# 1 Proposal Title

ANALYSIS & VISUALIZATION OF IP CONNECTIVITY

# 2 Primary Investigators

k claffy, Ph.D. UCSD Computer Science & Engineering, 1994
Bradley Huffaker, M.S. UCSD Computing Science & Engineering, 1998

# 3 Project Summary

We would like to build on the success of our last two years of research and analysis of Internet connectivity, which Cisco has found useful from both a research and operational perspective for the last two years.

For 2004-2005 our goal will be to derive three new connectivity information maps, which will involve analysis, and visualization components, as well as creating publically available software and databases that will support the community in a wide variety of operational analysis and research tasks. First, we would like to analyze and depict inter-AS connectivity based on organization (ASes under common ownership[1]) granularity rather than AS granularity. This task will use URB-funded (2003-4) software, in conjunction with new supporting CAIDA software that intelligently synthesizes registry information from several disparate sources. Second, we plan to derive a pop-level map of the Internet. This ambitious task will rely on tools we are building under URB auspices this year [1][2], as well as results from CAIDA's 2004 NSF/DHS-sponsored research in topology measurement. Finally, we will derive and update monthly an IPv6 topology map as the native IPv6 infrastructure grows throughout the year. This task will build on a formal collaboration between CAIDA and WIDE [2] and require resources on CAIDA's side to gather and distill a representative set of data. Each of these topology mapping tasks: (1) relies on cost-sharing with other existing funding sources for supporting software; (2) yields tangible deliverables to the community, in the form of both analysis software as well as three of the most highly relevant Internet topology maps that are possible to reliably estimate today.

---

[1] Or *stewardship* for those on the left. :)

[2] These tools analyze peering, transit, and customer relationships from routing table data.

# 4 Description of Research and Goals

Mapping of the Internet is complex, i.e., demanding creativity and insight, as well as hard, i.e., demanding persistence and endurance. Even the process of gathering a well-defined subset of topology data involves a measurement space littered with methodological landmines. Measurement tools carry inherent limitations [3], and analysis approaches require estimation and approximation that rely on – as often as they yield – insights into topology characteristics [4, 5, 6, 7, 8]. Further, while gathering a terabyte of topology data requires minimal conceptual effort, gathering the same amount of legitimate topology *information* is as challenging as it is important. Without concerted persistent effort in this area, our understanding of the structure and growth of Internet topology becomes fainter by the day.

In the meantime, a wide variety of topology and routing research persists, from building models of Internet topology growth based on economics and organizational science foundations [9], to inferring policy relationships from routing data using a variety of techniques [4, 7, 10, 11], to developing a rigorous mathematical framework for routing over Internet-like graphs [12]. But without exception, macroscopic Internet topology and interdomain routing research activities fall uncomfortably short of basic scientific standards for empirical validation.

To not only leverage Cisco's investment in this research but also facilitate our continued pursuit of unprecedented granularity and accuracy in Internet topology mapping, CAIDA proposes to build three of the most structurally interesting and timely maps, and more importantly the associated topology databases, of Internet infrastructure.

1. Depict inter-AS connectivity at an organizational (common AS administration) granularity as well as AS granularity, which will require new supporting CAIDA software that intelligently synthesizes registry information from several disparate sources. We already have research agreements with the four main address registries for bulk access to their registry data.

2. Develop a pop-level map of the Internet with as much policy structure as we can directly gather and indirectly infer. CAIDA will use, and extend where necessary, tools developed in last year's URB project [1] to gather the data for this task.

3. Build a hierchically structured topology map of the IPv6 Internet and correlate structure and growth patterns with that of IPv4 topology. This task relies

on: (1) an active WIDE/CAIDA collaboration on IPv6 macroscopic topology measurement [2]; (2) years of previous CAIDA work in IPv4 topology analysis [13, 14, 15, 16, 17, 18, 19].

