Public Real Data Repositories and Measurement Tools

Michalis Faloutsos
UC Riverside
Dpt. of CS
Riverside, CA
michalis@cs.ucr.edu

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Misc

General Terms

Experimentation

Keywords

Data and Measurements

Real network data is valuable and scarce. The efforts below are trying to do something for this.
We provide a list of available tools and data for measurements in both wireless and wireline networks. To avoid lawsuits, we pledge ignorance: the sites are listed as is with no implicit guarantees of content. We just browsed through them briefly.
For convenience, we created UCRchive, an electronic and extended version of the resources that are presented below. The original descriptions of the available resources was too large to include here.
Site: http://networks.cs.ucr.edu/ucrchive/+ measurement.htm

Here "+" after a "/" means that it continues to next line. Ugly notation but it makes the lines look better.
If you wish to have your trace files and software included in UCRchive, please send an email to anirban@cs.ucr.edu, cc-ed to michalis@cs.ucr.edu, with a small description of what your project is about and a link to your site from where users can download traces or software.

A. Wireless Data Resources

The CONAN (Congestion Analysis of Wireless Networks) project analyzes and makes available wireless network data traces that were collected from the IETF 62 meeting held in Minneapolis in March 2005 . The data sets contain anonymised packet traces from channels 1, 6, and 11 during one day session and one evening (plenary) session.
Site: http://moment.cs.ucsb.edu/conan.

The MOMENT lab at UCSB has a suite of open-source tools to help in the management of a testbed (visualization, monitoring) and open-source routing protocol implementations. In addition, they have a repository of wireless network traces collected at a recent IETF meeting.
Site: http://moment.cs.ucsb.edu/

CRAWDAD (Community Resource for Archiving Wireless Data At Dartmouth) will be a publicly-available community resource for collecting and working with data from wireless networks. We are hoping to include data from lots of different networks (currently mainly campus 802.11networks.
Site: http://crawdad.cs.dartmouth.edu/about.php

MOVE allows users to rapidly generate realistic mobility models for VANET simulations. MOVE is built on top of an open source micro-traffic simulator SUMO. The output of MOVE is a realistic mobility model and can be immediately used by popular network simulators such as ns-2 and qualnet.
Site: http://www.cse.unsw.edu.au/∼klan/move/

USC wireless LAN traces archive includes six-months of association patterns of on-campus WLAN users, including VPN session logs, DHCP logs, traps at the access switch, authentication server logs, flow size traces. These traces can be used to understand and model user mobility patterns and usage of WLANs.
Site: http://nile.usc.edu/MobiLib/

The WIRE1x is an implementation of IEEE 802.1x client. WIRE1x supports various EAP-based authentication mechanisms, including EAP-MD5, EAP-TLS, EAP-TTLS, and PEAP. It can work with various versions of MS Windows, including Windows XP, 2000, ME, and 98.
Site: http://wire.cs.nthu.edu.tw/wire1x/

StarEast is a multi-radio, multi-channel wireless platform which can host upto 4 wireless cards, enabling a novel multi-radio per node MESH network. They provide documentation and some drivers for setting up a testbed, particulary for Intel 2200/2915 cards and the supported features of the IPw2200 driver to configure wireless tools. They intend to release data traces in the near future.
Site: http://commnet.ee.washington.edu/funlab/

The TCP traces at the networks lab in Chungnam National University are the first public availble packet traces collected in CDMA 1x EV-DO network, Korea. The

traces could be used to study the throughput, RTT, and loss rate of TCP performance over 3G wireless network.
Site: http://networks.cnu.ac.kr/measurement/cdma-1x-evdo/

