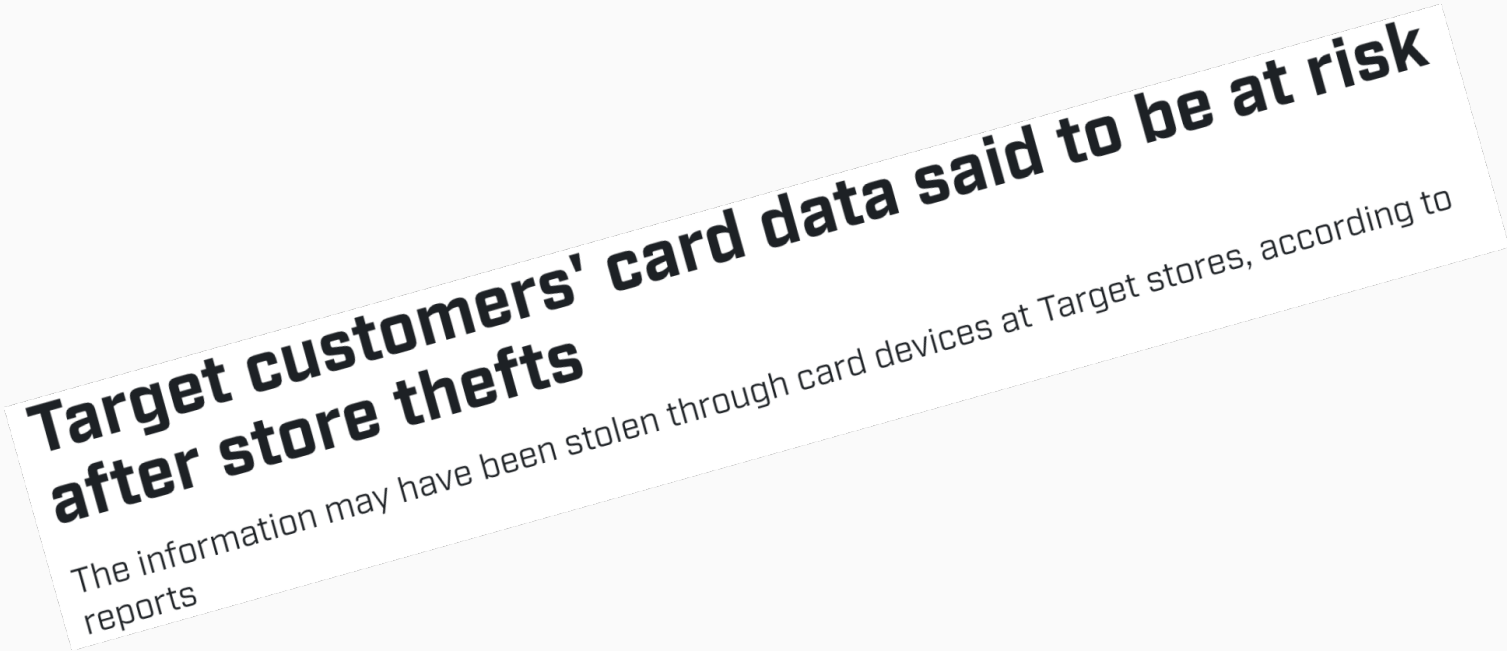
Risk Assessment of A Corporate Network

using Bayesian Networks

Cyber Warriors
Elizabeth Archer, Patrick Seise, Matthew Risley

December 2018

Cybersecurity in the News

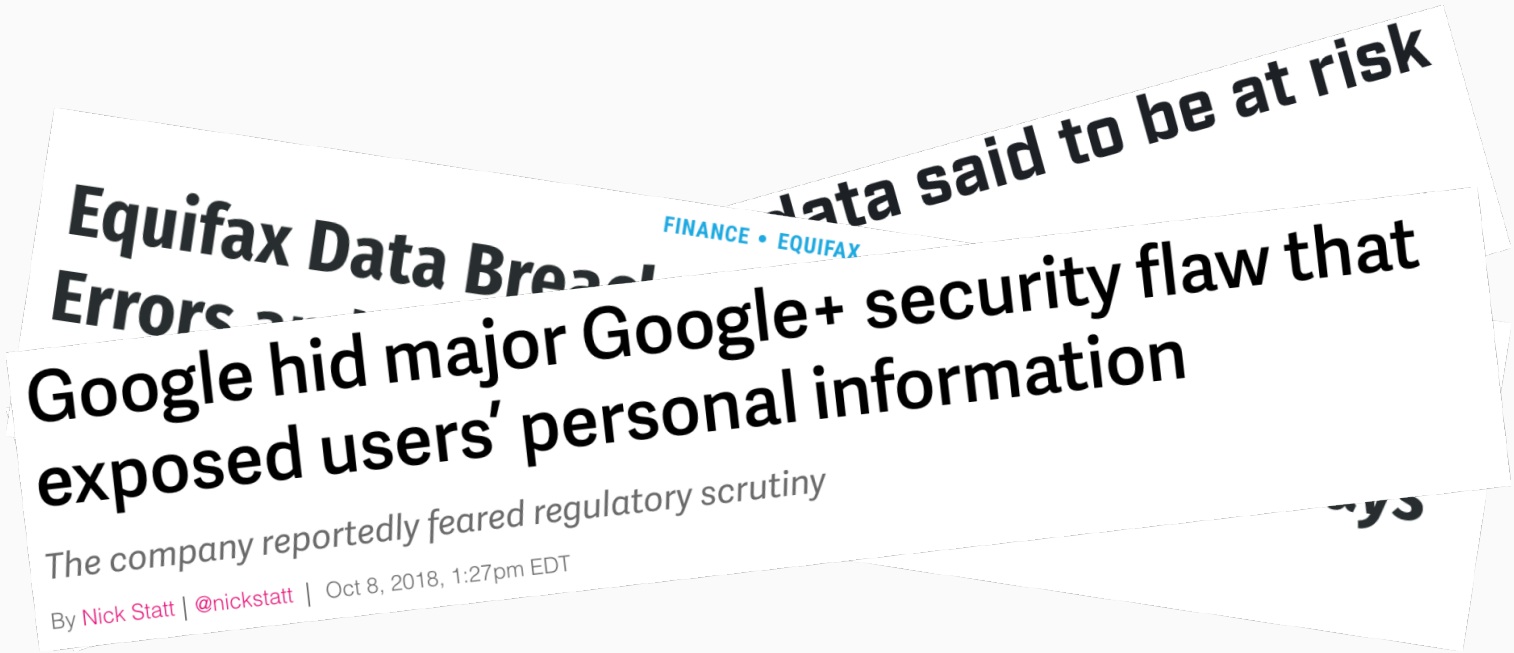


Target customers' card data said to be at risk after store thefts

The information may have been stolen through card devices at Target stores, according to reports

