Decision Rules in Cyber Security Behavioral Models

Anton Mobley
GE Healthcare Product Security

All slides are opinion of the speaker and do not reflect the opinions of GE Healthcare

Talk Outline

1. Facts about “Big Data”
2. Overview of the Detection/Decision Problem
3. Decision Rule Selection
4. Cyber Security Use Cases
5. Model Use Cases
6. Practical Implementation of Models

Facts

- IBM estimates 2.5 quintillion bytes of data are generated per day
- Data mining techniques are use across many domains such as
  1. Targeted Advertising
  2. Financial Analysis / Trading
  3. Remote Sensing
  4. Cyber Security
  5. Health Information
  6. Social Media Analysis
  7. Political Campaigns
  8. Image Processing
- Humans need an army of machines to make most of their decisions

The Problem: Wrong Decisions Cost Money, And Machines Are Sometimes Wrong

False Alarms (Type 1 Error)	Missed Targets (Type 2 Error)
An advertisement is shown to the wrong demographic repeatedly resulting in inefficient ad spend and loss of a customer for the ad tech company	An unidentified undecided voter is not visited resulting in a vote for the opposition party
An employee's computer is taken for a week to be reimaged due to being mistakenly identified as a compromised device	A company's critical files are encrypted with ransomware resulting in lost data or a significant fine.
A medical test leaves a patient in fear of having contracted an extremely rare disease	A trading algorithm is biased toward no action resulting in financial loss
The military puts resources at risk due to flawed intelligence generated from a detector	A medical operation is interrupted because a security control didn't't detect an attack resulting in a failed procedure and patient injury

The Goal Is To Minimize Cost

Intuitively we need the following to calculate risk:

1. Detector performance:

   $P_D(\tau)$, $P_{FA}(\tau)$

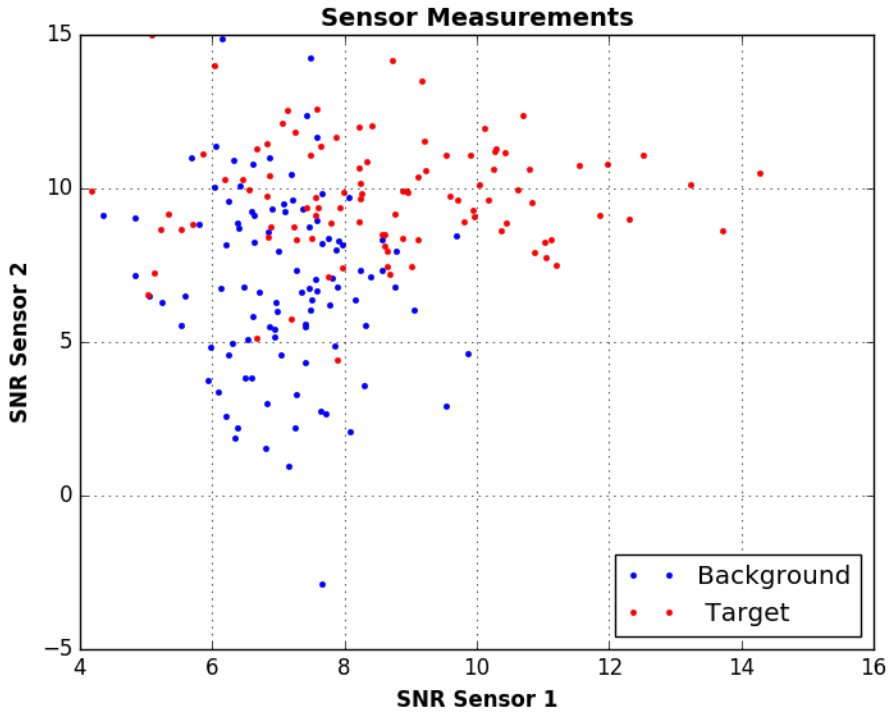
2. Cost of making an incorrect decision for a given target type:

   c_{FN}, c_{FA}

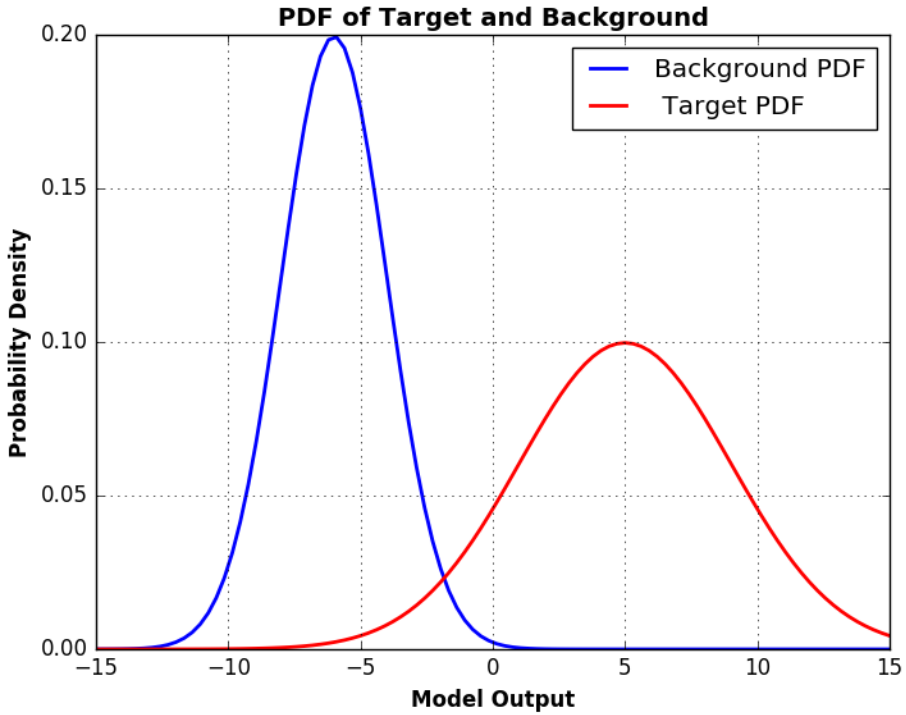
3. Rate at which target events occur relative to background events:

   π_{Target} ,$\pi_{Background}$

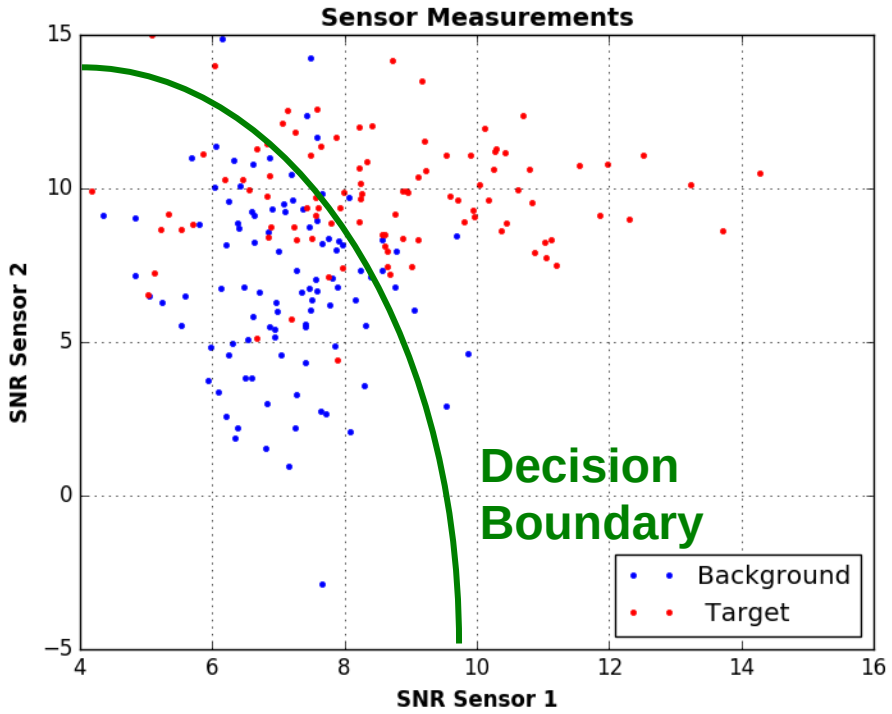
Detector Performance: Mapping to a Decision Space



