splunk> .conf2017

Critical Syslog Tricks

(That No One Seems to Know About)

Jonathan Margulies | Security Expert/ Splunk Professional Svcs Consultant, Rational Cyber
George Barrett | Security Expert/ Splunk Professional Svcs Consultant, Rational Cyber

September 2017 | Washington, DC

Forward-Looking Statements

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 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.

splunk> .conf2017

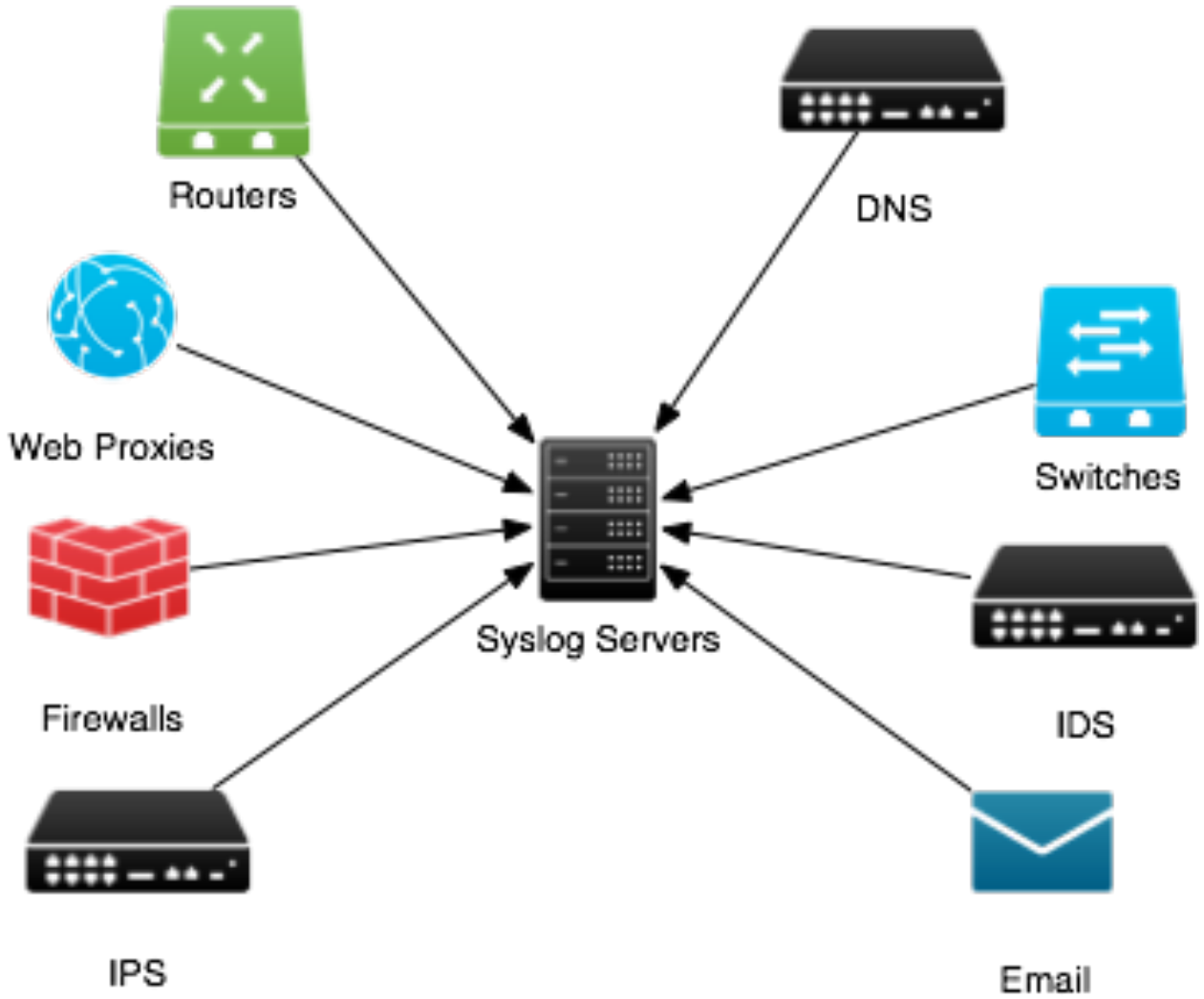
Do You Have A Syslog Collection Problem?

