

Network Working Group  
Request for Comments: 5194  
Category: Informational

A. van Wijk, Ed.  
G. Gybels, Ed.  
June 2008

Framework for Real-Time Text over IP Using  
the Session Initiation Protocol (SIP)

Status of This Memo

This memo provides information for the Internet community. It does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

Abstract

This document lists the essential requirements for real-time Text-over-IP (ToIP) and defines a framework for implementation of all required functions based on the Session Initiation Protocol (SIP) and the Real-Time Transport Protocol (RTP). This includes interworking between Text-over-IP and existing text telephony on the Public Switched Telephone Network (PSTN) and other networks.

## Table of Contents

1. Introduction .....	3
2. Scope .....	4
3. Terminology .....	4
4. Definitions .....	4
5. Requirements .....	6
5.1. General Requirements for ToIP .....	6
5.2. Detailed Requirements for ToIP .....	8
5.2.1. Session Setup and Control Requirements .....	9
5.2.2. Transport Requirements .....	10
5.2.3. Transcoding Service Requirements .....	10
5.2.4. Presentation and User Control Requirements .....	11
5.2.5. Interworking Requirements .....	13
5.2.5.1. PSTN Interworking Requirements .....	13
5.2.5.2. Cellular Interworking Requirements .....	14
5.2.5.3. Instant Messaging Interworking Requirements .....	14
6. Implementation Framework .....	15
6.1. General Implementation Framework .....	15
6.2. Detailed Implementation Framework .....	15
6.2.1. Session Control and Setup .....	15
6.2.1.1. Pre-Session Setup .....	15
6.2.1.2. Session Negotiations .....	16
6.2.2. Transport .....	17
6.2.3. Transcoding Services .....	18
6.2.4. Presentation and User Control Functions .....	18
6.2.4.1. Progress and Status Information .....	18
6.2.4.2. Alerting .....	18
6.2.4.3. Text Presentation .....	19
6.2.4.4. File Storage .....	19
6.2.5. Interworking Functions .....	19
6.2.5.1. PSTN Interworking .....	20
6.2.5.2. Mobile Interworking .....	22
6.2.5.2.1. Cellular "No-gain" .....	22
6.2.5.2.2. Cellular Text Telephone Modem (CTM) .....	22
6.2.5.2.3. Cellular "Baudot mode" .....	22
6.2.5.2.4. Mobile Data Channel Mode .....	23
6.2.5.2.5. Mobile ToIP .....	23
6.2.5.3. Instant Messaging Interworking .....	23
6.2.5.4. Multi-Functional Combination Gateways .....	24
6.2.5.5. Character Set Transcoding .....	25
7. Further Recommendations for Implementers and Service Providers .....	25
7.1. Access to Emergency Services .....	25
7.2. Home Gateways or Analog Terminal Adapters .....	25
7.3. User Mobility .....	26

7.4. Firewalls and NATs .....	26
7.5. Quality of Service .....	26
8. Security Considerations .....	26
9. Contributors .....	27
10. References .....	27
10.1. Normative References .....	27
10.2. Informative References .....	29

## 1. Introduction

For many years, real-time text has been in use as a medium for conversational, interactive dialogue between users in a similar way to how voice telephony is used. Such interactive text is different from messaging and semi-interactive solutions like Instant Messaging in that it offers an equivalent conversational experience to users who cannot, or do not wish to, use voice. It therefore meets a different set of requirements from other text-based solutions already available on IP networks.

Traditionally, deaf, hard-of-hearing, and speech-impaired people are amongst the most prolific users of real-time, conversational, text but, because of its interactivity, it is becoming popular amongst mainstream users as well. Real-time text conversation can be combined with other conversational media like video or voice.

This document describes how existing IETF protocols can be used to implement a Text-over-IP solution (ToIP). Therefore, this document describes how to use a set of existing components and protocols and provides the requirements and rules for that resulting structure, which is why it is called a "framework", fitting commonly accepted dictionary definitions of that term.

This ToIP framework is specifically designed to be compatible with Voice-over-IP (VoIP), Video-over-IP, and Multimedia-over-IP (MoIP) environments. This ToIP framework also builds upon, and is compatible with, the high-level user requirements of deaf, hard-of-hearing and speech-impaired users as described in RFC3351 [22]. It also meets real-time text requirements of mainstream users.

ToIP also offers an IP equivalent of analog text telephony services as used by deaf, hard-of-hearing, speech-impaired, and mainstream users.

The Session Initiation Protocol (SIP) [2] is the protocol of choice for control of Multimedia communications and Voice-over-IP (VoIP) in particular. It offers all the necessary control and signalling required for the ToIP framework.

The Real-Time Transport Protocol (RTP) [3] is the protocol of choice for real-time data transmission, and its use for real-time text payloads is described in RFC 4103 [4].

This document defines a framework for ToIP to be used either by itself or as part of integrated, multi-media services, including Total Conversation [5].

## 2. Scope

This document defines a framework for the implementation of real-time ToIP, either stand-alone or as a part of multimedia services, including Total Conversation [5]. It provides the:

- a. requirements for real-time text;
- b. requirements for ToIP interworking;
- c. description of ToIP implementation using SIP and RTP;
- d. description of ToIP interworking with other text services.

## 3. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [6] and indicate requirement levels for compliant implementations.

## 4. Definitions

**Audio bridging:** a function of an audio media bridge server, gateway, or relay service that sends to each destination the combination of audio from all participants in a conference, excluding the participant(s) at that destination. At the RTP level, this is an instance of the mixer function as defined in RFC 3550 [3].

**Cellular:** a telecommunication network that has wireless access and can support voice and data services over very large geographical areas. Also called Mobile.

**Full duplex:** media is sent independently in both directions.

**Half duplex:** media can only be sent in one direction at a time, or if an attempt to send information in both directions is made, errors may be introduced into the presented media.

Interactive text: another term for real-time text, as defined below.

