Assessing IPv6 Through Web Access A Measurement Study and Its Findings

Mehdi Nikkhah, Roch Guérin

Dept. Elec. & Sys. Eng
University of Pennsylvania

Yiu Lee, Richard Woundy

Comcast Corporation

Penn

University of Pennsylvania

Outline

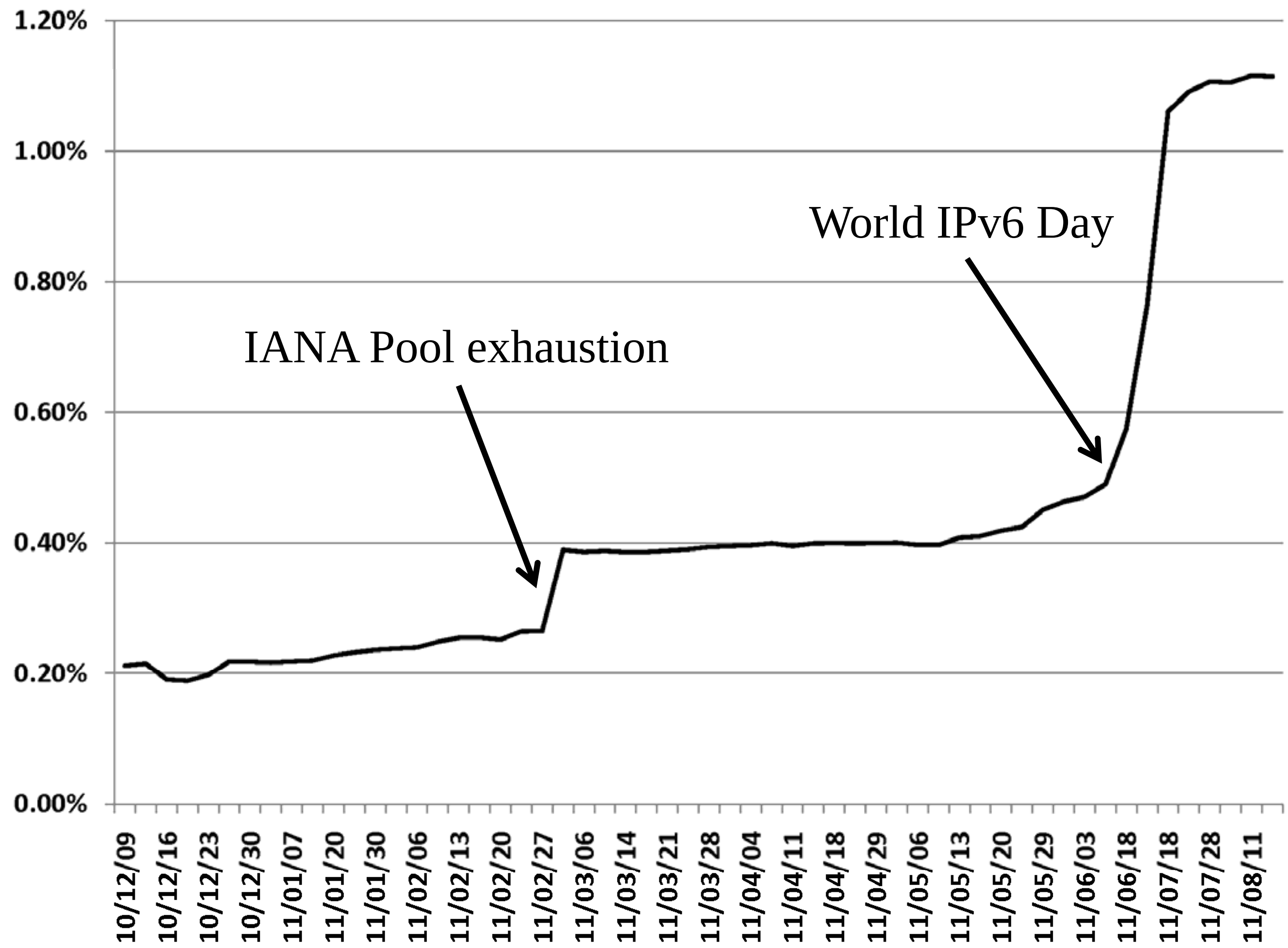
- Background and motivations
- Measurement infrastructure
- Measurement methodology

