We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

3,900
Open access books available

116,000
International authors and editors

120M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
1. Introduction

Telemedicine can be defined as the use of information and communication technology (ICT) to deliver medical services and information from one location to another. In other words, telemedicine can be seen as a way of distributing medical expertise and services to medically underserviced areas such as remote and rural areas using ICT as a communication platform. Though any communication system can be used in telemedicine, rapid development in computer technology and easiness to purchase has led to more amenability to computer-based telemedicine technologies which are IP-based.

Services offered by telemedicine are designed to help improve healthcare access and information service while reducing the isolation of healthcare providers and residents in rural areas. Telemedicine can also reduce the time and allay the costs of rural patient transportation significantly. Telemedicine includes applications in areas such as pathology and radiology, as well as consultations in specialties such as neurology, dermatology, cardiology, and general medicine. Telemedicine is also used for Continued Medical Education (CME), administration, research and development. Table 1 shows a summary of telemedicine application categories.

- Initial urgent evaluation of patients, triage decisions, and pre-transfer arrangements.
- Medical and surgical follow-up and medication checks.
- Supervision and consultation for primary care encounters in sites where a physician is not available.
- Routine consultation and second opinions based on history, physical exam findings, and available test data.
- Transmission of diagnostic images.
- Extended diagnostic work-ups or short-term management of self-limited conditions.
- Management of chronic diseases and conditions requiring a specialist not available locally.
- Transmission of medical data.
- Public health, preventive medicine, and patient education.

From the applications listed in Table 1, the services offered by telemedicine can be grouped as follows:

- Tele-education
- Tele-consultation
- Tele-monitoring
- Tele-management (or combination of tele-monitoring and tele-consultation)
- Tele-diagnosis
- Medical data exchange

The scope of this chapter is to present a holistic approach to the provision of quality of service in an IP-based telemedicine system. Methods of mitigating some of the problems that degrade the quality of real-time traffic when transmitted over packet networks are discussed, which include DiffServ/MPLS, error concealment and low-bit rate coding.

2. Need for QoS in telemedicine

Telemedicine makes it possible for rural patients in health clinics who need special medical care to have face-to-face consultations with specialists that are situated in a hospital or another medical institution that is far away. In other words, it enables the nurse in a primary health center to send the medical sounds, images and video, which are captured using medical peripherals such as medical video scopes (Octoscope; Colposcope; Dermoscope; Rectoscope; Iriscope; Endscope, etc.), Stethoscope, cameras and other medical instruments, to a doctor in a secondary or tertiary medical center to be used to make a diagnosis for a patient.

For the telemedicine to be effective, it must be able to emulate the onsite face-to-face consultation experience. That is, the telemedicine system must provide quality video, image and audio or sound transmission in both real-time and store-and-forward modes. Quality audiovisual communication is also essential in order for the physician at the distant end to better approximate an on-site physical examination so as to make a correct diagnosis from the received medical sounds, images and/or video.

According to Nanda and Fernandes (2007), it is essential in a critical medical environment that the networking applications perform with a surgical precision; otherwise the outcome could be fatal, both for patients as well as the future of networking in telemedicine. For this to be possible, it is necessary that the information signals reach the end location with high degree of reliability and predictability. Such systems are said to have stringent real-time QoS constraints, of which if not met can lead to disaster; for example, the unbounded delay and jitter in the control system of telerobotic surgery can lead to control loop instability and mission failure (Szymanski and Gilbert, 2010). Lastly, the availability of resources is also imperative in health networks, because the generated traffic may be crucial for the patients' health and life (Zvikhachevskaya et al., 2009).

2.1 Quality of service

According to ITU-T recommendation E.800, QoS can be defined as the collective effect of the service performances, which determine the degree of satisfaction of the user of service (Afullo, 2004). Muhammed et al. (2006), describes QoS as the ability of the network to provide a service with an assured service level, and it is building block for reaching quality end-user experience (QoE); that is, QoS in the network ensures that the user gets quality usable services. These two definitions show clearly, that there is a relationship between the
performance of the system and the satisfaction of the user of the services provided by the system.

In the technical context, QoS is a set of attributes that can be used to define the network’s capability to meet the requirements of the user and application (Kilkki, 2008). QoS parameters include delay (or latency), jitter, bit rate (throughput) and packet loss rate (Aidarous & Plevyak, 2003; Malindi & Kahn, 2008; Tulu & Chatterjee, 2008), and according to Salatian et al. (2011), network resource availability also plays a role in QoS, where delay represents the maximum delay bound that is acceptable to the application; jitter reflects variations in delay; bit rate, also called throughput, refers to data rate that the system can transfer or the rate at which packets are transmitted in a network, packet loss rate refers to the percentage of packet lost among all the delivered packets in a given time interval, and network resource availability is the infrastructure associated with the transmission of data (Salatian et al., 2011).

