

RTCWEB  
Internet-Draft  
Intended status: Standards Track  
Expires: January 9, 2013

T. Reddy  
Muthu A M. Perumal  
R. Ram Mohan  
D. Wing  
Cisco  
July 8, 2012

STUN Extensions for Authenticated Firewall Traversal  
draft-reddy-rtcweb-stun-auth-fw-traversal-00

## Abstract

Some networks deploy firewalls configured to block UDP traffic. When SIP user agents or WebRTC endpoints are deployed behind such firewalls, media cannot be sent over UDP across the firewall, but must be sent using TCP (which causes a different user experience) or through a session border controller.

This draft describes an alternate model wherein extensions to ICE connectivity checks can be examined by the firewall to permit outgoing UDP flows across the firewall.

## Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

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."

This Internet-Draft will expire on January 9, 2013.

## Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of

Internet-Draft

STUN extn for FW Traversal

July 2012

publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

<a href="#">1.</a>	<a href="#">Introduction . . . . .</a>	<a href="#">3</a>
<a href="#">2.</a>	<a href="#">Notational Conventions . . . . .</a>	<a href="#">4</a>
<a href="#">3.</a>	<a href="#">Problem Statement . . . . .</a>	<a href="#">4</a>
<a href="#">4.</a>	<a href="#">Solution Overview . . . . .</a>	<a href="#">5</a>
<a href="#">4.1.</a>	<a href="#">Different Components and the Trust model . . . . .</a>	<a href="#">5</a>
<a href="#">5.</a>	<a href="#">Usage and Processing . . . . .</a>	<a href="#">6</a>
<a href="#">5.1.</a>	<a href="#">Generating FW-FLOWDATA Attribute . . . . .</a>	<a href="#">6</a>
<a href="#">5.2.</a>	<a href="#">Sending FW-FLOWDATA Attribute in Binding Request . . . . .</a>	<a href="#">6</a>
<a href="#">5.3.</a>	<a href="#">Firewalls processing FW-FLOWDATA Attribute . . . . .</a>	<a href="#">7</a>
<a href="#">6.</a>	<a href="#">STUN Attribute Format . . . . .</a>	<a href="#">9</a>
<a href="#">7.</a>	<a href="#">Key Provisioning . . . . .</a>	<a href="#">11</a>
<a href="#">8.</a>	<a href="#">Security Considerations . . . . .</a>	<a href="#">11</a>
<a href="#">9.</a>	<a href="#">IANA Considerations . . . . .</a>	<a href="#">12</a>
<a href="#">10.</a>	<a href="#">Acknowledgements . . . . .</a>	<a href="#">12</a>
<a href="#">11.</a>	<a href="#">References . . . . .</a>	<a href="#">12</a>
<a href="#">11.1.</a>	<a href="#">Normative References . . . . .</a>	<a href="#">12</a>
<a href="#">11.2.</a>	<a href="#">Informative References . . . . .</a>	<a href="#">13</a>
	<a href="#">Authors' Addresses . . . . .</a>	<a href="#">13</a>

## [1.](#) Introduction

To protect networks using real-time communications, firewalls or session border controllers are typically deployed.

Firewalls include Application Layer Gateway functionality, which intercepts and analyzes the session signaling traffic such as the Session Initiation Protocol (SIP) traffic and creates dynamic mapping to permit the media traffic. In particular, firewall extracts the media transport addresses and ports from the session description and creates dynamic mapping for media to flow through. This model has the following problems:

1. It does not work if the session signaling is end-to-end encrypted (say, using TLS).
2. It does not work if a non-standard session signaling is used that the firewall does not understand.
3. It does not work if the session signaling and media traverse different firewalls.

When an enterprise deploys WebRTC, the above problems are relevant because:

1. The session signaling between the WebRTC application running in the browser and the web server could be using TLS.
2. WebRTC does not enforce a particular session signaling protocol to be used. So, the firewall may not be able to understand it.
3. This session signaling and the peer-to-peer media may traverse different firewalls.

As a result the firewall may block ICE connectivity checks and media traffic.

Session Border Controllers (SBCs) are active participants with call signaling. Like firewalls, they also create dynamic mappings to permit media traffic. This forces call signaling and media through specific IP addresses, belonging to the SBC or an SBC-controlled media relay device.

