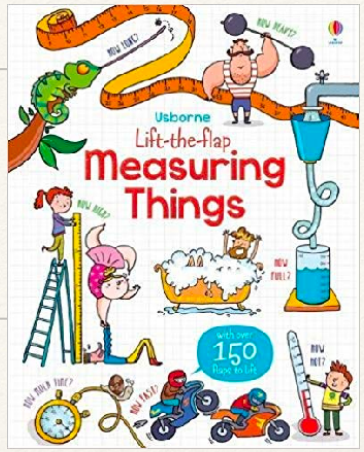
Internet Measurement Primer

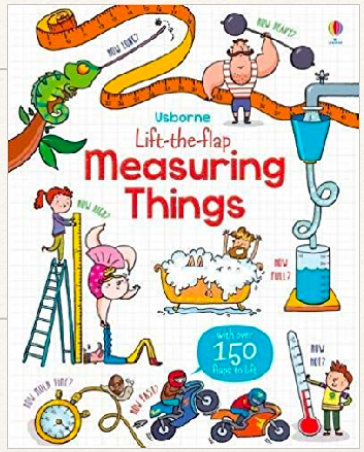
Robert Beverly
rbeverly@nps.edu
NSDI 2020
February 24, 2020

You get paid to *measure* the Internet?



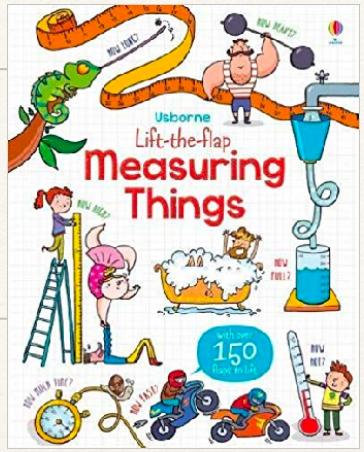
- The Internet is pervasive and crucial to society

You get paid to *measure* the Internet?



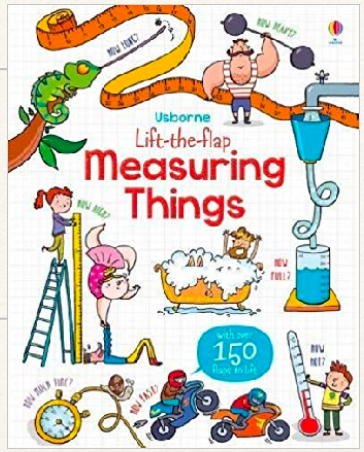
- The Internet is <u>pervasive</u> and <u>crucial</u> to society
- **But**:

You get paid to *measure* the Internet?



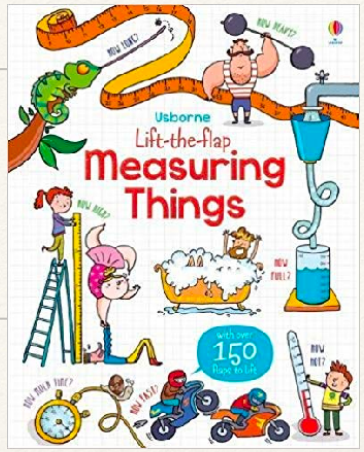
- The Internet is pervasive and crucial to society
- **But**:
  - Constantly evolving use, structure, and protocols

You get paid to *measure* the Internet?



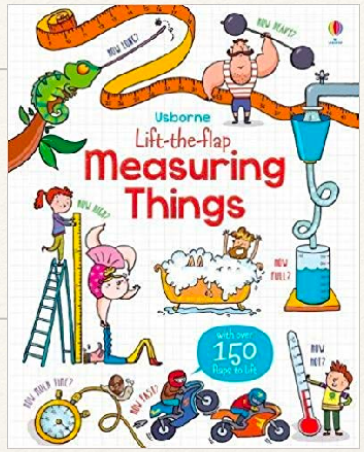
- The Internet is pervasive and crucial to society
- **But**:
  - Constantly evolving use, structure, and protocols
  - Distributed, multi-party, and economically driven

You get paid to *measure* the Internet?



