Agenda for Aug. 21, 2020 Meeting

Discussions and Planning for all three Funded Projects in AY 2020-21

GenCyber:

- Dates: February 6, 7, 13, 14, and 20. These are three Saturdays and two Sundays.
- Website at www.iup.edu/iupgencyber was updated and application site will open soon, see the link: https://form.jotform.com/202217692899165.
- LattePanda issues: QC, board getting enough power, WiFi connectivity, overheating, and bad customer service. ***Need to decide based on that whether to go with it or replace it with Raspberry Pi????***
- We need to start working on the modules per my previous email ASAP.
- We need to decide on a bi-weekly meeting time that will fit all of us as well as the teachers Gentile and Lent.
- Student helpers see the NCAE-C Research grant below.

NCAE-C Research grant:

- We have just received yesterday the partially executed grant. I am working with RI and SGSR to ensure it gets signed and sent to NSA soon.
- Project should officially start at the beginning of this Sept.
- We need to start on building the model we described in our grant, Xinwen input?
- We need to hire two RELIABLE undergraduate students to start with research work and also help with some other related assignments.
- If we have a graduate student, we can also give a GA. Otherwise, we can hire him/her hourly. Raj input??
- We need to decide on a time for weekly meeting that all three of us and all hired students attend to assign tasks and report on progress. **This grant has a mid-year report requirement and its renewal for second year is based on performance.**
- Any other feedback.

DoD Scholarship:

- We are still waiting to receive the partially executed grant. I contacted them a number of times to try to speed up this process up.
- For almost a month, I have been working with various offices at IUP to have everything prepared (tuition and fees waivers, stipends, books, etc.). We are almost ready just waiting to receive the official grant.
- Website at www.iup.edu/DoDScholarship will be updated once the grant is officially received.
- I am in contact with all students regarding the registration for courses.
- We will need to decide on a time for our monthly meeting for all twelve of us??
- Any other feedback of input??

Agenda for Sept. 4, 2020 Meeting

Discussions and Planning for all three Funded Projects in AY 2020-21

GenCyber:

- Dates: February 6, 7, 13, 14, and 20. These are three Saturdays and two Sundays.
- Website at www.iup.edu/iupgencyber now has the following:
  - Online Application on Jotform (we actually started receiving applications)
  - Camp Flyer
  - Camp Brochure
- We agreed last meeting to replace LattePanda with Raspberry Pi Kit. This info was updated in the flyer and website.
- We extended an offer for Daniel Saylor to work on GenCyber. He has never worked at IUP before thus he needs to go through the formal hiring process via HireAHawk. I have already sent him instructions and asked him to apply.
- We need to start working on the modules per my previous email ASAP.
- Based on input from the K-12 teachers, best time to meet is M/W after 6 PM. I will set up the first meeting soon based on this time.

NCAE-C Research grant:

- Fully executed grant was received, Sept 1. IUP award went on Sept. 2. I am still waiting for RI to hold a kick-off meeting then we will establish the required cost center to be able to start hiring students.
- After discussing this with both of you, we decide to hire John Conlen. I have already contacted him and will complete his profile sheet once we have the cost center.
- We will still need to hire one other RELIABLE undergraduate student, but we can hold on this for a while till we initiate the project.
- Raj indicated that we can use the help of some of his graduate students working with him this semester and if one is qualified, we can hire him or give him a GA. Raj input??
- We need to start on building the model we described in our grant, Xinwen input?
- We need to decide on a time for our weekly meeting????
- **This grant has a mid-year report requirement and its renewal for second year is based on performance.**

DoD Scholarship:

- We are still waiting to receive the fully executed grant. I am doing all what I can.
- I had just completed our first monthly meeting with CySP students in which we discussed: tuition and fees waivers, stipends, books, health insurance, registration for courses etc.
- Website at www.iup.edu/DoDScholarship will be updated once the grant is officially awarded.
- What is the best time for our future monthly meeting for all twelve of us???
- Any other feedback of input??

Agenda for Sept. 11, 2020 Meeting

Discussions and Planning for all three Funded Projects in AY 2020-21

GenCyber:

- Dates: February 6, 7, 13, 14, and 20. These are three Saturdays and two Sundays.
- Website at www.iup.edu/iupgencyber now has flyer, brochure, and open application.
- I have contacted the IASD superintendent office to request permission to distribute flyers at IJH and ISH.
- I have requested the creation of a D2L course to be used in delivering the online camp as per our previous discussion.
- I asked Dan Saylor to apply via HireAHawk for a Cybersecurity helper position. He applied and I have selected and approved his application. He has to go through few other steps with HR to be eligible to start working on the project.
- I will meet with Dan today to give him the Raspberry PI so he can start experimenting with it.
- We need to start working on the modules ASAP and assign tasks to Dan (10 hours).
- Based on input from the K-12 teachers, best time to meet is M/W after 6 PM. Are you OK with Monday the 14th or Wednesday the 16th at 6 PM.

NCAE-C Research grant:

- Fully executed grant was received, Sept 1. IUP award went on Sept. 2.
- I have requested an expedited establishment of the cost center and we now have it.
- I have asked John Conlen to complete the profile sheet and will sent it to HR once I receive it from him. John will be working for 10 hours per week for now.
- Discussion: Possible other candidates, Cassandra Lefever, Ethan Buhl, Reuben, graduate student(s) suggested by Xinwen/Cao.
- We need specific tasks to give to each student that we will hire.
- Xinwen, updates regarding the model and initial steps??
- It looks like this time is fine for our weekly meeting. Any comments.

DoD Scholarship:

- Just received the fully executed grant yesterday (9/10/2020) but effective date is Sept. 15. We asked for a pre-award authorization and we do expect to get that soon.
- After receiving the award, I submit right away the tuition waiver forms and grant all nine students their stipends on IUP NextGen.
- Website at www.iup.edu/DoDScholarship will be updated very soon.
- I had a meeting with Michelle Fryling to publicize the grant at IUP.
- I need to assign each student a specific task as a fulfillment to their obligations to the CySP program. I need your help coming up with these assignments and supervising these students.
- Any other feedback of input??

Agenda for Sept. 21, 2020 Meeting

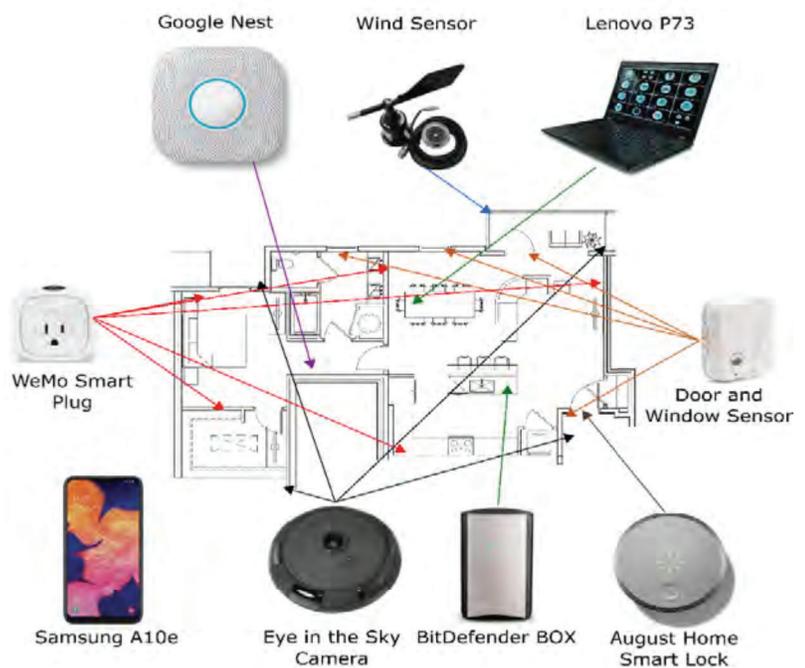## CAE-C Research Project in AY 2020-21

Announcements:

- We need one volunteer to record the meeting minutes, who is willing to do that??
- Project Title: ***Investigating Effective and Efficient Anomaly Detection on IoT Systems via a Novel Fusion of Deep Learning Techniques***.
- The project is supported by the NCAE-C Cyber Curriculum and Research 2020 Program.
- The main goal of this project is to ***develop a practical framework (that is based on a sound theoretical foundation) for the IoT anomaly detection in order to assess and improve the impact IoT systems have on the security of various networks***.
- Fully executed grant was received, earlier this Sept. and WF has requested an expedited establishment of the cost center which we now have.
- All three faculty were working over the past month on recruiting students to the research team.
- Project Team Members: PI: Dr. Waleed Farag, Co-PIs: Drs. Raj Ezekiel and Xinwen Wu, and Student Researchers: Ethan Buhl, John Conlen, Cassandra Lefever, and Alex Robbins.
- We will be meeting weekly to report on progress and discuss any issues.
- ***Given the nature of this project, it is essential that all of us respond promptly to ALL email communications.***

Actions:

- Our project will focus on developing an IoT framework for a smart building IoT system (see the Figure). Please make sure you carefully review this model.
- We need to decide on a weekly time that will fit all of us.
- Please refer to attachment #1 prepared by Dr. Wu for description of project timeline, milestones, and specific tasks.
- We will assign each student researcher a weekly task that s/he will work on, document all activities then report on progress during our next weekly meeting.
- ***Task assignments.***

Questions and Adjournment:

Agenda for Sept. 28, 2020 Meeting

## CAE-C Research Project in AY 2020-21

Announcements:

- We agreed that we will be meeting weekly on Mondays, at 5 PM. Please use the Zoom link provided to you via email.

Discussions:

- Below is a brief description that outlines the tasks (focusing on smart buildings) assigned to student researchers last week:

1. What are the major applications for a typical IoT based smart building, for each of these applications, describe the data format – the applications and their data formats will be critical for us to select anomaly detection algorithms and adapt them if necessary.

2. What devices are used for network switching and routing purposes? What communication media are used for a typical smart building? What operating systems are used?

3. In the perception layer (see the given Figure), what devices are connected to the IoT? For each type of these devices, give a description including:

    - What operating system is used?

    - How data are collected and communicated?

    - What is the data format?

    - What are the basic security mechanisms? Any authentication? Any ID management?

    - A vulnerability analysis, including, for example, what are the most vulnerable devices against malicious intrusion?
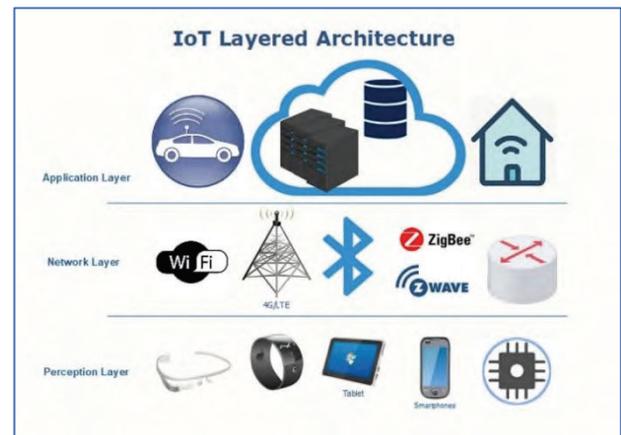
- Discussion of first group work progress (Alex and Cassandra).

- Discussion of second group work progress (John and Ethan).

- Discussion of issues and/or challenges.

Questions and Adjournment:

CAE Research Project Meeting Minutes

9/28/2020

- Participants
    - Dr. Waleed Farag (Project Director)
    - Dr. Soundararajan Ezekiel (Co-Director)
    - Dr. Xin-Wen Wu (Co-Director)
    - Ethan Buhl
    - Cassandra Lefever
    - John Conlen
    - Alex Robbins
- Meeting starts at 5:03 p.m.
- Main Points:
    - Weekly meetings are set for 5 p.m. every Monday via Zoom
    - Steps for 1.1, review of tasks (please see project layout)
    - Is everyone able to access the shared OneDrive? ACM Digital Library?
    - Group progress
        - Alex and Cassandra
            - Look into Zigbee more, shift focus to hub and how data is formatted and processed
        - John and Ethan
    - Consider scalability: Automobile IoT, Body Area Network, different types of LANs (IoAT)
    - Consistent format for documentation (IEEE) will be posted to OneDrive
- Adjournment 5:45 p.m.
- Next meeting will be held on Monday, October 5, 2020 via Dr. Farag Zoom room.

