Internet Architecture and Experimentation

To do ...

- Internet architecture
- Principles
- Experimentation
A bit of history – Early days

- ~1960 ARPA sponsored research on networking to let researchers share computers remotely
  - Electronic computers were a scarce resource
  - Renting an IBM System/360 - $5K/month ($35K/month 2016)

- 1969 – First four ARPANET nodes connected
  - UCLA, Stanford Research Institute, UCSB, U. of Utah
  - Key design decision – Packet switching
    - Data broken into packets, each routed separately to destination
    - Simpler than circuit switching
A bit of history – Early days

- From 1975 to 1980s
  - Successful ARPANET ~ 100 nodes
  - ARPA research on packet switching over radio and satellite
  - New LANs connected via gateways
  - TCP/IP conversion in 1983
  - Autonomous Systems and backbone AS for scalability
A bit of history – NSF takes over

- Late 1980s NSF takes over
  - Focus on expanding the backbone
- NSF encourage development of regional networks
  - Three tiers: backbone, regional, enterprise
- Enterprises wanted to connect their networks
  - They were building TCP/IP networks and wanted to connect
  - But NSF charter prohibited them from using NSFNET
  - 1987 first commercial ISP, many follow shortly
A bit of history – To commercial operation

- By 1990, service providers were interconnected
  - Congress lets NSFNET interconnect with commercial networks
  - By 1995, NSFNET was retired
    - No single default backbone anymore
    - Many backbones interconnected through Network Access Points

- ~1995 Web
  - Easier to use Internet
  - Million of non-academic users

- Now …
  - Large ISPs interconnected, regional ISPs, mid-size ISP and eyeballs
Internet model (traditional and outdated)

- National Backbone Operators: Sprint, MCI, AGIS, …
- Regional Access Providers
  - NAP
  - ISP1
  - ISP2
- Local Access Providers
- Customer IP Networks
Updated Internet model

- Global Internet Core
- Regional / Tier 2 Providers
- Customer IP Networks

- Global Transit/ National Backbones
- “Hyper Giants” Large Content, Consumer, Hosting CDN
- IXP
- ISP

- Flatter and much more densely interconnected Internet
- Disintermediation between content and “eyeball” networks
- New commercial models between content, consumer and transit

Labovitz et al., SIGCOMM 2010
Networks and protocols

- Networks’ basic components
  - End-systems, routers and links – media moving packets

- For communication to work, parties must agree on several things – protocol
  - How does the receiver know it got the last bit of a message?
  - How longs are integers?
  - ...

- Each protocol – a set of rules and formats to use
  - Sequence of messages to exchange, format of messages, …
  - Implemented by pair of software modules on each end
  - Typically arranged in layers, to handle complexity – protocol stack
Protocols in communication

- Internet protocols – four layers
  - ISO OSI, 7 layers, not used in practice but good for understanding

- As a message is
  - Pass down, each layer adds a header (sometimes a trailer)
  - On the receiver side the message is push upward with each layer stripping off and examining their own headers
Internet protocols

- **Link layer – Ethernet, WiFi, …**
  - Move packets between locations
  - Physically interface with communication medium

- **Network layer – IP**
  - Move packets between hosts, over sequences of links
  - Provided mostly by routers as they do packet forwarding

- **Transport layer – TCP, UDP**
  - Transfer msgs between clients, including breaking them into packets, controlling flow, re-ordering, etc

- **Application layer – SMTP, HTTP, P2Ps, DNS,**
  - Implements each particular application (email, WWW, …)
Internet protocols – TCP/IP

- IP – Internet Protocol – primary network layer protocol
- A best effort service – no guarantees
- Somebody else does reliability – TCP
- Connection between host & network is called interface
  - Each interface has an address; format defined by IP
  - In IPv4 addresses – 32b expressed as four decimals
  - Assigned in a manner that reflects network organization
- IP addresses often grouped by prefixes – initial bits
  - e.g. 165.124.0.0/16 – addresses with first 16b = 165.124
Internet protocols – TCP/IP

- **TCP** – to ensure data reliability

- **Basic service** – a comm. channel between two hosts
  - Protocol specifies how connections are initiated (three-way handshake), used and released

- **General approach to reliability** – ACKs
  - After sending each segment, source sets a timer waiting for an ACK

- **Congestion avoidance** by varying number of un-ACKs packets the sender allows – window size
  - How the window size is adjusted? Congestion control algo
Internet as a set of ASes

- **Internet**
  - Collection of separately, usually competing, managed networks

- **Autonomous system (AS)**
  - Set of network elements under a single organization’s control
  - 1 ISP can operate N ASes; no AS is managed by >1 ISP

