Networking - Civil engineering for the 21st century

Henning Schulzrinne
Dept. of Computer Science
Columbia University
New York, NY
Overview

• Network as core infrastructure
• The illusion of a next-generation Internet
  – Interfaces persist, implementations change
  – Towards the two-port Internet
  – What you learned in Networking 101 is (mostly) wrong
• Challenges – 2 examples:
  – diagnostics → DYSWIS
  – opportunistic and store-carry-forward networks → 7DS

Sarnoff 2009 (Princeton, NJ)
IP as a core infrastructure interface
The great infrastructure

- Technical structures that support a society → “civil infrastructure”
  - Large
  - Constructed over generations
  - Not often replaced as a whole system
  - Continual refurbishment of components
  - Interdependent components with well-defined interfaces
  - High initial cost

water  energy  transportation

Sarnoff 2009 (Princeton, NJ)
The Internet as core civil infrastructure

• Involved in all information exchange
  – (in a few years)

• Crucial to
  – commerce
  – governance
  – coordination
  – inter-personal communication

• Assumed to just be there
  – “plumbing”, “pipes”, …

Sarnoff 2009 (Princeton, NJ)
Interfaces: Energy

• Lots of other (niche) interfaces
• Replaced in a few applications

Sarnoff 2009 (Princeton, NJ)
Interfaces: Paper-based information

Sarnoff 2009 (Princeton, NJ)
Interfaces: Transportation

About 60% of world railroad mileage

1435 mm
1830 (Stephenson)
1846 UK Gauge Act

12’
Sarnoff 2009 (Princeton, NJ)
1949
Modular: 1975-

4 kHz spectrum
48 V off-hook
275 mV audio

Sarnoff 2009 (Princeton, NJ)
Other long-lived interfaces

- Cigarette lighter (1956)
- Fuel nozzle
- SQL (1974)
- Sarnoff 2009 (Princeton, NJ)
What makes interfaces permanent?

- Widely distributed, uncoordinated participants
- Capital-intensive
  - depreciated over 5+ years
  - see Y2K problem
- Allocation of cost vs. savings
  - ISP saves money, end user pays
- Hard to have multiple at once
  - “natural monopoly”
Extrapolating from history

• IP now “the” data interface
• Unclear that any packet-based system can be
  – ≥ 10 times cheaper
  – ≥ 10 times more functionality
  – ≥ 10 times more secure
• Replacing phone system due to generality, not performance
  – IP offers general channel
• → We’re stuck with IPv4/IPv6
  – except for niche applications (car networks, BlueTooth, USB, …)

Sarnoff 2009 (Princeton, NJ)
Integrating infrastructures: Energy

- Much of the improvement in civil infrastructure needs networks → information networks complement other networks
  - transportation
  - energy

- Energy time management
  - Plug-in hybrid is notified when it should charge
  - Dishwasher, water heater run after midnight
  - “when can I get 100 kW?”

- Utility requests load reduction
  - “please reduce load by 1 MW”

- Energy management
  - “Dear fridge, how many kWh have you used?”

Sarnoff 2009 (Princeton, NJ)
Example: Possible IETF RECIPE effort

- Discover controllers and elements
  - Utility (gas, electric)
  - Local controllers
- Authenticate
  - Prices and actions may depend on customer contract
- Control
- Information

“charge at 2300”
“wash at 1900”
“What’s the projected cost of a kWh at 1500?”

Sarnoff 2009 (Princeton, NJ)
What role does research need to play?
Quest for Fundamental Understanding?

Yes

No

Considerations of Use?

