The bridge to possible

Diagnose Network Issues from Telemetry Data with AI-ML Methods

Frank Brockners, Distinguished Engineer,
Outshift by Cisco

BRKNWT-2404

cisco Live!

Cisco Webex App

https://ciscolive.ciscoevents.com/ciscolivebot/#BRKNWT-2404

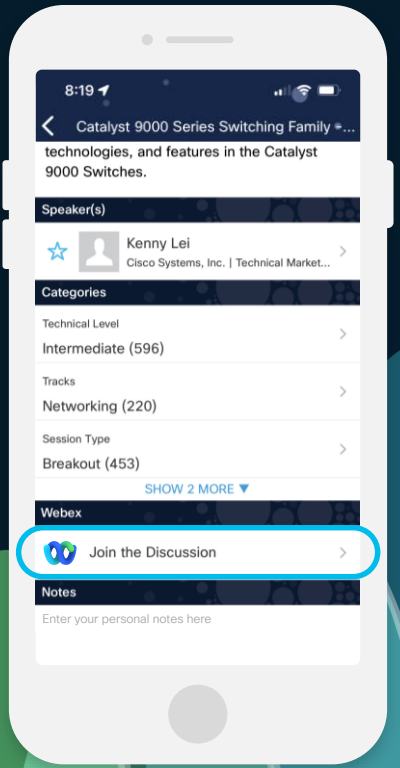
Questions?

Use Cisco Webex App to chat with the speaker after the session

How

1. Find this session in the Cisco Live Mobile App
2. Click “Join the Discussion”
3. Install the Webex App or go directly to the Webex space
4. Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until June 7, 2024.



Outshift by Cisco is the **incubation** engine delivering what's next and new for Cisco: **Emerging** technologies that target new markets and **personas** to build **meaningful** businesses and achieve innovative results.

We turn ideas into awesome products – with solutions for modern cloud native applications, edge, quantum, and AI.

Innovation, full speed ahead

Products
Cloud Application Security | Generative AI Infra

Cisco
DevNet

Cisco
Research

Cisco
Open Source Program Office

Cisco
Responsible AI Committee

I mentioned about the ability to interact in a humanlike capability. One of the big opportunities for **AI is to replace services with software** and if that's the TAM that we're going after, the starting point is not hundreds of billions the starting point is **possibly tens of trillions**.[3]

2025: 40% of services engagements will include GenAI-enabled delivery, triggering a shift in human-delivered services for strategy, change, and training organizations[1]

2028: GenAI technology will be used for 35% of network configuration and troubleshooting activities, up from near zero in 2023[2]

1. Gartner: WW IT Spending
2. Gartner: Forecast AI Services 2023-27
3. Pat Grady, Sequoia, https://youtu.be/TDPqt7ONUCY?si=ikudM8J4hl42nTHH&t=160

Understanding the state of an IT system

- IT “entities” – applications, runtime-systems (like K8s), servers, network devices offer a lot of operational data,
  - Example: In addition to logs, contemporary routers offer more than 1,000,000 counters via model driven telemetry

- Change detection methods alert operators about changes on one or multiple devices
- Can we leverage the wealth of information available in a device to have it tell us in natural language what its state is?
- Can we treat logs and even feature-names in timeseries as natural language and leverage associated methods to interpret them?

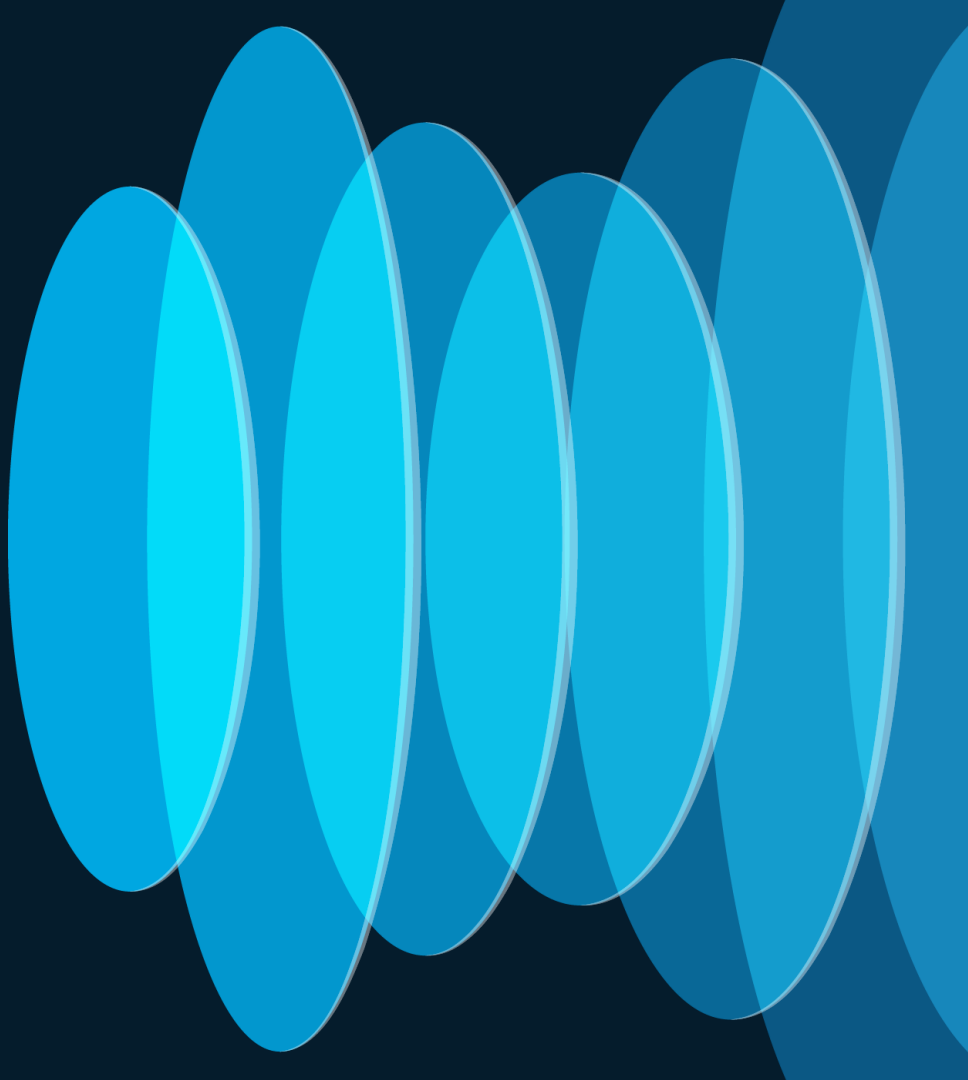
Agenda

Diagnosing cross-domain, multi-layer, multi-source telemetry with the help of AI

Diagnose streaming network telemetry with the help of AI

Diagnose logs with the help of AI

Diagnosing Logs with the help of AI



Imagine...

You are an IT Ops engineer and someone provided you with a log file “kube-flannel-ds-xd5wp.log” and asked:

“What is the issue?”

kube-flannel-ds-xd5wp.log

frank@CSCO-W-PF2XAQ7F: /r

```
I0315 19:05:00.332435       1 main.go:209] CLI flags config: {etcdEndpoints:http://127.0.0.1:4001,http://127.0.0.1:2379
etcdPrefix:/coreos.com/network etcdKeyfile: etcdCertfile: etcdCAFile: etcdUsername: etcdPassword: version:false kubeSubn
etMgr:true kubeApiUrl: kubeAnnotationPrefix:flannel.alpha.coreos.com kubeConfigFile: iface:[] ifaceRegex:[] ipMasq:true
ifaceCanReach: subnetFile:/run/flannel/subnet.env publicIP: publicIPv6: subnetLeaseRenewMargin:60 healthzIP:0.0.0.0 heal
thzPort:0 iptablesResyncSeconds:5 iptablesForwardRules:true netConfPath:/etc/kube-flannel/net-conf.json setNodeNetworkUn
available:true}
W0315 19:05:00.332497       1 client_config.go:618] Neither --kubeconfig nor --master was specified.  Using the inCluste
rConfig.  This might not work.
I0315 19:05:00.361865       1 kube.go:139] Waiting 10m0s for node controller to sync
I0315 19:05:00.361980       1 kube.go:461] Starting kube subnet manager
I0315 19:05:00.370473       1 kube.go:482] Creating the node lease for IPv4. This is the n.Spec.PodCIDRs: [10.244.0.0/24
]
I0315 19:05:00.370490       1 kube.go:482] Creating the node lease for IPv4. This is the n.Spec.PodCIDRs: [10.244.1.0/24
]
I0315 19:05:00.370510       1 kube.go:482] Creating the node lease for IPv4. This is the n.Spec.PodCIDRs: [10.244.2.0/24
]
I0315 19:05:01.362771       1 kube.go:146] Node controller sync successful
I0315 19:05:01.362790       1 main.go:229] Created subnet manager: Kubernetes Subnet Manager - xp02-vm21
I0315 19:05:01.362794       1 main.go:232] Installing signal handlers
I0315 19:05:01.362951       1 main.go:452] Found network config - Backend type: vxlan
I0315 19:05:01.363354       1 match.go:210] Determining IP address of default interface
I0315 19:05:01.363666       1 match.go:263] Using interface with name ens33 and address 172.16.0.90
I0315 19:05:01.363689       1 match.go:285] Defaulting external address to interface address (172.16.0.90)
I0315 19:05:01.363818       1 vxlan.go:141] VXLAN config: VNI=1 Port=0 GBP=false Learning=false DirectRouting=false
I0315 19:05:01.367882       1 kube.go:627] List of node(xp02-vm21) annotations: map[string]string{"flannel.alpha.coreos.
com/backend-data":"{\"VNI\":1,\"VtepMAC\":\"22:da:29:ff:7c:09\"}", "flannel.alpha.coreos.com/backend-type":"vxlan", "fla
nnel.alpha.coreos.com/kube-subnet-manager":"true", "flannel.alpha.coreos.com/public-ip":"172.16.0.90", "kubeadm.alpha.ku
bernetes.io/cri-socket":"unix:///var/run/containerd/containerd.sock", "node.alpha.kubernetes.io/ttl":"0", "volumes.kuber
netes.io/controller-managed-attach-detach":"true"}
--More--(0%)
```

You might...

- Browse through the file with your favorite editor
- Run a couple of grep commands using terms that you know are commonly hinting at issues

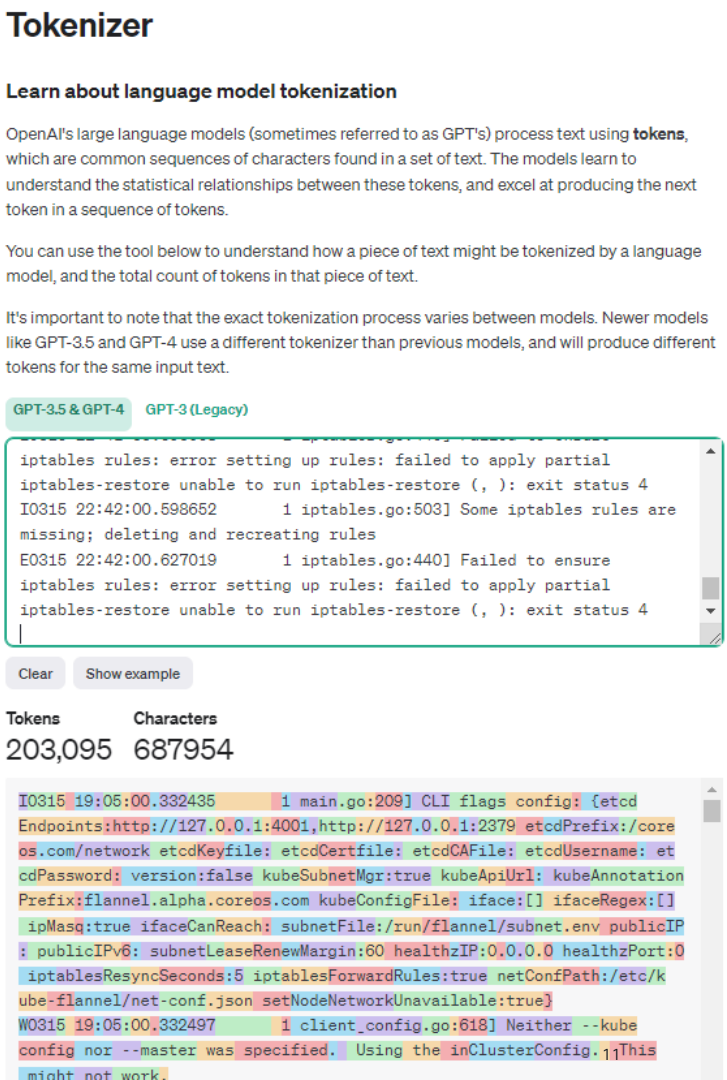
Could we delegate this task to AI?

Log files are often large...

The example file has 4490 lines, 89657 words and 687954 characters

```
$ wc kube-flannel-ds-xd5wp.log
 4490  89657 687954 kube-flannel-ds-xd5wp.log
```

... too large to fit into the context window of an LLM like e.g., GPT-4



We need to filter our input file...

- Different from “log anomaly detection”, our objective is only to “shrink” the dataset to a level, where we can fit things into a context window of an LLM – without losing the key information.
- Methods for “log anomaly detection” can be leveraged, e.g.,
  - Statistical approaches, using token frequencies (Detecting Anomalies in Logs by Combining NLP features with Embedding or TF-IDF, Log File reduction with TFIDF ranking, Group similar log-lines together using TF-IDF, ...)
  - Embedding-based approaches (e.g., LogBERT, LogBD, LSADNET, ...)

... Let’s explore a simple statistical method using TF-IDF to score.

