

# Internet Measurement

Summer Semester 2019

Prof. Georgios Smaragdakis, Ph.D.

# Instructor

Prof. Georgios Smaragdakis, Ph.D.

[georgios@inet.tu-berlin.de](mailto:georgios@inet.tu-berlin.de)

<http://www.smaragdakis.net>

Office: MAR 4.030

(please send me an email to arrange an appointment)

# Requirements

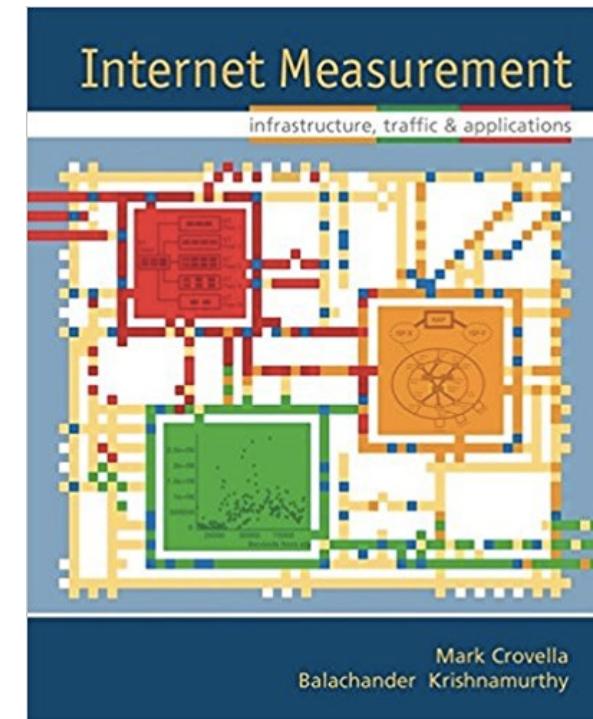
- BSc / MSc level
- Good understanding of basic Networking principles, e.g., Network Protocols and Architectures (prerequisite), Internet Control Plane.

# Mission

- To understand how the Internet **really** works and how it can be hacked, protected, and improved using **real** network measurements, i.e., data you collect or analyze.

# Readings

- All the course material will be available in ISIS
  - Papers from the Internet Measurement literature
  - “Internet Measurement:  
infrastructure, traffic, and  
applications”  
M. Crovella and  
B. Krishnamurthy



# Lecture structure

- Lectures (please participate!)
- Practical exercises in class  
(how to use tools and analyze data)
- 2-3 invited lectures by experts in Internet Measurements
- First invited lecture:
  - **Walter Willinger**, Niksun Inc.,  
May 16, 2pm – 4pm

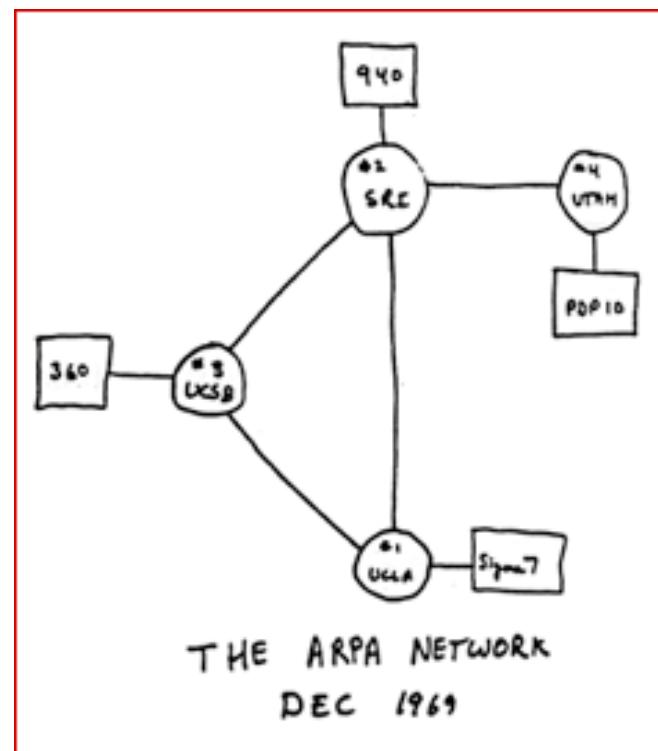
# Exam

- Exam after the end of the semester  
(oral exam for the Small/Large module)
- If you are a visiting student, e.g., Erasmus student, please contact me!

# Why Measure the Internet?

# Why Measure the Internet?

- The Internet is not any more a small academic network!



