Log infrastructure & Zabbix
logging tools integration
About me

- **Me**
  - Linux System Architect @ ICTRA
  - from Belgium (...)
  - IT : Linux & SysAdmin work, Security,

- **ICTRA**
  - ICT for Rail
    - for Transport – Mobility – Security
  - 1800 IT Professionals – engineers - technicians
  - Facts :
    - 5.500 KM fibre optic
    - 3 main datacenters, a lot of 'technical' locations
    - 2.600 camera's in 51 major railway stations
    - ...

ICTRA, ICT for Rail

Ticketing solutions

Information systems

Train info in real time

ICT network

Management computer hardware NMBS Group

Fleet management system

GSM for Rail

Integrated security solutions

Monitoring of trains
Our zabbix installation

- Used by different teams
  - Linux team → use of automation (Puppet)
  - Solaris team → heavy use of scripts and API
  - Train announcement system team
- 1 master server in active-slave (Pacemaker)
- proxies
- MySQL master-slave cluster (different story...) with MasterHA

<table>
<thead>
<tr>
<th>Number of hosts (monitored/not monitored/tempates)</th>
<th>1665</th>
<th>1441 / 125 / 99</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of items (monitored/disabled/not supported)</td>
<td>195174</td>
<td>167027 / 22943 / 5204</td>
</tr>
<tr>
<td>Number of triggers (enabled/disabled)[problem/unknown/ok]</td>
<td>131430</td>
<td>131222 / 208 [496 / 0 / 130726]</td>
</tr>
<tr>
<td>Number of users (online)</td>
<td>106</td>
<td>8</td>
</tr>
<tr>
<td>Required server performance, new values per second</td>
<td>962.69</td>
<td>-</td>
</tr>
</tbody>
</table>
Why do we log?

- **Goal**
  - legal reasons
  - central storage
  - analysis
    - metrics
    - security (compliance)
  - anomaly and fault detection → monitoring

- **Requirements**
  - average number of events/second, peak load
  - resiliency against cracking attempts needed?
  - central / de-central ?
  - remote locations ?
  - search performance
Typical reasons (SANS)

- Detect/Prevent Unauthorized Access and insider Abuse
- **Meet Regulatory Requirement**
- Forensic Analysis and Correlation
- Ensure Regulatory Compliance
- Track Suspicious Behavior
- IT Troubleshooting and Network Operation
- Monitor User Activity
- **Best Practices/Frameworks such as COBIT, ISO, ITIL, etc.**
- **Deliver Reports to Departments**
- Measure Application Performance
- Achieve ROI or Cost Reduction in System Maintenance
Where does Zabbix fit in?

- **Zabbix**
  - perfect for monitoring resources
  - good in alerting
- **Zabbix is NOT**
  - aimed at analyzing a huge amount of log files
  - transformation of log files
  - storing log files
- Other tools are perfect for gathering, transformation and analysis
- And can use Zabbix for alerting when condition x,x,z happens
Although...

- due to popular demand, file content and logfile parsing has been added in **Zabbix 2.2**
- changes for
  - `vfs.file.regexp[ ]`
  - `vfs.file.regmatch[ ]`
  - `log[ ]` and `logrt[ ]`
- can now return a part of a string
- or a part – 'an interesting number' using regexp subgroups
- Read the zabbix blog post by Richard :-}
Parts available

- Zabbix
  - zabbix-sender
- Syslog
  - rsyslog
  - syslog-ng
- Search technology
  - ElasticSearch
  - Solr
  - Sphinx
- Storage (DB, ES, Hbase...)
- Security
  - ELSA
  - Snorby
  - OSSEC
- Queuing (amqp, key-value...)

And...

- Splunk ($)
- Logstash
  → zabbix-sender
- Graylog2
- Kibana
- Octopussy
  → zabbix-sender
- Flume (ETL!)
- Fluentd
  → zabbix-sender
- ... audit
  ... systemd journal
One of my metrics “idea's”

Monitoring - Metrics and Logging network for Linux

CPU
DISK
DB's
Context Switches
Connection tracking
Apache
Nginx
NFS
IPTables
Swap
Varnish
Processes
Perl LDAP monitoring ...

Log files (decentral or central log server)
Logstash
Elastic Search (search cluster)
Kibana
Graphite frontend

Collectd
Graphite server
Zabbix Server
Zabbix frontend

Zabbix Agent
RabbitMQ
Alerting and forwarding to other systems

Stated
Riemann (event processing, decision engine)

Application code

IPMI
SNMP
SSH
Telnet ...

February 2013 - PieterB
If you have money...