CAE Research Project Meeting Minutes
10/5/2020

- Participants
    - Dr. Waleed Farag (Project Director)
    - Dr. Soundararajan Ezekiel (Co-Director)
    - Dr. Xin-Wen Wu (Co-Director)
    - Ethan Buhl
    - Cassandra Lefever
    - John Conlen
    - Alex Robbins
- Meeting start 5:02 p.m.
- Main Points
    - Group progress
        - John and Ethan – network components and smart building operating systems
            - Goals/Plans for next week: elaborate on XBOS and other IoT OSs, pick some off the list below (Windows, Amazon, TinyOS, etc.)
            - https://www.g2.com/categories/iot-operating-systems
        - Alex and Cassandra
            - Goals/Plans for next week: Investigate Thread networking
- Adjournment 5:55 p.m.
- Next meeting will be held on Monday, October 12, 2020 via Dr. Farag Zoom room.

CAE Research Project Meeting Minutes
10/12/2020

- Participants
  - Dr. Waleed Farag (Project Director)
  - Dr. Soundararajan Ezekiel (Co-Director)
  - Dr. Xin-Wen Wu (Co-Director)
  - Ethan Buhl
  - Cassandra Lefever
  - John Conlen
  - Alex Robbins
- Meeting start 5:01 p.m.
- Minutes of last meeting held on Oct. 5, 2020 were unanimously approved.
- Main Points
  - Group progress
    - John and Ethan – IoT OSs
      - AmazonFreeRTOS
        - Seems promising, extensive documentation
      - Windows for IoT
      - Contiki
      - TinyOS
      - Tasks for next week – find boards that work with AmazonFreeRTOS, research popularity of these OSs, minimum number of OSs we may need
    - Alex and Cassandra
      - Thread
  - Plans for next week and beyond
    - Wrap up Phase 1.1
      - Focus on Zigbee and AmazonFreeRTOS, address three bullets in project outline, produce a technical report (2-3 pages maybe more), appendix can have details
    - Meet as a group to determine which standard and OSs we should use, check compatibility requirements
- Adjournment 5:50 p.m.
- Next meeting will be held on Monday, October 19, 2020 via Dr. Farag Zoom room.

CAE Research Project Meeting Minutes
10/19/2020

- Participants
  - Dr. Waleed Farag (Project Director)
  - Dr. Soundararajan Ezekiel (Co-Director)
  - Dr. Xin-Wen Wu (Co-Director)
  - Ethan Buhl
  - Cassandra Lefever
  - John Conlen
  - Alex Robbins
- Meeting start 5:03 p.m.
- Main Points
  - Group Progress
    - Alex and Cassandra
      - Stage 1.1 summary report
      - Goals for next week:
        - Expand on security part of paper: reasons attacks occur, vulnerabilities, implementation problems, any unusual behaviors when a device is compromised
    - John and Ethan
      - Zigbee and AmazonFreeRTOS are compatible, xGM210P Development Kit is a possible candidate for our project
      - Goals for next week: Stage 1.1 summary report, investigate data collection on devices and by OSs (ex: login credentials)
  - Stage 1.2
    - See research work breakdown for details
    - We will keep the group setup the same
  - Sometime in the next few weeks we will discuss IEEE format and clean up our reports
  - Cyber Security Day is tomorrow, 10/20, starting at 9 a.m., see you there!
- Adjournment 5:45 p.m.
- Next meeting will be held on Monday, October 26, 2020 over Zoom via Dr. Farag Zoom room.

CAE Research Project Meeting Minutes
10/26/2020

- Participants
    - Dr. Waleed Farag (Project Director)
    - Dr. Soundararajan Ezekiel (Co-Director)
    - Dr. Xin-Wen Wu (Co-Director)
    - Ethan Buhl
    - Cassandra Lefever
    - Alex Robbins
    - John Conlen was excused due to sickness
- Meeting start 5:03 p.m.
- Main Points
    - The abstracts from last week were submitted successfully to SPIE
    - Alex and Cassandra
        - Reviewed stage 1.1 write-up, IoT hubs and gateways
        - For Next Week: Research computational capabilities of IoT hub devices (processors, RAM, local storage), issues with efficiency in cloud storage, bottom-up anomaly detection
    - Ethan (John was sick)
        - Terminology, trying to define anomalies
        - For next week: Supervised vs Unsupervised anomaly detection, items in 1.2, 1.1 write-up
- Adjournment 5:50 p.m.
- Next meeting will be held on Monday, November 2, 2020 via Dr. Farag Zoom room.

CAE Research Project Meeting Minutes

11/2/2020

- Participants
    - Dr. Waleed Farag (Project Director)
    - Dr. Soundararajan Ezekiel (Co-Director)
    - Dr. Xin-Wen Wu (Co-Director)
    - Ethan Buhl
    - Cassandra Lefever
    - John Conlen
    - Alex Robbins
- Minutes of the previous meeting dated Oct. 26, 2020 were unanimously approved.
- Meeting start 5:05 p.m.
- Main Points
    - Alex and Cassandra
        - IoT hub capabilities and logging options, IDS diagram, anomaly detection
        - For Next Week: compare data sets we have with the data that can be gathered using the logging options found, maybe start drafting a paper
    - John and Ethan
        - Finalized report for phase 1.1, Network traffic collection, data sets to consider
        - For Next Week: Download the datasets and have them visualized for next meeting
- Adjournment 6:10 p.m.
- Next meeting will be held on Monday, November 9, 2020 via Dr. Farag Zoom room.

CAE Research Project Meeting Minutes
11/9/2020

- Participants
    - Dr. Waleed Farag (Project Director)
    - Dr. Soundararajan Ezekiel (Co-Director)
    - Cassandra Lefever
    - John Conlen
    - Alex Robbins
- Unanimously approved minutes for meeting on November 2, 2020
- Meeting start 5:05 p.m.
- Main Points
    - Alex and Cassandra
        - Draft paper for stage 1.2, sample screenshots of databases, generative adversarial networks
        - For Next Week: Look into adding more journal references to the paper, continue to look into GANs and virtual machines, and working on the paper
    - John
        - Downloaded and sorting through datasets, network traffic collection
        - For Next Week: extract a few columns through MATLAB to search for anomalies
- Adjournment 5:45 p.m.
- Next meeting will be held on Monday, November 16, 2020 at 5 p.m. via Dr. Farag Zoom room.

CAE Research Project Meeting Minutes
11/16/2020

- Participants
    - Dr. Waleed Farag (Project Director)
    - Dr. Soundararajan Ezekiel (Co-Director)
    - Dr. Xin-Wen Wu (Co-Director)
    - Ethan Buhl
    - John Conlen
    - Cassandra Lefever
    - Alex Robbins
- Meeting start 5:03 p.m.
- Unanimously approved minutes for meeting on November 9, 2020
- Main Points
    - John and Ethan
        - Datasets for anomaly detection, packet sniffing software/hardware
        - For next week: contact owners of datasets, try to find a way to access the larger ones, start working on and finalizing write-up for stage 1.2
    - Alex and Cassandra
        - Stage 1.2 write-up update
        - For Next Week: work on finalizing stage 1.2 paper
    - CAE Scholarship Update
    - Moving into stage 1.3 after this week, see task breakdown document
    - Working Thanksgiving break or take the week off? We will discuss on Monday.
- Adjournment 6:05 p.m.
- Next meeting will be held on Monday, November 23, 2020 at 5 p.m. via Dr. Farag Zoom room.

CAE Research Project Meeting Minutes

11/23/2020

- Participants
  - Dr. Waleed Farag (Project Director)
  - Dr. Soundararajan Ezekiel (Co-Director)
  - Cassandra Lefever
  - Alex Robbins
- Meeting start 5:03 p.m.
- Unanimously approved minutes for meeting on November 16, 2020
- Main Points
  - Alex and Cassandra
    - Update on stage 1.2 paper: GANs, Fence GANs, virtual machines
    - For Next Meeting: finish up 1.2 paper, Alex: investigate side channel attacks and Cassandra: meet with Dr. Ezekiel about GANs at your convenience
  - We will take the week of Thanksgiving off, there will be no meeting this upcoming Monday. Enjoy the break!
- Adjournment 5:40 p.m.
- Next meeting will be held on Monday, December 7, 2020 at 5 p.m. via Dr. Farag Zoom room.

CAE Research Project Meeting Minutes
12/14/2020

- Participants
  - Dr. Waleed Farag (Project Director)
  - Dr. Soundararajan Ezekiel
  - Cassandra Lefever
  - Alex Robbins
- Meeting starts at 11 AM
- Main Points
  - Alex and Casandra
    - Dr. Pearlstein: Fence GAN could work, but it would take a lot of extra effort. Open discussion to the team.
    - Topics Covered: Side channel attacks, ghost attacks related to attribute selection
- Adjournment
- Next meeting will be held on Tuesday, December 22, 2020 via Dr. Farag Zoom room.

CAE Research Project Meeting Minutes
12/22/2020

- Participants
  - Dr. Waleed Farag (PI)
  - Dr. Soundararajan Ezekiel (Co-PI)
  - Dr. Xin-Wen Wu (Co-PI)
  - Ethan Buhl
  - Cassandra Lefever
  - Alex Robbins
- Meeting start 11:05 AM
- Main Points
  - Next meeting with Dr. Pearlstein will be Saturday or Sunday at 10 a.m. Dr. Ezekiel will send out an email with this information.
  - Group Updates
    - Ethan
      - Found four datasets, hoping to use the Titan machine to start processing and graphing data
      - For next time: get familiar with the datasets that were found and the structure of them, focus on what dataset is most suitable for our application; get in contact with Adam and Rohith to get access to datasets that are already on the machine
      - Titan machine needs to be turned on (it is in Stright Hall room 305), Dr. Ezekiel will take care of this; Dr. Pearlstein also has a machine that we can use
    - Alex and Cassandra
      - Investigated: feature engineering, PCA, GANs, WGANs, Python scripting
- Adjournment 12:05 p.m.
- Next meeting will be held on Tuesday, December 29th, 2020 at 11:00 a.m. via Dr. Farag Zoom room.

CAE Research Project Meeting Minutes
12/29/2020

- Participants
  - o Dr. Waleed Farag (PI)
  - o Dr. Soundararajan Ezekiel (Co-PI)
  - o Dr. Xin-Wen Wu (Co-PI)
  - o Ethan Buhl
  - o Cassandra Lefever
  - o Alex Robbins
- Meeting start at 11:04 AM
- Corrected and unanimously approved minutes for meetings on December 14, 2020 and December 22, 2020
- Main Points
  - o For now, the Titan machine is down. We will work to get it up and running ASAP. In the meantime, if a machine is necessary, Dr. Pearlstein is willing to let us use his machine.
  - o Group Updates:
    - Update on "IoT Device Network Logs" dataset and "IoT Network Intrusion" dataset
    - IP address embedding
    - **For next week: all students work together to identify a best-fit dataset for our application, contact Dr. Ezekiel as soon as one is found and we will meet with Dr. Pearlstein to discuss next steps**
- Adjournment at 12:05 PM
- Next meeting will be held on Tuesday, January 5th, 2021 at 11 a.m. via Dr. Farag Zoom room.

CAE Research Project Meeting Minutes
1/5/2021

- Participants
  - Dr. Waleed Farag (PI)
  - Dr. Soundararajan Ezekiel (Co-PI)
  - Dr. Xin-Wen Wu (Co-PI)
  - Ethan Buhl
  - Cassandra Lefever
  - Alex Robbins
- Meeting starts at 11:05 AM
- Unanimously approved minutes for meeting on December 29, 2020
- Main Points
  - Update on COMNETS dataset
    - We can convert pcap to csv
    - Since the data is unlabeled, we may want to proceed with using the autoencoder that is set up on the Titan machine as our next step to process the data
  - Update on VTech thesis paper for encoding network information
  - Tasks:
    - Alex and Dr. Ezekiel will meet at some point this week to start working with the data. Which parameters should be used, extracted, etc.?
    - Ethan - We want to get access to the MatLab and Python code that is on the machine. Adam may have a copy of the script if we cannot fix the machine. Dr. Ezekiel will be emailing updates.
    - Cassie – Continue to read and try to understand the master thesis on encoding network information.
  - Students: Your assistance may be required for another project on February 6, 7, and 13. Details will be posted soon.
- Adjournment at 12:05 PM
- Next meeting will be held on Tuesday, January 12, 2021 at 11:00 a.m. via Dr. Farag Zoom room.

