

# Monitoring IPv6 Content Accessibility and Reachability

Contact: R. Guerin ([guerin@ee.upenn.edu](mailto:guerin@ee.upenn.edu))  
University of Pennsylvania



# Outline

- Goals and scope
- Software overview
  - Functionality, performance, and requirements
- Initial findings
  - Preliminary measurement results
  - Some interpretations
- Next steps
  - Wider scale deployments and data accessibility

# Acknowledgments

- This is a joint project between Comcast and the University of Pennsylvania, and supported in part by Comcast
- Software currently deployed at Penn and Comcast (see “References” slide for links to the monitors web front-end and other relevant URLs)

# Background

- By most accounts we are going to run out of IPv4 addresses soon (from <http://www.potaroo.net/tools/ipv4> - and many others)

- Projected IANA Unallocated Address Pool Exhaustion: **Sep-2011**
  - Projected RIR Unallocated Address Pool Exhaustion: **Jul-2012**

⇒ Although we have had IPv6 for 15 years and have not really bothered migrating to it or using it, this is about to change

- This raises two basic questions
  1. What can I access with (only) an IPv6 address?
  2. How different is it accessing it over IPv6 or IPv4?
- Answer to 1. determines how much “translation” will be needed
- Answer to 2. influences how much “translation” will be needed (see paper on this issue in the “References” slide at the end of this presentation)

# Motivations

- In spite of a number of efforts (see again “References” slide for a list or related efforts), we don’t have much/enough information when it comes to the two previous questions
- Obtaining visibility (*i.e.*, data) into answers to those questions is the primary goal of this project

# Basic Approach

- To answer the questions
  1. What can I access with (only) an IPv6 address?
  2. How different is it accessing it over IPv6 or IPv4?
- We need a methodology to
  - Identify what in the Internet is IPv6 accessible
  - Compare the performance of IPv6 access to that of IPv4
- A three-prong approach
  1. Probing the Internet for IPv6 *“accessibility”*
  2. Systematically evaluating this accessibility, *i.e.*, *“reachability”*
  3. Identifying reasons for differences between IPv6 and IPv4 access
- Monitoring system (this talk) focuses on providing information about 1. and 2.
- Data from monitoring system is a key input to the investigation of 3. (some preliminary tidbits)

# Monitoring System

- A software client that runs in hosts and probes “the Internet” for IPv6 accessibility and reachability
  - *Accessibility*: Site has a registered IPv6 (and IPv4) address
  - *Reachability*: An `http` query to the site’s IPv6 address succeeds and returns the “same” content as a query to the site’s IPv4 address

**Caveat:** Focus is on web access as opposed to other services
- A `mysql` backend database that stores the information retrieved by the monitoring client
  - A schema that keeps time-series of retrieved information and supports various structured queries
- A web front-end that displays some of the data obtained by the monitoring client
  - Continuously updated plots that reflect new data as it becomes available

# Monitoring System

## Functional Overview

- Assessing IPv6 accessibility (DNS queries)
  - Input: List of sites – top 1M from Alexa + others (standard API to import list of sites)
  - Process: DNS queries for A and AAAA records
  - Output: `mysql` database records IPv4 and IPv6 sites accessibility (status, addresses, etc.)
- Assessing IPv6 reachability (web queries)
  - Input: Sites that are IPv4 and IPv6 accessible
  - Process(1):
    - Query sites for content (`http get` of main page), and compare content (based on page size)
    - Query sites with “identical” IPv4 and IPv6 content multiple times to compare download performance
  - Output: Store results in `mysql` database
  - Process(2): `traceroute` (optional) to all pairs of site addresses with identical content
    - Note: Many fail or are incomplete – Alternative uses local LookingGlass server to obtain AS-level path
  - Output: Store `traceroute` results in `mysql` database
- Displaying IPv6 reachability
  - Percentage of sites that are IPv6 reachable
    - Top 1M (Alexa), all monitored sites, split by ranking categories, etc.
  - IPv6 vs. IPv4 “performance” measures
    - Scatterplots of IPv4 vs. IPv6 download times and speeds, ranking based comparisons
    - Raw data in table format for most of the plots



