

Network Working Group
Request for Comments: 3578
Category: Standards Track

G. Camarillo
Ericsson
A. B. Roach
dynamicsoft
J. Peterson
NeuStar
L. Ong
Ciena
August 2003

Mapping of Integrated Services Digital Network (ISDN)
User Part (ISUP) Overlap Signalling
to the Session Initiation Protocol (SIP)

Status of this Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Copyright Notice

Copyright (C) The Internet Society (2003). All Rights Reserved.

Abstract

This document describes a way to map Integrated Services Digital Network User Part (ISUP) overlap signalling to Session Initiation Protocol (SIP). This mechanism might be implemented when using SIP in an environment where part of the call involves interworking with the Public Switched Telephone Network (PSTN).

Table of Contents

1. Introduction	3
2. Conversion of ISUP Overlap Signalling into SIP en-bloc Signalling	3
2.1. Waiting for the Minimum Amount of Digits	4
2.2. The Minimum Amount of Digits has been Received	4
3. Sending Overlap Signalling to a SIP Network.	5
3.1. One vs. Several Transactions	5
3.2. Generating Multiple INVITEs.	6
3.3. Receiving Multiple Responses	8
3.4. Canceling Pending INVITE Transactions.	9
3.5. SIP to ISUP.	9
4. Security Considerations.	10
5. Acknowledgments.	10
6. Normative References	10
7. Intellectual Property Statement.	11
8. Authors' Addresses	12
9. Full Copyright Statement	13

1. Introduction

A mapping between the Session Initiation Protocol (SIP) [1] and the ISDN User Part (ISUP) [2] of SS7 is described in RFC 3398 [3]. However, RFC 3398 only takes into consideration ISUP en-bloc signalling. En-bloc signalling consists of sending the complete telephone number of the callee in the first signalling message. Although modern switches always use en-bloc signalling, some parts of the PSTN still use overlap signalling.

Overlap signalling consists of sending only some digits of the callee's number in the first signalling message. Further digits are sent in subsequent signalling messages. Although overlap signalling in the PSTN is the source of much additional complexity, it is still in use in some countries.

Like modern switches, SIP uses en-bloc signalling. The Request-URI of an INVITE request always contains the whole address of the callee. Native SIP end-points never generate overlap signalling.

Therefore, the preferred solution for a gateway handling PSTN overlap signalling and SIP is to convert the PSTN overlap signalling into SIP en-bloc signalling using number analysis and timers. The gateway waits until all the signalling messages carrying parts of the callee's number arrive, and only then, it generates a SIP INVITE request. Section 2 describes how to convert ISUP overlap signalling into en-bloc SIP this way.

However, although it is the preferred solution, conversion of overlap to en-bloc signalling sometimes results in unacceptable (multiple second) call setup delays to human users. In these situations, some form of overlap signalling has to be used in the SIP network to minimize the call setup delay. However, introducing overlap signalling in SIP introduces complexity and brings some issues. Section 3 analyzes the issues related to the use of overlap signalling in a SIP network and describe ways to deal with them in some particular network scenarios. Section 3 also describes in which particular network scenarios those issues make the use of overlap signalling in the SIP network unacceptable.

2. Conversion of ISUP Overlap Signalling into SIP en-bloc Signalling

In this scenario, the gateway receives an IAM (Initial Address Message) that contains only a portion of the called number. The rest of the digits dialed arrive later in one or more SAMs (Subsequent Address Message).

2.1. Waiting for the Minimum Amount of Digits

If the IAM contains less than the minimum amount of digits to route a call, the gateway starts T35 and waits until the minimum amount of digits that can represent a telephone number is received (or a stop digit is received). If T35 expires before the minimum amount of digits (or a stop digit) has been received, a REL with cause value 28 is sent to the ISUP side. T35 is defined in Q.764 [4] as 15-20 seconds.

If a stop digit is received, the gateway can already generate an INVITE request with the complete called number. Therefore, the call proceeds as usual.

2.2. The Minimum Amount of Digits has been Received

Once the minimum amount of digits that can represent a telephone number has been received, the gateway should use number analysis to decide if the number that has been received so far is a complete number. If it is, the gateway can generate an INVITE request with the complete called number. Therefore, the call proceeds as usual.

However, there are cases when the gateway cannot know whether the number received is a complete number or not. In this case, the gateway should collect digits until a timer (T10) expires or a stop digit (such as, #) is entered by the user (note that T10 is refreshed every time a new digit is received).

When T10 expires, an INVITE with the digits collected so far is sent to the SIP side. After this, any SAM received is ignored.

