MIS|TI™ PRESENTS

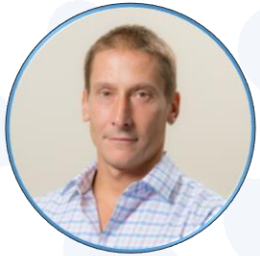
InfoSecWorld

Conference & Expo 2018

FROM TRAPPING TO HUNTING: INTELLIGENTLY ANALYZING ANOMALIES TO DETECT NETWORK COMPROMISES

Giovanni Vigna, CTO

Lastline, Inc.



WHO AM I?

Professor at the University of California in Santa Barbara (UCSB)

- Research focused on system security
- Made tools available through public portals (Anubis, Wepawet, Revolver, angr…)

CTO at Lastline, Inc.

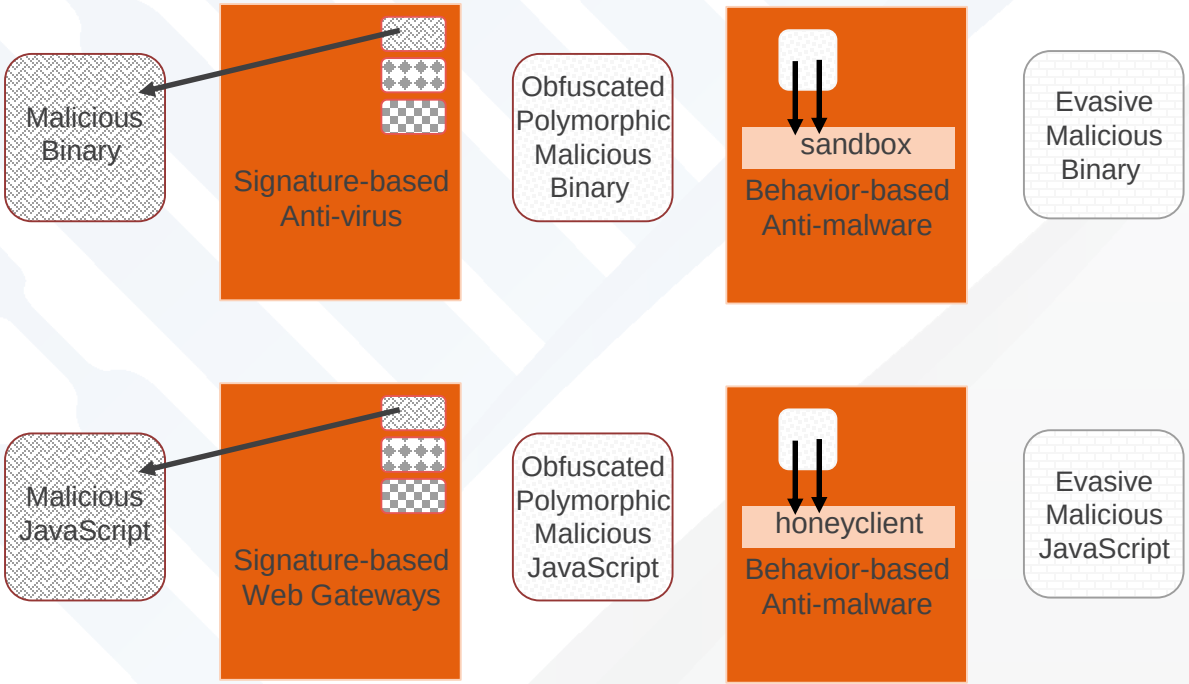
- Company focused on advanced malware detection

Founder of the Shellphish hacking team

- Longest running team at DEF CON CTF
- Developed Mechanical Phish, an autonomous hacking system for the DARPA Cyber Grand Challenge

