Arista AVA: The Power of AI-Driven Networking

Table of contents

Introduction

Artificial intelligence (AI) has been all over the news especially as large language models like GPT-4 capture the imagination of even the average internet user, let alone the technophiles. These solutions have the potential to change the very definition of the enterprise, what it means to work, and lower the cost of doing business. In parallel, the world is producing data at a blistering pace. A World Economic Forum[1] study estimates that by 2025, we will create 463 exabytes of data daily. For context, that's 463, followed by 18 zeros! Or as the authors put it, "Forty times more bytes than there are stars in the observable universe." Of course, what isn't likely to happen in that same time frame is that humans will suddenly evolve to consider all of that available data, factor probabilities, and make optimal operational choices. That is not to suggest that humans are becoming expendable. Far from it, we have highly tuned abilities to recognize patterns[2], understand abstract relationships, and generalize. For instance, we don't need hundreds or thousands of training samples to know that a user being targeted by a phishing attempt is the CFO of the organization. We recognize the name right away and our instincts take over to drive remediation next steps. The task at hand then for the AI tools, is to surface just the right information that enables operators to make decisions which ultimately lead to the most favorable business outcomes.

When it comes to IT and security use cases, the network can be foundational to generating the decision support data operators and analysts need. But at the same time, as networks have expanded from campus to data center and cloud, the amount of data to be processed has grown exponentially. So how do you get to the Goldilocks balance of "just right" information? It comes down to two key capabilities:

- A system to efficiently and in real-time collect the ground truth data.
- Intelligence to extract the information and context buried within the raw data and present it to the operator.

Arista is uniquely positioned to deliver on both of these capabilities. Arista EOS® based on our network data lake (NetDL™)[3], provides a multi-modal, multi-tenant-capable data lake that offers real-time network telemetry to other Arista solutions as well as those from our partners. Arista Autonomous Virtual Assist (AVA) uses an AI-driven approach to anticipate operator questions, extract answers from NetDL and deliver the insights necessary for effective human decision-making. With this combination, Arista is providing networking for the data-driven enterprise.

In this paper, we will share how AVA uses cutting-edge AI models to solve real-world challenges for network and security operators and present specific case studies that illustrate the business value the approach can generate.

[1] https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/

[2] https://arxiv.org/ftp/arxiv/papers/1801/1801.00631.pdf

[3] https://www.arista.com/en/solutions/cloud-networking

AVA Architecture

Arista AVA is an AI-enabled decision support system that combines cloud scalability with the codified expertise of real-world network and security operations experts. Real-time, complete data, including application, in-band network telemetry, flow visibility, and complete control plane state incorporated into the Arista EOS and NetDL stack, constantly serve as the basis to train the AVA AI/ML models. These data originate from a wide range of sensors and devices, physical or virtual, or through direct integrations with network and cloud infrastructure components such as the Arista DANZ Monitoring Fabric (DMF) or span ports on Arista EOS platforms.

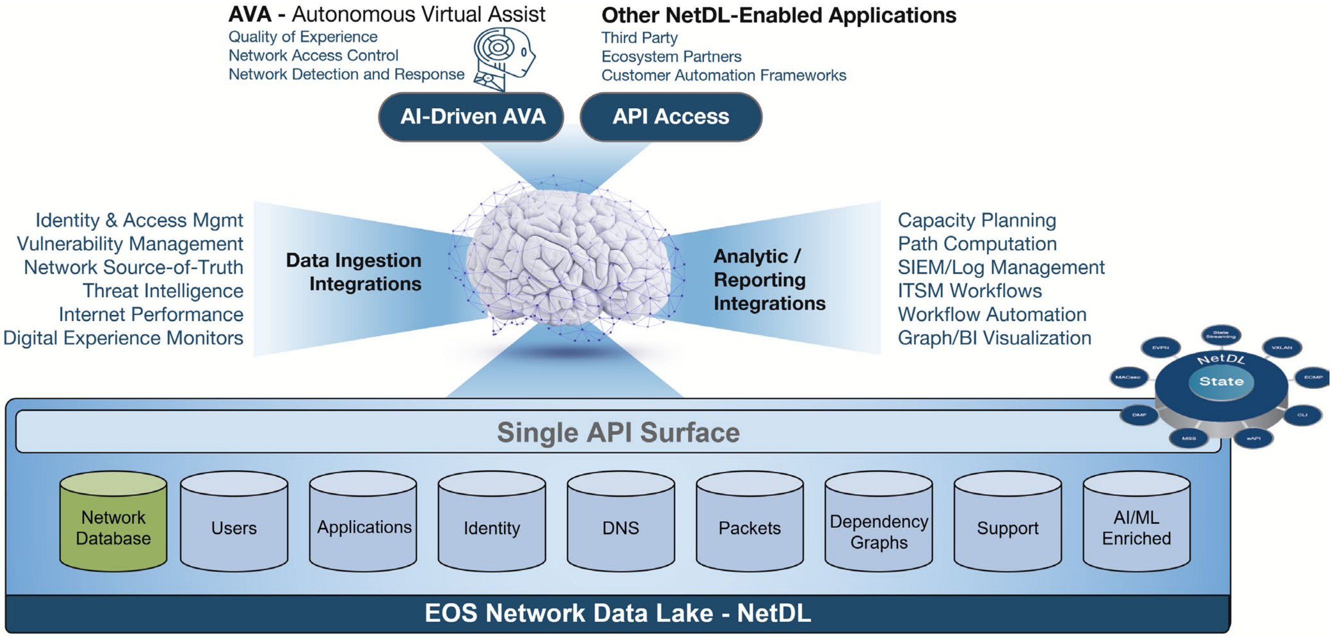


Figure 1: The Arista Data-Driven Cognitive Architecture

Based on the specific use case, AVA processes ground-truth data about the network devices' state and, if required, raw packets to pre-compute answers for questions a highly skilled analyst would ask. AVA thus surfaces the weak and early signals of a network issue, along with corroborating evidence to establish conviction. By that same token, AVA also eliminates those signals that cannot be corroborated so that human analysts avoid wasting valuable cycles. The net result is that the operations team is in a far better position to act—whether by proactively identifying connectivity issues or disrupting an adversary's objectives at the outset because AVA identified the initial warning signs of a ransomware attack. Our field results show that AVA frequently finds more incident-related activity than a senior human operator analyzing the same activity.

The AVA Advantage

Legacy data science approaches run into some significant challenges given the scale and diversity of the modern network. Any AI model is only as effective as the labeled samples used to train it. However, volumes of data and the types of conditions on even the simplest of today's networks make effective labels scarce. In addition, given how aspects such as threats evolve rapidly, the labels themselves become obsolete quickly, requiring frequent retraining and operational costs entailed in that effort. Even if these challenges are overcome, the resulting models are often massive, monolithic, and opaque, making them less useful, as they are slow and not explainable. In other words, a human analyst seeing the result often has no idea why something is flagged as a network or security issue. Consequently, neither do they feel confident in the veracity of the information, nor do they know what they should do next (Figure 2).