# The (Academic) Internet Growth

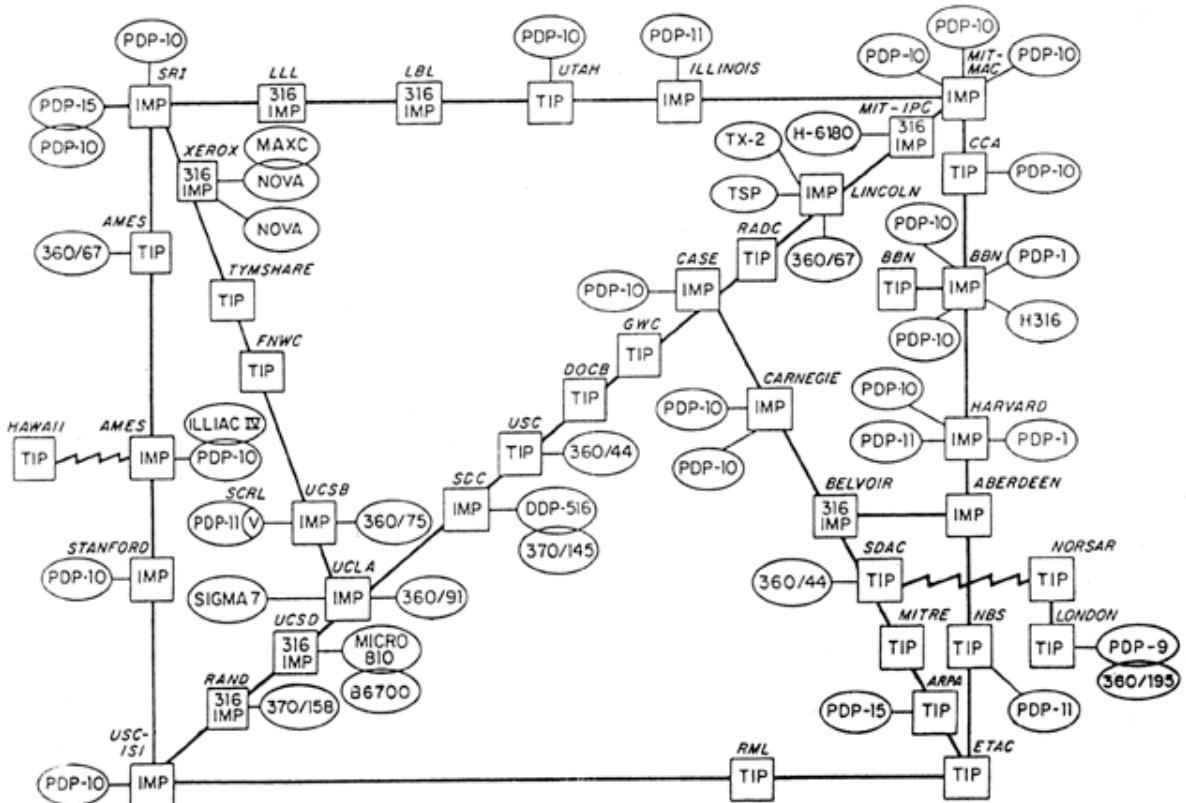
- Basic Principles  
(end-to-end)

## END-TO-END ARGUMENTS IN SYSTEM DESIGN

J.H. Saltzer, D.P. Reed and D.D. Clark\*  
M.I.T. Laboratory for Computer Science

- Inter-networking  
(OSI, initial protocols)
- First nodes outside  
Continental US
- First scalability  
and operational issues  
(congestion)

ARPA NETWORK, LOGICAL MAP, SEPTEMBER 1973

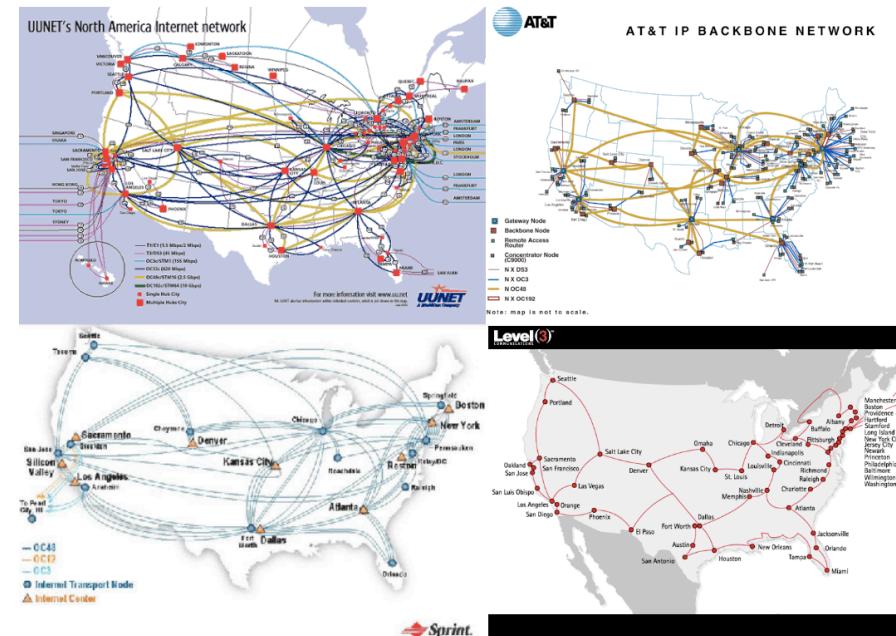


# Commercialization of the Internet 1980s-1990s

- Congestion control (TCP/IP)
- Security issues (Morris worm)
- Network management and economics

<https://www.vox.com/a/internet-maps>

- Researchers start measuring the Internet



End-to-End Internet Packet Dynamics

Vern Paxson  
Network Research Group  
Lawrence Berkeley National Laboratory\*  
University of California, Berkeley  
vern@ee.lbl.gov