No

Yes

Pure basic research (Bohr)

Use-inspired basic research (Pasteur)

Guessing at problems (Infocom)

Pure applied research (Edison)

Most networking research wants to be here

Most networking research is here

Sarnoff 2009 (Princeton, NJ)

Pasteur’s Quadrant: Basic Science and Technological Innovation, Stokes 1997 (modified)
Network research $\rightarrow$ reality

rarely read

“too much effort”

13,000 QoS papers

Sarnoff 2009 (Princeton, NJ)
# Planning vs. Evolution

<table>
<thead>
<tr>
<th>Planning</th>
<th>Evolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>requirements analysis</td>
<td>start small</td>
</tr>
<tr>
<td>describe all features</td>
<td>outline architecture</td>
</tr>
<tr>
<td>ATM &amp; B-ISDN NGN</td>
<td>Ethernet &amp; web</td>
</tr>
</tbody>
</table>

*see also CACM 12/08*
Are we an engineering discipline?

- Reasonable set of rules and tools for designing networks
- But:
  - no easy way to predict service capabilities
  - no formal protocol engineering
    - mostly passed-down “wisdom” and (IETF/ITU) culture
  - no (formal) learning from mistakes
  - no “Professional Engineering” (PE) exams
    - just (Cisco/Novell/Microsoft) certification

Sarnoff 2009 (Princeton, NJ)
Completing the migration of comm. applications

Sarnoff 2009 (Princeton, NJ)
Migration of applications, cont’d.

<table>
<thead>
<tr>
<th></th>
<th>text, still images</th>
<th>audio</th>
<th>video</th>
</tr>
</thead>
<tbody>
<tr>
<td>synchronous</td>
<td>IM</td>
<td>VoIP</td>
<td>video conferencing</td>
</tr>
<tr>
<td>asynchronous</td>
<td>email</td>
<td>email, voicemail</td>
<td>YouTube</td>
</tr>
</tbody>
</table>

Sarnoff 2009 (Princeton, NJ)
Aside: technology evolution

• Early technology stages:
  – make it work
  – make it cheap
  – make it fashionable
    – This happened in the auto industry. Early cars barely worked at all, every journey was an adventure. In the 1920s Ford broke the automobile patent and built a car for the common man, a car that did not need the skills of a mechanic to drive. Reliability improved gradually until the 1970s when there was a sudden realization that consumers would pay more for a car that was not designed to rust. Today most cars will go 10,000 miles between services and not need major repairs beyond a clutch plate for 50,000 or even 100,000 miles

• Completion of conversion from analog to digital/packet media

• Patterson: Security, Privacy, Usability, Reliability
  – phishing attacks, DDOS
  – cost of purchase vs. cost of ownership
  – dependability (crashes & reboots)

Sarnoff 2009 (Princeton, NJ)
Why is the Internet ossifying?

• Lack of network transparency
  – NATs: only UDP + TCP; only client-server
  – Firewalls

• Standardization delays
  – No major new application-layer protocol since 1998
  – Protocols routinely take 5+ years

• Deployed base
  – Major OS upgrade every 7-8 years
  – But: automatic software updates

Sarnoff 2009 (Princeton, NJ)
Building Internet applications

- C/C++ with sockets
- custom protocols on UDP, TCP
- extensible CMS, Wiki
  (Drupal, Mambo, Joomla, ...)
- Ruby on Rails, Spring, ...
  Ajax, SOAP
- PHP, Java w/libraries
  Java RMI, HTTP

80% care about this level

taught in Networking 101

Sarnoff 2009 (Princeton, NJ)
Which Internet are you connected to?

port 80 + 25

IPv4 NAT
IPv4 DHCP
IPv4 PIA
IPv4 IPv6
QoS multicast

Sarnoff 2009 (Princeton, NJ)
## Cause of death for the next big thing

<table>
<thead>
<tr>
<th>Issue</th>
<th>QoS</th>
<th>multicast</th>
<th>mobile IP</th>
<th>active networks</th>
<th>IPsec</th>
<th>IPv6</th>
</tr>
</thead>
<tbody>
<tr>
<td>not manageable across competing domains</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td></td>
</tr>
<tr>
<td>not configurable by normal users (or apps writers)</td>
<td>✗</td>
<td></td>
<td></td>
<td>✗</td>
<td>✗</td>
<td></td>
</tr>
<tr>
<td>no business model for ISPs</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>no initial gain</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td></td>
</tr>
<tr>
<td>80% solution in existing system</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td></td>
</tr>
<tr>
<td>increase system vulnerability</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td></td>
</tr>
</tbody>
</table>

