LATERAL MOVEMENT DETECTION

Leveraging data in the cybersecurity industry

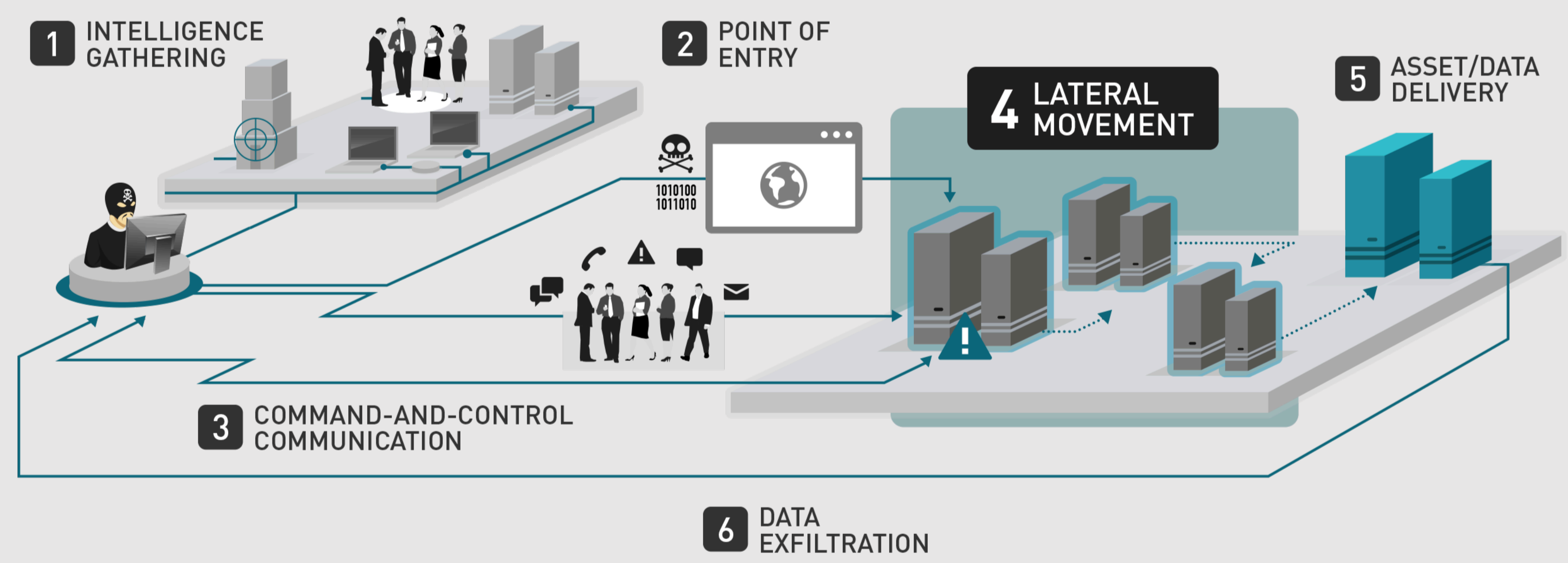
AT A GLANCE

The Team

MBAn | Rapid7 | Faculty Advisor

Raphaelle Delpont, Gabrielle Rappaport (MBAn) · Roy Hodgman, Vasudha Shivamoggi, Katie Wilbur (Rapid7) · Rahul Mazumder (Faculty Advisor)

Capstone Company: Rapid 7 **Location:** Boston, MA



Problem Statement

Develop and implement an algorithm that detects lateral movement attacks within network data and generates alerts when unexpected behaviour is detected.

DATASET

Rapid7 has deployed sensors capable of gathering network communication. We used this data to conduct our lateral movement detection analysis.

