TeliaSonera

Streaming in Mobile Networks

White Paper

August 2004

MediaLab
TeliaSonera Finland

www.medialab.sonera.fi
info@medialab.sonera.fi

Contents

01. Introduction

The mobile phones of today have enough computing capacity, memory, and multimedia features for advanced multimedia applications such as playing audio-visual content. These features, together with packet-switched data services provided by GPRS, EDGE, and WCDMA networks and advanced compression algorithms have made streaming of audio-visual content to mobile phones possible.

Streaming is a method of transferring digital data with real-time characteristics in such a way that the recipient can view the content while receiving the data. The data can be basically any content, but in this paper we concentrate only on the streaming of audio and video content. The advantage of streaming compared to downloading is that it makes it possible for the recipient to start viewing the content almost immediately, and an entire file does not have to be downloaded and stored on the client device. On the other hand, the quality of a presentation is constrained by the underlying network.

Since streaming is sensitive to interruptions, it is important to understand the properties of the transport network when developing streaming services for mobile networks. This paper provides a brief introduction to 2G/3G mobile systems capable of streaming media content. These technologies include GPRS, EDGE, and UMTS mobile systems.

In mobile streaming, interoperability between different streaming components is very important. The components can be divided in three categories: servers, encoders, and players. To ensure interoperability, standardized file formats, codecs, and protocols are needed. Most major software providers have adopted standardized technologies instead of supporting proprietary solutions. The MPEG-4 standard can be used for file formats and codecs, and the 3GPP PSS standard for the entire mobile streaming framework. The 3GPP PSS streaming standard is introduced briefly in this paper.

This paper is based on two publications written at TeliaSonera Finland MediaLab: *Streaming Tester Software for Mobile Systems Master's thesis* [1] and *Packet Switched Streaming Service White Paper* [2]. The information on packet-switched mobile networks has been gathered mainly from [3], [4], and [5].

02. Packet-Switched Data in Mobile Networks

Global System for Mobile communications (GSM) is a digital mobile phone system taken into use in 1991. The focus was to develop a modern, standardized, digital mobile system to replace old analog systems incompatible with each other. The GSM technology family is constantly evolving, and there have been many additions to the basic technology. Today many of these improvements concern mobile data services and the mobile Internet. Figure 1 presents the architecture of a dual 2G/3G network built on GSM network technology.

Basics of GSM Technology

GSM is based on a cellular radio network architecture. In cellular networks, the whole coverage area is divided into numerous smaller regions called cells. A cell is basically defined as the geographical area in the radio coverage of one Base Transceiver Station (BTS). Mobile Stations (MS) (usually mobile phones) communicate with the mobile network through the radio interface between the MS and the BTS. Mobile stations can seamlessly move from one cell to another. The situation where a mobile station changes cells is referred to as a handover.

A GSM network can be roughly divided to Network and Switching Subsystem (NSS) and Base Station Subsystem (BSS). Mobile stations are connected to NSS through BSS's radio interface. NSS is usually furthermore connected to other networks such as a fixed-line telephone network.

BSS consists of numerous Base Transceiver Stations (BTS) controlled by Base Station Controllers (BSC). BTS is typically a radio antenna tower combined with a small shack containing the equipment. In rural areas, BTSs are typically located in high locations to maximize radio coverage and thus cell size. In dense urban areas, BTSs are often located on walls or rooftops to maximize cell capacity. BTSs are connected to BSCs using fixed-line connections. One BSC can control multiple base stations. BSC's function is to control the usage of radio interface resources in its area.

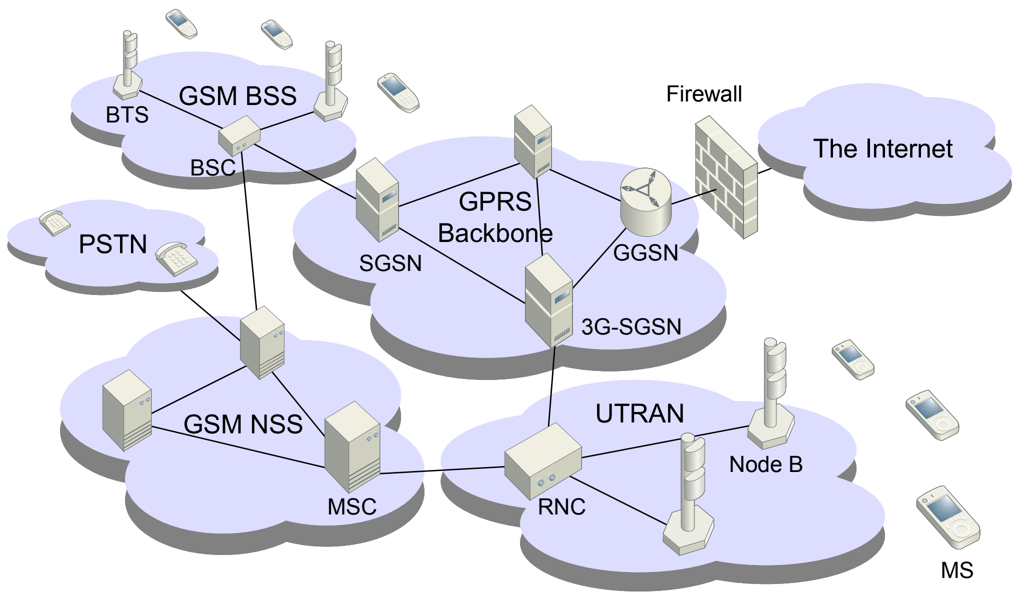

Figure 1. The 2G/3G dual mobile network architecture

NSS consists of Mobile services Switching Centers (MSC) and numerous registers related to it. Every BSC is connected to a MSC with a fixed-line connection or a radio transmission line. MSCs are responsible for switching and routing connections between mobile stations and devices in other networks. They also route and switch connections inside the GSM core network between MSCs, as well as in the MSC between different BSCs. MSCs can be connected to external networks such as the Public Switched Telephone Network (PSTN) or other operators' GSM networks.

