TSGS#3(99)202

3G TS 26.110 V1.0.0 (1999-04)

Technical Specification

3rd Generation Partnership Project (3GPP);
TSG-SA Codec Working Group;
Codec for Circuit Switched Multimedia Telephony Service;
General Description

3GPP

Reference
TSG-SA4-W2 (<Shortfilename>.PDF)

Keywords
Multimedia Codec, Circuit Switched, Description

3GPP

Postal address

Office address

Internet
secretariat@3gpp.org
Individual copies of this deliverable
can be downloaded from
http://www.3gpp.org

3GPP

Contents

3GPP

Intellectual Property Rights

Foreword

This Technical Specification has been produced by the 3rd Generation Partnership Project, Technical Specification Group Services and System Aspects, CODEC Working Group (TSG-S4).

The contents of this TS may be subject to continuing work within the 3GPP and may change following formal TSG approval. Should the TSG modify the contents of this TS, it will be re-released with an identifying change of release date and an increase in version number as follows:

Version m.t.e

where:

m   indicates [major version number]

x   the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

y   the third digit is incremented when editorial only changes have been incorporated into the specification.

Introduction

This document contains a specification for H.324 based multimedia codecs for circuit switched 3GPP networks. The term codec is usually associated with a single media type. However, many multimedia services require a close integration of disparate media types. In this sense, the representations of these media types (in the form of media streams) are at least logically bound into a single multimedia stream. As such, a H.324 based multimedia codec must handle multiplexing/de-multiplexing and skew. It will also have to provide codecs for each of the derived media streams. End-to-end, in-band control is also required for the purposes of configuration and establishing individual media streams. Finally, since 3GPP networks are inherently error prone, error detection and/or correction must also be provided by the multimedia codec since it has a comprehensive view of the bit stream it produces and therefore can apply the most efficient form of error detection and/or correction.

Scope

This specification introduces the set of specifications which apply to 3G H.324 multimedia terminals.

References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.

- For a specific reference, subsequent revisions do not apply.

- For a non-specific reference, the latest version applies.

- A non-specific reference to an ETS shall also be taken to refer to later versions published as an EN with the same number.

[1]   ITU-T Recommendation H.223: "Multiplexing protocol for low bitrate multimedia communication"

[2]   ITU-T Recommendation H.223 — Annex A: "Multiplexing protocol for low bitrate multimedia communication over low error-prone channels"

[3]   ITU-T Recommendation H.223 — Annex B: "Multiplexing protocol for low bitrate multimedia communication over moderate error-prone channels"

[4]   ITU-T Recommendation H.223 — Annex C: "Multiplexing protocol for low bitrate multimedia communication over highly error-prone channels"

[5]   ITU-T Recommendation H.223 — Annex D: "Optional multiplexing protocol for low bitrate multimedia communication over highly error-prone channels"

[6]   ITU-T Recommendation H.245: "Control protocol for multimedia communication"

[7]   ITU-T Recommendation G.723.1: "Dual rate speech coder for multimedia communication transmitting at 5.3 & 6.3 kbit/s"

[8]   ITU-T Recommendation H.263: "Video coding for low bitrate communication"

[9]   ITU-T Recommendation H.261: "Video CODEC for audiovisual services at p X 64 kbit/s"

[10]   ITU-T Recommendation H.324: "Terminal for low bitrate multimedia communication"

[11]   3GPP Technical Specification 3G TS 26.111: "Modifications to H.324"

[12]   3GPP Technical Specification 3G TS 26.112: "Call Set Up Requirements"

[13]   3GPP Technical Reference 3G TR 26.911: "Terminal Implementor's Guide"

[14]   ITU-T Recommendation X.691: "Information Technology - ASN.1 Encoding Rules - Specification of Packed Encoding Rules (PER)"

[15]   International Standard ISO/IEC 14494-2: "Information technology — Generic coding of audio-visual object — Part 2: Visual, 1999"

[16]   3GPP Technical Specification 3G TS 26.071: "Mandatory Speech Codec; General Description"

[17]   3GPP Technical Specification 3G TS 26.090: "Mandatory Speech Codec; Speech Transcoding Functions"

3GPP

[18]   3GPP Technical Specification 3G TS 26.073: "Mandatory Speech Codec; ANSI C-Code"

Definitions and abbreviations

Definitions

For the purposes of the present document, the following terms and definitions apply.

