3GPP2 S.R0106-A
Version 1.0
Version Date: 15 May 2008

Packet-Switched Video Telephony

Stage 1 Requirements

EDITORS

Version 1.0: Aleksandar Gogic, +1 858 651 5386, agogic@qualcomm.com

Version 2.0: Mike Dolan, +1 630 979 1033, mfdolan@alcatel-lucent.com

REVISION HISTORY

REVISION HISTORY		
Ver. 1.0	*Initial Publication*	*21 July 2005*
Ver. 2.0	*Version Update*	*15 February 2007*
Rev. A v. 1.0	*Revision A Initial Publication*	*15 May 2008*

Table of Contents

1 Scope

This document defines the functional characteristics and the requirements of the packet-switched Video Telephony (VT) services, sometimes also referred to as Multimedia Conversational Services (MCS). Video Telephony is defined as one-to-one video/speech communication capability. VT (MCS) should be used as a basis for multiparty multimedia conferencing, but this is out of scope for the initial document revision.

This initial effort is to be mindful of the continued development in subsequent phases of the following enhancements (without implying any relative priority):

- Interoperability support to other video telephony systems;
- Capability of multipoint communication.

Hence the following two phases in the development of video telephony are defined:

Phase 1: Development of basic one-on-one video telephony functionality (Subject to Stage 1 Revision A or earlier)

Phase 2: Enhancement for support of other video telephony systems and multipoint communication (Subject to Stage 1 Revision B or later)

2 References

[1] ITU-T Recommendation H.323: “Visual Telephone Systems and Equipment for Local Area Networks which Provide a Non-Guaranteed Quality of Service”

[2] ITU-T Recommendation H.324: “Terminal for Low Bitrate Multimedia Communication”

[3] C.S0009-0 v1.0 Speech Service Option Standard for Wideband Spread Spectrum Systems

[4] C.S0014-A v1.0 Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems

[5] C.S0020-A v1.0 High Rate Speech Service Option 17 for Wide Band Spread Spectrum Communication Systems

[6] C.S0030-0 v3.0 Selectable Mode Vocoder (SMV) Service Option for Wideband Spread Spectrum Communication Systems

[7] C.S0052-A v1.0 Source-Controlled Variable-Rate Multimode Wideband Speech Codec (VMR-WB), Service Options 62 and 63 for Spread Spectrum Systems

[8] IETF RFC 3261: “Session Initiation Protocol” (SIP)

[9] IETF RFC 3550: “RTP: A Transport Protocol for Real-Time Applications” (a.k.a. Real Time Protocol) - IETF, July 2003

[10] S.R0108 HRPD-cdma2000 1x Interoperability for Voice and Data SRD

[11] C.S0014-B v1.0 Enhanced Variable Rate Codec, Speech Service Option 3 and 68 for Wideband Spread Spectrum Digital Systems,

[12] C.S0014-C v1.0 Enhanced Variable Rate Codec, Speech Service Option 3, 68 and 70 for Wideband Spread Spectrum Digital Systems

3 Abbreviations

3G	Third Generation system
3GPP2	Third Generation Partnership Project 2
BER	Bit Error Rate
CIF	ITU-T Common Intermediate Format (352 pixels x 288 lines)
FER	Frame Error Rate
IETF	Internet Engineering Task Force
ISO	International Standards Organization
ITU-T	International Telecommunication Union - Telecommunication Sector
MCS	Multimedia Conversational Services
MDN	Mobile Directory Number
MOS	Mean Opinion Score
PSNR	Peak Signal-to-Noise Ratio – a quality metric for video codec on a lossy transmission channel
QCIF	ITU-T Quarter Common Intermediate Format (176 pixels x 144 lines)
QoS	Quality of Service
RFC	Request for Comments
RLP	Radio Link Protocol
RTP	Real Time Protocol (See RFC 3550)
SIP	Session Initiation Protocol
URL	Uniform Resource Locator
VPOP	Voice Precedence over Packet
VT	Video Telephony

4 Introduction

The potential benefits of communicating via visual media, in addition to speech, have long been recognized, as greatly enhancing the potential for users to communicate and to convey information. Video telephony consists of ability of a user to simultaneously talk to another party, view video images from the other party, and send to the other party video images captured by local camera, stored in local device, etc. User can stop video communication without interrupting voice conversation.