Idea: “Watch out for an influx of rare things”

- **Intuition**: A term which is rare across all the document but frequent in a specific document/selection is more important
- **TF-IDF** (Term-Frequency, Inverse Document Frequency) allows us to score the importance of words in a “document”, based on how frequently they appear on multiple “documents”. TF-IDF is the product of two main statistics – term frequency and inverse document frequency
  - If the **word appears frequently in a document** – assign a **high score** to that word (term frequency – TF)
  - If the **word appears in a lot of documents** – assign a **low score** to that word (inverse document frequency – IDF)

TF-IDF applied to our case

- Approach:
  - A line in the log-file represents a “document”
  - The entire log-file represents all documents (a.k.a. the corpus)
- Sort the lines based on highest per-word weight, i.e., sort the log file in decreasing order of rarity.
  - *“Rarely occurring informative lines at the top. Repetitive ‘everything-works’ at the bottom.”*
- Stemming (make words of the same stem count as one), or different tokenization approaches can improve the results

See also https://anadi-sharma.medium.com/declutter-log-files-using-tf-idf-transformation-358e8b61efa8

TF-IDF applied to kube-flannel-ds-xd5wp.log

Original log file:

```
I0315 19:05:00.332435       1 main.go:209] CLI flags config: {etcdEndpoints:http://127.0.0.1:4001,http://127.0.0.1:2379
etcdPrefix:/coreos.com/network etcdKeyfile: etcdCertfile: etcdCAFile: etcdUsername: etcdPassword: version:false
kubeSubnetMgr:true kubeApiUrl: kubeAnnotationPrefix:flannel.alpha.coreos.com kubeConfigFile: iface:[] ifaceRegex:[] ipMasq:true
ifaceCanReach: subnetFile:/run/flannel/subnet.env publicIP: publicIPv6: subnetLeaseRenewMargin:60 healthzIP:0.0.0.0
healthzPort:0 iptablesResyncSeconds:5 iptablesForwardRules:true netConfPath:/etc/kube-flannel/net-conf.json
setNodeNetworkUnavailable:true}
W0315 19:05:00.332497       1 client_config.go:618] Neither --kubeconfig nor --master was specified.  Using the
inClusterConfig.  This might not work.
I0315 19:05:00.361865       1 kube.go:139] Waiting 10m0s for node controller to sync
I0315 19:05:00.361980       1 kube.go:461] Starting kube subnet manager
I0315 19:05:00.370473       1 kube.go:482] Creating the node lease for IPv4. This is the n.Spec.PodCIDRs: [10.244.0.0/24]
I0315 19:05:00.370490       1 kube.go:482] Creating the node lease for IPv4. This is the n.Spec.PodCIDRs: [10.244.1.0/24]
I0315 19:05:00.370510       1 kube.go:482] Creating the node lease for IPv4. This is the n.Spec.PodCIDRs: [10.244.2.0/24]
I0315 19:05:01.362771       1 kube.go:146] Node controller sync successful
I0315 19:05:01.362790       1 main.go:229] Created subnet manager: Kubernetes Subnet Manager - xp02-vm21
I0315 19:05:01.362794       1 main.go:232] Installing signal handlers
I0315 19:05:01.362951       1 main.go:452] Found network config - Backend type: vxlan
I0315 19:05:01.363354       1 match.go:210] Determining IP address of default interface
I0315 19:05:01.363666       1 match.go:263] Using interface with name ens33 and address 172.16.0.90
I0315 19:05:01.363689       1 match.go:285] Defaulting external address to interface address (172.16.0.90)
I0315 19:05:01.363818       1 vxlan.go:141] VXLAN config: VNI=1 Port=0 GBP=false Learning=false DirectRouting=false
I0315 19:05:01.367882       1 kube.go:627] List of node(xp02-vm21) annotations: map[string]
string{"flannel.alpha.coreos.com/backend-data":"{\"VNI\":1,\"VtepMAC\":\"22:da:29:ff:7c:09\"}",
"flannel.alpha.coreos.com/backend-type":"vxlan", "flannel.alpha.coreos.com/kube-subnet-manager":"true",
"flannel.alpha.coreos.com/public-ip":"172.16.0.90", "kubeadm.alpha.kubernetes.io/cri-
socket":"unix:///var/run/containerd/containerd.sock", "node.alpha.kubernetes.io/ttl":"0", "volumes.kubernetes.io/controller-
managed-attach-detach":"true"}
I0315 19:05:01.367988       1 vxlan.go:155] Setup flannel.1 mac address to 22:da:29:ff:7c:09 when flannel restarts
W0315 19:05:01.378785       1 main.go:505] no subnet found for key: FLANNEL_SUBNET in file: /run/flannel/subnet.env
W0315 19:05:01.378801       1 main.go:540] no subnet found for key: FLANNEL_IPV6_SUBNET in file: /run/flannel/subnet.env
I0315 19:05:01.378813       1 iptables.go:64] Current network or subnet (10.244.0.0/16, 10.244.1.0/24) is not
one (0.0.0.0/0, 0.0.0.0/0), trying to recycle old iptables rules
I0315 19:05:01.405058       1 iptables.go:74] Setting up masking rules
I0315 19:05:01.407044       1 iptables.go:103] Setting up masking rules for
I0315 19:05:01.408320       1 iptables.go:372] generated 7 rules
I0315 19:05:01.409231       1 iptables.go:213] Changing def
I0315 19:05:01.411936       1 iptables.go:221] IPv
```

Line-sorted doc with TF-IDF metric:

```
033  I0315 19:05:01.434308       1 iptables.go:365] bootstrap done
034  I0315 19:05:01.438309       1 iptables.go:365] bootstrap done
035  I0315 19:05:01.446828       1 iptables.go:365] bootstrap done
015  I0315 19:05:01.363818       1 vxlan.go:141] VXLAN config: VNI=1 Port=0 GBP=false Learning=false DirectRouting=false
028  I0315 19:05:01.414661       1 main.go:400] Running backend.
010  I0315 19:05:01.362794       1 main.go:232] Installing signal handlers
013  I0315 19:05:01.363666       1 match.go:263] Using interface with name ens33 and address 172.16.0.90
019  W0315 19:05:01.378801       1 main.go:540] no subnet found for key: FLANNEL_IPV6_SUBNET in file: /run/flannel/subnet.env
018  W0315 19:05:01.378785       1 main.go:505] no subnet found for key: FLANNEL_SUBNET in file: /run/flannel/subnet.env
027  I0315 19:05:01.414648       1 main.go:396] Wrote subnet file to /run/flannel/subnet.env
023  I0315 19:05:01.408320       1 iptables.go:372] generated 7 rules
026  I0315 19:05:01.411371       1 iptables.go:372] generated 3 rules
032  I0315 19:05:01.425153       1 main.go:421] Waiting for all goroutines to exit
007  I0315 19:05:00.370510       1 kube.go:482] Creating the node lease for IPv4. This is the n.Spec.PodCIDRs: [10.244.2.0/24]
006  I0315 19:05:00.370490       1 kube.go:482] Creating the node lease for IPv4. This is the n.Spec.PodCIDRs: [10.244.1.0/24]
005  I0315 19:05:00.370473       1 kube.go:482] Creating the node lease for IPv4. This is the n.Spec.PodCIDRs: [10.244.0.0/24]
003  I0315 19:05:00.361865       1 kube.go:139] Waiting 10m0s for node controller to sync
008  I0315 19:05:01.362771       1 kube.go:146] Node controller sync successful
004  I0315 19:05:00.361980       1 kube.go:461] Starting kube subnet manager
029  I0315 19:05:01.415768       1 vxlan_network.go:65] watching for new subnet leases
009  I0315 19:05:01.362790       1 main.go:229] Created subnet manager: Kubernetes Subnet Manager - xp02-vm21
014  I0315 19:05:01.363689       1 match.go:285] Defaulting external address to interface address (172.16.0.90)
012  I0315 19:05:01.363354       1 match.go:210] Determining IP address of default interface
017  I0315 19:05:01.367988       1 vxlan.go:155] Setup flannel.1 mac address to 22:da:29:ff:7c:09 when flannel restarts
011  I0315 19:05:01.362951       1 main.go:452] Found network config - Backend type: vxlan
2511  I0315 21:19:00.766758       1 iptables.go:503] Some iptables rules are missing; deleting and recreating rules
2509  I0315 21:18:55.725654       1 iptables.go:503] Some iptables rules are missing; deleting and recreating rules
2245  I0315 21:07:51.170791       1 iptables.go:503] Some iptables rules are missing; deleting and recreating rules
2507  I0315 21:18:50.692843       1 iptables.go:503] Some iptables rules are missing; deleting and recreating rules
2505  I0315 21:18:45.658245       1 iptables.go:503] Some iptables rules are missing; deleting and recreating rules
2503  I0315 21:18:40.625697       1 iptables.go:503] Some iptables rules are missing; deleting and recreating rules
2501  I0315 21:18:35.585531       1 iptables.go:503] Some iptables rules are missing; deleting and
2499  I0315 21:18:30.545976       1 iptables.go:503] Some iptables rules are missing; del
2515  I0315 21:19:10.843118       1 iptables.go:503] Some iptables rules are mi
2497  I0315 21:18:25.514795       1 iptables.go:503] Some iptables rule
2513  I0315 21:19:05.802291       1 iptables.go:503] Some iptable
I0315 21:19:36.018942       1 iptables.go:503] Some
I0315 19:05:01.881928       1 iptables.go
```

We're almost ready to ask an LLM for help

```
033  I0315 19:05:01.434308       1 iptables.go:365] bootstrap done
034  I0315 19:05:01.438309       1 iptables.go:365] bootstrap done
035  I0315 19:05:01.446828       1 iptables.go:365] bootstrap done
015  I0315 19:05:01.363818       1 vxlan.go:141] VXLAN config: VNI=1 Port=0 GBP=false Learning=false DirectRouting=false
028  I0315 19:05:01.414661       1 main.go:400] Running backend.
010  I0315 19:05:01.362794       1 main.go:232] Installing signal handlers
013  I0315 19:05:01.363666       1 match.go:263] Using interface with name ens33 and address 172.16.0.90
019  W0315 19:05:01.378801       1 main.go:540] no subnet found for key: FLANNEL_IPV6_SUBNET in file: /run/flannel/subnet.env
018  W0315 19:05:01.378785       1 main.go:505] no subnet found for key: FLANNEL_SUBNET in file: /run/flannel/subnet.env
027  I0315 19:05:01.414648       1 main.go:396] Wrote subnet file to /run/flannel/subnet.env
```

- We used a TF-IDF metric to sort our log-file
- To create a diagnosis of the log file, leverage an LLM
  - Choose the top-n lines of the log-file (here: n=10)
  - Use those as part of a prompt, asking an LLM to describe the issue and propose a resolution

The Prompt (for GPT-4)

You are an IT operations engineer. Your task is to diagnose the issue described by an excerpt from a log file included between three backticks (```) and the file name below. The excerpt from the log file shows the most relevant log lines. Perform the following tasks one after the other:

1. Create a technical description of what this issue is in no more than 100 words. Be sure to keep technical details in the description (such as host names, IP addresses, interfaces, ...). Also consider the "syslog file name" to determine the host name.

2. Propose a resolution for the issue, composed of a single paragraph detailing your suggested next steps to fix the provided issue description. Be technical and specific.

File-name: kube-flannel-ds-xd5wp.log
```
033  I0315 19:05:01.434308       1 iptables.go:365] bootstrap done
034  I0315 19:05:01.438309       1 iptables.go:365] bootstrap done
035  I0315 19:05:01.446828       1 iptables.go:365] bootstrap done
015  I0315 19:05:01.363818       1 vxlan.go:141] VXLAN config: VNI=1 Port=0 GBP=false Learning=false DirectRouting=false
028  I0315 19:05:01.414661       1 main.go:400] Running backend.
010  I0315 19:05:01.362794       1 main.go:232] Installing signal handlers
013  I0315 19:05:01.363666       1 match.go:263] Using interface with name ens33 and address 172.16.0.90
019  W0315 19:05:01.378801         1 main.go:540] no subnet found for key: FLANNEL_IPV6_SUBNET in file: /run/flannel/subnet.env
018  W0315 19:05:01.378785         1 main.go:505] no subnet found for key: FLANNEL_SUBNET in file: /run/flannel/subnet.env
027  I0315 19:05:01.414648       1 main.go:396] Wrote subnet file to /run/flannel/subnet.env
```

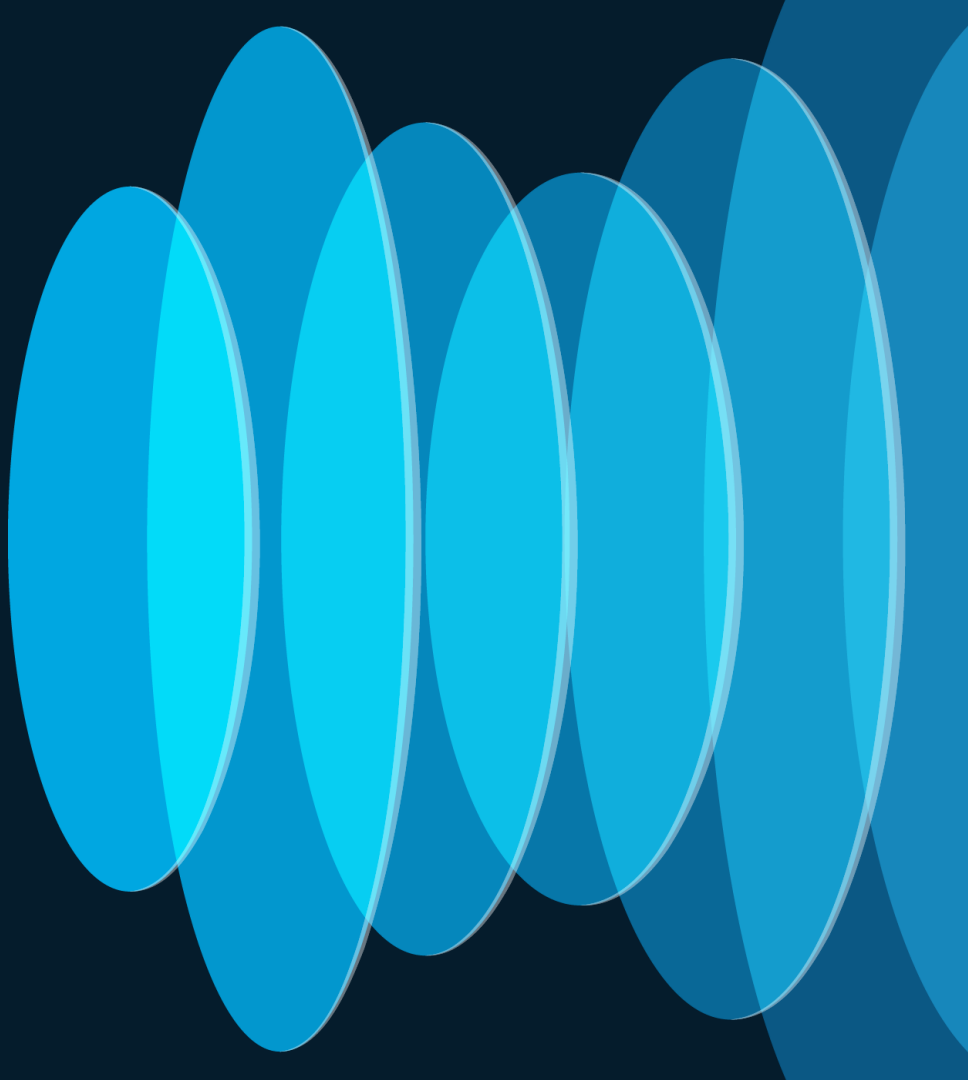
The Response

1. The issue presented in the log file `kube-flannel-ds-xd5wp.log` from a Kubernetes node pertains to the Flannel CNI (Container Network Interface) plugin, which is used for providing networking for pods. Specifically, the log entries indicate that Flannel has completed its initialization (`bootstrap done`) and is running the backend. It is using the network interface `ens33` with IP address `172.16.0.90`. However, it reports warnings that it could not find the required subnet configuration for both IPv4 (`FLANNEL_SUBNET`) and IPv6 (`FLANNEL_IPV6_SUBNET`) in the file `/run/flannel/subnet.env`, which is a critical part of the network setup for pods.