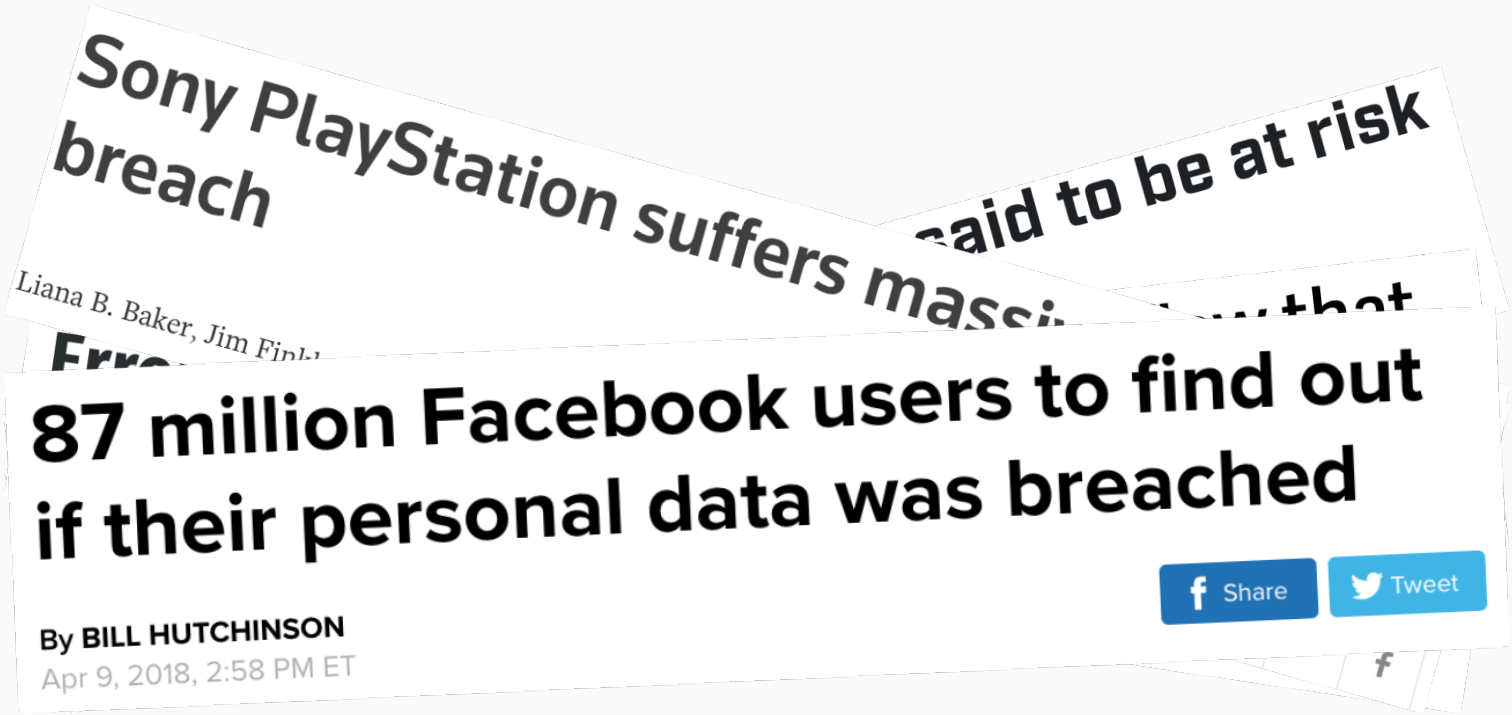
Cybersecurity in the News

Cybersecurity in the News

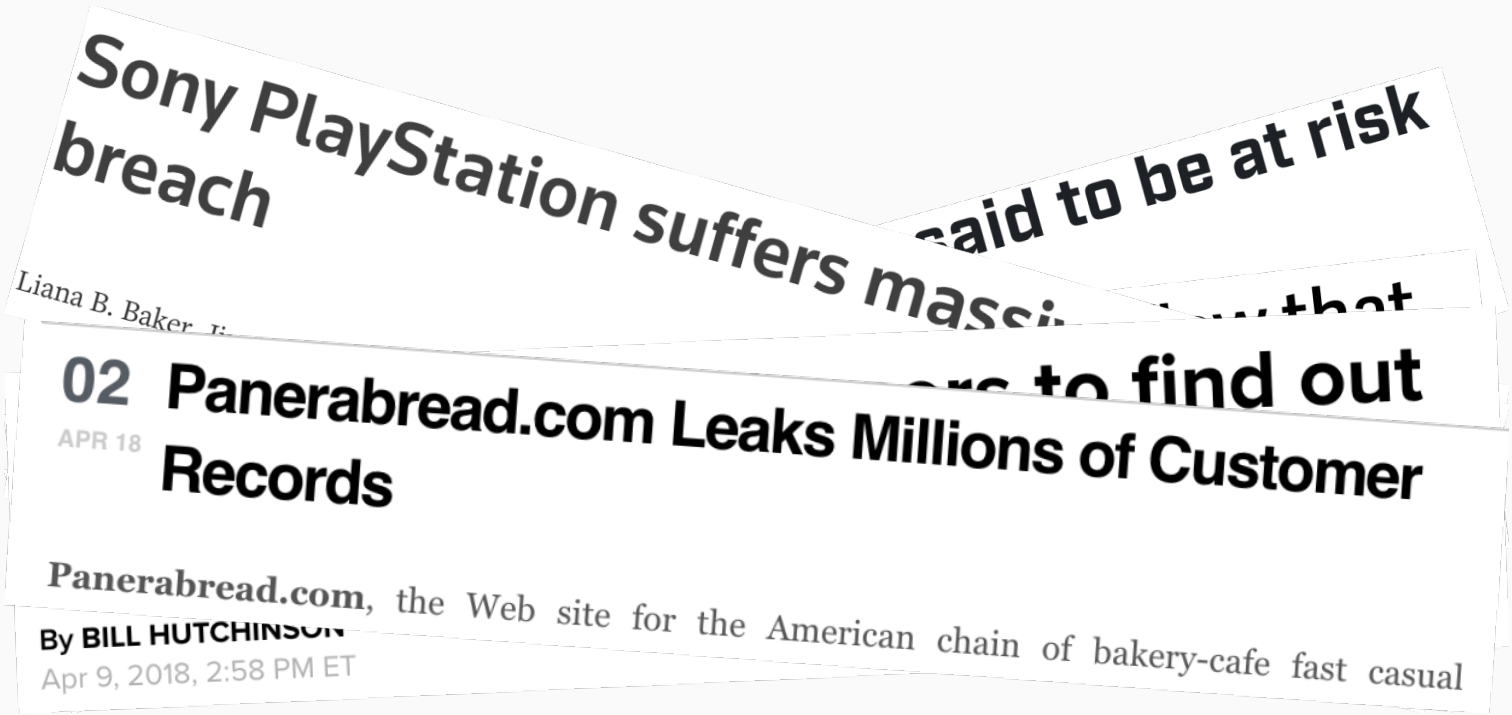


Cybersecurity in the News

Cybersecurity in the News



Cybersecurity in the News



Cybersecurity in the News

Cybersecurity in the News

Cybersecurity in the News