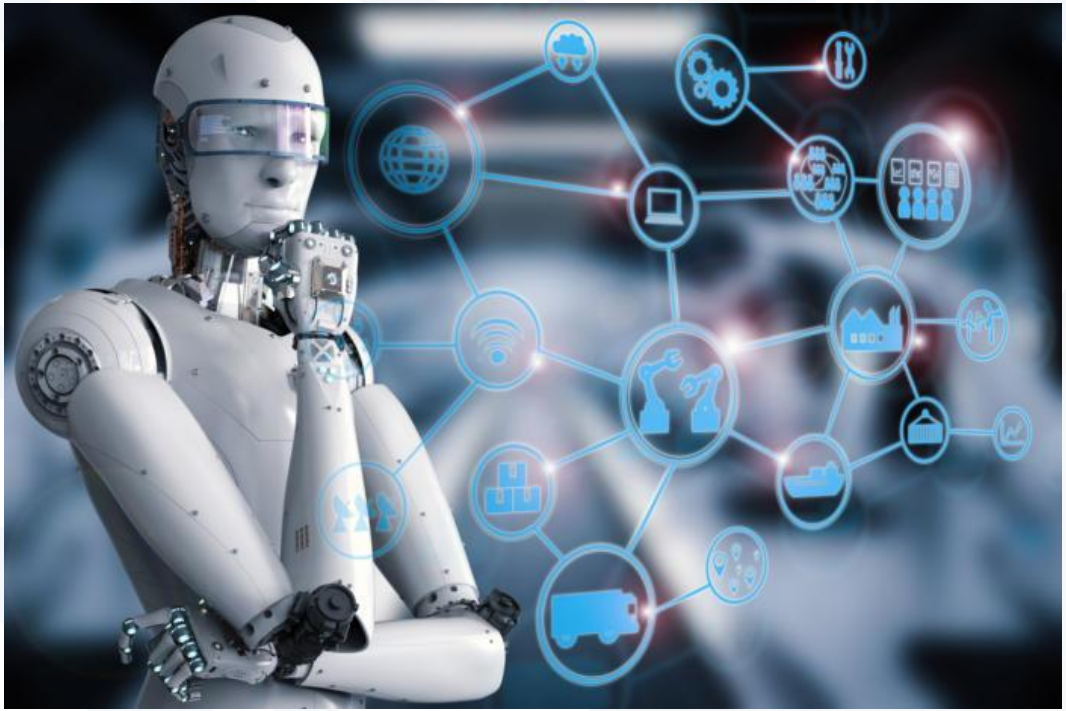
TRAPS

ARMS RACE(S)

