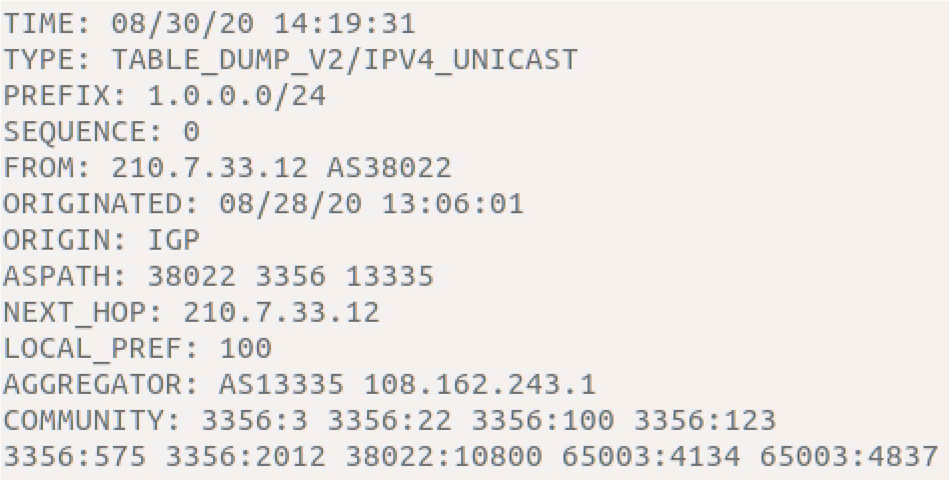
Appendices

Appendix A

A.1 BGP Update Example



```
TIME: 08/30/20 14:19:31
TYPE: TABLE_DUMP_V2/IPV4_UNICAST
PREFIX: 1.0.0.0/24
SEQUENCE: 0
FROM: 210.7.33.12 AS38022
ORIGINATED: 08/28/20 13:06:01
ORIGIN: IGP
ASPATH: 38022 3356 13335
NEXT_HOP: 210.7.33.12
LOCAL_PREF: 100
AGGREGATOR: AS13335 108.162.243.1
COMMUNITY: 3356:3 3356:22 3356:100 3356:123
3356:575 3356:2012 38022:10800 65003:4134 65003:4837
```

Figure A.1 BGP Update from AS38022

Appendix B

B.1 Network Graph Before Anomaly Event

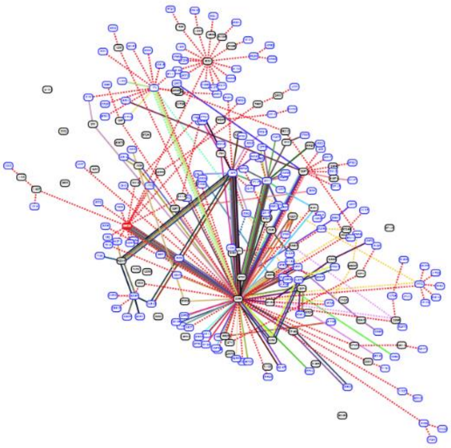


Figure B.1 Graph at 9:04 2020-08-30 (1 hour before anomaly event)

B.2 Network Graph During Anomaly Event

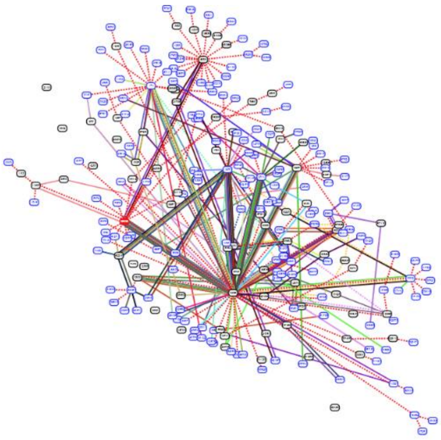


Figure B.2 Graph at 10:05 2020-08-30 (during anomaly event)

B.3 Network Graph After Anomaly Event

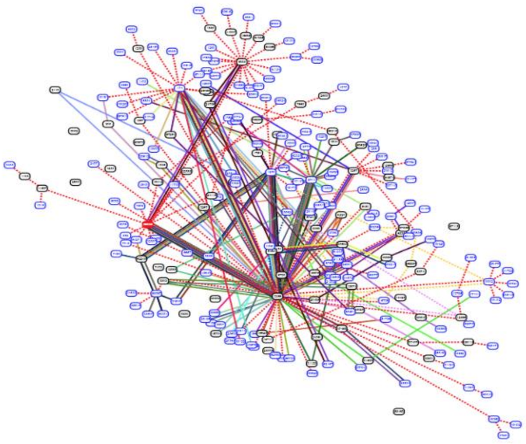


Figure B.3 Graph at 16:19 2020-08-30 (after anomaly event)