CAE Research Project Meeting Minutes

1/12/2021

- Participants
    - Dr. Waleed Farag (PI)
    - Dr. Soundararajan Ezekiel (Co-PI)
    - Dr. Xin-Wen Wu (Co-PI)
    - Ethan Buhl
    - Cassandra Lefever
    - Alex Robbins
- Meeting starts at 11:05 AM
- Unanimously approved minutes for meeting on January 5, 2021
- Main Points
    - Updates:
        - The replacement parts have been ordered for the Titan machine; it should be ready by Friday.
        - SPIE conference – students will need to work on the papers, PowerPoint slides, and a video explanation for the submitted abstracts. Dr. Ezekiel will advise students on next steps for this.
    - Update on dataset:
        - Alex – worked on feature extraction from raw PCAP files
    - Update on Vtech thesis for embedding network information.
        - For next time: Dr. Ezekiel will find the notes on One-Hot encoding from previous research and send these to Cassie to complement her work.
        - Cassie - Also, try to get a better understanding of the IP embedding that was being done in the paper.
    - Ethan and Alex: While we are waiting for the machine to be repaired, Dr. Wu will send a link to a resource we can use to start training an algorithm. Try to download/run some algorithms for next week.
    - GenCyber Camp for Middle and High School Students:
        - Students in the project will help by creating/co-hosting Zoom sessions, monitoring Zoom sessions, providing technical support for individual students, etc.
        - Sessions will be held:
        - February 6, 7, 13, 14, 20
        - 9 a.m. to 3 p.m.
- Adjournment at 12:15 PM
- Next meeting will be held on Tuesday, January 19, 2021 at 2:00 p.m. (via Dr. Farag Zoom room) pending approval from Dr. Ezekiel.

CAE Research Project Meeting Minutes
1/21/2021

- Participants
  - Dr. Waleed Farag (PI)
  - Dr. Soundararajan Ezekiel (Co-PI)
  - Dr. Xin-Wen Wu (Co-PI)
  - Ethan Buhl
  - Cassandra Lefever
  - Alex Robbins
- Meeting starts at 5:00 p.m.
- Unanimously approved minutes for meeting on January 12, 2021
- Main Points
  - Updates:
    - Alex Update on h2o scripts: Kmeans are running for all CSV files, PCA runs on Xbee_Normal only
    - Ethan: ran scripts and started looking into tutorials on h2o AutoML
    - Meeting with Dr. Pearlstein, Dr. Ezekiel, and Casey:
      - Looked into IP2vec – Dr. Pearlstein does not believe this method is useful for us, recommended converting IP addresses to latitude/longitude – APIs exist to help with this
  - Tasks:
    - Cassie – for now, continuing to investigate the alternate approaches that Dr. Pearlstein recommended for encoding
    - Ethan – look through the h2o and AutoML documentation to find more details on the output interpretation
    - Tomorrow, Friday, Dr. Ezekiel, Ethan, and Alex have to meet to discuss the outputs of the algorithms and start on next steps. Cassie can join too if not busy.
  - GenCyber Camp for Middle and High School Students:
    - **Meeting for student event helpers will be held on Sunday, January 24, 2021 tentatively at 10 a.m.**
    - They will be oriented during the meeting on their duties for the camp sessions.
    - Students in the project will help by creating/co-hosting Zoom sessions, monitoring Zoom sessions, providing technical support for individual students, etc.
    - Sessions will be held:
    - February 6, 7, 13, 14, 20
    - 9 a.m. to 3 p.m.
- Adjournment at 6:00 p.m.
- Next meeting will be held on Thursday, January 28, 2021 at 4:30 p.m. via Dr. Farag Zoom room.

CAE Research Project Meeting Minutes

1/28/2021

- Participants
  - Dr. Waleed Farag (PI)
  - Dr. Soundararajan Ezekiel (Co-PI)
  - Dr. Xin-Wen Wu (Co-PI)
  - Ethan Buhl
  - Cassandra Lefever
  - Alex Robbins
- Meeting start at 4:50 p.m.
- Unanimously approved minutes for meeting on January 21, 2021
- Main Points
  - Updates:
    - The Titan machine is up and running. Cassie and Ethan can access the machine now.
    - Everyone who uses Titan, please pay attention to the temperatures and let Dr. Ezekiel know if there are any problems (ex: if it goes above 50 degrees)
    - Cassie – looked into IP addresses, what attributes are important in other anomaly detection algorithms for IoT
    - Ethan - We are having issues installing h2o on the Titan. Perhaps we can get help from Jackson on that.
  - Tasks:
    - Ethan and Jackson – Try to get h2o working and perform some inventory/housekeeping on the Titan – what resources we have, what files are there, etc. Be sure to make a backup of the GAN code and store in the cloud (ex: GitHub, Bitbucket)
    - Cassie – Continue your research and also help Ethan with the GAN
  - GenCyber Camp for Middle and High School Students:
    - Sessions will be held:
      - February 6, 7, 13, 14, 20
      - 9 a.m. to 3 p.m.
    - Students be on the lookout for a message from Dan and Alex. We will be holding a dry run of the camp materials so we can test them and make everyone aware of their roles.
- Adjournment at 5:30 PM
- Next meeting will be held on Thursday, February 4, 2021 at 4:30 p.m. via Dr. Farag Zoom room.

CAE Research Project Meeting Minutes

2/11/2021

- Participants
  - Dr. Waleed Farag (PI)
  - Dr. Soundararajan Ezekiel (Co-PI)
  - Dr. Xin-Wen Wu (Co-PI)
  - Ethan Buhl
  - Cassandra Lefever
  - Alex Robbins
  - Maria Balega
  - Adam Lutz (guest)
- Meeting starts at 4:00 p.m.
- Unanimously approved minutes for meeting on January 28, 2021
- Main Points
  - Updates:
    - The mid-report deadline for the project is due at the end of March
    - Titan is up and running smoothly
    - Titan machine environment has many algorithms for us to use: GAN, DCNN, clustering
    - Cassie – Preprocessing, embedding network information
    - Ethan – Set up a virtual environment for running h2o on Titan, but getting errors running the script
  - Tasks:
    - Alex, Cassie, Ethan, Maria – will be meeting with Adam as soon as Adam gets the Titan TensorFlow environment fixed
    - Start getting results from the GAN by end-of-day Monday
    - Adam, Ethan - Need to inventory and backup important code and documentation on the machine; start documenting the algorithms we have
    - Ethan - On Monday, we will meet to discuss the status of the PCA scripts and follow up with Dr. Pearlstein
    - Cassie, Maria – work together, continue looking into preprocessing, meet with Dr. Ezekiel for help
  - GenCyber Camp
    - Sessions will be held:
      - February 13, 14, 20
      - 9 a.m. to 3 p.m.
    - Be on the lookout for emails from Dr. Farag and Dan regarding this project. Announcements will be posted in the Slack as well.
- Adjournment at
- Next meeting will be held on Thursday, February 18, 2021 at 4:30 p.m. via Dr. Farag Zoom room.

Autoencoder – Meeting 2/16/2021

- Autoencoders are mainly used for labeling
    - Unsupervised learning, data is unlabeled, used for training purposes
- Needs a ground truth:
    - Something for the autoencoder to train on, so it has something to compare other data to
    - Ground truth data represents the normal conditions
- There is a hidden layer in this approach
- To run autoencoder on the Titan:
    1. Open Anaconda Navigator – "anaconda-navigator" on the terminal
    2. Go to Environments, select the tf environment
    3. Open Jupyter Notebook
    4. File is IoTAE
    5. Steps to ML
        a. Add CSV file for normal data
        b. Add CSV file for test data
        c. Normalize data (ln [42]) – everything is scaled between 0 and 1 so its all in the same range
        d. The model is in ln [50] – 3 layers with a hidden layer in the middle
        e. Model gets trained, the loss gets lowered (which is what we want)
        f. Plot the distribution of the loss, want a good distribution to determine a threshold
        g. To get better results, we can adjust the NUM_EPOCHS
    6. The reconstruction error/score is what we will use to separate anomalous data from normal data
    7. X_Test – first it is normal, then after you change it (because the model has been made) ONLY run lines 64 and 65

See comments – the script deals with one file that is split, we can get around this my removing that step

Should create our own python notebook file

Meeting 2/17/2021

- Algorithms we will consider
    - Autoencoder
    - DCNN
    - GAN
    - Peak finding algorithm
    - RANSAC
- Algorithms go into a black box which takes in anomalous data and output flagged anomalies

-
- Autoencoder
    - Compute mean and variance
    - Mahalanobis distance between vectors
    - Our autoencoder looks at loss mean and variance
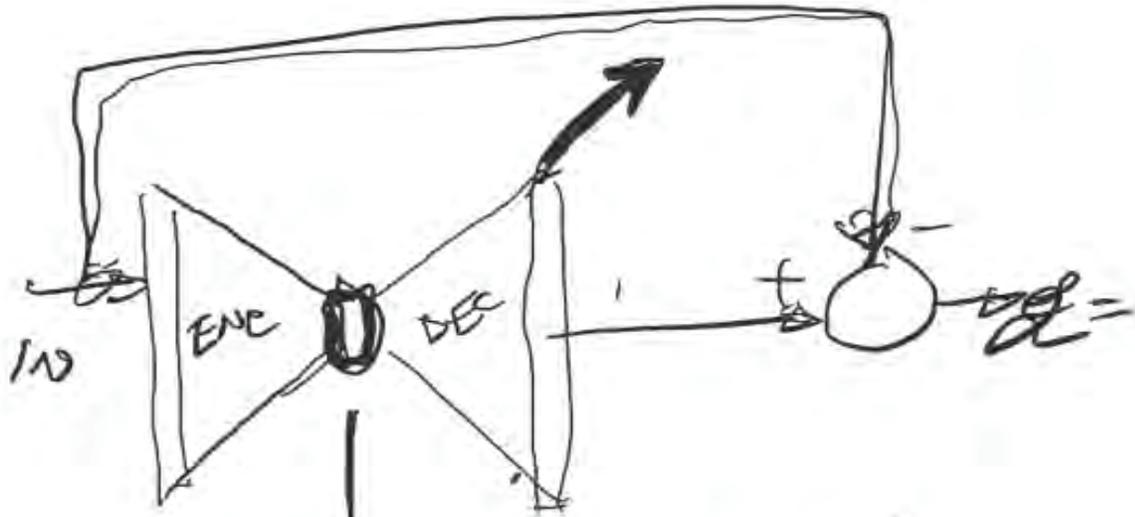
compute mean &
variance

Mahalanobis dist

Look @ £ p.d.f. mean & variance

- Limit the error, anomalous data is very infrequent – can allow for some anomalies in training

-
- If the error is very far out, reject it
- Use a clipped MSE
- https://stats.stackexchange.com/questions/350211/loss-function-autoencoder-vs-variational-autoencoder-or-mse-loss-vs-binary-cross ???

- Peak finding
  - Medium filters – used to try to reject peaks
- We also want to be able to compare the results of each algorithm
- Dr. Pearlstein can provide us with more energy usage data
  - It is unlabeled though

TODO:

- Clean up COMNETS dataset
  - Remove bad columns
  - Convert everything to decimal – remove/parse out colons and extra commas between values in the same field
- Modify Adam's script
  - Take in two files separately instead of splitting one file into two (as discussed above and in code comments)
- Run the script with the COMNETS data, get output
- Meet with Raj, Adam, and Jitu to discuss results

CAE Research Project Meeting Minutes

2/18/2021

- Participants
  - Dr. Waleed Farag (PI)
  - Dr. Soundararajan Ezekiel (Co-PI)
  - Dr. Xin-Wen Wu (Co-PI)
  - Ethan Buhl
  - Cassandra Lefever
  - Alex Robbins
  - Maria Balega
- Meeting starts at 4:30 p.m.
- Unanimously approved minutes for meeting on February 11, 2021
- Main Points
  - Updates:
    - Recap on previous meetings with Adam and Dr. Pearlstein
  - Tasks:
    - Alex – Clean up the dataset – using MatLab, convert everything to hexadecimal so the autoencoder can run it
    - Ethan – Look into the RANSAC algorithm – any pre-written code?
    - Cassie and Maria – Look into peak finding algorithms and after Alex is finished start running tests on the autoencoder
  - GenCyber Camp
    - Our last session will be held:
      - February 20
      - 9 a.m. to 3 p.m.
    - Be on the lookout for emails from Dr. Farag and Dan regarding this project. Announcements will be posted in the Slack as well.
- Adjournment at
- Next meeting will be held on Thursday, February 25, 2021 at 4:30 p.m. via Dr. Farag Zoom room.