- Measurement data and findings
- Summary and next steps

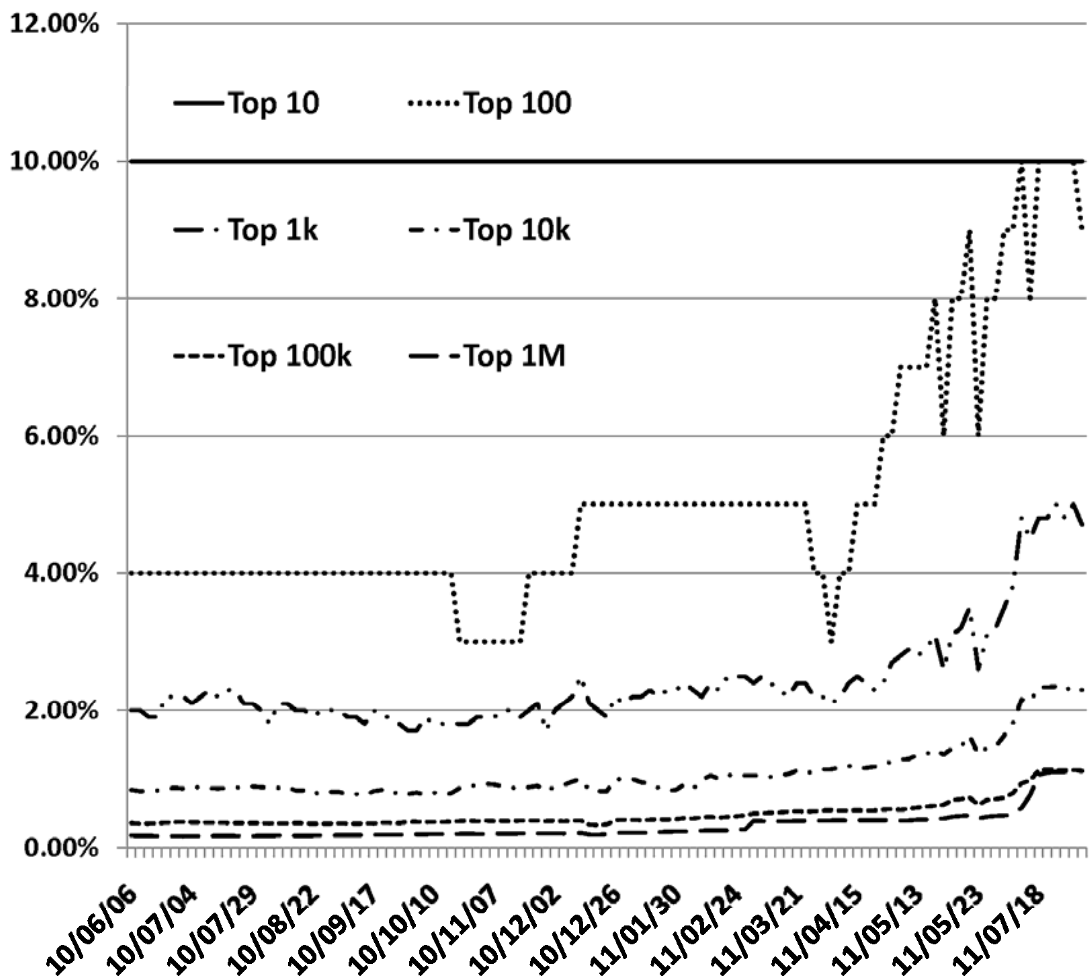
Motivations

- We “ran out” of IPv4 addresses in Feb. 2011
  - This was not unexpected and did not bring the Internet to a screeching halt, but it is a clear indication that we have entered a new period where a key Internet resource (addresses) will become scarce
- We’ve had a solution to the problem for over 15 years – It’s called IPv6
  - But for that solution to work, it has to be enabled across the Internet, and that has so far not really been the case…