Appendix C

C.1 Closeness Centrality Equation

$$Closeness\ Centrality(u) = \frac{n-1}{\sum_{v=1}^{n-1} d(u,v)}$$

Where *u* represents a node in the graph, *n* represents the number of reachable neighbours of node *u* in the graph and *d(u,v)* represents the shortest-path distance between nodes *v* and *u*.

Appendix D

D.1 Error Convergence

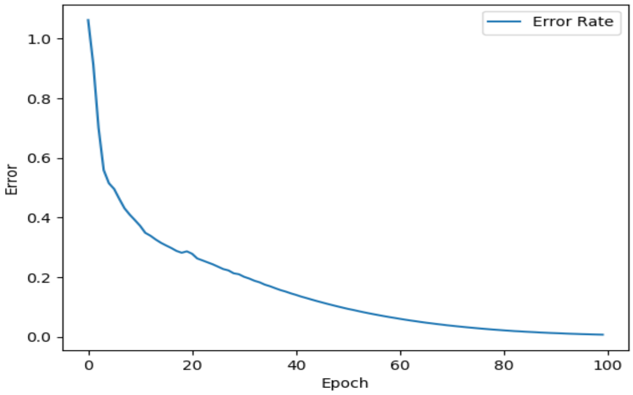


Figure D.1 Error Convergence During Training

D.2 Threshold

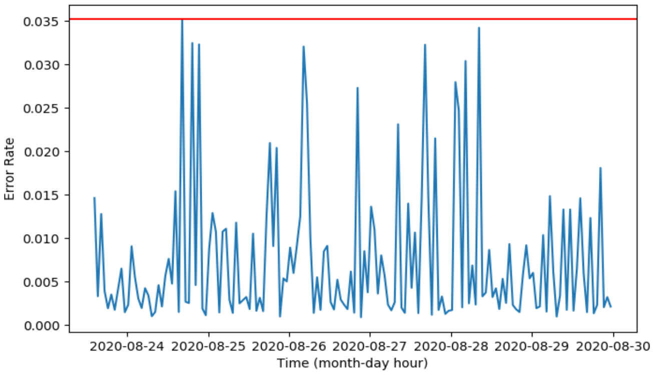


Figure D.2 Threshold of Normal Data

D.3 Anomaly Detection without Normalisation

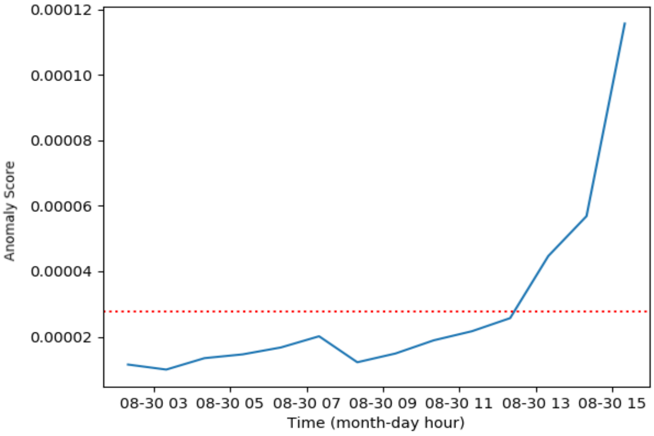


Figure D.3 Anomaly Detection without Normalisation

Figure D.4 Anomaly Detection with Normalisation

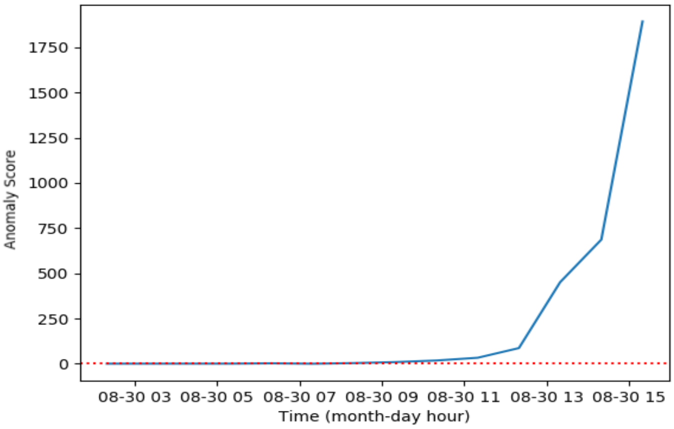


Figure D.4 Anomaly Detection with Normalisation

Appendix E

E.1 NZ Closeness Centrality using Full Joint UG

