Internet Engineering Task Force (IETF)
Request for Comments: 7584
Category: Standards Track
ISSN: 2070-1721

R. Ravindranath
T. Reddy
G. Salgueiro
Cisco

July 2015

Session Traversal Utilities for NAT (STUN) Message Handling for SIP Back-to-Back User Agents (B2BUAs)

Abstract

Session Initiation Protocol (SIP) Back-to-Back User Agents (B2BUAs) are often designed to be on the media path rather than just intercepting signaling. This means that B2BUAs often act on the media path leading to separate media legs that the B2BUA correlates and bridges together. When acting on the media path, B2BUAs are likely to receive Session Traversal Utilities for NAT (STUN) packets as part of Interactive Connectivity Establishment (ICE) processing.

This document defines behavior for a B2BUA performing ICE processing. The goal of this document is to ensure that B2BUAs properly handle SIP messages that carry ICE semantics in Session Description Protocol (SDP) and STUN messages received as part of the ICE procedures for NAT and Firewall traversal of multimedia sessions.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction 2
2. Terminology 4
3. SDP-Modifying Signaling-only B2BUA 5
4. Media Plane B2BUAs 5
  4.1. Overview 5
  4.2. Mandatory ICE Termination with B2BUA 6
  4.3. Optional ICE Termination with B2BUA 9
  4.4. STUN Handling in B2BUA with Forked Signaling 11
5. Security Considerations 12
6. References 12
  6.1. Normative References 12
  6.2. Informative References 13
Acknowledgements 13
Authors' Addresses 14

1. Introduction

In many SIP deployments, SIP entities exist in the SIP signaling and media path between the originating and final terminating endpoints, which go beyond the definition of a traditional SIP proxy. These SIP entities, commonly known as B2BUAs, are described in [RFC7092] and often perform functions not defined in Standards Track RFCs.

SIP [RFC3261] and other session control protocols that try to use a direct path for media are typically difficult to use across Network Address Translators (NATs). These protocols use IP addresses and transport port numbers encoded in the signaling, such as SDP [RFC4566] and, in the case of SIP, various header fields. Such addresses and ports are unreachable if any peers are separated by NATs.

Mechanisms such as STUN [RFC5389], Traversal Using Relays around NAT (TURN) [RFC5766], and ICE [RFC5245] did not exist when protocols like SIP began to be deployed. Some mechanisms, such as the early versions of STUN, started appearing, but they were unreliable and suffered a number of issues typical for UNilateral Self-Address Fixing (UNSAF) as described in [RFC3424]. For these reasons, B2BUAs are being used by SIP domains for SIP and media-related purposes. These B2BUAs use proprietary mechanisms to enable SIP devices behind NATs to communicate across the NAT.

[RFC7362] describes how B2BUAs can perform Hosted NAT Traversal (HNT) in certain deployments. Section 5 of [RFC7362] describes some of the issues with Session Border Controllers (SBCs) implementing HNT and offers some mitigation strategies. The most commonly used approach to solve these issues is "restricted-latching", defined in Section 5 of [RFC7362], whereby the B2BUA will not latch to any packets from a source public IP address other than the one the SIP User Agent (UA) uses for SIP signaling. However, this is susceptible to attacks where an attacker who is able to see the source IP address of the SIP UA may generate packets using the same IP address. There are other threats described in Section 5 of [RFC7362] for which Secure Real-time Transport Protocol (SRTP) [RFC3711] can be used as a solution. However, this would require the B2BUAs to terminate and reoriginate SRTP, which is not always desirable.

This document describes proper behavior of B2BUAs performing ICE processing. This includes defining consistent handling of SIP messages carrying ICE semantics in SDP and STUN messages received as part of the ICE procedures performed on the media path for NAT and Firewall traversal of multimedia sessions.