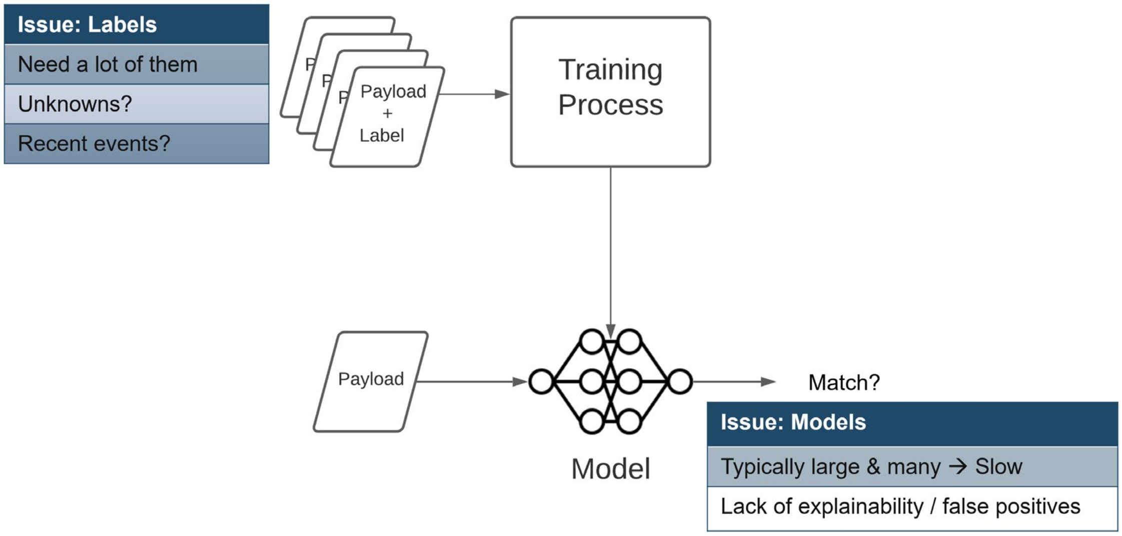


Figure 2: Challenges with legacy artificial intelligence approaches applied to network data

Even though one of the goals is to find the outliers and the anomalies, AVA doesn't approach this problem like legacy AI solutions. Instead, AVA starts by learning what is normal and, dare we say, mundane. In most networks, this includes traffic such as email, patching, and even department or workgroup-specific application usage. As you can imagine, there is an abundance of labels to train AVA's ML models to find these kinds of network behaviors. Unlike the approach described in Figure 2, which attempts to learn what "bad traffic" looks like, AVA works to eliminate much of the hay from the proverbial haystack of network data. This, in turn, leaves a significantly smaller consideration set within which AVA can then look for the needles.

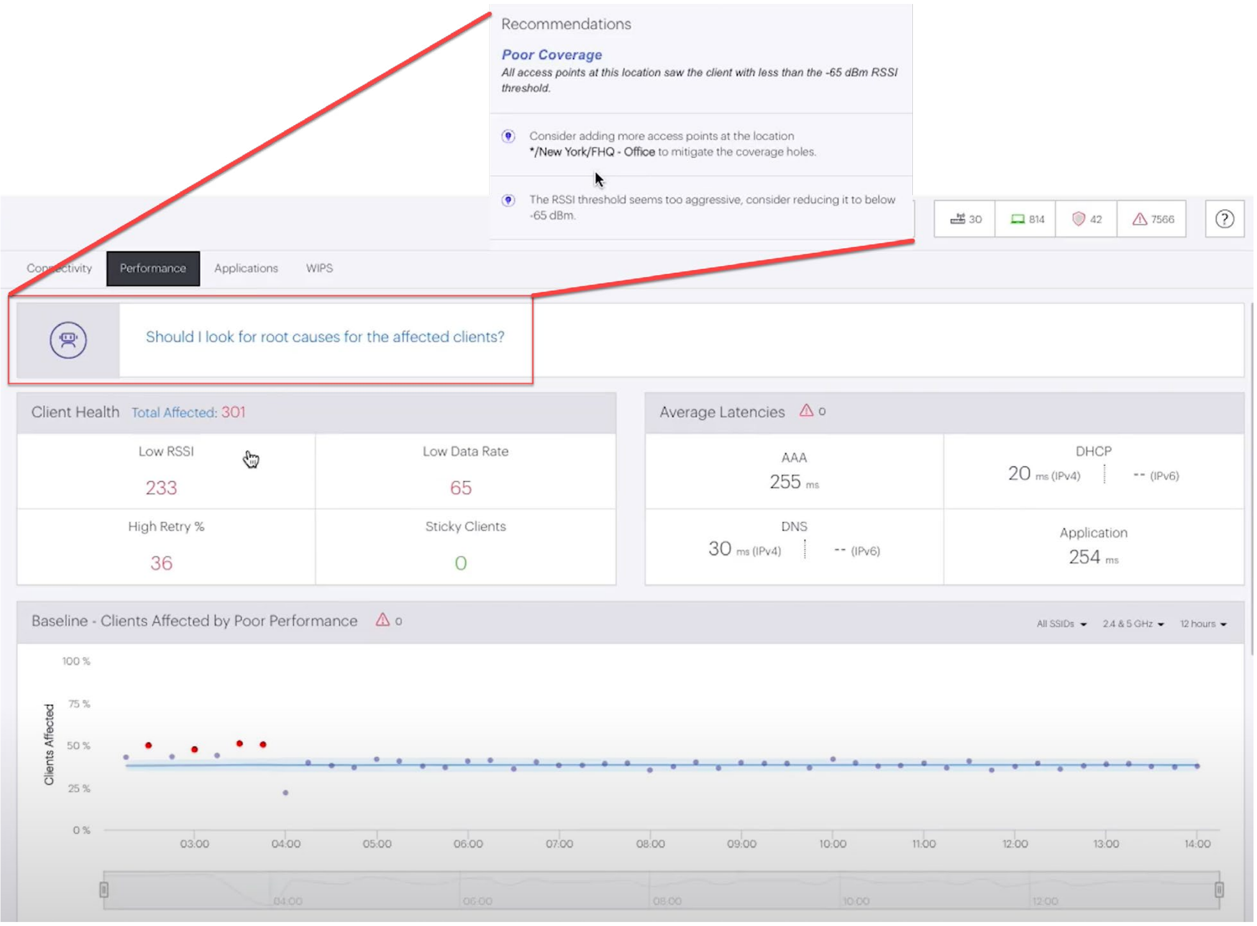
Core to AVA's success with processing the voluminous data in NetDL is the knowledge graph. The data AVA consumes is first processed with techniques to discover the entities—people, devices, VLANs, applications, etc.—and the relationships between them. This forms the core of the knowledge graph. The graph can then be enhanced iteratively by knowledge from domain experts in the form of heuristics, as well as ongoing feedback from human operators. Given the size of the knowledge graph is significantly smaller than the raw features, additional AI approaches can be applied iteratively. This not only allows for better AI analysis but, unlike traditional ML models, with AVA, the algorithmic outputs are captured as real-world entities and their properties. The human-readable and understandable outputs deliver explainable AI with clearly defined next steps for the analyst.