Sarnoff 2009 (Princeton, NJ)
The two-port Internet

• Many public access systems only allow port 80 (HTTP) and maybe 25 (SMTP)
  – e.g., public libraries

• Everything tunneled over HTTP
  – Web-based email
  – Flash video delivery (e.g., YouTube)
  – HTTP CONNECT for remote login
## More than just Internet Classic

<table>
<thead>
<tr>
<th>Network</th>
<th>wireless</th>
<th>mobility</th>
<th>path stability</th>
<th>data units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet “classic”</td>
<td>last hop</td>
<td>end systems</td>
<td>&gt; hours</td>
<td>IP datagrams</td>
</tr>
<tr>
<td>mesh networks</td>
<td>all links</td>
<td>end systems</td>
<td>&gt; hours</td>
<td></td>
</tr>
<tr>
<td>mobile ad-hoc</td>
<td>all links</td>
<td>all nodes, random</td>
<td>minutes</td>
<td></td>
</tr>
<tr>
<td>opportunistic</td>
<td>typical</td>
<td>single node</td>
<td>≈ minute</td>
<td></td>
</tr>
<tr>
<td>delay-tolerant</td>
<td>all links</td>
<td>some predictable</td>
<td>some predictable</td>
<td>bundles</td>
</tr>
<tr>
<td>store-carry-forward</td>
<td>all nodes</td>
<td>all nodes</td>
<td>no path</td>
<td>application data units</td>
</tr>
</tbody>
</table>

Sarnoff 2009 (Princeton, NJ)
<table>
<thead>
<tr>
<th>Network model</th>
<th>route stability</th>
<th>motion of data routers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet</td>
<td>minutes</td>
<td>unlikely</td>
</tr>
<tr>
<td>mobile ad-hoc</td>
<td>3 τ</td>
<td>disruptive</td>
</tr>
<tr>
<td>store-carry-forward</td>
<td>&lt; 3 τ</td>
<td>helpful</td>
</tr>
</tbody>
</table>

Sarnoff 2009 (Princeton, NJ)
IP model

application
upper-layer protocol
IP
link layer

application
upper-layer protocol
IP
link layer

Sarnoff 2009 (Princeton, NJ)

D. Thaler, IETF 7
Basic IP service model

- Unchanged since 1978
- Send without signaling
- Receive at provisioned address, without signaling
  - but: permission-based sending
- Variable-sized packets $< \approx 1,500$ bytes
- Packets may be lost, duplicated, re-ordered

Sarnoff 2009 (Princeton, NJ)
Myth #1: Addresses are global & constant

DHCP
128.59.16.28
128.59.16.14
10.0.1.2
192.168.0.1
10.0.1.1

STUN
1.2.3.4

also: identifier-locator split

128.59.16.12

128.59.16.28

Sarnoff 2009 (Princeton, NJ)
Myth #2: Connectivity commutes, associates

- Referals, call-backs, redirects
- Assumptions:
  - A connects to B → B can connect to A
  - A connects to B, B to C → C can connect to A
- May be time-dependent

Sarnoff 2009 (Princeton, NJ)
Myth #2a: Bidirectional connectivity

Sarnoff 2009 (Princeton, NJ)
Myth #3: End-to-end delay of 1st packet typical

- 1st packet may have additional latency
  - ARP, flow-based routers
- MIPv6, PIM-SM, MSDP: fixed path during initial data burst
- Choice of server may be suboptimal
  - higher delay, lower throughput, inefficient network usage

Sarnoff 2009 (Princeton, NJ)
Addressing assumptions

• A host has only one address & one interface
  – apps resolve name and use first one returned
  – address used to identify users and machines
  – machine-wide DHCP options

• Failing
  – multi-homing on hosts (WiFi + Ethernet + BlueTooth + 3G)

