implies more re-connections and corresponding overhead) and sampling intervals increase (which implies less data per connection). Especially in the lower left corner, quite unrealistic values appear. For t_R = 1 h, the overhead can be significant if long sampling intervals (t_R = 30 s…1 min) are applied.

Table 6.4. Total message lengths *L* (cf. Appendix B) for different combinations of reporting intervals/time horizons and sampling intervals; results for non-practicable time horizons in brackets.

t_S \ t_R (t_H)	1 min (0.3 s)	5 min (1.5 s)	1 h 18 s	8 h 2 min 24 s	1 day 7 min 12 s	3 days 21 min 36 s
1 s	(1.14 kB)	(3.56 kB)	36.7 kB	290 kB	869 kB	2.55 MB
4 s	(718 B)	(1.28 kB)	9.58 kB	72.9 kB	218 kB	652 kB
10 s	(622 B)	(872 B)	4.14 kB	29.4 kB	87.4 kB	261 kB
30 s	(590 B)	(670 B)	1.72 kB	10.2 kB	29.4 kB	87.4 kB
1 min	(574 B)	(622 B)	1.14 kB	5.35 kB	15.0 kB	43.9 kB

Table 6.4 illustrates that short sampling intervals in combination with long reporting intervals (upper-right corner) yields large files, which needs long stationary coverage in order to be transmitted in an efficient way, and thus might be difficult to be sent via a weak mobile link. On the other hand, long sampling intervals in combination with short reporting intervals (lower-left corner) implies quite small amounts of data, avoiding segmentation on the TCP level as in the streaming case. Here, a lot of effort is spent on establishing a connection for pretty little data.

Table 6.5. Required capacity *r'* per HGV (cf. Equation (6.5)) for different combinations of reporting intervals/time horizons and sampling intervals; results for non-practicable time horizons (in brackets) suppressed.

t_S \ t_R (t_H)	1 min (0.3 s)	5 min (1.5 s)	1 h 18 s	8 h 2 min 24 s	1 day 7 min 12 s	3 days 21 min 36 s
1 s	N/A	N/A	16.7 kbps	16.5 kbps	16.5 kbps	16.5 kbps
4 s	N/A	N/A	4.36 kbps	4.15 kbps	4.13 kbps	4.12 kbps
10 s	N/A	N/A	1.88 kbps	1.68 kbps	1.66 kbps	1.65 kbps
30 s	N/A	N/A	781 bps	578 bps	558 bps	552 bps
1 min	N/A	N/A	518 bps	304 bps	284 bps	278 kps

Table 6.5 reveals that the required capacity per HGV is mainly a function of the sampling interval, which determines the gross amount of data to be sent during the day. As the time horizon has been adapted to the reporting interval such that, on average, 500 connections exist simultaneously, the size of the reporting interval plays a subordinate role regarding the capacity requirements. Given a certain sampling interval, the differences in required capacities correlate with the different amounts of overhead, cf. Table 6.2.

Table 6.6. Estimated cost C per HGV (cf. Equation (6.4)) for different combinations of reporting intervals/time horizons and sampling intervals; results for non-practicable time horizons in brackets.

t_S \ t_R / t_H	1 min (0.3 s)	5 min (1.5 s)	1 h 18 s	8 h 2 min 24 s	1 day 7 min 12 s	3 days 21 min 36 s
1 s	(256 kr)	(160 kr)	138 kr	136 kr	136 kr	136 kr
4 s	(158 kr)	(58 kr)	36 kr	34 kr	34 kr	34 kr
10 s	(137 kr)	(38 kr)	16 kr	14 kr	14 kr	14 kr
30 s	(130 kr)	(29 kr)	6 kr	5 kr	5 kr	5 kr
1 min	(126 kr)	(27 kr)	4 kr	3 kr	2 kr	2 kr

As in the streaming scenario, we assume a price of 16 kr/MB for mobile connectivity (and limit our attention to the uplink traffic). Similarly to Table 6.5, Table 6.6 shows a primary dependency of the results on the duration of the sampling interval and a minor dependency on the duration of the reporting interval due to the overhead. Comparing Table 6.6 to Table 6.1, we can also see that bulk transfer is slightly more expensive than streaming, which is due to the increased amount of overhead per packet and for the handshakes.