Real-time text: a term for real-time transmission of text in a character-by-character fashion for use in conversational services, often as a text equivalent to voice-based conversational services. Conversational text is defined in the ITU-T Framework for multimedia services, Recommendation F.700 [21].

Text gateway: a function that transcodes between different forms of text transport methods, e.g., between ToIP in IP networks and Baudot or ITU-T V.21 text telephony in the PSTN.

Textphone: also "text telephone". A terminal device that allows end-to-end real-time text communication using analog transmission. A variety of PSTN textphone protocols exists world-wide. A textphone can often be combined with a voice telephone, or include voice communication functions for simultaneous or alternating use of text and voice in a call.

Text bridging: a function of the text media bridge server, gateway (including transcoding gateways), or relay service analogous to that of audio bridging as defined above, except that text is the medium of conversation.

Text relay service: a third-party or intermediary that enables communications between deaf, hard-of-hearing, and speech-impaired people and voice telephone users by translating between voice and real-time text in a call.

Text telephony: analog textphone service.

Total Conversation: a multimedia service offering real-time conversation in video, real-time text and voice according to interoperable standards. All media streams flow in real time. (See ITU-T F.703, "Multimedia conversational services" [5].)

Transcoding service: a service provided by a third-party User Agent that transcodes one stream into another. Transcoding can be done by human operators, in an automated manner, or by a combination of both methods. Within this document, the term particularly applies to conversion between different types of media. A text relay service is an example of a transcoding service that converts between real-time text and audio.

TTY: originally, an abbreviation for "teletype". Often used in North America as an alternative designation for a text telephone or textphone. Also called TDD, Telecommunication Device for the Deaf.

Video relay service: a service that enables communications between deaf and hard-of-hearing people and hearing persons with voice telephones by translating between sign language and spoken language in a call.

Acronyms:

2G	Second generation cellular (mobile)
2.5G	Enhanced second generation cellular (mobile)
3G	Third generation cellular (mobile)
ATA	Analog Telephone Adaptor
CDMA	Code Division Multiple Access
CLI	Calling Line Identification
CTM	Cellular Text Telephone Modem
ENUM	E.164 number storage in DNS (see RFC3761)
GSM	Global System for Mobile Communications
ISDN	Integrated Services Digital Network
ITU-T	International Telecommunications Union-Telecommunications Standardisation Sector
NAT	Network Address Translation
PSTN	Public Switched Telephone Network
RTP	Real-Time Transport Protocol
SDP	Session Description Protocol
SIP	Session Initiation Protocol
SRTP	Secure Real Time Transport Protocol
TDD	Telecommunication Device for the Deaf
TDMA	Time Division Multiple Access
TTY	Analog textphone (Teletypewriter)
ToIP	Real-time Text over Internet Protocol
URI	Uniform Resource Identifier
UTF-8	UCS/Unicode Transformation Format-8
VCO/HCO	Voice Carry Over/Hearing Carry Over
VoIP	Voice over Internet Protocol

## 5. Requirements

The framework described in Section 6 defines a real-time text-based conversational service that is the text equivalent of voice-based telephony. This section describes the requirements that the framework is designed to meet and the functionality it should offer.

### 5.1. General Requirements for ToIP

Any framework for ToIP must be derived from the requirements of RFC 3351 [22]. A basic requirement is that it must provide a standardized way for offering real-time text-based conversational services that can be used as an equivalent to voice telephony by deaf, hard-of-hearing, speech-impaired, and mainstream users.

It is important to understand that real-time text conversations are significantly different from other text-based communications like email or Instant Messaging. Real-time text conversations deliver an equivalent mode to voice conversations by providing transmission of text character by character as it is entered, so that the conversation can be followed closely and that immediate interaction takes place.

Store-and-forward systems like email or messaging on mobile networks, or non-streaming systems like instant messaging, are unable to provide that functionality. In particular, they do not allow for smooth communication through a Text Relay Service.

In order to make ToIP the text equivalent of voice services, ToIP needs to offer equivalent features in terms of conversationality to those provided by voice. To achieve that, ToIP needs to:

- a. offer real-time transport and presentation of the conversation;
- b. provide simultaneous transmission in both directions;
- c. support both point-to-point and multipoint communication;
- d. allow other media, like audio and video, to be used in conjunction with ToIP;
- e. ensure that the real-time text service is always available.

Real-time text is a useful subset of Total Conversation as defined in ITU-T F.703 [5]. Total Conversation allows participants to use multiple modes of communication during the conversation, either at the same time or by switching between modes, e.g., between real-time text and audio.

Deaf, hard-of-hearing, and mainstream users may invoke ToIP services for many different reasons:

- because they are in a noisy environment, e.g., in a machine room of a factory where listening is difficult;
- because they are busy with another call and want to participate in two calls at the same time;
- for implementing text and/or speech recording services (e.g., text documentation/audio recording) for legal purposes, for clarity, or for flexibility;

- to overcome language barriers through speech translation and/or transcoding services;
- because of hearing loss, deafness, or tinnitus as a result of the aging process or for any other reason, creating a need to replace or complement voice with real-time text in conversational sessions.

In many of the above examples, real-time text may accompany speech. The text could be displayed side by side, or in a manner similar to subtitling in broadcasting environments, or in any other suitable manner. This could occur with users who are hard of hearing and also for mixed media calls with both hearing and deaf people participating in the call.

A ToIP user may wish to call another ToIP user, join a conference session involving several users, or initiate or join a multimedia session, such as a Total Conversation session.

A common scenario for multipoint real-time text is conference calling with many participants. Implementers could, for example, use different colours to render different participants' text, or could create separate windows or rendering areas for each participant.

## 5.2. Detailed Requirements for ToIP

The following sections list individual requirements for ToIP. Each requirement has been given a unique identifier (R1, R2, etc.). Section 6 (Implementation Framework) describes how to implement ToIP based on these requirements by using existing protocols and techniques.

