<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joel Sommers</td>
<td>University of Wisconsin-Madison</td>
</tr>
<tr>
<td>Paul Barford</td>
<td></td>
</tr>
<tr>
<td>Walter Willinger</td>
<td>AT&amp;T Labs-Research</td>
</tr>
</tbody>
</table>
Analysis of Internet Data is Difficult

• Exploratory analysis of data is an important precursor to hypothesis-driven investigation

• But ... Internet data sets are large, multidimensional, and complex

• **Visualizations** are typically used for initial qualitative analysis
Visualization tools

• General-purpose
  • gnuplot, grace
  • GGobi, R, MatLab

• Special purpose
  • Protocol behavior, e.g., tcptrace, nam
  • Network monitoring, e.g., ethereal, ntop
  • Statistical or scaling properties, e.g., LRD plots
SPLAT

- A general-purpose tool with support for basic Internet data types
  - 2D scatter or phase plots
  - Animations along time dimension
  - zoom, pan, rotation of plotting space
- Filtering and pruning
  - Distributions or lists of correlated data
  - Auxiliary data sets, synchronized with main plot data
Demo 1: TCP packet traffic

- Laboratory trace data: long-lived TCP sources, dumbbell topology

- Phase plot of spacings of consecutive packets of a flow as they enter (x) and exit (y) a queue

- Filter example: time series of delay through the bottleneck queue (sawtooth behavior)
Demo 1: Phase Plot Interpretation

- **Multiplexing can cause expansion**
- **Vertical lines show regular pacing of ingress packets** (upstream queue, upstream bandwidth constraint, etc.)
- **Horizontal lines show regular pacing of packets egress from router** (congestion and queueing, bandwidth, application, etc.)
- **Queueing can cause compression**

Points along diagonal mean spacing was not disturbed. Points near origin indicate back-to-back packets. Points higher on diagonal indicate packets separated by RTT.
Demo 2: flow-level data

- Laboratory trace data: web-like TCP sources producing self-similar traffic using Harpoon, dumbbell topology
- Plot flow size \((x)\) and flow duration \((y)\)
- Filter example 1: round-trip times
- Filter example 2: time series of delay through the bottleneck queue
Summary

• SPLAT shows promise for general-purpose exploratory analysis of Internet data
  • We’ve been using various incarnations of it for 3 years
• Filtering/pruning mechanisms are important for large, multidimensional data sets
• Code will be available soon
the end

wail.cs.wisc.edu
Demo 3: IP address (spatial) data

- Abilene network flow records from Houston, TX router
- Phase plot of source address ($x$) and destination address ($y$)
- Filter example: distribution of amount of data transferred between source/destination

![Phase plot diagram]

- a popular destination prefix
- intra-prefix traffic
- an active source address or prefix
Splat configuration file example

<splat_data>
  <plot_data filename="flows.txt.gz"
    name="Test flow size/duration"
    xcol="6" ycol="8" zcol="7"
    xtype="int"
    xlabel="transfer size" xunits="bytes"
    xrange="0:1000000"
    xprecision="0"
    ytype="float"
    ylabel="transfer duration" yunits="seconds"
    yrange="0:10.0"
    yprecision="3" />

  <filter ftype="list" dtype="int" count="incr"
    col="9" name="round-trip time filter" />

  <filter ftype="distribution" dtype="string" count="6"
    col="0:1" name="src/dst distribution filter" />

  <auxfilter filename="flows_qlen.txt.gz" zlabel="time (sec)"
    wlabel="delay (millisec)" zcol="0" wcol="1" name="queuing delay" />
</splat_data>
# Splat data format example

