

Updates on the Internet emulation trials from the 9th CAIDA/WIDE workshop

Hiroaki Hazeyama

Assistant Prof. of Internet Engineering Lab.
Nara Institute of Science and Technology

Contact to : nerdbox-freaks_at_wide.ad.jp

Background

- We need a large-scale internet-like test environment for testing the actual running codes of internet scale applications
 - Because **a field trial experiment** spends money and human resources to measure data, operate experiments, negotiate and announce to interests, and so on
 - Because **Legacy Network Simulators** cannot run actual running codes, we have to rewrite for network simulators, and we cannot monitor effects from OS or hardware.
- **Network Emulation** may produce the internet-like test environment
 - Controllable, manageable, tractable
 - Observing everything
 - Reasonable cost

Rough Definition

- **Network Simulation**
 - Test an algorithm or calculate some probabilistic behaviors on one or more computers (CPU, Memory, Storage)
 - Abstract various components by simulation models
 - TCP/IP, OS, queuing, network topology, loss rate
- **Network Emulation**
 - Test running code or measure behaviors on one or more servers (CPU, Memory, Storage, Network)
 - Using TCP/IP or other OS resources on an actual OS
 - Abstract various components by emulators
 - WiFi link behaviors, delay, loss rate, network topology

Internet Emulation

- Emulation which reflects some / whole characteristics of the Internet
 - Network topology
 - DNS topology
 - Routing
 - Traffic
 - End node behavior
 - etc.

In the 9th CAIDA-WIDE Workshop

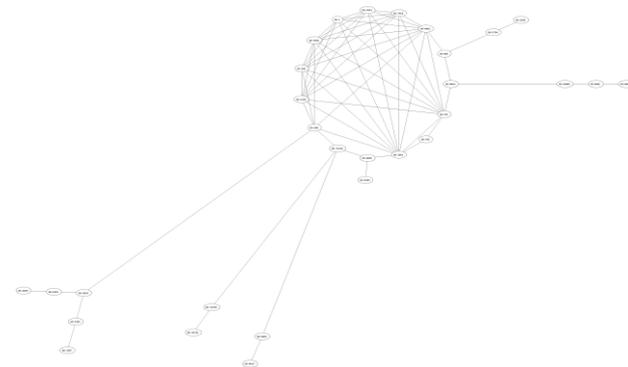
- I talked about our challenges on AS topology emulation
 - Official Records on Jan. 2008
 - 293 physical nodes with IP traceback experiments
 - 5,000 virtual nodes on XENebula
- See details in the presentation at 9th CAIDA/WIDE Workshop
 - <https://iplab.naist.jp/research/anybed/9th-caida-wide-workshop-hazeyama.pdf>

Our Current Work

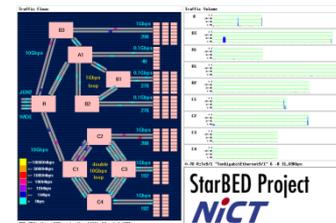
- Research
 - AS Topology Emulation
 - OSPF Topology Emulation
 - Traffic Emulation
- Release Engineering
 - Documentation
 - Hands-on Workshop
- Visualization

AS Topology Emulation

- Initially, we started AS topology emulation for our IP Traceback Experiments from 2001
- From 2006, we started the trial on full AS topology Emulation
 - We have learned experiences and TIPS on StarBED
 - We also started release engineering of the IP traceback implementation and testbed toolsets
- We collaborate with NICT HRC / SRC researchers



- Test AS Topology in 2002 -



- StarBED (NICT HRC)-

Our AS Topology Emulation Approach

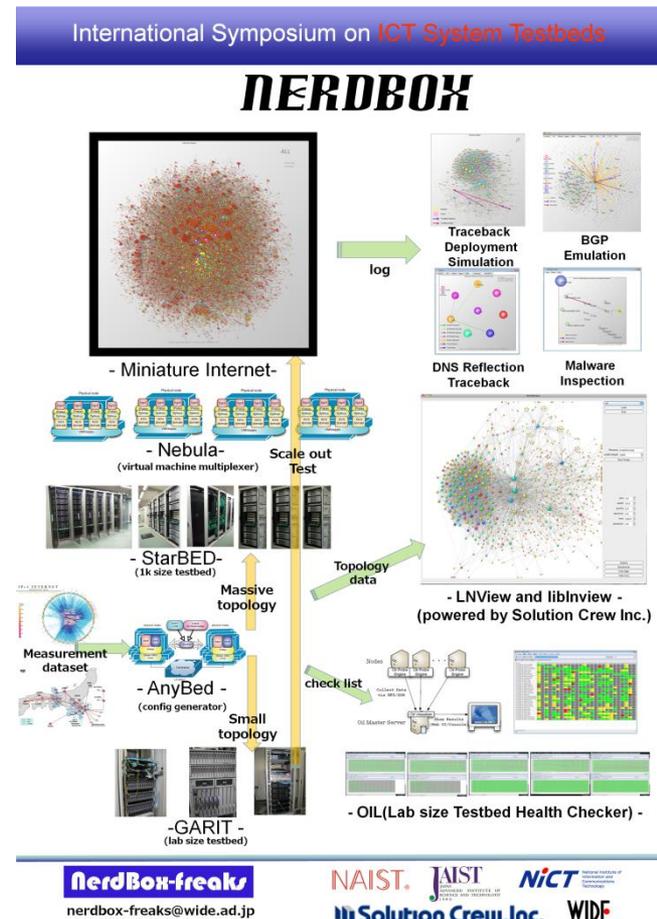
- Dataset
 - 2002 – 2005
 - Show ip BGP dataset on WIDE-BB
 - OregonIX RouteView dataset
 - 2006 -
 - CAIAD Project's IPv4 2byte ASN AS Relationship Dataset (ASRD)
- Abstraction Level
 - 1 physical / virtual node == 1 AS
 - An EBGP daemon runs on each node
 - Also, several running code to be evaluated run on each node
 - All EBGP sessions are established on a single flat VLAN
 - Due to the consumption of VLAN Numbers, unicast routing on the IP layer is our initial focus

