

---

# Detecting Lateral Movement

## A Data Analysis & Visualization Approach

**Maddy Longo, Jeff Martin, & Ian Vossoughi**

Steven Gomez  
MIT LL - 0558

George Heineman  
WPI

Diane Staheli  
MIT LL - 0551





# The Advanced Persistent Threat (APT)

- **APT is a contemporary form of cyberattack**
- **Stealthy, advanced infiltration to steal valuable data**
- **Critical Stage: Lateral Movement**
  - **Progress from initial compromised computer across network**
  - **Use available resources to steal user credentials and access new computers**
  - **Slowly and steadily increase network reach while remaining camouflaged in network traffic**



# Threat Detection: Process and Problems

- **Automated intrusion detectors learn normal activity**
  - Raise alerts for anomalous activity
- **Automated detectors often have high false positive rates**
  - Faced with many false positives, analysts lose focus
- **Malicious activity represents only a small portion of the total events**
  - Rare but devastating



# Where Are The Malicious Events?

## Los Alamos National Laboratory cybersecurity dataset

	Time	Source User	Destination User	Source Computer	Destination Computer	Auth. Type	Logon Type	Auth. Orientation	Success / Failure
190	769065	U66@DOM1	U66@DOM1	C2707	C2707	Kerberos	Network	LogOn	Success
191	769067	U1048@DOM1	U1048@DOM1	C17693	C2846	NTLM	Network	LogOn	Success
192	769067	U66@DOM1	U66@DOM1	C3430	C3430	Kerberos	Network	LogOn	Success
193	769069	U1048@DOM1	U1048@DOM1	C2846	C2846	?	Network	LogOff	Success
194	769069	U5254@DOM1	U5254@DOM1	C17693	C636	NTLM	Network	LogOn	Success
195	769069	U66@DOM1	U66@DOM1	C2892	C2892	?	Network	LogOff	Success
196	769069	U66@DOM1	U66@DOM1	C2892	C2892	Kerberos	Network	LogOn	Success
197	769069	U66@DOM1	U66@DOM1	C3331	C3331	?	Network	LogOff	Success
198	769069	U66@DOM1	U66@DOM1	C3331	C3331	Kerberos	Network	LogOn	Success
199	769070	U66@DOM1	U66@DOM1	C1823	C1028	Kerberos	Network	LogOn	Success
200	769071	U6146@DOM1	U6146@DOM1	C14053	C612	Kerberos	Network	LogOn	Success
201	769071	U6146@DOM1	U6146@DOM1	C612	C612	?	Network	LogOff	Success
202	769071	U66@DOM1	U66@DOM1	C2331	C2331	?	Network	LogOff	Success
203	769071	U66@DOM1	U66@DOM1	C2331	C2331	Kerberos	Network	LogOn	Success
204	769072	U66@DOM1	U66@DOM1	C3430	C3430	?	Network	LogOff	Success

