Macro- and Microscopic Analysis of the Internet Economy from Network Measurements

Jakub Mikians

UPC BarcelonaTech

March 3, 2015
1 Introduction

2 Macroscopic view: Interdomain Traffic Matrix
   - Introduction
   - Paper I: Characterizing ITM
   - Paper II: Synthesizing ITM

3 Microscopic view: Price Discrimination
   - Introduction
   - Paper III: Detecting price discrimination
   - Paper IV: Crowd assisted search of PD

4 Conclusions and Further Work

5 Contributions
Introduction

2 Macroscopic view: Interdomain Traffic Matrix
   • Introduction
   • Paper I: Characterizing ITM
   • Paper II: Synthesizing ITM

3 Microscopic view: Price Discrimination
   • Introduction
   • Paper III: Detecting price discrimination
   • Paper IV: Crowd assisted search of PD

4 Conclusions and Further Work

5 Contributions
Introduction

Scale of the Internet economy

- Scale of the Internet and its economy
  - One fifth of the global GDP growth in recent years
  - 75% of Internet economic impact comes from traditional industries
- Share in global GDP between 3.4% and 4.1%
  1. United States
  2. China
  3. Japan
  4. Germany
  5. INTERNET
  6. ...
Some biggest players present in the ICT\(^1\) market before the Internet
- AT&T
- Comcast
- ...

... other companies are children of the digital economy
- Google
- Amazon
- ...

Interactions between those largest players shape the Internet economy at the *macro scale*  

\(^1\)Information and Communications Technology
Introduction

Regular user

- On the other side – a regular user of the Internet
- For a large user base, Internet is an important place of work, retail and social interaction
- 40% world population online
- Cumulative decisions of the users impact network economics

- At the same time the users generate a wide spectrum of personal information. This information is desired by the online marketing companies and e-retailers
- Interactions between the users, retailers and service providers contribute the Internet economy at micro scale
In this thesis we look at the Internet economy from two different perspectives:

- **Macro-scale:**
  - Flow of the traffic is directly related to flow of the money between ASes.
  - We examine traffic flowing between AS-es.
  - Characterize traffic between AS-es.
  - Propose method to generate synthetic traffic matrices.

- **Micro-scale:**
  - Investigate economic phenomenon at the intersection of the user’s personal information and retail business - *price discrimination* (PD).
  - Will look for the empirical evidences that PD exists on the Internet.
  - We present a feasible and scalable approach to investigate PD – crowd sourcing.
1 Introduction

2 Macroscopic view: Interdomain Traffic Matrix
   - Introduction
     - Paper I: Characterizing ITM
     - Paper II: Synthesizing ITM

3 Microscopic view: Price Discrimination
   - Introduction
     - Paper III: Detecting price discrimination
     - Paper IV: Crowd assisted search of PD

4 Conclusions and Further Work

5 Contributions
Macroscopic view - ITM

Introduction

- AS-level, the highest level of organization of the Internet
- Traffic flowing between AS-es can be described by the Interdomain Traffic Matrix (ITM)
- ITM describes traffic between the largest business entities, therefore it is directly related to the network macroeconomics
- Insight into ITM → insight into the Internet macro economy

- Knowledge of the traffic - better peering decisions, be ahead of the competition
- Publicly available interdomain traffic data is a scarce resource - sensitive business information
- Need to be able to create synthetic matrices for research purposes
We investigate characteristics of the ITM, describe it quantitatively from a perspective of a large research network.

We analyse:

- Sparsity
- Statistical distribution of the traffic
- Observe that the distribution can be related to congestions in a network
Knowledge of ITM useful in other research areas - economics, peering, routing

We propose a novel method to generate synthetic traffic matrices:
- Stems from first-principles (connection-based approach)
- Recognizes the fact that the traffic is a mixture of different applications
- Regional artifacts – different popularities of the content
1 Introduction

2 Macroscopic view: Interdomain Traffic Matrix
   - Introduction
   - Paper I: Characterizing ITM
   - Paper II: Synthesizing ITM

3 Microscopic view: Price Discrimination
   - Introduction
   - Paper III: Detecting price discrimination
   - Paper IV: Crowd assisted search of PD

4 Conclusions and Further Work

5 Contributions
Characterizing ITM

- GÉANT - most complete source of direct measurements of interdomain traffic available to the researchers
- We focus on spatial properties
- Sampled NetFlow data
Characterizing ITM

