

FN1206: The Path to Operational Enlightenment

An Introduction to Wire Data



Simon O'Brien

Principal Sales Engineer | Splunk



Vinu Alazath

Software Engineer | Splunk

Forward-Looking Statements



During the course of this presentation, we may make forward-looking statements regarding future events or plans 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 may differ materially. 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, it may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements made herein.

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 functionalities described or to include any such feature or functionality in a future release.

Splunk, Splunk>, Turn Data Into Doing, The Engine for Machine Data, Splunk Cloud, Splunk Light and SPL are trademarks and registered trademarks of Splunk Inc. in the United States and other countries. All other brand names, product names, or trademarks belong to their respective owners. © 2019 Splunk Inc. All rights reserved.

Agenda

Challenges

Product Overview

Architecture and Deployment

Demo

Stream futures





Challenges

Challenges

A large orange circle containing text.

Lack of
Application
Visibility Impacts
Customer
Experience

A large yellow circle containing text.

Limited Cloud
Insights

A large pink circle containing text.

Long MTTR
Hurts the
Business


Challenges

A large orange circle containing text.

Lack of
Application
Visibility Impacts
Customer
Experience

A large yellow circle containing text.

Limited Cloud
Insights

A large pink circle containing text.

Long MTTR
Hurts the
Business

Splunk Stream Real-Time Intelligence

Lack of
Application
Visibility Impacts
Customer
Experience

Limited Cloud
Insights

Long MTTR
Hurts the
Business

splunk®



Product Overview

What's Wire Data?

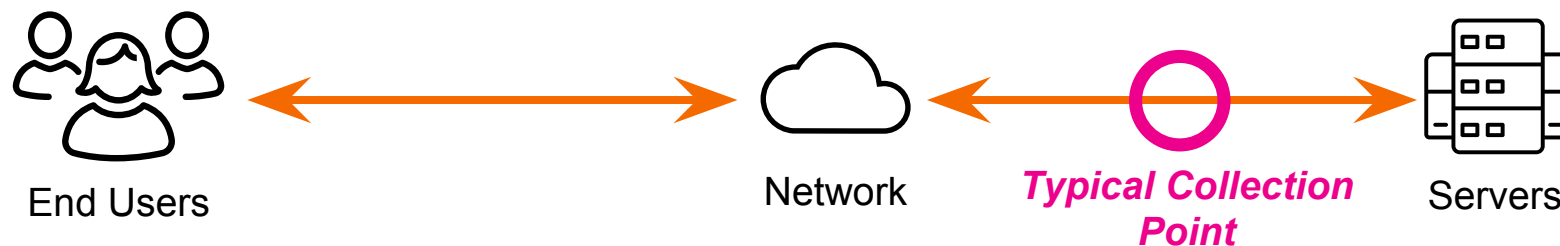
```
tcpdump -qns 0 -A -r blah.pcap
20:57:47.368107 IP 205.188.159.57.25 > 67.23.28.65.42385: tcp 480
0x0000: 4500 0214 834c 4000 3306 f649 cdbc 9f39  E....L@.3..l...9
0x0010: 4317 1c41 0019 a591 50fe 18ca 9da0 4681  C..A....P.....F.
0x0020: 8018 05a8 848f 0000 0101 080a ffd4 9bb0  .....
0x0030: 2e43 6bb9 3232 302d 726c 792d 6461 3033  .Ck.220-rlly-da03
0x0040: 2e6d 782e 616f 6c2e 636f 6d20 4553 4d54  .mx.aol.com.ESMT
0x0050: 5020 6d61 696c 5f72 656c 6179 5f69 6e2d  P.mail_relay_in-
0x0060: 6461 3033 2e34 3b20 5468 752c 2030 3920  da03.4;.Thu,.09.
0x0070: 4a75 6c20 3230 3039 2031 363a 3537 3a34  Jul.2009.16:57:4
0x0080: 3720 2d30 3430 300d 0a32 3230 2d41 6d65  7.-0400..220-Ame
0x0090: 7269 6361 204f 6e6c 696e 6520 2841 4f4c  rica.Online.(AOL
0x00a0: 2920 616e 6420 6974 7320 6166 6669 6c69  ).and.its.affili
0x00b0: 6174 6564 2063 6f6d 7061 6e69 6573 2064  ated.companies.d
```

Network Conversations

Machine data

Poly-structured data

Authoritative record of real-time and historical communication between machines and applications



How Will Wire Data Help Solve the Problem?

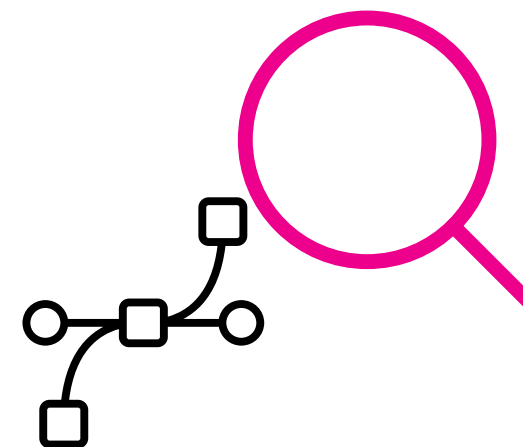
Wire data represents capture of true conversations between endpoints

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



Why Splunk Stream?

Traditional Wire Data flow-type records (such as NetFlow) generally contains only IP addresses and TCP or UDP ports.

While this can show host-host connections, it doesn't give any insight about the content of those conversations (like telephone call records)

Splunk Stream parses wire data all the way up the stack and generates Events with information at every level (more akin to a written transcript of a phone call)