splunk> .conf2017

You Might Have A Syslog Collection Problem If…

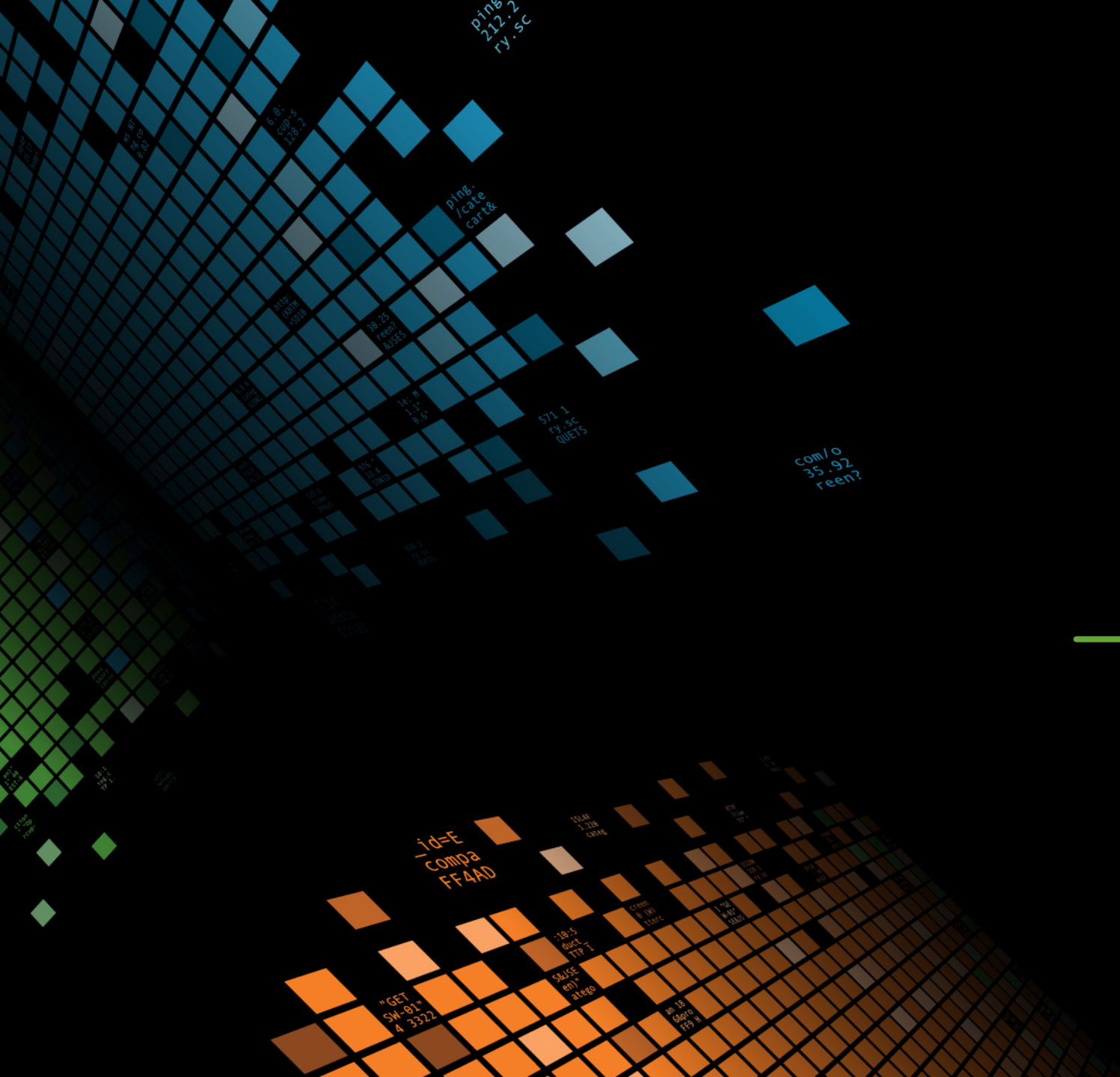
- Your syslog data arrives in Splunk more than a few seconds after the event time
- Syslog data that comes in while Splunk is restarting gets dropped
- You notice gaps or missing events in your syslog data feeds
- You need a new listening port every time you get a new syslog data source
- Your indexers or heavy forwarders have to look in raw events to figure out what index, sourcetype, or host to assign to those events
- Multiple hosts' syslog data are being aggregated under the same host because they came through the same syslog server
- Your IT people use grep instead of Splunk to troubleshoot live issues