AVA Delivers Solutions

The data-driven architecture, coupled with artificial intelligence, enables several network and security operations use cases. This section will provide a few examples of how AVA is optimizing workflows, speeding mean time to resolution, and improving security outcomes.

Case Study 1 – Proactive Network Operations

AVA enables network reachability modeling to deliver proactive notifications when a specific network service or application is experiencing reachability issues. Using dynamic anomaly detection, AVA identifies anomalies based on deviations from a learned reachability/latency baseline. The historical bounds and anomaly scores adapt to normal variations as time goes on. Access to NetDL's historical data supports both real-time and forensic troubleshooting of any issues identified.



Recommendations

Poor Coverage
All access points at this location saw the client with less than the -65 dBm RSSI threshold.

Consider adding more access points at the location ***/New York/FHQ - Office** to mitigate the coverage holes.

The RSSI threshold seems too aggressive, consider reducing it to below -65 dBm.

Should I look for root causes for the affected clients?

Client Health Total Affected: 301

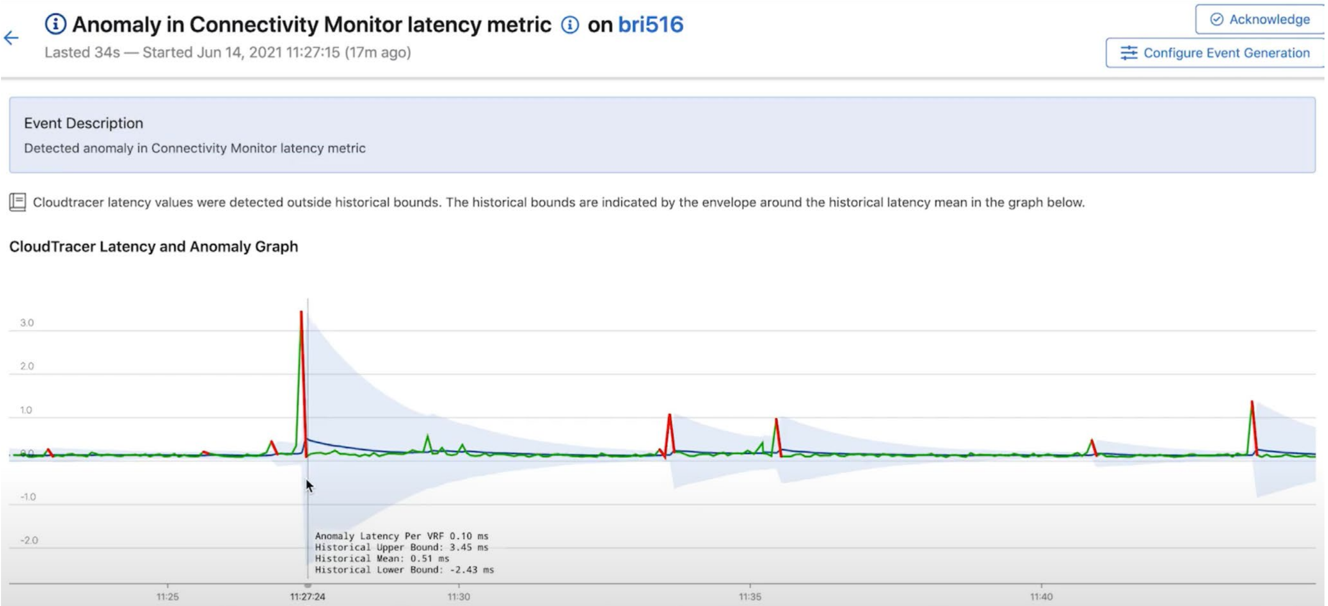
Low RSSI	Low Data Rate
233	65
High Retry %	**Sticky Clients**
36	0

Average Latencies ⚠ 0

AAA	DHCP
255 ms	20 ms (IPv4) -- (IPv6)
DNS	**Application**
30 ms (IPv4) -- (IPv6)	254 ms

Figure 3: AVA proactively identifies network and application reachability issues

AVA also assists with another common network operation challenge: switch table overflow, resulting in a network outage. With the myriad chipset implementations and associated hardware table capacities, it is difficult for operators to monitor, track and extrapolate utilization trends manually. Instead, AVA models hardware resource utilization growth trends, enabling predictive assessments and notification ahead of exceeding capacity. Customers use these notifications to take preventive measures and avert a crisis.



Anomaly in Connectivity Monitor latency metric on bri516

Lasted 34s — Started Jun 14, 2021 11:27:15 (17m ago)

Event Description

Detected anomaly in Connectivity Monitor latency metric

Cloudtracer latency values were detected outside historical bounds. The historical bounds are indicated by the envelope around the historical latency mean in the graph below.

CloudTracer Latency and Anomaly Graph

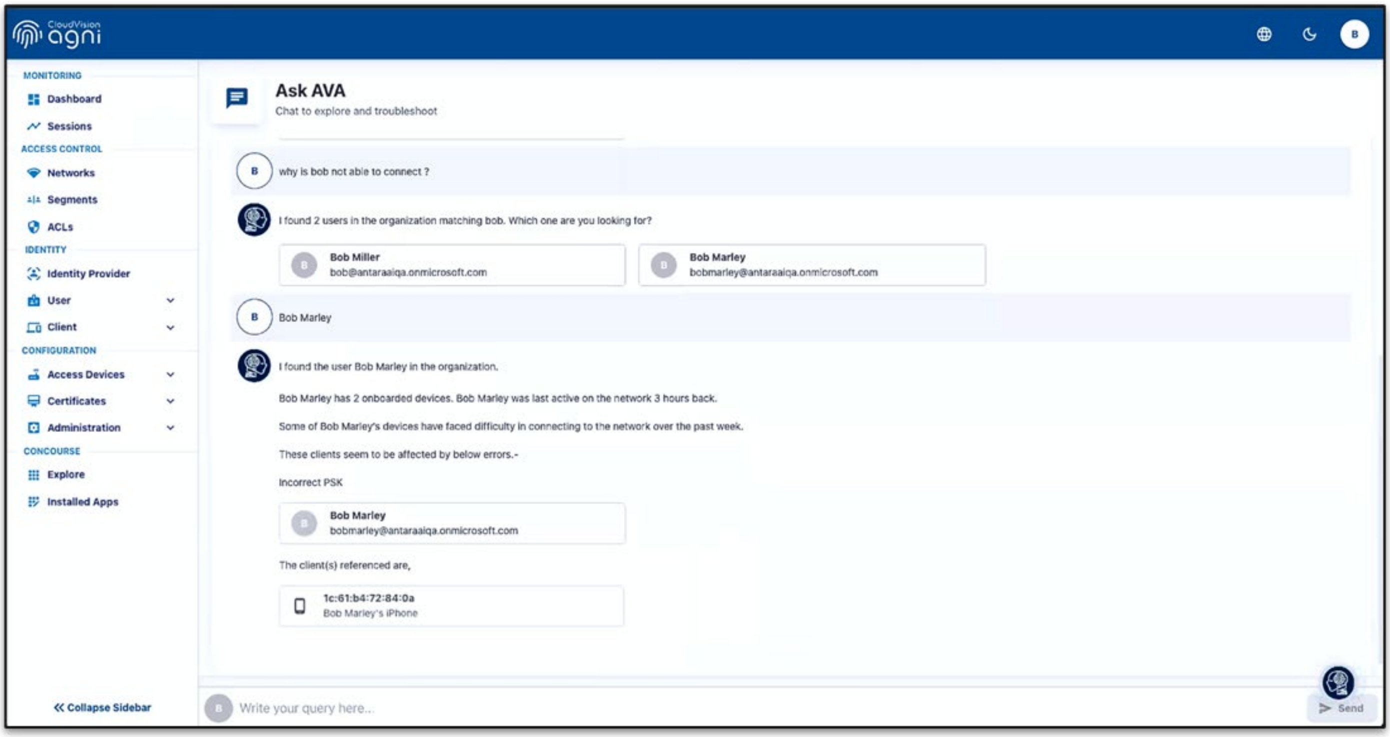
```
Anomaly Latency Per VRF 0.10 ms
Historical Upper Bound: 3.45 ms
Historical Mean: 0.51 ms
Historical Lower Bound: -2.43 ms
```

