NEICE Cyber-Security Analytics

Intern Presentation:
Aneesh Jonelagadda and Ron Gorai

Background

What is NEICE?

- Online System
    - Place children across state lines
    - Child Welfare

KOViD

Catalyst for Digital Evolution

Problem Statement

Required to create security reports of NEICE database usage

- Manually made

- Tedious

Resources

Server Logs

- Session ID
- IP
- Username
- Time Stamp
- Action

Computers

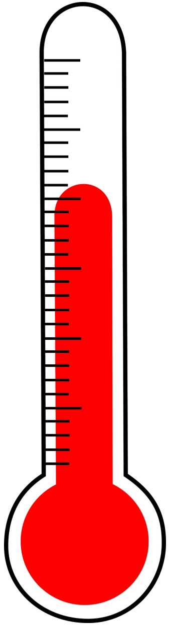
Fsczvzbvf3010pwapomdhjr2	166.94.34.233	aneesh.jonelagadda@atlantis.gov	2016-10-03 17:22:06.363	Search Child

Solutions

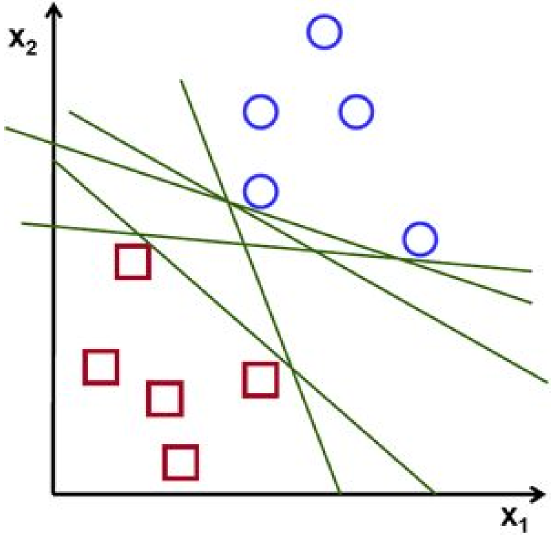
- Simple threshold-based identification

- Machine Learning - based identification

- Visualizations
    - Human nuance



or



Machine Learning

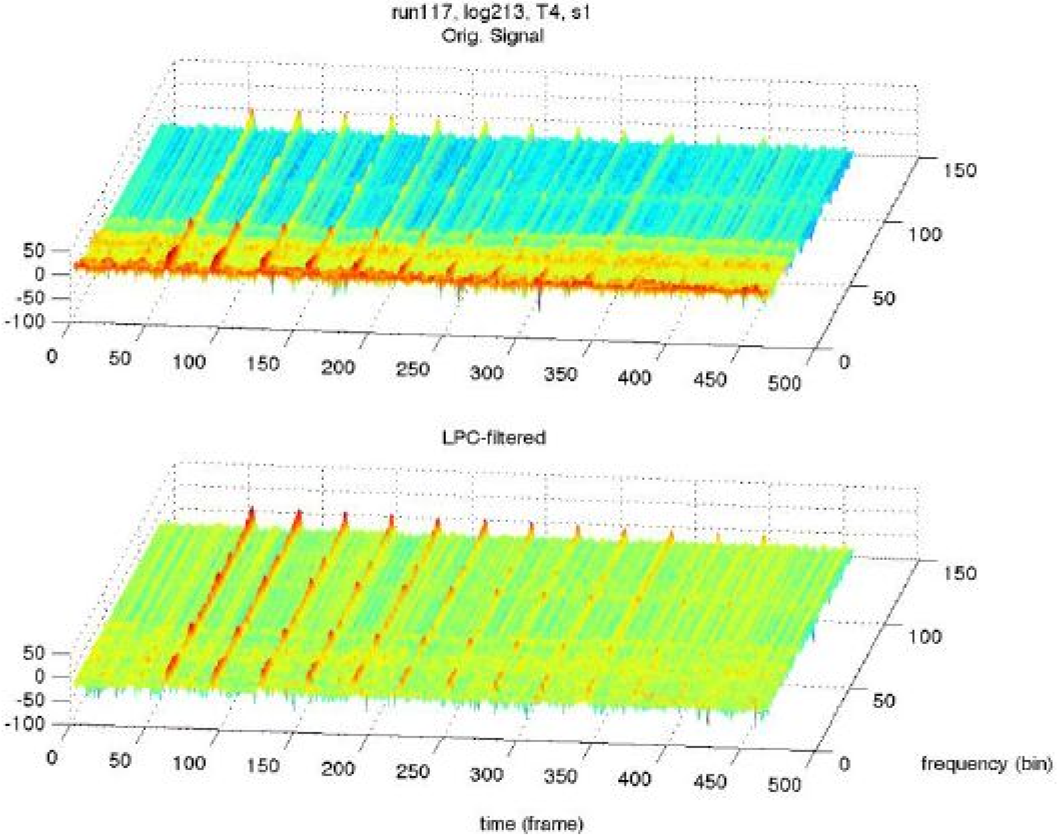
- Edge Detection

- Support Vector Machines (SVM)