Syslog Brings In Your Most Important Logs



What You'll Learn From This Presentation

- How to configure syslog-ng to collect all your syslog data for Splunk
- How to architect your syslog collection infrastructure
- How to configure Splunk to collect all the data from syslog-ng and index it in about 3 seconds
- How to find and troubleshoot syslog collection problems quickly



syslog-ng

splunk> .conf2017

A Few Things To Note About Syslog-ng

- It's free. There's a paid version, but this presentation assumes you didn't buy it.
- We recommend version 3.5 or higher, as that supports multithreading and some other useful features
- You can do everything we're recommending using rsyslog instead, but we don't recommend it
  - syslog-ng handles poorly formatted syslog events more gracefully

https://syslog-ng.org/

https://www.balabit.com/documents/syslog-ng-ose-latest-guides/en/syslog-ng-ose-guide-admin/html/index.html

Configuring syslog-ng (options)

```
options {
    flush_lines (100);
    time_reopen (10);
    log_fifo_size (1000);
    chain_hostnames (off);
    use_dns (no);
    use_fqdn (no);
    create_dirs (yes);
    keep_hostname (yes);
    threaded (yes);
};
```

https://gitlab.com/rationalcyber/syslog-ng-configuration/blob/master/syslog-ng.conf

Configuring syslog-ng (Listening And Writing)

```
source s_aggregation {
    udp(ip(0.0.0.0) port(514));
    tcp(ip(0.0.0.0) port(514));
};

destination d_splunkf {
     file(“/mnt/$LOGHOST/log/$R_YEAR-$R_MONTH-
$R_DAY/$HOST_FROM/$HOST/$FACILITY.log" dir-owner("splunk") dir-
group("splunk") owner("splunk") group("splunk"));
};
```

splunk> .conf2017

This Is The Most Important Line!

file("/mnt**/$LOGHOST/log/$R_YEAR-R_MONTH-R_DAY/$HOST_FROM/**$HOST/$FACILITY.log" dir-owner("splunk") dir-group("splunk") owner("splunk") group("splunk"));

- **/$LOGHOST**
  - Essentially, "the hostname of this syslog-ng server." You're going to be collecting syslog on more than one server, so this will help with troubleshooting.
- **/log/R_YEAR-R_MONTH-$R_DAY**
  - This is important for log rotation. We'll explain that on its own slide.
- **/$HOST_FROM**
  - "The host I received this feed from." It may be the same as the originating host, or it may be an intermediate syslog server. In the latter case, helps with troubleshooting.

splunk> .conf2017

The Rest Of That Line

file(“/mnt/$LOGHOST/log/$R_YEAR-R_MONTH-R_DAY/$HOST_FROM**/$HOST/$FACILITY.log” dir-owner("splunk") dir-group("splunk") owner("splunk") group(“splunk"))**;

- **/$HOST**
  - “The hostname from the syslog header.” This may be an actual hostname, FQDN, or IP address, but it’s always the most reliable source of the logs’ originating host.
- **/$FACILITY.log**
  - “The syslog facility setting.” This generally isn’t useful by itself, but it can almost always be used in combination with $HOST to separate different sourcetypes from the same host.
- **dir-owner("splunk") dir-group("splunk") owner("splunk") group(“splunk”))**
  - Splunk should never be running as root! Make sure the splunk user can read and rotate all the log files.

Rotating Logs

- **Do not use logrotate on a syslog server**
  - It will restart syslog-ng and you'll lose a couple of seconds of logs
  - https://www.balabit.com/documents/syslog-ng-ose-latest-guides/en/syslog-ng-ose-guide-admin/html/example-logrotate.html
- Use these cron jobs instead (adjust the times as needed):