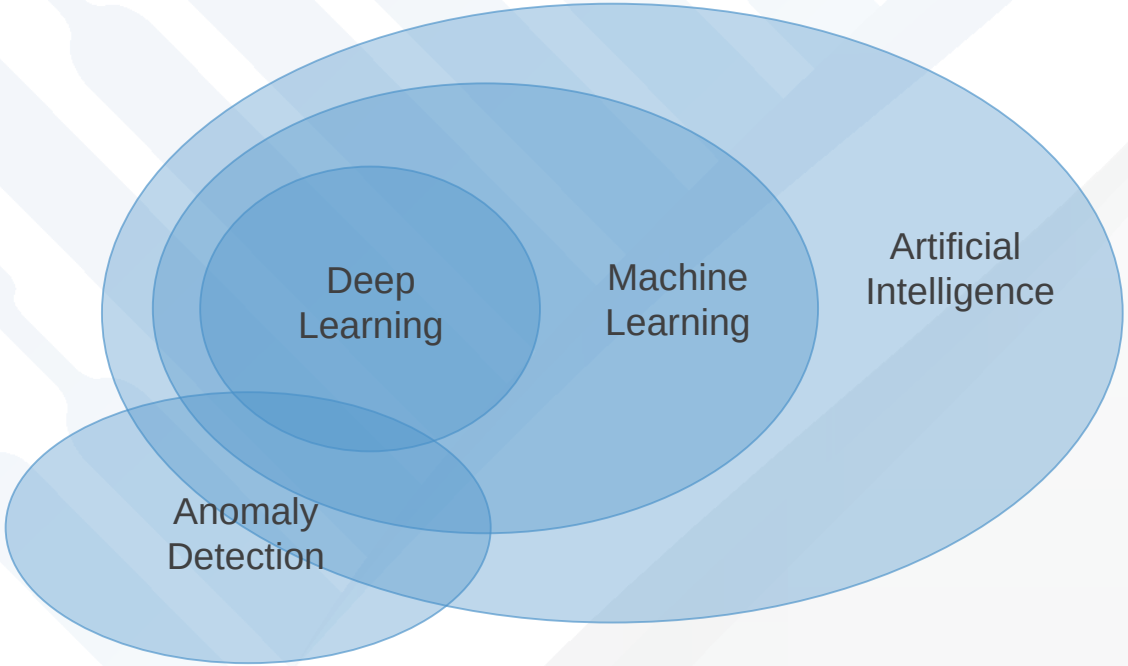


HUMANS

BEYOND HUMANS: ARTIFICIAL INTELLIGENCE



AI, ML, DEEP LEARNING, AND ANOMALY DETECTION



WHAT IS MACHINE LEARNING?

“Machine learning explores the study and construction of algorithms that can learn from and perform predictive analysis on data”

https://en.wikipedia.org/wiki/Machine_learning

WHY MACHINE LEARNING?

Because it supports:

- Data analysis
- Characterization
- Classification

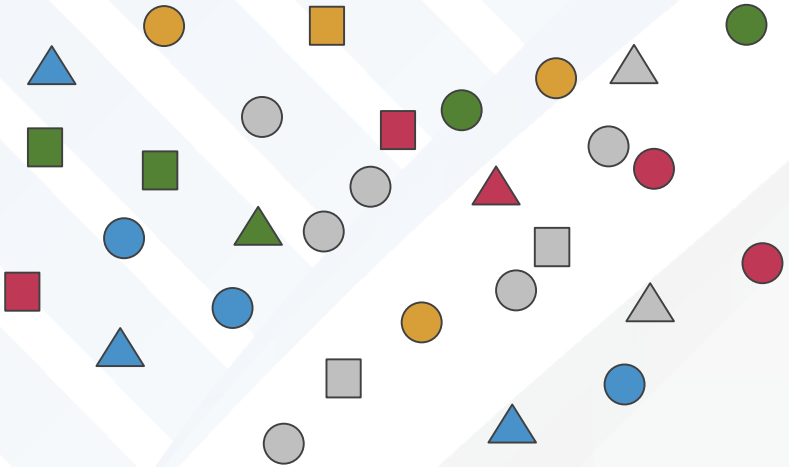
MACHINE LEARNING

?

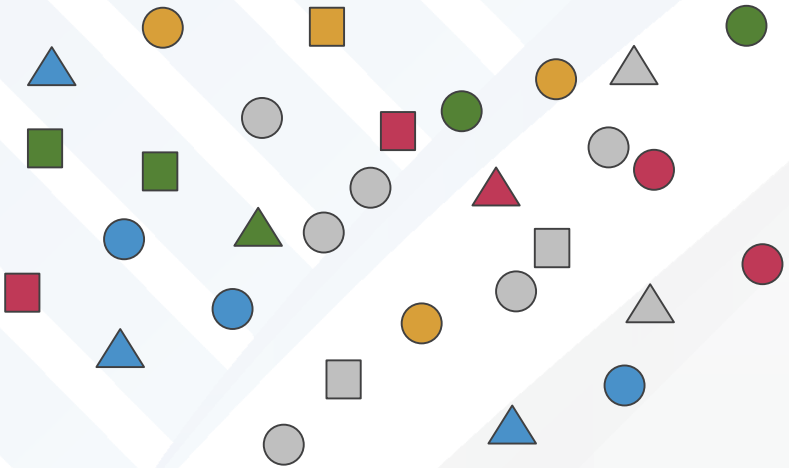
Round?

Has >3 sides?

