Network Analytics

1. Introduction

This document describes the project undertaken by PalC Networks for developing Network Analytics application and infrastructure for one of our clients.

1.1. Network Analytics Introduction

Network Analytics is a practice of the collecting and analyzing different types of network data like network events, state information, packets etc. to identify threats, trends, pattern of the network. This information will be later used to predict any real-time failures, better planning of the network for efficient usage of different components, securing the network and end points etc. Usage of Artificial Intelligence and Machine Language in network analysis helps the network be more adaptive with self-configuring, self-optimizing environment.

Growth in Data Centers:

- Multi Cloud / Hybrid clouds:

In near future, there is going to massive growth in cloud data centers. The evolving Software Defined Data, telecom, IoT networks are becoming too complex to manually configure and maintain. Networks are evolving as they now support range of other functions than just packet transfer.

- Web/Hyper scale Data centers:

There has been an explosion of wired and wireless data ranging from Social, Mobile, Video streaming, gaming and emerging heterogenous IoT sensors.

- Application and service visibility:

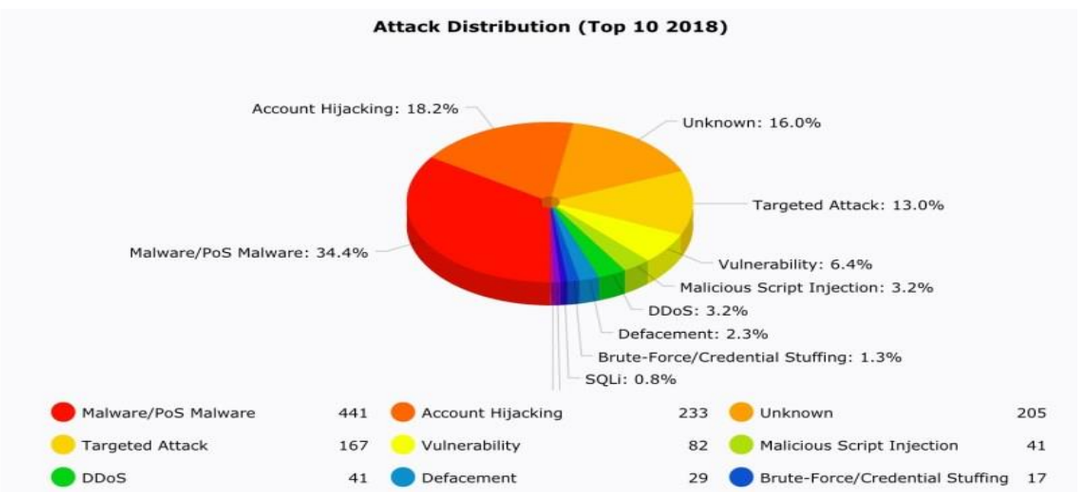
Effective usage of Network analytics will not only give service providers and customers an understanding of network components, but also help redefine service levels and contracts.

- security risks, Identifying & troubleshooting issues:

Numerous security threats (existing and new) introduced due to various solutions, protocols and applications. Network operational and maintenance challenges have been introduced due to the changing traffic patterns, rise of cloud services, rise of Big Data, consumerization of IT etc.

1.2. Current state of Network Attacks

Following is a sample of the top attacks seen in 2018:



1.3. Infrastructure and Cybersecurity Threat Profiles

This section outlines some of the emerging data trends and the associated cyber security threats

1.3.1. Recent Threat Profiles and Analysis

Cybersecurity breaches have been in the center stage for several years and has not exactly been a ride in the park for cybersecurity professionals. This is a living list that we actively track with more threats being added as they appear.