CAE Research Project Meeting Minutes
2/25/2021

- Participants
    - Dr. Waleed Farag (PI)
    - Dr. Soundararajan Ezekiel (Co-PI)
    - Dr. Xin-Wen Wu (Co-PI)
    - Ethan Buhl
    - Cassandra Lefever
    - Alex Robbins
- Meeting starts at 4:00 p.m.
- Unanimously approved minutes for meeting on February 18, 2021
- Main Points
    - Updates:
        - Alex –
            - Datasets – cleaned normal data, attack data is in the process of being cleaned
            - Autoencoder Script – copied the script into new file and removed excess, getting an error when trying to run it
        - Ethan –
            - RANSACs – found documentation on RANSAC and some frameworks for us to use in the future, found Python libraries on GitHub
        - Cassie –
            - Peak Finding Algorithms, worked with Dr. Ezekiel to discuss this and a wavelet program we have to work with
    - Tasks:
        - Alex –
            - Work on parallel processing with the GPUs in MATLAB to speed up the data cleaning process
            - Work with Adam and/or Jackson to troubleshoot the autoencoder
            - Send finalized SPIE paper to Dr. Farag, Dr. Wu, and Dr. Ezekiel for final approval and submission
        - Ethan –
            - Continue to explore RANSAC – start running some of the libraries that were found on RANSAC to get some initial results; compare them, keep track of datasets used, data sizes, values, results in tables
            - Start developing a means to compare RANSAC to the other algorithms
            - Dr. Wu will share a document with explanations of some algorithms to help out
        - Cassie –
            - Continue to work on peak finding and the wavelets program
- Adjournment at 4:55 p.m.
- Next meeting will be held on Thursday, March 4, 2021 at 4:30 p.m. via Dr. Farag Zoom room.

CAE Research Project Meeting Minutes

3/3/2021

- Participants
  - Dr. Waleed Farag (PI)
  - Dr. Soundararajan Ezekiel (Co-PI)
  - Dr. Xin-Wen Wu (Co-PI)
  - Ethan Buhl
  - Cassandra Lefever
  - Alex Robbins
  - Maria Balega
- Meeting starts at 4:00 p.m.
- Unanimously approved minutes for meeting on February 25, 2021
- Main Points
  - GenCyber:
    - Follow-up activity is on Saturday, March 6th, from 9 a.m. to 11 a.m.
    - Please see the email from Dr. Farag for specific instructions
  - Updates:
    - Ethan – More on RANSAC, MatLab has a package on RANSAC
    - Alex – update on autoencoder
    - Cassie and Maria –
      - found new dataset that could be useful, it is labelled
      - wrote a short peak finding algorithm on the Titan
  - Tasks:
    - Ethan – add the MatLab package to the Titan and start running it
    - Cassie, Maria – meet with Adam to try running the new dataset on DCNN in the Titan
    - Alex - We also need to keep looking for IoT-specific datasets, specifically with Zigbee
- Adjournment at 5:00 p.m.
- Next meeting will be held on TBD, check for email from Dr. Farag (via Dr. Farag Zoom room).

CAE Research Project Meeting Minutes
3/11/2021

- Participants
    - o Dr. Waleed Farag (PI)
    - o Dr. Xin-Wen Wu (Co-PI)
    - o Ethan Buhl
    - o Cassandra Lefever
    - o Alex Robbins
    - o Maria Balega
- Meeting starts at 4:00 p.m.
- Unanimously approved minutes for meeting on March 3, 2021
- Main Points
    - o Updates:
        - Cassie and Maria
            - Demonstrated Peak Finding test script using random data, could help us with unlabeled data?
            - Found three more datasets – Crawdad, IoT-23, GHOST
        - Alex
            - Update on Scholars Forum Poster
            - Zigbee data – found .pcap files with over 15,000 normal packets
        - Ethan
            - Demonstrated RANSAC test script
    - o Tasks:
        - Cassie and Maria
            - Continue exploring the new datasets that were found, Zigbee and non-Zigbee
            - Continue working on the peak finding algorithms
        - Alex
            - Continue working on Scholars Forum poster and SPIE PowerPoint and presentation
        - Ethan
            - Look into what we need to do for pre-processing data to input to the RANSAC algorithm (what the algorithm expects as input)
- Adjournment at 5:00 p.m.
- Next meeting will be held on TBD, check for email from Dr. Farag (via Dr. Farag Zoom room).

CAE Research Project Meeting Minutes
3/18/2021

- Participants
    - Dr. Waleed Farag (PI)
    - Dr. Xin-Wen Wu (Co-PI)
    - Dr. Soundararajan Ezekiel (Co-PI)
    - Ethan Buhl
    - Cassandra Lefever
    - Alex Robbins
    - Maria Balega
- Meeting starts at 4:30 p.m.
- Unanimously approved minutes for meeting on March 11, 2021
- Main Points
    - Updates:
        - Alex – Update on Scholars Forum and SPIE presentation
        - Cassie and Maria – Update on GHOST IoT dataset
        - Ethan – RANSAC preprocessing with Python
    - Tasks:
        - Alex
            - Finish up SPIE video presentation
            - Dr. Ezekiel will get Alex in touch with a grad student to finish formatting the paper
            - Finish mid-semester report
            - Help Cassie with converting the datasets to CSV
        - Cassie and Maria
            - Continue investigating datasets, contact authors, work with Alex to address issues
        - Ethan
            - Continue with RANSAC, need to decide if we should continue using it or move to something else
- Adjournment at 5:15 p.m.
- Next meeting will be held on TBD, check for email from Dr. Farag (via Dr. Farag Zoom room).

CAE Research Project Meeting Minutes

3/25/2021

- Participants
    - o Dr. Waleed Farag (PI)
    - o Dr. Soundararajan Ezekiel (Co-PI)
    - o Dr. Xin-Wen Wu (Co-PI)
    - o Cassandra Lefever
    - o Alex Robbins
    - o Maria Balega
- Meeting starts at 4:30 p.m.
- Unanimously approved minutes for meetings on March 11 and March 18, 2021
- Main Points
    - o Updates:
        - ▪ Alex
            - SPIE paper got submitted and passed quality check
        - ▪ Cassie and Maria
            - Datasets – CRAWDAD, IoT23
    - o Tasks:
        - ▪ Alex
            - Meet with Dr. Ezekiel tomorrow to discuss next steps for the project, get DCNN running
            - Select and clean up two moderate sized files (one benign, one malicious) from the dataset that Cassie shared, IoT23, on Titan
        - ▪ Maria
            - Research the IoT23 dataset a bit more, try to get a clear understanding of it, work with Alex so he can clean up the data
        - ▪ Cassie
            - Tasks on hold until Monday
        - ▪ We want to focus on three of the eight algorithms we talked about so far
            - GAN, autoencoder, DCNN, wavelet, …
            - We also need to pick a dataset to use for each of the three – for now we are using IoT23
    - o April 10th and 17th
        - ▪ We have a workshop, please keep these dates open, more details coming soon
- Adjournment at 5:15 p.m.
- Next meeting will be held on Thursday, April 1, 2021 at 4:30 p.m. via Dr. Farag Zoom room.

CAE Research Project Meeting Minutes
4/1/2021

- Participants
    - o Dr. Waleed Farag (PI)
    - o Dr. Soundararajan Ezekiel (Co-PI)
    - o Dr. Xin-Wen Wu (Co-PI)
    - o Cassandra Lefever
    - o Alex Robbins
    - o Maria Balega
- Meeting starts at 4:30 p.m.
- Unanimously approved minutes for meeting on March 25, 2021
- Main Points
    - o Updates:
        - ▪ Alex – Rescaled Range Analysis
            - • Hurst algorithm in MatLab
            - • Attempted recreation in Python for better graphing
        - ▪ Maria
            - • Update on IoT23 Dataset – millions of IoT captures
        - ▪ Cassie – Wavelets
    - o Tasks:
        - ▪ Alex –
            - • Convert part of the IoT23 dataset to CSV and run it through the h2o algorithms we have
            - • Dr. Farag, Dr. Ezekiel, Dr. Wu, and Alex need to meet with Jackson and Dr. Pearlstein sometime Friday or Saturday to discuss normalization techniques and clustering, we want to narrow down what technique we want to use
    - o April 10th and 17th
        - ▪ We have a workshop, please keep these dates open, more details coming soon
- Adjournment at 5:35 p.m.
- Next meeting will be held on Thursday, April 8, 2021 at 4:30 p.m. via Dr. Farag Zoom room.

CAE Research Project Meeting Minutes
4/8/2021

- Participants
    - Dr. Waleed Farag (PI)
    - Dr. Soundararajan Ezekiel (Co-PI)
    - Dr. Xin-Wen Wu (Co-PI)
    - Cassandra Lefever
    - Alex Robbins
    - Maria Balega
- Meeting started at 5:20 p.m.
- Unanimously approved minutes for meeting on April 1, 2021
- Main Points
    - Updates:
        - Alex –
            - H2o on IoT23 dataset
    - Tasks:
        - Alex
            - Try using kmeans on Matlab, see the cluster visually, see if any errors are produced, when you get a result email the professors
        - Cassie and Maria –
            - Understand the dataset more, go back to the source of IoT23 and try to get a better understanding, re-visit the COMNETs one, search for backup labeled datasets
    - BC3 Workshop - April 10th and 17th
        - Details - https://www.iup.edu/cybersecurity/grants/dod-capacity-building-project/bc3-workshop/
        - 8:45 a.m. – 3:15 p.m. on both days
        - Sessions will mostly be instructor-led with some time in breakout rooms, very similar to GenCyber
        - We will be using Slack for communicating during the event
        - Zoom link: https://iupvideo.zoom.us/j/94427492724
- Adjournment at 6:00 p.m.
- Next meeting will be held on Thursday, April 15, 2021 at 4:30 p.m. via Dr. Farag Zoom room.

CAE Research Project Meeting Minutes

4/14/2021

- Participants
    - o Dr. Waleed Farag (PI)
    - o Dr. Soundararajan Ezekiel (Co-PI)
    - o Dr. Xin-Wen Wu (Co-PI)
    - o Cassandra Lefever
    - o Alex Robbins
    - o Maria Balega
- Meeting started at 4:30 p.m.
- Unanimously approved minutes for meeting on April 8, 2021
- Main Points
    - o Updates:
        - ▪ Alex –
            - • Update on k-means in MatLab, cleaned up dataset
        - ▪ Cassie and Maria
            - • Updates on IoT23
    - o Tasks:
        - ▪ Alex –
            - • Think about how we can convert the data into a meaningful form
            - • Meet with Dr. Wu and Dr. Ezekiel at 7:00 p.m. on Thursday to discuss the data
        - ▪ Maria –
            - • Get room access from Dr. Ezekiel to access Titan, try updating TeamViewer to solve the freezing problem
        - ▪ Cassie –
            - • Send the dataset and all notes on it to professors and Dr. Pearlstein
        - ▪ We will be meeting with Dr. Pearlstein on Friday from 4:00 p.m. to 5:00 p.m.
    - o BC3 Workshop - April 17th
        - ▪ Details - https://www.iup.edu/cybersecurity/grants/dod-capacity-building-project/bc3-workshop/
        - ▪ 8:45 a.m. – 3:15 p.m.
        - ▪ Sessions will mostly be instructor-led with some time in breakout rooms, very similar to GenCyber
        - ▪ We will be using Slack for communicating during the event
        - ▪ Zoom link: https://iupvideo.zoom.us/j/94427492724
- Adjournment at 5:40 p.m.
- Next meeting will be held on Thursday, April 22, 2021 at 4:30 p.m.

CAE Research Project Meeting Minutes

4/29/2021

- Participants
    - o Dr. Waleed Farag (PI)
    - o Dr. Soundararajan Ezekiel (Co-PI)
    - o Dr. Xin-Wen Wu (Co-PI)
    - o Cassandra Lefever
    - o Alex Robbins
    - o Maria Balega
- Meeting started at 4:30 p.m.
- Unanimously approved minutes for meeting on April 14, 2021
- Main Points
    - o Updates:
        - Alex and Maria
            - Update on k-means clustering for IoT23 datasets
            - Starting with ports for now, working on one-hot encoding for network addresses and other columns
        - Cassie
            - Update on Zeek Flow documentation
    - o Tasks:
        - General comment – We should keep track of computational complexity when analyzing data (actual time to compute, CPU cycles, etc.) for future reports
        - We will be meeting with Dr. Pearlstein sometime to discuss one-hot encoding, we will be implementing our own one-hot encoding script after finals
        - Alex –
            - Remove clusters with amount of flows below 100 and re-run the bad analyses without those clusters, see if results improve
        - Maria-
            - Download full IoT23 dataset onto Titan, make a backup on one of Dr. Ezekiel's hard drives
- Adjournment at 5:45 p.m.
- Next meeting will be held May 6. Use DR. Farag Zoom link to join the meeting.

