

# Abilene Weekly NetFlow Reports

Stanislav Shalunov <shalunov@internet2.edu>

Campus-Focused Workshop, Houston, 2002-04-11

# What is NetFlow?

- Originally a Cisco proprietary technology
- Now supported by other vendors and being standardized in the IETF
- Used to be a method to speed up packet forwarding
  - Cache the next interface for given  $\langle \text{src}, \text{dst}, \text{proto}, \text{ports}, \text{tos} \rangle$
  - Look in the cache hash table before you consult the routing table
- It was realized it's useful for accounting purposes
- Not any longer used for optimization, but the accounting use is growing more widespread

## NetFlow version 5

- There are several versions of NetFlow (5, 7, 8 are widespread)
- Provide different fields and different levels of aggregation
- NetFlow v5 gives you records with the following:  
src\_ip, dst\_ip, packets, octets, start\_time, end\_time,  
src\_port, dst\_port, proto, tos, src\_as, dst\_as, if\_in, if\_out
- Can use for accounting purposes, but there's more

## Methodology of Data Collection

- Collect 1% sampled NetFlow data from all core Abilene routers
- Collection done at ITEC-Ohio with `flow-tools`
- Throw away data coming from interfaces between core nodes
- `flow-tools` now include SNMP hooks for that
- Concatenate the rest of the data
- Ship the resulting files (5–9 GB) to our RAID array daily
- Resulting view treats Abilene as a single data-forwarding unit

## Methodology of Data Processing

- The goal is to capture long-term trends
- Weekly averages for everything, hence weekly reports
- Daily averaging too volatile, monthly would take too long
- Two data sets: one the complete thing, one “bulk TCP”
- Bulk TCP = TCP connection that transferred  $> 10$  MB
- For full data set can do traffic composition
- For bulk TCP data set can do more, including throughput

## Data Presentation

- Find it at <http://netflow.internet2.edu/weekly/>
- Weekly: new report added, time-series graphs updated
- The heart: TCP throughput analysis (includes CDF)
- Time-series graphs already start to look interesting
- Traffic composition: finally you know what uses Abilene
- Numbers of interest:
  - Median bulk TCP throughput is slightly less than 2 Mb/s
  - 95th percentile is around 6 Mb/s
  - Roughly half of traffic is file sharing

## Costs, Tools Used

- Capacity overhead of data collection is negligible
- Need a machine with disk space (\$100 for 40 GB now)
- FOTS (free off the shelf) `flow-tools` for collection
- Custom-written stuff for analysis (around 2 man/month)
  - CWEB program to make a pass over complete data set
  - Perl programs to post-process and handle presentation
- CWEB part is available as documentation of classification
- Perl part can be released if there's interest

# Web Show