Student number: C1646404

Initial plan

Title: Detecting Network based attacks in Industrial control systems

Supervisor: Eirini S Anthi

Moderator: Neetesh Saxena

Module Details:

Module number: CM3203

Credits: 40-one semester final year project

Student number: C1646404

Table of Contents

Student number: C1646404

1 Project description

Operational technology (OT) refers to computers systems that monitor industrial operations/equipment. Industrial control systems (ICS) describe a branch of OT networks that control the monitoring of industrial operations using sensors actuators and a central control hub. ICS networks are found in critical national infrastructure e.g. water treatment, oil and gas as well as other types of industries like the food and beverage and discrete manufacturing industries. ICS systems can incorporate multiple types of control systems e.g. Supervisory control and data acquisition system (SCADA) and distributed control systems (DCS). SCADA systems gather site/plant information and transmit this information to a central location where an operator can analyse the data [2]. This process can also be automated. DCS use a number of control mechanisms distributed over the system which will be used to control different areas and tasks within the system.

The perceived level of security threats in regard to industrial control systems (ICS) has risen over recent years [3]. This indicates that those who viewed security threats on ICS as low are starting to understand the shift in focus and thus the importance of securing ICS networks. The notion of ICS networks being secure due to the use of physically secure locations to store system components while likely to reduce the chances of a local threat, do not provide any security for remote based network attacks. ICS networks are being adapted to perform more closely to typical IT networks for example, incorporating remote access and connecting components of an ICS to the "outside world" [3].Due to these adaptations ICS networks are vulnerable to a multitude of cyberattacks for example, the recent attack on Iranian nuclear power plants using Stuxnet [4]. Stuxnet is a malicious worm that is targeted towards SCADA systems. Stuxnet was built to manipulate computers controlling PLCs that were using Siemens Step7 software (protocol specific to ICS). Once infected, Stuxnet can send commands to PLCs that will cause the PLC to omit dangerous commands for example, to increase the rotation speed of a centrifuge in a nuclear power station causing the centrifuge to break. Another example of a cyberattack on an ICS network is the 2015 Ukraine power grid cyberattack. This attack disrupted the electrical supply to over 225,000 people [5]. This attack used spear-phishing emails using BlackEnergy malware to compromise the IT networks. Attackers were then able to gain control over the SCADA system, from here a simple set of commands shut off the power to numerous substations remotely. These are just two cases that highlight the growing importance of securing OT and thus ICS networks. The second example took advantage of the lack of OT isolation from the IT network, attackers gained access to the OT network via first compromising the IT network. In light of the rising level of cyberattacks directed at ICS networks research has moved towards securing an ICS network in a similar manner to that of an IT network. An example of this is the introduction of using intrusion detection systems (IDS) for ICS.

In the field of IDS, machine learning based models are being used to classify traffic as "benign" or "malicious" some IDS will further classify traffic to distinguish between types of attack. While this has been a popular topic within IT networks for a number of years, the resources and research for IDS tailored to ICS networks is only recently becoming more focused. Classifier models do exist within ICS but research into adversarial methods to alter attacks in a way to force misclassification is a less researched topic. Moreover, improving a model to handle these types of attempts at misclassification is also an area of research that has not been exhausted.

This project will build on the foundations of established machine learning algorithms to build a model that could become the basis of an IDS tailored to ICS networks. Once this model has been finalised adversarial techniques and methods will be used to create new adversarial examples that alter malicious traffic in an attempt to force the model to misclassify malicious traffic as benign. Following

this, improvements to the classifier model will be made to enhance the accuracy of correct classification when faced with adversarial examples.

The final deliverable of this project will be a research paper outlining the steps executed to first highlight the changes applied to the malicious data to force misclassification and then the steps necessary to improve the model to overcome such attacks. The complete project aim's and objectives will be discussed in the next section of this report.

