F

A complete system integration of stream-based IP flow-record querier

VAIBHAV BAJPAI

Masters Thesis

School of Engineering and Science
Jacobs University Bremen
Bremen, Germany

June 2012

[January 10, 2012 at 15:00]

ABSTRACT

Short summary of the contents in English...

We have seen that computer programming is an art,
because it applies accumulated knowledge to the world,
because it requires skill and ingenuity, and especially
because it produces objects of beauty.

— **?** [**?**]

ACKNOWLEDGMENTS

Put your acknowledgments here.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

ACRONYMS

Part I

INTRODUCTION

You can put some informational part preamble text here

1 TRAFFIC MEASUREMENT APPROACHES

1.1 CAPTURING PACKETS

1.2 CAPTURING FLOWS

1.3 REMOTE MONITORING

1.4 REMOTE METERING

2 FLOW EXPORT PROTOCOLS

2.1 NETFLOW

2.2 IPFIX

2.3 SFLOW

3 LANGUAGES AND TOOLS

3.1 SQL-BASED QUERY LANGUAGES

3.1.1 *NetFlow exports as relational DBMS*

3.1.2 *Data Stream Management System*

3.1.3 *Gigascope*

3.1.4 *Tribeca*

3.2 FILTERING LANGUAGES

3.2.1 *flow-tools*

3.2.2 *nfdump*

3.3 PROCEDURAL LANGUAGES

3.3.1 *FlowScan*

3.3.2 *Clustering NetFlow Exports*

3.3.3 *SiLK Analysis Suite*

4 LEGAL CONSIDERATION

Part II

STATE OF THE ART

You can put some informational part preamble text here

5 FLOWY

5.1 PROCESSING PIPELINE

5.1.1 *Splitter*

5.1.2 *Filter*

5.1.3 *Grouper*

5.1.4 *Group-Filter*

5.1.5 *Merger*

5.1.6 *Ungrouper*

5.2 PYTHON FRAMEWORK

5.2.1 *PyTables and PLY*

5.2.2 *Records*

5.2.3 *Filters and Rules*

5.2.4 *Branches and Branch Masks*

6 FLOWY IMPROVEMENTS USING MAP/REDUCE

7

FLOWY 2.0

8 FLOWY: APPLICATIONS

8.1 IPv6 transition failure identification

8.2 Cybermetrics: user identification

8.3 Application identification using flow signatures

8.4 TCP level spam detection

Part III

MOTIVATION

Part IV

WORK PLAN

You can put some informational part preamble text here

9 DESIGN

10 IMPLEMENTATION

11 PERFORMANCE EVALUATION

12 CONCLUSION

Part V

IMPLEMENTATION AND EVALUATION

You can put some informational part preamble text here

13 DESIGN

14 IMPLEMENTATION

15 PERFORMANCE EVALUATION

16

FUTURE WORK

17

CONCLUSION

Part VI

APPENDIX

A APPENDIX

Put your appendix here.

COLOPHON

This thesis was typeset with LaTeX 2_ε using Hermann Zapf's *Palatino* and *Euler* type faces (Type 1 PostScript fonts *URW Palladio L* and *FPL* were used). The listings are typeset in *Bera Mono*, originally developed by Bitstream, Inc. as "Bitstream Vera". (Type 1 PostScript fonts were made available by Malte Rosenau and Ulrich Dirr.)

The typographic style was inspired by **?** 's genius as presented in *The Elements of Typographic Style* [**?**]. It is available for LaTeX via CTAN as "`thesis`".

NOTE: The custom size of the textblock was calculated using the directions given by Mr. Bringhurst (pages 26–29 and 175/176). 10 pt Palatino needs 133.21 pt for the string "abcdefghijklmnopqrstuvwxyz". This yields a good line length between 24–26 pc (288–312 pt). Using a "*double square textblock*" with a 1:2 ratio this results in a textblock of 312:624 pt (which includes the headline in this design). A good alternative would be the "*golden section textblock*" with a ratio of 1:1.62, here 312:505.44 pt. For comparison, `DIV9` of the `typearea` package results in a line length of 389 pt (32.4 pc), which is by far too long. However, this information will only be of interest for hardcore pseudo-typographers like me.

To make your own calculations, use the following commands and look up the corresponding lengths in the book:

```
\settowidth{\abcd}{abcdefghijklmnopqrstuvwxyz}
\the\abcd\ % prints the value of the length
```

Please see the file `thesis.sty` for some precalculated values for Palatino and Minion.

145.86469pt

DECLARATION

Put your declaration here.

Bremen, Germany, June 2012

Vaibhav Bajpai