Early Media in SIP: Problem Statement, Requirements, and Analysis of Solutions
draft-barnes-sip-em-ps-req-sol-00

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with Section 6 of BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on July 19, 2007.

Copyright Notice

Copyright (C) The IETF Trust (2007).

Abstract

The Session Initiation Protocol (SIP) enables and permits endpoints in an INVITE transaction to send media packets before an SDP offer/answer exchange has completed; we refer to such media packets as early media. The use of early media is common in current SIP implementations, and causes several problems, which have been documented elsewhere. This document puts forward the goal of modifying SIP so that compliant implementations will complete an SDP
Internet-Draft                em-ps-req-sol                 January 2007

exchange before sending media, and lays out high-level requirements for such a solution. The set of possible solutions is then laid out, and each possibility examined in light of these requirements.

Table of Contents

1. Introduction .................................................. 3
2. Requirements .................................................. 5
2.1. Elimination of Early Media ................................. 5
2.2. Additional Messaging ......................................... 6
2.3. Interoperability with current SIP standards ............... 6
2.4. Interoperability with the PSTN ............................. 6
2.5. Denial of Service ........................................... 6
2.6. IPR Constraints ............................................. 6
3. Early Media Solutions ......................................... 6
3.1. Minimizing SDP completion time ............................. 7
3.1.1. Fast completion of INVITE transaction .................. 7
3.1.2. Elimination of offer from INVITE ....................... 8
3.1.3. Provisional Responses ..................................... 8
3.1.4. Separate INVITE transaction .............................. 9
3.1.5. Non-INVITE SIP Mechanisms ............................... 9
3.1.6. Non-SIP mechanisms ...................................... 10
3.2. Handling remaining media ................................... 11
3.2.1. Unreliable SDP answer ................................... 11
3.2.2. Backwards-compatibility .................................. 11
4. IANA Considerations .......................................... 11
5. Security Considerations ...................................... 12
6. Acknowledgements .............................................. 12
7. References ..................................................... 12
7.1. Normative References ....................................... 12
7.2. Informative References ..................................... 12
Authors’ Addresses ............................................... 13
Intellectual Property and Copyright Statements ................. 14
1. Introduction

A primary function of the Session Initiation Protocol (SIP, [RFC3261]) is to carry SDP messages that form an SDP offer/answer exchange [RFC3264]. The offer/answer exchange serves to negotiate the parameters necessary for the two end hosts to send multimedia to each other, including transports, port numbers, codecs, and security parameters (e.g., MIKEY [RFC3830]). In the most common SIP INVITE transaction, the calling party (UAC) is the SDP offerer and the called party (UAS) is the SDP answerer; in this scenario, the SDP offer is carried in the SIP INVITE request, and the SDP answer in the 200/OK response.

Because there is often a substantial delay between the sending of an INVITE and the receipt a 200/OK, this assignment of SDP messages to SIP messages creates a substantial delay between an SDP offer and an SDP answer. Consequently, there is a period of time between the two messages in which the two parties have asymmetric information: The called party has knowledge of both the SDP offered by the calling party and of his own intended response, while the calling party knows only the offer he sent. Given this information, the called party has all the information he needs to begin sending media to the calling party (because this was included in the SDP offer); moreover, he can send these media without sending an SDP answer to the calling party. We define early media as media sent before the completion of an SDP offer/answer exchange.

Ours is a different definition than is used in RFC 3960 [RFC3960], which defines early media as "media ... that is exchanged before a particular session is accepted by the called user." The definition in RFC 3960 is consistent with the current usage of SDP within the SIP INVITE transaction, but at the same time conflates two functions of SIP: Negotiation of session-layer parameters (through the SDP offer/answer) and communication of an application-layer request and acceptance of a session (by the INVITE and 200/OK messages). As suggested above, it is this assignment of SDP messages to SIP messages that creates the possibility of early media.

Early media is known to create several problems for SIP operation:

- **Call failure:** When endpoints use early media for an extended period of time, timers can fire to close the UAC INVITE transaction before the early media finish and the UAS sends a 200/OK. This causes the session proposed in the initial INVITE to be dropped by the UAC, and in some implementations, the flow of media is cut off.
Confidentiality: Techniques such as MIKEY [RFC3830] and Security Descriptions [RFC4568], which use SDP to establish keys for media security protocols, require a full SDP offer/answer exchange to negotiate keys. Such keying techniques are therefore incompatible with early media.