Sample IPv6 Accessibility Data (Penn) Top 1M Sites



Sample IPv6 Accessibility Data (Penn) Top 1M Sites by Rank



Motivations

- We "ran out" of IPv4 addresses in Feb. 2011
  - This was not unexpected and did not bring the Internet to a screeching halt, but it is a clear indication that we have entered a new period where a key Internet resource (addresses) will become scarce
- We've had a solution to the problem for over 15 years – It's called IPv6
  - But for that solution to work, it has to be enabled across the Internet, and that has so far not really been the case…
- There are many (good) reasons that have been put forward to explain the lack of IPv6 success to-date
- Our goal is ***NOT*** to explain why we are where we are
- Instead we want to understand
  - Where are we exactly when it comes to IPv6 deployment?
  - What are some remaining issues that may stand in the way?
  - Are there specific steps we can take to alleviate them?

A Measurement-Based Approach

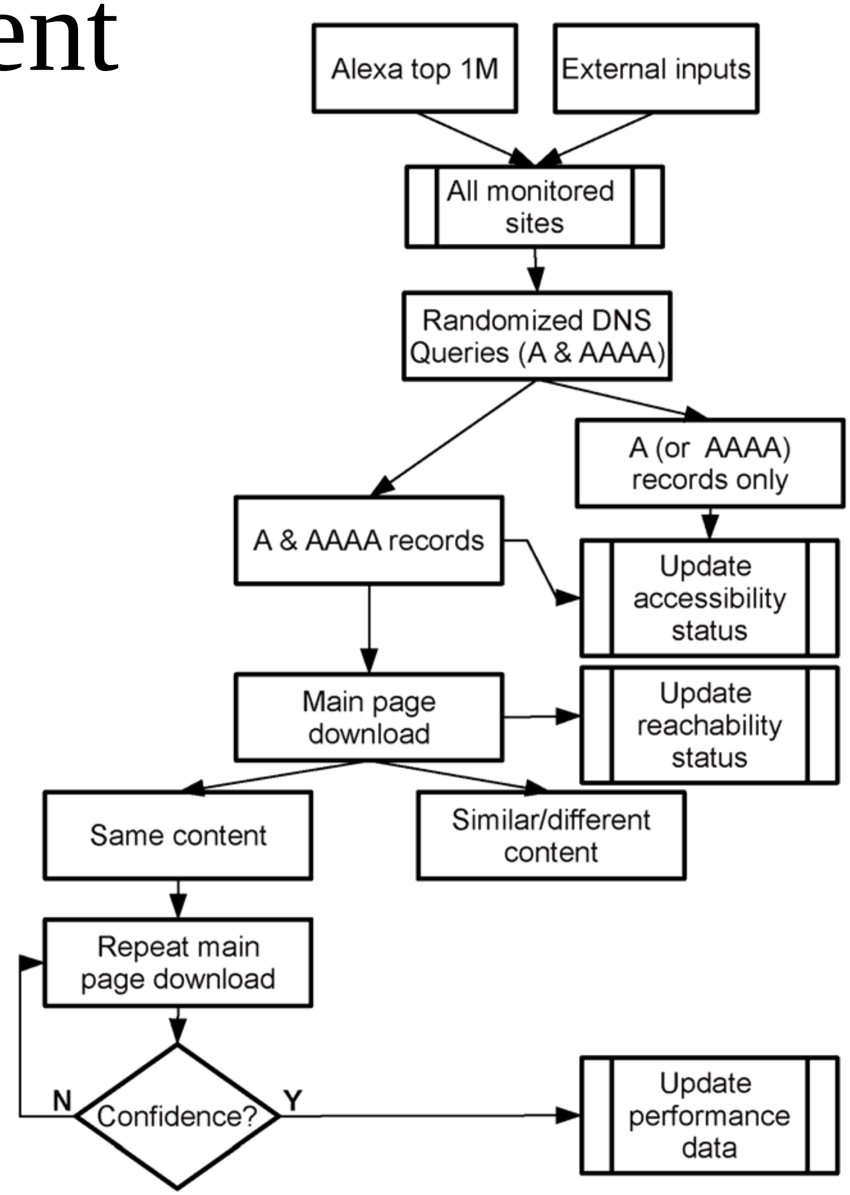
- Assessing IPv6 deployment status
  - There are many aspects and equally many metrics one could target
  - We'll focus on one, which is reasonably representative, *i.e.*, web access – how many web sites are ***natively*** accessible over IPv6 and how does IPv6 access compare to IPv4 access?
- Quantifying Internet-wide IPv6 web accessibility
  - A monitor *client* that regularly checks for IPv6 (and IPv4) accessibility of a large number of web sites
  - Multiple vantage points from which the monitor client is run
  - A common repository that aggregates measurement results across vantage points

