Use of DNS SRV and NAPTR Records for SPEERMINT
draft-ietf-speermint-srv-naptr-use-02.txt

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 May 11, 2008.

Abstract

   The objective of this document is to specify the Best Current
   Practice (BCP) adopted by a service provider providing multimedia
   communication services such as Voice over Internet Protocol(VoIP) in
   order to locate another service provider to peer with in the context
   of Session PEERing for Multimedia INTerconnect.

Table of Contents

1. Introduction

   A service provider needs to identify the ingress Session Initiation
   Protocol (SIP) (RFC 3261 [1]) server of a peering network before it
   can signal and route SIP based real-time communication sessions.
   This function of locating the ingress SIP server of peering network
   is typically performed by the egress SIP server of the service
   provider originating the SIP session. Also, the ingress server in
   the peering network needs to locate the originating service
   provider s egress server in situations where the peering connection
   to it gets terminated after receiving the SIP requests or if the
   egress SIP server of originating service provider fails. The SIP
   servers at originating and peering side use the DNS procedures, using
   both SRV [2] and NAPTR [3] records, in order to locate each other.

2. Terminology

   In this document, the key words "MUST", "MUST NOT", "REQUIRED",
   "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY",
   and "OPTIONAL" are to be interpreted as described in RFC 2119 [2] and
   indicate requirement levels for compliant implementations.

3. Session Peering Setup

   SIP systems are represented by user agents (UA). The diagram below
   shows the case of direct peering where a user agent (UA1), hosted by

a service provider SP1, initiates a SIP session to a User Agent
(UA2), hosted by service provider SP2. The egress SIP server of SP1
is a SIP signaling path border element (SBE) as defined in section 3
of[6], called SBE1, that interfaces with session peering service
provider SP2. The SIP session initiated by UA1 is received by this
network element SBE1. The resource to which the SIP request needs to
be routed by SBE1 is identified by a SIP or SIPS URI. This could be
the SIP URI of UA2 found in the Request-URI of the SIP request
received by SBE1, or the next hop from SBE1 found in the topmost
Route header of SIP request. In order to determine the resource to
route the request to, SP1 MAY make use of ENUM [4] lookup services or
an internal lookup to determine the SIP URI of the resource. This
lookup MAY be performed by SBE1 or another network element of SP1.

```
............................        ............................
.                  +------+ .        . +------+                  .
.                  |      | .        . |      |                  .
.                  | SBE 1|--------------| SBE 2|                  .
.                 / |      | .        . |      | \                .
.                /  +------+ .        . +------+  \               .
.   +------+   /      ||     .        .    ||      \  +------+     .
.   |      | /        ||     .        .    ||       \ |      |     .
.   | UA 1 |          ||     .        .    ||         | UA 2 |     .
.   |      |          ||     .        .    ||         |      |     .
.   +------+          ||     .        .    ||         +------+     .
.                  +-------+ .        . +-------+                  .
.                  |       | .        . |       |                  .
.                  | DNS 1 | .        . | DNS 2 |                  .
.                  |       | .        . |       |                  .
.                  +-------+ .        . +-------+                  .
.                           .        .                           .
.                 SP 1      .        .      SP 2                  .
............................        ............................
```
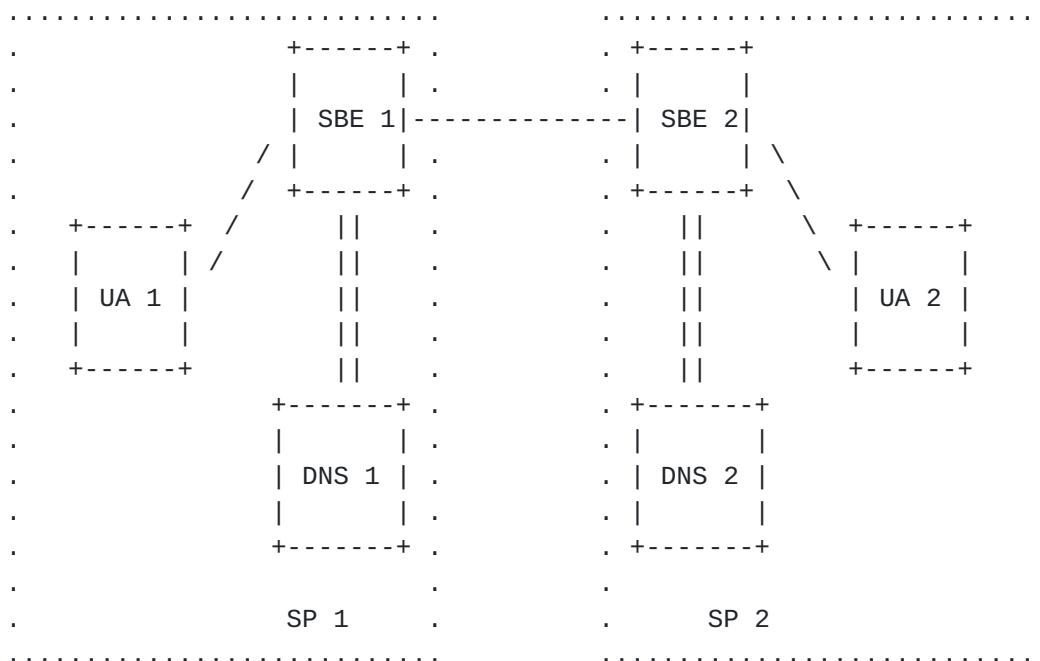
Figure 1: Logical Peering Scenario (direct peering)