# Monitoring System

## Requirements and Characteristics

- Current system configuration (probably bare minimum)
  - Intel Core2 2.66GHz with 2GB RAM and 160GB HD (ATA 7200 rpm)
  - OS: Linux (Ubuntu 9.04 or OpenSuse 11.2)
- Required software packages (monitoring client is written in java)
  - JDK 6.0 or higher
  - mysql 5.1 or higher
  - Python 3
  - Tomcat server 5.5
- Network connectivity
  - 1GB/s E/N
  - Native IPv4 and IPv6 connectivity
- Operational characteristics
  - Monitoring IPv6 accessibility of ~3.5M sites and IPv6 reachability of ~5,000 sites takes approximately 1.5 days (software architecture allows distribution across multiple machines)
  - Storage requirements
    - Accessibility: ~275 bytes/site (135 bytes of data and 140 bytes of indices), *i.e.*, storage requirements for 3.5M sites is ~1 GB
    - Reachability: ~60 bytes/site (50 bytes of data and 10 bytes of indices), *i.e.*, storage requirements for 5k sites is ~300 kB/monitoring round
    - traceroute: 476 bytes/site (460 bytes of data and 16 bytes of indices), *i.e.*, storage requirements for 5k sites is ~ 2.4 MB/monitoring round

# Initial Deployment - Penn

IPv6 Adoption Monitor - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://mnlab-ipv6.seas.upenn.edu/monitor/index2.html

Most Visited openSUSE Getting Started Latest Headlines Mozilla Firefox

Google Calendar IPv6 Adoption Monitor Roch Guerin's Web Page

## IPv6 Adoption Monitor

### Menu

- [Home](#)
- [Figure-1](#)
- [Figure-1.2](#)
- [Figure-2](#)
- [Figure-3](#)
- [Figure-4](#)
- [Figure-4.1](#)
- [Figure-4.2](#)
- [Figure-5](#)
- [Figure-5.1](#)

\*Mouseover each menu item to view figure description

### Project Goals

As we rapidly approach the date at which the current set of IPv4 addresses will have been exhausted, i.e., in less than a year according to the latest [estimates](#), migrating to IPv6 is becoming increasingly important. This migration is, however, largely dependent on ensuring that the current IPv4 Internet, and in particular its content, becomes itself accessible over IPv6. Tracking the extent to which this is happening is the main purpose of this project.

This tracking is performed by a monitoring client that queries the Domain Name System (DNS) for IPv4 and IPv6 addresses (A and quad-A records) for a number of known sites. The list of sites queried includes the top one million (1M) web sites according to the ranking maintained by [Alexa](#), and possibly additional sites beyond the top 1M.

Content is deemed IPv6 *accessible* if DNS returns a quad-A record for the site. Sites identified as being both IPv4 and IPv6 accessible are then queried for content, and deemed IPv4 and IPv6 *reachable* if the same content can be retrieved over both. The relative performance of content retrieval over IPv4 and IPv6 is then compared based on a succession of queries.

In order to provide a comprehensive perspective on the level of IPv6 adoption across the Internet, the monitoring software is being deployed at multiple locations, with the goal of ultimately making the information gathered across locations publicly available to facilitate research and evaluation by others.

**Note:** Sites are deemed to offer the *same* content over IPv4 and IPv6 if they return pages of approximately the same size.

### Monitoring Data

The data gathered by the monitoring client is displayed in a number of figures as described below. Additional details are provided in the page associated with each figure. Figures can be scaled by selecting a corresponding area in the figure, and additional information can be obtained about individual data points by positioning the mouse cursor on top of them.

