# Methods for detecting cyber attacks

## Introduction

Dr. Ran Dubin

Lecture slides were taken from Prof. Asaf Shabtai

# Course Outline

- Email: rand@ariel.ac.il
- Sun 9:00-12:00
- Office hours: (or via email)
- Information: Slides ,articles, books, web.
- This course is inspired by Prof. Asaf Shabtai from Ben- Gurion university.

# Course Goals

- Howe AI can be  applied for detecting cyber attacks
- Learn the inherent challenges in applying AI in security domain
    - Concept Drift
    - Imbalanced datasets
- What are the techniques that can be used for approaching these challenges
- Better understand security risks and how AI can help to detect them.

# My Course Goals

- Introduce you to the domain of machine learning and security
- Overall picture of what has been done in these domains
- Touch and feel AI +security
- MSc/PhD research topics
- Third Year Projects

# Your Course Goals

- Experiment with real life cyber problems
- Read academic papers → transform them to your ideas
- First touch in advanced academic research
  - Is this for you ?
- Innovate

# Program

- 15% תרגיל תכנות ביתי - השנה יהיה בתחום Anomaly Detection
- 35% תרגיל תכנות ותחרות כיתתית – השנה יהיה בתחום Encrypted network malware detection
- פרויקט קורס: בחירת פרויקט מתחילת הקורס מרשימה מוגדרת
  - 10% מצגת על נושאי מתקדמים בלמידת מכונה בסייבר
  - 40% פרויקט מסכם על מימוש מאמר ושיפור מאמר בתחום הסייבר ההגנתי

# Course Plan

יתכנו שינוים קלים
עקב הרצאת אורח

| Lecture | Topic(s) |
|---|---|
| | הקדמה – שימוש בלמידת מכונה בעולם הסייבר סקיוריטי |
| | מבוא  למערכות למידה (הצגת פרויקטים) |
| | מבוא למערכות למידה 2 |
| | גילוי איומים בעזרת אנומליה ( הצגת תרגיל תכנות ביתי גילוי אנומליה) |
| | זיהוי נוזקות בעזרת מערכות למידה (הצגת תחרות כיתה) |
| | שימוש במערכות למידה לניתוח מידע  מוצפן |
| | ניתוח מידע, איסוף מידע ויצירת וקטורי מאפיינים (הצגת תחרות כיתה) |
| | תחרות כיתה והצגת ההישגים |
| | שיטות ensemble בסייבר סקיוריטי |
| | DLP |
| | פרויקט מסכם |
| | פרויקט מסכם |

# Course Projects Examples

**MSC+:**

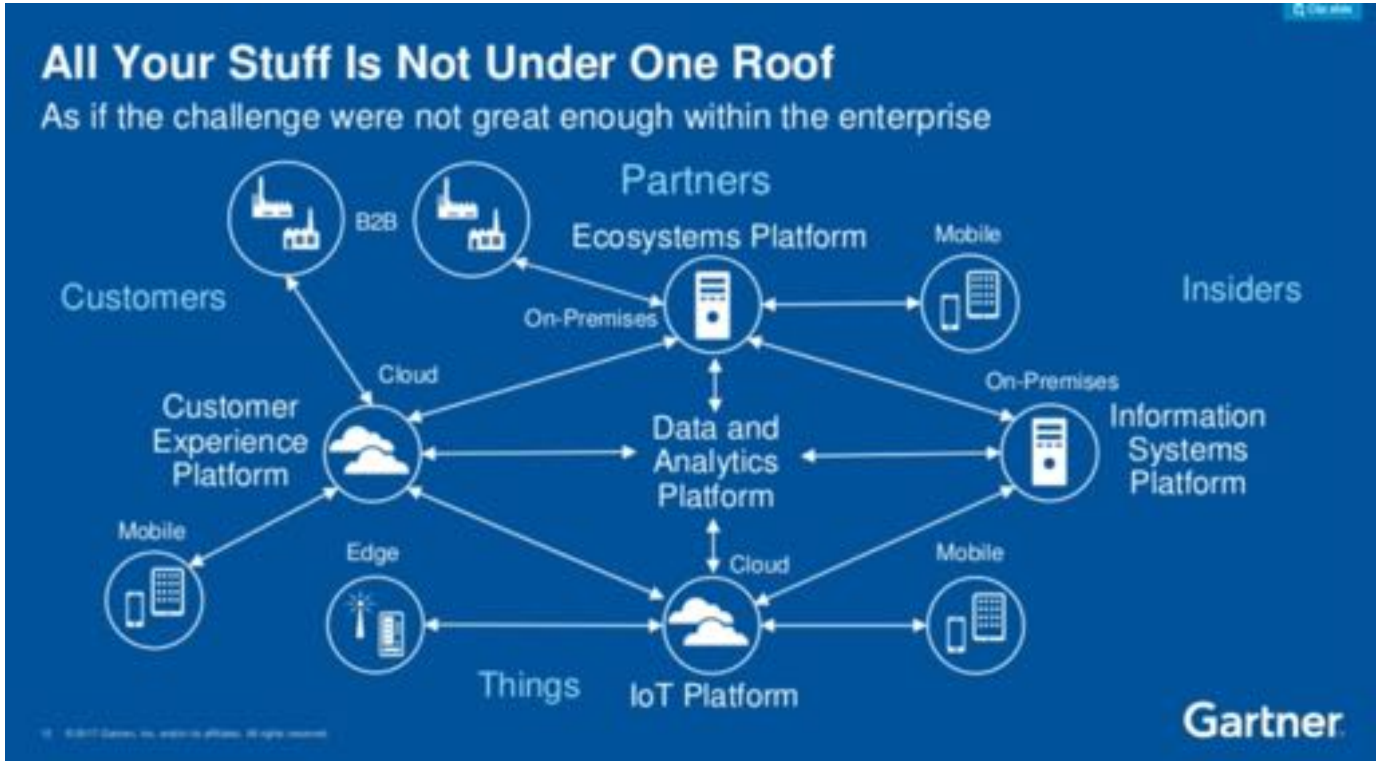
- **DATE: Detecting Anomalies in Text via Self-Supervision of Transformers**
  **Improve change BERT and compare it with and without their approach**
- **Detection zero days using unknown class**
  **Based on An Effective Baseline for Robustness to Distributional Shift**
- Zero shot detection for network malware classification
  Link\

BSC:

- Advanced DGA detection and (data collection or generator function creation)
- URL classification and collection
- Data loss prevention and collection
- Phishing Detection and Phishing data collection

# Security Challenges

- Intrusion detection
- Malware detection
- Authentication/verification
- Database security
- Data leakage detection
- Fraud detection
- SPAM/Phishing detection
- Privacy
- Social networks security
- Web content filtering
- Forensics
- Security analytics
- Malicious insider detection
- …