Navigation bar: whereIsTheMal + Ready

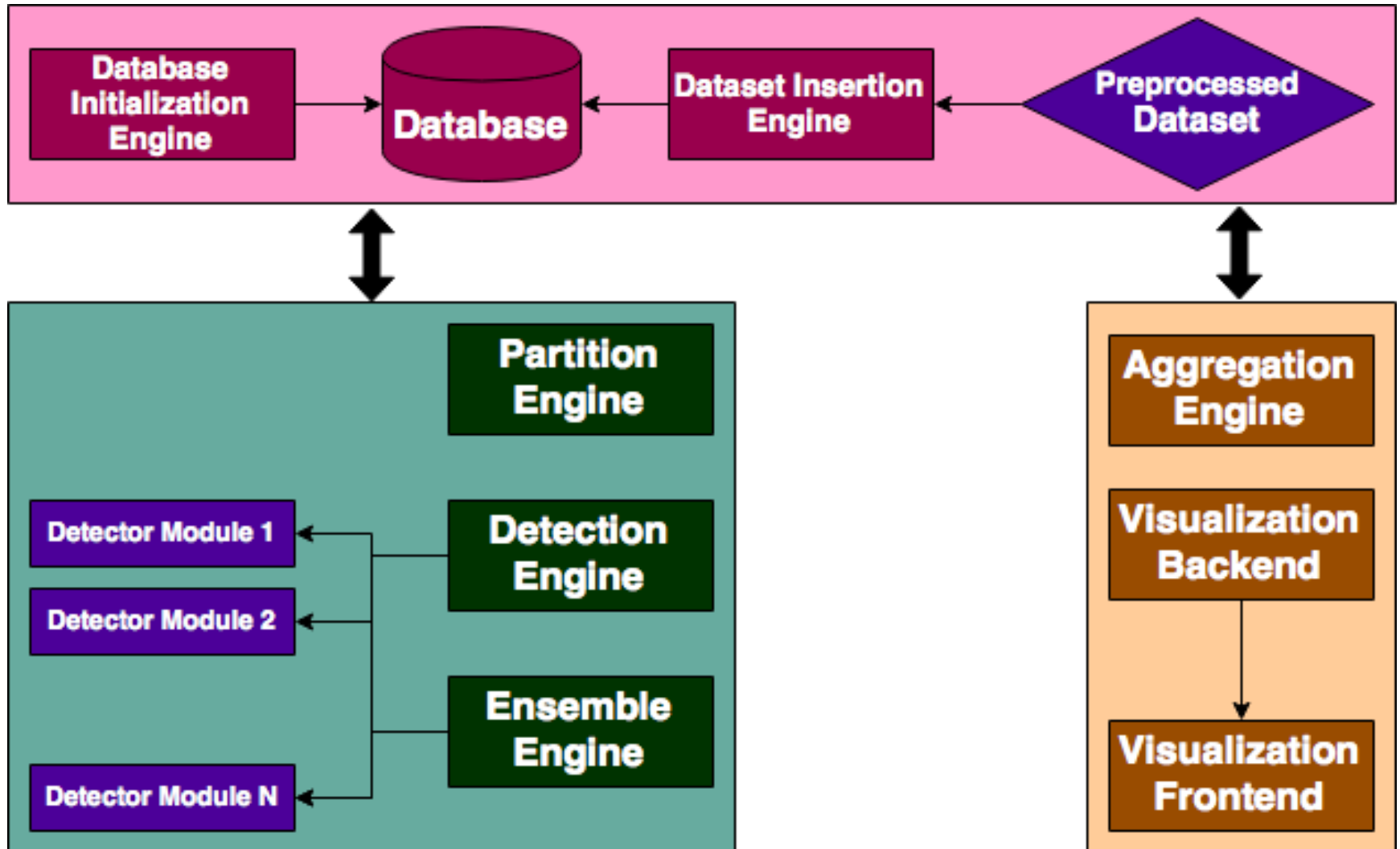


# Hypothesis Driven Solutions

- **Objective**
  - *Explore the use of data analysis and data visualization for the purpose of detecting lateral movement*
  - *Develop a proof-of-concept tool for lateral movement detection*
- **Hypothesis One**
  - Ensemble of anomaly detectors will improve accuracy
- **Hypothesis Two**
  - Visualization that uses event's time, location, and suspicion level will allow an analyst to isolate lateral movement
- **Framework developed as a proof-of-concept tool to evaluate both hypotheses**