Figure 4: Detecting anomalies from a baseline enable AVA to identify grey failures

Case Study 2 – AIOps for Network Access Control

Deploying and managing network access control (NAC) solutions have historically been cumbersome and error-prone, requiring expensive and hard-to-find human experts. Arista CloudVision AGNI innovates in this area by automating and optimizing key workflows, including providing a conversational AI capability called Ask AVA for configurations, troubleshooting, and simulations.

Ask AVA uses various AI techniques including a generative pre-trained transformer (GPT)-based approach for training purposes. AGNI has two sets of models: one based on classic machine learning uses support vector machines while the other uses ChatGPT's large language models via OpenAI APIs. Ask AVA uses a chat-like interface and natural language processing (NLP) to aid even a junior analyst with tasks such as configuring, troubleshooting, and analyzing policy configurations. From the NLP queries, AVA autonomously identifies the analyst's intent and assists in configuration, provides contextual output to troubleshoot problems, and analyzes the correctness of network and security policies. The net effect is a simpler and more secure NAC deployment.



Ask AVA

Chat to explore and troubleshoot

why is bob not able to connect ?

I found 2 users in the organization matching bob. Which one are you looking for?

Bob Miller
bob@antaraaiqa.onmicrosoft.com

Bob Marley
bobmarley@antaraaiqa.onmicrosoft.com

Bob Marley

I found the user Bob Marley in the organization.

Bob Marley has 2 onboarded devices. Bob Marley was last active on the network 3 hours back.

Some of Bob Marley's devices have faced difficulty in connecting to the network over the past week.

These clients seem to be affected by below errors.-

Incorrect PSK

Bob Marley
bobmarley@antaraaiqa.onmicrosoft.com

The client(s) referenced are,

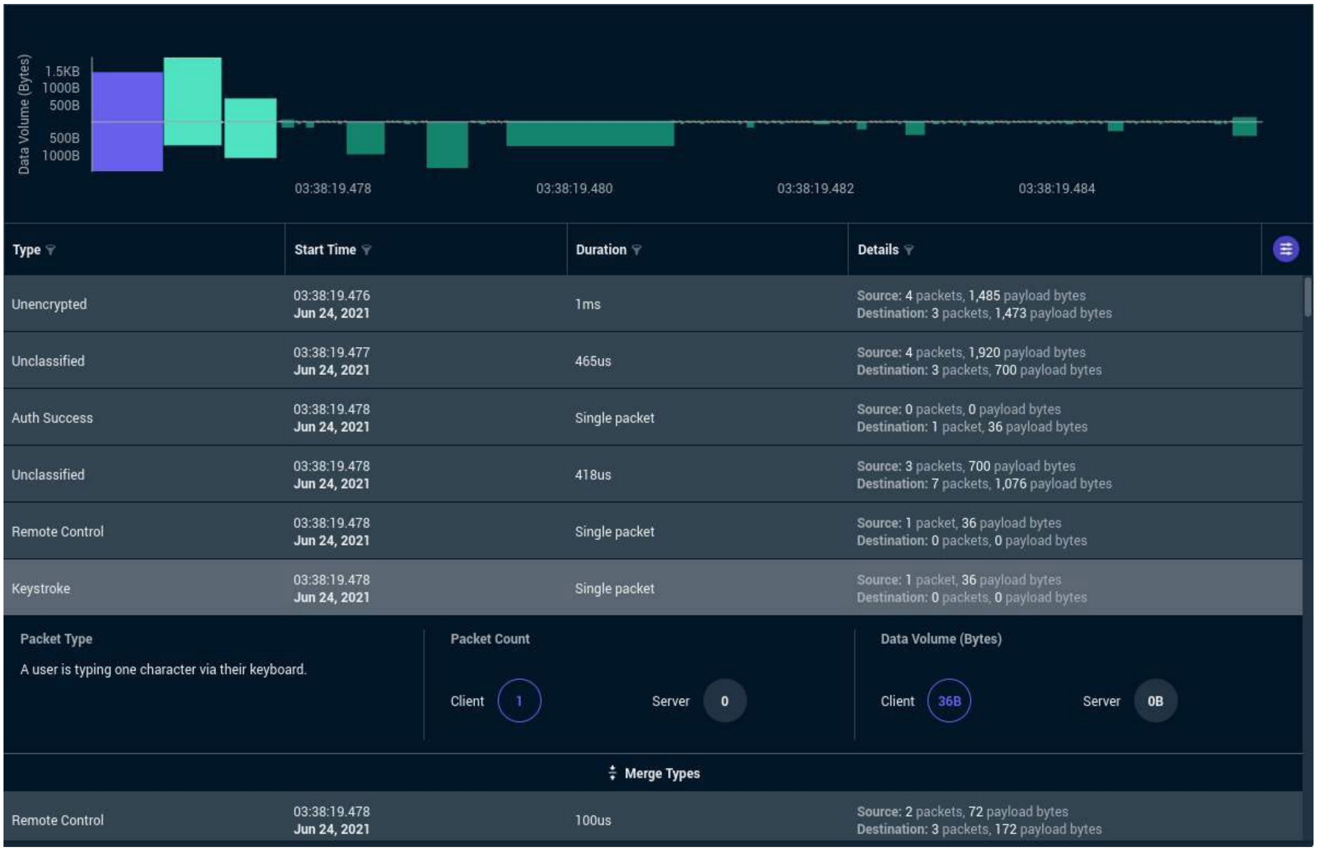
1c:61:b4:72:84:0a
Bob Marley's iPhone

Figure 5: AVA streamlines operations and reduces errors in network access control workflows using conversational AI

Case Study 3 – Autonomous Network Detection and Response

Like the network operations use cases above, AVA can also deliver day 0 value to the security operations team. For instance, AVA uses unsupervised machine learning to identify and track users, devices, applications, and other entities over time. AVA can also use this information to cluster similar entities. This presents a significant enhancement from legacy approaches that rely on unsupervised learning to spot anomalies from "normal" baselines for individual IP addresses rather than entities. Attributing behaviors to an IP address leads to high false positives and negatives, which translates to operational burdens on analysts.

