ITSA
INTERNET TRAFFIC STATISTICS ARCHIVE

Martijn Hoogesteger
Ricardo de O. Schmidt
Aiko Pras
HAVE YOU EVER...

- Assumed HTTP accounts for most of our Internet traffic?
- Wondered if HTTPS is slowly taking over more and more HTTP traffic?
- Been interested in the differences in Internet traffic between countries or continents?
- Thought about how Internet traffic grows?
INTRODUCTION & MOTIVATION

Internet2 NetFlow: Weekly Reports

NetFlow data from all core routers of the Internet2 network are analyzed to produce weekly reports of use of the network. In essence, this is a weekly version of the "Bulk TCP Use and Performance on Internet2" by Stanislav Shalunov and Benjamin Teitelbaum (note, however, that the paper used one day of data while the reports use a week of data each and that the paper used data from the busiest router while the reports present a network-wide view).

The reports are produced automatically using programs written by Arne E. Jarp and Stanislav Shalunov. The CWEB program nfstat.w reads large daily files and produces intermediate results that are further digested by a collection of Perl programs.

Weekly reports for weeks starting on the following dates are available:

- 20100426
- 20100419
- 20100412
- 20100405
- 20100328
- 20100321
- 20100314
- 20100308

A new source of information is needed.
INTERNET TRAFFIC STATISTICS ARCHIVE (ITSA)
A NEW SYSTEM

ISP's routers
University's routers

NetFlow
IPFIX

Flow data collection is always done on the data provider's end.

Executed every 7 / 14 / 30 days

Scripts to collect flow data
Process flow data and generate traffic statistics

JSON Report File

Statistics storage in local database

Scripts to retrieve statistics-data from database

Publish traffic statistics in public website (tables and charts)

The generated report file, and consequently, the eventual statistics, contain no privacy-sensitive data.

UNIVERSITY OF TWENTE.
ITSA
DATA PROVIDERS

ISP's routers
University's routers
NetFlow
IPFIX

Flow data collection is always done on the data provider's end.

Executed every 7/14/30 days

Scripts to collect flow data
Process flow data and generate traffic statistics

JSON Report File

Statistics storage in local database

Published traffic statistics in public website (tables and charts)

Scripts to retrieve statistics-data from database

The generated report file, and consequently, the eventual statistics, contain no privacy-sensitive data.
ITSA
DATA PROVIDERS
More data sources

- AARNet
- NorduNet
- AmLight
- ESNet
- More?
ITSA
REPORT GENERATION

Execution cycle:
1. Collect flow data using NetFlow and IPFIX.
2. Process flow data to generate traffic statistics.
4. Retrieve statistics from the database.
5. Publish traffic statistics in a public website (tables and charts).

Flow data collection is always done on the data provider's end. Processing the flow-level data and generating the report file can either be done on the data provider's side, or on our end. The generated report file, and consequently, the eventual statistics, contain no privacy-sensitive data.
ITSA REPORT GENERATION

- Most used ports
- Most used IP-Protocols
- Flow duration
- Packet count
- Byte count
- Transfer rate
- Packets per second
- Bytes per packet
ITSA
TO THE CENTRAL SERVER

Flow data collection is always done on the data provider’s end.

Processing the flow-level data and generating the report file can either be done on the data provider’s side, or on our end.

The generated report file, and consequently, the eventual statistics, contain no privacy-sensitive data.
IT'S A WEBSITE

Report for Week 18

Download report JSON file

Ports distribution (ports with share < 0.5% aggregated in 'Other')

IP Protocols distribution high-end

IP Protocols distribution low-end
- IPv6
- More data sources
- Historical views
- Comparative views
- Aggregated views
QUESTIONS? COMMENTS?

- Martijn Hoogesteger
  m.hoogesteger@student.utwente.nl
- Ricardo de O. Schmidt
  r.schmidt@utwente.nl
- stats.simpleweb.org