# Framework Architecture





# The Analyst Workflow

Cybersecurity Dataset

1

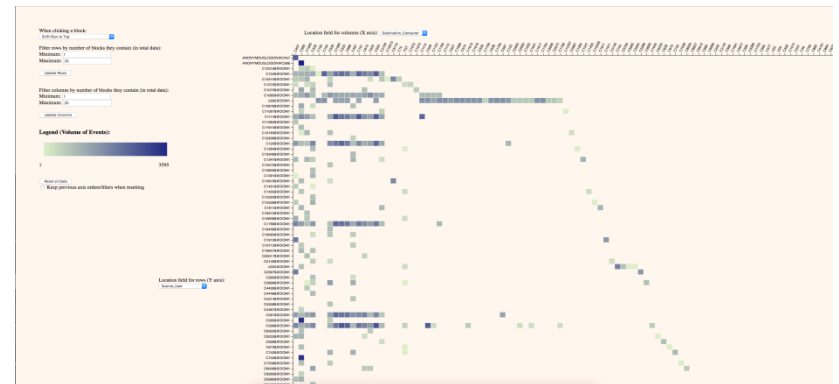
2

3

Shell

```
je27416@490079-mitll: ~/MQP/code/framework
File Edit View Search Terminal Help
root@localhost > workspace -use experiment
root@localhost experiment > status
Status of Workspace 'experiment':
|Clear| (Are all tables in the workspace empty?)
-False
|Events| (EventType [event_count])
-Authentication [161344]
|Partitions| (PartitionType [event_count])
-Test [48403]
-Train [112941]
|Aggregates| (name [total_entries])
-Destination_User x Source_Computer [137091]
-Destination_User x Destination_Computer [127312]
-Source_User x Destination_User [107841]
-Source_User x Destination_Computer [127120]
-Source_User x Source_Computer [136895]
-Source_Computer x Destination_Computer [91090]
|Detectors| (name [prediction_count : AUROC])
-Random_Forest [161344 : 0.785714]
-ImperfectOracle [161344 : 0.798289]
|Ensembles| (name [prediction_count : AUROC : diversity])
-Random_Forest_X_ImperfectOracle [161344 : 0.956776 : -1.0]
root@localhost experiment >
```

Visualization



Cybersecurity Analyst



# Visualization – Overview

When clicking a block:  
Shift Row to Top

Filter rows by number of blocks they contain (in total data):

Minimum: 1  
Maximum: 20

Update Rows

Filter columns by number of blocks they contain (in total data):

Minimum: 1  
Maximum: 20

Update Columns

Legend (Volume of Events):



