Collecting and Analysing Traffic Profiles within a Commercial Network

Prof. D. J. Parish
P. Sandford
High Speed Networks Group

Loughborough University

Presentation Summary

- Overview of the ‘Detecting and Preventing Criminal Activities on the Internet’ project
- Discussion of Architecture
- Example Network Patterns and Anomalies

Overview of the Network Abuse Detection Project

- EPSRC Funded
- Partnered by NTL, CESG and SPSS
- 3 Years, starting April 2004

Loughborough University ntl: SPSS CESG

Project Objectives

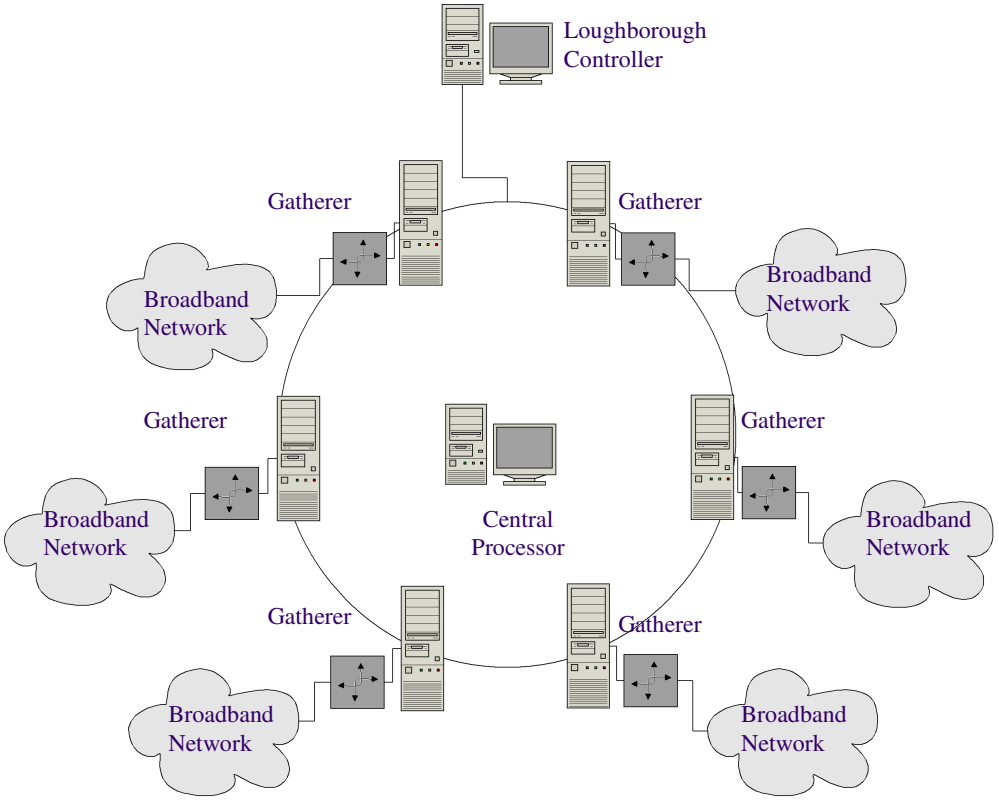
- Identify illegal activity inside the network core
  - Sometimes Necessary
  - Some prevention best done here
- Use statistical traffic summaries of headers to identify anomalies
  - Processing Overhead
  - High Throughput
  - No user data
  - Cheap Equipment

Loughborough University ntl: SPSS CESG

Approach Summary

- Anomaly Based
  - Describing Normality
  - Classifying Deviation
- DataMining
  - Identify Relationships
  - Update view of normality

Approach Summary (Cont.)



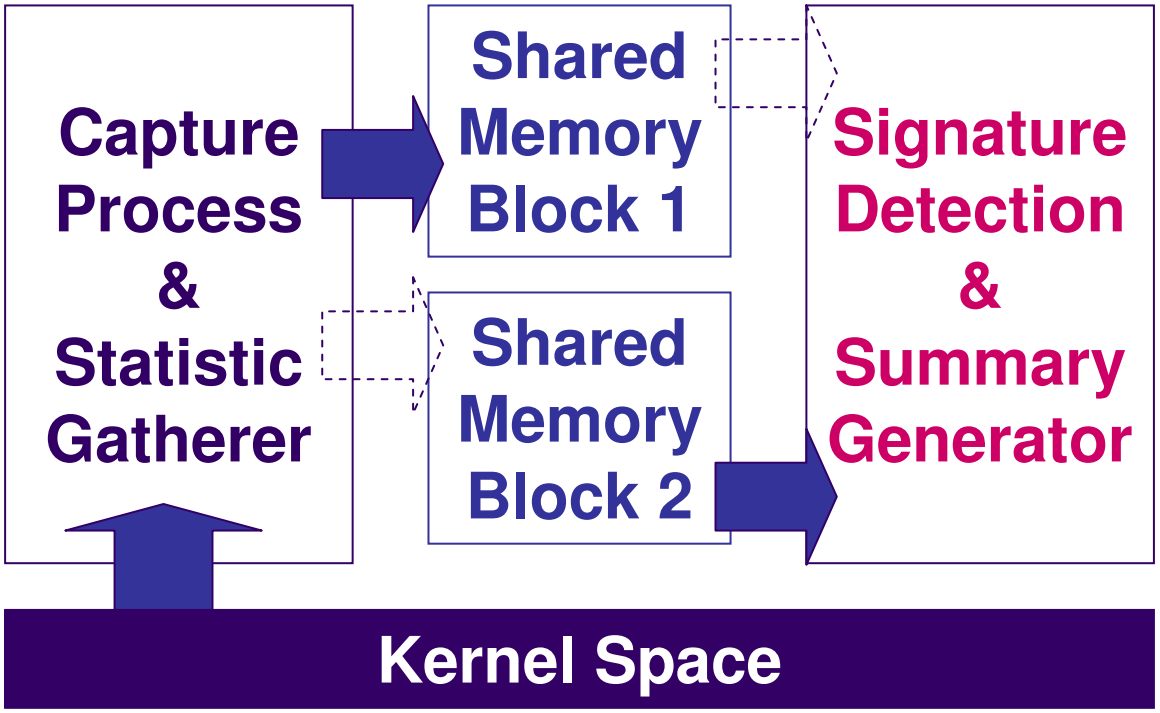
Loughborough University ntl: SPSS CESG

Current Status

- Hardware/Basic software installed
  - 6 Monitored PoPs
  - 1 Data-Mining Engine
  - Control from Loughborough/NTL
- Modules In Place For
  - Summary Gathering
  - Base lining data
  - Simple Signature Detection
  - Basic Alerting
  - Outgoing spam email detection