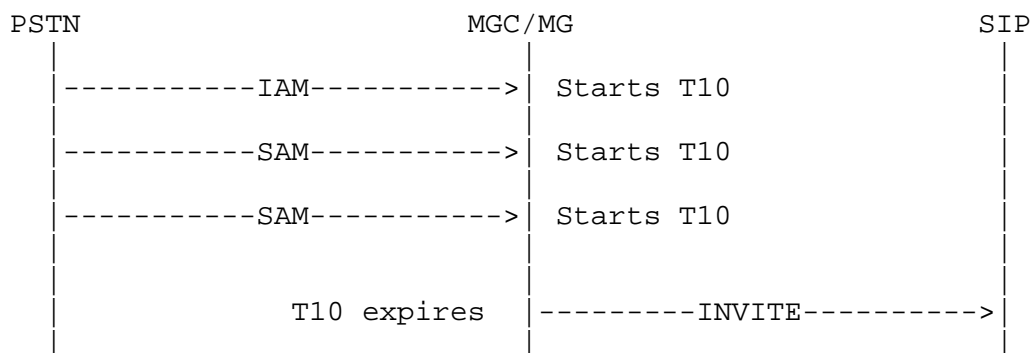


Figure 1: Use of T10 to convert overlap signalling to en-bloc

Note that T10 is defined for conversion between overlap signalling (e.g., CAS) and en-bloc ISUP. PSTN switches usually implement a locally defined value of timer T10 -- which may not be within the 4-6 second range recommended by Q.764 [4] -- to convert overlap ISUP to en-bloc ISUP. This document uses T10 and recommends the range of values defined in Q.764 [4], which seems suitable for conversion from overlap to en-bloc SIP operation. The actual choice of the timer value is a matter of local policy.

3. Sending Overlap Signalling to a SIP Network

This section analyzes the issues related to the use of overlap signalling in a SIP network and describes a possible solution and its applicability scope. It is important to note that, if used outside its applicability scope, this solution could cause a set of problems, which are identified in this section.

3.1. One vs. Several Transactions

An ingress gateway receiving ISUP overlap signalling (i.e., one IAM and one or more SAMs) needs to map it into SIP signalling. One possible approach would consist of sending an INVITE with the digits received in the IAM, and once an early dialog is established, sending the digits received in SAMs in a SIP request (e.g., INFO) within that early dialog.

This approach has several problems. It requires that the remote SIP user agent (which might be a gateway) sends a non-100 provisional response as soon as it receives the initial INVITE to establish the early dialog. Current gateways, following the procedures in RFC 3398 [3], do not generate such a provisional response. Having gateways generate such a response (e.g., 183 Session Progress) would cause ingress gateways to generate early ACMs, confusing the PSTN state machine even in calls that do not use overlap signalling.

In this approach, once the initial INVITE request is routed, all the subsequent requests sent within the early dialog follow the same path. That is, they cannot be re-routed to take advantage of SIP-based services. Therefore, we do not recommend using this approach.

An alternative approach consists of sending a new INVITE that contains all the digits received so far every time a new SAM is received. Since every new INVITE sent represents a new transaction, they can be routed in different ways. This way, every new INVITE can take advantage of any SIP service that the network may provide.

However, having subsequent INVITES routed in different ways brings some problems as well. The first INVITE, for instance, might be routed to a particular gateway, and a subsequent INVITE, to another. The result is that both gateways generate an IAM. Since one of the IAMs (or both) has an incomplete number, it would fail, having already consumed PSTN resources. It could even happen that both IAMs contained complete, but different numbers (i.e., one number is the prefix of the other one).

Routing in SIP can be controlled by the administrator of the network. Therefore, a gateway can be configured to generate SIP overlap signalling in the way described below only if the SIP routing infrastructure ensures that INVITES will only reach one gateway. When the routing infrastructure is not under the control of the administrator of the gateway, the procedures of Section 2 have to be used instead.

Within some dialing plans in the PSTN, a phone number might be a prefix of another one. This situation is not common, but it can occur. Where en-bloc signalling is used, this ambiguity is resolved before the digits are placed in the en-bloc signalling. If overlap signaling was used in this situation, a different user than the one the caller intended to call might be contacted. That is why in the parts of the PSTN where overlap is used, a prefix of a telephone number never identifies another valid number. Therefore, SIP overlap signalling should not be used when attempting to reach parts of the PSTN where it is possible for a number and some shorter prefix of the same number to both be valid addresses of different terminals.

3.2. Generating Multiple INVITES

In this scenario, the gateway receives an IAM (Initial Address Message) and possibly one or more SAMs (Subsequent Address Message) that provide more than the minimum amount of digits that can represent a phone number.

As soon as the minimum amount of digits is received, the gateway sends an INVITE and starts T10. This INVITE is built following the procedures described in RFC 3398 [3].