Monitors Locations

Vantage Points		Date on-line	AS_PATH	Type
Comcast	(B)	2/4/11	Y	Commercial
Loughborough U.	(D)	4/29/11	Y	Academic
Penn	(A)	7/22/09	Y	Academic
UPC Broadband	(C)	2/28/11	Y	Commercial
Go6-Slovenia	(E)	5/19/11	N	Commercial
Tsinghua U.	(F)	3/22/11	N	Academic

Monitoring Client

- Inputs: Alexa top 1M and imported sites
- DNS queries for A and AAAA records
- For sites with A and AAAA records
  - Initial query to determine content similarity
    - Query order randomized in each monitoring round
  - Subsequent queries compare IPv6 and IPv4 download times
    - Target confidence interval to minimize impact of transient fluctuations
- IPv6 and IPv4 AS_PATHS retrieved
- Final results are stored to `mysql` database and uploaded to common repository (at Penn)



Measurement Data Overview

- From each vantage-point
  - Download times + page size (download speed) for all web sites accessible over IPv6 and IPv4
  - One or two monitoring rounds per week for several months
  - AS_PATH information when available
- Slightly different lists of monitored sites at each vantage point
  - Different start dates
  - Asynchronous sampling of Alexa (Alexa churn)
  - Local additions (Penn)
- Download speed averaged over entire monitoring period
  - Sites that fail to meet confidence targets are eliminated

Vantage Points	# Sites (unique IPs)
Comcast	844,355
Loughborough U.	883,413
Penn	1,633,606
UPC Broadband	946,977
Go6-Slovenia	850,954
Tsinghua U.	917,582

Comparing IPv6 and IPv4 Web Access

IPv4 is better (faster) over 60% of the time

WHY?

Measurement Data Scope

# IPv6+IPv4	Comcast	LU	Penn	UPCB	All
Sites (total)	4,568	5,069	12,385	7,843	-
Sites (kept)	3,525	3,906	7,994	4,418	-
Dest. ASes (IPv4)	724	801	1,047	766	1,364
Dest. ASes (IPv6)	592	642	727	609	1,010
ASes crossed (IPv4)	922	1,019	1,332	988	1,785
ASes crossed (IPv6)	742	764	849	746	1,208

Causes of Measurement Inaccuracies

	Insufficient Samples	↑	↓	↗	↘
Comcast	251	83	52	530	127
Loughborough U.	258	49	63	419	374
Penn	2,807	180	103	732	569
UPCB	1,146	233	214	1,033	799

- No performance bias identified among sites removed because of unstable performance
  - Does not favor either IPv6 or IPv4 nor does it display strong association with a specific type of connectivity

Causes for IPv6-IPv4 Differences

- There are four major factors that can affect how IPv6 and IPv4 perform
  - **(E)** The client **E**nd-system
  - **(S)** The **S**erver end-system and its access network
  - **(D)** The network **D**ata plane
  - **(C)** The network **C**ontrol plane

The main focus is on assessing **(D)** and **(C)**, *i.e.*, the ***network***, and the findings are that