**All Your Stuff Is Not Under One Roof**

As if the challenge were not great enough within the enterprise

# Security Challenges

- Intrusion detection
- Malware detection
- Authentication/verification
- Database security
- Data leakage detection
- Fraud detection
- SPAM/Phishing detection
- Privacy
- Social networks security
- Web content filtering
- Forensics
- Security analytics
- Malicious insider detection
- …

Solutions

- Anomaly detection
- Classification
- Prediction
- Clustering

# Security Challenges

- Intrusion detection
- Malware detection
- Authentication/verification
- Database security
- Data leakage detection
- Fraud detection
- SPAM/Phishing detection
- Privacy
- Social networks security
- Web content filtering
- Forensics
- Security analytics
- Malicious insider detection
- …

Challenges

- Big data
- Unlabeled data/incomplete data
- Real time vs. offline analysis
- Concept drift
- Imbalanced datasets
- One/multi-class/label
- Performance measures

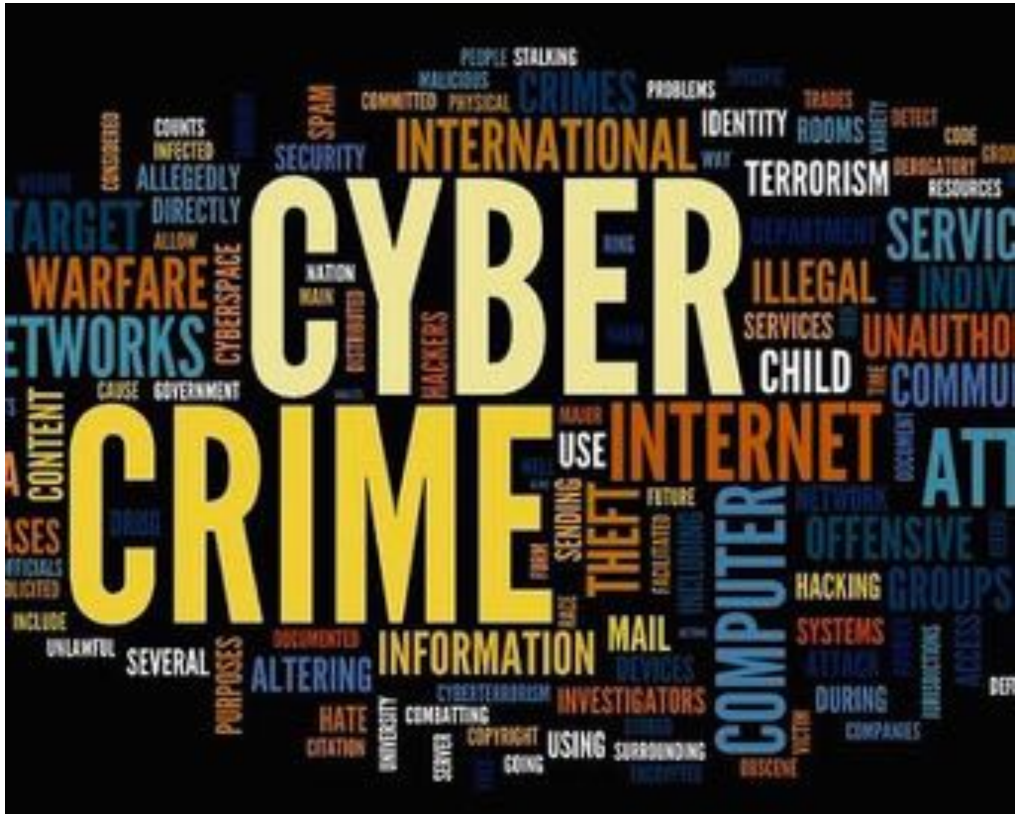
# Security Challenges

- Intrusion detection
- Malware detection
- Authentication/verification
- Database security
- Data leakage detection
- Fraud detection
- SPAM/Phishing detection
- Privacy
- Social networks security
- Web content filtering
- Forensics
- Security analytics
- Malicious insider detection
- …

Challenges
And solutions

- Big data
  - *Dimensionality reduction (sampling, feature selection…)*
- Unlabeled data/incomplete data
  - *co-training or active learning*
  - *Anomaly detection*
- Real time vs. offline analysis
  - *Incremental learning*
  - *Online learning*
- Concept drift
  - *Model and threshold update*
- Imbalanced datasets
  - *Over/under sampling*
- Out Of Distribution Detection

# Introduction to Cyber Detection



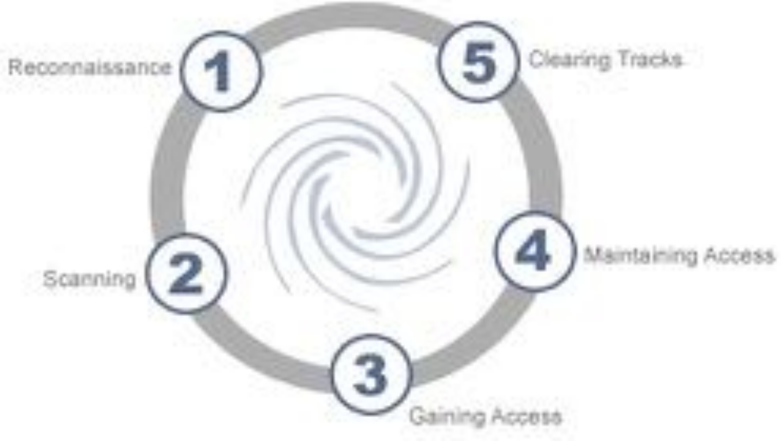
[This Photo](#) by Unknown Author is licensed under [CC BY-SA-NC](#)

# What Is Cyber Crime?

- Any illegal act involving a computer its system or its applications
  - Computer used as a tool
  - Computerized device is used as a target
  - Computer that contain evidences of the crime
- Examples:
  - Identity theft, Internet fraud, ATM fraud, Bank transfer fraud, file sharing and piracy, hacking ,computer virus, denial of service, spam and more ..