In order to route the SIP request to this resource in SP2, SBE1 needs
to determine the ingress SIP signaling path border element for SP2,
called SBE2, by resolving the SIP or SIPS URI in DNS. SBE1 makes use
of the NAPTR and DNS SRV mechanism defined in [5] to determine the IP
address, port, and transport protocol for peering with the SP2
ingress SIP proxy server (i.e. SBE2). SBE1 and SBE2 which are
involved in the session peering, support a set of protocols and have
list of preferences for these protocols. UDP, TCP and TLS MUST be
supported by these proxies.

As a best current practice, SBE1 and SBE2 SHOULD be deployed in a
highly scalable and highly available manner, such as a cluster of

servers. These servers are of different prioritization and weight, to
ensure capacity-based load balancing.

The figure below shows the case of indirect/transit peering where
SBE2 is the ingress SIP server of a transit service provider. The
mechanism to locate SBE2 is the same as described for direct peering
scenario.

```
.............................     .............       .              .
.                   +------+ .     . +------+    .     .              .
.                   |      | .     . |      |    .     .              .
.                   | SBE 1|---------| SBE 2|    .     .              .
.               /  |      | .     . |      |    .     .              .
.              /   +------+ .     . +------+    .     .              .
.   +------+  /        ||    .     .     ||      .     . +------+     .
.   |      | /         ||    .     .     ||      .     . |      |     .
.   | UA 1 |           ||    .     .     ||      .     . | UA 2 |     .
.   |      |           ||    .     .     ||      .     . |      |     .
.   +------+           ||    .     .     ||      .     . +------+     .
.                  +-------+ .     . +-------+   .     .              .
.                  |       | .     . |       |   .     .              .
.                  | DNS 1 | .     . | DNS 2 |   .     .              .
.                  |       | .     . |       |   .     .              .
.                  +-------+ .     . +-------+   .     .              .
.                           .     .             .     .              .
.                  SP 1     .     . Transit SP .     .    SP 2      .
.............................     .............       ...............
```
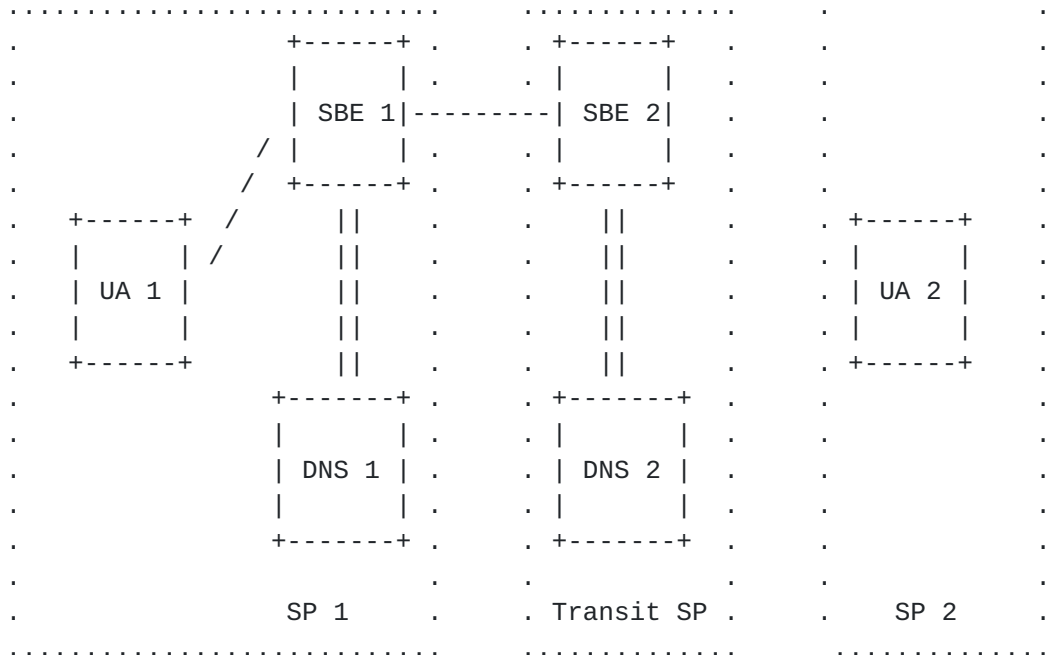
                Figure 1: Logical Peering Scenario (indirect peering)