Transmitting a video stream has proven to be a very challenging goal, since it requires significant resources, and therefore can place a heavy burden on the system. Because a video stream contains much more information than voice alone, it demands much higher data rate. In order to reduce the data rate, many video codecs are optimized for maximum compression alone, and thus are sensitive to transmission errors.

With the development of 3G wireless communications systems, the data rate available to each user can be considerably increased. The available network throughput has reached the threshold where reasonable quality video telephony services can be realized.

New developments in packet (IP based) networks as of late, have made video telephony a more viable service due to some of the following:

- Standardization of comprehensive QoS capabilities in all variants of cdma2000®[1] wireless networks;
- Payload flexibility of packet switching and routing, which enables asymmetric and variable rate bearers;
- Optimization of wireline protocols for wireless networks (e.g. protocol header compression/removal);
- Increase in number of devices (PCs and laptop computers) frequently attached to a network (wide area wireless, WLAN, or wired), in addition to the camera-equipped wireless phones.

5 Services and Features

5.1 Service Characteristics

Packet-Switched Video Telephony provides two-way transmission of real-time video, speech, combined speech/video, and other services in the wireless communications systems. These services can be used in videophone, multimedia video conferencing, long-distance classroom, video communications in fieldwork, etc. Latency (round-trip delay between the parties) needs to be kept to a minimum in order to support real-time information exchange between the parties. In general, it is not feasible to recover transmission errors by retransmission techniques, because of the delays involved in requesting retransmission. Occasionally, video and speech quality may need to be compromised to meet latency requirements. Video and speech codecs must be robust enough to cope with the possibilities of errors in wireless transmission channels. Intelligent use of error-recovery and error-concealment techniques is desirable to ensure graceful performance degradation on noisy channels.

5.2 Service Structure

A VT terminal supports the capture, transport, and playback of real-time full-duplex speech and video. The VT terminal contains a speech codec that produces good voice quality and a video codec that produces reasonable video quality. The terminal sets up, maintains, and releases video telephony sessions, and transmits and receives the following data flows:

- a low data rate packet loss sensitive control flow;
- a low data rate and delay sensitive speech flow; and
- a medium data rate delay sensitive video flow.

[1] cdma2000® is the trademark for the technical nomenclature for certain specifications and standards of the Organizational Partners (OPs) of 3GPP2. Geographically (and as of the date of publication), cdma2000® is a registered trademark of the Telecommunications Industry Association (TIA-USA) in the United States.

Assuming provision of adequate network resource (e.g. sufficient radio resource capacity to handle the traffic load), video telephony services can in principle exist alongside other wireless communication services without mutual impact.

VT services are symmetric between the parties in terms of available functionality, which includes encoding, transmission, and decoding components. However, to increase flexibility and optimize radio spectrum efficiency, the services can be implemented without requiring symmetric bearers on the forward and the reverse link, as well as to allow the user to add or drop a multimedia component, and the operator to manage access based on the network traffic load, QoS, etc.

Video and speech codecs encode the video and speech signals acquired by the video terminal. After packetization, the encoded data is transmitted by the wireless network, and delivered to the other participant of the VT call using the Internet for transport and routing. At the same time, the VT terminal processes incoming video and speech data from the remote participant by de-packetizing them, then decoding this data with corresponding video and speech codecs. The output is then played back at the local video/audio device.

Video telephony also includes system control protocols for setting up calls between parties, exchanging and negotiating various options and capabilities, and communicating with and controlling the various codecs used.

6 Service Requirements

6.1 Terminology

Some of the features and requirements listed below, though standardized, are optional for a product implementation. Standards specifications are to be developed regardless of this optional nature of a feature, so that those products supporting them would operate consistently. To make optional/mandatory distinction clear, use of terms is clarified as follows:

- The term "it shall be possible" refers to optional features that must be allowed by the service. Subject feature is optional and its implementation in the service is a decision for the operator.
- The term "shall" means that the feature is not optional to the service, and must be supported in every implementation.

