High performance. Delivered.

SIP

accenture

Agenda

- Scenario
- Authentication
- Services
  - VPN
  - CLIR
  - Emergency Call
- Scalability
- Reliability
- Rerouting
- NAT
- General Compatibility

accenture

Some Scenarios

- SIP proxy with direct SIP clients
  - 2 Million subscribers
- SIP proxy with PSTN/VoIP gateways
  - 1500 small ISDN gateways
  - 20 large ISUP gateways
- Services
  - Replace VPN service
  - Implement mandatory CLIR
  - Implement mandatory Emergency Service
- NAT

accenture

Agenda

- Scenario
- Authentication
- Services
  - VPN
  - CLIR
  - Emergency Call
- Scalability
- Reliability
- Rerouting
- NAT
- General Compatibility

accenture

Authentication

- 2 paradigms
  - Direct subscriber authentication using REGISTER
  - Per INVITE screening when UA don’t register
    - For instance if it is a PSTN gateway
  - How to support both at same time?
- Should I integrate with existing subscribers DB?
  - Where will the subscriber password be?
  - In Wireless there is always a HLR…

Agenda

- Scenario
- Authentication
- Services
  - VPN
  - CLIR
  - Emergency Call
- Scalability
- Reliability
- Rerouting
- NAT
- General Compatibility

Services

- VPN
  - INVITE (telephony address) routing dependent on the origin node
    - Just like our friend RFC 2547
      - Should keep one RIB per VPN
    - Could use domains as VPN tags
  - TON preservation from ISDN to ISUP not possible in SIP
    - Could use number prefixes to signal TON to the gateways

Services

- CLIR
  - Legal requirement
  - Privacy header is the way to go… but
    - RFC3323 is the standard
    - …and Cisco (main VoIP gw around) implements draft-ietf-sip-privacy02 (deprecated draft)
  - When services interact (for instance CLIR and Emergency) there are other RFCs to follow
    - For instance RFC 3325…

Services

- Emergency Call (a.k.a. 911)
  - This call should be handled to the network with high priority
    - Never drop them under load
    - CLID should be present regardless and the user can't mask it (RFC 3324 and 3325)

- Emergency Call
  - However, where should I send the call?
    - SIP server in Madrid, ISUP gateways at Madrid, Barcelona, Valencia, Bilbao
    - I should send the call to the gateway closest to the subscriber…
      - Where is the user? I only have an IP address…
  - PSAP (Public Safety Answering Point) should be able to check call originating location

accenture

Agenda

- Scenario
- Authentication
- Services
  - VPN
  - CLIR
  - Emergency Call
- Scalability
- Reliability
- Rerouting
- NAT
- General Compatibility

Scalability
Let's use it for real

- 1 voice call establishment = 1 INVITE transaction = statefully takes at least 8 SIP packets
- Typical requirements are 100K BHCA
  - BH = Busy Hour
  - BHCA = BH Call Attempts

Scalability

- 100K BHCA
  - ~30 voice calls/sec
    - ~300 pkt/sec
    - You are not thinking in going into a DB during each INVITE transaction, right?
      - Therefore we need an in memory object to hold all subscribers
      - 2 Million subs, 400 bytes/subs = 800 MB memory object

Agenda

- Scenario
- Authentication
- Services
  - VPN
  - CLIR
  - Emergency Call
- Scalability
- Reliability
- Rerouting
- NAT
- General Compatibility

accenture

Reliability

aka Are the end point still there?

- How to know that an end point is still valid?
  – We don’t want to wait 30 secs to discover that we can’t complete the call (the user will not wait…)
- What about:
  – Ping using ICMP the gateways from the management console
    - IP stack is there, is the SIP stack ok as well?

Reliability

- What about:
  - Ping using SIP OPTIONS
    - Ok, SIP stack is there
    - But, 1000 gateways with 15 sec pooling interval = 200 pkt/sec for pooling
    - 15 sec! This pooling interval is too short!
      - But INVITE transaction timeout in 30 secs…

Reliability

- What about:
  - Use TCP as transport (persistent connections) and don’t send INVITEs to disconnected gateways
    - TCP support is mandatory by RFC 3261
      - Few people use it
      - There are different implementations out there
      - Mostly don’t work very well
    - RFC recommends TCP to be closed after ~30 secs of inactivity
      - We want persistent connections for this case

accenture

Reliability

- What about:
  - Use TGREP
    - Part of TRIP (a.k.a. TRIP-GW)
      - BGP like protocol for exchange of telephony route information
    - Excellent idea, but operators are afraid of it
      - What, dynamic telephony routing???
    - Engineers always want the leading edge
      - But the boss say, How much it will cost? How long it will take? Are we the first to do this?

Reliability

… what if the SIP server fails?