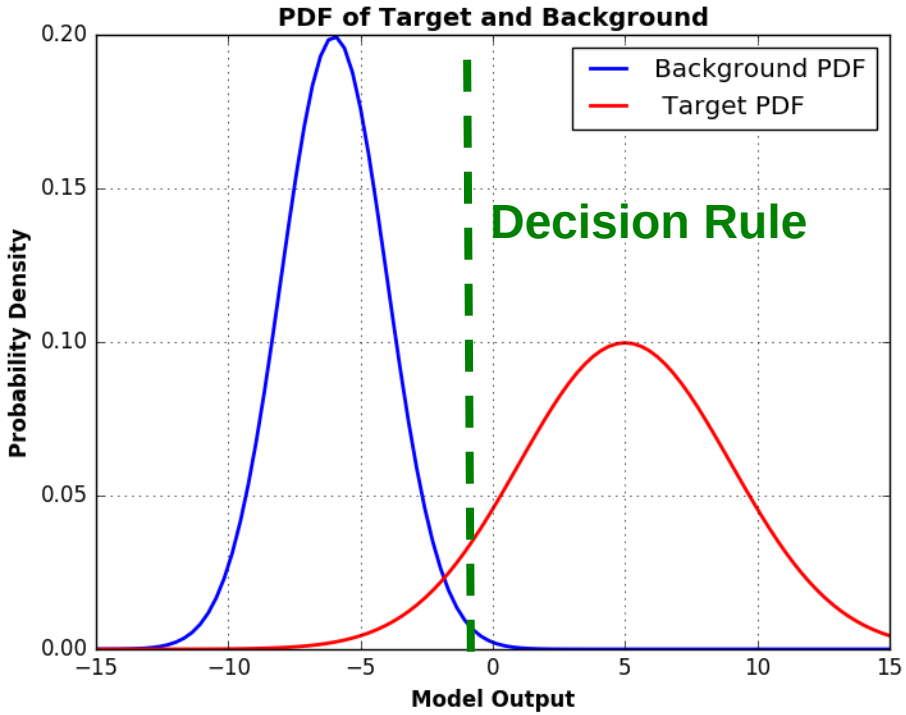
Transform



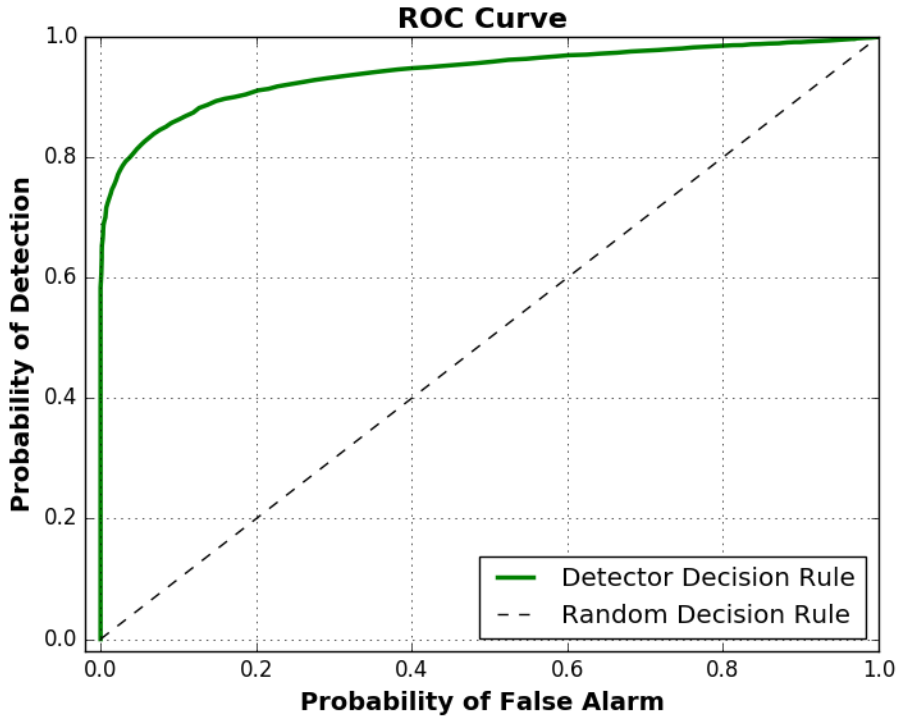
Detector Performance: Making A Decision

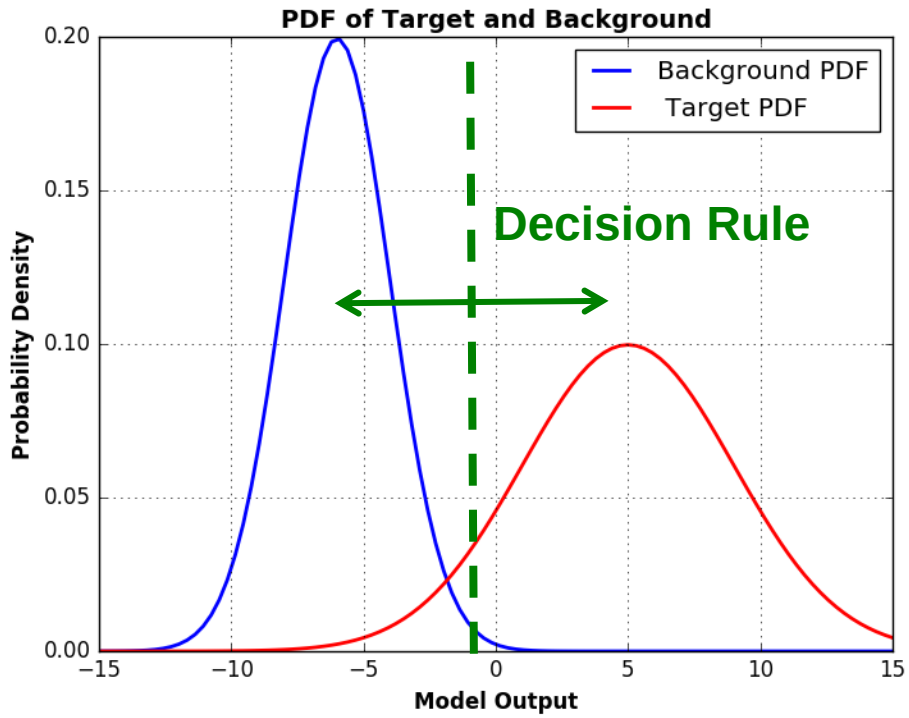


Transform

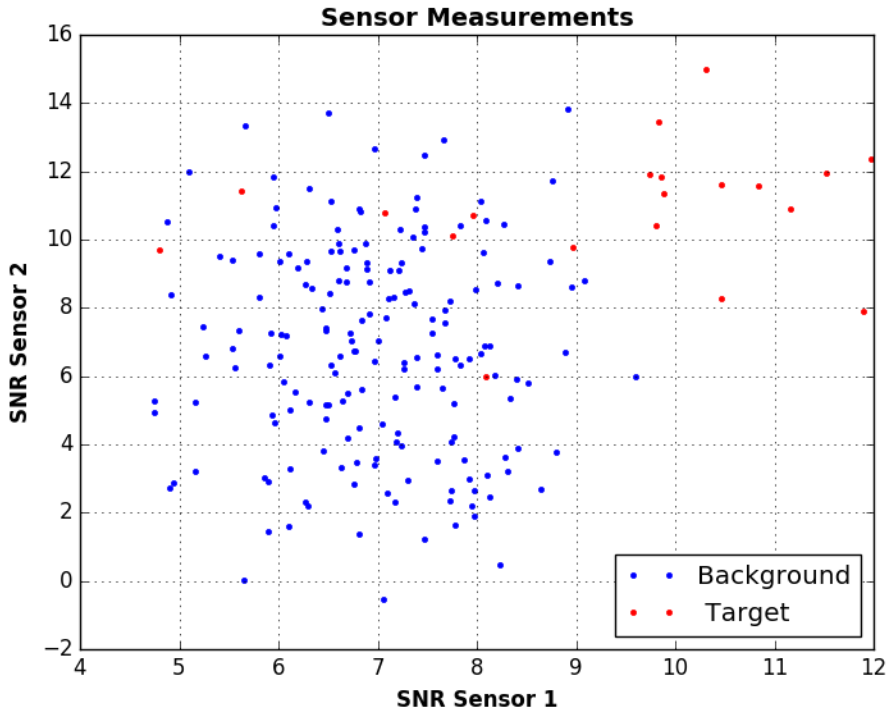


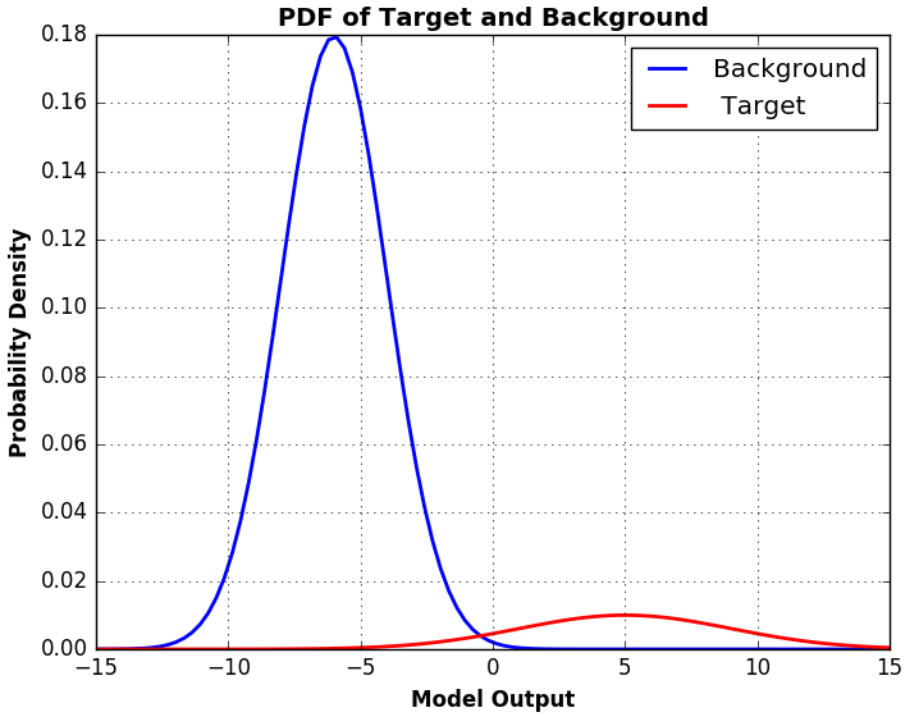
Detector Performance (P_D,P_{FA})





Priors (π_0, π_1)