Loughborough University ntl: SPSS CESG

PoP Software Summary



Loughborough University ntl: SPSS CESG

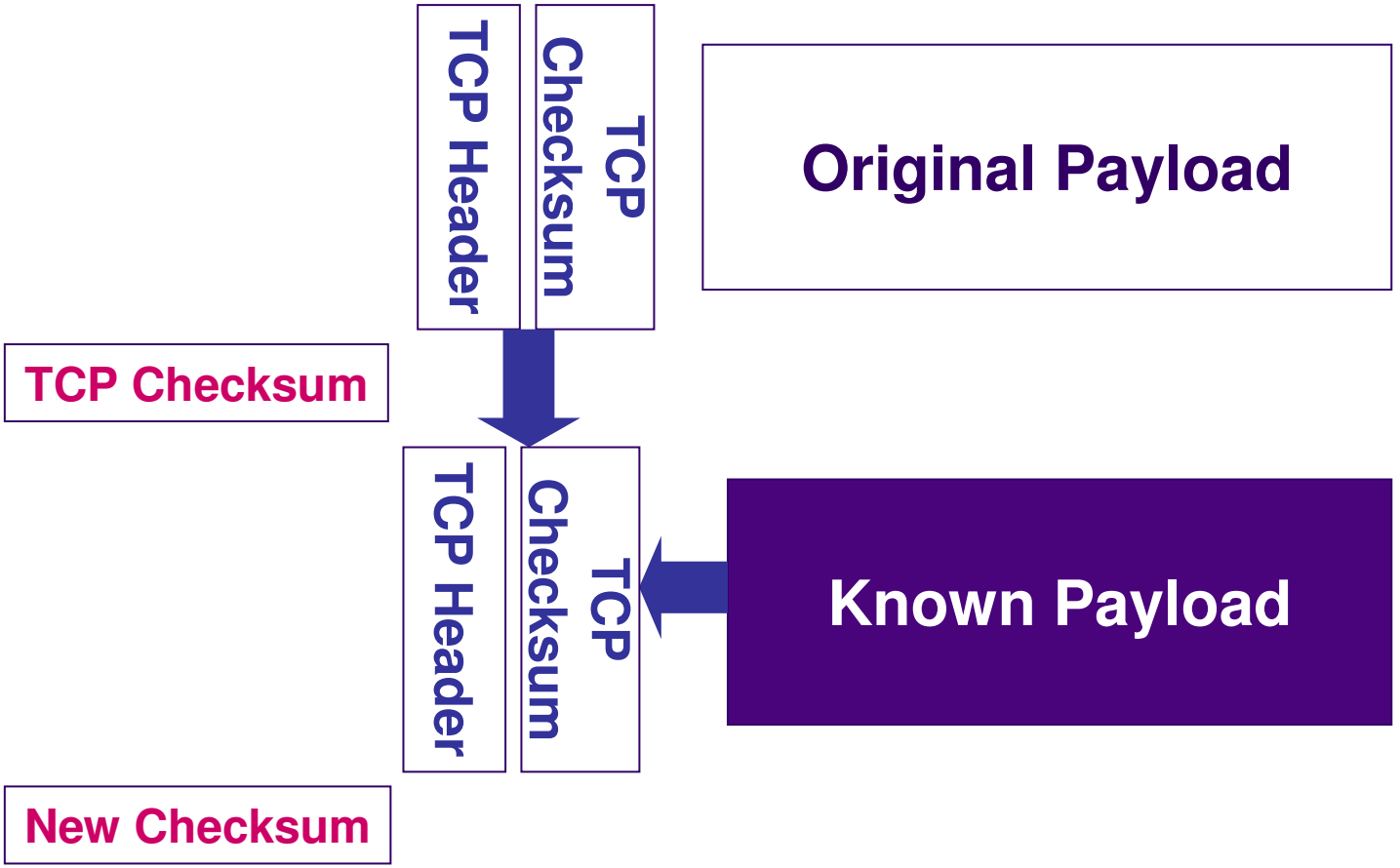
PoP Software Summary

- Fast Packet Processing
  - Purely Counter Increments
- Pseudo Real-Time Statistics
  - Processing Time Constant
  - Variable Statistic Window Size

Note on Capture Interfaces

- Dealing With Interrupts
  - Poll
  - Buffer
- Context Switching
  - Memory Mapping

Signature Detection Without Data



Loughborough University ntl: SPSS CESG

Data Mining - Example

- TTL Field
  - Initially Thought to be Limited Use
  - Data Mining Highlighted TTL in Lab Tests

