The Truthiness of Wire Data: Using Splunk Stream for Performance Monitoring

David J. Cavuto
Principal Product Manager, Splunk App for Stream
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.
Cavuto Bio

- General Awesomeness
Presentation Overview

1. Problem Statement
2. What is Wire Data? What is Splunk Stream?
3. Using Wire data from Stream in ITSI
4. Automatic Anomaly Detection using Wire Data in ITSI
5. Where to go from here
Problem Statement

- Applications may not accurately report their own performance accurately, for a number of reasons
  - Heavily loaded
  - Mistaken measurements
  - Developer error
  - It Died!
  - No access to data (politically, organizationally)
  - They’re lying! 😊

- Insert quote about trusting something to monitor itself
Why Wire Data with Splunk Stream?

- Use Wire Data provided by Splunk Stream to monitor application and network performance
- Direct ingest into Splunk (no props/transforms) makes it simple
- Stream is not a dedicated APM / NPM tool, but has aspects of both
- In most cases, Stream will do everything you need to isolate faults and pinpoint the trouble spot(s)
- It’s Free!
What’s Network (Wire) Data?

- Machine data
- Poly-structured data
- Authoritative record of real-time and historical communication between machines and applications

```
tcpdump -qns 0 -A -r blah.pcap
20:57:47.368107 IP 205.188.159.57:23346 > 67.23.28.65:481: tcp 480
0x0000: 4500 0214 834c 0000 3306 f649 cdb3 9f39 E...L@.3..I...9
0x0010: 4317 1c41 0019 a591 50fe 18ca 9da0 4681 C..A....P......F.
0x0020: 8018 05a8 848f 0000 0101 080a ffd4 9bb0 ................
0x0030: 2e43 6bb9 3232 302d 726c 792d 6461 3033 .Ck.220-raly-da03
0x0040: 0x0050: 0x0060: 0x0070: 0x0080: 0x0090: 0x00a0: 0x00b0: tcpdump -qns 0 -A -r blah.pcap
```
Splunk App for Stream (6.6)

- **Metadata Collection**
  - Collects essential elements of the application conversation
  - Eliminates redundancy of duplicate packet headers

- **Live Interface Collection Option**
  - Collect directly on hosts
  - Also from a tap or SPAN port

- **Estimate Mode**
  - Deploy Stream without collecting data (or affecting license)
  - Test data volume

- **Aggregation Mode**
  - Statistics generated at endpoint
  - Similar to “stats sum(x)” in SPL

- **Filtering at Endpoint**

- **Out-of-Box Content**
  - Dashboards for common protocols

- **Distributed Forwarder Mgt**
  - Similar to Splunk UF mgt
  - All config centrally managed
  - Forwarder Groups
How will Wire Data help Solve Problem?

- Wire data represents capture of true conversations between endpoints
- It has the “omniscient view” of what actually transpired
- The conversations contain the details about each transaction, including the time of occurrence
- Less chance of interference
  - Intentional / Malicious
  - Load or resource based
## Protocols Parsed with Stream 6.6

<table>
<thead>
<tr>
<th>Simple Transport</th>
<th>File Service</th>
<th>Authentication</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCP</td>
<td>NFS</td>
<td>Diameter</td>
</tr>
<tr>
<td>UDP</td>
<td>SMB</td>
<td>LDAP</td>
</tr>
<tr>
<td>IP</td>
<td></td>
<td>RADIUS</td>
</tr>
<tr>
<td><strong>Infrastructure</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNMP</td>
<td>IMAP</td>
<td></td>
</tr>
<tr>
<td>DHCP</td>
<td>MAPI</td>
<td></td>
</tr>
<tr>
<td>DNS</td>
<td>POP3</td>
<td></td>
</tr>
<tr>
<td>ICMP</td>
<td>SMTP</td>
<td></td>
</tr>
<tr>
<td><strong>File Transfer</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FTP</td>
<td>AMQP</td>
<td></td>
</tr>
<tr>
<td>HTTP</td>
<td>IRC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SMPP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>XMPP</td>
<td></td>
</tr>
<tr>
<td><strong>Email</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Messaging</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Voice over IP (VoIP)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Examples of What’s Available From the Streaming Network/Wire Data