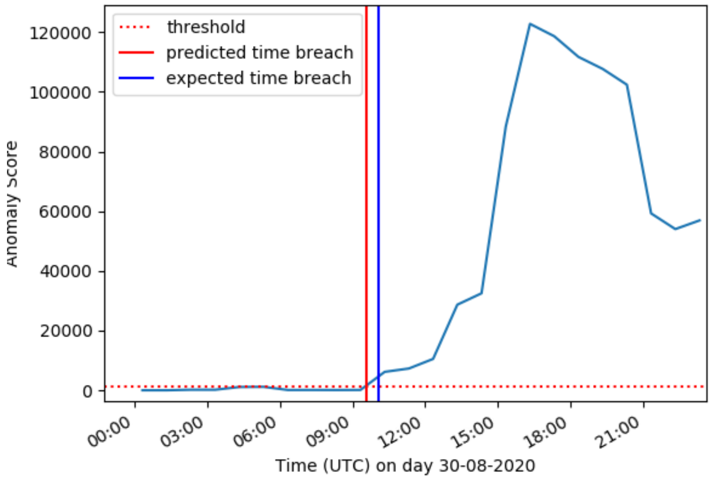


Figure E.1 NZ Entire Network Closeness Centrality with Full Joint UG

E.2 NZ Degree Centrality using Full Joint UG

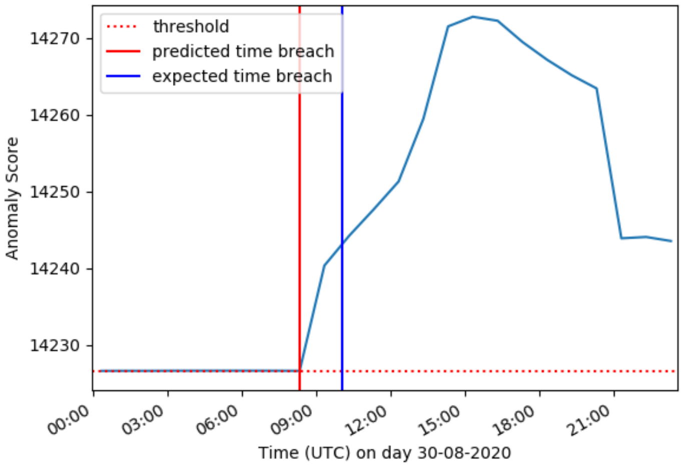


Figure E.2 NZ Entire Network Degree Centrality with Full Joint UG

E.3 SOXRS Closeness Centrality using Full Joint UG

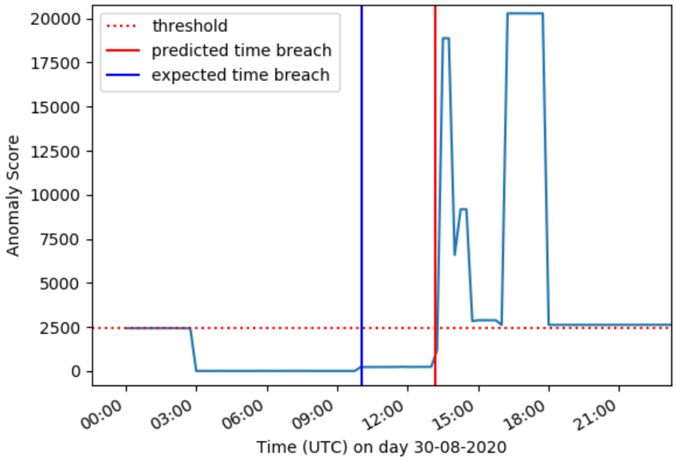


Figure E.3 SOXRS Entire Network Closeness Centrality with Full Joint UG

E.4 SOXRS Degree Centrality using Full Joint UG

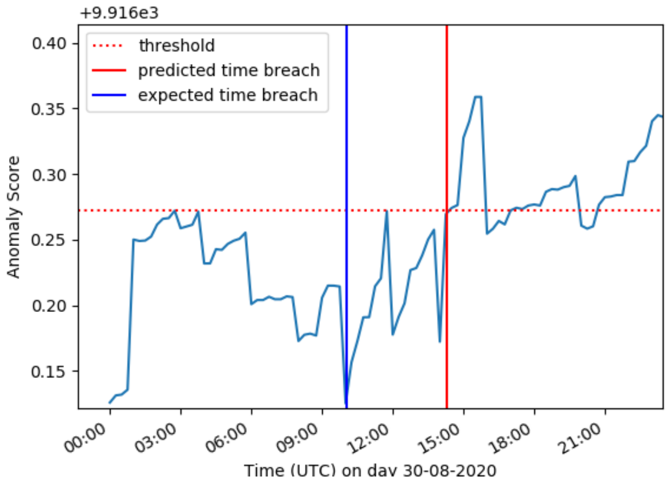


Figure E.4 SOXRS Entire Network Degree Centrality with Full Joint UG

E.5 WIDE Closeness Centrality using Full Joint UG

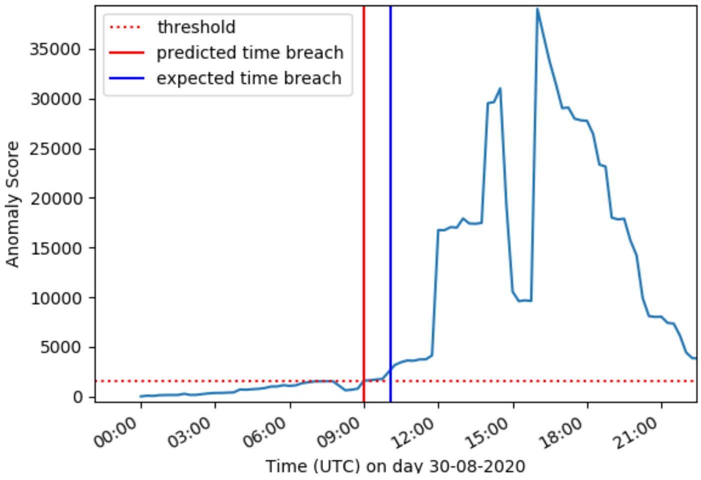


Figure E.5 WIDE Entire Network Closeness Centrality with Full Joint UG

E.6 WIDE Degree Centrality using Full Joint UG

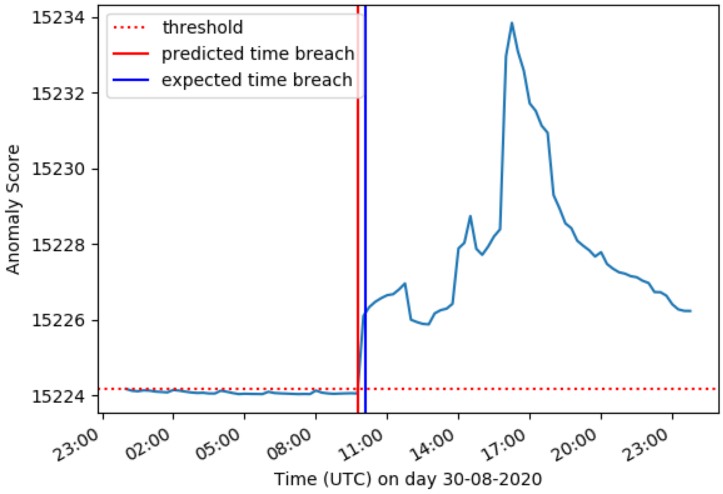