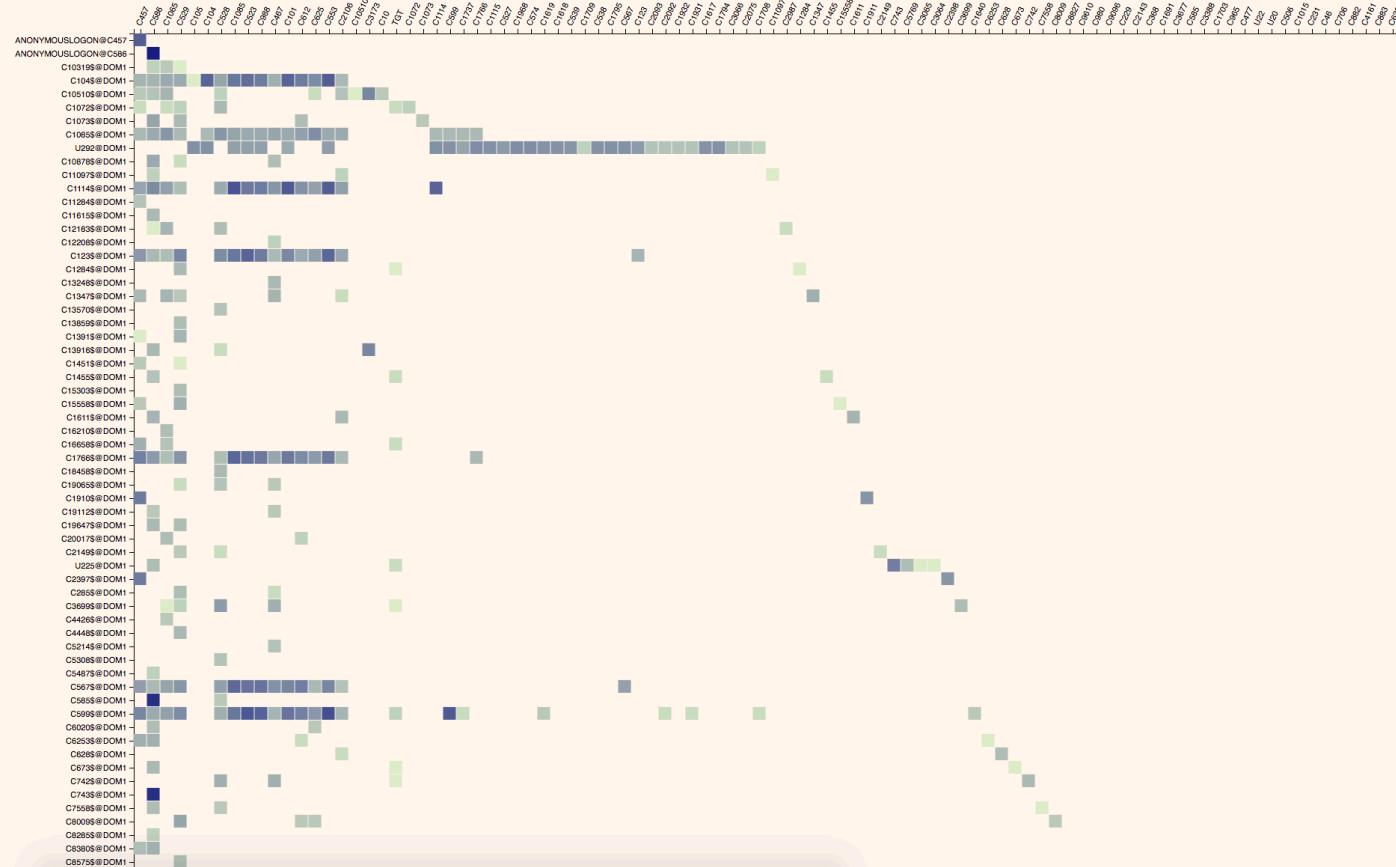
Reset all Data

Keep previous axis orders/filters when resetting

Location field for rows (Y axis):

