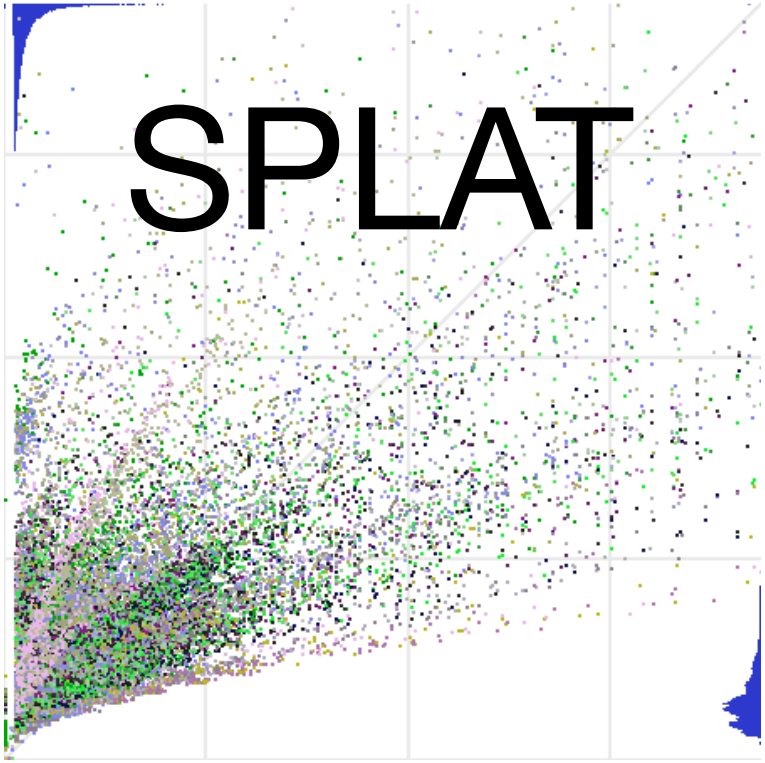


A Visualization Tool for Mining Internet Measurements

Joel Sommers — University of Wisconsin-Madison
Paul Barford

Walter Willinger — AT&T Labs-Research

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

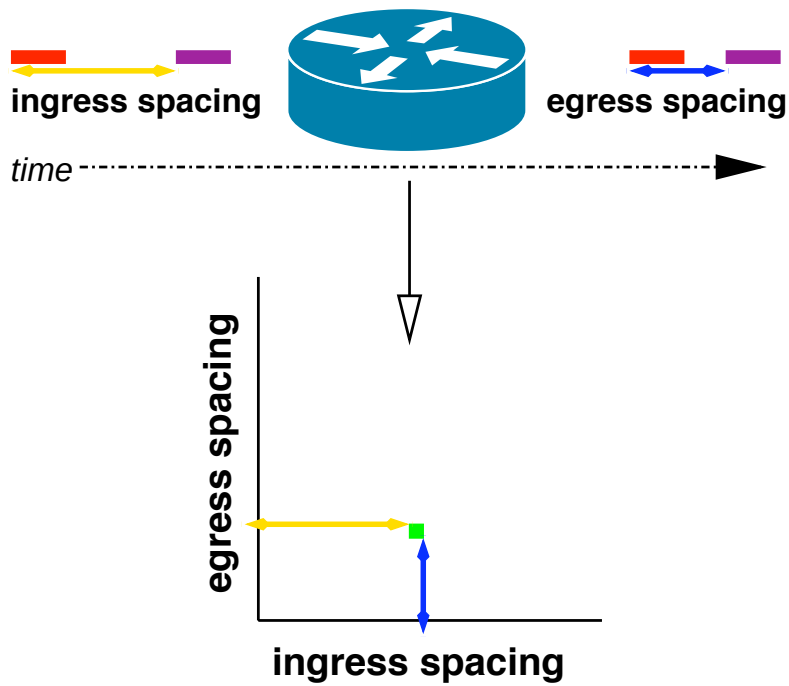
- General-purpose
  - gnuplot, grace
  - GGobi, R, MatLab
- Special purpose
  - Protocol behavior, e.g., tcptrace, nam
  - Network monitoring, e.g., ethereal, ntop
  - Statistical or scaling properties, e.g., LRD plots

