Date of approval Grade

Advanced Services over IP networks.

Markku Suistola

suistola@iki.fi

Helsinki 21.11.2000

Seminar: Ubiquitous computing - our Future ?

University of Helsinki

Department of Computer Science

Advanced Services over IP networks.

Markku Suistola
Seminar: Ubiquitous computing - our Future ?
Department of Computer Science
University of Helsinki
21.11.2000, 31 pages

This article describes the modern advanced IP network services and the evolution trends that lie ahead. This article will also describe the extra protocols needed for the performance optimisation and the quality improvement at the different layers of the OSI-model. The predefined Quality of Services (QoS) and support for a real-time content are the most essential requirements for any system transporting a multimedia over IP network. The QoS parameters are the building blocks of all multimedia systems. Multimedia content places real-time requirements, such as low jitter and latency. Increased bandwidth sometimes solves the problem, but the bottleneck always remains somewhere in the network.

The real-time (RTP, RTCP, RTSP) and the resource reservation (RSVP) protocols are needed for a successful transportation of the real-time sensitive content media over the IP networks. Traditionally the protocols of the Internet are considered as "best-effort"-protocols meaning that they cannot meet the requirements placed by the new services and especially the multimedia applications. Services deploying the network performance, such as the teleconferencing over network, the voice calls over the Internet (VoIP) or even the large scale data storage systems (NAS/SAN), still use IP-based network model. Some extra concern has to be paid on because the Internet today is far from a secure platform for these applications. Many of the services are still quite vulnerable for the attacks from the Internet. As we should clearly see, the future will bring the security aspects to the design of the new protocols.

Keywords: QoS, Quality of Service, CoS, real-time, multimedia, RSVP, RTP, RTCP, RTSP, ATM, Video-Telephony, VoIP, IPSec, VPN, Multicast, Video-on-demand, IPTV, Distance Learning, SAN, AIN

Table of Contents

1. Introduction

The Internet is built on the IP technology. The evolution of the IP technique is coming to certain point where it is almost possible to see into the world of new unused possibilities. Business and industry are moving towards the kind of Intranets and Internet that are capable to support the new rising IP services and the electronic business (e-Commerce) models.

Traditionally the networks have been seen as a slow and high-cost part of the infrastructure with only limited amount of useful functions. At the traditional model the possible problems have nearly always been solved by increasing the raw performance of the network connections. This sometimes solves the problem, but not always. Many of the new services that the networks are supposed to support are sensible to the end-to-end connection quality. For Example audio and video (multimedia) need both performance and quality, be because the content suffers greatly from delays, errors and lost particles. It is not anymore about what the network may offer but more of what the applications need for the proper operation. This Article describes the two ways of dealing with this problem: Class of Service (CoS) and Quality of Service (QoS).

Chapter 2 describes the most essential issues of the multimedia oriented content transportation such as requirements for a predefined Quality of Service, Real-Time delivery, and resource reservation. Chapter also describes the main differences between the ways of dealing with Quality at the networks (CoS and QoS).

Chapter 3 describes the protocols designed for multimedia oriented content transportation on the different layers of the OSI Model. Chapter describes briefly the Resource ReSerVation protocol (RSVP), the Real-Time Transport Protocol (RTP), Real-Time Content Protocol (RTCP) and Real-Time Streaming Protocol (RTSP).

Chapter 4 describes the new services that the Advanced Intelligent Network (AIN) may provide. The chapter also describes the security issues involved with the newly developing services. The different multimedia services (such as IPTV and VoIP) are briefly described as well. At the end of chapter 4, the large scale storage solutions, SAN/NAS are described.

2. Quality for the IP-networks

Traditionally the Internet has been seen as a network providing millions of slow connections. The traditional services, such as file transfer and electronic mail (email), have been designed for slow, open, secure and basically unreliable networks. As the network performance is increasing the new ways of deploying the Internet are rising. Multimedia is coming to the Internet together with many new and inexperienced users. Typical multimedia is mixed combination of voice and video with an advanced packing system. The nature of the multimedia comes from the content, because transfer rates should be high at the same time when the transmission delays must be kept small. Every packet of multimedia system has to arrive it's destination in order and in certain fixed time (packet has to come before predefined deadline or it should be discharged). [AKR94].

The use of Internet for multimedia transportation is tempting. The required infrastructure already exists and the investments won't have to be too hard. The WAN technologies used offer low-cost, multipurpose and shared bandwidth for large networks when comparing to the dedicated point-to-point connections [Jai98]. The Integrated Services working group of the Internet Engineering Task Force (IETF) has several research projects on the multimedia transportation over the public Internet.

There are essentially two types of Quality services available:

- Prioritization (Class of Services, CoS): network traffic is classified and apportioned network resources according to the bandwidth management policy criteria. To enable CoS, classifications give preferential treatment to applications identified as having more demanding requirements.

- Resource reservation (real Quality of Service, QoS): network resources are apportioned according to an application's QoS request, and subject to bandwidth management policy. RSVP provides this mechanisms together with the other multimedia protocols.

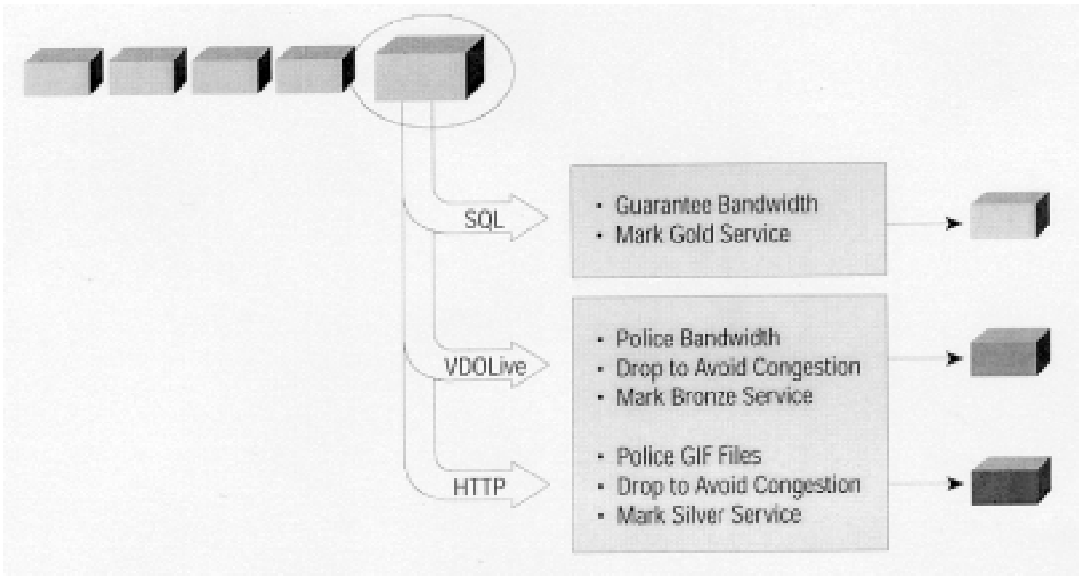

Figure 2.1.a: One example of using the CoS to produce Quality for network.

2.1 Class of Service

Rapid growth in the Internet and Intranet deployment and usage has resulted in a major shift in both corporate and consumer computing paradigms. This shift has resulted in massive increases in demand for network bandwidth, performance and flexibility to support both existing and emerging applications and services. However, this demand has often left Internet Service Providers (ISPs) with insufficient network capabilities to fully leverage the opportunity. In response to these demands, the new Class of Service (CoS) have been introduced to network services. This set of services will enable a new Intelligent Internet business model wherein the Internet quality can be implemented by rapidly defining, deploying and charging for the differentiated services targeted at the specific requirements including efficient handling of both mission critical and bandwidth hungry web applications.