- The Internet is pervasive and crucial to society
- **But**:
  - Constantly evolving use, structure, and protocols
  - Distributed, multi-party, and economically driven
  - Massive scale, w/ abstraction and info hiding

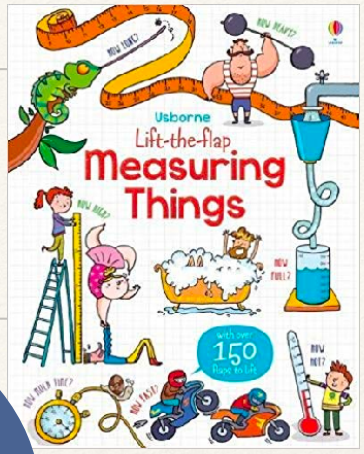
You get paid to *measure* the Internet?



How many hosts are there on the Internet?

- e and crucial to society
- D
  - Constantly evolving use, structure, and protocols
  - Distributed, multi-party, and economically driven
  - Massive scale, w/ abstraction and info hiding

You get paid to *measure* the Internet?

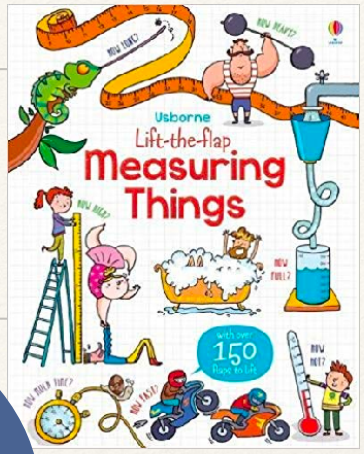


How many hosts are there on the Internet?

Where is there congestion in the network?

-
-
  - Constantly evolving use, structure, and protocols
  - Distributed, multi-party, and economically driven
  - Massive scale, w/ abstraction and info hiding

You get paid to *measure* the Internet?



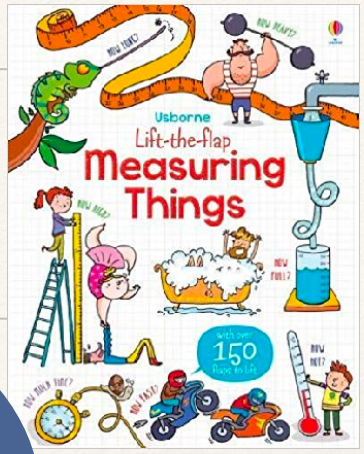
How many hosts are there on the Internet?

Where is there congestion in the network?

-
- B
  - Co ng use, structure, and protocols
  - , and economically driven
  - raction and info hiding

Is this packet legitimate, or spoofed?

You get paid to *measure* the Internet?



How many hosts are there on the Internet?

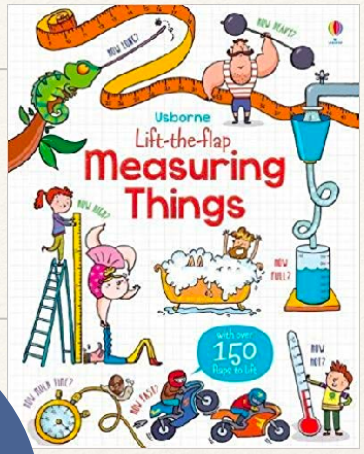
Where is there congestion in the network?

- C… use, st…
- …
- …tractio…

Is this packet legitimate, or spoofed?

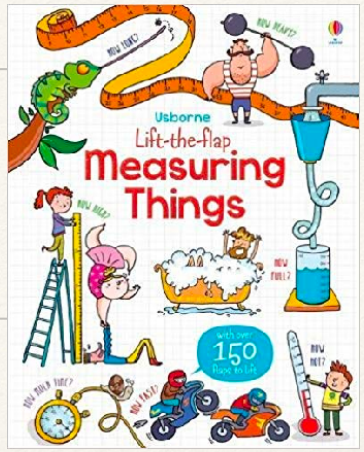
Why can't I reach a website?

You get paid to *measure* the Internet?



- How many hosts are there on the Inter
- Where is there congestion the network?
- What is the topology of the Internet?
- Is this packet legitimate, or spoofed?
- can't I reach a website?