# Advanced Presist Threat

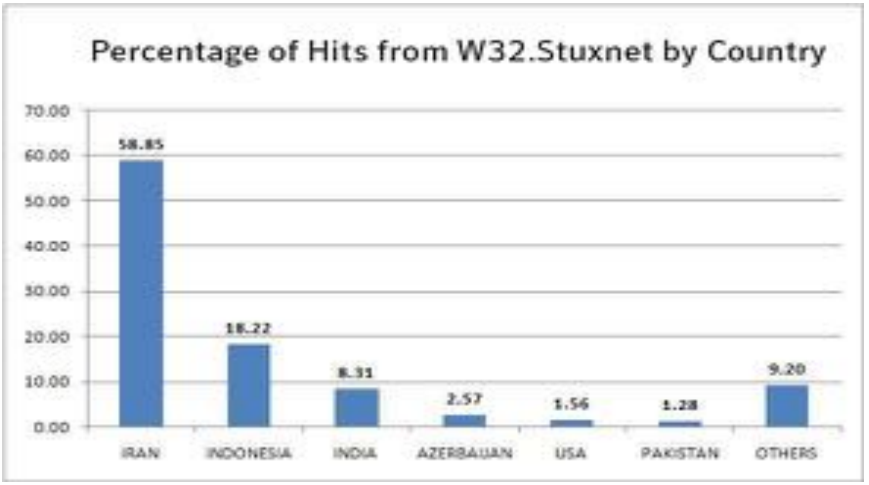
- A cyber attack mounted by organizational teams (e.g., foreign government) that have **deep resources**, **advanced penetration skills**, **specific target** profiles (political or commercial) and are remarkably **persistent** in their efforts



- Operates in a stealth mode over a prolonged period of time in order to achieve their objectives
- Reconnaissance- explore /learn the target

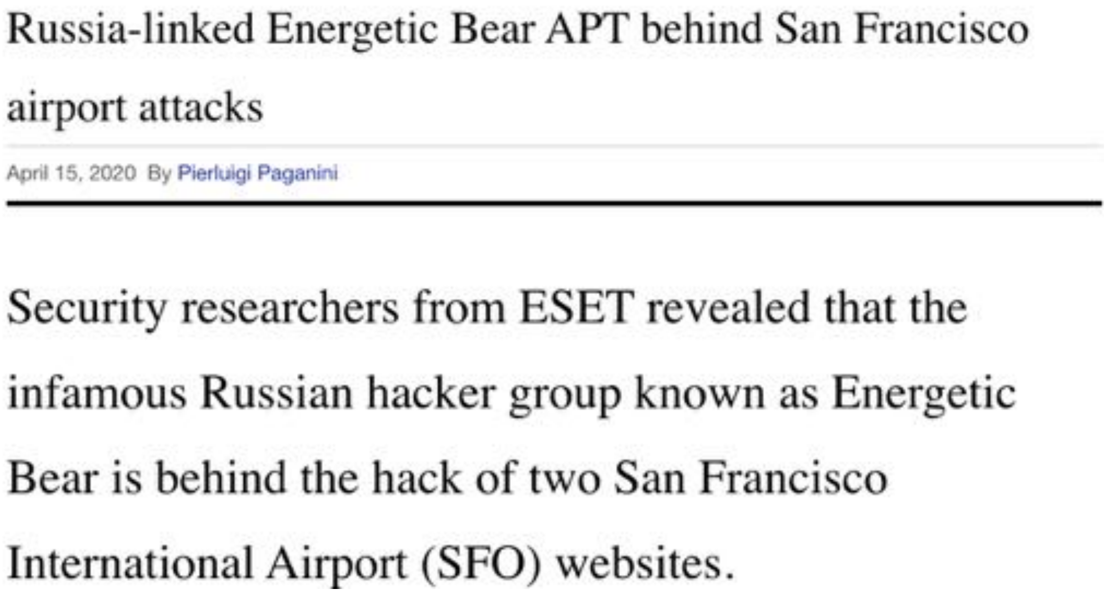
# APT- StuxNet

- July 2010 – **Stuxnet,** a Microsoft Windows computer worm, targets industrial software and equipment
- Attackers used exploit code on vulnerable Windows systems and a **zero-day bug in the Print Spooler Service that makes malicious code passed**. This code spies on the system and sends information back to the hackers



# APT Example –Watering Hole

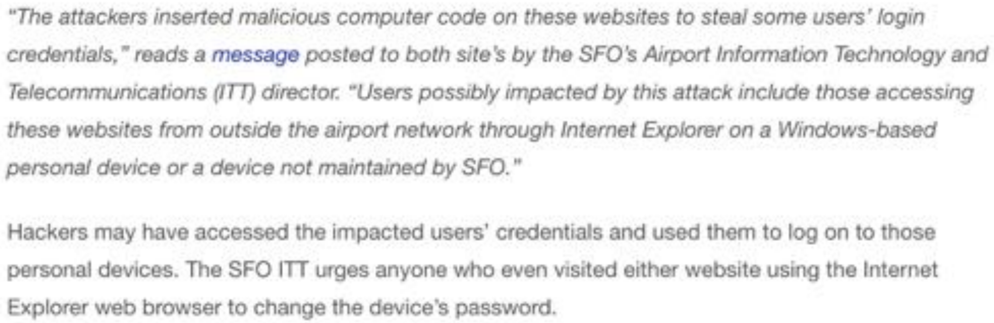
Designed to steal windows user login credentials



## Russia-linked Energetic Bear APT behind San Francisco airport attacks

April 15, 2020 By Pierluigi Paganini

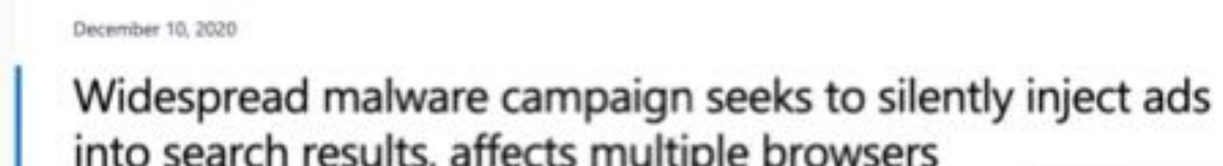
Security researchers from ESET revealed that the infamous Russian hacker group known as Energetic Bear is behind the hack of two San Francisco International Airport (SFO) websites.



*"The attackers inserted malicious computer code on these websites to steal some users' login credentials," reads a message posted to both site's by the SFO's Airport Information Technology and Telecommunications (ITT) director. "Users possibly impacted by this attack include those accessing these websites from outside the airport network through Internet Explorer on a Windows-based personal device or a device not maintained by SFO."*

Hackers may have accessed the impacted users' credentials and used them to log on to those personal devices. The SFO ITT urges anyone who even visited either website using the Internet Explorer web browser to change the device's password.