Relevant features:
Source asset | Destination asset | Communication timestamp | Protocol

Data statistics:
- 20,000 SSH internal communications in 4 months in Rapid7 Boston office
- 100% unlabelled without prior examples of intrusions

IMPACT

1. DIRECT LABOR SAVINGS

+$1M

Impact to Security Analysts:
Created machine learning models to classify 99% of the data as "normal", significantly reducing manual review of client network data at a projected cost of $1M+

2. AVOIDANCE SAVINGS

+$36M

Impact to Clients:
By laying the foundation for modern machine learning in cybersecurity and driving initial findings, Rapid7 and their clients can avoid costs of at least $36M+ annually

3. NOVEL ML TOOL

Patentable research

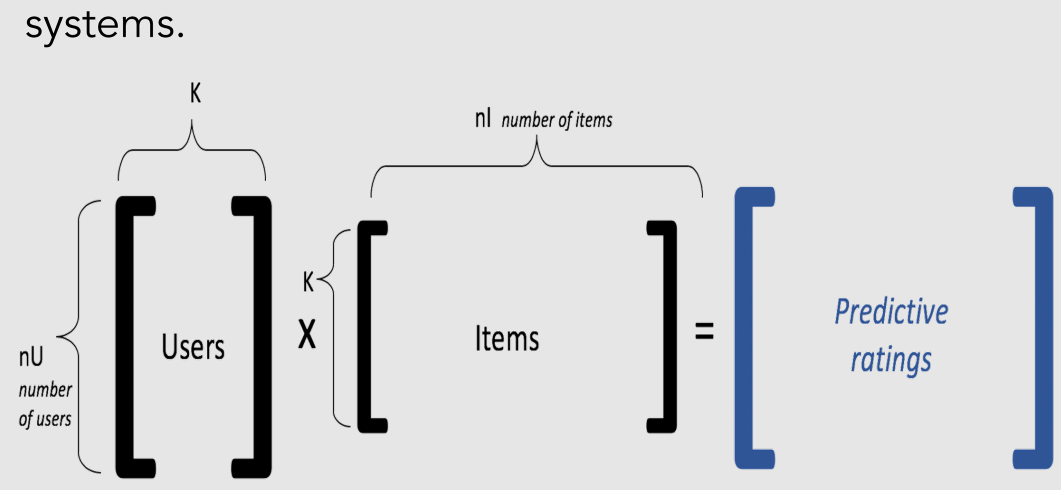
Impact to Rapid7:
Packaged flexible, scalable, online and auto-tuned machine learning pipeline to be used on network communication dataset to detect lateral movement

THREE STEPS ALGORITHM

1 Scoring each connection in the network to flag anomalies

MATRIX FACTORIZATION [1]

Learns the communication habits between assets based on source-destination pairs as in recommender systems.

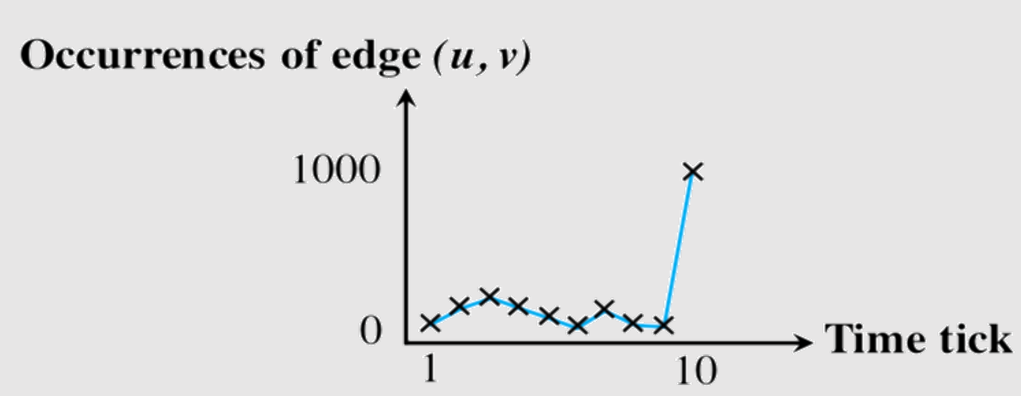


Learn and update the behaviours of network assets over time.