There are four main versions of the GSM radio interface: GSM-900, GSM-1800, GSM-850, and GSM-1900. The number indicates the radio frequency (MHz) used. Europe and most of the world uses GSM-900 and GSM-1800, but in the United States and Canada, GSM-850 and GSM-1900 are used. Modern GSM phones usually support two or three frequency bands. Time-Division Multiple Access (TDMA) technology is used to share radio channels between multiple mobile stations. Radio channels are divided to frames which are furthermore divided to short timeslots (TS). Each user has their own timeslots and this way they can have access to the full channel bandwidth for a short period at a time.

Circuit-Switched Data Services

In addition to voice calls, GSM technology provides support for data calls. By using a Circuit-Switched Data (CSD) service, a GSM phone can be used like a modem to transfer arbitrary digital data. Because GSM is already a digital system, there is no need to modulate data from digital to analogue and back, like traditional modems do. Circuit switching means that after the data call has been established, the user has a synchronous data channel in continuous, exclusive use until either the caller or the recipient disconnects it. The bit rate for user data in GSM is 9.6 kbps.

Because the original bit rate of CSD is quite low, some improvements have been developed to achieve higher bit rates. By using High-Speed Circuit-Switched Data (HSCSD) technology, a mobile station can use multiple timeslots (usually 2 or 4) grouped to one logical channel. By using 4 timeslots, bit rates as high as 38.4 kbps can be achieved. In addition, the bit rate per one timeslot can be raised from 9.6 to 14.4 kbps by using enhanced channel coding technique which lowers the number of bits used for error correction. Therefore, the maximum bit rate using 14.4 kbps channel coding and 4 timeslots is 57.6 kbps.

Packet-Switched Data Services

Circuit-switched connections are well suited for voice calls, but when they are used for typical data services, some resources are usually wasted. This happens because in CSD every data channel is in the exclusive use of only one user, and there are no means to share it with others. Since typical data services (e.g. web browsing) do not fully utilize the channel most of the time, a lot of capacity is wasted. This is a disadvantage for both mobile operators and customers. Customers have to pay for the connection time, and the mobile network's resources are used even if no data is being transferred. Also, before transferring any data, a new connection must be opened by placing and establishing a data call, which takes at least a few seconds.

Typically, data services are used to transfer packet data such as Internet Protocol (IP). It is obvious that transferring packet data on a circuit-switched channel wastes network resources. To solve the major drawbacks of CSD, packet-switched data services were added to the GSM technology family. Packet switching means that transferred data is split to data packets that are individually routed and transferred between network nodes. There is no established connection between nodes like in CSD. Because the data is transferred in packets and no connections are established, network resources can be better shared between multiple users. Users can be billed only for actual the amount of transferred data and the network's resources are used more effectively. Transmission can also be started almost immediately when there is a need for it.

03. General Packet Radio Service (GPRS)

General Packet Radio Service (GPRS) is an add-on to existing GSM networks providing an option to use packet-switched protocols (usually IP) for transferring data. GPRS is focused only on transferring packet data, while the existing GSM network can still be used for circuit-switched voice and data calls. GPRS requires modifications and additions to mobile network elements, as well as support from mobile terminals. GPRS is an integral part of third-generation (3G) mobile networks, and it is used for both voice and data services in 3G mobile networks such as UMTS.

The main features of GPRS are speed, immediacy, and better utilization of network resources. Higher data rates are achieved using multiple timeslots simultaneously and using more efficient channel coding algorithms. Immediacy means that there is no need for dial-up procedures as when using CSD. Data can be transferred almost immediately upon need for it and the mobile station is in radio coverage. Users are not necessarily billed for connection time anymore, but by the actual amount of transferred data. Also, the usage of radio resources is more efficient and flexible since GPRS users can share timeslots left over from circuit-switched connections. This way GPRS also improves the peak time capacity of base stations.

GPRS Network Architecture

GPRS has been designed so that it can be added to an existing GSM network with as few changes to base stations as possible. This is important to network operators, as the base station subsystem represents a huge share of the network hardware. BTSs require usually only software updates and existing connections to BSCs can also be used with GPRS.

BSCs require both new software and hardware. Packet Control Unit (PCU) is a new unit in BSC used for separating packet data from circuit-switched connections, and forwarding them to the proper networks. Circuit-switched connections are still directed to the GSM network through MSC, while packet data is directed to the GPRS backbone network through a new network element called Serving GPRS Support Node (SGSN). The GPRS backbone is a high-speed packet-switched IP network connecting Serving GPRS Support Nodes (SGSN), Gateway GPRS Support Nodes (GGSN), operators' GPRS networks together.

SGSN is a GPRS network element on the same logical level as MSC in the GSM network. The primary function of SGSN is to deliver packets between the network and mobile stations within its service area. SGSN also tracks the mobile stations' locations in the network, processes registrations, and checks network access for mobile stations. SGSN is connected to the base stations subsystem by a Frame Relay connection between SGSN and the PCU unit of BSC. SGSN can be physically integrated to BSC.

