LIBRARY
Te Tumu Herenga
THE UNIVERSITY OF AUCKLAND

RESEARCHSPACE@AUCKLAND

http://researchspace.auckland.ac.nz

ResearchSpace@Auckland

Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

To request permissions please use the Feedback form on our webpage. http://researchspace.auckland.ac.nz/feedback

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form.

Network Event Detection with Entropy Measures

Raimund E. A. Eimann

A thesis submitted in partial fulfilment of the requirements
for the degree of
Doctor of Philosophy
in
Computer Science

The University of Auckland, 2008

Abstract

Information measures may be used to estimate the amount of information emitted by discrete information sources. Network streams are an example for such discrete information sources. This thesis investigates the use of information measures for the detection of events in network streams.

Starting with the fundamental entropy and complexity measures proposed by Shannon and Kolmogorov, it reviews a range of candidate information measures for network event detection, including algorithms from the Lempel-Ziv family and a relative newcomer, the T-entropy. Using network trace data from the University of Auckland, the thesis demonstrates experimentally that these measures are in principle suitable for the detection of a wide range of network events.

Several key parameters influence the detectability of network events with information measures. These include the amount of data considered in each traffic sample and the choice of observables. Among others, a study of the entropy behaviour of individual observables in event and non-event scenarios investigates the optimisation of these parameters.

The thesis also examines the impact of some of the detected events on different information measures. This motivates a discussion on the sensitivity of various measures.

A set of experiments demonstrating multi-dimensional network event classification with multiple observables and multiple information measures concludes the thesis.

Preface and Acknowledgments

This Ph.D. thesis applies T-entropy, a relatively new information measure, for network event detection for the first time. If the author is correct, it is also the first Ph.D. thesis to review the expected entropies of common network protocol fields.

Now looking back this project, I remember many occasions on which I could have gotten sidetracked. I would like to express my gratitude to colleagues and friends for their support during the sometimes quite challenging periods of this project.

My academic supervisor Ulrich Speidel made himself available at short notice for discussions and was always ready with good suggestions when I ran into difficulties. On countless occasions Ulrich helped me to ship around difficulties by providing new insights and ideas. Ulrich's support was not limited to the project work itself: He also helped me come to terms with turbulent times in my personal life, the loss of my father being the most difficult.

Nevil Brownlee has been my co-supervisor during this project. At the beginning of the project I had the opportunity of visiting Nevil for three days at the Cooperative Association for Internet Data Analysis (CAIDA) in San Diego, USA. It was a privilege to meet many interesting people that work in the area of network measurement and data analysis face-to-face.

Both Ulrich and Nevil have provided essential help with the research papers that were published in the course of this project.

On the technical side, I owe thanks to the staff at ITSS and the Department of Computer Science. In particular, Russell Fulton and James Harper have greatly helped with data interpretation and acquisition.

I very much appreciated the feedback and suggestions from my fellow Ph.D. students Jia Yang and DongJin Lee.

With English being my second language, this thesis depended heavily on the watchful eyes of my proofreaders, Caroline and Peter Seddon, as well as Michael Taylor to whom I feel deeply indebted.

This project would not have been possible without a three year stipend from the Department of Computer Science at the University of Auckland provided. I would also like to thank Microsoft New Zealand for their interest in this project and their generous scholarship in support of it.

Last but not least I want to thank my family for their continuous love and support during this project. I am particularly grateful for my mother's support with the household and her loving care for my daughters Melanie and Bianca while I was busy writing.

Auckland, April 2008 Raimund Eimann

Contents

1 Introduction **13**

1.1 Structure of this thesis 15

1.2 Other work in the area/literature 17

2 What's in a string? - Information measures **21**

2.1 General Remarks 21

2.2 Complexity measures 23

2.2.1 Kolmogorov-Chaitin complexity 23

2.2.2 LZ production complexity 24

2.2.3 LZ77 26

2.2.4 LZ78 28

2.2.5 T-complexity 29

2.3 Entropy measures 34

2.3.1 Entropy in Thermodynamics and Information Theory 34

2.3.2 Shannon entropy 34

2.3.3 N-gram entropy 35

2.3.4 Entropy estimators 36

2.4 Commonalities and differences between the two approaches 36

2.5 T-information and T-entropy 38

2.6 Concluding remarks 40

3 Network Events 41

3.1 Possible information measurement points within a network 41

3.2 Events and non-events 42

3.3 Networks considered in this thesis 43

3.3.1 Trace files collected at the UoA border gateway 44

3.3.2 Additional trace files collected at an internal campus gateway 45

3.3.3 Third-party trace files: NLANR PMA Special Traces Archive 46

3.4 Events considered in this thesis 46

3.4.1 Distributed Denial of Service (DDoS) attacks 48

3.4.2 Spam 49

3.4.3 Windows Messenger spam 50

3.4.4 Scans 51

3.4.5 Worms 52

3.4.6 SSH dictionary attack 52

3.4.7 SYN flood attacks 53

3.4.8 ICMP floods 54

3.4.9 Failure of network equipment 54

3.5 Conclusions 55

4 Entropy of Network Traffic 57

4.1 Capturing network traffic 58

4.2 T-entropy of raw IPv4 traffic 59

4.3 T-entropy of filtered IPv4 traffic 63

4.4 T-entropy of IPv4 traffic with added artificial events 65

4.4.1 Generating artificial events for injection into a trace file 65

4.4.2 Example: injection of an artificial SYN flood 68

4.5 Conclusions 69

5 Noise 71

5.1 Types of noise 71

5.2 Signal-to-noise ratio 73

6 Packet Mapping 77

6.1 Motivation 77

6.2 Field selection 80

6.3 Simple mapping 82

6.4 Fields available for mapping 84

6.4.1 PCAP packet header fields 85

6.4.2 Ethernet header fields 87

6.4.3 IPv4 header fields 88

6.4.4 TCP header fields 97

6.4.5 UDP header fields 100

6.4.6 Payload . . . 100

6.5 Conditional field selection for mappings . . . 100

6.6 More complex mappings . . . 102

6.7 Conclusions . . . 106

7 Traffic sampling 107

7.1 Motivation . . . 107

7.2 Volume-based sampling . . . 108

7.3 Interval-based sampling . . . 110

7.4 Combo sampling . . . 111

7.5 Window size considerations . . . 113

7.5.1 Effects of window size variation . . . 113

7.5.2 Window size optimisation . . . 116

8 T-Entropy Analysis of Network Traffic 119

8.1 SYN flood attacks . . . 120

8.2 DDoS attacks . . . 121

8.3 Windows Messenger spam . . . 121

8.4 Worms . . . 122

8.5 Scans . . . 125

8.6 SSH dictionary attacks . . . 126

8.7 ICMP floods . . . 128

8.8 Simulation of network equipment failures . . . 129

8.9 Conclusions 134

9 Sensitivity 135

9.1 Sensitivity of T-entropy to network events at different packet rates 135

9.2 Sensitivity comparisons between different measures 137

9.3 Conclusions 140

10 Multi-Dimensional Traffic Analysis 143

10.1 Motivation 143

10.2 Traffic analysis with multiple mappings 144

10.3 Traffic analysis with multiple measures 146

10.4 Concluding remarks 153

11 Conclusions 155

11.1 Results 156

11.2 Open problems 157

11.3 Future work 157

A Appendix 159

A.1 Software 159

A.1.1 Third party software used for trace file capture and processing 160

A.1.2 Software developed in the course of the project 160

A.2 Filter mechanism used in the software developed during this project 170

A.3 Details of mappings used in this thesis 171