SPLAT

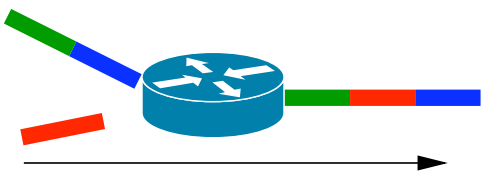
- A general-purpose tool with support for basic Internet data types
  - 2D scatter or phase plots
  - Animations along time dimension
  - zoom, pan, rotation of plotting space
- Filtering and pruning
  - Distributions or lists of correlated data
  - Auxiliary data sets, synchronized with main plot data

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)



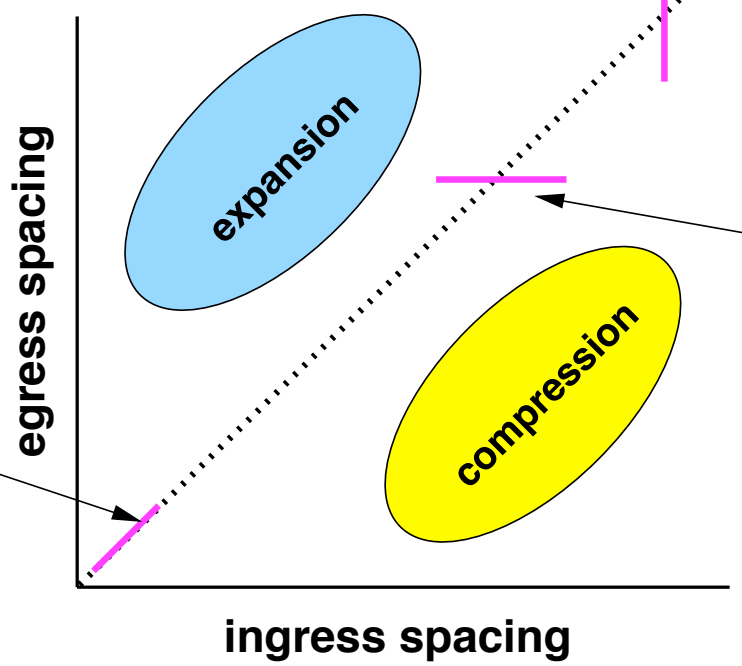
Demo 1: Phase Plot Interpretation



multiplexing can cause expansion

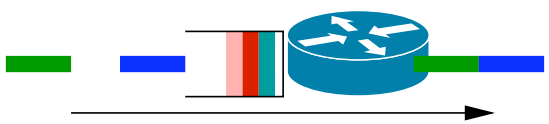
vertical lines show regular pacing of ingress packets (upstream queue, upstream bandwidth constraint, etc.)

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.



horizontal lines show regular pacing of packets egress from router (congestion and queueing, bandwidth, application, etc.)