Beyond their utility to the operational and research communities as accurate sources of core Internet topology data, these semantically rich maps also bear acute relevance to interests in public policy, national cybersecurity, and commercial technology market analysis. CAIDA's depth of experience with collection and analysis of topology and routing data makes it the ideal candidate to construct these maps, and Cisco's towering topological presence renders it the ideal candidate to support construction of these maps. The methodology and tools to capture, analyze, and depict the resulting data will be at least as important as the data itself, and we intend to prioritize maintainability of the software for future use in topology research. Thus this project involves three levels of contributions: the visual maps, and the insights they reveal to non-experts; the associated topology knowledge bases; and the supporting measurement, analysis, and presentation software.

Note that this work naturally extends CAIDA's URB-funded work for the last two years. In 2002-3 we developed an analysis methodology for ranking the richness of connectivity of ASes [20, 21] based on massive topology data as measured by CAIDA's macroscopic IP topology monitoring project [22] and observed by RouteViews BGP table snapshots [23, 24]. The coverage of these topology probes was unprecedented, dramatically higher than any previous work in this area, and the data yielded significant insight into the relative richness of IP connectivity of different ASes. This work was enthusiastically received both internally at Cisco[3] and by the wider community, as there continues to be wild marketing-inspired speculation and little sound methodology regarding macroscopic Internet connectivity analysis. URB funding allows (and would continue to allow) CAIDA to support daily update of the interactive web page for AS topological ranking. [4], [5]

We list development milestones below in section 5.

---

[3] Both the PSIRT and CIAG [25] groups at Cisco have expressed strong interest in continuation of this work.

[4] `http://as-rank.caida.org/cgi-bin/main.pl`

[5] Claffy presented this work at a Cisco 'Nerd Lunch' seminar at Cisco on 29 April 2003.

## 4.1   Impact

Cisco specifically requested the AS ranking analysis from CAIDA last year, and has made use of it each year since. From Wendy Garvin of Cisco's PSIRT team:

> Cisco's PSIRT team depends on the accurate and regularly updated data provided from the as-rank tool in order to properly identify the 'core' providers in the Internet in order to identify critical infrastructure in a non-prejudiced fashion. 'Skitter Core' has now become an industry wide tag for impartial ranking of transit traffic carriers. This is valuable research, and an extension of it into the IPv6 arena would provide us much needed data on a previously poorly explored technology.

In addition to specifically requesting above the third map we propose, she also emphasizes Cisco's need for the first map we propose.

> Many AS's are grouped together by one provider - the AS-RANK tool has no way of collating that data right now, which leads to some inaccuracies. The data is valid, but the assumption that it shows the picture in a way that people think about the problem is inaccurate. People think 'AT&T carries that traffic.' AT&T maintains at least 3 very large AS's. To group them together should show a proportionately higher amount of transit traffic for AT&T, but right now there's no way to correlate that data.

> This is a rough problem, as the tool is dependent on whois data right now, and would require human input to deal with this problem. The only suggestion I have is to have an interface which would allow a user to input 'groupings' of AS's of some form.

Note that we have a more automated, scalable strategy for dealing with aggregation at organizational granularity; the public database we create for this task will have utility far beyond its use for the proposed organizational topology map.

# 5 Timeframes for Funding and Research Completion

Funding ($100,000) to begin 1 July 2004. (or whenever possible)
Research Milestones

- **1 July 04** Begin implementation of organization database that incorporates several whois databases. (We already have agreements with the largest three registries to access their data in bulk.)

- **1 October 04** Initial pass at organizational map of the Internet, with interactive web access to database

- **1 November 04** First scamper-derived topology database and map of the IPv6 Internet

- **1 January 05** Automate process for generation of IPv6 and organizational maps

- **30 Jan 05** Initial draft of pop-level topology database and map of the Internet

- **30 Mar 05** Depiction of differences between IPv6 and IPv4 topological structure and growth patterns

- **30 June 05** Deliver final versions of all three maps to the community via web site. Present to relevant venues as appropriate.