GGSN is a gateway element used for connecting the GPRS backbone to external IP networks such as the public Internet, other operators' GPRS networks, or intranets. GGSN is shown to an external network as an IP router. GGSNs are separated from external networks by firewalls. GGSNs also maintain the routing information necessary to route packets to mobile stations through correct SGSN nodes.

GPRS Mobile Stations

New mobile terminals are required for taking advantage of GPRS, because old GSM phones are not able to handle packet data or the new enhanced air interface. All GPRS terminals must be backward compatible with GSM which they use for circuit-switched services. GPRS terminals are categorized to Class A, B, and C devices. Class A devices support simultaneous usage of GPRS packet data and circuit-switched voice calls. Class B devices can automatically switch between GPRS packet data and GSM circuit-switched services, but they cannot be used simultaneously for transferring data. GPRS is switched to busy mode while there is an active circuit-switched connection. For Class C devices, the user must select the used service manually and the other service is unreachable. Currently available terminals are mostly Class B devices. Most GPRS devices are mobile phones, but there are also more specialized devices such as PDAs with an embedded GPRS phone and card modems for laptops.

Channel Coding Schemes and Multislot Classes

GPRS provides data-rates up to a theoretical maximum of 171 kbps. The highest data-rates are achieved using up to eight timeslots simultaneously and channel coding algorithms with reduced error correction, when possible.

There are four different channel coding schemes called CS-1, CS-2, CS-3, and CS-4. The coding scheme used depends on the quality of radio link between the mobile station and the base station. CS-1 is used in poor radio environments, as it has the most effective error correction. CS-4 does not have any error correction and it can be only used in very good conditions. CS-2 is equivalent to the channel coding used for GSM CSD services. The bit rates for coding schemes are presented in Table 2.1. GPRS mobile stations must support all coding schemes, but for base stations only CS-1 is mandatory. Currently, mostly CS-1 and CS-2 are used in practice in actual mobile networks.

The device's multislot class determines the maximum number of timeslots for downstream, upstream, and total. For example, a device with multislot class 10 can utilize up to 4 timeslots for downstream, up to 2 for upstream, but only 5 timeslots can be used at a time. This makes configurations with 4+1 or 3+2 timeslots possible for downstream and upstream, respectively. The theoretical maximum of simultaneous timeslots is eight, but currently there are no devices capable of that. Using more timeslots requires more processing capability and power consumption also becomes higher. Devices with higher multislot classes are also more complex and thus more expensive.

Recent devices can typically utilize 3 or 4 timeslots for downstream and 1 or 2 for upstream. Using the normal CS-2 coding scheme, theoretical data rates up to 40.2 kbps (3 timeslots) and 53.6 kbps (4 timeslots) can be achieved. It must be noticed that these bit rates are not effective transfer rates for user data, but raw bit rates that can be transferred on a radio link using CS-2 channel coding. Bit rates for user data are lower due to packet overhead. Theoretical maximum bit rates for different coding schemes and timeslots are presented in Table 1. Currently uncommon combinations are greyed out.

Table 1. Theoretical maximum transfer rates for GPRS (kbps)

Timeslots	1 TS	2 TS	3 TS	4 TS	5 TS	6 TS	7 TS	8 TS
CS-1	9.05	18.10	27.15	36.20	45.25	54.30	63.35	72.40
CS-2	13.40	26.80	40.20	53.60	67.00	80.40	93.80	107.20
CS-3	15.60	31.20	46.80	62.40	78.00	93.60	109.20	124.80
CS-4	21.40	42.80	64.20	85.60	107.00	128.40	149.80	171.20

04. Enhanced Data Rates for Global Evolution (EDGE)

Enhanced Data Rates for Global Evolution (EDGE) is an upgrade to GSM/GPRS mobile networks improving the performance of the air interface between a mobile station and BTS. EDGE is not directly involved with the GPRS core network and thus only the base station subsystem requires updates. The improvement in speed can be utilized by both packet-switched and circuit-switched data services. These enhanced data services are called Enhanced GPRS (EGPRS) and Enhanced CSD (ECSD). EDGE is an important phase in the development of mobile networks towards real 3G networks, because it can be added to an existing network infrastructure with few modifications. It seems that EDGE is going to be used alongside 3G networks in the future, since it provides relatively high bit rates and wider radio coverage is easier and cheaper to achieve than on real 3G networks.

Differences to GPRS

EDGE is an add-on to the air interface between mobile stations and base stations. Adding EDGE to existing GSM/GPRS network requires software updates for BSCs and BTSs, and possibly hardware upgrades for BTSs depending on their model. The core network still behaves like in GPRS, and does not require any changes. New mobile equipment with EDGE support is required to take advantage of EDGE features. Mobile devices with EDGE support are backwards compatible with GPRS and GSM, and can use old services if EDGE is not available.

It is not necessary for operators to add EDGE support to the entire network, as it is possible to offer it only in cities and use normal GPRS in rural areas. EDGE not only improves the data-rate for each user, but also grows the network's capacity by allowing more users to share the same timeslots. It is also important that EDGE users can share timeslots with conventional GPRS users.

EDGE operates on the same radio frequency as GSM/GPRS, but its radio channel modulations and protocols are different. While conventional GSM/GPRS uses Gaussian Minimum Shift Keying (GMSK) for modulating digital information to the radio channel, EDGE uses more effective 8-Phase Shift Keying (8PSK) modulation. 8PSK is usually more efficient than GMSK, excluding very poor radio conditions.