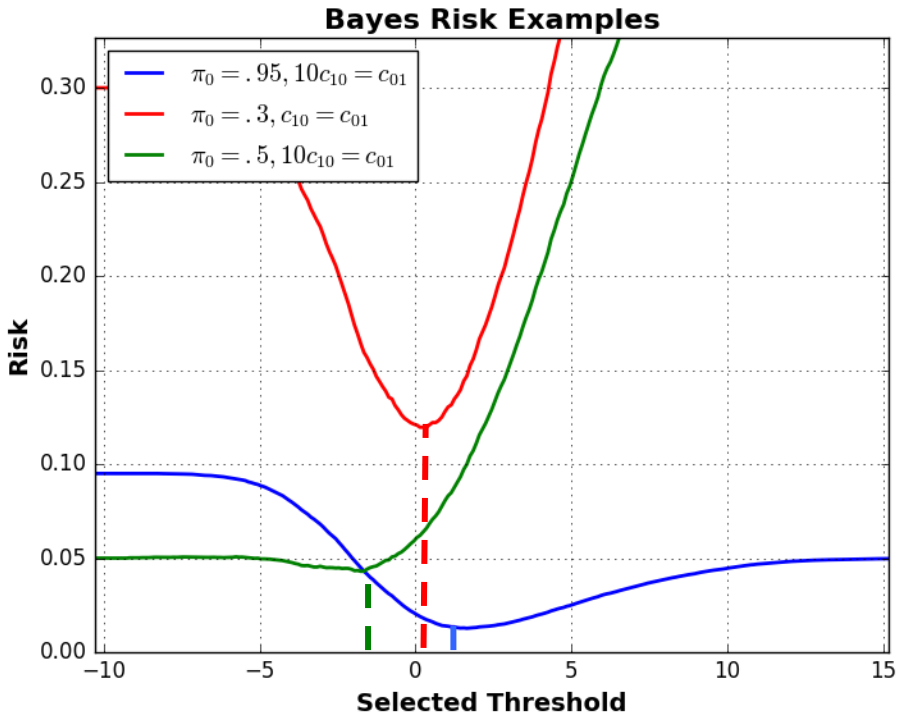


π_0 is the probability that a sample is background
π_1 is the probability that a sample is target
$\pi_0 + \pi_1 = 1$

Costs (c_{FA}, c_{FN})

False Alarms (Type 1 Error)	Cost ($)	Missed Targets (Type 2 Error)	Cost ($)
An advertisement is shown to the wrong demographic repeatedly resulting in inefficient ad spend and loss of a customer for the ad tech company	Ad budget from a company (1,000s+)	An unidentified undecided voter is not visited resulting in a vote for the opposition party	1,000s to the employee
An employee's computer is taken for a week to be reimaged due to being mistakenly identified as a compromised device	Portion of a week's salary	A company's critical files are encrypted with ransomware resulting in lost data or a significant fine.	500-50,000 plus risk of no recovery
A medical test leaves a patient in fear of having contracted an extremely rare disease	Patients trust, possibly 1000s if the test is abandoned	A trading algorithm is biased toward no action resulting in financial loss	1,000s-1,000,000s
The military puts resources at risk due to flawed intelligence generated from a detector	1,000s – 1,000,000,000s, Potential loss of life	A medical operation is interrupted because a security control didn't't detect an attack resulting in a failed procedure and patient injury	1,000s-1,000,000s, Potential loss of life

Expected Risk Minimization

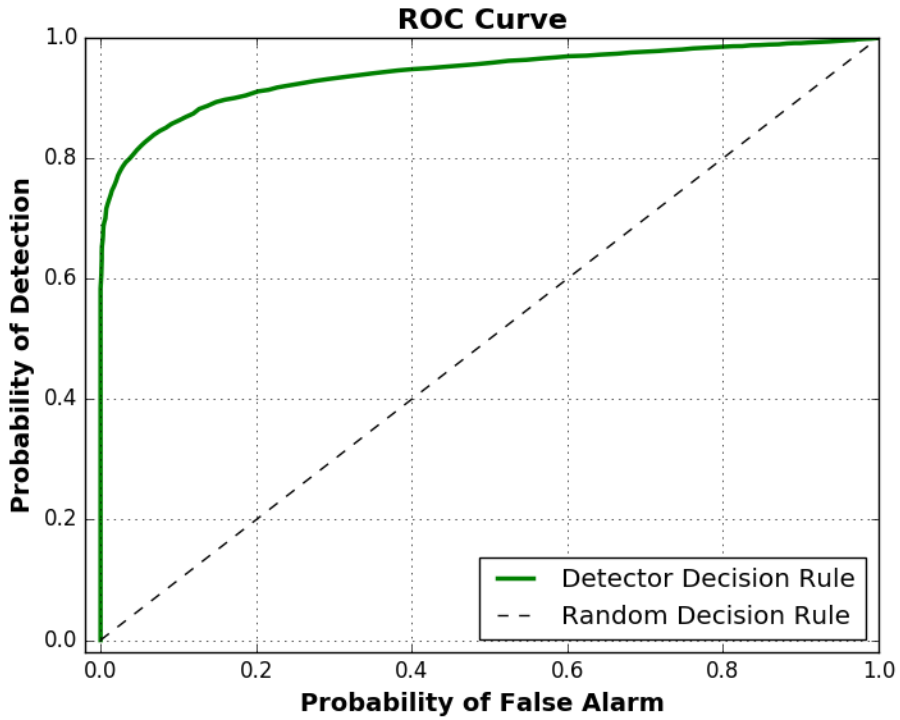


Assume **both** costs and priors are known

Select the threshold that minimizes the risk function:

$$R(\tau) = \underbrace{\pi_0 c_{FA} P_{FA}(\tau)}_{\text{Risk From False Alarm}} + \underbrace{\pi_1 c_{FN} P_D(\tau)}_{\text{Risk From Missed Detection}}$$

Neyman-Pearson Threshold Selection



Assume costs and priors are **unknown**

Select the threshold that maximizes the detection probability for a tolerable false alarm rate

This is just choosing a point on the ROC Curve!

Most often used since priors and costs are rarely precisely known (and costs change) but risk is not really known

Minimax selection can be used when costs are known but priors unknown, however, not as common

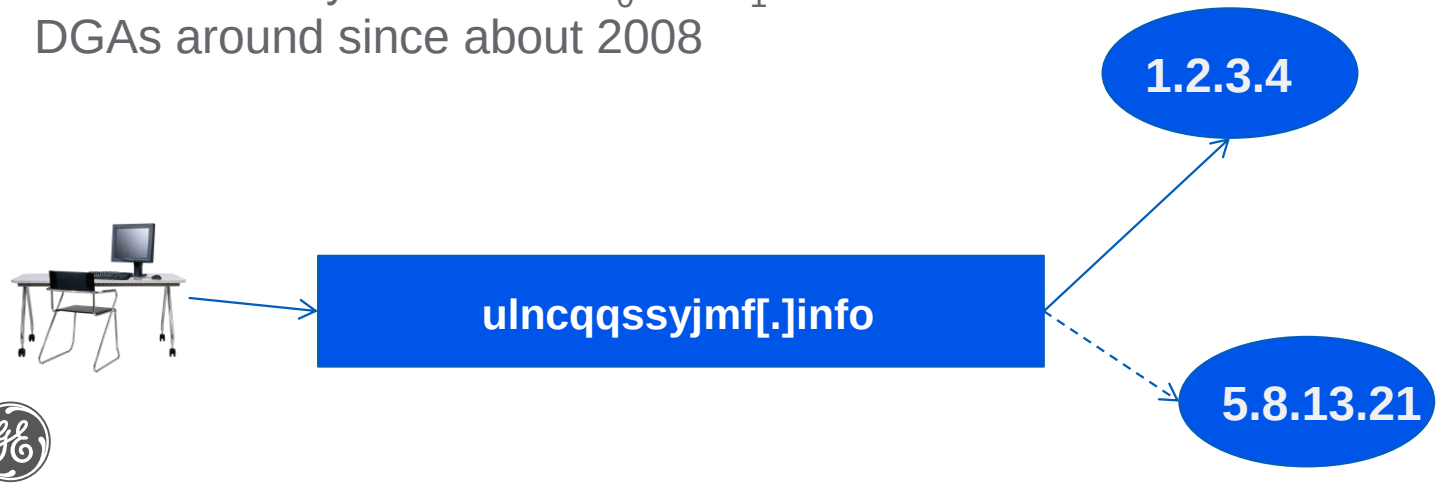
Cyber Security
Command And Control Detection
Use Cases