…



MACHINE LEARNING



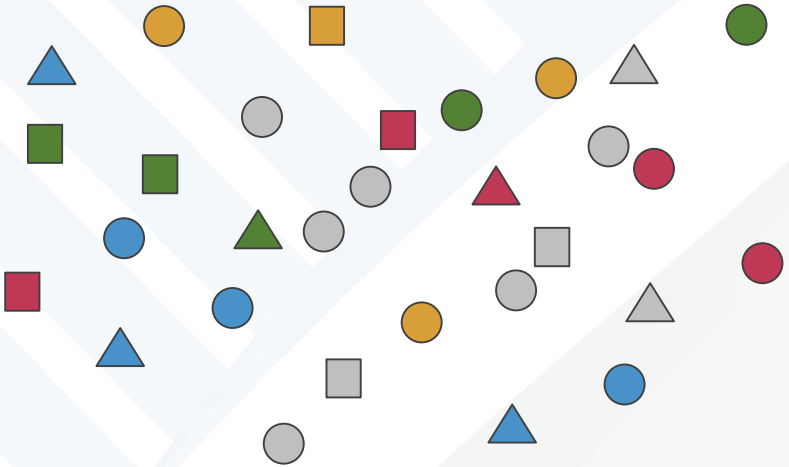
MACHINE LEARNING

?

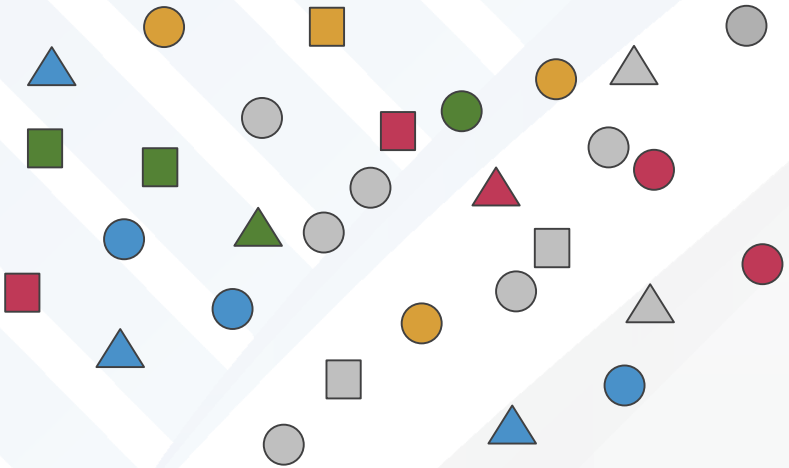
Reds are bad

Blues, greens, oranges are good

What about greys?



MACHINE LEARNING



MACHINE LEARNING AND SECURITY

Supports the clustering of events and objects

- Can be labeled as a group
- Can be handled together

Supports the classification of events and objects

- Can create classifiers based on historical ground truth

Takes the human (mostly) out of the loop

CHALLENGES IN MACHINE LEARNING

How good are your datasets?

How good are your features?

How good are your algorithms?

How change-resistant is your approach?

CHALLENGES IN MACHINE LEARNING

https://imgs.xkcd.com/comics/machine_learning.png

THE GHOST IN THE MACHINE

= cat!

THE GHOST IN THE MACHINE




= cat!

THE GHOST IN THE MACHINE

= cat!

ADVERSARIAL MACHINE LEARNING

- Learning (labeling, classifying) in the presence of data samples that do not want to be classified correctly
- An adversary can pollute your dataset
- An adversary can steal your learned models
- An adversary can modify a sample to be misclassified

