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 21:59]

ABSTRACT

Short summary of the contents in English...

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

ACRONYMS

IPFIX Internet Protocol Flow Information Export

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

Flowy [2][3] is the first prototype implementation of a stream-based flow record query language [4][1][5]. The query language allows to describe patterns in flow-records in a declarative and orthogonal fashion, making it easy to read and flexible enough to describe complex relationships among a given set of flows.

5.1 PROCESSING PIPELINE

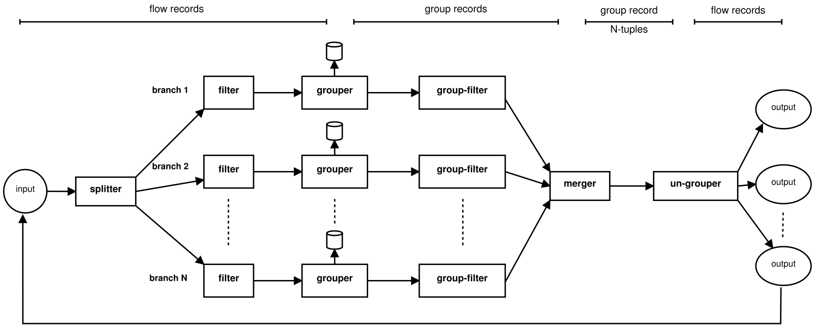

Figure 1: Flowy: Processing Pipeline [1]

The pipeline consists of a number of independent processing elements that are connected to one another using UNIX-based pipes. Each element receives the content from the previous pipe, performs an operation and pushes it to the next element in the pipeline. Figure 1 shows an overview of the processing pipeline. The flow record attributes used in this pipeline exactly correlate with the attributes defines in the Internet Protocol Flow Information Export (IPFIX) Information Model specified in RFC 5102 [6]. A complete description on the semantics of each element in the pipeline can be found in [4]

5.1.1 *Splitter*

The `splitter` takes the flow-records data as input in the `flow-tools` compatible format. It is responsible to duplicate the input data out to several branches without any processing whatsoever. This allows each of the branches to have an identical copy of the flow data to process it independently.

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.

BIBLIOGRAPHY

[1] Vladislav Marinov and Jürgen Schönwälder. Design of a Stream-Based IP Flow Record Query Language. In *Proceedings of the 20th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management: Integrated Management of Systems, Services, Processes and People in IT*, DSOM '09, pages 15–28, Berlin, Heidelberg, 2009. Springer-Verlag.

[2] Kaloyan Kanev. Flowy - Network Flow Analysis Application. Master's thesis, Jacobs University Bremen, Campus Ring 1, 28759 Bremen, Germany, 2009.

[3] Kaloyan Kanev, Nikolay Melnikov, and Jürgen Schönwälder. Implementation of a stream-based IP flow record query language. In *Proceedings of the Mechanisms for autonomous management of networks and services, and 4th international conference on Autonomous infrastructure, management and security*, AIMS'10, pages 147–158, Berlin, Heidelberg, 2010. Springer-Verlag.

[4] Vladislav Marinov. Design of an IP Flow Record Query Language. Master's thesis, Jacobs University Bremen, Campus Ring 1, 28759 Bremen, Germany, 2009.

[5] Vladislav Marinov and Jürgen Schönwälder. Design of an IP Flow Record Query Language. In *Proceedings of the 2nd international conference on Autonomous Infrastructure, Management and Security: Resilient Networks and Services*, AIMS '08, pages 205–210, Berlin, Heidelberg, 2008. Springer-Verlag.

[6] J. Quittek, S. Bryant, B. Claise, P. Aitken, and J. Meyer. Information Model for IP Flow Information Export. RFC 5102 (Proposed Standard), January 2008. Updated by RFC 6313.

DECLARATION

Put your declaration here.

Bremen, Germany, June 2012

Vaibhav Bajpai