Flow-type Data

7. Application
6. Presentation
5. Session

4. Transport
3. Network

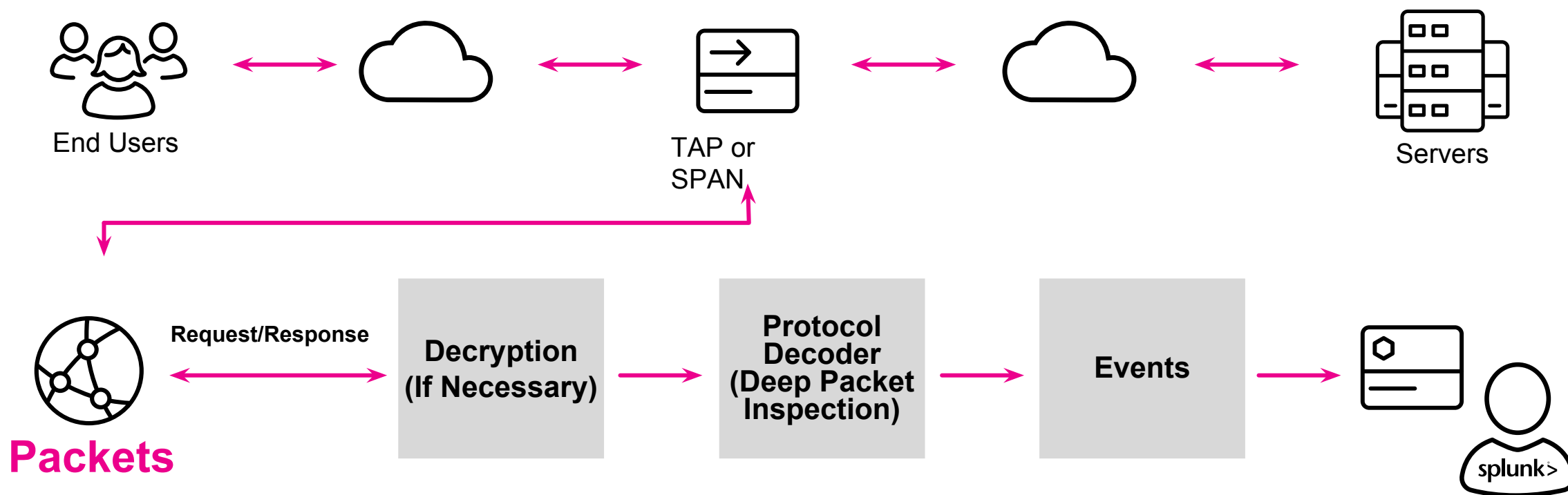
2. Data Link

1. Physical

Splunk Stream

7. Application
6. Presentation
5. Session
4. Transport
3. Network
2. Data Link
1. Physical

Wire Data Collection / Metadata Generation



What's Available In Splunk Stream Data?

Performance Metrics

Round Trip Time
Client Request Time
Server Reply Time
Server Send Time
Total Time Taken
Base HTML Load Time
Page Content Load Time
Total Page Load Time

Application Data

POST Content
AJAX Data
Section
Sub-Section
Page Title
Session Cookie
Proxied IP Address
Error Message

Business Data

Product ID
Customer ID
Shopping Cart ID
Cart Items
Cart Values
Discounts
Order ID
Abandoned?

Splunk Stream

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)

File extraction capabilities ++ MD5 hashing

Aggregation Mode

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

Filtering at Endpoint

Out-of-Box Content

- Dashboards for common protocols

1GbE and 10GbE link options

- 10 GbE uses DPDK SDK (dpdk.org)

Protocols Parsed with Stream – to be updated?

Simple Transport

- TCP
- UDP
- IP

Infrastructure

- ARP
- DHCP
- SNMP
- DNS
- ICMP
- IGMP
- Netflow/sflow

File Transfer

- FTP
- HTTP

File Service

- NFS
- SMB

Email

- IMAP
- MAPI
- POP3
- SMTP

Messaging

- AMQP
- IRC
- SMPP
- XMPP

Authentication

- Diameter
- LDAP
- RADIUS

Database

- MYSQL
- Postgres
- TDS (Sybase / MS-SQL)
- TNS (Oracle SQL*Net)

VoIP

- SIP
- RTP

Commercial Application Detection

Add the many hundreds of applications to be detected to the TCP stream type existing “app” field

Help diagnose the problem of “what is going over port 80”? And also “what’s taking all of my bandwidth?”

DOES NOT PARSE applications, simply detects them

- Will detect encrypted protocols!
- Will detect vendor-proprietary protocols!
- Uses empirical patterns, DNS, Cert CNs and other methods

Current feature supports 300+ applications, many more to be added

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 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 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.com 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 CCProxy ChatON Chatroulette.com Chrome Update Cisco Discovery Protocol Cisco MeetingPlace Cisco Netflow Common Unix Printer System Crackle craigslist Data Stream Interface DB2 Debian/Ubuntu Update Dropbox Download Dropbox Upload Dropbox.com eBay.com Edonkey Evernote.com EverQuest - EverQuest II Facebook Facebook Messenger FarmVille Find My iPhone Firefox Update Flickr Generic Routing Encapsulation GitHub Gmail Basic Gmail drive Gmail Mobile GNUnet Gnutella Google Accounts Google Analytics Google App Engine Google Cache Google Calendar Google Chat Google Cloud Messaging Google Cloud Storage Google Documents (aka Google Drive) Google Earth Google Generic Google groups Google GStatic Google Hangouts (formerly Google Talk) Google Mail Google Maps Google Picasa Google Play Music, Google Play Musique Google Play Store Google Plus Google Safe Browsing Google Tag Manager Google Toolbar Google Translate Google.com GoToDevice Remote Administration GoToMeeting Online Meeting GoToMyPC Remote Access GPRS Tunneling Protocol GPRS Tunneling Protocol version 2 Half-Life Hi5.com 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 iCloud (Apple) iHeartRADIO iMessage File Download Imgur.com Independant Computing Architecture (Citrix) Instagram 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 LogMeIn Rescue magicJack Mail.ru Agent Maktoob 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 Multi Protocol Label Switching data-carrying mechanism Nagios Remote Data Processor Nagios Remote Plugin Executor Name Service Provider Interface Netflix.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 Pastebin_posting PCAnywhere Photobucket.com Pinterest.com Playstation Network Plenty Of Fish QIK Video QQ QQ File Transfer QQ Games QQ Mail QQ WeiBo QQ.com QQDownload QQLive Network Player QQMusic QQStream Quake quick VOD 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 Internet 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 Spotify SquirrelMail Steampowered.com Symantec Norton AntiVirus Updates Syslog Systems Network Architecture Teamspeak v2 TeamSpeak v3 TeamViewer Telnet Teredo protocol Terminal Access Controller Access-Control System Plus TIBCO RendezVous Protocol Tor2web Tumblr Twitch Twitpic Twitter UStream uTorrent uTP (Micro Transport Protocol) UUSee Protocol VEVO Viber Vimeo.com Vine Virtual Router Redundancy Protocol VMWare 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 Xbox Live Xbox Live Marketplace Xbox Music Xbox Video (Microsoft Movies and Tv) xHamster.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 Passwd Yellow Pages Server Youtube.com

