Understanding the Impact of Network Infrastructure Changes using Large Scale Measurement Platforms

Vaibhav Bajpai
Jacobs University, Bremen

TUM Seminar, Raitenhaslach

Dissertation Committee

Prof. Dr. Jürgen Schönwälder
Jacobs University Bremen, Germany

Dr. Kinga Lipskoch
Jacobs University Bremen, Germany

Prof. Dr. Filip De Turck
University of Ghent, Belgium

Oct 11, 2016

Supported by:
Flamingo Project: flamingo-project.eu
Leone Project: leone-project.eu

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

This thesis would not have been possible without these amazing people!

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

Introduction | Motivation

- Large-scale Internet Measurement Platform

  An infrastructure of dedicated hardware probes that periodically runs Internet measurement tests to satisfy specific use-case requirements.

- Use-cases

  Since '98: Topology Mapping (Mature)

  CAIDA Ark [1], DIMES [2], iPlane [3] *et al.*

  Since '08: Network Performance

  SamKnows [4], BISmark [5], RIPE Atlas [6, 4], PerfSONAR [4] *et al.*

  - Measure performance and reliability of *broadband* access networks
  - Facilitate *regulators* to make better policy decisions.

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

Introduction | Motivation

- This dissertation *expands* the goal:

  Since '98: Topology Mapping

  Since '08: Network Performance

  - Measure performance and reliability of *broadband* access networks
  - Facilitate *regulators* to make better policy decisions.
  - Understand the impact of *network infrastructure changes*

---

1. Measuring IPv6 Performance
2. Measuring Access Network Performance

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

Introduction | Research Contributions

1. Survey on Internet Performance Measurement Platforms [COMST '15]

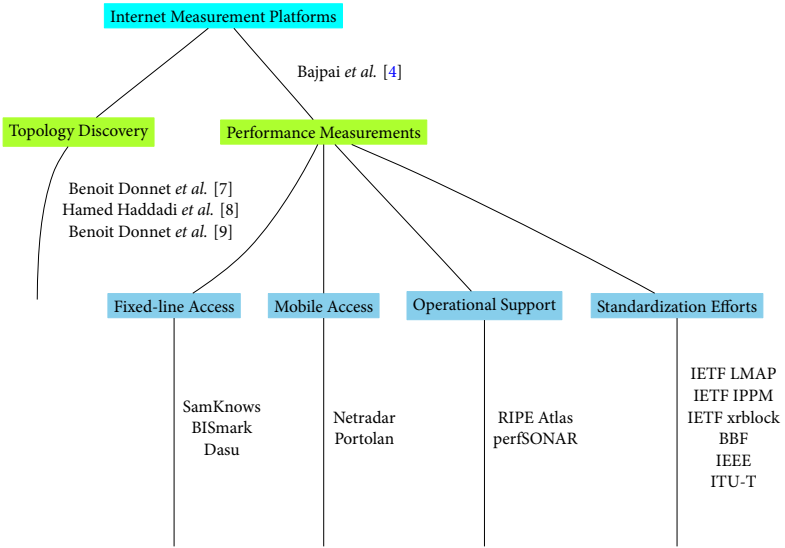
2. Measuring IPv6 Performance
   - Measuring TCP Connect Times [NETWORKING '15]
   - Measuring YouTube Performance [PAM '15]
   - Measuring Effects of Happy Eyeballs [ANRW '16]
   - Measuring Web Similarity [CNSM '16]

3. Measuring Access Network Performance
   - RIPE Atlas Vantage Point Selection [*]
   - Dissecting Last-mile Latency Characteristics [*]
   - Lessons Learned from using RIPE Atlas [CCR '15]

* entries are papers currently under review.