Introduction

Client :

GENERAL DYNAMICS
Mission Systems

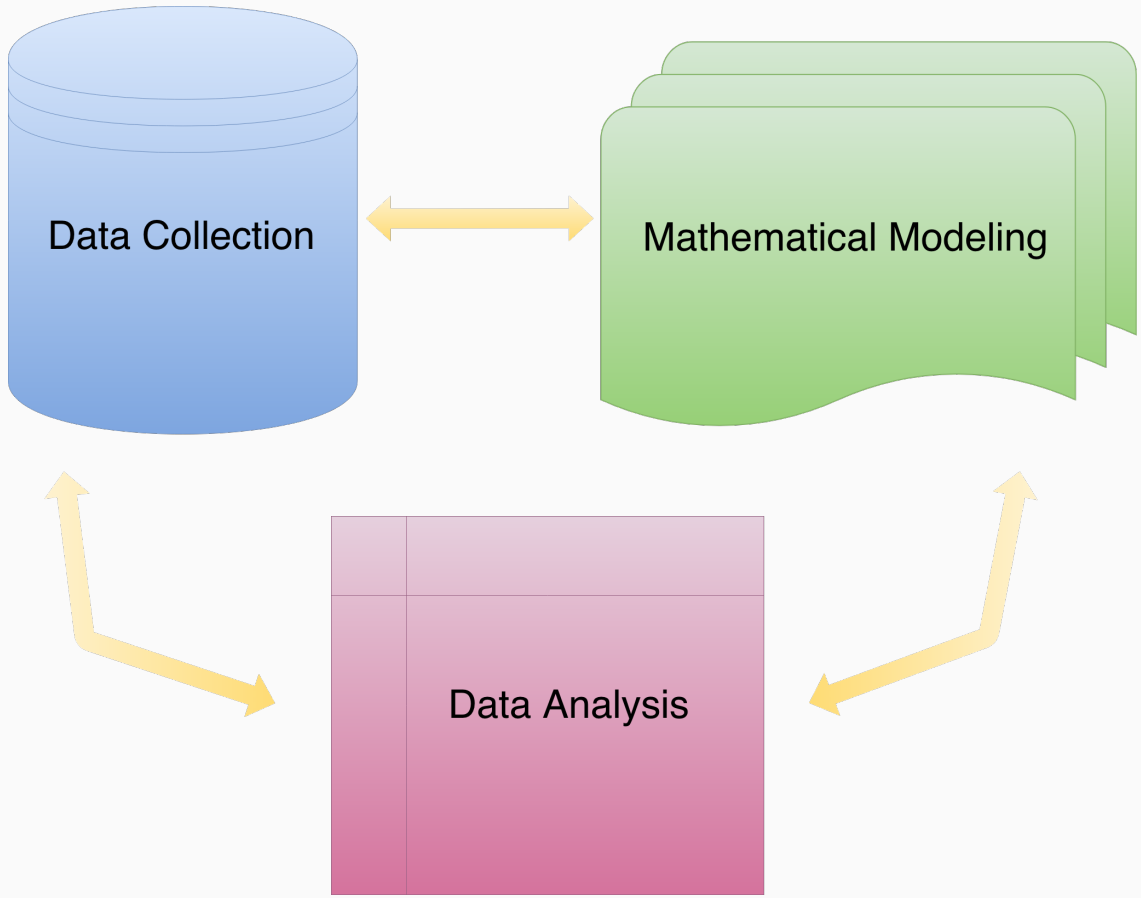
Problem Statement:

How do we secure our data?
How do we prevent attackers from accessing our systems?
How do we assess the risk of a network through associated vulnerabilities contained within a network?