<table>
<thead>
<tr>
<th>Performance Metrics</th>
<th>Application Data</th>
<th>Business Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round Trip Time</td>
<td>POST Content</td>
<td>Product ID</td>
</tr>
<tr>
<td>Client Request Time</td>
<td>AJAX Data</td>
<td>Customer ID</td>
</tr>
<tr>
<td>Server Reply Time</td>
<td>Section</td>
<td>Shopping Cart ID</td>
</tr>
<tr>
<td>Server Send Time</td>
<td>Sub-Section</td>
<td>Cart Items</td>
</tr>
<tr>
<td>Total Time Taken</td>
<td>Page Title</td>
<td>Cart Values</td>
</tr>
<tr>
<td>Base HTML Load Time</td>
<td>Session Cookie</td>
<td>Discounts</td>
</tr>
<tr>
<td>Page Content Load Time</td>
<td>Proxied IP Address</td>
<td>Order ID</td>
</tr>
<tr>
<td>Total Page Load Time</td>
<td>Error Message</td>
<td>Abandoned?</td>
</tr>
</tbody>
</table>
How to Collect and Monitor Data

- Stream has two deployment architectures and two collection methodologies

**Deployment:**
- Out-of-band (stub) with tap or SPAN port
- In-line directly on monitored host

**Collection:**
- Technical Add-On (TA) with Splunk Universal Forwarder (UF)
- Independent Stream Forwarder using HTTP Event Collector (HEC)
Deployment: Dedicated Collector

End Users → Internet → Firewall → TAP or SPAN → Servers

End Users

Search Head

Splunk Indexers

Linux Forwarder

Splunk_TA_Stream
Deployment: Run on Servers

- End Users
- Internet
- Firewall
- Physical or Virtual Servers
- Universal Forwarder
- Splunk_TA_stream
- Physical Datacenter, Public or Private Cloud
- Search Head
- Splunk Indexers
- End Users
Stream Forwarder Options

Makes it easy to add Stream anywhere in your environment

1. Stream TA

- Stream deploys as a modular input on top of your Splunk Forwarders.

2. Independent Stream Forwarder

- Stream deploys as a stand-alone binary and communicates via HEC.
- Requires >= Splunk 6.3.1
300+ Commercial Applications Detected 😊

