Keeping The Junk Out Of Splunk (V2.0)

Sandy D. Voellinger

Principal Consultant, The Crypsis Group

1 Keeping the Bean Counters Happy

2 Getting the most out of your Splunk license

Introduction And Backstory

Disclaimer

During the course of this presentation, we may make forward looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC. The forward-looking statements made in the this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not, be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

Agenda

- The 5th "V"
- Component Overview
- Selectivity
- Filtering
- Limits
- Other

The 5th “V”

The Accelerating Pace Of Data

Volume | Velocity | **?????** | Variety | Variability

Machine data is the fastest growing, most complex, most valuable area of big data

The 5th "V"

Volume | Velocity | **VALUE** | Variety | Variability

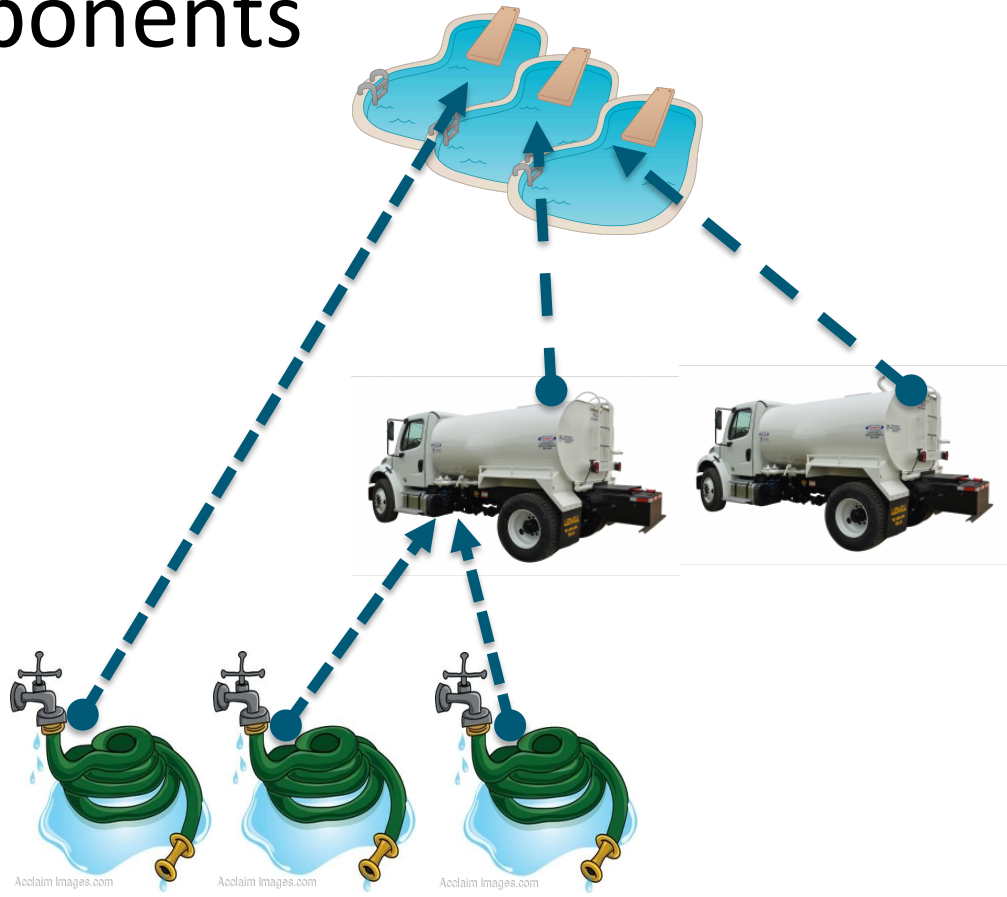
Component Overview

Components

- Props.conf
  - Defines what rules are applied to any event and when they are applied
- Transforms.conf
  - Actually defines the rules from props.conf
- Inputs.conf
  - Defines what we're actually monitoring on the edge host
- Limits.conf
  - Set various limits and varies depending on role of system you set it on

