Network traffic characterization

A historical perspective
## Incoming AT&T traffic by port
(18 hours of traffic to AT&T dial clients on July 22, 1997)

<table>
<thead>
<tr>
<th>Name</th>
<th>port</th>
<th>%bytes</th>
<th>%packets</th>
<th>bytes per packet</th>
</tr>
</thead>
<tbody>
<tr>
<td>world-wide-web</td>
<td>80</td>
<td>56.75</td>
<td>44.79</td>
<td>819</td>
</tr>
<tr>
<td>netnews</td>
<td>119</td>
<td>24.65</td>
<td>12.90</td>
<td>1235</td>
</tr>
<tr>
<td>pop-3 mail</td>
<td>110</td>
<td>1.88</td>
<td>3.17</td>
<td>384</td>
</tr>
<tr>
<td>cuseeme</td>
<td>7648</td>
<td>0.95</td>
<td>1.85</td>
<td>333</td>
</tr>
<tr>
<td>secure web</td>
<td>443</td>
<td>0.74</td>
<td>0.79</td>
<td>603</td>
</tr>
<tr>
<td>internet chat</td>
<td>6667</td>
<td>0.27</td>
<td>0.74</td>
<td>239</td>
</tr>
<tr>
<td>file transfer</td>
<td>20</td>
<td>0.65</td>
<td>0.64</td>
<td>659</td>
</tr>
<tr>
<td>domain name</td>
<td>53</td>
<td>0.19</td>
<td>0.58</td>
<td>210</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

World Wide Web traffic dominates traffic mix
# MWN traffic by port

(24 hours of traffic to/from MWN clients in 2006)

<table>
<thead>
<tr>
<th>Port</th>
<th>% Conns</th>
<th>% Success</th>
<th>% Payload</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web</td>
<td>80</td>
<td>70.82%</td>
<td>68.13%</td>
</tr>
<tr>
<td></td>
<td>445</td>
<td>3.53%</td>
<td>0.01%</td>
</tr>
<tr>
<td>Web</td>
<td>443</td>
<td>2.34%</td>
<td>2.08%</td>
</tr>
<tr>
<td>SSH</td>
<td>22</td>
<td>2.12%</td>
<td>1.75%</td>
</tr>
<tr>
<td>Mail</td>
<td>25</td>
<td>1.85%</td>
<td>1.05%</td>
</tr>
<tr>
<td></td>
<td>1042</td>
<td>1.66%</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>1433</td>
<td>1.06%</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>135</td>
<td>1.04%</td>
<td>0.00%</td>
</tr>
<tr>
<td>&lt; 1024</td>
<td>83.68%</td>
<td>73.73%</td>
<td>79.05%</td>
</tr>
<tr>
<td>&gt; 1024</td>
<td>16.32%</td>
<td>4.08%</td>
<td>20.95%</td>
</tr>
</tbody>
</table>
**Grouping IP Packets Into Flows**

- **Group packets with the “same” address**
  - **Application-level:** single transfer web server to client
  - **Host-level:** multiple transfers from server to client
  - **Subnet-level:** multiple transfers to a group of clients

- **Group packets that are “close” in time**
  - 60-second spacing between consecutive packets
**Incoming WorldNet traffic by port**  
(18 hours of traffic to WorldNet dial clients on July 22, 1997)

<table>
<thead>
<tr>
<th>Name</th>
<th>port</th>
<th>%bytes</th>
<th>%pkts</th>
<th>%flows per flow</th>
<th>pkts per packet</th>
<th>duration (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>world-wide-web</td>
<td>80</td>
<td>56.75</td>
<td>44.79</td>
<td>74.58</td>
<td>12</td>
<td>819</td>
</tr>
<tr>
<td>netnews</td>
<td>119</td>
<td>24.65</td>
<td>12.90</td>
<td>1.20</td>
<td>210</td>
<td>1235</td>
</tr>
<tr>
<td>pop-3 mail</td>
<td>110</td>
<td>1.88</td>
<td>3.17</td>
<td>2.80</td>
<td>22</td>
<td>384</td>
</tr>
<tr>
<td>cuseeme</td>
<td>7648</td>
<td>0.95</td>
<td>1.85</td>
<td>0.03</td>
<td>1375</td>
<td>333</td>
</tr>
<tr>
<td>secure web</td>
<td>443</td>
<td>0.74</td>
<td>0.79</td>
<td>0.99</td>
<td>16</td>
<td>603</td>
</tr>
<tr>
<td>internet chat</td>
<td>6667</td>
<td>0.27</td>
<td>0.74</td>
<td>0.16</td>
<td>89</td>
<td>239</td>
</tr>
<tr>
<td>file transfer</td>
<td>20</td>
<td>0.65</td>
<td>0.64</td>
<td>0.26</td>
<td>47</td>
<td>659</td>
</tr>
<tr>
<td>domain name</td>
<td>53</td>
<td>0.19</td>
<td>0.58</td>
<td>10.69</td>
<td>1</td>
<td>210</td>
</tr>
<tr>
<td>. . .</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Incoming application flows with a 60-second timeout
- Diverse flow characteristics across different protocols
Short-vs. long-lived Web flows

Many very short flows (30% are less than 300 bytes)
Many medium-sized flows (short web transfers)
Most bytes belong to long flows (large images, files)

**Flow densities are signatures**
Traffic measurements: Pre-1990

- Early Telephony: Importance of measurements (e.g., Erlang, Palm, Wilkinson, ...)
- Modern Telephony: Measurements are a scarce commodity; supposedly „well-understood“ characteristics
- Early data networking: Importance of measurements (e.g., ARPANET measurements by Kleinrock et al.)
- Modern data networking: No data or only a few small data sets are available
Traffic measurements: Pre-1990

- Traffic data analysis
  - Strictly traditional inference techniques
  - Focus on choosing best-fitting model
  - Obsession with „Squeezing a data set dry“

- Traffic and performance modeling
  - Black-box or operational models dominate
  - No real need to talk to subject-matter experts
  - Traffic is viewed as „just another time series…”
  - Main objective: „What can be analyzed?“
Post-1990: What has changed?