Source\_User

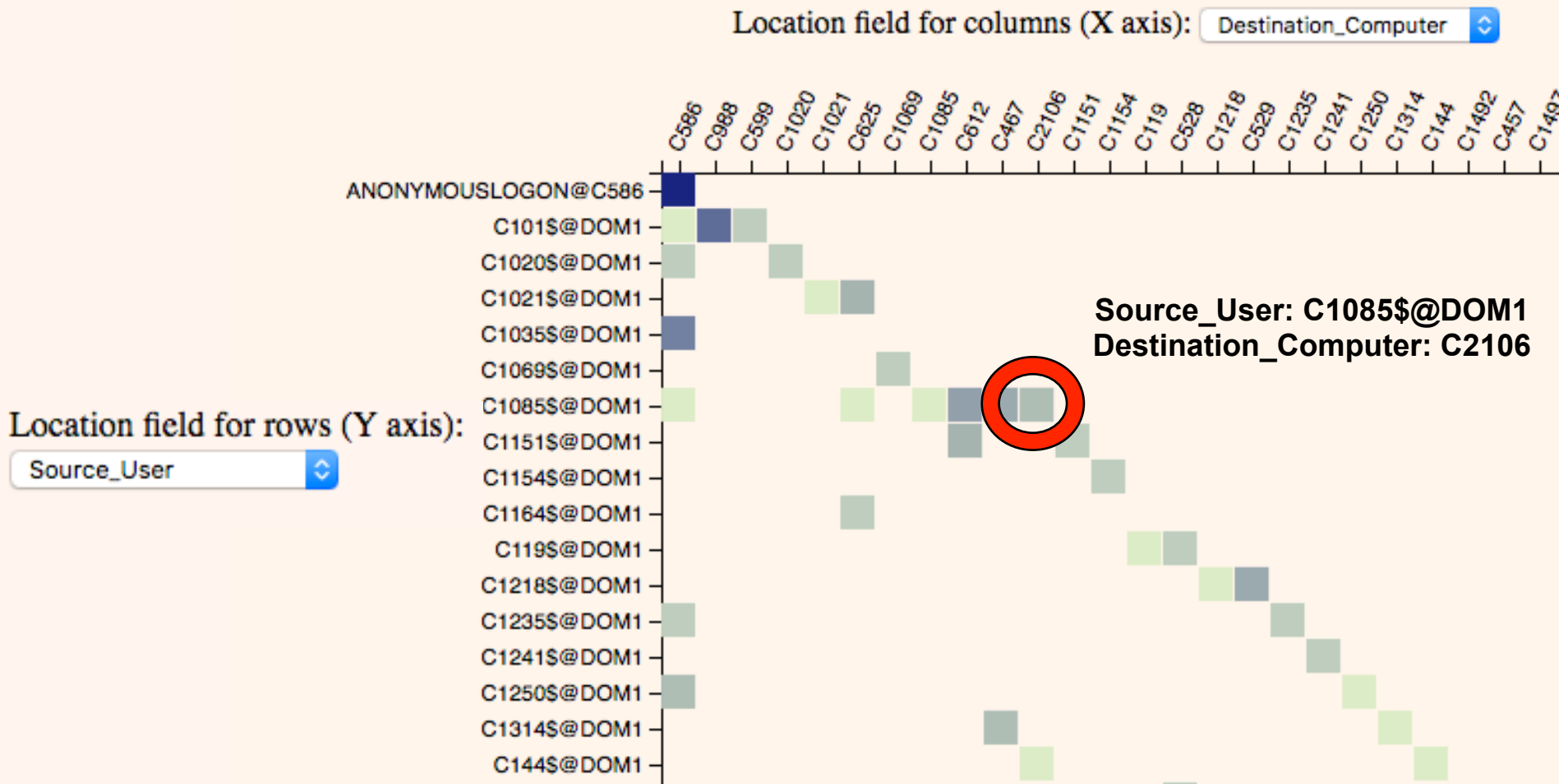
Location field for columns (X axis): Destination\_Computer