2.2 Profile of the telemedicine traffic

Typical telemedicine application includes the transfers of basic patient information, transfers of high resolution images such as radiographs, computer tomography (CT) scans, magnetic resonance imaging (MRI) pictures, ultrasound, pathology images, video images of endoscopes or other procedures, patient interviews and examinations, consultations with medical specialists and health care educational activities; therefore, telemedicine traffic includes medical images, video, audio and data from different telemedicine applications or services. These applications may be classified as store-and-forward, near real-time and real-time, where store-and-forward is used for non-emergent situations where the consultation may be made within the next 24 – 48 hours, real time involves transmission of the information as it is acquired and is sometimes interactive like two-way telephone conversation, or two-way face-to-face video conferencing consultation, and near real-time is used for emergent situations where certain data is transmitted immediately to help during consultation that is in process or to be in process soon.

Each of these telemedicine applications has its own QoS requirements (Malindi and Kahn, 2008; Skorin-Kapov and Matijasevic, 2010), ranging from low to high bandwidth, delay tolerance to delay intolerance, and tolerance to packet losses to packet loss intolerance. These QoS requirements vary depending on the type of traffic, type of service, and the context in which the service is invoked. For example, in an emergency situation, a remote specialist diagnosis may require near real-time transmission of medical data, while in a different, non-emergency situation, the patient data is transferred (with tolerance for delay) to a remote location to be analyzed by a specialist (Skorin-Kapov and Matijasevic, 2010). Often the real time applications have stringent network requirements than other applications. According to Feng et al. (2010), the variety of telemedicine applications produces traffic with diverse network requirements and some telemedicine traffic, such as tele-diagnostic with interactive audio and video transmission, may require very high bandwidth, and strict delay and jitter requirements. For example, interactive video communication requires low delay of 200 to 300 ms round-trip and an average jitter that is not more than 30 ms, and for speech the latency is also 200-300 ms and jitter must be limited to 50ms (Cisco Systems, 2002; Sze et al., 2002; Hassan et al., 2005; Tobagi, 2005). According to Szymanski and Gilbert (2010), telerobotic surgery may tolerate maximum delays of up 250 ms and a relatively small jitter of the order of 10s of milliseconds. Some telemedicine applications such as remote surgery may also require guaranteed level of availability in
addition to latency and jitter requirements mentioned above. Such applications are referred to as safety-critical or life-critical applications and the type of QoS they require from the network is sometimes referred to as hard QoS. According to Vergados et al. (2006), Vouyioukas et al. (2007), Zvikhachevskaya et al. (2009), Skorin-Kapov and Matijasevic, (2010), the basic requirements for the different types of telemedicine are presented in Table 2.

<table>
<thead>
<tr>
<th>Application Type</th>
<th>Required throughput</th>
<th>Small delay</th>
<th>Small jitter</th>
<th>Context sensitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tele-consultation</td>
<td>High</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Tele-diagnosis</td>
<td>High</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Tele-monitoring</td>
<td>Low</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Tele-education</td>
<td>High</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Access to DB</td>
<td>Low/High</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 2. QoS Requirements for telemedicine services

2.3 Factors affecting the quality of service

Quality transmission of video and sound depends on four variants: acquisition (or recording), coding, transmission and reproduction. For example, in video communication, the type and the resolution of camera used, the coding standard used, the communication network used, and the resolution of the display at the receiver will contribute to the overall quality of the video at the receiver, whereas in sound transmission it is the acoustic response of the microphone or the stethoscope used, the coding standard used, the communication network used, and the frequency response of the speakers at the receiver will contribute to the quality of the sound at the receiver. Each of these variants has some limitations, which make it impossible to achieve the ideal quality of the signal being transmitted as a result in most cases a compromise is reached between the ideal and the practical quality transmission.

As alluded to in the introduction, most of the telemedicine system are IP-based and the default and traditional datagram delivery service used in IP-based networks is called best-effort service. This type of service delivers the packets as fast as possible, using, in most cases, first-in-first-out (FIFO) scheduling mechanism, which delivers the datagram to the output port in the same order as they came from the input port without giving preference to any type of traffic. For slightly loaded network this service can be sufficient, but if the network is congested queueing losses and queueing delays can make the network not to be suitable for real-time and critical traffic. Another problem with IP network is the link-state routing protocols that are used to distribute IP routing information throughout a single autonomous system (AS) in an IP network; that is, Interior Gateway Protocols (IGPs). These IGPs like Intermediate System-Intermediate System (IS-IS) protocol, and Open Shortest Path First (OSPF) protocol create a shortest path matrix for each router and then forward traffic along this shortest path. Unlike ATM’s PNNI, IP routing protocols do not have a Connection Admission Control (CAC) mechanism to balance network traffic along multiple paths. The result is under-utilization of certain links, creation of hot spots in the network and congestion along the shortest path route. These congestions can cause long queues, resulting in serious degradation for applications that are running due to packet loss, latency, and jitter thus compromising the quality of service. Additional to IP debilities, other factors, such as transmission errors and limited bandwidths, can also jeopardize the quality of service required by telemedicine applications.
3. Provision of quality of service

IP network is a packet-based network, which is an unpredictable system with parameters that are variable according to network usage. In order to overcome the challenges posed by such a system and still be able to provide quality transmissions the following are proposed:

i. Traffic differentiation and prioritizing
ii. Traffic engineering and protection
iii. Error control and concealment
iv. Using encoding schemes that will produce low bit rates without compromising much on quality in order to cater for low capacity systems.