Use Case #1: Domain Generation Algorithm

- Each Domain is assigned a score based on character randomness
- False alarms can typically be disregarded with less than a minute of research
  - Costs of infection going untreated is much higher than the cost of a false alarm
  - $c_{01} > c_{10}$
- The priors are very uneven
  - In a large network millions of domains are requested each day but a small fraction are caused by malware. $\pi_0 >> \pi_1$
- DGAs around since about 2008

Crypto Locker Beacon Attempts

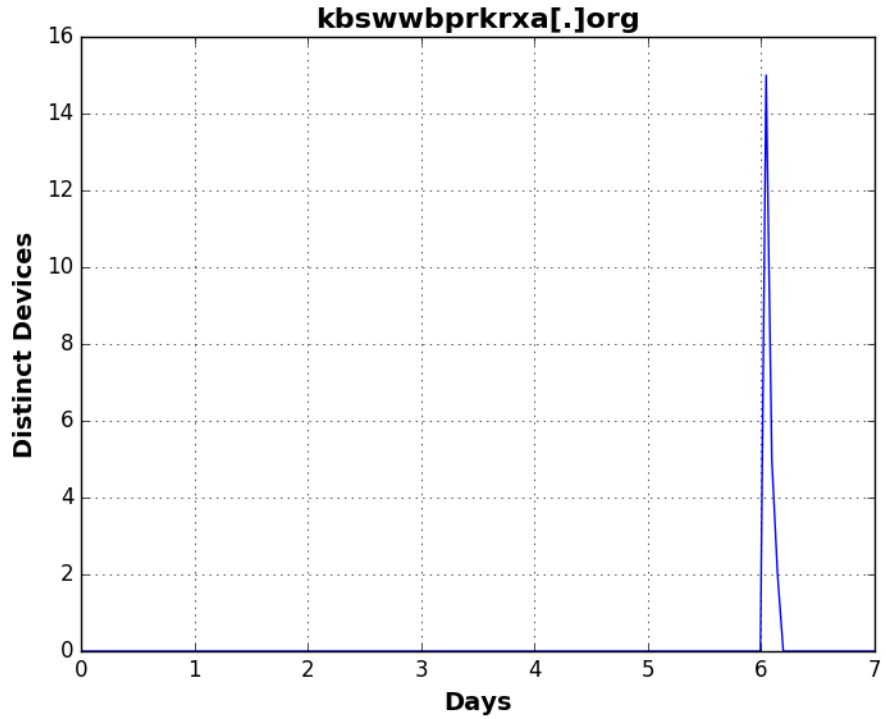
nqmdljcgsqxk[.]org
odnbfevppsqc[.]co[.]uk
ulncqqssyjmf[.]info
ijooyifnmftw[.]com
wtxxcggdcqog[.]net
krykkxsxpmvx[.]biz
wdrkojdwwvqi[.]ru
kbswwbprkrxa[.]org
ylcgayqhadsj[.]co.uk
mjdsiqdcnyab[.]info
dpxhcblijiiy[.]com
ecyfvsvuwpqm[.]net
fxidnqysmpka[.]biz



DGA SIP Demo

Use Case # 2: Synchronous Beaconing

- Many types of malware use command and control domains that are only active for several minutes as a tactic to avoid security researcher blacklisting
- Simple SNR calculation works for detecting this signal
- When many devices connect to a domain that is rarely visited its usually explained by
  - Malware
  - Niche News
  - Software Update



Synchronous Beaconing SIP Demo

Use Case #3: Character Mask Clustering

Full URL
wtxxcggdcqogt[.]net/content/ap1.php?f=b6863
nqmdljcgsqxkj[.]org/content/ap2.php?f=b6863
wtxxcggdcqogt[.]net/content/fdp2.php?f=50
ijooyifnmftwq[.]com/content/field.swf
nqmdljcgsqxkj[.]org/content/score.swf
greateratlantagahomesales.com/engine/listing_detail/standard/3197/en/20276557/
westsiderealtors.com/engine/listing_detail/standard/243/en/26818746/
mollymcgrory.com/engine/listing_results/standard/3252/en/46531933
cnn.com/2017/02/09/us/snowstorm-northeast-weather/index.html

URL Mask Transform

URL Character Mask
/xxxxxxx/xx#.xxx?x=x####
/xxxxxxx/xx#.xxx?x=x####
/xxxxxxx/xxx#.xxx?x=##
/xxxxxxx/xxxxx.xxx
/xxxxxxx/xxxxx.xxx
greateratlantagahomesales.com/xxxxxx/xxxxxxx_xxxxxx/xxxxxxxx/####/xx/#########/
westsiderealtors.com/xxxxxx/xxxxxxx_xxxxxx/xxxxxxxx/###/xx/#########/
mollymcgrory.com/xxxxxx/xxxxxxx_xxxxxx/xxxxxxxx/####/xx/#########/
cnn.com/####/##/##/xx/xxxxxxxxx-xxxxxxxxx-xxxxxxx/xxxxx.xxxx

- Domains that have similar URL formats (GET requests) are often times generated by the same underlying code
  - These patterns can be expressed as character masks
- Use other characteristics such as IP addresses and fuzzy mask matching to group smaller clusters together

