| YAHOO! | TechCrunch | IHUFFPOSTI | YAHOO! | MAKERS | YAHOO! | verizon ⁷ digital media services | tumblr. | | YAHOO! | RYOT | engadget | ҮАНОО! | Aol. | BrightRoll |
|--------|------------|------------|--------|--------|--------|---|---------|--------|--------|------|----------|--------|------|------------|
| | | | | | | | | by AoL | | | | | | |

Panoptes: A Network Telemetry Ecosystem - Part Deux

Varun Varma, Sr. Principal Engineer vvarun@oath.com

Oath:

Panoptes is:

- Greenfield Python based network telemetry platform that provides real time telemetry and analytics @ Yahoo
- Implements discovery, polling, distribution bus and numerous consumers



Architecture

•



Core Concepts

- Plugins
- Resources
- Metrics
- Enrichments
- Data encoding and distribution

4

Plugins:

- Are Python classes conforming to a well defined API
- Can collect/process and transform data from any source
 - SNMP
 - API
 - CLI
 - 0
- May be of three types:
 - Discovery
 - Enrichment
 - Metrics

Resources:

- Are abstract representations of *what* should be monitored
 - In the context of network telemetry, these would be usually be the network devices to monitor
- Are 'discovered' using discovery plugins
 - Which, usually, would would talk to a Configuration Management Database but could also be from topology walks
- Have and id, endpoint and various metadata
 - For example, the vendor name or operating system version of a device would be it's metadata
- Are specified within Panoptes with a DSL
 - Example: "resource_class" = "network" AND "resource_subclass" = "switch" AND "resource_type" = "cisco" AND "resource_metadata.os_version" LIKE "12.2%"

Metrics:

- Are numbers that can be measured and plotted
 - Example is the bytes in/bytes out counter of an interface
- Are generally fast changing
 - Or have the potential to do so
- Can be collected through various means:
 - SNMP
 - API
 - CLI
 - Streaming

Enrichments:

- Are metadata in addition to metrics
 - For interfaces, we collect metrics like bytes in and bytes out and enrichments like interface name and description
- Can be any data type
 - Unlike metrics which can only be numeric
- Can come from sources other than the device being monitored
 - The geo location of the device or the ASN number to name mapping

Enrichments - pt. 2:

- Usually are more expensive to process than metrics
 - Might need complex transformations
 - And therefore:
 - Are collected at a rate less than those for metrics
 - We collect interface metrics every 60 seconds, but enrichments every 30 minute
 - Are cached
- Overall, let us scale more by being efficient about data collection

Data Encoding and Distribution

- Panoptes is a distributed system
 - Discovery, enrichment and polling are all decoupled from each other
- Kafka and/or Redis are used to pass data between all subsystems
 - This makes it so that you can extend or introspect any subsystem
- JSON is used to encode all data within Panoptes
 - It's non-performant but developer/operator friendly

Workflow



Scaling and Operations

Scale: Orders of Magnitude

100K 10K Network Devices

100 Network Sites



Servers

60 Seconds

10 Systems Replaced

•

Scaling Issues

- Panoptes was built to be horizontally scalable and free of single points of failure from day one
 - Performance or high-availability are not easy to bolt on afterwards
- We chose Python to be developer friendly it was every bit as slow as we thought
 - High throughput actions are delegated to C extension modules
 - Ditto for JSON serialization for all data
- We broke everything Redis, Zookeeper, Kafka
 - Redis allows 'only' 10,000 clients to be connected by default :)

Divide and Conquer: Federated API

- Due to availability concerns, each site has its own MySQL cluster
 - Telemetry data must be available during a network partition
 - Centralized telemetry store might not be reachable in all cases
- Each API endpoint acts as a tribe node
 - If a tribe node doesn't have the requested data, it returns a pointer to the node that does through a find API



Covered Systems

- Interface metrics for Arista, Cisco, Juniper
- System metrics for A10 (AX, TH), Arista EOS, Brocade TrafficWorks, Cisco IOS-XE, Juniper (MX, SRX)
- Functional metrics for VIPs (A10 AX, TH, Brocade), A10 LSN, Juniper SRX
- Ethernet state and topology for Cisco IOS, NX-OS

Operational Experiences

- Metrics across different platforms or versions of even the same OS from vendors aren't consistent
 - Normalizing these metrics was our single biggest time drain
- SNMP has it's faults, but is still ubiquitous
 - Specially so in a multi-vendor, multi-platform and multi-generational networks such as ours
- Performance of APIs was much better than SNMP
- We knew that we didn't know how we would use the data
 - \circ Using Kafka proved to be the right choice we already have 3 separate consumers

Operational Experiences - Pt. 2

- We don't expose 'raw' data to external systems
 - It's tempting to give access to external teams via Kafka, but that would lead to friction if we want to change our internals
 - Instead, we expose APIs which abstract away all our internals
- Custom UIs are useful
 - And enabled by APIs

