Toward a theory of harms in the Internet

David Clark—MIT CSAIL

kc claffy—UCSD CAIDA

Some context

- The goal of this paper is operational and pragmatic.
  - We are concerned with what is wrong with the Internet and how to fix it.
  - CAIDA has spent 20+ years measuring aspects of the Internet.
    - Good data leads to good science.
    - Good data can lead to evidence-based policy-making.
  - What should we be measuring?
    - How can we think about this methodically? How set priorities?
- This paper is a work in progress.
- Our expertise is variable.

General organization

- Organized based on where the harm arises.
  - Layers or system elements.
  - Alternative—what is the consequence of the harm?
  - Goal is mitigation.
- Roughly—work up through the layers.
- Derived from earlier exercise to collect aspirations for the Internet.
  - A harm would seem to imply a failure to achieve an aspiration.

Our list of aspirations

- Reach
- Ubiquity
- Evolution
- Uptake
- Affordable
- Trustworthy
- Lawful
- National security
- Innovation
- Generality
- Unblocked
- Choice
- Redistribution
- Unification
- Local values
- Universal values
- Global

Defining a *harm*

- There are lots of reasons why aspirations are not achieved.
  - The failure is not automatically a harm.
    - I try to start a company and don't get funding. My aspiration to innovate is thwarted, but this is not a harm.
    - A large incumbent smushes me using anticompetitive methods. That is a harm.
- Harm: an impairment—either with respect to an individual, a firm or society—to an entity's welfare interests, relative to the normal expectations of the time and context.
  - Operationally, a harm is an impairment that rises to a level that some sort of intervention is warranted to remedy it.

Broad categories of harm

- Lack of effective availability.
- Loss of trust in the Internet experience.
- Erosion of privacy.
- Harmful barriers to innovation.
- Harms to discourse.

Broad categories of harm

- Lack of effective availability.
- Loss of trust in the Internet experience.
- Erosion of privacy.
- Harmful barriers to innovation.
- Harms to discourse.

Availability

- Harms that relate to the Internet service.
  - It's not available.
  - It costs too much.
  - It fails.
  - It offers inadequate service.
    - Both at a point in time and over time.
  - People choose not to use it.
    - When is this a harm?

It's not available

- Either absolutely, or with inadequate service.
- Measuring this harm:
  - Maps of deployment. FCC 477.
    - Begs a definitional question: deployment of *what?*
  - Maps of mobile service.
  - Measurements of service quality. FCC MBA
    - How should expectations evolve over time?
    - Dynamic definition. Emergent answer.

Measuring other harms

- Measure failures and outages.
  - Should reporting of outages be mandatory?
  - Do “we” care about outages in other parts of the world?
  - Resilience is very tricky to measure.
- Collect data on costs.
  - A little tricky.
- Survey users and non-users. Ask why.
  - Pew does this. Cost, fear, no perceived value…
  - When do refusniks impose a harm on society?

Broad categories of harm

- Lack of effective availability.
- Loss of trust in the Internet experience.
- Erosion of privacy.
- Harmful barriers to innovation.
- Harms to discourse.

Trustworthy/integrity

- High-level statement:
  - "I used the Internet, bad things happened."
  - "I worry that bad things happen, so I limit my use."
- Challenge: harms seem unbounded.
  - Look for "organizational baskets".
  - Look at layers/system elements.
- Relates to security, but broader.

Organize by system element

- Network services
- The end-nodes
- The applications

Network services

- Start with the packet carriage service:
  - Very simple service model, so harms are (somewhat) easy to classify.
- Packets are dropped, delivered to the wrong destinations, routed past malicious actors, delivered with inadequate performance, or you receive traffic you did not want.
  - The service objective is availability.

What causes harms at this layer?

1. Your ISP is malicious or has adverse interests.
2. The routing protocols of the Internet are attacked.
   - The interdomain routing protocol of the internet, Border Gateway Protocol, or BGP, has known vulnerabilities that can (and do) cause harmful consequences.
   - Massive measurement challenge.
     - ~750k distinct routing assertions. Is one of them wrong? For how long?
     - 20+ year dispute about how to remedy this harm.

Harm: Fooling the user

3. The user is fooled into using the wrong IP address as a destination.
   - The Domain Name System (DNS) maps names to addresses.
     - Harm: The user is fooled into using the wrong name.
     - Harm: The DNS is corrupted and gives the wrong address.
   - A massive measurement and tracking challenge.
     - ~350M Domain names. Are some of them corrupted? For how long?

Harm: Unwelcome traffic

- The goal of the Internet is availability.
  - It delivers what it is given.
  - Including malicious traffic.
  - Perhaps an alternate design would have required that the receiver first give permission, but that is not the Internet.
- Traffic can range from scans and probes to massive floods (DDoS attacks).
  - Typical scan rates. 2-3k per day.
  - Annoyance or harm?
  - DDoS is a harm (and hard to mitigate).