A B2BUA can use ICE [RFC5245], which provides authentication tokens (conveyed in the ice-ufrag and ice-pwd attributes) that allow the identity of a peer to be confirmed before engaging in media exchange. This can solve some of the security concerns with HNT solution. Further, ICE has other benefits like selecting an address when more than one address is available (e.g., a dual-stack environment where the host can have both IPv4 and IPv6 addresses), verifying that a path works before connecting the call, etc. For these reasons, endpoints often use ICE to pick a candidate pair for media traffic between two agents.

B2BUAs often operate on the media path and have the ability to modify SIP headers and SDP bodies as part of their normal operation. Such entities, when present on the media path, are likely to take an active role in the session signaling depending on their level of activity on the media path. For example, some B2BUAs modify portions of the SDP body (e.g., IP address, port) and subsequently modify the media packet headers as well. Section 18.6 of ICE [RFC5245] explains two different behaviors when B2BUAs are present. Some B2BUAs are likely to remove all the SDP ICE attributes before sending the SDP across to the other side. Consequently, the call will appear to both endpoints as though the other side doesn’t support ICE. There are other types of B2BUAs that pass the ICE attributes without modification, yet modify the default destination for media contained in the "m=" and "c=" lines and the RTCP attribute (defined in [RFC3605]). This will be detected as an ice-mismatch, and ICE processing will be aborted for the session. The session may continue if the endpoints are able to reach each other over the default candidate (sent in "m=" and "c=" lines).

Section 3.1.3 of [RFC7092] defines a SDP-Modifying Signaling-only B2BUA that operates in the signaling plane only and is not in the media path, but it does modify SDP bodies and is thus aware of and understands SDP syntax and semantics. Such B2BUA MUST follow the behavior mentioned in Section 3.

Section 3.2 of [RFC7092] describes three different categories of B2BUAs that operate on both the signaling (SIP and SDP) and media planes according to the level of involvement and active participation in the media plane:

- A B2BUA that acts as a simple media relay. It is effectively unaware of anything that is transported and only modifies the transport header (could be UDP/IP) of the media packets.

- A B2BUA that performs a media-aware role. It inspects and potentially modifies RTP or RTP Control Protocol (RTCP) headers; but it does not modify the payload of RTP/RTCP.

- A B2BUA that performs a media-termination role and operates at the media payload layer, such as RTP/RTCP payload (e.g., a transcoder).

When B2BUAs that operate on the media plane (media relay, media aware, or media termination) are involved in a session between two endpoints performing ICE, then it MUST follow the behavior described in Section 4.

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

All of the pertinent B2BUA terminology and taxonomy used in this document is defined in [RFC7092].

NATs are widely used in the Internet by consumers and organizations. Although specific NAT behaviors vary, this document uses the term "NAT", which maps to NAT and Network Address Port Translation (NAPT) terms from [RFC3022], for devices that map any IPv4 or IPv6 address and transport port number to another IPv4 or IPv6 address and transport port number. This includes consumer NATs, Firewall-NATs, IPv4-IPv6 NATs, Carrier-Grade NATs (CGNs) [RFC6888], etc.

3. SDP-Modifying Signaling-only B2BUA

An SDP-Modifying Signaling-only B2BUA is one that operates in the signaling plane only and is not in the media path, but it modifies SDP bodies as described in Section 3.1.3 of [RFC7092]. Such B2BUAs MUST NOT change the IP address in the "c=" line, the port in the "m=" line, and the ICE semantics of SDP, as doing so can cause an ice-mismatch.

4. Media Plane B2BUAs

4.1. Overview

When one or both of the endpoints are behind a NAT, and there is a B2BUA between the endpoints, the B2BUAs MUST support ICE or at a minimum support ICE lite functionality as described in [RFC5245]. Such B2BUAs MUST use the mechanism described in Section 2.2 of [RFC5245] to demultiplex STUN packets that arrive on the RTP/RTCP port.