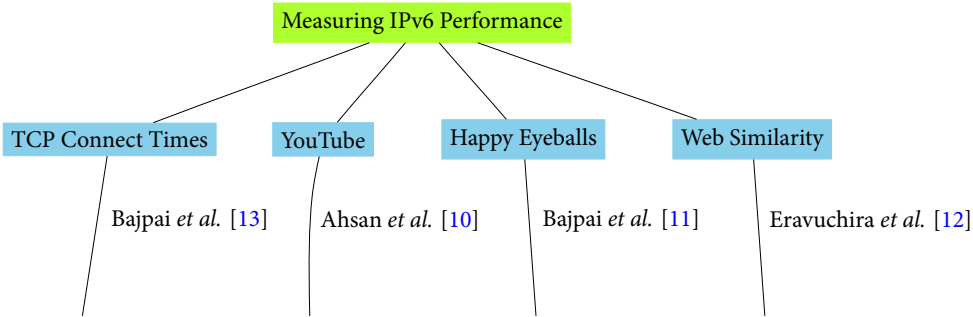
Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

Part I



Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

Part II | Overview



Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

Part II | Motivation

- Literature has *largely* focussed on measuring IPv6 adoption [14, 15, 16] ('10 –'14).
  - Addressing
  - Naming
  - Routing
  - Reachability
- Very little work [17] on measuring performance of service delivery over IPv6.
- Largely due to lack of available content over IPv6.
- A number of *significant* events occured during the span of this dissertation.
  - IANA IPv4 Address Exhaustion [18]
  - World IPv6 Day '11 [19]
  - World IPv6 Launch Day '12 [20]
  - RIR IPv4 Address Exhaustion [18]

APNIC	Apr'11
RIPE	Sep'12
LACNIC	Jun'14
ARIN	Sep'15

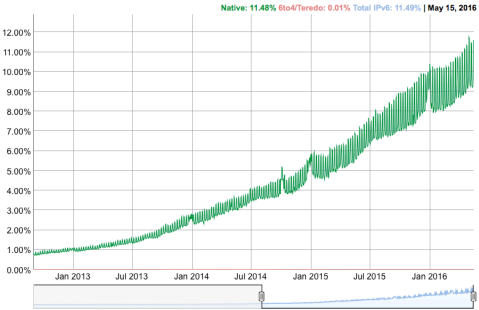
Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

Part II | Motivation

- Large IPv6 broadband rollouts[1] [13].
- Global IPv6 adoption [21].

09/2012 0.85%
05/2016 11.48%

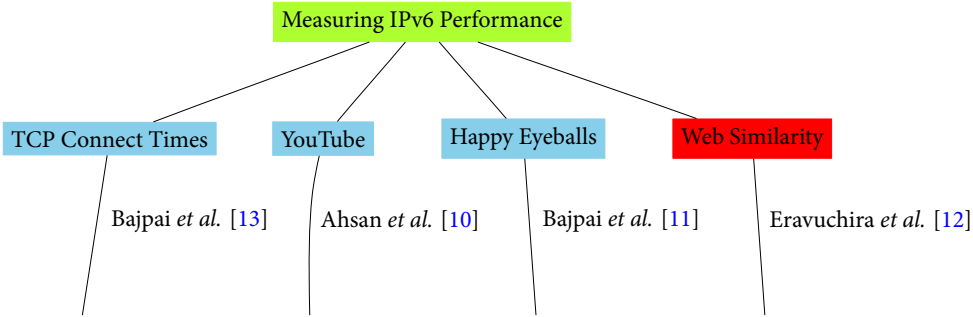
Belgium	40.49%
Switzerland	27.38%
United States	23.62%
Germany	21.41%



- This study *closes* the gap.
- It measures IPv6 performance of *operational* dual-stacked content delivery services.

[1] Comcast, Deutsche Telekom AG, AT&T, Verizon Wireless, T-Mobile USA

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A



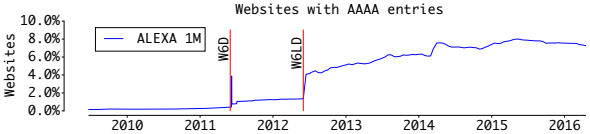
* entries are papers currently under review.

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

Web Similarity | Introduction

Recent work [13], [17], [15] has compared performance of dual-stacked websites over IPv4 and IPv6.

No study comparing web similarity over IPv4 / IPv6.



http://www.employees.org/~dwing/aaaa-stats

We want to know:

- *How similar are webpages accessed over IPv6 to their IPv4 counterparts?*
- *What factors contribute to the dissimilarity over IPv4 and IPv6?*

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

Web Similarity | Introduction

We measure against ALEXA top 100 dual-stacked websites.