H.324 terminal: ITU-T H.324 recommendation, including Annex C

3G H.324 terminal: Based on ITU-T H.324 recommendation modified by 3GPP for purposes of 3GPP circuit switched network based video telephony

Abbreviations

For the purposes of the present document, the following abbreviations apply:

ACELP	Algebraic-Code-Excited Linear-Prediction
AEC	Acoustic Echo Cancellation
ADC	Analogue Digital Converter
CT	Correlation Threshold
DAC	Digital Analogue Converter
DCT	Discrete Cosine Transformation
EI	Error Indication
EOB	End Of Block
FEC	Forward Error Correction
GOB	Group Of Blocks
GQUANT	Group Quantizer information
HDLC	High-Level Data Link Control
HEC	Header Error Control
LC	Logical Channel
MC	Multiplex Code
MP-MLQ	Multipulse Maximum Likelihood Quantization
MPL	Multiplex Payload Length
PC	Personal Computer
MCU	Multipoint Conference Unit
MUX	H.223 Multiplex layer
PDU	Protocol Data Unit
SN	Sequence Number
VLC	Variable Length Code

General

3G H.324 terminals provide real-time video, audio, or data, in any combination, including none, over 3GPP circuit-switched, radio networks. They are based on ITU-T H.324 with Annex C. Communication may be either 1-way or 2-way. Such terminals may be part of a portable device or integrated into an automobile or other non fixed location device. They may also be fixed, stand-alone devices; for example, a video telephone or kiosk. 3G H.324 terminals may also be integrated into PCs and workstations.

In addition to 3G H.324 to 3G H.324 communication, interoperation with other types of multimedia telephone terminals is possible, however a gateway may be required.

Multipoint communication between more than two 3G H.324 terminals is possible using an Mutipoint Communication Unit (MCU). MCU functionality is for further study.

3G H.324 terminals are based on ITU-T H.324 with Annex C. For performance reasons and to define the call set-up procedures, some modifications to H.324 were made. These are described in 3G TS 26.111, except call set-up procedures are described in 3G TS 26.112. 3G H.324 terminals shall conform to these two specifications. Because of the many options in H.324, an implementor's guide, 3G TR 26.911, provides preferred options for 3G H.324 implementations.

Figure 1 below shows the functional components of a generic 3GPP multimedia terminal. The video, speech and data are optional. If a media type is supported, the standards indicated are mandatory except those enclosed in square brackets are optional.
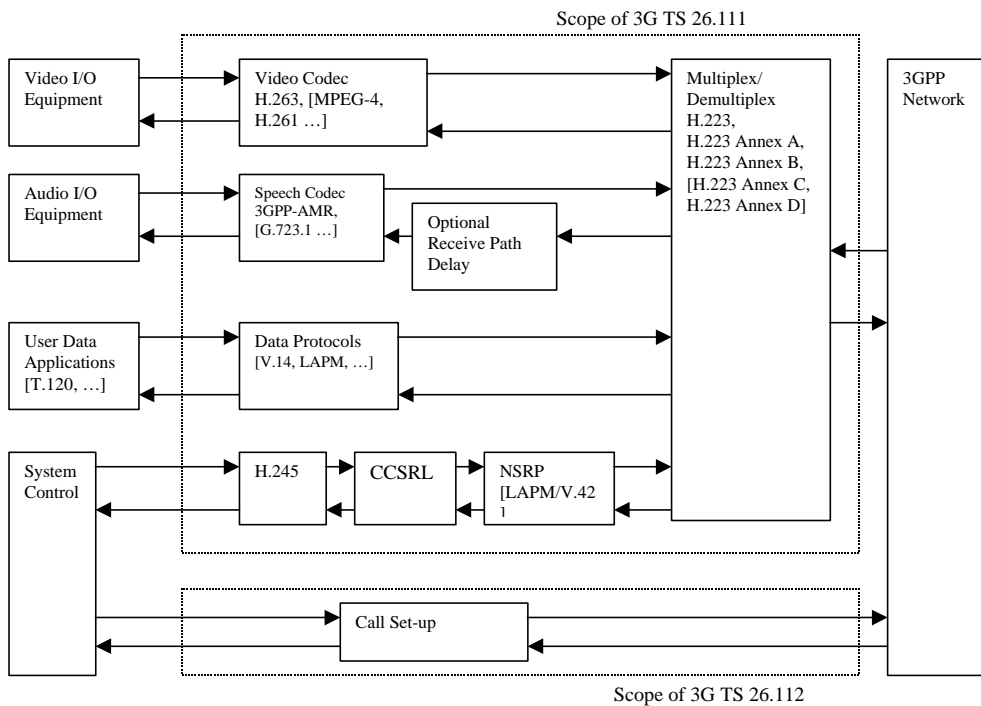