<table>
<thead>
<tr>
<th></th>
<th>trace W</th>
<th>trace M</th>
<th>trace Y</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>period</strong></td>
<td>1 week Nov 22–28, 2010</td>
<td>1 month Nov 1–30, 2010</td>
<td>52 weeks from Jan 4, 2010</td>
</tr>
<tr>
<td><strong>flows</strong></td>
<td>$3.91 \times 10^9$</td>
<td>$1.99 \times 10^{10}$</td>
<td>$2.17 \times 10^{11}$</td>
</tr>
<tr>
<td><strong>packets</strong></td>
<td>$3.61 \times 10^{12}$</td>
<td>$1.74 \times 10^{13}$</td>
<td>$1.70 \times 10^{14}$</td>
</tr>
<tr>
<td><strong>bytes</strong></td>
<td>$3.26 \times 10^{15}$</td>
<td>$1.55 \times 10^{16}$</td>
<td>$1.45 \times 10^{17}$</td>
</tr>
<tr>
<td><strong>NetFlow data volume</strong></td>
<td>111 GB</td>
<td>476 GB</td>
<td>5.75 TB</td>
</tr>
</tbody>
</table>

**Table:** Parameters of the GÉANT NetFlow traces.
We define sparsity as a ratio of number of zeros in the ITM to all observable items in the matrix.

Challenge - how to know if there is no traffic between ASes or the traffic is not routed through GÉANT?

Only lower bound of the sparsity can be estimated.

Observed sparsity > 45%
Characterizing ITM

Statistical distributions

- We find that 94% of the rows is heavy-tailed (top 15% entries in each row account for 95% of the traffic)
- Distributions resemble LogNormal or Pareto

![Graphs showing CCDF data for LogNormal and Pareto distributions with different D values](image.png)

(a) Pareto-like ($D = 0.88$)  
(b) LogNormal-like ($D = 0.27$)  
(c) In the middle ($D = 0.43$)

**Figure:** Instances of the generated traffic distribution. The tail of the distribution varies between the “straight” Pareto-like to the “bent” LogNormal-like.
Characterizing ITM
Shape and throughput

**Figure:** Type of the distribution tail and average throughput. Each dot is a separate AS. The dot size indicates the number of visible non-zero prefixes.
Characterizing ITM

Congestion

- Shape of the distribution can be caused by congestions in the observed network
- Traffic bottlenecks can cause “tail truncation” effect
- Hypothesis: in a congested network every new connection would compete for bandwidth, so there should be negative correlation between # of flows and throughput

---

Characterizing ITM

Congestion

Figure: Number of flows and the median throughput for a LogNormal-like (a) and Pareto-like (b) AS. 22–24 Nov 2010, 10:00–20:00. A few extreme outliers in (b) are not drawn.

- Correlation $[-0.77, -0.85]$ for LogNormal (presumably congested) and insignificant for Pareto (presumably not congested)
Characterizing ITM

Correlations between rows

- Measured for 15k+ pairs of rows
- Calculated Spearman correlations
- 99% of the correlations positive, up to 0.85, 0.28 average
Characterizing ITM

Low effective rank

- Matrix with low effective rank can be approximated by a linear combination of small number of rows and columns
- Analysis of the eigenvalues confirms low rankness

Figure: Eigenvalues of the submatrices (relative magnitudes). Only a small number of the values is significant, what indicates a low effective rank.
1 Introduction

2 Macroscopic view: Interdomain Traffic Matrix
   - Introduction
   - Paper I: Characterizing ITM
   - Paper II: Synthesizing ITM

3 Microscopic view: Price Discrimination
   - Introduction
   - Paper III: Detecting price discrimination
   - Paper IV: Crowd assisted search of PD

4 Conclusions and Further Work

5 Contributions
Synthesizing ITM

- Research on internet economy, e.g. peering decisions, requires models of ITM
- Scarce data - need to build synthetic matrices that preserve properties of real ITMs

- Our algorithm (ITMgen) is:
  - Connection-level
  - Takes into account different applications and its relative popularity
  - Different content types (application types, forward / reverse traffic ratio)
  - Regional effects (content is popular in one region and not in other region)
  - Alternative to the existing gravity model

---

Synthesizing ITM

Traffic model

Traffic from AS \( i \) to AS \( j \) can be expressed as

\[
T_{i,j} = \sum_{\kappa} m_\kappa \left( S_i p_{i\kappa}^{\kappa}(j) + d_\kappa S_j p_{j\kappa}^{\kappa}(i) \right)
\]  

