PCH Data-Collection and the Academic Research Community

Version 1.0

February, 2004

Bill Woodcock

Packet Clearing House

Data we collect

- **Exchange participation**
  - in-addr
  - Reachability
- **Netflow destination matrixes**
- **Responsible parties**
  - Contact information
  - Reachability
- **Routing table snapshots**
- **DNS integrity**

Limitations on what we can publish

- **Basically unlimited, but…**

- **Size**
  - Need either queries or physical shipment of disks

- **Trade Secrecy**
  - Unaggregated Netflow exports

- **Privacy**
  - Bulk responsible-party contact information

- **Cost of publishing**
  - Staffing to deal with requests from the academic community

What we'd like to be collecting

- **RTP quality data**
  - Currently discarded rather than sent with SIP tear-down

- **Broader peering agreement visibility**
  - Large providers aren't interested in letting this data out

- **Netflow exports from more providers**
  - Internal bottleneck on labor

- **Anycast/topology/geography correlation**
  - Internal bottleneck on labor

How researchers work with us

- **Introductions or anonymization**

  - Giving researchers access to the network and people who run it

- **Anycast-platform hosted experiments**

  - Cron-jobs send data back to experimenter's collector

- **Anycast-platform hosted machines**

  - Ship Sun B100s blades to each location

- **Data exports from existing datasets**

  - Queries or copies of datasets

Thanks, and Questions?

Copies of this presentation can be found
in Keynote, PDF, QuickTime and PowerPoint formats at:

http:// www.pch.net / resources / papers / data-collection

Bill Woodcock
Research Director
Packet Clearing House
woody@pch.net