The requirements are organized under the following headings:

- session setup and session control;
- transport;
- use of transcoding services;
- presentation and user control;
- interworking.



### 5.2.1. Session Setup and Control Requirements

Conversations could be started using a mode other than real-time text. Simultaneous or alternating voice and real-time text is used by a large number of people who can send voice but must receive text (due to a hearing impairment), or who can hear but must send text (due to a speech impairment).

R1: It SHOULD be possible to start conversations in any mode (real-time text, voice, video) or combination of modes.

R2: It MUST be possible for the users to switch to real-time text, or add real-time text as an additional modality, during the conversation.

R3: Systems supporting ToIP MUST allow users to select any of the supported conversation modes at any time, including in mid-conversation.

R4: Systems SHOULD allow the user to specify a preferred mode of communication in each direction, with the ability to fall back to alternatives that the user has indicated are acceptable.

R5: If the user requests simultaneous use of real-time text and audio, and this is not possible because of constraints in the network, the system SHOULD try to establish text-only communication if that is what the user has specified as his/her preference.

R6: If the user has expressed a preference for real-time text, establishment of a connection including real-time text MUST have priority over other outcomes of the session setup.

R7: It MUST be possible to use real-time text in conferences both as a medium of discussion between individual participants (for example, for sidebar discussions in real-time text while listening to the main conference audio) and for central support of the conference with real-time text interpretation of speech.

R8: Session setup and negotiation of modalities MUST allow users to specify the language of the real-time text to be used. (It is RECOMMENDED that similar functionality be provided for the video part of the conversation, i.e., to specify the sign language being used).

R9: Where certain session services are available for the audio media part of a session, these functions MUST also be supported for the real-time text media part of the same session. For example, call transfer must act on all media in the session.

### 5.2.2. Transport Requirements

ToIP will often be used to access a relay service [24], allowing real-time text users to communicate with voice users. With relay services, as well as in direct user-to-user conversation, it is crucial that text characters are sent as soon as possible after they are entered. While buffering may be done to improve efficiency, the delays SHOULD be kept minimal. In particular, buffering of whole lines of text will not meet character delay requirements.

R10: Characters must be transmitted soon after entry of each character so that the maximum delay requirement can be met. An end-to-end delay time of one second is regarded as good, while users note and appreciate shorter delays, down to 300ms. A delay of up to two seconds is possible to use.

R11: Real-time text transmission from a terminal SHALL be performed character by character as entered, or in small groups of characters, so that no character is delayed from entry to transmission by more than 300 milliseconds.

R12: It MUST be possible to transmit characters at a rate sufficient to support fast human typing as well as speech-to-text methods of generating real-time text. A rate of 30 characters per second is regarded as sufficient.

R13: A ToIP service MUST be able to deal with international character sets.

R14: Where it is possible, loss or corruption of real-time text during transport SHOULD be detected and the user should be informed.

R15: Transport of real-time text SHOULD be as robust as possible, so as to minimize loss of characters.

R16: It SHOULD be possible to send and receive real-time text simultaneously.

### 5.2.3. Transcoding Service Requirements

If the User Agents of different participants indicate that there is an incompatibility between their capabilities to support certain media types, e.g., one User Agent only offering T.140 over IP, as described in RFC 4103 [4], and the other one only supporting audio, the user might want to invoke a transcoding service.

Some users may indicate their preferred modality to be audio while others may indicate real-time text. In this case, transcoding

services might be needed for text-to-speech (TTS) and speech-to-text (STT). Other examples of possible scenarios for including a relay service in the conversation are: text bridging after conversion from speech, audio bridging after conversion from real-time text, etc.

A number of requirements, motivations, and implementation guidelines for relay service invocation can be found in RFC 3351 [22].

R17: It MUST be possible for users to invoke a transcoding service where such service is available.

R18: It MUST be possible for users to indicate their preferred modality (e.g., ToIP).

R19: It MUST be possible to negotiate the requirements for transcoding services in real time in the process of setting up a call.

R20: It MUST be possible to negotiate the requirements for transcoding services in mid-call, for the immediate addition of those services to the call.

R21: Communication between the end participants SHOULD continue after the addition or removal of a text relay service, and the effect of the change should be limited in the users' perception to the direct effect of having or not having the transcoding service in the connection.

R22: When setting up a session, it MUST be possible for a user to specify the type of relay service requested (e.g., speech to text or text to speech). The specification of a type of relay SHOULD include a language specifier.

R23: It SHOULD be possible to route the session to a preferred relay service even if the user invokes the session from another region or network than that usually used.

R24: It is RECOMMENDED that ToIP implementations make the invocation and use of relay services as easy as possible.

#### 5.2.4. Presentation and User Control Requirements

A user should never be in doubt about the status of the session, even if the user is unable to make use of the audio or visual indication. For example, tactile indications could be used by deaf-blind individuals.

R25: User Agents for ToIP services MUST have alerting methods (e.g., for incoming sessions) that can be used by deaf and hard-of-hearing people or provide a range of alternative, but equivalent, alerting methods that can be selected by all users, regardless of their abilities.

R26: Where real-time text is used in conjunction with other media, exposure of user control functions through the User Interface needs to be done in an equivalent manner for all supported media. For example, it must be possible for the user to select between audio, visual, or tactile prompts, or all must be supplied.

R27: If available, identification of the originating party (e.g., in the form of a URI or a Calling Line Identification (CLI)) MUST be clearly presented to the user in a form suitable for the user BEFORE the session invitation is answered.

R28: When a session invitation involving ToIP originates from a Public Switched Telephone Network (PSTN) text telephone (e.g., transcoded via a text gateway), this SHOULD be indicated to the user. The ToIP client MAY adjust the presentation of the real-time text to the user as a consequence.

R29: An indication SHOULD be given to the user when real-time text is available during the call, even if it is not invoked at call setup (e.g., when only voice and/or video is used initially).

R30: The user MUST be informed of any change in modalities.