- **Traffic measurements**
  - Abundance of traffic measurements; reproducibility

- **Traffic data analysis**
  - Data exhibits unusual features
  - From statistical inference to scientific inference
  - Networks are complex; need for subject-matter expertise

- **Traffic and performance modeling**
  - Need for physical-based or structural models
  - Main objective: „What matters for performance?“
Traffic measurement challenges

- Telephone networks are static entities
  - Have hardly changed for years and decades (exception cellular phone systems...)
  - Have evolved in a predictable manner

- Modern data networks are highly dynamic entities
  - User population, services and applications
  - Traffic mix, protocols, ...
  - Data networks that don’t change are suspicious
  - Internet as an example of extreme heterogeneity
Traffic measurement challenges

- Measuring high-speed network traffic
  - **High-quality**: Special-purpose traffic recorders
  - **High-volume**: Terabyte storage devices
  - **Diversity**: many large datasets from
    - Different networks
    - Different times
    - Different points in the network
  - **Sensitivity**: Who can record and collect what data?

- High-speed network traffic is complex
  - Unusual behavior, constant surprises, ...
  - What are interesting/relevant measurements?
Sample data trace

Measured Data Traffic (Ethernet LAN)
Netdynamics – „Killer application“

- WWW and the Internet
  - 1993: ... Hardly any WWW traffic on the Internet
  - 1994: ... About 10% of total Internet traffic is WWW
  - 95/96: ... Up to 60-70% of overall Internet traffic is WWW
  - 06/07: ... Up to 60-70% of overall Internet traffic P2P

- New applications and services
  - Games? IPTV?

- New network protocols
Network dynamics: User population

- **Number of Internet hosts**
  - Early 1989: 80,000
  - Early 1992: 727,000
  - Oct. 1993: 2,056,000
  - Late 1996: 10,000,000
  - Now: \(100\times\text{xxxxxx}xxxx\)

- **Internet traffic volume (Merit; Inc.)**
  - March 1991: \(1.3\times10^{12}\) bytes/month
  - March 1994: \(1.1\times10^{13}\) bytes/month
High-volume measurements

- 1 hour of ETHERNET LAN traffic (10 Mbits)
  - About 1 million packets
- 1 day of uninterrupted ETHERNET LAN
  - About 2 Gigabytes of data
- 1 hour of ATM traffic (155 Mbits)
  - About 100 million packets
- 1 day of uninterrupted ATM measurements
  - About 1 Terabyte of data
- 1 day of uninterrupted 1 Gigabit measurements
  - About 10 Terabyte of data
High-quality measurements

- Timestamp accuracy
  - From millisecond to microsecond accuracy
- More than just another time series
  - Information about all layers in network hierarchy
    - TCP/IP header information
    - Payload
    - Higher level protocol information
- Active measurements
  - Actively injecting traffic into the network
- Passive measurements
  - Passively monitoring network information
Plane old telephony (POTS)

- Billing data
  - Signaling for each phone call
  - Billing on a call by call basis
  - Source, destination, start time, duration

- Studies
  - Call arrival process
  - Call holding time distributions
  - Spatial calling patterns

- Application
  - Network planning, Dimensioning, etc.
CCS/SS7 measurements

- Common Channel Signaling (CCS) Network
  - Slow but mature packet network: 56 Kbps
  - Running Signaling System 7 (SS7) protocol
  - Measurements at the level of individual SS7 messages
  - Variable length messages
  - Days/weeks worth of data
  - Hundreds of millions of messages

- Study of SS7 traffic at message-level

- Study of telephone traffic (POTS)
  - Call arrival process
  - Call holding time distributions
  - Spatial calling patterns
Data sources in IP networks

- Configuration data
  - Network
  - Service
  - Customer registration

- Usage data
  - Network data for each
    - Packet, flow, dial session
    - Routers MIB: utilization, loss statistics
    - Routing tables
    - Active probes

- Servers
  - Customer care
  - Email, Web hosting, E-commerce
Measurement design considerations

- Network operation has priority
  - Unless crucial for billing

- Network measurement as an afterthought
  - Design of new protocols
  - Design of network hardware
  - Design of networks

- Security
  - Who
  - Where
  - How
  - Impact on network
Measured Data Traffic (Ethernet LAN)

Time Unit = 100 Seconds

Time Unit = 10 Seconds
Measured Data Traffic (Ethernet LAN)

Packet/Time Unit

Time Unit = 100 Seconds

Time Unit = 10 Seconds

Time Unit = 1 Second
Time Series

Example

- # of packets (bytes) per 10 milliseconds
- # of TCP connections arriving per second
- # of modem sessions arriving per second

Definitions

- Time series: $X_1, X_2, \ldots, X_n$
- Aggregated process: $X^{(m)}$

$$X^{(m)}(k) = \frac{1}{m} (X_{(k-1)m+1} + \ldots + X_{km}), k \geq 1$$

- Stationary time series: distribution of $X$ independent of time