- Fig. 1: Plots the percentage of IPv6 accessible sites among all the sites queried by the monitoring client.
- Fig. 1.2: Plots the percentage of IPv6 accessible sites among the top 1M sites based on Alexa's ranking.
- Fig. 2: Identifies among sites that could be reached over both IPv4 and IPv6, how many offer the same, similar, or different content. The notions of same, similar, and different content are based on page sizes.
- Fig. 3: Percentage of IPv6 accessible sites by rank.
- Fig. 4: Scatter plot of IPv4 vs. IPv6 download times for all accessible sites.
- Fig. 4.1: Percentage by rank of sites with a lower content download time over IPv6 than over IPv4.
- Fig. 4.2: Difference by rank between IPv4 and IPv6 download times.
- Fig. 5: Scatter plot of normalized (by content size) IPv4 vs. IPv6 download times for all accessible sites.
- Fig. 5.1: Average difference between IPv4 and IPv6 download speeds by rank.
- Fig. 6: Evolution over time of IPv6 routing table size (number of prefixes).

### Want to Become Involved?

If you are interested in becoming involved in this monitoring effort, send an email to [Prof. R. Guerin](#) at the University of Pennsylvania. The software is made available under an [open BSD copyright agreement](#). However, there is an implicit understanding that sites which receive a copy of the monitoring software will be willing to make their results available to the common repository maintained by the [University of Pennsylvania](#). In particular, this implies the agreement not to disable the upload functionality of the client, which is responsible for transferring monitoring results to the Penn repository after each round of measurement.

-- Sites ranking, when used, are derived from information provided by [Alexa](#) --

Monitoring IPv6 Content Accessibility and Reachability

Done

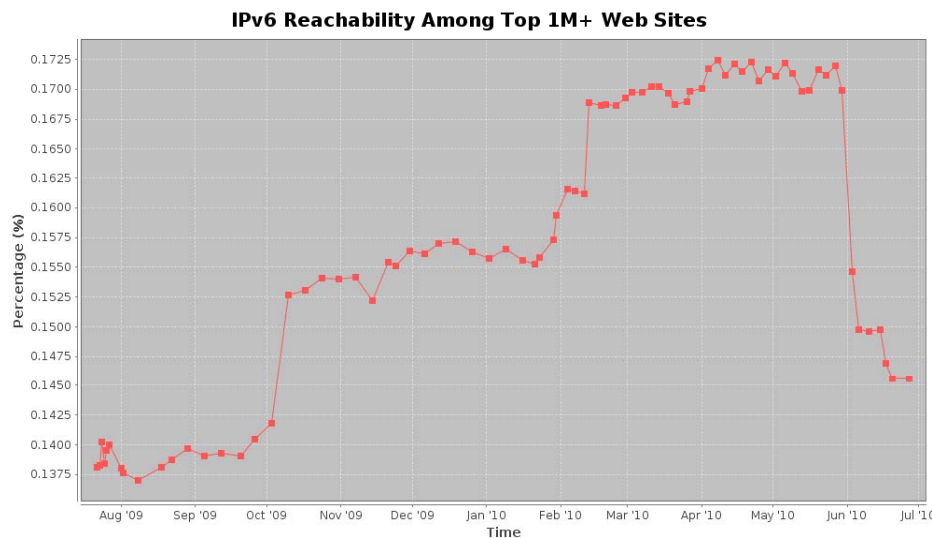
guerin@M367PC1...on/i paper.pdf - KPDF ToN - Konqueror June 2010.xls - OpenOffice