Figure E.6 WIDE Entire Network Degree Centrality with Full Joint UG

Appendix F

F.1 Autoencoders Architecture

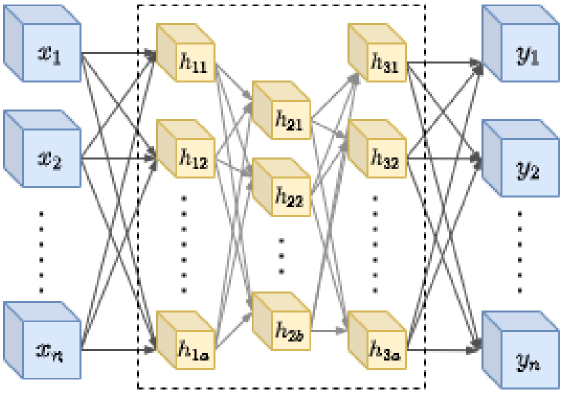


Figure F.1 Autoencoders Architecture

F.2 UG Architecture

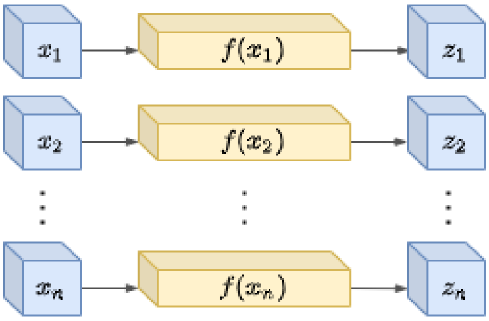


Figure F.2 UG Architecture

Appendix G

G.1 Project Timeline

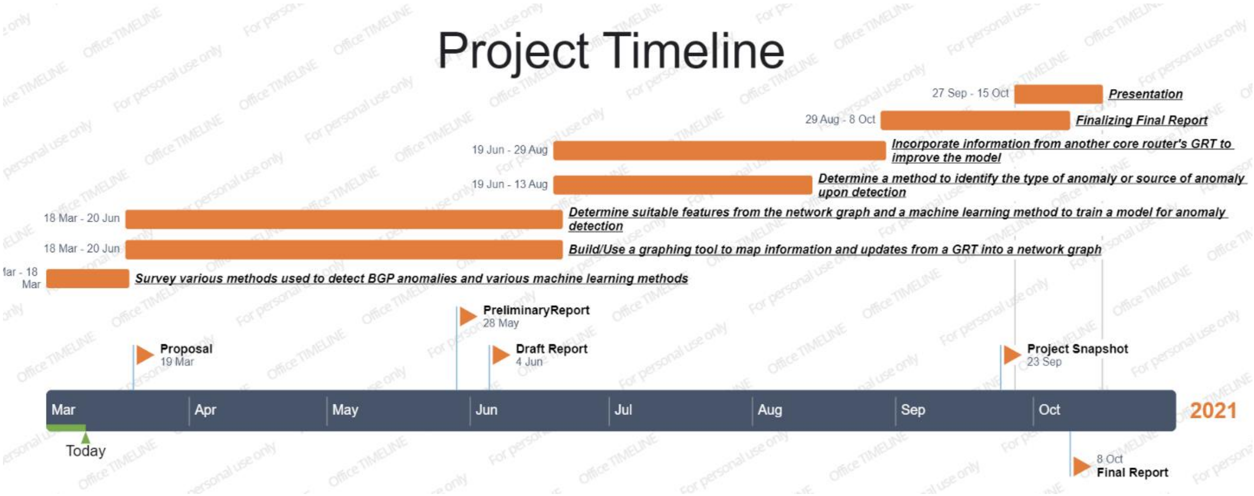


Figure G.1 Project Timeline

Appendix H

H.1 Project Proposal

VICTORIA UNIVERSITY OF WELLINGTON

Te Whare Wananga o te Upoko o te Ika a Maui

School of Engineering and Computer Science

Te Kura Mātai Pūkaha, Pūrorohiko

PO Box 600
Wellington
New Zealand

Tel: +64 4 463 5341
Fax: +64 4 463 5045
Internet: office@ecs.vuw.ac.nz

Modelling BGP Updates for Anomaly Detection using Machine Learning

Janel Hang

Supervisor(s): Winston Seah and Marcus Frean

Submitted in partial fulfilment of the requirements for
ENGR489

Abstract

This proposal documents the engineering problem to be solved, the proposed solution and evaluation methods. Furthermore, the project resource requirements are documented to state the necessary resources to complete the project.

1. Introduction

BGP hijacking is a form of a Distributed Denial of Service (DDOS) attack where an attacker can impersonate a network to reroute Internet traffic. To detect BGP hijacking, modelling BGP updates using machine learning can be used. Therefore, this project aims to use a graphing tool and determine the features from a BGP core router data to train a machine learning model for anomaly detection.

The evaluation of the proposed solution will be based on the detection of the anomalies in the given datasets, where each dataset contains the number of anomalies.