- Adobe Flash Plugin Update
- Adobe Update Manager
- aim express
- AIM Transfer
- AllMusic.com
- Altiris
- Amazon Ad System
- Amazon Cloud Drive
- Amazon Generic Services
- Amazon MP3
- Amazon Video
- Amazon Web Services
- Cloudfront CDN
- Android connectivity Manager
- Aol Instant Messenger
- formerly OSCAR
- Apple AirPlay
- Apple Airport
- Apple AirPrint
- Apple App Store
- Apple FaceTime
- Apple Generic Services
- Apple HTTP
- Live Streaming
- Apple Location Map
- Apple Maps
- Apple Music
- Apple Push Notification Service
- Apple Siri
- Apple Update
- ASProxy
- Atlassian Background Intelligent Transfer Service
- Baidu Player
- Baidu_wallet
- Baidu.com
- Bet365
- Bitcoin
- client
- BitTorrent
- Bittorrent Apps
- BitTorrent Bleep
- (aka BitTorrent Chat)
- BlackBerry Locate
- BlackBerry Messenger
- BlackBerry Messenger Audio
- BlackBerry Messenger Video
- BlackBerry.com
- Border Gateway Protocol
- CARBONITE
- CProxy
- ChatON
- Chatroulette.com
- Cisco
- Unique
- Networking
- Cisco MeetingPlace
-Cisco Netflow
- Common Unix Printer System
- Crackle craigslist
- Data Stream Interface
- DB2
- Debian
- Ubuntu
- Update
- Dropbox
- Download
- Dropbox.com
- eBay.com
- Edonkey
- Evernote.com
- EverQuest
- Facebook
- Firefox
- Generic Routing
- Encapsulation
- GitHub
- Gmail
- Basic
- Gmail
- Mobile
- GNUnet
- Gnusellla
- Google Analytics
- Google App Engine
- Google Cache
- Google Calendar
- Google Chat
- Google Cloud
- Google Cloud Storage
- Google Documents
- (aka Google Drive)
- Google Earth
- Google Generic
- Google groups
- GStatic
- Google Maps
- Google Picasa
- Google Play Music,
- Google Play Music
- Google Play Store
- Google Plus
- Google Safe Browsing
- Google Tag Manager
- Google Toolbar
- Google Translate
- .com
- GoTo Mystic
- Remote Management
- Online Meeting
- GoToMyPC
- Remote Access
- GPRS Tunneling Protocol
- GPRS Tunneling Protocol
- version 2
- Half-Life
- HiS
- High Entropy
- Hot Standby
- Router Protocol
- HP Printer Job
- Language
- Hulu
- HyperText Transfer Protocol
- version 2
- HTTP/2
- I2P
- Invisible Internet Project
- IBM
- Informix
- IBM Lotus
- Sametime
- IBM SmartCloud
- IBM Websphere
- MQ
- (Cloud (Apple)
- iHeartRADIO
- iMessage
- File
- Download
- Imgur.com
- Independent Computing Architecture
- (Citrix)
- Internet Group Management Protocol
- Internet Printing Protocol
- Internet Security Association
- and Key Management Protocol
- Internet Small Computer Systems Interface
- IOS over the-air (OTA)
- update
- IP Payload Compression
- Protocol IP-in-IP
- tunneling
- IPSec
- Encapsulating
- Security Payload
- IRC File
- Transfer Data
- iTunes
- Jabber File Transfer
- Java Update
- JEDI
- (Citrix)
- Kazaa (FastTrack protocol)
- KIK Messenger
- King Digital Entertainment
- LinkedIn.com
- Live
- hotmail for mobile
- Livestream.com
- LogMeln
- Rescue magicJack
- Mail.ru
- Agent Maxtoob
- mail
- Media Gateway Control Protocol
- Message Session Relay Protocol
- Microsoft ActiveSync
- Microsoft Lync
- Microsoft Lync Online
- Microsoft Office
- 365
- Microsoft Remote Procedure Call
- Microsoft Service Control
- Microsoft SharePoint
- Microsoft SharePoint Administration
- Application
- Microsoft SharePoint Blog
- Management Application
- Microsoft SharePoint Calendar Management Application
- Microsoft SharePoint Document Management Application
- Microsoft Multi Protocol Label Switching
data-carrying mechanism
- Nagios
- Remote Data Processor
- Nagios Remote Plugin
- Executive Name Service
- Provider Interface
- Netflikx
- .com
- NetMeeting
- ILS
- Network Time Protocol
- Nintendo
- Wi-Fi Connection
- Nortel/SynOptics
- Network Management Protocol
- OkCupid
- Online Certificate
- Status Protocol
- Oovoo
- Open Shortest Path First
- Opera Update
- Orkut.com
- Outlook
- Web Access
- (Office 365)
- Outlook Web App
- PalTalk
- Paltalk
- audio
- chat
- PalTalk Transfer Protocol
- Paltalk video
- Pandora Radio
- Pastebin
- patching
- PCAnywhere
- Photobucket.com
- Pinterest.com
- PlayStation Network
- Plenty Of Fish
- QIK Video
- QQ File Transfer
- QQ Games
- QQ Mail
- QQ Weibo
- QQ.com
- QDownload
- QQlive
- Network Player
- QQMusic
- QQStream
- Quake quic
- QVOD
- Player RapidShare.com
- Real Time Streaming Protocol
- Remote Desktop Protocol
- (Windows Terminal Server)
- Remote Procedure Call
- RetroShare Routing Information Protocol
- V1 Routing Information Protocol
- V2 Routing Information Protocol
- ng1
- Rovio Entertainment
- RSS
- Salesforce.com
- SAP
- SecondLife.com
- Secure
- Shell Session
- Traversal Utilities for
- NAT Sharepoint
- Online Silverlight
- (Microsoft Smooth Streaming)
- Simple Object Access Protocol
- Skinny Client Control Protocol
- Slacker Radio
- Slingbox
- Snapchat
- SOCket Secure
- v5
- SoMud BitTorrent tracker
- SoundCloud SourceForge
- SPDY
- SquirrelMail
- Steampowered
- Symantec Norton AntiVirus
- Updates Syslog
- Systems Network Architecture
- TeamSpeak v2
- TeamSpeak v3
- TeamViewer
- Telnet
- Teroedo protocol
- Terminal Access Controller
- Access-Control System
- Plus TIBCO RendezVous Protocol
- Tor2web
- Tumblr
- Twitch
- Tumblr
- Ustream
- UTorrent
- (Micro Transport Protocol)
- UUSee protocol
- VEVO
- Viber
- Vimeo.com
- Vine
- Virtual Router
- Redundancy Protocol
- VMare
- Vmware
- horizon_view
- Waze
- Social GPS Maps
- & Traffic WebEx
- WhatsApp
- Messenger
- WHOIS
- WiiConnect24
- Wikipedia.com
- Windows Azure
- CDN
- Windows Internet Naming Service
- Windows Live File Storage
- Windows Live Groups
- Windows Live Hotmail
- Windows Live Hotmail Attachments
- Windows Live SkyDrive
- Windows Live SkyDrive Login
- Windows Marketplace
- Windows Update
- WordPress.com
- World of Warcraft
- XB:Live
- Xbox Live
- Marketplace
- Xbox Music
- Xbox Video
- (Microsoft Movies and TV)
- xHam.com
- Yahoo groups
- Yahoo mail
- classic
- Yahoo Mail
- v.2.0
- Yahoo Messenger
- Yahoo Messenger conference
- service
- Yahoo Messenger Transfer Protocol
- Yahoo Messenger Video
- Yahoo Search
- Yahoo webmail
- for mobile
- Yahoo Webmessenger
- Yahoo.com
- Yellow Page
- Bind
- Yellow Page
- Password
- Yellow Pages
- Server
- Youtube.com