Figure 1   Scope of circuit switched multimedia 3GPP specification.

Short descriptions of ITU-T H.324, 3G TS 26.111, 3G TS 26.112, and 3G TR 26.911 are given below.

3GPP

ITU-T H.324

ITU-T H.324 describes terminals for low bitrate multimedia communication. That ITU-T recommendation contains "ANNEX C, Multimedia Telephone Terminals Over Error Prone Channels" (sometimes referred to as H.324/M). This annex is considered an integral part of the recommendation. Therefore, herewith H.324 shall mean ITU-T H.324 with Annex C.

Originally designed for V.34 modems, H.324 now supports ISDN and wireless networks. Therefore, it is well suited as a basis for 3GPP multimedia codecs. Relevant to wireless networks, H.324 describes the overall system architecture and introduces control (H.245), mux (H.223), video (H.261 and H.263), and audio (G.723.1).

Annex A provides a short overview of H.324 and multimedia codecs.

Modifications to H.324 (3G TS 26.111)

To enable cost-effective, high-quality H.324 terminals for 3GPP networks, some modifications were made to H.324. These modifications are described in 3G TS 26.111. Terminals adhering to this specification and 3G TS 26.112 (see below) are herewith known as 3G H.324 terminals. 3G H.324 terminals shall conform to 3G TS 26.111 and 3G TS 26.112.

Call set-up requirements (3G TS 26.112)

H.324 does not describe call set-up procedures for 3GPP networks. These are described in 3G TS 26.112 and shall be used for 3G H.324 terminals.

Terminal implementor's guide (3G TR 26.911)

A successful 3G H.324 terminal will have to function well at bandwidths as low as 32 KBPS and in potentially high error rate environments. 3G H.324 contains many options that may be employed by an implementor. To help choose which options and combinations of options are useful, an implementor's guide is provided in 3G TR 26.911.

Annex A — Background Information

The section is intended for informational purposes only. This is not an integral part of this specification. Each section below relates to the functional components in figure 1.

1.1 Video I/O Equipment

For a video telephone this would most likely consist of a video camera and display monitor. Other possible input sources could be a VCR or disk drive. While most applicable I/O equipment relies on a standard format for the video signal or bit stream, this format is likely to differ from that mandated by the video codec. In such cases, circuitry or software is used to transcode between the two formats.

Video Codec

ITU-R 601 (NTSC or PAL) is a typical video input signal and represents a bit stream of 20.7 Mbyte/s for the actual image (excluding blanking intervals). The first order of compression occurs by reducing the resolution of the input signal.[1] For example, CIF resolution at 30 fps produces a bit stream 4.6 Mbyte/s. Additional savings occur by dropping frames. In a videoconference, where motion is relatively slow, 10 fps is considered adequate. Thus, the original signal of 20.7 Mbytes/s could be reduced to 1.5 Mbyte/s with just these techniques. However, this is still 188 times greater than can be transmitted on, for example, a 64 KBPS channel. Substantial compression is still required, especially considering that framing, control, and audio would as well require a portion of the available bandwidth.

To achieve the degree of compression required for video telephony, all of the video codecs that can currently be employed in a 3GPP multimedia codec use a combination of spatial and temporal redundancy reduction to reduce the bandwidth required by the video media stream. Spatial redundancy can be reduced by converting the input signal from the time domain to the frequency domain using a DCT. This produces a DC value and other coefficients, where most of the scene energy is concentrated in the coefficients corresponding to the lower frequencies. Next, a coarse quantizer is applied (which, in this domain, has little effect on image quality). This results in many of the coefficients being encoded to 0. The significant coefficients are encoded to a much smaller range of values. The coefficients are then reordered so that, typically, the larger magnitude values will occur first followed by 0 value coefficients. Finally, the coefficients are replaced with a count of the number of zero value coefficients followed by the value of a nonzero coefficient. This combination is translated into a VLC. Applying this type of compression to the entire video frame produces an intra frame.