whitespace delimited text

<table>
<thead>
<tr>
<th>#</th>
<th>src</th>
<th>dst</th>
<th>srcport</th>
<th>dstport</th>
<th>proto</th>
<th>pkts</th>
<th>bytes</th>
<th>time</th>
<th>duration</th>
<th>rtt</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.52.0.190</td>
<td>10.52.0.242</td>
<td>10000</td>
<td>45764</td>
<td>6</td>
<td>7</td>
<td>5747</td>
<td>0.000</td>
<td>0.103591</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>10.52.0.162</td>
<td>10.52.0.250</td>
<td>10000</td>
<td>39019</td>
<td>6</td>
<td>12</td>
<td>13181</td>
<td>0.020</td>
<td>0.109305</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>10.52.0.186</td>
<td>10.52.0.250</td>
<td>10000</td>
<td>39020</td>
<td>6</td>
<td>23</td>
<td>29016</td>
<td>0.068</td>
<td>0.184029</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>10.52.0.162</td>
<td>10.52.0.250</td>
<td>10000</td>
<td>39021</td>
<td>6</td>
<td>26</td>
<td>32403</td>
<td>0.105</td>
<td>0.178167</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>10.52.0.150</td>
<td>10.52.0.242</td>
<td>10000</td>
<td>45765</td>
<td>6</td>
<td>10</td>
<td>10720</td>
<td>0.112</td>
<td>0.077207</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>10.52.0.162</td>
<td>10.52.0.142</td>
<td>10000</td>
<td>50229</td>
<td>6</td>
<td>24</td>
<td>29068</td>
<td>0.131</td>
<td>0.178153</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>10.52.0.158</td>
<td>10.52.0.254</td>
<td>10000</td>
<td>41913</td>
<td>6</td>
<td>25</td>
<td>32351</td>
<td>0.178</td>
<td>0.123730</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>10.52.0.230</td>
<td>10.52.0.250</td>
<td>10000</td>
<td>39022</td>
<td>6</td>
<td>45</td>
<td>62236</td>
<td>0.189</td>
<td>0.424942</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>10.52.0.234</td>
<td>10.52.0.250</td>
<td>10000</td>
<td>39023</td>
<td>6</td>
<td>7</td>
<td>5747</td>
<td>0.210</td>
<td>0.151866</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>10.52.0.162</td>
<td>10.52.0.250</td>
<td>10000</td>
<td>39024</td>
<td>6</td>
<td>64</td>
<td>90924</td>
<td>0.212</td>
<td>0.215752</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>10.52.0.230</td>
<td>10.52.0.242</td>
<td>10000</td>
<td>45766</td>
<td>6</td>
<td>9</td>
<td>7810</td>
<td>0.213</td>
<td>0.211501</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>10.52.0.134</td>
<td>10.52.0.242</td>
<td>10000</td>
<td>45768</td>
<td>6</td>
<td>10</td>
<td>10720</td>
<td>0.214</td>
<td>0.082178</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>10.52.0.230</td>
<td>10.52.0.254</td>
<td>10000</td>
<td>41913</td>
<td>6</td>
<td>7</td>
<td>5747</td>
<td>0.216</td>
<td>0.141256</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>10.52.0.162</td>
<td>10.52.0.130</td>
<td>10000</td>
<td>56819</td>
<td>6</td>
<td>9</td>
<td>7877</td>
<td>0.226</td>
<td>0.108570</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>10.52.0.214</td>
<td>10.52.0.210</td>
<td>10000</td>
<td>40742</td>
<td>6</td>
<td>10</td>
<td>10720</td>
<td>0.233</td>
<td>0.242824</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>10.52.0.234</td>
<td>10.52.0.210</td>
<td>10000</td>
<td>40741</td>
<td>6</td>
<td>25</td>
<td>32351</td>
<td>0.234</td>
<td>0.302991</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>10.52.0.186</td>
<td>10.52.0.210</td>
<td>10000</td>
<td>41643</td>
<td>6</td>
<td>7</td>
<td>5747</td>
<td>0.253</td>
<td>0.091417</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>10.52.0.234</td>
<td>10.52.0.242</td>
<td>10000</td>
<td>45767</td>
<td>6</td>
<td>12</td>
<td>13181</td>
<td>0.257</td>
<td>0.226785</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>10.52.0.162</td>
<td>10.52.0.142</td>
<td>10000</td>
<td>50230</td>
<td>6</td>
<td>10</td>
<td>9339</td>
<td>0.263</td>
<td>0.106570</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>10.52.0.198</td>
<td>10.52.0.242</td>
<td>10000</td>
<td>45769</td>
<td>6</td>
<td>10</td>
<td>10720</td>
<td>0.265</td>
<td>0.222600</td>
<td>55</td>
<td></td>
</tr>
</tbody>
</table>