# Adware



December 10, 2020

Widespread malware campaign seeks to silently inject ads into search results, affects multiple browsers

Microsoft 365 Defender Research Team

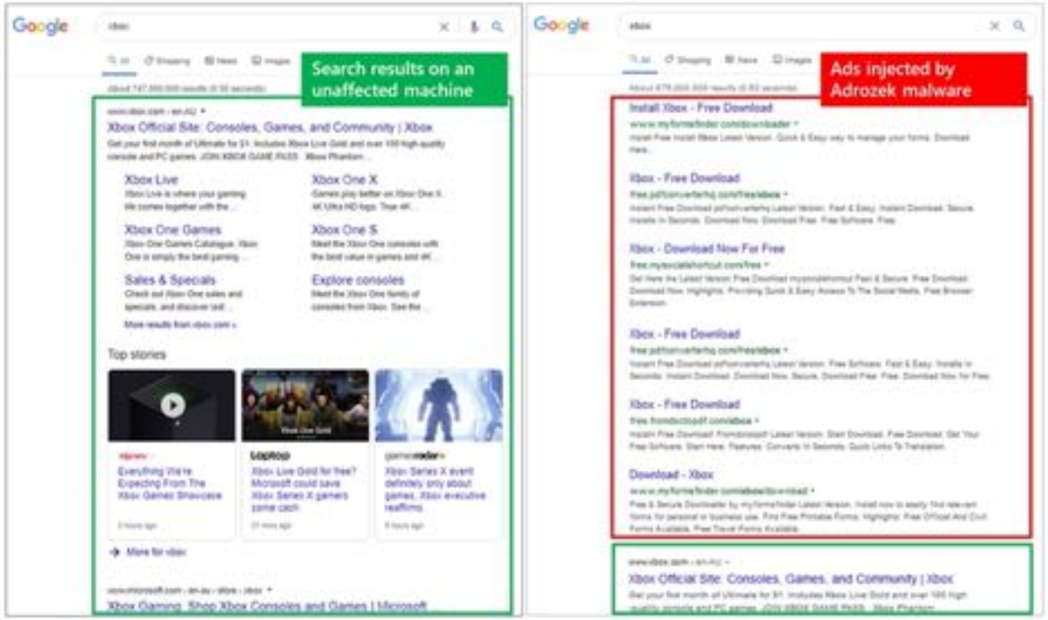

Google
Search results on an unaffected machine
Google
Ads injected by Adrozek malware

*Figure 1. Comparison of search results pages on an affected machine and one with Adrozek running.*

Adrozek adds browser extensions, modifies a specific DLL per target browser, and changes browser settings to insert additional, unauthorized ads into web pages, often on top of legitimate ads from search engines

# Cyber Security Domain

- Evolving domain – endless game
- Plenty of data
- Practical contribution
- Strong support of the stakeholders

# Why Machine Learning

- Cyber security is a domain problem, not a domain solution, thus it seeks solutions from other areas

- Traditionally, security problems were aided by mathematical model. e.g.,
  - Password security – using cryptography

# Why Machine Learning – Modern Cyber Security

- Signatures are expensive and not robust enough.

- Scale our insights and data

- Deals with abstract threats which cannot be solved only by using mathematical models:
  - Malware detection
  - Intrusion detection
  - Data leakage
  - …

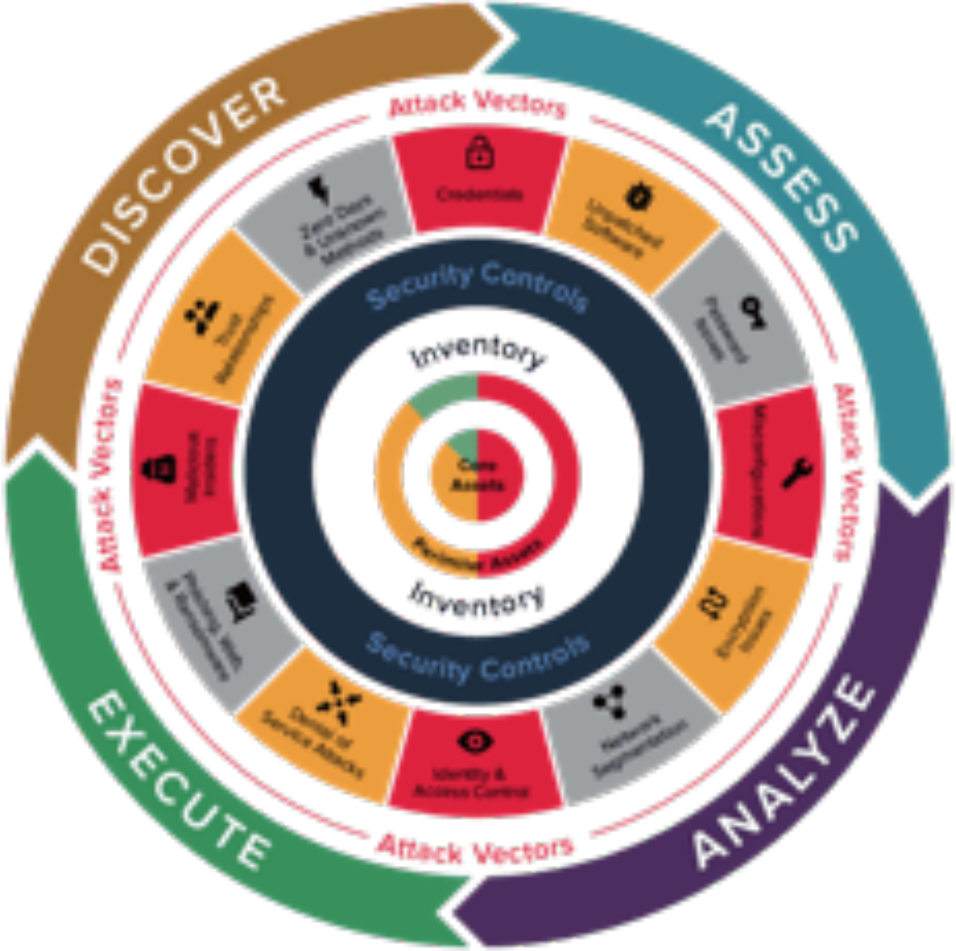
- Need for other research methods

# Why Machine Learning

- Complex task
  - Protect against whom? employees, business partners, hacker/attacker, customers…
  - Intentional vs. accidental
  - Complex systems
  - Evolving technologies (smartphones, cloud…)
  - Increasing amounts of data
  - Detecting complex links between attributes