B. Wireline Data and Tools

We have to start from CAIDA, the temple of measurements. CAIDA has been collecting active and passive measurements of the Internet since 1997. They have too many tools and data sets to describe in detail.
For active measurements of Internet topology and performance, CAIDA uses the traceroute-based tool skitter that captures forward IP paths and round-trip times from skitter hosts to a specified list of destinations. Collection of this data started on January 17, 1998 and is ongoing [1] Finally, they significantly improved the accuracy of business AS relationship inference techniques, and developed a procedure to rank ASes based on results of this inference. Starting from January 2006, they archive on a weekly basis and make publicly available complete AS-level Internet topologies enriched with AS relationship information for every pair of neighboring ASes (AS-links).
For passive measurements of the Internet traffic, CAIDA has monitoring locations in several large ISPs in the United States where OC48 taps collect packet headers at large peering points. They make these data available to the research community to the extent possible while preserving the privacy of individuals and organizations who donate data or network access. This collection, started on August 14, 2002, is at the moment on hold pending the deployment of OC192 monitoring hardware. For security data, they use the UCSD Network Telescope that provides a unique view of anomalous traffic with no legitimate destination carried on the Internet. This ongoing collection started on January 31, 2001. Data monitoring malicious events (such as Denial-of-Service attack backscatter, Internet worms, and host scanning) are available for qualified researchers.
Recently, CAIDA has been implementing the Internet Measurement Data Catalog (DatCat), which will be a repository to archive meta-data for many heterogeneous data sets and will facilitate searching for and sharing of data among researchers. They expect to open this catalog for public browsing in the summer of 2006. The DatCat is a database architecture completed with a flexible annotation system. As of February 2006, they are populating and testing this database with our own annotated data sets. Upon completion of this testing phase, they will contact owners of large Internet data sets and collaborate with them to catalog their data.
Site: http://www.caida.org/data/

At University of Napoli "Federico II", they developed open source tools for Internet traffic capture, analysis and generation, and for performing active measurements of QoS parameters. The tools are called: Plab and D-ITG. The software and archives with some of the traffic and data traces are publicly available.
Site: http://www.grid.unina.it/Traffic

At UIUC, several measuring tools have been developed: (a) NVisionIP is a network state visualization tool providing an operator a multi-level interactive view into a Class B address space using NetFlows source data. NVisionIP can be used for security event monitoring but has many other network management uses including network profiling. (b) VisFlowConnect-IP is a tool for visualizing network traffic connectivity using NetFlows as source data. It can be used to see connectivity between an edge network and the Internet or solely within an internal network. (c) CANINE is a Converter and ANonymizer for Investigating Netflow Events. CANINE can convert between the following NetFlow formats: Cisco NetFlow v5, v7; NFDUMP; and Argus. As an anonymizer, CANINE provides independent anonymization for each NetFlows fields.
Site: http: //security.ncsa.uiuc.edu/distribution/+VisFlowConnectDownLoad.html
Site: http://security.ncsa.uiuc.edu/distribution/+NVisionIPDownLoad.html
Site: http://security.ncsa.uiuc.edu/distribution/+CanineDownLoad.html

NETI@home (NETwork Intelligence at home) is an open-source software package that collects network performance statistics from end-systems. NETI@home is designed to run on end-user machines and collect various statistics about Internet performance. These statistics are then sent to a server at the Georgia Institute of Technology (Georgia Tech), where they are collected, studied, and anonymized to be made publicly available.
Site: http://www.neti.gatech.edu

DIMES is a distributed scientific research project, aimed to study the structure and topology of the Internet, with the help of a volunteer community (similar in spirit to projects such as SETI@Home). The DIMES agent performs Internet measurements such as TRACEROUTE and PING at a low rate, consuming at peak 1KB/S. Running the DIMES agent will also provide you with maps of how the Internet looks from your home (currently) and will (in the future) provide you with a personalized 'Internet weather report' and other user-focused features.
Site: www.netdimes.org/

IGen is a network topology generator. It is designed to study the construction of a router-level Internet topology. IGen belongs to the family of structural topology generators. It relies on network design heuristics and operational practice to build the internal structure of each domain.
Site: http://www.info.ucl.ac.be/∼bqu/igen

TOTEM, a Toolbox of Traffic Engineering Methods, is an open-source toolbox, which provides a fully integrated set of traffic engineering (TE) methods for intra-

[1] CAIDA also offers a single instance of the router-level Internet topology graph (derived from May 2003 skitter data) and single instances of AS-level graphs derived BGP tables and from the RIPE WHOIS database.

domain, inter-domain, IP-based and MPLS-based TE. It is suitable for network optimisation, load balancing, protection/restoration, etc. The toolbox is designed to be used both as off-line in simulation mode and on-line.
Site: http://totem.run.montefiore.ulg.ac.be/

The GeoLIM project aims at providing the geographic location of an Internet host using solely its IP address. The key element of GeoLIM is its ability to transform delay measurements between landmarks (probe machines) and a target host into geographic distance constraints. Then it uses multilateration—alike GPS—to estimate the geographic location of the target host.
Site: http://planetlab-01.ipv6.lip6.fr:10000/cbg.php