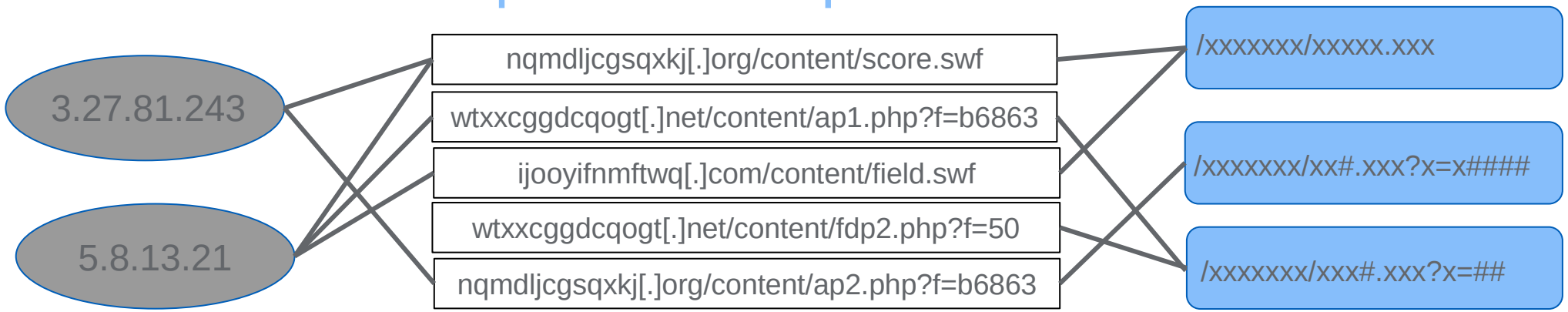
Mask	Count
/xxxxxxx/xxxxx.xxx	2
/xxxxxxx/xx#.xxx?x=x####	2
/xxxxxxx/xxx#.xxx?x=##	1
/xxxxxx/xxxxxxx_xxxxxx/xxxxxxxx/####/xx/#########/	2
/xxxxxx/xxxxxxx_xxxxxx/xxxxxxxx/###/xx/#########/	1
/####/##/##/xx/xxxxxxxxx-xxxxxxxxx-xxxxxxx/xxxxx.xxxx	1

**Blackhole EK example from https://nakedsecurity.sophos.com/exploring-the-blackhole-exploit-kit-7/

Character Mask Clustering Findings

- Domains are grouped by GET request patterns and IP addresses used
- Due to the limited output of the algorithm the priors are similar
  - $\pi_0 \sim= \pi_1$
- Many devices may be detected at once meaning the cost of a false alarm is disproportionately large
- c0 ~= c1
- Limited number of model outputs per run, new patterns are rare
- Character mask patterns are statistically whitelisted by eliminating the most common GET patterns, for example WordPress generated sites

Sample Blackhole Exploit Kit Cluster

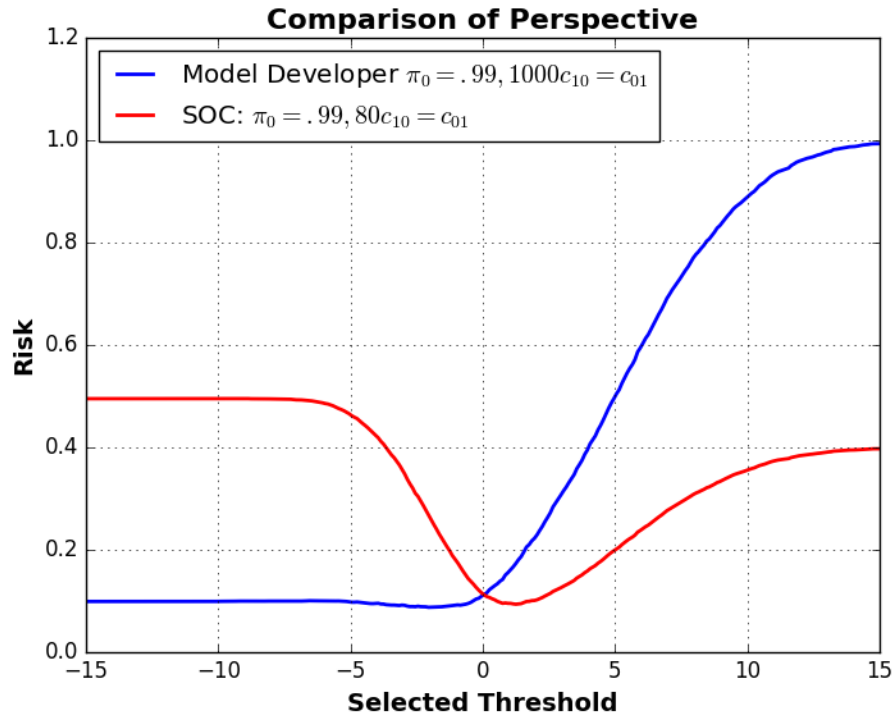


Character Mask Clustering SIP Demo

Notes on Implementation of Cyber Security Use Cases

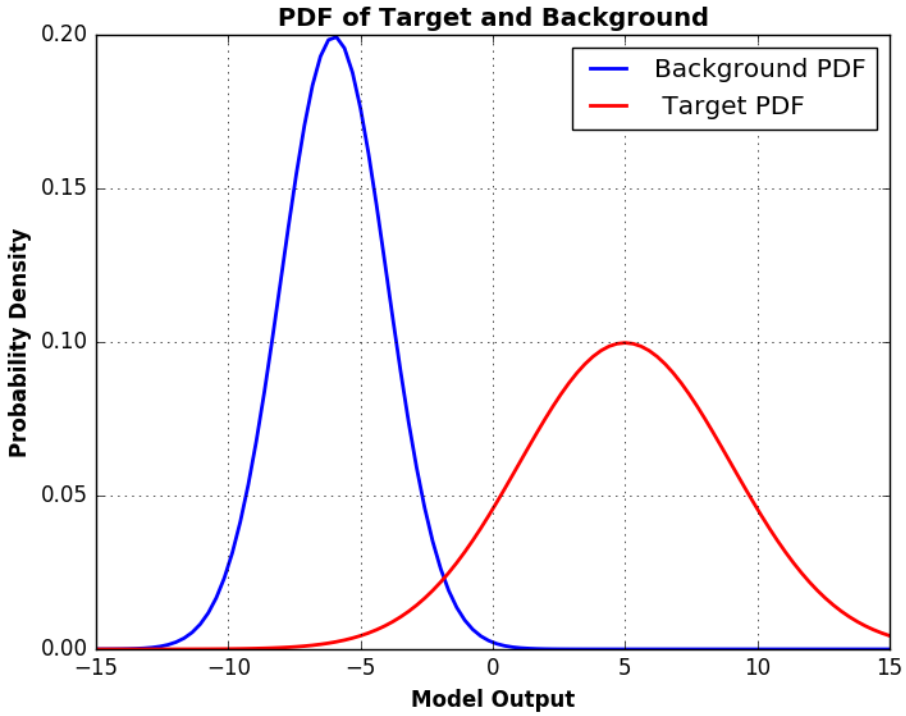
Common Currency and Quantifying Cost

- The true cost of a false alarms / missed positive are difficult to accurately gauge for the people building the models
  - Makes Bayes Risk threshold selection difficult
  - Costs can be very different between decision makers at varying levels
- Costs will be a function of data being protected
- The cost of false alarms and missed positives isn't necessarily dollars
  - Individual utility functions vary
  - The trust in a model is reduced every time someone has to action a false alarm