Similarly, AVA uses supervised machine learning to identify patterns of activity that relate to attacker tactics, techniques, and procedures. For instance, AVA can classify remote access tools, reverse shells, unauthorized applications used for command and control, etc., without the need for decrypting the underlying data. This encrypted traffic analysis eliminates the privacy, policy, and technology challenges of decrypting data for analysis.



Type	Start Time	Duration	Details
Unencrypted	03:38:19.476 **Jun 24, 2021**	1ms	**Source:** **4** packets, **1,485** payload bytes **Destination:** **3** packets, **1,473** payload bytes
Unclassified	03:38:19.477 **Jun 24, 2021**	465us	**Source:** **4** packets, **1,920** payload bytes **Destination:** **3** packets, **700** payload bytes
Auth Success	03:38:19.478 **Jun 24, 2021**	Single packet	**Source:** **0** packets, **0** payload bytes **Destination:** **1** packet, **36** payload bytes
Unclassified	03:38:19.478 **Jun 24, 2021**	418us	**Source:** **3** packets, **700** payload bytes **Destination:** **7** packets, **1,076** payload bytes
Remote Control	03:38:19.478 **Jun 24, 2021**	Single packet	**Source:** **1** packet, **36** payload bytes **Destination:** **0** packets, **0** payload bytes
Keystroke	03:38:19.478 **Jun 24, 2021**	Single packet	**Source:** **1** packet, **36** payload bytes **Destination:** **0** packets, **0** payload bytes

Packet Type

A user is typing one character via their keyboard.

Packet Count

Client 1 Server 0

Data Volume (Bytes)

Client 36B Server 0B

Merge Types

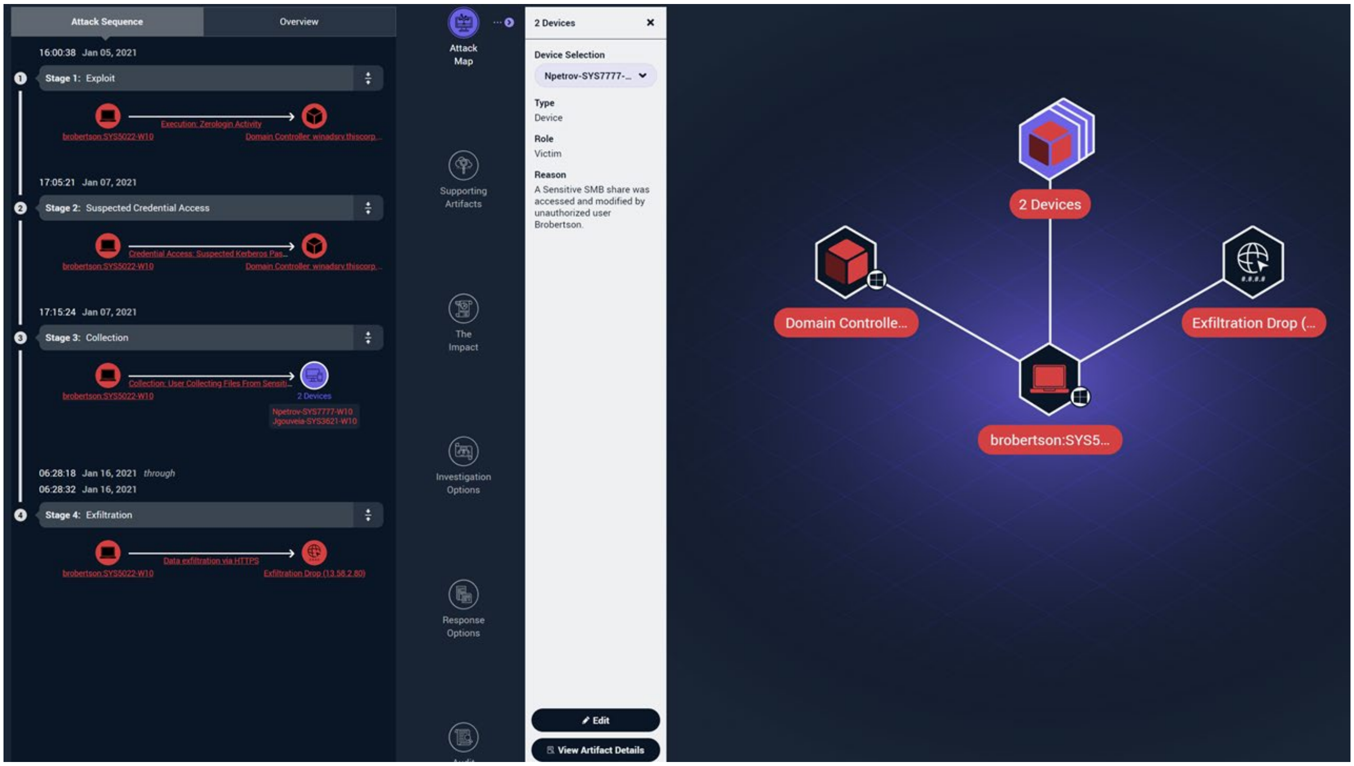
Type	Start Time	Duration	Details
Remote Control	03:38:19.478 **Jun 24, 2021**	100us	**Source:** **2** packets, **72** payload bytes **Destination:** **3** packets, **172** payload bytes

Figure 6: AVA flushes out the entire scope of an attack, enabling a decisive and rapid response

Another instance of AVA driving value for the security team is the ability to autonomously pull open-source and threat intelligence and thus, contextualize a potential threat uncovered in the environment. For example, when confronted with a suspect domain or IP address, much like an experienced security expert would, AVA pre-computes answers to questions such as:

- What other domains or destinations first showed up on the network at approximately the same time as the initial suspect domain?
- Did other devices attempt to connect to any of the same domains?
- Was there any trace of lateral movement activity beyond the initial victim?

Answering questions like these requires analyses of both internal data sources and external information via search engines, threat, and vulnerability databases, etc. AVA analyzes these results using natural language processing techniques such as entity extraction and topic modeling. For the operator, the benefit is that AVA helps uncover all potential victims within the organization and different parts of the attacker infrastructure, e.g., multiple command and control domains and Ips, all on a single screen.



16:00:38 Jan 05, 2021
1 Stage 1: Exploit
brobertson:SYS5022-W10 → Execution: Zerologin Activity → Domain Controller: winadsrv.thiscorp...

17:05:21 Jan 07, 2021
2 Stage 2: Suspected Credential Access
brobertson:SYS5022-W10 → Credential Access: Suspected Kerberos Pas... → Domain Controller: winadsrv.thiscorp...

17:15:24 Jan 07, 2021
3 Stage 3: Collection
brobertson:SYS5022-W10 → Collection: User Collecting Files From Sensiti... → 2 Devices
Npetrov-SYS7777-W10
Jgouveia-SYS3621-W10

06:28:18 Jan 16, 2021 *through*
06:28:32 Jan 16, 2021
4 Stage 4: Exfiltration
brobertson:SYS5022-W10 → Data exfiltration via HTTPS → Exfiltration Drop (13.58.2.80)

2 Devices

Device Selection
Npetrov-SYS7777-...