The subsequent sections describe the behavior B2BUAs MUST follow for handling ICE messages. A B2BUA can terminate ICE and thus have two ICE contexts with either endpoint. The reason for ICE termination could be due to the need for B2BUA to be in the media path (e.g., address hiding for privacy, interworking between ICE to no-ICE, etc.). A B2BUA can also be in optional ICE termination mode and passes across the candidate list and STUN short-term credentials (ice-ufrag and ice-pwd attributes) from one endpoint to the other side after adding its own candidates. A B2BUA can be in optional ICE termination mode when it does not have a need to be on the media path. The below sections describe the behaviors for these two cases.

4.2. Mandatory ICE Termination with B2BUA

A B2BUA that wishes to always be in the media path follows these steps:

- When a B2BUA sends out the SDP, it MUST advertise support for ICE and MAY include B2BUA's candidates of different types for each component of each media stream.

- If the B2BUA is in ICE lite mode as described in Section 2.7 of [RFC5245], then it MUST send an a=ice-lite attribute and MUST include B2BUA host candidates for each component of each media stream.

- If the B2BUA supports full ICE, then it MAY include B2BUA's candidates of different types for each component of each media stream.

- The B2BUA MUST generate new username and password values for ice-ufrag and ice-pwd attributes when it sends out the SDP and MUST NOT propagate the ufrag password values it received in the incoming SDP. This ensures that the short-term credentials used for both the legs are different. The short-term credentials include authentication tokens (conveyed in the ice-ufrag and ice-pwd attributes), which the B2BUA can use to verify the identity of the peer. The B2BUA terminates the ICE messages on each leg and does not propagate them.

- The B2BUA MUST NOT propagate the candidate list received in the incoming SDP to the outbound SDP and instead only advertise its candidate list. The B2BUA MUST also add its default candidate in the "c=" line (IP address) and "m=" line (port). In this way, the B2BUA will be always in the media path.

- Depending on whether the B2BUA supports ICE lite or full ICE, it implements the appropriate procedures mentioned in [RFC5245] for ICE connectivity checks.

```
       +-------+            +-------------------+             +-----+
       | Alice |            | Media Plane B2BUA |             | Bob |
       +-------+            +-------------------+             +-----+
           |(1) INVITE               |(3) INVITE                 |
           |    a=ice-ufrag1         |    a=ice-ufrag2           |
           |    a=ice-pwd1           |    a=ice-pwd2             |
           |   (Alice’s IP, port)    |   (B2BUA’s IP, port)      |
           | (Alice’s candidate list)|   (B2BUA’s candidate list)|
           |------------------------>|-------------------------->|
           |                         |                           |
           |(2) 100 trying           |                           |
           |<------------------------|                           |
           |                         |(4) 100 trying             |
           |                         |<--------------------------|
           |                         |                           |
           |                         |(5) 200 OK                 |
           |                         |    a=ice-ufrag3           |
           |                         |    a=ice-pwd3             |
           |                         |    (Bob’s IP, port)       |
           |                         |    (Bob’s candidate list) |
           |                         |<--------------------------|
           |(6) 200 OK               |                           |
           |    a=ice-ufrag4         |-----------ACK------------>|
           |    a=ice-pwd4           |           (7)             |
           | (B2BUA’s IP, port)      |                           |
           |(B2BUA’s candidate list1)|                           |
           |<------------------------|                           |
           |                         |                           |
           |--------ACK------------->|                           |
           |        (8)              |                           |
           |                         |                           |
           |<----ICE Connectivity 1->|                           |
           |      checks+conclusion  |<-----ICE Connectivity 2-->|
           |           (9)           |      checks+conclusion    |
           |                         |           (10)            |
           |                         |                           |
           |<-------Media packets--->|<----Media packets-------->|
           |           (11)          |         (12)              |
           |                         |                           |
           |<---ICE keepalives 1---->|                           |
           |          (13)           |<----ICE keepalives 2----->|
                                                (13)
```

Figure 1: INVITE with SDP Having ICE and with a
Media Plane B2BUA Terminating ICE

The above figure shows an example call flow with two endpoints, Alice and Bob, using ICE processing, and a B2BUA handing STUN messages from both the endpoints. For the sake of brevity, the entire list of ICE SDP attributes are not shown. Also, the STUN messages exchanged as part of ICE connectivity checks are not shown. Key steps to note from the call flow are:

- Alice sends an INVITE with SDP having ICE candidates.

