IETF 96, 20 July 2016

# Supporting SIP in a Dual-Stack Network

## Happy Eyeballs for SIP

Dale R. Worley

# With the help of

- Olle E. Johansson
- Gonzalo Salgueiro
- Roman Shpount

# Problems to solve

- The core SIP RFCs are written for both IPv4 and IPv6
- Dual-stack deployments aren’t cleanly handled
- RFC 6157 handles IPv6 transition, but doesn’t solve Happy Eyeballs. In fact, it states it as a problem.
- As discussed extensively on the list, RFC 3263 is not clear on DNS procedures in dual-stack networks
- Happy Eyeballs-like problem in SIP due to long transaction timeouts when there is no response

# draft-ietf-sipcore-dns-dual-stack

- Addresses the narrow problems of DNS SRV record look-ups of SIP servers in dual-stack environments
- Addresses two issues by:
  - Requiring lookup of both A and AAAA records by dual-stack devices
  - Documenting that as a consequence, DNS SRV records can indicate server preference of address family
- Now in IESG

# The next increment of work

- Allow devices to change the target order prescribed by RFC 3263/2782
- Encourage UAs to maintain flows or probe targets before sending requests
- Reduce client transaction timeouts, which default to 32 sec
- Include current UDP approaches for large-scale dual-stack deployments

A preliminary version is available as draft-worley-sipcore-dual-stack.

# Changing target order

- Strictly, allowed by RFC 3261 section 8.1.2
- Detailed guidance for reordering targets
- May simultaneously initiate *flows* with multiple targets, but only one outstanding *request* at a time
- If targets are accessible for a long time, behavior MUST approximate 3263.  In particular, cached "non-reachable" status must time out.

# Flows, probes, and keepalives

- UAs should maintain current flow status with the targets of its home proxy
  - Many different methods
  - SIP Outbound (RFC 5626) is the best known
- Alternative is probing targets before sending request

# Reduce transaction timeouts

- Timer B and Timer F are currently 64*T1, which default to 32 sec
- Reducing T1 is probably not a good idea, because it affects many timers
- Reducing B/F without reducing T1 lowers the number of retransmissions
- Shorter timeouts are more vulnerable to intermittent connectivity

# Large-scale dual-stack deployment 1

- Dual-stack must support current UDP approaches
- SIP Outbound (RFC 5626) is understood as the "correct" technique for many of these problems, but is not widely deployed
  - Need to understand why
  - Perception that overhead is excessive
  - Should we define a subset/revision of Outbound?

# Large-scale dual-stack deployment 2

- Need low-overhead, low-state signaling encryption
  - Protect signaling from ALGs
    - ALGs cause many problems in practice
  - Possibly a subset or variant of DTLS

# Recovery in unstable networks

- Instability
  - Changing connectivity (IP address and family)
  - Intermittent network connectivity
- Failure detection
  - Signaling keep-alives
  - Absence of media (RTCP, actually)
  - Network layer status
- Call recovery
- Transaction recovery (especially INVITE)
- Interaction with session timers

# Work plan (very preliminary) 1

- 2017 Jan
  - Target reordering/flow probing/reduced timeouts
- 2017 Jul
  - Outbound Lite
  - low-overhead NAT & keepalive for large-scale deployments

# Work plan (very preliminary) 2

- 2018 Jan
  - DTLS Lite
  - low-overhead UDP encryption for large-scale deployments
  - Is this a Security topic?
- 2019 Jan
  - Call/transaction recovery

# Questions?