# Understanding The Threat Landscape

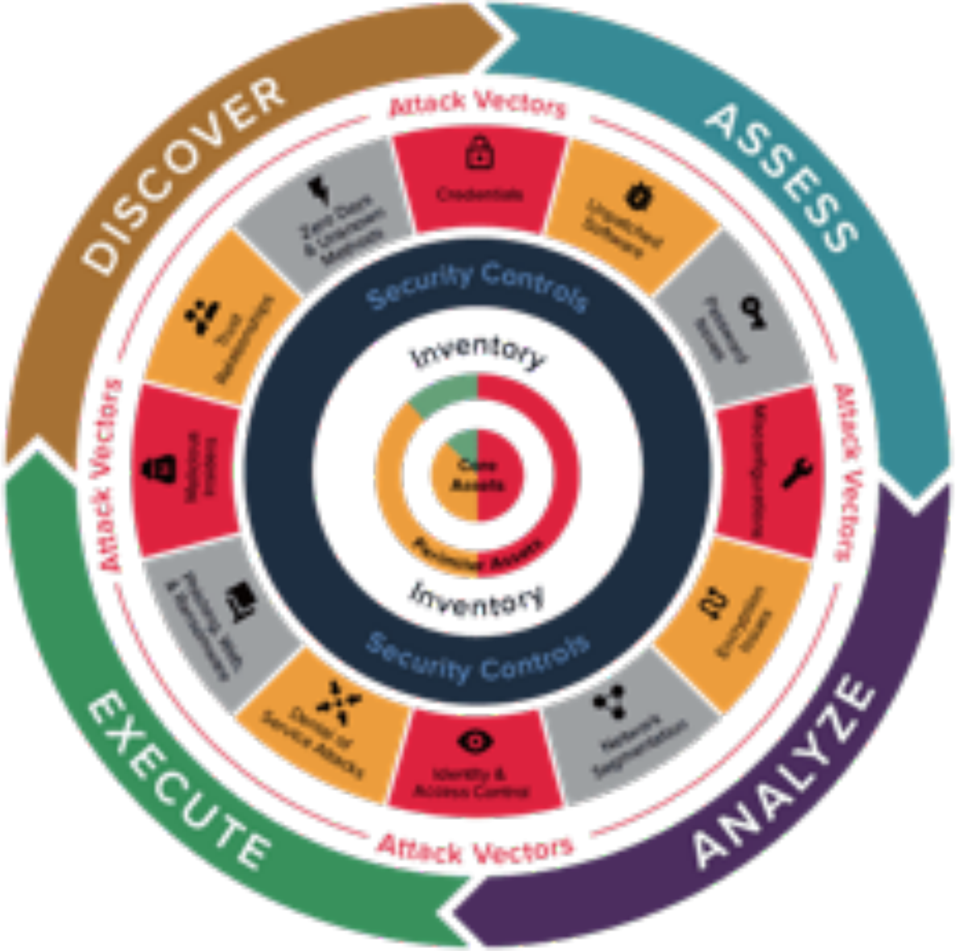
- Let’s look at this diagram, starting from the center and going outwards to discuss the concentric circles.
- At the very center of your security posture is an inventory of all your assets:
  - including core and perimeter assets
  - on-prem, cloud, mobile, and 3rd party assets;
  - managed and unmanaged assets
  - applications and infrastructure, catalogued based on geographic location.



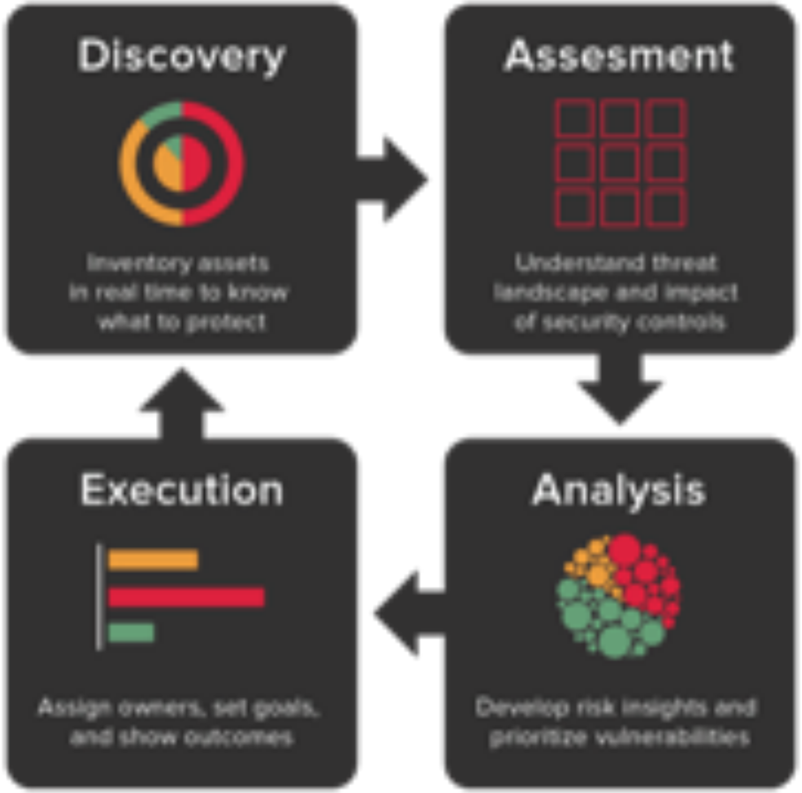
Slide source: https://www.balbix.com/blog/how-to-picture-your-enterprise-security-posture/

# Understanding The Threat Landscape

- Attack vectors are the methods that adversaries use to breach or infiltrate your network.
- Attack vectors:
  - malware and ransomware
  - man-in-the-middle attacks
  - compromised credentials
  - phishing (91% of the attacks starts with an email).
  - Some attack vectors target weaknesses in your security and overall infrastructure, others target weaknesses in the humans that have access to your network.



# Understanding The Threat Landscape



Discovery

Inventory assets in real time to know what to protect

Assesment

Understand threat landscape and impact of security controls

Analysis

Develop risk insights and prioritize vulnerabilities

Execution

Assign owners, set goals, and show outcomes

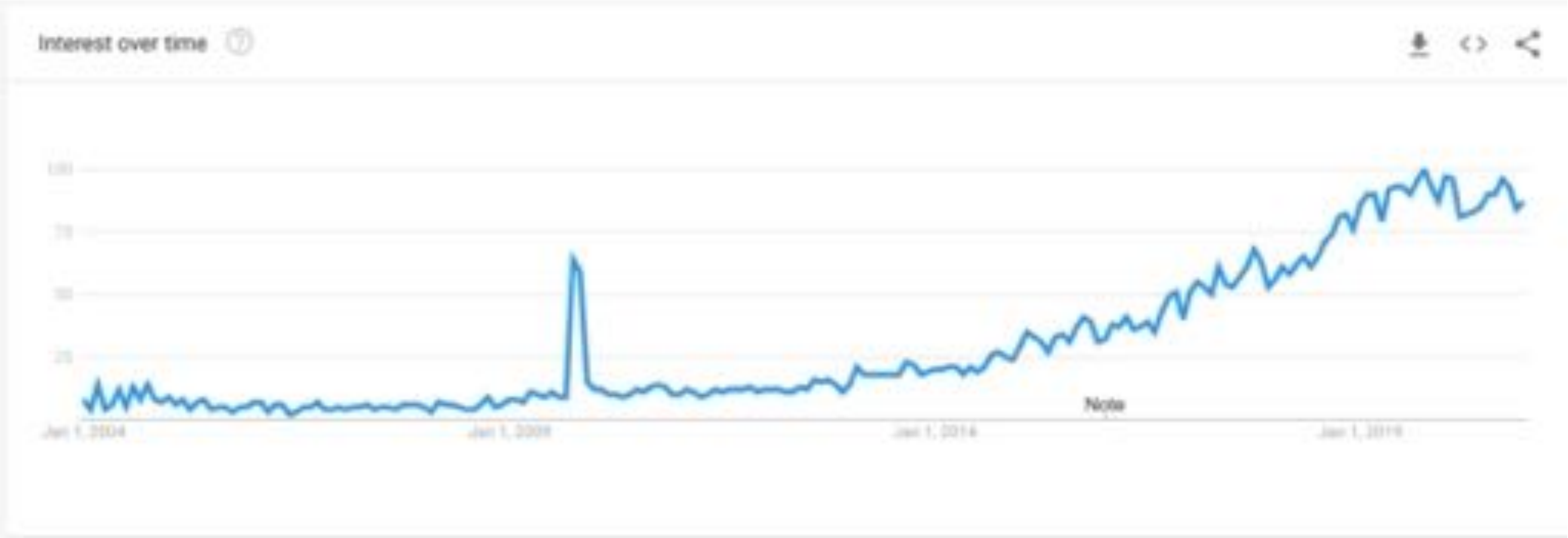
Image source::https://www.balbix.com/blog/how-to-picture-your-enterprise-security-posture/