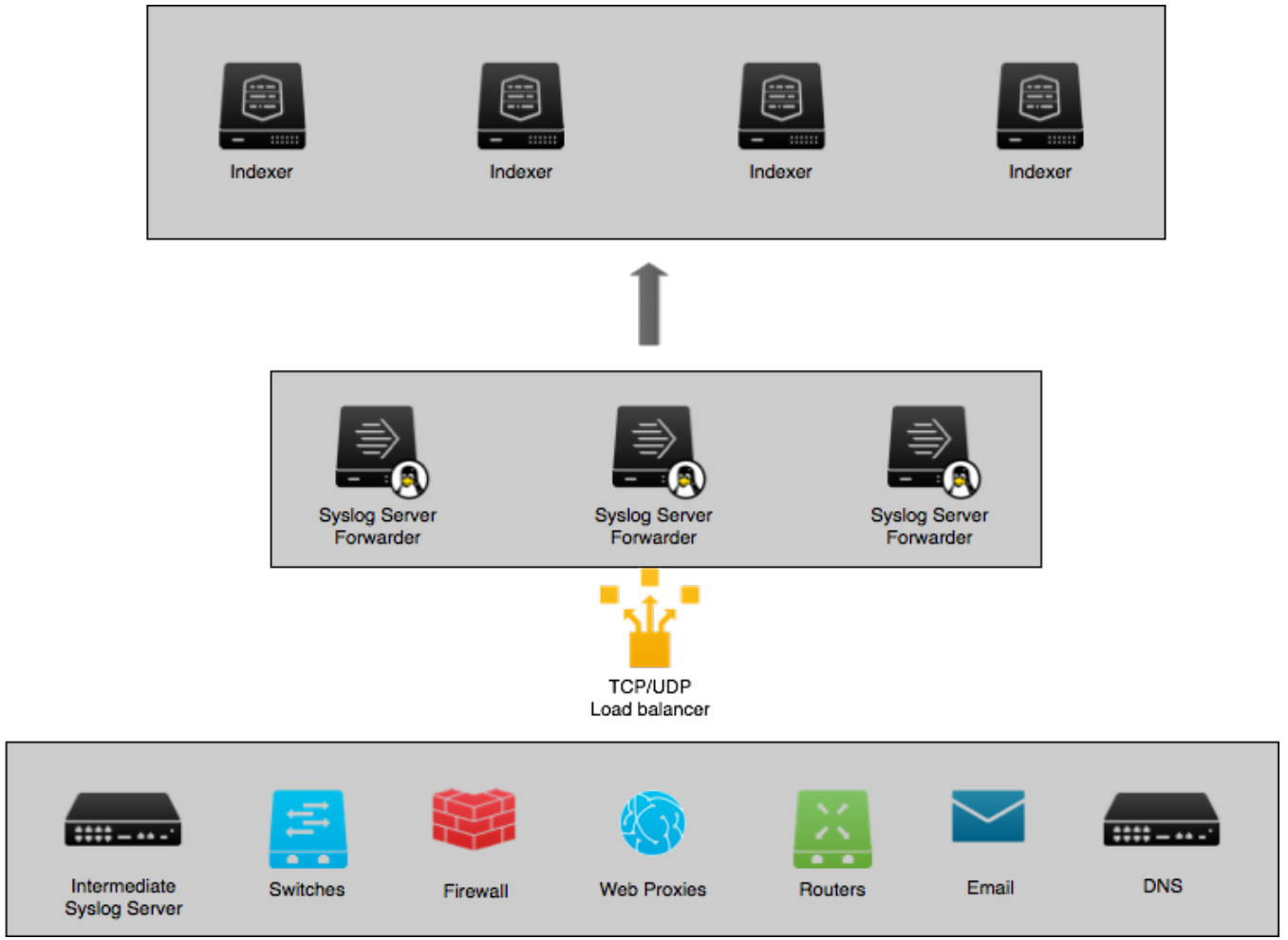
```
#cron job 1: at 5am, find yesterday's logs, and move them to old_logs
0 5 * * * /usr/bin/find /mnt/*/log/????-??-?? -maxdepth 0 -type d ! -mmin -
300 -exec bash -c 'dir={}; old=${dir/\/log\//\/old_logs\/}; mv ${dir}
${old}' \;

#cron job 2: find any files older than 5 days, 23 hours, and delete them
0 4 * * * /usr/bin/find /mnt/*/old_logs/????-??-?? -maxdepth 0 -type d ! -
mmin -8580 -exec rm -rf {} \;
```

splunk> .conf2017

Architecting Syslog Infrastructure For Splunk

splunk> .conf2017

Network Architecture



What Kind of Forwarder?