(1)

- Two terms in the summation represent the traffic generated from a user due to application \( \kappa \), and the traffic produced by that application in the reverse direction
- \( \kappa \) – application
- \( p_{i\kappa}^{\kappa}(j) \) – relative popularity content related to application \( \kappa \), subjective to \( i \)
- (a)symmetry in the two directions of traffic due to application \( \kappa \) is denoted by \( d_\kappa \), and this parameter is application-dependent.
- \( m_\kappa \) – contribution of each application to the overall traffic mix.
- \( S_i \) – relative size of an AS (number of users in AS \( i \))
Synthesizing ITM
Parametrizing the model

- Parametrizing is definitely challenging...
- We focus only on WEB and P2P content

- Alexa.com – popularity of content per country, top 1 mln pages
- openbittorrent.com for P2P statistics
- Open marketing reports to model sizes of ASes (number of users), combined with P2P data and whois information
- Packet level traces from CESCA to model application-level characteristics
Synthesizing ITM
Parametrizing popularity of Web content

Figure: WEB popularity distribution of ASes, globally and for three example regions.

- “Popularity” of the content hosted by different ASes in different countries is similar across regions.
- Top 1mln pages from Alexa
- Does not reflect real traffic between ASes, but can serve as a basis of comparison
Synthesizing ITM
Parametrizing popularity of P2P content

Figure: P2P activity distribution.

- Data gathered from BitTorrent crawls
Synthesizing ITM
Parametrizing application mix

- Application level characteristics obtained from direct monitoring of CESCA link for 14 days.
- WEB traffic ratio $\log_{10}(d_K)$: (0.4, 1.5)
- P2P traffic ratio: (−0.87, 1.25).
- $m_{P2P} = 0.65$, $m_{WEB} = 0.35$. 
Synthesizing ITM
Validating – traffic distribution

Figure: Statistical distribution of the traffic produced and consumed by the observed ASes, for the Telefonica data (dashed line) and the model (solid) for the synthetic ITMs of different sizes.
Synthesizing ITM
Validating – regional effects

(a) Traffic exchanged with ASes within same region; matrices of 4 different sizes are shown.

(b) Regional traffic of content providers.

Figure: Regional traffic exchange.

- Regional exchange of the traffic is achieved by grouping ASes in 10 “regions” with equal number of ASes and properly parametrizing $p_i^\kappa(j)$.
- Gravity model consistently underestimates regional traffic.
Synthesizing ITM
Use case - cloud storage

- The model can be used to evaluate what-if scenarios
- Arbitrary, common sense parameters (for sake of the exercise)

- Users generate an additional 5% of upstream traffic
- Traffic is skewed, traffic ratio $\log_{10}(d_{ST})$ with a normal distribution $N(0.7, 0.2)$
- Simulation suggests that ASes providing cloud storage will increase traffic from 16% to 20%
- Overall traffic generated by all ASes will increase by 9.1%

- ITMgen can be used in evaluating what-if scenarios – important in economical research
1 Introduction

2 Macroscopic view: Interdomain Traffic Matrix
   - Introduction
   - Paper I: Characterizing ITM
   - Paper II: Synthesizing ITM

3 Microscopic view: Price Discrimination
   - Introduction
   - Paper III: Detecting price discrimination
   - Paper IV: Crowd assisted search of PD

4 Conclusions and Further Work

5 Contributions
Detecting price discrimination

- Price discrimination - setting the price of a given product for each customer individually according to the customer’s valuation of the product. The same product is offered to the customers with different prices.

- Economic phenomenon at the very “bottom” of the Internet economy (micro scale)

- Research at the intersection of e-retail and user personal information

- Empirically demonstrate existence of PD

- Analyse information vectors facilitating PD
Detecting price discrimination
Economic model of Internet services

Economic model behind the Internet services:
- Offer service for "free"
- Attract users
- Collect information
- Monetize information

What happens with all that information?
Popular answer - it is used for targeted advertising

Investigate alternative hypothesis - the information is used for *price discrimination* (PD)

E-commerce market size: $961 billion

*Does price discrimination, facilitated by personal information, exist in the Internet?*
Develop methodology to investigate PD in the Internet
Find empirical evidences that PD exists in the Internet
Present examples of PD on multiple e-commerce websites
Uncover information vectors facilitating PD