guerin@M367PC1...N F guerin@M367PC1...RO IPv6 Adoption Monitor

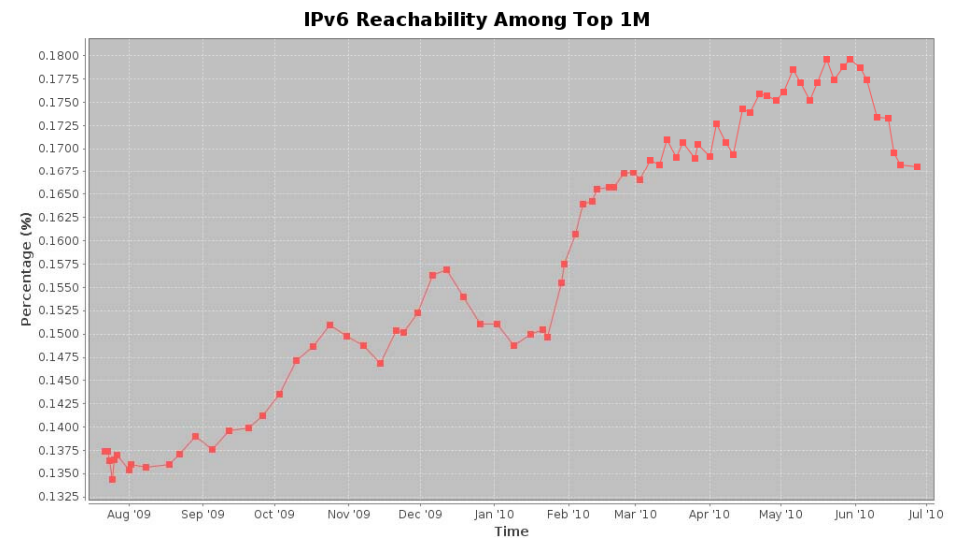
1 4 06:10 pm Tuesday 06/29/10

# Initial Data – IPv6 Penetration

- Still a long way to go, *i.e.*,  $< 0.2\%$  penetration!
  - Note: The vast majority of sites offer the same content over IPv6 and IPv4 (less than 7% have different content)
- Side Note: Recent drop apparently caused by one hosting site (in AS 28677) that stopped registering IPv6 addresses



Options: [See RawData](#)

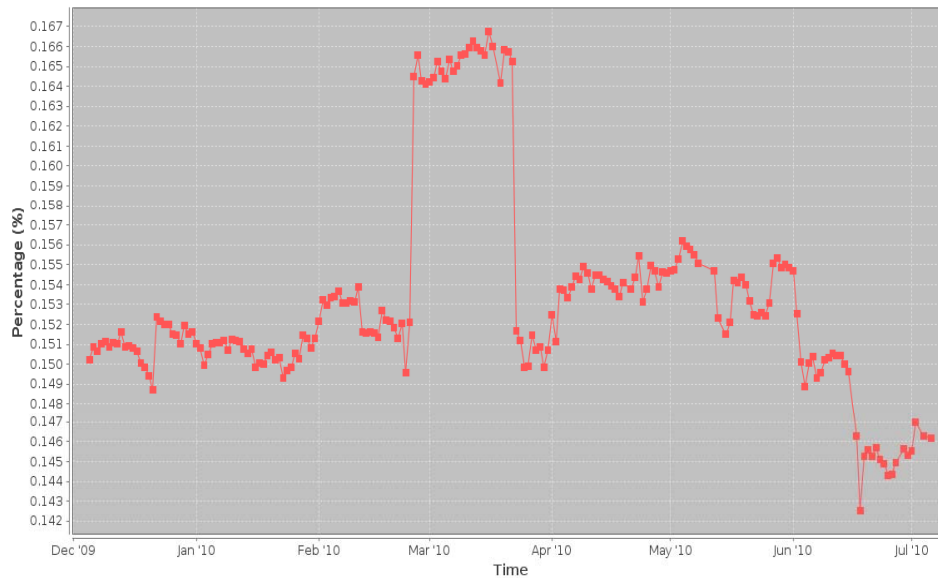


Options: [See RawData](#)