Traffic classification and prioritization are the key factors of the CoS. The network must be able to efficiently sort and classify packets into traffic classes or service levels for the appropriate network handling to meet the application requirements and the willingness to pay for the services. Both service providers as well as the customer applications must have the capability to classify traffic but the service provider must be able to override customer classifications under the appropriate conditions [CNM99].

IP Precedence for Traffic Classification

IP precedence provides the capability to partition the traffic into multiple classes of service. The network operator may define for example 6 classes of service and then use network policy ruleset to define network policies in terms of congestion handling and bandwidth allocation for each class. The IP Precedence uses the 3 precedence bits in the Type-of-Service field in the IP header to specify class of service assignment for each packet. The IP Precedence feature provides considerable flexibility for precedence assignment including network content class assignment based on IP or MAC address, physical port, or application (see Figure 2.1.a for an example).

The IP Precedence feature enables the network to act either in passive mode (accepting precedence assigned by the customer) or in active mode using defined policies to either set or override the precedence assignment. IP Precedence can be mapped into adjacent technologies (e.g. Tag Switching, Frame Relay, Ethernet or ATM) to deliver end-to-end CoS policies in a heterogeneous network environment. Thus, IP Precedence enables service classes to be established with no changes to the existing applications and with no complicated network signaling requirements.

2.2 Quality of Service

Quality of Service (QoS) is a ability of a network element to have some level of predefined assurance that its traffic and service requirements can be satisfied. Enabling QoS requires cooperation of all the OSI network layers from top-to-bottom, as well as every network element from end-to-end. QoS assurances are only as good as the weakest link in the “chain” between the sender and the receiver.

QoS does not create bandwidth. QoS only manages bandwidth according to application demands and network management settings, and in that regard it cannot provide certainty if it involves sharing [VKB95]. Hence, QoS with a guaranteed service level requires resource allocation to individual data streams. Considering that bandwidth is a finite resource, a priority for QoS designers has been to ensure that best-effort traffic is not starved after the reservations are made.

CoS and QoS compared

Class of Service (CoS) is not Quality of Service (QoS). Network classification solves some of the problems, but only if one has a total control over the Network. This is not possible at the Internet, since it consists of huge amount of networks owned by different organisations. Some of the new breed of Internet applications include multimedia and require significant bandwidth. Others have strict timing requirements, or function one-to-many or many-to-many (multicast). These require network services beyond the simple "best-effort" service that CoS delivers. CoS is still “best-effort” service and it contains only a one fragment of Internet even in the best case [CNM99].

IP Telephony (VoIP) is today's "killer application." More than any other, the desire to provide telephone service over the Internet is driving the convergence of the telephone and Internet industries. This is quite interesting, since the design principles behind the telephone networks are almost exactly the opposite as those behind IP and Internet. (Whereas IP uses packet-switching with “best effort” services, telephone networks use circuit-switching to provide provisioned service). [Sta99_nq]

Distributed QoS

Multimedia content available over the Internet is often considered as distributed. This kind of multimedia can be either presentational or conversational. Presentational multimedia is based on some predefined content, which cannot be altered (e.g. Video-on-demand) or adjusted to fit the available network resources. Conversational multimedia is produced at the time of request by the interaction between multimedia system and user. Conversational multimedia requires a separate on-demand connection for controlling the data streams and a multicast service to save the finite network resources [VBK95].

The Quality depends on the actual time taken by the end-to-end travel and also on the amount of the successful transmissions [AKR94]. This is while usually a small jitter and latency are required on QoS systems. Every part of the end-to-end system has to participate on fulfilling the requirements, because the result is almost always compromised with the resources available and with the services ordered. This is a service that should be offered by the operating system of QoS system client [AFK95].

Compression influences the QoS-requirements. Packing methods, such as intraframe compression, interframe compression and layered compression, have all different effects on the QoS. The idea is that the layered compression combines both the interframe compression and the intraframe compression. Intraframe compression (e.g. JPEG compression) is used for single object and interframe compression (e.g. MPEG compression) for series of intraframe compressed objects [VKB95]. The QoS parameters that may be changed with this kind of distribute compressed multimedia are for example picture resolution, colour depth, mono/stereo-audio and sampling rate [AFK95]. General QoS parameters are frame sizes, system performance and largest allowed jitter and latency. Communication protocols offer QoS services at four different layers of OSI-model: lower layer, network and transport layer and application layer [VKB95].

2.3 Bandwidth, latency and jitter of Real-Time systems

The amount of bandwidth that an Real-Time application requires varies greatly depending on the QoS scheme used. Today's multimedia applications have a wide range of bandwidth requirements and a corresponding wide range of price points. Many useful multimedia applications are possible at reasonable data rates. Among them are the following:

- 64 kbps, telephone-quality audio,
- 100 kbps, simple application sharing,
- 128 kbps to 1 Mbps, videoconferencing,
- 1.54 Mbps, MPEG video,
- 8 Mbps to 100 Mbps, imaging,
- More than 100 Mbps, virtual reality,

Most of today's business computers are on a shared LAN such as Ethernet or a Token Ring. There are many networks that have between 10 and 100 users per LAN segment, meaning that each host on these LANs can use 50 to 100 kbps of bandwidth. This bandwidth allows them to run simple multimedia applications such as shared applications and transfers of simple multimedia imagefiles.

Latency

Real-time, interactive applications such as desktop conferencing are sensitive to accumulated delay (latency). For example, telephone networks are engineered to provide less than 400 milliseconds round-trip latency. Multimedia networks that support desktop audio/video conferencing also need to be engineered with a latency budget that is less than 400 ms round-trip [CNM99].

The round-trip latency budget is consumed by the sending computer, the network, and the receiving computer. As a rule of thumb, the sending computer will take a few milliseconds to send a packet. The network contributes to latency in several ways, including propagation delay, transmission delay, store-and-forward delay, and processing delay.

- Propagation delay is the length of time it takes information to travel the distance of the line. This is essentially controlled by the speed of light and is independent of the networking technology used. As a rule of thumb, it takes approximately 20 ms to send information between San Francisco and New York.

- Transmission delay is the length of time it takes to put a packet on the media. Transmission delay is determined by the speed of the media and the size of the packet.

- Store-and-forward delay is the length of time it takes for an internetworking device such as a switch, bridge, or router to receive a packet before it can send it.

- Most internetworking devices receive a packet before sending it out on another interface. The amount of delay that is introduced depends on the size of the packet and the speed of the media. Processing delay consists of steps such as looking up a route and changing a header. When a packet comes in, the networking device (bridge, router, or switch) needs to decide which interface it should be sent out on. In some cases, the packet also needs to be manipulated (by changing the data link layer encapsulation, changing the hop count, etc.).

Voice communication requires low latency in order to maintain audio quality. Telephone networks are typically engineered to keep end-to-end latency below 400 ms. Data networks that carry voice traffic will need to be engineered similarly.

Jitter

When a network provides variable latency for different packets, it introduces jitter. Jitter is particularly bad for audio streams, because it can cause audible pops and clicks that can be disruptive to communications. Multimedia networks must provide techniques that minimize jitter for traffic that is adversely affected by it [CNM99].

Priority queuing. Network administrators can request that high-priority traffic be queued ahead of other traffic. Multimedia traffic that is sensitive to jitter should be treated as high-priority traffic.