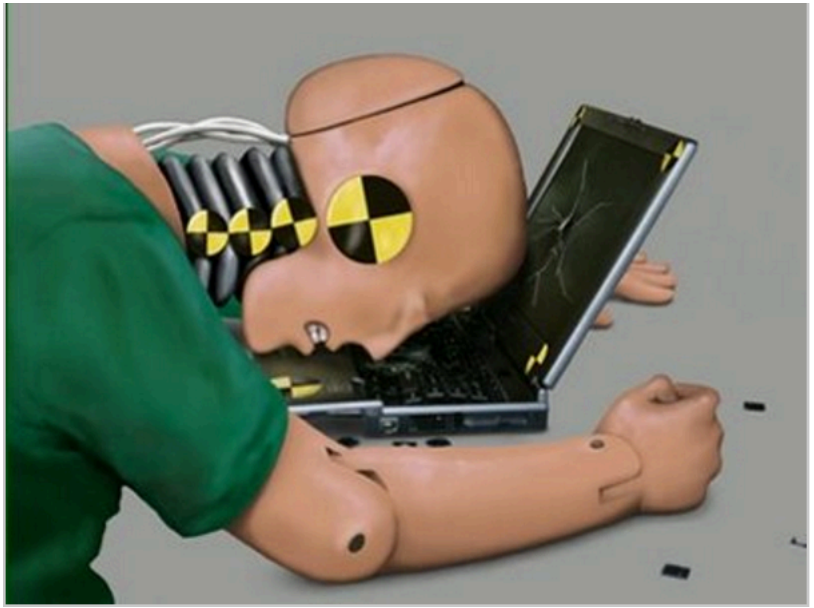
Components

- Indexers
  - Fully capable
  - Deposits data to disk
- Forwarder
  - Heavy weight (HWF)
    - Fully capable
  - Universal (UF)
    - Light weight agent

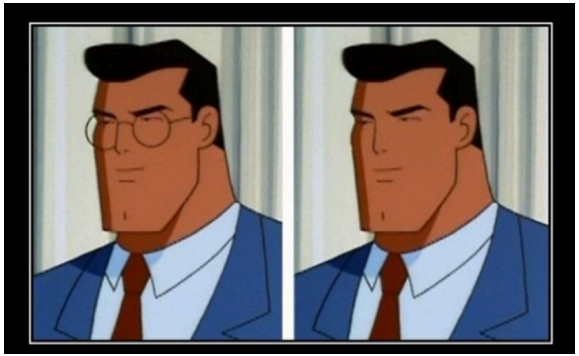


Selectivity

Don’t Be A Dummy: TEST



Look Closely, And Be PICKY!



Pay attention to what is Splunk is actually consuming

- Forwarders:
  - Consume by the file, not the entire directory by default
    - Future unknowns: Where did that file come from?
    - Manual copy's/changes cause Re-Indexing
    - Avoids Whitelisting/Blacklisting
  - Granular SourceTyping means easier to find