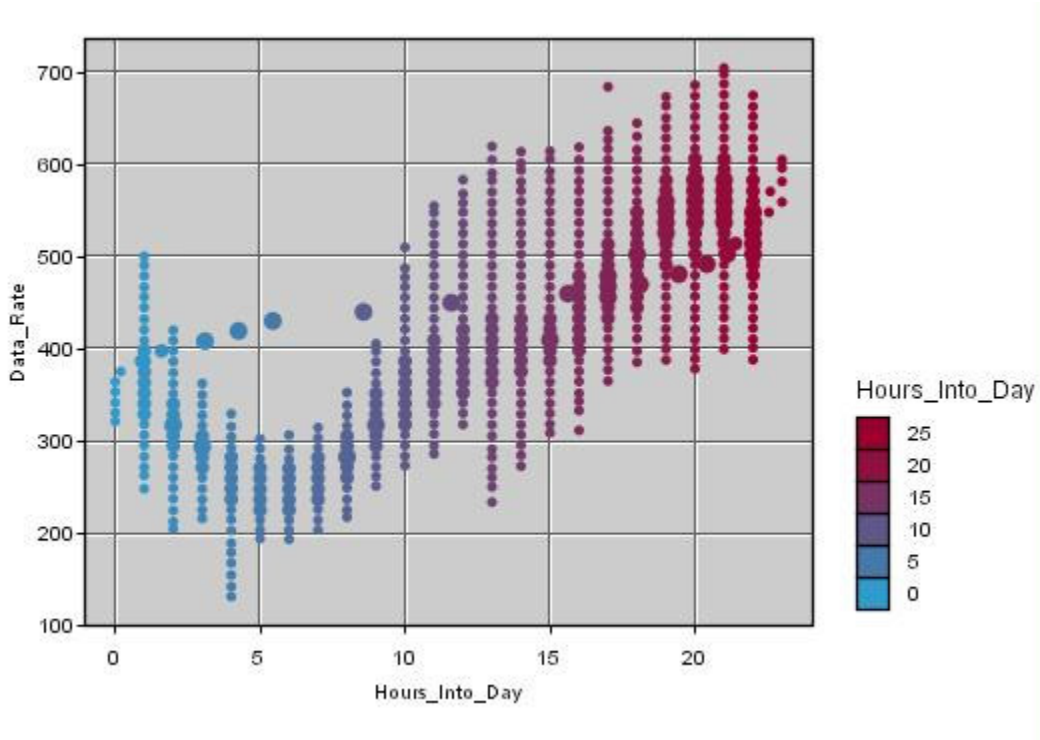
Data Mining - Example

- TTL Field
  - Default Values (Depending on OS)
  - Consistent by Default
  - Can be used to Identify Spoofing
  - Shows Daily Pattern (Windows / Linux, File Sharing / Web Browsing)

Patterns

- Data Rates
- Port Numbers (Applications)
- An Anomaly Example

Data Rates

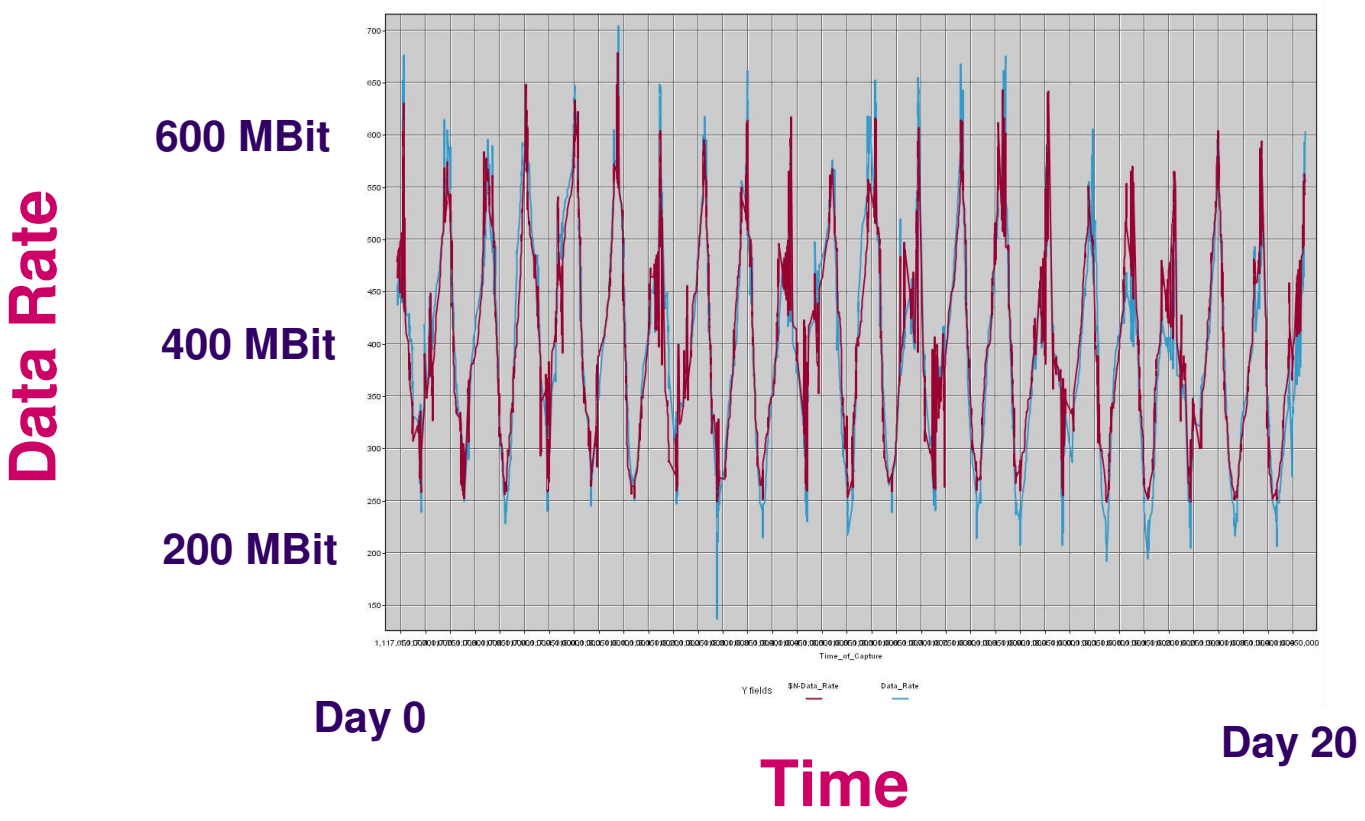


Loughborough University ntl: SPSS CESG

Data Rates (Cont.)