- The B2BUA modifies the received SDP from Alice by removing the received candidate list, gathering its own candidates, and generating new username and password values for ice-ufrag and ice-pwd attributes. The B2BUA also changes the "c=" line and "m=" line to have its default candidate and forwards the INVITE (Step 3) to Bob.

- Bob responds (Step 5) to the INVITE with his own list of candidates.

- The B2BUA responds to the INVITE from Alice with SDP having a B2BUA candidate list. The B2BUA generates new username and password values for ice-ufrag and ice-pwd attributes in the 200 OK response (Step 6).

- ICE Connectivity checks happen between Alice and the B2BUA in Step 9. Depending on whether the B2BUA supports ICE or ICE lite, it will follow the appropriate procedures mentioned in [RFC5245]. ICE Connectivity checks also happen between Bob and the B2BUA in Step 10. Steps 9 and 10 happen in parallel. The B2BUA always terminates the ICE messages on each leg and has two independent ICE contexts running.

- Media flows between Alice and Bob via B2BUA (Steps 11 and 12).

- STUN keepalives would be used between Alice and B2BUA (Step 13) and between Bob and B2BUA (Step 14) to keep NAT and Firewall bindings alive.

Since there are two independent ICE contexts on either side of the B2BUA, it is possible that ICE checks will conclude on one side before concluding on the other side. This could result in an ongoing media session for one end while the other is still being set up. Any such media received by the B2BUA would continue to be sent to the other side on the default candidate address (that was sent in "c=" line).

4.3. Optional ICE Termination with B2BUA

A B2BUA willing to be in the media path only for NAT traversal, but that does not otherwise require to be in the media path, can do the following steps mentioned in this section.

- When a B2BUA receives an incoming SDP with ICE semantics, it copies the received candidate list and appends its own candidate list in the outgoing SDP. The B2BUA also copies the ufrag/password values it received in the incoming SDP to the outgoing SDP and then sends out the SDP.
- The B2BUA’s candidates MAY have lower priority than the candidates provided by the endpoint, this way the endpoint and remote peer candidate pairs are tested first before trying candidate pairs with B2BUA’s candidates.
- After offer/answer is complete, the endpoints will have both the B2BUAs and remote peer candidates. It will then use ICE procedures described in Section 8 of [RFC5245] to nominate a candidate pair for sending and receiving media streams.
- With this approach, the B2BUA will be in the media path only if the ICE checks between all the candidate pairs formed from both the endpoints fail.

```
       +-------+            +-------------------+             +-----+
       | Alice |            | Media Plane B2BUA |             | Bob |
       +-------+            +-------------------+             +-----+
           |(1) INVITE               |(3)  INVITE                |
           |   a=ice-ufrag1          |     a=ice-ufrag1          |
           |   a=ice-pwd1            |     a=ice-pwd1            |
           |  (Alice’s IP, port)     | (Alice’s IP, port)        |
           |(Alice’s candidate list) | (Alice’s candidate list + |
           |                         |  B2BUA’s candidate list1) |
           |------------------------>|-------------------------->|
           |                         |                           |
           |(2)  100 trying          |                           |
           |<------------------------|                           |
           |                         |(4) 100 trying             |
           |                         |<--------------------------|
           |                         |                           |
           |                         |(5) 200 OK                 |
           |                         |    a=ice-ufrag2           |
           |                         |    a=ice-pwd2             |
           |                         | (Bob’s IP, port)          |
           |                         | (Bob’s candidate list)    |
           |                         |<--------------------------|
           |(6) 200 OK               |                           |
           |    a=ice-ufrag2         |-----------ACK------------>|
           |    a=ice-pwd2           |           (7)             |
           |  (Bob’s IP,port)        |                           |
           |(B2BUA’s candidate list2 |                           |
           | + Bob’s candidate list) |                           |
           |<------------------------|                           |
           |                         |                           |
           |----------ACK----------->|                           |
           |          (8)            |                           |
           |                         |                           |
           |<----ICE Connectivity 1 (9)------------------------->|
           |                         |                           |
           |<----ICE Connectivity 2->|                           |
           |      checks+conclusion  |<-----ICE Connectivity 2-->|
           |           (10)          |      checks+conclusion    |
           |                         |           (11)            |
           |                         |                           |
           |<-------------------Media packets------------------->|
           |                       (12)                          |
           |                         |                           |
           |<------------------ICE keepalives------------------->|
                                   (13)
```