6.2.3 Capacity considerations

The total required capacity for N HGVs uploading their data simultaneously within their time horizon is obtained from a multiplication of the values presented in Table 6.5 by factor N. The corresponding results are shown in Figure 6.5.

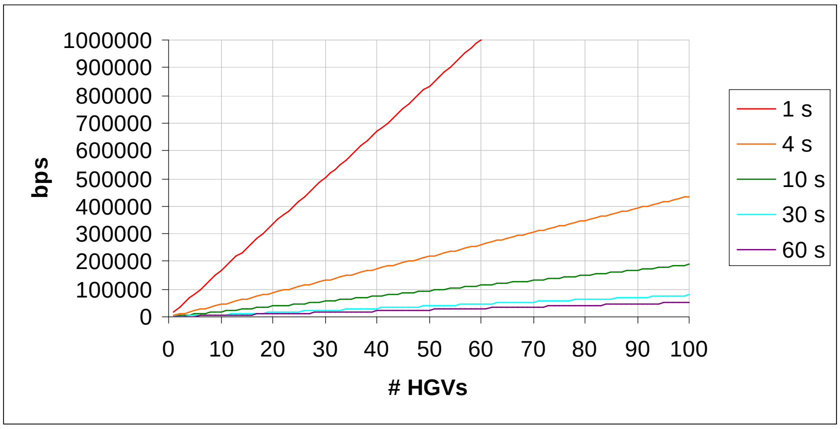

Figure 6.5. Required capacity versus the number of HGVs for a bulk transfer with a reporting interval t_R = 1 h and various sampling intervals.

Given a sampling interval of 1 s, we reach 1 Mbps for 60 HGVs; for a sampling interval of 4 s, we reach almost 450 kbps for 100 HGVs. A WLAN scenario is supposed to be able to handle such capacity demands, in particular because the geographical spread of HGVs implies the need for several access points, offering some tens of Mbps in total.

On the other hand, mobile capacity that typically spans up to several tens of kbps may not be sufficient in these cases involving many HGVs in one spot reporting at the same time. If there were some few HGVs reporting, they might get own codes assigned in UMTS, yielding guaranteed throughput up to of 64 kbps per HGV. Comparing this to the values reported in Table 6.4, we can see that such a capacity assignment reduces the use of the time horizon by at least factor four. In other words, the HGV uploads and thus releases its connection quicker, which is positive from the viewpoint of server scalability. Some few HGVs could even be handled by GPRS in case large sampling intervals (30 s…1 min) were used.

6.2.4 Recommended sending pattern

Let us turn back to the problem of simultaneous reporting during the time horizon. If 100 HGVs were to report one hour of positioning data each, the time horizon is given by 18 s, see Section 6.2.2. Obviously, those 100 HGVs do not need to report at the same time; indeed, their total time horizon amounts to 1800 s, i.e. half an hour, within one hour. That means that if the HGV bulk transfers were perfectly synchronized in time, only one HGV was reporting at a time at maximum. Such a feature would open up for the use of mobile technology (mainly UMTS due to better capacity offerings) even in hotspots. 100000 HGVs and a limitation to say 500 simultaneous connections require 200 time slots within the reporting interval. Following the approach presented for streaming traffic in Section 6.1.6, we let OBU i send in slot (i mod 200) $\in \{0, 1, \ldots, 199\}$. Again, we end up with a pseudo-random discrete-uniform-type distribution of time slots in the access networks and even on the link towards the server, balancing the load in time and space. As outlined in Section 6.1.6, a potential alternative consists in using a Random Backoff Scheme in order to cope with temporal overload situations.