CAE Research Project Meeting Minutes

5/12/2021

- Participants
    - o Dr. Waleed Farag (PI)
    - o Dr. Soundararajan Ezekiel (Co-PI)
    - o Dr. Xin-Wen Wu (Co-PI)
    - o Cassandra Lefever
    - o Maria Balega
    - o Alex Robbins
- Meeting started at 1:00 p.m.
- Unanimously approved minutes for meeting on April 29, 2021
- Main Points
    - o Tasks:
        - One-hot encoding on IP addresses and Zeek history labels
        - Set up DNN/DCNN
            - 1 input layer
            - Start with 2 fully convolutional layers
            - 3 fully connected layers
        - Support Vector Machine (SVM)
        - Extreme Gradient Boost (EGB)
        - Results should be added to the OneDrive folder labeled "DCNN Results" in Dr. Farag's OneDrive, email group with results
        - Maria – use a flash drive to upload dataset onto Titan machine
    - o Upcoming Workshops
        - May 17th and 18th
        - June 1st and 2nd
        - See emails from Dr. Farag for details
- Adjournment at 1:50 p.m.
- Next meeting will be held on Wednesday April 19th at noon via Dr. Farag's Zoom room

CAE Research Project Meeting Minutes

5/19/2021

- Participants
  - Dr. Waleed Farag (PI)
  - Dr. Soundararajan Ezekiel (Co-PI)
  - Dr. Xin-Wen Wu (Co-PI)
  - Cassandra Lefever
  - Maria Balega
  - Malcom Holden
  - Alicia Deak
- Meeting started at 12:00 p.m.
- Unanimously approved minutes for meeting on May 12, 2021
- Main Points
  - Discussed
    - One Hot Encoding
    - DCNN
    - SVM
    - Full dataset uploaded on Titan
  - Tasks:
    - Cassie – Continue with One Hot Encoding, learn from Alex
    - Maria – Data visualization, EGB, learn from Alex
    - Malcolm - Talk to Alex and SVM
    - Alex – Continue with DCNN
- Adjournment at 1 p.m.
- Next meeting will be held on Wednesday May 26th at noon via Dr. Farag's Zoom room

CAE Research Project Meeting Minutes
5/26/2021

- Participants
    - Dr. Waleed Farag (PI)
    - Dr. Soundararajan Ezekiel (Co-PI)
    - Dr. Xin-Wen Wu (Co-PI)
    - Cassandra Lefever
    - Maria Balega
    - Alicia Deak
    - Alex Robbins
- Meeting started at 12:00 p.m.
- Unanimously approved minutes for meeting on May 19, 2021
- Main Points
    - NCC Workshop
        - June 1st and 2nd 8:45 a.m. – 3 p.m.
        - See email from Dr. Farag for details
    - Discussed:
        - Alicia – Update from research facility in Rome, NY
            - Anomaly detection based on machine learning, broken down into different algorithms
                - SVM, DCNN, GAN, autoencoder, etc.
            - Using data from satellites for the research - https://www.celestrak.com/NORAD/elements/
            - Currently in the preprocessing stage, having trouble because the dataset is unique (two lines, 14 groupings, nothing is labeled)
            - Getting help from Adam
        - Malcom –
            - Will be back from vacation on Monday night
            - Worked with Alicia to get the basis of the SVM code and its RBF function
        - Alex and Cassie –
            - Worked on DCNN, input data is ready, getting an error in the function
        - Maria – Extreme Gradient Boosting, notes and an example program on the Titan using sample Iris dataset

    - Tasks:
        - Alicia –
            - Help teach SVM and GAN to Cassie, Maria, and Malcom after a model is built
        - Malcom –
            - Will continue to work with Alicia
        - Alex and Cassie –
            - Work with Alicia and Adam to troubleshoot the error in the DCNN

- - ▪ Maria –
      - Work on getting a portion of the IoT23 dataset running in the xgboost code to see if it can work, then gradually try more of the dataset
- Adjournment at 1:10 p.m.
- Next meeting will be tentatively held on Wednesday June 2nd at 5 p.m. via Dr. Farag's Zoom room

CAE Research Project Meeting Minutes

6/2/2021

- Participants
    - Dr. Waleed Farag (PI)
    - Dr. Soundararajan Ezekiel (Co-PI)
    - Dr. Xin-Wen Wu (Co-PI)
    - Cassandra Lefever
    - Maria Balega
    - Alicia Deak
    - Malcolm Holden
    - Alex Robbins
- Meeting started at 5:00 p.m.
- Unanimously approved minutes for meeting on May 26, 2021
- Main Points
    - Discussed:
        - Maria - Update on XGBoost algorithm
            - Worked with Dr. Ezekiel and Dr. Pearlstein on normalizing the data
            - Started with 200 lines of the data
            - Ran data through code, code ran and got an initial accuracy of 100%, but our testing is not ideal as the training and testing data overlap
        - Malcolm –
            - Working with Alicia, reading her notes, and learning MatLab
        - Alicia – SVM, one-hot encoding, cyclic learning
            - Working on converting our data into images to fit into DCNN
            - Working on satellite datasets, trying to get permission to access another dataset used by the Air Force
            - Shared PowerPoint slides summarizing current work with Starlink data and novelty detection
    - Tasks:
        - Maria –
            - Work with Alicia to run more of the dataset through the XGBoost Algorithm
            - Work on setting the parameters of the XGBoost algorithm
        - Dr. Ezekiel will email a PowerPoint presentation to the students explaining XGBoost in more detail
        - Alicia will train Maria on running the SVM on our dataset
        - Alex –
            - Search for a more traditional DNN for MatLab to use instead
            - Meet with Dr. Ezekiel and Dr. Pearlstein to discuss this
            - Sort and document files created on the Titan
        - Alicia, Jackson, Adam, and Alex will continue to work with DCNN

- One-hot encode the data -> convert to images -> run it through the DCNN
- We want Adam to help us find his original DCNN code on the Titan
  - Cassie -
    - Sort and document files created on the Titan
  - Malcom –
    - Continue working with Alicia on MatLab, SVM
- Adjournment at 6:10 p.m.
- Next meeting will be held on Monday June 7th at 5 p.m. via Dr. Farag's Zoom room

CAE Research Project Meeting Minutes

6/7/2021

- Participants
  - Dr. Waleed Farag (PI)
  - Dr. Soundararajan Ezekiel (Co-PI)
  - Dr. Xin-Wen Wu (Co-PI)
  - Cassandra Lefever
  - Maria Balega
  - Malcolm Holden
  - Alex Robbins
- Meeting started at 5:00 p.m.
- Unanimously approved minutes for meeting on June 2, 2021
- Main Points
  - Discussed:
    - Alex –
      - Rewrote DCNN/DNN MatLab code, organized files on Titan
    - Alicia and Dr. Ezekiel -
      - Worked with Adam on SVM and cyclic learning in Python, running into errors with the large dataset, will work on debugging
    - Maria –
      - Worked with one of the files for the IoT23 dataset, Mirai 34, after running the file through XGBoost achieved an accuracy score close to 100% after trial-and-error with parameters, need to verify this result
      - Also tried using the Mirai 35 file, file is very large and causing some issues
    - Malcolm –
      - Met with Alicia, Alicia was able to troubleshoot her code
    - Cassie –
      - GenCyber folder is up-to-date
    - Malcolm will take minutes for future meetings
  - Tasks:
    - Alex –
      - Meet with Malcom and Maria, discuss progress and the current state of the project, transitioning leadership role primarily to Maria for now
      - Work with Dr. Ezekiel on DCNN/DNN code
    - Maria –
      - Break up Mirai 35 file into 4 or 5 smaller files to put into MatLab, change the ratio of benign : malicious data from 8 : 2 to 9 : 1
      - Document parameters, results, outcome, timing, memory used, etc.
    - Malcolm –
      - Meet with Alicia to learn SVM
      - Meet with Alex to discuss state of DCNN/DNN

- Adjournment at 5:50 p.m.
- Next meeting will be held on Monday June 14th at 5 p.m. via Dr. Farag's Zoom room

CAE Research Project Meeting Minutes

6/14/2021

- Participants
    - Dr. Waleed Farag (PI)
    - Dr. Soundararajan Ezekiel (Co-PI)
    - Dr. Xin-Wen Wu (Co-PI)
    - Maria Balega
    - Alex Robbins
    - Alicia Deak
- Meeting started at 5:00 p.m.
- Unanimously approved minutes for meeting on June 7, 2021
- Main Points
    - Discussed:
        - Maria
            - Ran multiple datasets on xgboost algorithm with different parameters
            - Had some issues normalizing dataset, used Python normalization function to attempt to solve this
            - Ran Mirai34 dataset file again and measured time, changing parameters such as max depth for training
            - Results have been summarized in spreadsheet files
        - Alicia
            - Obtained some results from running SVM on datasets other than Mirai
            - Xgboost is working better for some datasets, SVM is working better for other datasets
    - Tasks:
        - Maria and Alicia
            - Work on TSNE
            - Remember to record CPU time while running algorithms on different datasets
            - Work on running size 10 million dataset file
            - Documentation of work
            - Keep visualization of results in mind (making plots for final report)
- Adjournment at 5:40 p.m.
- Next meeting will be held on Monday June 21st at 5 p.m. via Dr. Farag's Zoom room

CAE Research Project Meeting Minutes

6/21/2021

- Participants
    - Dr. Soundararajan Ezekiel (Co-PI)
    - Dr. Xin-Wen Wu (Co-PI)
    - Maria Balega
    - Alex Robbins
    - Alicia Deak
- Meeting started at 5:00 p.m.
- Unanimously approved minutes for meeting on June 14, 2021
- Main Points
    - Discussed:
        - Maria
            - Worked on preprocessing Mirai35 dataset with 10 million lines
            - Ran the Mirai35 dataset in XGBoost algorithm and got warning message about mixed datatypes
            - Working on fixing the warning and still running it
        - Alicia
            - Looked at different types of distance metrics in TSNE
            - Ran different datasets to compare metrics
    - Tasks:
        - Maria and Alicia
            - Continue working on running Mirai35 dataset
            - Work on TSNE
            - Work on image conversion
    - General Comments:
        - Think about which columns to test for datasets and run multiple samples with different columns to compare results
        - Keep visualization of results in mind
        - Dr. Ezekiel made suggestion to increase students hourly pay and recommends students claim hours
- Adjournment at 6 p.m.
- Next meeting will be held on Monday June 28th at 5 p.m. via Dr. Farag's Zoom room

CAE Research Project Meeting Minutes

6/28/2021

- Participants
  - Dr. Waleed Farag (PI)
  - Dr. Soundararajan Ezekiel (Co-PI)
  - Dr. Xin-Wen Wu (Co-PI)
  - Maria Balega
- Meeting started at 5:00 p.m.
- Unanimously approved minutes for meeting on June 21, 2021
- Main Points
  - Discussed:
    - Maria
      - Ran the Mirai35 dataset with 10 million lines in XGBoost and got accuracy of 99.99%
      - Had issues with missing values as dashes and changed them to blanks
      - Added metrics including precision, F1, and recall
  - Tasks:
    - Maria
      - Run a different IoT23 capture in XGBoost that isn't dominated by DDoS attacks
      - Work with Adam to learn autoencoder
      - Work with Alicia to run datasets in SVM to see comparison of results with XGBoost
      - Meet with new students and train them for GenCyber
- Adjournment at 6 p.m.
- Next meeting will be held on Wednesday July 7th at 12 p.m. via Dr. Farag's Zoom room