2 Project aims and Objectives

2.1 Project Aim

To use an established classification model as the basis of an IDS tailored to ICS networks. The system must use publicly available datasets which will contain both benign and malicious traffic. The system must use a machine learning model to classify and distinguish between different types of attacks. The use of adversarial methods will be implemented on the dataset to enforce misclassification. The classifier and thus, the IDS will be improved as a result of extensive evaluation and retested to further indicate the performance improvements made to the model to overcome these adversarial examples.

2.2 Project Objectives

- Select publicly available dataset (ICS Cyber Attack Datasets webpage [1]).
    - Dataset must be publicly available and specific to Industrial control systems which will include benign and malicious traffic. This must include proprietary protocols used in ICS along with appropriate documentation outlining the type of attacks the dataset contains along with any known flaws in the dataset that need to be taken into consideration when evaluating the efficiency and effectiveness of a machine learning model-based IDS.
    - The dataset chosen must not include excessive amounts of "noisy" data that is not relevant to ICS. Dataset must be large enough to allow the model to not closely resemble a specific dataset that will then not be transferable to unseen datasets (Overfitting).
    - Dataset must be in a format compatible with Weka (CSV or ARFF).
- Research machine learning techniques and classifier models and learn how to evaluate them.
    - Gain a better understanding of machine learning and how to utilise Weka to produce an appropriate model that will become the foundation of the IDS.
    - Understand the importance of not introducing overfitting by allowing the model to produce great results for training dataset but less promising results for unseen data.
    - To examine the likelihood of overfitting in the IDS I will split the dataset up into a training set and then an unseen test set (this may be done using cross-validation or percentage splitting methods), comparing the two results using performance measures discussed later in this report. If there is a drastic drop in performance between the training set and unseen dataset this will indicate the influence of overfitting.
- Identify and select the appropriate machine learning algorithm I will use as the basis of the IDS. This decision will be a reflection of the analysis of multiple machine learning algorithms available within Weka that most suit my project.
    - I will endeavour to create a baseline performance to then compare each algorithm against.

    - I will use recognised methods of evaluating the overall performance of a classifier model namely, precision, recall and F-measure.
    - These measures will be used to evaluate the effectiveness of correctly identifying malicious traffic along with using these measures to distinguish between a model that can be used on unseen data reliably and a model that too closely resembles the training set and cannot adapt to unseen data.
- Finalise model
    - The resulting model produced using the previous objective will then be finalised and form the basis of the IDS.
- Build scripts that use adversarial methods to carefully alter malicious traffic in an attempt to force the model to misclassify attacks as being benign.
    - Research current adversarial techniques being used to bypass IDS's for IT networks to gain an understanding of the fundamentals and principles of such methods.
    - Adapt existing/create new adversarial methods designed specifically for ICS protocols using a scripting language e.g. Python. Transform these methods into an adversarial example(s) that can be introduced to the model to re-evaluate the accuracy of correct classification.
- Evaluate and improve the classifier model in order to ensure accuracy when faced with the adversarial example(s) designed in the previous stage of the project.
    - Focus on methods of improving the classification model to handle examples similar to those created in the previous stage improving the overall robustness and effectiveness of the model and thus, the IDS.
- Converge findings and write a comprehensive report detailing the steps taken to achieve each deliverable and the final conclusions. This paper could be used as a reference for future work in this area of IDS and classifier models.

3 Work Plan

Week 1: Initial Plan 27/01/2020 – 02/02/2020

1. Research current state of the art in regards to IDS for ICS
2. Create initial plan
3. **Milestone:** submit initial plan on PATS. Achieved date: 03/02/2020

Week 2: Choose Dataset 03/02/2020 – 07/02/2020

1. Select publicly available dataset
2. Analyse dataset for potential flaws and relevancy.
3. Use cross-validation or percentage split methods to partition dataset up into training sets and a test set
4. **Milestone:** secured an appropriate dataset to use as training and (unseen) test data for classifier model. Achieved date: 07/02/2020