- Time Series vs Cumulative

Edge Detection

- Training Data classification



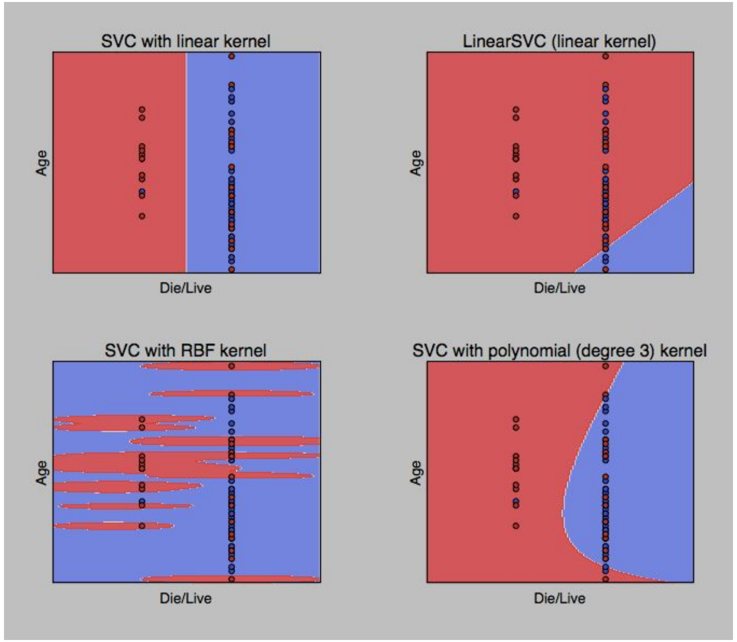
Support Vector Machines

- Game of Separation

- Kernels

- Draws boundaries based on training data

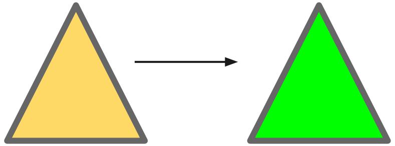
  Suspicious vs Not Suspicious



Two Models, Two Benefits

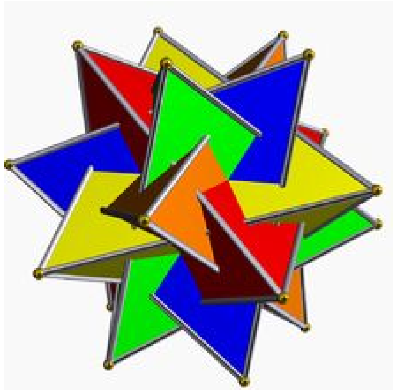
Model 1 = Time Series

- Pro: Change within role
- Con: Not compared to others



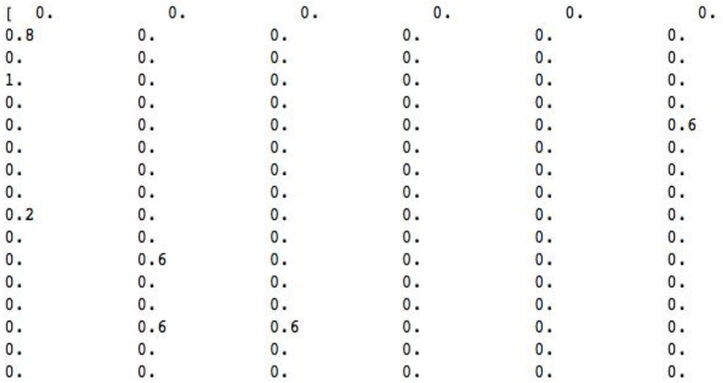
Model 2 = Cumulative

- Pro: "Normal User"
- Con: How does user change?