Background

- **Corporate Network** A series of machines (computers, routers, servers, etc.) that a corporation uses
- **Attacker** A malicious player with intentions to gain information or data from a network
- **Compromise** The state of an attacker gaining unauthorized access to a machine
- **Privilege Level** Controls the access to resources (i.e. memory regions, ports)
- **Privilege Escalation** The change from one privilege level to a more privileged level (i.e. user, admin, root)

Components



Survey Of Solution Methods

1. Network of Servers
   - Provide an assessment of how to better organize a network to prevent further privilege escalation.
2. Network of Users
   - Access the shortest path an attack would take based on vulnerabilities available to gain access to a user with the highest level of privilege.
3. Markov Chains
   - Represent an infection model that would provide insight to a attackers infiltration latency.
4. Regression Analysis
   - Require large amounts of data to attribute to specific factors such as user access level, number of users, etc.

Data Collection

Common Vulnerabilities and Exposures Database

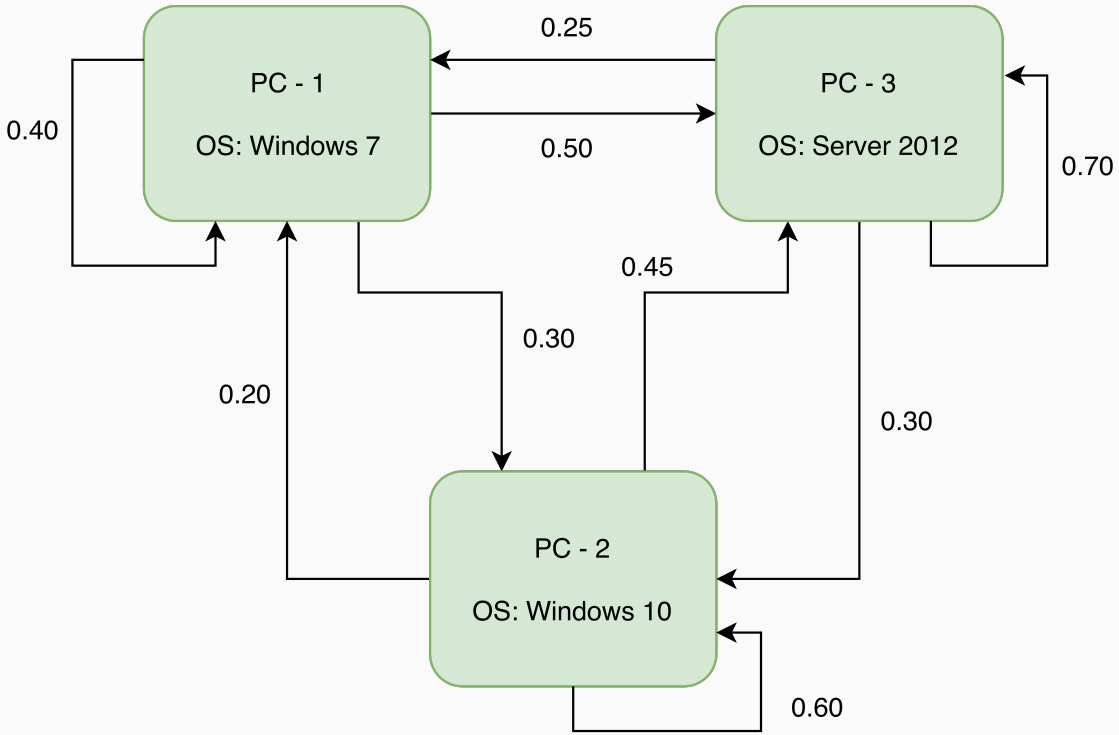
Common Vulnerabilities Scoring System

Data Collection

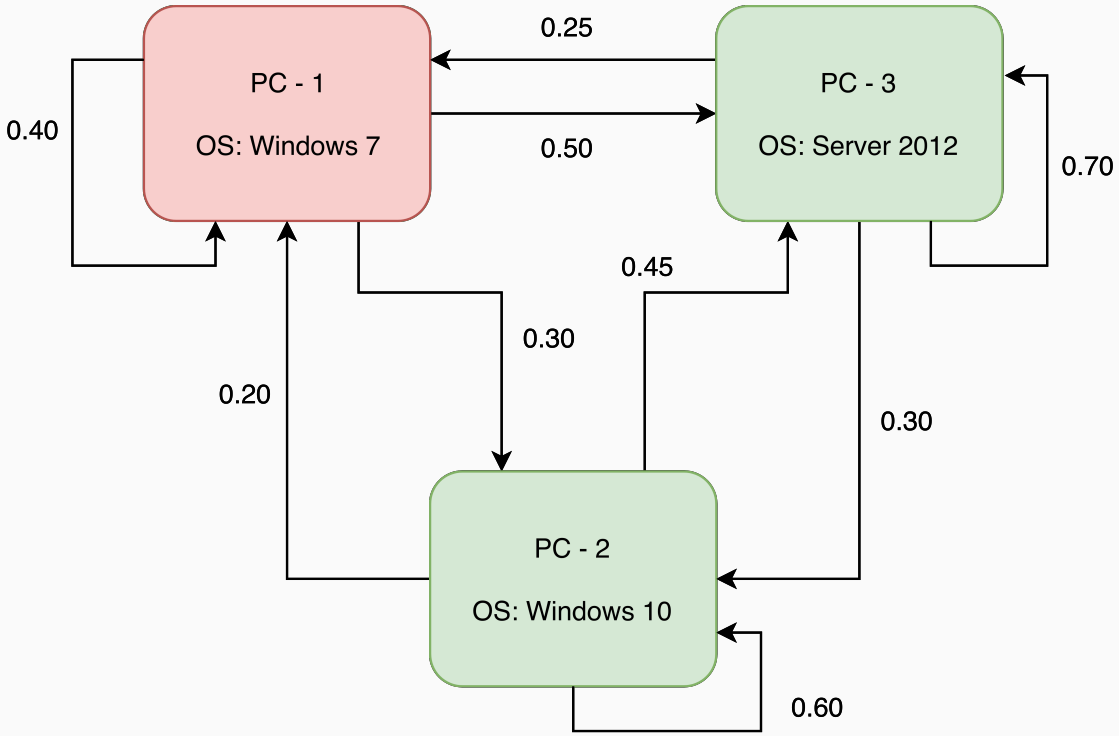
Vulnerability Type(s)	CVSS Score	Access	Complexity	Authentication	Confidentiality	Integrity	Availability
DoS +Priv	7.2	Local	Low	Not Required	Complete	Complete	Complete
Priv	4.4	Local	Medium	Not Required	Partial	Partial	Partial
Priv	7.2	Local	Low	Not Required	Complete	Complete	Complete
Priv	9.3	Remote	Medium	Not Required	Complete	Complete	Complete
Info	1.9	Remote	Medium	Not Required	Partial	None	None