Example of Applications in Search

```
sourcetype=stream:* | stats count by app
```

amazon_aws	31	krb5	30
apple	5	live_hotmail	6
apple_location	2	norton_update	5
dhcp	6	ntp	2
facebook	6	ocsp	81
flickr	1	pinterest	1
google	58	skype	1411
google_analytics	4	smb	12
google_gen	29	spdy	4
google_safebrowsing	8	spotify	3
google_tags	3	teredo	15
gstatic	11	tumblr	28
http	7945	twitter	11
http2	11	yahoo	129
https	214	yahoo_search	1
icloud	8	ymsg_webmessenger	3
imgur	9	youtube	1



Architecture and Deployment && Best practices

Collect and Monitor Data with Stream

Stream has two deployment architectures and two collection methodologies

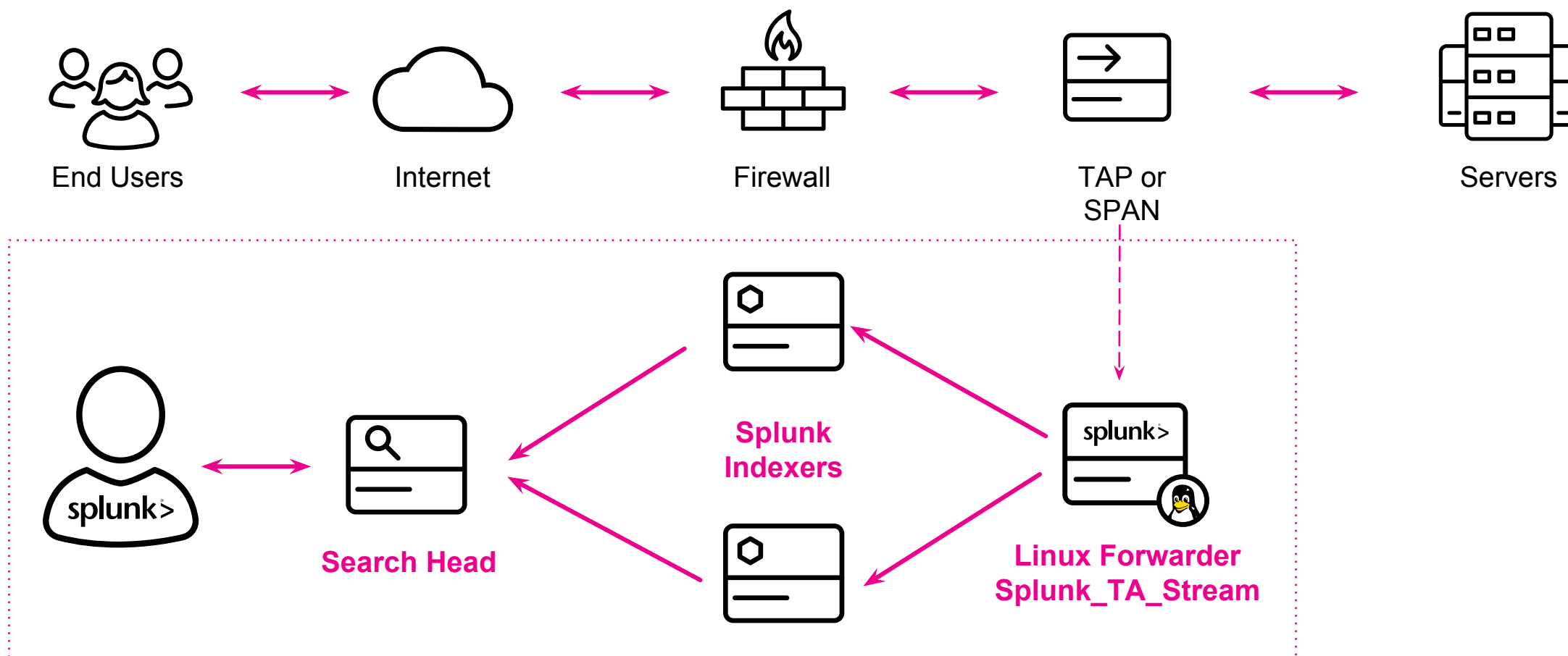
Deployment:

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

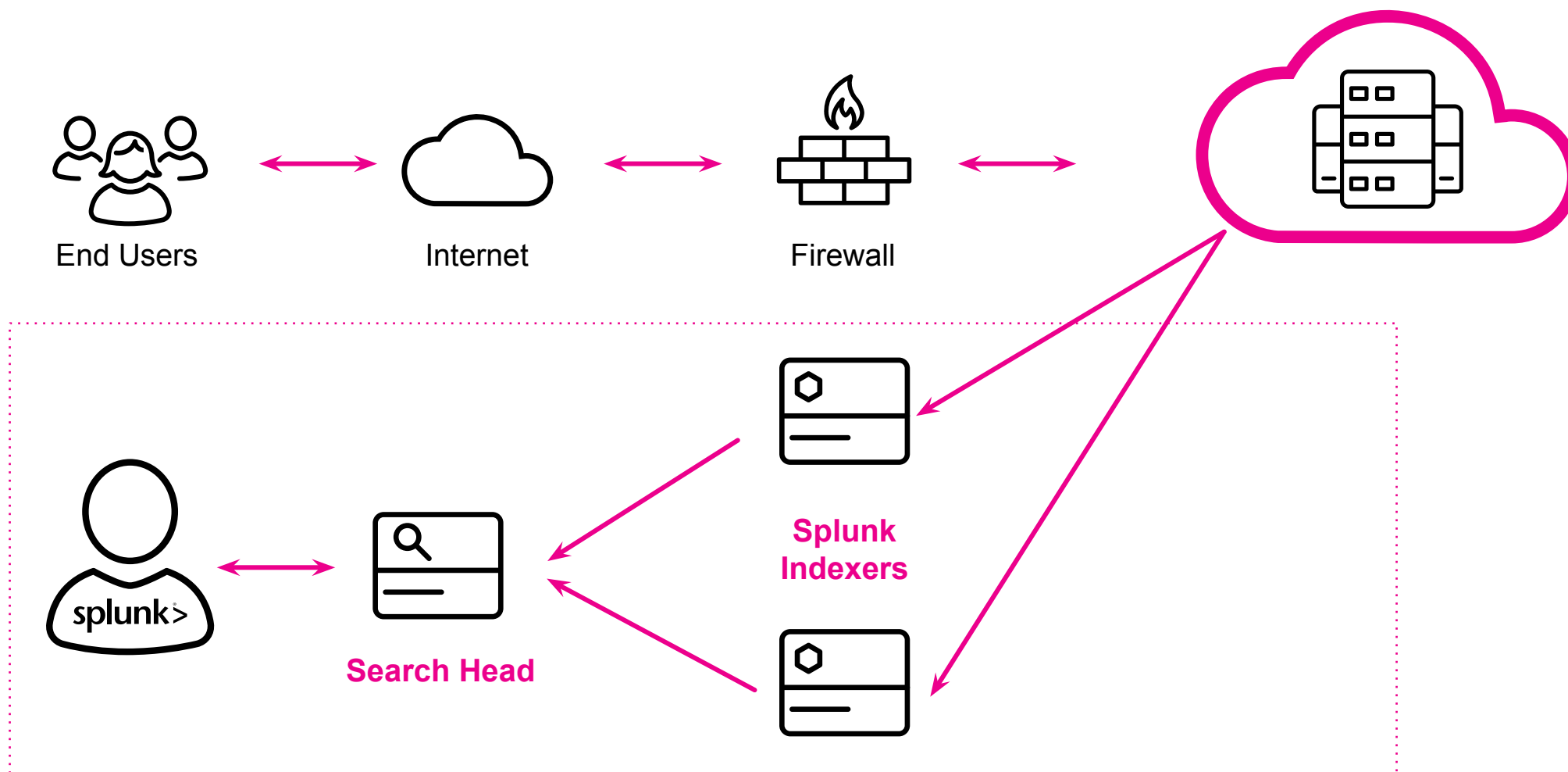
Collection:

- Technical Add-On (TA) with Splunk Universal Forwarder (UF)
- Independent Stream Forwarder using HTTP Event Collector (HEC)

Deployment: Dedicated Collector



Deployment: Run on Servers

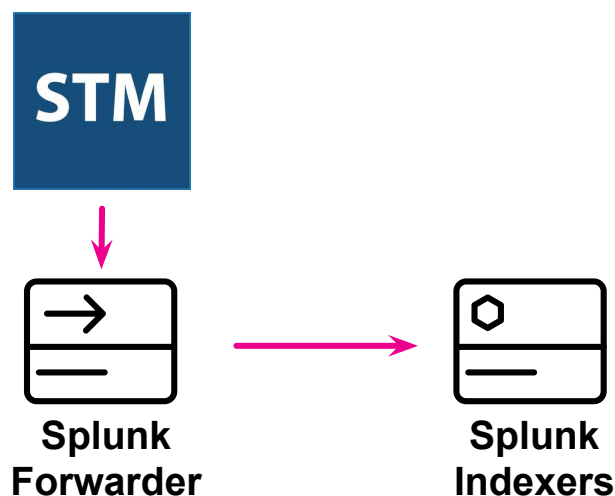


Stream Collection Options

Makes it easy to add Stream anywhere in your environment

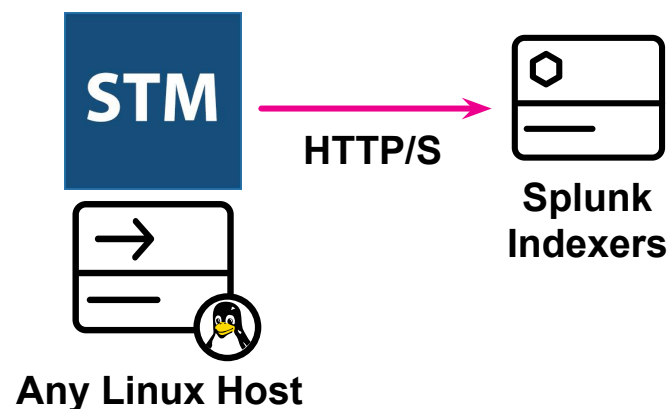
1. Stream TA

- 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 \geq Splunk 6.3.1

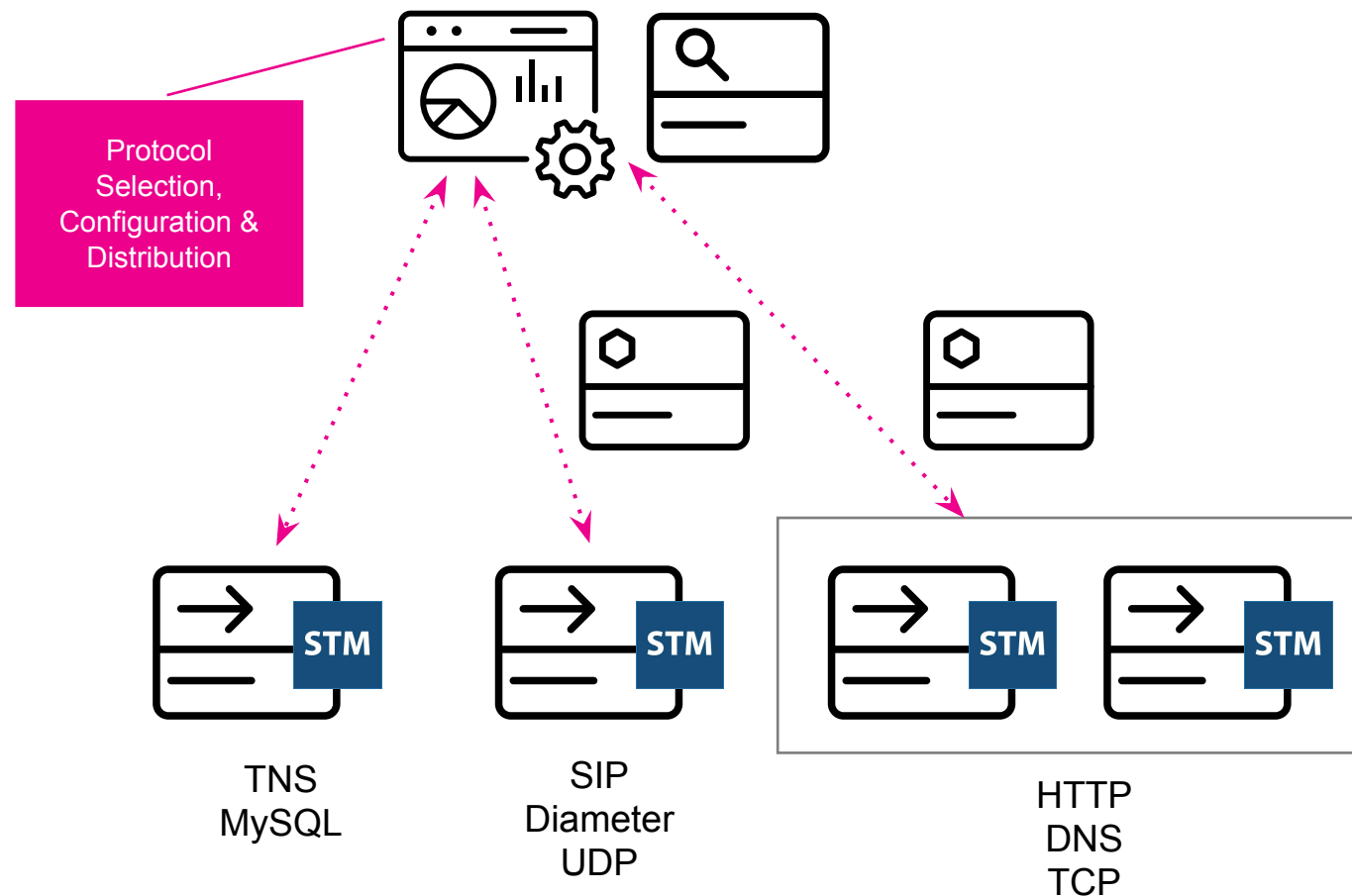


Distributed Forwarder Management

Gain more deployment flexibility

Increase management efficiency with per-forwarder protocol control

Tailor data collection by assigning different sets of protocols to groups of forwarders



Demo!



Roadmap and futures

Organizational Challenges



Lack of
Application
Visibility Impacts
Customer
Experience

New protocol library –
extra application
detection and prioritised
full protocol decoding



Limited Cloud
Insights

Support for AWS
VPC port mirroring –
via VXLAN
decapsulation
support



Long MTTR
Hurts the
Business

Enhanced DPDK
support – Broader set
of NIC capture

Call to action

1. Download Splunk Essentials for Wire Data!
2. Please stay and chat or come to the booth after the session!



Tell us what you think or what you want!

ponypoll.com/wiredata





Q&A



Simon O'Brien | Principal Sales Engineer
Vinu Alazath | Senior Engineer



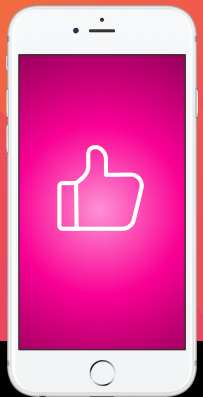
Thank

You



Go to the .conf19 mobile app to

RATE THIS SESSION





Appendix

Takeaway slides



FAQ and Summary

FAQ

Can I limit the amount of data collected with Stream?

- Yes. The app enables capture of only the relevant network/wire data for analytics, through filters and aggregation rules