The online nature of the algorithm makes it scalable and doesn't require data storage.

MIDAS [2]

Detects micro-cluster anomalies within the network connections, or suddenly arriving groups of suspiciously similar edges.



We rely on the hypothesis that the average number of connections between two assets stays stable over time.

Implementation of Count Min Sketch Data Structures
Computing anomaly score based on Chi-squared statistics:

$$\chi^2 = \left(a_{uv} - \frac{s_{uv}}{t}\right)^2 * \frac{t^2}{s_{uv}(t-1)}$$

s_{uv} : the total number of edges from u to v up to the current time
a_{uv} : the number of edges from u to v in the current time tick

Detects local bursts of activity in the network and temporal anomalies
Online algorithm enabling scalability

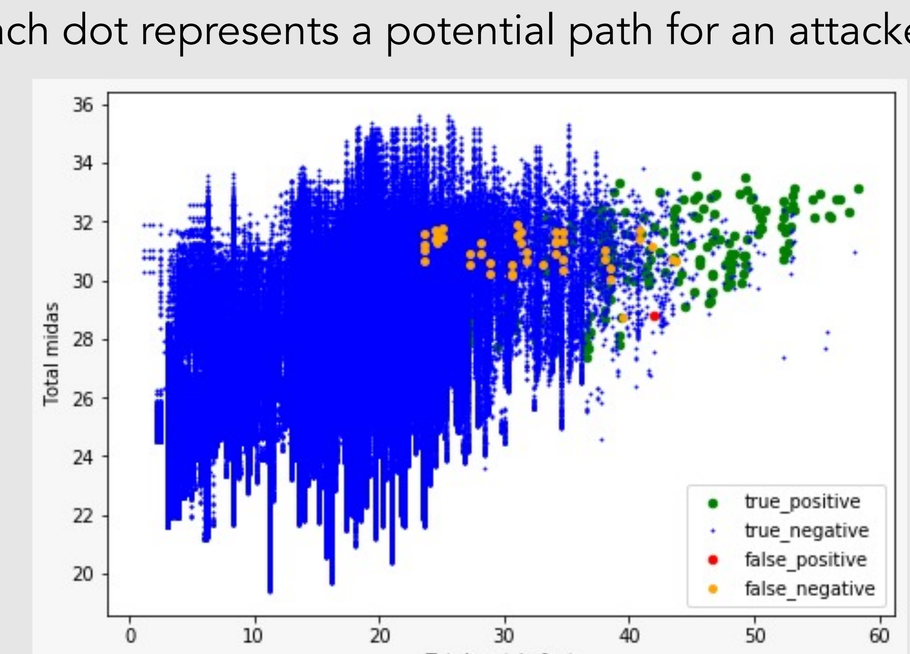
2 Joining communications together to simulate the attacker's potential paths in the network

- Connections chronologically ordered
- Paths are constituted of unique assets

From 20,000 connections to 1,000,000 paths

Exacerbates consecutive anomalies and helps detect anomalous paths

Each dot represents a potential path for an attacker



3 Flagging the abnormal paths to detect lateral movement attacks

	LSTM	XGBoost	Classification rules based on quantiles
IN SAMPLE F1-SCORE:	0.09	0.76	0.37
OUT SAMPLE F1-SCORE:	0.02	0.28	0.71

Need for a custom-made classification model specific to the problem of lateral movement detection

Generating the alerts from the abnormal paths

We built classes of equivalence to group similar flagged paths together.
We sent alerts to the security team containing all the paths linked to the attack.

[1] *João Vinagre, Alípio Jorge, and João Gama. Fast incremental matrix factorization forrecommendation with positive-only feedback, 07 2014*

[2] Siddharth Bhatia, Bryan Hooi, Minji Yoon, Kijung Shin, and Christos Faloutsos. Midas:Microcluster-based detector of anomalies in edge streams, 2019.