3.1 Traffic differentiation and prioritizing

IETF has proposed some mechanisms to compensate for the best effort deilities of IP-based networks in order to make IP suitable for almost all the different multimedia applications. These mechanisms or technologies, which include IntServ, DiffServ, and MPLS, have been proposed to provide QoS in IP-based multiservice networks. All the three mechanisms are able to guarantee the appropriate QoS treatment to real-time traffic concerning its special delay, jitter, packet loss and bandwidth requirements. However, the drawback of IntServ is that it suffers from scalability problems (Aidarous & Plevyak, 2003; Leon-Garcia & Widjaja, 2004). Hence most of the literature on provision of QoS has been concentrating on the combination of DiffServ and MPLS as the enablers of QoS in IP networks. Where DiffServ is used to divide traffic into a small number of classes, and MPLS is used for traffic engineering and to offer IP QoS services efficiently over diverse layer-two backbone networks (Patrikakis et al., 2003).

One of the reasons for the adoption of a combination of DiffServ and MPLS, is that, though MPLS offers traffic engineering and great savings in terms of switching speed, it still lacks an effort to distinguish among the flows of different delay characteristics, and drop precedence in order to meet the two conditions that are necessary for true QoS as proposed by Fineberg (2003), which are guaranteed bandwidth, and class-related scheduling and packet discarding treatment. Secondly, using DiffServ alone cannot guarantee QoS since DiffServ does not influence a packet path, and therefore during a congestion or failure, even high-priority packets would not get guaranteed bandwidth. Pairing MPLS with DiffServ can compensate these downsides, so that DiffServ can be responsible for classifying the traffic according to different flow characteristics, while MPLS is responsible for providing a connection-oriented environment that enables traffic engineering. According to Ghazel and Saidane (2008), the integration of DiffServ with MPLS guarantees the QoS for a broad range of multiservice traffic by offering traffic policy and classification in diverse QoS classes of the transport data plane.

3.1.1 Traffic differentiation

The diversity in network resources that are needed for different e-health application, as well as various levels of urgency in medical situations makes the DiffServ model an appropriate architecture for QoS provision (Vergados, 2006). DiffServ divides traffic into a small number of classes with each class being identified by a mark called DSCP, which is carried in the six-bit differentiated field of the IP header. DiffServ defines Classes of Service (CoS), called aggregates, and QoS resource management functions with node-based, or per-hop, operation. The CoS definitions include a behaviour aggregate (BA) which has a specific requirements for scheduling and packet discarding, and an ordered aggregate (OA) which
performs classification based on scheduling requirements only, and may include several drop precedence values. Thus, OA is a coarser classification than a BA and may include several BAs. DiffServ offers per hop behaviour (PHB) to BA, which includes scheduling and packet discarding, and PHB scheduling class (PSC) to OA, which only concerns scheduling. The value of the DSCP field is used to specify a BA (or class), which is used by DiffServ-compliant nodes for choosing the appropriate PHB. Per-Hop Behaviors refers to specific forwarding treatments of traffic aggregates or packets in a DiffServ network, which includes packet scheduling, queuing, policing or shaping behavior of a node on any given packet belonging to a BA. There are four standard PHBs that have been defined: the assured forwarding (AF) PHB, expedited forwarding (EF) PHB, class-selector PHB, and default PHB. Where AF PHB is designed to give a reliable service even in times of congestion; EF PHB is designed for traffic that is required to be guaranteed enough resources to ensure that it receives its minimum guaranteed rate (Ibe, 2002, p.227); class-selector PHBs are designed to preserve backward compatibility with the IP-Precedence scheme defined in RFC1812 and default PHB is the standard best-effort treatment that nodes (routers) perform when forwarding traffic (Ibe, 2002, p.228; Aidarous and Plevyak, 2003, p.119; Leon-Garcia and Widjaja, 2004, p.719). AF PHB can be further subdivided into twelve PHBs to make the total of fourteen PHBs in all. The twelve AF PHBs are divided into four PSCs, and each of the PSCs consists of three sub-behaviours, which are related to different packet discarding treatment.