6.2 System

VT-01. **Backward Compatibility:** Packet-switched VT system in this document should be interoperable with the packet-switched VT system defined in S.R0106-0.

VT-02. **Interworking.** The VT system should be interoperable with international standardized video conferencing systems (e.g., ITU-T H.324M [2]/ H.323 [1], IETF RFC 3261 SIP [8]). It should be able to interconnect with 3GPP circuit-switched VT. It should be able to exchange controlling information with peers in format complying with international standardized protocol (e.g., ITU-T H.245, H.225.0). These requirements are applicable for Phase 2 of packet-switched VT.

VT-03. **Media Component Transition.** It shall be possible to add or release a VT multimedia component (e.g. video) without interrupting other multimedia components (e.g. speech).

VT-04. **Media Component Transition Management.** When a terminal adds or releases a multimedia component, this information shall be conveyed to the correspondent terminal.

VT-05. **Addressing.** Both MDN and SIP URI addressing shall be supported.

VT-06. **Calling Party ID:** The system shall be able to provide the calling party ID and an indication of the video telephony call to the called party, subject to restrictions imposed by the calling party.

VT-07. **Mobility Management.** Provided system resources are available, it shall be possible to maintain an uninterrupted a VT call when the VT user travels within a PDSN coverage area and across multiple PDSN coverage areas that support VT.

VT-08. **Roaming.** User shall be able to obtain VT service in the coverage area of a visited PDSN that supports VT service.

VT-09. **Resource Optimization.** The system should be optimized to conserve the processing power and other resources for the mobile station and base station equipment, such as memory, power, spectrum (radio resources) consumption, etc.

VT-10. Void.

VT-11. It shall be possible to provide VT service using the functionality of IMS.

VT-12. If the VT service is based on IMS, the VT user shall be a subscriber of IMS.

VT-13. It shall be possible for the VT user to originate a VT call on an HRPD access network.

VT-14. It shall be possible for the VT user to receive a VT call on an HRPD access network.

VT-15. It shall be possible to complete (accept) a VT call attempt made to a user that is authorized for VT service but does not have VT service coverage by connecting only the audio portion of the VT call.

6.3 User Configuration

VT-16. **Configuration Options.** User shall be able to configure the following on a per-call basis:

a. Whether video or voice call is the preferred call origination type;

b. What multimedia components will be transmitted by the called party upon accepting an incoming call;

VT-17. On-screen Display Mode: The user should be able to configure the display mode as incoming or outgoing video, or both (e.g. picture-in-picture).

6.4 Service Interaction

VT-18. **Precedence over Data.** VT Service shall support the user ability to control VT precedence over other delay tolerant data services of the same user. (See

also VT-24 QoS Control.)

VT-19. **Call Waiting Notification of Incoming Voice Call during a VT Call.** It should be possible to notify a VT user of an incoming voice call. The VT service should be able to maintain an ongoing VT call if an incoming voice call notification occurs.

VT-20. **Call Waiting Notification of Incoming Video Telephony Call during a Voice Call.** It should be possible to notify a voice call user of an incoming VT call. An ongoing voice call should be maintained if an incoming VT call notification occurs.

VT-21. **Priority.** VT service and voice service should have the same priority. The VT service shall be able to maintain an ongoing VT call if an additional incoming VT call occurs. The VT service shall be able to maintain an ongoing VT call if an incoming voice call occurs.

Note: Voice service includes both circuit-switched cdma2000-1X voice service and packet-switched VoIP service.

VT-22. Void.

VT-23. Void.

VT-24. Void.

6.5 Video Codec

VT-25. **Codec Selection.** A default codec should be recommended to minimize the impact on Mobile Stations, and to eliminate or minimize the need for transcoding schemes.

VT-26. **Picture Size.** The service should be able to provide services for international standardized picture sizes, such as CIF and QCIF. It should also support user agent (application) defined non-standard picture sizes. Picture size shall be negotiable at initiation and during the VT call.

6.6 Speech Codec

VT-27. **Codec Selection.** A native 3GPP2 codec (see [3][4][5][6][7][11][12]) should be the default codec in phase 1. A native 3GPP2 wideband codec (e.g. [12] in Service Option 70 wideband mode) should be the default codec for wideband speech capable terminals.