- Select or deselect protocols and associated attributes with fine-grained precision within the app interface

How can I estimate my indexing volume?

- Data volume can vary based upon the number of selected protocols, attributes and the amount of network traffic
- Use Stream Estimate to understand the indexing impact

Where is Stream typically installed?

- Stream can be installed on any physical or virtual host running supported OS, on premises or in the cloud
- It can be installed off of TAP and SPAN ports
- It can be deployed in combination with TAP aggregation or visibility switches



Stream Capabilities

This thing looks awesome – what can it actually do?

Stream Capabilities

NetFlow Collector

- NetFlow v5, v9 (with template support), IPFIX (with vendor extensions)

Flow Visualization for all IPv4 space

PCAP Upload via SH and Continuous Directory Monitoring via Forwarder

Configuration Templates

- Easier integration with other Splunk products

Stream Capabilities

Splunk Stream 7.1 was released GA in April 2018 – 7.1.3 in April 2019. Stream 7.2 – Oct 2019

New targeted packet capture

- Supports capture of full network packets.
- Configure packet streams based on targeted protocol fields.
- Enables search against raw packet data.

New file extraction for metadata streams

- Extract content files from network traffic, including email attachments, images, pdfs, etc.
- Supports search for extracted files.

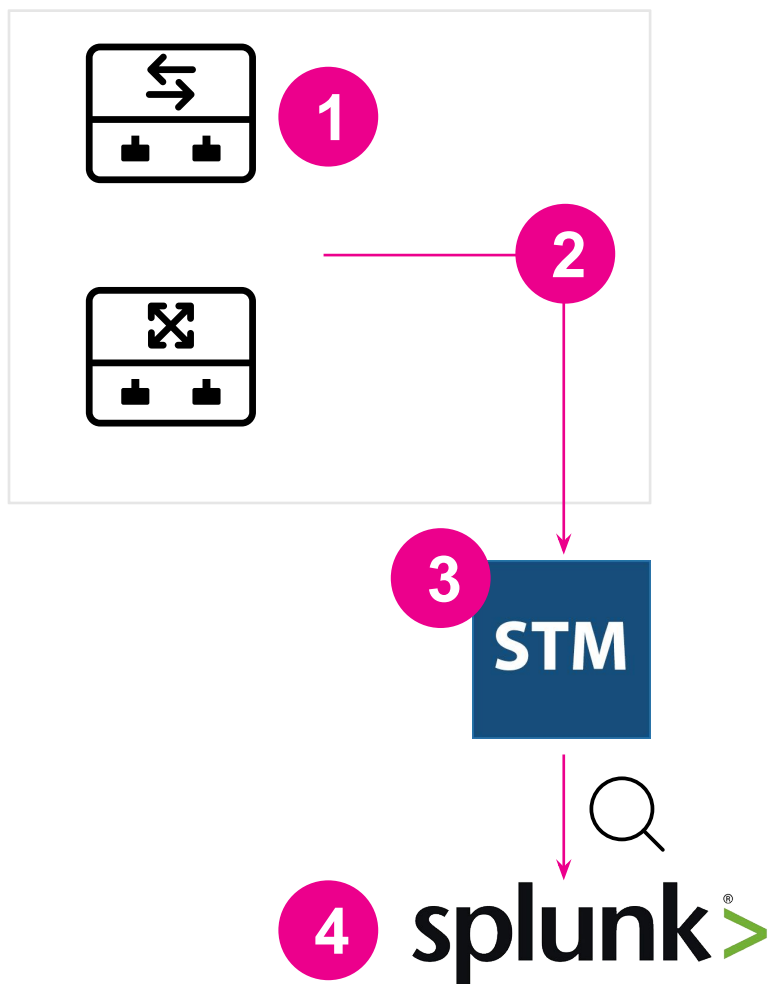
New SQL query parsing

- Capture SQL statement elements, including table names and SQL commands.
- New fields added to all database protocols.
- Use for fine-grained analysis of database activity.