Despite the efficiency of intra coding, significantly more compression is required. In addition to removing spatial redundancy, all video codecs apply temporal reduction as well. This is achieved by comparing the current frame to the previous and estimating the set of vectors which when applied to their respective areas of the scene would create the new, current frame based on the old, previous frame. The match is usually not perfect, so an error component is transmitted as well. The error component is also transformed to the frequency domain, so the same compression efficiency achieved in the intra frame is achieved here as well — enhanced by the fact the range of error coefficients is less than intra coefficients. Since generally only a few areas of a scene change from frame to frame, high compression can be achieved by sending a series of inter frames. If the error component for a particular block is too large, it can be encoded as an intra block.

Since, by their nature, VLCs are not fixed length, a single bit error can make it impossible to decode an entire frame. Unfortunately, each inter coded frame relies on its previous frame to be decoded. Thus, a single bit error can destroy the entire remaining bit stream. Video codecs have various ways of handling errors. The simplest is to use error detection to determine if a frame contains an error. The transmitter is then signalled that an error occurred. It then sends an intra coded frame, which does not depend on any previous frames. This approach consumes considerable bandwidth and is only practical for very low error networks. Other, more sophisticated schemes are available using the video codecs available to 3G H.324 terminals.

---

[1] Note that ITU-R 601 represents 16 bit precision colour, whereas true colour is usually considered to require 24 bit precision. Also, the spatial resolution of ITU-R 601 is substantially less than can be achieved with normal human vision.

H.261

H.261 supports CIF and QCIF images as input. It provides good video quality at 64 kbit/s or higher. It uses BCH codes for Forward Error Correction (FEC). However, this is not recommended for H.324.

H.263

H.263 is an extension of H.261. It allows sub-QCIF, 4CIF and 16CIF as additional input formats. H.263, in its original version, provides four annexes that describe optional modes for enhanced coding.

Advanced prediction mode (Annex F) provides half-pel motion estimation, median-based motion vector prediction, 4 motion vectors per macroblock (one per block), and overlapped block motion compensation

Unrestricted motion vectors (Annex D) work in conjunction with advanced prediction mode and allow motion vectors to point outside the picture area

Arithmetic coding (Annex E) can be used instead of variable length coding

PB-frames (Annex G) allow bi-directional prediction similar to MPEG

Other significant differences exist, but require a level of detail to explain that renders them outside the scope of this document.

A second version of H.263 (known as H.263+) adds annexes I through T, some of which address error prone environments and are therefore of special interest to 3GPP multimedia codecs.

MPEG-4

MPEG-4 Visual (ISO/IEC 14496-2) is a generic video codec. One of its target areas is mobile communications. Error resiliency and high efficiency make this codec particularly well suited for 3G H.324.

MPEG-4 Visual is organised into Profiles. Within a Profile, various Levels are defined. Profiles define subsets of tool sets. Levels are related to computational complexity. Among these Profiles, Simple Visual Profile provides error resilience (through data partitioning, RVLC, resynchronisation marker and header extension code) and low complexity.

MPEG-4 allows various input formats, including general formats such as QCIF and CIF. It is also baseline compatible with H.263.

Audio I/O Codec

Generally, a video telephone would require a handset, headset, or microphone and speaker. Often, integrated circuits are employed that convert the typically analogue input signal to a PCM format bit stream (ADC) and convert PCM to an analogue signal for acoustic output (DAC). This is helpful since many speech codecs use PCM for input and output. Video telephones often use a separate microphone and speaker. This allows the user to be seen without a handset or headset. However, if this is so, AEC will be required.