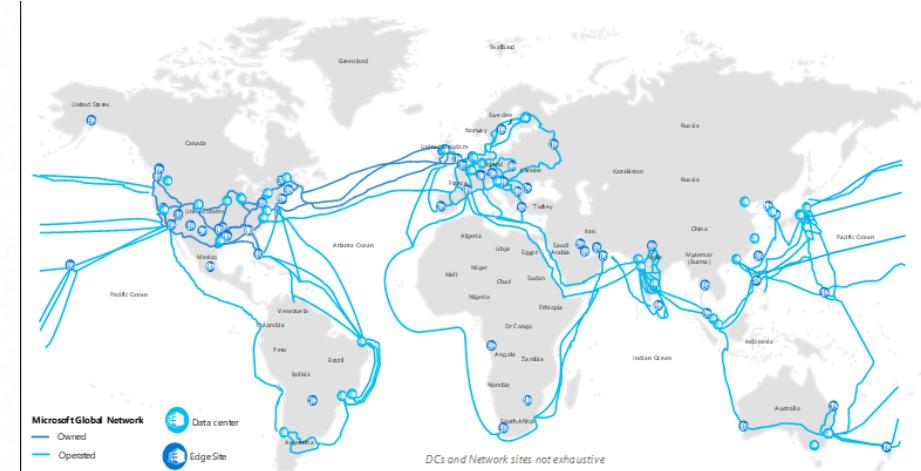
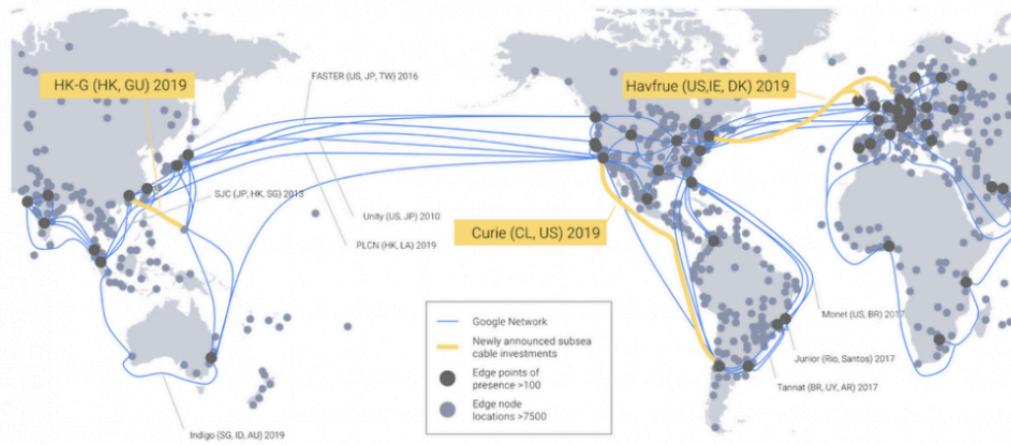
LBNL-40488

June 23, 1997

# The Internet is the Substrate of the Web

## Google Network

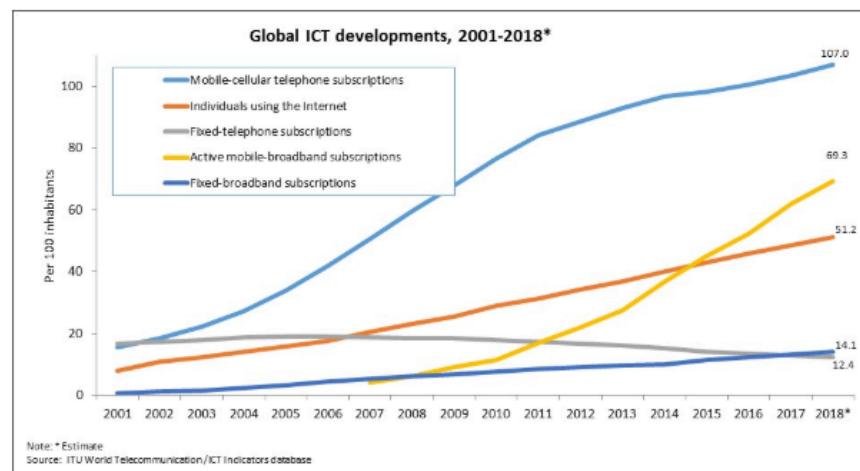
The largest cloud network, comprised of more than 100 points of presence



**Network Deployment:** Akamai has deployed the most pervasive, highly-distributed content delivery network (CDN) with approximately 239,000 servers in 139 countries and nearly 1,600 networks around the world.

# Why Measure the Internet?

- Internet and increasingly Mobile Internet is becoming the **de-facto means** for communication, commerce, information, and entertainment
- How many servers/IoTs/Smartphones are connected to the Internet?



## A Multi-perspective Analysis of Carrier-Grade NAT Deployment

Philipp Richter<sup>1</sup>, Florian Wohlfart<sup>2</sup>, Narseo Vallina-Rodriguez<sup>3</sup>,

Mark Allman<sup>3</sup>, Randy Bush<sup>5</sup>, Anja Feldmann<sup>1</sup>, Christian Kreibich<sup>3,6</sup>,

Nicholas Weaver<sup>3</sup>, Vern Paxson<sup>3,4</sup>

<sup>1</sup>TU Berlin, <sup>2</sup>TU München, <sup>3</sup>ICSI, <sup>4</sup>UC Berkeley, <sup>5</sup>Internet Initiative Japan, <sup>6</sup>Lastline

## ZMap: Fast Internet-Wide Scanning and its Security Applications

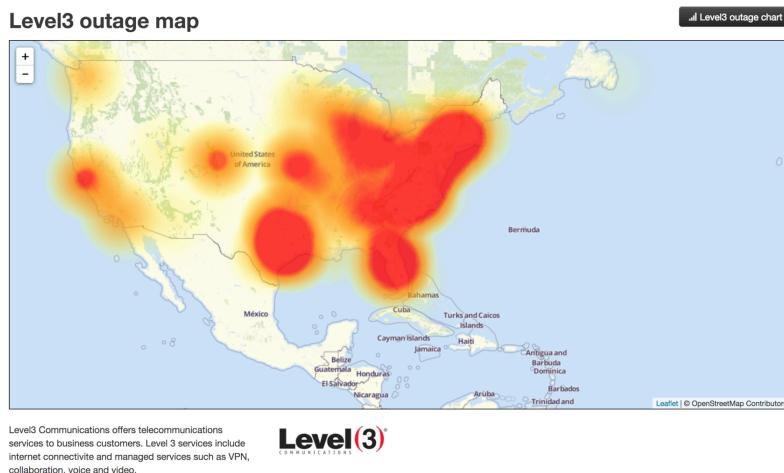
Zakir Durumeric  
University of Michigan  
zakir@umich.edu