Custom queuing. Bandwidth can be reserved so that different streams of data are guaranteed a minimum quantity of bandwidth. This feature allows multimedia traffic to have low jitter while still permitting other network applications to run effectively.

In addition, applications frequently provide techniques that minimize jitter. The most common technique is to have the network place the data into a buffer that the display software or hardware pulls data from. The insulating buffer reduces the effect of jitter the same way that a shock absorber reduces the effect of road irregularities on a car; variations on the input side are smaller than the total buffer size and are therefore not observable on the output side.

2.4 Network technologies and Quality

Network layer (Layer II) protocols (such as Ethernet, FDDI, Token Ring, ATM and ISDN) offer various Quality services. The Network layer (Layer II) protocols ATM and ISDN have a native support for QoS requirements, when Ethernet, FDDI and Token Ring networks have nothing to offer (except than some CoS functions) at layer II [VKB95]. The Quality services for all the Ethernet networks are handled at the Layer III, where the most common way is to use link optimization techniques best suited for that specific content [Cha97].

ATM

One technology for extending integrated services networks is to use ATM (B-ISDN). This solution holds the promise of guaranteed quality of service for different classes of traffic. There are three problems with using ATM for quality-of-service guarantees today:

- ATM technology is unlikely to be available end to end in most networks (the major LAN media is Ethernet, FDDI and Token Ring). If ATM is not used end to end, the applications do not receive the benefit of the ATM quality-of-service guarantees.
- Today's ATM standards for handling flow control and congestion are not complete. As a result, ATM only works well if bandwidth is overengineered and if the ATM switches have sufficient buffering to handle the traffic pattern. Many early ATM switches have inadequate buffering and primitive queuing algorithms.
- Today's multimedia applications are being written for a generic network interface. They do not include the ability to make quality-of-service requests. As a result, the ATM switches do not know what quality of service is required for different applications.

When comparing the Ethernet and the ATM techniques we easily find the ATM suitable for transporting mixed content of the IP and voice traffic. ATM is also more suitable for teleoperators for it's better load balancing and high availability preferences. Ethernet is then again better and cheaper solution for Local Area Networks (LANs) where the network is propably used by one organisation only. VLAN techniques can be used to expand the LANs wider area but the underlaying network must support the ethernet frames (ATM networks may consist of different kind of fragments and the PVCs can still be constructed).

3. Multimedia protocols

Traditional network functions, such as file transfers, are not sensitive to delay. While network users may prefer that a file transfer occur quickly, the transfer will take place regardless of the amount of time it takes. Traffic generated by these applications is called elastic, because it can stretch to work under any delay conditions. New multimedia network applications, such as audio and video, require that certain minimum numbers of bits be transferred within a specific timeframe or they lose all coherence. To arrive successfully, the inelastic traffic generated by these applications requires the network to allocate specific resources for them and to pass information about the nature of the transferred content.

This chapter describes the new protocols designed by the Internet Engineering Task Force (IETF). The new protocols - RSVP, RTP RTCP and RTSP - are all described by their preferences and by their actual use. The mission of RSVP is to allow network nodes to communicate among themselves and with end systems so that they can reserve end-to-end network resources for inelastic applications. RTCP, the Real Time Control Protocol, is used to establish the audio channels themselves. RTP, the Real Time Transport Protocol, is used for transport of the real-time content stream. RTSP is an application-level protocol designed to work with the lower-level protocols like RTP, RSVP to provide a complete streaming service over the Internet [Jai98].

3.1 Resource ReSerVation Protocol (RSVP)

The Resource Reservation Protocol (RSVP) is a resource reservation setup protocol designed for an integrated services internetwork. An application invokes RSVP to request a specific QoS for a data stream. Hosts and routers use RSVP to deliver these requests to the routers along the path(s) of the data stream and to maintain router and host state to provide the requested service. RSVP does not participate in routing decision and it does not provide any routing information for the routers [Jai98].

RSVP is designed with tight co-operation with the RTP (RSVP is an open standard which means that the new highly compatible services can be built on the RSVP). As the time goes by the RSVP will be integrated as a part of Internet's infrastructure. RSVP

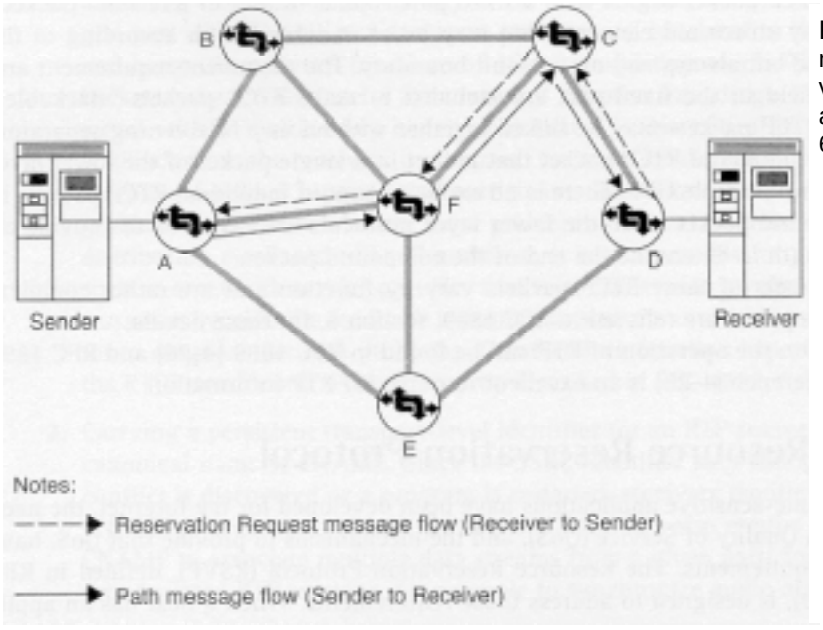

Figure 3.1.a: There is a network resource reservation with one sender and one receiver e.g. for 600-kbps video feed.

causes overhead and shortage of the network resources and therefore it's use is limited only to edges of the Internet and to the separate Intranets.

Preferences of RSVP

RSVP permits participants in flows (potentially any flow, but primarily targeting multimedia flows) to advise the network of their needs, and for the network to configure itself to meet them. The participants are identified as senders, receivers, and network elements (see Figure 3.1.a for an example of the RSVP system).

At each node (router or host) along the path, RSVP passes a new resource reservation request to an admission control routine to determine whether there are sufficient resources available. If there are, the node reserves the resources and updates its packet scheduler and classifier control parameters to provide the requested Quality of Service. The RSVP is receiver oriented which means that the receiver is responsible for taking care of the resource reservation [Whi97]. RSVP's design goals include:

- Support for multicast or unicast data delivery (multicast operation saves the network resources effectively and enables the different service levels for the different receivers.) [ZDE93],
- Reserve resources for simplex data streams,
- Minimize the application state,

- Resource maintenance for every flow,
- A "soft state" maintenance in the routers,
- Support for dynamic membership changes,
- Automatically adapt on routing changes,
- Provide several reservation models to fit a variety of applications,
- Operate transparently through nodes that do not support RSVP.

Minimisation of application state is a goal that drives many aspects of the architecture of the RSVP. In a receiver-oriented design, each participant in the flow caches just enough information to play its role. The originator of the RSVP channel (at the multicast environment) may decide if the other receivers are allowed to participate the channel. The options include possibilities to reserve the connection just for one receivers or predefined group of receivers with data filters or to allow others to share the same channel. The latter option is better for the network performance, because the data has to be transmitted only once for whole multicast group [ZDE93]. RSVP also maintains actively the connection.