3.1.2 Traffic prioritization

Once the traffic has been classified, it needs to be ranked according to the order of priority it deserves and thereafter set the per-hop behavior (PHB) that closely suite its QoS requirements. Usually, the traffic is classified according to the content and not the context, however, Skorin-Kapov and Matijasevic, (2010) have proposed that the service context in terms of emergency or patient critical versus non-emergency and non-critical is also crucial in traffic scheduling mechanisms. Vouyioukas et al. (2007) alluded to the importance to distinguish between the requirements for real-time and non-real time, medical for diagnosis and non-diagnosis. Emergency sections need to be prioritized over non-emergency sessions to ensure the best coordination (Gouveia et al. 2009). The following table can be used for prioritizing:

<table>
<thead>
<tr>
<th>Traffic</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emergency services and Tele-surgery</td>
<td>Highest</td>
</tr>
<tr>
<td>Real-time Tele-consultation and Tele-management</td>
<td>High</td>
</tr>
<tr>
<td>Tele-diagnosis and Tele-monitoring</td>
<td>Medium high</td>
</tr>
<tr>
<td>Tele-education</td>
<td>Medium low</td>
</tr>
<tr>
<td>Medical data exchange</td>
<td>Low</td>
</tr>
</tbody>
</table>

Table 3. Prioritization of telemedicine traffic

3.2 Traffic engineering and protection

Once the traffic has been classified and arranged according to its priority levels, traffic engineering (TE) is needed, which is a process of controlling how the traffic flows through the network as to optimize resource utilization (Xiao et al., 2000). This is done in order to
prevent uneven network traffic distribution which can cause high utilization or congestion in some part of a network, even when total capacity of the network is greater than total demand (Xiao et al., 2002). Uneven distribution may be due to the fact that packets tend to follow the shortest path of route, thus resulting in the shortest path being highly utilized and congested, while a longer path remains under-utilized. This uneven traffic distribution in the core network can result in QoS problems such as packet loss, latency, and jitter increase as the average load on a router rises, even when the network utilization is low. Additional to congestion, another problem that can be encountered is route failure, which can result in packet loss and quality of service degradation if not attended to quickly. The IntServ and Diffserv QoS schemes, mentioned above, do not eliminate traffic congestion or provide fast rerouting, but provides differentiated performance degradation for different traffic during network congestion. Therefore, additional to Diffserv, there is still a need for managing network resources such as bandwidth, and controlling routing and traffic in order to minimize congestion and provide fast re-route in the event of route failures so that traffic oriented performances, such as latency, jitter, and packet loss, can be improved.

3.2.1 Traffic engineering using MPLS

In order to facilitate efficient and reliable network operations while simultaneously optimizing network resource utilization and traffic performance, the IETF has introduced multi-protocol label switching (MPLS), constraint-based routing (CR) and an enhanced link state IGP.

MPLS is an IETF specified framework that provides for efficient designation, routing, forwarding, and switching of traffic flows through the network. It approaches the QoS by attempting to address the shortcomings of the IP routing, which include speed, scalability, QoS management and traffic engineering. In traditional IP networks each router makes an independent forwarding decision on each packet as it traverses the network. This decision usually involves a complex manipulation of a large routing table to determine the next hop of a packet. MPLS also uses IP addresses either IPv4 or IPv6, to identify end points and intermediate switches and routers. This makes MPLS networks IP-compatible and easily integrated with traditional IP networks. However, unlike traditional IP, MPLS flows are connection-oriented and packets are routed along pre-configured Label Switched Paths (LSPs). MPLS also slightly modifies the IP packet format to include a new Shim Label header, called the MPLS header, which contains a label field. MPLS works by tagging each packet with an identifier to distinguish the LSPs. This identifier is contained within the label field and it consists of fixed-length value, called label, which is used as index into table, which specify the next hop, and new label. When a packet is received, the MPLS router uses this label, instead of the traditional IP destination address, to deliver the packet along the selected path or LSP, and then looks up the LSP in its own forwarding table to determine the best link over which to forward the packet, and the label to use on this next hop. The 32-bit MPLS header sits between the layer 2 (data link layer) header and the layer 3 (network layer).

In a Diffserv/MPLS network MPLS edge nodes need to be Diffserv-enabled in order to engineer the network with traffic engineering that is applied on per class basis. This type of traffic engineering is referred to as Diffserv-aware MPLS traffic engineering (DS-TE) and its essential goal is to guarantee network resources separately for each traffic in order to improve and optimise its compliance with QoS requirements. The essential goal of DS-TE is to guarantee network resources separately for each type of traffic in order to improve and optimise its compliance with QoS requirements (Minei, 2004). This makes it possible to
establish separate LSPs for different classes, taking into consideration resources available to each class. For example, a separate LSP can be established for each type of real-time or premium traffic, and these LSPs can be given higher priority than other LSPs (Goode, 2002). DS-TE enables the introduction of the concept of LSP priorities, where some LSPs are marked as more important than others, and it also allows the more important LSPs to be able to confiscate resources from the less important LSPs; that is, pre-empting the less important LSPs. This is done to ensure that important LSP always establishes along the optimal path regardless of the existing reservations, and when LSP needs to reroute in the event of node or link failure, important LSPs have a better chance of finding an alternative path. It is also done to ensure that when high-priority-bandwidth is not needed for real-time traffic, it can be used for lower priority class of traffic that is carried on less important LSPs (Goode, 2002; Minei, 2004). Each LSP has two priorities associated with it: a setup priority and a hold priority. The setup priority controls access to the resources for an LSP at the time of LSP establishment, and the hold priority control access to the resources for an LSP that is already established. The fundamental requirement for DS-TE is to be able to enforce different limits on the percentage of a link’s bandwidth for different sets of aggregation of traffic flows of the same class, which are placed inside an LSP; that is, DS-TE must enforce different bandwidth constraints (BCs) to different sets of traffic trunks (TT). This implies keeping track of how much bandwidth is available for each type of traffic at a given time on all the routers throughout the MPLS network. To address this, Le Faucheur and Lai introduced the concept of a class type (CT), which is the set of traffic trunks crossing a link, that are governed by a specific set of bandwidth constraints. This CT is used for the purpose of link bandwidth allocation, constraint based routing and admission control. Once a traffic trunk has been assigned to a CT it will remain an element of that same CT on all links. Each CT may carry traffic from more than one class of service (RFC 3564, 2003). According to RFC 3564 (2003), the DS-TE solution must support up to 8 CTs. Those 8 CTs are referred to as CTc, 0 ≤ c ≤ MaxCT-1 = 7; that is, CT0 through CT7. The DS-TE solution must enforce a different set of bandwidth constraint for each CT, and a DS-TE implementation must support at least 2 CTs. LSPs that are traffic engineered to guarantee bandwidth from a particular CT are referred to as DS-TE LSPs, and in current IETF model, a DS-TE LSP can only carry traffic from on CT. By convention the best effort traffic is mapped to CT0.