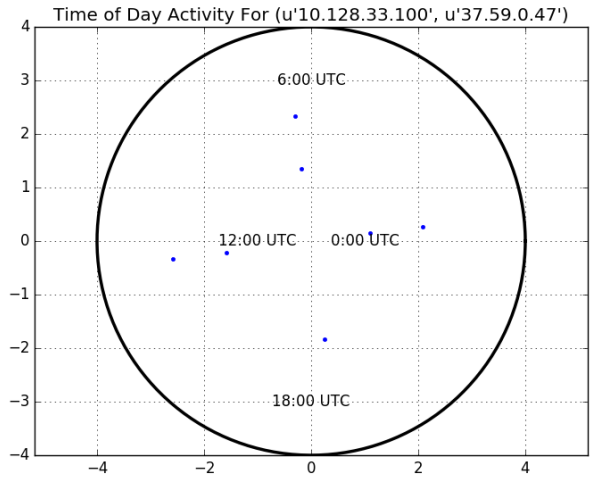
Machine Learning is Based on Empirical Data Sets

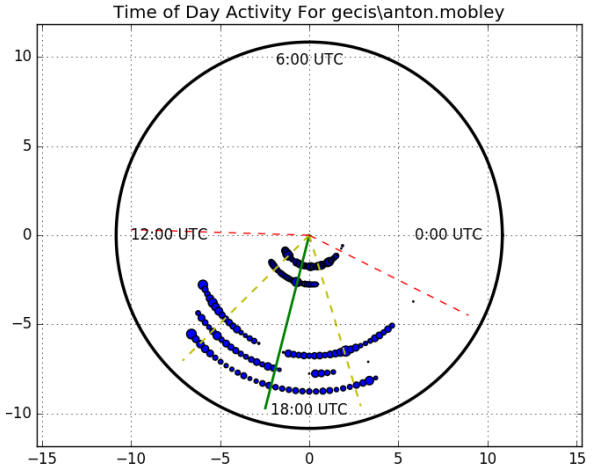
- Need appropriate levels of false alarms, missed targets, background, and target across input features
  - **ROC curves have error bars**
  - Need to setup experiments (A/B, Field test, labs)
- In cyber security there is a tendency to only deal in false alarms when evaluating a tool
- Over-fitting is an issue if model parameters are abused
  - Cross validation is great but often training sets are small



Explainability

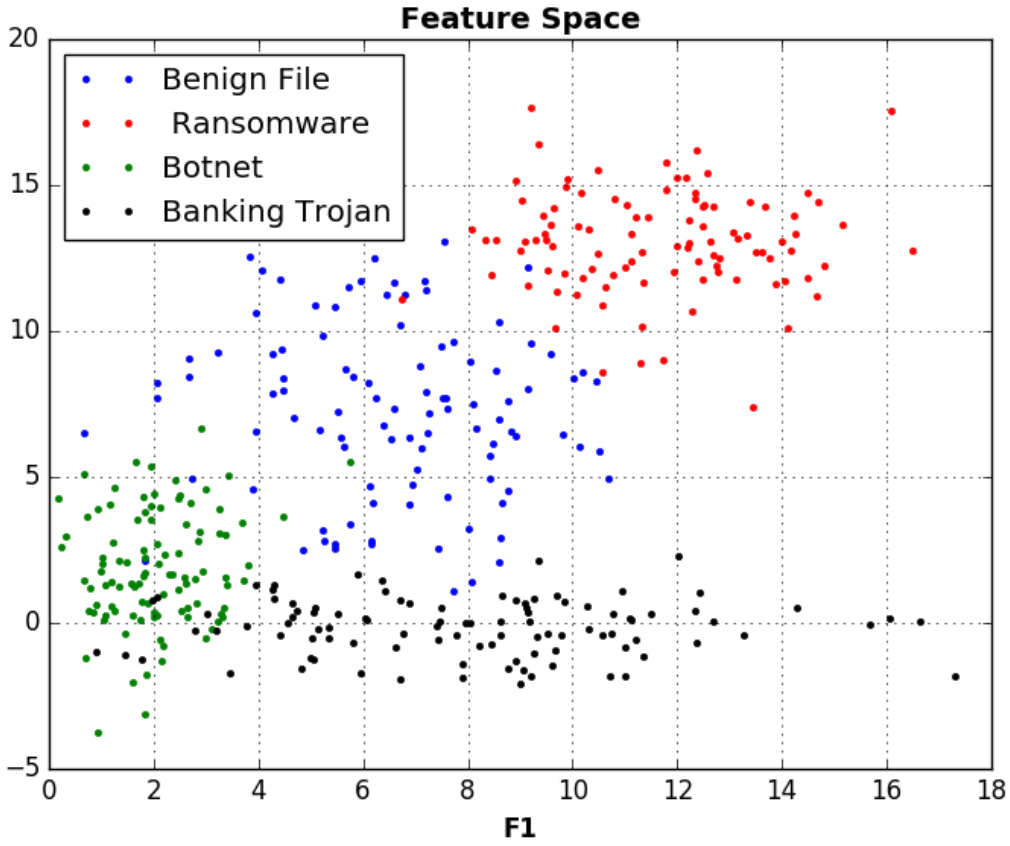
- Security action is often limited by what an investigator understands and observes
- Behavioral models can be combined to produce a score but sometimes individual behaviors can’t occur simultaneously
- Different domains may not care about explainability to perform action, but if there is a human in the loop this is often important





Scope Creep

- Need to be very careful about over-generalizing security models
  - Two class solutions typically easiest
- Tendency to want to score everything as malicious vs benign



Summary

1. There is a lot of data being generated daily
2. Wrong decisions will be made
3. Risk can be minimized by understanding priors, costs, and detector performance
4. Minimizing risk in practice is often less straightforward