RSVP in use

RSVP system consists of autonomous nodes that are all capable of resource reservation. Each node uses local procedures for reservation setup (see Figure 3.1.b). Policy control procedure determines whether the user has administrative permission to make the reservation. Admission control procedure keeps track of the system resources and determines whether the node has sufficient resources to supply the requested QoS. The RSVP daemon is a middleman that returns an error notification to the application that originated the request. If both checks succeed, the RSVP daemon sets parameters in the packet classifier and packet scheduler to obtain the requested QoS. The packet classifier determines the QoS class for each packet and the packet scheduler orders packet transmission to achieve the promised QoS for each stream. RSVP daemon also communicates with the routing process to choose the right path for reservation requests and to handle changing memberships and routes.

All the routers along the reverse data stream path use two types of RSVP messages: PATH and RESV. The PATH messages are sent periodically from the sender to the

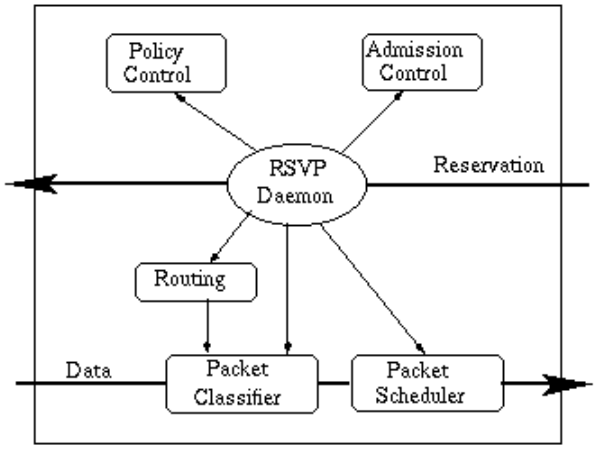

Figure 3.1.b: reservation at a node on the data flow path [Jai98]

multicast address. A PATH message contains information of sender's template (data format, source address, source port) and traffic characteristics. This information is used by receivers to find the reverse path to the sender and to determine what resources should be reserved. RESV messages contain the data filter and they are generated by the receivers. The data filter contain instructions for the packet classifier and packet scheduler procedures of RSVP. RESV messages follow the exact reverse path of PATH messages, setting up reservations for one or more senders at every node. The RSVP daemon needs to send refresh messages periodically to maintain the reservation states. Soft states enable the RSVP to easily handle membership and route changes.

RSVP guarantees that the network resources are available when the transmission actually takes place. Although RSVP sits on top of the IP in the protocol stack, it is not a routing protocol, but rather an Internet control protocol. Actually, RSVP relies on the underlying routing protocols to find where it should deliver the reservation requests [Sta99_pa].

3.2 Real-Time Protocol (RTP)

Real-time Transport Protocol (RTP) is an IP-based protocol providing support for the transport of real-time data such as video and audio streams. The services provided by RTP include time reconstruction, loss detection, security and content identification. RTP is primarily designed for multicast of real-time data, but it can be also used in unicast. It

can be used for one-way transport such as video-on-demand as well as interactive services such as the VoIP [Mil00 Pages 124-140].

RTP is designed to work in conjunction with the auxiliary control protocol RTCP to get feedback on quality of data transmission and information about participants in the on-going session. RTP uses UDP as a transport mechanism because it has lower delay than transmission control protocol (TCP), and because the actual voice traffic, unlike data traffic or signaling, tolerates low levels of loss and cannot effectively exploit retransmission [SCF96].

Preferences of RTP

Multimedia applications require appropriate timing in data transmission and playing back. RTP provides timestamping, sequence numbering, and other mechanisms to take care of the timing issues. Through these mechanisms, RTP provides end-to-end transport for real-time data over datagram network (IP network).

- RTP provides end-to-end delivery services for data with real-time characteristics, such as interactive audio and video. RTP itself does not provide any mechanism to ensure timely delivery (support is needed from lower layers such as RSVP for QoS)
- RTP assumes that the underlaying network provides framing.
- RTP does not offer any form of reliability or flow/congestion control. RTP passes the information, but the implementation is totally left to the application.

Timestamping is used when the sender sets the timestamp according to the first octet of the whole packet. The receiver uses the timestamp to reconstruct the original timing and synchronisation of different sources in order to construct the whole multimedia session. UDP does not deliver packets in timely order or guarantee that they should ever reach

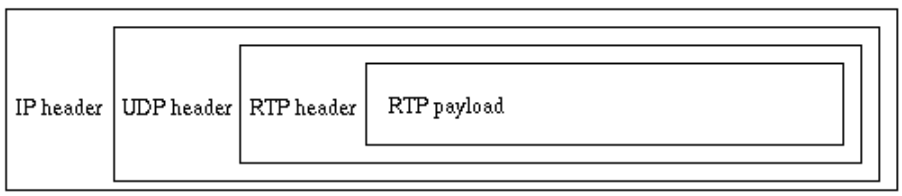

Figure 3.2.a: RTP data in an IP packet

their destination, so sequence numbers are used to place the incoming data packets in the correct order.

The RTP payload type identifier specifies the payload format as well as the encoding/compression schemes. Payload type identifier enables the receiving application to know how to interpret and present the content (RTP supports PCM, MPEG1/MPEG2 audio and video, JPEG video, Sun CellB video and H.261 video streams by default). Custom payload types can be added by providing a profile and payload format specification (RTP sender can only send one type of payload for one RTP session at a time, although the payload type may change during transmission).

Source identification allows the receiving application to know where the data is coming from. For example, in an audio conference, from the source identifier a user could tell who is talking.

RTP in use

RTP is usually implemented within the application, because issues such as lost packet recovery and congestion control have to be implemented in the application level. To set up an RTP session, the application defines a particular pair of destination transport addresses (one network address plus a pair of UDP ports for RTP and RTCP). In a multimedia session, each medium is carried in a separate RTP session, with its own RTCP packets reporting the reception quality for that session. For example, audio and video would travel on separate RTP sessions, enabling a receiver to select whether or not to receive a particular medium.

The RTP header is encapsulated in a UDP/IP packet (See Figure 3.2.a). RTP is typically run on top of UDP to make use of its multiplexing and checksum functions. UDP was chosen as the target transport protocol for RTP because of two reasons. First, RTP is primarily designed for multicast, the connection-oriented TCP does not scale well and therefore is not suitable. Second, for real-time data, reliability is not as important as timely delivery (retransmission of the TCP are not desirable). Efforts have been made to make RTP and RTCP transport-independent so they can be also run on CLNP

(Connectionless Network Protocol), IPX (Internetwork Packet Exchange), AAL5/ATM or other protocols.

Lets suppose that each participant of a multimedia conference sends audio data in segments of 20ms duration. Each segment of the audio data is preceded by an RTP header, and then the resulting RTP message is placed in a UDP packet. The RTP header indicates the type of audio encoding that is used (e.g., PCM). The encoding can be changed during the conference in reaction to network congestion or to accommodate low-bandwidth requirements when e.g. new participant arrives. Timing information and a sequence number in the RTP header are used by the receivers to reconstruct the timing produced by the source, so that in this example, audio segments are contiguously played out at the receiver every 20 ms.

3.3 Real-Time Control Protocol (RTCP)

Real-Time Control Protocol (RTCP) is a control protocol designed to work in conjunction with RTP. Within an RTP session, participants periodically send RTCP packets to convey feedback on quality of data delivery and information of membership. RTCP supports the multicast-to-unicast conversions [SCF96]. The RTCP can be used for the performance managing and the diagnostic purposes [Mil00 Pages 124-140]. .

