Table of Contents

Preliminary Content

Acknowledgements

I want to thank my Advisor, Professor Peter Hoff, and the Director of Undergraduate Studies, Professor Mine Cetinkaya-Rundel, for their guidance in this project. I also want to thank my parents for their continued unwavering support in all my endeavors.

Abstract

The goal of this project is to identify novel methods for detecting anomalies in network IP data. The space is represented as a 3-dimensional tensor of the continuous features (source bytes, destination bytes, source packets, destination packets) divided by their respective source port and destination port combinations. This project implements and assesses the validity of principal component analysis and matrix completion via singular value decomposition (more methods pending) in determining anomalous entries in the tensor.

Chapter 1

Introduction

1.1 Anomaly Detection

1.1.1 Supervised versus Unsupervised

1.2 Network Attacks

Network security is becoming increasingly relevant as the flow of data, bandwith of transactions, and user dependency on hosted networks increase. As entire networks grow in nodes and complexity, attackers gain easier entry points of access to the network. The most benign of attackers attempt to shutdown networks (e.g. causing a website to shutdown with repeated pings to its server), while more malicious attempts involve hijacking the server to publish the attacker's own content or stealing unsecured data from the server, thus compromising the privacy of the network's users.

Attackers follow a specific three step strategy when gathering intelligence on a network, the most important component of which is scanning. Network scanning is a procedure for identifying active hosts on a network, the attacker uses it to find information about the specific IP addresses that can be accessed over the Internet, their target's operating systems, system architecture, and the services running on each node/computer in the network. Scanning procedures, such as ping sweeps and port scans, return information about which IP addresses map to live hosts that are active on the Internet and what services they offer. Another scanning method, inverse mapping, returns information about what IP addresses do not map to live hosts; this enables an attacker to make assumptions about viable addresses.

All three of these scanning methods leave digital signatures in the networks they evaluate because they apply specific pings that are then stored in the network logs. Most scanners use a specific combination of bytes, packets, flags (in TCP protocol), and ports in a sequence of pings to a network. Identifying a scanner's often many IP addresses from the set of pings available in the network's logs is thus an unsupervised anomaly detection problem.

This particular dataset is from Duke University's Office of Information Technology, and it covers all transactions in their network traffic during a five minute period in

February 2017.

Chapter 2

Literature Review

2.1 Name of Current OIT Method

2.2 Kernel Principal Component Analysis

2.3 Matrix Completion via Singular Value Decomposition

Ask Mine what goes into a literature review
–>

Chapter 3

Network Attacks Dataset

3.1 Features

3.1.1 Argus

3.1.2 Categorical Features

3.1.3 Continuous Features

3.2 Exploratory Data Analysis

3.3 Relationships

3.4 Correlation

3.5 Other Stuff

Does EDA even go into a thesis or is it appendix?

–>

Chapter 4

Matrix Techniques for Anomaly Detection

4.1 Ports Combination Matrix/Tensor

4.2 Principal Component Analysis

4.3 Matrix Completion via Singular Value Decomposition

->
  ->
  ->
  ->

Chapter 5

Statistical Model

5.1 Uneven Variances

Discuss AMMI model

5.2 Determined Model

5.3

–>

–>

–>

–>

Conclusion

If we don't want Conclusion to have a chapter number next to it, we can add the `{-}` attribute.

More info

And here's some other random info: the first paragraph after a chapter title or section head *shouldn't be* indented, because indents are to tell the reader that you're starting a new paragraph. Since that's obvious after a chapter or section title, proper typesetting doesn't add an indent there.

Appendix A

The First Appendix

This first appendix includes all of the R chunks of code that were hidden throughout the document (using the `include = FALSE` chunk tag) to help with readibility and/or setup.

In the main Rmd file

```
# This chunk ensures that the thesisdowndss package is
# installed and loaded. This thesisdowndss package includes
# the template files for the thesis.
if(!require(devtools))
  install.packages("devtools", repos = "http://cran.rstudio.com")
if(!require(thesisdowndss))
  devtools::install_github("mine-cetinkaya-rundel/thesisdowndss")
library(thesisdowndss)
```

In Chapter ??:

Appendix B

The Second Appendix, for Fun

References

Angel, Edward. 2000. *Interactive Computer Graphics : A Top-down Approach with Opengl*. Boston, MA: Addison Wesley Longman.

———. 2001a. *Batch-File Computer Graphics : A Bottom-up Approach with Quicktime*. Boston, MA: Wesley Addison Longman.

———. 2001b. *Test Second Book by Angel*. Boston, MA: Wesley Addison Longman.