While GPRS defined four coding schemes, EDGE defines nine different modulation coding schemes named MCS-1 to MCS-9. The four first schemes use GMSK modulation and are targeted for poor radio conditions offering data-rates from 8.8 kbps to 17.6 kbps per timeslot. The other five schemes use 8PSK modulation and offer data-rates from 22.4 kbps to maximum of 59.2 kbps per timeslot. Therefore, the theoretical maximum data-rate with eight timeslots would be 473.6 kbps.

05. Universal Mobile Telecommunications System (UMTS)

Universal Mobile Telecommunications System (UMTS) is one of third generation (3G) mobile systems defined in International Mobile Telecommunications 2000 (IMT-2000) standard by International Telecommunications Union (ITU). UMTS is designed especially for the needs of broadband mobile Internet and other packet-switched services offering data-rates up to 2 Mbps in optimal radio conditions indoor.

UMTS uses completely different radio interface than GSM/GPRS called Wideband Code Division Multiple Access (WCDMA). Mobile stations share the same radio frequency, and they are separated from each other by hash codes. Time division, such as in TDMA, is not used and thus there is no concept of a timeslot anymore. Because the WCDMA radio interface is not compatible with TDMA, UMTS networks require a completely different base station subsystem than GSM/GPRS networks. It is called UMTS Terrestrial Radio Access Network (UTRAN). However, UMTS uses the same GPRS core network as GSM/GPRS systems. UMTS compatible mobile stations are also backward compatible with GSM/GPRS and they can be used if UMTS is not available. UMTS also features handovers between UMTS and GSM/GPRS networks, so mobile stations can change networks without a noticeable break in the connection. It is believed that UMTS will be implemented only in cities in the first phase, while rural areas will be handled with existing GSM/GPRS/EDGE systems. Handover between the different network technologies is called inter-system handover (ISHO).

UMTS Features

One of the key features of UMTS is speed. UMTS improves data-rates clearly from earlier 2G mobile systems. UMTS target data rates are 144 kbps for rural outdoor areas, 384 kbps for urban outdoor areas, and 2048 kbps for indoor and close range outdoor areas.

One of the new features, Quality of Service (QoS), allows the network to provide different class of service for different needs. UMTS specifies four QoS classes. Conversational class is for highly interactive and real-time services with low response times and high throughput like voice, video telephony, and real-time gaming. Streaming class is for transferring multimedia content to mobile stations. Interactive class is for web browsing, non-real-time network gaming, and other services that do not need low response times. Finally, background class is for services that do not have strict demands for performance. By using QoS, network resources can be better shared between users taking their requirements into account. QoS classes can also be used as a basis for billing. Because higher QoS classes likely reserve more network resources and offer better user experience they can be priced differently.

UMTS Network Architecture

UMTS Terrestrial Radio Access Network (UTRAN) can be added to an existing 2G network in parallel to a GSM radio access network. UTRAN and GSM BSS share the same GPRS core network. The UTRAN architecture is quite similar to the GSM radio access network. UTRAN contains Radio Network Systems (RNS) each controlled by Radio Network Controller (RNC). RNC basically corresponds to BSC in GSM. RNCs are connected to MSC and 3G-SGSN. Normal SGSN cannot be used since RNC has a different interface from BSC. Base station in UTRAN is called Node B. Each RNC is connected to multiple base stations. Each base station can provide service for multiple cells.

06. Practical Data Rates

In order to create usable streaming services on mobile networks, developers must have a realistic view of network performance. Theoretical maximum values for GSM/GPRS transfer rates using different number of timeslots and different coding schemes are presented in Table 1. These values define the absolute upper limit for throughput, but they cannot be directly used for representing the actual data rate for user data. In addition to the number of timeslots and coding scheme used, throughput is highly affected by used protocols, packet sizes, and quality of radio channel.

It is possible to mathematically analyze the actual throughput by taking packet size and quality of the radio channel into account. Theoretical analysis of GPRS performance is presented in [6]. The paper includes a graph that represents throughput (kbps) as a function of radio channel quality (dB). In good radio conditions using CS-2 coding, the throughput for one timeslot is calculated to be around 11 kbps. It must be noticed that in real life, packet overhead reduces the bandwidth a little more. Realistic throughput can also be estimated by field tests which can be performed running actual applications (such as file transfer) and measuring their performance by using e.g. packet capturing software, or using special performance tester software.

In [7], GPRS performance was estimated by transferring a 1.2 MB file six times using file transfer protocol (FTP) on a GPRS device capable of utilizing 3 timeslots simultaneously. The paper includes a graph representing the amount of data (MB) as a function of time (s). It is hard to estimate maximum and minimum bit rates (slope) from the graph, but the average bit rate of the whole transmission (over 8.5 minutes) can be estimated to be around 20 kbps (around 6.7 kbps per timeslot). The paper also reveals that mobility has a significant effect on performance.

In [8], GPRS throughput was tested running dedicated performance tester software. Tests were run using a device capable of utilizing 2 timeslots. According to the paper, throughput varied between the minimum of 2.6 kbps and the maximum of 20 kbps with the average of 12.5 kbps (around 6.25 kbps per timeslot). This result seems to correlate well with the results presented in [7]. Also, the maximum bit rate of 20 kbps seems realistic when compared to the theoretical throughput (11 kbps per timeslot) presented in [6].

