MENA
INFORMATION SECURITY
CONFERENCE 2018

Leveraging Machine Learning & AI in Cyber Defense

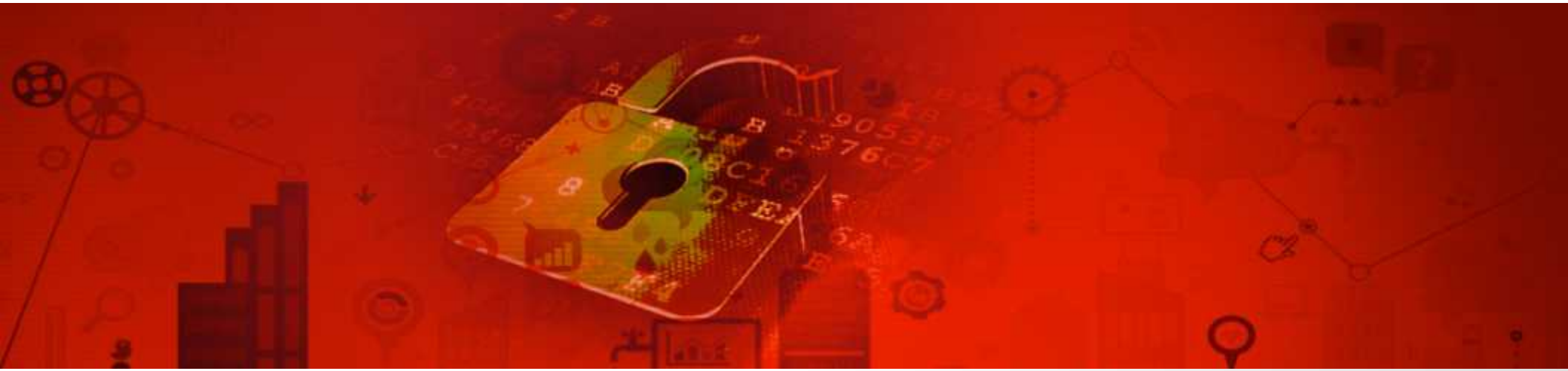
Rayan Mohtasib, Major Accounts Manager

Security Fabric and AI

VIRTUPORT

@VirtuPortMEA

www.menaisc.com



AI

Artificial Intelligence

is the simulation of human intelligence processes by machines, especially computer systems. These processes include learning, reasoning, and self-correction

MENA INFORMATION SECURITY CONFERENCE 2018

FORTINET

Early AI Defined

Alan Turing called an infant's mind an 'unorganized machine' in 1930s

Created early definitions of machine learning

- First type (A) consists of simple NAND (negative – AND gates
- Second type (B) is combination of A types with modifiers added – results in weighted input/variable output method
- Saw the need for:
  - **Seeded solution set of accurate or known potential output**
  - **Population of variably weighted pieces or functions**
  - **A method for removing the worst solutions while retaining the best**

Major inhibitor of his research – was far ahead of available capabilities in terms of computing power.

Artificial Intelligence – Nearing a Century

History of AI Outside Cyber Security Industry

1930s — **Turing, Kleene and Church** propose machine learning solution

1943 — **McCullouch and Pitts** create formal design of Turing's 'artificial neurons'

1956 — AI research formally founded as a discipline at **Dartmouth College**

1960s — AI research heavily funded by the **U. S. military**

1974 — **The First AI Winter** Difficulty resulted in funding cuts in **US and Britain**

1980s — **Proliferation of Expert Systems** Lisp vs. PC

1990s — **AI applied to data mining, medical diagnosis** with increased CPU power

1997 — **IBM Deep Blue** beats Grand Master Kasparov in chess

2003 — **Deep learning** is achieved using faster computing, large data structures

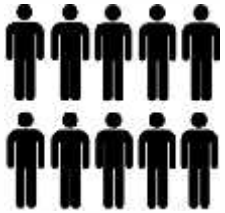
2011 — **IBM Watson** defeats Jeopardy champions Brad Rutter and Ken Jennings

2013 — **IBM Watson** application for management decisions of lung cancer treatment

2015 — **2,700+** AI projects in place at **Google**

2017 — **Elon Musk** calls for the regulation of AI before we hit 100 years

Why AI in Security ?