# Use Expirience From Other Domains

Multimedia
Security
Handwriting recognition
Recommender systems
Face detection
CRM (Customer relationship management)
Personalization
Image processing
Computer vision
Natural language processing
Marketing
Text summarization
Manufacturing
Computer Security
Search engine
Sentiment analysis
Face tracking
Market basket analysis
Information retrieval
Face recognition
Diagnosis
Bioinformatics
Game
Speech recognition
Gene expression
E-commerce
Medicine
Anomaly detection
Human interaction
Intrusion detection system
Collaborative filtering
Object recognition
Fraud detection
Spam

Example : Intezer – genetic software mapping for malware analysis

# Cyber Security Search



# The Concept of Learning In a ML System

- Learning = Improving with experience at some task
  - Improve over task *T*,
  - With respect to performance measure, *P*
  - Based on experience, *E*

# SPAM Use Case

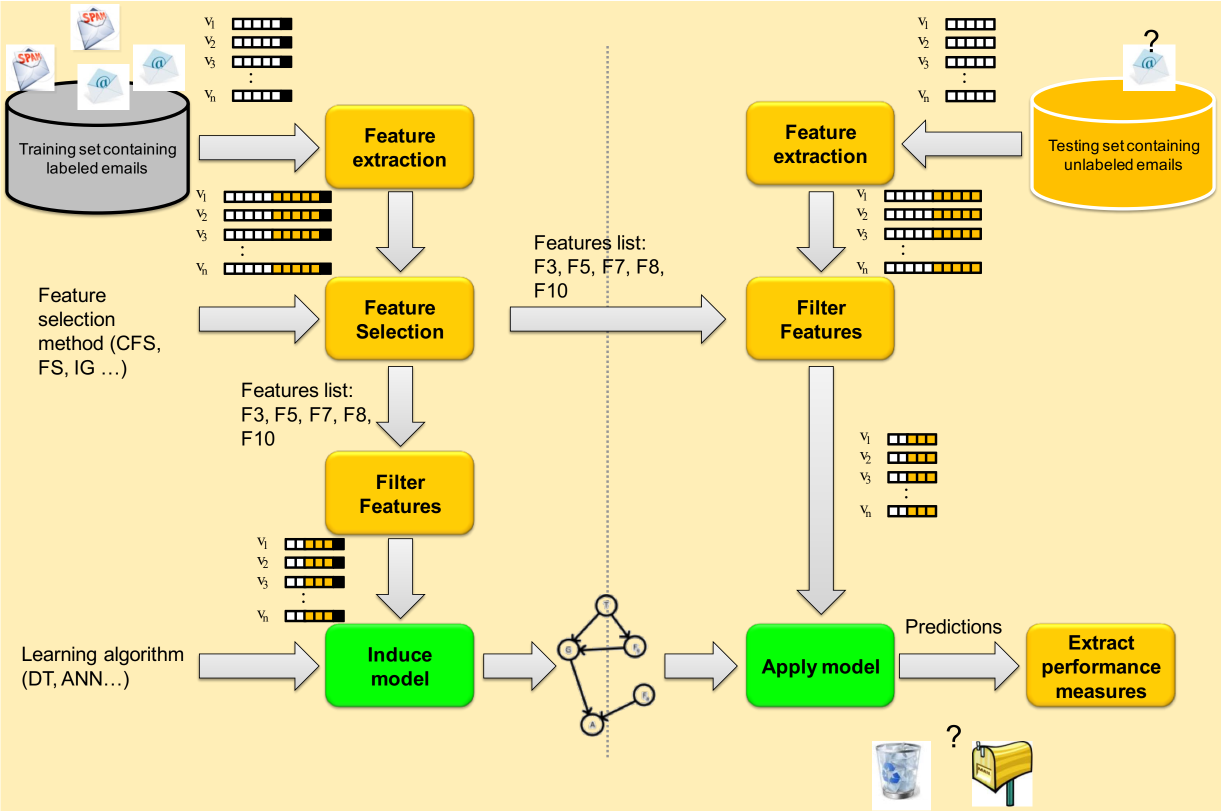
- Spam - an email that the user does not want to receive and has not asked to receive
  - ***Improve*** *task T*: Identify Spam Emails
  - ***Performance*** *metric P*:
    - % of spam emails that were filtered
    - % of ham (non-spam) emails that were incorrectly filtered-out
  - *Based on* ***experience*** *E*:
    a database of emails that were labelled by users/experts

# Dataset

Instances

| Input Attributes | | | | Target Attribute |
|---|---|---|---|---|
| Number of new Recipients | Email Length (K) | Country (IP) | Customer Type | Email Type |
| 0 | 2 | Germany | Gold | Ham |
| 1 | 4 | Germany | Silver | Ham |
| 5 | 2 | Nigeria | Bronze | Spam |
| 2 | 4 | Russia | Bronze | Spam |
| 3 | 4 | Germany | Bronze | Ham |
| 0 | 1 | USA | Silver | Ham |
| 4 | 2 | USA | Silver | Spam |