You get paid to *measure* the Internet?



- The Internet is pervasive and crucial to society
- **But**:
  - Constantly evolving use, structure, and protocols
  - Distributed, multi-party, and economically driven
  - Massive scale, w/ abstraction and info hiding

Hard! Lots we don't understand!
Natural fit for experimental science

Whither Measurement?

1 METRE, Paris

Whither Measurement?

1 METRE, Paris

- Inform Internet evolution:
  - E.g., IPv6, DASH, QUIC, DNSSEC, IoT, and app du jour

Whither Measurement?

1 METRE, Paris

- Inform Internet evolution:
  - E.g., IPv6, DASH, QUIC, DNSSEC, IoT, and app du jour
- Make the Internet better:
  - E.g., Security, resilience, accountability, privacy

Whither Measurement?

1 METRE, Paris

- Inform Internet evolution:
  - E.g., IPv6, DASH, QUIC, DNSSEC, IoT, and app du jour
- Make the Internet better:
  - E.g., Security, resilience, accountability, privacy
- Inform policy:
  - E.g., access, freedom, neutrality

Whither Measurement?

1 METRE, Paris

- Inform Internet evolution:
  - E.g., IPv6, DASH, QUIC, DNSSEC, IoT, and app du jour
- Make the Internet better:
  - E.g., Security, resilience, accountability, privacy
- Inform policy:
  - E.g., access, freedom, neutrality
- Business + economics:
  - E.g., improve performance -> better user experience and/or more time for backend processing -> more revenue

(major) hurdles community faces today?

(major) hurdles community faces today?

- **Scale**:

  …there is much to measure

(major) hurdles community faces today?

- **Scale**:

  …there is much to measure

- **Access**:

  …and it's hard to get / not designed to be measured

(major) hurdles community faces today?

- **Scale**:

  …there is much to measure

- **Access**:

  …and it's hard to get / not designed to be measured

- **Ground-truth:**

  …and harder to validate inferences

(major) hurdles community faces today?

- **Scale**:

  …there is much to measure

- **Access**:

  …and it's hard to get / not designed to be measured

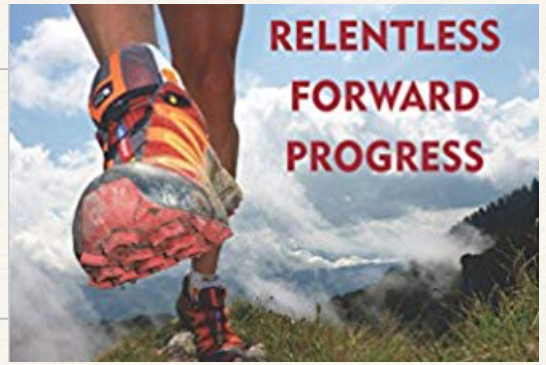
- **Ground-truth:**

  …and harder to validate inferences

- **Reproducibility:**

  …and best data and results are hoarded for privacy, policy, and self-serving reasons

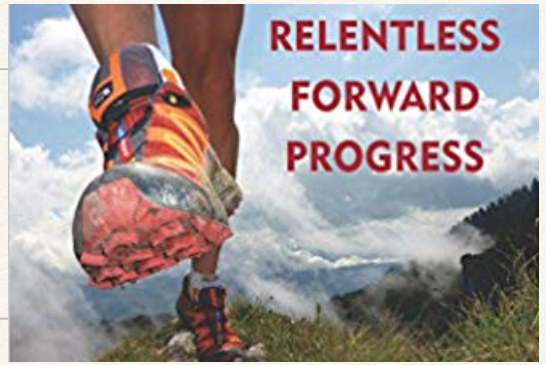
Relentless Forward Progress



Relentless Forward Progress

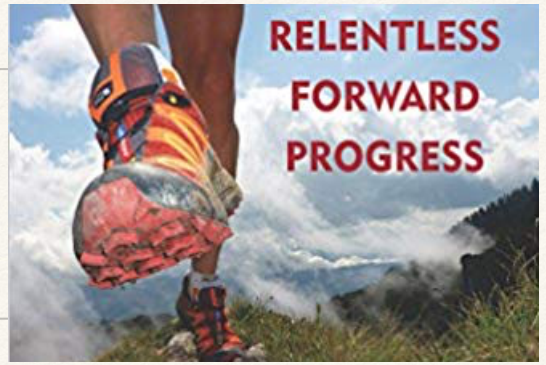
- Despite these hurdles, much success:

Relentless Forward Progress



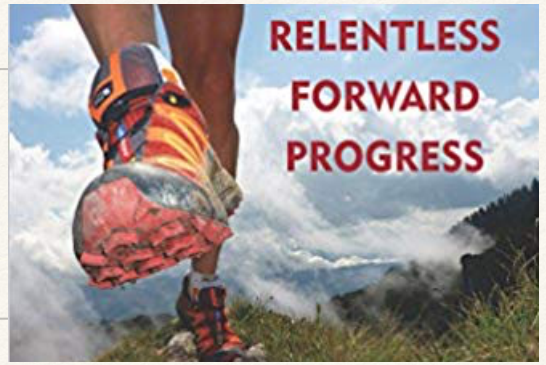
- Despite these hurdles, much success:
  - Better idea of the network topology than ever before

Relentless Forward Progress



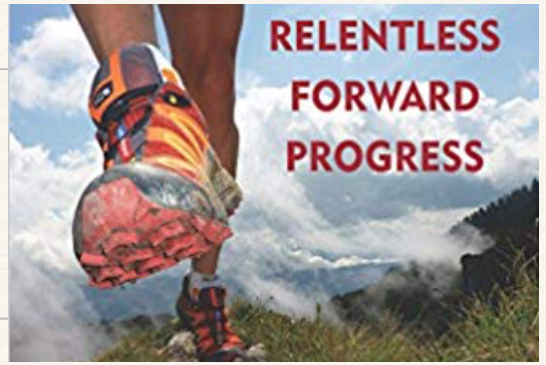
- Despite these hurdles, much success:
  - Better idea of the network topology than ever before
  - Measurements that drive protocol improvements

Relentless Forward Progress



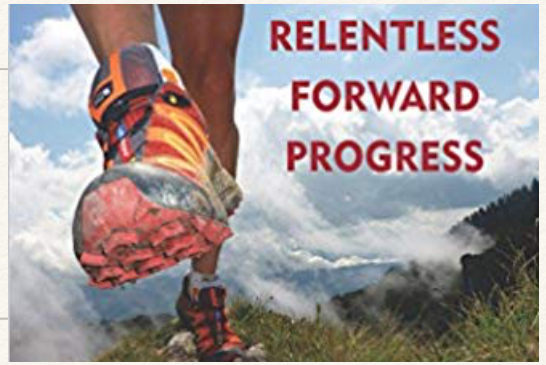
- Despite these hurdles, much success:
  - Better idea of the network topology than ever before
  - Measurements that drive protocol improvements
  - Measurements that drive security (DNS, routing, etc)