6.2.5 Connection admission control

The sending pattern approach described above helps to control the spread of data uploads in time, however it cannot fully prevent the server from overload due to the fact that some HGVs may have accumulated data during much longer time than given by their reporting interval due to coverage problems. On the server side, a simple connection admission control (CAC) could be implemented by breaking TCP handshakes as a certain number of TCP connections are exceeded. An HGV which gets blocked by the CAC should be given the opportunity to report even outside of its designated interval, which is similar to the Random Backoff Scheme. However, the stability of such a concept needs to be studied in detail.

6.2.6 Impact of compression and loss

The impact of compression and loss in terms of capacity is basically comparable to the streaming case, cf. Section 6.1.7. Some particularities need special attention:

- The impact of compression is biggest in those cases having a comparably small amount of overhead, i.e. large reporting intervals and rather short sampling intervals
- TCP reacts to loss with a reduced sending speed. This prolongs the transfer process, which means more ongoing connections at the server side.
- If many TCP connections share the same bottleneck, unfairness in terms of throughput can arise and (unnecessary) loss may arise. Therefore, in the access networks, the number of TCP connections using the same (mobile or wireless) access point should be limited.

6.3 Conclusions and recommendations

From the results presented in Sections 6.1 and 6.2, we can draw a couple of conclusions:

- Streaming is slightly cheaper than bulk transfer, which is due to less overhead. However, they complement each other; data that cannot be streamed immediately can be stored in the OBU for later bulk transfer at a place with good stationary connectivity (e.g. a parking lot well covered by UMTS, or a ferry terminal offering WLAN connectivity).
- Both streaming and bulk transfer can handle even small position sampling intervals (in the order of seconds). Larger position sampling intervals (in the order of a minute) limit the traffic to such small rates that even GPRS seems to be able to handle the bulk transfer.

- “Heavy” uncoordinated data dumping is only advisable in WLAN hotspots involving several access points. On the other hand, “light” dumping by one or some few HGVs via UMTS and GPRS might work well, given that the coverage is stable.
- In general, the position sampling intervals should be much smaller than the reporting intervals, allowing for a large number of positions per report and thus reducing the capacity-consuming and cost-driving overhead per report. “Good” combinations are for instance a sampling interval of 4 s, combined with a reporting interval of 5 min, or a sampling interval of 30 s, combined with a reporting interval of 1 h.
- The most critical and limiting factor for bulk transfer is the number of parallel sessions at the server, which has to be kept within reasonable limits. Over-dimensioning is no way out, as it still cannot prevent peak loads. Some mechanism controlling the sending pattern or blocking excessive requests will help to avoid overload effects on the server side.
- In order to warrant a continuous flow of data and to avoid excessive resource competition in the access networks and at the server, both streaming and bulk transfers need some kind of time slots assigned to different HGVs, during which the transmissions are supposed to take place. Such time slots could be based upon the identification number of the OBU, leading to a pseudo-random distribution of transmission time slots both across the access networks and towards the server.

From these results, we issue the following recommendations with regards to system design and future work:

- Streaming should be the preferred way of sending positioning data and should be applied whenever possible. It allows for shorter dead times in the payment control loop. Moreover, data that has been streamed successfully does not bother the OBU any more. Accumulated data on the OBU must be seen as some kind of debt within the system. As this debt is growing over time, the need for good connectivity rises in order to “get rid of it” in reasonable time without stressing the system, e.g. by an unnecessarily long connection time towards the server.
- An investigation of the effects of the introduction of time slots for sending data could be studied by means of discrete-event simulation, modeling the data transfers as flows and investigating the load of access networks and server over time.
- Another simulation project might address “heavy data dumping” at a ferry terminal. Here, we could investigate differences between WLAN and UMTS and also investigate the effects of different system parameters (e.g. reporting interval) and control actions (e.g. the introduction of transmission time slots).
- Finally, the implementation of the real system should allow for detailed measurements of network performance and for the collection of end-to-end statistics regarding success and failure of position reporting.