- Very easy to install
- Scales, integrates, Big Data...
- Splunk free: 500MB/day indexing volume :-(
- Missing some features as well
- Good enough for a test
- Integration using zabbix-sender
Open source logging infrastructure @ ICTRA

- General
  - Rsyslog
  - Logstash
  - ElasticSearch
  - Kibana

- Other
  - ELSA (Splunk alike)
    - Security related
  - Graylog
Rsyslog

- Used for our central log repository
- Reliability:
  - Use on disk-queues
  - Use RELP → application level reliability
    - TLS available in recent versions
- Output & input modules
- Filter
- Output format can be configured
- Use high-precision timestamps
- Future: CEE/Lumberjack
Alerting from Rsyslog

- **ommail**
  - if $msg contains 'hard disk fatal failure'
    then :ommail:;mailBody

- **omprog**
  - → to zabbix_sender

- **omsnmp**
Logstash

- Collect logs, parse and store for later use
- Written in JRuby
- Easy to deploy
- Inputs
  - file, log4j, queues, SNMP, syslog, RELP, GELF...
- Use logstash when you need filters
  - kv, grep, grok, mutate, xml, multiline
- With logstash you can parse all those weird log formats and get something useful
Logstash components

- **Shipper**
  - collect and forward events to other instances
  - remote or on the central syslog servers

- **Broker**
  - Redis
  - RabbitMQ

- **Indexer**
  - Receives and indexes events
  - From Redis to ElasticSearch

- **Kibana**
  - Webinterface for ElasticSearch and Logstash
Logstash → zabbix_sender examples

- Keepalived (HAPProxy HA)
- OpenDJ
  - OpenDJ → multiple backend instances → multiple access logs
  - performance counters “etime”
  - counts of user x logins (for patterns)
  - MILD_ERR or worse in log file → alert to respective level in zbx
- Java Applications: parse xml, warning on condition x
Keepalived example

- Input file: messages
- Filter to work only with interesting messages
  - Grep
  - Or grok
    - pattern => "{SYSLOGLINE}"  
  - Grok on “program”
- I prefer to work with booleans when possible
  - Mutate:
    tags => 'keepalived_state_master'
    replace => ["@message", "1"]
    add_tag => [ "zabbix_sender" ]
    add_field => [ "zabbix_host", "@source_host", "zabbix_item", "keepalived.status" ]
Logstash example

- Send it to zabbix
- OpenDJ
  - Access log
    - Entries as "BIND RES conn=1 op=2 msgID=3 result=0 authDN="uid=a" etime=102
  - First I tried grok & multiline
    - but.... a simple kv filter for key=value formats exists
- NOTES:
  - Test
    - `java -jar /opt/logstash/logstash.jar agent -f /etc/logstash/conf.d/x.conf`
  - Try different approaches → what offers the best performance?
  - exclude _grokparsefailure when necessary
  - know the available filters
ElasticSearch

- Indexing and searching logs
- distributed RESTful search and analytics
- Scales horizontally
- What about long-term storage?
  - Use an archiving platform?
- Discovery: multicast or unicast
  - discovery.zen.ping.multicast.enabled: false
- Solr?
  - Compared on http://solr-vs-elasticsearch.com/
A frontend: Kibana

Or try the new version (Kibana 3) on http://demo.kibana.org
Another logging tool: Graylog

- Uses AMQP, GELF...
- LDAP integration
- Very good for applications
  - libraries
- For syslog: use Logix
Octopussy

- Described as “open source log management”
- Based on Perl...
- Features nice for enterprise usage:
  - LDAP
  - A lot of templates are already included
    - Bind, Cisco Router, Cisco Switch, DenyAll Reverse Proxy, Drbd, F5 BigIP, Fortinet FW, Ironport MailServer, Linux Kernel/System, Linux IPTables, Monit, MySQL, Nagios, NetApp NetCache, Juniper Netscreen FW, Juniper Netscreen NSM, Postfix, PostgreSQL, Samhain, Snmpd, Squid, Sshd, Syslog-ng, Windows Snare Agent, Xen...
  - Sends alerts with zabbix_sender :-)}
ELSA – Enterprise Log Search and Archive

- Uses MySQL + Sphinx
- Syslog-ng instead of rsyslog → patterndb
- LDAP
- Normalization
  - open-source IDS (Bro/Suricata/OSSEC)
  - Cisco
  - Email alerts possible →
    should be trivial to call zabbix_sender
- Had some issues with installation script
- Use Security Onion for a testdrive
- Beware of the specific query language
Not tested: fluentd

- Documentation seems complete
- Performance in the line of other tools?
  - “largest user currently collects logs from 5000+ servers, 5 TB of daily data, handling 50,000 msgs/sec”
- Japanese community?
Q&A

- Any questions?