Relentless Forward Progress



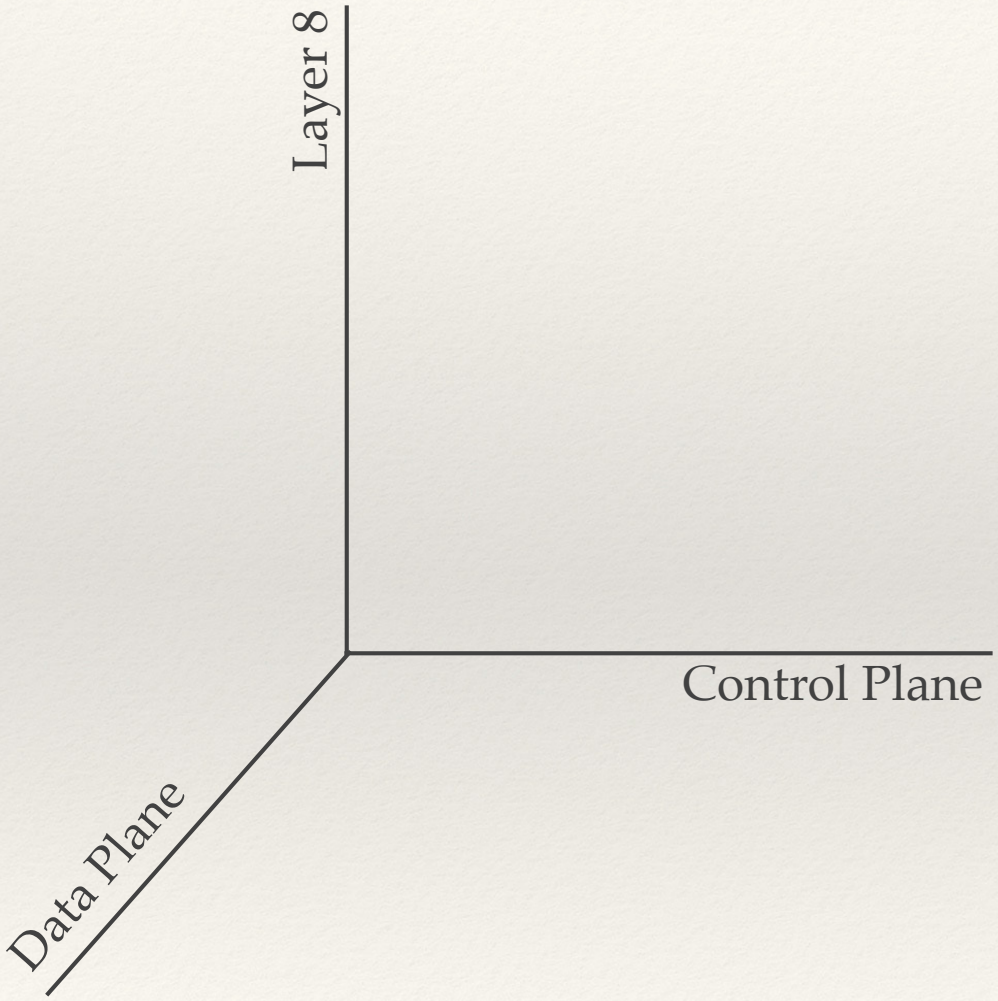
- Despite these hurdles, much success:
  - Better idea of the network topology than ever before
  - Measurements that drive protocol improvements
  - Measurements that drive security (DNS, routing, etc)
  - Variety of passive and active measurement platforms

Relentless Forward Progress



- Despite these hurdles, much success:
  - Better idea of the network topology than ever before
  - Measurements that drive protocol improvements
  - Measurements that drive security (DNS, routing, etc)
  - Variety of passive and active measurement platforms
  - Reproducibility and artifacts emphasis in ACM IMC, CCR