1. `simweb` : A tool for measuring web similarity over IPv4 and IPv6.
2. Websites (27%) have some fraction of webpage elements failing over IPv6.
3. Failure rates over IPv6 are largely due to DNS resolution error on images, js and CSS.
4. Both same-origin and cross-origin sources contribute to the failure rates over IPv6.

To the best of our knowledge, this is the first study to:

- *Measure webpage similarity over IPv4 and IPv6.*
- *Investigate IPv6 adoption that goes beyond the root page of a dual-stacked website.*

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

Web Similarity | Methodology

We use 2 well-known webpage complexity metrics from literature [22, 23]:

1. Content Complexity
   The number & size of fetched webpage elements.

2. Service Complexity
   The number of same-origin & cross-origin sources.

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

Web Similarity | Methodology

- We use the ALEXA top 100 dual-stacked websites as measurement targets [13].

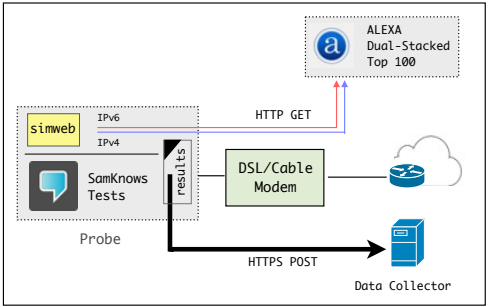
1. www.google.com
2. www.facebook.com
3. www.youtube.com
4. www.yahoo.com
5. www.wikipedia.org
6. www.qq.com
7. www.blogspot.com
8. ...

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

Web Similarity | Methodology

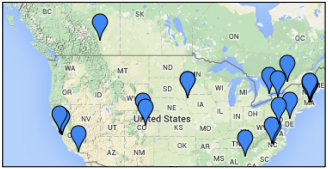
The `simweb` test:

- runs twice (once for each AF).
- repeats every hour.
- uses user-agent string: `Mozilla/4.0`



Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

Web Similarity | Methodology



NETWORK TYPE	#
RESIDENTIAL	55
NREN / RESEARCH	11
BUSINESS / DATACENTER	09
OPERATOR LAB	04
IXP	01

RIR	#
RIPE	42
ARIN	29
APNIC	07
AFRINIC	01
LACNIC	01

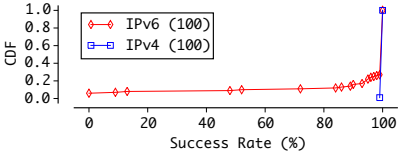
We measure from 80 dual-stacked SamKnows probes.

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

Web Similarity | Results

Can we fetch all webpage elements over IPv6?

- 27% of websites show some rate of failure over IPv6.
- 9% exhibit more than 50% failures over IPv6.
- 6% show complete failure (0% success) over IPv6.



#	Webpage	Success Rate (%) IPv6(↓)	Success Rate (%) IPv4	W6LD
01	www.bing.com	0	100	✓
02	www.detik.com	0	100	✓
03	www.engadget.com	0	100	✓
04	www.nifty.com	0	100	
05	www.qq.com	0	100	
06	www.sakura.ne.jp	0	100	
07	www.flipkart.com	09	99	✓
08	www.folha.uol.com.br	13	100	
09	www.aol.com	48	100	✓
10	www.comcast.net	52	100	✓
11	www.yahoo.com	72	100	✓
12	www.mozilla.org	84	100	✓
13	www.orange.fr	86	100	✓
14	www.seznam.cz	89	100	✓
15	www.mobile.de	90	100	✓
16	www.wikimedia.org	90	100	
17	www.t-online.de	93	100	✓
18	www.free.fr	95	100	
19	www.usps.com	95	100	
20	www.vk.com	95	100	✓
21	www.wikipedia.org	95	100	✓
22	www.wiktionary.org	95	100	
23	www.elmundo.es	96	100	✓
24	www.uol.com.br	96	100	✓
25	www.marca.com	97	100	✓
26	www.terra.com.br	98	100	✓
27	www.youm7.com	99	100	