2. To resolve the issue, first ensure that Flannel is properly configured with the necessary subnet information. This typically involves setting the `FLANNEL_SUBNET` and `FLANNEL_IPV6_SUBNET` environment variables. You can do this by checking the Kubernetes ConfigMap for Flannel and ensuring it contains the correct subnet definitions. If the ConfigMap is correct, you may need to investigate why the Flannel daemon set is not able to write or read these keys to/from `/run/flannel/subnet.env`. Ensure the Flannel daemon has the proper permissions to access and modify this file. If the problem persists, consider restarting the Flannel pod (`kubectl delete pod kube-flannel-ds-xd5wp --namespace=kube-system`) to trigger a reconfiguration, and then check the logs again to see if the error is resolved.

The LLM “understands” logs of IT systems and can help us diagnose issues.

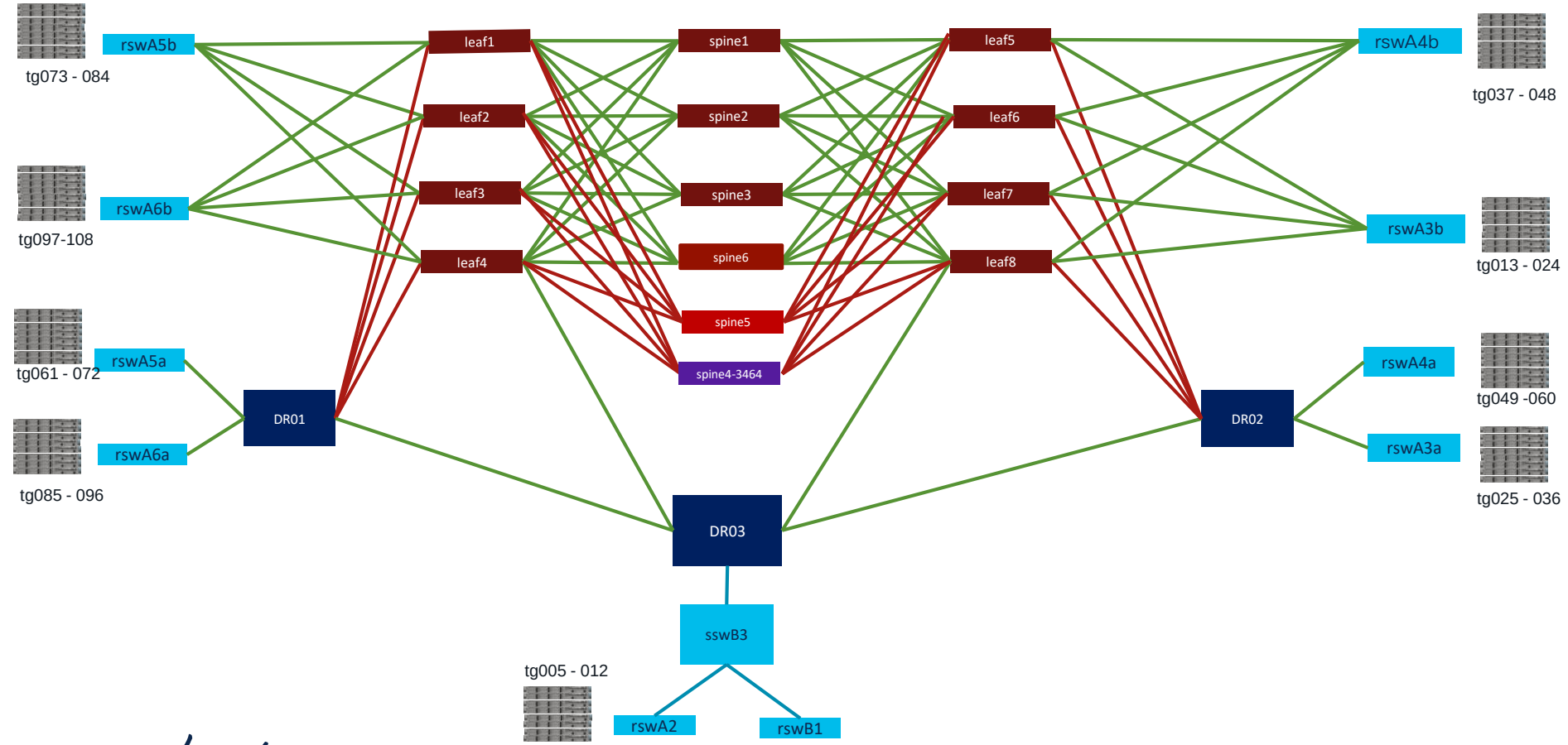
What about Model Driven Telemetry, i.e., timeseries from network devices?



Imagine...

You are an IT Ops engineer tasked with troubleshooting a network issue.

You have gathered streaming telemetry information and are now trying to draw insights from the data.



Example Data-Set: Model-Drive Telemetry Sensor Paths Collected

Cisco-IOS-XR-nto-misc-oper:memory-summary/nodes/node/summary * 1
Cisco-IOS-XR-wdsysmon-fd-oper:system-monitoring/cpu-utilization * 1
Cisco-IOS-XR-ipv6-io-oper:ipv6-io * 1
Cisco-IOS-XR-ipv4-io-oper:ipv4-network/nodes/node/statistics/traffic * 1
Cisco-IOS-XR-ip-tcp-oper:tcp * 1
Cisco-IOS-XR-fib-common-oper:fib-statistics * 1
Cisco-IOS-XR-ip-tcp-oper:tcp-connection * 1
Cisco-IOS-XR-ip-tcp-oper:tcp-nsr * 1
Cisco-IOS-XR-clns-isis-oper:isis/instances/instance/neighbors/neighbor * 1
Cisco-IOS-XR-clns-isis-oper:isis/instances/instance/interfaces/interface * 1
Cisco-IOS-XR-ip-bfd-oper:bfd/session-briefs * 1
Cisco-IOS-XR-ipv4-bgp-oper:bgp/instances/instance/instance-active/default-vrf/afs/af/af-process-info/global * 1
Cisco-IOS-XR-ip-bfd-oper:bfd/counters * 1
Cisco-IOS-XR-pfi-im-cmd-oper:interfaces/interface-briefs/interface-brief * 1
Cisco-IOS-XR-ip-bfd-oper:bfd/client-details/client-detail/brief * 1
Cisco-IOS-XR-ip-bfd-oper:bfd/summary * 1
Cisco-IOS-XR-pfi-im-cmd-oper:interfaces/interface-summary/interface-type * 1
Cisco-IOS-XR-clns-isis-oper:isis/instances/instance/topologies/topology/ipv4-routes/ipv4-route * 1
Cisco-IOS-XR-ipv4-bgp-oper:bgp/instances/instance/instance-active/default-vrf/process-info * 1
Cisco-IOS-XR-infra-statsd-oper:infra-statistics/interfaces/interface[interface-name=Hu*] * 1
Cisco-IOS-XR-ip-iarm-v6-oper:ipv6arm/addresses/vrfs/vrf/interfaces * 1
Cisco-IOS-XR-ip-iarm-v4-oper:ipv4arm/addresses/vrfs/vrf/interfaces * 1
Cisco-IOS-XR-drivers-media-eth-oper:ethernet-interface/statistics/statistic * 1
Cisco-IOS-XR-pfi-im-cmd-oper:interfaces/interface-xr/interface * 1

Example Data-Set: Synthetic Events

Absolute Timestamp	Relative Timestamp	Synthetic Event	Device	Detail
1.558.249.381.658.610	0.000s	ixchariot_traffic		
1.558.249.387.826.110	6.168s	break_bfd	spine4-3464	eth1/25
1.558.250.581.677.300	1200.019s	shutdown_interface	leaf7	HundredGigE0/0/0/10
1.558.251.787.725.630	2406.067s	enable_bfd	spine4-3464	eth1/25
1.558.252.981.645.430	3599.987s	enable_interface	leaf7	HundredGigE0/0/0/10
1.558.254.187.737.670	4806.079s	break_bfd	spine4-3464	eth1/25
1.558.255.381.645.900	5999.987s	shutdown_interface	leaf7	HundredGigE0/0/0/10
1.558.256.587.759.050	7206.100s	enable_bfd	spine4-3464	eth1/25
1.558.257.781.742.200	8400.084s	enable_interface	leaf7	HundredGigE0/0/0/10
1.558.258.987.739.670	9606.081s	break_bfd	spine4-3464	eth1/25
1.558.260.181.642.780	10799.984s	shutdown_interface	leaf7	HundredGigE0/0/0/10
1.558.260.189.485.350	10807.827s	ixchariot_traffic_stopped		
1558260295	10913.341s	collect_telemetry		

Diagnosing the telemetry data

- Can we discover and diagnose the synthetic events solely by analyzing the data using an unsupervised method?
- Approach – like what we did for diagnosing logs
  - Pre-processing: **Clean** the data
  - **Detect** change points
  - **Diagnose** a change point using an LLM
    - Too much data to fit into a context window: Choose relevant data for the change point
    - Leverage an LLM to diagnose the change point

Resulting Data For “Leaf7”

```
import pandas as pd
import traiage.utils.utils as utils

# TODO ensure the selected dataset dir contains a merged.csv file.
merged_data_fn, _ = datasets.get_input_data_file("merged.csv")

df = pd.read_csv(open(merged_data_fn, 'rb'))

# show number of rows and columns - dimensionality
shape = df.shape
print("dataset dimensions: rows={}, columns={}".format(shape[0], shape[1]))
# display a sample of the dataset, first 10 rows with first 10 columns for each row.
utils.displayDataFrame(df.iloc[0:9,0:9])
```

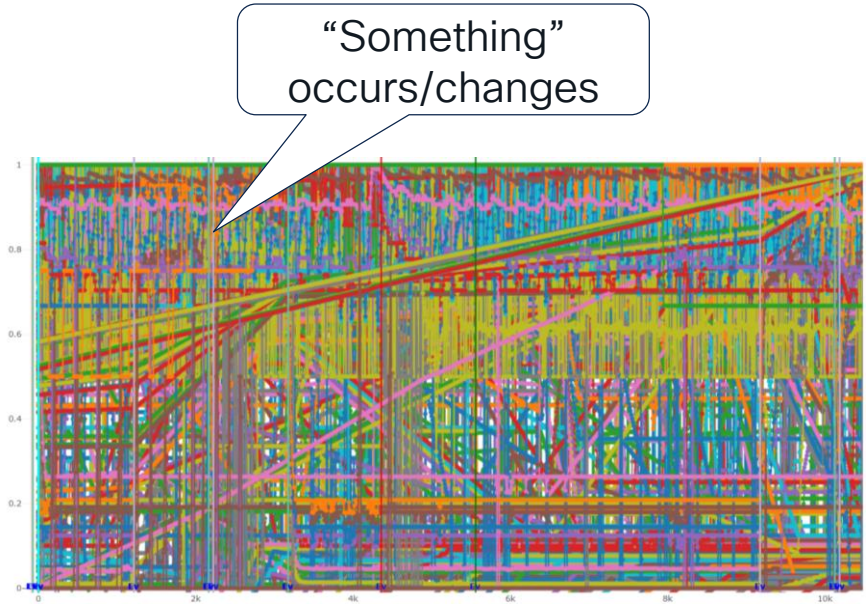
dataset dimensions: rows=1079, columns=7334

dataset dimensions: rows=1079, columns=7334

ts.V1	n0:Cisco-IOS-XR-drivers-media-eth-oper:ethernet-interface_statistics_statistic.csv:HundredGigE0/0/0/0:received-good-bytes	n0:Cisco-IOS-XR-drivers-media-eth-oper:ethernet-interface_statistics_statistic.csv:HundredGigE0/0/0/0:received-good-frames	n0:Cisco-IOS-XR-drivers-media-eth-oper:ethernet-interface_statistics_statistic.csv:HundredGigE0/0/0/0:received-multicast-frames	n0:Cisco-IOS-XR-drivers-media-eth-oper:ethernet-interface_statistics_statistic.csv:HundredGigE0/0/0/0:received-total-bytes	n0:Cisco-IOS-X… interface_stati… total-frames
1558249381.658611	513408648445952	121428366854.500000	70062.976587	513408648445952	121428366854
1558249391.658611	513493882415104	121439268000.000000	70063.975488	513493882415104	121439268000
1558249401.658611	513570679283712	121449025040.250000	70064.000000	513570679283712	121449025040
1558249411.658611	513647466450944	121458776629.500000	70064.000000	513647466450944	121458776629
1558249421.658611	513724222164992	121468531715.750000	70064.911685	513724222164992	121468531715
1558249431.658611	513800108363776	121478179924.000000	70065.000000	513800108363776	121478179924
1558249441.658611	513876775303168	121487920800.750000	70065.909252	513876775303168	121487920800
1558249451.658611	513952962519040	121497604673.250000	70066.908253	513952962519040	121497604673
1558249461.658611	514032355700736	121507688129.500000	70067.000000	514032355700736	121507688129

The “nature” of our data-set

- Time-series data
- High dimensional
  - High number of features
  - Variable number of available features
    - Over time
    - Over different devices
- Different in nature
  - Units
  - Behavior
  - Interpretation



Data Pre-Processing:
Consolidate various data sources

Cisco-IOS-XR-infra-statsd-oper:infra-statistics_interfaces_interface_latest_interfaces-mib-counters.csv

Cisco-IOS-XR-infra-statsd-oper:infra-statistics_interfaces_interface_latest_protocols_protocol.csv

Cisco-IOS-XR-ip-tcp-oper:tcp-connection_nodes_node_brief-informations_brief-information.csv

Cisco-IOS-XR-ip-tcp-oper:tcp-connection_nodes_node_detail-informations_detail-information.csv

Cisco-IOS-XR-ip-tcp-oper:tcp-connection_nodes_node_extended-information_display-types_display-type_connection-id.csv