According to the references and our own field tests [1], a maximum throughput of around 10 kbps per timeslot can be used as a realistic estimate for user data in GPRS. It is assumed here that the data is transferred in good radio conditions using the CS-2 coding scheme. However, in streaming applications data is transferred over RTP/UDP/IP protocol stack, which causes some extra packet overhead. A typical packet overhead caused by packet headers is around 9% when average packet size is around 400 bytes. This makes 9 kbps per timeslot a realistic estimate of the maximum throughput for payload data in streaming applications.

EDGE and UMTS networks provide higher data rates than GSM/GPRS. According to our preliminary test results, media content can be streamed at 50 kbps to an EDGE device utilizing 2 time slots (at least MCS-6 channel coding) or to a UMTS device using 64 kbps data bearer. Higher bit rates require more EDGE timeslots or a more efficient data bearer in UMTS. Currently EDGE devices can utilize up to 4 timeslots and UMTS networks provide data bearers up to 384 kbps for downstream.

07. Streaming in Mobile Networks

The term *mobile streaming* is used if the content is streamed to a terminal over a mobile phone network. A terminal is usually a mobile phone, PDA or laptop with the packet-switched data capabilities and streaming media player software. Streaming over short-range wireless technologies such as Bluetooth or Wireless Local Area Network (WLAN) are not considered as mobile streaming in this paper. A typical mobile streaming architecture is presented in Figure 2. The architecture is basically the same as used in broadband streaming, but the clients are connected to the Internet and the streaming media server through a mobile network. A typical mobile streaming architecture is presented in Figure 2.

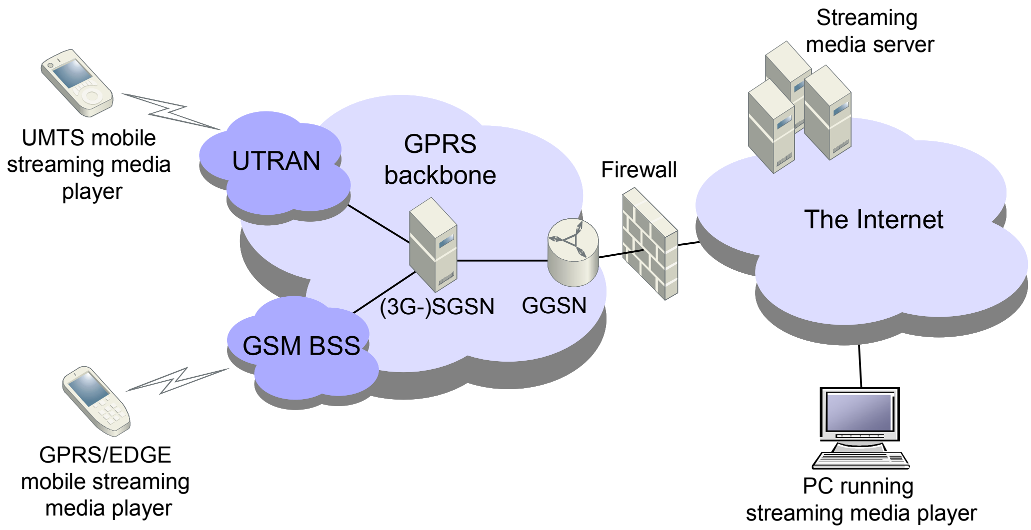

Figure 2. Typical mobile streaming architecture [1]

A stream is a flow of data packets containing media content. The packets are normally generated by a streaming media server from an arbitrary data source, which can be media content stored on the server or captured from a live source (e.g. camera, microphone, television broadcast, etc.). Streaming of previously stored data is called on-demand streaming, while streaming of live content is called live streaming or webcasting.

The content data is usually packed using a codec targeted for compressing such type of data. The bit rate of the stream specifies how much compressed payload data is sent in a time unit. Bit rates are typically measured in bits per second. The stream's bit rate can be either constant or variable. The used bit rate and codec highly affect to the quality of the encoded content.

The generated data packets are continuously sent to the recipient over a packet-switched network using some streaming protocol. The recipient is running streaming media player software, which receives the packets, decodes the content data with an appropriate codec, and finally shows the presentation to the user.

Streaming is sensitive to errors and delays in the transmission, because a continuous flow of data is required for an uninterrupted presentation. If some data packets are lost or delayed during the transmission, the media player may not be able to decode the data correctly, and some errors or interrupts may occur. For compensating possible delays, streaming media players usually receive some amount of packets before starting to play the content. This is referred as buffering. Buffering can also occur while streaming if the player runs short of the data due to lack of bandwidth. Therefore, properties of the underlying mobile network have an significant effect on streaming quality and reliability.

08. The Mobile Streaming Standard: 3GPP PSS

Third Generation Partnership Project (3GPP) is a collaboration agreement between a number of telecommunications standardization bodies established in 1998. The original scope of the project was to produce globally used technical specifications for 3G mobile networks evolved from GSM technology, but currently also the maintenance and development of the original GSM technology is included in the project. The information provided in this section is mainly based on [2].

3GPP Packet Switched Streaming (3GPP PSS) is a specification defining a framework for interoperable, end-to-end streaming services in packet-switched mobile networks. 3GPP PSS highly reuses the work done by the organizations such as The Internet Engineering Task Force (IETF), World Wide Web Consortium (W3C), Motion Picture Experts Group (MPEG), International Organization for Standardization (ISO), and International Telecommunication Union (ITU). Packet-switched streaming was first introduced in 3GPP Release 4.