Enhanced protocol support

- New IGMP transport layer (Layer 4) multicast protocol support.
- New RTCP protocol support adds qualitative analysis capabilities for streaming media and VoIP services.

Flow Collector Data Flow



- 1 NetFlow enabled devices
- 2 Export NetFlow (over UDP)
- 3 NetFlow Metadata captured by Stream
- 4 Events in Splunk Indexer / Search Head

NetFlow Collector

- NetFlow listening sockets (UDP ports)
- Actively capture Flows from NetFlow v5, v9, IPFIX
- Creates Splunk-compatible Flow Records
- Management from Stream Centralized UI

Flow Collection

Active Flow listening socket on Stream Forwarder

Flexible Configuration Options

- Selectable fields and filtering
- Can configure multiple, distinct listening ports on each Stream Forwarder

Supports most common versions of Flow protocols

- Cisco NetFlow, Juniper jFlow, HP sFlow, cFlowd
- NetFlow v5, v9, IPFIX

- V9 with templates (standard and custom)
- IPFIX with vendor extensions

Aggregation of Flow records (pre-indexing) can dramatically reduce the number of Splunk Events created

Performance > 465,000 flows/second (on a single Independent Stream Forwarder)

NetFlow and sFlow Streams UX

The screenshot shows the 'Configure Stream - netflow' page in the Splunk App for Stream. The page has a dark blue header with the Splunk logo and navigation links. Below the header, there's a sub-header 'Configure Stream - netflow' with a 'Netflow Protocol Events' subtitle. On the right, there are buttons for 'Clone', 'Delete', 'Cancel', and 'Save'. The main configuration area includes a 'Mode' section with 'Enabled', 'Estimate', and 'Disabled' buttons. A message states: 'To enable receiving NetFlow or sFlow, you must specify the netflowReceiver settings in streamfwd.conf on each Stream Forwarder. Read more'. Below this, there's a 'Splunk Index' dropdown set to 'default', a 'Protocol' dropdown set to 'Netflow', and an 'Aggregation' section with 'No', 'Yes, every', and a text input field for 'seconds'. At the bottom, there's a section for 'Fields (130 enabled)' and 'Filters (0 filters)'. A search bar is present with the text 'Enable the fields to collect events on.' and an 'Extract New Field' button. Below this is a table of fields.

Enable	Name	Description	Type	Actions
<input checked="" type="checkbox"/>	app	Specifies the name of an application	Original	Edit
<input checked="" type="checkbox"/>	app_tag	Application Id	Original	Edit
<input checked="" type="checkbox"/>	bgp_next_hop_ip	IP address of the next (adjacent) BGP hop	Original	Edit
<input checked="" type="checkbox"/>	bytes	Total number of Layer 3 bytes in the flow	Original	Edit
<input checked="" type="checkbox"/>	bytes_in	Incoming counter for number of bytes associated with an IP Flow	Original	Edit
<input checked="" type="checkbox"/>	bytes_out	Outgoing counter for number of bytes associated with an IP Flow	Original	Edit
<input checked="" type="checkbox"/>	channel	Identifier of the 802.11 (Wi-Fi) channel	Original	Edit
<input checked="" type="checkbox"/>	dest_ip	Destination address of flow	Original	Edit
<input checked="" type="checkbox"/>	dest_ip_prefix	Destination address prefix	Original	Edit

MD5 Hashing of Files

File Hashing provides integrity verification of files, can be used for a number of security use cases

- inbound malware detection
- outbound data loss prevention

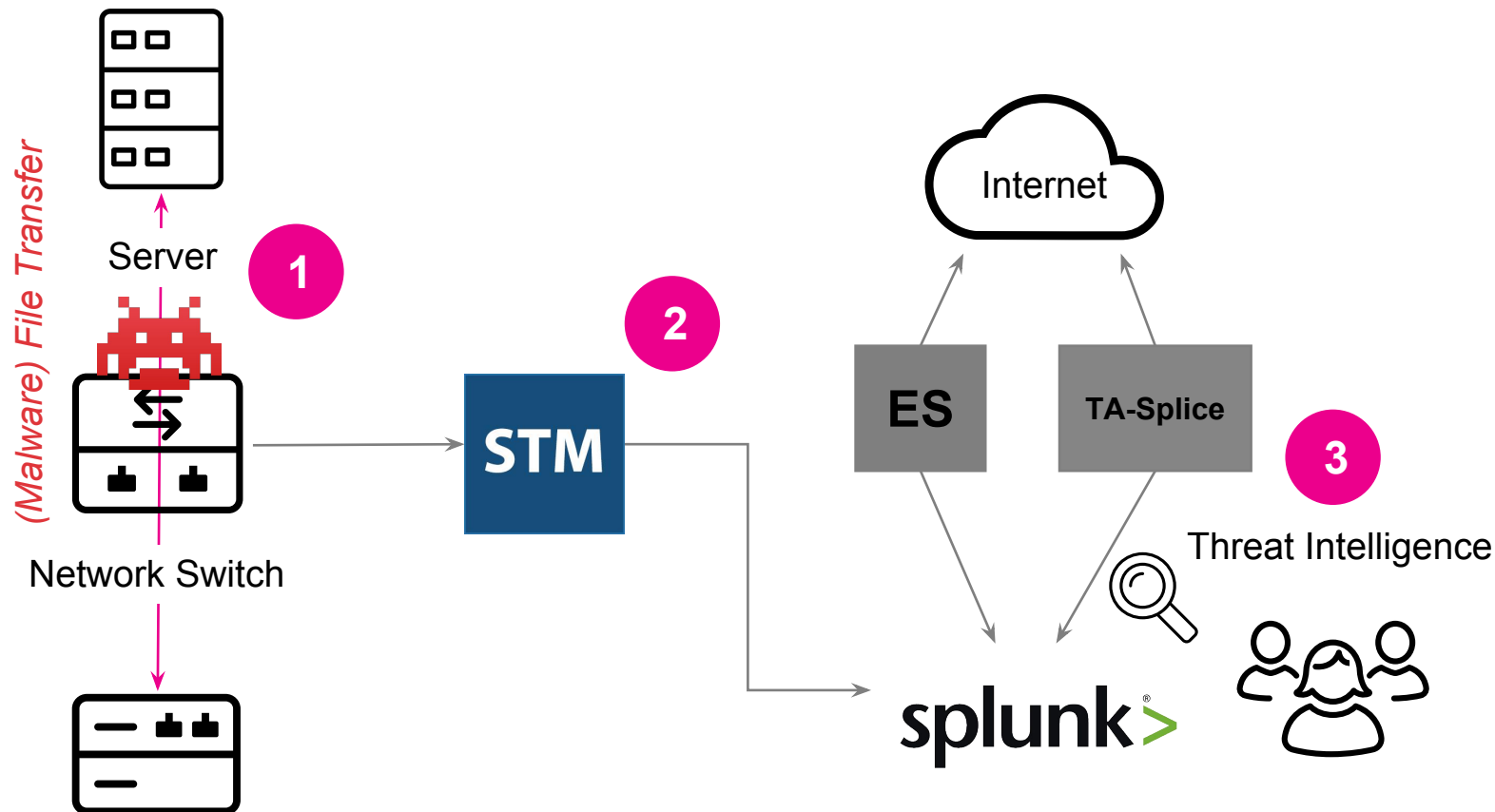
Stream generates MD5 hashes equivalent to “md5sum” unix command after decoding content back to binary