Week 3: Evaluate Machine Learning Algorithms Using Weka. 10/02/2020 – 14/02/2020

1. Select and run appropriate machine learning algorithms available on Weka on training set
2. Analyse results using baseline performance to highlight the "best" algorithm for this problem/dataset.
3. **Milestone:** acquire classifier model that could be used as basis of IDS. Achieved date: 14/02/2020

Student number: C1646404

Week 4, 5 and 6: Research and Develop an adversarial example and use this example on dataset to evaluate classifier model in current state. 14/02/2020 – 06/03/2020

1. Research current adversarial examples that are used on traditional IT networks
2. Modify existing/create new example in python
3. Re-evaluate classifier model using new modified data (adversarial example).
4. **Milestone:** Created an adversarial example that is deployed on dataset and have a baseline of classifier performance using modified dataset. This baseline will be used to measure the effectiveness of the subsequent task's objectives (improving the classifier models performance on new modified dataset. Achieved date: 06/03/2020

Week 7,8,9 and 10: Improve performance of classification model regarding modified dataset produced by previous task. 09/03/2020 – 03/04/2020

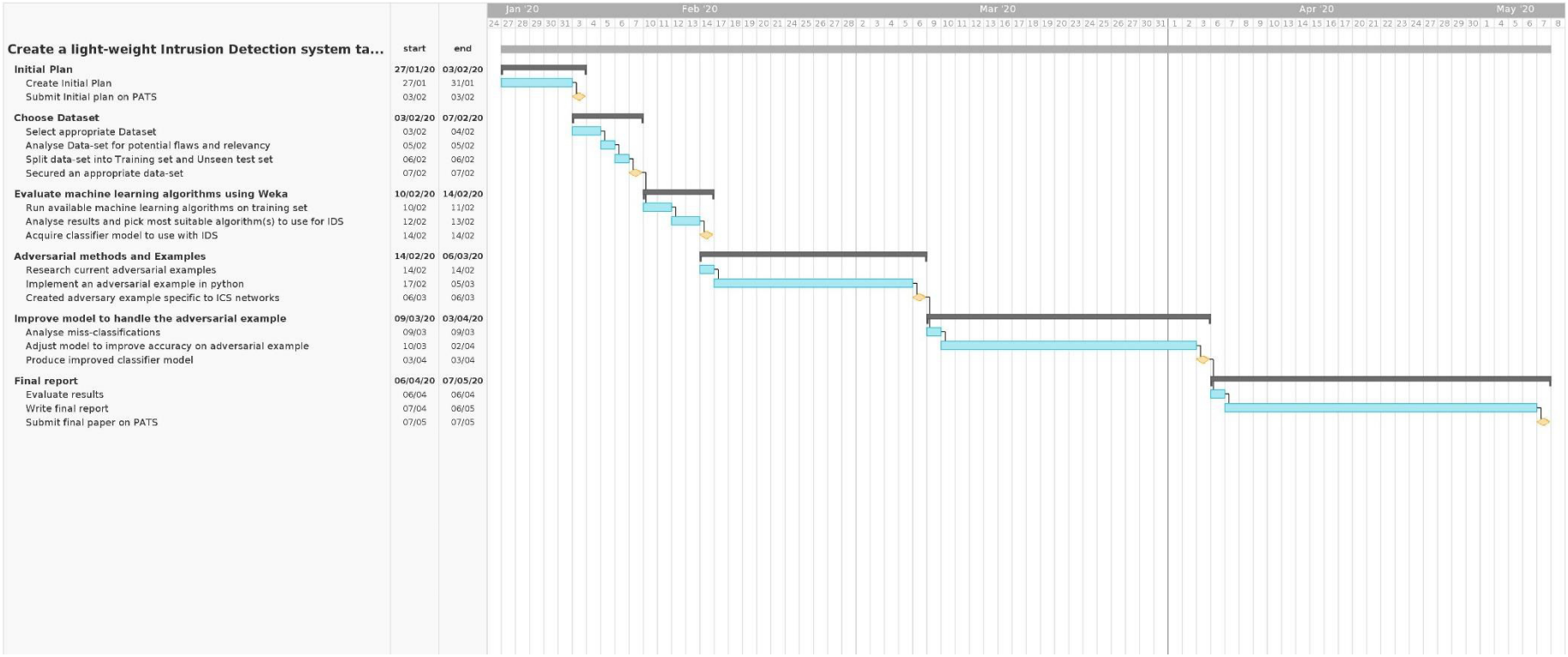
1. Analyse why the model misclassified traffic, extract potential correlation/theory as to what exactly caused the misclassification.
2. Adjust the classification model to improve the performance and therefore improving the robustness and effectiveness of an IDS that used this model as the basis of the system.
3. **Milestone:** Have an Improved model that handles modified data that was being misclassified but now produces better accuracy and precision when classifying traffic that has been altered in an attempt to "fool" the model to misclassify malicious traffic as benign. Achieved date: 03/04/2020

