Methods

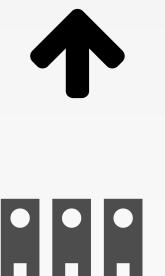
The method adapted is to utilize the Amazon Machine Learning software, it will be trained on a dataset comprised of normal user situations, crafted mistakes, and malicious activity. After providing the training dataset, the software will be instructed to make predictions against similar datasets to verify accuracy. It will then be tested against a human actor to test given to the model for batch prediction and testing dataset of 100 commands will be accuracy. predictions in a high-fidelity simulation. will be evaluated for statistical

Introduction and Motivation

Insider threats are on the rise, current commercial software can monitor, log, and prevent access to designated files and directories it remains difficult to predict and prevent unin research in the area, the focus of this study is to more accurately predict insider authorized insider usage. Due to the threats in a server environment.

Connection Setup

Set up Amazon Web Services connection
client = boto3.client('machinelearning')



Proa 10

School of Informatics, Computing, and Cyber Systems Northern Arizona University

Why the linux terminal?

ubiquity of the operating systems based on it are connected to the internet in servers around the globe. It is also transferable to all POSIX-compatible terminal environments that are in use today. Linux has been chosen because of the

Mistake of Warn

http://joshuabowen.info

jdb567@nau.edu

Bowen

Joshua

do not execute command
75 and ev >= 0.25:
I think your command may
or contain a mistake, please try agian.') # if mistake, do elif ev <= 0.75 a print('\nI the

Activity Malicious Prevent

stop session # if determined malicious, stop s
elif ev < 0.25 and ev >= 0:
 print('\nIntruder detected!')
 sys.exit(1)

Use Allow Normal

if good, allow command
ev > 0.75:
os.system(com)



data showing malicious and mistake commands, the vast majority of data points falls into the normal category.

unnecessary lines. Due to the lack of

gathered datasets I wrote simp

le programs to remove

research. Finally is the UNIX User Dataset from Purdue University. Due to the nature of of the two

ond is a list of commands assembled from my

an un-edited history file from my laptop. Sec-

Data SourcesThere are three main sources of data used for the machine learning model. The first is

The first is

User Get

from working dir # Update username, hostname, and work
username = os.environ.get('USER')
hostname = os.environ.get('HOSTNAME')
wdir = os.environ.get('PWD')

results of the high-fidelity simulation show that not only is the software I wrote ineffective as a terminal but that it is also easy to bypass. the participants were able to escape it and after that became untraceable to my

The

system, erall the process has not proved anything of statical significance.

Overall

prediction it has been shown about 20% when predicting

Results

mistake commands.

Or

completion of the predic the model is off by abou malicious

+ # - # - 0 + ind prompt
username).rstrip() +
+ str(wdir).rstrip()

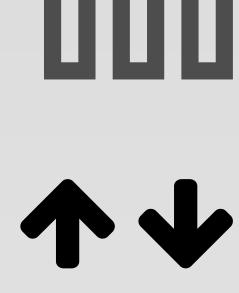
str(hostname).rstrip()

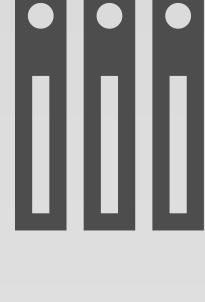
understndable # Get user input
cmd = input(prompt) # Convert to and
cmds = str(cmd)

string

return cmds

0 earnin Evaluatio Machine Send to





get_p

def

Correctly
request = {'

response Send

Conclusions

ased on the results of the tests it has been determined that the model isn't that accurate, even when given special weighting. This is likely due to the lack of data in the malicious and mistake categories. Future research in the area needs more open data. One solution is to ask organizations to publish the cleaned logs after an attack. Based

Another problem identified during this study is the programming needed to accurately capture information from a live terminal. While I was able to create a fake prompt it was severely limited and that skewed the data from the simulation.

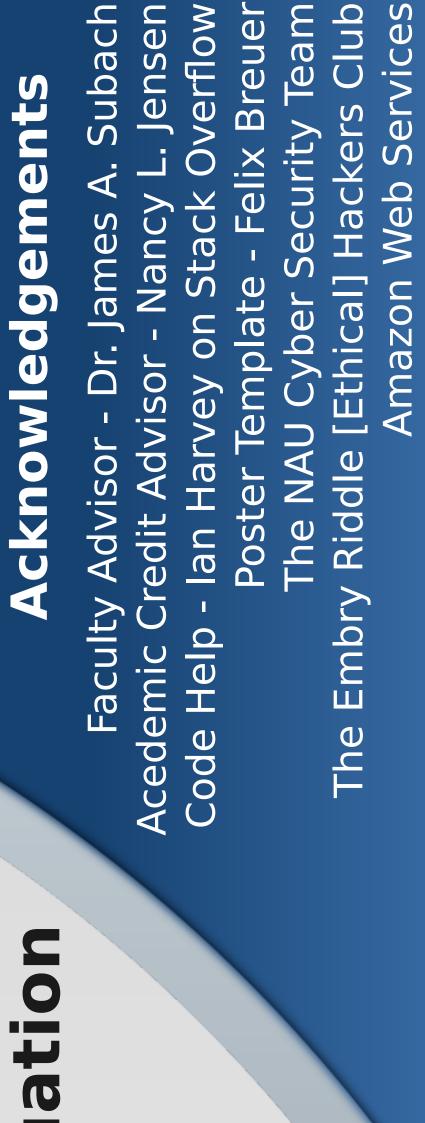
Future research will be needed in order to determine how to best approach the problem of practical proactive intruder detection. I theorize that larger datasets comprising of more points such as timestamps and keyboard information would help better track user behavior in an attempt to establish recognizable patterns by neural networks.

and Command 60-









Dr. James A. Subach

Font Awesome GitHub Material Source Code, Refrences,

http://github.com/Skraelingjar/cyberml