Monitor the actual file

- Beware Logrotate
  - access_log, not access_log*
- Testing
  - oneshot, spool

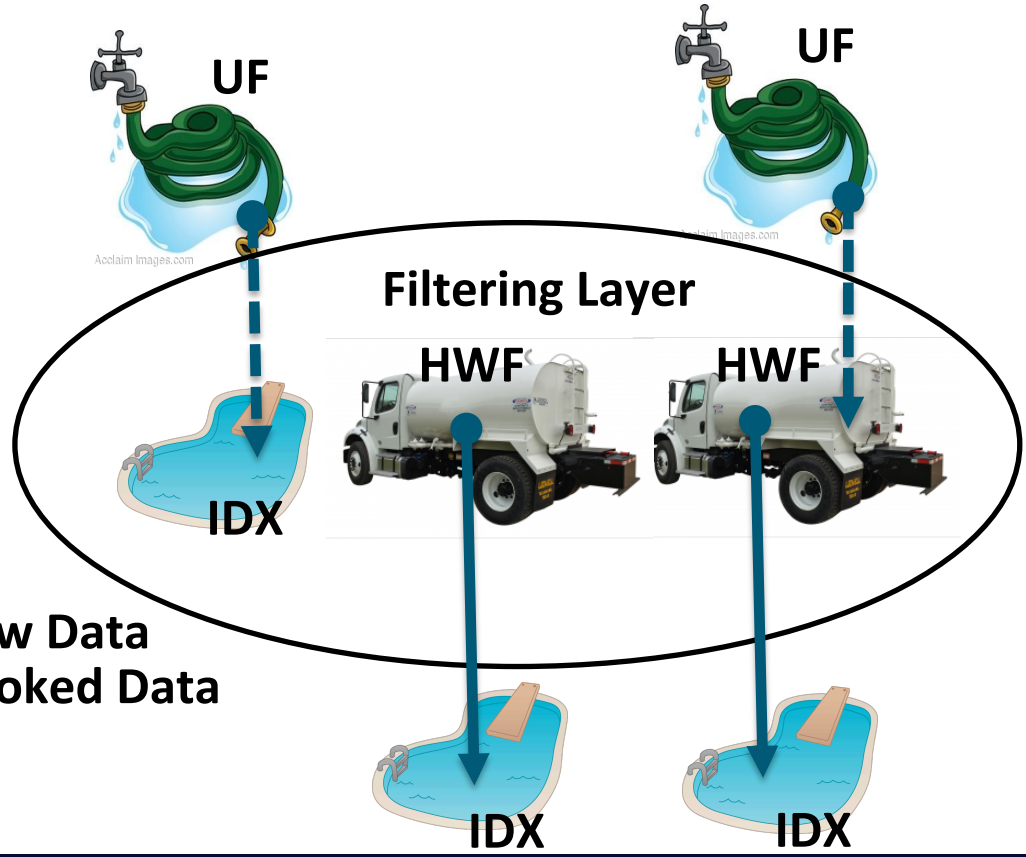
```
[monitor:///var/log/access_log]
sourcetype=access_log_mycom
disabled=false
index=web_logs
```

Filtering

Powerful

Where?

- HWF at the edge
- Intermediate HWF with UF at the edge
  - CPU
  - Bits
- Indexers
  - Core exhaustion



Aug 6, 2015 10:52:27 PM org.apache.catalina.core.ApplicationContext log DEBUG: ContextListener: contextInitialized()
Aug 6, 2015 10:52:27 PM org.apache.catalina.core.ApplicationContext log DEBUG: SessionListener: contextInitialized()
Aug 6, 2015 10:52:27 PM org.apache.catalina.core.ApplicationContext log INFO: ContextListener: attributeAdded('org.apache.jasper.compiler.TldLocationsCache', 'org.apache.jasper.compiler.TldLocationsCache@67c78f98', 'password: 780380dd7bc5ba URL: http://example.com')
Aug 6, 2015 10:55:50 PM org.apache.catalina.core.ApplicationContext log DEBUG: SessionListener: contextDestroyed()
Aug 6, 2015 10:55:50 PM org.apache.catalina.core.ApplicationContext log DEBUG: ContextListener: contextDestroyed()
Aug 6, 2015 10:55:54 PM org.apache.catalina.core.ApplicationContext log DEBUG: ContextListener: contextInitialized()
Aug 6, 2015 10:55:54 PM org.apache.catalina.core.ApplicationContext log DEBUG: SessionListener: contextInitialized()
Aug 6, 2015 10:55:54 PM org.apache.catalina.core.ApplicationContext log INFO: ContextListener: attributeAdded('org.apache.jasper.compiler.TldLocationsCache', 'org.apache.jasper.compiler.TldLocationsCache@4963d86c', 'password: 5139bdda898af3b85 URL: http://example.com')
Aug 6, 2015 10:55:55 PM org.apache.catalina.core.ApplicationContext log DEBUG: SessionListener: contextDestroyed()
Aug 6, 2015 10:55:55 PM org.apache.catalina.core.ApplicationContext log DEBUG: ContextListener: contextDestroyed()
Aug 6, 2015 10:56:09 PM org.apache.catalina.core.ApplicationContext log DEBUG: ContextListener: contextInitialized()
Aug 6, 2015 10:56:09 PM org.apache.catalina.core.ApplicationContext log DEBUG: SessionListener: contextInitialized()
Aug 6, 2015 10:56:09 PM org.apache.catalina.core.ApplicationContext log INFO: ContextListener: attributeAdded('org.apache.jasper.compiler.TldLocationsCache', 'org.apache.jasper.compiler.TldLocationsCache@cd74b31', 'password: d41d8cd98f00b URL: http://example.com')