# IPv6 Reachability– Comparing Perspectives

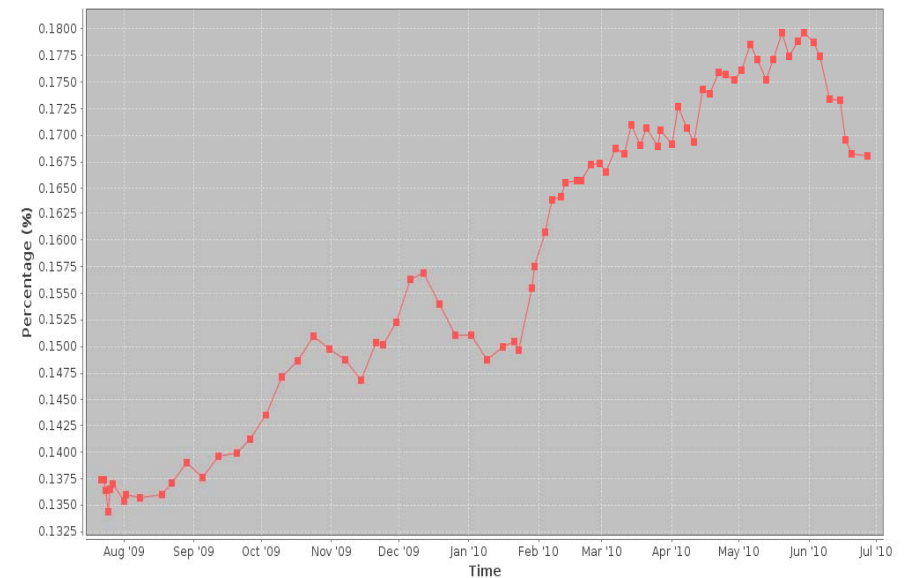
- Penn vs. Comcast monitors: IPv6 reachability for top 1M sites
  - Obvious differences!
  - In the process of exploring where they come from

Comcast



Options: [See RawData](#)