Table: Windows 10 Vulnerabilities

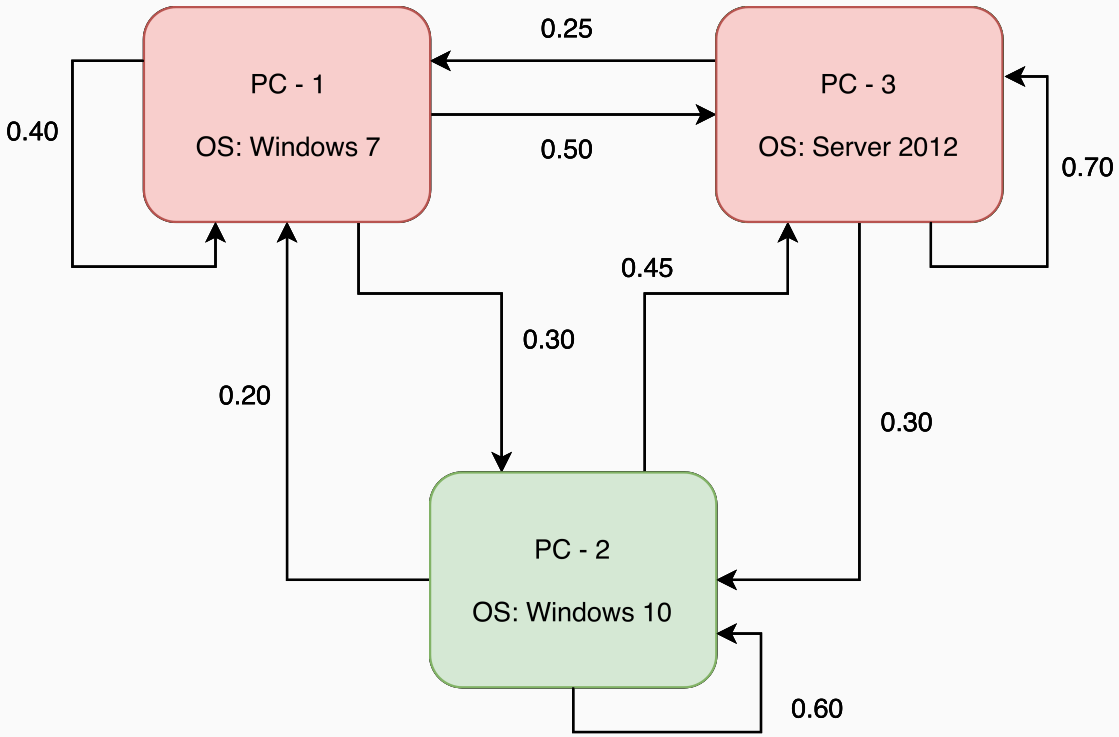
Preliminary Approach - Markov Chains



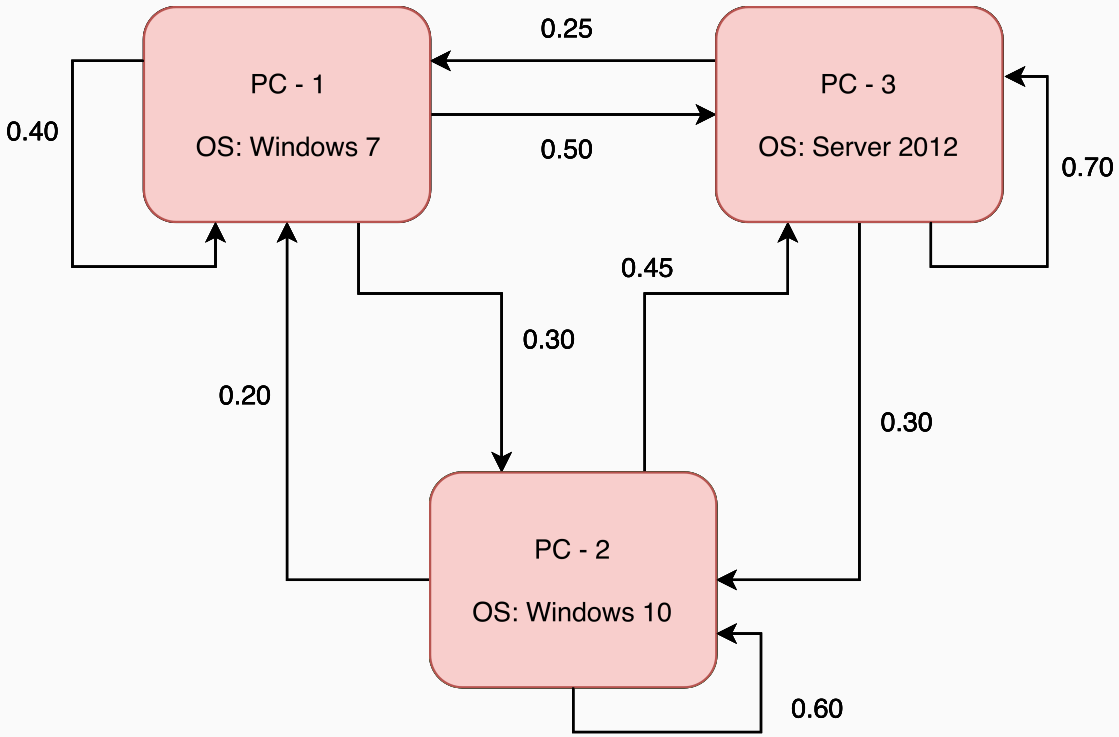
Preliminary Approach - Markov Chains



Preliminary Approach - Markov Chains



Preliminary Approach - Markov Chains



Limitations

- Not Realistic
  - Assume that an attacker would only be on one node at a time
  - Each node would represent a "state" in the Markov chain
  - Attackers should be able to access multiple nodes at once.

- Data Availability
  - Difficult to utilize data of a compromised corporate network
  - Initial factors proposed by our client wouldn't be feasible without such available data.
  - No open source data pertaining to specific privilege attacks (i.e. active accounts, usage levels, number of hosts, privilege available, network structure, etc.)

Literature Review

1. Exiting the Risk Assemement Maze: A Meta Survey (2018)
   Authors: D. Gritzalis, G. Iseppi, et. al
   - Various model approaches to simulate lateral movement within a network
2. Biologically Inspired Risk Assessment in Cyber Security using Neural Networks (2014)
   Authors: M. Ionit and V. Patriciu
   - The effectiveness of neural networks in assessing risk (number of factors, accuracy, etc)
3. Risk Analysis and Probabilistic Survivability Assessment (RAPSA): An Assessment Approach for Power Substation Hardening (2014)
   Authors: C. Taylor, A. Krings, and J. Alves-Foss
   - Methods using more quantitative information and corresponding accuracy

Challenges

Data, Data, Data, $\cdots$

- What data is available to support purposed models?
  - Corporations are reluctant to provide proprietary data about own network let alone security breaches.
- Which models support limited amounts of data but provide valuable analysis?
  - Bayesian networks requires limited probabilistic terms representing connections between nodes but need an overall network layout to be efficient.