# SPAM -Learning

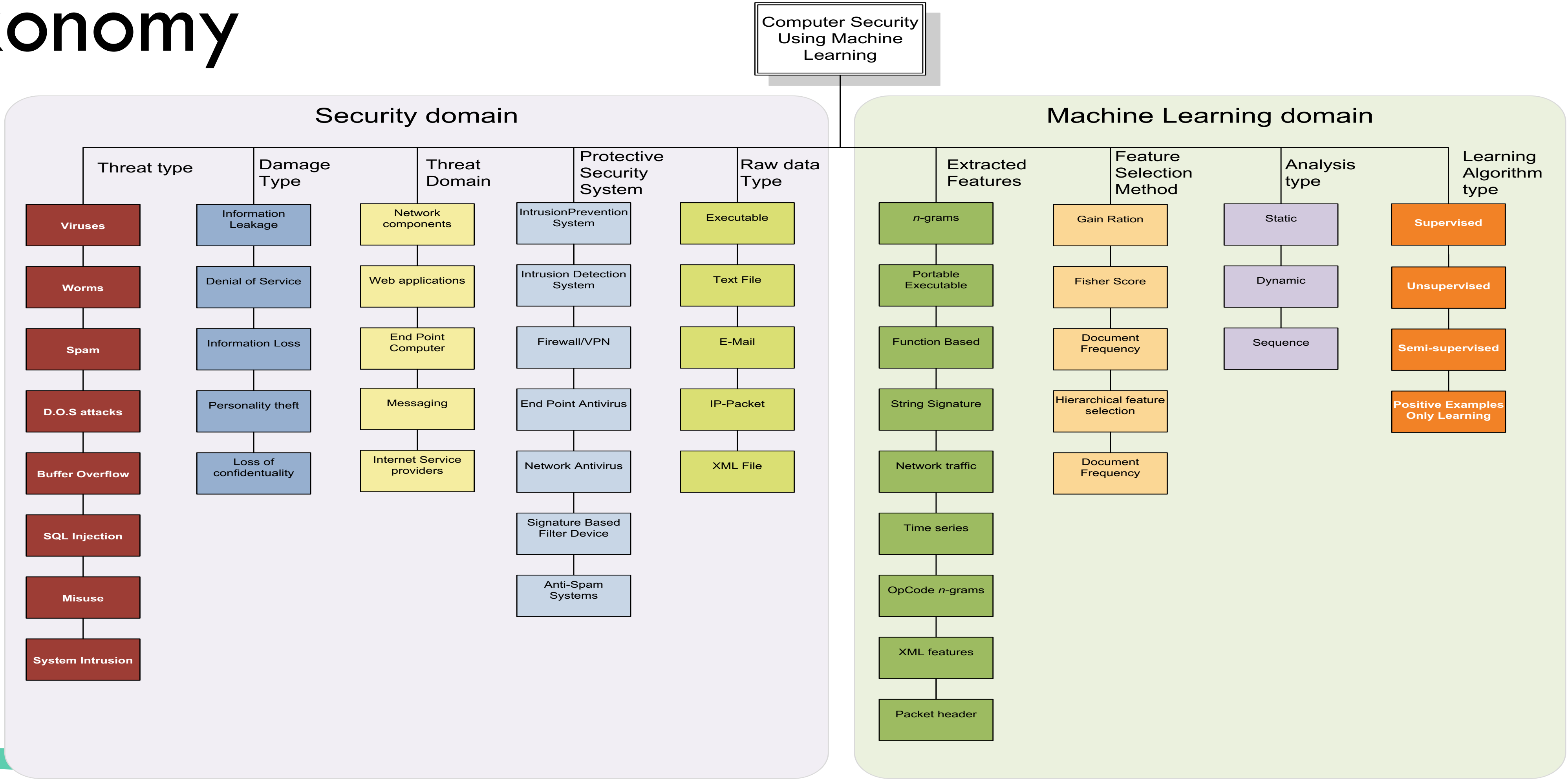

SPAM
SPAM
Training set containing labeled emails
$V_1$
$V_2$
$V_3$
$V_n$
Feature extraction
Feature selection method (CFS, FS, IG …)
Feature Selection
Features list: F3, F5, F7, F8, F10
Filter Features
Learning algorithm (DT, ANN…)
Induce model
Features list: F3, F5, F7, F8, F10
?
Testing set containing unlabeled emails
Feature extraction
Filter Features
Apply model
Predictions
Extract performance measures
?

# Information Security and Machine Learning: Taxonomy

- Problem Domain: Information Security – the problems we need to solve:
    - Malware detection
    - Intrusion detection
    - SPAM mitigation
    - etc.

- Solution Domain: Machine-Learning – from which solutions are drawn
    - Artificial Neural Networks
    - Decision Trees
    - Clustering
    - …

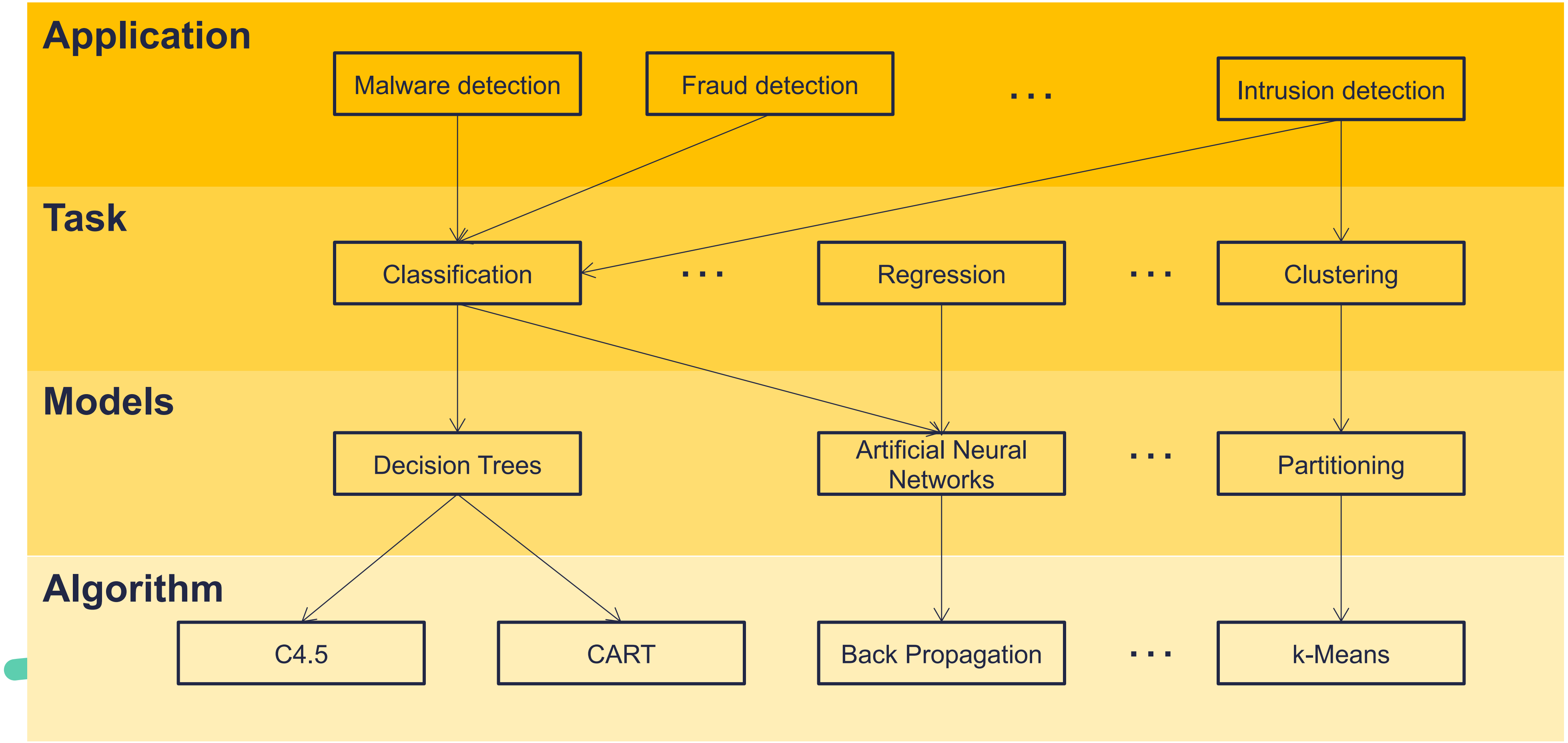
# Information Security and Machine Learning: Taxonomy