Type
Device

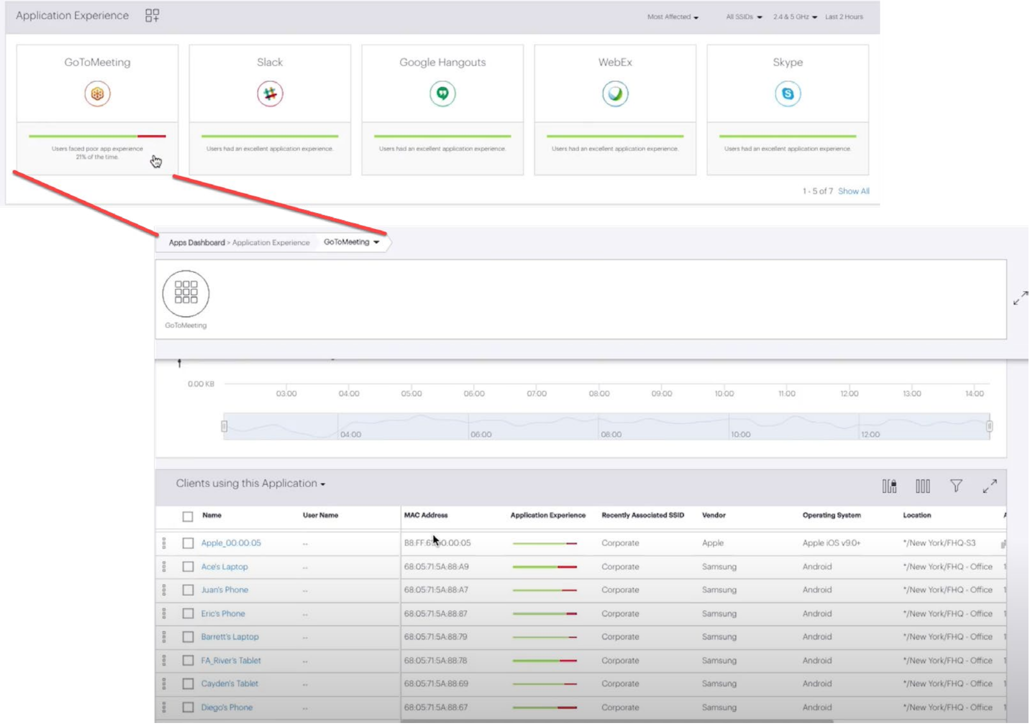
Role
Victim

Reason
A Sensitive SMB share was accessed and modified by unauthorized user Brobertson.

Figure 7: AVA flushes out the entire scope of an attack, enabling decisive and rapid response

Case Study 4 – Quality of Experience (QoE)

AVA helps provide the network operator with a clear view of the root causes of poor user experience and what remedial actions can be taken to improve that experience. This day 0 capability analyzes real-time data with the benefit of lab-trained models that understand the causes of network QoE issues, their interaction, and their effects. As a specific example, AVA employs a support vector machine (SVM) classifier to determine the performance of voice and video collaboration applications. The model is trained using a vast library of voice and video call flows labeled as a good or bad experience.



GoToMeeting	Slack	Google Hangouts	WebEx	Skype
Users faced poor app experience 21% of the time.	Users had an excellent application experience.	Users had an excellent application experience.	Users had an excellent application experience.	Users had an excellent application experience.

Clients using this Application

Name	User Name	MAC Address	Application Experience	Recently Associated SSID	Vendor	Operating System	Location
Apple_00:00:05	--	B8:FF:6[illegible]0:00:05		Corporate	Apple	Apple iOS v9.0+	*/New York/FHQ-S3
Ace's Laptop	--	68:05:71:5A:88:A9		Corporate	Samsung	Android	*/New York/FHQ - Office
Juan's Phone	--	68:05:71:5A:88:A7		Corporate	Samsung	Android	*/New York/FHQ - Office
Eric's Phone	--	68:05:71:5A:88:87		Corporate	Samsung	Android	*/New York/FHQ - Office
Barrett's Laptop	--	68:05:71:5A:88:79		Corporate	Samsung	Android	*/New York/FHQ - Office
FA_River's Tablet	--	68:05:71:5A:88:78		Corporate	Samsung	Android	*/New York/FHQ - Office
Cayden's Tablet	--	68:05:71:5A:88:69		Corporate	Samsung	Android	*/New York/FHQ - Office
Diego's Phone	--	68:05:71:5A:88:67		Corporate	Samsung	Android	*/New York/FHQ - Office

Figure 8: AVA proactively surfaces application quality of experience issues and recommends fixes

Case Study 5 – IoT Observability and Security

Detecting unknown IoT devices on the network offers a great example of how AVA automates human expertise.

A human expert intuitively describes an IoT device as one that most often doesn't have a browser, doesn't use enterprise protocols like SMB and Kerberos, and typically communicates with a small set of destinations. Using the information in NetDL, AVA can infer and index properties like these for the devices on the network, thereby easily highlighting the IoT devices. AVA then goes further by using recommendation systems algorithms to tag other IoT devices that are not captured by the originally encoded human intuition. This iterative approach is both quick and comprehensive at the task of eliminating an important blind spot for customers dealing with an explosion of devices on the network.