• Attempts to restore
  – MIP: attachment-independent address
  – HIP: cryptographic host identify

Sarnoff 2009 (Princeton, NJ)
Other assumptions

- Multicast supported on link
- IPv4 broadcast
- Broadcast/multicast << replicated unicast
- Reordering is rare
- Loss is rare and random
- An end-to-end path exists at a single time point

D. Thaler, draft-iab-ip-model-evolution

Sarnoff 2009 (Princeton, NJ)
Causes

• Link-layer technologies
  – satellite, DSL
  – NBMA

• Network-layer technologies
  – security: broken by design vs. broken by accident?
  – NATs
  – Ill-defined meaning of IP addresses and names
    • theoretically, single network interface
    • practically, often more than that
      – virtualization
      – multi-homing
      – fail-over

Sarnoff 2009 (Princeton, NJ)
Research challenges
User challenges vs. research challenges

• Are we addressing real user needs?
  – Engineering vs. sports

• My guesses

  - reliability
  - cost
  - ease of use
  - no manual
  - integration
  - no re-entry
  - no duplication
  - phishing
  - data loss
  - limited risk

Sarnoff 2009 (Princeton, NJ)
A^7: Anytime Anywhere Affordable Access to Anything by Anyone Authorized

- Anytime and anywhere
  - From chip-level and biological networks to global scale

- Anything
  - Digital artifacts to services

- Anyone
  - “young and old, rich and poor, abled and disabled, literate and illiterate”

- Access
  - “Only authorized users will have the relevant access rights.”

- Affordable

- Authorized

Jeanette Wing, NSF, Assistant Director for CISE

http://www.cra.org/CRN/articles/nov08/Wing-A7.html

Sarnoff 2009 (Princeton, NJ)
Network challenges:

- Multi-homing
- Routing table explosion

- 99.9 → 99.999%

Zero configuration:

- +2 years
- +5 years
- +8 years

Sarnoff 2009 (Princeton, NJ)
Example: BGP growth

http://bgp.potaroo.net/  Sarnoff 2009 (Princeton, NJ)
Network of the (near) future

Homes passed by multiple networks → increase reliability by connecting to all (“reliable system out of unreliable components”)

Sarnoff 2009 (Princeton, NJ)
Need for new network abstractions

- Need to isolate applications from gritty network reality
- Name-based
  - multiple end points for one service
    - extend DNS MX and SIP NAPTR/SRV model to all services
  - IPv4 = IPv6
  - local vs. global address space
  - TCP = SCTP
  - multi-homing

Sarnoff 2009 (Princeton, NJ)
What about security?

- “The future Internet must be secure”
- Most security-related problems are not network problems
  - spam: identity and access, not SMTP
  - web: (mostly) not TLS, but distinguishing real bank from fake one
  - web: cross-domain scripting, code injection
  - browser vulnerabilities & keyboard sniffers
- Automated tools
  - better languages, taint tracking, automated input checking, stack protection, memory randomization, …
- Probably need more trust mediation

Sarnoff 2009 (Princeton, NJ)
What about security?

Technologies (mostly) available, but use & deployment hard

Sarnoff 2009 (Princeton, NJ)
Usability: Email configuration

- Application configuration for (mobile) devices painful
- SMTP port 25 vs. 587
- IMAP vs. POP
- TLS vs. SSL vs. “secure authentication”
- Worse for SIP...

Sarnoff 2009 (Princeton, NJ)
Usability: SIP configuration

- highly technical parameters, with differing names
- inconsistent conventions for user and realm
- made worse by limited end systems (configure by multi-tap)
- usually fails with some cryptic error message and no indication which parameter
- out-of-box experience not good