We show that crowd sourcing is a feasible method to investigate PD at large scale
We build, deploy and evaluate a working (!) system
1 Introduction

2 Macroscopic view: Interdomain Traffic Matrix
   • Introduction
   • Paper I: Characterizing ITM
   • Paper II: Synthesizing ITM

3 Microscopic view: Price Discrimination
   • Introduction
   • Paper III: Detecting price discrimination
   • Paper IV: Crowd assisted search of PD

4 Conclusions and Further Work

5 Contributions
Detected price discrimination

Information vectors

Examined information vectors:

- System and browser differences
- Geographical location of the originating query
- Certain traits of the user (e.g. affluent vs budget conscious)
Detecting price discrimination
Findings

- No evidence for system/browser based PD
- We find location-based PD
- We find PD based on user traits (i.e. originating URL revealing that user is budget-sensitive)
- We find search discrimination\textsuperscript{4} based on user traits (personas)

\textsuperscript{4}E.g. returning more expensive products to buyers with a higher willingness to pay. It operates on multiple products trying to \textit{steer} buyers towards an appropriate price range
Detecting price discrimination

Setup

- Machines with different browsers and systems (no wget)
- Proxy servers in different geographical locations (PlanetLab)
- Trained “personas” (user profiles with particular traits)
- 35 product categories, 200 distinct vendors
- Over 600 concrete products

Manually checked that the differences cannot be explained by differences in taxation, shipping costs or custom duties
Detecting price discrimination
Different locations – different countries

Figure: Price differences at Amazon based on the customer’s geographic location using the prices in New York, USA as reference. For each of the considered products there exist at least two locations with different prices.

- Differences of prices for Kindle e-books at Amazon
- Top 100 e-books (21% difference in majority of the cases, up to 166%)
- Steam (store.steampowered.com, e-entertainment online retailer) – 300 products, differences for 20% of products between Germany and Spain, 3.5% products differed between US, Brazil and Korea
Detecting price discrimination
Different locations – same US state

**Figure**: Price differences at staples.com. The dot sizes mark the mean price surplus for the locations, from 0% (small dots) up to 3.9% (large dots)

- staples.com, single state, 29 products, 200 ZIP codes
- Outskirts have larger prices than big cities!
Detecting price discrimination

Personal information – trained profiles

(a) Prices (mean/min/max) shown by Google to the different personas. The median number of products in each category per persona is 12.

(b) Mean prices (with std. deviations) of top-10 results from Cheaptickets.com returned to affluent and budget personas. The mean difference is 15%, and can be even as high as 50%.

- Difference in search results – search discrimination, steering users towards different prices according to their preferences
- No price discrimination observed
Detecting price discrimination
Personal information – URL of origin

Figure: Price difference at the Shoplet.com online retailer site, with- and without redirection from a price aggregator.

- URL of origin can indicate user’s sensitiveness to prices (e.g. customer comes from a discount site).
- nextag.com – price aggregator, shoplet.com – retailer
- Mean difference between prices with- and without redirection: 23%
1 Introduction

2 Macroscopic view: Interdomain Traffic Matrix
   - Introduction
   - Paper I: Characterizing ITM
   - Paper II: Synthesizing ITM

3 Microscopic view: Price Discrimination
   - Introduction
   - Paper III: Detecting price discrimination
   - Paper IV: Crowd assisted search of PD

4 Conclusions and Further Work

5 Contributions
Crowd assisted search of PD
User-centric approach

How can a real user know if she is subject to PD?

- Manually sampling selected sites is not enough
- Crowdsourcing:
  - Allows end-users to point the examples of sites engaging in PD
  - Allows extracting prices without requiring our manual intervention
  - Broadens scope of the measurement
Crowd assisted search of PD
The Price Sheriff

- A browser plug-in installed by a user
- User highlights a price on an e-retailer site
- Confirm check for the price (click a pop-up)
- Back-end service will access the examined page from 14 different vantage points
- Price is extracted from the contacted pages, compared and presented to the user
Crowd assisted search of PD
The Price $heriff – Results

- 340 users from 18 different countries
- Products from 600 domains

- Afterwards we crawled selected 21 domains with up to 100 products per retailer, on a daily basis for a week
Crowd assisted search of PD
Results – Crowd-collected dataset

Figure: Magnitude of price differences per domains

- Prices vary between 15% and 40%, and even up to 200% in extreme cases
- Some retailers not very popular - crowd sourcing is useful
Crowd assisted search of PD
Results – Crawled dataset