Preferences of RTCP

RTCP passes information of the quality of the data distribution for the application. The information can be used for tuning the QoS parameters. RTCP is very important part of system, because the sender needs feedback directly and on-line from the receivers.

- RTCP provides together with RTP functionality and control mechanisms necessary for carrying real-time content. But RTCP itself is not responsible for the higher-level tasks like assembly and synchronisation. These have to be done at application level.
- The flow and congestion control information of RTP is provided by RTCP sender and receiver reports.

The control is needed for scalability. The information passed by the RTCP can be used for limiting the use of the finite resources of some part of the system. RTCP may also

carry small messages to the users, which is nice feature if for example the system keeps track on opening and closing the connections (the information can then be transported easily to other users with only little overhead) [Jai98].

RTCP in use

RTCP keeps track on every user of Real-Time system by using canonical name (CNAME). CNAME is key value that identifies the user and enables the synchronisation of different RTP sessions at application level [SCF96]. RTCP send control packets every 5 seconds to all the members of the system. Specific rapport packets are used deliver traffic information: Sender reports, SR, for those who send information and receiver reports, RR, to those who only receive information from the system. The sender informs the members of Real-Time system with SDES-control packets of a new recipients and with the BYE-control packets of the terminated connections [SCF96].

3.4 Real-Time Streaming Protocol (RTSP)

Real-Time Streaming Protocol (RTSP) is a protocol for controlling the delivery of data with real time properties such as audio and video streams. It is useful for large-scale broadcasts and audio or video on demand streaming, and is supported by a variety of vendors of streaming audio and video multimedia (including Cisco IP/TV, RealNetworks RealAudio G2 Player, and Apple QuickTime 4 software).

Preferences of RTSP

RTSP is a client-server multimedia presentation protocol to enable controlled delivery of streamed multimedia data over IP network. It provides advance remote control functionality for audio and video streams (just like VCR) [SRL97].

- RTSP is an application level protocol with syntax and operations similar to HTTP, but works for audio and video. It uses URLs like those in HTTP. Unlike HTTP, in RTSP both servers and clients can issue requests [SRL97].
- RTSP messages are be carried out-of-band. The protocol for RTSP may be different from the data delivery protocol. An RTSP server needs to maintain states, using SETUP, TEARDOWN and other methods.

- RTSP is implemented on multiple operating system platforms and it allows interoperability between clients and servers from different manufacturers [Jai98].

RTSP in use

RTSP is allowed to run over either UDP or TCP, though all commercial RTSP servers are TCP-based. RTSP establishes a TCP-based control connection, or channel, between the multimedia client and server. RTSP uses this channel to control commands such as "play" and "pause" between the client and server. These requests and responses are text-based and are similar to HTTP.

RTSP does not typically deliver continuous data streams over the control channel, usually relying on a UDP-based data transport protocol such as standard RTP to open separate channels for data and for RTCP messages. Typically, RTP and RTCP channels occur in pairs, with RTP being an even numbered port, and RTCP channel being the next consecutive port.

RTSP is easily expandable. The HTTP and MIME server can read the RTSP directly which increases the interoperability with the old systems. RTSP request can be used through a proxy services and they can also be used over secure tunnel (VPN or PPTP). RTSP contains mechanism for creating a multi-server environment, where the RTSP may act as a load balancer. RTSP in bidirectional which enables the clients to receive information about the service updates and get the most recent information without delays [SRL97]. RTSP can use both unicasts and multicast for transmitting the data to the clients. If the RTSP uses multicast for sending the data the server chooses the correct multicast group for the client (his happens typically if the service is not interactive, but rather presentative such as near-media-on-demand or IPTV) [SRL97].

4. Advanced Services over IP

No communications innovation in history has been adopted as quickly as the Internet. The World Wide Web has reached more than 50 million users. Compare this rapid adoption rate with that of the personal computer, which took 16 years, and with television, which took 13 years to reach the same penetration level.

Internet adoption is a tornado, roaring into the 21st century and catching up millions of users with its inherent power. There are expectations of 300 million worldwide users to be riding that tornado by 2002. This whirlwind adoption rate will hasten the adoption of IP-based services ubiquitously delivered through the communication vehicle known as the Internet.

Basic Internet access secures a large customer base of both consumers and business users. The increased competition has commoditized the service resulting in low prices and low margins with little differentiation. The challenge of the ISPs is to create competition by building differentiation on top of their Advanced Intelligent Network (AIN) infrastructure.

4.1 Advanced Intelligent Network

In the recent years the Advance Intelligent Network (AIN) has been widely accepted by the telecom and computer industries. The AIN is merely a network-based service-independent architecture for all telecom networks. It’s merit is to introduce and develop the network-based information services intelligently, rapidly and economically. [KuL98]

SS7 Technology Overview

Currently the information services created on the AIN are quite a few and simple, which can hardly reveal the potential of the AIN. The SS7 is a set of standards for the Common Channel Signaling (CCS) system. SS7 defines the architecture, network elements, interfaces, protocols, and management procedures for a Public Switched Telephone Network (PSTN) that transports control information between network switches and between switches and databases. SS7 is used between the PSTN switches, replacing per-trunk, in-band signaling (Typically on a separate PSTN data network). The

SS7 network is used for network control and the only data sent over it is signaling messages [KuL98].

Personal Communications Service

Personal Communications Service (PCS) is a Advanced Intelligent Network architecture that provides personal, terminal, and service mobility. PCS has been allocated for broadband, narrowband, and unlicensed services. Together the SS7 and the PCS will integrate the wired and wireless networks together to provide users with personal and mobile end-to-end connections. The basic needs of users (such as voice communications, email, file transfer and remote login) are there even today. However the need for more specific service features in different places at different times under different cisrcumstances are the reasons for PCS evolution. PCS should provide end-user-controlled capabilities for service customisation, association and differentiation in addition to location mobility and calling number uniqueness. The crucial technical issues of the PCS are:

- Internetworking of different networks (AIN, Internet, UMTS, GPRS),
- Network control (integration of the SS7 for controlling purposes),
- Network resource management (QoS negotiations, high availability of distributed networks resources such as Databases and function specific server),
- Integration and co-ordination (enables the multimedia support integration on the e.g. ATM networks),
- Network Security (enables the use of third party security associations from the Internet together with the telecom network security features) [KuL98].

Personalised Internet

As the Internet becomes more widespread and the user base becomes increasingly diverse, many users are demanding a Personalised Internet service that caters to their individual preferences and requirements. Service providers can build upon their Premium Internet service to further move up the value chain with their Internet customers. Business users, parents of children who are on-line, and consumers require different Internet experiences to address their requirements for performance, reliability, content, and price. Service providers can deliver higher performance and reliability to business