- Are there ways to generate realistic data given current availability limitations?
  - Monte Carlo randomization suggests a reasonable approach.

Literature Review

Network Vulnerability Assessment using Bayesian Networks
Authors: Yu Liu & Hong Man

- Bayesian Network Approach
- More compact representation of attack paths than conventional methods.
- Utilized a Maximum Probability Explanation algorithm to compute an optimal subset of attack paths relative to prior knowledge on attackers and attack mechanisms.

Mathematical Modeling

- Assumptions
  - Each node in the network represents a privilege level of a machine on a network (computer, router, server, etc.).
  - Each edge of the network represents the probability that an attacker can escalate their privilege on that machine or move to another machine with some privilege level on the network.
  - Probabilities correspond to CVE data of known vulnerabilities related to privilege escalation.

Mathematical Modeling

Given a network with k - nodes we define the state of the network with a vector

$$\underline{\mathbf{N}} = \left[X_1 \cdots X_k\right]^T$$

Each node within the network will be represented as x_i where $x_i \sim \text{Bern}(P_i)$

$$x_i \ \epsilon \ \{0, 1\}, \text{ where } \begin{cases} 0 = \text{not compromised} \\ 1 = \text{compromised} \end{cases}$$

At each stage [iteration], each node has the probability p_i of being compromised:

$$P(x_i = 1) = p_i$$
$$P(x_i = 0) = 1 - p_i$$

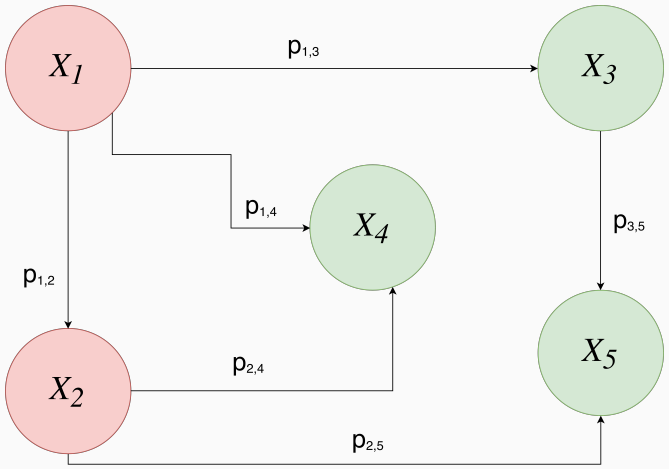
Mathematical Modeling

Each corresponding parent node will be denoted as Pnt_i, from this we can assume that the probability that a node will be compromised in the $n+1$ state as

$$P(x_i^{(n+1)} = 1 \mid \text{Pnt}_i^{(n)}) = 1 - \prod_j (1 - P(x_i^{(n+1)} = 1 \mid x_j^{(n)}))$$
$$\text{where } j \ \epsilon \ \text{Pnt}_i$$