- Time of Day Variation
  - Peak Times
    - Late evening (8pm)
  - Low Times
    - Early Morning (5am)

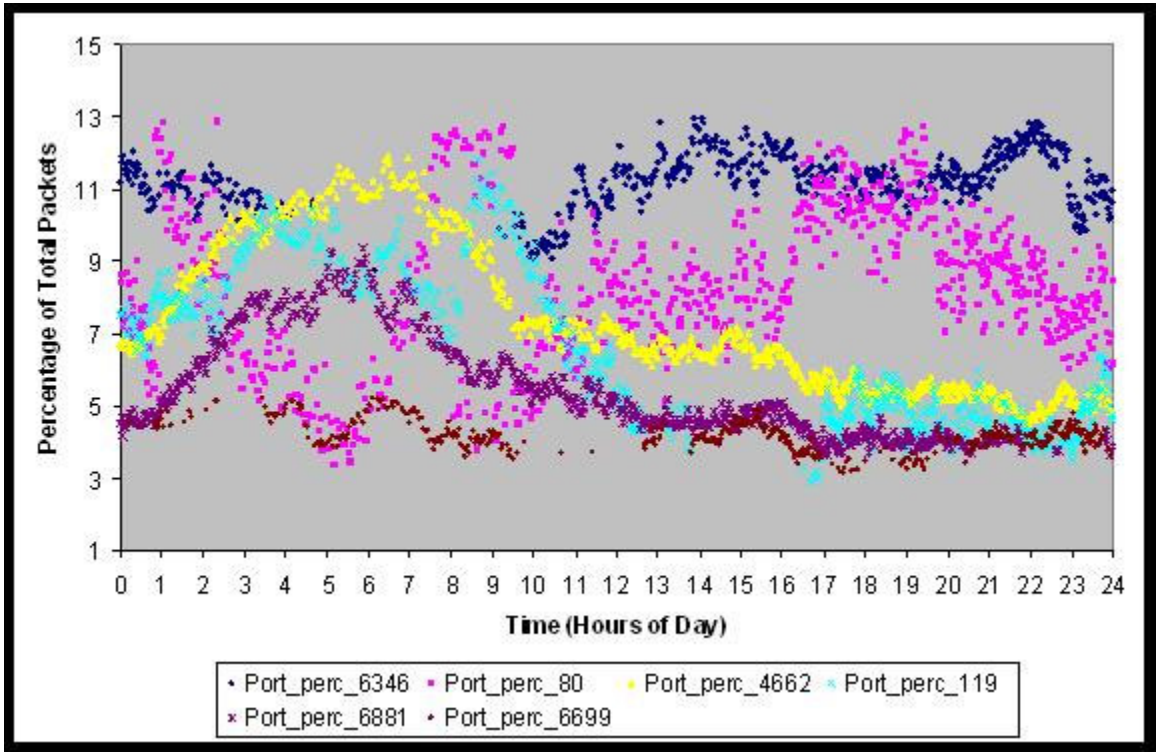
Model of Data Rate



Port Numbers

Port	Common Application	Percentage of All Packets	
		June 05	January 05
80	HTTP	8.8%	6.5%
6346	Gnutella	7.9%	4.893%
4662	eMule	6.4%	10.269%
119	NNTP / UseNet	5.9%	5.3%
6881	BitTorrent	4.1%	3.55%
6699	WinMX	2.3%	7.178%
6348	Gnutella (Secondary Port)	0.1%	0.01%

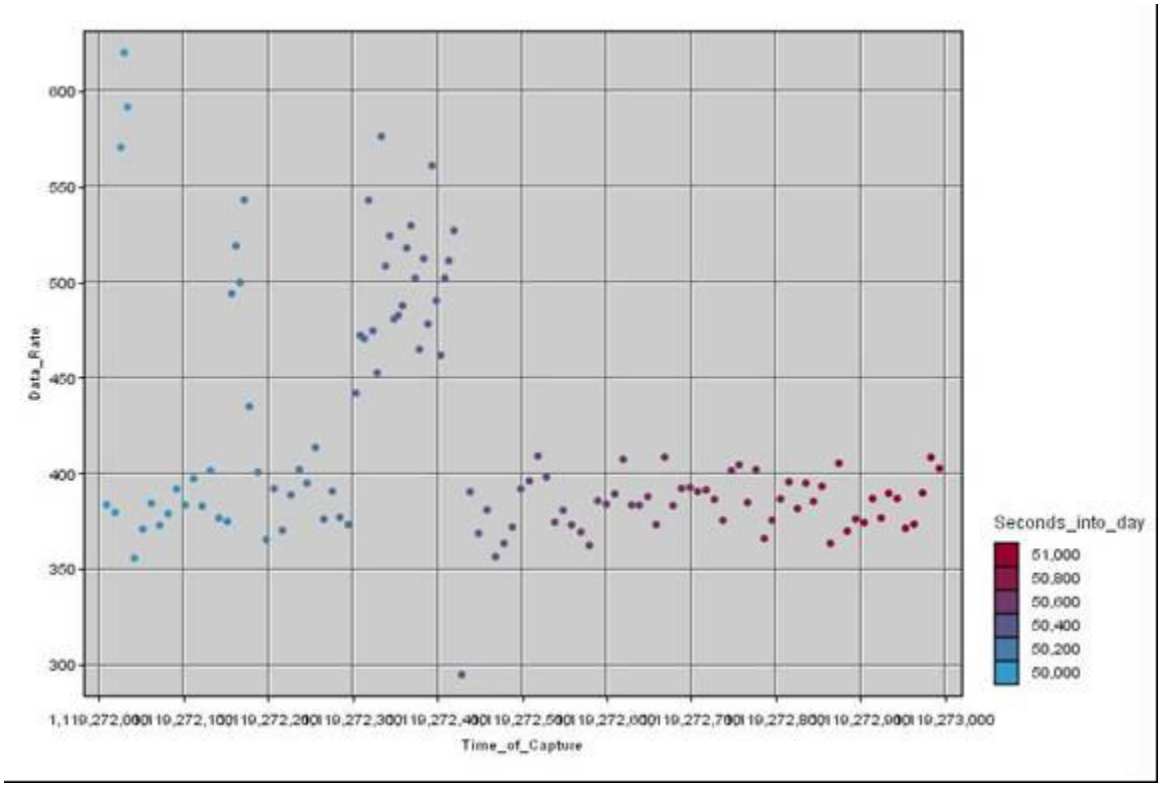
Port Numbers (Cont.)