   The figure below shows a high level SIP call flow setting up a direct
   SIP peering session between SP1 and SP2. In this call flow a VoIP
   session is eastablished between a caller, Bob (sip:bob@sp.com), in
   SP1 and callee, Alice(sip:alice@sp2.com), in SP2 using SIP INVITE
   request. All SIP signaling MUST go through the SBE1 and SBE2 as these
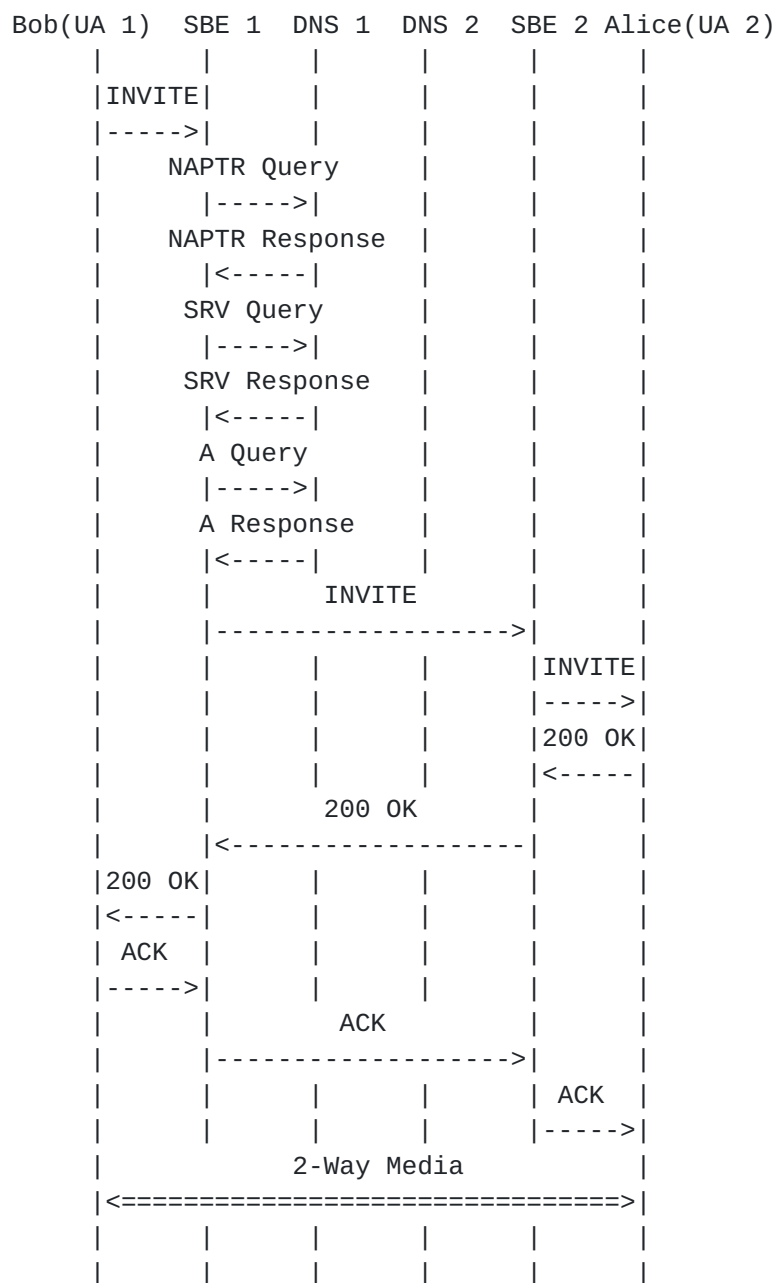   are the ingress and egress points in SP1 and SP2 network.

```
              Bob(UA 1)  SBE 1  DNS 1  DNS 2  SBE 2 Alice(UA 2)
                  |       |       |       |       |       |
                  |INVITE |       |       |       |       |
                  |----->|        |       |       |       |
                  |     NAPTR Query |      |       |       |
                  |       |----->|        |       |       |
                  |     NAPTR Response |   |       |       |
                  |       |<-----|        |       |       |
                  |      SRV Query  |      |       |       |
                  |       |----->|        |       |       |
                  |      SRV Response |    |       |       |
                  |       |<-----|        |       |       |
                  |       A Query   |      |       |       |
                  |       |----->|        |       |       |
                  |       A Response |     |       |       |
                  |       |<-----|        |       |       |
                  |       |        INVITE  |       |       |
                  |       |------------------->|   |       |
                  |       |       |       |       |INVITE |
                  |       |       |       |       |----->|
                  |       |       |       |       |200 OK|
                  |       |       |       |       |<-----|
                  |       |        200 OK  |       |       |
                  |       |<-------------------|   |       |
                  |200 OK |       |       |       |       |
                  |<-----|        |       |       |       |
                  | ACK  |        |       |       |       |
                  |----->|        |       |       |       |
                  |       |         ACK    |       |       |
                  |       |------------------->|   |       |
                  |       |       |       |       | ACK  |
                  |       |       |       |       |----->|
                  |        2-Way Media              |       |
                  |<===============================>|
                  |       |       |       |       |       |
                  |       |       |       |       |       |
```