# Visualization – Heatmap





# Visualization – Dashboard

When clicking a block:

Isolate Row of Selected Block

Filter rows by number of blocks they contain (in total data):

Minimum:


Maximum:

Filter columns by number of blocks they contain (in total data):

Minimum:

Maximum:

**Legend (Volume of Events):**



1 3595

Keep previous axis orders/filters when resetting



# Visualization – Sorting and Filtering

When clicking a block:

Shift Row to Top

Filter rows by number of blocks they contain (in total data):

Minimum: 10  
Maximum: 20

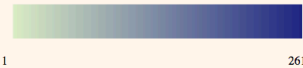
Update Rows

Filter columns by number of blocks they contain (in total data):

Minimum: 1  
Maximum: 5

Update Columns

Legend (Volume of Events):



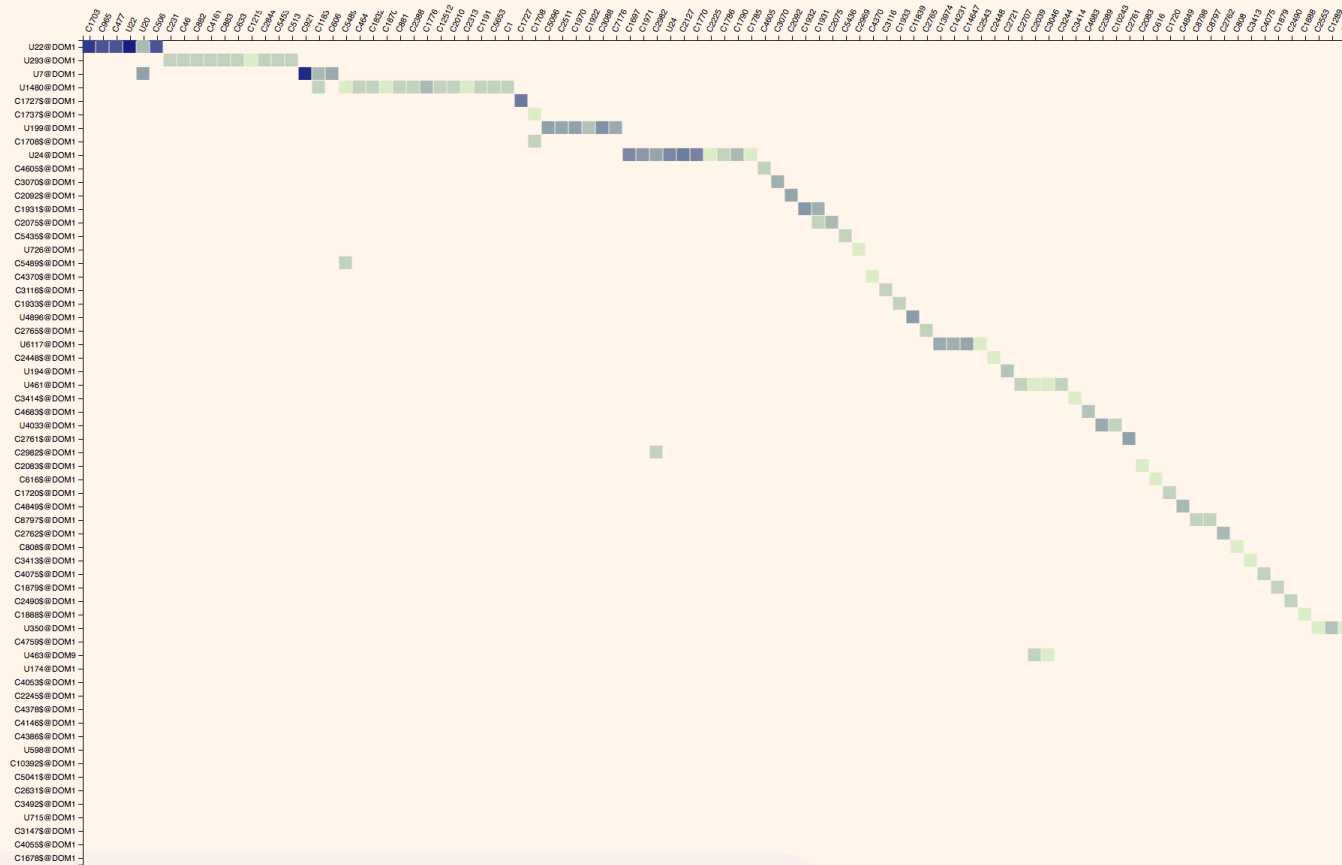
Reset all Data

Keep previous axis orders/filters when resetting

Location field for rows (Y axis):

Source\_User

Location field for columns (X axis): Destination\_Computer





# Visualization – Isolated Timeplot

When clicking a block:

Shift Column to Left

Isolated Row:

**U22@DOM1**

Ensembler/Detector to apply as block gradient:

ImperfectOracle\_X\_Random\_Forest  Reshade Gradients

Raw Data Dump

Legend (Suspicion of Events):