Cisco-IOS-XR-ip-tcp-oper:tcp-connection_nodes_node_statistics_clients_client.csv

Cisco-IOS-XR-ip-tcp-oper:tcp-connection_nodes_node_statistics_pcbs_pcb.csv

Cisco-IOS-XR-ip-tcp-oper:tcp-connection_nodes_node_statistics_summary.csv

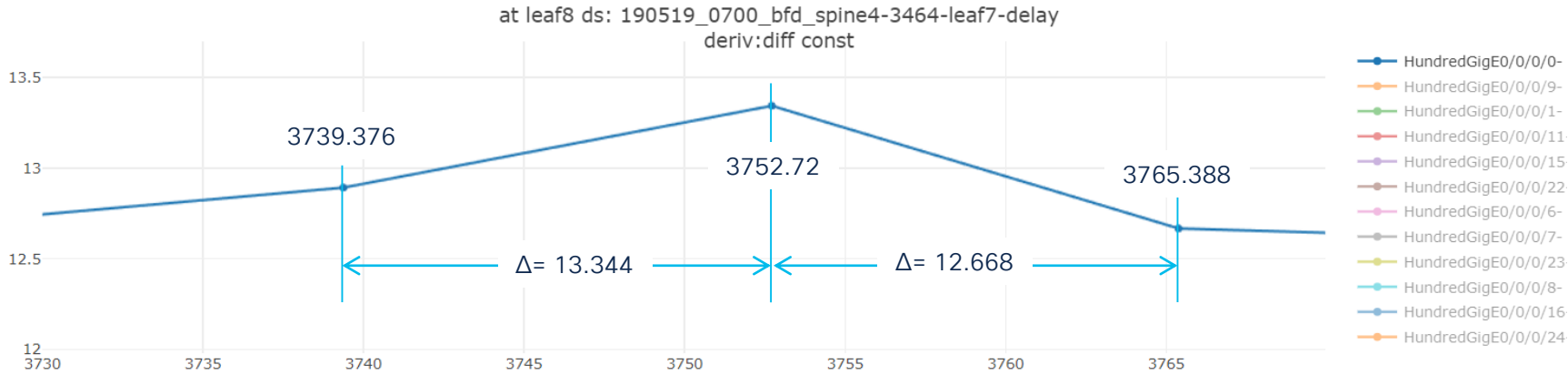
Cisco-IOS-XR-ip-tcp-oper:tcp-nsr_nodes_node_client_brief-clients_brief-client.csv

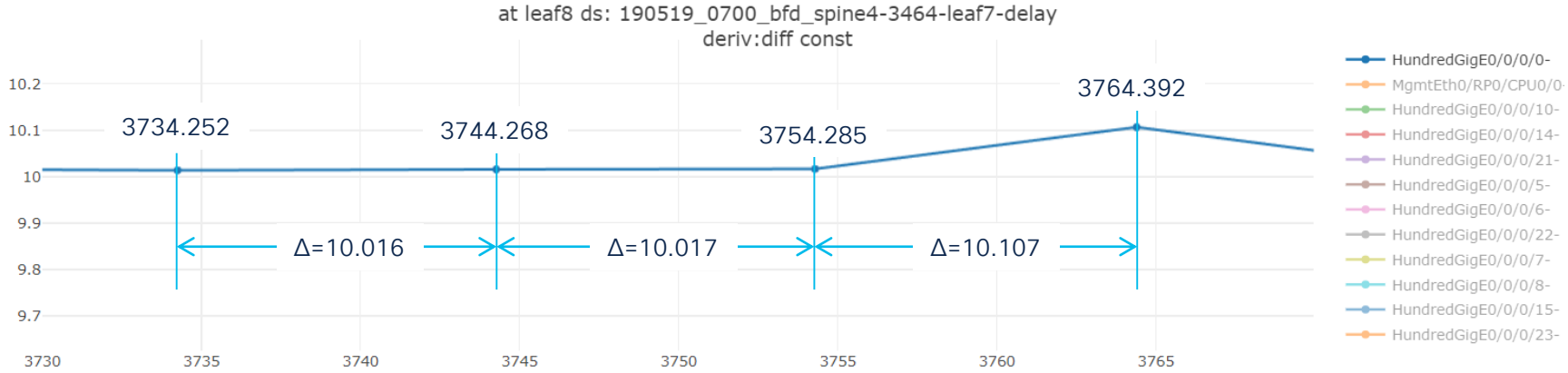
Cisco-IOS-XR-ip-tcp-oper:tcp-nsr_nodes_node_client_detail-clients_detail-client.csv

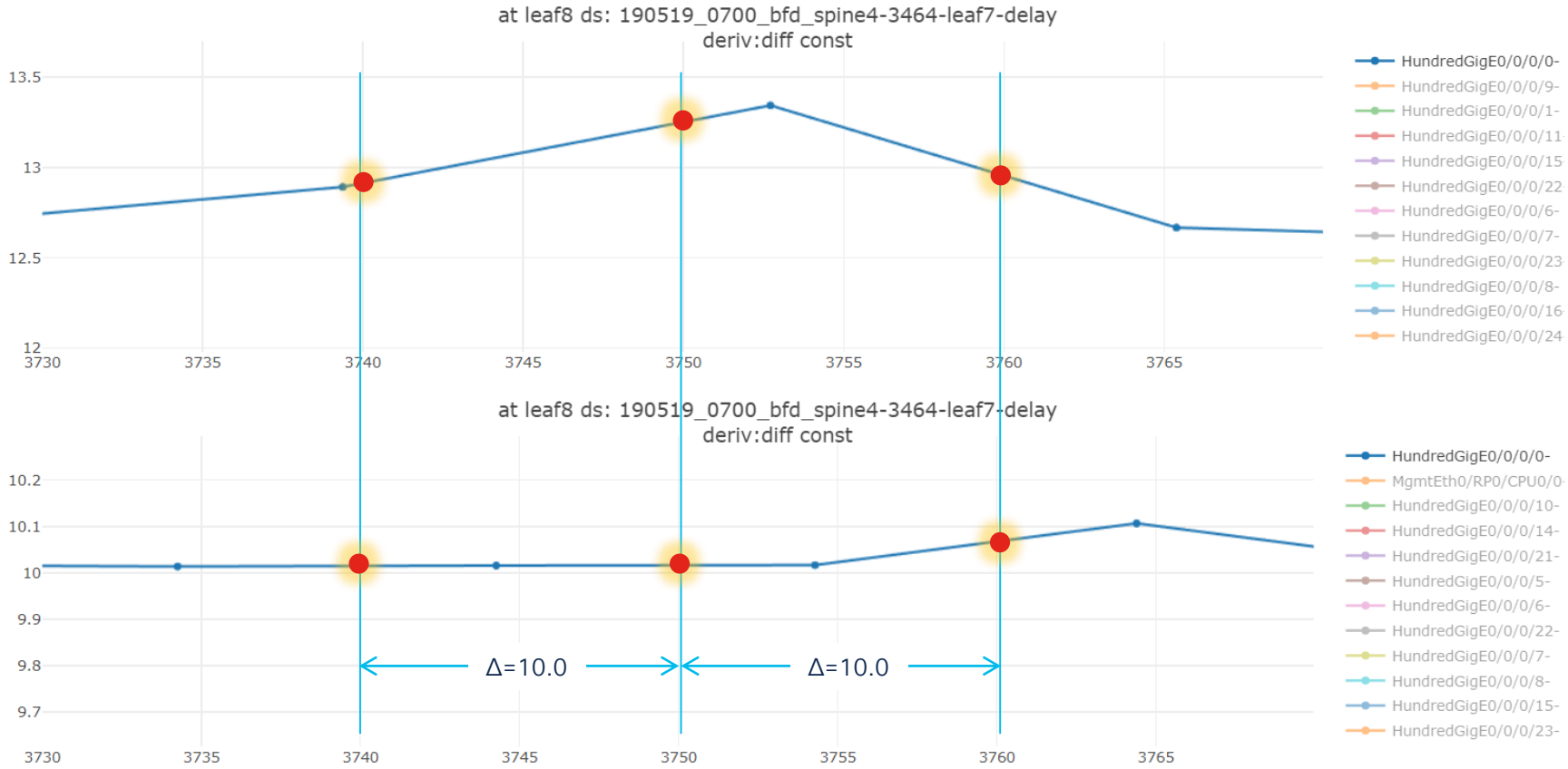
Cisco-IOS-XR-ip-tcp-oper:tcp-nsr_nodes_node_session-set_brief-sets_brief-set.csv

Cisco-IOS-XR-ip-tcp-oper:tcp-nsr_nodes_node_session-set_detail-sets_detail-set

Cisco-IOS-XR-ip-tcp-oper:tcp-nsr_nodes_node_statistics_stat

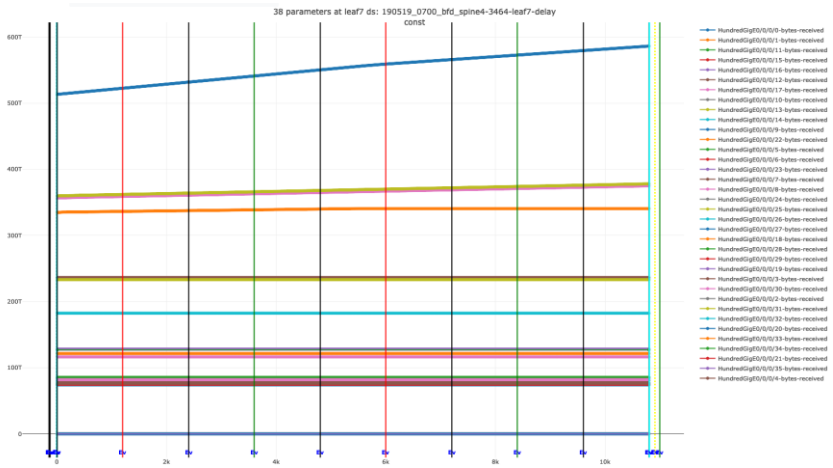




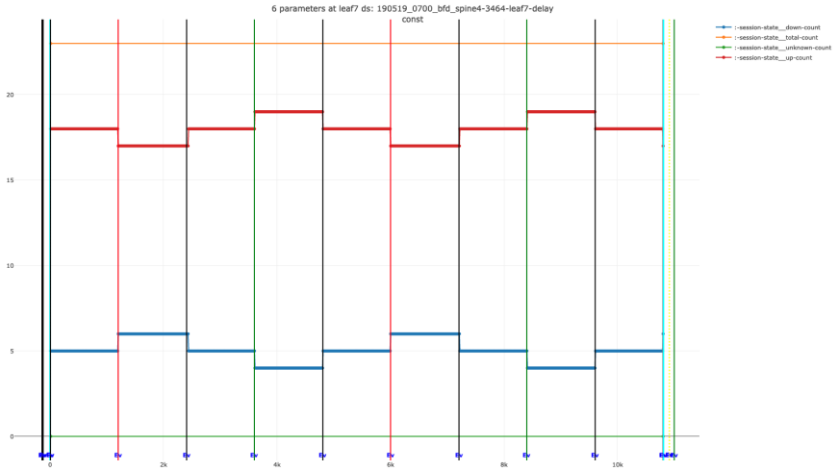


Data Pre-Processing:
Make heterogenous timeseries comparable

Interface bytes received



BFD session counts



Data Pre-Processing

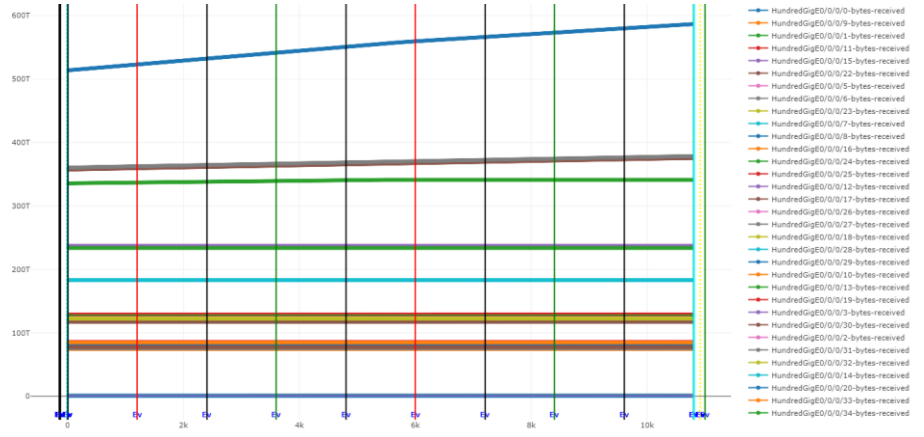
$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

$$x' = \frac{x - \bar{x}}{\sigma}$$

- **Scaling** methods include Min-Max and Z-Normalizing. By scaling the data, it allows every univariate time-series to evolve within comparable ranges, which can be necessary for processing method later on.
- **Smoothing** like the exponential moving average, in order to focus on the baseline of the data only and remove background variance, which processing methods aren't interested in.
- **Differentiating** the incremental time-series is a solution to compare the incremental and non incremental data. This goes hand in hand with the scaling to be able to compare data of different types and ranges.

