Aruba AMON based analytics 1

Rasa cloud questions: 1

Rasa Cloud project challenges 2

Aruba Cisco Analytics 2

Theory of operations 2

**Check out Google-Drive/Tech/Projects folders for more details**

# Aruba AMON based analytics

1. AMON messages were getting generated from different modules like following.

* Dot1x/wpa-exchange modules on controller
* DHCP module in controller
* 801.11 Auth/Assoc/Roaming module in AP
* DNS query response
* RX/TX stats from Aps for station

1. OPMS at customer premises will collect these statistics from the controller/AP, filter out unwanted one, compress and forward the rest to the cloud over open-vpn tunnel.
2. Scala based gateway and AWS cloud will listen for these traffic and put these on kafka topic partitioned on controller IP. Also it will save these info into a separate seq file.
   1. What is the format, is it google protobuf format ?
   2. How the kafka topic stuff works ?
3. Conn\_amon scala application will subscribe to the live Kafka topic
   1. How the state machine is run

# Rasa cloud questions:

1. What is the storage we have for Cassandra DB(40 node, each 1TB, doesn’t use hadoop, You can add more capacity by adding one more node), Parquet(hadoop storage, here 15TB is available along with processing node), processing node (total 30 nodes, each has 1/2 TB storage so total 15TB storage(this storage is used by hadoop)). Processing nodes (30 Nodes used in Hadoop) are copy job, batch analytics job, Streaming mode/Kafka topic jobs.
2. At what rate it is getting depleted ?
3. At what rate sequence files are depleting the storage (it is also on hadoop)
4. What all technologies are used ?
   1. Cassandra, Parquet for DB
   2. Scala for analytics clients
   3. Spark – big data processing
   4. Kafka for streaming data
   5. Zoo keeper (i think I used to run kafka using zoo keeper)
   6. Oozie for activity management, do you write Ansible script for this ?````````````````````
   7. Yarn (for scheduling/restarting batch jobs)
   8. Hadoop for log file, analytics client code deployment, parquet storage
   9. Data dog (used for monitoring the resource usage, number of DB transactions, CPU utilization, memory utilization.
   10. Amazon AWS vr S3
5. What rate each site are pumping the data to our cloud (compressed data 7-8 mbps, uncompressed around 25-30 mbps from each site ?)
6. How many nodes of cassandra (40) and analytics client(30) are there
7. What is the interface between OPMS and Cloud - Open ssl session(encrypted channel), our own TCP connection.
8. Check cloud architecture document

# Rasa Cloud project challenges

## Scaling issues faced

30 days connectivity data aggregation for insight table. Insight table was saving connectivity insight events for every 1 hour. We need to aggregate to 24 hour, then 30 days. Once we were doing 30 days event aggregation, task was taking 6 hr of computation. The reason following.

1. Each insight (say dot1x failure) has to aggregate (how many occurrences of this incident/insight/issue) that failure incidents on following parameters
   1. For each Aps, how many occurrence of this inside(failure)
   2. Based on AP model
   3. BSS (i.e. each radio in each AP)
   4. For each Channel
   5. For each station
   6. For each location
   7. For radio type
   8. For particular RADIUS server
   9. For particular controller

For hourly aggregation itself was 100 Aps, 1000 stations. Aggregating to 24 hr was fine, however then 30 days aggregation was breaking the system. We need to report how many unique stations were affected, how many unique Aps/BSSIDs were affected during 30 days.

Deployment of following size for one university i.e University of Washington

1. 120k stations
2. 10k Aps
3. 100 controllers

Each of above site was pushing 30mbps of Data to cloud (compressed 8 mbps).

Like above we have 5 University deployment and then Aruba’s internal networks. (multitenancy)

**Not running separate job for each tenant/timezone:**

Since it was multitenancy, each folk’s start and end time was different. We don’t want to run separate jobs for each tenant or timezone. So there was complexity in doing this kind of computation. Able to do this using aggregation, union, set, filter functionalities of RDD of spark infra.

## Complexity faced

1. SNMP integration
2. Network sniffing (needs DPDK)
3. CLI input
4. Logs input
5. State machine design for each connection attempt based on above events

There was no transaction-id or session-id concept in above all messages to relate all these events to apply to particular Station FSM

# Aruba Cisco Analytics

This analytics software development was done with Cisco controller. Following were sources of information.

1. SNMP MIBs (Not used in Aruba solution)
2. CLI show commands
3. Capturing north bound traffic of the controller
4. Aruba had AMON data coming out from AP/Controllers (Aruba specific)
5. DHCP/CPPM/RADIUS log collectoin

For #3, we put one on-premises equipment to capture Cisco controller traffic (Controller’s all RX/TX packets will be mirrored to this equipment). This equipment was cento-os and software were using pcap. Later plan was to use Intel DPDK. On premises software was written in C, the analytics software was in python. DB was mongo-db.

For Aruba setup, #2,#4, #5 traffic was collected by a Linux box in the premises. This box’s job to filter out required info and send to the AWS cloud.

## Theory of operations

Capture following packet at controller’s north bound

1. Station’s rx/tx 802.11 management frame sent over CAPWAP header (between Controller/AP)
2. Station’s EAP/DHCP/DNS frames sent over CAPWAP header(between Controller/AP)
3. Station’s Data packets
4. Controller and RADIUS/DHCP/ interaction

## Challenges

1. Scalability
2. Accuracy of the data as relating different events from DHCP/Dot1x/AP/Controller/DNS/RADIUS to one incident/attempt is difficult. There are no transaction id etc.

# Technical Design decision taken by me, which was not good

Doing FSM for all station for all attempts. The FSM was to take care of whole cases in details. This caused major issue in actual UW deployment as many message from above modules were missing.