                      Figure 2: Example Call Flow

   The target, to which the request is sent, is determined by SBE1 as
   follows:

3.1 TARGET Determination

   The target resource is identified with a SIP or SIPS URI. This is
   the URI in the Route header, if present, or the URI from the request
   URI of the SIP request received by SBE1.
   The host value of the hostport component of the URI is the TARGET.
   This TARGET is the domain to be contacted. The NAPTR/SRV/A lookup as
   described in the following section should be skipped if
   transport/port/ip address is already specified for the target URI.

3.2 NAPTR Lookup

   Next the SBE1 determines the transport protocol of the TARGET by
   performing a NAPTR query for the TARGET. NAPTR processing as
   described in [3] will result in the discovery of the most preferred
   transport protocol of a server instance of SBE2 and SRV records.

   Considering our example call flow setup above, SBE1 wishes to resolve
   sip:alice@sp2.com and performs a NAPTR query for that TARGET domain
   sp2.com, and the following NAPTR records are returned:

```
;           order pref flags service      regexp  replacement
   IN NAPTR 50   50  "s"  "SIPS+D2T"    ""    _sips._tcp.sp2.com
   IN NAPTR 90   50  "s"  "SIP+D2T"     ""    _sip._tcp.sp2.com
   IN NAPTR 100  50  "s"  "SIP+D2U"     ""    _sip._udp.sp2.com
```

   DNS MUST return at least three records - one with "SIP+D2T", one with
   "SIP+D2U" and one with "SIPS+D2T" service type for the case of direct
   and indirect peering (section 4.3 in [6]). For indirect (transit)
   peering (section 4.4 in [6]) since domain validation as specified in
   section 26.3.2.2 of [1] for TLS at layer 5 will not work, SIPS over
   TLS cannot be used.

3.3 SRV Lookup

   Depending on what transport protocols SBE1 supports, SBE1 selects one
   from the preference list of NAPTR results and performs the SRV lookup
   to obtain a list of available server instances for SBE2. TLS SHOULD
   be the preferred transport protocol for peering between SBE1 and
   SBE2.

   In our example SBE1 uses TCP, the SRV lookup for _sip._tcp.sp2.com
   would return list of available servers :

      ;;   Priority Weight Port   Target
          IN SRV 0   1   5060   server1.sp2.com
          IN SRV 0   2   5060   server2.sp2.com

   Alternatively, if no NAPTR records are found, then SBE1 uses the
   preferred transport protocol and issues an SRV query for that
   specific transport using "sips" for SIPS URI and SIP URI with TLS and
   "sip" for SIP URI as the SRV domain prefix.

   In our example, SBE1 prefers to use TCP and target SIP URI of SP2 is
   sip:alice@sp2.com, it sends a SRV query for _sip._tcp.sp2.com.

   The SRV responses MAY also include A records with it.

3.4 Using SRV Results

   If A records are not returned with the SRV responses, procedure from
   RFC 2782 describes how to use and interpret the results obtained from
   the SRV query. The target entry of the SRV RRs is looked up by
   querying the DNS for address records. If the SRV response from DNS
   includes A records with it, it will cut down on round trips and
   lookup of DNS again for target entry. On determining the transport
   protocol, service, port and address record from the SRV RRs as
   described above, the SBE1 will try to connect to the (protocol,
   address, service). Once the connection is established to an
   available instance of SBE2, SBE1 sends the SIP request to SBE2. SBE1
   MUST act in a stateful manner and any retransmission of SIP requests
   for a specific SIP transaction, including ACKS for non-2xx response
   or CANCEL for that SIP transaction MUST go to the same server
   instance of SBE2.