AS Topology Emulation Toolsets

- **AnyBed**
 - XML-based Configuration Generator and AS/OSPF topology emulation experiment toolsets
 - <http://sourceforge.net/projects/anybed/>
- **OIL**
 - Health checker tools on AS topology emulation
 - Several parts of OIL are included in AnyBed
- **XENebula**
 - VM Multiplexer based on XEN
 - <http://tbn.starbed.org/XENebula/>
- **LNView**
 - QT4-based AS topology visualization and log animation tool
 - <https://iplab.naist.jp/research/traceback/>

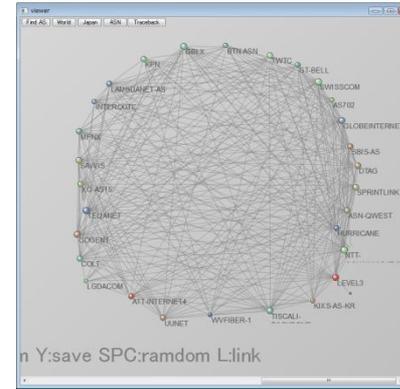
Nerdbox Toolset Chain for AS Topology Emulation

1. Get AS Relationship dataset (ASRD) from CAIDA
2. Outfit a subgraph from ASRD by AnyBed
3. Check a subgraph topology by LNView
4. Create Quagga and linux configurations on each node by AnyBed
5. Create the BGP topology and run experiments by AnyBed
6. Visualize log or status by OIL or LNview
7. Scale the test topology by XENebula over 1,000 nodes
8. Control Scenarios by AnyBed, SpringOS or XBurner

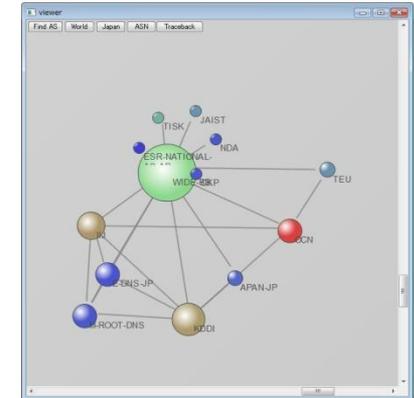


Outfitting ASRD to testbed

- Outfitting Filter Rules of ASRD
 - Top N ASes
 - M Hops ASes from Root point AS
 - Regional ASes with 1hop around ASes
 - Shortest Paths among listed ASes
 - Combination

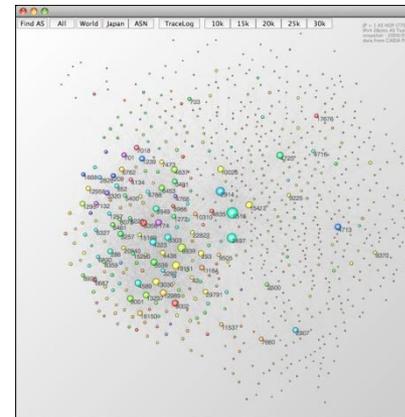


Top 30

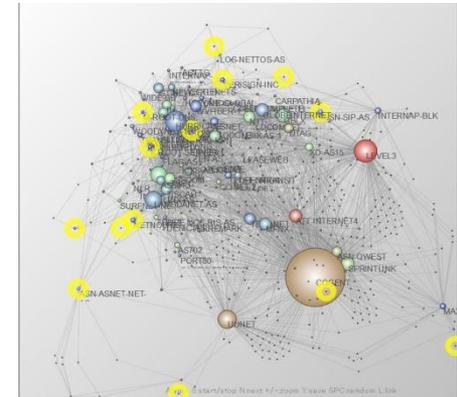


1 hop from AS 2500

- http://www.usenix.org/event/cset08/tech/full_papers/hazeyama/hazeyama_html/



JP + 1 hop around



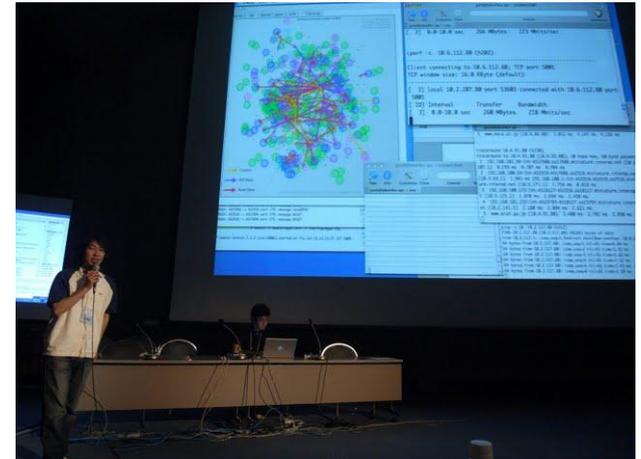
AS Paths among Root DNS

Feasibility Studies of AS topology Emulation with Virtual Machines

- ◆ Dr. Miwa (NICT SRC) reported the feasibility of AS topology Emulation with XENebula in VISA 2009
- Throughput tests 250 nodes on 5 physical nodes
 - Down to 1/10 throughput and RTT performance
- Merging emulated AS topology into WIDE-Camp network
 - Emulated JP AS Network (445ASes) on StarPOD was Stable
 - Throughput was 30-60Mbps in netperf, 98Mbps in iperf
- The first 10K AS topology Emulation
 - 150 physical nodes of Group F on StarBED
 - Tuning on Linux kernel, XEN loopback device, Quagga, XENebula's resource allocation algorithm
 - Very unstable due to the ARP storm, overflow FDB entries
 - Lack of control and measurement schemes

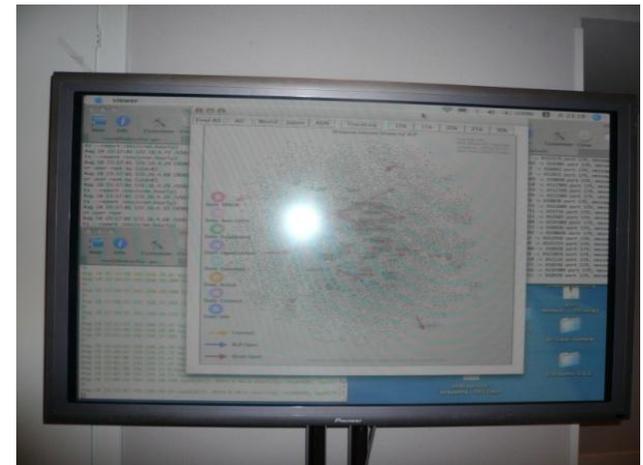
Measurement and control schemes for XENebula