- Cluster required
- Availability x Fault Tolerance
- Redundancy and Load balancing
  - Manual load balance – distributing clients
    - Puts the burden on the UA configuration
    - Error-prone
  - Multiple DNS SRV records
  - Virtual IP
    - Allow IP “roaming”

accenture

Reliability

- Do I synchronize the servers within the cluster?
  - Sync location information, of course
    - Even for non-registered end points?
  - Active session information for stateful proxies?
- If I sync, how to do that?
  - Do persistency to a DB? Slow
  - Do persistency to a shared file system (Veritas)? Better
  - Do in-memory tables replication resending protocol messages? Hmmm…
    - Such as SER t_replicate() or FreeRadius rad_relay
  - Do in-memory tables replication using reliable multicast? This sound good. Is it reliable?

accenture

Agenda

- Scenario
- Authentication
- Services
  - VPN
  - CLIR
  - Emergency Call
- Scalability
- Reliability
- Rerouting
- NAT
- General Compatibility

accenture

Rerouting

- Let’s say you want backup routes to customers/PSTN links
  - However, when do you reroute from the primary end point to the backup?
    - RFC 3261 says that INVITE transaction fails after ~30 secs
  - Are you going to wait >30 secs for a call to complete?

Rerouting

- Let’s reroute within an INVITE transaction
  - For instance, after the first INVITE packet timeout (3.5 secs)
    - Still large… but ok, it is better
- What’s happen when I receive a…
  - 603, 486 – no reroute, right?
  - 5xx – reroute, right?

accenture

Agenda

- Scenario
- Authentication
- Services
  - VPN
  - CLIR
  - Emergency Call
- Scalability
- Reliability
- Rerouting
- NAT
- General Compatibility

accenture

NAT

Let's don't use NAT, right?

- As usual NAT adds a mess to the Internet
- Run SIP over TCP
  - Should keep TCP connection open and ensure incoming transactions will reuse connection
  - If not should provide some keep alive packet to keep NAT bidings
    - Short REGISTER messages?
    - Proprietary solutions? (Nortel SIP PING message)
- What about the VoIP RTP sessions?
  - UDP, dynamically created ports

accenture

NAT

- Several solutions (as usual):
  - RTP proxy
    - Managed by the SIP proxy at each NAT demarcation
    - Resource intensive, all the calls goes through here (50 packet/sec per call with 20ms sampling)
  - STUN (RFC3689)
    - The better idea so far, but…
      …mostly don’t work with internal to internal calls[1]
  - Many others…
    - In fact there is a effort in MMUSIC to define a single method for NAT transversal (ICE)

1 - draft-jennings-midcom-stun-results-00.html

Agenda

- Scenario
- Authentication
- Services
  - VPN
  - CLIR
  - Emergency Call
- Scalability
- Reliability
- Rerouting
- NAT
- General Compatibility

accenture

General Compatibility

aka Annoying compatibility…

- H.323 Legacy devices
  - Mostly of the H.323 to SIP is focused on voice calls
    - Video calls don’t work because H.323 conveys much details not supported on SIP
    - Some H.323 devices have specific codecs (H.261, for instance) not supported in some SIP clients

accenture

General Compatibility

- SIP implementations
  - Let’s say you want to do 3rd Party Call Control and redirect calls (draft-ietf-sipping-3pcc-06) there are at least 4 basic ways (each one with small variations)
  - Proxy does one way and UA does another way… You know where I am going…

General Compatibilty

- QoS
  - You feel mighty… Let's deploy QoS
    - First forget IntServ
  - So we have:
    - DiffServ
    - 802.1p
    - FR/ATM/MPLS circuits

General Compatibility

- QoS
  - How to map one to another?
    - How to put everything together?
    - Let's say the IP phone support 802.1p and DiffServ, Ethernet switch support 802.1p, router support DiffServ and multiple vc to ATM switch
  - You are not thinking on per call, dynamically, apply the policies using COPS right?
    - … you may or may not, but mostly of the operators are.

General Compatibility

- Let’s say that you need an IVR…
  - How to receive DTMF tones?
    - Inside RTP flow as sounds – fine for G.711
      - A problem for 3.4 Kbps codec…
    - Inside RTP but decoded – RFC 2833
      - Much better
    - Outside the RTP (OOB) in the SIP messages
      - Using subscribe/notify or messages…
      - Now we could use the DTMF to provide mid-call events from the SIP proxies…
  - Did I say that the SIP UAs should agree in the method?

accenture

Questions?

?

accenture

Don’t forget to:

- Read RFC 3261, it is worth the effort
- Use SIP
  - Install SER from www.iptel.org
    - a good idea for a server on your site
    - you may need STUN
  - Install a SIP client
    - Windows Messenger (not MSN Messenger) already supports SIP
    - Some really good Linux SIP clients are out there (KPhone)
- Discuss with your peers (using VoIP, of course)
- Call me or email:
  - <sip: manta at iptel dot org>
  - <emailto: manta at ieee dot org>