This draft has the WebRTC server generate a cryptographic token which is passed to the WebRTC endpoint. The endpoint includes the token in its ICE connectivity checks. The firewall intercepts the ICE connectivity checks containing the token, validates it, and permits the ICE connectivity checks and the subsequent media flow through the

firewall.

## [2.](#) Notational Conventions

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](#)].

This note uses terminology defined in [[RFC5245](#)].

## [3.](#) Problem Statement

In the below topology, a WebRTC server is deployed in the enterprise Data Center. Alice makes a WebRTC call to Bob. For the two endpoints to successfully establish media sessions, firewalls FW1 and FW2 need to permit the ICE connectivity checks and media traffic. In such scenarios the mechanism described in this draft proposes a new comprehension-optional FW-FLOWDATA STUN attribute to be included in STUN Bind requests sent during ICE connectivity checks so that firewalls will permit media traffic between internal peers. This STUN attribute is created by the trusted WebRTC server and sent to the endpoints to be propagated by the respective ICE agents during ICE connectivity checks.

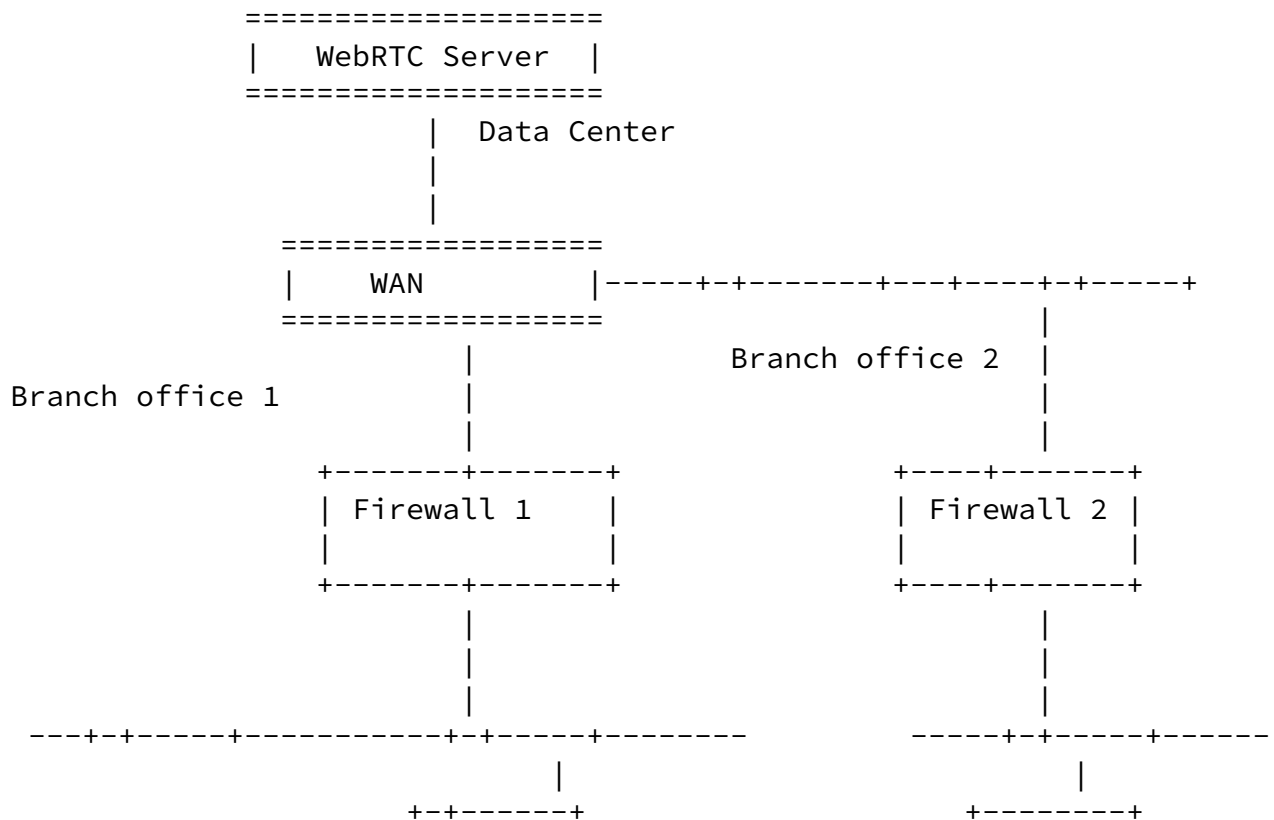




Figure 1: WebRTC service in enterprise - internal call

## 4. Solution Overview

This section gives an overview of the solution and the different components involved in the solution and the role of each component.