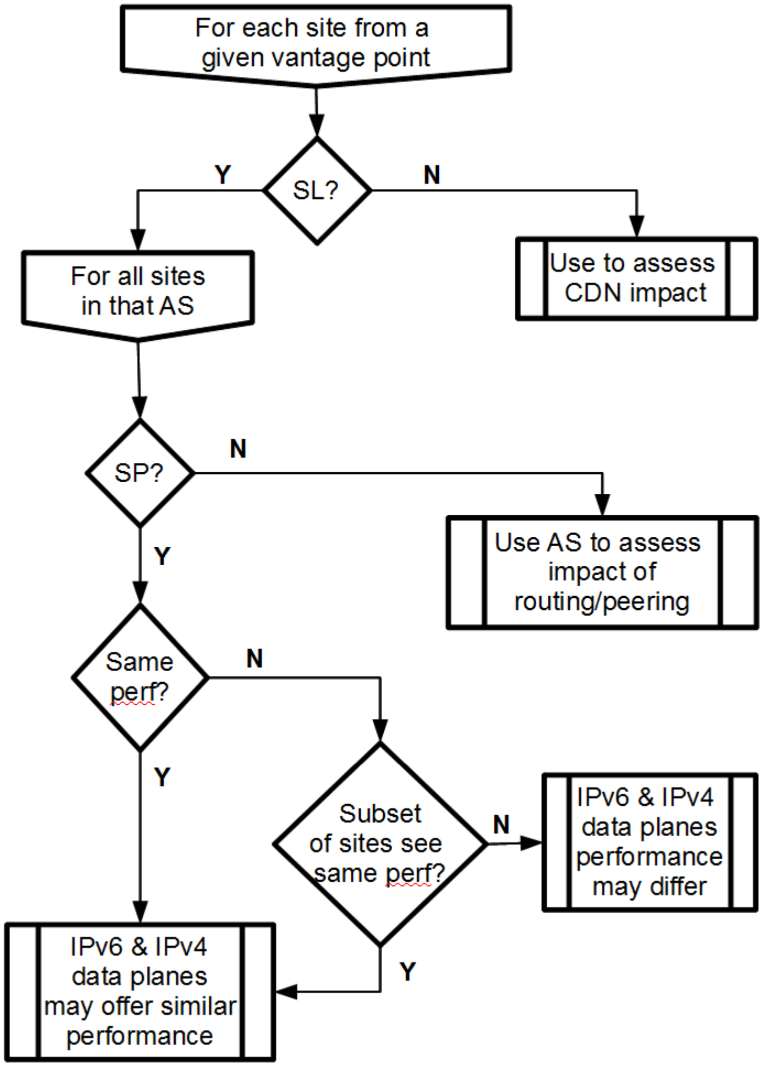
- **(D)** does not appear to be an issue (anymore)
- **(C)** is the main cause behind performance differences

General Methodology

- Given our focus on the "network", the goal is to eliminate (E) and (S) to the extent possible, and then identify when either (C) or (D) are responsible for performance differences
  - The (monitoring) client s/w runs on machines we control, so that (E) can be altogether eliminated
  - We don't have much visibility into (web) servers and access networks, so that ruling (S) out calls for mostly indirect methods

- The general approach we use relies on classifying sites as a function of differences in IPv6 and IPv4 ***"locations"*** and ***"paths"***
  - Same location ≡ Same destination AS
  - Same path ≡ Same AS_PATH

Classifying Sites' Destination ASes



- DL ≡ Different Location(s)
- SL ≡ Same location
  - SP ≡ Same AS Path
  - DP ≡ Different AS Path

# sites	Comcast	LU	Penn	UPCB
DL	450	352	784	485
SP	1,113	2,291	424	2,597
DP	1,962	1,263	6,786	1,336
IPv6 ≈ IPv4	82.8%	82.2%	41%	84.8%

IPv6 ≈ IPv4: IPv6 performance is within 10% confidence interval of IPv4 performance, or IPv6 outperforms IPv4

Identical IPv6 and IPv4 AS Paths

	Comcast	LU	Penn	UPCB
IPv6 ≈ IPv4	80.7%	70.2%	81.3%	79.8%
Zero mode	6%	10.8%	9.4%	7.3%
Small # sites	13.3%	19%	9.3%	12.9%
# ASes	233	248	75	124
Cross-check ☑	129	164	47	82
Cross-check ☒	0	0	0	0

Positive (negative) cross-checks for ASes in the same “category” from different vantage points