Loughborough University ntl: SPSS CESG

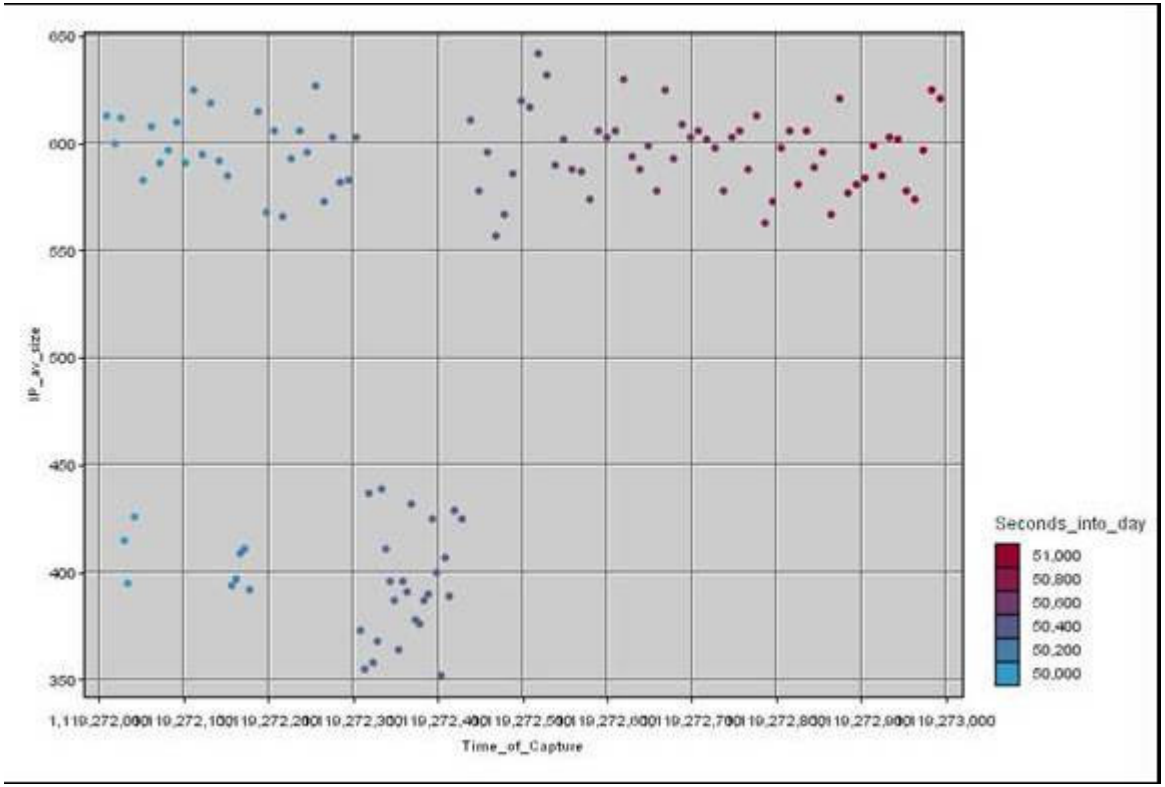
Anomaly Example

Anomaly Example – Data Rate



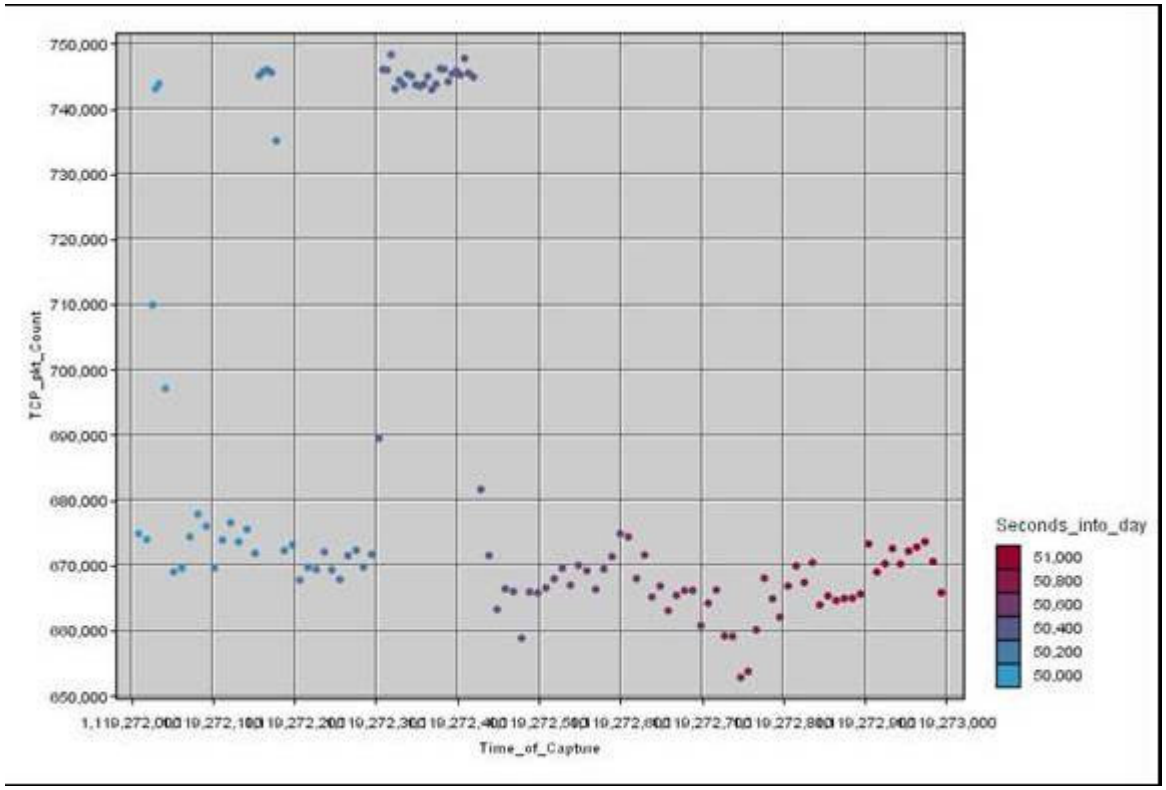
Loughborough University ntl: SPSS CESG

Anomaly Example – Average Packet Size



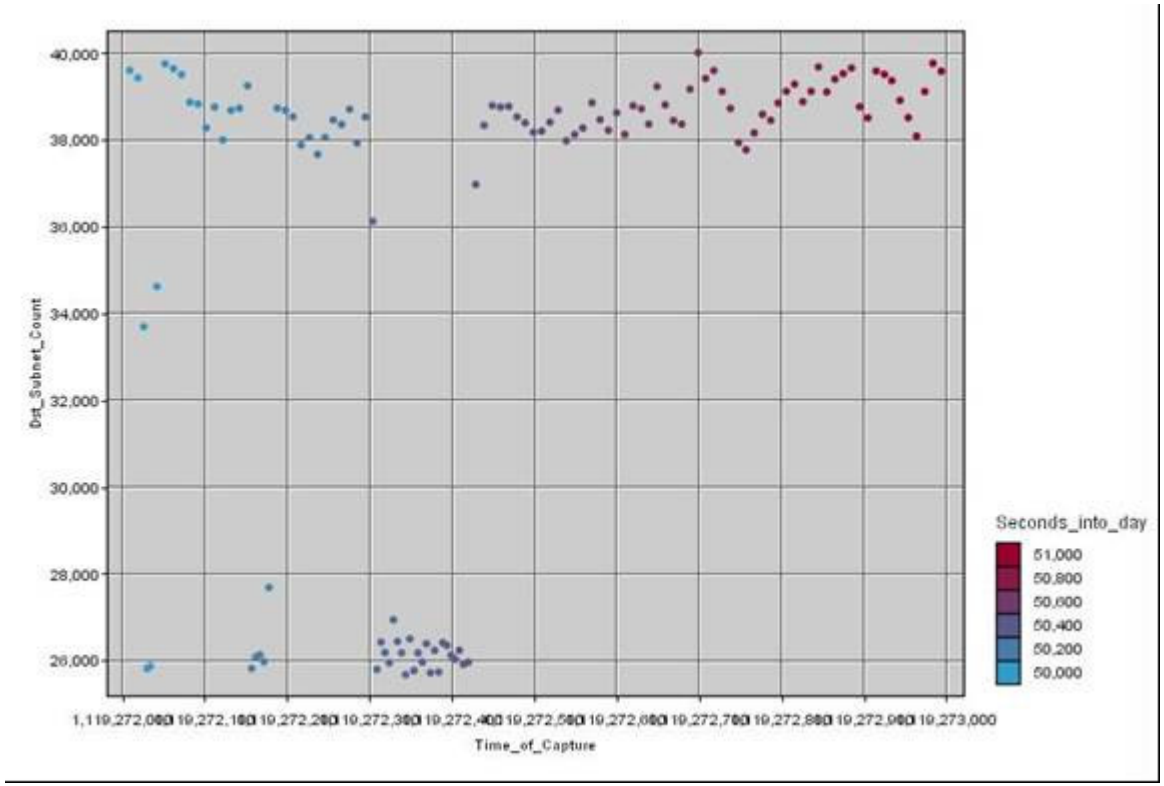
Loughborough University ntl: SPSS CESG