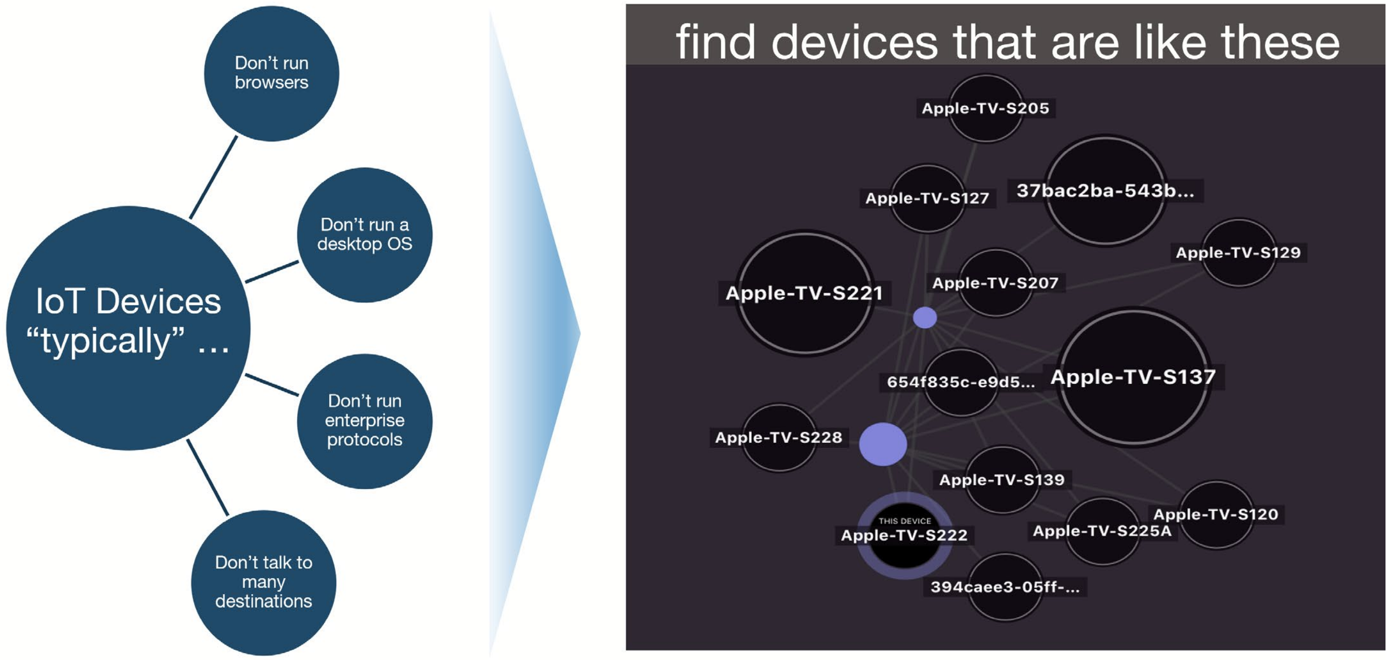
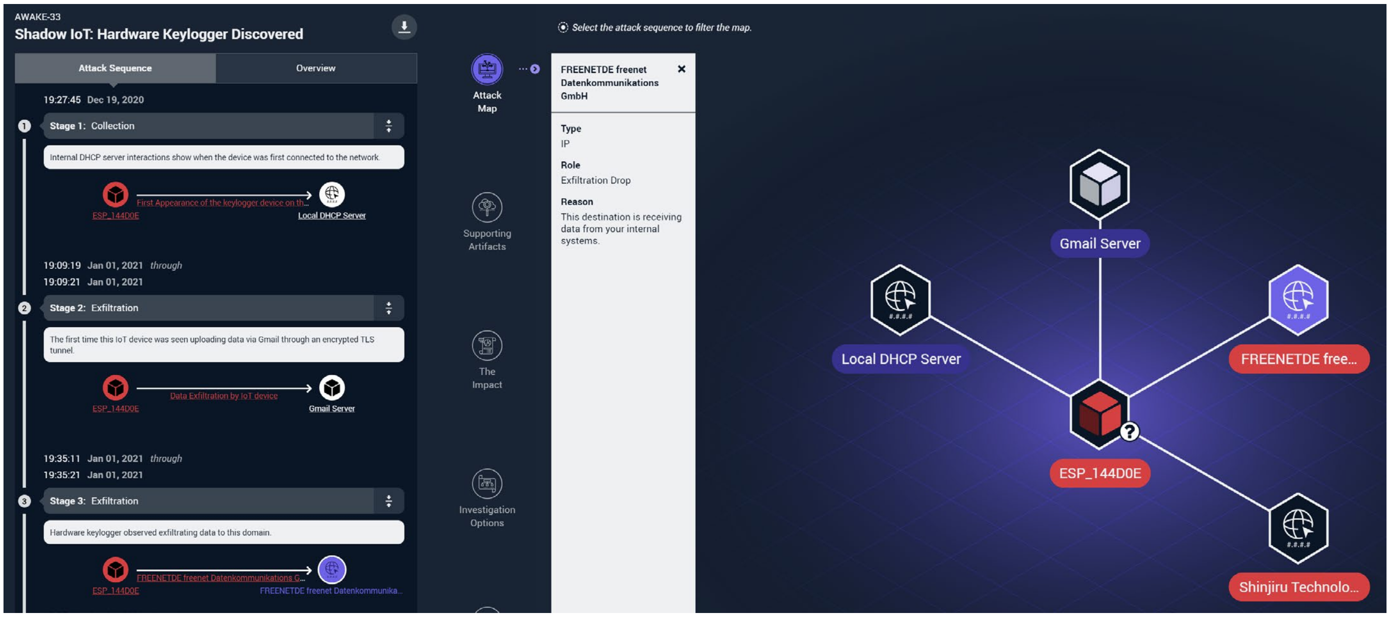


Figure 9: AVA encodes human intuition to discover IoT devices on the network

Consider the following real-world customer example which illustrates the benefits of this IoT observability. An IoT device plugged into a critical device's USB port was being used to intercept keystrokes between the keyboard and the computer. Detecting this threat first required identifying the shadow IoT device. As Figure 9 shows, the device had been sending encrypted emails and a custom UDP stream to locations in Germany and Malaysia. These "weak signals" added further conviction, allowing AVA to trigger a device quarantine via CloudVision AGNI, Arista's NAC solution.



AWAKE-33

Shadow IoT: Hardware Keylogger Discovered

Attack Sequence | Overview

19:27:45 Dec 19, 2020

1 Stage 1: Collection

Internal DHCP server interactions show when the device was first connected to the network.

ESP_144D0E → First Appearance of the keylogger device on th... → Local DHCP Server

19:09:19 Jan 01, 2021 *through*
19:09:21 Jan 01, 2021

2 Stage 2: Exfiltration

The first time this IoT device was seen uploading data via Gmail through an encrypted TLS tunnel.

ESP_144D0E → Data Exfiltration by IoT device → Gmail Server

19:35:11 Jan 01, 2021 *through*
19:35:21 Jan 01, 2021

3 Stage 3: Exfiltration

Hardware keylogger observed exfiltrating data to this domain.

ESP_144D0E → FREENETDE freenet Datenkommunikations G... → FREENETDE freenet Datenkommunika...

Select the attack sequence to filter the map.