Figure: Magnitude of price variability per domain

- Selected retailers found due to crowd sourcing, examined with a systematic crawl
- Price variation was a repeatable phenomenon – observed in up to 100% requests
- Prices vary between 10% and 30% for most of the retailers
Crowd assisted search of PD
Price variations for all products

Figure: Maximal ratio of price differences per product price (all stores)

- Characterizing variations from the perspective of products
- Is there any correlation between price of the product and the price variation?
- PD in entire range of products, from $10 to $10k
Crowd assisted search of PD
Investigating specific retailers

Figure: Ratio of price differences per product price

- Examples of differences of prices with multiplicative term or additive term

(a) www.digitalrev.com  
(b) www.energie.it
Crowd assisted search of PD
Differences per location – US

Figure: Magnitude of price difference per location – www.homedepot.com

- Pair-wise comparisons of how the prices differ between two locations for a specific retailer. Prices relative to a minimum price of the product across all the vantage points.
- E.g. New York is consistently more expensive than Chicago, but there are also mixed pairs (Boston and Lincoln)
Crowd assisted search of PD
Differences per location – per country

Figure: Magnitude of price difference per location

- Per-country differences
- Not consistent for different retailers (e.g. US and Germany)
Crowd assisted search of PD
Logged in users

Figure: The impact of login on the price of Kindle ebooks at www.amazon.com

- Users logged in and not logged in to Amazon
1 Introduction

2 Macroscopic view: Interdomain Traffic Matrix
   - Introduction
   - Paper I: Characterizing ITM
   - Paper II: Synthesizing ITM

3 Microscopic view: Price Discrimination
   - Introduction
   - Paper III: Detecting price discrimination
   - Paper IV: Crowd assisted search of PD

4 Conclusions and Further Work

5 Contributions
Conclusions and Further Work
Macroscopic view – Interdomain Traffic Matrix

- Scarcity of the inter-AS data makes the ITM research difficult but not impossible – value in measuring qualitative properties of ITM
- In Paper I - we analysed GÉANT dataset. Research should be extended with other sources of data, preferably commercial system
- In I and II focused on spatial properties of ITM – research on temporal properties could shed light on long term evolution of ITM
- There might exist correlation between inter-AS traffic and congestion (I). Especially valuable for network operators
- Openly available data combined with direct measurement (II) can be used to generate snapshots of ITM
- Synthesis can be improved by better parametrization of the model
- Model generates static snapshots, temporal aspect should be included as well
- Our model is topology agnostic – explore how to generate synthetic topologies that match the synthetic matrices
Conclusions and Further Work
Microscopic view – Price Discrimination

- In III and IV we show empirically that PD exists in the Internet
- In III we analysed different information vectors leading to PD and show that prices can change according to user’s location or personal traits
- In IV we argue that crowd-sourcing is a feasible method to conduct research on PD
- We present an Internet user-oriented online tool
- Scaling the experiment (non-trivial engineering effort)
- Use crowdsourcing to not only gather, but also assess gathered data
- Uncover economical and technical mechanisms behind PD
Introduction

Macroscopic view: Interdomain Traffic Matrix
- Introduction
- Paper I: Characterizing ITM
- Paper II: Synthesizing ITM

Microscopic view: Price Discrimination
- Introduction
- Paper III: Detecting price discrimination
- Paper IV: Crowd assisted search of PD

Conclusions and Further Work

Contributions
Introduction

Contributions

Macroeconomic aspects - analysing and synthesizing ITM


Microeconomic aspects - price discrimination


Contributions

Other activities

- Our work on price discrimination on the internet, presented in Paper III, was mentioned in several The Wall Street Journal articles:
  - “How the Journal Tested Prices and Deals Online, WSJ, 2012, Dec
  - “Websites Vary Prices, Deals Based on Users’ Information”, WSJ, 2012, Dec
  - “Want a Deal Online? Pose as a Bargain Shopper”, WSJ, 2013, Jan

- Research described in Papers III and IV was presented at LAP/CPC/ICPEN conference (Antwerp, 16-17 April 2013) as invited talk.

London Action Plan (LAP) is a network of anti-spam government authorities and leading technologists. International Consumer Protection Enforcement Network (ICPEN) and EU Consumer Protection Cooperation Network (CPC) are focused on broad enforcement and policy consumer protection initiatives.