- **ASes exchanged traffic at connection points, aka peering or exchange points**
  - Connections formed by establishing a link between “gateway” routers in each AS
Internet routing

- Packets must be forward to dest as specified in each packet’s header
  - Routing is a graph – find the lowest-cost path between two nodes

- Routers decide outgoing interface for each packet
  - Decision – routing
  - Placing it in the correct outgoing interface – forwarding
  - For routing, routers keep info on network configuration and state

- Internet routing is hierarchical
  - Two levels intra-AS (intra-domain) and inter-AS
  - Both (1) for scalability (too large, too many routers) and (2) for independence (each AS mange their network independently)
Intra-domain routing

- Common intra-AS protocols
  - Distance vector, Bellman-Ford, algorithms – RIP
  - Link state algorithms – OSPF, IS-IS

- Open Shortest Path First (OSPF)
  - Routers maintain a network map, updated when links change
  - Each router monitor the links to which it is connected and initiates a flooding protocol upon change
  - Router talks to everyone, but only tells them what it knows for sure (state of directly connected links)

- To compute path to destination
  - Each computes best path from it to every destination – path with the lowest sum of link weights → routing table
Intra-domain routing

- Inter-domain routing is a bit more challenging
  - Economic and policy issues
  - e.g. “I prefer to send traffic via AS X than Y, but I’ll use Y if it is the only path, and don’t want to carry traffic between X and Y”

- Border Gateway Protocol (in v4) – some highlights
  - Internet is an arbitrary interconnected set of ASes
  - BGP advertises complete paths as an enumerated AS list
    - Necessary to enable policy decision and detect loops
  - Designed to run on top of TCP
    - So no need to retransmit, but need keep-alive messages
  - BGP updates
    - A BGP router will advertise (no obligation) one of several routes to a destination, selected according to its policies
    - It can also send a negative advertisement – withdrawn route
A 5’ brake

There are few forces more powerful than geeks desperately trying to get internet in a new apartment.

Okay, the Pringles cantenna has let us patch in to the WiFi network across the road.

And they have internet?

No, but I think the cable van will hook up their house first.

xkcd
Design principles of the Internet

- To meet functional goals of the Internet, several design decisions have been made – Inferred principles

- Decentralized design and operation
  - A loose interconnection of networks, not really “one” network
  - Connecting a node to the Internet does not require the consent of any global entity
  - Key factor for its rapid growth

- IP hourglass or IP over everything
  - Internet overarching goal – to provide connectivity – IP is key
  - Easy to incorporate both new applications and new communication media
Stateless switching
- Switches are expected to be stateless wrt connections
- When packet arrives, router inspect packet IP’s header and a table to determine how to forward the packet
- Results in very simple routers
- Simplicity of router element is related to …

End-to-end
- Insight – many network functions require cooperation form end-systems for correct and complete operation
  - Don’t try to provide it within the network (except for performance), this will not simplify end-systems
  - E.g. reliable transfer will always need to be checked by end-systems
- Some challenges to end-to-end: untrustworthy world, more demanding apps (use of CDNs), less sophisticated users, …
Design principles and measurements

- Decentralized design and operation
  - Hard to learn the current configuration of the Internet

- IP over everything
  - Makes measuring hard since it hides details of the particular physical medium

- Stateless switching
  - Complicates measurements since routers do not capture or track any aspect of the traffic going by

- End-to-end argument
  - Lack of instrumentation at many points in the network, since it encourages the design of network elements with minimal functionality
Measurement and experimentation

- At every layer
  - Network infrastructure and routing
  - Traffic
  - Applications

- At higher-layers, other concerns
  - Censorship
  - Ethical considerations

- Where can we take measurements?
Where do we measure?

Measurement locations in an ISP

- **ISP X**: Gateway routers route traffic between ISP and other AS
- **ISP Y**: Access routers provide connectivity to customer networks over an access link
- **NAP (Network access point)**: An exchange point of multiple ISPs that can exchange traffic

- **Gateway routers**: Route traffic between ISP and other AS
- **Backbone router**: Central to the ISP's network
- **Access router**: Supplies access to customer networks
- **Access link**: Connection between access routers and customers
- **Peering links**: Connect different ASes for traffic exchange
Measurement in the wide area

- Measurement at a single or few locations are hard to generalize from …

- Measurements across the wide-area
  - Many of the same places, but across a wider area
  - Done by researchers and companies
  - Using distributed platforms for coordinated measurements

- Some measurement platforms and datasets
  - Platforms: NIMI, Planetlab, RIPE Atlas, Archipelago, Dasu, …
  - Datasets: NLANR/CAIDA, Route Views, CRAWDAD …
Wide area measurements – some issues...