Eric Wustrow  
University of Michigan  
ewust@umich.edu

J. Alex Halderman  
University of Michigan  
jhalderm@umich.edu

# Why Measure the Internet?

- The Internet is a “critical infrastructure”. Failures, attacks, and outages in the Internet will effect many other activities.



## Detecting Peering Infrastructure Outages in the Wild

Vasileios Gotsas  
CAIDA/TU Berlin  
vgotsas@ucsd.edu

Anja Feldmann  
TU Berlin  
anja@inet.tu-berlin.de

Christoph Dietzel  
TU Berlin/DE-CIX  
christoph@inet.tu-berlin.de

Arthur Berger  
MIT/Akamai  
awberger@csail.mit.edu

Georgios Smaragdakis  
MIT/TU Berlin  
gsmaragd@csail.mit.edu

Emile Aben  
RIPE NCC  
emile.aben@ripe.net

## Mirai botnet: Three admit creating and running attack tool

13 December 2017

Share



Web-connected security cameras were among the many devices hijacked by botnet

A US-based man has pleaded guilty to creating a giant botnet that was used to disrupt access to much of the web in October 2016.

## Understanding the Mirai Botnet

Manos Antonakakis<sup>◦</sup> Tim April<sup>‡</sup> Michael Bailey<sup>†</sup> Matthew Bernhard<sup>¤</sup> Elie Bursztein<sup>◦</sup>  
Jaime Cochran<sup>¤</sup> Zakir Durumeric<sup>¤</sup> J. Alex Halderman<sup>¤</sup> Luca Invernizzi<sup>◦</sup>  
Michalis Kallitsis<sup>§</sup> Deepak Kumar<sup>†</sup> Chaz Lever<sup>◦</sup> Zane Ma<sup>†\*</sup> Joshua Mason<sup>†</sup>  
Damian Menscher<sup>◦</sup> Chad Seaman<sup>‡</sup> Nick Sullivan<sup>¤</sup> Kurt Thomas<sup>◦</sup> Yi Zhou<sup>†</sup>

<sup>†</sup>Akamai Technologies <sup>¤</sup>Cloudflare <sup>◦</sup>Georgia Institute of Technology <sup>¤</sup>Google  
<sup>§</sup>Merit Network <sup>‡</sup>University of Illinois Urbana-Champaign <sup>¤</sup>University of Michigan

# Why Measure the Internet?

- There is a **big gap** between the design and implementation/adoption of protocols!

The screenshot shows the IETF Datatracker interface for the QUIC (quic) working group. The top navigation bar includes links for About, Documents, Meetings, History, Photos, Email expansions, List archive, and Tools. Below the navigation, a table lists nine active Internet-Drafts. The columns include Document, Date, and Status. The table header indicates there are 9 hits.

Document	Date	Status
draft-ietf-quic-applicability-03 Applicability of the QUIC Transport Protocol	2018-10-22 15 pages	I-D Exists WG Document: Informational Jul 2019
draft-ietf-quic-http-19 Hypertext Transfer Protocol Version 3 (HTTP/3)	2019-03-11 53 pages	I-D Exists WG Document: Proposed Standard Jul 2019
draft-ietf-quic-invariants-03 Version-Independent Properties of QUIC	2018-10-09 9 pages <span style="background-color: orange;">Expires soon</span>	I-D Exists WG Consensus: Waiting for Write-Up: Proposed Standard
draft-ietf-quic-manageability-03 Manageability of the QUIC Transport Protocol	2018-10-22 19 pages	I-D Exists WG Document: Informational Jul 2019
draft-ietf-quic-qpack-07 QPACK: Header Compression for HTTP over QUIC	2019-03-11 37 pages	I-D Exists WG Document: Proposed Standard Jul 2019
draft-ietf-quic-recovery-19 QUIC Loss Detection and Congestion Control	2019-03-11 38 pages	I-D Exists WG Document: Proposed Standard Jul 2019
draft-ietf-quic-spin-exp-01 The QUIC Latency Spin Bit	2018-10-22 8 pages	I-D Exists WG Document
draft-ietf-quic-tls-19 Using TLS to Secure QUIC	2019-03-11 43 pages	I-D Exists WG Document: Proposed Standard Jul 2019

## A First Look at QUIC in the Wild

Jan Rüth<sup>1</sup>, Ingmar Poese<sup>2</sup>, Christoph Dietzel<sup>3</sup>, and Oliver Hohlfeld<sup>1</sup>

<sup>1</sup> RWTH Aachen University {rueth, hohlfeld}@comsys.rwth-aachen.de

<sup>2</sup> Benocs GmbH ipoese@benocs.com

<sup>3</sup> TU Berlin / DE-CIX christoph@inet.tu-berlin.de

## The QUIC Transport Protocol: Design and Internet-Scale Deployment

Adam Langley, Alistair Riddoch, Alyssa Wilk, Antonio Vicente, Charles Krasic, Dan Zhang, Fan Yang, Fedor Kouranov, Ian Swett, Janardhan Iyengar, Jeff Bailey, Jeremy Dorfman, Jim Roskind, Joanna Kulik, Patrik Westin, Raman Tenneti, Robbie Shade, Ryan Hamilton, Victor Vasiliev, Wan-Teh Chang, Zhongyi Shi \*

Google

quic-sigcomm@google.com

# Why Measure the Internet?

- Understanding how the Internet works can inform **policymaking** and **improve** the life and protect **privacy** of Billion of people.