### 4.1. Different Components and the Trust model

Figure 1 above shows typical components involved in a WebRTC scenario. As part of the call setup, the WebRTC endpoint would have to gather its candidates from a STUN/TURN server, send the candidates in the offer to the peer endpoint. On receiving the answer from the peer endpoint it starts the ICE connectivity checks. As discussed in the problem statement, firewalls would typically block these ICE connectivity checks and media flowing there after. To allow this traffic a firewall needs to authorize the flow.

- o A new comprehensive optional STUN attribute called FW-FLOWDATA is defined as part of this draft. This is used by WebRTC endpoints requiring firewall traversal.

- o This STUN attribute FW-FLOWDATA is generated by the WebRTC server in co-ordination with the WebRTC endpoint.
- o Once the WebRTC session ends the firewall's dynamic mappings are closed after timeout.
- o DISCUSSION: Could we could have a FW-FLOWDATA attribute sent in a STUN message to close the dynamic mappings in the firewalls?

## 5. Usage and Processing

An RTP endpoint which generates media can include the FW-FLOWDATA attribute in its STUN Binding requests used in ICE connectivity checks, to inform on-path firewalls to permit the flow.

### [5.1.](#) Generating FW-FLOWDATA Attribute

The WebRTC server after processing the OFFER/ANSWER sends the FW-FLOWDATA STUN attribute to both the peers to be included in the ICE connectivity checks. The Authentication Tag field in the FW-FLOWDATA attribute contains the digest of the FW-FLOWDATA attribute for data origin authentication and integrity protection. The server first selects the candidate address info based on OFFER/ANSWER exchange and generates other fields of this attribute. The server then computes a digest for the FW-FLOWDATA attribute using HMAC-SHA1. The key for HMAC-SHA1 is provisioned using the technique in [Section 7](#). The result of which is truncated to 96 bits (retaining the left most bits) to produce HMAC-SHA-1-96 and input into the Authentication Tag field. The mechanism to send FW-FLOWDATA attribute from the WebRTC server to the client is outside the scope of this draft. But it is assumed that signalling protocols used for WebRTC call setup will be enhanced to deliver this new attribute to the WebRTC client. The WebRTC server MUST provide a new FW-FLOWDATA to allow the media session to continue before Lifetime expires.

### [5.2.](#) Sending FW-FLOWDATA Attribute in Binding Request

Once a WebRTC endpoint receives the FW-FLOWDATA, it is responsible for generating the STUN message and retransmitting the transactions per the STUN specification. The FW-FLOWDATA attribute should be placed before the FINGERPRINT attribute (if present) and after the MESSAGE-INTEGRITY attribute. The STUN length field is adjusted to point to the new end of the STUN message; that is, the STUN length field always accurately indicates the length of the STUN message (including the MESSAGE-INTEGRITY, FINGERPRINT, and FW-FLOWDATA attributes). This does not interfere with 3rd party receivers of the STUN message, as they will adjust the STUN length field to point to

the end of the MESSAGE-INTEGRITY field. Receivers that do not understand the FW-FLOWDATA will ignore it.

FW-FLOWDATA attribute received by the WebRTC client is passed to the web browser's ICE agent (API to be added in in W3C WebRTC-API specification [I.D.w3c-webrtc]). The ICE agent includes the FW-FLOWDATA attribute with all ICE connectivity checks, so that on-path firewalls can validate and permit the ICE connectivity checks and

forthcoming media.

For the FW-FLOWDATA attribute to be visible to the firewalls between the client and the TURN server, the FW-FLOWDATA should be included in the ALLOCATE request, channel bind or refresh messages going to the TURN server. This is to avoid firewalls having to look for STUN packets within STUN (TURN) packets.