3GPP PSS Release 4 (frozen in March 2001) defines a basic framework, protocols, codecs, and the 3GPP file format. 3GPP PSS itself does not specify the coding of content data, but utilizes already standardized codecs, formats, and data types. By 3GPP Release 4 specification, the AMR and MPEG-4 AAC codecs are used for encoding audio, and the H.263 and MPEG-4 codecs for encoding video. Release 6 will introduce some additional codecs. 3GPP PSS is widely supported by a majority of the streaming platform providers such as RealNetworks, Apple, and Packet Video. It is also implemented in many mobile phones on the market. This makes 3GPP currently the most important mobile streaming standard applicable today.

In this white paper, we concentrate on streaming protocols. The MPEG-4 white paper [9] and Packet-switched Streaming white paper [2] cover file formats and codecs.

Streaming Protocols

There are some protocols intended especially for streaming applications. Some of the protocols are targeted for initializing and controlling streaming sessions (RTCP, RTSP, SDP), while others are transport protocols for transferring the payload data (RTP). The 3GPP PSS protocol stack is presented in Figure 3. As seen in the figure, protocols are transported over the TCP/IP or UDP/IP protocol stack through the packet-switched mobile network and the Internet. It should be noted that the actual streaming of audio and video content must be done over UDP protocol. Streaming over TCP and HTTP is often used in the fixed-liner Internet, but it is not included in the 3GPP PSS standard.

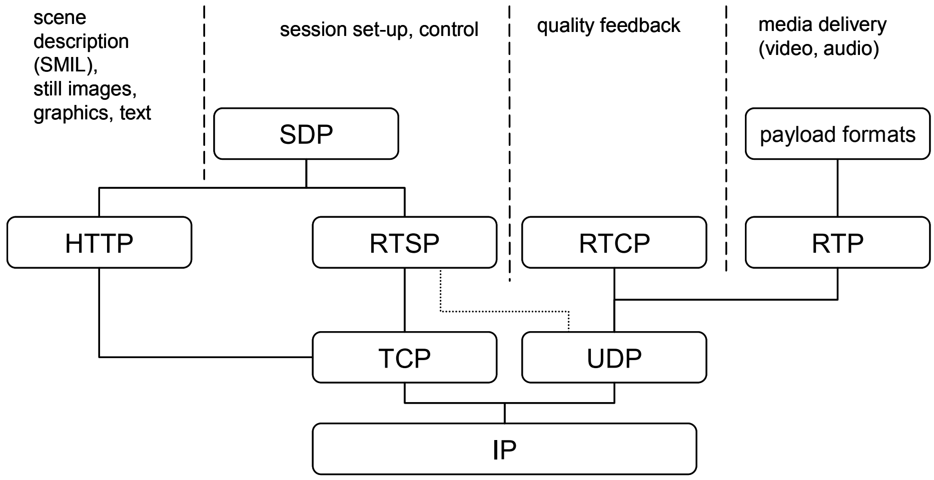

Figure 3. The 3GPP PSS protocol stack [2]

Real-Time Transport Protocol (RTP)

Real-time Transport Protocol (RTP) [10] provides functions for end-to-end transport of real-time data, such as audio, video, multimedia, or other content. RTP supports both unicast and multicast transmissions. RTP is only a transport protocol, and thus it does not guarantee any quality of service for the transported services. RTP is independent of the underlying transport protocols and networks, but when it is used for streaming audio and video content in IP networks, it is usually transferred over UDP/IP protocol.

Real-Time Control Protocol (RTCP)

Real-Time Control Protocol (RTCP) [10] is used in association with the RTP protocol to provide feedback on the quality of the transport, and for adding minimal identification and control functions. RTCP uses the same distribution channel as RTP, so the underlying transport protocol must provide some kind of multiplexing for the RTP data and the RTCP control packets. For example, over UDP this is typically done simply directing the RTP and RTCP packets to different UDP ports.

The primary function of RTCP is to provide feedback for other participants of the streaming session. Adaptive encoders and streaming servers can utilize the feedback information for adjusting the stream to match the current transport quality. Feedback is delivered in RTCP sender and receiver reports.

Real-Time Streaming Protocol (RTSP)

Real-Time Streaming Protocol (RTSP) [11] is an application-level protocol used for establishing and controlling either a single or several time-synchronized streams of continuous media content, such as audio and video. RTSP is not typically used for delivering the payload data itself, although it is possible. To interleave the payload data with RTSP, usually protocols such as RTP are used. Basically RTSP can be thought as a "network remote control" for multimedia servers. The basic operation of RTSP in association with RTP and RTCP is presented in Figure 4.

RTSP protocol is a text-based protocol resembling Hypertext Transfer Protocol (HTTP), but there are also some major differences. The biggest differences are that RTSP is not a stateless protocol like HTTP, since an RTSP server has to maintain its state in almost all cases, and that both the server and the client can issue requests. RTSP is highly independent of the used transport protocol and thus the RTSP session is not related to e.g. TCP, UDP or other connections used for the transportation.