Sarnoff 2009 (Princeton, NJ)
Usability: Interconnected devices

opens doors

incoming call
generates TAN

time, location

updates location

address book

any weather service
school closings

acoustic alerts

alert, events

Sarnoff 2009 (Princeton, NJ)
Mobile why’s

• Not research, but examples of real annoyances
• Why does each mobile device need its own power supply?
• Why do I have to adjust the clock on my camera each time I travel?
• Why do I have to know what my IMAP server is and whether it uses TLS or SSL?
• Why do I have to type in my address book?
• Why do I have to “synchronize” my PDA?
• Why do I have to manually update software?
• Why is connecting a laptop to a projector a gamble?
• Why do we use USB memory sticks when all laptops have 802.11b?
Examples of “invisible” behavior

- MP3 player in car automatically picks up new files in home server
- A new email with vcard attachment automatically updates my cell phone address book
- The display of my laptop appears on the local projector
  - without cable or configuration
- I can call people I just met at COMSNETS
  - without exchanging business cards
- My car key opens my front door
- My cell phone serves as a TAN (one-time password) generator
- My cell phone automatically turns itself off during a lecture
- My camera knows where the picture was taken
Protocol & UI design guidelines

• Users should never be exposed to protocol names, ports or cryptographic protocols.
• If the network does not support an option, the UI should not show it.
• Every application protocol must allow the discovery of the domain-appropriate server and any backups.
• User-specific parameters must have reasonable defaults; others must be obtained automatically.
• A UI must make it clear why a protocol failed and indicate who is likely responsible.
• Protocols must work with (reasonable) NATs or fail with a clear indication that a NAT is the likely culprit.

Sarnoff 2009 (Princeton, NJ)
Increasing reliability and usability through end system diagnostics

with Kyung-Hwa Kim, Vishal Singh and Kai Miao

Sarnoff 2009 (Princeton, NJ)
Circle of blame

probably packet loss in your Internet connection → reboot your DSL modem

ISP

probably a gateway fault → choose us as provider

VSP

must be a Windows registry problem → re-install Windows

app vendor

must be your software → upgrade

Sarnoff 2009 (Princeton, NJ)
Diagnostic undecidability

- symptom: “cannot reach server”
- more precise: send packet, but no response
- causes:
  - NAT problem (return packet dropped)?
  - firewall problem?
  - path to server broken?
  - outdated server information (moved)?
  - server dead?
- 5 causes → very different remedies
  - no good way for non-technical user to tell
- Whom do you call?

Sarnoff 2009 (Princeton, NJ)
Traditional network management model

"management from the center"

Sarnoff 2009 (Princeton, NJ)
Old assumptions, now wrong

• Single provider (enterprise, carrier)
  – has access to most path elements
  – professionally managed
• Problems are hard failures & elements operate correctly
  – element failures (“link dead”)
  – substantial packet loss
• Mostly L2 and L3 elements
  – switches, routers
  – rarely 802.11 APs
• Problems are specific to a protocol
  – “IP is not working”
• Indirect detection
  – MIB variable vs. actual protocol performance
• End systems don’t need management
  – DMI & SNMP never succeeded
  – each application does its own updates

Sarnoff 2009 (Princeton, NJ)
Managing the protocol stack

- **media**
  - echo gain problems
  - VAD action

- **RTP**
  - protocol problem
  - playout errors

- **UDP/TCP**
  - TCP neg. failure
  - NAT time-out
  - firewall policy

- **IP**
  - no route
  - packet loss

- **SIP**
  - protocol problem
  - authorization
  - asymmetric conn (NAT)

Sarnoff 2009 (Princeton, NJ)
Types of failures

- **Hard failures**
  - connection attempt fails
  - no media connection
  - NAT time-out

- **Soft failures (degradation)**
  - packet loss (bursts)
    - access network? backbone? remote access?
  - delay (bursts)
    - OS? access networks?
  - acoustic problems (microphone gain, echo)
  - a software bug (poor voice quality)
    - protocol stack? Codec? Software framework?