Techniques



- A variety of techniques
- Here's one simple taxonomy…
- And some examples

Techniques: Control Plane

Techniques: Control Plane

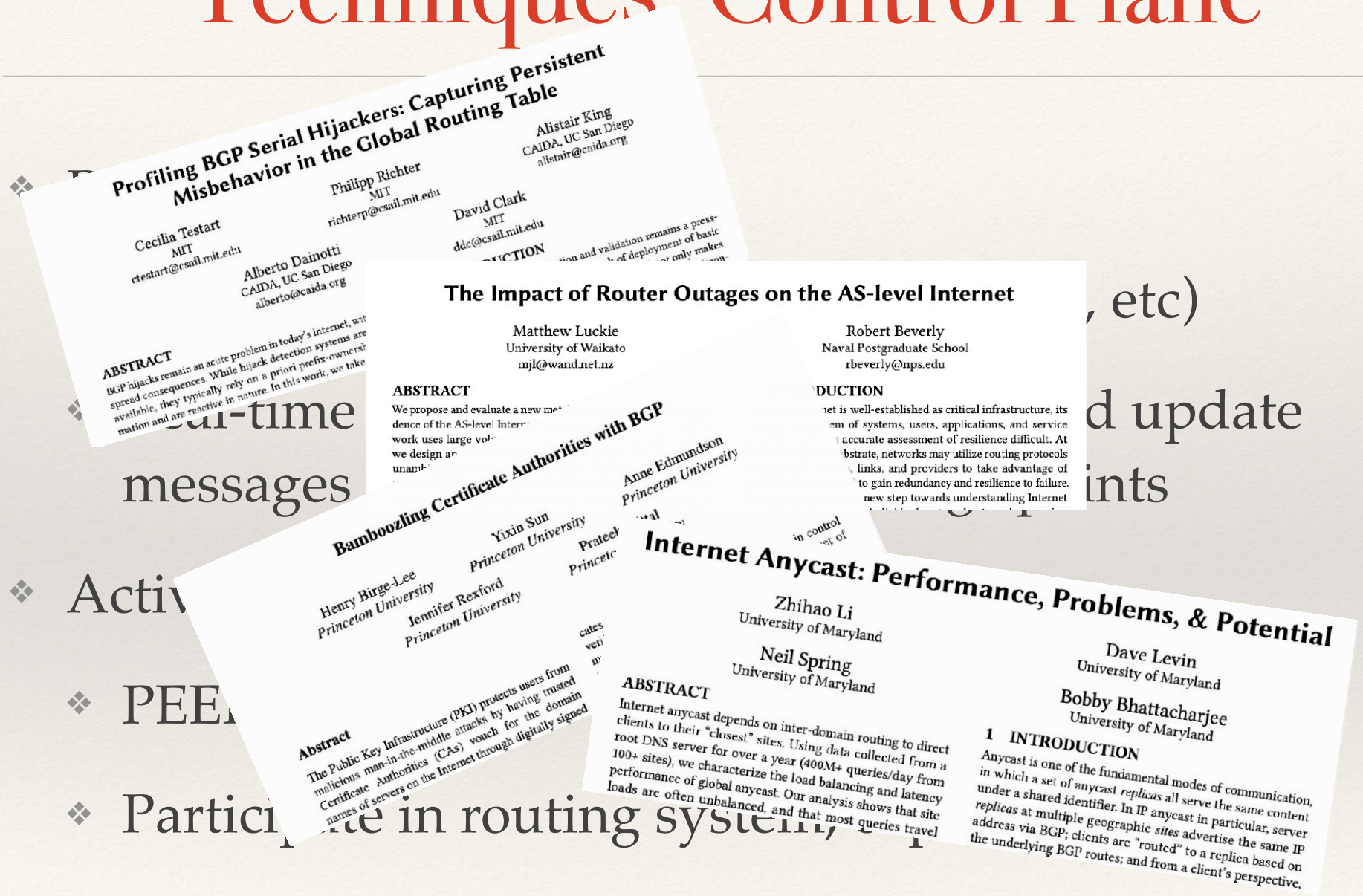
- Passive:
  - Looking glasses (RouteViews, RIPE RIS, etc)
  - Real-time and historic routing tables and update messages from hundreds of vantage points

Techniques: Control Plane

- Passive:
  - Looking glasses (RouteViews, RIPE RIS, etc)
  - Real-time and historic routing tables and update messages from hundreds of vantage points
- Active:
  - PEERING
  - Participate in routing system, experiment