### 5.3. Firewalls processing FW-FLOWDATA Attribute

Firewalls can reliably determine a UDP message is a STUN message because all STUN messages sent as ICE connectivity checks include the 32-bit STUN magic cookie and the FINGERPRINT attribute. STUN messages which are authenticated also include a MESSAGE-INTEGRITY attribute which authenticates the fields prior to the MESSAGE-INTEGRITY.

When the firewall receives a STUN binding request with FW-FLOWDATA attribute it stores the Authentication Tag in the FW-FLOWDATA attribute. The firewall then generates a digest for the FW-FLOWDATA attribute using HMAC-SHA1. The result of which is truncated to 96 bits (retaining the left most bits) to produce HMAC-SHA-1-96. If the value of the newly generated digest HMAC-SHA-1-96 is identical to the stored one, the firewall can ensure that the FW-FLOWDATA attribute has not been tampered with. Otherwise the packet is discarded.

To facilitate timestamp checking for replay attacks, each firewall SHOULD store the following information for each host: (The timestamp check is mostly brought from SEND [[RFC3971](#)])

- o The receive time of the last received and accepted STUN message with FW-FLOWDATA Attribute. This is called RDlast.
- o The timestamp in the last received and accepted STUN message with FW-FLOWDATA Attribute. This is called TSlast.

When a message is received from a new host (i.e., one that is not stored in the cache), the received timestamp, TSnew, is checked, and the packet is accepted if the timestamp is recent enough to the reception time of the packet, RDnew:



The recommended value for the allowed Delta is 180 seconds. If the timestamp is NOT within the boundaries then discard the STUN message.

When a message is received from a known host (i.e., one that already has an entry in the cache), the timestamp is checked against the previously received STUN message with FW-FLOWDATA Attribute:

$$TS_{new} + fuzz > TS_{last} + (RD_{new} - RD_{last}) \times (1 - drift) - fuzz$$

If this inequality does not hold, firewall SHOULD discard the message. If, on the other hand, the inequality holds, then firewall SHOULD process the message. Moreover, if the above inequality holds and  $TS_{new} > TS_{last}$ , the receiver SHOULD update  $RD_{last}$  and  $TS_{last}$ . Otherwise, the receiver MUST NOT update  $RD_{last}$  or  $TS_{last}$ . The default value of fuzz is 1 second and drift 1 % (0.01).

The firewall also performs the following checks:

- o Ensures that the source IP address and UDP port of the packet matches with one of the local CAI entries or remote CAI entries in the payload unless the firewall is configured to ignore this check.
- o Ensures the destination IP address and UDP port of the packet matches with one of the remote CAI entries or local CAI entries in the packet payload.

If all the above checks pass then the firewall creates the 5-tuple dynamic mapping using the local candidate IP address, local candidate port, remote candidate IP address, remote candidate port, transport protocol. The session time of the dynamic mapping will be set to a short lifetime (default value of 60 seconds).

- o If the initial ICE connectivity check includes the ICE-CONTROLLING attribute but does not include USE-CANDIDATE and a subsequent ICE connectivity check includes both these attributes, the firewall can determine that the ICE agent is the controlling agent using regular nomination and this candidate pair is nominated for media flow. The firewall then sets the session time of the dynamic mapping equal to the Lifetime field in FW-FLOWDATA attribute.

- o If the initial ICE connectivity check includes the ICE-CONTROLLED attribute but does not include USE-CANDIDATE and a subsequent ICE connectivity check includes both the attributes, the firewall can determine that the ICE agent is the controlled agent using regular nomination and this candidate pair is nominated for media flow. The

firewall then sets the session time of the dynamic mapping equal to the Lifetime field in FW-FLOWDATA attribute.

- o If the initial ICE connectivity check includes the ICE-CONTROLLING attribute and the USE-CANDIDATE attribute, firewall can determine that the ICE agent is the controlling agent using aggressive nomination. It then waits for the media traffic to flow before setting the session time of the dynamic mapping equal to Lifetime field in FW-FLOWDATA attribute.

- o If the initial ICE connectivity check includes the ICE-CONTROLLED attribute and the USE-CANDIDATE attribute, the firewall can determine that the ICE agent is the controlled agent using aggressive nomination. It then waits for the media traffic to flow before setting session time of the dynamic mapping equal to Lifetime field in FW-FLOWDATA Attribute.