CAE Research Project Meeting Minutes

7/7/2021

- Participants
    - Dr. Soundararajan Ezekiel (Co-PI)
    - Dr. Xin-Wen Wu (Co-PI)
    - Maria Balega
    - Zaryn Good
    - Franklin May
    - Alicia Deak
- Meeting started at 12:00 p.m.
- Unanimously approved minutes for meeting on June 28, 2021
- Main Points
    - Discussed:
        - Maria
            - Ran a different capture from the IoT23 dataset (Okiru36) with 13 million lines in XGBoost and got 100% accuracy
            - Tried running dataset in SVM but got memory error
        - Maria and Alicia
            - Ran different datasets through XGBoost, SVM, and t-SNE to see comparison of results, XGBoost had slightly better accuracies than SVM
    - Tasks:
        - Maria and Alicia
            - Continue running datasets in XGBoost, SVM, t-SNE, image conversion, and one-hot encoding
            - Graph results from datasets to compare each algorithm
        - Zaryn and Franklin
            - Meet with Dr. Ezekiel to discuss wavelets
- Adjournment at 12:55 p.m.
- Next meeting TBD via Dr. Farag's Zoom room

CAE Research Project Meeting Minutes

7/19/2021

- Participants
  - Dr. Waleed Farag (PI)
  - Dr. Soundararajan Ezekiel (Co-PI)
  - Dr. Xin-Wen Wu (Co-PI)
  - Maria Balega
  - Zaryn Good
  - Franklin May
  - Alicia Deak
- Meeting started at 12:00 p.m.
- Unanimously approved minutes for meeting on July 7, 2021
- Main Points
  - Discussed:
    - Maria and Alicia
      - Shared comparison of results from smaller datasets that were run through SVM, t-SNE, and XGBoost
      - Ran into memory error when trying to run bigger datasets
    - Zaryn and Franklin
      - Worked on running GAN to compare results with XGBoost, ran into memory error
  - Tasks:
    - Maria and Alicia
      - Talk to Cassie and Alex to identify previous IoT papers and datasets
      - Run IoT datasets through XGBoost, SVM, and t-SNE
      - Record results and store them in shared OneDrive folder
    - Zaryn and Franklin
      - Work on running GAN and image conversion, record results
- Adjournment at 1:10 p.m.
- Next meeting will be held on Monday July 26 at 12 p.m. via Dr. Farag's Zoom room

CAE Research Project Meeting Minutes

7/26/2021

- Participants
    - o Dr. Waleed Farag (PI)
    - o Dr. Soundararajan Ezekiel (Co-PI)
    - o Maria Balega
    - o Zaryn Good
    - o Alicia Deak
- Meeting started at 12:00 p.m.
- Unanimously approved minutes for meeting on July 19, 2021
- Main Points
    - o Discussed:
        - Maria and Alicia
            - Met with Alex to discuss Comnet dataset, not ideal for XGBoost since it is not labeled
            - Ran IoT23 datasets in SVM and XGBoost, got accuracies up to 99.99%
        - Zaryn
            - Working on running datasets through image conversion
    - o Tasks:
        - Maria and Alicia
            - Explore Comnet dataset further, try to find an IoT dataset that is labeled
            - Work on generating a synthetic IoT dataset
            - Run IoT23 datasets in XGBoost and SVM without binary classification
        - Zaryn
            - Continue working on image conversion
        - Franklin
            - Continue working on GAN
- Adjournment at 12:45 p.m.
- Next meeting TBD via Dr. Farag's Zoom room

CAE Research Project Meeting Minutes
8/10/2021

- Participants
    - Dr. Waleed Farag (PI)
    - Dr. Soundararajan Ezekiel (Co-PI)
    - Maria Balega
    - Zaryn Good
    - Alicia Deak
    - Franklin May
- Meeting started at 12:00 p.m.
- Unanimously approved minutes for meeting on July 26, 2021
- Main Points
    - Discussed:
        - Maria and Alicia
            - Ran IoT23 datasets in SVM and XGBoost without binary classification, got accuracies up to 99.99%
            - Looking into other IoT datasets
        - Zaryn
            - Ran various datasets through image conversion
        - Franklin
            - Ran various datasets using SVM, XGBoost, t-SNE and averaged the results
    - Tasks:
        - Maria and Alicia
            - Analyze IoT23 datasets and XGBoost code, data samples are not balanced
            - Continue searching for a labeled IoT dataset with publications to compare results
            - Work on generating a synthetic IoT dataset
        - Zaryn
            - Work on running IoT23 datasets through image conversion
            - Get script for histogram
        - Franklin
            - Work on running IoT23 datasets through SVM, XGBoost, t-SNE and averaging results
- Adjournment at 12:45 p.m.
- Next meeting will be held on Monday August 16 at 12 p.m. via Dr. Farag's Zoom room

CAE Research Project Meeting Minutes

8/16/2021

- Participants
    - Dr. Waleed Farag (PI)
    - Dr. Soundararajan Ezekiel (Co-PI)
    - Maria Balega
    - Zaryn Good
    - Alicia Deak
    - Franklin May
    - Ryan Gress
- Meeting started at 12:00 p.m.
- Unanimously approved minutes for meeting on August 10, 2021
- Main Points
    - Discussed:
        - Maria
            - Looking into Deep Belief Network (DBN), working on coding the algorithm
            - Found a paper using IoT23 dataset that we can compare our results with
        - Alicia
            - Working on creating synthetic data
        - Zaryn
            - Working on image conversion, has the code running and samples, need to run on titan since it takes hours
        - Franklin
            - Looking into DGCNN, results are not as good as DBN, not going to focus on this algorithm
    - Tasks:
        - Maria
            - Continue working on DBN
            - Work on preprocessing IoT23 data similar to the published paper to compare results
        - Alicia
            - Continue working on creating synthetic data
        - Ryan
            - Look into 4 other IoT datasets from the published IoT23 paper
        - Franklin
            - Working on averaging IoT23 datasets using XGBoost, SVM, etc.
            - Run IoT datasets from published paper using XGBoost, SVM, etc.
        - Zaryn
            - Continue working on image conversion using IoT23 data
            - Work with Jackson to run DCNN on titan
- Adjournment at 1 p.m.

- Next meeting will be held on Monday August 23 at 11:30 a.m. via Dr. Farag's Zoom room

CAE Research Project Meeting Minutes

8/23/2021

- Participants
    - Dr. Waleed Farag (PI)
    - Dr. Soundararajan Ezekiel (Co-PI)
    - Maria Balega
    - Zaryn Good
    - Alicia Deak
    - Franklin May
    - Ryan Gress
    - Cassie Lefever
    - Eliot Hocke
- Meeting started at 11:30 a.m.
- Unanimously approved minutes for meeting on August 16, 2021
- Main Points
    - Discussed:
        - Maria
            - Working on DBN algorithm, having issues with code/ data, DBN works best with images
            - Working on preprocessing the IoT23 data similar to the published paper
        - Ryan
            - Found 2 datasets used from the published paper
        - Zaryn
            - Met with Jackson and got DCNN code, converted images and is planning to run them in DCNN
        - Franklin
            - Started working on preprocessing the KDD dataset, having issues with the code
    - Tasks:
        - Maria & Cassie
            - Continue working on DBN
            - Work on preprocessing IoT23 data and running it in XGBoost, SVM, etc.
        - Alicia
            - Continue working on creating synthetic data
        - Ryan & Franklin
            - Look further into the datasets found and keep looking for others
            - Continue preprocessing the KDD dataset and run datasets in XGBoost, SVM, etc.
        - Zaryn
            - Run images through DCNN using binary classification
            - Compare results with XGBoost and other algorithms
- Adjournment at 12:15 p.m.

- Next meeting will be held on Thursday September 2 at 3:45 p.m. via Dr. Farag's Zoom room

CAE Research Project Meeting Minutes

9/2/2021

- Participants
  - Dr. Waleed Farag (PI)
  - Dr. Soundararajan Ezekiel (Co-PI)
  - Dr. Xinwen Wu
  - Maria Balega
  - Zaryn Good
  - Alicia Deak
  - Franklin May
  - Ryan Gress
  - Cassie Lefever
  - Eliot Hocke
  - Sky Semone
- Meeting started at 3:45 p.m.
- Unanimously approved minutes for meeting on August 23, 2021
- Main Points
  - Discussed:
    - Maria & Cassie
      - Preprocessed all the individual captures from IoT23 dataset, working on combining captures together
    - Alicia
      - DBN had high error rate, its best for image datasets
      - Created a synthetic dataset using Avocado data
    - Zaryn
      - Ran into issues with DCNN running, wasn't reading images correctly
      - Working with Jackson to convert code to fix images
    - Franklin
      - Started implementing confusion matrix to make sure data isn't skewed
    - Ryan
      - Looked further into datasets, found another possible dataset that has a published paper for comparison
  - Tasks:
    - Maria & Cassie
      - Continue preprocessing IoT23 dataset similar to the published paper and run it in XGBoost
    - Alicia
      - Try running images in DBN
      - Work on synthetic dataset using IoT23 data
    - Zaryn & Eliot
      - Continue working on fixing the image issue, work with Sky on DCNN
    - Franklin

- Continue working on confusion matrix
  - Ryan
    - Continue looking into datasets and find published paper so we can compare our results
  - General tasks:
    - Work on abstracts and continue revising them
      - Franklin is working on DCNN abstract
      - Maria is working on XGBoost abstract
    - Back up data on titan
- Adjournment at 4:50 p.m.
- Next meeting will be held on Thursday September 9 at 3:45 p.m. via Dr. Farag's Zoom room

CAE Research Project Meeting Minutes
9/9/2021

- Participants
  - Dr. Waleed Farag (PI)
  - Dr. Soundararajan Ezekiel (Co-PI)
  - Dr. Xinwen Wu
  - Maria Balega
  - Zaryn Good
  - Franklin May
  - Ryan Gress
  - Cassie Lefever
- Meeting started at 3:55 p.m.
- Unanimously approved minutes for meeting on September 2, 2021
- Main Points
  - Discussed:
    - Maria & Cassie
      - Ran into issues combining and splitting the IoT23 captures, getting memory error
      - Working on XGBoost abstract
    - Zaryn
      - Ran IoT23 data converted to images through DCNN, got 60-70% accuracy
    - Franklin
      - Added confusion matrix into XGBoost and SVM algorithms
      - Rewrote preprocessing for KDD dataset
      - Working on DCNN abstract
    - Ryan
      - Found published papers to compare our results with
      - Looking into TON-IoT dataset
  - Tasks:
    - Maria & Cassie
      - Continue preprocessing IoT23 dataset similar to the published paper and run it in XGBoost
      - Try combining as many csv as possible to avoid memory error, try using MATLAB to preprocess to determine if it's a language or hardware issue
    - Alicia
      - Try running images in DBN
      - Work on synthetic dataset using IoT23 data
    - Zaryn
      - Work on running the full IoT23 through various DCNN, record all results
    - Franklin
      - Continue preprocessing KDD and run it in XGBoost and SVM

- Ryan
  - Work with Maria and Cassie to see if the TON-IoT dataset can work with XGBoost
  - Work with Zaryn to see if TON-IoT dataset can work with DCNN
- General tasks:
  - Work on abstracts and continue revising them
    - Franklin is working on DCNN abstract
    - Maria is working on XGBoost abstract
- Adjournment at 4:55 p.m.
- Next meeting will be held on Thursday September 16 at 3:15 p.m. via Dr. Farag's Zoom room

CAE Research Project Meeting Minutes

9/16/2021

- Participants
    - Dr. Waleed Farag (PI)
    - Dr. Soundararajan Ezekiel (Co-PI)
    - Maria Balega
    - Zaryn Good
    - Franklin May
    - Ryan Gress
    - Cassie Lefever
    - Alicia Deak
    - Eliot Hocke
    - Sky Semone
- Meeting started at 3:10 p.m.
- Unanimously approved minutes for meeting on September 9, 2021
- Main Points
    - Discussed:
        - Maria & Cassie
            - Tried to combine files using MATLAB, still getting memory error
            - Could only combine 6 csv files in python, not reading all classifications from the captures when running XGBoost
        - Alicia
            - Working on synthetic data
        - Zaryn
            - Running the Mirai35 through Alex Net, averaging 80% accuracy, focusing on running each capture at time is giving better results
        - Franklin
            - Ran 20% of KDD dataset in XGBoost and SVM using binary classification, achieved 95% accuracy in XGBoost, achieved 93% accuracy in SVM
        - Ryan
            - Ran a capture from TON-IoT dataset in XGBoost using binary classification, achieved 78% accuracy
            - Working with Zaryn to run DCNN
    - Tasks:
        - Maria & Cassie
            - Work on fixing the memory error, try to take classification label separate and concatenate after that when preprocessing
        - Alicia & Eliot
            - Convert IoT23 csv files to dat or txt file to see if it uses less memory
            - Continue working on synthetic data
        - Zaryn
            - Continue running IoT datasets in DCNN

