Workshop on Overcoming Measurement Barriers in Internet Research
WOMBIR-2

15-16 April, 20 April 2021

Workshop Goals

- Gather input from research community on measurement-related barriers to Internet research

    - Identify critical research questions that warrant a call for network measurement

    - Articulate in detail multiple dimensions of barriers

    - Discuss and evaluate potential directions to overcome barriers

Workshop Motivations

- Measurement, including longitudinal measurement, is essential to understanding the Internet's behavior and evolution

- Persistent infrastructure to support long-term data collection and sharing foster reproducibility and repeatability, robustness, and extensibility of research results

- Such infrastructure is expensive to create, deploy, and maintain

- Need to balance cost of maintaining such infrastructure against the potential benefits of the generated data

- Other questions not amenable to third-party measurement..

WOMBIR Part 1: January 2021

- 2 days (in a row..)

- 4 hours/day (in a row..)

- 67 participants

- ~40 position papers (1-page each, published)
  - *Identify critical research question*
  - *What data/infrastructure is needed?*
  - *How can NSF support barriers to access/sharing the data?*

- 25 talks (5-min, 3-4 per session)

- 7 sessions (50 minutes each)

WOMBIR Part 1: January 2021

- Session 1: Challenge: Understanding Properties of Internet Access

- Session 2: Broader Goals for Measurement-Based Internet Research

- Session 3: Cooperation with industry to get relevant data

- Session 4: New methods and analytics (e.g., ML/AI, diff privacy)

- Session 5A/B: Challenges of data collection and curation

- Session 6: Ethics of measurement and analysis

- Session 7: Closing Discussion and Directions (Plan WOMBIR-2)

WOMBIR Part 1: Session 1

- Session 1: Challenge: Understanding Properties of Internet Access

  - Several papers addressed same question: What measurement will be useful as part of addressing the digital divide -- the served and the un(der)served?

  - Assumed that the challenge was well-understood, so focused on range of measurements mentioned to address the challenge

  - Learned: challenge was not well-understood

  - Led to new structure for WOMBIR-2: level-setting talks, discussion

WOMBIR Part 1: Session 2

- Session 2: Broader Goals for Measurement-Based Internet Research

  - What should NSF's role be in different types of network measurement?

  - How should NSF relate to other government agencies (FCC, FTC, new efforts, etc.)?

  - What research questions might drive federal investment in the area of network measurement?

  - Learned: this meta-challenge was not well-understood

  - WOMBIR-2: Session 2 and 5

WOMBIR Part 1: Session 3

- Session 3: Cooperation with industry to get relevant data

  - Many papers emphasized need to cooperate with industry to get data

  - Industry data often not available for peer-review or subsequent research, impeding scientific progress and reproducibility

  - Learned: this meta-challenge is too well-understood, but need new ways of thinking about it

  - Last session of WOMBIR-2

WOMBIR Part 1: Session 4

- Session 4: New methods and analytics

  - AI/ML research (labels, labels, labels)

  - Daunting gap between privacy technology to support data accessibility (differential privacy, SMPC, GANs) and network research community

  - Lesson learned: privacy tech folks needed to address this gap

  - WOMBIR-2: Session 3: Role of Data Disclosure Control Technology

WOMBIR Part 1: Session 5

- Session 5: Challenges of data collection and curation

  - No shortage of measurement projects

  - Shortage of open, sustained, adaptive infrastructures

  - Not just funding, but also incentive barriers

  - Community needs to seriously consider what it needs to scale

  - Map platforms to range of research questions they can support

  - Lessons learned: many positive directions, need to articulate

  - WOMBIR-2: Session 2 and 5

WOMBIR Part 1: Session 6

- Session 6: Ethics of measurement and analysis

  - Increasingly ethical issues in data collection and use.

  - What roles should NSF and peer review have?

  - Does the community need an advisory IRB?

  - Lessons learned: Yes, probably.

  - WOMBIR-2: Session 2 and 5

WOMBIR Part 1: Feedback

- Strong community interest, but also diffuse conversations
  - WOMBIR-2: focus on specific topics targeting report text
- Need session on mobile/wireless, specific barriers
  - WOMBIR-2: Session 3
- Need more time for focused discussion
- (but..) Parallel breakouts limit participation
- People want time between sessions and summary
- Need to encourage participation by all attendees

WOMBIR Part 2: Format (Days 1-2)

- Days 1-2: Serial "breakouts" — topic-specific discussion

  - Anyone can attend any or all — except for Session 5

  - Session 5 (Day 2): parallel breakouts: one at a time

  - Recording will not be public

  - Topics do leak into eachother: repetition is reinforcement

- Each session: 1-3 speakers frame (5-10 min) and moderate discussion

  - Also responsible for getting you to help write summary notes

  - At least 45 minutes discussion + 15 min note writing

WOMBIR-2: Session 5: parallel breakouts
(Will be labelled in zoom)

- (A) Access: Home, Developing world (vantage points)
  Moderators: Ellen Zegura (Georgia Tech), David Clark (MIT)

- (B) Mobile/wireless
  Moderator: kc claffy (CAIDA / UC San Diego)

- (C) Cloud/ISP
  Moderator: John Heidemann (USC)

- (D) Layer2/optical
  Moderator: Fabian Bustamante (Northwestern)

- (E) Encryption/protocols, e.g., TLS, DOH
  Moderator: Mattijs Jonker (U Twente)

WOMBIR-2: Day 3: breakout summaries
(Tuesday 20 April 2021 10AM ET)

- Session 1: Understanding properties of Internet access: deployment, adoption, performance, pricing

- Session 2: Advancing Longitudinal Internet Research

- Session 3: Mobile and Wireless Measurement Challenges

- Session 4: The Role of Data Disclosure Control Technology in Enabling Internet Research

- Session 5: Overcoming Trends Toward Invisibility of Infrastructure

- Session 6: Industry / academia / government collaboration and roles

WOMBIR-2: Hosts

- David Clark (MIT)

- Fabian Bustamante (Northwestern)

- John Heidemann (USC/ISI)

- Mattijs Jonker (Twente)

- UCSD/CAIDA support

  - Alex Ma (webmaster)

  - Josh Polterock

  - Jennifer Sun

  - kc