API Examples

```
Realtime – Purpose
                                                                                                                                                                                                                                                                                                                        Bulk/Historical - Generic
       Specific
▼ {
              "members_metrics": [
                                                                                                                                                                                                                                                                                                                             - {
                                                                                                                                                                                                                                                                                                                                          - aggregateTags: [
                            V {
                                                                                                                                                                                                                                                                                                                                                              "_aggregate",
                                                                                                                                                                                                                                                                                                                                                           "resource_endpoint",
                                                     "load balancer model": "
                                                                                                                                                                                                                                                                                                                                                          "resource_site",
"vip_type",
                                                     "weight": 1,
                                                                                                                                                                                                                                                                                                                                                          "real port",
                                                     "site": "
                                                                                                                                                                                                                                                                                                                                                          "vip_property"
                                                                                                                                                                                                                                                                                                                                               1,
                                                     "vip": "not planting of the second of the se
                                                                                                                                                                                                                                                                                                                                       - dps: {
                                                                                                                                                                                                                                                                                                                                                         1525809840: 100000
                                                    },
                                                     "vip property": """,
                                                                                                                                                                                                                                                                                                                                             metric: "${Panoptes.network-load-balancer-vip.real_max_connections}",
                                                                                                                                                                                                                                                                                                                                        - tags: {
                                                     "max connections": 100000,
                                                                                                                                                                                                                                                                                                                                                         vip_protocol: "tcp",
                                                                                                                                                                                                                                                                                                                                                         vip_ip_address_version: "4",
                                                     "bytes_in_gauge": 802742,
                                                                                                                                                                                                                                                                                                                                                        vip_port: "9999",
real_dns_name: """,
                                                     "bytes out gauge": 0,
                                                                                                                                                                                                                                                                                                                                                         vip_dns_name: "
                                                     ١.
                                                                                                                                                                                                                                                                                                                                              __groupId_: "real_dns_name: |vip_dns_name: |vip_ip_address_version:4|vip_port:9999|vip_protocol:tcp"
                                                     "polling interval": 60,
                                                                                                                                                                                                                                                                                                                                   },
                                                     "active_connections_gauge": 24307,
                                                     "vip_port": 443,
                                                     "status": 0,
                                                     "pool name": "generation of the second secon
                                                     "packets out gauge": 0,
                                                     "timestamp": 1496772838,
                                                     "real port": 443,
                                                     "vip_type": "13dsr",
                                                     "packets in gauge": 4221,
                                                     "cache age": 41,
                                                     "name": "applicable play hell parameters",
                                                     "connections per second gauge": 281,
                                                     "total connections counter": 746440138,
```

Centralized Telemetry

- We push metrics to our in-house time series database and alerting service (centralized telemetry)
- Custom dashboard service our user base is familiar with
- Economies of scale no need to provision new hardware or software



Here we see control and data plane CPU statistics for a load balancer in one of our West Coast data centers.

Custom UI Examples

| sibi11g | er pr alle a calendaria | | ٩ | | | | | |
|-------------|--|--|--|----------|--|--|--|--|
| Links | 8 | SYSTEM SERVICE GROUPS | | | | | | |
| 5 0 1 | Sentem Metica 🔍 XESMan Sentex Lesues S Gusta Saluek - Presetes | Leaf Brances Wirksp * Device Merics | Proceedings of the second seco | | | | | |
| © ₩ | RTMLE while souther an and See Mi Make | correction to the second secon | | | | | | |
| 0 | Model | 06:00 01:00 11:0 | 88 12:00 14:00 16:00 | | | | | |
| × . | 05: | Data CPU % Fars OK % | | 0000 | | | | |
| - | Collector: Environment | | | | | | | |
| BD ► | Automation: ENVAREED Banner: DEDICATED BID VIP | VIPs | | | | | | |
| Prop | Serties Constall-Advigs 500 125 % | Name 🛧 | Search Propeny | Q | | | | |
| | Malening Grandel Gill 113 | | Addenving PEP (US) Isign (US) Curvetall AdvApp (US) Addiening Curvetall (US) | | | | | |
| | RC_Control_Times (TW): 12.5 S | NAMES OF A DECK | EC, Central, Tech (TN) Addserving PBP (US) | | | | | |
| | Regin (US): 25.0 N | and desired many | AdvServing PBP (US) | | | | | |
| | | Sphilars.com | login (JS) Rova per paper 💌 1-0 | iora < > | | | | |



•

Future: Streaming Telemetry

Proposed Architecture



A recap: we were here at NANOG 70...

Since then:

- Added an enrichment subsystem
- Built many plugins to poll system and functional metrics
- Worked out scaling issues
- And most importantly...

I'm here to announce open sourcing of Panoptes

MVP

- Core Platform including discovery, enrichment and polling
- Interface metrics and enrichment plugins
 - Also heartbeat plugins
- Integration with InfluxDB

An example dashboard



Get it at: https://github.com/yahoo/ panoptes

Questions?