- Malware with worm capabilities
- Release of more Shadow Brokers tools
- Vulnerability of mobile carriers
- Growing Information Overload
- Adapting the firewall to face new threats
- Monitoring Cloud Configuration and Security
- High Impact Attacks
- Insider Threats

1.4. Emerging Trends are Driving Cyber-security Change

The key computing trends driving the need for a new cybersecurity paradigm are as follows:

- Changing attack patterns
- The “Cloud Insecurity”
- The rise of cloud services
- The influence of Big Data
- Rapid proliferation of the Internet of Things (IoT)
- Complexity that leads to stasis
- Inability to scale
- Vendor dependence

1.5. Key Business Challenges for Cyber-security operators

The key business challenges facing cyber-security operators are as follows:

- Ransomware Evolution
- AI Expansion
- IoT threats
- Big Data Store

1.6. Requirement

The requirement is to bring the infrastructure support and the security application for the Analytics solution.

2. Analytics Solution

Analytics solution will be a learning system capable of running in an adaptive mode by leveraging the advanced streaming analytics provided by a combination of platform and suite of predictive analytics applications using machine learning and artificial intelligence.

This allows the operator to move from the current trial-and-error model of security to an automated proactive approach using unsupervised machine learning models in conjunction with advanced real time protocol analysis.

The software defined data, IIOT, IOT security platform can perform tasks for which, it has not been explicitly pre-provisioned or programmed, by dynamically adjusting the security behavior based on a series of real-time empirical observations and behavior analysis.

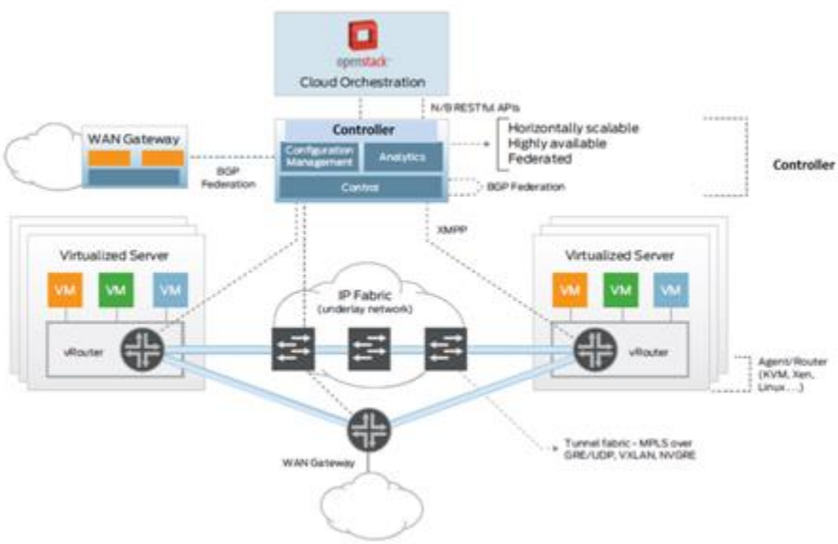
The analytics and security platform also allows the operators the ability to provide this capability in an “analytics as a service (AaaS)” to their internal and external customers.

This case study will provide a comprehensive set of capabilities that includes data acquisition (streaming & batch), data cleansing and data loading on a massive data lake. The platform also provides a rich set of machine learning (ML/AI) libraries that will enable the Cyber Security department to build applications for security.

It provides real time visibility across all communication stack in your physical and virtual network. The Network Analytics platform provides behavior-based application insight using large scale unsupervised machine learning to build dynamic policy models which is used to automate real time policy enforcement.

2.1. Analytics Network Controller

Network Controller automates network sensor resource provisioning and orchestration to dynamically create highly scalable virtual security analytics networks and to chain a rich set of virtualized network functions (VNFs) and physical network functions (PNFs) to form differentiated analytic service chains on demand. Integrated with a cloud management platform such as OpenStack, the Network Controller enables the agile creation and dynamic scaling of service instances with high availability and reliability. The Controller also makes it really simple to onboard analytic network functions onto the platform without requiring any API integration or modifications to third-party service software. The Controller’s advanced analytics capabilities provide deep insights into application and infrastructure performance for better visualization, easier diagnostics, rich reporting, custom application development and machine automation. The controller also interfaces with network policy engines and network elements and sensors to enforce network policy.