## On the Origins of Memes by Means of Fringe Web Communities

Savvas Zannettou\*, Tristan Caulfield†, Jeremy Blackburn†, Emiliano De Cristofaro†,  
Michael Sirivianos\*, Gianluca Stringhini°, and Guillermo Suarez-Tangil†

\*Cyprus University of Technology, †University College London,  
°University of Alabama at Birmingham, °Boston University, †King's College London  
sa.zannettou@edu.cut.ac.cy, {t.caulfield,e.dechristofaro}@ucl.ac.uk, blackburn@uab.edu,  
michael.sirivianos@cut.ac.cy, gian@bu.edu, guillermo.suarez-tangil@kcl.ac.uk



## Tracing Cross Border Web Tracking

Costas Iordanou  
TU Berlin / UC3M

Ingmar Poese  
BENOCS

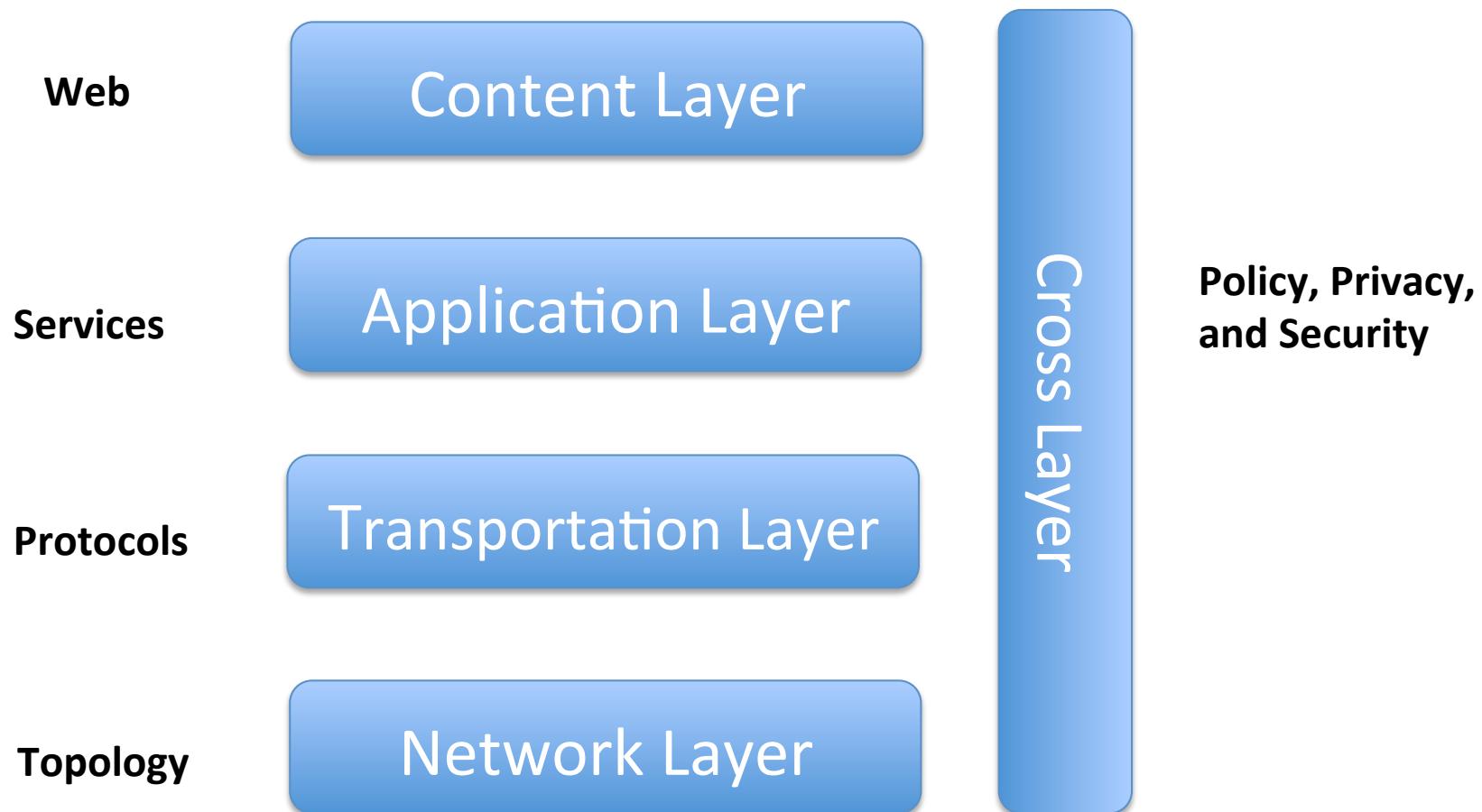
Georgios Smaragdakis  
TU Berlin

Nikolaos Laoutaris  
Data Transparency Lab / Eurecat

# Why Measure the Internet?

- Despite the importance of the Internet, our ability to answer even simple questions about the Internet is quite **limited!**
- No central authority
- Lack of “ground truth”
- Fast evolving Internet and Web!