- We add measurement and control schemes into XENebula for demonstrations on Interop Tokyo 2009 and ACM SIGCOMM 2009
 - syslogd
 - sshd(dropbear)
 - Shell script OIL tools
 - LNView
- In Interop tokyo 2009, We required **1hour and half** to create a stable JP AS topology (680ASes) over 50 physical diskless nodes



Challenges on 10K AS topology

- The 2nd 10K AS topology Emulation in ACM SIGCOMM 2009
 - LNView animation of BGP messages
 - Static ARP entries
 - Slow booting through ssh
- ◆ We used 222 physical nodes on StarBED
 - Each of Top 20 ASes were assigned to a whole physical node
 - Unstable 2 middle ASes on the rehearsal were assigned to a whole physical node
 - Other 9978 nodes were assigned virtual machines



ACM SIGCOMM 2009 Demo

- Result: Unstable
 - We spent 3 days to boot up 10K bgpd
 - 101 bgpd nodes didn't boot up
 - 7 nodes did not any response from network
 - On other 94 nodes, domU remounted /dev/sda1 in Read Only, therefore, bgpd could not create /var/run/bgpd.pid
 - Numerous RESET messages
- Now we are tackling to **the resource allocation problems** on AS topology Emulation **with virtual machines**
 - Memory, CPU, Network, I/O interrupts, and so on

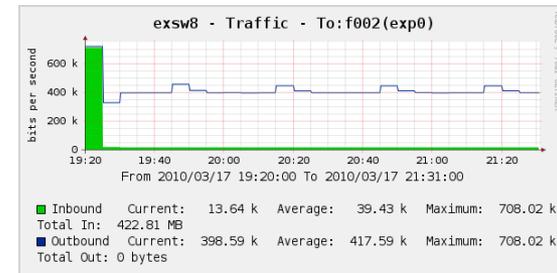
Engineering toward Stable AS topology

Emulation on Virtual Machines

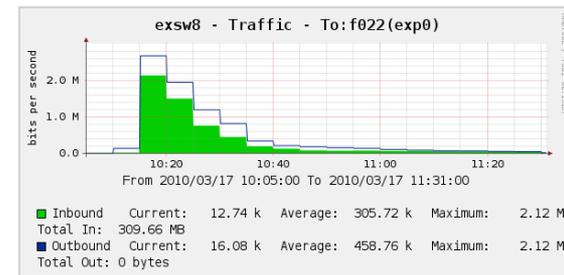
- Now, we can cook the top 500 AS in 3 min.
 - over 100 physical nodes
 - Tuning on ttylinux
 - Auto static ARP registering tool
 - Tuning on boot sequence methods
 - Measurement-based memory allocation algorithm on Quagga

◆ a stable 2,500 ASes topology can be created in 2 hours and half

◆ In 4,000 ASes over 100 physical nodes, the bgp network was never stable



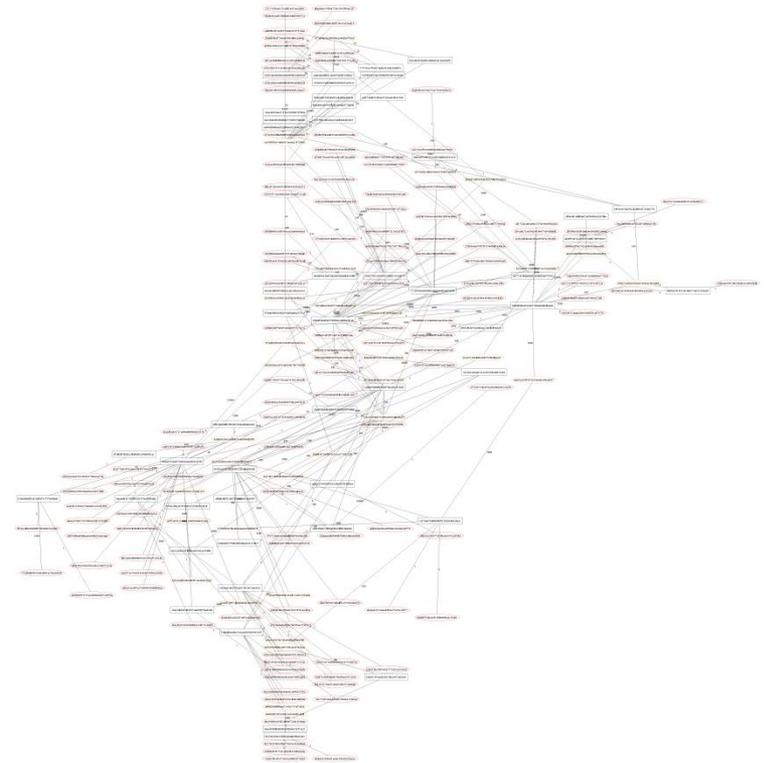
Boot time on Top 500 ASes



Boot time on Top 2,500 ASes

OSPF Topology Emulation

- WIDE-BB IPv4 OSPF topology Emulation
 - We estimates links and cost settings from OSPF LSDB, and emulates the OSPF topology by AnyBed
- ◆ Differences on actual settings of Backbone OSPF routers
 - point-to-point links
 - redistributed routes from BGP or Static



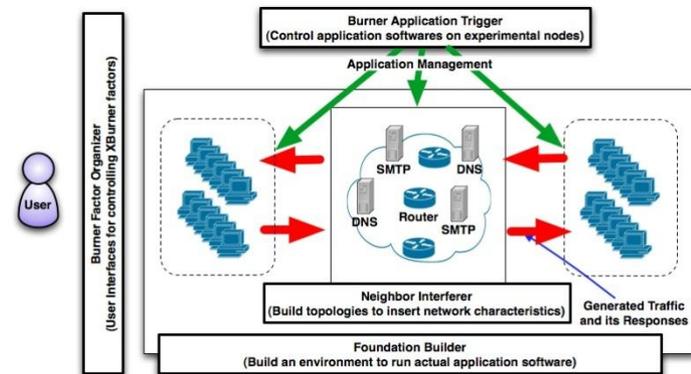
Emulated WIDE- BB IPv4 OSPF Topology
(addresses are hashed)

Traffic Emulation and Mapping

- XBurner : Dr. Miyachi (NICT HRC)
 - A Scenario-based Traffic Emulator based on SpringOS, AnyBed and XENebula
- COSMO : Dr. Miyamoto (NICT SRC)
 - Traffic Dataset Mapping Tool based on AnyBed