C-BGP is a BGP routing solver that can be used with large-scale topologies. By BGP routing solver, we refer to a tool that computes the routes known and selected by each router in the topology once BGP has converged. C-BGP accurately reproduces the BGP decision process and it allows the definition of complex routing policies.
Site: http://cbgp.info.ucl.ac.be

Tstat provides information about classic and novel performance indexes and statistical data about Internet traffic. Tstat collects indexes at both the network (IP) layer and transport (TCP/UDP) layer. Real time protocols (RTP/RTCP) are also analyzed, allowing you to get statistical measurements on VOIP traffic, for example. Tstat analyzes either real-time captured packet traces, or previously recorded packet-level traces in various dump formats. Tstat can be used to persistently monitor links. Results can be browsed from a simple web interface.
Site: http://tstat.tlc.polito.it

The video trace library provides a large publicly available library of traces of encoded (compressed) video. The traces have been generated from encodings of over 50 videos of typically one hour length each. For each video, the library includes traces for approximately 60 different encodings in MPEG-4 (both in a single layer and two layers), H.261, and H.263, for a total of over 3000 hours worth of traces of encoded video.
Site: http://trace.eas.asu.edu

Open-source IP spoofing tester active-measurement tool used to collect data from around the Internet. The results of our study are both academically and operationally relevant to current work in network architecture, policy and security. To this end, we have anonymized our data set for the wider community.
Site: http://spoofer.csail.mit.edu
Site: http://ana.lcs.mit.edu/spoofer/

The BGP++ simulation module provides a BGP implementation for ns-2 and GTNetS network simulators. BGP++ is based on Zebra bgpd, which is modified to work in simulated ns-2 and GTNetS routers. As such, it provides a full-versed implementation of BGP that supports most of the features of Zebra bgpd including the Zebra CISCO-like configuration language. Moreover, BGP++ is designed with scalability in mind. It supports parallel distributed simulations and employs efficient memory sharing data structures, which enable its user to perform detailed simulations of up to few thousands of BGP speakers.
Site: www.ece.gatech.edu/research/labs/MANIACS/+BGP++/

An AS taxonomy repository provides: a) a classification of ASes into many different classes, b) several AS attributes calculated for every AS. The repository aims to promote deeper analysis of the macroscopic Internet structure and to inspire more adequate Internet modeling.
Site: www.ece.gatech.edu/research/labs/MANIACS/+as_taxonomy/

C. Open Source Implementation Efforts

TKN Hierarchical Mobile IPv6 Implementation on Linux. This experimental software provides a first open source implementation for the IETF Hierarchical MIPv6 protocol. It is based on the MIPL implementation. MAP discovery is based on the Linux IPv6 router advertisement daemon (radvd).
Site: http://www.tkn.tu-berlin.de/research/hmip/
FabiHTTP: an Open Source Implementation of HTTP/1.0 Webserver. This HTTP/1.0 Webserver implementation is available for both Windows and Linux versions. The code is free software and released under the GNU General Public License (GPL). Its performance has been preliminary evaluated and compared with other implementations.
Site: user.informatik.uni-goettingen.de/~teleprak/+SS2005/webperf/

The IETF NSIS Working Group is developing an extensible IP signaling framework which can be used to manage various control states in network nodes for end-to-end communications, which intends to be a replacement of the current Resource Reservation Protocol (RSVP). This site provides an open source implementation for the NSIS protocol suite, released under GPL and assisted with a user manual. This GIST implementation is able to support more than 50,000 sessions simultaneously in a low-end PC testbed.
Site: http://user.informatik.uni-goettingen.de/~nsis

GoCASP is an implementation of the Cross-Application Signaling Protocol (CASP) on the Linux operating system at the Georg-August University of Goettingen. Together with QoS signaling client protocol, it forms the generic signaling framework which was developed by researchers from Columbia University, Siemens and University of Goettingen. It is released under the GNU General Public License and user manual is provided. In addition, a software provides a measurement environment to analyze the network performance of Linux kernel 2.6.9 network stack. It includes a kernel patch and a packet generator, assisted with a user manual. The code is released under GPL.
Site: http://user.informatik.uni-goettingen.de/~casp/
Site http://user.informatik.uni-goettingen.de/~kperf/