|  |
| --- |
| Image result for data center |
| ASDN – Automated Software Defined Networking  Technical Specification – CA400 |
| |  |  |  | | --- | --- | --- | | FILIP NIKOLIC – 14470852 | 17/05/2018 | Supervisor: Brian Stone | |

Table of Contents

[1. Introduction 1](#_Toc514321209)

[1.1. Overview 1](#_Toc514321210)

[1.2. Glossary 1](#_Toc514321211)

[2. System Architecture 2](#_Toc514321212)

[2.1. Statistical Machine Learning 2](#_Toc514321213)

[2.1.1. Metrics Used 2](#_Toc514321214)

[2.1.2. Data Visualisation 3](#_Toc514321215)

[2.1.3. Algorithm Evaluation and Building the Best Model 3](#_Toc514321216)

[2.1.4. Making Predictions using the CART Algorithm 5](#_Toc514321217)

[3. High-Level Design 6](#_Toc514321218)

[4. Problems and Solutions 7](#_Toc514321219)

[5. Installation Guide 8](#_Toc514321220)

[6. References 9](#_Toc514321221)

# Introduction

## Overview

This document will outline the system architecture, high-level design overview, as well as the problems encountered while developing the application.

The Network Deployment Automation & Maintenance Tool has been developed to automate the most common/repetitive task carried out by network engineers.

The application is dependent on highly specialised hardware/software to be used to its full potential. The three main areas the program automates are device \*deployment\* ([Juniper networks] (http://www.juniper.net/us/en/) Routers, Switches and Firewalls), \*maintenance\* (version control configuration files-GitHub like feature) and \*troubleshooting\* - a network mapping tool. It communicates with other systems on the Local Area Network using various

protocols, such as SSH, FTP, TCP, etc. To accomplish this, there are

various requirements needed to be satisfied by the user (Network Engineer) in

order to use some program features, mainly the deployment module.

## Glossary

# System Architecture

### Statistical Machine Learning

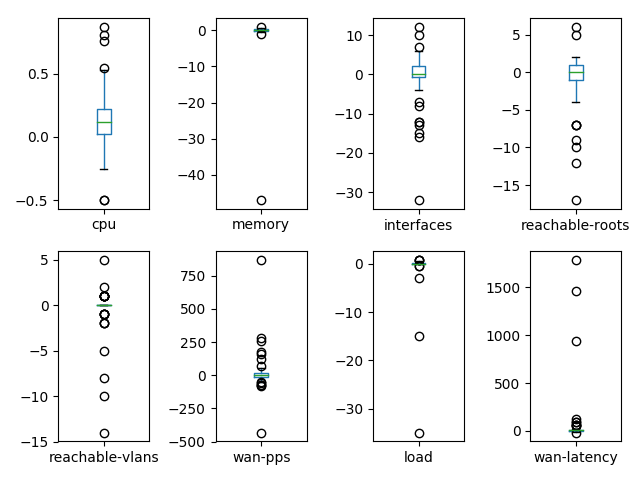
Each metric is a delta of two values, the first taken before the configuration has been committed and the second after the commit. They must all be considered by the system when arriving at the result.

### Metrics Used

* + **CPU Utilisation** 
    - Indicating the potential overhead increase on the device’s processor.
    - In many cases, a dramatic change of this value (both positive and negative) suggests a negative network state change.
  + **Memory Utilisation**
    - This metric is used in conjunction with CPU utilisation and can further help determine whether an error on the network occurred.
    - It should be an even more stable metric then CPU utilisation, where extreme fluctuation almost guarantees unwanted behaviour.
  + **Number of Interfaces**
    - By considering the number of physical and virtual interfaces that came online or went offline, we can determine whether the configuration has added or removed devices from the network.
    - When multiple interfaces go offline, this is most commonly a negative state change, alternatively a single added interface in conjunction with other attributes such as increase in system load, latency, etc may imply a loop introduced on the network.
  + **Reachable Roots**
    - Using this metric, we can determine how many more/less Layer 3 devices we can reach after the configuration change.
    - In most situations a decrease is negative and increase negative.
  + **VLAN Availability** 
    - This metric is used to show how many more/less Virtual Local Area Networks are reachable by the device.
    - Akin to the reachable roots metric, a decrease is in most cases negative and increase positive.
  + **Wide Area Network packets per second (pps)**
    - This distinct metric may help determine whether a loop has been introduced or indicate a significant decrease in user traffic.
    - Most commonly, a drastic positive or negative change in this value hints negative impacts of the new configuration.
  + **System Load Average** 
    - Unlike CPU and memory utilisation metrics, the system load average takes into consideration many more points, such as number of processes running, system response times, processes on hold, etc.
    - A sharp change of this value is ordinarily negative indication.
  + **WAN Latency measured in milliseconds (ms)**
    - To measure WAN latency in milliseconds, we take the transit delays of packets to the default device route.
    - A significant increase of this value is undesirable.
  + **Result Class**
    - The two classes are used to indicate the dataset result. The “valid” class implies that the changes made to the network had no negative impacts. The “invalid” class indicates potential unwanted behaviour, whereby one or all network clients are experiencing reduced performance or have lost connection completely.
    - They are output results of running an evaluation process on the set of above mentioned metrics.

