Syslog

Robert Petkus
NSLS-II Controls Group
2010 Fall EPICS Collaboration Meeting
October 12, 2010
Outline

- Syslog
- Rsyslog
- Syslog-ng
- Splunk
- LogZilla
- Test Bed
Syslog

- is the standard logging solution on UNIX/LINUX systems and network routers/switches
- has evolved over time with several implementations => syslog, rsyslog, syslog-ng
- employs a layered architecture – separation of message content from transport
- reads and logs messages to log files, a console, and/or other systems
- supports output to named pipes (FIFOs) and remote logging (traditionally UDP/514)
- generates messages composed of (5) parts: Time Stamp, Program name, Facility, Priority, Log message
Syslog Configuration Rules

- **Facility**:
  - auth
  - authpriv
  - cron
  - daemon
  - ftp
  - kern
  - lpr
  - mail
  - mark
  - news
  - uucp
  - local7

- **Priority**:
  - debug
  - info
  - notice
  - warning
  - err
  - crit
  - alert

- **Selector**

- **Action (Log)**:
  - File
  - Database
  - Named Pipes
  - Terminal/Console
  - Remote Machine
  - User
  - Everyone
RSyslog & Syslog-ng

RSyslog improves upon syslog with

• native support to write logs to a database => MySQL, Postgres, OpenTDS, SQLLite, libdbi
• the ability to send email based on a trigger
• support for TCP (improved reliability over UDP) and RELP (improved reliability over TCP)
• Encryption (SSL/TLS)
• filters supporting regular expressions
• data compression (zlib) on the fly (send & receive)
• On-demand disk spooling for both scheduled log processing and data buffering
Syslog-ng

Syslog-ng competes with Rsyslog and offers

- direct database access (MSSQL, MySQL, Oracle, Postgres, SQLite3)
- high performance => 75k messages/s real time and >24GB raw logs/hour
- robust TCP / encryption
- advanced configurability => message sorting, parsing, rewriting, classification in real time
- human readable pattern matching (and regex)
- precision time-stamping => millisecond resolution
Log Analysis => Splunk

What is Splunk? A system administrator search engine

- Search and analyze data from servers, apps, network appliances indexed in real time
- Generate reports, audits, sign data
- Data sources can be logs, alerts, scripts, archive files, SNMP trap data, etc.
- Configure alerts to send emails/daily reports/SNMP messages and trigger scripts
- Ability to forward data from one/many Splunk instance(s) to another (forwarder – receiver)
  - Data centralization, load-balancing, data cloning, data routing, distributed search
  - (2) flavors: Regular (forwards raw or parsed data) & Light (raw or unparsed)
- Timestamp modification/manipulation; Train to recognize new Timestamp formats
- Creation of tags to cluster groups of hosts, fields, sourcetypes, etc.
- LDAP authentication
Splunk Indexing

- break events into searchable segments
- build index data structures
- write raw data & index files to disk

Indexing

SPLUNK INDEX PROCESSING

- RAW DATA

Parsing

Parsing Pipeline Data Chunks

- extract fields host, source, sourcetype
- configure character encoding
- identify line termination
- identify timestamps create if none
- mask data if configured

Post-Indexing Compression
Splunk Search

Example 1: Keep only search results that have the specified "src" or "dst" values.

```
src="10.9.165.*" OR dst="10.9.165.8"
```

Example 2: Search for events with either codes 10 or 29, and a host that isn't "localhost" and an xqpt that is greater than 5

```
(code=10 OR code=29) host!="localhost" xqpt>5
```

Example 3: Search for events with "404" and from host "webserver1"

```
404 host="webserver1"
```
Log Analysis => LogZilla, etc.

LogZilla

• Web front-end providing real-time access to syslog messages logged to MySQL

• Customized searches/report generation based on host, facility, priority, etc.

• Fast search via Sphinx => MySQL batch index and data search
  • 60+ MB/sec indexing performance

• Limited functionality compared to Splunk
LogZilla Web Interface
Prototype Environment at NSLS-II

In preparation of deploying server infrastructure at the production facility, we’ve

- Deployed a central log server (syslog-ng) collecting logs from all internal systems (~20)
  - (2) streams (to simultaneously run Splunk & LogZilla)
    - Stream A => TCP forked to both ASCII text and MySQL (LogZilla)
    - Stream B => TCP direct to Splunk DB
      - No performance bottlenecks (GbE, private net) but scale-out will require RAID array
  - Splunkd configured as a “collector”

- On client-side
  - Syslog-ng packages and configs pushed to clients via Puppet
  - Noisy logs (DHCP, Iptables, etc.) filtered-out locally but sent over wire to central log
  - Interesting clients with non-syslog app logs (NX, Virtualbox, conserver, Apache) run Splunk as a “light forwarder” to the Splunk collector on central log.
Resources

- Rsyslog (http://www.rsyslog.com/)
- Splunk (http://www.splunk.com)
- LogZilla (http://nms.gdd.net/index.php/LogZilla)
- Sphinx open-source SQL full-text search engine (http://sphinxsearch.com/)
Thanks

Questions – Comments ?