XBurner

- XBurner generates background traffic by running actual applications
 - Apache and wget, FTP, ping, etc.
- Componets of XBurner
 - TinyCoreLinux
 - SpringOS
 - XENebula
 - AnyBed
 - Apps



COSMO

- COSMO maps traffic dataset to an AS topology
 - Converts src / dst addresses to src / dst AS address
 - Using CAIDA's pfx2as-files and AnyBed assigned topology configuration XML data
 - Replays mapped traffic by pcap
- Current Problems on developing COSMO
 - We cannot map anonymised traffic dataset on AS topology
 - For example, the address information of MAWI dataset is masked
 - We need traffic dataset with src/dst AS information

Release Engineering

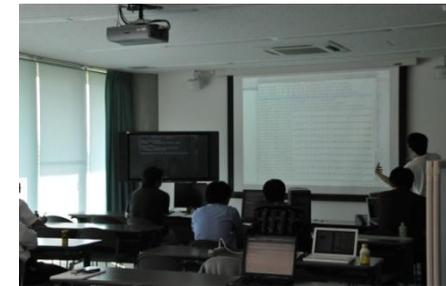
- Documentation
- Hands-on Workshops

Documentation

- Now tutorial manuals of our testbed tools are available in English or in Japanese
 - AnyBed
 - <http://iplab.naist.jp/research/anybed/>
 - Spring OS
 - <http://www.starbed.org/>
 - XENebula
 - <http://tbn.starbed.org/XENebula-en/>

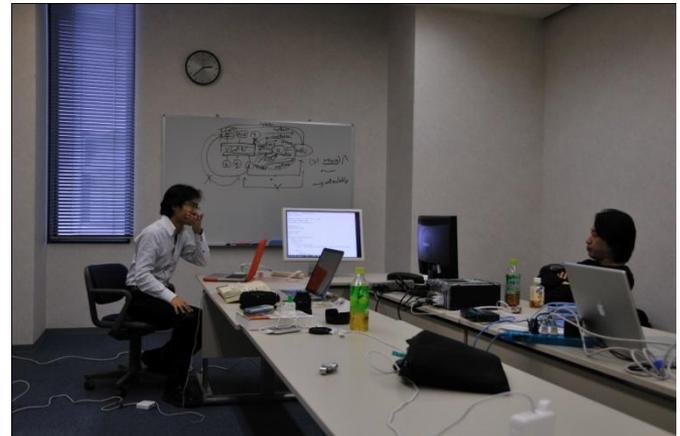
Baby StarBED Workshop

- Hands on Workshop to construct lab-size (baby) StarBED and to release each testbed toolset
- 1st workshop (May 2008) @ NAIST
 - Tutorial of SpringOS
- 2nd workshop (July 2008) @ HRC
 - Tutorial of AnyBed
- 3rd workshop (Dec. 2008) @ Keio
 - Tutorial of XENebula



Cloud Computing Developers Workshop

- Hands on workshop for Developers
 - Organized by Solution Crew Inc.
 - Supported by NAIST, NICT HRC/SRC
- 1st workshop (Oct. 2009)
 - Tutorial on XDT, ADMD
 - Developing XBurner
- 2nd workshop (Apr. 2010)
 - Tutorial on VRDF, TNF, Cybex
 - Developing LNView version 2

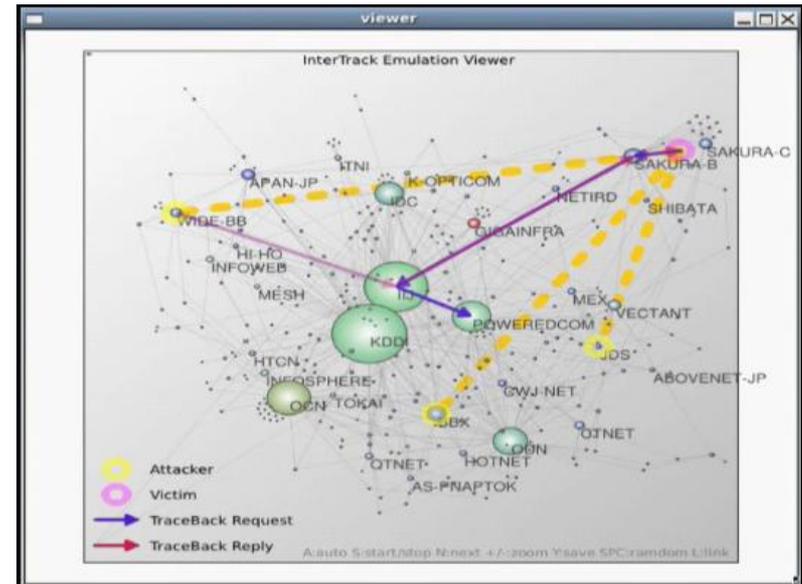


Visualization

- LNView
- Whole IPv4 2bytes AS topology picture files
- Demo of LNView

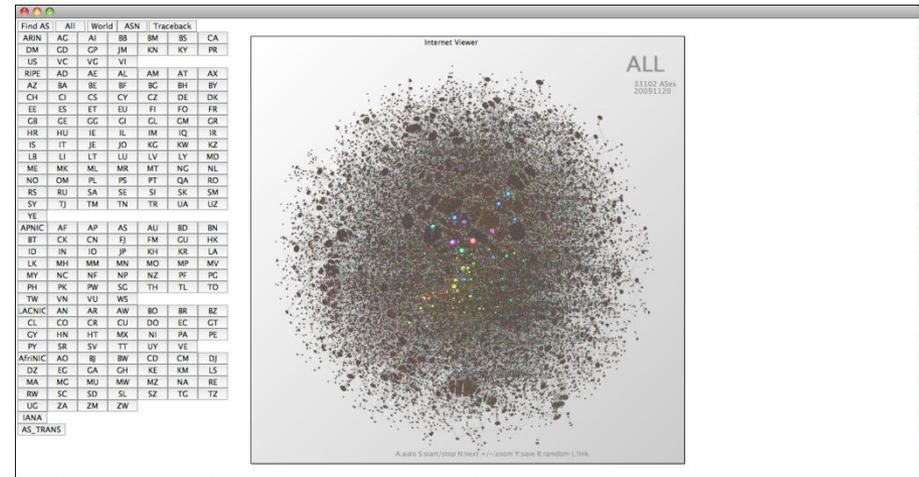
LNView

- Version 0 (- 2006)
 - Prototype for visualizing only JP AS topology
 - Yoshihide Matsumoto (NEC Software) Developed
 - Animation of IP Traceback log
 - Very heavy gravity calculation