### Data Visualisation

There were three datasets with 50 entries each and a validation algorithm is used to determine the most accurate one. To better understand the data and decide on the best suited dataset, we use a number of visualisation techniques. The one I found most helpful is a box plot graph:



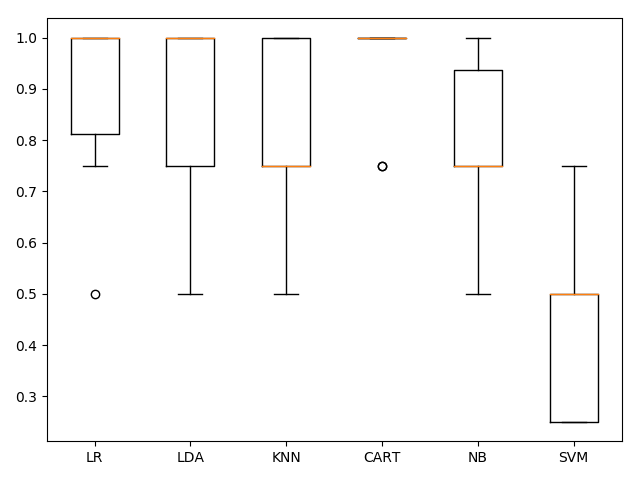
### Algorithm Evaluation and Building the Best Model

The data must be separated out to form a validation dataset, 80% of which will be used to train our models and 20% that held back as a validation dataset. This will then be used to run the 10-fold cross validation.

Six algorithms were evaluated to choose the best one that will work with this dataset. Their accuracy results tell us in how many cases the evaluation of the algorithm was correct and in how many incorrect. Listed below are the results of those algorithms ran on the “best” dataset:

1. Logistic Regression (LR) is on average correct 90% of the time, with a mean deviation of score of 0.165831.
2. Linear Discriminant Analysis (LDA) is on average correct 87.5% with a mean deviation of score of 0.167705.
3. K-Nearest Neighbours (KNN) is on average correct 80% of the time, with a mean deviation of score of 0.187083.
4. Classification and Regression Trees (CART) is on average correct 92.5% of the time, with a mean deviation of score of 0.114564.
5. Gaussian Naive Bayes (NB) is on average correct 80% of the time, with a mean deviation of score of 0.150000.
6. Support Vector Machines (SVM) is on average correct 45% of the time, with a mean deviation of score of 0.187083.

Above is a good combination of simple linear (LR and LDA), nonlinear (KNN, CART, NB and SVM) algorithms. To ensure the evaluation of each is done using the same data splits, we reset the random number seed before every run. This makes the results directly comparable.



Using the graph above, we can see that the most suitable algorithm for this dataset is the Classification and Regression Trees Algorithm. Therefore, we will use it to further build and improve upon the model accuracy and performance.

### Making Predictions using the CART Algorithm

After running the algorithm on a set of 10 data points, an accuracy of 100% was observed. Whereby, CART correctly classified six invalid and four valid configuration changes correctly.   
The following metrics were observed:

* **Precision** is the ratio of correctly predicted positive observations to the total predicted positive observations.
* **Recall** is the ratio of correctly predicted positive observations to the all observations.
* **F1-Score** is the weighted average of Precision and Recall.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | precision | recall | f1-score | support |
|  |  |  |  |  |
| Invalid | 1.0 | 1.00 | 1.00 | 6 |
| Valid | 1.00 | 1.00 | 1.00 | 4 |
| avg / total | 1.00 | 1.00 | 1.00 | 10 |

# High-Level Design

# Problems and Solutions

# Installation Guide

# References

Cover Image, (2018), Datacentre [ONLINE]. Available at: http://ccscleaning.com/wp-content/uploads/2017/04/DataCentre.jpg [Accessed 17 April 2018].