R31: Users MUST be presented with appropriate session progress information at all times.

R32: Systems for ToIP SHOULD support an answering machine function, equivalent to answering machines on telephony networks.

R33: If an answering machine function is supported, it MUST support at least 160 characters for the greeting message. It MUST support incoming text message storage of a minimum of 4096 characters, although systems MAY support much larger storage. It is RECOMMENDED that systems support storage of at least 20 incoming messages of up to 16000 characters per message.

R34: When the answering machine is activated, user alerting SHOULD still take place. The user SHOULD be allowed to monitor the auto-answer progress, and where this is provided, the user SHOULD be allowed to intervene during any stage of the answering machine procedure and take control of the session.

R35: It SHOULD be possible to save the text portion of a conversation.

R36: The presentation of the conversation SHOULD be done in such a way that users can easily identify which party generated any given portion of text.

R37: ToIP SHOULD handle characters such as new line, erasure, and alerting during a session as specified in ITU-T T.140 [8].

#### 5.2.5. Interworking Requirements

There is a range of existing real-time text services. There is also a range of network technologies that could support real-time text services.

Real-time/interactive texting facilities exist already in various forms and on various networks. In the PSTN, they are commonly referred to as text telephony.

Text gateways are used for converting between different protocols for text conversation. They can be used between networks or within networks where different transport technologies are used.

R38: ToIP SHOULD provide interoperability with text conversation features in other networks, for instance the PSTN.

R39: When communicating via a gateway to other networks and protocols, the ToIP service SHOULD support the functionality for alternating or simultaneous use of modalities as offered by the interworking network.

R40: Calling party identification information, such as CLI, MUST be passed by gateways and converted to an appropriate form, if required.

R41: When interworking with other networks and services, the ToIP service SHOULD provide buffering mechanisms to deal with delays in call setup and with differences in transmission speeds, and/or to interwork with half-duplex services.

##### 5.2.5.1. PSTN Interworking Requirements

Analog text telephony is used in many countries, mainly by deaf, hard-of-hearing and speech-impaired individuals.

R42: ToIP services MUST provide interworking with PSTN legacy text telephony devices.

R43: When interworking with PSTN legacy text telephony services, alternating text and voice function MAY be supported. (Called "voice carry over (VCO) and hearing carry over (HCO)").

#### 5.2.5.2. Cellular Interworking Requirements

As mobile communications have been adopted widely, various solutions for real-time texting while on the move were developed. ToIP services should provide interworking with such services as well.

Alternative means of transferring the text telephony data have been developed when TTY services over cellular were mandated by the FCC in the USA. They are the a) "No-gain" codec solution, and b) the Cellular Text Telephony Modem (CTM) solution [7], both collectively called "Baudot mode" solution in the USA.

The GSM and 3G standards from 3GPP make use of the CTM modem in the voice channel for text telephony. However, implementations also exist that use the data channel to provide such functionality. Interworking with these solutions should be done using text gateways that set up the data channel connection at the GSM side and provide ToIP at the other side.

R44: a ToIP service SHOULD provide interworking with mobile text conversation services.

#### 5.2.5.3. Instant Messaging Interworking Requirements

Many people use Instant Messaging to communicate via the Internet using text. Instant Messaging usually transfers blocks of text rather than streaming as is used by ToIP. Usually a specific action is required by the user to activate transmission, such as pressing the ENTER key or a send button. As such, it is not a replacement for ToIP; in particular, it does not meet the needs for real-time conversations including those of deaf, hard-of-hearing, and speech-impaired users as defined in RFC 3351 [22]. It is less suitable for communications through a relay service [24].

The streaming nature of ToIP provides a more direct conversational user experience and, when given the choice, users may prefer ToIP.

R45: a ToIP service MAY provide interworking with Instant Messaging services.

## 6. Implementation Framework

This section describes an implementation framework for ToIP that meets the requirements and offers the functionality as set out in Section 5. The framework presented here uses existing standards that are already commonly used for voice-based conversational services on IP networks.

### 6.1. General Implementation Framework

This framework specifies the use of the Session Initiation Protocol (SIP) [2] to set up, control, and tear down the connections between ToIP users whilst the media is transported using the Real-Time Transport Protocol (RTP) [3] as described in RFC 4103 [4].

RFC 4504 describes how to implement support for real-time text in SIP telephony devices [23].

### 6.2. Detailed Implementation Framework

#### 6.2.1. Session Control and Setup

ToIP services MUST use the Session Initiation Protocol (SIP) [2] for setting up, controlling, and terminating sessions for real-time text conversation with one or more participants and possibly including other media like video or audio. The Session Description Protocol (SDP) used in SIP to describe the session is used to express the attributes of the session and to negotiate a set of compatible media types.

SIP [2] allows participants to negotiate all media, including real-time text conversation [4]. ToIP services can provide the ability to set up conversation sessions from any location as well as provision for privacy and security through the application of standard SIP techniques.

##### 6.2.1.1. Pre-Session Setup

The requirements of the user to be reached at a consistent address and to store preferences for evaluation at session setup are met by pre-session setup actions. That includes storing of registration information in the SIP registrar to provide information about how a user can be contacted. This will allow sessions to be set up rapidly and with proper routing and addressing.

The need to use real-time text as a medium of communications can be expressed by users during registration time. Two situations need to be considered in the pre-session setup environment:

- a. **User Preferences:** It **MUST** be possible for a user to indicate a preference for real-time text by registering that preference with a SIP server that is part of the ToIP service.
- b. **Server Support of User Preferences:** SIP servers that support ToIP services **MUST** have the capability to act on calling user preferences for real-time text in order to accept or reject the session. The actions taken can be based on the called users preferences defined as part of the pre-session setup registration. For example, if the user is called by another party, and it is determined that a transcoding server is needed, the session should be re-directed or otherwise handled accordingly.

The ability to include a transcoding service **MUST NOT** require user registration in any specific SIP registrar, but **MAY** require authorisation of the SIP registrar to invoke the service.