Figure 2: INVITE with SDP Having ICE and with a
Media Plane B2BUA in Optional ICE Termination Mode

The above figure shows a sample call flow with two endpoints, Alice and Bob, doing ICE, and a B2BUA handing STUN messages from both the endpoints. For the sake of brevity, the entire ICE SDP attributes are not shown. Also, the STUN messages exchanged as part of the ICE connectivity checks are not shown. Key steps to note from the call flow are:

- Alice sends an INVITE with an SDP having its own candidate list.

- The B2BUA propagates the received candidate list in incoming SDP from Alice after adding its own candidate list. The B2BUA also propagates the received ice-ufrag and ice-pwd attributes from Alice in the INVITE (Step 3) to Bob. In this example, the B2BUA does not modify the default candidate sent in the "c=" line and "m=" line and retains the values sent originally from Alice. If B2BUA wants to be in the media path when ICE connectivity checks between endpoints fails or one of the endpoints does not support ICE, then it overwrites its candidate address and port as a default candidate in the "m=" and "c=" lines.

- Bob responds (Step 5) to the INVITE with his own list of candidates.

- The B2BUA responds to the INVITE from Alice with an SDP having a B2BUA’s candidate list and the candidate list received from Bob. The B2BUA would also propagate the received ice-ufrag and ice-pwd attributes from Bob in (Step 5) to Alice in the 200 OK response (Step 6).

- ICE Connectivity checks happen between Alice and Bob in (Step 9). ICE Connectivity checks also happen between Alice and the B2BUA and Bob and the B2BUA as shown in Steps 10 and 11. Steps 9, 10, and 11 happen in parallel. In this example, Alice and Bob conclude ICE with a candidate pair that enables them to send media directly.

- Media flows between Alice and Bob in Step 12.

4.4. STUN Handling in B2BUA with Forked Signaling

Because of forking, a B2BUA might receive multiple answers for a single outbound INVITE. When this occurs, the B2BUA SHOULD follow Sections 3.2 or 3.3 for all of those received answers.

5. Security Considerations

As described in Section 2.5 of [RFC5245], ICE uses the STUN short-term credential mechanism for authentication and message integrity. STUN connectivity checks include the MESSAGE-INTEGRITY attribute that contains HMAC-SHA1 of the STUN message, and the Hashed Message Authentication Code (HMAC) is computed using the key exchanged in the signaling channel. The signaling channel between the endpoints and B2BUA MUST be encrypted so that the key is not visible to eavesdroppers, otherwise the security benefits of short-term authentication would be lost.

6. References

6.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC3711] Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K. Norrman, "The Secure Real-time Transport Protocol (SRTP)", RFC 3711, DOI 10.17487/RFC3711, March 2004, <http://www.rfc-editor.org/info/rfc3711>.

[RFC5245] Rosenberg, J., "Interactive Connectivity Establishment (ICE): A Protocol for Network Address Translator (NAT) Traversal for Offer/Answer Protocols", RFC 5245, DOI 10.17487/RFC5245, April 2010, <http://www.rfc-editor.org/info/rfc5245>.