- Franklin
  - Run full KDD dataset in XGBoost and SVM
- Ryan
  - Run other captures from TON-IoT dataset in XGBoost
  - Reformat the TON-IoT data to run through DCNN
- General tasks:
  - Work on abstracts and continue revising them
    - Franklin is working on DCNN abstract
    - Maria is working on XGBoost abstract
- Adjournment at 4:00 p.m.
- Next meeting will be held on Wednesday September 22 at 3:15 p.m. via Dr. Farag's Zoom room

CAE Research Project Meeting Minutes
9/23/2021

- Participants
  - Dr. Waleed Farag (PI)
  - Dr. Soundararajan Ezekiel (Co-PI)
  - Dr. Xinwen Wu
  - Maria Balega
  - Zaryn Good
  - Franklin May
  - Cassie Lefever
  - Alicia Deak
  - Eliot Hocke
- Meeting started at 4:00 p.m.
- Unanimously approved minutes for meeting on September 16, 2021
- Main Points
  - Discussed:
    - Maria & Cassie
      - Got 19 captures from IoT23 data preprocessed using MATLAB, split into 6 csv files and running them in XGBoost
    - Alicia & Eliot
      - Converted all IoT23 csv files to dat extension, synthetic data is ready
    - Zaryn
      - Waiting for DCNN files and code from Jackson to run data
      - DCNN is having trouble finding malicious, only getting 50% accuracy
    - Franklin
      - Ran full KDD in XGBoost with nonbinary classification, achieved up to 98% accuracy, having trouble with SVM because of dataset size
  - Tasks:
    - Maria & Cassie
      - Run preprocessed IoT23 data in XGBoost, work on accessing Dr. Pearlstein's machine
    - Alicia & Eliot
      - Run through .dat files to see how much memory it uses, generate synthetic data subset and run it in XGBoost, SVM, etc.
    - Zaryn
      - Continue running DCNN with Jackson, look into the why the accuracy for anomalies is low
    - Franklin
      - Try running KDD full dataset in SVM
    - Ryan
      - Run other captures from TON-IoT dataset in XGBoost
      - Reformat the TON-IoT data to run through DCNN

- o General tasks:
    - Work on abstracts and continue revising them
        - Franklin is working on DCNN abstract
        - Maria is working on XGBoost abstract
- Adjournment at 4:50 p.m.
- Next meeting will be held on Thursday September 30 at 3:00 p.m. via Dr. Farag's Zoom room

CAE Research Project Meeting Minutes

9/30/2021

- Participants
  - o Dr. Waleed Farag (PI)
  - o Dr. Soundararajan Ezekiel (Co-PI)
  - o Dr. Xinwen Wu
  - o Maria Balega
  - o Zaryn Good
  - o Franklin May
  - o Cassie Lefever
  - o Alicia Deak
  - o Eliot Hocke
  - o Ryan Gress
  - o Dr. Larry Pearlstein
- Meeting started at 3:00 p.m.
- Unanimously approved minutes for meeting on September 23, 2021
- Main Points
  - o Discussed:
    - Maria & Cassie
      - Got one fourth of a merged IoT23 csv file to run in XGBoost, achieved up to 99% accuracy but two classifications were missing so this may have affected results
      - Working with Dr. Pearlstein to connect to TCNJ's machine
    - Alicia & Eliot
      - Found that dat files were same size as csv, running synthetic data through various algorithms and its performing better than original data
    - Zaryn
      - Got the DCNN code from Jackson
    - Franklin
      - Achieved up to 98% accuracy in SVM when running full KDD dataset
  - o Tasks:
    - Maria & Cassie
      - Continue running more of the split IoT23 data in XGBoost, try to get all nine classifications, try running in local machine or Linux machine
    - Alicia & Eliot
      - Run dat files through XGBoost
      - Continue working with synthetic data to compare it against original data
    - Zaryn
      - Run the new code in DCNN, continue to look into why the accuracy for anomalies is low
    - Franklin
      - Work with OneVSRest and OneVSOne classifier in SVM

- Ryan
  - Run other captures from TON-IoT dataset in XGBoost
  - Reformat ToN-IoT data to run through DCNN
  - General tasks:
    - Work on abstracts and continue revising them
      - Franklin is working on DCNN abstract
      - Maria is working on XGBoost abstract
- Adjournment at 4:00 p.m.
- Next meeting will be held on Thursday October 7 at 4:00 p.m. via Dr. Farag's Zoom room

CAE Research Project Meeting Minutes

10/7/2021

- Participants
    - Dr. Waleed Farag (PI)
    - Dr. Soundararajan Ezekiel (Co-PI)
    - Dr. Xinwen Wu
    - Maria Balega
    - Zaryn Good
    - Franklin May
    - Cassie Lefever
    - Alicia Deak
    - Eliot Hocke
    - Ryan Gress
- Meeting started at 4:00 p.m.
- Unanimously approved minutes for meeting on September 30, 2021
- Main Points
    - Discussed:
        - Maria & Cassie
            - Got access to TCNJ machine and Redhawk machine at IUP
            - Ran through the five IoT23 files in XGBoost, all had 9 classifications except one, achieved up to 99% accuracy, data distribution is uneven
        - Alicia & Eliot
            - Ran the dat file through XGBoost to see if it would run better than csv, getting stuck in an endless loop
            - Can create synthetic data to accommodate nan values in datasets
        - Zaryn
            - Running DCNN, training is taking a while
        - Franklin
            - Running KDD dataset in SVM, getting an endless loop
            - Added histogram in XGBoost algorithm
        - Ryan
            - Ran TON-IoT in XGBoost using binary classification, 14% anomalous in each capture, achieved up to 98% accuracy, max_depth was 100 so it may be overfitting data
    - Tasks:
        - Maria & Cassie
            - Run the five IoT23 files in SVM to compare results
            - Look into published papers for comparing results
        - Alicia & Eliot
            - Continue to run dat files through XGBoost and solve endless loop error
            - Find and run a balanced dataset in XGBoost to see results
        - Zaryn

- Continue working on DCNN with IoT23 dataset, start working on converting TON-IoT dataset to images
  - Franklin
    - Run a program in titan and TCNJ machine and compare CPU time
    - Continue working with SVM trying to solve endless loop
  - Ryan
    - Continue running the TON-IoT captures in XGBoost, try to stay below 10 on max_depth
- General tasks:
  - Work on abstracts and continue revising them
    - Franklin is working on DCNN abstract
    - Maria is working on XGBoost abstract
- Adjournment at 5:10 p.m.
- Next meeting will be held on Thursday October 14 at 3:00 p.m. Via Dr.Farag's Zoom Room

CAE Research Project Meeting Minutes

10/14/2021

- Participants
  - Dr. Waleed Farag (PI)
  - Dr. Soundararajan Ezekiel (Co-PI)
  - Dr. Xinwen Wu
  - Maria Balega
  - Zaryn Good
  - Franklin May
  - Cassie Lefever
  - Alicia Deak
  - Ryan Gress
  - Sky Semone
- Meeting started at 3:00 p.m.
- Unanimously approved minutes for meeting on October 7, 2021
- Main Points
  - Discussed:
    - Maria & Cassie
      - Achieved up to 89% accuracy running SVM with IoT23, issues with NaN values so these rows were dropped, lost a lot of data and classifications
    - Alicia & Eliot
      - Working with dat files, only classifying one label in algorithms
      - Solved endless loop error in python, issue was in the data itself
      - Ran two balanced datasets, with binary SVM performed better achieving up to 92% accuracy, with multiclass XGBoost performed better achieving accuracy in the high 90s
    - Zaryn
      - DCNN is running more efficiently on Titan, found that malicious data is getting poor accuracy due to sorting of images
    - Franklin
      - Ran XGBoost on Titan and TCNJ machine to compare CPU times, TCNJ is faster
    - Ryan
      - Ran TON-Iot data in XGBoost, got up to 89% accuracy, XGBoost was mostly better than other algorithms used in published paper
  - Tasks:
    - Maria & Cassie
      - Work on writing report using IEEE format, create graph/table of results
    - Alicia & Eliot
      - Look into comparing SVM, XGBoost, etc. with classical methods
    - Zaryn & Sky
      - Continue working on DCNN

- Franklin
  - Work on running KDD in SVM
  - Look into the time unit for CPU comparison of TCNJ and Titan
- Ryan
  - Work on combining captures from TON-IoT dataset, run binary and multiclassification to compare results with published paper
  - Get XGBoost results from Franklin to compare KDD and TON-IoT
- Adjournment at 4:00 p.m.
- Next meeting will be held on Thursday October 21 at 3:00 p.m.
-

CAE Research Project Meeting Minutes

10/21/2021

- Participants
    - o Dr. Waleed Farag (PI)
    - o Dr. Soundararajan Ezekiel (Co-PI)
    - o Dr. Xinwen Wu
    - o Maria Balega
    - o Zaryn Good
    - o Franklin May
    - o Cassie Lefever
    - o Ryan Gress
    - o Sky Semone
- Meeting started at 4:00 p.m.
- Unanimously approved minutes for meeting on October 14, 2021
- Main Points
    - o Discussed:
        - ▪ Maria & Cassie
            - • Working on writing report, created a graph and tables of results
            - • Working to even out the distribution of anomalies using 75% benign and 25% malicious for IoT23 dataset, SVM accuracy went down
        - ▪ Zaryn & Sky
            - • Ran IoT23 captures through VGG-19 using binary classification, achieved validation accuracy up to 98-99%
            - • Working on training data using ResNet-101
        - ▪ Franklin
            - • Compared CPU time with TCNJ and Titan, Titan was faster
            - • Rewrote KDD preprocessing, currently running SVM
        - ▪ Ryan
            - • Combined TON-IoT captures together similar to published paper and ran through XGBoost, achieved up to 94% accuracy
    - o Tasks:
        - ▪ Maria & Cassie
            - • Continue working on writing report and evening out the distribution of anomalies
        - ▪ Alicia & Eliot
            - • Look into comparing SVM, XGBoost, etc. with classical methods
        - ▪ Zaryn & Sky
            - • Continue working with VGG-19 and ResNet-101, try using TCNJ machine to see if performance improves
        - ▪ Franklin
            - • Run a few more scripts in TCNJ and Titan to compare CPU time
            - • Continue running SVM

- Ryan
  - Work on running each capture and the combined TON-IoT dataset in XGBoost using multiclassification
- Adjournment at 5:00 p.m.
- Next meeting will be held on Thursday October 28 at 3:00 p.m.

CAE Research Project Meeting Minutes

10/28/2021

- Participants
    - o Dr. Waleed Farag (PI)
    - o Dr. Soundararajan Ezekiel (Co-PI)
    - o Dr. Xinwen Wu
    - o Maria Balega
    - o Zaryn Good
    - o Franklin May
    - o Ryan Gress
    - o Sky Semone
    - o Alicia Deak
- Meeting started at 4:00 p.m.
- Unanimously approved minutes for meeting on October 21, 2021
- Main Points
    - o Discussed:
        - Maria & Cassie
            - Started writing report
            - Split IoT-23 data into 60% benign and 40% malicious and created synthetic data, achieved accuracies up to 99% with XGBoost and SVM
        - Zaryn & Sky
            - Getting accuracy up to 100% with DCNN
        - Franklin
            - Achieved higher accuracies with SVM than XGBoost using KDD
        - Ryan
            - Running multiclassification of ToN-IoT captures in XGBoost, achieving up to 92% accuracy
    - o Tasks:
        - Maria & Cassie
            - Continue working on report, look further into SVM results
            - Run the same dataset into two XGBoost algorithms to verify they are working correctly
        - Alicia
            - Look into the algorithms to compare them and the datasets to try and find issues, create visualizations
        - Zaryn & Sky
            - Work on image conversion of KDD, look further into DCNN results
        - Franklin
            - Continue working on abstract, try running KDD through SVM in Matlab to compare results
        - Ryan
            - Continue working on multiclassification for the ToN-IoT captures

- Adjournment at 5:00 p.m.
- Next meeting will be held on Wednesday November 3 at 3:00 p.m.