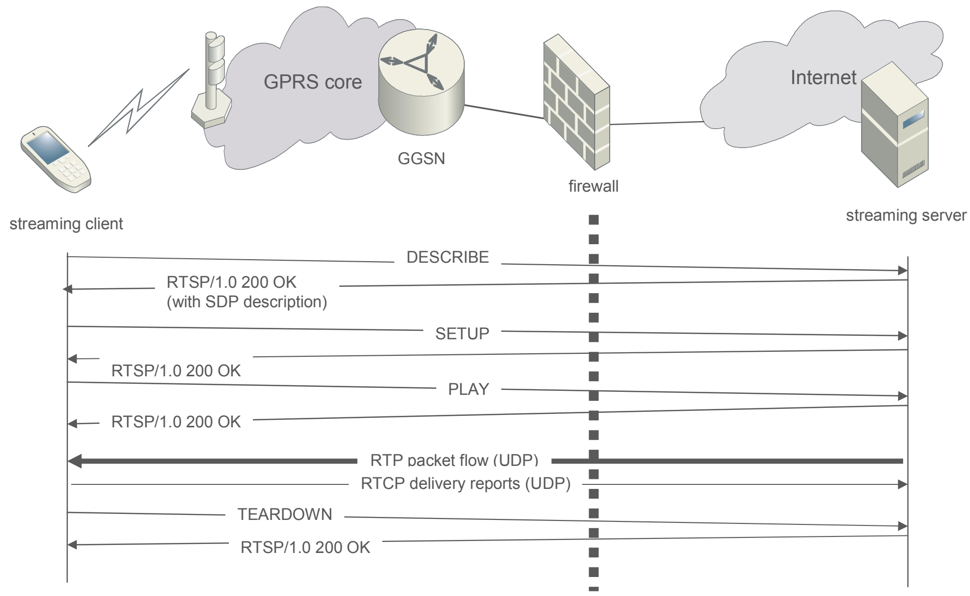

Figure 4. Basic RTSP/RTP Operation [2]

Other Protocols Related to Streaming

Although RTP, RTCP, and RTSP are currently the most important and widely used streaming protocols, there are also other protocols used for different purposes by streaming services.

Session Description Protocol (SDP) [12] is a text-based protocol for defining the name, purpose, media, protocol, codec, timing, and transport information of a streaming session. SDP is used only for describing streaming sessions and thus it does not play a part in actual data streaming.

Hypertext Transfer Protocol (HTTP) [13] is a stateless, text-based, application-level protocol intended mainly for transferring hypertext and hypermedia information. Due to the protocol's generic nature, HTTP can also be used for other purposes, such as file transfers. In streaming, HTTP is widely used for finding and browsing media content and delivering the streaming session descriptions. HTTP can also be used for transporting media content, although it can be more likely thought as progressive downloading than streaming.

09. Challenges for Networks in Mobile Streaming

There are some major challenges that need to be solved or bypassed before mobile streaming services can be brought to the market. Some are the same as in fixed-line networks a few years ago, but some are new challenges dealing with mobile network specific issues.

In GPRS networks, one of the biggest challenges is the limited bandwidth since only approximately 9 kbps per timeslot of actual payload data can be streamed when using CS-2 coding. This translates into a maximum of 27 kbps for 3 timeslot devices and 36 kbps for 4 timeslot devices. In EDGE and UMTS networks, higher bit rates can be achieved. The limited bandwidth problem can be reduced by using an effective codec for the data compression. Unfortunately, complex compression algorithms require much computation power for encoding and decoding the content, which can be a problem on mobile devices with limited hardware resources. Also, handling streams with the higher bit rates possible in UMTS networks requires more processor capacity. Adding more powerful processors to mobile devices makes them more complex and thus more expensive. Powerful processors also increase the devices' power consumption and cause heating problems.

Low throughput is not the only problem with current mobile networks. In all radio networks, the quality of the radio channel may vary basically any time. Because lower quality channels require more bits for error correction, the bandwidth for the actual payload data may drop. Also retransmissions on the radio link layer affect the bit rate. The bandwidth may also drop if the same network resources are shared between many users. Currently, there are no means to provide users any guaranteed quality of service in GPRS networks.

In cellular networks, mobile devices often change base stations which they use for communicating with the network. Changing a cell during streaming causes a short break in the transmission and some data packets to be delayed or even lost. Because the current GPRS networks have quite long roundtrip times, resending the lost packets is often impossible within an acceptable time. Also, available network resources may vary between the cells and thus the handovers may result in a drop of bandwidth if there are not enough resources available for streaming.

10. Conclusions

Mobile networks' ability to transfer data has increased in the recent years, and at the same time, advanced multimedia capabilities have been introduced in mobile phones. When combined, these features have made totally new mobile services possible. Mobile streaming of audio and video content is one of the applications of 2G/3G mobile systems.

Mobile streaming media platforms have moved from proprietary solutions towards standardized technologies such as 3GPP PSS. Using standardized technologies and protocols ensures that the same mobile streaming services are usable on different manufacturers' devices and software.

Mobile streaming is highly dependent on the performance of the underlying network. In practice, the restrictions of current GSM/GPRS networks and devices limit the bit rate of the media stream to a maximum of 9 kbps per timeslot. Thus, the maximum bit rate for media content streamed to devices capable of utilizing 3 timeslots to receive data would be 27 kbps. On the other hand, the minimum practical bit rate for an audio/video stream targeted to mobile devices is around 20 kbps. This is barely enough for stationary scenes such as "talking heads", but insufficient for full-motion content. EDGE and UMTS technologies improve network throughput and enable better quality for streamed audio-visual content. A 50 kbps media stream can be transferred to an EDGE device utilizing 2 timeslots or to an UMTS device using a 64 kbps data bearer. Higher bit rates require more EDGE timeslots or a higher bandwidth UMTS data bearer.