[RFC5389] Rosenberg, J., Mahy, R., Matthews, P., and D. Wing, "Session Traversal Utilities for NAT (STUN)", RFC 5389, DOI 10.17487/RFC5389, October 2008, <http://www.rfc-editor.org/info/rfc5389>.

[RFC5766] Mahy, R., Matthews, P., and J. Rosenberg, "Traversal Using Relays around NAT (TURN): Relay Extensions to Session Traversal Utilities for NAT (STUN)", RFC 5766, DOI 10.17487/RFC5766, April 2010, <http://www.rfc-editor.org/info/rfc5766>.

[RFC7092] Kaplan, H. and V. Pascual, "A Taxonomy of Session Initiation Protocol (SIP) Back-to-Back User Agents", RFC 7092, DOI 10.17487/RFC7092, December 2013, <http://www.rfc-editor.org/info/rfc7092>.

6.2. Informative References

[RFC3022] Srisuresh, P. and K. Egevang, "Traditional IP Network Address Translator (Traditional NAT)", RFC 3022, DOI 10.17487/RFC3022, January 2001, <http://www.rfc-editor.org/info/rfc3022>.

[RFC3261] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, DOI 10.17487/RFC3261, June 2002, <http://www.rfc-editor.org/info/rfc3261>.

[RFC3424] Daigle, L., Ed. and IAB, "IAB Considerations for UNilateral Self-Address Fixing (UNSAF) Across Network Address Translation", RFC 3424, DOI 10.17487/RFC3424, November 2002, <http://www.rfc-editor.org/info/rfc3424>.

[RFC3605] Huitema, C., "Real Time Control Protocol (RTCP) attribute in Session Description Protocol (SDP)", RFC 3605, DOI 10.17487/RFC3605, October 2003, <http://www.rfc-editor.org/info/rfc3605>.

[RFC4566] Handley, M., Jacobson, V., and C. Perkins, "SDP: Session Description Protocol", RFC 4566, DOI 10.17487/RFC4566, July 2006, <http://www.rfc-editor.org/info/rfc4566>.

[RFC6888] Perreault, S., Ed., Yamagata, I., Miyakawa, S., Nakagawa, A., and H. Ashida, "Common Requirements for Carrier-Grade NATs (CGNs)", BCP 127, RFC 6888, DOI 10.17487/RFC6888, April 2013, <http://www.rfc-editor.org/info/rfc6888>.

[RFC7362] Ivov, E., Kaplan, H., and D. Wing, "Latching: Hosted NAT Traversal (HNT) for Media in Real-Time Communication", RFC 7362, DOI 10.17487/RFC7362, September 2014, <http://www.rfc-editor.org/info/rfc7362>.

Acknowledgements

Special thanks to Dan Wing, Pal Martinsen, Charles Eckel, Marc Petit-Huguenin, Simon Perreault, Lorenzo Miniero, Ari Keranen, and Parthasarathi Ravindran for their constructive comments, suggestions, and early reviews that were critical to the formulation and refinement of this document.

Thanks to Ben Campbell, Barry Leiba, Nevil Brownlee, Spencer Dawkins, Sam Hartman, Stephen Farrell, Kathleen Moriarty, and Francis Dupont for their thoughtful review comments.

Authors' Addresses

Ram Mohan Ravindranath
Cisco Systems, Inc.
Cessna Business Park, Varthur Hobli
Sarjapur Marathahalli Outer Ring Road
Bangalore, Karnataka 560103
India

Email: rmohanr@cisco.com

Tirumaleswar Reddy
Cisco Systems, Inc.
Cessna Business Park, Varthur Hobli
Sarjapur Marathalli Outer Ring Road
Bangalore, Karnataka 560103
India

Email: tireddy@cisco.com

Gonzalo Salgueiro
Cisco Systems, Inc.
7200-12 Kit Creek Road
Research Triangle Park, NC 27709
United States

Email: gsalguei@cisco.com