A point-to-point session takes place between two parties. For ToIP, one or both of the communicating parties will indicate real-time text as a possible or preferred medium for conversation using SIP in the session setup.

The following features **MAY** be implemented to facilitate the session establishment using ToIP:

- a. **Caller Preferences:** SIP headers (e.g., Contact) [10] can be used to show that real-time text is the medium of choice for communications.
- b. **Called Party Preferences [11]:** The called party being passive can formulate a clear rule indicating how a session should be handled, either using real-time text as a preferred medium or not, and whether this session needs to be handled by a designated SIP proxy or the SIP User Agent.
- c. **SIP Server Support for User Preferences:** It is **RECOMMENDED** that SIP servers also handle the incoming sessions in accordance with preferences expressed for real-time text. The SIP server can also enforce ToIP policy rules for communications (e.g., use of the transcoding server for ToIP).

#### 6.2.1.2. Session Negotiations

The Session Description Protocol (SDP) used in SIP [2] provides the capabilities to indicate real-time text as a medium in the session setup. RFC 4103 [4] uses the RTP payload types "text/red" and "text/t140" for support of ToIP, which can be indicated in the SDP as a part of the SIP INVITE, OK, and SIP/200/ACK media negotiations. In



addition, SIP's offer/answer model [12] can also be used in conjunction with other capabilities, including the use of a transcoding server for enhanced session negotiations [28,29,13].

#### 6.2.2. Transport

ToIP services MUST support the Real-Time Transport Protocol (RTP) [3] according to the specification of RFC 4103 [4] for the transport of real-time text between participants.

RFC 4103 describes the transmission of T.140 [8] real-time text on IP networks.

In order to enable the use of international character sets, the transmission format for real-time text conversation SHALL be UTF-8 [14], in accordance with ITU-T T.140.

If real-time text is detected to be missing after transmission, there SHOULD be a "text loss" indication in the real-time text as specified in T.140 Addendum 1 [8].

The redundancy method of RFC 4103 [4] SHOULD be used to significantly increase the reliability of the real-time text transmission. A redundancy level using 2 generations gives very reliable results and is therefore strongly RECOMMENDED.

In order to avoid exceeding the capabilities of the sender, receiver, or network (congestion), the transmission rate SHOULD be kept at or below 30 characters per second, which is the default maximum rate specified in RFC 4103 [4]. Lower rates MAY be negotiated when needed through the "cps" parameter as specified in RFC 4103 [4].

Real-time text capability is announced in SDP by a declaration similar to this example:

```
m=text 11000 RTP/AVP 100 98
a=rtpmap:98 t140/1000
a=rtpmap:100 red/1000
a=fmtp:100 98/98/98
```

By having this single coding and transmission scheme for real-time text defined in the SIP session control environment, the opportunity for interoperability is optimized. However, if good reasons exist, other transport mechanisms MAY be offered and used for the T.140-coded text, provided that proper negotiation is introduced, but the RFC 4103 [4] transport MUST be used as both the default and the fallback transport.

### 6.2.3. Transcoding Services

Invocation of a transcoding service MAY happen automatically when the session is being set up based on any valid indication or negotiation of supported or preferred media types. A transcoding framework document using SIP [28] describes invoking relay services, where the relay acts as a conference bridge or uses the third-party control mechanism. ToIP implementations SHOULD support this transcoding framework.

### 6.2.4. Presentation and User Control Functions

#### 6.2.4.1. Progress and Status Information

Session progress information SHOULD use simple language so that as many users as possible can understand it. The use of jargon or ambiguous terminology SHOULD be avoided. It is RECOMMENDED that text information be used together with icons to symbolise the session progress information.

In summary, it SHOULD be possible to observe indicators about:

- Incoming session
- Availability of real-time text, voice, and video channels
- Session progress
- Incoming real-time text
- Any loss in incoming real-time text
- Typed and transmitted real-time text

#### 6.2.4.2. Alerting

For users who cannot use the audible alerter for incoming sessions, it is RECOMMENDED to include a tactile, as well as a visual, indicator.

Among the alerting options are alerting by the User Agent's User Interface and specific alerting User Agents registered to the same registrar as the main User Agent.

It should be noted that external alerting systems exist and one common interface for triggering the alerting action is a contact closure between two conductors.

#### 6.2.4.3. Text Presentation

Requirement R32 states that, in the display of text conversations, users must be able to distinguish easily between different speakers. This could be done using color, positioning of the text (i.e., incoming real-time text and outgoing real-time text in different display areas), in-band identifiers of the parties, or a combination of any of these techniques.

#### 6.2.4.4. File Storage

Requirement R31 recommends that ToIP systems allow the user to save text conversations. This SHOULD be done using a standard file format. For example: a UTF-8 text file in XHTML format [15], including timestamps, party names (or addresses), and the conversation text.

#### 6.2.5. Interworking Functions

A number of systems for real-time text conversation already exist as well as a number of message-oriented text communication systems. Interoperability is of interest between ToIP and some of these systems.

Interoperation of half-duplex and full-duplex protocols, and between protocols that have different data rates, may require text buffering. Some intelligence will be needed to determine when to change direction when operating in half-duplex mode. Identification may be required of half-duplex operation either at the "user" level (i.e., users must inform each other) or at the "protocol" level (where an indication must be sent back to the gateway). However, special care needs to be taken to provide the best possible real-time performance.

Buffering schemes SHOULD be dimensioned to adjust for receiving at 30 characters per second and transmitting at 6 characters per second for up to 4 minutes (i.e., less than 3000 characters).

When converting between simultaneous voice and text on the IP side, and alternating voice and text on the other side of a gateway, a conflict can occur if the IP user transmits both audio and text at the same time. In such situations, text transmission SHOULD have precedence, so that while text is transmitted, audio is lost.

Transcoding of text to and from other coding formats may need to take place in gateways between ToIP and other forms of text conversation, for example, to connect to a PSTN text telephone.