## 5.1 Required/expected Research Cooperation with Cisco

Researchers are available to meet with Cisco staff to discuss methodologies and the implications of analysis results. In particular we already enjoy active levels of communication with Cisco engineers (Barry Greene, Wendy Garvin, Fred Baker) and funded researchers (Rob Thomas, Cymru), and expect to continue to benefit from their expertise. Barry Greene has offered to coordinate review of the code for BGP and other features that may enhance the research.[6]

---

[6]For example, Barry notes that there are new BGP peering tools that make it easier to add "BGP probes" to a network. Cisco also supports protocols such as DRP and SAA which can provide information for remote route table looksups and RTTs without needing a box physically installed.

# 6   SUPPORT REQUIREMENTS

## 6.1   Total Budget

$100,000

## 6.2   Duration

1 July 2004 - 30 June 2005 (or whatever Cisco can do)

## 6.3   Breakdown

SALARIES & BENEFITS:

| k claffy, Ph.D., P.I. | 10% effort | $15,500 |
| Brad Huffaker, M.S. | 25% effort | $20,760 |
| 1 grad student | 49% (9mo); 100% (3mo) | $46,677 |
| 1 grad student | (summer) 100% (3mo) | $12,193 |

DOMESTIC TRAVEL:
2 CAIDA researchers to present at Cisco as well as mutually agreed venue), 2 @ $1218 == $4,870

GRADUATE TUITION:
$16,560

# 7   GRADUATE STUDENTS INVOLVED

Two graduate students will be selected by October 2004.

# 8   OTHER CURRENT OR ANTICIPATED MATCHING FUNDS

Research on this project will draw cost-sharing support from NCS (now under DHS), NSF, and WIDE.

## 9    Short biographies of the researchers

**kc claffy, Ph.D.** is principal investigator for the distributed Cooperative Association for Internet Data Analysis (CAIDA), and resident research scientist based at the University of California's San Diego Supercomputer Center.  kc's research interests include Internet workload/performance data collection, analysis and visualization, particularly with respect to commercial ISP collaboration/cooperation and sharing of analysis resources. kc received her Ph.D. in Computer Science from UCSD in 1994.
**Brad Huffaker, M.S.** serves as technical manager for several tool development and traffic analysis efforts at CAIDA. He specializes on efforts to develop analytical and visualization techniques suitable for insight on the configuration, evolution and occurrence of network events in large network topologies. Brad received both the B.S. and M.S. degrees in Computer Science from UCSD.

## 10    Name of Cisco Champion

Fred Baker, Barry Greene, Wendy Garvin

## 11    Relevant University Administrative Contact

(Check may be sent here. Only wording required is 'unrestricted gift' per the URB web page.)

```
Darlene Piche
CAIDA SDSC 0505
9500 Gilman Drive
La Jolla, CA 92093-0505
858.534.5109
piche@sdsc.edu
```

# References

[1] kc claffy and B. Huffaker, "Cisco URB 2003 proposal: Connectivity Ranking of Autonomous Systems," April 2002. `http://www.caida.org/projects/cisco03asrank/`.

[2] B. Huffaker and M. Luckie, "Scamper macroscopic Internet topology measurement tool," 2004. `http://www.caida.org/tools/measurement/scamper/`.

[3] David Moore, "Pitfalls and problems with Internet data," April 2002. `http://www.caida.org/outreach/presentations/2002/ipam0203/`.

[4] L. Gao, "Inferring and Characterizaing Internet Routing Policies," in *ACM SIGCOMM Internet measurement workshop*, April 2003.

[5] G. Siganos, "Analyzing BGP policies: Methodology and tool," in *Proc. IEEE INFOCOM*, April 2004. `http://www.cs.ucr.edu/~michalis/PAPERS/siganos-info04.pdf`.

[6] L.Subramanian and S. Agarwal and J. Rexford and R. H. Katz, "Characterizing the Internet hierarchy from multiple vantage points," in *Proc. IEEE INFOCOM*, June 2002.

[7] Z. Morley Mao and David Johnson and Jennifer Rexford and Jia Wang and Randy Kaths, "Scalable and Accurate Identification of AS-Level Forwarding Paths," in *Proc. IEEE INFOCOM*, March 2004.