2.2. Presentation layer

The Portal enables consumption of this data through an easy-to-navigate and scalable web GUI and through representational state transfer (REST) APIs. The platform also provides Apache Kafka based push notification to which northbound systems can subscribe to receive notifications about policy compliance deviations, flow anomalies, etc. Advanced users have access to the built in data lake and can write custom applications using programming languages such as Python and Scala that are run on the platform using the powerful computing resources available.

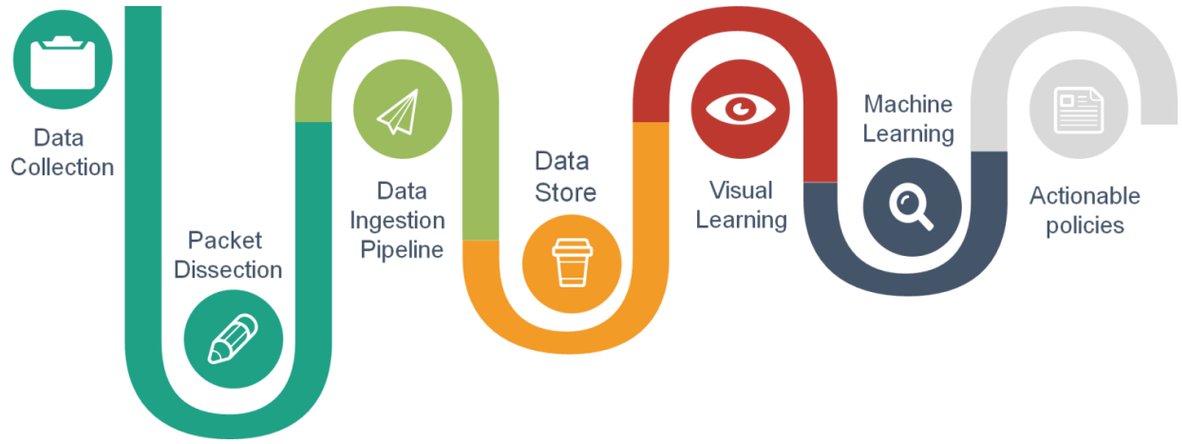
It also provides a state-of-the art, customizable, visual dashboard that presents information and insights in a clear and understandable format, providing high-level summaries while still allowing a user to go as deep into the data as required.

3. Features to be supported

- High-performance, real-time analytics with high scalability and low latency for various sources of data.
- Collect real-time data from application components and apply behavior-analysis algorithms to identify application groups and their communication patterns and service dependencies. This is in turn used to automate whitelist policy recommendations for zero-trust security.
- The telemetry data from every packet in the data center is collected and analysis is performed on millions of events to provide comprehensive actionable insight from billions of records within seconds. Long term data retention and playback without loss of detail is also available to simulate and analyze problems after the fact for future use case scenarios.
- Run analytics on the big data to acquire real-time actionable insights of what is happening in the data-center and display it on the GUI as shown above.

4. Approach

The workflow for Analytics platform is shown below:



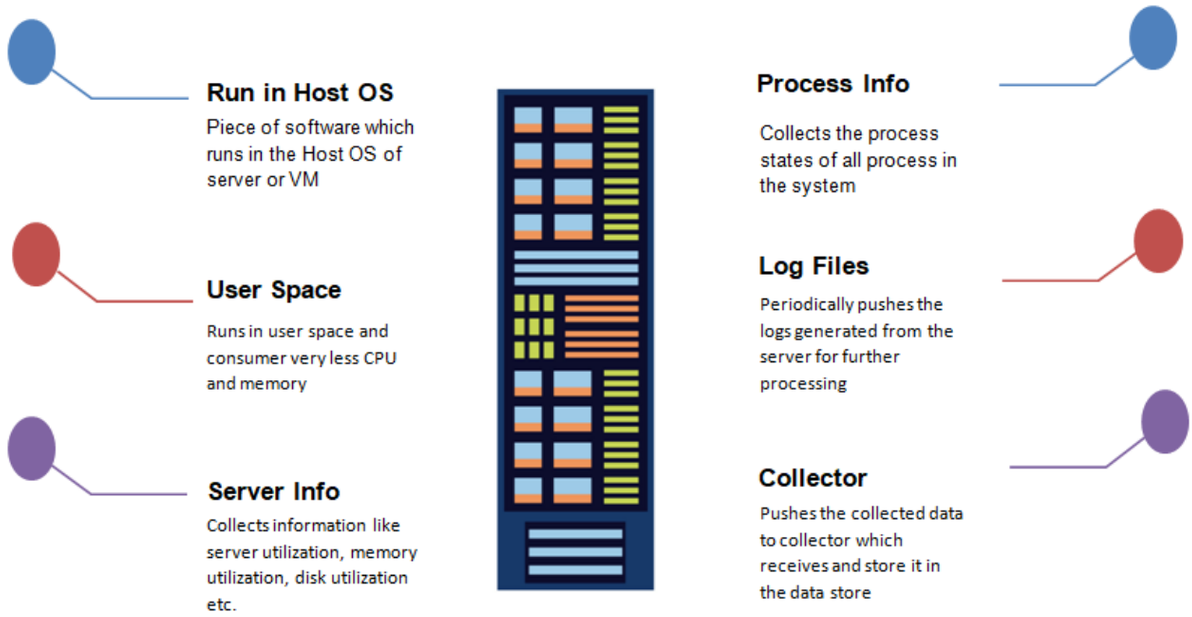
Network Telemetry data is collected using sensors and enforcement points. Two types of sensors are used: hardware sensors and software (endpoint) sensors. These sensors allow the network analytics and enforcement solution to support both existing (brownfield) and new (greenfield) network infrastructure.

4.1. Software and Hardware Sensors

The Network Analytics platform has the following main functional layers:

Telemetry DATA Ingestion Layer: This layer consists primarily of sensor functions. Sensors are the eyes and ears of the analytics platform. Two types of sensors are used:

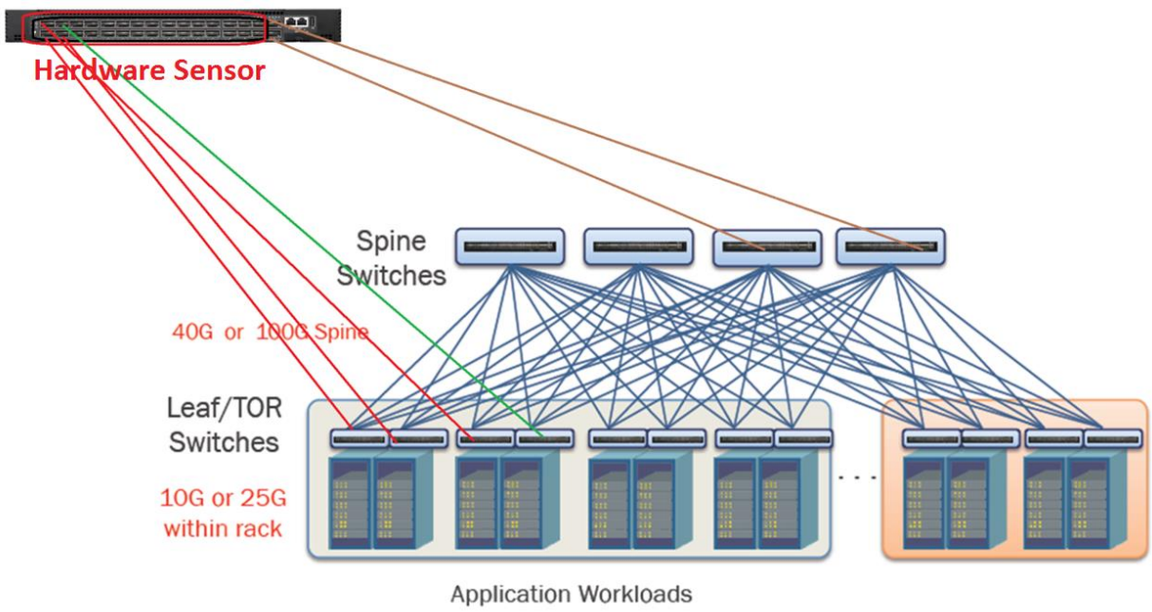
Software sensors: These lightweight sensors run as user processes and can be installed on any server (virtualized or bare metal). Two types of software sensors are used: full-visibility sensors and limited-visibility sensors. Limited-visibility software sensors are designed to collect connection information to support specific IoT Security Analytics use cases. These sensors are not designed to provide comprehensive telemetry data. The sensors also enforce network security on the network, IoT and IIOT edge gateways in real time working with network firewalls and security applications. It requires 1.5kB DRAM and about 5-6 kB for storage, so it is very resource friendly.