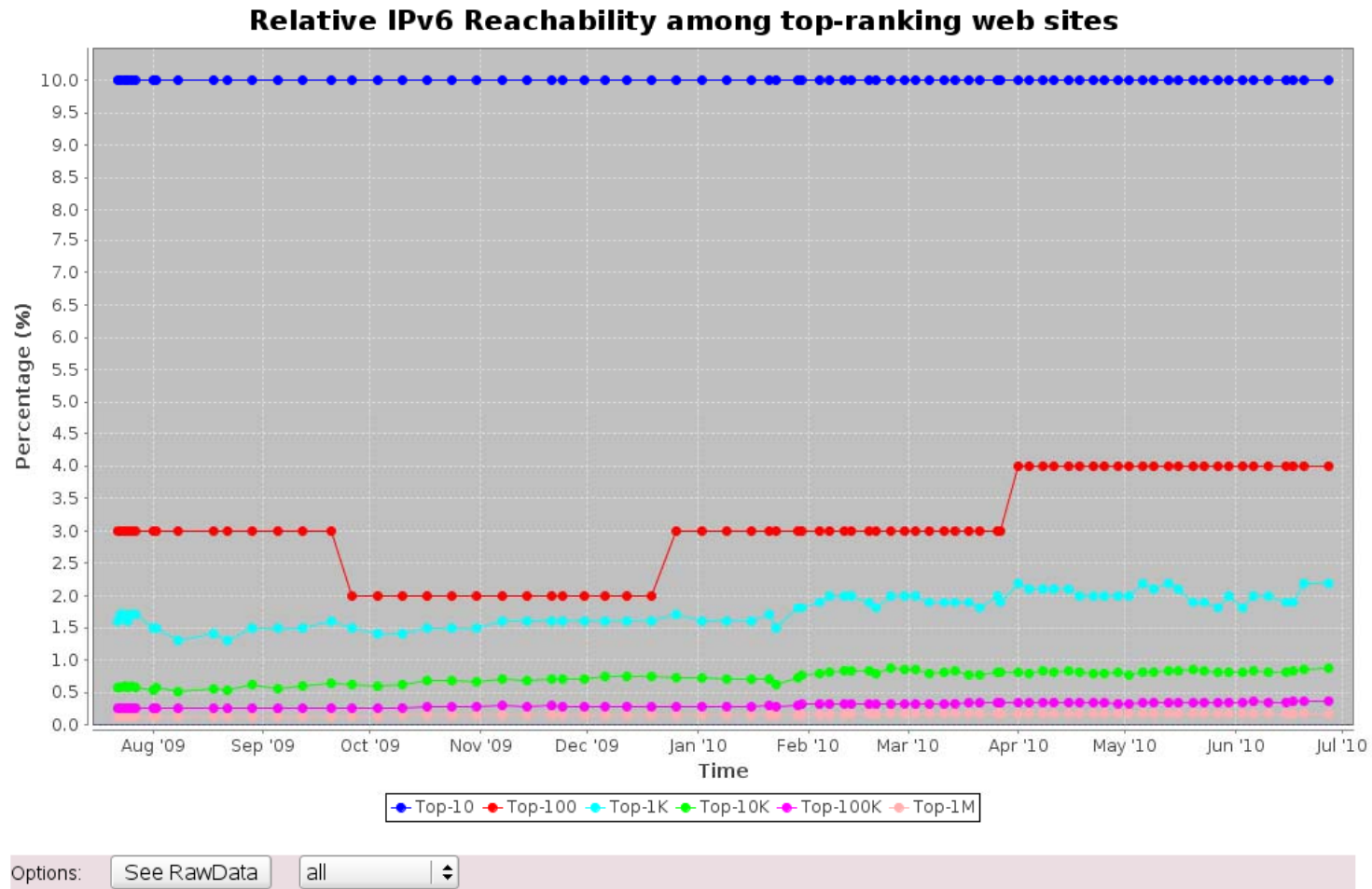
Penn



Options: [See RawData](#)

# Initial Data – IPv6 Penetration by Rank

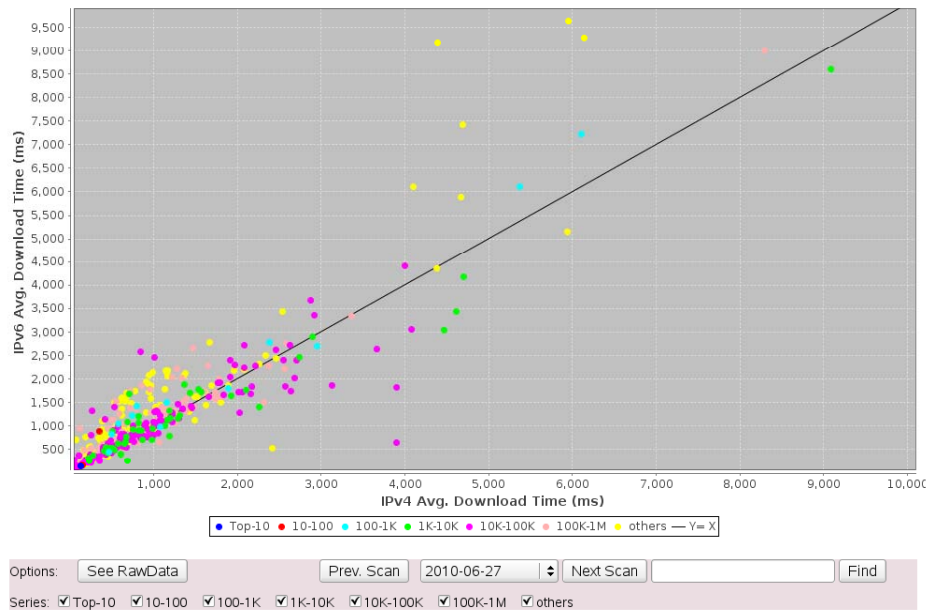
- More popular sites more likely to be IPv6 reachable