Specifically **for SMTP file attachments** and **HTTP**

MD5 hashes generated with Stream integrate directly into the Threat Intelligence framework of Enterprise Security, and has been tested with ES

As a bonus, **any** non-numeric field can be MD5 hashed using the “Extract New Field” option. Field can be length-truncated if desired

MD5 Hashing Data Flow



MD5 hashing

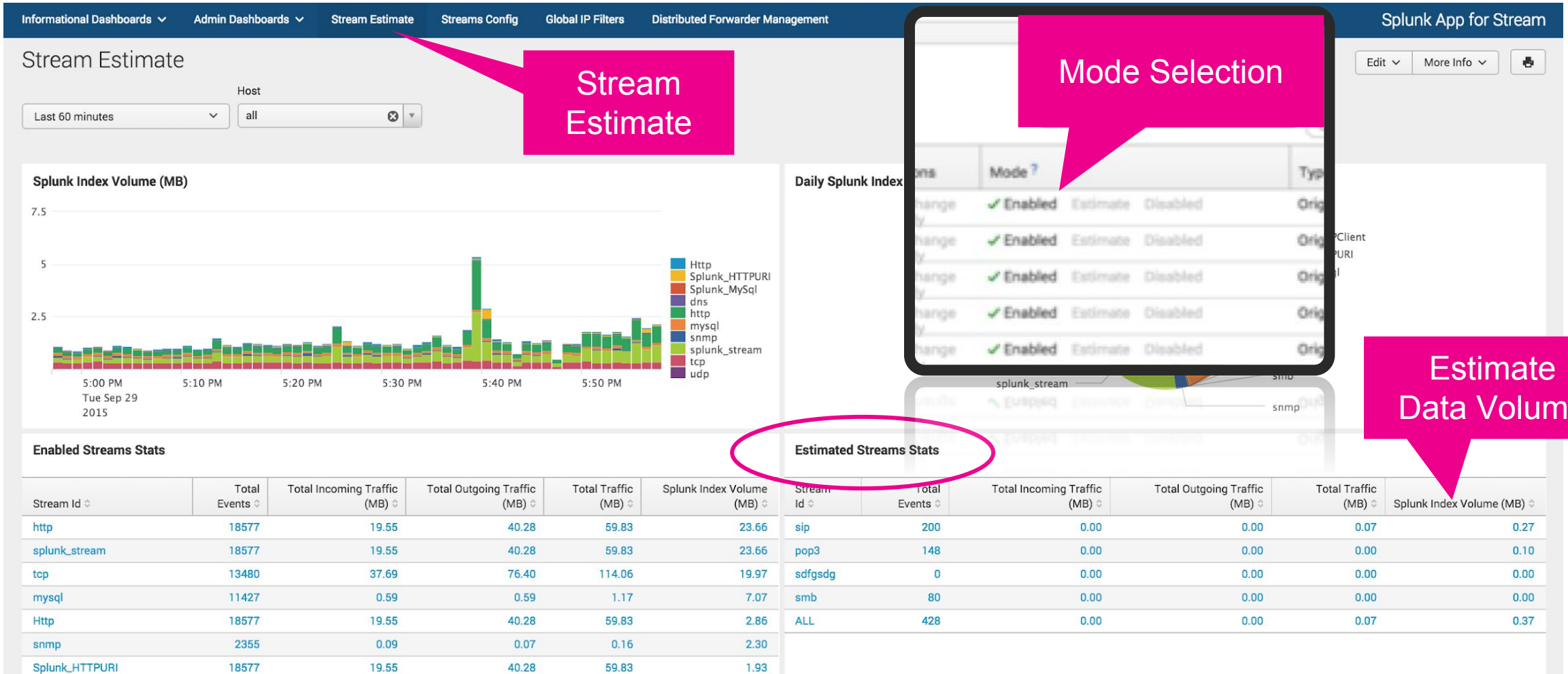
- Used to enable DLP and Security use cases
- Examines both inbound and outbound data transfer
- Can be used to find IOCs as well as data exfiltration
- Better metric than file names or file types