Scales Security Ops
With an increasing volume and sophistication of attacks organizations need a force multiplier for their security teams

Assist Decisions
Given a broader threat surface area security teams need assistance in effective and accurate data driven decision making

Improve Responsiveness
Due to increasing risk of compromise, institutions and individuals will demand faster response to breaches

Be Proactive
The goal is to instrument proactive security controls to minimize exposure to emerging threats

MENA INFORMATION SECURITY CONFERENCE 2018

FORTINET

ML

[Machine Learning]

Machine learning is a branch of artificial intelligence (AI) that refers to technologies that enable computers to learn and adapt through experience.

MENA INFORMATION SECURITY CONFERENCE 2018

FORTINET

Classification of Machine Learning for Cyber Security

Shallow Learning (SL)

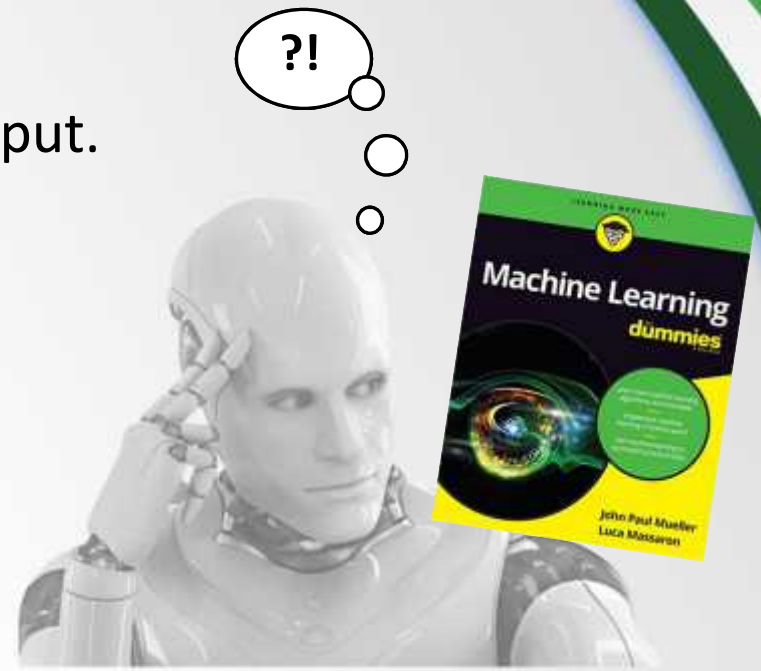
requires a domain expert (that is, a *feature engineer*) who can perform the critical task of identifying the relevant data characteristics before executing the SL algorithm

Deep Learning (DL)

relies on a multi-layered representation of the input data and can perform feature selection autonomously through a process defined *representation learning*

Types of Problem Solving

- **Supervised Learning** – Using known solution sets to embed proper functions and create proper output.
  - **Reinforcement** – action on an environment triggers an observation resulting in a defined state.
- **Unsupervised** Learning – unknown solution sets
  - **Clustering** – group according to similarities.
  - **Dimensionality Reduction** – deductive reasoning.
  - **Structured Prediction** – random fields are analyzed to predict according to defined output probabilities.
  - **Anomaly Detection** – input does not match expectations.



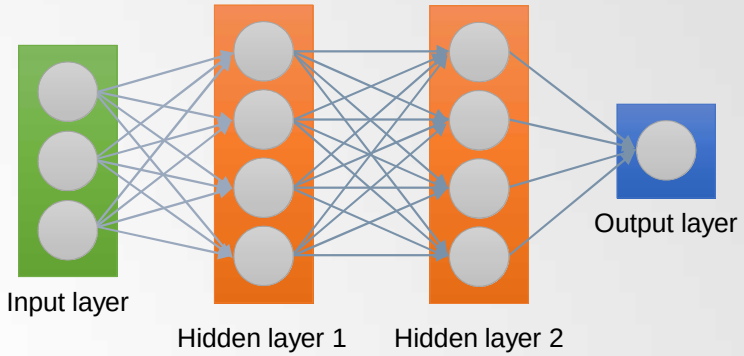
Artificial Neural Networks (ANNs)

Large collections of simple interconnected nodes (neurons), each with a weighted input and output value.