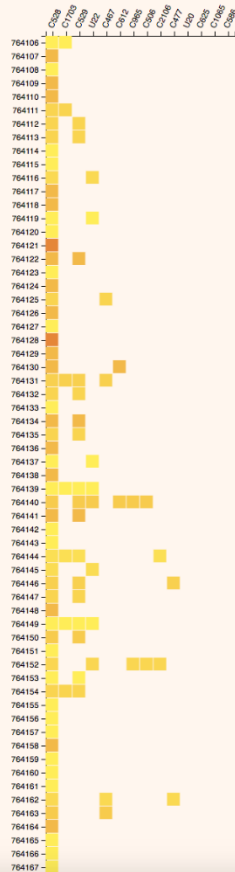


0

1

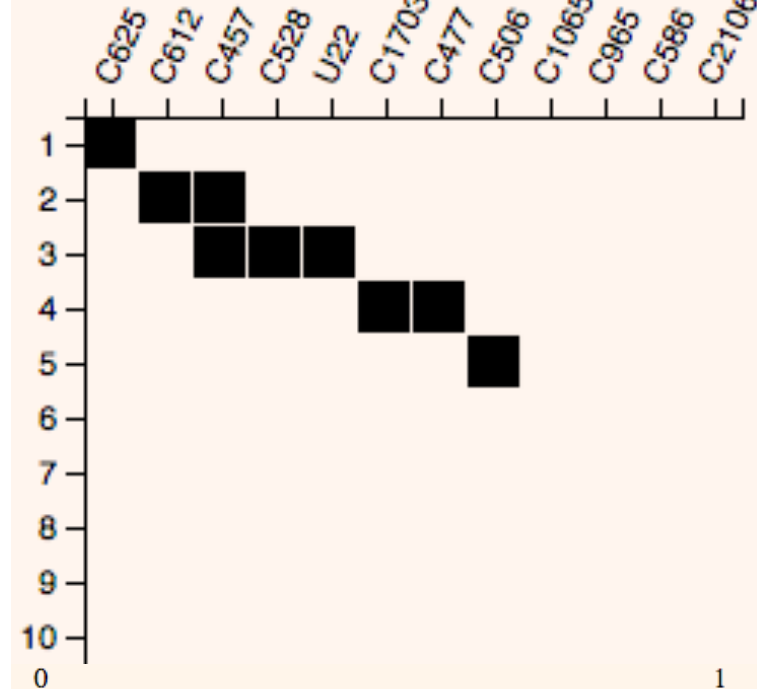
Reset all Data

Keep previous axis orders/filters when resetting



When clicking a block: **Destination Computer**

Shift Column to Left



Time

Reset all Data

Keep previous axis orders/filters when resetting



# Evaluating Hypothesis One - Experiment

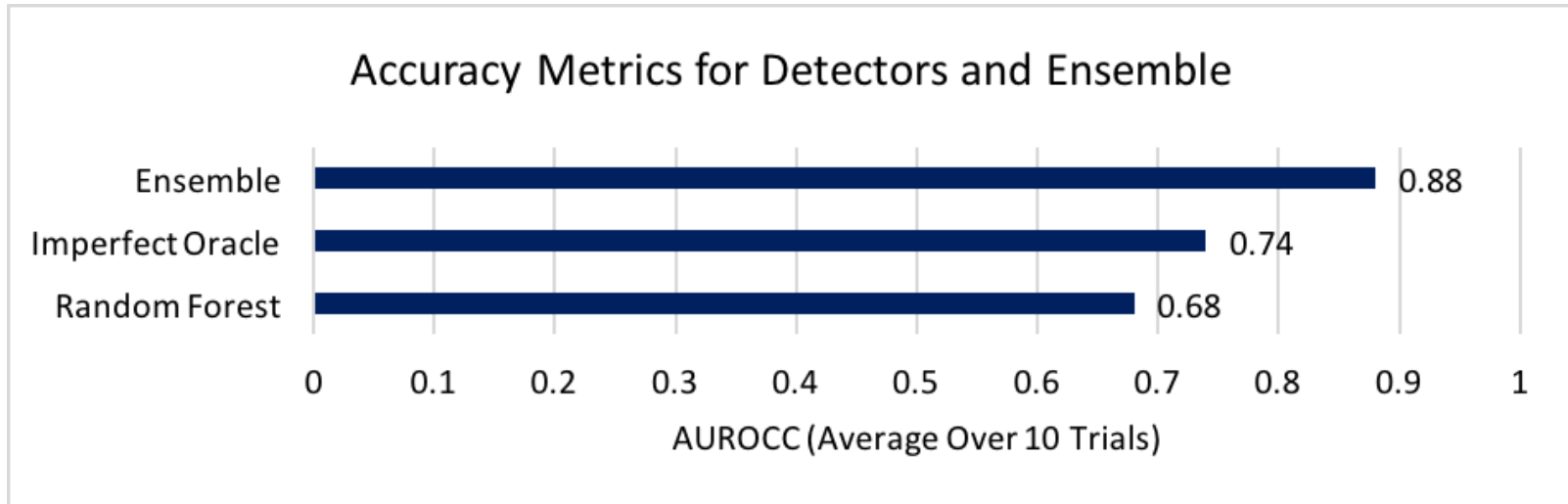
**Hypothesis:** *Ensemble of anomaly detectors will improve accuracy*

- 1) Preprocessed Los Alamos dataset (~1.5B → ~150K events)**
- 2) Implemented machine learning (random forest) and synthetic detectors**
- 3) Trained, tested, and computed performance metrics for each detector and ensemble across 10 trials**



# Evaluating Hypothesis One - Results

**Hypothesis:** *Ensemble of anomaly detectors will improve accuracy*



- **Ensemble AUROCC (Area Under the Receiver Operating Characteristic Curve) was 19% greater than the best detector's AUROCC**
- **This evidence supports Hypothesis One**