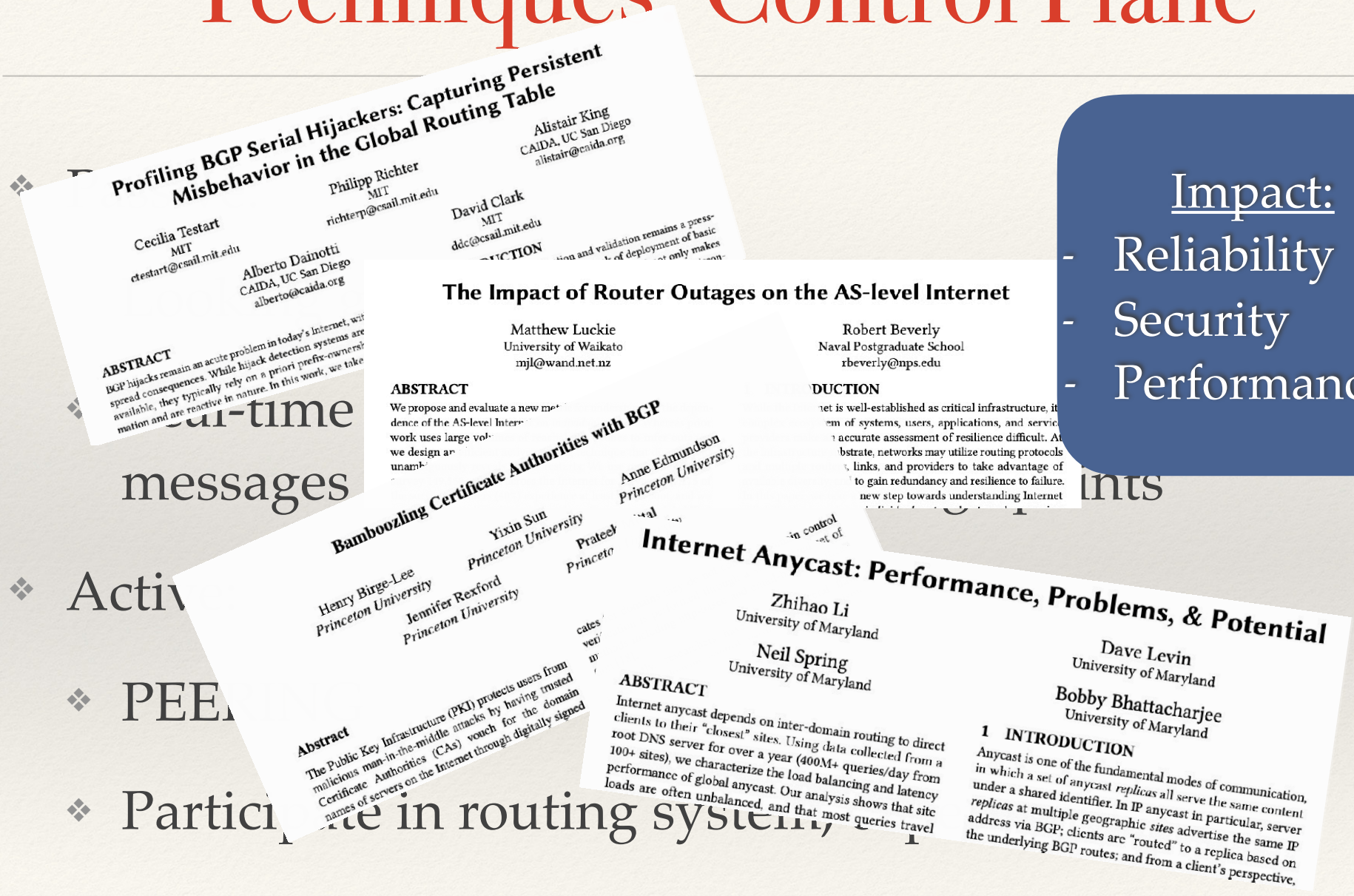
Techniques: Control Plane

- R
- Real-time , etc)
- Activ d update messages ints
- PEE
- Partici te in routing system, e



Techniques: Control Plane

- messages
- Activ
- PEE
- Partici te in routing system, [illegible]



Impact:
- Reliability
- Security
- Performance

Techniques: Data Plane

Techniques: Data Plane

- Active:
  - High-speed exhaustive IPv4 Internet-wide probing now common
  - Platforms: Ark, Atlas, scans.io, PlanetLab
  - Regularly performed and archived (rich datasets)

Techniques: Data Plane

- Active:
  - High-speed exhaustive IPv4 Internet-wide probing now common
  - Platforms: Ark, Atlas, scans.io, PlanetLab
  - Regularly performed and archived (rich datasets)
- Passive:
  - Network telescopes, packet captures of subnetwork without hosts
  - No hosts = backscatter, scans, and misconfig = security insights
  - Archived captures from large telescopes

Techniques: Data Plane

- Active:
  - de probing now common
  - R
- Passi
  - Network telescope
  - No hosts = backsca
  - Archived captures

Techniques: Data Plane

- Active:
  - [illegible]
  - [illegible]
  - R[illegible]
- Passi[illegible]
  - Network telescope
  - No hosts = backsca[illegible]
  - Archived captures