Hop-Count Level Comparison (Same IPv6 and IPv4 AS Paths)

		1 hop	# sites	2 hop	# sites	3 hop	# sites	4 hop	# sites	≥ 5 hops	# sites
Comcast	IPv4	64.2	137	41.6	632	36.0	304	36.8	10	-	0
	IPv6	59.9		42.1		35.4		34.0		-	
LU	IPv4	50.3	229	62.5	1829	42.7	115	21.3	16	-	0
	IPv6	57.3		62.2		39.2		19.4		-	
Penn	IPv4	-	0	-	0	36.0	23	29.5	203	29.1	169
	IPv6	-		-		34.4		27.6		29.5	
UPCB	IPv4	-	0	43.7	168	62.8	2,202	50.3	38	13.4	1
	IPv6	-		41.4		64.7		47.6		13.7	

Download speeds in kbytes/sec

World IPv6 Day Validation (Same IPv6 and IPv4 AS Path)

	LU	Penn	UPCB
IPv6 ≈ IPv4	85.7%	92.3%	72.2%
Other	14.3%	7.7%	27.8%
#ASes	42	13	36
Cross-check ☑	17	8	13

Conclusions From Same AS_PATH Comparisons

- When IPv6 and IPv4 web access requests are forwarded along the “same” path, they see mostly comparable network performance

$\Rightarrow$ The IPv6 and IPv4 data planes perform mostly similarly

- Next step focuses on sites (ASes) reachable over different IPv6 and IPv4 AS paths

Different IPv6 and IPv4 AS Paths

IPv6 ≈ IPv4: IPv6 performance is within 10% confidence interval of IPv4 performance, or IPv6 outperforms IPv4

	Comcast	LU	Penn	UPCB
IPv6 ≈ IPv4	11%	10%	3%	8%
Zero mode	5%	3%	12%	6%
# ASes	233	248	75	124

- World IPv6 Day Results

	LU	Penn	UPCB
IPv6 ≈ IPv4 (DP)	48.9%	53.5%	51.0%
#ASes	92	114	102
IPv6 ≈ IPv4(SP)	85.7%	92.3%	72.2%

Recall SP figures → IPv6 ≈ IPv4(SP)

“Ruling Out” Bad AS Paths

- Could the poorer performance of IPv6 be caused by sub-par data plane IPv6 performance in some (transit) ASes?
  - Checking for “bad apples” (ASes that display higher correlation with bad IPv6 performance), did not reveal any such AS
  - Many (though not all) ASes in DP paths were found present in “good” SP paths

% good ASes	Comcast	LU	Penn	UPCB
100%	11.1%	6.4%	3.2%	17.2%
[75%,100%]	20.8%	0.9%	20.8%	22.4%
[50%,75%]	45.8%	68.8%	58.8%	52.6%
[25%,50%]	27.8%	19.3%	15.8%	7.8%
[0%,25%]	6.9%	4.6%	1.4%	0%

Conclusions From Different AS_PATH Comparisons

- When IPv6 and IPv4 web access requests are forwarded along "different" paths, IPv6 often sees worse network performance
  - No "bad" ASes were identified as possible culprits
- Comparison of equal hop-count DP paths revealed similar IPv6 and IPv4 performance, at least for reasonable hop count values for which tunnels are less likely

$\Rightarrow$ Differences in performance can be reasonably attributed to differences in routing (peering) choices

Summary and Miscellaneous Findings

- **Observations and recommendations**
  1. The IPv6 data plane does not appear to be an issue any more
  2. The sparser IPv6 topology restricts IPv6 routing choices, which can in turn have a substantial impact on performance

  ⇒ **Ensuring peering parity between IPv6 and IPv4 is probably the most effective step to eliminate performance differences**

- The lack of commercial IPv6 CDN offering also has an impact
  - Across vantage points, IPv4 outperformed IPv6 over 90% of the time, when web requests were sent to ***different ASes*** (likely CDN instances)
  - Performance differences though were relatively small (around 15%), but this could change as the load of IPv6 requests increases

  ⇒ IPv6 CDN offerings could further improve IPv6 standing