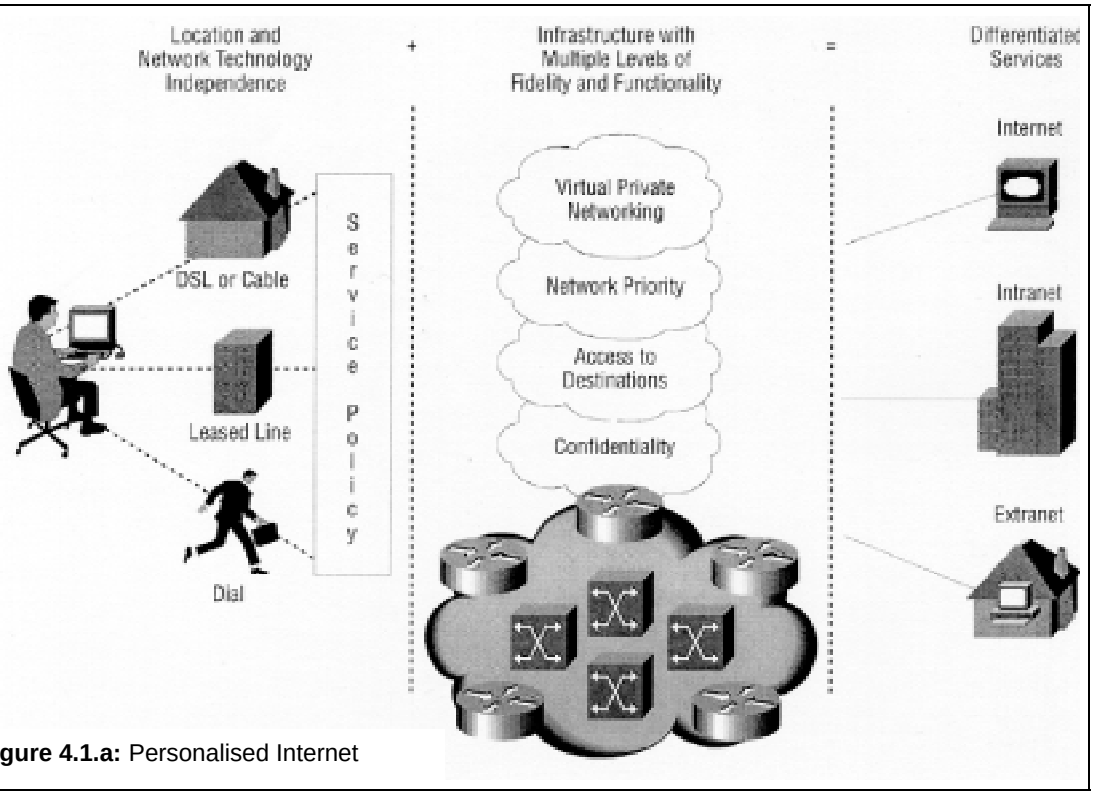

Figure 4.1.a: Personalised Internet

users, offer safe Web surfing to parents, and provide specialised content for a specific consumer community, perhaps defined by religious or ethnic affiliations or school associations.

Personalised Internet leverages the higher service levels of Premium Internet and it uses authentication, authorisation and directory-enabled networking to define a user's privileges matching user's networks and applications. For specialised content, caching and load balancing optimises performance when Web surfing. Further, flexible bandwidth choices including dial, DSL, cable, and wireless ensures users can enjoy a high-quality Internet experience regardless of application or location (See Figure 4.1.a).

4.2 IPSec Network Security

IPSec is a framework of open standards developed by the Internet Engineering Task Force (IETF). IPSec provides security for transmission of sensitive information over unprotected networks such as the Internet. IPSec acts at the network layer, protecting and authenticating IP packets between participating IPSec devices ("peers"). IPSec provides the following network security services. These services are optional. In general, local security policy will dictate the use of one or more of these services:

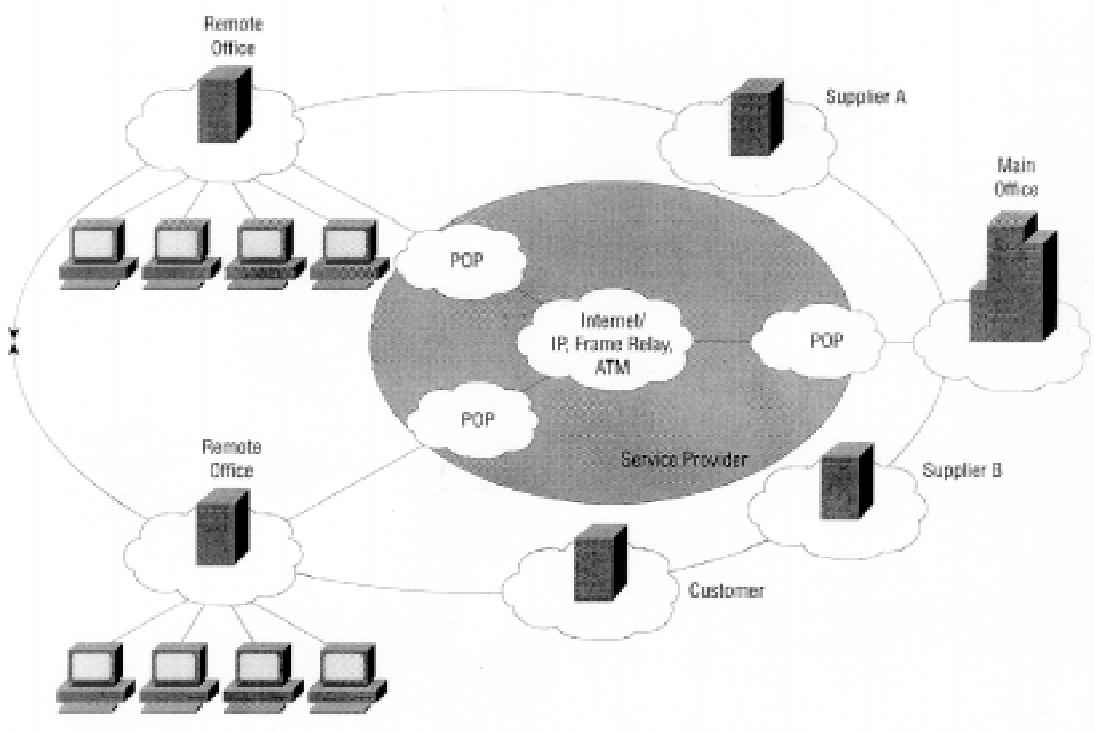

Figure 4.2.a: Example of the Internet with the IPSec.

- Data Confidentiality (The IPSec sender can encrypt packets before transmitting them across a network),
- Data Integrity (The IPSec receiver can authenticate packets sent by the IPSec sender to ensure that the data has not been altered during transmission).
- Data Origin Authentication (The IPSec receiver can authenticate the source of the IPSec packets sent. This service is dependent upon the data integrity service).
- Anti-Replay (The IPSec receiver can detect and reject replayed packets). [DoH99, Pages 41-56]

With IPSec, data can be transmitted across a public network without fear of observation, modification, or spoofing. This enables use of the virtual private networks (VPNs), including intranets, extranets, and remote user access. IPSec provides a more robust security solution and is standards-based. IPSec also provides data authentication and anti-replay services in addition to the data confidentiality services (see Figure 4.2.a).

How IPSec Works

In simple terms, IPSec provides secure connections (tunnels) between two peers, such as two routers of Firewalls [Opp97]. You define which packets are considered sensitive and should be sent through these secure tunnels, and you define the parameters which should be used to protect these sensitive packets, by specifying characteristics of these tunnels (encryption algorithm and authentication mechanism). Then, when the IPSec peer sees such a sensitive packet, it sets up the appropriate secure tunnel and sends the packet through the tunnel to the remote peer.

More accurately, these tunnels are sets of security associations that are established between two IPSec peers. The security associations define which protocols and algorithms should be applied to sensitive packets, and also specify the keying material to be used by the two peers. Security associations are unidirectional and are established per security protocol (AH or ESP). The IPSec enables the selection of the traffic based on source and destination address and even optionally on application used (OSI Layer 4). If no security association exists that IPSec can use to protect this traffic to the peer, IPSec uses Internet Key Exchange (IKE) to negotiate with the remote peer to set up the necessary IPSec security associations on behalf of the data flow (the parameters can usually be given manually instead of using the IKE, but it easily creates an awful situation to manage).