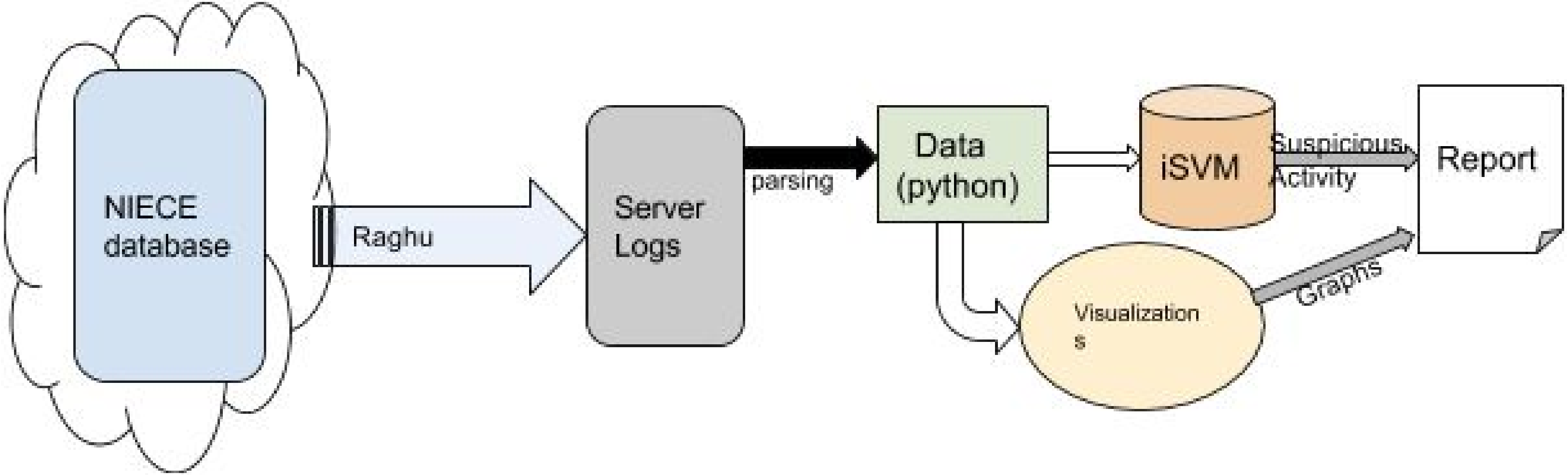
Action Profile

- Describes average actions of user
- Action Space



```
[   0.          0.          0.          0.          0.          0.
 0.8         0.          0.          0.          0.          0.
 0.          0.          0.          0.          0.          0.
 1.          0.          0.          0.          0.          0.
 0.          0.          0.          0.          0.          0.
 0.          0.          0.          0.          0.          0.6
 0.          0.          0.          0.          0.          0.
 0.          0.          0.          0.          0.          0.
 0.          0.          0.          0.          0.          0.
 0.2         0.          0.          0.          0.          0.
 0.          0.          0.          0.          0.          0.
 0.          0.6         0.          0.          0.          0.
 0.          0.          0.          0.          0.          0.
 0.          0.          0.          0.          0.          0.
 0.          0.6         0.6         0.          0.          0.
 0.          0.          0.          0.          0.          0.
 0.          0.          0.          0.          0.          0.
```

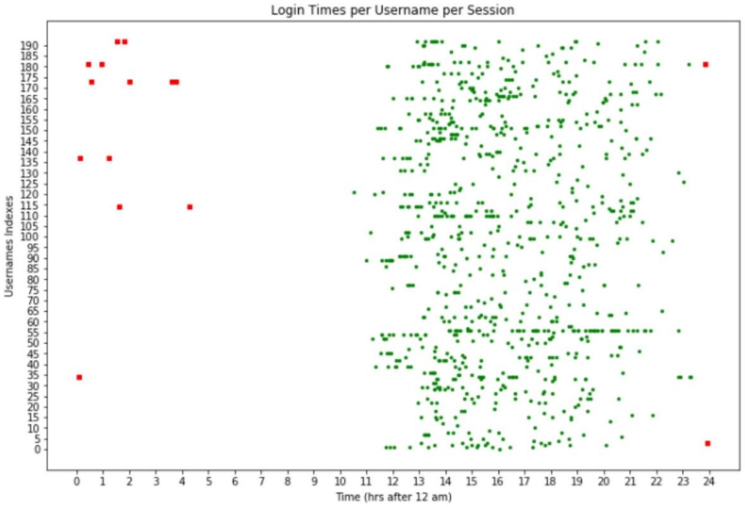
Pipeline

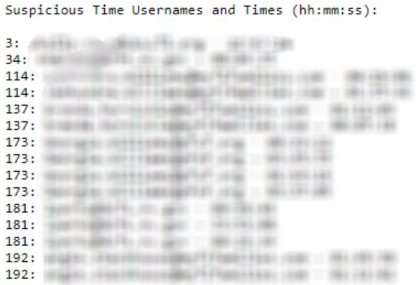

NIECE database
Raghu
Server Logs
parsing
Data (python)
iSVM
Suspicious Activity
Report
Visualization
s
Graphs