3.2.2 Traffic protection
Once the high-priority traffic has been classified and the important LSPs have been established, the high-priority or premium traffic can be mapped on these important LSP to allow it to utilize the resources available for the premium class. To make sure that this premium traffic is less affected by route failures, the important LSP will have Fast Reroute enabled so that when there is a route failure the premium traffic can be switched to the protection LSP, which is established around the ‘working’ or operational LSP as an alternative route in the event of any link/s or router/s along the working LSP fail. This traffic protection and fast rerouting is essential for applications that cannot tolerate packet loss in order to make sure that premium traffic is less affected by route failures.

3.3 Error control and concealment
One of the inherent problems with any communication channel is the introduction of errors to the data being transmitted through it. According to Wang and Zhu (1998), transmission errors can be roughly classified into two categories: random bit errors and erasure errors. Random bit errors are caused by the imperfections of physical channels, which result in bit

www.intechopen.com
inversion, bit insertion, and bit deletion. Erasure errors, on the other hand, can be caused by packet losses, or system failures for a short time (Wang and Zhu, 1998). These errors of transmission channel can be effectively corrected by channel coding methods such as forward error correction (FEC) and automatic repeat request (ARQ) or a mixture of them (Sullivan and Wiegand, 2005). Both FEC and ARQ channel coding methods improve the reliability of the information being transferred by adding additional bits to the transmitted data stream, which of course increases the amount of data being sent. However, these two traditional methods do not provide the bounded delay that is required for real-time traffic as a result other types of error-control mechanisms such as error resilience and error concealment have been proposed in addition to ARQ and FEC schemes (Wu et al., 2000; Stanković et al., 2003).

3.3.1 Error resilience

Error resilient schemes deal with packet loss on the compression layer by considering the semantic meaning of the compression layer and attempt to limit the scope of damage caused by the packet loss on the compression layer. The standard error resilient schemes are resynchronization, data partitioning, and data recovery (Arankalle, 2003; Wu et al., 2000).

In resynchronization the encoder must introduce resynchronization markers in the video bitstream at various locations (e.g. beginning of the frame in MPEG –1/2, H.261/263, or periodically in MPEG 4) so that when the decoder detects an error, it can hunt for this resynchronization marker to regain synchronization with the encoder, thereby preventing error propagation across different segments of the video bitstream, which are separated by the markers. These markers are designed such that they can be easily distinguished from all other codewords and a small perturbation of these codewords (Villasenor et al., 1999; Wang et al., 2000; Du et al., 2003).

Data partitioning is a technique, which involves segmenting the bit stream into segments. This is based on the fact that some bits in the encoded video bitstream are more important than others within the same bitstream. For example, DCT coefficients are less important than GOB headers, motion vectors, and Macroblock information bits. Hence an error occurred in the received DCT information has less impact on the decoded quality than bit errors received in group of block (GOB) headers, motion vectors, and Macroblock information bits. Because of this unequal data importance within the encoded video bitstream, it is possible to separate the bitstream into two layers of importance to allow each to be protected according to its level of importance; that is, unequal error protection (UEP). This helps during the decoding to prevent the errors occurring in one segment from affecting other segment (Katsaggelos, Ishtiaq, Kondi, Hong, Banham and Brailean, 1998). One of the shortcomings of using resynchronization markers alone is that, after detecting an error in a video bit stream and resynchronizing to the next resynchronization marker, the video decoder discards all the data between the two resynchronization markers, where the error has occurred. Reversible variable-length codes (RVLCs) alleviates this problem and enable the decoder to better isolate the error location by enabling data recovery in the presence of errors. RVLCs are VLCs that are wrapped in an error correcting code; that is, they are having the prefix property when decoding them in both the forward and reverse directions. This enables RVLCs to be uniquely decoded in both forward and reverse direction. So when a decoder detects an error while decoding a bit stream in a forward direction, it jumps to the next resynchronization marker and decodes
the bit stream in the backward direction until it encounters an error, and based on the locations of these two errors, the decoder can recover some of the data between the two consecutive resynchronization markers that would otherwise have been discarded (Talluri, 1998). According to Li et al. (2000), RVLC can be used in conjunction with data partitioning, where data partitioning is used to separate the less important information bits from the most important information bits and RVLC is used in coding the most important information bits so that the decoder is able to recover data within the erroneous segments (Villasenor et al., 1999; Li et al., 2000). RVLC has been adopted in both MPEG-4 and H.263, in conjunction with insertion of synchronized markers (Wang et al., 2000).