Speech Codec

3GPP AMR

The AMR codec uses eight source codecs with bit-rates of 12.2, 10.2, 7.95, 7.40, 6.70, 5.90, 5.15 and 4.75 kbit/s. The coder operates on speech frames of 20 ms corresponding to 160 samples at the sampling frequency of 8000 sample/s. It performs the mapping from input blocks of 160 speech samples in 13-bit uniform PCM format to encoded blocks of 95, 103, 118, 134, 148, 159, 204, and 244 bits and from encoded blocks of 95, 103, 118, 134, 148, 159, 204, and 244 bits to output blocks of 160 reconstructed speech samples. The coding scheme for the multi-rate coding modes is the so-called Algebraic Code Excited Linear Prediction Coder (ACELP). The multi-rate ACELP coder is referred to as MR-ACELP. At each 160 speech samples, the speech signal is analysed to extract the parameters of the CELP model (LP filter coefficients, adaptive and fixed codebooks' indices and gains). These parameters are encoded and

3GPP

transmitted. At the decoder, these parameters are decoded and speech is synthesised by filtering the reconstructed excitation signal through the LP synthesis filter.

The adaptive multi-rate speech codec is described in a bit-exact arithmetic in form of a fixed-point ANSI-C code to allow for easy type approval as well as general testing purposes of the adaptive multi-rate speech codec.

The DTX mechanism includes a Voice Activity Detector (VAD) on the TX side; evaluation of the background acoustic noise on the TX side, in order to transmit characteristic parameters to the RX side; and generation of comfort noise on the RX side during periods where the radio transmission is turned off.

The AMR specification contains error concealment. The purpose of frame substitution is to conceal the effect of lost AMR speech frames. The purpose of muting the output in the case of several lost frames is to indicate the breakdown of the channel to the user and to avoid generating possible annoying sounds as a result from the frame substitution procedure.

G.723.1

G.723.1 can be used for compressing the speech or other audio signal component of multimedia services at a very low bitrate as part of H.324. This coder has two bit-rates associated with it, 5.3 and 6.3 kbit/s. The higher bitrate has greater quality. The lower bit-rate gives good quality and provides system designers with additional flexibility. Both rates are a mandatory part of the encoder and decoder. It is possible to switch between the two rates at any frame boundary. An option for variable rate operation using discontinuous transmission and noise fill during non-speech intervals is also possible using a series of silence frames or a single silence frame followed by no frames until speech is detected.

G.723.1 encodes speech or other audio signals in frames using linear predictive analysis-by-synthesis coding. The excitation signal for the high rate coder is Multipulse Maximum Likelihood Quantization (MP-MLQ) and for the low rate coder is Algebraic-Code-Excited Linear-Prediction (ACELP). The frame size is 30 ms and there is an additional look ahead of 7.5 msec,. This coder is designed to operate with a digital signal obtained by first performing telephone bandwidth filtering (ITU-T Recommendation G.712) of the analogue input, then sampling at 8000 Hz and then converting to 16-bit linear PCM for the input to the encoder. The output of the decoder is converted back to analogue by similar means.

G.723.1 has been designed to be robust for indicated frame erasures. An error concealment strategy for frame erasures has been included in the decoder. However, this strategy must be triggered by an external indication that the bit stream for the current frame has been erased. This can be achieved in H.324 using the AL2 Error Indication (EI) flag and the optional AL2 Sequence Number (SN). Because the coder was designed for burst errors, there is no error correction mechanism provided for random bit errors. If a frame erasure has occurred, the decoder switches from regular decoding to frame erasure concealment mode.

G.723.1 contains three annexes. Annex A describes the silence compression system designed for the G.723.1 speech coder (mentioned above). Annex B describes an alternative implementation of G.723.1 contained in floating point C source code. Annex C specifies a channel coding scheme which can be used with the triple rate speech codec G.723.1. The channel codec is scalable in bit-rate and is designed for mobile multimedia applications as a part of the overall H.324 family of standards.

User Data Applications

An example of a User Data Application is T.120. This protocol allows multipoint data conferencing that includes data and image transferral. Other functions, such as shared whiteboards and applications, are possible.

Data Protocols

Various data protocols can be supported. These always support data applications (see 0, User Data Applications). A specific protocol or set of protocols is often stipulated by the data application. Each protocol provides varying degrees of error detection and/or correction.

System Control

In general, system control constitutes the overall state machine for the terminal. It usually has to be aware of when a connection has been established. At that point it can begin H.245 procedures such as master/slave determination, capabilities exchange, and opening logical channels. Upon call termination, either initiated at the near or far ends, system control generally initiates H.245 end session procedures.