The software sensor provides actionable insight as follows.

- What is happening inside the server?
- Why my server is very slow?
- Which application is taking more CPU resources?
- Is there any application which has any memory leaks?
- Is there any application running inside the server is compromised and can cause potential harm to the network
- Can I provision more VMs / application containers inside a server?
- Whether the mission critical application has availability of right set of resources?
- How the OS treats this application in terms of scheduling and allocating the memories?
- How well the application program is written?
- Whether it has too many I/O operations which causes poor performance of application?
- What is the relation between this application and other applications running in the network?
- Whether the application is targeted for any attacks?

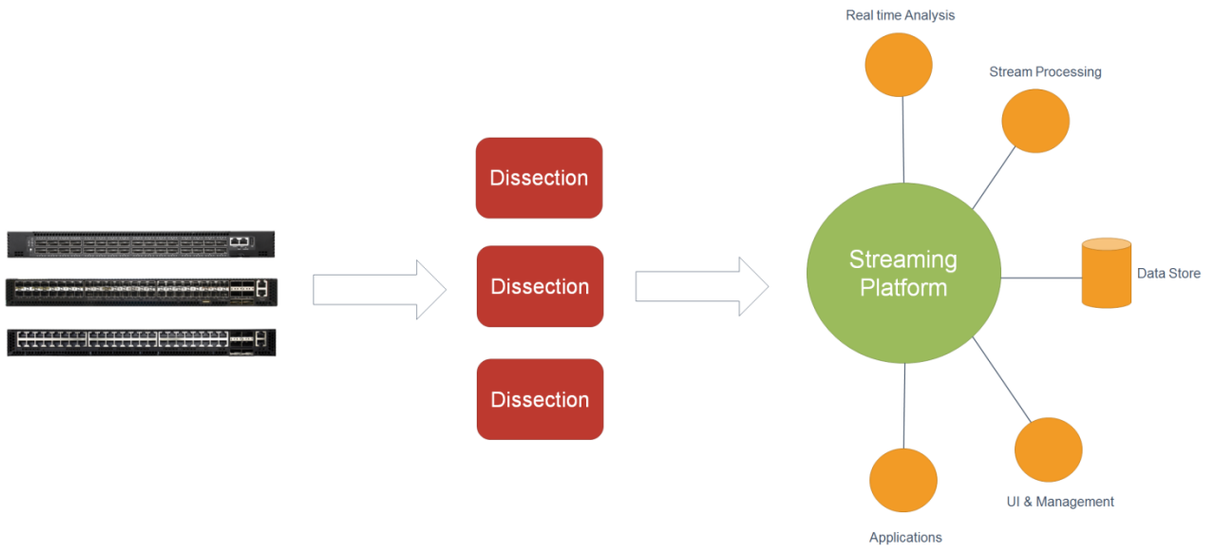
Hardware sensors: These sensors are embedded in 10/40/100 Gbps open compute switches from DELL, ACCTON, HPE, DELTA for large scale industrial settings or on an Appliance form factor in remote settings for IoT field deployments. It is usually placed adjacent to ToRs and Leaf node in the Data Center.