Access Control: In the presence of early media, a UAC must anticipate that media will arrive without any signaling, and thus is obliged to receive (and presumably render) media received from any host on the Internet.

Interaction with forking/retargeting: When an INVITE from a UAC is forked or retarget to UASs that send early media, because these UASs send no signaling, the UAC has no information besides the media itself on which to base a decision as to which media stream to render.

Denial of Service: Early media is the basis of the "voice hammer" attack, described in [draft-ietf-mmusic-ice].

To date, attempts to deal with early media have failed to solve all of the problems caused by early media. Such work typically has started from the assumption that early media is a necessary component of SIP (largely due to RFC 3398 [RFC3398] and RFC 3578 [RFC3578]) and has generally avoided the issue of whether key use-cases (such as PSTN interoperability) can be satisfied without the use of early media. RFC 3960 presents two models for managing early media streams, which mitigate early-media related problems in some cases. However, the document neither specifies the models in sufficient detail nor makes their implementation a mandatory part of SIP. More recently, Stucker has written a draft extending 3960 by detailing a more current set of early media coping mechanisms, and proposing a solution that uses additional SDP attributes to mitigate the problems caused by early media [draft-stucker-sipping-early-media-coping]. Although Stucker’s proposed solution addresses several of the above problems, it stops short of requiring an SDP exchange before media transmission. Ejzak has also authored a recent draft regarding early media [draft-ejzak-sipping-p-em-auth]. However, his draft proposes the use of a p-header to prevent fraudulent uses of early media, which are largely orthogonal to the early media problems considered in this draft.

In contrast to previous approaches, this draft proposes to address the problems of early media by normatively modifying the SIP protocol to preclude the transmission of early media. In order to remove the possibility of early media, this modification must ensure (to the greatest extent possible) that the SDP offer/answer exchange has completed before any media are sent. At a high level, this goal can
be approached from two directions: To complete the SDP exchange as quickly and reliably as possible, and to forbid sending of media before completion. A successful solution will combine these two approaches, since as progress is made on the former, the latter becomes less onerous.

The goal of this draft is three-fold: First, in the above, we have provided a consolidated statement of the early media problem, namely that the presence of media before the completion of an SDP offer/answer exchange is harmful to the functioning of SIP calls. Second, Section 2 specifies requirements for an update to SIP that eliminates early media, in an effort to consolidate the requirements that have been presented to date and enable discussion of solutions. Third, Section 3 lays out the space of possible solutions, enumerating the trade-offs entailed by each so that consensus can be reached on the direction that should be taken by the SIP working group.

2.  Requirements

There have been several informal proposals to date that would mitigate or eliminate early media, but each was disqualified in turn by requirements as perceived by the SIP and SIPPING working groups. In this section, we seek to create a consolidated set of requirements so that a consistent evaluation of solutions can move forward. Below, we use the term "proposed solution" to refer to a protocol that normatively updates the relevant standards (e.g., RFCs 3261, 3264, and 3398) in such a way as to eliminate early media, while we use the term "current SIP" to refer to SIP as defined in RFC 3261 and related documents.

2.1.  Elimination of Early Media

The proposed solution must define a normative modification to the SIP protocol (and other relevant protocols) that minimizes or eliminates the possibility of early media in SIP sessions; i.e., it should minimize or eliminate the possibility that a compliant implementation could send media before an SDP offer/answer exchange has completed. Solutions in which receipt of SDP-bearing messages is explicitly acknowledged (e.g., an INVITE/200/ACK or INVITE/180/PRACK exchange) are preferred, since these acknowledgements are the only way to completely eliminate early media (otherwise the answerer cannot know with certainty that the offerer has received his answer). If a proposed solution does not include such acknowledgements, it must define UAS behaviors that minimize the probability of the UAS sending media before the UAC has received the SDP answer.
2.2. Additional Messaging

In order to minimize the impact of the proposed solution on call-setup times, the proposed solution should minimize the number of signaling messages that are required in addition to those used by current SIP. Proposed solutions should accommodate the fact that messages sent "on the media path" (i.e. directly between the two SIP endpoints) will often be unable to reach their destinations due to current media gating techniques.