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

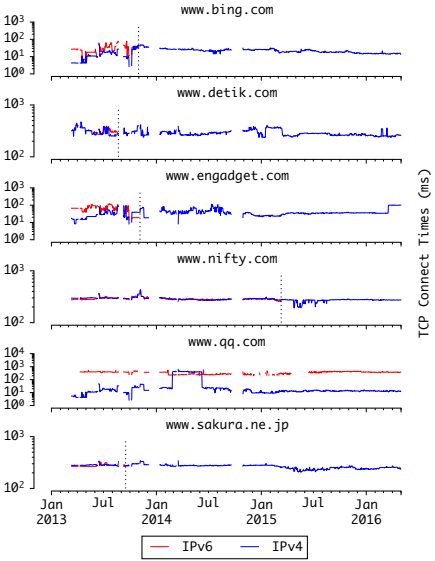
Web Similarity | Results

ALEXA top 100 dual-stacked websites:

- 6% show complete failure over IPv6.

#	Webpage	Success Rate (%) IPv6(↓)	Success Rate (%) IPv4	W6LD
01	www.bing.com	0	100	✓
02	www.detik.com	0	100	✓
03	www.engadget.com	0	100	✓
04	www.nifty.com	0	100	
05	www.qq.com	0	100	
06	www.sakura.ne.jp	0	100	

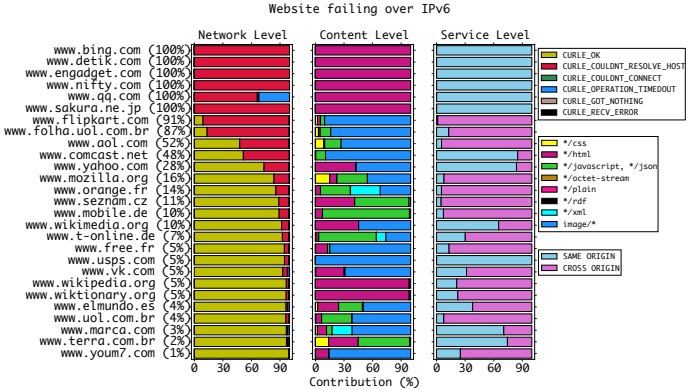
- Metrics that measure IPv6 adoption should account for *changes* in IPv6-readiness.



Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

Web Similarity | Results

Where in the network does the failure occur?

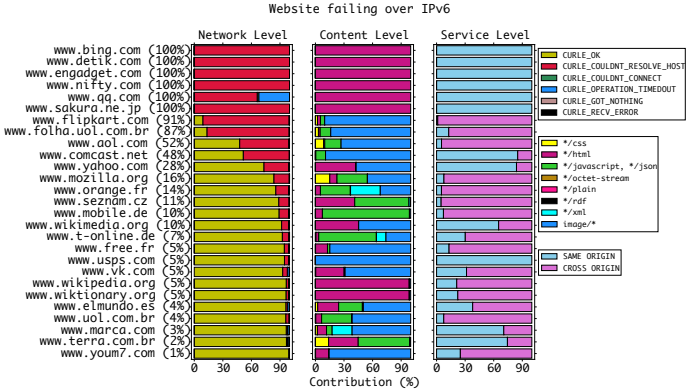


- `CURLE_COULDNT_RESOLVE_HOST` is the major contributor to failure rates.
- AAAA entries missing for these webpage elements in the DNS.

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

Web Similarity | Results

Which type of objects fail more than others?

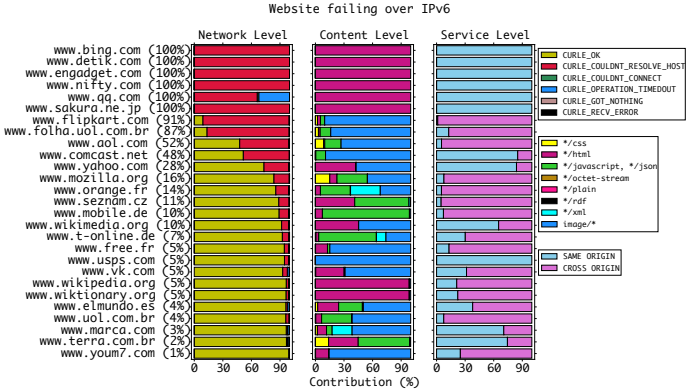


- `image/*`, `*/javascript`, `*/json` and `*/css` content contribute to the majority of the failure over IPv6.