- Designing multi-site measurement platforms
- Vantage points locations and layers
  - Placing landmarks ensuring representativeness
  - Networks, applications, clients …
  - Moving up the stack …
- Platform programmability
- Measurement coordination

- Observation and control
  - Observation studies
  - Controlled experiments
  - Natural experiments
More issues ...

- Main problem
  - A decentralized and distributed architecture
  - No support for third-party measurements
- So, measurement efforts
  - have limited visibility (and getting worst)
  - rely on hacks, rarely validated
  - More often that not … what we can measure is not what we want and, worst, what we think we measure
- Key question for any measurement-based research

*Do the available measurements, their analysis and modeling efforts support the claims that are made?*
Standards? What standards?

- Basis for validation of measurement-based research
  - There are no standards, so we repeat common errors

- A Socratic approach
  - Q1: Are the measurements being used of good enough quality for the purpose of the study? We need metadata for this
  
  - Q2: Is the level of statistical rigor used in the analysis commensurate with the quality of the measurements?
  
  - Q3: Have alternative models been considered and what criteria have been used to rule them out?
  
  - Q4: Does model validation reduce to showing that the proposed model can reproduce certain statistics of the data?
Topology studies as (counter-)examples

- What does the Internet look like? Why do we care?
  - Performance of networks critically dependent on topology
  - Modeling of topology needed to generate test topologies
  - ...

- Internet topology at different levels
  - Router-level reflect physical connectivity between nodes
    - Inferred from tools like traceroute or well known public measurement projects like CAIDA’s Ark
  - AS-level reflects a peering relationship between two providers/clients
    - Inferred from inter-domain routers that run BGP and public projects like Oregon Route Views
Trends in topology modeling

- (Observation ➔ modeling approach)
- Long-range links are expensive
  - Random graph (Waxman ’88)
- Real networks are not random, but have obvious hierarchy
  - Structural models (GT-ITM, Zegura et al. ‘96)
- Internet topologies exhibit power law degree distributions (Faloutsos et al., ‘99)
  - Degree-based models replicate power-law degree sequences
- Physical networks have hard technological (and economic) constraints
  - Optimization-driven models topologies consistent with design tradeoffs of network engineers

Power laws and Internet topology

$R(d) = P(D > d) \times \#\text{nodes}$

- Most nodes have few connections
- A few nodes have lots of connections

- Router-level and AS graphs
- Led to research in degree-based network models
Many complex systems are error tolerant because of the redundantly wired network underneath – but not all

Two types

- Random graphs and small-world, homogenous (E)
  - Nodes connected to each other with some probability
  - Connectivity follows a Poisson and decreases exponentially

- Scale-free networks, in-homogeneous (SF)
  - Nodes are preferentially connected to high connected nodes

*Albert, Jong and Barabasi, Error and attack tolerance of complex networks, Nature, July 2000 (6,300 citations)
Error and attacks on complex networks

- Failures and attacks
  - Failures – randomly take a node, look at diameter
    - E networks - monotonically increases
    - SF networks – unchanged even with 5% of nodes gone
  - Attack – take the high connected nodes first
    - E networks – monotonically increases
    - SF – diameter increases rapidly and the network breaks in pieces

- Analysis done with synthetic networks and

- Two datasets
  - For the Internet same as Faloutsos
  - For WWW, generated from a sample of 350k nodes

- Achilles’ heel: robust to random failure, fragile to specific attack

*Albert, Jong and Barabasi, Error and attack tolerance of complex networks, Nature, July 2000*
Does the Internet have these features?

- No … focus on degree distribution, ignoring structure
  - Real Internet very structured
  - Evolution of graph is highly constrained

- Answering key questions
  - (Q1) Are the available measurements good enough ….  
    • Original goal of data collection “shape of multicast trees”, collected with traceroute … and BGP, which is about routing, not connectivity
  - (Q2) Given the answer to Q1, fitting a particular parameterized distribution is overkill
  - (Q3) There are other models, consistent with the data, with different features
    • Seek a theory for Internet topology that is explanatory, no just descriptive
  - (Q4) Model validation reduced to showing that the proposed model can reproduce certain statistics of the available data
Coming up

- Experimental platforms and experimental design
  - PlanetLab, Dasu, RIPE, mobile platforms …
- Measurement experiments and experimental designs through examples
- Topology, routing, traffic …