If IKE is used to establish the security associations, the security associations will have lifetimes so that they will periodically expire and require renegotiation (This provides an additional level of security) [DoH99, Pages 99-128] Multiple IPSec tunnels can exist between two peers to secure different data streams, and each tunnel uses a separate set of security associations. For example, some data streams might be just authenticated while other data streams are both encrypted and authenticated.

You can nest IPSec traffic to a series of IPSec peers. For example, in order for traffic to traverse multiple firewalls (and these firewalls have a policy of not letting through traffic that they themselves have not authenticated), the router needs to establish IPSec tunnels with each firewall in turn. The "nearest" firewall becomes the "outermost" IPSec peer.

4.3 Voice and Multimedia

Over the last ten years, client/server model computing has had a great impact on the information technology. Client/server model computing has enhanced users' productivity, build challenges for computer networking and restructured the computer industry. Today, another new technology is poised to impact business computing in an equally dramatic way. Networked multimedia and voice computer applications will significantly affect users and network managers and have a tremendous impact on computing and network infrastructures [HSK98].

Voice-over-IP (VoIP)

Voice-over-IP (VoIP) enables Network nodes to carry voice traffic (for example, telephone calls and faxes) over an IP network. In VoIP hardware, usually the digital signal processor (DSP) segments the voice signal into frames and stores them in voice packets. These voice packets are transported using IP in compliance with the International Telecommunications Union-Telecommunications (ITU-T) specification H.323 [Mil00, Pages 177-181], the specification for transmitting multimedia (voice, video, and data) across a network. Because it is a delay-sensitive application, you need to have a well-engineered, end-to-end network witch supports QoS (sometimes the CoS satisfies the needs) to successfully use VoIP (See Figure 4.3.a).

Voice-over-IP explained

VoIP equipment can be software feature of the router, VoIP telephone (with DSP circuit) or software installed on users workstation. The general flow of a two-party voice call using VoIP is as follows:

1. The user picks up the handset; this signals an off-hook condition to the signaling application part of VoIP in the router.
2. The session application part of VoIP issues a dial tone and waits for the user to dial a telephone number.
3. The user dials the telephone number; those numbers are accumulated and stored by the session application.

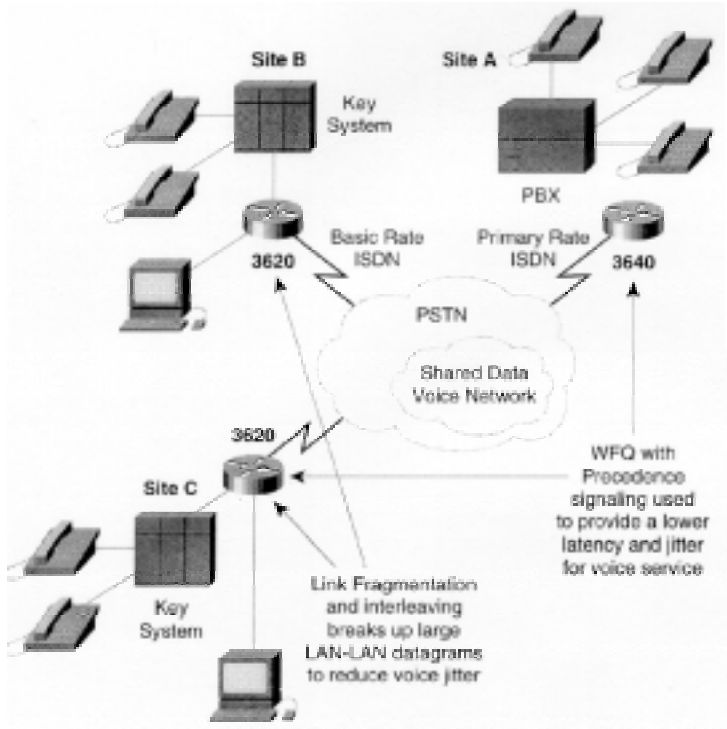

Figure 4.3.a: This diagram provides an overview of a QoS and VoIP solutions. VoIP solutions are possible only with end-to-end QoS.

4. After enough digits are accumulated to match a configured destination pattern, the telephone number is mapped to an IP host via the dial plan mapper. The IP host has a direct connection to either the destination telephone number or a Private Branch Exchange (PBX) that is responsible for completing the call to the configured destination pattern.
5. The session application then runs the H.323 session protocol to establish a transmission and a reception channel for each direction over the IP network. If the call is being handled by a PBX, the PBX forwards the call to the destination telephone. If Resource Reservation Protocol (RSVP) has been configured, the RSVP reservations are put into effect to achieve the desired QoS over the IP network.
6. The coder-decoder compression schemes (CODECs) are enabled for both ends of the connection and the conversation proceeds using Real-Time Transport Protocol (RTP/UDP/IP) as the protocol stack.
7. Any call-progress indications (or other signals that can be carried inband) are cut through the voice path as soon as end-to-end audio channel is established. Signaling that can be detected by the voice ports is also carried over the IP network encapsulated in the Real-Time Transport Control Protocol (RTCP).
8. When either end of the call hangs up, the RSVP reservations are torn down. Then each end waits for the next off-hook condition to trigger another call setup.

Figure 4.3.b: The Cisco IP Phone 7960 voice instrument: a second-generation full-featured IP phone. Features: Messages, Directories (including the firm phone book), Settings (different ringing tones) and Services (supports for example XML).

VoIP enables the companies to save from the telecom costs by combining the existing data connections with voice data. The new technology enables also the videoconferencing features as well as combining the Customer Relation Management (CRM) applications with the on-line information provided by the VoIP infrastructure (for example of IP technology see Figure 4.3.b). [Mil00, Pages 141-172]

Case study: Cisco IP/TV

The Cisco IP/TV Product Family is a network video streaming solution. It delivers TV-quality video programming to desktop PCs by leveraging IP/TV software and IP/TV Server. It's a full solution that offers organisations a fast and effective way to deliver business communications to their employees. This product enables the networked users to watch management broadcasts, training programs, university classes, business TV and other programs from the convenience of their own desktops.

The Cisco IP/TV family offers high-quality video broadcasting and video-on-demand services using IP multicast. Cisco IP/TV software is built on Microsoft's Windows Media Technologies, which enable Windows Media Tools for creating and editing live and on-demand ASF content; advanced compression software such as Windows Media Audio and MPEG-4; and Windows Media On-Demand Producer which lets users synchronise Web pages or an audio or video track to create a multimedia presentation. Cisco IP/TV products use standard protocols running on existing IP networks. A quick look at some of the ways businesses are using IP/TV:

- Training and Distance Learning: For the employees at brand office who need product training, but the training centre is far away. Bring the training centre to the employee; it's much less expensive than sending the employee to the training centre.
- Business TV to the Desktop: For organisations which include knowledge workers who need instant access to international business or financial developments (stock market trends, financial news or satellite broadcasts to knowledge workers).
- Corporate Communications: To every employee who wishes have the same information [CIT00].

4.4 Large Scale Storage solutions

Driven by the explosive growth of data-intensive applications and businesses' escalating reliance on information as a competitive differentiator, storage is rapidly becoming a central component of corporate technology strategies. The new storage-centric infrastructure that industries envision includes an open, modular, scalable storage network not tied to any one server or application.

The objective is a single, massive data repository accessible by any person, application or system across the entire organisation. Emerging technologies such as network-attached storage (NAS) and the storage area network (SAN) promise both a single source of data and improved performance and availability.