The Network Analytics platform can work with only software or only hardware sensors. Standard deployment consists of hardware and software sensors and they provide the following functions:

- Hardware sensors provide full visibility into application process-related context details.
- Hardware sensors act as enforcement points to enable application segmentation.

Hardware sensors provide packet level details, buffer details, tunnel endpoint mappings, detect traffic bursts. Hardware and Software sensors provide measurement of network latency and application latency. Software sensors and the hardware sensors collect three types of telemetry information:



Flow information: This information contains details about flow endpoints, protocols, and ports; when the flow started; how long the flow was active; etc.

Inter-packet variation: This information captures any inter-packet variations seen within the flow. Examples include variations in the packet's time to live (TTL), IP/TCP flags, and payload length.

Context details: Context information is derived outside the packet header. In the case of a software sensor, this information includes details about the process, including which process generated the flow, the process ID, and the user associated with the process.

4.2. Analytics Layer

Proprietary machine learning based algorithms enable signature-less flagging of anomalous traffic patterns, only using Network Flow data. Get an alarm when the network is under attack.

The machine learning based algorithms perform behavioral analysis on the incoming Network Flow data, and highlight traffic patterns that are unusual, extreme, or threatening.

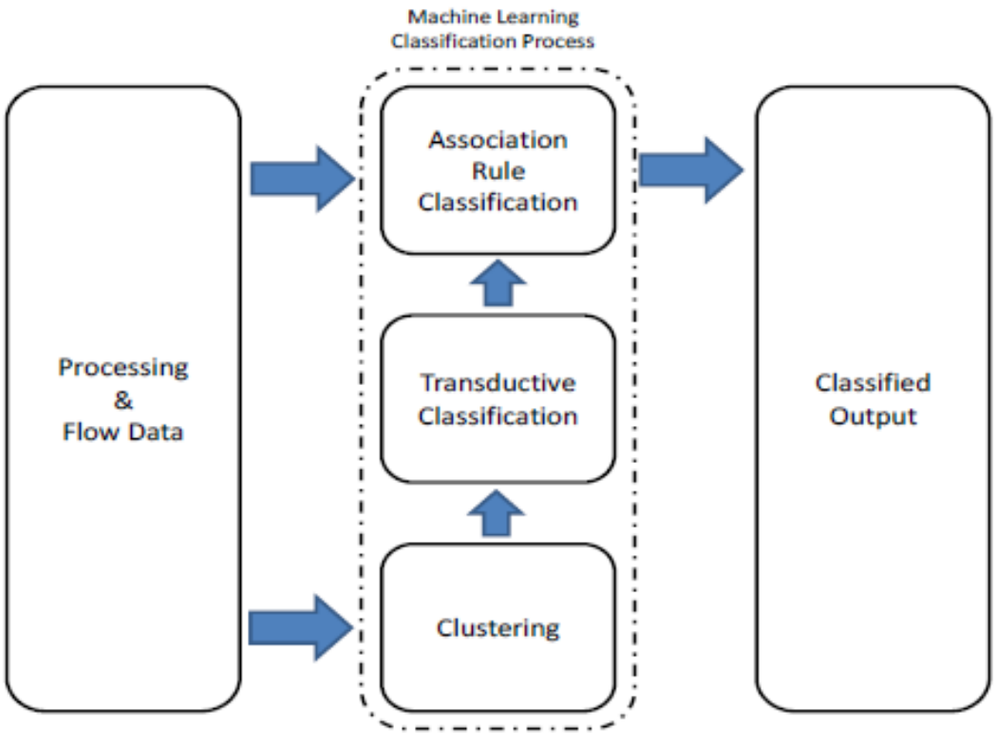
Our analysis learns what services exist in your network, and avoids raising false alarms for their usual activities. Since we do not rely on pre-programmed traffic signatures, you don't need to fine-tune them endlessly.