3.3.2 Error concealment
Error concealment is implemented at the receiver by the decoder, usually after data recovery, in order to hide the effects of errors once they have occurred. It improves the quality of the decoded sequence in the presence of errors when the error rate is not that high. There are three types of concealment: temporal concealment, spatial concealment, and motion-compensated concealment. Temporal concealment replaces a corrupted area with a pixel values in the same location in the previous decoded frame, and it is only effective if there is a little change from the previous frame. Spatial concealment replaces the distorted block by a spatial interpolation between adjacent error-free blocks. It is used for those situations where temporal concealment is not effective, for example, where there is high motion or scene change. Motion-compensated concealment estimates the motion vectors for the lost block rather than simple assuming no motion. (Riley and Richardson, 1997, p.124).

3.4 Coding for low-bit rates
Rural and cellular-based telemedicine systems have limited bandwidths (Zvikhachevskaya et al., 2009; Salatian et al., 2011), which can affect the QoS. One approach to deal with low bandwidth and service contention in a rural setup is to use data compression to shrink larger files to smaller files so that they can occupy less space and become faster to transfer over the network. This decreases the delay, and jitter, while increasing the throughput (Salatian et al., 2011). This enables the transmission of bandwidth-hungry traffic like video using a system with a limited capacity.

The Digital Imaging and Communication in Medicine (DICOM) committee has adopted various JPEG variants, such as lossy and lossless JPEG, JPEG-LS, and JPEG 2000. For medical images, which are having a diagnostic use, medical image compression techniques have primarily focused on lossless methods, where the signal can be reconstructed exactly from its uncompressed format (Pattichis et al., 2002). For digital video, the DICOM committee has adopted MPEG 2 for diagnosis video. However, due to its high bandwidth requirements and frame synchronization problem, MPEG 4 standard can be used for real-time diagnosis video and for non-diagnosis video applications such as video conferencing. H.263 may be acceptable (Pattichis et al., 2002). The advancement in video coding has led to the development of H.264 by Joint Video Team (JVT), which is a joint venture between ISO’s MPEG and ITU-T’s VCEG. The official title of the H.264 standard is Advanced Video Coding (AVC); however, it is widely known as by its old working title, H.26L, and by its ITU document number, H.264 or as ISO MPEG 4 Part 10, or just MPEG AVC. The main objective of H.264 standard is to provide means to achieve substantially higher video quality as compared to what could be achieved using any of the existing video coding standard. It
QoS in Telemedicine

has a number of advantages that distinguish it from existing standards, while at the same time, sharing common features with other existing standards. These advantages include up to 50% in bit rate saving compared to H.263 or MPEG 4, high quality video that is consistent at high and low bit rates, error resilience, and network friendliness that makes it possible to transport H.264 bit streams over different networks. These advantages make H.264 an ideal standard for several video applications such as video conferencing, storage and broadcast video. According to Panayides et al. (2010), H.264/AVC provides for efficient (size wise) and timely (real time) encoding.

4. Performance evaluation of the proposed QoS

In the preceding section techniques for providing a holistic QoS in IP-based telemedicine system were discussed and here the performance of the proposed solutions are evaluated. First will be the network functions (DiffServ and MPLS), second will be the coding standard, and thirdly will be the error control and concealment.

4.1 DiffServ/MPLS evaluation

First is the evaluation of the performance of DiffServ with MPLS in a multiservice IP-based network. The simulation is performed with only DiffServ first and thereafter with a combination of both DiffServ and MPLS, and as part of this performance evaluation different schedulers were evaluated to determine the most appropriate one for real-time traffic, which is delay and delay variation sensitive, in a differentiated services environment. Topology used is as shown in Figure 1.

Fig. 1. Simulation setup

Each site in the network is capable of communicating using video, voice or data and these applications can either run one at a time or simultaneously. Since most of rural systems have a limited bandwidth, the core links are represented by bidirectional E1 links (2.048 Mbps).