Results

- Folder created
    - Reports generated
        - Graphs
        - Suspicious Activity

Results (graphs)

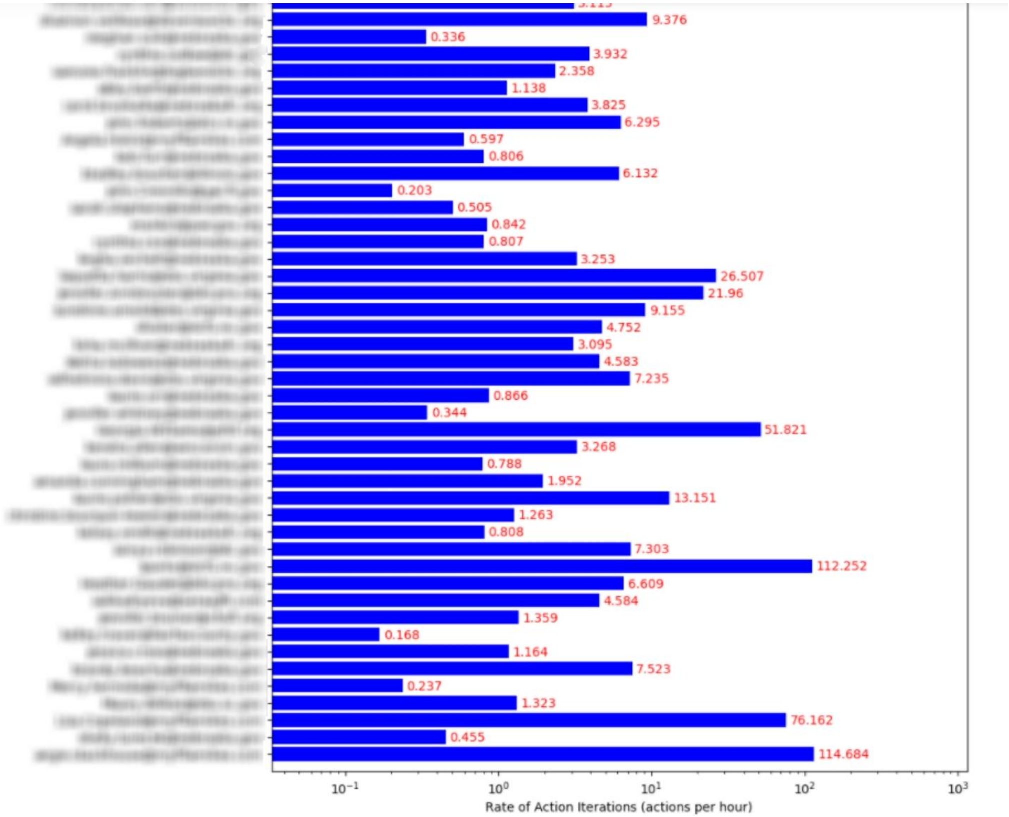




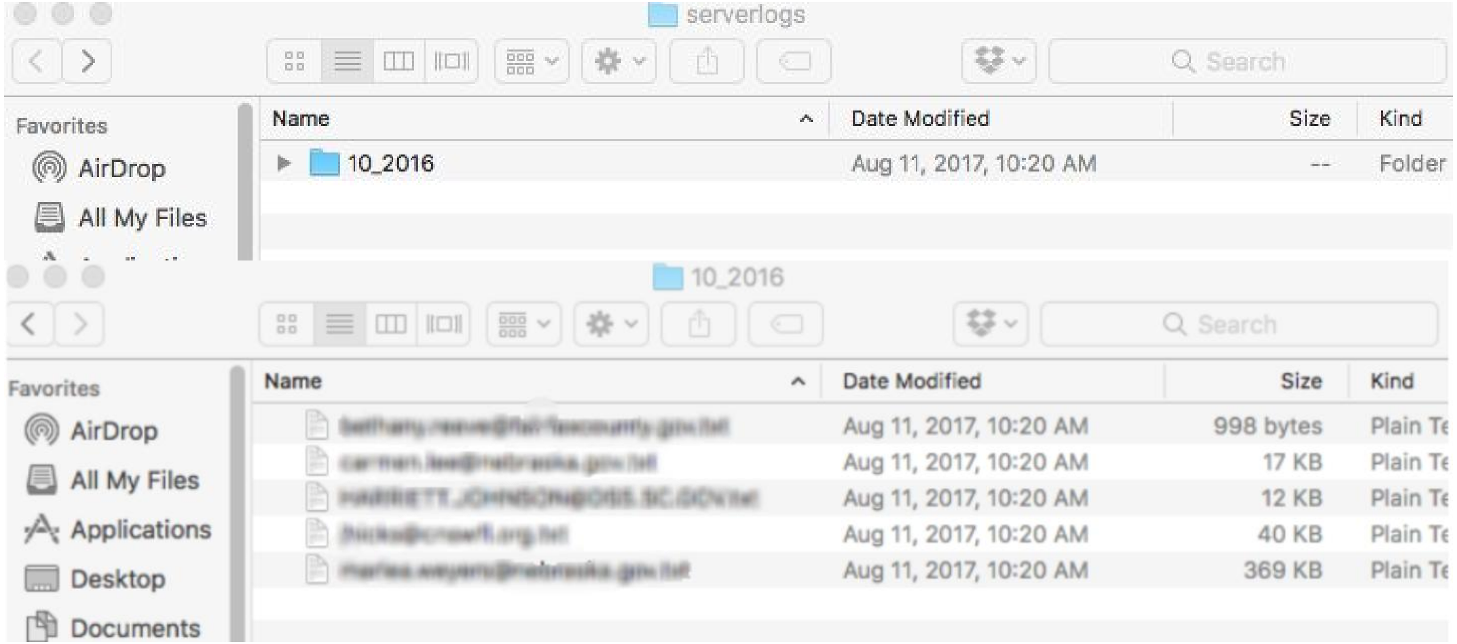
```
Suspicious Time Usernames and Times (hh:mm:ss):

3: [redacted]
34: [redacted]
114: [redacted]
114: [redacted]
137: [redacted]
137: [redacted]
173: [redacted]
173: [redacted]
173: [redacted]
173: [redacted]
181: [redacted]
181: [redacted]
181: [redacted]
192: [redacted]
192: [redacted]
```

Results (graphs)



Results (machine learning reports)



Name	Date Modified	Size	Kind
10_2016	Aug 11, 2017, 10:20 AM	--	Folder