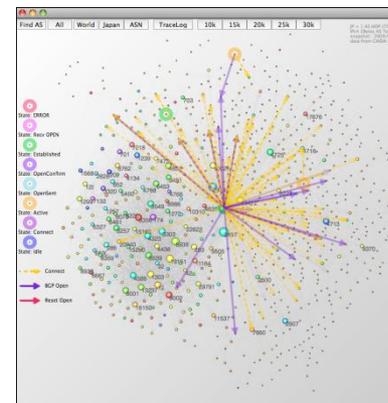
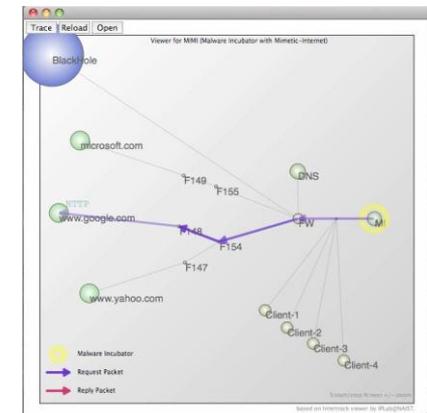


LNView

- Version 1 (2007 – 2009)
 - Visualization on the whole AS topology
 - powered by Hiroyuki Fujiwara (Solution Crew Inc.)
 - Animation of BGP log, etc.
 - Picture Printing
 - gravity calculation has been tuned up, but, it is still heavy on the whole AS topology
 - 1 Frame in 2 minutes

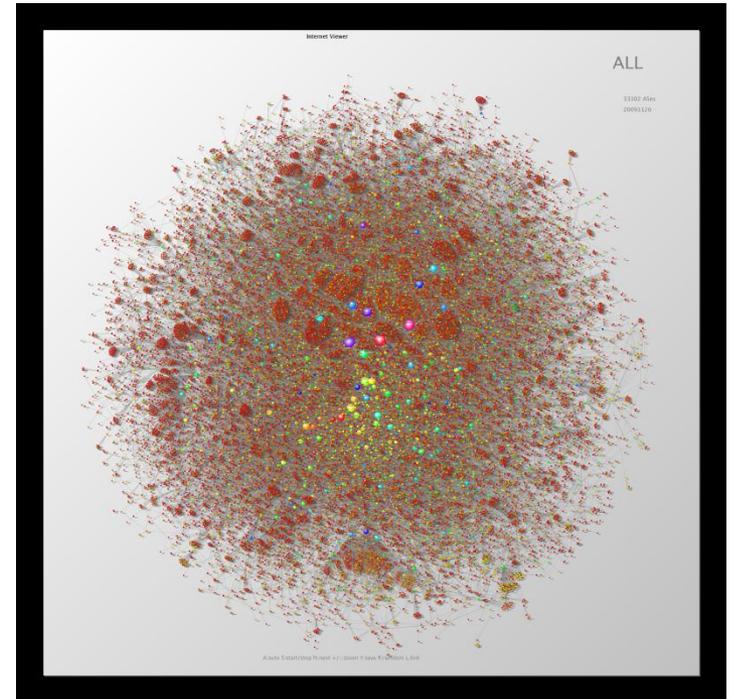


Regional AS topologies by Hazeyama

BGP log Viewer
by HazeyamaMalware inspection
by Dr. Miwa (NICT SRC)

Whole IPv4 2bytes AS topology picture files

- PNG
 - High Quality
 - 23170 x 23170 pixel
 - 200MB
 - Need 64bit OS and 8GB memory to print out
 - Medium Quality
 - 5,000 x 5,000 pixel
 - 20 MB
- PDF (vector)
 - Up to 10,000 AS topology
 - mmap error on Acrobat print cap ?
 - Top 10,000 AS : 608MB
 - JP AS topology : 162MB



5,000 x 5000 pixel PNG

Next Challenges

- Challenges on 2010
 - AS Topology Emulation with more reality
 - Mapping actual assigned global addresses
 - Mapping BGP filter rules on IRR
 - Whole IPv4 2bytes AS topology Emulation
 - Traffic Emulation
 - Replaying MAWI / CAIDA / CASFI traffic dataset

Next Challenges

- Future (We need dataset and collaborators)
 - IPv6 / 4bytes AS topology Emulation
 - DNS topology and anycast emulation
 - Multicast on AS topology emulation
 - Router level topology emulation
 - Mapping various measurement data to AS topology emulation

We need following dataset

- DNS Topology dataset
- Anonymous traffic dataset
 - with source AS and destination AS information
- IPv6 AS Relationship dataset
- 4 bytes ASN AS Relationship dataset

Mille Stones on 2010

- 2010 Q2
 - Evaluation and Demonstration of 3 minutes cooking
 - Interop Tokyo 2010 Cloud Computing Competition
- 2010 Q3
 - 3rd trial on 10,000 AS topology emulation
 - ACM SIGCOMM 2010 Demo??, SC'10 Demo??
- 2010 Q4 - 2011 Q1
 - Full IPv4 2byte ASN AS topology Emulation