Week 11,12,13 and 14: Final report. 06/04/2020 – 07/05/2020

1. Aggregate all documentation of processes executed during the life cycle of this project
2. Converge analysis and evidence of all results including performance measures of the classifier.
3. Write a comprehensive document detailing all steps needed to achieve the results found during this project. Include all steps relating to creating an adversarial example directed to ICS network traffic and how the model was improved to deal with this "attack".
4. include learning outcomes of the project and the effect this has had on personal/professional development.
5. **Milestone:** submit final report to PATS. Achieved date: 07/05/2020
6. **Milestone:** if results provide new insights to this field of study, submit a slightly modified version of the final report for possible publication.

Student number: C1646404

3.1 Gantt chart



3.2 Risk analysis

#	Risk	Likelihood	Risk severity	Mitigation strategies
1	Machine learning classifier model is biased to training dataset and cannot be reliably used on unseen data	Low	High	-From the outset of the project overfitting is a known concern and will therefore the model will be tested early in development for signs of overfitting. -I will split the chosen dataset up into two subsets allowing one set to be unseen. Allowing a comparison of performance indicating the likelihood of overfitting.
2	Project fails to produce an output that can be assessed and submitted	Low	High	-An appropriate size dataset will be chosen allowing minimal time to train model -Development will start early in the project lifecycle to ensure adequate testing. -
3	Required personal learning for this project is unachievable (Learning about machine learning models, how to utilise Weka and adversarial techniques and methods)	Low	Medium	-Background research and implementation using unfamiliar concepts and techniques will be started early allowing ample time for learning and personal development.
4	Illness	Moderate	Low	-I have accounted for illness in my Gantt chart which should limit the effect of this potential risk.
5	Data loss	Low	High	-Ensure that multiple backups of all code and important documentation is stored in different locations. An external hard drive could be used as a secondary storage medium.

3.3 Ethics

I will only use data that is freely available online for this project, so I do not need to take make any ethical considerations.

References

[1] Industrial Control System (ICS) Cyber Attack Datasets - Tommy Morris. [online] Available at: https://sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets [Accessed 3 Feb. 2020].

[2] What is SCADA? Webopedia Definition. [online] Available at: https://www.webopedia.com/TERM/S/SCADA.html [Accessed 3 Feb. 2020].

[3] [online] Available at: https://www.belden.com/hubfs/resources/knowledge/white-papers/sans-survey-report-ics-security.pdf [Accessed 3 Feb. 2020]. Page 8.

[4] Zetter, K., Zetter, K., Greenberg, A., Barrett, B., Barrett, B., Barrett, B., Gilbertson, S. and Newman, L. (2020). An Unprecedented Look at Stuxnet, the World's First Digital Weapon. [online] WIRED. Available at: https://www.wired.com/2014/11/countdown-to-zero-day-stuxnet/ [Accessed 3 Feb. 2020].

[5] BBC News. (2020). Ukraine power cut 'was cyber-attack'. [online] Available at: https://www.bbc.co.uk/news/technology-38573074 [Accessed 3 Feb. 2020].