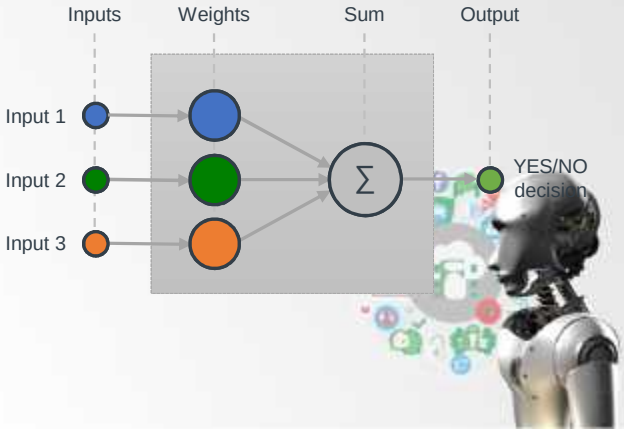
Type of AI – Artificial Neural Network (Multilayer Perceptron)

- Consists of three or more layers
  - Input layer
  - One or more hidden layers
  - Output layer
- Layers are made of up nodes
  - Connected to every node in the previous and subsequent layer
  - Provide discrete processing of input information (files and features)
  - Produces an output value based on inputs, function, and weighted valuation

The Multilayer Perceptron approach provides deep machine learning capabilities.



MP behavior is similar to human neurons - if input is strong enough, signal is passed according to weighted value



FORTINET

Layers + Nodes + Features = Learning

- System is fed initial data sets for analysis
  - Supervised machine learning approach
- Information (features) extracted during the learning phase.
  - Binary present within the files
  - Represents behavioral activation
- The system learns to weight features based on
  - Surety of indicators (‘tells’)
  - Frequency of observation

Source accuracy of the population input A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.

FORTINET

Antivirus Evolution

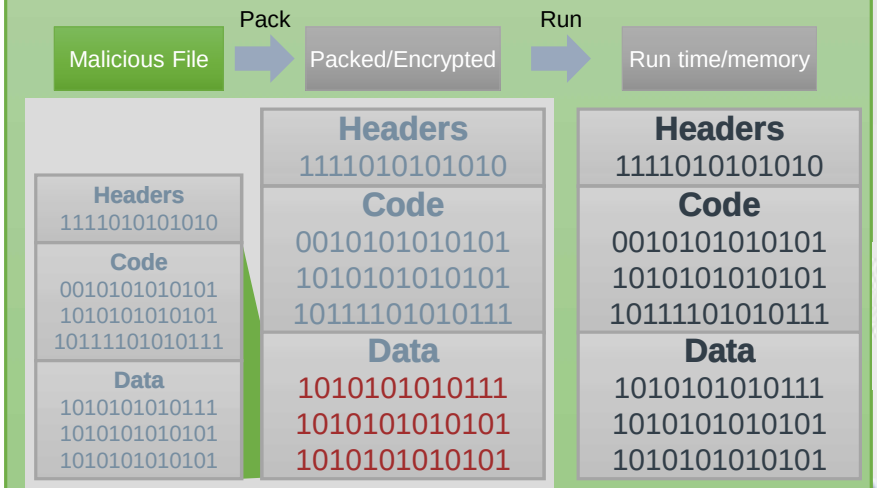
LEVEL I

- Simple MD5 / SHA 56 computations
- Resulted in large DBs for file comparisons
- One signature – one piece of malware
- Reactive and non-responsive to mutations

C:\Md5summalware.exe
5e3830ee3282a53920e00784fec44cf
d (malware.exe)

LEVEL II

- Content Pattern Recognition Language
- Looks at wrappers and payload for repeats
- Handles large volumes of permutations
- Proactive in nature