CAE Research Project Meeting Minutes

11/2/2021

- Participants
  - o Dr. Waleed Farag (PI)
  - o Dr. Soundararajan Ezekiel (Co-PI)
  - o Dr. Xinwen Wu
  - o Maria Balega
  - o Zaryn Good
  - o Franklin May
  - o Ryan Gress
  - o Sky Semone
  - o Alicia Deak
  - o Cassie Lefever
- Meeting started at 3:00 p.m.
- Unanimously approved minutes for meeting on October 28, 2021
- Main Points
  - o Discussed:
    - Maria & Cassie
      - Ran two XGBoost algorithms for comparison, achieved 99.99% and 100% accuracies
    - Alicia
      - Looking into the datasets and algorithms to find issues
    - Zaryn & Sky
      - Working on image conversion of KDD and ToN-IoT datasets
      - Created guidelines to use for preprocessing
    - Franklin
      - Working on writing SVM code in Matlab
    - Ryan
      - Ran ToN-IoT using multiclassification, having an issue of getting the same results
  - o Tasks:
    - Maria & Cassie
      - Split IoT-23 data into 50% benign and 50% malicious, calculate different errors (MAE, log loss function), continue running SVM
    - Alicia
      - Continue looking into datasets and algorithms
      - Work on report for NSA, work with others to organize folders
    - Zaryn & Sky
      - Work on writing XGBoost algorithm in Matlab
      - Continue working on image conversion and running DCNN
    - Franklin
      - Continue working on writing code for SVM in Matlab

- ▪ Ryan
  - Continue running ToN-IoT dataset in XGBoost, fix seed issues
- Adjournment at 3:50 p.m.
- Next meeting will be held on Thursday November 11 at 3:00 p.m.

CAE Research Project Meeting Minutes

11/11/2021

- Participants
    - o Dr. Soundararajan Ezekiel (Co-PI)
    - o Zaryn Good
    - o Franklin May
    - o Ryan Gress
    - o Sky Semone
    - o Alicia Deak
    - o Cassie Lefever
- Meeting started at 3:00 p.m.
- Unanimously approved minutes for meeting on November 2, 2021
- Main Points
    - o Discussed:
        - ▪ Maria & Cassie
            - Remerged all the CSV files for the IoT23 dataset.
            - Restarted preprocessing for IoT23 dataset.
        - ▪ Alicia & Sky
            - Rearranged content in the Titan.
                - o Data files are being moved into Storage 6 to save space on Titan.
            - Gathering content for report.
                - o Discussed writing the entire report, which will be sent to Dr. Farag when done.
        - ▪ Zaryn & Sky
            - Increase the speed of data to image conversion to factor of 30.
            - Converted more data to images.
            - Ran DCNN on KDD to compare to XGBoost.
            - Reviewed Matlab codes with Dr. Ezekiel.
        - ▪ Franklin
            - SVM is mostly working but there has not been much time to run it on the Titan since it has been busy.
        - ▪ Ryan
            - Found what the issue was with why he was getting the same results with XGBoost on ToN-IoT dataset.
    - o Tasks:
        - ▪ Maria & Cassie
            - Put documentation within your code so that anyone else could look at it and understand it.
            - Finish preprocessing IoT23 dataset and run through XGBoost to get results.
        - ▪ Alicia

- Create one XGBoost file that everyone will use to process their results.
- Continue to clean up the Titan and the report
    - Zaryn & Sky
        - Put documentation within your code so that anyone else could look at it and understand it.
        - Discuss DCNN and goals with Franklin for the abstract.
    - Franklin
        - Put documentation within your code so that anyone else could look at it and understand it.
        - KDD dataset will be rerun through the new XGBoost script.
        - Clean up code.
    - Ryan
        - Put documentation within your code so that anyone else could look at it and understand it.
        - Take your code from your personal machine and put it onto the Titan
- Adjournment at 3:45 p.m.
- Next meeting will be held on Thursday November 18 at 3:00 p.m.

CAE Research Project Meeting Minutes

11/16/2021

- Participants
  - Dr. Waleed Farag (PI)
  - Dr. Soundararajan Ezekiel (Co-PI)
  - Dr. Xinwen Wu
  - Zaryn Good
  - Franklin May
  - Ryan Gress
  - Sky Semone
  - Cassie Lefever
  - Maria Balega
- Meeting started at 3:00 p.m.
- Unanimously approved minutes for meeting on November 11, 2021
- Main Points
  - Discussed:
    - Maria & Cassie
      - Working on preprocessing the IoT-23 dataset
    - Zaryn & Sky
      - Testing KDD for different test/train ratios, accuracies are decreasing
      - Updated image conversion
    - Franklin
      - Cleaned up KDD preprocessing, running it in XGBoost and SVM, got 99% for SVM which may be from skewed data
    - Ryan
      - Cleaned up TON-IoT preprocessing, ran through XGBoost function that Alicia created and got 99% accuracy which is different than usual
  - Tasks:
    - Maria & Cassie
      - Finish preprocessing IoT-23 dataset and run through XGBoost
      - Create graph of results for writeup
    - Alicia
      - Continue working on report and XGBoost function
    - Zaryn & Sky
      - Continue running datasets in DCNN
    - Franklin
      - Continue working on running KDD in XGBoost and SVM, look into execution time
      - Work on DCNN abstract with Zaryn
    - Ryan
      - Look into XGBoost function and figure out why new results differ from the original

- Adjournment at 3:50 p.m.
- Next meeting will be held on Tuesday November 23 at 3:00 p.m.

CAE Research Project Meeting Minutes

11/23/2021

- Participants
  - o Dr. Waleed Farag (PI)
  - o Dr. Soundararajan Ezekiel (Co-PI)
  - o Dr. Xinwen Wu
  - o Zaryn Good
  - o Ryan Gress
  - o Sky Semone
  - o Cassie Lefever
  - o Maria Balega
  - o Alicia Deak
- Meeting started at 3:00 p.m.
- Unanimously approved minutes for meeting on November 16, 2021
- Main Points
  - o Discussed:
    - General Comments:
      - Issues connecting to TCNJ machine, Maria contacted Dr. Pearlstein
      - Submit updated results and figures for writeup
      - Validate algorithms and frameworks
    - Maria & Cassie
      - Finished preprocessing the IoT-23 dataset
      - XGBoost split 80/20 achieved 85% accuracy, 65/35 split achieved 99% accuracy, 75/25 split achieved 95% accuracy
    - Alicia
      - Changed weighting on metrics in XGBoost algorithm
      - Ran KDD in XGBoost, an extra column was created in preprocessing, with extra column it is performing 80% accuracies and without it is performing 60% accuracies
    - Zaryn & Sky
      - Compared KDD dataset with different architectures, best performances with higher train to test ratio, ResNet50 and VGG19 outperformed
    - Ryan
      - Fixed issues with preprocessing ToN-IoT dataset
      - Ran XGBoost algorithm, it's not finding malicious activity
  - o Tasks:
    - Maria & Cassie
      - Update report with graphs, results, and abstract
      - Determine what to do with classifications that have low representation
    - Alicia & Franklin
      - Look further into KDD dataset and why there is an extra column
      - Update report with graphs, results, and abstract

- Zaryn & Sky
  - Work on image conversion and training for IoT23 dataset
  - Update report with graphs and results
- Ryan & Alicia
  - Look further into why XGBoost is not finding malicious activity
  - Update report with graphs and results
- Adjournment at 3:50 p.m.
- Next meeting will be held on Tuesday November 30 at 3:00 p.m.

CAE Research Project Meeting Minutes

11/30/2021

- Participants
    - o Dr. Waleed Farag (PI)
    - o Dr. Soundararajan Ezekiel (Co-PI)
    - o Dr. Xinwen Wu
    - o Zaryn Good
    - o Ryan Gress
    - o Cassie Lefever
    - o Maria Balega
    - o Alicia Deak
    - o Franklin May
- Meeting started at 3:00 p.m.
- Unanimously approved minutes for meeting on November 23, 2021
- Main Points
    - o Discussed:
        - ▪ Maria & Cassie
            - Added graphs, results, and XGBoost abstract to report
            - Dropped 3 labels from preprocessed IoT23 data, ran XGBoost and achieved accuracies up to 99% with 65/35 split performing best
        - ▪ Alicia & Franklin
            - Added graphs and results to report, XGBoost precision seems to rise with 65/35 split with the extra column in the preprocessed data
        - ▪ Zaryn
            - Working on updating report
        - ▪ Ryan
            - Split ToN-IoT data into 50% malicious and 50% benign, XGBoost accuracies were still low around 50%
    - o Tasks:
        - ▪ Maria & Cassie
            - Look further into XGBoost results and why 65/35 split accuracy is higher
            - Add figures and results to report
        - ▪ Alicia & Franklin
            - Begin running KDD dataset in SVM, work on fine tuning XGBoost for better results, consider a graph to compare XGBoost and SVM
        - ▪ Zaryn
            - Add figures and results to report
            - Convert images for IoT23 dataset
        - ▪ Ryan & Alicia
            - Run ToN-IoT in SVM to compare results with published paper and XGBoost
- Adjournment at 4:00 p.m.

- Next meeting will be held on Tuesday December 14 at 3:00 p.m.

CAE Research Project Meeting Minutes

12/16/2021

- Participants
    - o Dr. Waleed Farag (PI)
    - o Dr. Soundararajan Ezekiel (Co-PI)
    - o Dr. Xinwen Wu
    - o Zaryn Good
    - o Franklin May
    - o Ryan Gress
    - o Alicia Deak
- Meeting started at 11:00am.
- Unanimously approved minutes for meeting on November 30, 2021
- Main Points
    - o Discussed:
        - Maria & Cassie
            - Finish preprocessing IoT-23 dataset and run through XGBoost
            - Create graph of results for writeup
        - Alicia
            - Continue working on report and XGBoost function
            - Create SVM code
        - Zaryn & Sky
            - Continue running datasets in DCNN
        - Franklin
            - Continue working on running KDD in XGBoost and SVM, look into execution time
            - Work on DCNN abstract with Zaryn
        - Ryan
            - Look into XGBoost function and figure out why new results differ from the original
    - o Tasks:
        - Maria & Cassie
            - Need to remove the NaN values to run through SVM
            - Organize OneDrive folder to contain any links (in a wordpad) of sites referenced. If possible, include the papers referenced
        - Alicia
            - Create the scatter plot function for SVM
            - Copy codes from titan to removeable drive
        - Zaryn
            - Figure out why the current dataset is running 1,000 images/hr instead of 20,000/hr.
            - Continue running the dataset
        - Franklin

- Send the graphs with a short description for each graph to Dr. Farag
- Update the report on the one drive
  - Ryan & Alicia
    - Run the multiclass through SVM.
    - Compare the results with the papers
- Adjournment at 11:50am.
- Next meeting will be held on Thursday December 23 11:00am.

CAE Research Project Meeting Minutes

Thursday December 23, 2021

- Participants
    - o Dr. Waleed Farag (PI)
    - o Dr. Soundararajan Ezekiel (Co-PI)
    - o Dr. Xinwen Wu
    - o Zaryn Good
    - o Franklin May
    - o Alicia Deak
- Meeting started at 11:00am.
- Unanimously approved minutes for meeting on December 16, 2021
- Main Points
    - o Discussed:
        - ▪ Maria
            - • Need to remove the NaN values to run through SVM
            - • Organize OneDrive folder to contain any links (in a wordpad) of sites referenced. If possible, include the papers referenced
        - ▪ Alicia
            - • Create the scatter plot function for SVM
            - • Copy codes from titan to removeable drive
        - ▪ Zaryn
            - • Figure out why the current dataset is running 1,000 images/hr instead of 20,000/hr.
            - • Continue running the IoT23 dataset
        - ▪ Franklin
            - • Send the graphs with a short description for each graph to Dr. Farag
            - • Update the report on the one drive
        - ▪ Ryan
            - • Run the multiclass through SVM.
            - • Compare the results with the papers
    - o Tasks:
        - ▪ Maria
            - • Need to remove the NaN values to run through SVM
            - • Organize OneDrive folder to contain any links (in a wordpad) of sites referenced. If possible, include the papers referenced
        - ▪ Alicia
            - • Copy codes from titan to removeable drive
            - • Use the temp files from IoT23 and create a smaller distribution for the DCNN
        - ▪ Zaryn
            - • Run the smaller subset of the IoT23 dataset.
        - ▪ Franklin

- Work with Zaryn on the DCNN report.
- Alicia, Zaryn, Franklin will meet to discuss DCNN and the code
  - Ryan
    - Compare the datasets through various SVM kernels and in a file.
    - See Franklins for an example.
- Adjournment at 11:50am.
- Next meeting will be held on Thursday January 6, 2022 11:00am.