   When SBE1 sends the SIP request to SBE2, it SHOULD set the sent-by
   parameter of the topmost Via header in the SIP request to a domain
   that identifies SBE1. It MUST NOT specify the port.

4. High Availability

   High Availability is ensured by detecting failures in the ability to
   connect to SBE1 and SBE2 server instances. In the event of a
   failure, when SBE1 tries to send SIP INVITE to SBE2, the following
   failures could occur:

4.1 SBE1 Fails to Reach SBE2

   A 503 error response is reported by the transaction layer, or failure
   can occur at the transport layer due to TCP disconnect in connection,

ICMP error in UDP or time out at transport layer or SIP layer timeout
when its not receiving any SIP response. In such situations, SBE1
tries a new SIP request transaction to the next available server
instance of SBE2 as determined by SRV RRs entry. The SIP T1 timer on
SBE1 SHOULD be configurable with a upper limit value of 500ms. A
shorter value of T1, say 100ms, reflects a faser failover support.

4.2 SBE2 Fails to Reach SBE1

Failure may also occur after the request is received by SBE2 from
SBE1 due to closure of the transport connection the request came in
on at SBE2, before the response can be sent back to SBE1. In this
situation, SBE2 uses the domain value present in the 'sent-by'
parameter in the top most Via header of the received SIP INVITE, and
queries for SRV records at this domain name using the service
identifier "_sips" if the Via transport is "TLS", "_sip" otherwise.
The sorted list of SRV RRs are obtained and used as described in [2]
to send the response back to SBE1. If the topmost element in the
list of server instances of SBE1 fails, the next available one is
tried.

[FOR NEXT REV: SHOULD WE ADD CALL FLOW FOR FAILURE SCENARIO DESCRIBED
IN 4.1 AND 4.2]

5. Caching/TTL

5.1 Caching
SBE SHOULD use caching of DNS results to eliminate unnecessary DNS
queries.

5.2 TTL

SRV RRs have a TTL value based on which the SBE1 caches the entry
for that duration, if it supports caching, and any further requests
to the same TARGET domain are delivered to the cached server
instance. The TTL recommended for SRV is about 1 hr. The TTL for
NAPTR is much higher, about 1 day (24hrs) since the NAPTR records do
not vary that often as compared to SRV.

6. Acknowledgements

Special thanks go to Yiu Lee for his valuable input to this document.

7. Security Considerations

This document introduces no new security considerations.

8. IANA Considerations

This document creates no new requirements on IANA namespaces
[RFC2434].

9. References

9.1 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]   Gulbrandsen, A., Vixie, P. and L. Esibov, "A DNS RR for
       Specifying the Location of Services (DNS SRV)", RFC 2782,
       February 2000.

[3]   Mealling, M. and R. Daniel, "The Naming Authority Pointer
       (NAPTR) DNS Resource Record", RFC 2915, September 2000.

[4]   Faltstrom, P. and M. Mealling, "The E.164 to Uniform
       Resource Identifiers (URI) Dynamic Delegation Discovery
       System (DDDS) Application (ENUM)", RFC 3761, April2004.

[5]   Rosenberg, J., Schulzrinne, H., "Session Initiation Protocol
       (SIP): Locating SIP Servers", RFC 3263, June 2002.

[6]   Meyer, D., "SPEERMINT Terminology", draft-ietf-speermint-
       terminology-12, August 2007.

9.2 Informative References

[RFC2434] Narten, T. and H. Alvestrand, "Guidelines for Writing an
           IANA Considerations Section in RFCs", BCP 26, RFC 2434,
           October 1998.

Authors' Addresses

   Tom Creighton
   Comcast Cable Communications
   1500 Market Street
   Philadelphia, PA 19102
   USA

   Phone: +1-215-320-8617
   Email: tom_creighton@cable.comcast.com

   Jason Livingood
   Comcast Cable Communications

     1500 Market Street
     Philadelphia, PA 19102
     USA

     Phone: +1-215-981-7813
     Email: jason_livingood@cable.comcast.com

Intellectual Property and Copyright Statements

     Full Copyright Statement

     Copyright (C) The IETF Trust (2007).

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

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

     Intellectual Property

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

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

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

Acknowledgment