Pre-Processing: Example

Raw Data



After Scaling and Differentiating

Pre-Processed Data for “Leaf 7”

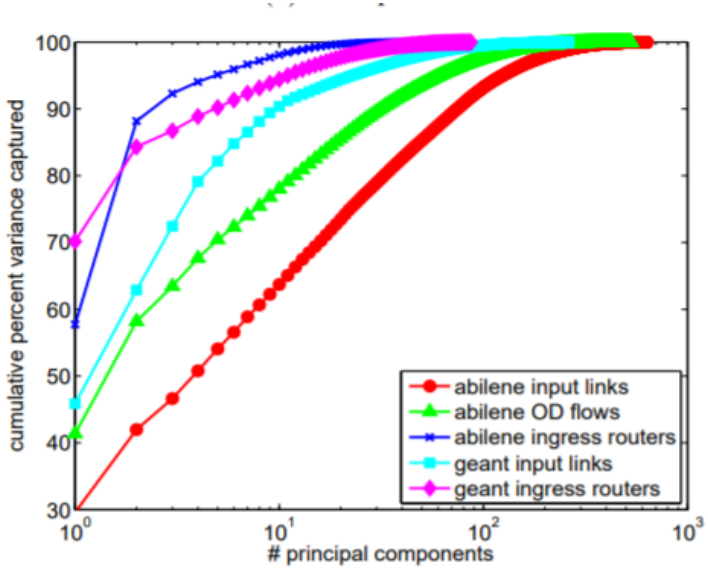
```
# TODO ensure the selected dataset dir contains a preprocessed_offline.csv file.
preprocessed_data_fn, _ = datasets.get_input_data_file("preprocessed_offline.csv")

df = pd.read_csv(open(preprocessed_data_fn, 'rb'))

# show number of rows and columns - dimensionality
shape = df.shape
print("dataset dimensions: rows={}, columns={}".format(shape[0], shape[1]))
# display a sample of the dataset, first 10 rows with first 10 columns for each row.
utils.displayDataFrame(df.iloc[0:9,0:9])
```

```
dataset dimensions: rows=1079, columns=7334
```

ts	n0:Cisco-IOS-XR-drivers-media-eth-oper:ethernet-interface_statistics_statistic.csv:HundredGigE0/0/0/0:received-good-bytes	n0:Cisco-IOS-XR-drivers-media-eth-oper:ethernet-interface_statistics_statistic.csv:HundredGigE0/0/0/0:received-good-frames	n0:Cisco-IOS-XR-drivers-media-eth-oper:ethernet-interface_statistics_statistic.csv:HundredGigE0/0/0/0:received-multicast-frames	n0:Cisco-IOS-XR-drivers-media-eth-oper:ethernet-interface_statistics_statistic.csv:HundredGigE0/0/0/0:received-total-bytes	n0:Cisco-IOS-X interface_stati total-frames
1558249381.658611	0.681327	0.687531	0.504115	0.681327	0.687531
1558249391.658611	0.681327	0.687531	0.504115	0.681327	0.687531
1558249401.658611	0.644663	0.648323	0.258243	0.644663	0.648323
1558249411.658611	0.626289	0.628532	0.129121	0.626289	0.628532
1558249421.658611	0.616965	0.618757	0.294610	0.616965	0.618757
1558249431.658611	0.608525	0.610206	0.169590	0.608525	0.610206
1558249441.658611	0.607697	0.609107	0.314231	0.607697	0.609107
1558249451.658611	0.605199	0.606604	0.409198	0.605199	0.606604
1558249461.658611	0.617882	0.619045	0.227750	0.617882	0.619045

Dealing With High Dimensionality

- We are given a collection of N high dimensional objects x_1, x_2, ... x_n
- How can we get a feel for how these objects are arranged in the data space? How can we visualize very high-dimensional data?
- Example methods:
  - PCA
  - t-SNE
  - UMAP



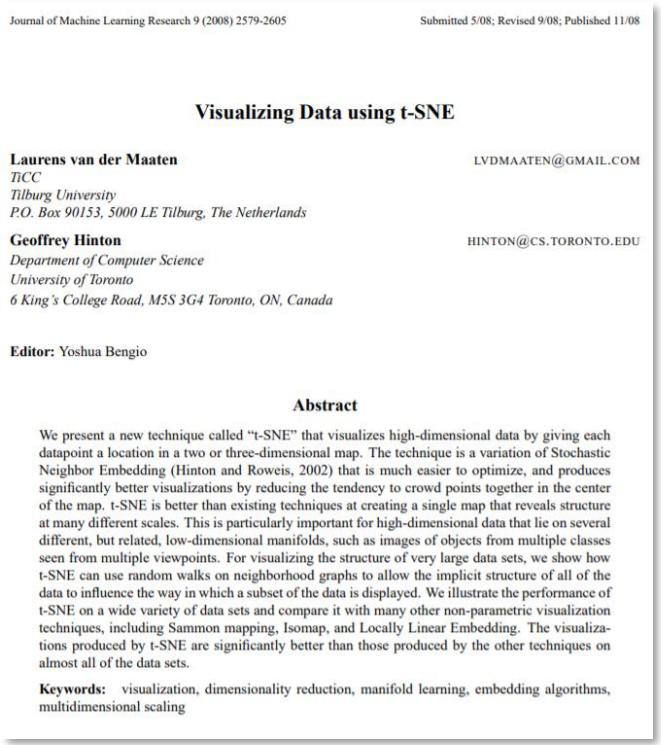
"Sensitivity of PCA for Traffic Anomaly Detection"
Haakon Ringberg, Augustin Soule, Jennifer Rexford, Christophe Diot
https://ics.forth.gr/netlab/mobile/Bibliography/LoadBalancing/LB/PCA_Anomaly_Deytection.pdf

t-Stochastic Neighbor Embedding (t-SNE)

t-SNE Idea

Build a map in which distances between points reflect similarities in the data:

Model each high-dimensional object by a 2- or 3-dimensional point in such a way, that similar objects are modeled by nearby points and dissimilar objects are modeled by distant points with high probability



Journal of Machine Learning Research 9 (2008) 2579-2605 Submitted 5/08; Revised 9/08; Published 11/08

Visualizing Data using t-SNE

Laurens van der Maaten LVDMAATEN@GMAIL.COM
TiCC
Tilburg University
P.O. Box 90153, 5000 LE Tilburg, The Netherlands

Geoffrey Hinton HINTON@CS.TORONTO.EDU
Department of Computer Science
University of Toronto
6 King's College Road, M5S 3G4 Toronto, ON, Canada

Editor: Yoshua Bengio

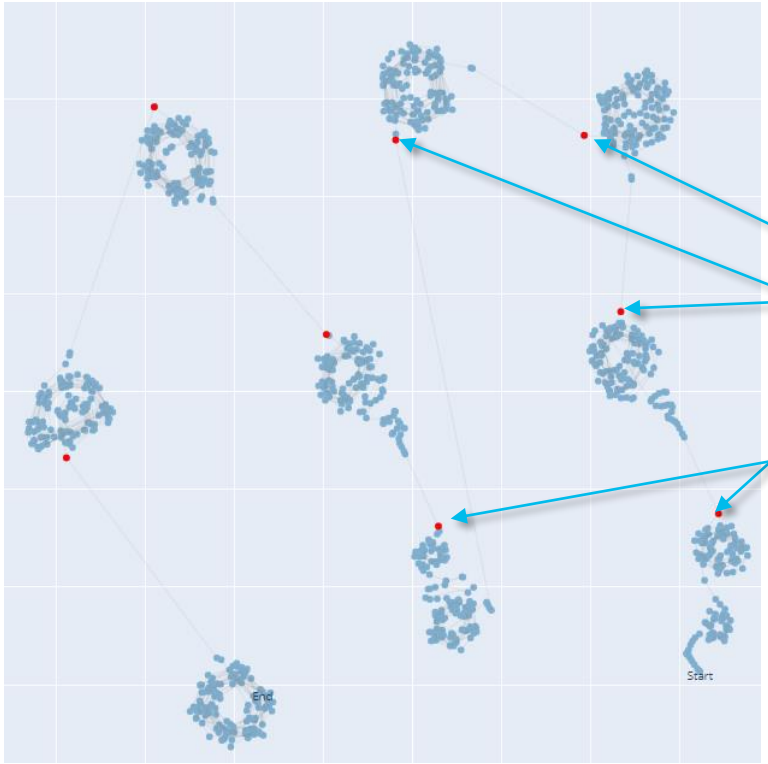
Abstract

We present a new technique called "t-SNE" that visualizes high-dimensional data by giving each datapoint a location in a two or three-dimensional map. The technique is a variation of Stochastic Neighbor Embedding (Hinton and Roweis, 2002) that is much easier to optimize, and produces significantly better visualizations by reducing the tendency to crowd points together in the center of the map. t-SNE is better than existing techniques at creating a single map that reveals structure at many different scales. This is particularly important for high-dimensional data that lie on several different, but related, low-dimensional manifolds, such as images of objects from multiple classes seen from multiple viewpoints. For visualizing the structure of very large data sets, we show how t-SNE can use random walks on neighborhood graphs to allow the implicit structure of all of the data to influence the way in which a subset of the data is displayed. We illustrate the performance of t-SNE on a wide variety of data sets and compare it with many other non-parametric visualization techniques, including Sammon mapping, Isomap, and Locally Linear Embedding. The visualizations produced by t-SNE are significantly better than those produced by the other techniques on almost all of the data sets.

Keywords: visualization, dimensionality reduction, manifold learning, embedding algorithms, multidimensional scaling

http://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf

t-SNE Visualization of our Data Set

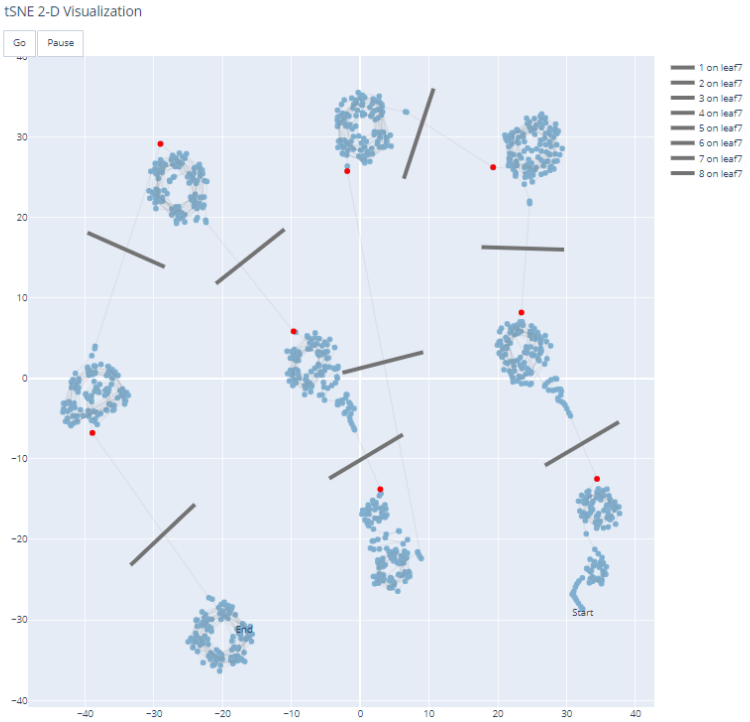


Relative Timestamp	Synthetic Event
0.000s	ixchariot_traffic
6.168s	break_bfd
1200.019s	shutdown_interface
2406.067s	enable_bfd
3599.987s	enable_interface
4806.079s	break_bfd
5999.987s	shutdown_interface
7206.100s	enable_bfd
8400.084s	enable_interface
9606.081s	break_bfd
10799.984s	shutdown_interface
10807.827s	ixchariot_traffic_stopped

https://github.com/CiscoDevNet/telemetry-experiments

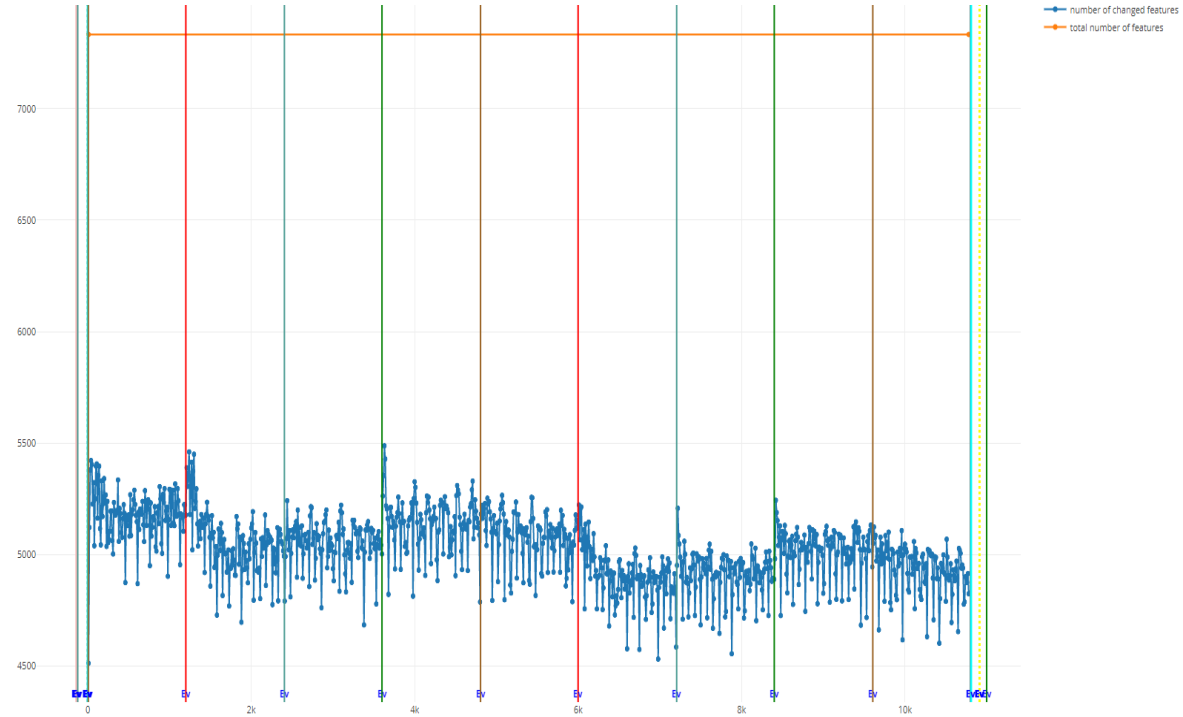
Transitions between clusters describe state changes: DBSCAN to detect clusters

```
[20]: detector.plot(withEvents=True)
      detector.select_changepoints()
```



DBSCAN original paper:
Ester, Martin; Kriegel, Hans-Peter; Sander, Jörg; Xu, Xiaowei (1996). Simoudis, Evangelos; Han, Jiawei; Fayyad, Usama M. (eds.). A density-based algorithm for discovering clusters in large spatial databases with noise . Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press. pp. 226–231. CiteSeerX 10.1.1.121.9220. ISBN 1-57735-004-9.

