75th IETF, Stockholm, Sweden
July 26-31, 2009

Session Initiation Protocol (SIP) Common Log Format (CLF)

Vijay K. Gurbani <vkg@bell-labs.com> Bell Laboratories/Alcatel-Lucent

Contributors

Humberto Abdelnur <Humberto.Abdelnur@loria.fr>
Tricha Anjali <tricha@ece.iit.edu>
Eric Burger <eburger@standardstrack.com>
Oliver Festor <Olivier.Festor@loria.fr>
Vijay K. Gurbani <vkg@bell-labs.com>
Hadriel Kaplan <hkaplan@acmepacket.com>
Adam Roach <adam@nostrum.com>
Theo Zourzouvillys <theo@voip.co.uk>

What is CLF

Common Log Format (CLF):
A summary of an application layer PDU*

* (To paraphrase from RjS)

What is CLF

Example HTTP CLF:

```
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] \
 "GET /apache_pb.gif HTTP/1.0" 200 2326
```

SIP CLF borrows a bit from Apache CLF and Squid CLF.

What is CLF

SIP CLF example (NOTE: just an example, not a normative representation)

```
<allOneLine>
     1230756560 192.168.1.10 - INVITE sip:bob@example.net
     sip:alice@example.com;tag=iu8u76 sip:bob@example.net
     i98ju@example.com "<sip:bob@home.example.net>“
    y6y78u -
 </allOneLine>
```

What SIP CLF is and is not ...

SIP CLF is NOT…

- … a replacement for a CDR (Call Detail Record).
- … a billing tool.
- … a QoS measurement tool.

SIP CLF IS:

- … a standardized format that can be used by all SIP entities.
- … an easily digestible log of past and current transactions.
- … a format that allows quick parsing to discover relation-ships between transactions

```
$ grep yuhyt6 sip-clf.txt
```

gets all transactions with this label.

- … amenable for easy parsing and creating other innovative tools.

Use cases

- Trend analysis (“I want to find out which geographical area are the most calls coming from at 2:00 AM”).
- Troubleshooting (“How long did it take to generate a final response to an INVITE?”)
- Message correlation across transactions (“Find all messages corresponding to Call-ID X, including all forked branches”)
- Transaction correlation across dialogs (“Find all messages for dialog created by Call-ID X and tags A and B”).
- Establish concise and standardized diagnostic trail of a SIP session locally and globally.
- Establish concise and standardized format for training automata (anomaly detection.)

Benefits of a SIP CLF

- Establishes a common reference for logging SIP messages across vendor/open-source implementations.
- Correlate SIP messages across transactions and dialogs.
- Easily search, merge, and summarize log records.
- Train anomaly detection systems to trigger alarms.
- Allow independent tool providers to provide innovative tools for trend analysis and traffic reports.
- Common diagnostic trail from testing of SIP equipment.
- Can be used for off-line analysis (trend analysis) as well as real-time analysis.

Challenges in defining SIP CLF

- SIP is not a *linear* request-reply protocol
  - HTTP is *linear*: pipelining okay, one request = one response.

- Complexity inherent in the protocol:
  - Serial and parallel forking elicit multiple responses.
  - Delays between getting a request and sending a response (outside of “long polling” in HTTP, servers respond quickly; not quite so in SIP. Impact on proxies.)
  - Multiple transactions grouped in a dialog; dialog persists for a long time, transactions short-lived (e.g., BYE comes much later, but relation between INVITE and BYE should be preserved in a log file.)

Challenges in defining SIP CLF

- ACK requests need careful considerations:
  - Only tied to an INVITE.
  - No responses for ACKs.
  - For non-2xx, ACKs hop-by-hop (part of INVITE transaction.)
  - For 2xx, ACK end-to-end.

- CANCEL requests need careful considerations:
  - Only tied to an INVITE.
  - Requires exactly one response.
  - Is propagated hop-by-hop.

Challenges in defining SIP CLF

- INVITE can pend, resulting in a 1xx response (200ms rule.) This 1xx response needs to be captured to train automata.

- SIP has a richer set of actors: UAS, UAC, B2BUA, proxy, registrar, redirect server, ...

- Need to take SIP extensibility in account.

- Preserve user privacy in CLF (through anonymization, etc.)

Progress so far

- SIPPING-sponsored BoF in IETF 74 (San Francisco.)
- Problem statement, motivation scenarios defined in http://tools.ietf.org/html/draft-gurbani-sipping-clf-01
- Mailing list formed (sip-clf@ietf.org):
  - https://www.ietf.org/mailman/listinfo/sip-clf
- Initial discussions on dispatch lead to proposal of chartering a working group; charter sent out by RAI AD (see http://www.ietf.org/mail-archive/web/sip-clf/current/msg00019.html)
- Much discussion has taken place on sip-clf mailing list.

Progress so far

- An ASCII mapping defined in

http://tools.ietf.org/html/draft-gurbani-sipping-clf-01

- A binary mapping defined in

http://tools.ietf.org/html/draft-roach-sipping-clf-syntax-01

- A PCAP-compatible binary syntax defined in

http://tools.ietf.org/html/draft-kaplan-sipping-clf-pcap-00

WG-to-be charter

- In scope:
  - WG to produce CLF suitable for logging at any SIP element, taking SIP's extensibility model into consideration.
  - WG not pre-constrained to producing either a bit-field oriented or text-oriented format, and may choose to provide both. If the group chooses to specify both, it must be possible to mechanically translate between the formats without loss of information.

WG-to-be charter

- Out of scope:
  - Specifying the mechanics of exchanging, transporting, and storing SIP Common Log Format records is explicitly out of scope.
  - Specifying a real-time transfer mechanism for heuristic analysis is explicitly out of scope.

WG-to-be charter

- Deliverables:
  - A problem statement enunciating the motivation, and use cases for a SIP Common Log Format. This analysis will identify the required minimal information that must appear in any record.
  - A specification of the SIP Common Log Format record.

Next steps

- Create WG (token: RAI AD).
- Socialize work with other IETF WGs:
  - opsarea
  - syslog
  - ipfix