2.3. Interoperability with current SIP standards

The proposed solution must define how entities implementing it will interact with entities that implement the current SIP protocol. The proposed solution should minimize the set of conditions under which a call would fail in such a hybrid scenario, but not in a scenario in which both endpoints use current SIP.

2.4. Interoperability with the PSTN

The proposed solution must be compatible with the creation of gateways to the PSTN, and may define normative updates as necessary to RFC 3398 and RFC 3578.

2.5. Denial of Service

The proposed solution must minimize the possibility of new denial of service attacks by minimizing the amount of state that is kept by a UAS prior to the completion of an offer/answer exchange. Ideally, no state in addition to the SIP INVITE state machine will be required (i.e., the initial SDP offer/answer will be conducted statelessly).

2.6. IPR Constraints

The proposed solution must, to the extent possible, avoid dependence on technologies on which there are known IPR constraints, as this tends to hinder adoption.

3. Early Media Solutions

The fundamental goal of a system to eliminate early media is to ensure that an SDP offer/answer exchange has completed before media packets are sent. Such a system will combine two approaches: Completing the SDP exchange as quickly and reliably as possible, and providing restrictions on the sending and receiving of media before such completion. A priori, neither approach is at odds with the requirements stated above, so we explore the possible mechanisms for
accomplishing each. Many of the solutions below have been suggested on the SIP and SIPPING mailing lists already, sometimes by multiple people independently; we have not made an effort to attribute them.

3.1. Minimizing SDP completion time

The simplest way to preclude early media is for the SIP endpoints to ensure that an SDP exchange has completed before media packets are generated. In order to convey the SDP messages comprising this exchange, it is necessary to modify the way that SDP messages are carried within SIP session establishment. There are a variety of mechanisms for achieving this goal, including modifications to messages within the INVITE transaction, other non-INVITE SIP mechanisms, and protocols outside of SIP.

3.1.1. Fast completion of INVITE transaction

Perhaps the most straightforward approach to quickly completing an SDP offer/answer exchange would be to simply complete the SDP exchange already inherent in the SIP INVITE transaction more quickly: To do this, the UAS would send a 200/OK response as soon as an INVITE message is received, rather than awaiting an indication from a higher layer to do so (formally, this moves both the client and server INVITE transactions to the Terminated state as quickly as possible). In essence, this is a change to the semantic meaning of the "200/OK" response; rather than indicating that an application-layer session has begun, it would indicate that an media association had been established, in the sense that all necessary parameters are established for media to flow. A separate message (for instance, an UPDATE [RFC3311]) could be used to indicate that an application-layer user is ready to begin exchanging media.

Quickly completing the INVITE transaction has the potential to eliminate early media entirely: Because the 200 response is explicitly acknowledged by the ACK message, both endpoints can be sure that the SDP exchange is complete when the ACK is transmitted. In ISUP terms, a 200/OK response used in this way would correspond to an ACM (rather than to an ANM, as in RFC 3398), and the UPDATE (or similar message) would correspond to the ANM. The failing of this approach is its interaction with SIP forking: Since a 200/OK message moves both endpoints to the Terminated state, the first UAS to respond would automatically cut off all other branches of the call. Although this issue could likely be resolved with a modification to rules for SIP forking, it could inhibit backward compatibility in the interim.
3.1.2. Elimination of offer from INVITE

One way to ensure that a UAS cannot send media until an SDP exchange has completed would be to forbid a UAC from including an SDP offer in its INVITE message, so that the SDP offer would be sent by the UAS in either a reliable provisional response or the 200/OK message, and the answer in the UAC’s PRACK or ACK message. This behavior is allowed by current SIP, but does not eliminate early media because it is not mandatory. At first blush, reusing the existing INVITE transaction in this way seems appealing: Because it is a subset of behaviors allowed by current SIP, it will interact well with current SIP endpoints, and it introduces no additional messaging or state machinery. However, if either endpoint does not support reliable provisional responses [RFC3262], this solution would prevent any media flows before the sending of a 200/OK (distinct from an SDP answer), which would cause many existing SIP applications to fail (including a large number of PSTN applications as they pass through gateways). In addition, the ACK message is not a reliable mechanism for transporting SDP answers, since the UAC (as an SDP answerer) receives no notification that the UAS has received its answer.