Case Study: IBM Storage Area Network

A SAN is a high-speed network dedicated to information management. More formally, a SAN is a combination of technologies (including hardware, software and networking components) that provides any-to-any interconnection of server and storage elements.

By separating information management from information processing, a SAN provides the flexibility required to meet the new computing requirements defined above. More to the point, by enabling the free, immediate exchange of information, a SAN provides the foundation for a "zero latency" computing environment.

SANs are based on a "fabric" of Fibre Channel hubs, switches and gateways connecting storage devices (such as disk arrays, optical disks or tape libraries) and servers on a

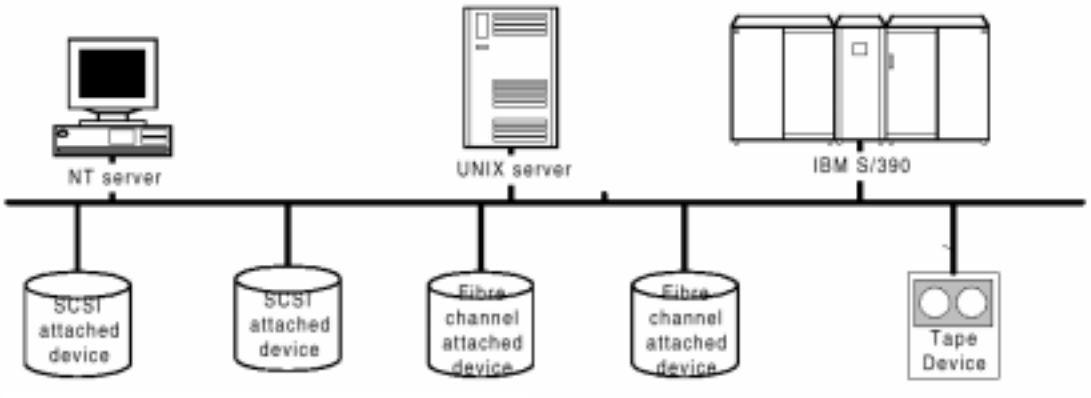

Figure 4.4.a: The SAN "Data Bus"

many-to-many basis (see Figure 4.4.a for an example). Application and transaction servers are attached to both the SAN and to Local Area Networks (LANs) or Wide Area Networks (WANs), creating what appears to be a massive pool of data. By reducing the amount of traffic that must travel along corporate networks, installing a SAN has the effect of increasing available bandwidth. The result is response times that are well matched to the requirements of e-business.

To flourish in the Information Age, a company must be able to make information as pervasive and easily accessible as electricity or telephone service. Such a corporate "information utility" would permit on-demand exchange of information across functional and departmental lines, forging new connections to customers and suppliers. Finding practical and affordable ways to link everyone to the flow of information has become a strategic necessity [Orr99].

5. Conclusions

Internet community around the world is busy getting themselves to be ready for the 21st century. The Internet is growing exponentially. The well established LAN and WAN technologies based on IP protocol suite connect bigger and bigger networks all over the world to the Internet. In fact, Internet has become the platform of most networking activities. This is the primary reason to develop multimedia protocols over the Internet. Another benefit of running multimedia over IP is that users can have integrated data and multimedia service over one single network, without investing on another network hardware and building the interface between two networks.

QoS is a fundamental service required for next-generation business traffic. As enterprises shift mission-critical applications to the network and create a unified multiservice architecture for data, voice, and video, the ability to manage delivery terms becomes increasingly critical. Enterprises need QoS to deploy on-demand bandwidth-intensive applications such as videoconferencing, as well as time-sensitive information delivery applications such as stock transactions.

The same kind of expand frenzy is currently going on in office buildings and plants around the world. Today, organisations are moving as quickly as they can all of their knowledge workers and business partners with high-speed connections. Through advances in Internet and Intranet communications and network technology the new world with the new IP-based infrastructure is going to expand into new fields of unused business potential - the homes of people.

Multimedia networking applications such as videoconferencing, distance learning, kiosks, multimedia books, multimedia mail, and simulations can add significant value to today's businesses. Adding these applications to the existing data network with the IPSecurity model is a cost-effective way of obtaining these benefits. The only question that remains unsolvable is that will the speed of technical evolution keep up and will the people (the consumers) buy this new technology.

References

AFK95 Ronnie T. Apteker, James A. Fisher, Valentin S. Kisimov, Hanoch Neishlos, Video Acceptability and Frame Rate. IEEE Multimedia Fall (1995), 32-40

AKR94 Caglan M. Aras, James F. Kurose, Douglas S. Reeves, Henning Schulzrinne, Real-time communication in the Packet-Switched Networks. Proceedings of the IEEE 82,1 (1994), 122-139.

Cha97 Samir Chatterjee, Requirements for Success in Gigabit Networking, Communications of the ACM, July 1997 / Vol 40, no. 7, 64-73.

CIT00 Cisco Systems, Cisco Products & Technologies, Cisco IP/TV 3400 Video Servers, Internet article made available at http://www.cisco.com/warp/public/cc/pd/mxsv/iptv3400/index.shtml

CNM99 Cisco White Paper, Networked Multimedia Overview, published Jun 14 1999, Internet article made available at http://www.cisco.com/warp/public/614/19.html

DoH99 Naganand Doraswamy, Dan Harkins, IPSec: The New Security Standard for the Internet, Intranets, and Virtual Private Networks, Prentice Hall PTR , Internet Infrastructure Series, ISBN 0-13-011898-2, (1999)

HSK98 Vicky Hardman, Martina Angela Sasse, Isidor Kouvelas, Successful Multiparty Audio Communication over the Internet, Communications of the ACM, May 1998 / Vol 41, no. 5, 74-80.

Jai98 Raj Jain, Multimedia over IP: RSVP, RTP, RTCP, RTSP. Internet article made available at http://www.cis.ohio-state.edu/~cliu/ipmultimedia/

KuL98 Geng-Sheng Kuo, Jing-Pei Lin, New design concepts for an Intelligent Internet, Communications of the ACM, November 1998 / Vol 41, no. 11, 93-98.

Mil00 Mark A. Miller, Voice over IP: Strategies for the converged Network. The M&T Books, ISBN 0-7645-4617-1 (2000).

Opp97 Rolf Oppliger, Internet Security: Firewalls and Beyond, Communications of the ACM, May 1997 / Vol 40, no. 5, 92-102.

Orr99 Ken Orr, Enterprise Storage Area Networks: "Data Wiring" the Enterprise. Internet article made available at http://www.storage.ibm.com/ibmsan/whitepaper/datawiring.htm

SCF96 H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, RTP: A Transport Protocol for Real-time Applications. Internet Engineering Task Force, RFC 1889 (1996).

SRL97 H. Schulzrinne, A. Rao, R. Lanphier, Real Time Streaming Protocol (RTSP). Internet Engineering Task Force, IETF (1997).

Sta99_nq Stardust.com, White Paper - The Need for QoS, Stardust.com Inc (1999)

Sta99_pa Stardust.com, White Paper - QoS protocols & architecture, Stardust.com Inc (1999)

Whi97 Paul P. White. RSVP and Integrated Services in the Internet: A Tutorial. IEEE Communications Magazine 35,5 (1997), 100-107.

VKB95 Andreas Vogel, Brigitte Kerherve, Gregor Von Bochmann, Jan Gecsei, Distributed Multimedia and QoS: A Survey. IEEE Multimedia Summer (1995), 10-19.

ZDE93 Lixia Zhang, Stephen Deering, Deborah Estrin, Scott Shenker, Daniel Zappala. RSVP: A New Resource ReSerVation Protocol. IEEE Network Magazine 7,5 (1993), 8-18.