Starting byte count: 1964

Low Hanging Fruit

- Start with the easiest, filter out entire line

props.conf:

```
TRANSFORMS-setNull = NukeDEBUGLines
```

transforms.conf:

```
[NukeDEBUGLines]
REGEX = DEBUG:
DEST_KEY = queue
FORMAT = nullQueue
```

Result #1

Aug 6, 2015 10:52:27 PM org.apache.catalina.core.ApplicationContext log INFO: ContextListener: attributeAdded('org.apache.jasper.compiler.TldLocationsCache', 'org.apache.jasper.compiler.TldLocationsCache@67c78f98', 'password: 780380dd7bc5ba URL: http://example.com')
Aug 6, 2015 10:55:54 PM org.apache.catalina.core.ApplicationContext log INFO: ContextListener: attributeAdded('org.apache.jasper.compiler.TldLocationsCache', 'org.apache.jasper.compiler.TldLocationsCache@4963d86c', 'password: 5139bdda898af3b85 URL: http://example.com')
Aug 6, 2015 10:56:09 PM org.apache.catalina.core.ApplicationContext log INFO: ContextListener: attributeAdded('org.apache.jasper.compiler.TldLocationsCache', 'org.apache.jasper.compiler.TldLocationsCache@cd74b31', 'password: d41d8cd98f00b URL: http://example.com')

Byte count: 803. A 60% reduction from 1964 bytes.

Focus On Targeted Opportunities

- Surgically replace or remove text

props.conf:

```
SEDCMD-obf-pass = s/password: .* Url:/password: x Url:/g
```

- Gotcha: SEDCMD comes before TRANSFORMS

Result #2

Aug 6, 2015 10:52:27 PM org.apache.catalina.core.ApplicationContext log INFO: ContextListener: attributeAdded('org.apache.jasper.compiler.TldLocationsCache', 'org.apache.jasper.compiler.TldLocationsCache@67c78f98', 'password: x URL: http://example.com')
Aug 6, 2015 10:55:54 PM org.apache.catalina.core.ApplicationContext log INFO: ContextListener: attributeAdded('org.apache.jasper.compiler.TldLocationsCache', 'org.apache.jasper.compiler.TldLocationsCache@4963d86c', 'password: x URL: http://example.com')
Aug 6, 2015 10:56:09 PM org.apache.catalina.core.ApplicationContext log INFO: ContextListener: attributeAdded('org.apache.jasper.compiler.TldLocationsCache', 'org.apache.jasper.compiler.TldLocationsCache@cd74b31', 'password: x URL: http://example.com')

Byte count from password sub: 762. A 63% reduction from 1964 bytes.

Continued…

- Surgically replace or remove text

props.conf:

```
SEDCMD-javaclass = s/org\.apache\.catalina\.core\.ApplicationContext\s+log\s+//g
```

- Gotcha: SEDCMD comes before TRANSFORMS

Result #3

Aug 6, 2015 10:52:27 PM INFO: ContextListener: attributeAdded('org.apache.jasper.compiler.TldLocationsCache', 'org.apache.jasper.compiler.TldLocationsCache@67c78f98', 'password: x URL: http://example.com')
Aug 6, 2015 10:55:54 PM INFO: ContextListener: attributeAdded('org.apache.jasper.compiler.TldLocationsCache', 'org.apache.jasper.compiler.TldLocationsCache@4963d86c', 'password: x URL: http://example.com')
Aug 6, 2015 10:56:09 PM INFO: ContextListener: attributeAdded('org.apache.jasper.compiler.TldLocationsCache', 'org.apache.jasper.compiler.TldLocationsCache@cd74b31', 'password: x URL: http://example.com')

Byte count from removing java class: 618. A 68.5% reduction from 1964 bytes!