Not your average machine learning gig

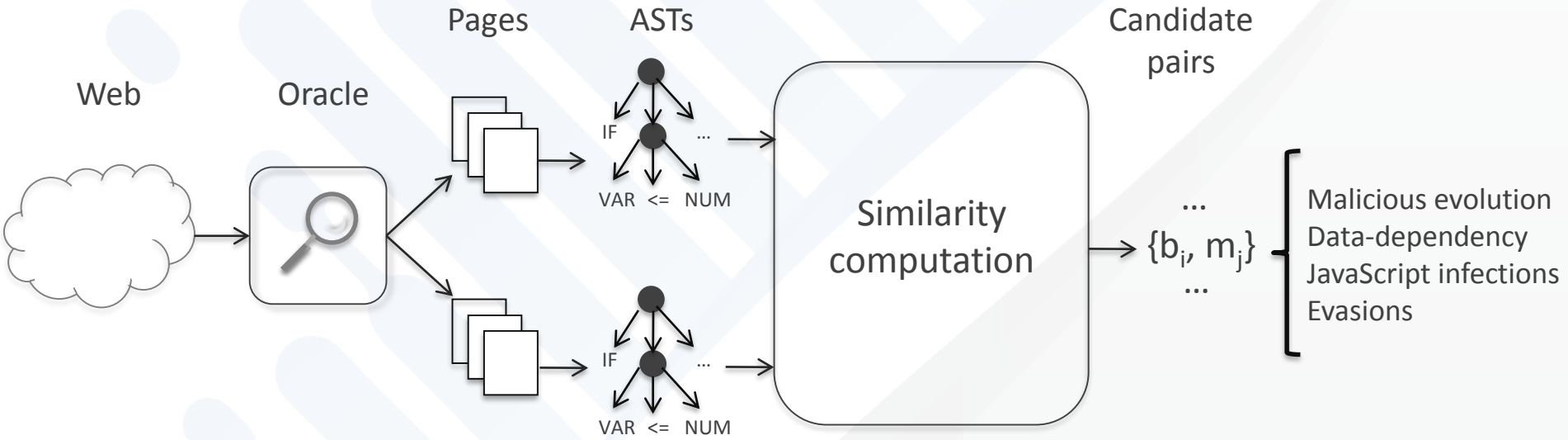
ADVERSARIAL MACHINE LEARNING IS HERE!

- Adversarial Perturbations Against Deep Neural Networks for Malware Classification
- Practical Black-Box Attacks against Deep Learning Systems using Adversarial Examples
- Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition
- Towards Evaluating the Robustness of Neural Networks
- Stealing Machine Learning Models via Prediction APIs

FIGHTING BACK

- Filter data to avoid pollution
- Elicit behavior that reveals the actual characteristics of the sample
- Use multi-classification and clustering to identify evasive samples

IDENTIFYING ADVERSARIES: MALICIOUS EVASIVE JAVASCRIPT



DETECTING EVASIONS

FROM TRAPPING TO HUNTING

NEEDLE AND HAYSTACK

ANOMALY DETECTION

“In data mining, anomaly detection (also outlier detection) is the identification of items, events, or observations which do not conform to an expected pattern or other items in a dataset.”

https://en.wikipedia.org/wiki/Anomaly_detection

CORRELATION

“In statistics, dependence or association is any statistical relationship, whether causal or not, between two random variables or bivariate data. Correlation is any of a broad class of statistical relationships involving dependence, though in common usage it most often refers to how close two variables are to having a linear relationship with each other.”

https://en.wikipedia.org/wiki/Correlation_and_dependence

MISUSE VS. ANOMALY DETECTION

Model What's Bad

Model What's Good

MODELING GOOD BEHAVIOR

- Time consuming
- Requires expert knowledge
- Incomplete
- Constantly outdated

LEARNING GOOD BEHAVIOR

- Automated
- Continued
- Comprehensive

LEARN WHAT YOUR NETWORK DOES

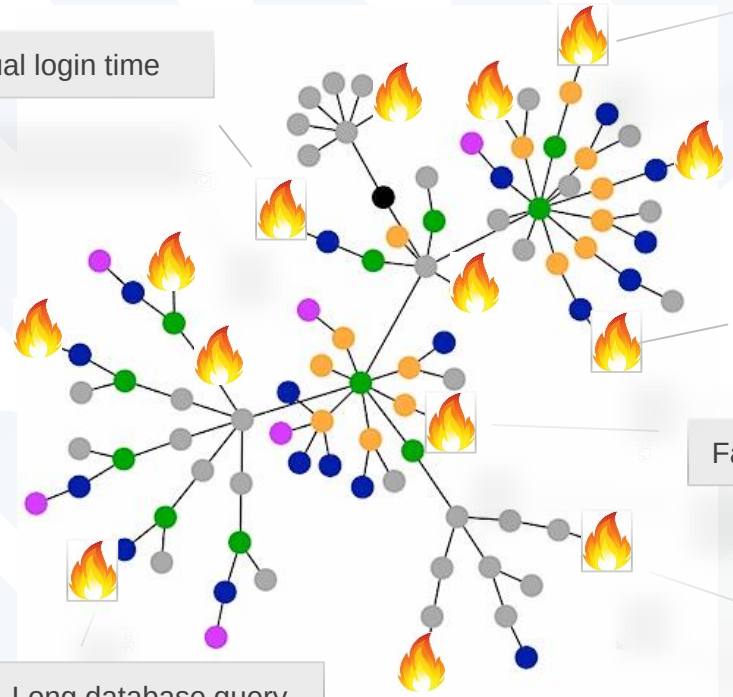
- Ports open
- Name resolutions
- HTTP requests
- Destinations of flows
- Amount of data sent/received
- Time of activity, logins
- Fan-in and fan-out