Q and A

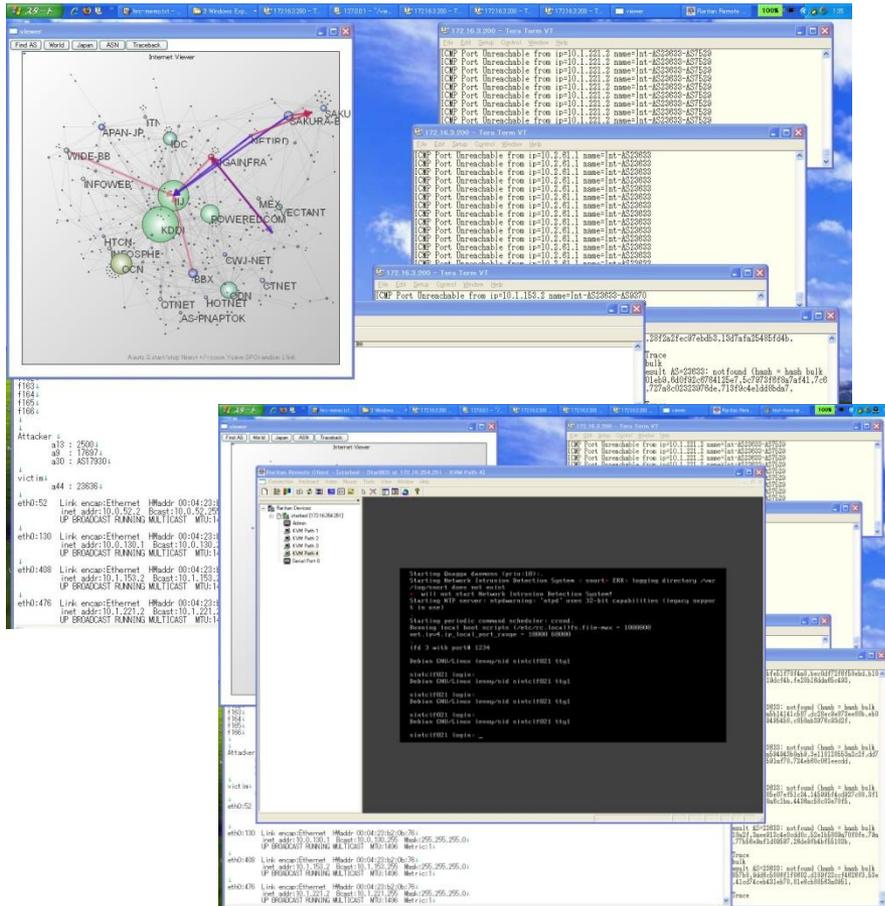
- Contact to : `nerdbox-freaks_at_wide.ad.jp`
- Software
 - AnyBed
 - <http://sourceforge.net/projects/anybed/>
 - XENebula
 - <http://tbn.starbed.org/XENebula/>
 - LNView
 - <https://iplab.naist.jp/research/traceback/>

Appendix

Demonstrations

- ACM SIGCOMM 2008
- Interop Tokyo 2009
- ACM SIGCOMM 2009
- International Symposium on ICT System Testbeds 2010

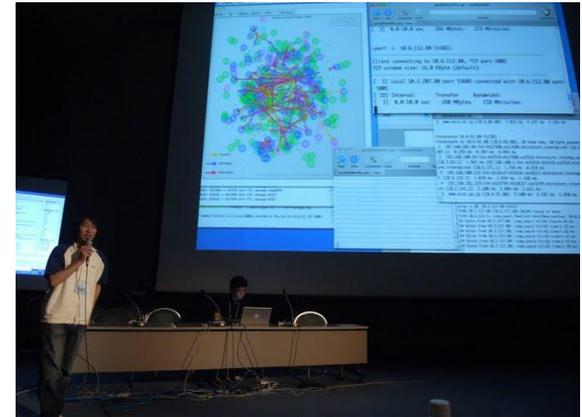
ACM SIGCOMM 2008 Demo



I used this demo environment for debugging of InterTrack (IP traceback implementation) toward the field trials with ISPs during Sept. 2008 to Sept. 2009

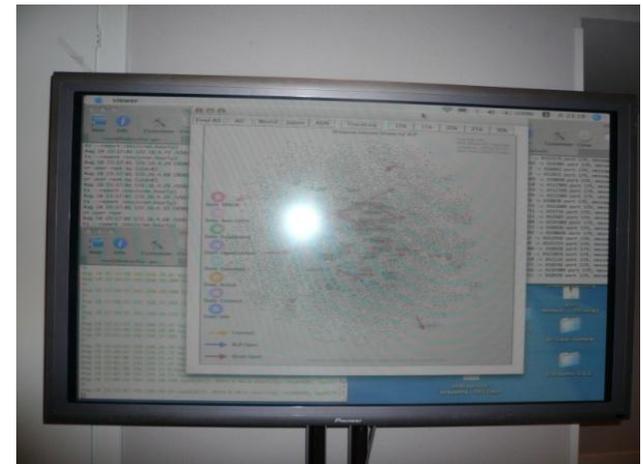
Interop Tokyo 2009

- JP domain AS topology Emulation
 - 680 ASes were created by XENebula over 50 diskless physical nodes
 - We recorded Quagga bgpd debug messages by syslog
- Animation of BGP log
 - I will run demo in visualization part
- Collaboration with P4P algorithm Demo
 - Each P4P agent connected as end node of the emulated AS topology.



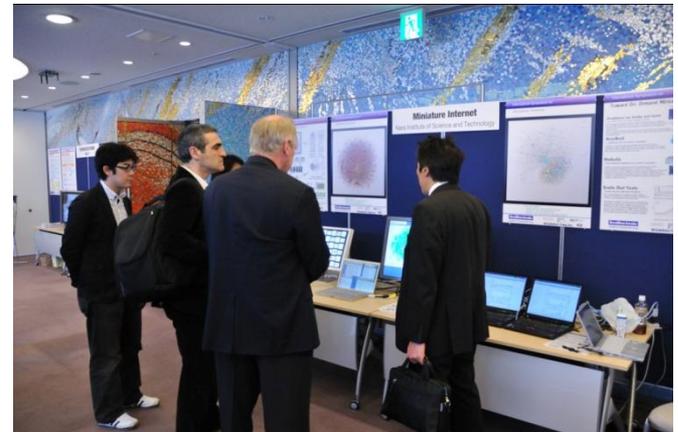
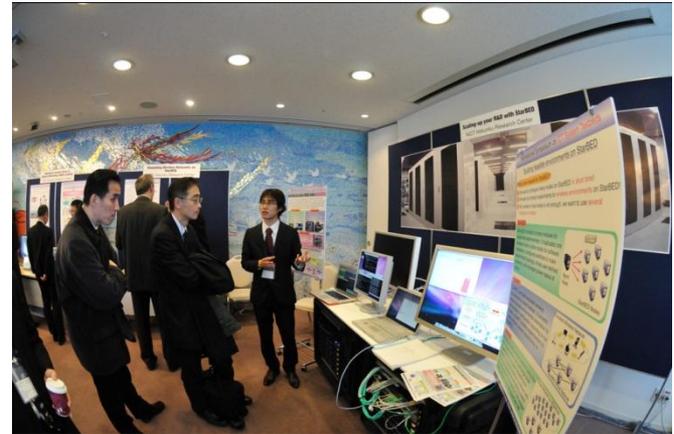
ACM SIGCOMM 2009 Demo

- The 2nd 10K AS topology Emulation
 - LNView animation of BGP messages
 - Static ARP entries
 - Slow booting through ssh
- ◆ Feedbacks from audiences
 - I'd like to create my Internet for analyzing bogon ASes and testing Secure BGP
 - I'd like to use StarBED
 - Can I run my routing daemon instead of Quagga bgpd ?