Mathematical Modeling - Example

Based off of the example network we build the following matrix:



$$M = \begin{bmatrix} 1 & p_{12} & p_{13} & p_{14} & 0 \\ 0 & 1 & 0 & p_{24} & p_{25} \\ 0 & 0 & 1 & p_{34} & 0 \\ 0 & 0 & 0 & 1 & p_{35} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

We calculate p_i for the $n+1$ state:

$$p_1 = 1$$
$$p_2 = 1 - (1 - p_{12})$$
$$p_4 = 1 - (1 - p_{14})(1 - p_{23})(1 - p_{34})$$

Mathematical Modeling

- Privilege escalation attacks imply a progressive movement through a network (i.e. node x_j will never point to node x_i).
  - Generated network will always yield an upper triangular matrix

- Given that the CVE scores are stored as a difficulty level they are converted into probabilistic terms .

$$p_{ij} = 1 - (\text{score} \cdot 0.10)$$

where i and j represent the position in the network matrix

- For simulations, we performed Monte Carlo sampling using the p_i values. Once calculated, random sample to determine if node x_i is compromised at the subsequent time stage.

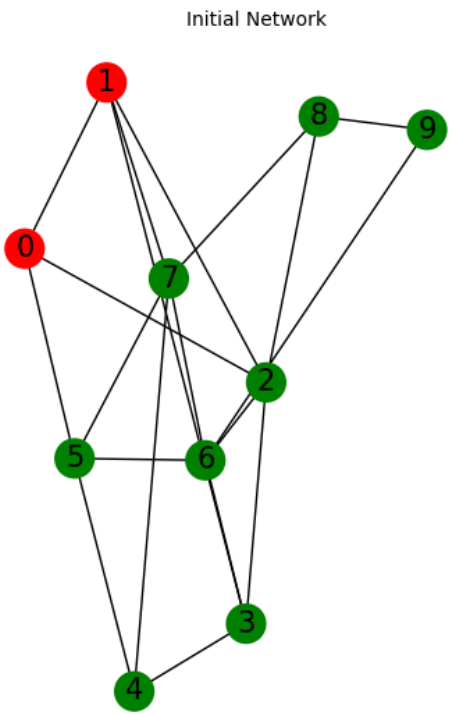
Implementation

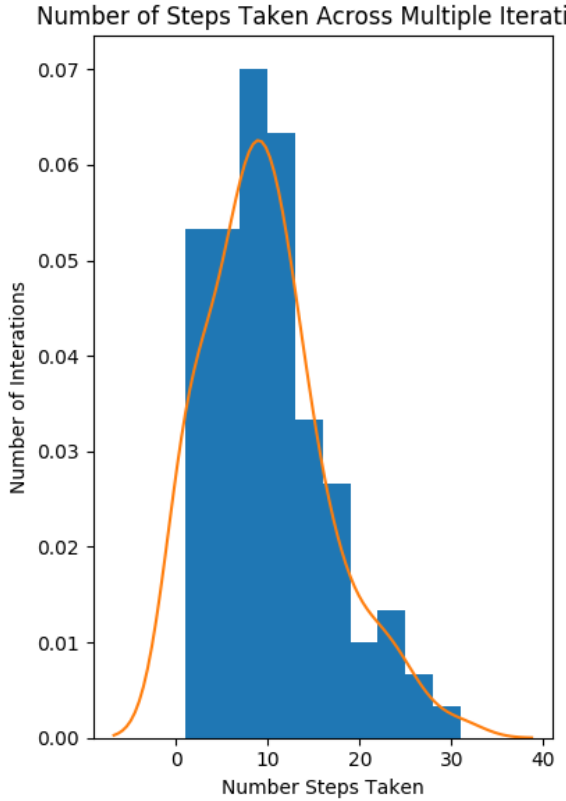
- Language: Python
- Library: `pandas`, `multiprocessing`, `networkx`, `numpy`
- User defined network size and number of iterations
- Simulation
  - Randomization and sparsity prioritized due to limitations of data.
  - Integrated stopping criteria indicated the highest level of access has been breached
  - Initial conditions are set
    $\underline{\mathbf{N}} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
  - Stopping criteria
    $\underline{\mathbf{watch}} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}$

Algorithm

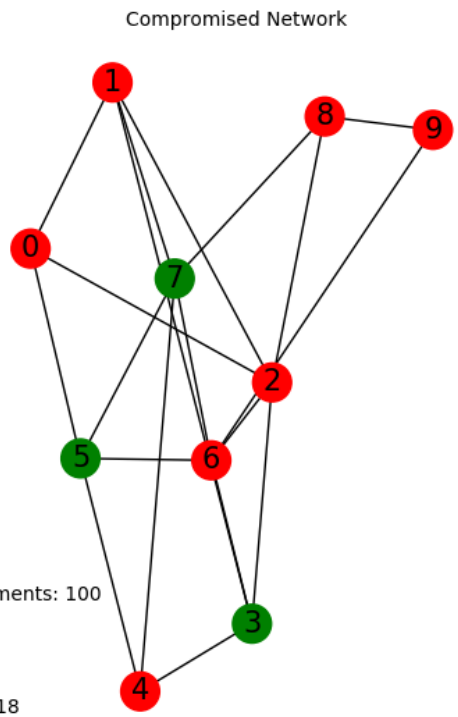
Data: *M* & **N** - Matrix & Initial State Vector
Result: **steps** - num of steps taken to become compromised
for *number of iterations* ▷ *Done in parallel* **do**
  while *N not Compromised check* **watch** **do**
    p = CalculateProb(M, N) ▷ p prob. of next state compromised
    N = randomDraw(N) ▷ 0 or 1s if indicators in p were compromised
    update count ▷ count used for values in results vector
  end
end