Anomaly Example – TCP Packet Count



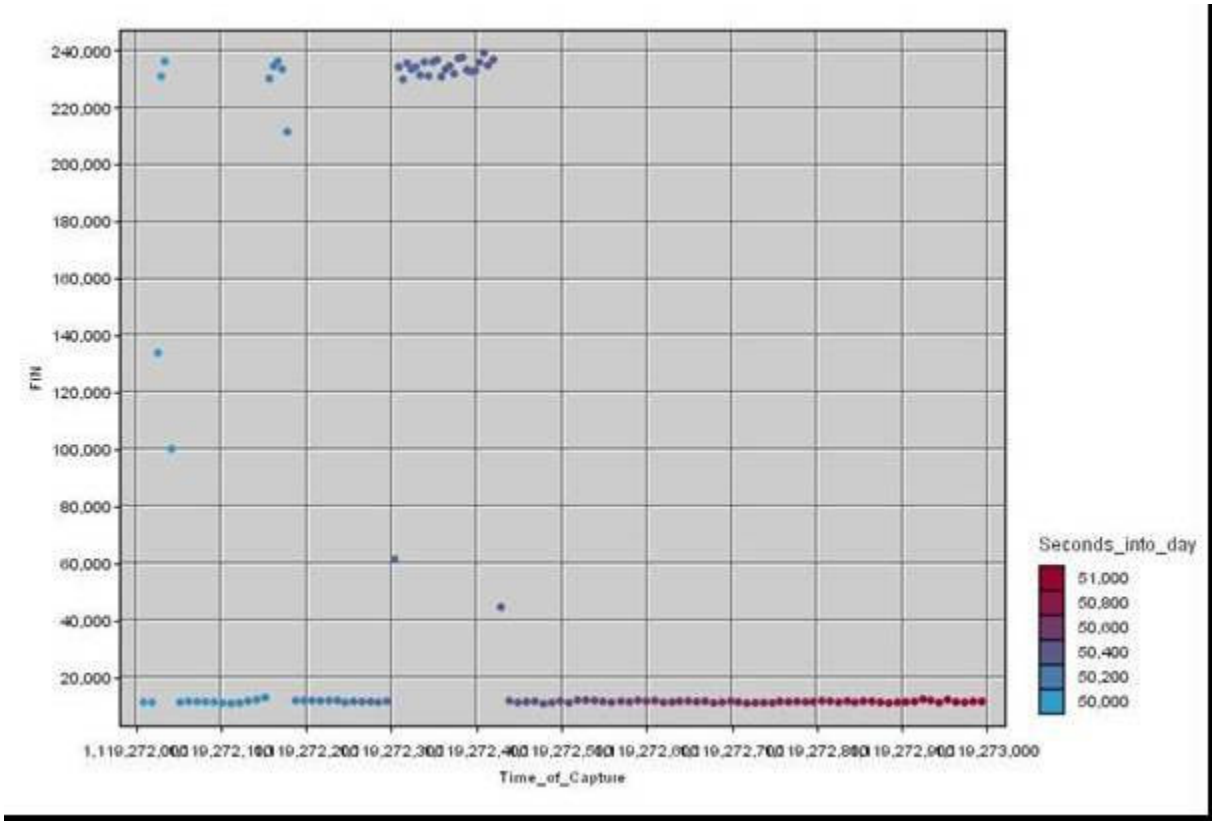
Loughborough University ntl: SPSS CESG

Anomaly Example – Destination Count



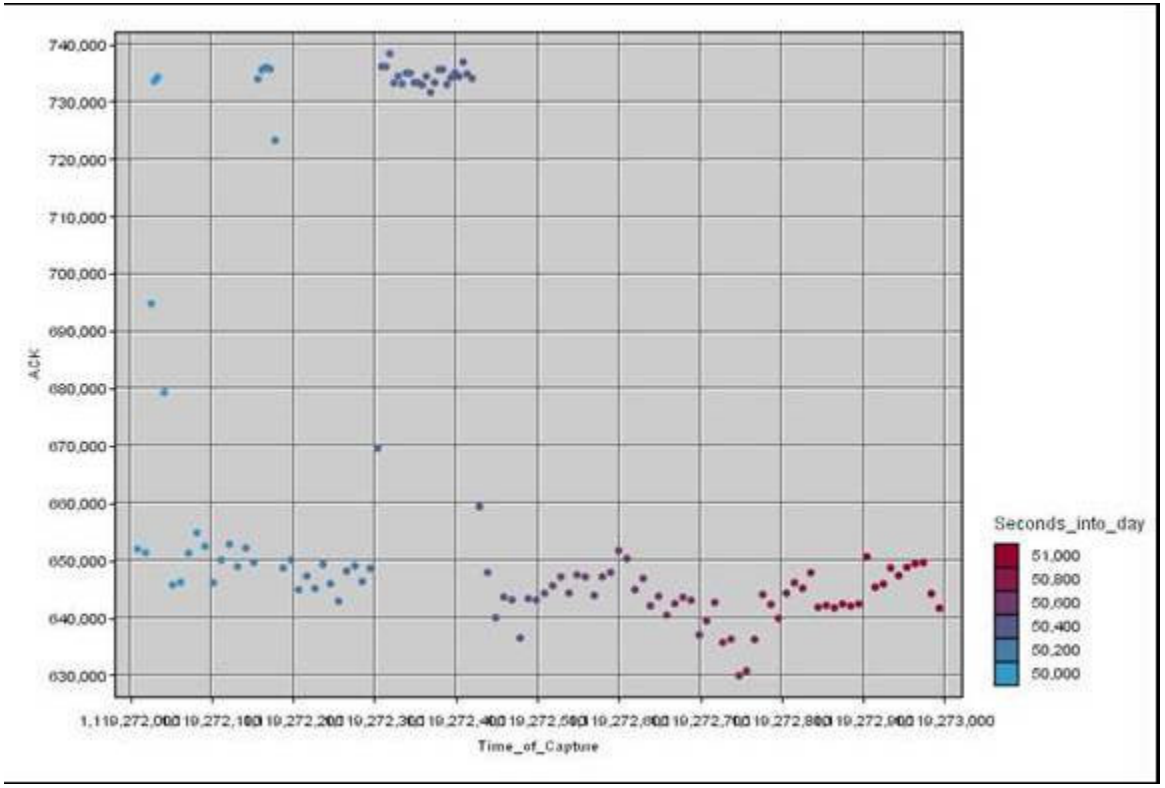
Loughborough University ntl: SPSS CESG

Anomaly Example – FIN Count



Loughborough University ntl: SPSS CESG