The Nuclear Option

- Remove multi-line text

props.conf:

```
SEDCMD-NukeXMLPayload = s/(?s)XML:.*$//
```

Aug 07, 2015 12:00:06 AM Lorem ipsum dolor sit amet

XML:<xml-ns=2014-09-01-Splunk Rules><note>
  <to>Tove</to>
  <from>Jani</from>
  <heading>Reminder</heading>
  <body>Don't forget me this weekend!</body>
</note>

Blank Slate

- Start with nothing, selectively accept events

props.conf:

```
[source::/var/log/messages]
TRANSFORMS-set = setnull,setparsing
```

```
Aug 17 04:13:26 li123-45 clamd[2891]: SelfCheck: Database status OK.
Aug 17 04:23:01 li123-45 ntpd[2883]: kernel time sync enabled 0001
Aug 17 04:23:26 li123-45 clamd[2891]: SelfCheck: Database status OK.
Aug 17 04:28:19 li123-45 sshd[7588]: PAM 6 more authentication failures;
logname= uid=0 euid=0 tty=ssh ruser= rhost=183.57.57.163  user=root
Aug 17 04:33:26 li123-45 clamd[2891]: SelfCheck: Database status OK.
Aug 17 04:43:26 li123-45 clamd[2891]: SelfCheck: Database status OK.
Aug 17 04:53:26 li123-45 clamd[2891]: SelfCheck: Database status OK.
Aug 17 05:12:56 li123-45 clamd[2891]: SelfCheck: Database status OK.
```

A Bite Of The Apple At A Time

transforms.conf:

```
[setnull]
REGEX = .
DEST_KEY = queue
FORMAT = nullQueue

[setparsing]
REGEX = \s+sshd\[\d+\]:
DEST_KEY = queue
FORMAT = indexQueue
```

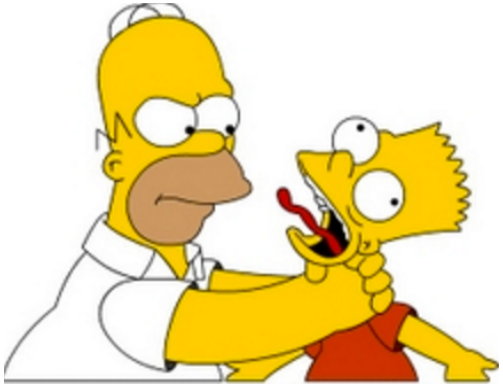
More docs on Route and Filter Data:

http://goo.gl/MhRxLV

Don’t Choke On Your Windows Data

Windows Event Logs

- Inputs.conf



http://goo.gl/YwBg0d

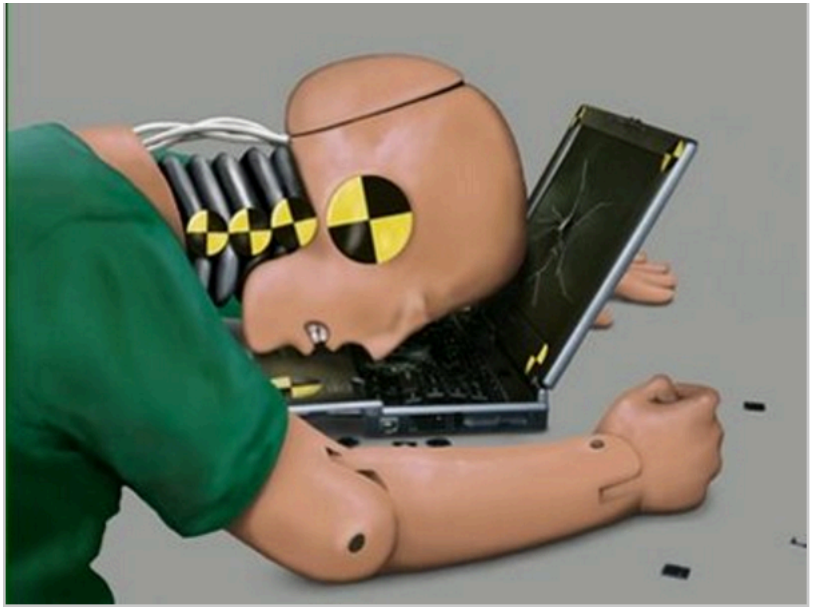
```
[WinEventLog://Application]
disabled = true
blacklist = 1,1001
index = windows

[WinEventLog://Security]
disabled = false
whitelist = 560,562,578,4663,4670,4768
index = windows

[WinEventLog://System]
disabled = true
blacklist = EventCode=%^1([8-9])$%"
index = windows
```