The Network Analytics ML (Machine Learning) layer brings in the following functionalities to the user

- Model Generation
- White-list and Black-list policy generation
- Zero-trust model
- Application dependency matrix
- Network and application performance insights
- Traffic Anomaly Detection
- Insider Threat
- Cyber Network Attacks
    - DOS & DDOS
    - Man in the middle attack
    - Database Ingestion Attack
    - Malware
- Protocol Based Attacks
    - DNS
    - Fork Bomb
    - HTTP Flood
    - IP Fragmentation Attacks
    - NTP Amplification
    - Flood Ping / Death Ping
    - SMURF Attack
    - TCP SYN Flood
    - UDP Flood
    - Etc.

There are five stages in the analysis of network packet data. The first two stages are taken care by Network analytics ingestion pipeline. The last three stages will be done on SAP HANA.

- Baselining
- Transductive Classification
- Clustering
- Association Rule Mining
- Anomaly Detection



Baselining the data involves converting all the collected data from different sensors and store it in a database. The platform has three different type of databases as

Baselining: The ingestion pipeline takes all the packets which are flowing in the network using Analytical switch and dissects the packet headers information. The dissected packet header information are then stored in the appropriate tables in database after performing pre-processing of packet, based on the type of packets.

- Hot Storage
- Warm Storage
- Cold Storage

Classification: Along with saving the packet header information, the ingestion pipeline also does the transductive flow classification by identifying the application and application group using Deep Packet Inspection methodologies. The traffic classification can be broadly divided into

- Port and packet payload based classification
- Statistical measurement based approaches
- Unsupervised ML
- Supervised ML
- Behavioral identification techniques

This pre-processed classified data is used for next step which is clustering

Clustering: One of the prominent unsupervised clustering techniques is the K-means clustering algorithm preferred over other methods such as hierarchical clustering, due to its enhanced computational efficiency. Using unsupervised K-means, flows belonging to individual applications are

separately cluster analyzed to extract unique subclasses per application, offering a finer granularity of the classification

4.3. Enforcement Layer

Software sensors act as the enforcement point for the detailed application policy generated by the platform, helping enable application segmentation. Using the data from the sensors, the Network Analytics platform provides consistent enforcement across public, private, and on-premises deployments. This layer also helps ensure that policy moves along with the workload, even when an application component is migrated from a bare-metal server to a virtualized environment. In addition, the enforcement layer helps ensure scalability, with consistent policy implemented for thousands of applications spanning tens of thousands of workloads.

Sample Use Case Pattern

Use Optimized algorithmic support for:

- Common stream data processing
- Complex event processing

Created Sensor Data fusions

- Generate Whitelist and policy mapping
- Detect Unusual Events occurring in stream(s) of data
- Post Detection, Isolate & Analyze Root Cause

Anomaly / Outlier Detection using unsupervised machine learning

Commonalities / Root Cause Forensics

- Event Chaining
- Prediction / Forecasting
- Actions / Alerts / Notifications

Create Automated Targeted Control Actions

- Distribute Whitelist policy

Create filter specification dynamically when zero-day attack occurs and send it via routing protocol updates to network elements for control action to prevent malicious traffic

IoT Security Solutions work together to provide protection throughout the attack continuum and also that can be integrated with complementary solutions for an overall security system:

Before an attack: Discover threats, and enforce and harden policies with existing Firewalls, Identity Services appliances, and Network Access Control (NAC) products.

During an attack: Detect, block, and defend against attacks that have already penetrated the network and are in progress with existing and next generation intrusion prevention web security systems.

Post attack: Scope, contain and remediate an attack to minimize damage in conjunction with deployed malware protection using real time network behavior analysis

5. GLOSSARY

NAC Network Access Control

DOS Denial of Service

TOR Top of the Rack

SDN Software Defined Networks