1 File Transfer Traffic between Client and Server directed towards Stream

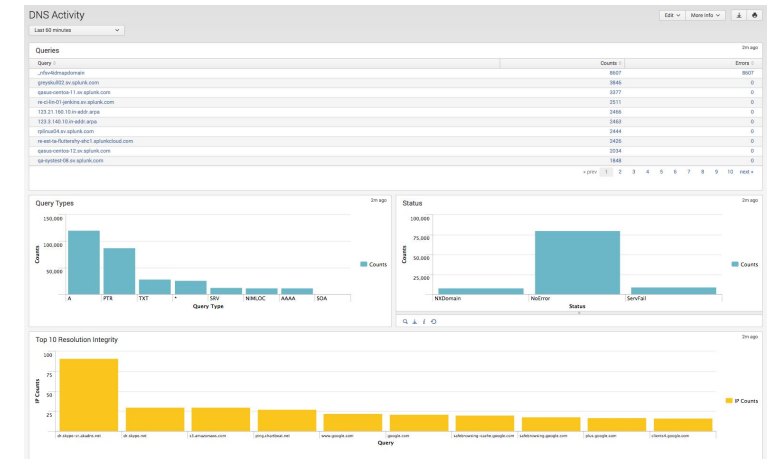
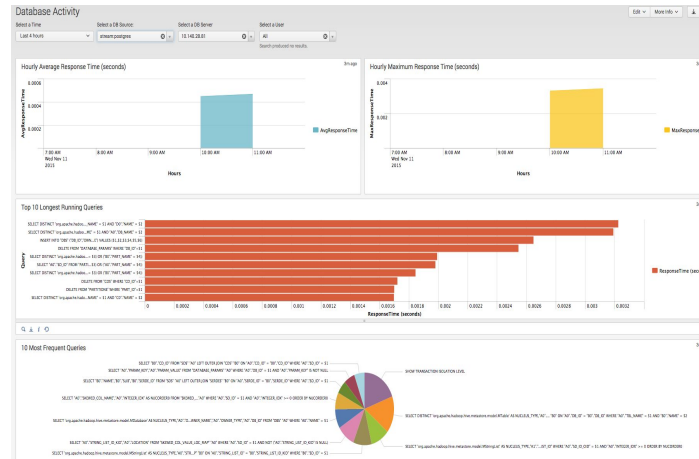
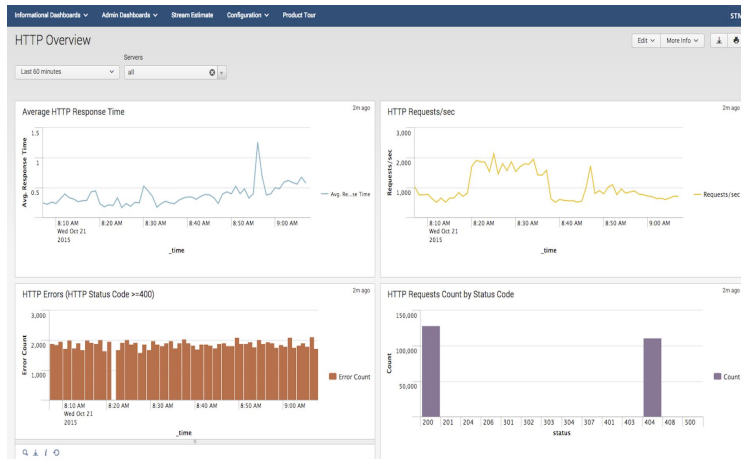
2 Stream generates MD5 hashes of files, sends to Splunk Indexers

3 MD5 hashes compared against Threat Intel from public databases

Tailor Data Collection to Your Monitoring Needs



Prebuilt Reporting



Get visibility into
applications performance
and user experience

Understand database
activity and performance
without impacting
database operation

Improve security and
application
intelligence with DNS
analytics

OSI Stack Model

Open Systems Interconnect (OSI) model

Published in 1984 by ISO and CCITT (now ITU-T)

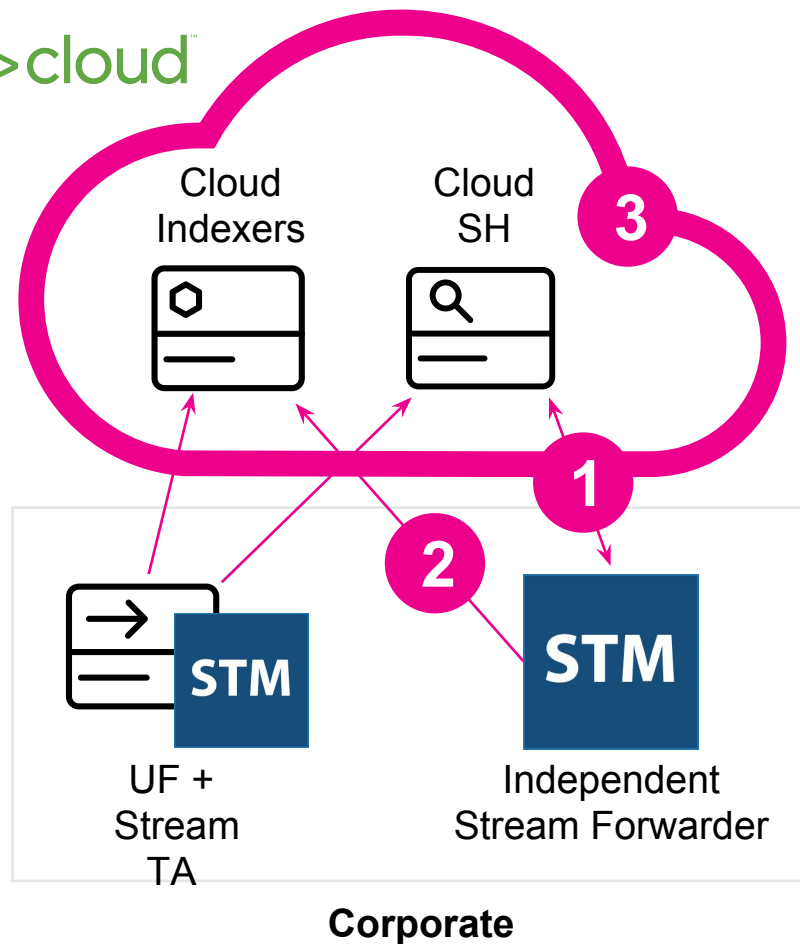
Forms the basis for all modern network communication models

Hierarchical messages encapsulated as they go down the stack, and get decapsulated as they go up the stack

Layer	Examples
7. Application	HTTP, SMTP
6. Presentation	TLS
5. Session	SCP
4. Transport	TCP, UDP
3. Network	IPv4, IPv6
2. Data Link	Ethernet
1. Physical	Ethernet, WiFi

Splunk Cloud Support for Stream

splunk>cloud



1

Stream forwarders fetch their configuration from the Cloud SH (authenticated)

2

Stream sends metadata back to Cloud indexers via the UF or HEC

3

Analysts connect to Cloud SH to explore the data collected by Stream