Next system element: end-node

- Harms:
  - Fails,
  - Allows unauthorized access,
    - -Which actor(s) should prevent this?
  - Blocks desired applications,
  - Allows malicious actor to hijack the update process,
  - Has harmful adverse interests.
    - Steals PII, etc.

End-node: responsibilities

- Compensate for failures and limitations (potential harms) at the Internet service layer.
  - Internet does not protect from observation and modification of packets.
    - End-node responsibility: encrypt the traffic. Triggers tussle.
  - Internet may mis-deliver traffic.
    - End-node responsibility: confirm end-point identity.
    - Depends on the Certificate Authority (CA) system.

Certificate Authority system

- Intended to give communicants the public key of the other parties.
  - Scale: perhaps a few hundred million certificates in the system.
  - CA system may itself be untrustworthy.
    - Provide the wrong key for a destination.
  - Authorities can be:
    - Corrupt
    - Subverted
    - Adversarial

Harms to trust: network and end-node

- The harms we identify can be traced (in the large) to flaws and vulnerabilities in three major systems in the Internet:
  - The Border Gateway Protocol
  - The Domain Name System
  - The Certificate Authority system.
- These systems are a major focus for technologists, and should be a major focus for policy makers.

Application layer

- Malicious applications
- Abusive applications
  - The next section
- Over-permissive applications
  - Applications that allow unconstrained interaction and don't include means to limit harms from that interaction.
  - Harmful interaction is a huge basket of harms.

Harmful interactions

- Arise at the application layer, and must be mitigated at the application layer.
- Need to get beyond whack-a-mole.
  - Some harms do deserve specific remedy.
- Two general approaches:
  - Constrain behavior.
  - Discipline participants. Implies attention to identity.

A list from the U.K.

- Child sexual exploitation and abuse.
- Terrorist propaganda and recruitment.
- Glamorizing gang life.
- Content illegally uploaded from prisons.
- Sale of opioids and other illegal drugs.
- Anonymous abuse.
- Cyber-bullying.
- Facilitating self-harm and suicide.
- Underage sharing of sexual imagery.
- Online disinformation.
- Online manipulation.
- Online abuse of public figures.

U.K. HM Government. Online Harms White Paper. April 2019

Broad categories of harm

- Lack of effective availability.
- Loss of trust in the Internet experience.
- Erosion of privacy.
- Harmful barriers to innovation.
- Harms to discourse.

Privacy/confidentiality

- Challenge: the concept of privacy can be defined abstractly, but those definitions are hard to operationalize.
  - Law, regulation and codes of conduct create “proxy harms”—specific limits on data collection and use.
- The advertising-funded ecosystem brings together targeted advertising, data collection and privacy.

What are the harms?

- Is excessive persuasion and manipulation a harm?
- Data collection in support of targeted advertising is used for other harmful purposes.
  - The advertising-funded ecosystem may be unsustainable.
- Pervasive surveillance is harmful to society.

Broad categories of harm

- Lack of effective availability.
- Loss of trust in the Internet experience.
- Erosion of privacy.
- Harmful barriers to innovation.
- Harms to discourse.

Innovation and choice

- An interconnected set of aspirations:
  - Innovation
  - Competition and choice
  - Economic growth.
- The U.S. has a deep faith in innovation and competition.
  - Not all innovation is pro-consumer, or lead to economic growth.

The fundamental harm

- Anticompetitive behavior.
  - Nothing new about this observation.
- What is new?
  - The layered platform character of the Internet ecosystem.
- Need to study the structure of the ecosystem to find new ways that anticompetitive behavior may manifest.
  - Are current laws and enforcement practices adequate?

Broad categories of harm

- Lack of effective availability.
- Loss of trust in the Internet experience.
- Erosion of privacy.
- Harmful barriers to innovation.
- Harms to discourse.

Harms to discourse

- Journalism
- The marketplace of ideas
- Our political processes

- What is new here? How are harms different?
  - The amplifying effects of the scale-free applications.
  - The pernicious consequences of the advertising ecosystem.
  - The inability to discipline misbehavior in the global space.

Theory of harm?

- We focus on which actors contribute to the making of the harm.
- The actor that is best positioned to mitigate the harm is not necessarily the one that allows/makes it.
- In general, in a layered system, actors that are well-positioned to mitigate a harm will not be at a lower layer than the layer in which the harm manifests.

Review

- Our goal was operational and pragmatic.
  - Our background is architectural and empirical.
- Good data is critical to good decision-making.
- There are many barriers to data collection.
  - Scale of the problem
  - Global scope of the problems
  - Proprietary aspect of key data
  - Legal barriers to data collection
- The governments may/will need to play a role in allowing and supporting the collection of adequate data.