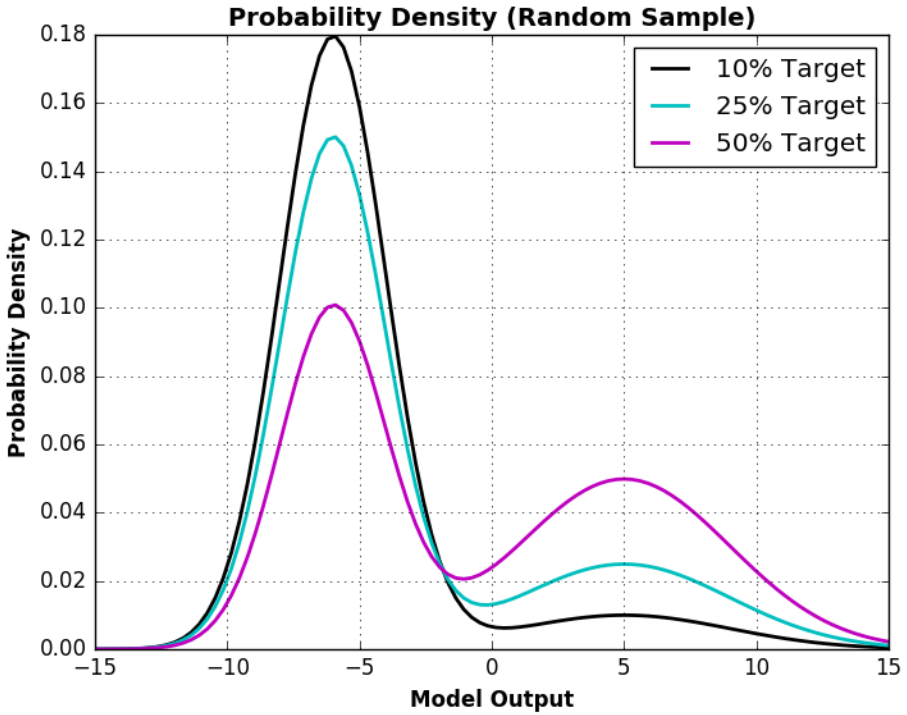
QUESTIONS???

Extra Slides

Talk Outline

1. Overview of the detection problem
2. Overview of threshold selection / cost minimization methods
3. Discuss cyber security use cases
4. Apply use cases to SIP model
5. Provide thoughts on practical implementation of models

Detection Problem: Building the Optimal Decision Rule



π_0 ,π_1 : Prior probability a random sample is the null/test hypothesis

c_{ij} : The cost of choosing *i* when the sample is actually *j*. Note that typically c_{ii} is taken to be 0 and will be assumed

R_i : Conditional Risk from choosing *i*

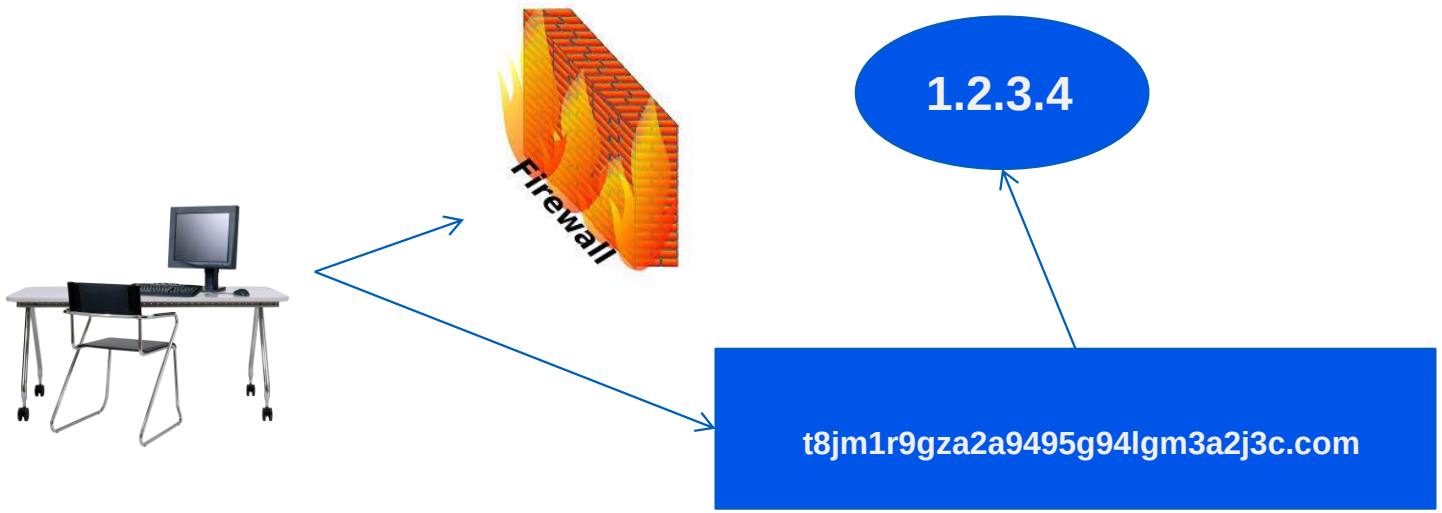
P_{FA}, P_D : Probability of false alarm and probability of detection

δ(τ) : The decision rule as a function of threshold τ

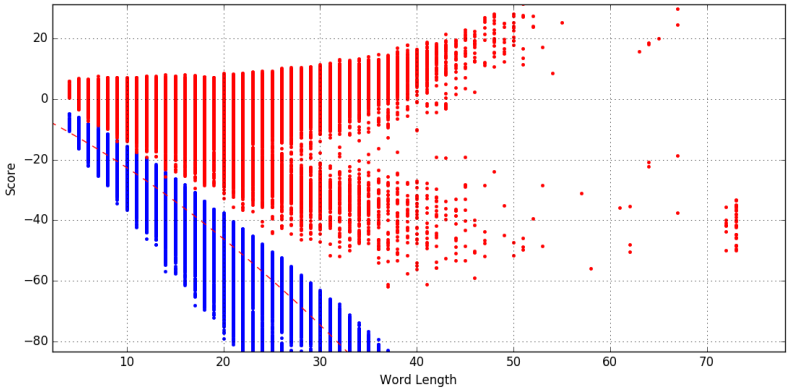
Use Case #1: Domain Generation Algorithms

Domain Name	Score
3lqjnuhra3xf585jgtkhk71exuhu6yrkna.com	-40.0964236335
t8jm1r9gza2a9495g94lgm3a2j3c.com	-143.407177799
wwwwwwwwwwwwwwwww.net	-40.4781346082
Cc59d96f87c7690f498f9945fcd946897786fd21.com	-49.5016279624

- Many types of malware use domain generation algorithms as a means of generating command and control domains to stay ahead of security blacklisting
- These algorithmically generated domains are detectable by measuring the character string entropy



Word Length VS Entropy Score Correction



Explainablility

- Behavioral detectors are in a sense just advanced features for a machine learning model
  - The behaviors are modeled because they correlate with malicious activity but the correlations should be understood
- If the results aren't understood, action typically wont happen