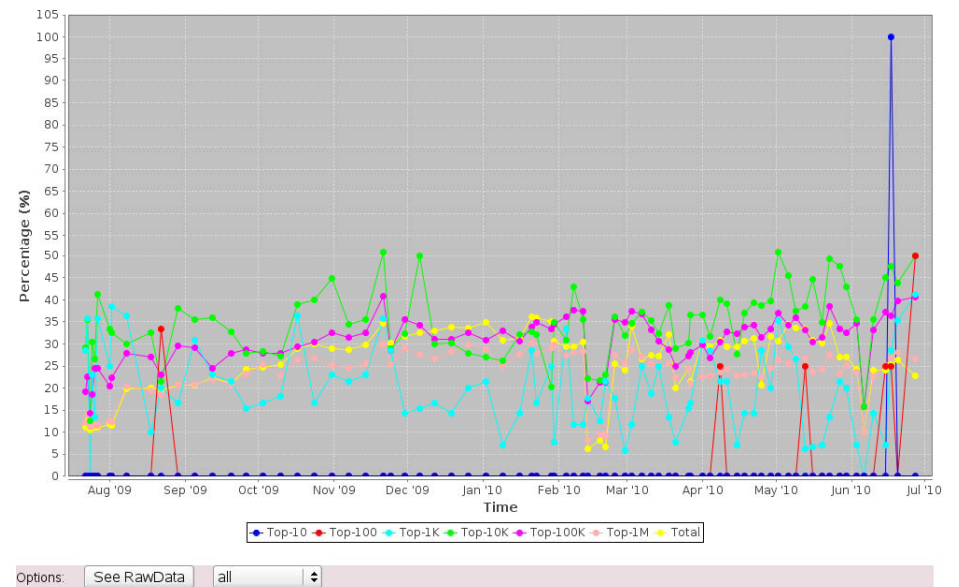
# Initial Data – Performance

- Room for improvement, *i.e.*, IPv6 yields better performance in only about 25% of the cases

Above (below)  $y=x$  line  
IPv6 is worse (better)



Percentage of sites for which IPv6 is better



# A Closer Look at Performance Differences (June 2010 Data)

- The more popular websites seem to fare slightly better than average except for the very top ones
- When IPv6 is better it is usually marginally better, while IPv4 can be significantly better

	Top 1M	Top 100k	Top 10k	Top 1k	Top 100	All Websites
IPv4 Better	951 (79.11%)	148 (64.06%)	26 (53.06%)	8 (61.53%)	3 (100%)	3516 (77.96%)
IPv6 Better	251 (20.88%)	83 (35.93%)	23 (46.93%)	5 (38.46%)	0 (0.00%)	994 (22.03%)
IPv4 Better	Top 1M	Top 100k	Top 10k	Top 1k	Top 100	All Websites
> 100%	143 (11.89%)	14 (6.06%)	1 (2.04%)	2 (15.38%)	2 (66.67%)	598 (13.25%)
50% to 100%	192 (15.97%)	7 (3.03%)	2 (4.08%)	1 (7.69%)	0 (0.00%)	629 (13.94%)
25% to 50%	141 (11.73%)	28 (12.12%)	3 (6.12%)	2 (15.38%)	0 (0.00%)	512 (11.35%)
0% to 25%	475 (39.51%)	99 (42.85%)	20 (40.81%)	3 (23.07%)	1 (33.33%)	1777 (39.40%)
IPv6 Better	Top 1M	Top 100k	Top 10k	Top 1k	Top 100	All Websites
> 100%	5 (0.41%)	2 (0.86%)	1 (2.04%)	0 (0.00%)	0 (0.00%)	21 (0.46%)
50% to 100%	7 (0.58%)	3 (1.29%)	2 (4.08%)	0 (0.00%)	0 (0.00%)	27 (0.59%)
25% to 50%	16 (1.33%)	9 (3.89%)	5 (10.20%)	0 (0.00%)	0 (0.00%)	71 (1.57%)
0% to 25%	223 (18.55%)	69 (29.87%)	15 (30.61%)	5 (38.46%)	0 (0.00%)	875 (19.00%)