PITFALLS IN ANOMALY DETECTION

Bad is anomalous

Anomalous is bad

PITFALLS IN ANOMALY DETECTION



Unusual login time

Unusual data transfer

Long TCP session

Failed DNS resolutions

Abnormal CPU activity

Long database query

FALSES!

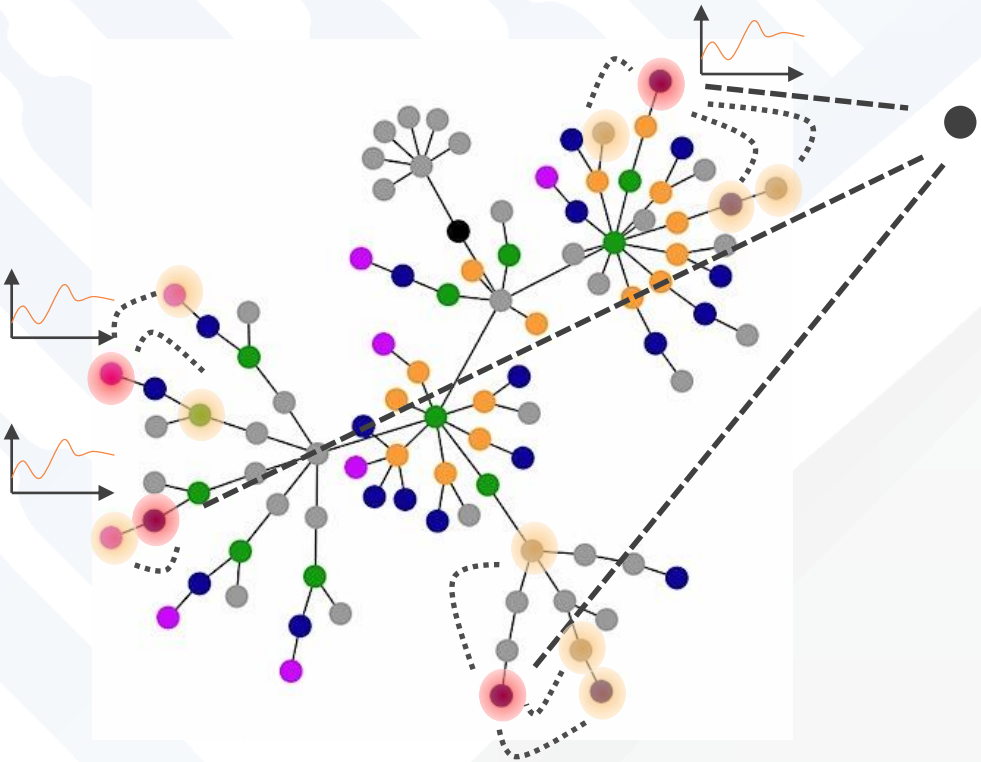
False **Positives**

False **Negatives**

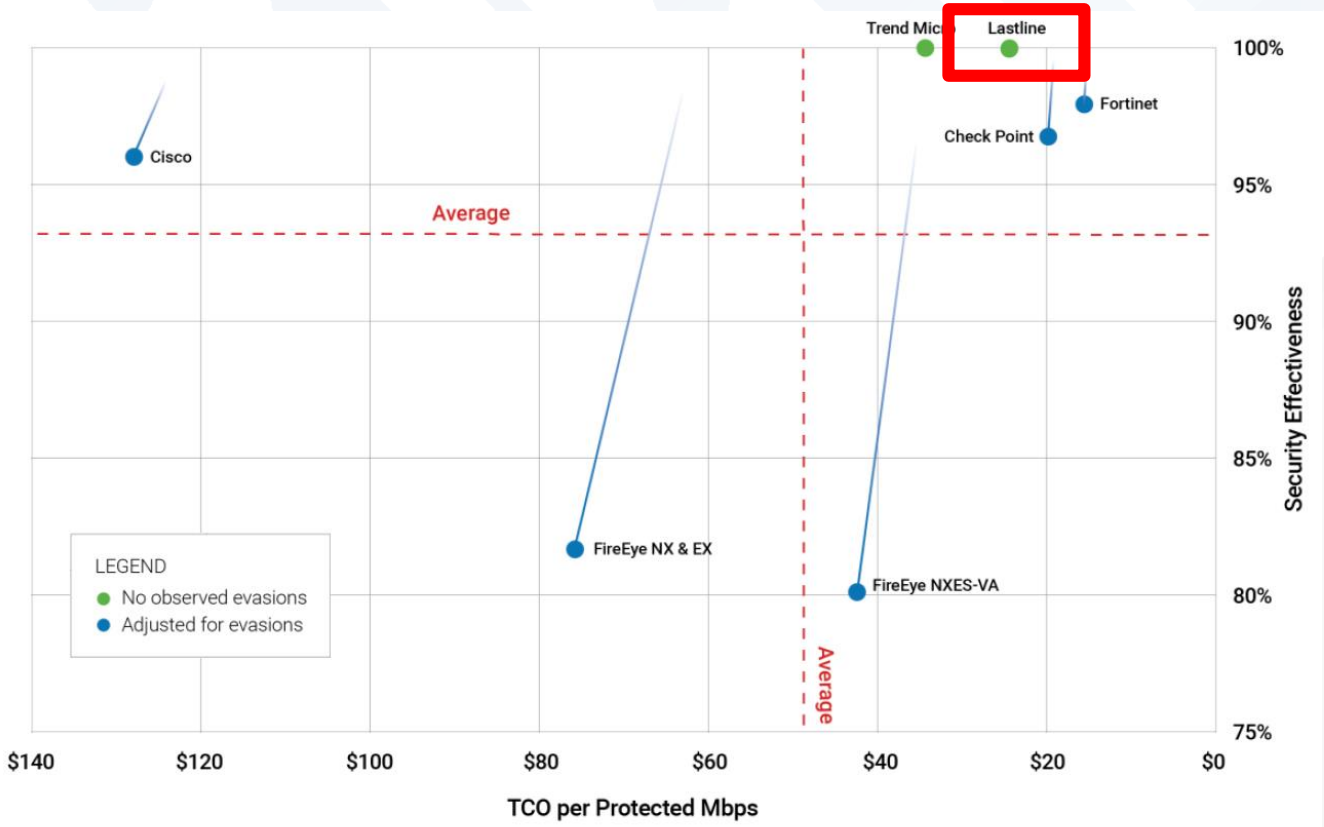
GROUNDING MACHINE LEARNING

- Use anomaly detection as an event generator
- Use known compromised hosts to build models of unknown malicious behavior
- Find similar instances of behavior using clustering
- Explore associations

ML-SUPPORTED THREAT HUNTING



YES, MACHINE LEARNING HELPS



NSS LABS
TRUSTED SECURITY INSIGHT

InfoSecWorld
Conference & Expo 2018

CONCLUSIONS

- Machine learning and network anomaly detection can be valuable tools in identifying breaches
- Adversarial machine learning poses new challenges to traditional machine learning
- Anomaly detection needs to be composed with breach data to avoid false positives
- Innovation over classic techniques is necessary to win against an ever-evolving threat

MIS|TI™ PRESENTS

InfoSecWorld

Conference & Expo 2018

THANK YOU

PLEASE FILL OUT YOUR EVALUATIONS!

Giovanni Vigna, CTO

Lastline, Inc.

SECTION SLIDE

BACKUP SLIDES