The resource requirements for the project will include the Research and Education Advanced Network New Zealand (REANZ) and Reseaux IP Europeens (RIPE) dataset [3] and the graphing tool software.

2. The Problem

As of 2021, people in separate locations around the world can communicate with one another using the Internet. It is commonly seen that younger generations use the Internet for vlogging, social networking and online shopping. The industry is also growing to use the Internet for communicating or transferring information. However, as more applications are delivered via the Internet, the greater the assumption is for the secure transmission of information through the Internet. More specifically, the routing protocol of the Internet, Border Gateway Protocol (BGP) must be secure to transfer information between different networks.

Over the last decade, there have been incidents of BGP hijacking. This is where attackers impersonate networks or Autonomous Systems (AS) by advertising false BGP routes to maliciously reroute Internet traffic. In worst cases, BGP hijacking led to loss of connectivity for domains such as panix.com [1], an Internet service provider in USA. Unintentional BGP hijacking are more commonly seen where in the Pakistan Telecom incident, invalid BGP routes were advertised in attempt to ban youtube.com [2]. However, a misconfiguration led to redirecting multiple ASes youtube.com traffic to the Pakistan AS, causing a Denial of Service for its own AS and a loss of connectivity for youtube.com for the affected ASes. Therefore, the detection of BGP hijacking is crucial when routing Internet traffic.

The detection of BGP hijacking can be accomplished using a machine learning algorithm. This is where a model can be trained using existing BGP events to detect anomalies within the network. However, with the node level features extracted from the raw BGP data and parsed into a machine learning model, the machine learning algorithms are incapable of detecting anomalies and are unable to so in real time [4]. Therefore, the aim of this project is to implement a graphing tool to map a core router’s BGP data and determine the features from the network graph to train a machine learning model for anomaly detection.

3. Proposed Solution

To model the BGP updates for anomaly detection using machine learning, the project is divided into the following tasks:

1. Survey various methods used to detect BGP anomalies and various machine learning methods
2. Build or use a graphing tool to map information and updates from a Global Routing Table (GRT) into a network graph
3. Determine suitable features from the network graph and machine learning method to train a model for anomaly detection
4. Determine a method to identify the type of anomaly or source of anomaly upon detection
5. (Extra) Incorporate information from another core router's GRT to improve the model

Appendix 1 includes a Gantt Chart that summarises the project timeline. Note that the project report will be written alongside the deliverables to document the deliverables within the project.

A suitable graphing tool will be chosen to map the BGP routing information. If an existing graphing tool is not found, a graphing tool will be built to map the BGP routing information.

4. Evaluation of Proposed Solution

The solution will be evaluated based on whether the anomalies in the REANZ dataset can be detected, where the number of anomalies is reflected within each dataset. The passing threshold for the successful detection of anomalies will be selected based on the threshold of the normal behaviour datasets. The detection of anomalies would be divided into the following categories of true positive, true negative, false positive and false negative.

The evaluation of anomalies will be using the measurements of internet forwarding density as recommended by [5]. For example, if there are unusual peaks of internet density or BGP updates at times suggested by the normal dataset, this will be classified as abnormal. To further enhance the evaluation, a time series analysis could be applied to evaluate the anomality of BGP events over time. The analysis should be conducted at prefix, AS and route levels to determine the instability and hence the potential anomalies within the network.

5. Ethics and Resourcing

5.1. Ethics

There are no ethical considerations within the design or the evaluation of this project.

5.2. Dataset

The project will use the REANNZ and RIPE datasets [3]. This will be used as a dataset for producing a graph that maps the BGP updates.

5.3. Budget

The resources used within this project are open-sourced. Hence, no budget would be required for this project.

5.4. Space and Access

This project can be done remotely and hence no space and access requirements would be required for this project.

5.5. Intellectual Property

The results of this project will be the intellectual property of Janel Huang, Winston Seah, Marcus Frean, Murugaraj Odiathevar and VUW.

6. COVID Alert Level Plans

6.1. Alert 1 and 2

During COVID alert levels, 1 and 2, project meetings will be held in person. During alert level 2, social distancing will be observed. The project will be conducted in or out of campus with social distancing observed at alert level 2.

6.2. Alert 3 and 4

During COVID alert levels, 3 and 4, project meetings will be held remotely on zoom. The project will be conducted remotely (out of campus) to reduce the likelihood of social gatherings.