4.1.1 Traffic source definition

Multi-service traffic involves data, voice and video traffic. Data traffic is usually transmitted using Transport Control Protocol (TCP) while real-time traffic such as voice and interactive video are transmitted using User Datagram Protocol (UDP). To simulate data applications,
Telnet and FTP traffic sources are going to be considered. To simulate speech traffic, which is characterized by alternating spurts and periods of silence, an ON/OFF process is going to be used. Since the ON (or OFF) periods do not have a heavy-tailed distribution, a non-fractal version of the ON/OFF model with an exponential distribution of on and off times is suitable. So voice traffic is simulated by the G.711 audio codec, which transmits data over RTP. The G.711 codec outputs data at a constant rate of 64 kbps with a frame size of 240 bytes, and in the case of IP-based networks, these frames are encapsulated by the IP/UDP/RTP protocols that augment the basic frame size with their headers resulting in packet sizes of 280 bytes for IPv4 and 300 bytes for IPv6, and transmission rates of 74.667 kbps for IPv4 and 80 kbps for IPv6. The parameters for speech traffic ON/OFF source are given in Table 4.

<table>
<thead>
<tr>
<th>Traffic type</th>
<th>Mean ON duration (seconds)</th>
<th>Mean OFF duration (seconds)</th>
<th>Peak rate (Packets/second)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voice</td>
<td>0.35</td>
<td>0.65</td>
<td>80k</td>
</tr>
</tbody>
</table>

Table 4. Simulation parameters for the voice

These ON/OFF parameters for VoIP are widely used in literature (Rakocevic, 2003) and are taken from measurements by Deng (1995). For data sources, the packet size is 1500 bytes for each type.

The real-time video is simulated using a trace-driven source. These traces are generated from 'Mr Bones' (a movie by a South African comedian, Leon Schuster), which is encoded using H.264 at 600 kbps.

4.1.2 Service definition

To support service differentiation, the simulated model specifies the following class:

- Premium for real-time traffic that requires low latency, low jitter and low loss, which is classified as EF.
- Silver for data services like telnet and FTP, which is classified as AF11 for Telnet and AF12 for FTP
- Best effort for non-emergent traffic such as e-mail

4.1.3 Simulation results

Nine schedulers: PQ, SCFQ, SFQ, WFQ, WF2Q+, LLQ, WIRR, RR, and WRR, were compared using the three QoS parameters: packet loss, [one way] delay, and jitter. Two traffic types were used for simulation: video trace and constant bit rate (CBR) traffic for other traffic source. Two queues were configured to support both EF and BE aggregates. Video was transmitted as premium traffic using EF services, while CBR was transmitted using BE services. To avoid synchronization problems, background traffic was generated with several CBR sources whose rate was chosen from a uniform random distribution in the range [10 kbps, 100 kbps], while the starting time was chosen from a uniform random distribution in the range [0 s, 5 s]. The simulation was done using packet sizes ranging from 64 to 1600 bytes.

The simulation results revealed that WFQ has the lowest average jitter values which are less than 1.55 ms, followed by LLQ with average values which are less than 1.65 ms, PQ and RR with average values that are less than 1.7 ms, and then others. The simulation results also

www.intechopen.com
revealed that when it comes to one way delay (OWD) PQ and LLQ have the lowest average one-way latency values which are less than 17.5 ms, followed by RR with average values of less than 19 ms, and then others.

The second scenario simulated was to check the ability of DiffServ to provide QoS of service when the network is heavy loaded and also its ability to do fast rerouting in the event of node or link failure when it is used alone.

The results revealed that as the number of video traffic is increased the one way delay and delay variations also increase as shown in Figure 2. However, when MPLS was incorporated in the simulation it was found that both the one way delay and jitter were reduced as shown in Figure 3. The simulation results also revealed that in the event of link or node failure, MPLS has an ability to do fast rerouting.

Fig. 2. One Way Delay (OWD) and Jitter (IPDV) for DiffServ only
Fig. 3. One Way Delay (OWD) and Jitter (IPDV) for DiffServ enabled MPLS

4.2 Low bit-rate coding standard evaluation
In a study by Zoha (2010), the applicability of H.264 was examined in telemedicine reference model and a comparison between MPEG4 and H.264 was made to assess which of the two video coding standards perform better in telemedicine application. MPEG was implemented with MoMuSys codec, and H.264 was implemented with H.264 reference model version JM6.1e.
With the test conditions for MPEG 4 set to match those of H.264, the results achieved showed that H.264 performs much better than MPEG 4 (Zoha, 2010).

4.3 Error control and concealment evaluation
Tests of the performance of the H.264 in an erroneous environment were also performed by the author to check the error resiliency of the H.264 using evaluation process steps as depicted in Figure 4.
A raw YUV digital video sequence, called “Mother & Daughter”, was downloaded in CIF format from the site http://www.tkn.tu-berlin.de/research/evalvid/cif.html, and processed by JM1.7 codec to generate a H.264 bitstream at 213.95 kbps. After encoding the trace files were generated, then followed by all the perceptual video evaluation process
steps depicted in Figure 4. Errors were introduced in the communication process in order to simulate link-level errors and losses. Then the quality of video was evaluated before and after transmission using peak signal-to-noise ratio (PSNR), Video quality measurement (VQM), and Double stimulus impairment scale (DSIS) measurements. Common compression artifacts such as blockiness (or block distortion), and blur (loss in the detail that occurs around edges, typically due to the quantization of transform coefficients representing higher frequencies) were also evaluated.