This technique would ensure that dynamic mappings created for pairs in ICE checklist which are not nominated for sending media will be removed after a short duration of time.

DISCUSSION: If WebRTC implementations of RTP support multiplexing of multiple media sessions onto a single RTP session, FW-FLOWDATA attribute can be enhanced to carry a flag indicating the same so that firewall can immediately close the dynamic mapping created for other pairs in the ICE checklist once media starts flowing on one the candidate pairs. In case of multi-homing firewalls can track multiple host IP addresses using authentication supplicant or, for hosts lacking the supplicant, use address-based authentication method.

## [6.](#) STUN Attribute Format

Internet-Draft

STUN extn for FW Traversal

July 2012

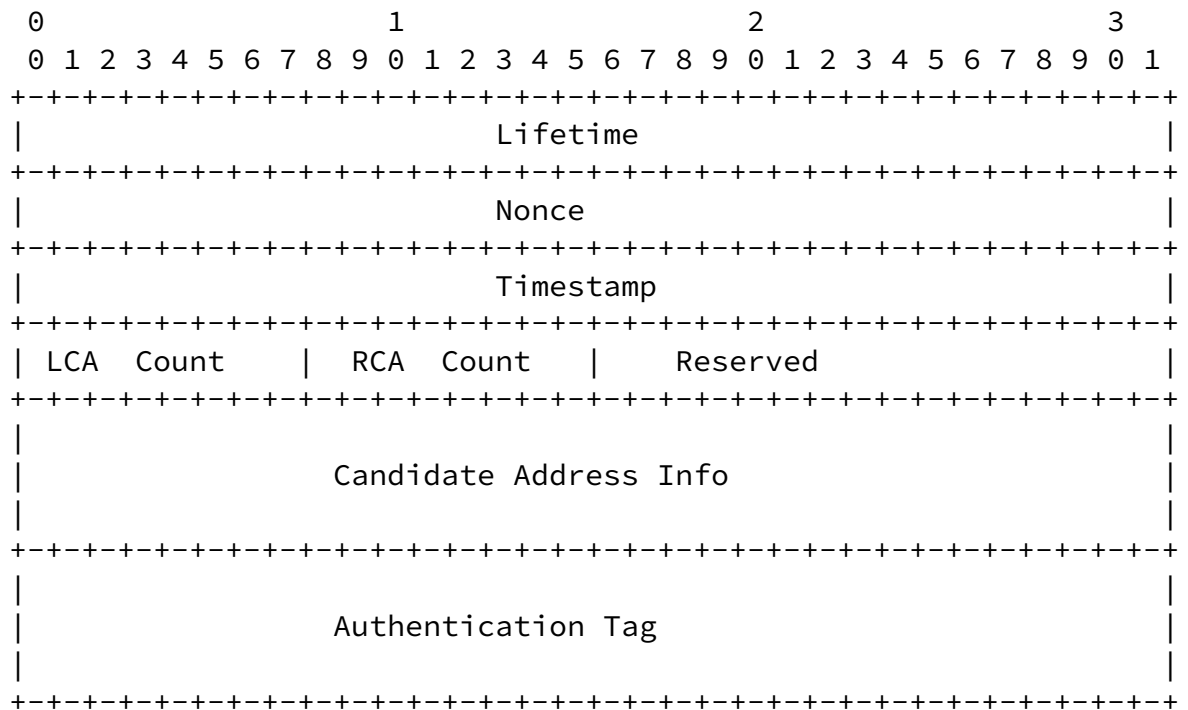


Figure 2: FW-FLOWDATA Attribute

**Lifetime:** 32-bit unsigned integer. The length of time in seconds that the STUN attribute is valid for the purpose of firewall creating dynamic mapping. The lifetime of the firewall dynamic mapping is set to this value. After the lifetime expires the mapping is deleted, unless the lifetime is extended using a another FW-FLOWDATA attribute.

**Nonce:** 96-bit unsigned integer. Random value chosen by the WebRTC Server that uniquely identifies the STUN attribute.

**Timestamp:** 64-bit unsigned integer field containing a timestamp. The value indicates the number of seconds since January 1, 1970, 00:00 UTC, by using a fixed point format. In this format, the integer number of seconds is contained in the first 48 bits of the field, and the remaining 16 bits indicate the number of 1/64K fractions of a second.