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

Web Similarity | Results

Where do the failing objects originate from?

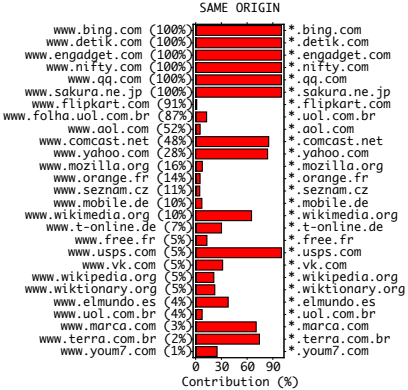


- Both same and cross origin sources contribute to the failure of webpage elements over IPv6.

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

Web Similarity | Results

What is failure contribution of same-origin sources?



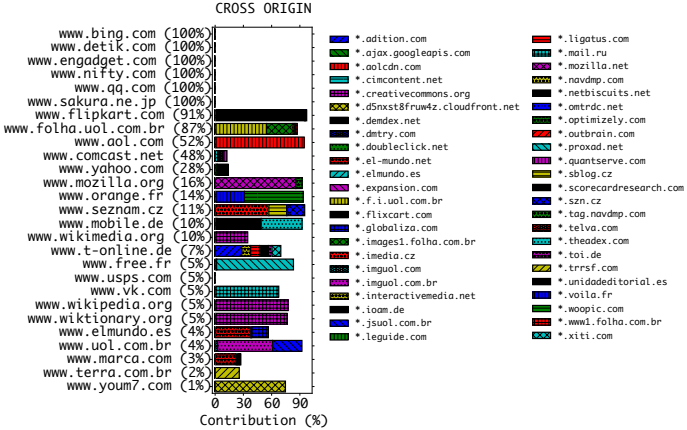
- 12% of websites have more than 50% webpage elements that belong to the same origin source and fail over IPv6.

#	Webpage	Same Origin (↓)
01	www.bing.com	100%
02	www.detik.com	100%
03	www.engadget.com	100%
04	www.nifty.com	100%
05	www.usps.com	100%
06	www.qq.com	100%
07	www.sakura.ne.jp	100%
08	www.comcast.net	85%
09	www.yahoo.com	83%
10	www.terra.com.br	74%
11	www.marca.com	70%
12	www.wikimedia.org	65%
13	www.elmundo.es	37%
14	www.vk.com	31%
15	www.t-online.de	30%
16	www.youm7.com	24%
17	www.wiktionary.org	22%
18	www.wikipedia.org	22%
19	www.free.fr	13%
20	www.folha.uol.com.br	12%
21	www.mozilla.org	7%
22	www.uol.com.br	7%
23	www.mobile.de	7%
24	www.aol.com	5%
25	www.orange.fr	5%
26	www.seznam.cz	4%
27	www.flipkart.com	1%

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

Web Similarity | Results

What is failure contribution of cross-origin sources?

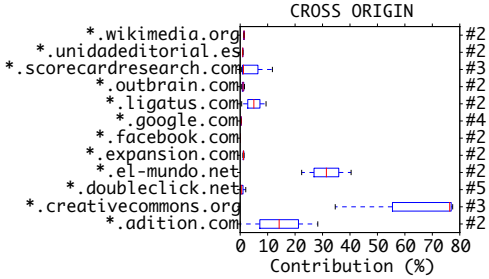


- Some of the cross-origin sources contribute to the failure of multiple websites.

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

Web Similarity | Results

Which cross-origin sources span across multiple failing websites?



CROSS ORIGIN	MEDIAN
*.creativecommons.org	76.33%
*.el-mundo.net	31.41%
*.adition.com	14.20%
*.ligatus.com	4.98%
*.wikimedia.org	1.40%
*.expansion.com	1.21%
*.scorecardresearch.com	1.19%
*.outbrain.com	1.06%
*.unidadeditorial.es	0.94%
*.doubleclick.net	0.54%
*.google.com	0.31%
*.facebook.com	0.06%

- doubleclick.net spans 5 websites with a 0.54% median contribution to failure rates.
- creativecommons.org has 76% median contribution to the failure rate of 3 websites.