# Measuring the Different Internet Layers



# Topics that we will cover in the Lectures:

- Internet Topology Discovery
  - Using passive measurements (BGP)
  - Using active measurements (Traceroute)
- Measuring the New Internet (ISPs, IXPs)
- Measuring the Web Infrastructure (CDNs, Cloud, DNS)
- Measuring Web Performance and Privacy
- Modeling of Network Traffic (Traffic flows)
- Assessing the Internet's Resilience and Security
- Topics: IP Geolocation, Network Attack Mitigation, etc.

# Internet Measurement Conference

IMC 2018   Home   Calls ▾   Shadow PC   Committees   Program   Posters   Awards   Local Info ▾   Supporters

ACM IMC 2018  
Oct 31 - Nov 2, 2018  
Boston, MA, USA

A photograph of the Boston skyline at sunset, viewed from across a river. In the foreground, there's a small island with sailboats docked along its shore. The background features several prominent skyscrapers, including the John Hancock Tower and the Prudential Center. The sky is clear with some light clouds.

IMC 2017   Home   Calls ▾   Shadow PC   Committees   Program   Posters   Awards   Local Info ▾   Supporters   Impressions

ACM IMC 2017  
November 1-3  
London, UK

A photograph of a modern urban landscape in London. It shows a mix of architectural styles, including a large brick building with a prominent dome and a modern glass and steel structure. The area appears to be a park or plaza with people walking around. The sky is overcast.

18 years conference <https://www.sigcomm.org/events/imc-conference>

# Ethical Consideration

- Be a good Internet citizen!
  - Do not try create harm
- If you find a vulnerability, report it!
- Be very careful with private data!!

# Example: IPv4 activity

- How many IPv4 addresses exist?

# Example: IPv4 activity

- How many IPv4 addresses exist?
  - from your Networking (NPA) course:  
32 bit addresses, i.e.,  
 $2^{32}$  (around 4.3 Billion addresses)

# Example: IPv4 activity

- How many IPv4 addresses exist?
  - from your Networking (NPA) course:  
 $2^{32}$  (around 4.3 Billion addresses)
  - How many of them have been allocated to networks?
  - Are there differences around the world?
  - Are there differences across networks?
  - How many of them are really used?
  - How about IPv6?

# Are there public data we can use?

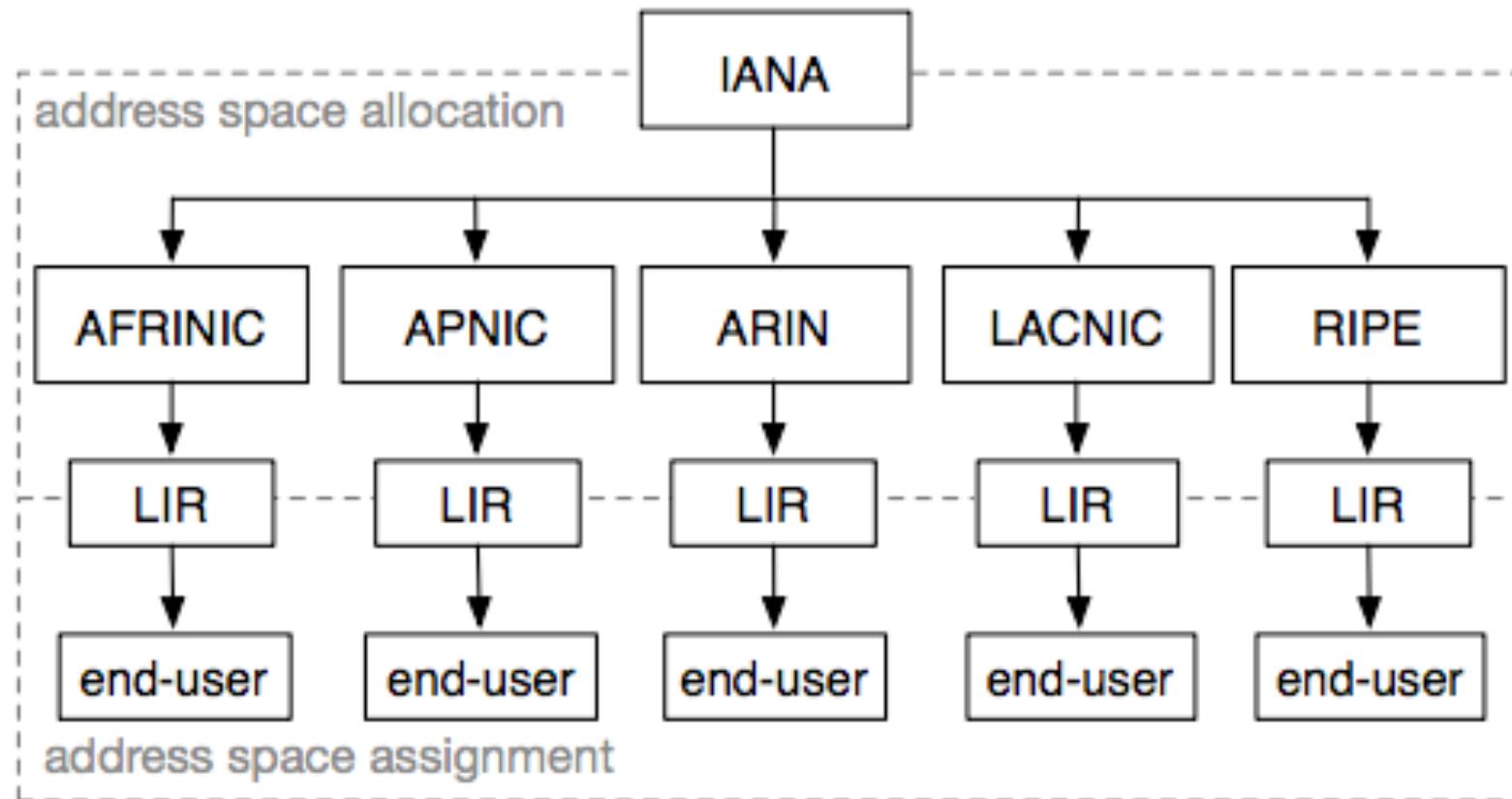
# Are there public data we can use? (Address Allocation)