Computer Security Using Machine Learning

- Security domain
  - Threat type
    - Viruses
    - Worms
    - Spam
    - D.O.S attacks
    - Buffer Overflow
    - SQL Injection
    - Misuse
    - System Intrusion
  - Damage Type
    - Information Leakage
    - Denial of Service
    - Information Loss
    - Personality theft
    - Loss of confidentuality
  - Threat Domain
    - Network components
    - Web applications
    - End Point Computer
    - Messaging
    - Internet Service providers
  - Protective Security System
    - IntrusionPrevention System
    - Intrusion Detection System
    - Firewall/VPN
    - End Point Antivirus
    - Network Antivirus
    - Signature Based Filter Device
    - Anti-Spam Systems
  - Raw data Type
    - Executable
    - Text File
    - E-Mail
    - IP-Packet
    - XML File
- Machine Learning domain
  - Extracted Features
    - *n*-grams
    - Portable Executable
    - Function Based
    - String Signature
    - Network traffic
    - Time series
    - OpCode *n*-grams
    - XML features
    - Packet header
  - Feature Selection Method
    - Gain Ration
    - Fisher Score
    - Document Frequency
    - Hierarchical feature selection
    - Document Frequency
  - Analysis type
    - Static
    - Dynamic
    - Sequence
  - Learning Algorithm type
    - Supervised
    - Unsupervised
    - Semi-supervised
    - Positive Examples Only Learning

# Four-tier Model



# Data Mining Applications

- Classification
  - Credit scoring/fraud detection
  - Malware detection
  - Intrusion detection
- Clustering
  - User profiling
  - Text analysis
  - Security analytics
- Association rules
  - Intrusion detection
- Regression
  - Forecasting
  - Credit scoring
- Anomaly detection
  - Information security
  - Security analytics
- Sequence mining
  - Intrusion detection
  - User profiling
- …

# Intrusion Detection

- Detecting security policy violations
- Predicting system failures
- Usually by outsiders; i.e., hackers
- Data for analysis
  - System
  - Applications/processes
  - Network

# Malware Detection

- Identifying files containing malicious code/functionality
- Acquiring data; different platforms (Windows, Linux, Java, pdf, Android, Web-based…)
- Verifying clean/malicious files
- Timestamps
- Static vs. Dynamic analysis

# Detecting Malicious Insider

- Detecting insiders that misuse their privileges
- Abnormal access to data
- Abnormal user activity
- Abnormal access to resources (services, servers…)

# Authentication / Verification

- Tracking user activity
- Keyboard / mouse / touch screen dynamics
- Access to files and directories
- Application usage patterns
- Software Defined Perimeter (SDP) – Zero Trust Authentication

# Data Leakage

- Identifying sensitive/private content
- Representing sensitive/private content
- eMail leakage
- Database security

# Forensics

- Authorship verification
  - Determine if given texts were or were not written by suspected author

- Steganography tool detection

- Security analytics; SIEM/SOC systems

# Dataset Challenges

- What data to collect?
  - Identify the threat we are targeting?
    - Detecting email leakage
    - Detecting malware types
  - Understanding the domain
    - Data leakage – encrypted files
    - Malware detection – encrypted files
    - Fraud detection
  - Using domain expert
    - An insider entering a facility on Friday evening and then a new account is created in the active directory

# Dataset Challenges

- Measuring/collecting data
  - Do we need access to the internal system? Simulated environment?
  - Access to data sources
    - Collecting network traffic
    - Email content for email analysis
  - Dedicated (resource aware) utilities
    - Keyboard/mouse activity
    - Application activity
  - Availability of data
    - Detecting database misuse

# Dataset Challenge

- Handling huge amounts of data (big data big problem)
  - Many users, many data sources, many features
  - Dimensionality reduction
  - Preprocessing data (e.g., aggregations)
    - Anti virus/Firewall alerts
  - Performance issues
  - Tuning algorithms
  - Training set selection

# Dataset Challenges

- Privacy issues
  - Accessing sensitive data
    - Email content for email analysis
    - User/device activity for detecting intrusions/verification
    - Social network data
  - Anonymizing data
  - Sharing data

# Dataset Challenges

- Not enough (positive) examples for training
  - Rare events
    - Intrusion or fraud detection
  - Unlabeled data
    - Sensitive data for data leakage detection
    - Example labeling using co training and active learning
  - Imbalanced datasets
    - Malware detection
  - Quality and reliability of the data

# Dataset Challenges – Selecting The Right Technique

- Application; what would we like to achieve?
  - Real-time intrusion detection or fraud detection in online payments vs. offline analysis of suspicious files
- Nature of the data
  - Using anomaly detection for detecting intrusions
- Algorithm/process design
  - Labeling, sequence mining, feature extraction, classification
- Evaluation framework
  - Evaluating various techniques

# Designing & Evaluating

- Resemble real-life scenario
  - Chronological experiments
    - Malware detection
  - Imbalance
  - Data processing
    - Malware detection – encrypted files
    - Email leakage –
  - Simulated data
  - Concept drift – התקיפות משתנות צריך לשנות את ההערכה
    - New platforms
    - New types of attacks
    - System affecting the data

# Challenges
# Performance & Productization

- How to analyze the results
  - Accuracy?
    - What if 99% of the network traffic is free of intrusion attempts?
  - Costs

- Integrating the experimental results into an operational system/environment
  - Resources
  - Data accessibility
  - Acceptable performance
  - Comprehensibility of the model

# Challenges

- Weaknesses of Machine Learning techniques
    - Can an attacker learn the normal behavior?
    - Replay attack
    - DoS
    - Influence on the learning algorithm

# Data Types

- Measurements
- Events
- Text / XML / Tables / Code
- Graphs
- Pictures
- Logs
- Sequences
- Time series
- …

# Data Processing

- Effective feature extraction (may combine several techniques)
- Normalizing attributes
- Discretization
- Predict missing values
- Labeling
- Removing useless attributes
- Selecting examples/attributes relevant for learning

THANK YOU