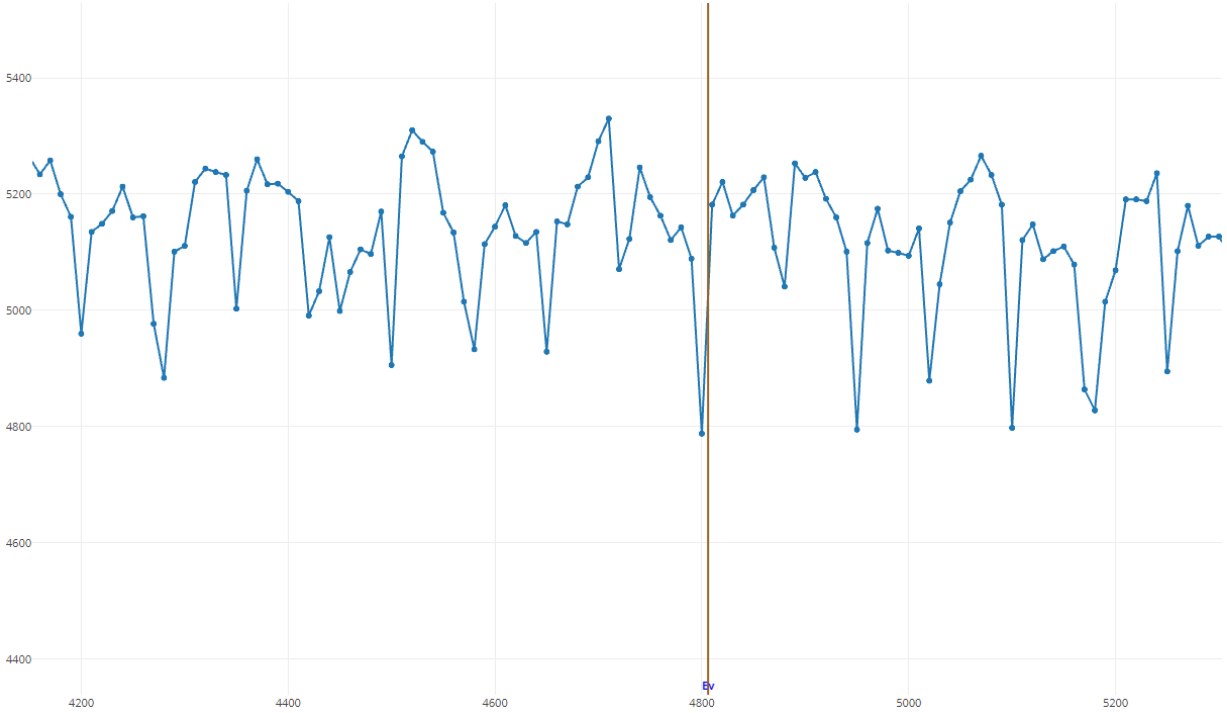
Which features are the most descriptive for a changepoint?

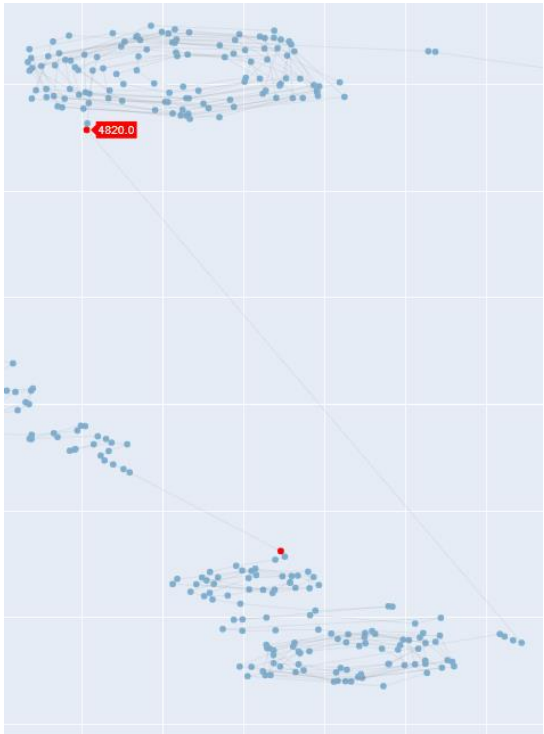


Many features change all the time.

“Change” isn’t sufficient as a filtering criteria.

Change-Point t=4820: Number of Changed features between timestamps

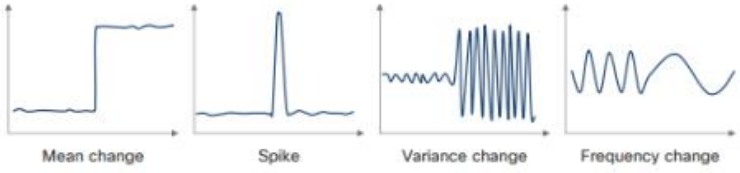




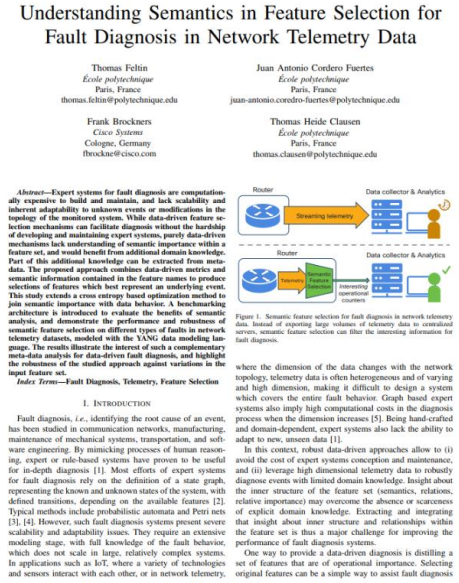
Which features are the most descriptive for a changepoint?

Combine

- **Change metric** (looking for steps, spikes, changes in variance)



- **Estimation of importance of features** based on approximation of domain knowledge
  - “TF-IDF style” ranking of features based on the importance of tokens/substrings of their name (i.e., the sensor path)



Understanding Semantics in Feature Selection for Fault Diagnosis in Network Telemetry Data

Thomas Feltin, *École polytechnique*, Paris, France, thomas.feltin@polytechnique.edu
Juan Antonio Cordero Fuertes, *École polytechnique*, Paris, France, juan-antonio.coredro-fuertes@polytechnique.edu
Frank Brockners, *Cisco Systems*, Cologne, Germany, fbrockne@cisco.com
Thomas Heide Clausen, *École polytechnique*, Paris, France, thomas.clausen@polytechnique.edu

***Abstract*—Expert systems for fault diagnosis are computationally expensive to build and maintain, and lack scalability and inherent adaptability to unknown events or modifications in the topology of the monitored system. While data-driven feature selection mechanisms can facilitate diagnosis without the hardship of developing and maintaining expert systems, purely data-driven mechanisms lack understanding of semantic importance within a feature set, and would benefit from additional domain knowledge. Part of this additional knowledge can be extracted from metadata. The proposed approach combines data-driven metrics and semantic information contained in the feature names to produce selections of features which best represent an underlying event. This study extends a cross entropy based optimization method to join semantic importance with data behavior. A benchmarking architecture is introduced to evaluate the benefits of semantic analysis, and demonstrate the performance and robustness of semantic feature selection on different types of faults in network telemetry datasets, modeled with the YANG data modeling language. The results illustrate the interest of such a complementary meta-data analysis for data-driven fault diagnosis, and highlight the robustness of the studied approach against variations in the input feature set.**

***Index Terms*—Fault Diagnosis, Telemetry, Feature Selection**

Figure 1. Semantic feature selection for fault diagnosis in network telemetry data. Instead of exporting large volumes of telemetry data to centralized servers, semantic feature selection can filter the interesting information for fault diagnosis.

I. Introduction

Fault diagnosis, *i.e.*, identifying the root cause of an event, has been studied in communication networks, manufacturing, maintenance of mechanical systems, transportation, and software engineering. By mimicking processes of human reasoning, expert or rule-based systems have proven to be useful for in-depth diagnosis [1]. Most efforts of expert systems for fault diagnosis rely on the definition of a state graph, representing the known and unknown states of the system, with defined transitions, depending on the available features [2]. Typical methods include probabilistic automata and Petri nets [3], [4]. However, such fault diagnosis systems present severe scalability and adaptability issues. They require an extensive modeling stage, with full knowledge of the fault behavior, which does not scale in large, relatively complex systems. In applications such as IoT, where a variety of technologies and sensors interact with each other, or in network telemetry, where the dimension of the data changes with the network topology, telemetry data is often heterogeneous and of varying and high dimension, making it difficult to design a system which covers the entire fault behavior. Graph based expert systems also imply high computational costs in the diagnosis process when the dimension increases [5]. Being hand-crafted and domain-dependent, expert systems also lack the ability to adapt to new, unseen data [1].

In this context, robust data-driven approaches allow to (i) avoid the cost of expert systems conception and maintenance, and (ii) leverage high dimensional telemetry data to robustly diagnose events with limited domain knowledge. Insight about the inner structure of the feature set (semantics, relations, relative importance) may overcome the absence or scarceness of explicit domain knowledge. Extracting and integrating that insight about inner structure and relationships within the feature set is thus a major challenge for improving the performance of fault diagnosis systems.

One way to provide a data-driven diagnosis is distilling a set of features that are of operational importance. Selecting original features can be a simple way to assist fault diagnosis

T. Feltin, J. A. C. Fuertes, F. Brockners and T. H. Clausen, "*Understanding Semantics in Feature Selection for Fault Diagnosis in Network Telemetry Data*", NOMS 2023 - 2023 IEEE/IFIP Network Operations and Management Symposium

Leveraging the semantics of the sensor path

Importance metric based on TF-IDF leveraging token distributions

The **importance** of a feature is approximated by its **rareness** in the feature set

Rare tokens often describe aggregated or summary information

tcp_connection_statistics:HundredGigE0/0/0/1:bytes-sent

Frequent tokens: many references to these tokens in the dataset, low importance

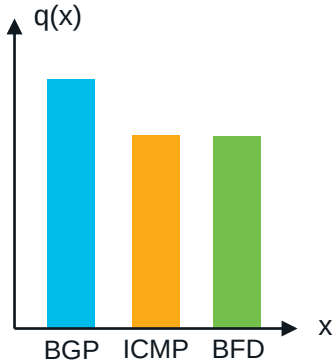
interfaces_interface-summary::interface-counts_admin-down-interface-count

Rare tokens: few references to these tokens in the dataset, high importance

Quantifying importance in a token distribution

(1) First problem: Specificity

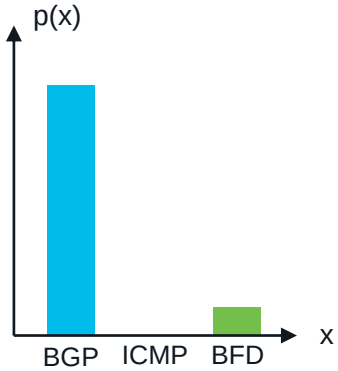
BGP_last_RIB_version
ICMP_received
ICMP_unreachable_address
BFD_hello_transmit_count
BGP_freed_path_count
BFD_session_up_count
BFD_session_down_count
BGP_neighbors_count
BGP_table_version
ICMP_sent



"anything could be impacted by the change"

Entropy : H(q) = 0.47

BGP_last_RIB_version
BGP_update_message_sent
BGP_freed_path_count
BFD_session_down_count
BGP_neighbors_count
BGP_data_bytes_read



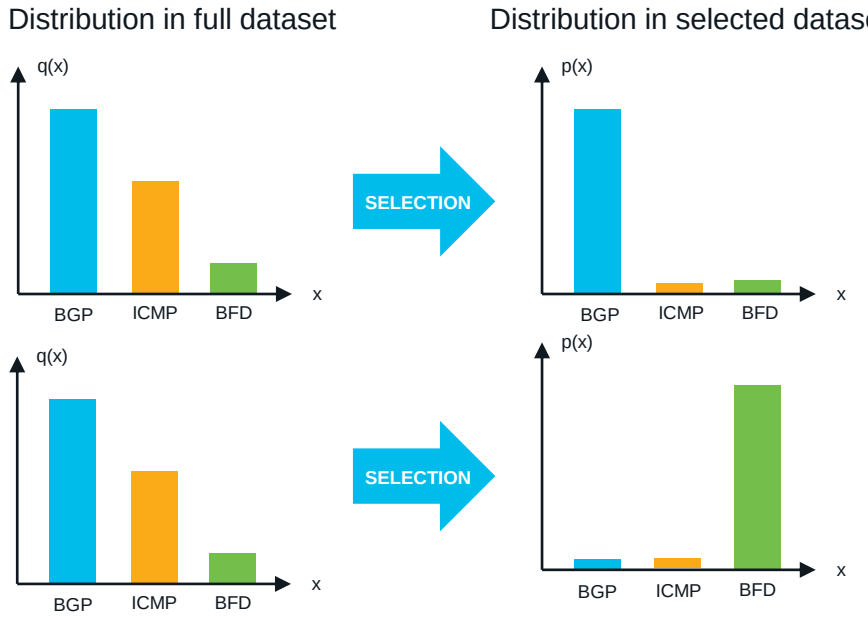
"something has happened which concerns BGP"

Entropy : H(p) = 0.22

Entropy describes a degree of focus of a selection

Quantifying importance in a token distribution

(2) Second Problem: Comparative importance of features



Cross-Entropy H(p,q)

(relative entropy compared to reference)

H(p,q) **high** because the selection is focused on **BGP** (which was **frequent** in the original set)

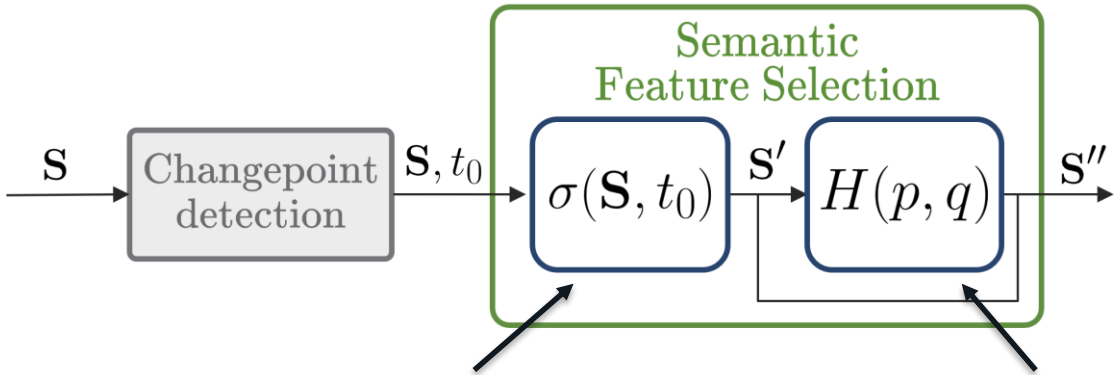
→ favors **focused** selections

H(p,q) **even higher** because the selection is focused on **BFD** (which was **rare** in the original set)

→ favors selections of **rare/important** sensor paths

Semantic feature selection for fault diagnosis

We present operators with the most meaningful counters among those that change



Change metric
(amplitude of steps, spikes, etc.)