3.1.3. Provisional Responses

Within the INVITE transaction, the only other messages sent from UAS to UAC (besides the INVITE, 200, and ACK messages) that could carry SDP information are provisional (101-199) responses (response type 100 is excepted because it is often not retransmitted by proxies). Upon receipt of an INVITE message, a UAS would respond with a 1xx message that would complete the SDP offer/answer. Some have suggested use of the 183 (Call Progress) message for this purpose, as is done in ICE, though another existing or new response type could in principle be used.

This approach has the advantage that if the UAS does not support this use of provisional messages, then the call will proceed in the same way as with current SIP. For PSTN gateways, this approach only requires the sending or understanding a single additional message. However, it would be difficult to use provisional responses to eliminate early media, since this would require them to be reliable. Though there exist both SIP (PRACK) and non-SIP (ICE binding request) messages that can acknowledge provisional responses, AT&T holds IPR related to reliable provisional responses. This has limited the deployment of PRACK, and may have the same effect on relevant portions of ICE.
3.1.4. Separate INVITE transaction

The Application Server model of RFC 3960 provides an SDP exchange for early media by conducting separate INVITE transactions for an "Early" session and for the "Final" session. In a recently proposed realization and extension of this model [draft-stucker-sipping-early-media-indirection], any entity along the signaling path can include an EMIND (Early Media Indirection) header that will prompt the UAC to initiate a separate INVITE transaction with a third entity, e.g., a media server to provide ring-back tones or balance announcements. This transaction establishes an "Early Session," which is terminated when the original INVITE transaction is completed and the "Final Session" begun.

Using a separate INVITE transaction for early media has several advantages: Since the answer in a 2xx response triggers an ACK in the early INVITE transaction, it can be used to completely prevent early media. Prospects for interoperability with current SIP implementations are good, since no new mechanisms are employed; in particular, the EMIND header will be ignored by implementations that do not support it. And the entire burden of state maintenance is placed on the UAC, which must manage both INVITE transactions. However, using two separate INVITE transactions incurs significant signaling overhead, and it is unclear that this approach, at least as currently specified, can eliminate early media from PSTN gateways that are unable to separate early and regular media.

3.1.5. Non-INVITE SIP Mechanisms

There are other SIP messages that could be used to carry an SDP answer that would be formally outside of the INVITE transaction, but could be contemporaneous with it. Because these messages would be outside the INVITE transaction, a separate mechanism may be required to associate the transmitted SDP answer with the SDP offer in the INVITE request.

3.1.5.1. INFO and UPDATE

The SIP INFO method [RFC2976] is currently used to convey additional signaling information in the middle of a session, after the INVITE transaction has terminated. However, it could also be used to convey information relevant to an INVITE transaction, subject to the constraint in RFC 2976 that an INFO request must not change the state of a SIP call. While receipt of an SDP answer in an INFO request would change state at the TU layer (namely, in the UAC core), it would not change the transaction-layer state of the call. INFO messages carrying SDP answers can be associated with the corresponding SDP offer by virtue of the fact that the INFO message
carrying the answer will carry the same Call-ID as the INVITE message carrying the offer. The SIP UPDATE method [RFC3311] could be likewise used (in fact, RFC 3311 lists early media as the main motivation for UPDATE, and RFC 3960 uses UPDATE messages as the basis for its gateway model).

Current standards require that the UAS in an INFO or UPDATE transaction to send a response to an INFO or UPDATE message, so these messages would provide a reliable transport for SDP answers. Just as with provisional responses, using INFO or UPDATE messages to convey SDP answers is backwards-compatible, and requires minimal modification to PSTN gateways. One potential drawback is that as currently specified, UPDATE and INFO messages cannot be sent before a dialog is established (i.e., a 101-199 response is sent). If not changed or lifted by the proposed solution, this restriction could limit the efficacy of UPDATE (or INFO) for quickly concluding an SDP exchange, since the required lxx message would introduce additional delay in the transmission of the SDP answer.