queueing can cause compression



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

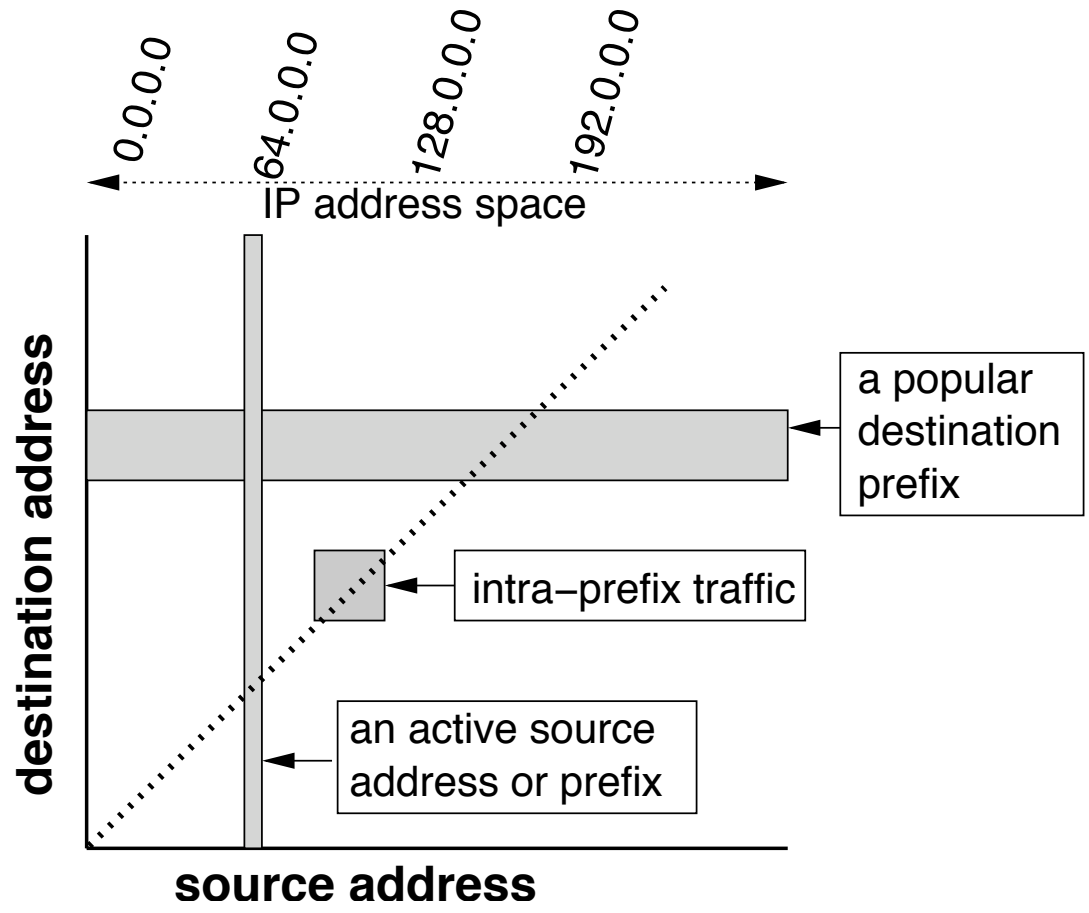
- SPLAT shows promise for general-purpose exploratory analysis of Internet data
  - We've been using various incarnations of it for 3 years
- Filtering/pruning mechanisms are important for large, multidimensional data sets
- Code will be available soon

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



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

```
#   src  dst  srcport  dstport  proto  pkts  bytes  time  duration  rtt
10.52.0.190 10.52.0.242 10000 45764 6 7 5747  0.000 0.103591 50
10.52.0.162 10.52.0.250 10000 39019 6 12 13181  0.020 0.109305 35
10.52.0.186 10.52.0.250 10000 39020 6 23 29016  0.068 0.184029 45
10.52.0.162 10.52.0.250 10000 39021 6 26 32403  0.105 0.178167 35
10.52.0.150 10.52.0.242 10000 45765 6 10 10720  0.112 0.077207 25
10.52.0.162 10.52.0.142 10000 50229 6 24 29068  0.131 0.178153 35
10.52.0.158 10.52.0.254 10000 41913 6 25 32351  0.178 0.123730 30
10.52.0.230 10.52.0.250 10000 39022 6 45 62236  0.189 0.424942 70
10.52.0.234 10.52.0.250 10000 39023 6 7 5747  0.210 0.151866 75
10.52.0.162 10.52.0.250 10000 39024 6 64 90924  0.212 0.215752 35
10.52.0.230 10.52.0.242 10000 45766 6 9 7810  0.213 0.211501 70
10.52.0.134 10.52.0.242 10000 45768 6 10 10720  0.214 0.082178 20
10.52.0.230 10.52.0.254 10000 41912 6 7 5747  0.216 0.141256 70
10.52.0.162 10.52.0.130 10000 56819 6 9 7877  0.226 0.108570 35
10.52.0.214 10.52.0.210 10000 40742 6 10 10720  0.233 0.242824 60
10.52.0.234 10.52.0.210 10000 40741 6 25 32351  0.234 0.302991 75
10.52.0.186 10.52.0.202 10000 41643 6 7 5747  0.253 0.091417 45
10.52.0.234 10.52.0.242 10000 45767 6 12 13181  0.257 0.226785 75
10.52.0.162 10.52.0.142 10000 50230 6 10 9339  0.263 0.106570 35
10.52.0.198 10.52.0.242 10000 45769 6 10 10720  0.265 0.222600 55
```