Note: Wideband speech refers to the increased fidelity of speech signal processing made possible by increasing speech signal frequency band to about 7 kHz (16 kHz sampling rate).

VT-28. **Error Concealment:** To improve voice quality, error concealment methods in speech decoder may be included for the cases of excessive speech frame loss and excessive jitter.

Note: In general, speech quality has more impact on user experience than video quality. Therefore, the speech decoder should use one of more error concealment methods. For example, when jitter limit is exceeded, buffer underflow may occur and consequently frame skip may occur. Loss of a speech frame due to radio transmission error has a similar effect. In those

situations, it is desirable to smooth the playback, not propagate to successive frames. Note that error concealment methods are contained solely within the mobile station receiver and are not subject to standardization. They are included here since their implementation affects user experience.

6.7 Packetization

VT-29. **Transport.** Protocol header compression or removal schemes should be supported on point-to-point transmission segments within the wireless network.

VT-30. **Payload.** RTP (see [9]) shall be used for media packetization. RTP payload schemes should be recognized and recorded in the IETF (Audio/Video Transport Working Group).

6.8 Media Adaptation

Media adaptation at the video terminals to dynamic changes in the network conditions can be used to achieve improved performance. These dynamic changes are observable at the end terminals and can be fed back from the receiver (video decoder) terminal to the sender, triggering an appropriate action at the sender (video encoder) terminal to adapt to such reported network conditions within the granted QoS levels of the network.

The terminal is not required to support such media adaptation. During the PSVT session setup, the end terminals signal their capabilities to support such feedback mechanisms that enable media adaptation.

VT-31. **Video Adaptation.** It shall be possible for the sender terminal to adapt source encoding to the network conditions reported from the receiver terminal. This capability shall be possible in either directions of media flow.

7 QoS Requirements

In order to provide end users with multimedia conversational services of acceptable quality, specific QoS requirements are imposed. Acceptable performance levels mandate specific minimum end-to-end QoS levels. To meet those end-to-end requirements, QoS requirements for the wireless communication portion of the link also need to be met. QoS parameter limits in the wireless portion need to be somewhat more stringent than the required end-to-end QoS, so as to apportion a part of the QoS "budget" to the rest of the link. Unless explicitly noted, the QoS specifications herein apply if the link is controlled by the wireless network end-to-end. For the case where one end of the communication link is outside of the wireless communications network, or it does not support QoS, the QoS requirements apply only to the portion that is within the wireless network.

7.1 General

VT-32. **Admission.** The network shall have the capability to admit or block an attempted VT service instance, based on QoS conditions in the network and radio channel conditions necessary for successful execution of the VT

service.

VT-33. **Data Rate Variability.** It should be possible to vary transmission rate for VT by adapting to speech/video source content, so as to maximize radio network capacity.

7.2 QoS Attributes

The QoS attributes relevant for video telephony are listed in 8.2.1 through 8.2.6.

VT-34. **QoS Attribute Control.** The video telephony service entities should be able to set up values of QoS attributes defined in sections 8.2.1 through 8.2.6 at call setup. If not specified during setup negotiation, default value set by the network should be used.

7.2.1 Maximum and Average Data Rate (kbps)

Definition of Maximum Data Rate: The maximum number of bits (summed over all multimedia components in a VT call) delivered in a frame, divided by the frame duration.

Definition of Average Data Rate (ADR): Number of bits summed over all multimedia components and over a segment of a VT call divided by the duration of the segment.

Average and Maximum Data Rates selection should be enabled to allow specification of the target grade of service. Maximum channel data rate should be a configurable parameter that operator can set.

7.2.2 Maximum Video Frame Rate (fps)

Definition: The maximum number of video frames per second (fps).

Frame rate may be changed during a VT call as a function of source content or other considerations.

7.2.3 Maximum Transfer Delay (sec)

Definition: The maximum delay for 95th percentile of the distribution of delay for all delivered data frames during a call. The delay is measured from the time of a video frame capture and from the beginning of an audio frame (segment) at the sending party, to the output of that same video/audio frame at the receiving party video and audio decoders.

