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 pervasive and crucial 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?

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

The Internet is pervasive and crucial to society.
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?
- Is this packet legitimate, or spoofed?

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

The Internet is pervasive and crucial to society.

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

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

The Internet is pervasive and crucial to society.
You get paid to measure the Internet?

- How many hosts are there on the Internet?
- Where is there congestion in the network?
- What is the topology of the Internet?
- Is this packet legitimate, or spoofed?
- Why 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?

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

❖ 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?

- 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?

- 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?
(重大) 障碍：

...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
Despite these hurdles, much success:
Relentless Forward Progress

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

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

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:**
  - 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

---

The Matter of Heartbleed

An Internet-Wide View of ICS Devices

Analysis of Country-wide Internet Outages Caused by Censorship

Millions of Targets Under Attack: a Macroscopic Characterization of the DoS Ecosystem

---

ABSTRACT

Industrial control systems (ICS) are a backbone of modern society, controlling the remote, often time-critical, processes in our society. Essentially, the protocols used by these devices are not designed with security as a primary concern. As a result, an alarming number of vulnerabilities are found in these protocols, which can be exploited by attackers.

ABSTRACT

In the first months of 2014, the vulnerability of the OpenSSL protocol suite, known as Heartbleed, was discovered. This bug allowed attackers to extract sensitive information from the server quickly and quietly.

ABSTRACT

In the first months of 2014, the vulnerability of the OpenSSL protocol suite, known as Heartbleed, was discovered. This bug allowed attackers to extract sensitive information from the server quickly and quietly.

ABSTRACT

The route of the Denial of Service (DoS) attacks has rapidly increased in frequency and intensity, with recent reports of attacks reaching “Tbps” and “Gbps” levels, and reports of noticeable degradation of Internet stability and reliability. However, a rigorous macroscopic characterization of this phenomenon and its root causes is missing. This paper presents a comprehensive analysis of the DoS ecosystem and identifies key trends and patterns that can guide future research.

ABSTRACT

In the first months of 2014, the vulnerability of the OpenSSL protocol suite, known as Heartbleed, was discovered. This bug allowed attackers to extract sensitive information from the server quickly and quietly.

1 INTRODUCTION

Denial-of-Service (DoS) attacks have rapidly increased in frequency and intensity, with recent reports of attacks reaching “Tbps” and “Gbps” levels, and reports of noticeable degradation of Internet stability and reliability. However, a rigorous macroscopic characterization of this phenomenon and its root causes is missing. This paper presents a comprehensive analysis of the DoS ecosystem and identifies key trends and patterns that can guide future research.
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

**Insights Into:**
- Emergent events
- Policies and changes
- Attacks and defenses

---

**The Matter of Heartbleed**

- Frank Li, Nicholas Weaver, Johanna Amann, Joetho Brackman, Matthias Payer, Vern Paxson
- UC Berkeley, University of California, Berkeley, Annenberg School for Communication

**An Internet-Wide View of ICS DDoS Attacks**

- Zakir Durumeric, David Adamic, Michael Schapire
- University of Notre Dame

**Analysis of Country-wide Internet Outages Caused by Censorship**

- Alberto Dainotti
  - University of Naples Federico II
  - alberto@unina.it

**Millions of Targets Under Attack: a Macroscopic Characterization of the DoS Ecosystem**

- Mattijs Jonker
  - University of Twente
  - m.jonker@utwente.nl

- Christian Rossow
  - CISPA, Saarland University

- Alistair King
  - CAIDA, UC San Diego

- Anna Sperotto
  - University of Twente

1 INTRODUCTION

Denial-of-Service (DoS) attacks have rapidly increased in frequency and intensity, with recent reports of attacks reaching millions of targets. The rise of the DoS-as-a-Service phenomenon (e.g., botnets) has dramatically expanded the threat to the stability and reliability of the Internet, posing new challenges for network operators and security researchers.
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:
  - Universities, colleagues, internships, partnerships
  - Crowdsourcing
  - Security data exchanges, e.g., SIE, DNSDB
- Ground-truth and validation:
  - NANOG, R&E networks, etc.
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.

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!”)
  - What (new) measurement techniques can you leverage?
  - Contribute output measurement data (and code) to the public?
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!")
  - What (new) measurement techniques can you leverage?
  - Contribute output measurement data (and code) to the public?

Thanks!

Questions / Discussion?

Rob Beverly: &lt;rbeverly@nps.edu&gt;