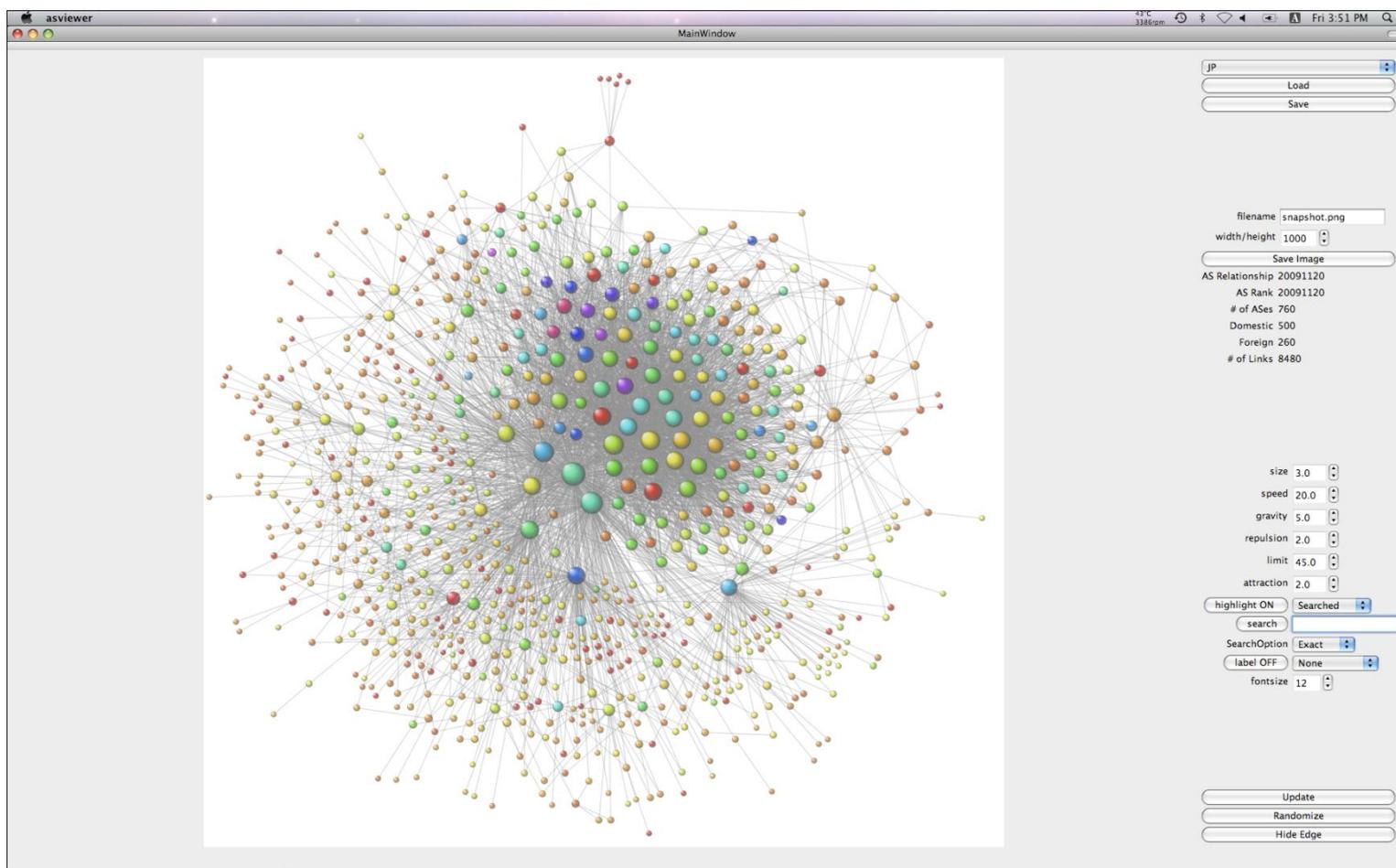
International Symposium on ICT System Testbeds 2010

- Demos by our collaborators
 - XBurner
 - 2,500 AS topology Emulation with live animation
- Other NICT HRC / SRC softwares
 - QOMET, RUNE, MAT
- Demos by heavy StarBED users
 - Panasonic Electric Works
 - Hitachi Information System