Call Set-up

All out-of-band network signalling for the purpose of call control is handled by call set-up, which is usually implemented as a state machine. This includes initiating, answering, and tearing down calls.

H.245

H.245 specifies the syntax and semantics for in-band, terminal-to-terminal control messages and the procedures for their use. Most importantly, H.245 is used for master/slave determination, capabilities exchange, H.223 mux table transmission, and opening and closing logical channels. There is also a large array of general control and indication messages. H.245 addresses a wide range of terminals and applications. Therefore, only a subset of the messages listed in H.245 pertain to 3G H.324 terminals. Messages fall into one of four categories: Request (requires a Response), Response (in response to a Request), Command (requires an action), and Indication (informative only).

H.245 messages are carried on a single logical channel within the H.223 mux. This channel is labelled LC 0 and is considered to be open upon establishing digital communications end-to-end and survives until digital communication is terminated. Due to the characteristics of the H.223 mux, bandwidth for H.245 messages is allocated on an as-needed basis. Since most H.245 traffic occurs and the beginning and end of the session, this conserves much needed bandwidth for video and audio. Error control is not provided for within H.245 and is specified elsewhere.

H.223

H.223 describes the multiplexing protocol used between H.324 terminals. It is packet oriented and each packet can contain a subset of a maximum of 65536 LCs. Each LC represents a single media, information, or control channel. The H.223 protocol is split into two layers, the lowest being the Multiplex Layer.

The Multiplex Layer exchanges data with the end terminal via MUX-PDUs. Multiplex table entries, of which there are 16 (and can be changed during a session), describe which octets from within the PDU are allocated to which logical channels. The multiplex table entry employed for a particular PDU is indicated by the 4 bit MC field in the MUX-PDU header. MUX-PDUs contain an integer number of octets. Errors within the MUX-PDU header are controlled using the HEC field in the MUX-PDU header. H.324 terminals utilising the V.34 transmission protocol frame MUX-PDUs with HDLC. Bit stuffing is used for data transparency in this case.

Above the Multiplex Layer is the Adaptation Layer, of which there are three different types.

1) AL1 is designed primarily for control information and data protocols. It can be either framed or unframed and does not provide any error control.

2) AL2 is designed primarily for the transfer of digital audio. AL2 PDUs contain 1 octet for an 8-bit CRC and an optional octet for a sequence number.

3) AL3 is designed primarily for the transfer of digital video. AL3 PDUs contain 2 octets for a 16-bit CRC. There is also optionally 1 or 2 octets for control. AL3 also allows limited retransmission.

For purposes of video telecommunications over wireless networks, four annexes to H.223 have been created. These create four levels of error detection and error correction.

Level 0

Level 0 applies to H.223 as described above.

Level 1

Level 1, described in Annex A, replaces HDLC framing with 1 or 2 16 bit flags. Unlike HDLC, Level 1 does not guarantee data transparency. However, if the MUX-PDU header is constructed in such a way as to make emulating the Level 1 framing flags impossible, data transparency can be achieved by correctly decoding the MUX-PDU. Should there be and error in the MUX-PDU header, resynchronisation techniques will have to be applied.

Level 2

Level 2, described in Annex B, uses the same framing as Level 1, but utilises a 3 octet header. This header starts with a 4 bit MC, which is the same as in Level 0. This is followed by an 8-bit MPL-field, with a range of values 0 – 254. Lastly, a 12 bit extended Golay code is used for parity bits. The PM in Level 2 is signalled through the polarity of the MUX-PDU flag. If the output of the correlator is greater than or equal to CT, the PM is 0. If it is less than or equal to - CT, the PM equals 1. After the parity bits, there can be an optional MUX-PDU header for the previous (corrupted) MUX-PDU. This 1 octet field uses the format described in Level 0. Level 2 also offers enhanced packet resynchronisation.

Level 3

Level 3, described in Annexes C and D, provides error correction capabilities at the mux level.

Bibliography

The following material, though not specifically referenced in the body of the present document (or not publicly available), gives supporting information.

<Publication>: "<Title>".

History

Document history		
0.1.0	April 20, 1999	Initial draft
0.2.0	April 21, 1999	Revised per TSG SA S4#4 discussion
1.0.0	April 22, 1999	Version presented for information to TSG-SA#3