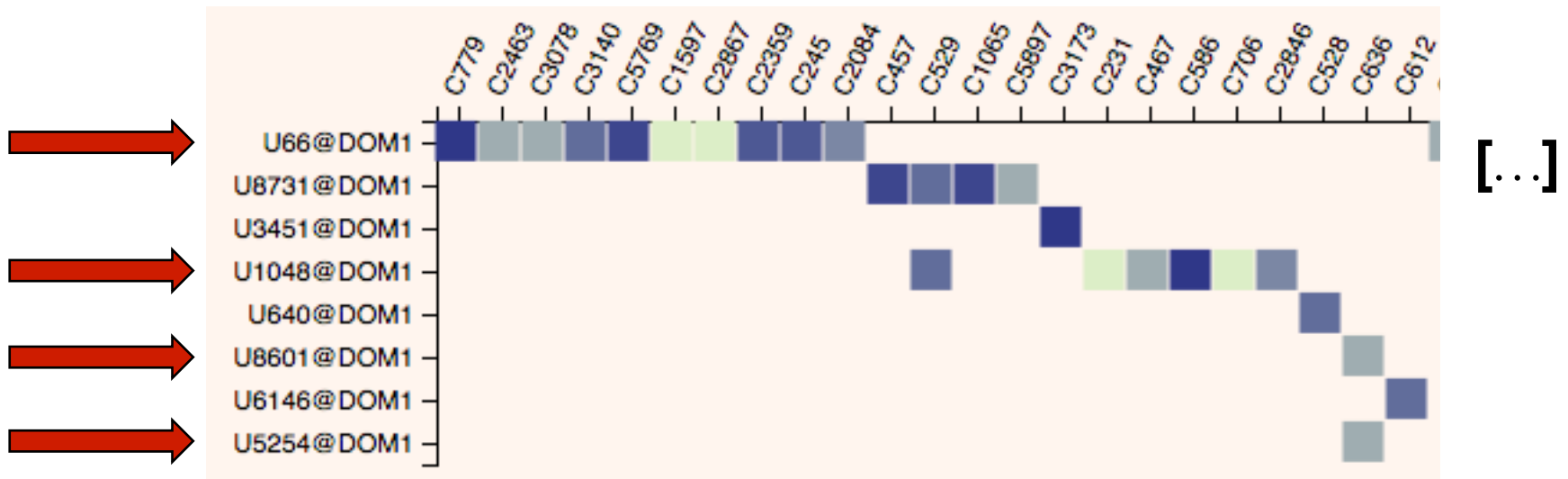


# Evaluating Hypothesis Two - Experiment

**Hypothesis:** *Visualization that uses event's time, location, and suspicion level will allow an analyst to isolate lateral movement*

- **Classification of User Intent**

- Given 100 seconds of authentication data from 8 users, determine which users had redteam activity
- Ground truth: 4 malicious and 4 benign users





# Evaluating Hypothesis Two – Results

**Hypothesis:** *Visualization that uses event's time, location, and suspicion level will allow an analyst to isolate lateral movement*

Username	U66	U5254	U8601	U1048	U8731	U3451	U640	U6146
S1 Predictions	Malicious	Benign	Benign	Benign	Benign	Benign	Benign	Benign
S2 Predictions	Benign	Benign	Benign	Benign	Malicious	Benign	Benign	Benign
S3 Predictions	Benign	Benign	Benign	Malicious	Malicious	Benign	Benign	Benign
S4 Predictions	Malicious	Benign	Benign	Malicious	Malicious	Benign	Benign	Benign
Ground Truth	Malicious	Malicious	Malicious	Malicious	Benign	Benign	Benign	Benign

- Each subject scored within one point of random guessing (4/8)
- This evidence does not support Hypothesis Two





# Evaluating Hypothesis Two – Results

- **MIT LL Information Security Department (ISD) Evaluation and Qualitative Feedback**
  - Visualization shows promise for quickly drawing attention toward anomalous activity
  - Should support detection of more complex anomalies, automated filtering:
    - Logins outside user's normal diurnal hours
    - Authentications into subnets outside user's normal access



# Conclusion

- **Framework**
  - Supports the analysis of a wide variety of cybersecurity datasets
  - Provides a modular interface for implementing and testing anomaly detectors and ensembles of detectors
  - Supports the aggregation and visualization of data in flexible ways
- **Ensemble**
  - Reduces false positive rate while maintaining detection rate
  - Will be more trustworthy to analysts
- **Visualization**
  - Visualization alone is insufficient to pinpoint malicious activity
  - Most useful for guiding the initial steps of the analyst
  - Allows the analyst to identify anomalies to investigate further



# Future Work

- **Improve framework security**
  - Sandbox detector execution
  - Improve authentication process
- **Improve framework efficiency**
  - Optimize SQL queries
  - Improve algorithm efficiency
- **Make detectors more trustworthy**
  - Develop signature-based detectors
  - Display justifications for event suspicion level
- **Add support for multi-layered datasets**
  - With subnets, user roles, computer roles, etc.
- **Evaluate an ensemble of more than two detectors**



# Acknowledgements

## Project Mentors

Steven Gomez  
MIT LL - 0558

George Heineman  
WPI

Diane Staheli  
MIT LL - 0551

## MIT LL Employees

Brian Desnoyers  
(58)

Sandeep  
Pisharody  
(58)

Rob Elkind  
(58)

Chad Meiners  
(51/LRNOC)

Tamara Yu (51/  
LRNOC)

Chris Redinger  
(ISD)

Richard  
Emerson  
(ISD)

Steve  
Castellarin  
(ISD)

Helga Wilde  
(ISD)

Gregory Burns  
(ISD)



# Questions?