7 Summary and outlook

This study presented an approach to dimension a Swedish RUC system, given the use of off-the-shelf wireless and mobile networks for reporting HGV positions towards a centralized server. For this purpose, two main scenarios were considered, streaming and bulk transfer, both augmented by authentication and encryption facilities. The dimensioning focused on the capacity requirements per access network for a given number of HGVs. The magnitude of the latter had to be estimated from traffic counts and the assumption of cell sizes in mobile networks, because such information is classified by the operators.

From the quantitative results presented, it can be seen that both streaming and bulk transfer are feasible, both in terms of network traffic and related cost. Streaming has the advantage that it avoids piling up data in the HGV and should be used as much as possible. However, some important boundary conditions need to be observed: If many HGVs are involved, uncoordinated streaming can induce overload on (low-bandwidth) access networks. Consequently, streaming should be limited to cases in which the number of reporting HGVs is rather small. During times of coverage problems, positioning data should be accumulated and transferred (as bulk) when the conditions improve. Lengthy bulk transfers, on the other hand, imply the risk of overloading the server in terms of open connections, which can seriously lower the server's performance. In the latter case, it has to be made sure that the bulk transfer is finished as quickly as possible, giving room to other bulk transfers. This puts forward the demands for stable connectivity and a high-bandwidth access network if many HGVs are involved. Natural collection and meeting points (parking lots; restaurants; ferry check-ins; toll stations) could be equipped with WLAN access for this purpose.

Capacity considerations on different time scales have however revealed a non-negligible risk. Too many simultaneous, uncoordinated streams and/or bulk transfers might turn the different parts (access network; server) into bottlenecks. Using simple "over-provisioning", this risk can be reduced, but not mitigated. A potentially wiser way to deal with this challenge is to try to minimize the risk of too many simultaneous connections by assigning suitable sending intervals for both streaming (here valid for each data packet) and bulk transfers (telling when the latter is to start). As GNSS is a vital ingredient of the solution, a time reference signal is available in each OBU and can serve as basis for a simple overload protection mechanism, in which each OBU uses a predetermined time slot in a given time interval to send the positioning data.

Future work should address the validation of the quantitative results and proposals derived in this study. This can be done using simulations ahead of implementation and real-world measurements once such a system would be available. Still, the estimations of the number of HGVs, based on statistical properties of HGV traffic, need to be improved, and corresponding traffic models need to be built. This puts the scene for future research work regarding the dimensioning of the Swedish RUC system. The outcomes of this report are expected to provide aids in determining key parameters of the RUC system, and they should also be discussed with potential service providers and network operators.

Acknowledgements

The authors would like to thank the participants of the ARENA project (www.arena-ruc.com), and in particular Ulrik Janusson and Thomas Sjöström from SWECO AB and Michael Forss from Netport, for fruitful discussions that helped to shape the contents of this document.

References

[1] J. Sundberg, U. Janusson, and T. Sjöström. A kilometre tax for heavy goods vehicles in Sweden – A conceptual systems design. Part 2: Proposal for systems design. Version 0.93, 2007-01-18.

[2] J. A. Persson, P. Davidsson, M. Boldt, B. Carlsson, and M. Fiedler. Evaluation of road user charging systems: the Swedish case. Proceedings of the 14th World Congress of Intelligent Transportation Systems (ITS 2007), Beijing, China, October 2007.

[3] M. Fiedler, S. Chevul, L. Isaksson, P. Lindberg, and J. Karlsson. Generic communication requirements of ITS-related mobile services as basis for seamless communications. Proceedings of NGI 2005 (published by IEEE), Rome, Italy, April 2005.

[4] J. Schiller. *Mobile Communications.* 2nd edition. Addison Wesley, 2003.

[5] M. Fiedler, L. Isaksson, S. Chevul, J. Karlsson, and P. Lindberg. Measurement and analysis of application-perceived throughput via mobile links. *The Olga Casals Lecture at HET-NETs 05* (*Tutorial T06*), Ilkley, UK, July 18–20, 2005.