FREENETDE freenet Datenkommunikations GmbH

Type
IP

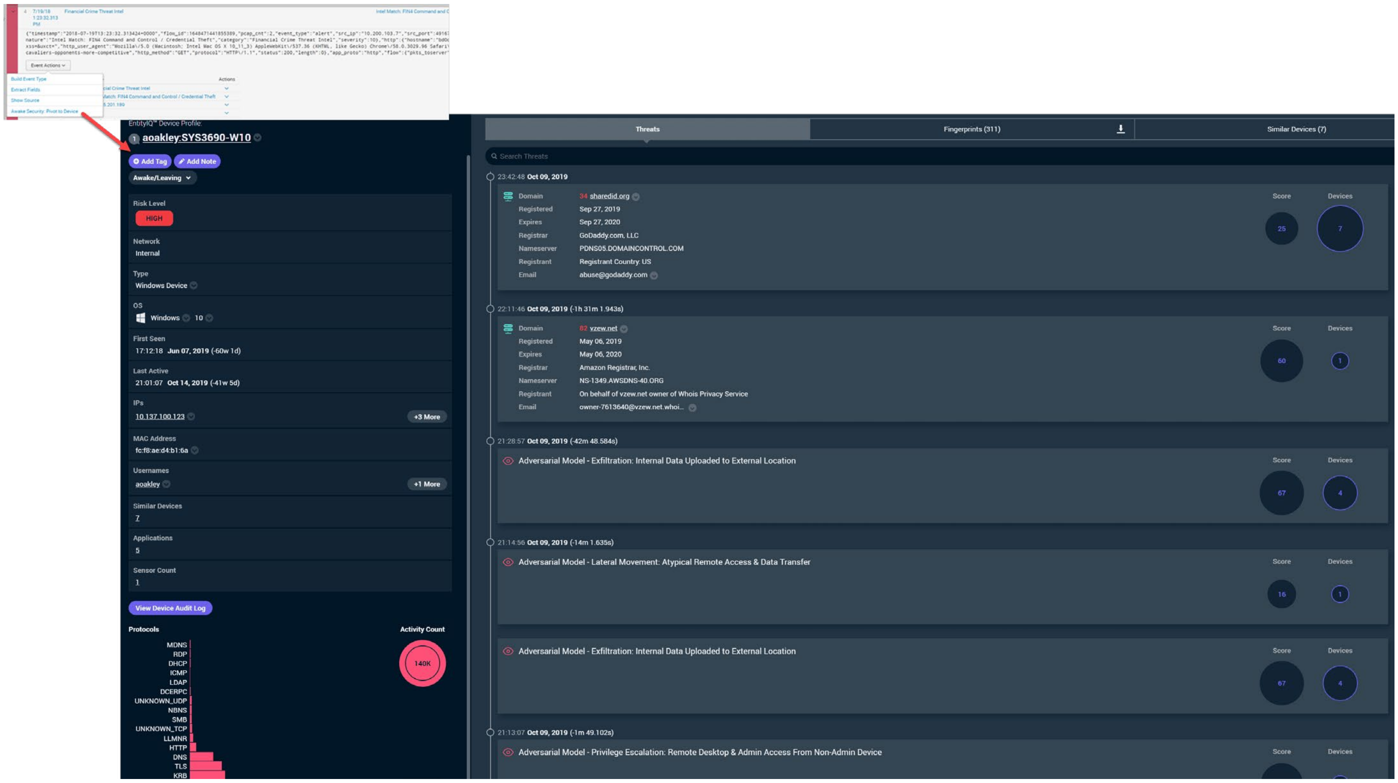
Role
Exfiltration Drop

Reason
This destination is receiving data from your internal systems.

Figure 9: Discovering and remediating rogue IoT devices on the network

Case Study 6 – Enriching the Ecosystem with AVA Insights

AVA insights can also be used to enrich other parts of the customers' IT and security ecosystem. For example, with one click, analysts using log aggregation and SIEM tools such as Splunk or Azure Sentinel can pivot from a meaningless IP address in those tools to an AVA-enriched profile that includes the name of the device, its primary user, applications running on it and other similar devices, as well as a forensic timeline of device activities. In addition, the analyst has a detailed timeline of that device's activities and can therefore make appropriate risk management decisions.



4 7/19/18 1:23:32.313 PM Financial Crime Threat Intel Intel Match: FIN4 Command and C

```
{"timestamp":"2018-07-19T13:23:32.313424+0000","flow_id":1648471441855389,"pcap_cnt":2,"event_type":"alert","src_ip":"10.200.103.7","src_port":49167
nature":"Intel Match: FIN4 Command and Control / Credential Theft","category":"Financial Crime Threat Intel","severity":10},"http":{"hostname":"bd0c
xss=&uxct=","http_user_agent":"Mozilla\/5.0 (Macintosh; Intel Mac OS X 10_11_3) AppleWebKit\/537.36 (KHTML, like Gecko) Chrome\/58.0.3029.96 Safari\
cavaliers-opponents-more-competitive","http_method":"GET","protocol":"HTTP\/1.1","status":200,"length":0},"app_proto":"http","flow":{"pkts_toserver"
```

EntityIQ™ Device Profile:

1 aoakley:SYS3690-W10

Risk Level
HIGH

Network
Internal

Type
Windows Device

OS
Windows 10

First Seen
17:12:18 Jun 07, 2019 (-60w 1d)

Last Active
21:01:07 Oct 14, 2019 (-41w 5d)

IPs
10.137.100.123 +3 More

MAC Address
fc:f8:ae:d4:b1:6a

Usernames
aoakley +1 More

Similar Devices
7

Applications
5

Sensor Count
1

23:42:48 Oct 09, 2019

Domain: 34 sharedid.org
Registered: Sep 27, 2019
Expires: Sep 27, 2020
Registrar: GoDaddy.com, LLC
Nameserver: PDNS05.DOMAINCONTROL.COM
Registrant: Registrant Country: US
Email: abuse@godaddy.com
Score: 25
Devices: 7

22:11:46 Oct 09, 2019 (-1h 31m 1.943s)

Domain: 82 vzew.net
Registered: May 06, 2019
Expires: May 06, 2020
Registrar: Amazon Registrar, Inc.
Nameserver: NS-1349.AWSDNS-40.ORG
Registrant: On behalf of vzew.net owner of Whois Privacy Service
Email: owner-7613640@vzew.net.whoi...
Score: 60
Devices: 1

21:28:57 Oct 09, 2019 (-42m 48.584s)

Adversarial Model - Exfiltration: Internal Data Uploaded to External Location
Score: 67
Devices: 4

21:14:56 Oct 09, 2019 (-14m 1.635s)

Adversarial Model - Lateral Movement: Atypical Remote Access & Data Transfer
Score: 16
Devices: 1

Adversarial Model - Exfiltration: Internal Data Uploaded to External Location
Score: 67
Devices: 4

21:13:07 Oct 09, 2019 (-1m 49.102s)

Adversarial Model - Privilege Escalation: Remote Desktop & Admin Access From Non-Admin Device
Score
Devices

Figure 11: Enriching Splunk with AVA context

Summary: Data-Driven Networking Made Possible

The combination of Arista AVA and EOS NetDL provides predictive and prescriptive intelligence for data-driven networks. AI/ML enrichment and analytics in combination with a broad ecosystem of vendors/partners, deliver market and customer-specific security, application, and network performance analysis, feeding continuous awareness and assurance. This provides a single source of truth and a decision-support architecture for Arista customers that ultimately delivers better business outcomes.

Santa Clara—Corporate Headquarters
5453 Great America Parkway,
Santa Clara, CA 95054

Phone: +1-408-547-5500
Fax: +1-408-538-8920
Email: info@arista.com

Ireland—International Headquarters
3130 Atlantic Avenue
Westpark Business Campus
Shannon, Co. Clare
Ireland

Vancouver—R&D Office
9200 Glenlyon Pkwy, Unit 300
Burnaby, British Columbia
Canada V5J 5J8

San Francisco—R&D and Sales Office 1390
Market Street, Suite 800
San Francisco, CA 94102

India—R&D Office
Global Tech Park, Tower A, 11th Floor
Marathahalli Outer Ring Road
Devarabeesanahalli Village, Varthur Hobli
Bangalore, India 560103

Singapore—APAC Administrative Office
9 Temasek Boulevard
#29-01, Suntec Tower Two
Singapore 038989

Nashua—R&D Office
10 Tara Boulevard
Nashua, NH 03062