Relatively low network throughput is not the only restriction to streaming quality. Field tests have revealed that the current GPRS mobile networks cannot always provide enough capacity required for reliable streaming. Usually, in suburban and rural areas outside of peak hours, more network capacity is available and streaming can be rather reliable, whereas in dense urban areas during peak hours problems are very likely to occur. Fast movement also decreases reliability, since handovers occur more often and the probability of getting a cell with insufficient resources increases. Handovers also cause short breaks in the transmission resulting in delayed or possibly lost packets. In addition, the quality of the radio channel varies more during fast movement. Media players can handle small numbers of delayed packets by having sufficient playback buffers and tolerate some lost packets. Longer delays and breaks cause interruptions in playback, or in the worst case, the whole connection to break. 2G/3G dual networks also introduce some new challenges such as inter-system handovers (ISHO) between GSM/EDGE and UMTS systems, and UMTS data bearer changes.

In conclusion, mobile streaming is already possible in GPRS and EDGE networks, but reliability and quality are not yet sufficient for high-quality commercial services. However, it is necessary to study mobile streaming in practice, since 3G mobile systems with improved performance are coming to the market.

References

[1] Streaming Tester Software for Mobile Systems Master's Thesis. Saku Tiainen, 2004. Available at: www.medialab.sonera.fi [Accessed June 14, 2004]

[2] Packet Switched Streaming Service White Paper, TeliaSonera Finland Medialab, 2003. Available at: www.medialab.sonera.fi [Accessed June 14, 2004]

[3] Andersson C. GPRS and 3G Wireless Applications. John Wiley & Sons, Inc. Canada, 2001.

[4] Holma H., Toskala A. WCDMA for UMTS – Radio Access for Third Generation Mobile Communications. John Wiley & Sons, Ltd. England, 2001.

[5] Penttinen J. GSM-tekniikka – Järjestelmän toiminta ja kehitys kohti UMTS-aikakautta. Werner Söderström Oy. Finland, 2001.

[6] Chen X, Goodman D. Theoretical Analysis of GPRS Throughput and Delay. USA, 2004. Available at: http://eeweb.poly.edu/dgoodman/Publications.html

[7] Kilpi J, Mannersalo P. Performance Analysis of GPRS/GSM from the Single User Point of View. VTT Information Technology. Finland, 2002.

[8] Korhonen J, Aalto O, Gurtov A, Laamanen H. Measured Performance of GSM HSCSD and GPRS. Sonera Corporation. Finland, 2002.

[9] MPEG-4 White Paper, TeliaSonera Finland Medialab, 2004. Available at: www.medialab.sonera.fi [Accessed June 21, 2004]

[10] RFC 1889. Real-time Transport Protocol. Available at: http://www.ietf.org/rfc/rfc1889.txt [Accessed June 14, 2004]

[11] RFC 2326 Real-Time Streaming Protocol. Available at: http://www.ietf.org/rfc/rfc2326.txt [Accessed June 14, 2004]

[12] RFC 2327 Session Description Protocol. Available at: http://www.ietf.org/rfc/rfc2327.txt [Accessed June 14, 2004]

[13] RFC 2616 Hypertext Transfer Protocol. Available at: http://www.ietf.org/rfc/rfc2616.txt [Accessed June 14, 2004]

Additional Resources

- 3GPP Home Page. 3GPP, 2004. http://www.3gpp.org
- 3GPP Specifications, 3G Partnership Project, 2004. Available at: http://www.3gpp.org/ftp/Specs/latest/

Definitions, Acronyms and Abbreviations

2G	2nd Generation mobile communications
3G	3rd Generation mobile communications
3GPP PSS	3GPP Packet-Switched Streaming
3GPP	Third Generation Partnership Project
3G-SGSN	3G Serving GPRS Support Node
8PSK	8-Phase Shift Keying
AAC	Advanced Audio Coding
AMR	Adaptive Multi-Rate
BSC	Base Station Controller
BSS	Base Station Subsystem
BTS	Base Transceiver Station
Codec	Coder-Decoder
CS	Coding Scheme
CSD	Circuit-Switched Data
ECSD	Enhanced Circuit-Switched Data

EDGE	Enhanced Data rates for GSM Evolution
EGPRS	Enhanced General Packet Radio Service
FTP	File Transfer Protocol
GGSN	Gateway GPRS Support Node
GGSN	Gateway GPRS Support Node (a GPRS network element).
GMSK	Gaussian Minimum Shift Keying
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
H.263	Video codec for low bit rates standardized by ITU-T
HSCSD	High-Speed Circuit-Switched Data
HTTP	Hypertext Transfer Protocol
IETF	The Internet Engineering Task Force
IMT	International Mobile Telecommunications
IP	Internet Protocol
ISHO	Inter System Handover
ISO	International Organization for Standardization
ITU	International Telecommunication Union
ITU-T	ITU Telecommunication Standardization Sector
MCS	Modulation Coding Scheme
MPEG	Moving Picture Experts Group
MS	Mobile Station
MSC	Mobile services Switching Centre
Node B	Base station in UMTS network
NSS	Network and Switching Subsystem
PCU	Packet Control Unit
PDA	Personal Digital Assistant
PSTN	Public Switched Telephone Network
QoS	Quality-of-Service
RNC	Radio Network Controller
RNS	Radio Network System
RTCP	Real-Time Control Protocol
RTP	Real-Time Transport Protocol
RTSP	Real-Time Streaming Protocol
SDP	Session Description Protocol
SGSN	Serving GPRS Support Node
TCP	Transmission Control Protocol
TDMA	Time Division Multiple Access
TS	Timeslot
UDP	User Datagram Protocol
UMTS	Universal Mobile Telecommunications System
UTRAN	UMTS Terrestrial Radio Access Network
W3C	World Wide Web Consortium
WCDMA	Wideband Code Division Multiple Access
WLAN	Wireless Local Area Network