MENA INFORMATION SECURITY CONFERENCE 2018

FORTINET

Cloud Assisted AI - a Self Evolving Detection System

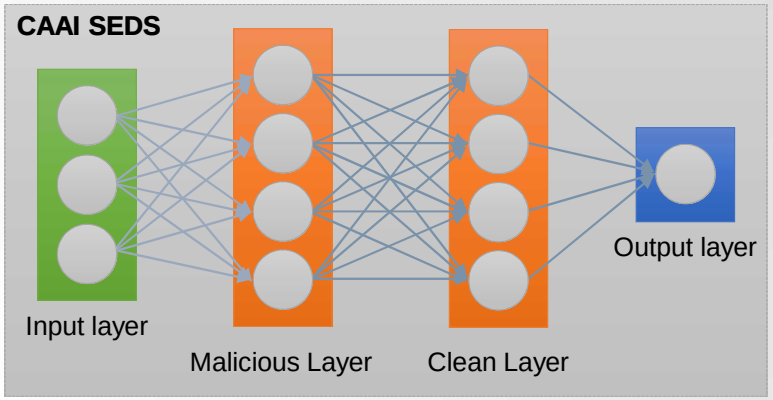
4 Layer Architecture

1 = process the input file

2 = 2.4 billion nodes analyzing potential malicious features

3 = 3.2 Billion nodes analyzing files for clean features

4 = output or decision layer (1 = malicious , 0 = clean)



Consists of separate layers for either malicious or clean feature processing
Mathematical models compare samples and features to decide output