[6] GSM world - 3GSM Platform. http://www.gsmworld.com/technology/3g/evolution.shtml (last seen Februari 10, 2008).

[7] S. Chevul. *On Application-Perceived Quality of Service in Wireless Networks*. Licentiate Thesis 2006:11. Blekinge Institute of Technology, ISBN: 91-7295-096-X, December 2006.

[8] G. Bianchi. *Performance Analysis of the IEEE 802.11 Distributed Coordination Function*. IEEE Journal on Selected Areas in Communications, 18(3):535–547, March 2000.

[9] IETF RFC 2960. Stream Control Transmission Control. http://www.ietf.org/rfc/rfc2960.txt

[10] IETF RFC 4346. The Transport Layer Security (TLS) Protocol. http://www.ietf.org/rfc/rfc4346.txt

[11] IETF RFC 2818. HTTP over TLS. http://www.ietf.org/rfc/rfc2818.txt

[12] Extensible Markup Language (XML).http://www.w3.org/XML/ (last seen February 10, 2008)

[13] M. Fiedler and K. Tutschku. Application of the stochastic fluid flow model for bottleneck identification and classification. *Proceedings of 2003 Design, Analysis, and Simulation of Distributed Systems* (*DASD 2003*), Orlando, FL, April 2003, pp 35–42.

[14] M. Hussain and S. Mahmood. Secure Software and Communication on Handheld Devices. M.Sc. Thesis in Computer Science MCS-2007:04, Blekinge Institute of Technology, Sweden, March 2007.

[15] J. Guitart, D. Carrera, V.Beltran, J. Torres, and E. Ayguadé. Session-Based Adaptive Overload Control for Secure Dynamic Web Applications. Proceedings of the 34th International Conference on Parallel Processing, Oslo, Norway, June 14–17, 2005, pp 341–349.

[16] V.Beltran, J. Guitart, D. Carrera, J. Torres, E. Ayguadé and J. Labarta. Performance Impact of Using SSL on Dynamic Web Applications. XV Jornadas de Paralelismo, Almeria, Spain, Sept 15–17, 2004, pp 471–476.

[17] C. Coarfa, P.Druschel, and D. Wallach. Performance Analysis of TLS Web Servers. ACM Transactions on Computer Science, Vol. 24, No.1, Feb 2006, pp 39–69.

[18] H. Johnson. Toward Adjustable Lightweight Authentication for Network Access Control. Ph.D. Thesis 2005:09, Blekinge Institute of Technology, School of Engineering, Dept. of Telecommunication Systems, ISBN 91-7295-077-3, December 2005.

[19] Vägverket Region Skåne Trafikinformation, http://193.181.63.7/ (last seen February 10, 2008)

[20] IETF RFC 3874. A 224-bit One-way Hash Function: SHA-224. http://www.ietf.org/rfc/rfc3874.txt

[21] B. Schneier, Description of a New Variable-Length Key, 64-bit Block Cipher (Blowfish). Fast Software Encryption 1993:191–204.

Appendix

A Calculation of *L* in the streaming scenario

The length L of a positioning message is obtained as follows: The application data consists out of p positioning samples, for which we assume a length of l_{pos} = 10 B, and an identification (ID) number of length 4 B. Authentication is performed by the SHA-224 algorithm [20], which adds 28 B. As encryption, we assume the block cipher Blowfish [21] that needs to apply padding to multiples of 16 B. Finally, UDP overhead amounts to 8 B and IP overhead to 20 B, respectively.

Regarding the limit of the number of positioning samples in order to avoid segmentation, we obtain the following results: Given a typical Maximum Transfer Unit of 1500 B, we arrive at a maximal UDP payload size of 1472 B, which means 92 cipher blocks à 16 B. Substracting 28 B authentication digest and the ID à 4 B, we find room for exactly p_{max} = 144 positioning samples à 10 B.