Heavy vs Universal

Heavy Forwarder Advantages

- Can handle timezone conversions
  - Keep your props and inputs together
- Takes load off your indexers
- PII masking

Universal Forwarder Advantages

- Need a lot less bandwidth to the indexing tier
  - Less metadata
- Less processor/memory load on the syslog servers

splunk> .conf2017

Configuring The Forwarder

splunk> .conf2017

inputs.conf

```
[monitor:///mnt/log/*/*/fireeye*/local2.log]
host_segment = 5
index = idps
sourcetype = fe_cef_syslog

[monitor:///mnt/log/*/*/mail*/*]
host_segment = 5
index = mail
sourcetype = sendmail_syslog
```

props.conf

```
[source::/mnt/log/*/*/fireeye*/local2.log]
SHOULD_LINEMERGE = false
TZ = UTC

[source::/mnt/log/*/*/mail*/*]
SHOULD_LINEMERGE = false
TZ = US/Eastern
```

outputs.conf

- Most of Splunk's pipeline queues default to a maximum size of 512KB. That's fine for a normal universal forwarder, but not for a syslog server
- Is your output queue too small?
  - `index=_internal host=<syslog_server> source=*metrics.log group=queue name=tcpout* | eval output_queue_pct=current_size/max_size*100 | timechart perc95(output_queue_pct) by host | eval Bad=80`
- A 64MB output queue works well for many enterprise syslog servers, but you may need more (if your 64MB queue is filling up) or less (if your RAM is filling up)
- Outputs.conf contents:

```
[tcpout]
maxQueueSize = 64MB
```

server.conf

- Like the outputs queue, most Splunk queues default to a maximum size of 512KB, which is often insufficient for a syslog server
- Contents of server.conf:

```
[queue]
maxSize = 64MB
```

Better Balance Across Indexers

- For better load balancing, have the forwarders change indexers often and mid-stream:
  - outputs.conf:

```
[tcpout]
autoLBFrequency = 5
forceTimebasedAutoLB = true
```

  - On UFs running 6.5+, do not use forceTimebasedAutoLB. Add this to props.conf for each data source (or in `[default]`) instead:

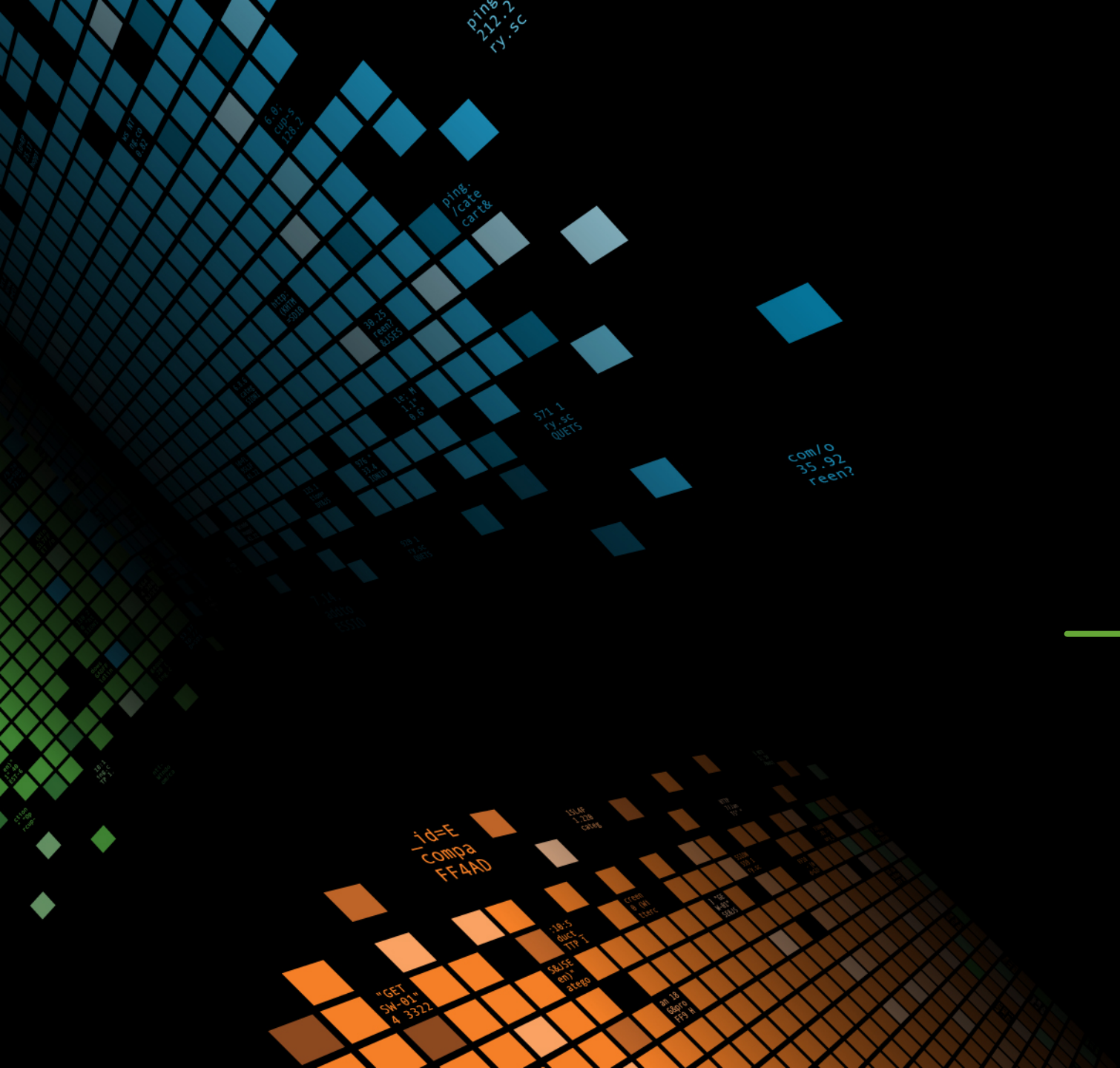
```
EVENT_BREAKER_ENABLE = true
```

https://www.splunk.com/blog/2014/03/18/time-based-load-balancing.html

Parallel Ingestion Pipelines

- Parallel ingestion pipelines allow Splunk to use more resources so it can ingest multiple streams of data at once
- Since these syslog servers are dedicated to Splunk data collection, they're excellent candidates for this feature
- The number of pipelines you set will depend on your hardware capacity and data rates. See notes on side effects of this setting at https://docs.splunk.com/Documentation/Splunk/latest/Admin/Serverconf
- Enabling parallel ingestion pipelines in server.conf:

```
[general]
parallelIngestionPipelines = 2
```



Automation

Syslog at Scale

- In a large enterprise, do not build syslog inputs and props manually!
  - With thousands of syslog feeds, they become impossible to manage
  - Small typos can cause massive failures
- We manage all of our syslog inputs in a CSV file and edit in Excel
- Find our script to auto-generate inputs.conf and props.conf for syslog servers at:
  - https://gitlab.com/rationalcyber/

Using A Catchall Index

- Sometimes upstream syslog sources start sending data you weren't expecting
- You want this data in Splunk, but you don't know what index or sourcetype to give it
- inputs.conf:

```
[monitor:///mnt/log/*]
blacklist = /mnt/log/*/*/((fireeye*/local2.log)|(mail*/*))
index = catchall
```

- This blacklist regex becomes unmanageable quickly; the script on the previous slide auto-generates it for you

Monitoring And Alerting

- Problems with one of the Splunk syslog servers (run every few minutes):
  - `| tstats count where source=/mnt/log/* by source | rex field=source "/mnt/log/(?<splunk_syslog_server>[^/]+)/" | stats sum(count), count by splunk_syslog_server`
- Problems with an upstream syslog server (run every few minutes):
  - `| tstats count where source=/mnt/log/* by source | rex field=source "/mnt/log/[^/]+/(?<upstream_syslog_server>[^/]+)/" | stats sum(count), count by splunk_syslog_server`
- Queues filling up and causing delays (observe daily—look for sustained issues):
  - `index=_internal host=<syslog_server> source=*metrics.log group=queue | eval queue_pct=if(isnull(current_size_kb), (current_size/max_size), (current_size_kb/max_size_kb)) | timechart limit=50 perc99(queue_pct) by name | eval Bad=80`
- Unknown syslog feeds (check weekly):
  - `| tstats count where index=catchall by source`

splunk> .conf2017

Thank You!

- george@rationalcyber.com
- jonathan@rationalcyber.com
- All of our open source projects, including all of our syslog resources:

https://gitlab.com/rationalcyber/

splunk> .conf2017

Thank You

Don't forget to rate this session in the .conf2017 mobile app

splunk> .conf2017