Name	Date Modified	Size	Kind
[redacted]	Aug 11, 2017, 10:20 AM	998 bytes	Plain Te
[redacted]	Aug 11, 2017, 10:20 AM	17 KB	Plain Te
[redacted]	Aug 11, 2017, 10:20 AM	12 KB	Plain Te
[redacted]	Aug 11, 2017, 10:20 AM	40 KB	Plain Te
[redacted]	Aug 11, 2017, 10:20 AM	369 KB	Plain Te

Challenges

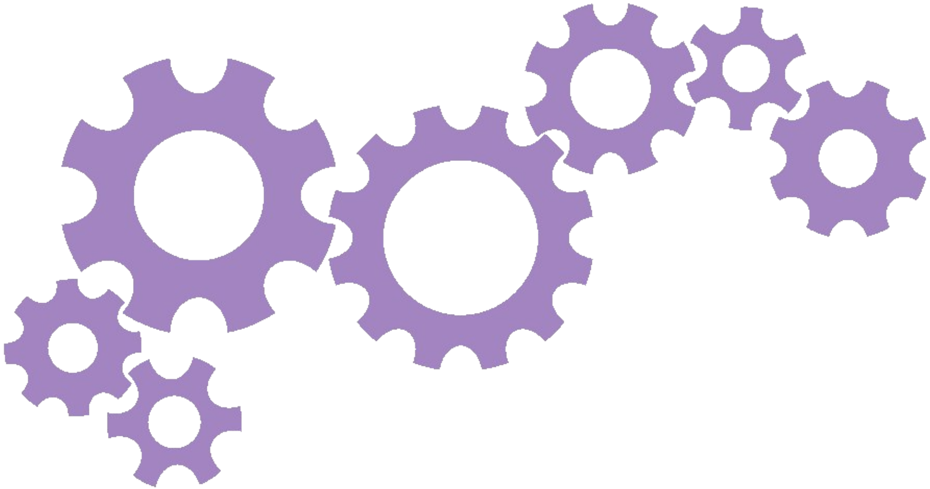
- Machine Learning data structure
- Parsing logs
- Parallelizing work

What I learned (Aneesh)

- Domain knowledge → Algorithms

- Colleague Interaction at Kovid

- Management



What I learned (Ron)

- Project Planning

- Parsing Log Data

- Constructing Visualizations