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

Web Similarity | Takeaway

- Metrics that measure IPv6 adoption should account for changes in IPv6-readiness.
- Limiting to root webpage can lead to overestimation of IPv6 adoption numbers.
- Unclear whether websites with failure rates can be deemed IPv6-ready.
- Few cross-origin sources once IPv6 enabled will help large number of websites at once.

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

1. Survey on Internet Performance Measurement Platforms [COMST '15]

2. Measuring IPv6 Performance

   - Measuring TCP Connect Times [NETWORKING '15]
   - Measuring YouTube Performance [PAM '15]
   - Measuring Effects of Happy Eyeballs [ANRW '16]
   - Measuring Web Similarity [CNSM '16]

3. Measuring Access Network Performance

   - RIPE Atlas Vantage Point Selection [*]
   - Dissecting Last-mile Latency Characteristics [*]
   - Lessons Learned from using RIPE Atlas [CCR '15]

www.vaibhavbajpai.com

v.bajpai@jacobs-university.de | @bajpaivaibhav

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

References

[1] kc claffy, "The 7th Workshop on Active Internet Measurements (AIMS7) Report," *Computer Communication Review*, vol. 46, no. 1, pp. 50–57, 2016. [Online]. Available: http://doi.acm.org/10.1145/2875951.2875960

[2] Y. Shavitt and E. Shir, "DIMES: let the internet measure itself," *SIGCOMM Comput. Commun. Rev.*, vol. 35, no. 5, pp. 71–74, Oct. 2005. [Online]. Available: http://doi.acm.org/10.1145/1096536.1096546

[3] H. V. Madhyastha, T. Isdal, M. Piatek, C. Dixon, T. Anderson, A. Krishnamurthy, and A. Venkataramani, "iPlane: An Information Plane for Distributed Services," in *Proceedings of the 7th Symposium on Operating Systems Design and Implementation*, ser. OSDI '06. Berkeley, CA, USA: USENIX Association, 2006, pp. 367–380. [Online]. Available: http://dl.acm.org/citation.cfm?id=1298455.1298490

[4] V. Bajpai and J. Schönwälder, "A survey on internet performance measurement platforms and related standardization efforts," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 3, pp. 1313–1341, 2015. [Online]. Available: http://dx.doi.org/10.1109/COMST.2015.2418435

[5] S. Sundaresan, S. Burnett, N. Feamster, and W. De Donato, "BISmark: A Testbed for Deploying Measurements and Applications in Broadband Access Networks," in *Proceedings of the 2014 USENIX Conference on USENIX Annual Technical Conference*, ser. USENIX ATC'14. Berkeley, CA, USA: USENIX Association, 2014, pp. 383–394. [Online]. Available: http://dl.acm.org/citation.cfm?id=2643634.2643673

[6] RIPE NCC Staff, "RIPE Atlas: A Global Internet Measurement Network," *Internet Protocol Journal*, Sep. 2015. [Online]. Available: http://ipj.dreamhosters.com/wp-content/uploads/2015/10/ipj18.3.pdf

[7] B. Donnet, "Internet topology discovery," in *Data Traffic Monitoring and Analysis*, ser. Lecture Notes in Computer Science, E. Biersack, C. Callegari, and M. Matijasevic, Eds. Springer Berlin Heidelberg, 2013, vol. 7754, pp. 44–81. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-36784-7_3

[8] H. Haddadi, M. Rio, G. Iannaccone, A. Moore, and R. Mortier, "Network topologies: Inference, modeling, and generation," *Commun. Surveys Tuts.*, vol. 10, no. 2, pp. 48–69, Apr. 2008. [Online]. Available: http://dx.doi.org/10.1109/COMST.2008.4564479

[9] B. Donnet and T. Friedman, "Internet topology discovery: A survey," *Commun. Surveys Tuts.*, vol. 9, no. 4, pp. 56–69, Oct. 2007. [Online]. Available: http://dx.doi.org/10.1109/COMST.2007.4444750