Fig. 4. Perceptual video evaluation process

The first test was done without any Intra-coded macro-blocks (MBs), and the second test was done with Intra-coded MBs, which are used in H.264 encoding to prevent error propagation (Shi et al., 2008). The results of the tests are presented in Figures 5, through 8, and Table 5.

Fig. 5. Original and encoded video before transmission

Fig. 6. Original and received video (without Intra-coded MBs) after transmission
Where Figure 5 shows the original (raw) and processed (encoded and decoded) video before transmission, Figure 6 shows the Original and received video (without Intra-coded MBs) after transmission, Figure 7 shows the original video and the received video (with Intra-coded MBs) after transmission, and Figure 8 shows the PSNR plots of the three conditions: before transmission, after transmission (without Intra-coded MBs), and after transmission (with Intra-coded MBs).

Fig. 7. Original video and the received video (with Intra-coded MBs) after transmission

Table 5 gives a summary of the values of the mean PSNR, VQM and DSIS scores, and impairments such as blurring, blockiness and jerkiness of the encoded video before and after transmission.

Fig. 8. PSNR quality plots before and after transmission
Table 5. Summary of mean PSNR, VQM score, DSIS Score, and perceptual effects

<table>
<thead>
<tr>
<th>Processed video under measurement</th>
<th>Mean PSNR (dB)</th>
<th>VQM score</th>
<th>DSIS score</th>
<th>Blurring</th>
<th>Jerkiness</th>
<th>Blockiness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before transmission</td>
<td>39.77</td>
<td></td>
<td></td>
<td>21</td>
<td>28</td>
<td>26</td>
</tr>
<tr>
<td>After transmission (without intra-MBs)</td>
<td>37.28</td>
<td>0.50</td>
<td>3.0</td>
<td>25</td>
<td>28</td>
<td>51</td>
</tr>
<tr>
<td>After transmission (with Intra-MBs)</td>
<td>38.35</td>
<td>0.38</td>
<td>3.5</td>
<td>22</td>
<td>28</td>
<td>37</td>
</tr>
</tbody>
</table>

From the PSNR plots, the received decoded pictures, and the results in Table 5 it can be seen that the errors introduced by the channel can have a profound effect on the quality of the transmitted video. However, using Intra-coded MBs, during coding can help minimize the effects of the errors. This is can be observed in the images of the received videos in Figures 6(right)-7(right). For example, in Figure 6(right) the fingers of the left hand, picture on the wall and the tip of the head of the mother is distorted due to the errors introduce by the channel, whereas in Figure 7(right), which is the video that is encoded with Intra-MBs, that only the hand of the mother has been affected by errors. This means then that the erasure effect that is caused by the packet loss and errors in the channel can be minimized by using resynchronization (or Intra-coded MBs), which prevents error propagation. This is also evident from the summary of the evaluation results in Table 5.

5. Conclusion

This chapter presents a holistic approach to the provision of quality of service in IP-based telemedicine system. Methods of mitigating some of the problems that degrade the quality of real-time video when transmitted over packet networks are be discussed. These include mechanisms to compensate for packet networks idiosyncrasy, especially IP best-effort debilities, in order to meet the latency and jitter requirements of real-time traffic. Secondly, to cater for erroneous channel and limited bandwidth, the solution includes adopting coding techniques that will provide coding efficiency, high quality video that is consistent at high and low bit rates, resilience to transmission errors, scalability, and network friendliness, which will result in perceived quality improvement, high-compression efficiency to suite the limited bandwidth, and possibility of transportation over different types of networks.

To facilitate quality of service within IP networks, DiffServ has been adopted in this study so as to provide differentiated services for different classes by using mechanisms that include packet classification, policing, class-based queuing and scheduling, and random early detection. To complement the DiffServ and provide a holistic approach to the provision of QoS, traffic engineering using MPLS was incorporated. MPLS provides connection-oriented paths in a connectionless IP environment. It also provides fast reroute during link or route failure, thus minimizing packet loss that can cause quality degradation and unavailability. The two technologies: DiffServ and MPLS compliments each other to minimize delay, jitter and packet loss.

Since almost every telemedicine system includes real-time video, which is bandwidth hungry, reduction of video transmission rate is required in order to transmit it in real-time.
over the band-limited channel that is available in most of the rural and mobile communication systems used in telemedicine. Tests have shown that H.264 is the best candidate for low bit rate video compression. Tests have also shown that H.264 has a superior built-in error resiliency to guarantee better quality even under erroneous transmission conditions.

6. References


Telemedicine is a rapidly evolving field as new technologies are implemented for example for the development of wireless sensors, quality data transmission. Using the Internet applications such as counseling, clinical consultation support and home care monitoring and management are more and more realized, which improves access to high level medical care in underserved areas. The 23 chapters of this book present manifold examples of telemedicine treating both theoretical and practical foundations and application scenarios.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License, which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.