Insights Into:

- Emergent events
- Policies and changes
- Attacks and defenses

Techniques: Layer 8

Techniques: Layer 8

- Data:
  - Universities, colleagues, internships, partnerships
  - Crowdsourcing
  - Security data exchanges, e.g., SIE, DNSDB

Techniques: Layer 8

- Data:
  - Universities, colleagues, internships, partnerships
  - Crowdsourcing
  - Security data exchanges, e.g., SIE, DNSDB
- Ground-truth and validation:
  - NANOG, R&E nets, providers, etc.

Techniques: Layer 8

- Data:
  - U… …ships
  - C…
  - Security da…
- Ground-truth and …
  - NANOG, R&E ne… …ders, etc.

Stellar: Network Attack Mitigation using Advanced Blackholing

Christoph Dietzel
TU Berlin/DE-CIX

Matthias Wichtlhuber
DE-CIX

Anja Feldmann
Max Planck Institute for Informatics

Georgios Sr…
TU B…

…has attracted

ABSTRACT

Network attacks, including Distri…
continuously increase in terms o…
(recent attacks exceed 1.7 Tbps)
the targeted companies/govern…
techniques, ranging from black…

DNS Observatory: The Big Picture of the DNS

Pawel Foremski
Farsight Security, Inc. / IITiS PAN
pjf@fsi.io

Oliver Gasser
Technical University of Munich
gasser@net.in.tum.de

Giovane C. M…

ABSTRACT

The Domain Name System (DNS) is thought of as having a…
sounding task of resolving domains into IP…
stub resolvers, different layers of…
nameservers, a multi…
ecosystem is a…

Network Hygiene, Incentives, and Regulation: Deployment of Source Address Validation in the Internet

Matthew Luckie
University of Waikato
mjl@wand.net.nz

Robert Beverly
Naval Postgraduate School
rbeverly@nps.edu

Ryan Koga
CAIDA, UC San Diego
rkoga@caida.org

Ken Keys
CAIDA, UC San Diego
kkeys@caida.org

Joshua A. Kroll
Naval Postgraduate School
jkroll@jkroll.com

k claffy
CAIDA, UC San Diego
kc@caida.org

ACM Reference Format:
Matthew Luckie, Robert Beverly, Ryan Koga, Ken Keys, Joshua A. Kroll, and k claffy. 2019. Network Hygiene, Incentives, and Regulation: Deployment of Source Address Validation in the Internet. In *2019 ACM SIGSAC*…

ABSTRACT

The Spoofer project has collected data on the deployment and characteristics of IP source address validation on the Internet since 2005. Data from the project comes from participants who install an active…

Techniques: Layer 8

- Data:
  - U
  - C
  - Security da
- Ground-truth and
  - NANOG, R&E ne ders, etc.

Result:

- Creative ways to obtain better data
- Higher standard of validation
- Realism and real-world impact

Greater than Sum of Parts

- Most research leverages multiple of these techniques
- Data fusion for insight and validation
- Huge value in continuous, archived measurements:
  - *Retroactive* understanding of important events

Parting Thoughts

Parting Thoughts

- Measurements underlie systems and science

Parting Thoughts

- Measurements underlie systems and science
- Internet measurement is hard, but not without progress

Parting Thoughts

- Measurements underlie systems and science
- Internet measurement is hard, but not without progress
- Measurement input/output in your own work:
  - Quality of input datasets for experimentation?
  - Skepticism of closed measurements ("believe us!")
  - Are you really measuring what you think you're measuring? ("*Strategies for Sound Internet Measurement*", Paxson 2004)
  - What (new) measurement techniques can you leverage?
  - Contribute output measurement data (and code) to the public?

Parting Thoughts

- Measurements un
- Internet measuren
- Measurement inp
  - Quality of inpu
  - Skepticism of c
  - Are you really
    (*"Strategies for S*
  - What (new) measurement techniques can you leverage?
  - Contribute output measurement data (and code) to the public?

Thanks!

Questions / Discussion?

Rob Beverly: <rbeverly@nps.edu>