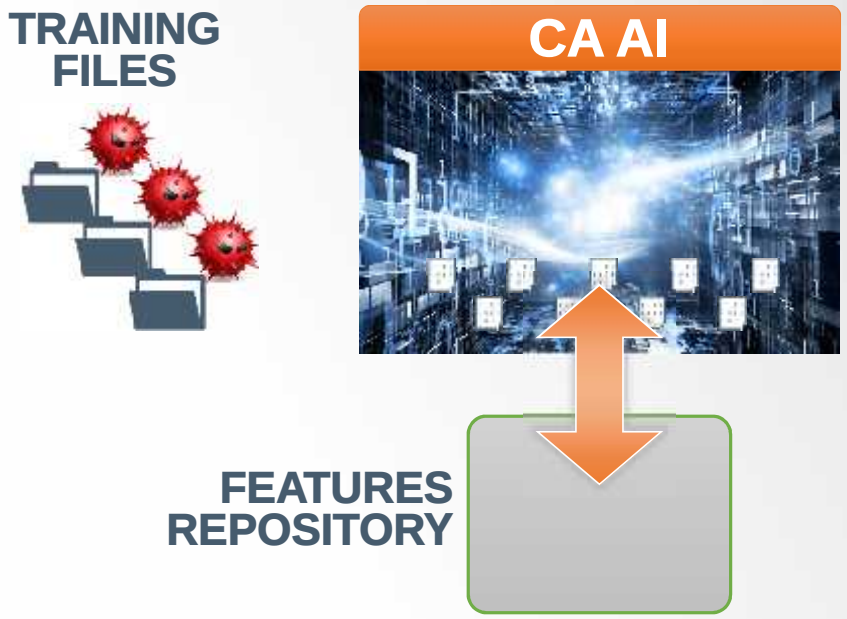
System Training

1. We start with CA AI and an empty feature repository

2. A training set of files are input, consisting of clean and malicious files. Files are labeled for initial training

3. CA AI logic determines commonality of files and builds the feature set.

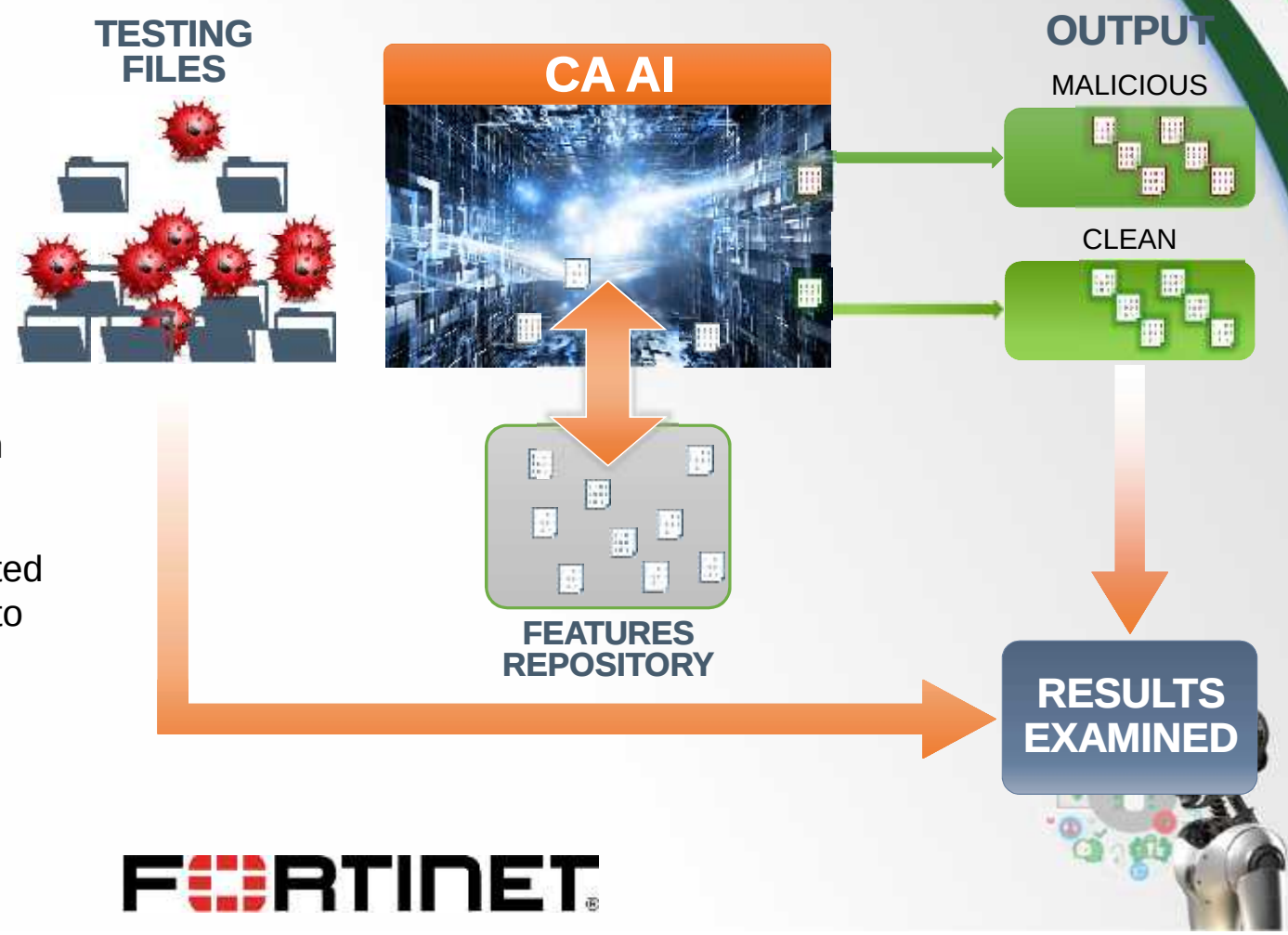
4. Features are modified as the system learns (weighting values, next phase)



FORTINET

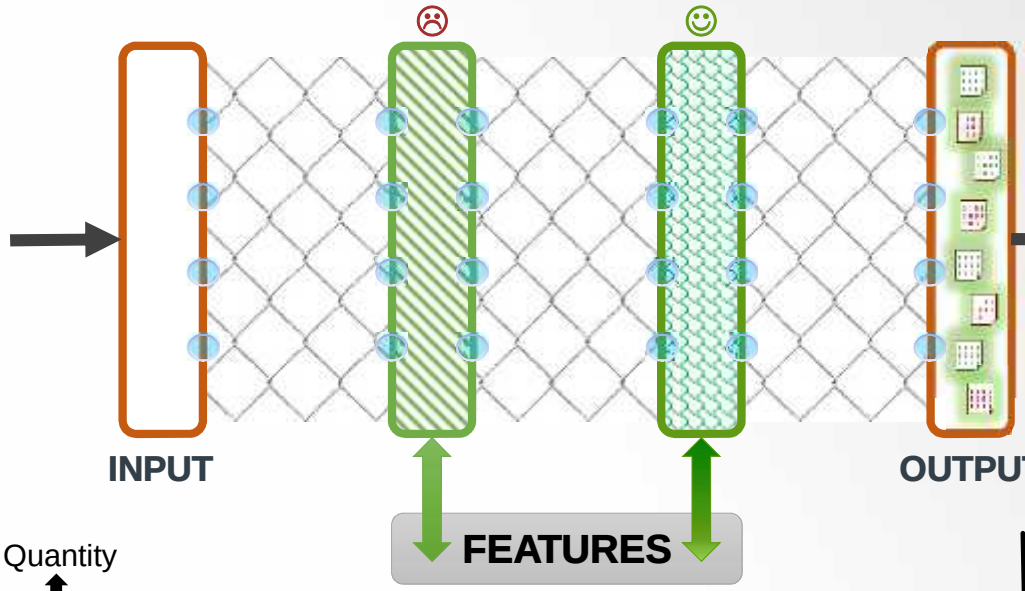
System Testing

1. Test samples are selected and input to the system
2. Using the feature repository, samples are analyzed
3. As this occurs, existing features may get modified or others added
4. The system determines clean or malicious output
5. Output is compared to expected results. If not accurate, reset to known point and retrain.