Machine Learning Mastery. 2016. Machine Learning Project in Python. [ONLINE] Available at: https://machinelearningmastery.com/machine-learning-in-python-step-by-step/. [Accessed 2 April 2018].

Machine Learning Mastery. 2016. Machine Learning Project in Python. [ONLINE] Available at: https://machinelearningmastery.com/machine-learning-in-python-step-by-step/. [Accessed 2 April 2018].

https://www.networkworld.com/article/3197119/lan-wan/machine-learning-are-we-there-yet.html

http://www.kddi-research.jp/english/newsrelease/2016/022201.html

\*\*\*Overview of ESXi

1. [Introduction](#1introduction)

1. [Overview](#11overview)

2. [Glossary](#12glossary)

2. [System Architecture](#2system-architecture)

3. [High-Level Design](#3high-level-design)

4. [Problems and Resolution](#4problems-and-resolution)

5. [Installation Guide](#5installation-guide)

\*\*1.Introduction\*\*

========

This document will outline the system architecture, high-level design overview,

as well as the problems encountered while developing the application.

The Network Deployment Automation & Maintenance Tool has been developed to

automate the most common/repetitive task carried out by network engineers.

### 1.2 Glossary

\*\*FTP Server\*\* – hardware/software using the File Transfer Protocol to store,

receive and share files over a network.

\*\*Console Server\*\* – a device containing one or multiple serial ports, allowing

it to

interface with other devices using various networking technologies. Mainly

deployed as a management device, as it enables the user to monitor and control devices

plugged in from a local or remote network.

\*\*Database Server\*\* - Hardware/software using languages such as MySQL to store and access

data from a local/remote location for tasks such as analysis, storage, manipulation and archiving.

\*\*Network Switch\*\* – connects various devices together on a computer network. It

uses packet switching to receive, process and forward data to the destination

device,

operating on the data link layer of the OSI Model.

\*\*Network Router\*\* – forwards data packets between different networks. Packets

are

usually forwarded between routers, until it reaches its destination node. They

operate on the Network Layer of the OSI Model.

\*\*Firewall\*\* – a network security device that monitors incoming and outgoing

traffic,

making decisions in real-time whether to allow or block traffic based on a

defined set

of security rules.

\*\*ICMP\*\* – network layer protocol that provides troubleshooting, control and

error message services. Allows for the usage of a network utility called Ping.

\*\*SSH\*\* – a cryptographic network protocol for operating network services

securely

over an unsecured network. The best-known example application is for remote

login

to computer systems by users.

\*\*XML\*\* – is a markup language that defines rules for encoding data in a human

readable and machine-readable code.

\*\*ARP\*\* – Address Resolution Protocol is used by the Internet Protocol (IPv4),

to map IP network addresses to the hardware addresses (MAC). It operates below

the network layer as a part of the interface between the OSI network and link

layer.

2.System Architecture

================

Network Architecture Required:

![lan](http://i.imgur.com/08NmJCK.png)

The system has been split up into \*\*three\*\* main modules:

1. \*\*Deployment\*\*

This module is used to automate the deployment of Juniper Networking

devices, by automatically updating the device’s operating system as well as

applying an initial configuration file.

2. \*\*Maintenance\*\*

This module is responsible for a version control of configuration files

applied to devices. It shows the user the lines added and removed from the

application file, as well as allowing for traceability of who has applied

the changes to the device configuration.

3. \*\*Troubleshooting\*\*

The main feature of this module is to create a map of a local area network,

by representing each device as a node on a graph. It also gives the user the

ability to easily check device details such as IPv4 address, MAC address and

manufacturer of device.

Finally, if there is a Juniper device on the network, additional information

will be shown such as CPU temperature, RAM usage, device uptime and active

alarms on a device.

\*\*System Architecture Diagram:\*\*

![system](http://i.imgur.com/pt2GNnc.png)

3rd party dependencies used:

- \*xmltodict\* – Used to encode and decode xml files while parsing output from

a Juniper device.

- \*paramiko\* – Used to communicate with various devices on the network via the

SSH protocol.

- \*PyQt5\* – A library used to develop the main application user interface.

- \*NetworkX\* & \*Matplotlib\* - Used to create and interactive graph in the

troubleshooting module, after completing a network scan.

- \*Pymysql\* – Used to communicate with the database server over the network.

3.High-Level Design

========

The three modules outlined previously in the document have no direct communication channels, all communication is carried out via the main GUI thread \*launcher.py\*.