Data behavior

Importance metric
(extracted from feature names)

Approximation of domain knowledge

Semantic feature selection for fault diagnosis

Definition of an optimization score, two main components:

1. The change metric (looking for steps, spikes, changes in variance)
   Data-driven, based only on **data behavior**

2. The cross-entropy between original distribution and selected distribution
   Estimation of importance of features based on **approximation of domain knowledge**

$$\mathcal{L}(\mathcal{S}, p, q) = \left(1 - e^{-\frac{|\mathcal{S}|}{\alpha}}\right) \sum_{k=1}^{K} H\left(p_k, q_k\right) \frac{1}{|\mathcal{S}|} \sum_{i=1}^{|\mathcal{S}|} \sigma_i$$

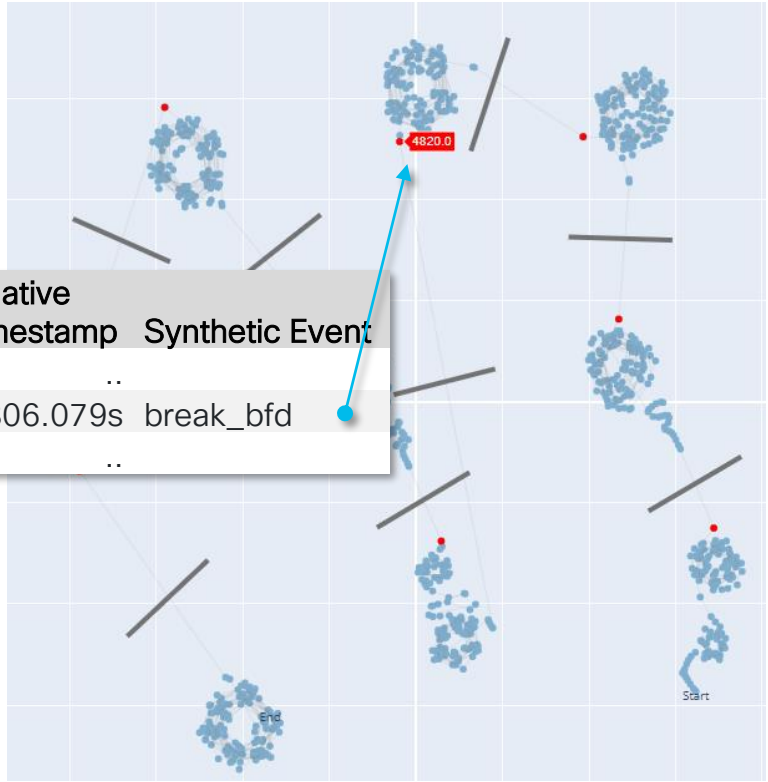
(α acts as verbosity parameter)

Regularization

(2) Feature importance

(1) Change amplitude

Semantic Feature selection: Top 5 Features



- Cisco-IOS-XR-ip-bfd-oper:bfd_counters_packet-counters_packet-counter.csv:bfd-mgmt-pkt-display-type-none:HundredGigE0/0/0/16:0/0/CPU0:hello-receive-count
  CHANGE: 1.0
- Cisco-IOS-XR-ip-bfd-oper:bfd_session-briefs_session-brief.csv:172.31.14.48:HundredGigE0/0/0/16:0/0/CPU0:0/0/CPU0:ip-single-hop:status-brief-information__async-interval-multiplier__negotiated-local-transmit-interval
  CHANGE: 1667000.0
- Cisco-IOS-XR-ip-bfd-oper:bfd_session-briefs_session-brief.csv:172.31.14.48:HundredGigE0/0/0/16:0/0/CPU0:0/0/CPU0:ip-single-hop:status-brief-information__async-interval-multiplier__negotiated-remote-transmit-interval
  CHANGE: 1667000.0
- Cisco-IOS-XR-ip-bfd-oper:bfd_summary.csv:::session-state__down-count
  CHANGE: 1.0
- Cisco-IOS-XR-ip-bfd-oper:bfd_summary.csv:::session-state__up-count
  CHANGE: -1.0

See also: https://github.com/CiscoDevNet/telemetry-experiments/blob/main/highlevel-pipelines/mdt_diagnosis_pipeline.ipynb

Ready to ask an LLM for a Diagnosis: Prompt for our example

You are a networking expert.
You are diagnosing a network issue based on telemetry information received from a Cisco router running IOS-XR 7.3.1.

Below, in the "List of sensor paths" section you find a list of YANG sensor path counters which have changed the most while the issue occurred.
Each line shows the name of the sensor path counter and the absolute amount that the sensor path counter has changed separated by the word CHANGE.
The sensor path counters are descriptive of the issue.

Perform the following two steps one after the other:

1. First, create a 'description', explain what is the issue with this router in a single paragraph. Be technical and specific.
2. Second, create a 'resolution', detailing your suggested next steps to fix the issue in a single paragraph. Be technical and specific.

List of sensor paths:
Cisco-IOS-XR-ip-bfd-oper:bfd_counters_packet-counters_packet-counter.csv:bfd-mgmt-pkt-display-type-none:HundredGigE0/0/0/16:0/0/CPU0:hello-receive-count CHANGE: 1.0
Cisco-IOS-XR-ip-bfd-oper:bfd_session-briefs_session-brief.csv:172.31.14.48:HundredGigE0/0/0/16:0/0/CPU0:0/0/CPU0:ip-single-hop:status-brief-information__async-interval-multiplier__negotiated-local-transmit-interval CHANGE: 1667000.0
Cisco-IOS-XR-ip-bfd-oper:bfd_session-briefs_session-brief.csv:172.31.14.48:HundredGigE0/0/0/16:0/0/CPU0:0/0/CPU0:ip-single-hop:status-brief-information__async-interval-multiplier__negotiated-remote-transmit-interval CHANGE: 1667000.0
Cisco-IOS-XR-ip-bfd-oper:bfd_summary.csv:::session-state__down-count CHANGE: 1.0
Cisco-IOS-XR-ip-bfd-oper:bfd_summary.csv:::session-state__up-count CHANGE: -1.0

Example: LLM Diagnosis (1/2): Issue Description

Description:
The telemetry data from the Cisco router indicates an issue with Bidirectional Forwarding Detection (BFD) on the interface HundredGigE0/0/0/16. BFD is a low-overhead, short-duration method for detecting failures in the path between adjacent forwarding engines, including the interfaces, data links, and forwarding planes. The `hello-receive-count` for BFD has incremented by 1, suggesting that a BFD control packet has been received, yet the `session-state__down-count` has increased by 1 while the `session-state__up-count` has decreased by 1, indicating a BFD session went down and did not come back up. Additionally, there has been a significant change in both the local and remote transmit interval settings, as indicated by the `negotiated-local-transmit-interval` and `negotiated-remote-transmit-interval` changing by 1,667,000 microseconds each. This drastic change suggests a renegotiation of BFD parameters which could be due to link flapping, misconfiguration, or an issue with the remote peer.

Example: LLM Diagnosis (2/2): Proposed Resolution

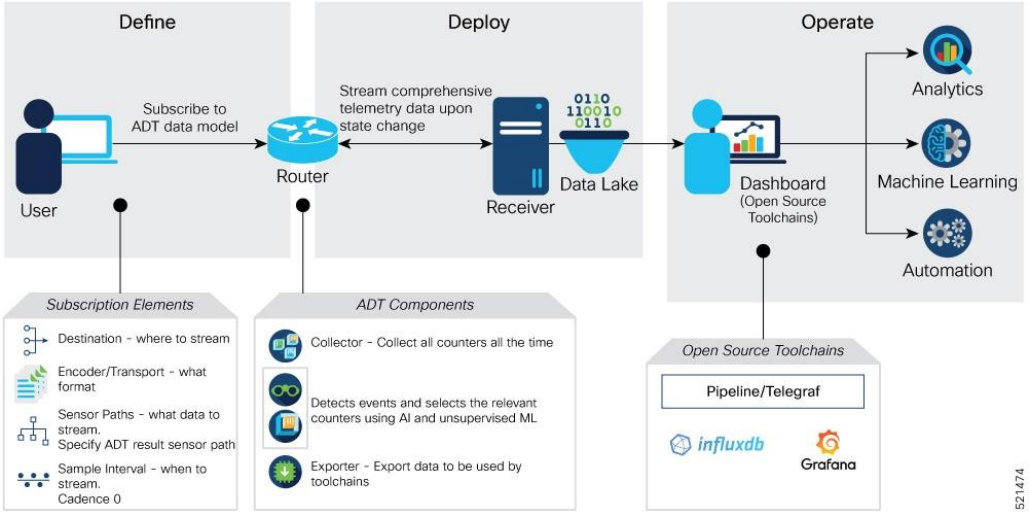
Resolution:
To resolve the issue, first, check the physical layer and ensure that the interface HundredGigE0/0/0/16 is stable without any flapping or physical errors. Review the BFD configuration on both ends of the session to ensure they are compatible and the intervals are not set too aggressively for the network's performance characteristics. Check the system logs and BFD event history for any indication of why the session went down. It is also advisable to check the CPU and memory usage of the device to ensure it is not being overloaded, as this could cause processing delays leading to BFD session drops. If the issue persists, perform a detailed packet capture on the interface to analyze BFD packet exchanges and identify if there are any irregularities or malformations in the packets. If necessary, contact Cisco support for further assistance, especially if there is suspicion of a bug in the IOS-XR version running on the router.

Interested in trying things yourself?

https://github.com/CiscoDevNet/telemetry-experiments

Side note:
AI Driven Telemetry – for Streaming Telemetry

- Introduced with IOS-XR 7.3.1
- Automatically select sensor paths, detect changepoints and on detecting a change-point export the set of sensor paths that describe the change the best.



https://www.cisco.com/c/en/us/td/docs/routers/asr9000/software/asr9k-r7-3/telemetry/configuration/guide/b-telemetry-cg-asr9000-73x/AI-driven-telemetry.html

ADT example

```
Router#show adt events
|-------------------------------------------------------------------------------|
| Number of Events :      5                                                     |
|-------------------------------------------------------------------------------|
|  Event id |                   Timestamp | Description                         |
|-------------------------------------------------------------------------------|
|       119 |  Tue 2020-12-01 13:41:56:141|                             Traffic|
|       120 |  Thu 2020-12-03 19:36:14:937|       Addressing & Reachability|
|       121 |  Thu 2020-12-03 20:00:48:015|       Addressing & Reachability|
|       122 |  Sun 2020-12-06 17:09:57:994|                             Traffic|
|       123 |  Tue 2020-12-08 18:21:45:419|                             Traffic|
|-------------------------------------------------------------------------------|
```

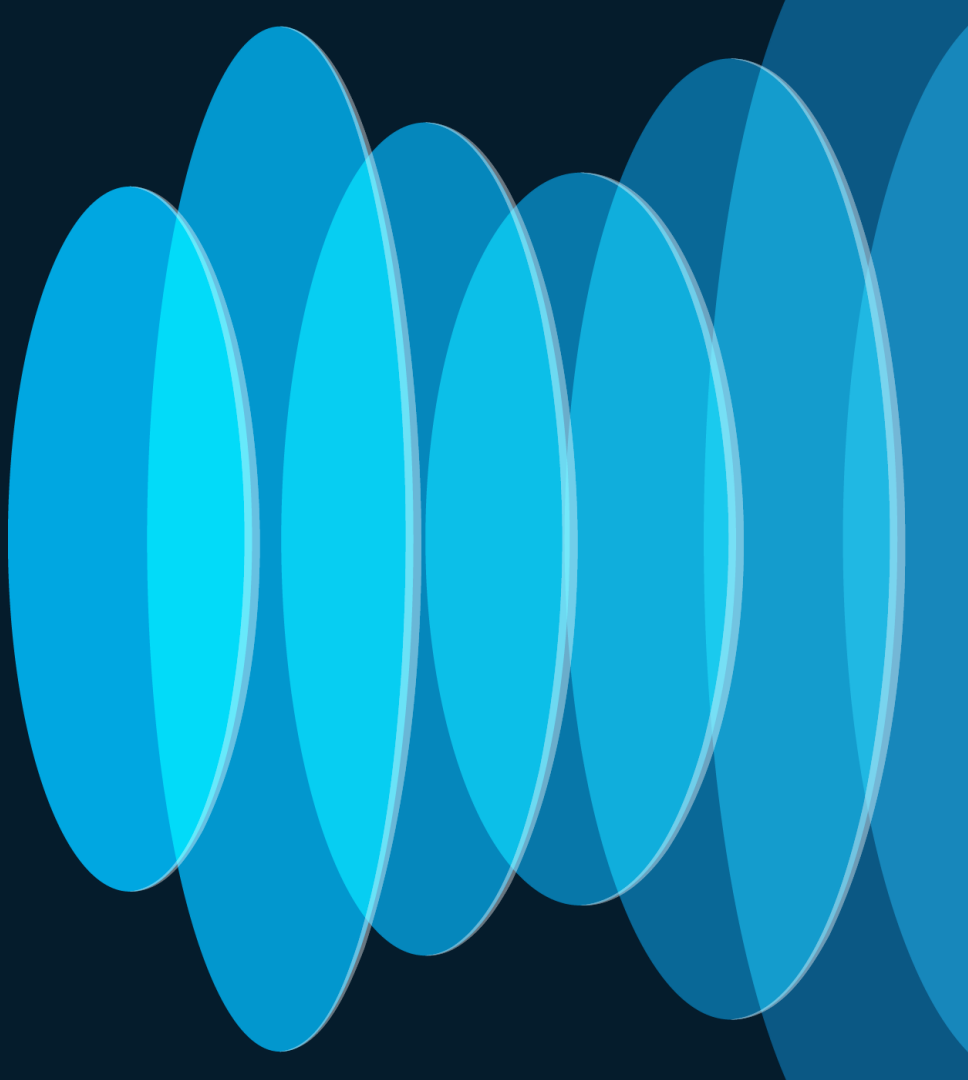
Sample ADT Event Output


```
"node_id_str": "PE4",
  "subscription_id_str": "app_1887_75f00000001",
  "encoding_path": "Cisco-IOS-XR-adt-oper:adt/adt-output",
  "collection_id": "9569581",
  "collection_start_time": "1607525488535",
  "msg_timestamp": "1607525488556",
  "data_json": [
    {
      "timestamp": "1607525488552",
      "keys": [],
      "content": {
        "adt-event": [
          {
            "event-id": 123,
            "change-description": "Traffic",
            "timestamp": "1607431905419",
            "change": [
              {
                "sensor-path": "Cisco-IOS-XR-infra-statsd-oper:infra-
statistics/interfaces/interface/latest/generic-counters/bytes-received",
                "sensor-path-tags": "interface-
name=GigabitEthernet0/3/0/19",
                "data": [
                  {
                    "value": {
                      "value-type": 8,
                      "val-counter64": "62808023132655"
                    },
                    "timestamp": "1607431545418"
                  },
                  ...
                  ...
                  {
                    "value": {
                      "value-type": 8,
                      "val-counter64": "62869633436614"
                    },
                    "timestamp": "1607432235421"
                  },
                  {
                    "value": {
                      "value-type": 8,
                      "val-counter64": "62872314602090"
                    },
                    "timestamp": "1607432265421"
                  }
                ]
              }
            ]
          }
        ]
      }
    }
  ],
  "collection_end_time": "1607525488556"
```

An LLM like GPT-4 “knows” IT and networking and can interpret different types of telemetry information.