# Initial Data Interpretation

## Where Do Differences Come From?

- Possible causes
  - Network data paths (tunnels, IPv6 forwarding, etc.)
  - Network control plane (routing, peering agreements, etc.)
  - Others
    - CDN type mechanisms (most don't offer an IPv6 service)
    - End-systems
- Preliminary classification
  - Same destination (AS) and  $\sim$  same (AS) path: Likely data plane issue
  - Same destination (AS) and  $\neq$  (AS) path: Likely control plane issue
  - Different destinations (ASes): Possible CDN (or maybe just configuration)



# Initial Analysis (End 2009 Data)

## Where Do Differences Come From?

- Differences between IPv6 and IPv4 paths seem to be the dominant reason (58% overall) for better IPv4 performance
- As expected, CDNs have a bigger impact among more popular web sites (23% of top 1k sites)
- Possible recommendations/conclusions)
  - Improve IPv6 peering (should affect overall performance)
  - Lobby for IPv6 support by CDNs to promote IPv6 adoption by more popular sites

	Same Destination ~ Same Path	Same Destination Different Paths	Different Destinations	Don't Know	Total
ALL Websites	207 (6.7%)	1799 (58%)	188 (6.1%)	898 (29%)	3092
Top 1 M	68 (7.8%)	565 (65%)	43 (5%)	194 (22.3%)	870
Top 100 k	24 (15.5%)	77 (49.7%)	4 (2.6%)	50 (32.3%)	155
Top 10 k	5 (10.6%)	19 (40.4%)	5 (10.6%)	18 (38.3%)	47
Top 1 k	0 (0%)	4 (30.8%)	3 (23.1%)	6 (46.2%)	13
Top 100	0 (0%)	1 (100%)	0 (0%)	0 (0%)	1
IPv4 Better	128 (61.8%)	1325 (73.7%)	127 (67.6%)	530 (59%)	2110 (68.2%)
IPv6 Better	79 (38.2%)	474 (26.3%)	61 (32.4%)	368 (41%)	982 (31.8%)

# Next Steps

- **Deployment:** Monitoring clients at additional locations
  - Software package available for distribution
    - Two versions: With and without `traceroute` component
  - OpenBSD copyright license (source code), **BUT** with the commitment to share data
    - Monitoring client includes built-in upload facility to common repository
  - Initially targeting 15-20 sites to provide global/diverse vantage points

Send email to [guerin@ee.upenn.edu](mailto:guerin@ee.upenn.edu) if interested and specify version (with or without `traceroute` – Preference to sites willing to do `traceroute`)

- **Development:** Open global repository of monitoring data (Hint: Need additional resources!)
  - Current version only setup for data uploads to ensure persistence of all monitoring information, with some data made available through web front-end
  - Final version to offer open access of full data (direct `mysql` access) to participants
- **Analysis:** Temporal and spatial analysis of data
  - Evolution over time for different categories of sites
  - Correlation of data and perspectives across monitoring sites

# References and Relevant Links

- This project
  - Paper on [“Fostering IPv6 Migration Through Network Quality Differentials”](#)
  - [Penn](#) and [Comcast](#) IPv6 monitors
- Other IPv6 resources (an obviously incomplete list)
  - OECD report on [“Internet Addressing: Measuring Deployment of IPv6”](#)
  - [RIPE IPv6 measurements compilation](#)
  - Geoff Huston [article](#) and [stats](#) on end-systems IPv6 abilities and preferences
  - Mark Prior [IPv6 status survey](#)
  - Mike Leber [Global IPv6 deployment progress Report](#)
  - [IPv6 Act Now](#)
  - [SixXS IPv6 prefix visibility](#)
  - [amsix](#) traffic statistics
  - [Hurricane Electric IPv6 service](#)