Session setup through gateways to other networks may require the use of specially formatted addresses or other mechanisms for invoking those gateways.

ToIP interworking requires a method to invoke a text gateway. These text gateways act as User Agents at the IP side. The capabilities of the gateway during the call will be determined by the call capabilities of the terminal that is using the gateway. For example, a PSTN textphone is generally only able to receive voice and real-time text, so the gateway will only allow ToIP and audio.

Examples of possible scenarios for invocation of the text gateway are:

- a. PSTN textphone users dial a prefix number before dialing out.
- b. Separate real-time text subscriptions, linked to the phone number or terminal identifier/ IP address.
- c. Real-time text capability indicators.
- d. Real-time text preference indicators.
- e. Listen for V.18 modem modulation text activity in all PSTN calls and routing of the call to an appropriate gateway.
- f. Call transfer request by the called user.
- g. Placing a call via the Web, and using one of the methods described here
- h. A text gateway with its own telephone number and/or SIP address (this requires user interaction with the gateway to place a call).
- i. ENUM address analysis and number plan.
- j. Number or address analysis leads to a gateway for all PSTN calls.

#### 6.2.5.1. PSTN Interworking

Analog text telephony is cumbersome because of incompatible national implementations where interworking was never considered. A large number of these implementations have been documented in ITU-T V.18 [16], which also defines the modem detection sequences for the different text protocols. In rare cases, the modem type identification may take considerable time, depending on user actions.

To resolve analog textphone incompatibilities, text telephone gateways are needed to transcode incoming analog signals into T.140 and vice versa. The modem capability exchange time can be reduced by the text telephone gateways initially assuming the analog text telephone protocol used in the region where the gateway is located. For example, in the USA, Baudot [25] might be tried as the initial protocol. If negotiation for Baudot fails, the full V.18 modem capability exchange will take place. In the UK, ITU-T V.21 [26] might be the first choice.

In particular, transmission of real-time text on PSTN networks takes place using a variety of codings and modulations, including ITU-T V.21 [26], Baudot [25], dual-tone multi-frequency (DTMF), V.23 [27], and others. Many difficulties have arisen as a result of this variety in text telephony protocols and the ITU-T V.18 [16] standard was developed to address some of these issues.

ITU-T V.18 [16] offers a native text telephony method, plus it defines interworking with current protocols. In the interworking mode, it will recognise one of the older protocols and fall back to that transmission method when required.

Text gateways MUST use the ITU-T V.18 [16] standard at the PSTN side. A text gateway MUST act as a SIP User Agent on the IP side and support RFC 4103 real-time text transport.

While ToIP allows receiving and sending real-time text simultaneously and is displayed on a split screen, many analog text telephones require users to take turns typing. This is because many text telephones operate strictly half duplex. Only one can transmit text at a time. The users apply strict turn-taking rules.

There are several text telephones which communicate in full duplex, but merge transmitted text and received text in the same line in the same display window. Here too the users apply strict turn taking rules.

Native V.18 text telephones support full duplex and separate display from reception and transmission so that the full duplex capability can be used fully. Such devices could use the ToIP split screen as well, but almost all text telephones use a restricted character set and many use low text transmission speeds (4 to 7 characters per second).

That is why it is important for the ToIP user to know that he or she is connected with an analog text telephone. The session description [9] SHOULD contain an indication that the other endpoint for the call

is a PSTN textphone (e.g., connected via an ATA or through a text gateway). This means that the textphone user may be used to formal turn taking during the call.

#### 6.2.5.2. Mobile Interworking

Mobile wireless (or cellular) circuit switched connections provide a digital real-time transport service for voice or data. The access technologies include GSM, CDMA, TDMA, iDen, and various 3G technologies, as well as WiFi or WiMAX.

ToIP may be supported over the cellular wireless packet-switched service. It interfaces to the Internet.

The following sections describe how mobile text telephony is supported.

##### 6.2.5.2.1. Cellular "No-gain"

The "No-gain" text telephone transporting technology uses specially modified Enhanced Full Rate (EFR) [17] and Enhanced Variable Rate (EVR) [18] speech vocoders in mobile terminals used to provide a text telephony call. It provides full duplex operation and supports alternating between voice and text ("VCO/HCO"). It is dedicated to CDMA and TDMA mobile technologies and the US Baudot (i.e., 45 bit/s) type of text telephones.

##### 6.2.5.2.2. Cellular Text Telephone Modem (CTM)

CTM [7] is a technology-independent modem technology that provides the transport of text telephone characters at up to 10 characters/sec using modem signals that can be carried by many voice codecs and uses a highly redundant encoding technique to overcome the fading and cell changing losses.

##### 6.2.5.2.3. Cellular "Baudot mode"

This term is often used by cellular terminal suppliers for a cellular phone mode that allows TTYS to operate into a cellular phone and to communicate with a fixed-line TTY. Thus it is a common name for the "No-Gain" and the CTM solutions when applied to the Baudot-type textphones.

#### 6.2.5.2.4. Mobile Data Channel Mode

Many mobile terminals allow the use of the circuit-switched data channel to transfer data in real time. Data rates of 9600 bit/s are usually supported on the 2G mobile network. Gateways provide interoperability with PSTN textphones.

#### 6.2.5.2.5. Mobile ToIP

ToIP could be supported over mobile wireless packet-switched services that interface to the Internet. For 3GPP 3G services, ToIP support is described in 3G TS 26.235 [19].

#### 6.2.5.3. Instant Messaging Interworking

Text gateways MAY be used to allow interworking between Instant Messaging systems and ToIP solutions. Because Instant Messaging is based on blocks of text, rather than on a continuous stream of characters like ToIP, gateways MUST transcode between the two formats. Text gateways for interworking between Instant Messaging and ToIP MUST apply a procedure for bridging the different conversational formats of real-time text versus text messaging. The following advice may improve user experience for both parties in a call through a messaging gateway.