Example audio delay: 0.5 sec.

Example video delay: 1.0 sec.

7.2.4 Inter-Media Skew (sec)

Definition: Level of synchronization between the video and the speech component in a VT call.

Example: 0.500 sec 90% of the time.

Note: The system design permits separate IP packet flows for speech and video. In general, it is not required for the network to coordinate

quality of service control across multiple IP packet flows. Since buffering can be used at the terminal to control the inter-media skew, the inter-media skew requirement is a terminal requirement rather than a network requirement.

7.2.5 Frame Error Rate

Definition: The rate of lost and corrupted data units (e.g., octets, RLP frames, video packets) in the transmitted data:

Note: The use of Frame Error Rate is expected to meet the following guidelines:

1. The frame error rate parameter(s) specifies the maximum acceptable lossiness in the radio network for the VT application.
2. The frame error rate parameter(s) can be correlated to the video quality perceived by the user (e.g. mean opinion score) for a given video codec.
3. The frame error rate parameter(s) is translatable to radio link parameters that the radio network can use to control the radio link lossiness.

7.2.6 Jitter (sec)

Definition: Fluctuation of arrival times of frames (video or speech). It is expressed by standard deviation of arrival times of frames relative to the expected (clocked) times.

7.3 Control of VT Parameters

VT-35. **Target Data Rate.** Under normal traffic/channel conditions, the video encoder should maintain a target peak and average data rate based on the default or originally negotiated service parameters (e.g. 56 kbps including RTP overhead).

VT-36. **Frame Rate.** Video encoder frame rate shall be configurable. The encoder shall adjust the frame rate and frame quality so as to not exceed the target bit rate. The encoder shall support the default frame rate.

VT-37. **Dynamic Setting of Parameters.** The service should be able to dynamically negotiate VT service parameters (e.g., bit rates, picture size, multimedia components, etc.) during the session, to adapt to the users' needs and/or channel conditions.

VT-38. **QoS Control.** User, User Agent, and/or network-resident proxy shall be able to specify QoS parameters, based on interaction between these VT service entities and the appropriate QoS control network entity. This is to enable the selection of the VT service quality that best fits user needs, and to permit the user to control the service, within bounds imposed by the network and/or subscription.

VT-39. **Service Parameter Control.** The user should be able to control certain service parameters any time during a VT call, subject to constraints of the link and network load conditions. These controls should include:

a. Negotiate to drop or add a multimedia component;

b. Negotiate to increase or decrease video frame rate;

c. Negotiate to increase or decrease picture resolution;

VT-40. Application Control of QoS: VT should make best effort to maintain an application level QoS.

Note: If the network is unable to deliver the requested level of QoS, the VT application in the terminal should adjust the application QoS to the network/link condition by means of one of the following:

a. Obtain RAN condition by referencing the appropriate network control entity;

b. Obtain end-to-end condition visible to the application, if the RAN information is not available.

When unable to maintain the requested service parameters, the VT application in the terminal should do one of the following:

c. Drop video/speech frames;

d. Instruct the local VT terminal to decrease the outgoing frame rate, lower per-frame quality, discontinue transmitting a multimedia component, or any combination thereof;

e. Request the remote VT terminal to decrease the outgoing frame rate, lower per-frame quality, discontinue transmitting a multimedia component, or any combination thereof;

f. Discontinue transmitting an incoming multimedia component to the local VT;

Both calling and called parties should be notified if transmission of a media component had been discontinued. Note that this notification may be implicit.

Steps to downgrade VT service QoS should occur only after services that do not support QoS (implicitly require only best-effort QoS) have been downgraded first.

Upon restoration of channel/traffic conditions, the VT application in the terminal should restore the originally negotiated service parameters by reversing the above mentioned controls.

8 Security

VT-41. Service Authentication. In addition to Access Authentication, the network should be able to perform VT service authentication at an appropriate time (e.g., VT origination, VT termination).

VT-42. Encryption. VT should support end-to-end encryption providing confidence for business customers to conduct sensitive communications using VT.