Contributions
Processing bulk of Internet traffic data

- Initial work on processing bulk of Internet traffic data:
Macro- and Microscopic Analysis of the Internet Economy from Network Measurements

Jakub Mikians

UPC BarcelonaTech

March 3, 2015
Characterizing ITM
Distribution of parameters

Figure: Distribution parameters as a function of throughput.
Shape of the tail of a distribution

To compare the shape of the previous distribution, we define a metric $D$ that indicates if the tail is LogNormal-like or Pareto-like. Let $F$ be an empirical CDF of the sample, and let $F_P$ and $F_L$ be the CDFs of the Pareto and LogNormal distributions that fit the tail of the sample. We measure the difference in the tail using the Kolmogorov-Smirnov metric:

$$KS(F_1, F_2) = \max|F_1(x) - F_2(x)| \text{ only for values of } x \text{ that are in the tail.}$$

We define $D$ as

$$D = \frac{KS(F, F_L)}{KS(F, F_L) + KS(F, F_P)} \quad (2)$$
Detecting price discrimination

Third party resources

Figure: Presence of third party resources on the sites used for training personas.
Crowd assisted search of PD
Results – Crowd-collected dataset

(a) Domains with the highest number of request where price differences occurred

(b) Magnitude of price differences per domains

- Prices vary between 15% and 40%, and even up to 200% in extreme cases
- Some retailers not very popular - crowd sourcing is useful
Crowd assisted search of PD

Results – Crawled dataset

(c) Measure extent of price variations for different domains

(d) Magnitude of price variability per domain

- Selected retailers found due to crowd sourcing, examined with a systematic crawl
- Price variation was a repeatable phenomenon – observed in up to 100% requests
- Prices vary between 10% and 30% for most of the retailers
Crowd assisted search of PD

Differences per locations

![Figure: Magnitude of price differences per location (all)](image)

- Do users from certain locations tend to pay more for the same product than others?
Crowd assisted search of PD
Differences per location – Finland

Figure: Magnitude of price differences per domains in Tampere, Finland
Introduction

Processing bulk traffic data

- Need to process bulk backbone data
- Excellence in using network traffic gathering and analysing tools
- As a result of the initial exercises we developed a novel portscan detection algorithm.  

\[\text{Paper V: “A Practical Approach to Portscan Detection in Very High-Speed Links”}\]
Early discard handshake packets that are not needed to detect portscans

A couple of Bloom filters effectively discards up to 85% of the packets

The method requires less than 1MB of memory to accurately monitor 10Gb/s link
Portscan Backup Slides

- DRAM - too slow
- Sampling - large impact on portscan detection
Bloom filters:

- Ignore legitimate handshakes (whitelist)
- Ignore failed connections that does not correspond to scans (e.g. TCP retransmissions)
- Drops 85% of the packets

- Finding scanners – known problem of finding top-k elements a data stream
Figure: Algorithm description.
Portscan Backup Slides

- bf_whitelist – track legitimate connections
- bf_syn – tracks repeated syn packets
- After that the connections are counted in an effective top-k counting data structure
Table: Statistics of the traces. trace C only accounts for Syn/SynAck packets.

<table>
<thead>
<tr>
<th></th>
<th>trace A 30min @ 1GigE</th>
<th>trace B 2h @ OC-3</th>
<th>trace C 30min @ 10GigE</th>
<th>trace A0 30min @ 1GigE</th>
</tr>
</thead>
<tbody>
<tr>
<td>date</td>
<td>2010-05-18</td>
<td>2010-04-16</td>
<td>2010-07-29</td>
<td>2010-05-18</td>
</tr>
<tr>
<td>TCP packets</td>
<td>228,848,927</td>
<td>144,885,865</td>
<td>13,978,845</td>
<td>97,380,742</td>
</tr>
<tr>
<td>TCP sources</td>
<td>188,136</td>
<td>263,055</td>
<td>467,264</td>
<td>89,086</td>
</tr>
<tr>
<td>TCP flows</td>
<td>2,892,334</td>
<td>5,199,928</td>
<td>11,526,323</td>
<td>1,133,392</td>
</tr>
<tr>
<td>average usage</td>
<td>879.1 Mb/s</td>
<td>185 Mb/s</td>
<td>3.5 Gb/s</td>
<td>n/a</td>
</tr>
</tbody>
</table>
Portscan Backup Slides

(a) trace A - 1 Gb/s UPC link  (b) trace B - MAWI traffic

(c) trace C - 10 Gb/s CESCA link  (d) online - 10GigE link