Limits

Don’t Be A Dummy: TEST

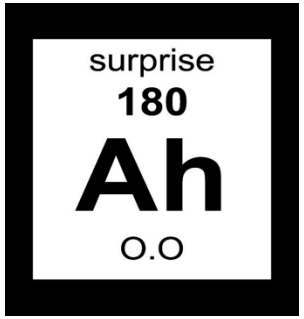


Throughput Limits Are Your Friend

- UF default: 256KB -- Thank you Splunk Product folks
  - Still, 22GB/day from a single UF possible
  - If small license, consider lowering

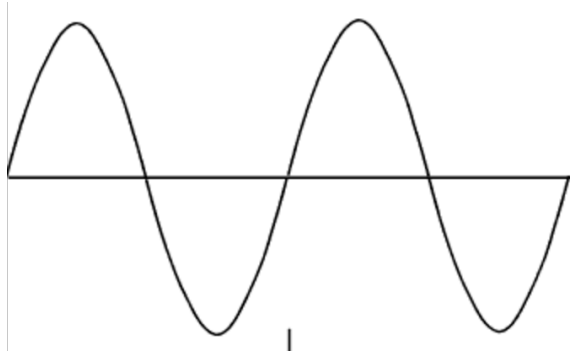
- limits.conf: [thruput] maxKBps=
  - Caveats: UF may never catch up, data may be lost
  - May need to filter closer to the edge, consider upgrading to HWF
  - Can be controlled via Deployment Server
  - Maybe check out queueSize in inputs.conf?

Throughput, Cont'd



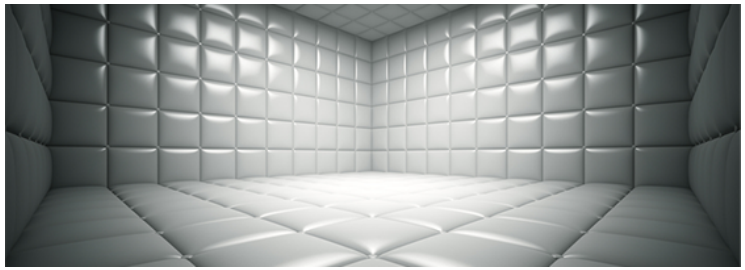
- HWF: No limit
  - Watch for surprises
  - Can be controlled by DS

- Indexer: limits.conf works here, too
  - Can be used to ensure you don't blow license
  - License size / # of indexers / 86400 = maxKBps
  - Use with ***extreme*** caution
  - SoS or DMC : indexingQueue provides visibility
  - Establish monitoring for sanity checking
    - stddev, outliers, hard limits



Other Options

Isolation



- Create separate indexers for consistently misbehaving forwarders
  - Only impacts search heads associated with that index or cluster
- Add maxKBps limit based on allowable consumption
  - Throttle their data
- License Pools
  - Carve up license by an allocated amount

Other Layers

- Packet filtering at OS: iptables, pf
  - Limit by src IP, subnet

- Firewall or switch rate limiting by src IP/subnet
  - PPS / BPS
    - *Example: switch(config)# hardware rate-limit port-security 1000*
  - Buy your NetEng/FWEng their favorite beverage(s) -- This will take some tweaking

Contact Information

- Sandy Voellinger
  - Principal Consultant @ www.crypsisgroup.com
  - https://www.linkedin.com/in/sandyv
  - Skype: svoellinger
    Splunk-UsersGroup on Slack: svoellinger
  - Sandy.Voellinger@CrypsisGroup.com

THANK YOU

splunk>