[8] Kuai Xu and Zhenhai Duan and Zhi-Li Lang and Jaideep Chandrashekar, "On Properties of Internet exchange points and their impact on AS topology and relationship," in *Networking*, 2004. `http://www.cs.fsu.edu/~duan/publications/networking04.ps`.

[9] Andrew Odlyzko *et al.*, "Internet economics and Internet Evolution," April 2004. `NSFITRproposal,undersubmission`.

[10] D. Vukadinovic and P. Huang, and T. Erlebach, "A spectral analysis of the Internet topology," in *ETH TIK-NR. 118*, 2001. http://www.tik.ee.ethz.ch/ vukadin/pubs/nls-tr.ps.gz.

[11] C. Gkantsidis, M. Mihail, and E. Zegura, "Spectral Analysis of Internet Topologies," in *Proc. IEEE INFOCOM*, 2003. `http://www.cc.gatech.edu/~gantsich/SpectralAnalysis/SpectralAnalysis03I%nfocom.pdf`.

[12] Dima Krioukov and Kevin Fall, "Compact routing on Internet-like graphs," in *Proc. IEEE INFOCOM*, April 2004. `http://www.krioukov.net/~dima/pub/crig-infocom.pdf`.

[13] B. Huffaker, D. J. Plummer, D. Moore, and k claffy, "Topology discovery by active probing," in *Symposium on Applications and the Internet (SAINT)*, Jan 2002. `http://www.caida.org/outreach/papers/2002/SkitterOverview/`.

[14] M. Fomenkov, k claffy, B. Huffaker, and D. Moore, "Macroscopic Internet topology and performance measurements from the dns root name servers," in *Usenix LISA, San Diego, CA 4-7*, Dec 2001. http://www.caida.org/outreach/papers/2001/Rssac2001a.

[15] A. Broido, E.Nemeth, and kc claffy, "Internet expansion, refinement and churn." European Transactions on Telecommunications, 13, No.1, Jan-Feb 2002, 33-51.

[16] A. Broido and k claffy, "Internet topology: connectivity of IP graphs," in *SPIE International Symposium on Convergence of IT and Communication, Denver, CO*, Aug 2001. http://www.caida.org/outreach/papers/2001/OSD/.

[17] A. Broido, Y. Hyun, R. Gao, and K. Claffy, "Their share: diversity and disparity in IP traffic," in *Passive and active measurement*, April 2004.

[18] P. Verkaik, A. Broido, E. Nemeth, and kc claffy, "Beyond CIDR Aggregation," April 2004. CAIDA Technical Report.

[19] M. Luckie and B. Huffaker, "Macroscopic IPv6 topology (and other) analysis," October 2003. WIDE/CAIDA workshop presentation, http://www.caida.org/outreach/presentations/2003/ipv6_wide0311/.

[20] kc claffy and B. Huffaker, "Cisco URB 2002 proposal: Connectivity Ranking of Autonomous Systems," April 2002. `http://www.caida.org/projects/cisco02autorank/`.

[21] kc claffy and B. Huffaker, "Connectivity Ranking of Autonomous Systems (using large topology sample) as viewed by CAIDA's macroscopic topology mapping project," 2003. (results of CAIDA 2002 URB grant) `http://www.caida.org/analysis/topology/rank_as/`.

[22] M. Fomenkov and B. Huffaker, "Macroscopic Topology measurements," 2003. `http://www.caida.org/analysis/topology/macroscopic/`.

[23] D. Meyer, "RouteViews," 2002. `http://www.routeviews.org`.

[24] A. Broido and k claffy, "Analysis of RouteViews BGP data: policy atoms," in *Network Resource Data Management Workshop, Santa Barbara, CA*, May 2001. `http://www.caida.org/outreach/papers/2001/NdrmBgp/`.

[25] C. System, "Critical Infrastructure Assurance Group (CIAG)," 2003. `http://www.cisco.com/security_services/ciag/`.