## **Internet Registries:**

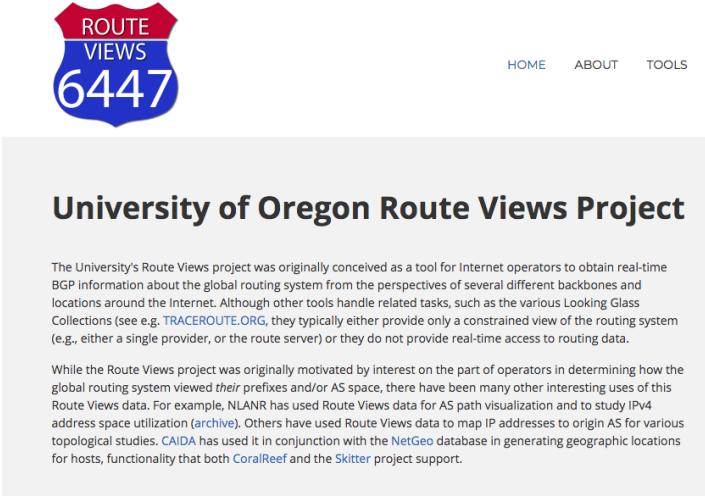
In recent years, address space that is allocated to networks is reported in databases (for accounting, to claim ownership of the address space, for filtering etc.).

Archives of Internet Registries are available.

# Internet Registries (Allocation of IP space)



# Are there public data to use? (Route Advertisements)

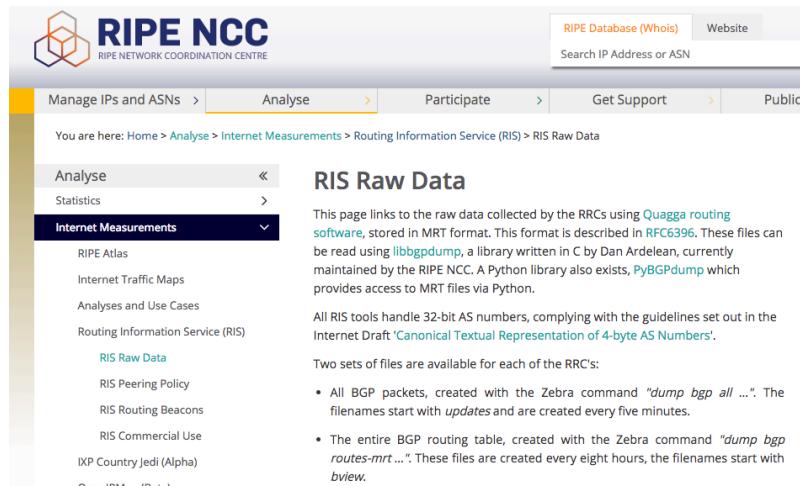


The screenshot shows the University of Oregon Route Views Project website. At the top is a red and blue logo with the text "ROUTE VIEWS 6447". Below it is a navigation bar with links to "HOME", "ABOUT", and "TOOLS". The main content area has a title "University of Oregon Route Views Project" and a paragraph of text describing the project's purpose and history. It also includes a sidebar with links to various resources like RIPE Atlas, Internet Traffic Maps, and Routing Information Service (RIS).

**University of Oregon Route Views Project**

The University's Route Views project was originally conceived as a tool for internet operators to obtain real-time BGP information about the global routing system from the perspectives of several different backbones and locations around the Internet. Although other tools handle related tasks, such as the various Looking Glass Collections (see e.g. [TRACEROUTE.ORG](#)), they typically either provide only a constrained view of the routing system (e.g., either a single provider, or the route server) or they do not provide real-time access to routing data.

While the Route Views project was originally motivated by interest on the part of operators in determining how the global routing system viewed *their* prefixes and/or AS space, there have been many other interesting uses of this Route Views data. For example, NLANR has used Route Views data for AS path visualization and to study IPv4 address space utilization ([archive](#)). Others have used Route Views data to map IP addresses to origin AS for various topological studies. CAIDA has used it in conjunction with the [NetGeo](#) database in generating geographic locations for hosts, functionality that both [CoralReef](#) and the [Skitter](#) project support.



The screenshot shows the RIPE NCC RIS Raw Data page. The header includes the RIPE NCC logo and links for "Manage IPs and ASNs", "Analyst", "Participate", "Get Support", and "Public". The main content area is titled "RIS Raw Data" and contains a detailed description of the raw data collected by the RRCs using Quagga routing software, stored in MRT format. It also mentions the libbgpdump library and PyBGPdump Python library. A sidebar on the left lists various RIS-related links: RIPE Atlas, Internet Traffic Maps, Analyses and Use Cases, Routing Information Service (RIS), RIS Raw Data (which is currently selected), RIS Peering Policy, RIS Routing Beacons, RIS Commercial Use, and IXP Country Jedi (Alpha).