- a. Concatenate individual characters originating at the ToIP side into blocks of text.
- b. When the length of the concatenated message becomes longer than 50 characters, the buffered text SHOULD be transmitted to the Instant Messaging side as soon as any non-alphanumeric character is received from the ToIP side.
- c. When a new line indicator is received from the ToIP side, the buffered characters up to that point, including the carriage return and/or line-feed characters, SHOULD be transmitted to the Instant Messaging side.
- d. When the ToIP side has been idle for at least 5 seconds, all buffered text up to that point SHOULD be transmitted to the Instant Messaging side.
- e. Text Gateways must be capable of maintaining the real-time performance for ToIP while providing the interworking services.

It is RECOMMENDED that during the session, both users be constantly updated on the progress of the text input. Many Instant Messaging protocols signal that a user is typing to the other party in the

conversation. Text gateways between such Instant Messaging protocols and ToIP MUST provide this signalling to the Instant Messaging side when characters start being received, or at the beginning of the conversation.

At the ToIP side, an indicator of writing the Instant Message MUST be present where the Instant Messaging protocol provides one. For example, the real-time text user MAY see ". . . waiting for replying IM. . . " and when 5 seconds have passed another . (dot) can be shown.

Those solutions will reduce the difficulties between streaming and blocked text services.

Even though the text gateway can connect Instant Messaging and ToIP, the best solution is to take advantage of the fact that the user interfaces and the user communities for instant messaging and ToIP telephony are very similar. After all, the character input, character display, Internet connectivity, and SIP stack can be the same for Instant Messaging (SIMPLE) and ToIP. Thus, the user may simply use different applications for ToIP and text messaging in the same terminal.

Devices that implement Instant Messaging SHOULD implement ToIP as described in this document so that a more complete text communication service can be provided.

#### 6.2.5.4. Multi-Functional Combination Gateways

In practice, many interworking gateways will be implemented as gateways that combine different functions. As such, a text gateway could be built to have modems to interwork with the PSTN and support both Instant Messaging as well as ToIP. Such interworking functions are called combination gateways.

Combination gateways could provide interworking between all of their supported text-based functions. For example, a text gateway that has modems to interwork with the PSTN and that support both Instant Messaging and ToIP could support the following interworking functions:

- PSTN text telephony to ToIP
- PSTN text telephony to Instant Messaging
- Instant Messaging to ToIP



#### 6.2.5.5. Character Set Transcoding

Gateways between the ToIP network and other networks MAY need to transcode text streams. ToIP makes use of the ISO 10646 character set. Most PSTN textphones use a 7-bit character set, or a character set that is converted to a 7-bit character set by the V.18 modem.

When transcoding between character sets and T.140 in gateways, special consideration MUST be given to the national variants of the 7-bit codes, with national characters mapping into different codes in the ISO 10646 code space. The national variant to be used could be selectable by the user on a per-call basis, or be configured as a national default for the gateway.

The indicator of missing text in T.140, specified in T.140 amendment 1, cannot be represented in the 7-bit character codes. Therefore the indicator of missing text SHOULD be transcoded to the ' (apostrophe) character in legacy text telephone systems, where this character exists. For legacy systems where the ' character does not exist, the . (full stop) character SHOULD be used instead.

### 7. Further Recommendations for Implementers and Service Providers

#### 7.1. Access to Emergency Services

It must be possible to place an emergency call using ToIP and it must be possible to use a relay service in such a call. The emergency service provided to users utilising the real-time text medium must be equivalent to the emergency service provided to users utilising speech or other media.

A text gateway must be able to route real-time text calls to emergency service providers when any of the recognised emergency numbers that support text communications for the country or region are called, e.g., "911" in the USA and "112" in Europe. Routing real-time text calls to emergency services may require the use of a transcoding service.

A text gateway with cellular wireless packet-switched services must be able to route real-time text calls to emergency service providers when any of the recognized emergency numbers that support real-time text communication for the country is called.

#### 7.2. Home Gateways or Analog Terminal Adapters

Analog terminal adapters (ATA) using SIP-based IP communication and RJ-11 connectors for connecting traditional PSTN devices SHOULD enable connection of legacy PSTN text telephones [23].

These adapters SHOULD contain V.18 modem functionality, voice handling functionality, and conversion functions to/from SIP-based ToIP with T.140 transported according to RFC 4103 [4], in a similar way as it provides interoperability for voice sessions.

If a session is set up and text/t140 capability is not declared by the destination endpoint (by the endpoint terminal or the text gateway in the network at the endpoint), a method for invoking a transcoding server SHALL be used. If no such server is available, the signals from the textphone MAY be transmitted in the voice channel as audio with a high quality of service.

NOTE: It is preferred that such analog terminal adaptors do use RFC 4103 [4] on board and thus act as a text gateway. Sending textphone signals over the voice channel is undesirable due to possible filtering and compression and packet loss between the endpoints. This can result in character loss in the textphone conversation or even not allowing the textphones to connect to each other.

### 7.3. User Mobility

ToIP User Agents SHOULD use the same mechanisms as other SIP User Agents to resolve mobility issues. It is RECOMMENDED that users use a SIP address, resolved by a SIP registrar, to enable basic user mobility. Further mechanisms are defined for all session types for 3G IP multimedia systems.

### 7.4. Firewalls and NATs

ToIP uses the same signalling and transport protocols as VoIP. Hence, the same firewall and NAT solutions and network functionality that apply to VoIP MUST also apply to ToIP.

### 7.5. Quality of Service

Where Quality of Service (QoS) mechanisms are used, the real-time text streams should be assigned appropriate QoS characteristics, so that the performance requirements can be met and the real-time text stream is not degraded unfavourably in comparison to voice performance in congested situations.