Alternatively using Dynamic Group Key Distribution, group keys are dynamically distributed among the WebRTC server and enterprise firewalls using GDOI [[RFC6407](#)]. In this way, each firewall requests a group key from a key server as part of an encrypted and integrity-protected key agreement protocol. Once the key server has authenticated and authorized the firewalls, it distributes a group key to the group member. The authentication in this model can be based on public key mechanisms, thereby avoiding the need for static key provisioning.

## [8.](#) Security Considerations

Hosts using WebRTC calls will see lot of FW-FLOWDATA attributes. They determine the key by trying a number of candidate keys and

Reddy, et al.

Expires January 9, 2013

[Page 11]

---

Internet-Draft

STUN extn for FW Traversal

July 2012

seeing if one of them is correct. The attack works when the keys have low entropy, such as a word from the dictionary. This attack can be mitigated by using strong keys with large entropy. In situations where even stronger mitigation is required, the keys can be dynamically changed using GDOI. The WebRTC server controls how long a firewall session is kept open via the Lifetime value and WebRTC server could use different Lifetime values depending on the anticipated level of trust of the device (e.g. company provided laptop might be trusted more than a Bring Your Own Device (BYOD)); the device with more trust need to obtain its authentication attribute less often). Firewalls in addition to timestamp checking can also maintain a cache of used Nonces, IP source addresses as an effective countermeasure against replay attacks.

All the security considerations applicable to STUN [[RFC5389](#)] and ICE [[RFC5245](#)] are applicable to this document as well.

## [9.](#) IANA Considerations

Allocate new STUN attribute value for FW-FLOWDATA from the [[STUN-ATTR](#)] registry.

## [10.](#) Acknowledgements

The authors would like to thank Prashanth Patil and Ramesh Nethi for review comments.

## 11. References

### 11.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC3971] Arkko, J., Kempf, J., Zill, B., and P. Nikander, "SEcure Neighbor Discovery (SEND)", [RFC 3971](#), March 2005.
- [RFC5245] Rosenberg, J., "Interactive Connectivity Establishment (ICE): A Protocol for Network Address Translator (NAT) Traversal for Offer/Answer Protocols", [RFC 5245](#), April 2010.
- [RFC5389] Rosenberg, J., Mahy, R., Matthews, P., and D. Wing, "Session Traversal Utilities for NAT (STUN)", [RFC 5389](#), October 2008.

Reddy, et al.	Expires January 9, 2013	[Page 12]
---------------	-------------------------	-----------

---

Internet-Draft	STUN extn for FW Traversal	July 2012
----------------	----------------------------	-----------

- [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](#), April 2010.
- [RFC6407] Weis, B., Rowles, S., and T. Hardjono, "The Group Domain of Interpretation", [RFC 6407](#), October 2011.

### 11.2. Informative References

- [I-D.ietf-rtcweb-use-cases-and-requirements]  
Holmberg, C., Hakansson, S., and G. Eriksson, "Web Real-Time Communication Use-cases and Requirements", [draft-ietf-rtcweb-use-cases-and-requirements-09](#) (work in progress), June 2012.
- [STUN-ATTR]  
IANA, "IANA: STUN Attributes", December 2011, <<http://www.iana.org/assignments/stun-parameters/>>

## Authors' Addresses

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

Email: [tiredy@cisco.com](mailto:tiredy@cisco.com)

Muthu Arul Mozhi Perumal  
Cisco Systems, Inc.  
Cessna Business Park, Varthur Hobli  
Sarjapur Marathalli Outer Ring Road  
Bangalore, Karnataka 560103  
India

Email: [mperumal@cisco.com](mailto:mperumal@cisco.com)

Reddy, et al.	Expires January 9, 2013	[Page 13]
---------------	-------------------------	-----------

---

Internet-Draft	STUN extn for FW Traversal	July 2012
----------------	----------------------------	-----------

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

Email: [rmohanr@cisco.com](mailto:rmohanr@cisco.com)

Dan Wing

Cisco Systems, Inc.  
170 West Tasman Drive  
San Jose, California 95134  
USA

Email: [dwing@cisco.com](mailto:dwing@cisco.com)