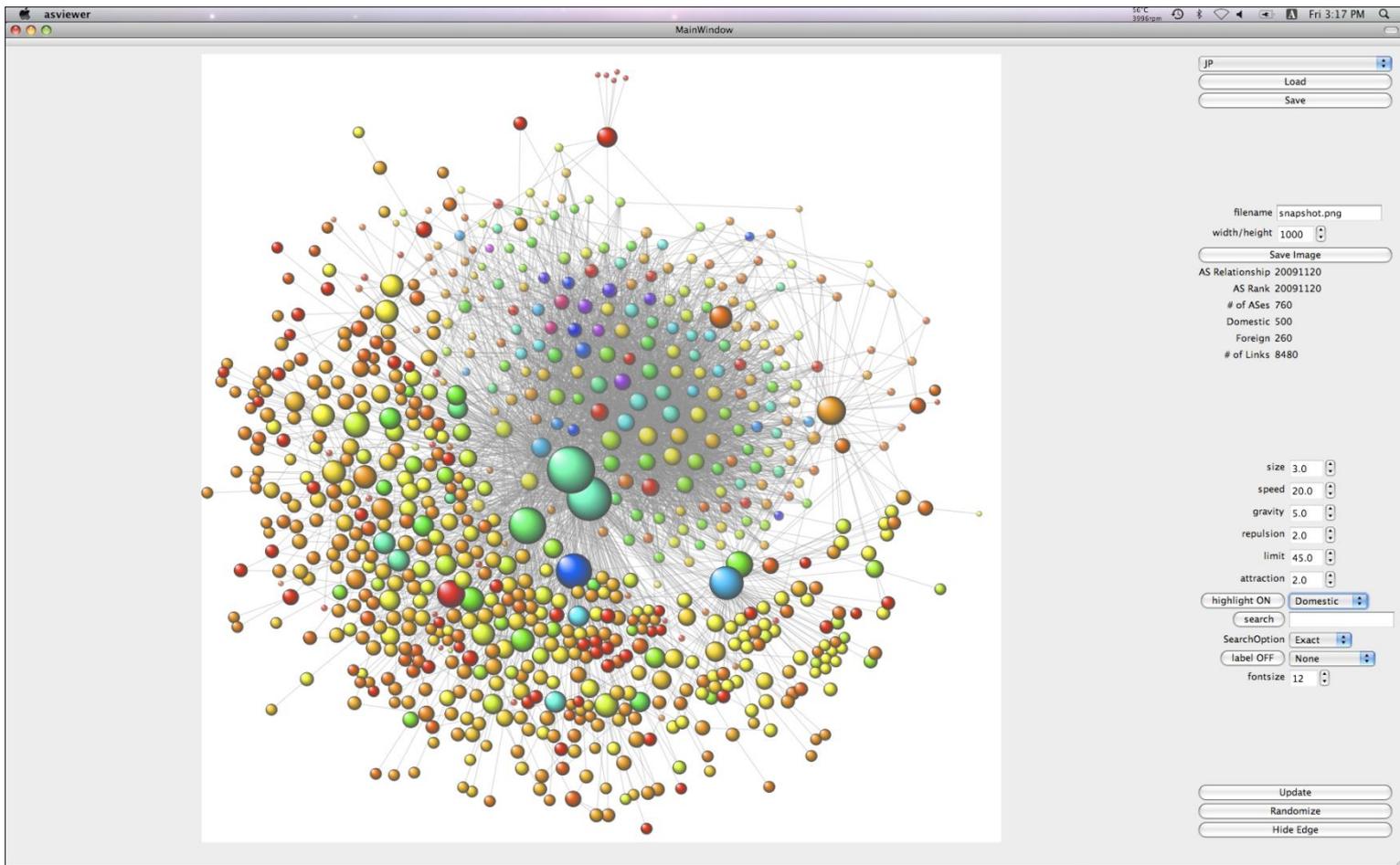


LNView's snapshot

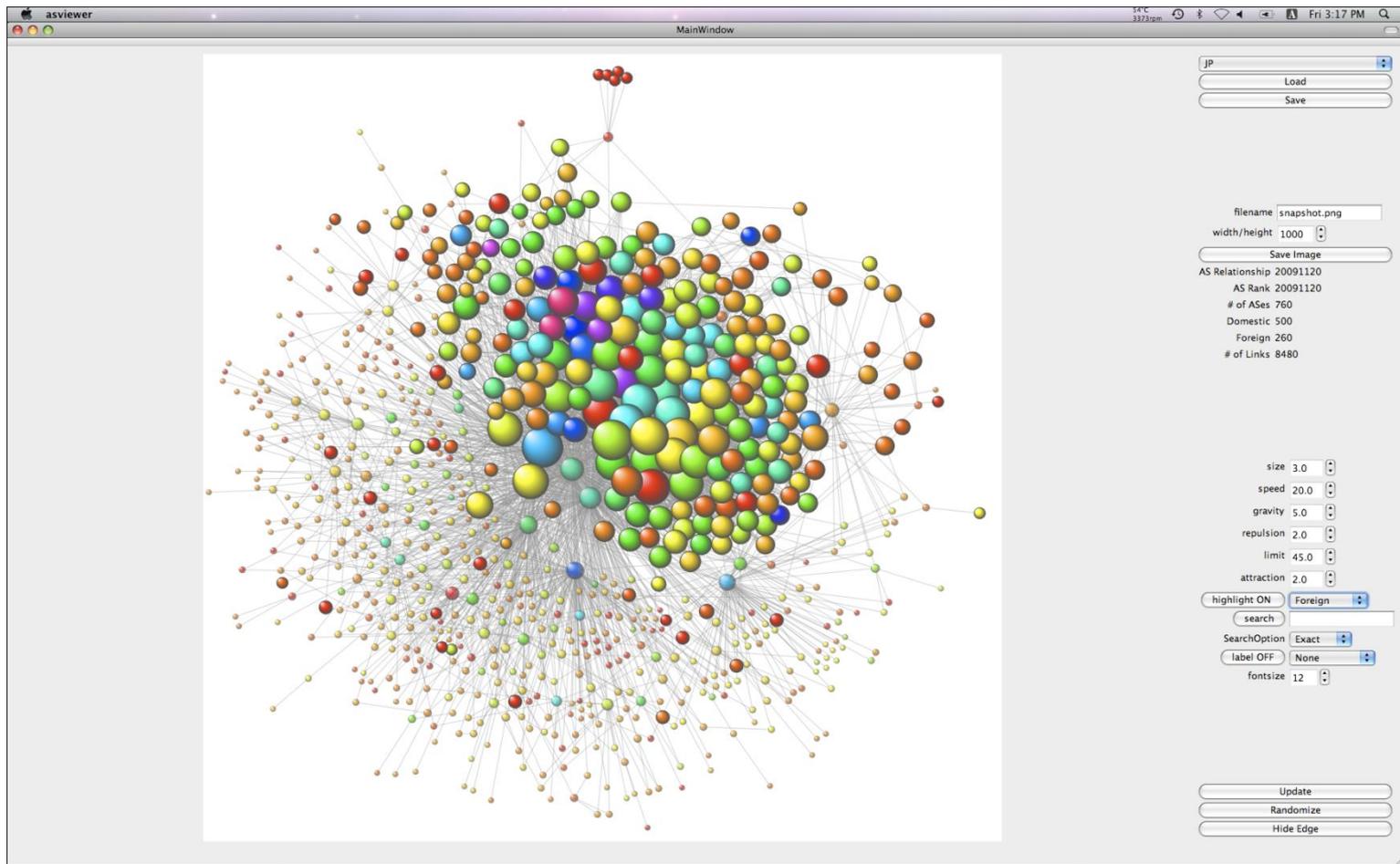
JP AS topology



Highlight Domestic ASes on JP



Highlight Foreign ASes on JP



Highlight AS 2500 (WIDE-BB) on JP