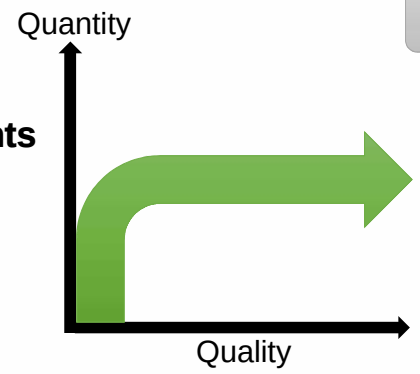
CA AI in Operation

RAW SAMPLES



Feature Set Improvements

- Quality
- Stabilized Number
- Weighting Confidence



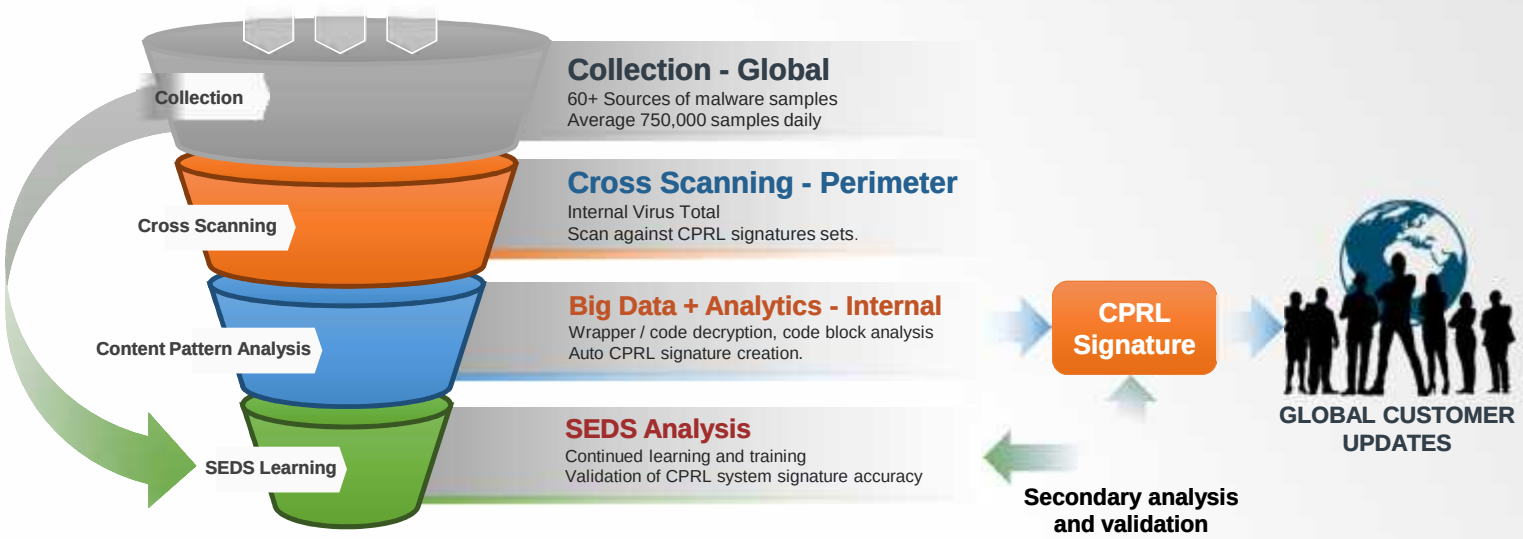
Continued Accuracy to a High Degree of Confidence

MENA INFORMATION SECURITY CONFERENCE 2018

FORTINET

Cloud Assisted AI – CURRENT OPERATIONS

- **Augmenting pattern recognition and automatic signature creation technology**
- **Continued learning and feature improvement – higher accuracy of the system**



FORTINET

Thank You

MENA
INFORMATION SECURITY
CONFERENCE 2018

FORTINET