AI for “open world, cross-domain insights”

The “TRAIAGE” Experiment



“Why doesn’t the Catalogue of my Webshop respond?”

Traditional Scenario

Customer/Partner IT Assistance Center

1st line Support Engineer

- Explore “bundle” that describes the issue and consult backend systems for more info
- Form initial diagnosis and dig further
- Domain-specific monitoring and expert systems assist with initial diagnosis

Customer LOB

“bundle”

App Monitor, Ops Assistance

Cloud Monitor, Ops Assistance

Network Monitor, Ops Assistance

Security Monitor, Ops Assistance

Data-Platforms

Applications

Cloud-Services (Databases, ..

Runtime (VMs, K8s, Docker, ..)

Server/Cloud - OS & Infra

Network Services (Firewall, ..

Core Network

Access/Wireless Network

Data-Sources

Can AI help the IT engineer with a rapid, triage-style diagnosis that is directionally accurate?

Customer/Partner IT Assistance Center

"Delegate" the task to create an initial diagnosis to a "virtual triage officer"

"Virtual Triage Officer"

- Analyze information sources (schema-less, multi-vendor)
- Identify entities, create hypothesis
- Determine dependencies between entities, refine hypothesis
- Hint at potential root cause

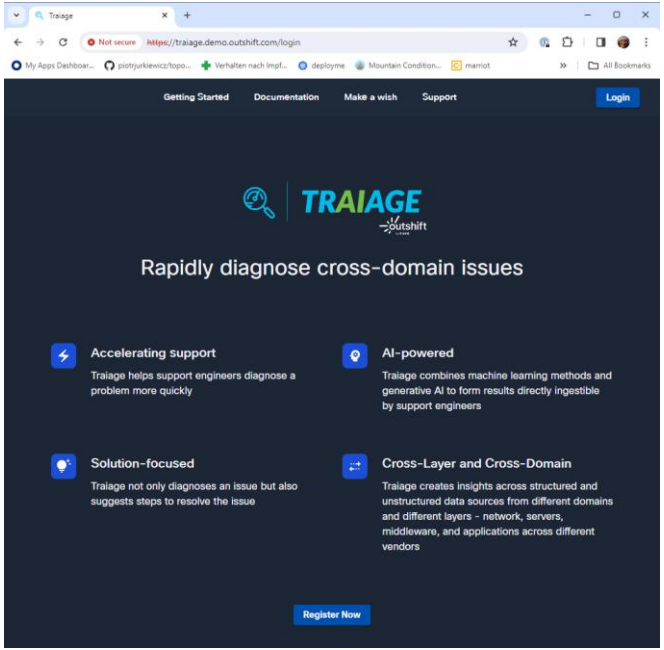
Customer LOB

"bundle"

- App Monitor, Ops Assistance
- Cloud Monitor, Ops Assistance
- Applications
- Cloud-Services (Databases, ..
- Runtime (VMs, K8s, Docker, ..)
- Server/Cloud - OS & Infra
- Network Services (Firewall, ..
- Core Network
- Access/Wireless Network
- Network Monitor, Ops Assistance
- Security Monitor, Ops Assistance

AI for Insights: Experimental Project TRAIAGE

- Rapid Response and Decision Support System (DSS) have historically been built using expert systems, that are limited to “known cases and scenarios” of the builder – and are expensive to maintain and expand.
- Artificial Intelligence can derive semantic dependencies dynamically and automatically.
- TRAIAGE infers relationships and dependencies from the data and arrives at insights, saving time and manual expert resources.
- Using Cisco data sources and purpose-trained models, TRAIAGE derives directionally accurate insights from structured and unstructured data.



TRAIAGE

Getting Started | Documentation | Make a wish | Support

FRANK BROCKNERS
CISCO

- Dashboard
- Datasets
- Settings

Dataset: cross-modality-example

Customer Problem Statement: The catalogue of the BestBags web application is not available.

- [ ] Use Cisco Data
- [ ] Specify Time Interval

ANALYZE

SOURCES

+ ADD SOURCES

Source	Vendor	Type	Entities	Actions
acb62039c3ec9356755776c95a550a179c87d8b3a6c28648cc3848d83d1a5a82.log	Cisco	log	docker	
d008ef4cf7f545b767ea425e03ef86e0c88cab2935fe88fe71c2ad4294936 2c9.log	Cisco	log	docker	
8250bd194b534dfcc1a4a5e8b68445a4813119544f3941fe800b9fea9f6698c2.log	Cisco	log	docker	
telegraf.conf	OSS	cfg	telegraf-plugin	
traiage-paris-k8s-n1.log	Cisco	log	traiage-paris-k8s-n1	
mysql.log	Oracle	log	MySQL Catalogue	
titan-install.log	Cisco	log	titan	
titan-manage.log	Cisco	log	titan	
telem-interface.txt	Cisco	cfg	titan	
vpp-config.cfg	Cisco	cfg	vpp	
vpp-custom-cmds.cfg	Cisco	cfg	vpp	
proxy.esl.cisco.com.log	Cisco	log	proxy.esl.cisco.com	

https://traiage.cisco.com:3000

Processing Approach

INPUT

Files Bundle

RELEVANCY

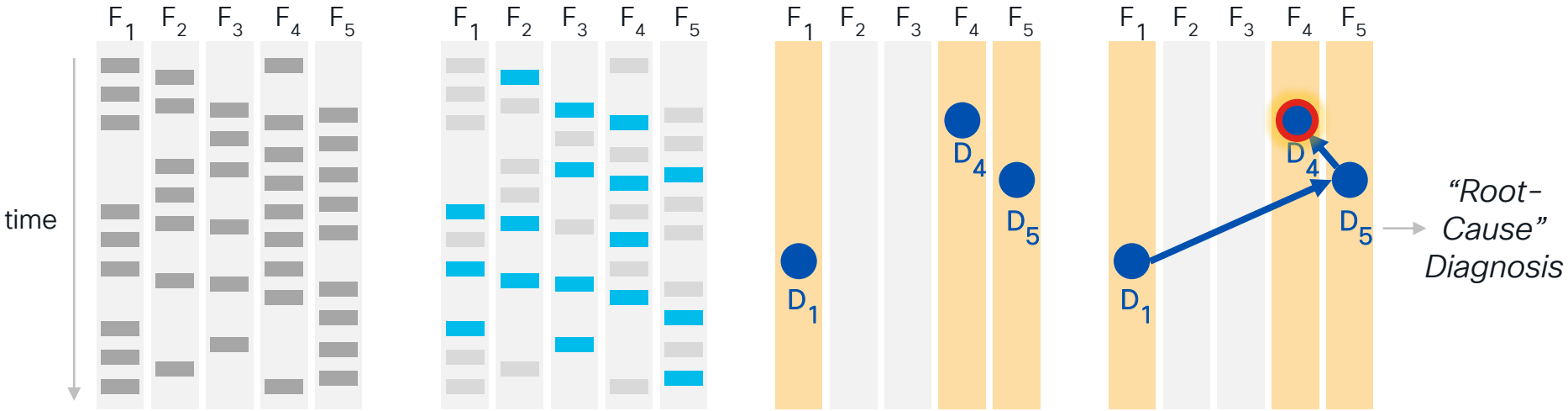
Which data (per file) is most information?

IMPORTANCE

Does the file describe an issue?

DEPENDENCY

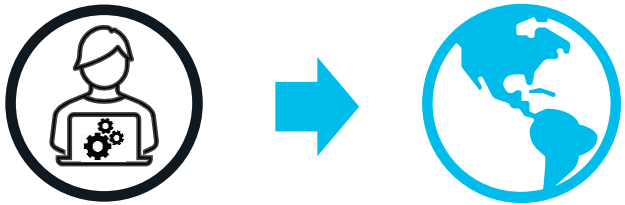
How do issues relate to each other?



TRAIAGE Approach

- Information retrieval from files in bundle
  - (for data w/o unknown schema): Combination of statistical and custom embeddings
  - (data w/ known schema): ELT platform loaders
- System of agents modeled as a series of small independent steps
  - Chain-of-Thought/ReAct-style reasoning
  - Task specific agents – grouped into voting ensembles
  - Programmatically defined pipelines to achieve stable, repeatable pipelines
- Cisco data sources to refine vanilla LLM responses
- Optimized system model using integrated benchmarking and RLHF

A glimpse at the future?!
LLMs can enable a paradigm shift in machine reasoning



Deductive
Fit the data to a human-defined schema
Upfront definition of structure and semantics

Closed World; Limited domain
Customized to use-case and deployment

Incremental extension of Ontologies

deterministic

Inductive
Derive the schema from data using GenAI
Natural language query of unstructured data

Open world
Generically applicable

Exponential growth of corpus (“world data”)

probabilistic

Do you want more?

BRKETI-2397 “Can we develop faster with CoPilot, ChatGPT & Co”?

Tuesday, June 4 – 2:30pm, L2, Room Mandalay Bay I

GenAI Solutions for the Enterprise

Explore the latest advancements in driving secure, private, and trustworthy GenAI adoption in the enterprise

START

Monday June 3 | 11:00 am
BRKETI-1005
An Introduction to GenAI Technologies & Solutions

Monday June 3 | 1:00 pm
CENETI-1001
Is Your Organization AI-Ready? How to Deploy and Manage AI-Powered Technologies Rapidly and Safely

Monday June 3 | 3:15 pm
ITLBRK-1113
Strategies for Safer, Faster Deployment of GenAI Apps & Services at Scale

Tuesday June 4 | 2:30 pm
BRKETI-2397
Can We Develop Faster with ChatGPT, CoPilot, and Co?

Wednesday June 5 | 11:00 am
PSOETI-2000
Turbocharge the delivery of secure, compliant, and responsible Generative AI capabilities across your organization -

Wednesday June 5 | 3:00 pm
DEVWKS-3003
Motific: Accelerating Delivery of Trustworthy GenAI Capabilities Across Your Organization

FINISH

Thursday June 6 | 9:30 am
BRKETI-2010
Beyond the Noise: Harnessing Generative AI for Telemetry Data

cisco Live!

Complete Your Session Evaluations

Complete a minimum of 4 session surveys and the Overall Event Survey to be entered in a drawing to **win 1 of 5 full conference passes** to Cisco Live 2025.

Earn 100 points per survey completed and compete on the Cisco Live Challenge leaderboard.

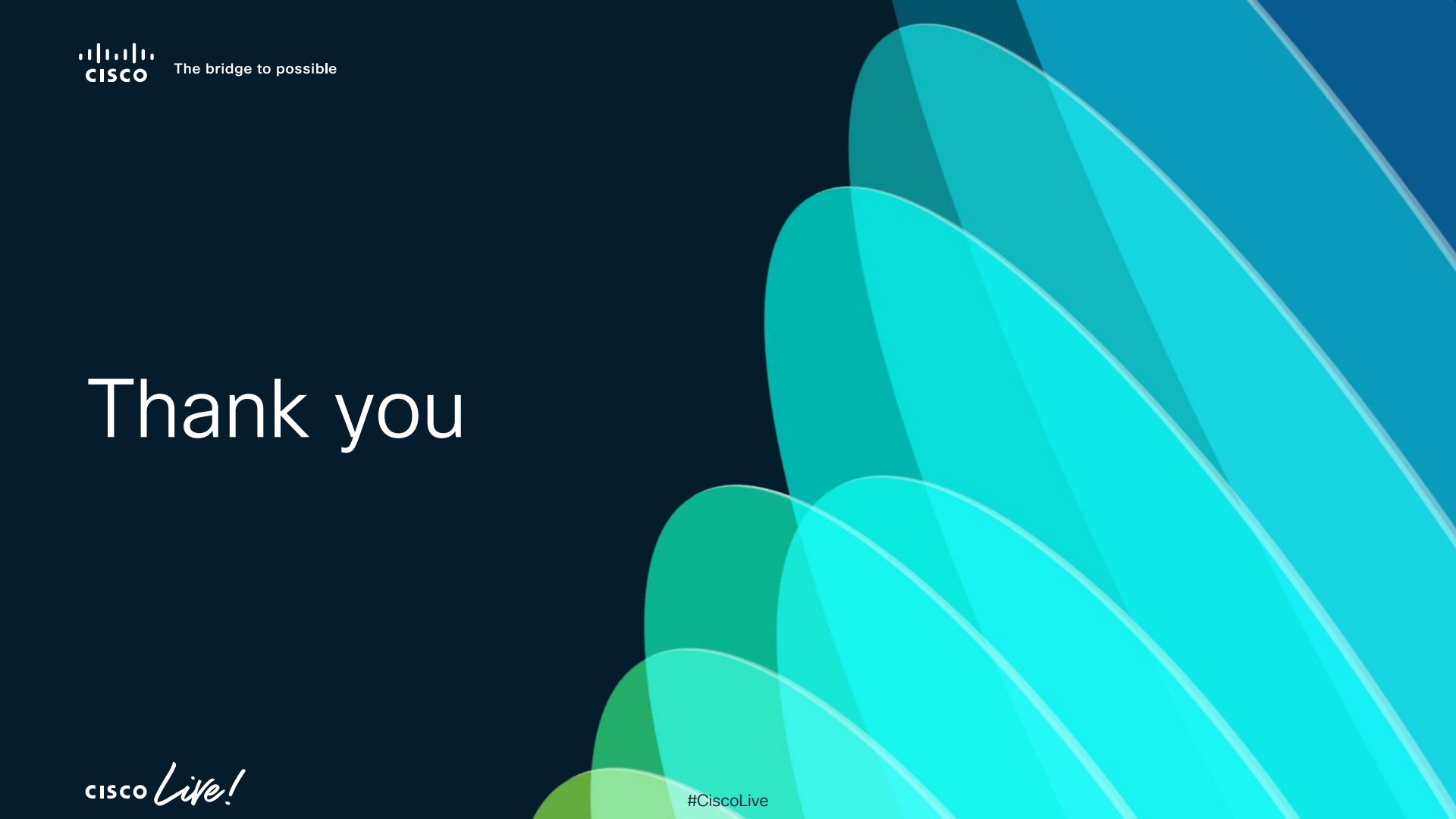
Level up and earn **exclusive prizes!**

Complete your surveys in the **Cisco Live mobile app.**

Continue your education

- Visit the Cisco Showcase for related demos
- Book your one-on-one Meet the Engineer meeting
- Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs
- Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

Contact me at fbrockne@cisco.com


The bridge to possible
Thank you
cisco Live!
#CiscoLive