In the absence if segmentation, we can express the length of the whole message, including all overheads, by

$$L = \mathrm{ceil}(p \cdot l_{pos} + 32\ \mathrm{B})_{16\ \mathrm{B}} + 28\ \mathrm{B} \leq 1500\ \mathrm{B} \qquad \text{with} \quad l_{pos} = 10\ \mathrm{B} \qquad \text{(A.1)}$$

where $\mathrm{ceil}(x)_y$ expresses rounding up the value x to integer multiples of y.

B Calculation of *L* in the bulk transfer scenario

Similarly to the streaming case, the application data consists of $p = t_S / t_R$ positioning samples, for which we assume a length of l_{pos} = 10 B, plus the ID à 4 B. The TLS/SSL MAC amounts to 28 B (SHA-224), and a Blowfish encryption implies a padding to multiples of 16 B. The volume of the TLS/SSL handshake is hard to quantify; we assume 400 B on the uplink. Furthermore, TLS/SSL applies a quantisation of 16 KB per PDU, i.e. longer messages are split into blocks of 16 KB at maximum. Each TLS/SSL PDU requires 5 + 1 B overhead. On the TCP level, segmentation takes place; given a typical MTU size of 1500 B, we obtain a TCP segment size of 1460 B. Each TCP segment implies 20 B TCP overhead and 20 B IP overhead. In contrast to the streaming scenario, segmentation is not an issue; it is indeed almost unavoidable, given the potentially large amounts of data to be sent. Additionally TCP handshake is assumed to take two packets à 40 B on the uplink. Finally, we arrive at the length L of the positioning data, including all kinds of overheads.

List of ARENA reports

ARENA REPORT 2008:1. “Road User Charging of Heavy Goods Vehicles in Sweden”. Final report ARENA 1., NetPort.Karlshamn

ARENA REPORT 2008:2. Sundberg, J., Janusson, U., and Sjöström., “A kilometre tax for heavy goods vehicles in Sweden – A conceptual systems design. Part 1: Requirements and preconditions”., SWECO VBB

ARENA REPORT 2008:3. Sundberg, J., Janusson, U., and Sjöström., “A kilometre tax for heavy goods vehicles in Sweden – A conceptual systems design. Part 2: Proposals for systems design”., SWECO VBB

ARENA REPORT 2008:4. Sundberg, J., “A New Approach to Control in the ARENA concept for HGV kilometre tax in Sweden”., SWECO VBB

ARENA REPORT 2008:5. Hamilton, C J. “A market based approach to achieve EFC interoperability in Europe”., Policy Technology

ARENA REPORT 2008:6. Eliasson, C and Fiedler, M., “Dimensioning study for road user charging”. Blekinge Institute of Technology.

ARENA REPORT 2008:7. Boldt, M and Carlsson, B,. “Hotanalys för positionsangivelsekedjan”. Blekinge Institute of Technology.

ARENA REPORT 2008:8. Davidsson, P and Persson, J., “A Criteria-Based Approach to Evaluating Road User Charging Systems”.,Blekinge Institute of Technology

ARENA REPORT 2008:9. Sundberg, J., “PM kring legala frågeställningar”., SWECO VBB

ARENA REPORT 2008:10. Janusson, U., Berg, P and Udin, C., “ARENA DEMO”., SWECO VBB

ARENA REPORT 2008:11. Sundberg, J., PM kring kostnadsberäkning”., SWECO VBB

ARENA REPORT 2008:12. Forss, M., Gustafsson, I., and Källström, L., “ARENA RUC Seminar 1 & 2 – Summary of the seminars”., NetPort.Karlshamn

ARENA REPORT 2008:13
Published papers produced within the project

ARENA
NetPort.Karlshamn
Biblioteksgatan 4 • 37435 Karlshamn • Sweden

Project partners:
Swedish Road Administration • SWECO • BMT Transport Solutions • Blekinge Institute of Technology • NetPort.Karlshamn

www.arena-ruc.se