If a SAM arrives to the gateway, T10 is refreshed and a new INVITE with the new digits received is sent. The new INVITE has the same Call-ID and the same From header field including the tag as the first INVITE sent, but has an updated Request-URI. The new Request-URI contains all the digits received so far. The To header field of the new INVITE contains all the digits as well, but has no tag.

Note that it is possible to receive a response to the first INVITE before having sent the second INVITE. In this case, the response received would contain a To tag and information (Record-Route and Contact) to build a Route header field. The new INVITE to be sent (containing new digits) should not use any of these headers. That is, the new INVITE does not contain neither To tag nor Route header field. This way, this new INVITE can be routed dynamically by the network providing services.

The new INVITE should, of course, contain a Cseq field. It is recommended that the Cseq of the new INVITE is higher than any of the previous Cseq that the gateway has generated for this Call-ID (no matter for which dialog the Cseq was generated).

When an INVITE forks, responses from different locations might arrive establishing one or more early dialogs. New requests such as, PRACK or UPDATE can be sent within every particular early dialog. This implies that the Cseq number spaces of different early dialogs are different. Sending a new INVITE with a Cseq that is still unused by any of the remote destinations avoids confusion at the destination.

If the gateway is encapsulating ISUP messages as SIP bodies, it should place the IAM and all the SAMs received so far in this INVITE.

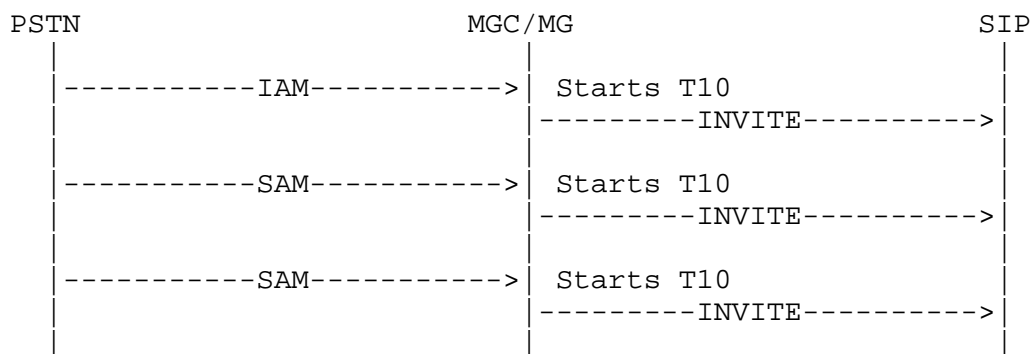


Figure 2: Overlap signalling in SIP

If 4xx, 5xx or 6xx final responses arrive (e.g., 484 address incomplete) for the pending INVITE transactions before T10 has expired, the gateway should not send any REL. A REL is sent only if no more SAMs arrive, T10 expires, and all the INVITES sent have been answered with a final response (different than 200 OK).

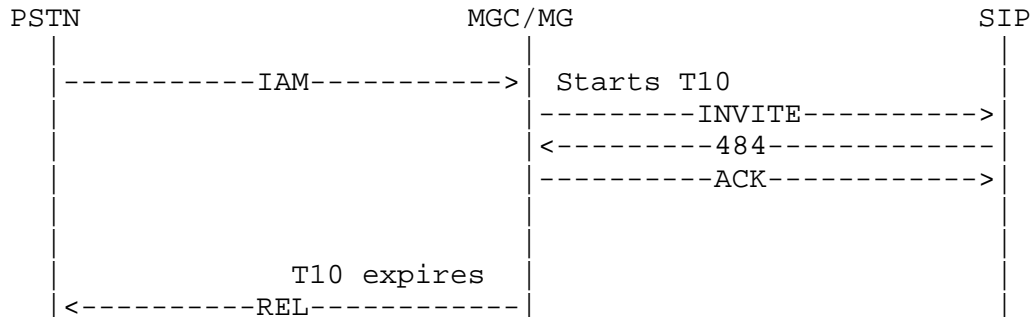


Figure 3: REL generation when overlap signalling is used

The best status code among all the responses received for all the INVITES that were generated is used to calculate the cause value of the REL as described in RFC 3398 [3].

The computation of the best response is done in the same way as forking proxies compute the best response to be returned to the client for a particular INVITE. Note that the best response is not always the response to the INVITE that contained more digits. If the user dials a particular number and then types an extra digit by mistake, a 486 (Busy Here) could be received for the first INVITE and a 484 (Address Incomplete) for the second one (which contained more digits).