9 Accounting

VT-43. Accounting Records. The following accounting records should be provided:

a. Negotiated QoS parameters or a relevant set thereof (some may be

default, others are subject to end-to-end QoS and may not be used in the accounting record analysis)

b. VT call duration

c. Time of day

d. Volume of data transmitted/received

e. IP packet count

VT-44. **Advice of Charge.** User should be advised of a possible change of tariff or tariff differential due to roaming status or time of day of the call. This notification may be implicit.

10 Use Cases and Call Scenarios (Informative)

Use cases describe PSVT situations in the context of a system that supports both the circuit-switched (CS) cdma2000 1x network, and the packet-switched (PS) HRPD network. Some call scenarios are described with the purpose of illustrating video telephony use cases and interactions with voice telephony services.

a. VT Call Origination on HRPD Network

Use Case 1: The VT subscriber camping on the HRPD network, which may or may not have previously established a packet data session, originates a VT call.

Use Case 2: The VT subscriber already engaged in a VT call places another VT call thus establishing a 3-way video conference.

Note: This use case is subject to Phase 2 of PSVT specifications.

b. VT Call Termination on HRPD Network

Use Case 3: The VT subscriber camping on the HRPD network, having previously established a packet data session and being registered on IMS, receives an incoming VT call.

Use Case 4: The VT subscriber is already engaged in a VT call. When an additional incoming VT call occurs, VT call waiting indication is presented to the subject subscriber.

c. VT Call Hand-down to Voice Call

Use Case 5: The VT subscriber is on a VT call on the HRPD network. When the HRPD network becomes unavailable to the VT subscriber, the video component of the call is removed, and the call is handed off to the cdma2000 1x network as a voice call.

d. VT Call Exceptions

Use Case 6: The calling party attempts a VT call to a user who is not an IMS subscriber, or does not have VT subscription. The call is established as a CS

call on cdma2000 1x.

Use Case 7: The calling party attempts a VT call to a VT subscriber who is out of the HRPD coverage. The call is established as the CS call on cdma2000 1x.

Use Case 8: A VT subscriber is active in a cdma2000 1x voice call. The incoming VT call is delivered as cdma2000 1x CS voice call. CS call waiting indication is presented to the subject subscriber.

Use Case 9: A VT subscriber is active in a cdma2000 1x packet data call. The network and the terminal device have VPOP capability. The incoming VT call is delivered as a cdma2000 1x CS voice call.

Use Case 10: A VT subscriber is active in a VT call. The incoming voice call is delivered as an HRPD VoIP call.

e. Normal VT Call Flow

The calling party originates a VT call. The called party is alerted of the incoming VT call. The calling party ID is displayed. The called party accepts the VT call, speech and video media start flowing, and the call is established successfully.

f. Unsuccessful VT Call

VT Call Rejection: The calling party originates a VT call. The called party is alerted of the incoming VT call. The calling party ID is displayed. The called party rejects the VT call. The calling party is notified that the VT call has failed due to the rejection by the called party, and the calling party is presented with the option to leave a video mail if supported.

Failure Due to Resource Limitation: The calling party originates a VT call. The VT call fails because the called party terminal does not support VT, or there are inadequate network resources to complete the call. The calling party is notified that the VT call has failed and the failure cause. The calling party is presented with an option to proceed by making a voice call.

g. Fallback of VT Call

If the network resources are inadequate to sustain a VT call in progress, VT terminal gracefully downgrades the video quality while maintaining speech flow until the video flow is discontinued.

When adequate network resource becomes again available, the video flow may resume.

A party engaged in a VT call can mute/fade outgoing speech/video during VT call.

h. Incoming Voice Call during VT Call

An incoming voice call does not automatically terminate an ongoing VT call. The system indicates to the VT user that a voice call is waiting. The VT user chooses whether to place the VT call on hold and accept the voice call.

i. Incoming VT Call during VT Call

If an incoming VT call is placed to a VT user during an ongoing VT call, the system indicates to the VT user that another VT call is waiting. The VT user chooses whether to place the ongoing VT call on hold and accept the VT call.

j. Teardown of VT Call

The calling party or the called party can tear down an ongoing VT call at any time.