[10] S. Ahsan, V. Bajpai, J. Ott, and J. Schönwälder, "Measuring YouTube from Dual-Stacked Hosts," in *Passive and Active Measurement - 16th International Conference, PAM 2015, New York, NY, USA, March 19-20, 2015, Proceedings*, 2015, pp. 249–261. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-15509-8_19

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A

[11] V. Bajpai and J. Schönwälder, "Measuring the effects of happy eyeballs," in *Proceedings of the 2016 Applied Networking Research Workshop*, ser. ANRW '16. New York, NY, USA: ACM, 2016, pp. 38–44. [Online]. Available: http://doi.acm.org/10.1145/2959424.2959429

[12] S. J. Eravuchira, V. Bajpai, J. Schönwälder, and S. Crawford, "Measuring web similarity from dual-stacked hosts," in *12th International Conference on Network and Service Management, CNSM 2016*, 2016.

[13] V. Bajpai and J. Schönwälder, "IPv4 versus IPv6 - who connects faster?" in *Proceedings of the 14th IFIP Networking Conference, Networking 2015, Toulouse, France, 20-22 May, 2015*, 2015, pp. 1–9. [Online]. Available: http://dx.doi.org/10.1109/IFIPNetworking.2015.7145323

[14] L. Colitti, S. H. Gunderson, E. Kline, and T. Refice, "Evaluating IPv6 Adoption in the Internet," ser. PAM '10, 2010, pp. 141–150. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-12334-4_15

[15] A. Dhamdhere, M. Luckie, B. Huffaker, k. claffy, A. Elmokashfi, and E. Aben, "Measuring the Deployment of IPv6: Topology, Routing and Performance," in *Proceedings of the 2012 ACM Conference on Internet Measurement Conference*, ser. IMC '12. New York, NY, USA: ACM, 2012, pp. 537–550. [Online]. Available: http://doi.acm.org/10.1145/2398776.2398832

[16] J. Czyz, M. Allman, J. Zhang, S. Iekel-Johnson, E. Osterweil, and M. Bailey, "Measuring IPv6 adoption," ser. ACM SIGCOMM '14, pp. 87–98. [Online]. Available: http://doi.acm.org/10.1145/2619239.2626295

[17] M. Nikkhah, R. Guérin, Y. Lee, and R. Woundy, "Assessing IPv6 Through Web Access a Measurement Study and Its Findings," in *Proceedings of the Seventh COnference on Emerging Networking EXperiments and Technologies*, ser. CoNEXT '11. New York, NY, USA: ACM, 2011, pp. 26:1–26:12. [Online]. Available: http://doi.acm.org/10.1145/2079296.2079322

[18] P. Richter, M. Allman, R. Bush, and V. Paxson, "A Primer on IPv4 Scarcity," *Computer Communication Review*, vol. 45, no. 2, pp. 21–31, 2015. [Online]. Available: http://doi.acm.org/10.1145/2766330.2766335

[19] Internet Society, "World IPv6 Day 2011," http://worldipv6day.org, [Online; accessed 25-January-2016].

[20] The Internet Society, "World IPv6 Launch," http://www.worldipv6launch.org, [Online; accessed 11-January-2016].

[21] "Google IPv6 Adoption Statistics," http://www.google.com/intl/en/ipv6/statistics.html, [Online; accessed 11-January-2016].

[22] M. Butkiewicz, H. V. Madhyastha, and V. Sekar, "Understanding Website Complexity: Measurements, Metrics, and Implications," in *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, ser. IMC '11. New York, NY, USA: ACM, 2011, pp. 313–328. [Online]. Available: http://doi.acm.org/10.1145/2068816.2068846

[23] ——, "Characterizing Web Page Complexity and Its Impact," *IEEE/ACM Trans. Netw.*, vol. 22, no. 3, pp. 943–956, Jun. 2014. [Online]. Available: http://dx.doi.org/10.1109/TNET.2013.2269999

Introduction
Motivation
Research Contributions
Part I
Part II
Web Similarity
Web Similarity
Introduction
Methodology
Results
Takeaway
Q/A