## 8. Security Considerations

User confidentiality and privacy need to be met as described in SIP [2]. For example, nothing should reveal in an obvious way the fact that the ToIP user might be a person with a hearing or speech impairment. It is up to the ToIP user to make his or her hearing or speech impairment public. If a transcoding server is being used,

this SHOULD be as transparent as possible. However, it might still be possible to discern that a user might be hearing or speech impaired based on the attributes present in SDP, although the intention is that mainstream users might also choose to use ToIP. Encryption SHOULD be used on an end-to-end or hop-by-hop basis as described in SIP [2] and SRTP [20].

Authentication MUST be provided for users in addition to message integrity and access control.

Protection against Denial-of-Service (DoS) attacks needs to be provided, considering the case that the ToIP users might need transcoding servers.

## 9. Contributors

The following people contributed to this document: Willem Dijkstra, Barry Dingle, Gunnar Hellstrom, Radhika R. Roy, Henry Sinnreich, and Gregg C. Vanderheiden.

The content and concepts within are a product of the SIPPING Working Group. Tom Taylor (Nortel) acted as independent reviewer and contributed significantly to the structure and content of this document.

## 10. References

### 10.1. Normative References

- [1] Bradner, S., Ed., "Intellectual Property Rights in IETF Technology", BCP 79, RFC 3979, March 2005.
- [2] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, June 2002.
- [3] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, July 2003.
- [4] Hellstrom, G. and P. Jones, "RTP Payload for Text Conversation", RFC 4103, June 2005.
- [5] ITU-T Recommendation F.703, "Multimedia Conversational Services", November 2000.
- [6] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

- [7] 3GPP TS 26.226, "Cellular Text Telephone Modem Description" (CTM).
- [8] ITU-T Recommendation T.140, "Protocol for Multimedia Application Text Conversation" (February 1998) and Addendum 1 (February 2000).
- [9] Handley, M., Jacobson, V., and C. Perkins, "SDP: Session Description Protocol", RFC 4566, July 2006.
- [10] Rosenberg, J., Schulzrinne, H., and P. Kyzivat, "Indicating User Agent Capabilities in the Session Initiation Protocol (SIP)", RFC 3840, August 2004.
- [11] Rosenberg, J., Schulzrinne, H., and P. Kyzivat, "Caller Preferences for the Session Initiation Protocol (SIP)", RFC 3841, August 2004.
- [12] Rosenberg, J. and H. Schulzrinne, "An Offer/Answer Model with Session Description Protocol (SDP)", RFC 3264, June 2002.
- [13] Camarillo, G., Burger, E., Schulzrinne, H., and A. van Wijk, "Transcoding Services Invocation in the Session Initiation Protocol (SIP) Using Third Party Call Control (3pcc)", RFC 4117, June 2005.
- [14] Yergeau, F., "UTF-8, a transformation format of ISO 10646", STD 63, RFC 3629, November 2003.
- [15] "XHTML 1.0: The Extensible HyperText Markup Language: A Reformulation of HTML 4 in XML 1.0", W3C Recommendation, Available at <http://www.w3.org/TR/xhtml1>.
- [16] ITU-T Recommendation V.18, "Operational and Interworking Requirements for DCEs operating in Text Telephone Mode", November 2000.
- [17] TIA/EIA/IS-823-A, "TTY/TDD Extension to TIA/EIA-136-410 Enhanced Full Rate Speech Codec (must used in conjunction with TIA/EIA/IS-840)"
- [18] TIA/EIA/IS-127-2, "Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems, Addendum 2."
- [19] "IP Multimedia default codecs", 3GPP TS 26.235

- [20] Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K. Norrman, "The Secure Real-time Transport Protocol (SRTP)", RFC 3711, March 2004.
- [21] ITU-T Recommendation F.700, "Framework Recommendation for Multimedia Services", November 2000.

## 10.2. Informative References

- [22] Charlton, N., Gasson, M., Gybels, G., Spanner, M., and A. van Wijk, "User Requirements for the Session Initiation Protocol (SIP) in Support of Deaf, Hard of Hearing and Speech-impaired Individuals", RFC 3351, August 2002.
- [23] Sinnreich, H., Ed., Lass, S., and C. Stredicke, "SIP Telephony Device Requirements and Configuration", RFC 4504, May 2006.
- [24] European Telecommunications Standards Institute (ETSI), "Human Factors (HF); Guidelines for Telecommunication Relay Services for Text Telephones". TR 101 806, June 2000.
- [25] TIA/EIA/825 "A Frequency Shift Keyed Modem for Use on the Public Switched Telephone Network." (The specification for 45.45 and 50 bit/s TTY modems.)
- [26] International Telecommunication Union (ITU), "300 bits per second duplex modem standardized for use in the general switched telephone network". ITU-T Recommendation V.21, November 1988.
- [27] International Telecommunication Union (ITU), "600/1200-baud modem standardized for use in the general switched telephone network", ITU-T Recommendation V.23, November 1988.
- [28] Camarillo, G., "Framework for Transcoding with the Session Initiation Protocol", Work in Progress, May 2006.
- [29] Camarillo, G., "The SIP Conference Bridge Transcoding Model", Work in Progress, January 2006.

## Authors' Addresses

Guido Gybels  
Department of New Technologies  
RNID, 19-23 Featherstone Street  
London EC1Y 8SL, UK

Tel +44-20-7294 3713  
Txt +44-20-7296 8001 Ext 3713  
Fax +44-20-7296 8069  
EMail: [guido.gybels@rnid.org.uk](mailto:guido.gybels@rnid.org.uk)  
<http://www.ictrnid.org.uk>

Arnoud A. T. van Wijk  
Real-Time Text Taskforce (R3TF)

EMail: [arnoud@realtimetext.org](mailto:arnoud@realtimetext.org)  
<http://www.realtimetext.org>

## Full Copyright Statement

Copyright (C) The IETF Trust (2008).

This document is subject to the rights, licenses and restrictions contained in BCP 78, and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

## Intellectual Property

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at [ietf-ipr@ietf.org](mailto:ietf-ipr@ietf.org).