Anomaly Example – Ack Count

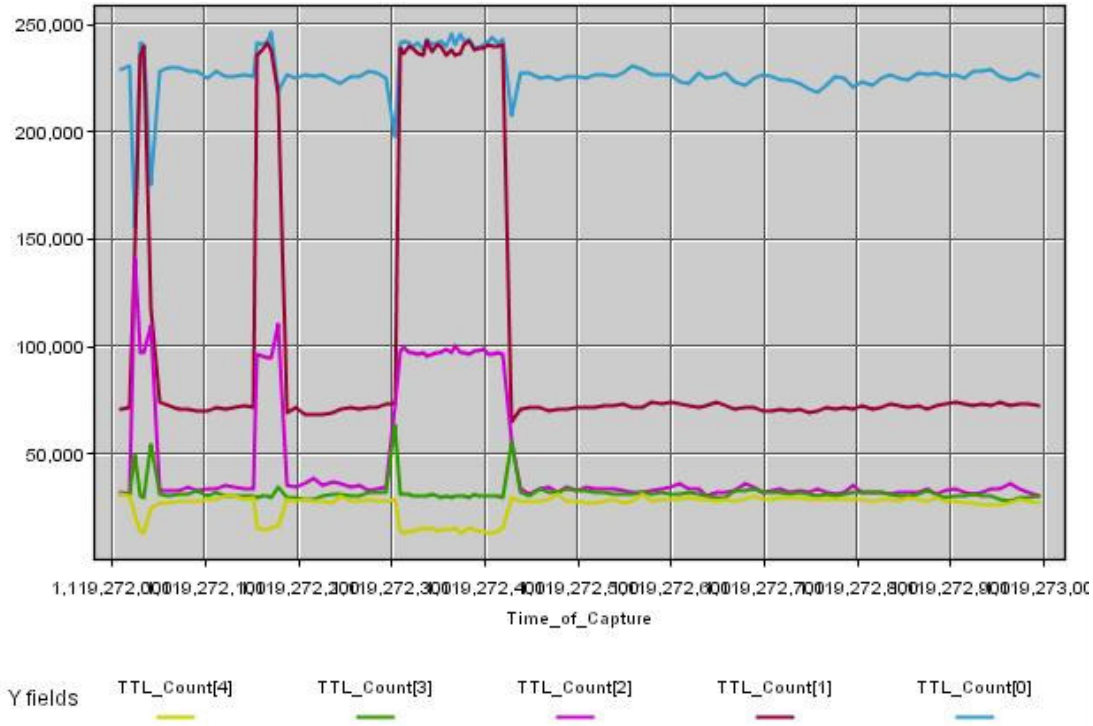


Loughborough University ntl: SPSS CESG

Anomaly - Conclusion

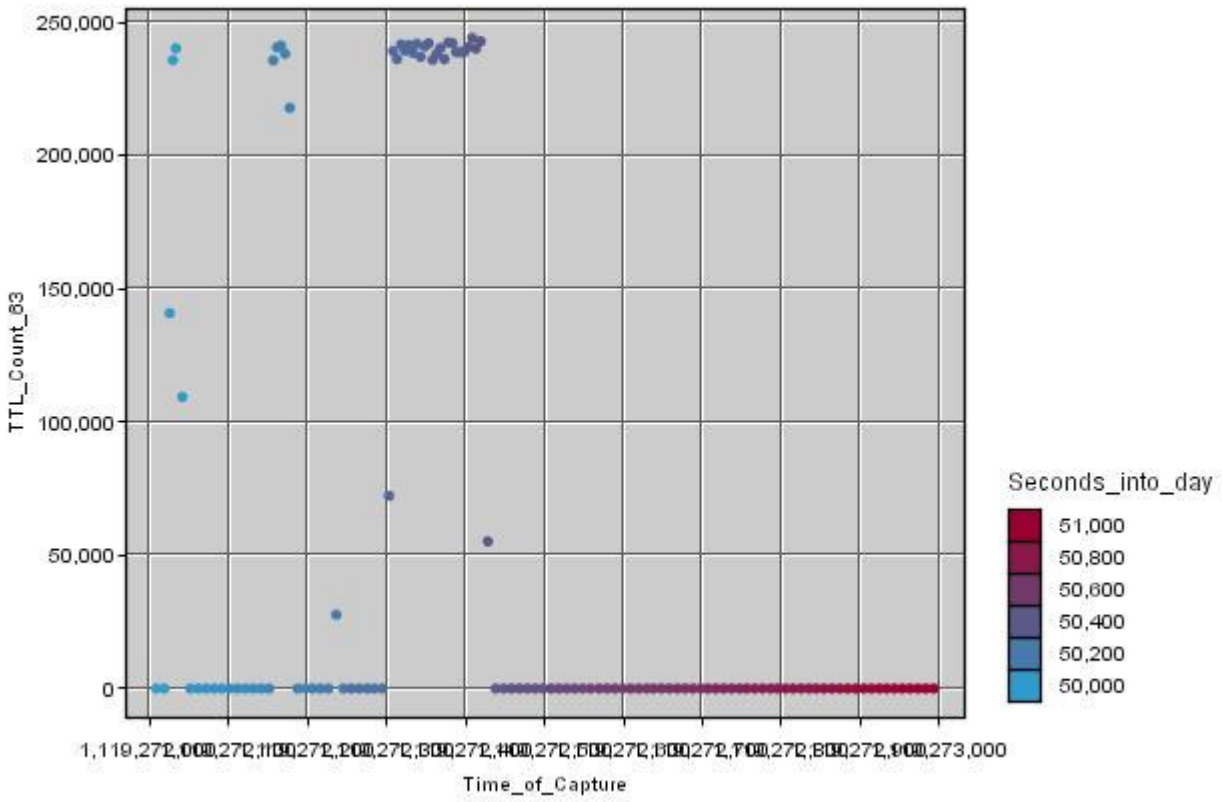
- Large Traffic Amounts
- Synchronised
- TCP from many sources
- Small number of destinations
- FIN/ACK Packets

Anomaly TTL



Loughborough University ntl: SPSS CESG

Anomaly TTL - 63



Summary

- System in Place
  - Monitoring High Data Rates
  - Modelling Network Patterns
  - Discovering Deviations from ‘Normal’
- Architecture Design
  - Scalable
  - Functional

Future Work

- Cross-site Correlation
- Automated Alerting
- Investigation of Mitigation Techniques

Questions