\*\*Data Flow Diagrams:\*\*

----------------------

\*Deployment\*

![deployment](http://i.imgur.com/zjOqKhi.png)

----------------------

\*Maintenance\*

![maintenance](http://i.imgur.com/UiOOSOJ.png)

----------------------

\*Troubleshooting\*

![troubleshooting](http://i.imgur.com/iT2VirJ.png)

----------------------

\*\*Constraints:\*\*

+ JUNOS - while interacting with a device to be configured from the deployment module, all commands sent and recieved are compatible with JUNOS (Juniper Networks Operating System).

+ Network Infrastructure - as stated previously in the document, a combination of specialised network hardware/software is required in order to utilise the application to its full potential.

+ Security - keeping the application dependency for an internet connection down to a minimum, as one of the main reasons for not using the GitHub platform for version control is the feedback gathered by the companies in the industry with security concerns.

All decisions during the design phase have been to accomodate the constraints stated above.

4.Problems and Resolution

========

This section should include a description of any major problems encountered

during the design and implementation of the system and the actions that were

taken to resolve them.

\*\*Problems:\*\*

1. Ensuring the application has as much cross platform client support as

possible (Windows, Linux, MacOS).

2. Taking a list of devices located on a local area network scanned and

construct an interactive visual representation from it.

3. Gathering device hardware information (basic/advanced).

4. PyQt5 user interface was a challenge due to the multithreaded nature of the

application.

5. The security profile of the application, as it handles very sensitive and

valuable data.

6. An efficient way of scanning files for differences, used by the maintenance

(version control) module.

7. Creating a simple application deployment and setup process.

\*\*Solutions:\*\*

1. We have achieved this by using commands and string parsing processes which

work correctly for each OS. In addition, we check which operating system the

software is running on. Commands and output layouts were always different,

making it quite a challenge.

2. This problem required multiple minor problems to be solved.

1. We have used a plotting library for Python which interacts well with

PyQt5 called \*Matplotlib.\* Allowing us to construct a node graph

specified in the Functional Specification.

2. Making a static graph interactive. As the graph is an image file,

overlaid it with a mouse event listener with dynamically determined

coordinates. This is connected to a window area, meaning the 0,0

coordinate is in the graph’s window. We save the coordinates of each

node. Allowing for a click on each node to trigger an event.

3. The calculations had to be done on a separate thread, therefore, an

additional one was needed to compute graph data at the same time. This

meant that we had to implement a producer-consumer architecture, whereby

a global variable is set, in order to notify the sleeping thread that

data is available.

3. Gathering hardware information came in two steps, basic and more advanced

which was applicable only to Juniper devices.

1. The basic information is produced by using the ARP table in order to

gather device MAC address live on the network.

We use this information to extrapolate the device vendor.

When online we do it by interfacing with an API which is connected to

the IEEE database.

If the user does not have access to the internet, we consult a static

page of 26,000+ entries for a device vendor.

2. The advanced information required Juniper OS detailed knowledge, as well

as how to parse the for the Information gathered from a device.

We had to use XML parsing to achieve this.

4. GUI updates must be done on the main thread, however any heavy

calculation/connection related activity needs its own thread to improve

efficiency, as well as to avoid application crashes.

This also meant that data processed during the time of execution had to be

handed back to the main thread.

This came in two forms, firstly the parent class was passed and then the

updated local or global variable.

This approach required the main thread to be aware of the one carrying out

the calculation.

Mutual exclusion is not an issue, as there was never more than one thread

working on a variable.

5. To achieve desired functionality, logging into various devices is required.

This means that usernames and passwords are being handled by the

application.

We have minimised our attack surface by prompting the user for a password

only at runtime, therefore not storing it anywhere.

In addition, we have followed best security practises, by requiring password

protected accounts to be set up.

The latest network protocols for communication, such as SSH were used.

6. The maintenance module is used to compare a local configuration file with

ones located on the FTP server.

This required another multithreaded component to be developed. Once both

files are available, a compare function will check their contents line by

line and add what was deleted and added to two separate arrays.

IO operations are only done on one file during the comparing process, as the

second file has already been pulled down and saved in memory.

7. We have developed an installation process that requires the user to run two

install files (Windows) and a single file (Linux).

This was the only solution we came across, as many of the dependencies we

have used do not work correctly when ran as an executable file.

5.Installation Guide

========

- Please refer to the [user manual](https://gitlab.computing.dcu.ie/mcneilc2/2017-CA326-Cillian-Network\_Deployment\_Automation\_Maintenance\_Tool/blob/master/user\_manual/user\_manual.md#3installation-client-side)

for a detailed installation guide on how to install our application