16
Analyzing Wire Data using Splunk

- Once data is collected, use normal SPL to generate dashboards to analyze data
- Pros:
  - Flexible and Detailed analytics
  - Simple drill-downs to raw data
- Cons:
  - SPL is hard!
  - Long Time-to-Value
Splunk ITSI

- IT Service Intelligence solution - Design “Services” comprised of KPIs
- Specifies, Extracts, and Monitors Key Performance Indicators (KPIs) in your Splunk data
- Glass tables allows you to visualize KPIs on an intuitive layout with color changes and other indicators
- Deep Dives illustrate time-series correlations between KPIs to show elements at fault or temporal causality
- Built in Anomaly Detection with Machine Learning allows
ITSI for Application and Network Monitoring

- **Stream + ITSI = APM / NPM ??**
- Not quite, but many of the elements
  - Monitor the Application
    - Response Time
    - Connections
    - Response Codes
    - Simultaneous
  - Monitor the Network
    - Bandwidth
    - Latency
    - Extraneous Traffic
Automatic Anomaly Detection with ITSI

- How do you know when something has gone wrong?
- Typically, anomaly detection takes much tuning, is delicate
- ITSI automatically learns the patterns in your data so anomaly detection happens automatically too!
- Dynamic Machine Learning – not simple static thresholds
Demo
Art of the Possible

What else can we do with Stream + ITSI?

- Automatically alert when a new Commercial Application is seen on the network and track the source user
- Dynamically let you know when an unusual number of customer transactions (high or low) are occurring on your eCommerce platform
- Alert when the delay accessing a critical server exceeds a threshold, and then help isolate the source of the trouble (network, host, storage, etc.)
THANK YOU