3.3. Receiving Multiple Responses

When overlap signalling in SIP is used, the ingress gateway sends multiple INVITES. Accordingly, it will receive multiple responses. The responses to all the INVITES sent, except for one (normally, but not necessarily the last one), are typically 400 class responses (e.g., 484 Address Incomplete) that terminate the INVITE transaction.

However, a 183 Session Progress response with a media description can also be received. The media stream will typically contain a message such as, "The number you have just dialed does not exist".

The issue of receiving different 183 Session Progress responses with media descriptions does not only apply to overlap signalling. When vanilla SIP is used, several responses can also arrive to a gateway if the INVITE forked. It is then up to the gateway to decide which media stream should be played to the user.

However, overlap signalling adds a requirement to this process. As a general rule, a media stream corresponding to the response to an INVITE with a greater number of digits should be given more priority than media streams from responses with less digits.

3.4. Canceling Pending INVITE Transactions

When a gateway sends a new INVITE containing new digits, it should not CANCEL the previous INVITE transaction. This CANCEL could arrive before the new INVITE to an egress gateway and trigger a REL before the new INVITE arrived. INVITE transactions are typically terminated by the reception of 4xx responses.

However, once a 200 OK response has been received, the gateway should CANCEL all the other INVITE transactions were generated. A particular gateway might implement a timer to wait for some time before sending any CANCEL. This gives time to all the previous INVITE transactions to terminate smoothly without generating more signalling traffic (CANCEL messages).

3.5. SIP to ISUP

In this scenario (the call originates in the SIP network), the gateway receives multiple INVITEs that have the same Call-ID but have different Request-URIs. Upon reception of the first INVITE, the gateway generates an IAM following the procedures described in RFC 3398 [3].

When a gateway receives a subsequent INVITE with the same Call-ID and From tag as the previous one, and an updated Request-URI, a SAM should be generated as opposed to a new IAM. Upon reception of a subsequent INVITE, the INVITE received previously is answered with 484 Address Incomplete.

If the gateway is attached to the PSTN in an area where en-bloc signalling is used, a REL for the previous IAM and a new IAM should be generated.

4. Security Considerations

When overlap signaling is employed, it is possible that an attacker could send multiple INVITES containing an incomplete address to the same gateway in an attempt to occupy all available ports and thereby deny service to legitimate callers. Since none of these partially addressed calls would ever complete, in a traditional billing scheme, the sender of the INVITES might never be charged. To address this threat, the authors recommend that gateway operators authenticate the senders of INVITE requests, first, in order to have some accountability for the source of calls (it is very imprudent to give gateway access to unknown users on the Internet), but second, so that the gateway can determine when multiple calls are originating from the same source in a short period of time. Some sort of threshold of hanging overlap calls should be tracked by the gateway, and after the limit is exceeded, the further similar calls should be rejected to prevent the saturation of gateway trunking resources.

5. Acknowledgments

Jonathan Rosenberg, Olli Hynonen, and Mike Pierce provided useful feedback on this document.

6. Normative References

- [1] 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.
- [2] "Application of the ISDN user part of CCITT signaling system no. 7 for international ISDN interconnections", ITU-T Q.767, February 1991.
- [3] Camarillo, G., Roach, A. B., Peterson, J. and L. Ong, "Integrated Services Digital Network (ISDN) User Part (ISUP) to Session Initiation Protocol (SIP) Mapping", RFC 3398, December 2002.
- [4] "Signalling system no. 7 - ISDN user part signalling procedures," ITU-T Q.764, December 1999.

7. Intellectual Property Statement

The IETF takes no position regarding the validity or scope of any intellectual property 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; neither does it represent that it has made any effort to identify any such rights. Information on the IETF's procedures with respect to rights in standards-track and standards-related documentation can be found in BCP-11. Copies of claims of rights made available for publication 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 implementors or users of this specification can be obtained from the IETF Secretariat.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights which may cover technology that may be required to practice this standard. Please address the information to the IETF Executive Director.

8. Authors' Addresses

Gonzalo Camarillo
Ericsson
Advanced Signalling Research Lab.
FIN-02420 Jorvas
Finland

EMail: Gonzalo.Camarillo@ericsson.com

Adam Roach
dynamicsoft
5100 Tennyson Parkway
Suite 1200
Plano, TX 75024
USA

EMail: adam@dynamicsoft.com

Jon Peterson
NeuStar, Inc.
1800 Sutter St
Suite 570
Concord, CA 94520
USA

EMail: jon.peterson@neustar.biz

Lyndon Ong
Ciena
5965 Silver Creek Valley Road
San Jose, CA 95138
USA

EMail: lyong@ciena.com

9. Full Copyright Statement

Copyright (C) The Internet Society (2003). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assignees.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS 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.

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.