Small Network Simulation





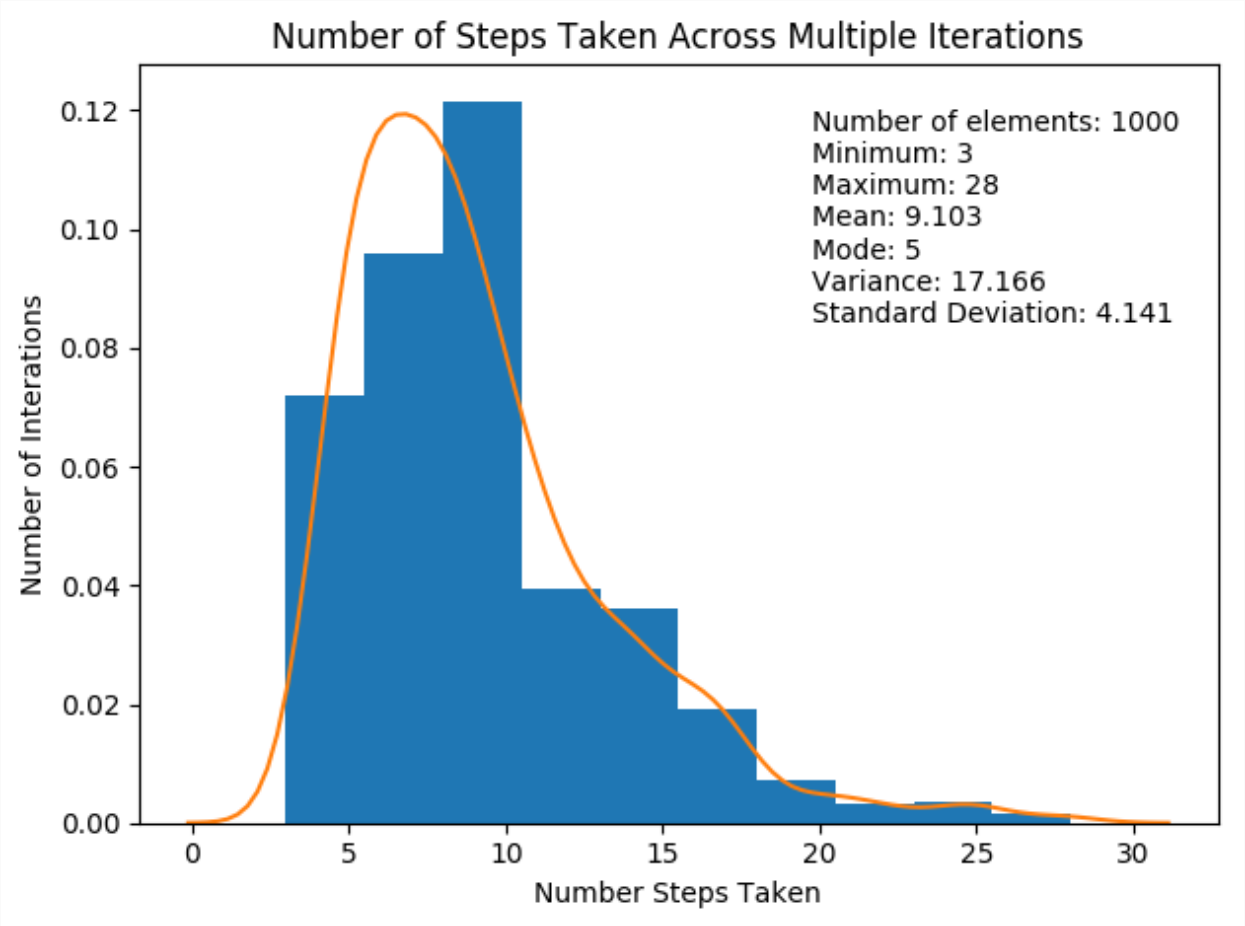
Number of elements: 100
Minimum: 1
Maximum: 31
Mean: 9.780
Mode: 1
Variance: 42.618
Standard Deviation: 6.496



Data Analysis

- Understanding the Network
  - Each node could represent a computer, router, server, etc.
  - We assumed that an attacker was already on the network
  - Given each simulation, how many steps would it take to achieve a compromised state.
  - The 8^{th} and 9^{th} nodes might represent the highest level privilege on the network.

Scaled Up


Number of Steps Taken Across Multiple Iterations
Number of elements: 1000
Minimum: 3
Maximum: 28
Mean: 9.103
Mode: 5
Variance: 17.166
Standard Deviation: 4.141
Number of Interations
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
5
10
15
20
25
30
Number Steps Taken

Model Limitations

- Mathematical Understanding
  - Utilizing Bayesian Statistics
  - Complexity of larger networks and calculating probabilities of n^{th} step
- Computational Complexities
  - Large network computation impractical on personal machine
  - Distribution across iterations
- Network Design
  - Sampling problem with too much randomization
  - Space filling designed after traditional network structure.

Future Work

- Create a corporate network that more accurately represents the clients network
- Factor in other security measures that impact security such as firewalls, and password requirements
- Identify path an attacker would most likely take to fully compromise a network
- Optimize the efficiency of the algorithm

Lessons Learned

1. Data can be really hard to find
2. There's always better models out there but what supports our data.
3. Utilizing GitHub allowed for collaboration when developing sub-components of code.
4. Complexities of performing analysis when it comes to Cyber Security.

Conclusion

Bayesian networks utilized Monte Carlo simulations to get realistic results.

Analyzing average number of steps taken to compromise a corporate network could inform security professionals how quick they would need to react.

In using our model corporations can assess the risks associated with their network and use that information to better secure their data!