3.1.5.2. Other Response Codes

In current SIP, if a UAS wishes to send early media in response to an INVITE, it will send the media without any SIP response. As an alternative, the UAS wishing to send early media could send a 401 response requiring http-auth with null authentication. The UAC would then signal its willingness to receive the media by sending a new INVITE (with the nonce for null authentication). In this model, the UAS’s SDP answer would be sent in its 401 response, and the UAC’s re-INVITE would confirm receipt of this answer.

Although this proposal does provide a reliable transport for SDP answers, it has significant drawbacks: In addition to adding significant messaging overhead, it requires both endpoints to maintain state across INVITE transactions (in addition to the nonce required by null authentication). And the prospects for backward compatibility with this solution are bleak: All calls to current SIP endpoints will fail unless the endpoints support the indicated authentication mechanisms, and forked calls would show unpredictable behavior due to the Heterogeneous Error Response Forking Problem (HERFP).

3.1.6. Non-SIP mechanisms

There are some proposals to use protocols other than SIP to resolve the problem of early media in SIP. For instance, a lower-level protocol could be defined that could negotiate and SDP session that could then be used by SIP, much as TLS negotiations are conducted before HTTPS data are sent. However, such a protocol would have to
duplicate many SIP features, and any such solution would require SIP entities to implement an additional protocol stack, which would have negative consequences for backward-compatibility and PSTN interoperability.

3.2. Handling remaining media

Early media can be completely eliminated only if neither party sends media until it is assured that the SDP offer/answer exchange is completed. Of course, this is only possible when both endpoints support the proposed solution and the SDP answer is carried in a reliable message: In this optimal case, both the offerer and the answerer must not send media until they have received assurance that the SDP exchange has completed. The proposed solution will specify the two remaining cases.

3.2.1. Unreliable SDP answer

If both endpoints support the solution, but the solution uses an unreliable message to carry the SDP answer, then the offerer can be assured that the SDP exchange has completed (because he has received the answer), but the answerer cannot. In this case, the offerer must not send media until the SDP exchange has completed. The proposed solution should provide an explicit, deterministic criterion for when the answerer may send media: For instance, the answerer may be free to send media after it receives media from the offerer, or after a timer fires that is set when the answer is sent.

3.2.2. Backwards-compatibility

Depending on the nature of the proposed solution, one party may not be able to know whether the other supports the solution. If a party cannot know, then it must behave as if the other party does support the solution (otherwise, the solution would never be used). If a party can be aware of the other party’s support, then the solution should allow a supporting party to knowingly send early media to a non-supporting party, since to do otherwise -- to mandate that a supporting party behave as it would if the other party supported the solution -- is incompatible with existing SIP applications that use early media. Nonetheless, an implementation of the solution should always avoid the use of early media; it should only send early media when it has received early media from a non-supporting endpoint.

4. IANA Considerations

This document makes no request of IANA.
5. Security Considerations

The only new security concern that arises from this draft is the possibility of introducing new denial of service attacks. The requirement addressing this concern is stated in Section 2.5 above. The performance of each candidate solution against this requirement is examined in the section where that solution's is proposed.

6. Acknowledgements

7. References

7.1. Normative References


7.2. Informative References


draft-ejzak-sipping-p-em-auth

draft-ietf-mmusic-ice

draft-stucker-sipping-early-media-coping
"Coping with Early Media in the Session Initiation Protocol (SIP)", October 2006.

draft-stucker-sipping-early-media-indirection
"Early Media INDirection (EMIND) in the Session Initiation Protocol (SIP)", October 2006.

Authors’ Addresses

Richard Barnes
BBN Technologies
9861 Broken Land Pkwy
Columbia, Maryland  21046
USA
Phone: +1-410-290-6169
Email: rbarnes@bbn.com

Matt Lepinski
BBN Technologies
10 Moulton St.
Cambridge, Massachusetts  02138
USA
Phone: +1-617-873-5939
Email: mlepinsk@bbn.com