**RIS Raw Data**

This page links to the raw data collected by the RRCs using Quagga routing [software](#), stored in MRT format. These files can be read using [libbgpdump](#), a library written in C by Dan Ardelean, currently maintained by the RIPE NCC. A Python library also exists, [PyBGPdump](#) which provides access to MRT files via Python.

All RIS tools handle 32-bit AS numbers, complying with the guidelines set out in the Internet Draft '[Canonical Textual Representation of 4-byte AS Numbers](#)'.

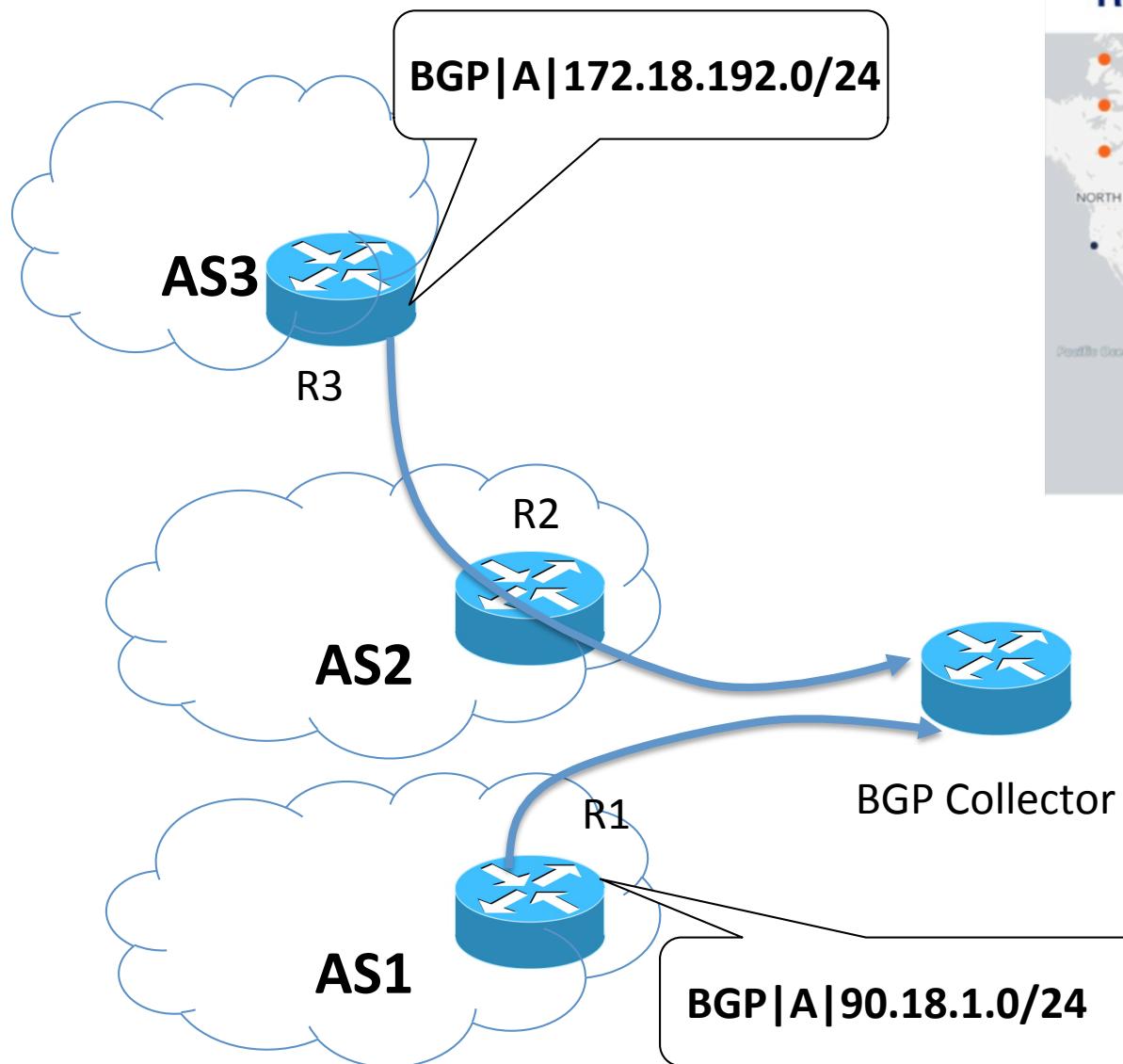
Two sets of files are available for each of the RRC's:

- All BGP packets, created with the Zebra command "`dump bgp all ...`". The filenames start with `updates` and are created every five minutes.
- The entire BGP routing table, created with the Zebra command "`dump bgp routes-mrt ...`". These files are created every eight hours, the filenames start with `bview`.

## Route Collectors:

Projects like RouteViews and RIPE RIS collect raw BGP data (announcements, withdrawals) at multiple locations around the globe that go back to 1999!

# How it works



## Routing Data (RIS)

- 18 BGP collectors and growing
- 600+ peers
- 150+ full-feed peers



...  
t1|R2|A|172.18.192.0/24|AS2 AS3  
t2|R3|A|90.18.1.0/24|AS1  
...

# Reading

“A Prime on IPv4 Scarcity”

P. Richter, M. Allman, R. Bush, and V. Paxson

Computer Communication Review, April 2015

(awarded best of CCR)

- Paper:

<https://dl.acm.org/citation.cfm?id=2766335>

- Presentation:

[https://people.csail.mit.edu/richterp/ipv4\\_primer\\_slides\\_sigcomm.pdf](https://people.csail.mit.edu/richterp/ipv4_primer_slides_sigcomm.pdf)