Sarnoff 2009 (Princeton, NJ)
Examples of additional problems

- **ping and traceroute no longer works reliably**
  - WinXP SP 2 turns off ICMP
  - some networks filter all ICMP messages
- **Early NAT binding time-out**
  - initial packet exchange succeeds, but then TCP binding is removed ("web-only Internet")
- **policy intent vs. failure**
  - “broken by design”
  - “we don’t allow port 25” vs. “SMTP server temporarily unreachable”

Sarnoff 2009 (Princeton, NJ)
Fault localization

- Fault classification – local vs. global
  - Does it affect only me or does it affect others also?

- Global failures
  - Server failure
    - e.g., SIP proxy, DNS failure, database failures
  - Network failures

- Local failures
  - Specific source failure
    - node A cannot make call to anyone
  - Specific destination or participant failure
    - no one can make call to node B
  - Locally observed, but global failures
    - DNS service failed, but only B observed it

Sarnoff 2009 (Princeton, NJ)
Do You See What I See?

Sarnoff 2009 (Princeton, NJ)
Project: “Do You See What I See?”

- Each node has a set of active and passive measurement tools
- Use intercept (NDIS, pcap)
  - to detect problems automatically
    - e.g., no response to SIP, HTTP or DNS request
    - deviation from normal protocol exchange behavior
  - gather performance statistics (packet jitter)
  - capture RTCP and similar measurement packets
- Nodes can ask others for their view
  - possibly also dedicated “weather stations”
- Iterative process, leading to:
  - user indication of cause of failure
  - in some cases, work-around (application-layer routing) → TURN server, use remote DNS servers
- Nodes collect statistical information on failures and their likely causes

Sarnoff 2009 (Princeton, NJ)
DYSWIS overview

XMLRPC
For Remote Function call

DHT for looking for remote node

Internet

Sarnoff 2009 (Princeton, NJ)
Architecture

Sensor node

Diagnosis node

inspect protocol requests (DNS, HTTP, RTCP, ...)

“not working” (notification)

orchestrate tests

contact others

ping 127.0.0.1
can buddy reach our resolver?

“DNS failure for 15m”

notify admin (email, IM, SIP events, ...)

request diagnostics

Sarnoff 2009 (Princeton, NJ)
Rule Example

(load-function ExMyUpcase)
(load-function SelfDiagnosis)
(load-function DnsConnection)
(load-function ProxyServer)
(load-function SipResult)
(defrule MAIN::SIP
 (declare (auto-focus TRUE))
 =>
 (process-sip void)
)

(deffunction process-sip (?args)
 "test dns and proxy server for sip"
 (bind ?result "NA")
 (bind ?result (self-diagnosis void))
 if (eq ?result "ok") then
 (bind ?result (dns-connection other))
 if (eq ?result "ok") then
 (bind ?result (proxy-connection void)))
Implementation

Sarnoff 2009 (Princeton, NJ)
http://wiki.cs.columbia.edu/display/res/DYSWIS
7DS and opportunistic networks: exploring networks beyond the Internet

with Suman Srinivasan, Arezu Moghadam

Sarnoff 2009 (Princeton, NJ)
Contacts are
- opportunistic
- intermittent

802.11 ad-hoc mode
BlueTooth

Sarnoff 2009 (Princeton, NJ)
Web Delivery Model

- 7DS core functionality: Emulation of web content access and e-mail delivery

Sarnoff 2009 (Princeton, NJ)
Search Engine

- Provides ability to query self for results
- Searches the cache index using Swish-e library
- Presents results in any of three formats: HTML, XML and plain text
- Similar in concept to Google Desktop

Sarnoff 2009 (Princeton, NJ)
Email exchange

Sarnoff 2009 (Princeton, NJ)
BonAHA framework

Node 1

- key11 = value11
- key12 = value12
- key13 = value13
- key14 = value14

Node 2

- key21 = value21
- key22 = value22
- key23 = value23
- key24 = value24

[1] node1.register()

[2] node1.get(key13)

[3] data = node1.fileGet(value13);

BonAHA
[CCNC 2009]

Sarnoff 2009 (Princeton, NJ)
Bulletin Board System

Written in Objective-C, for iPod Touch
Local Microblogging

Sarnoff 2009 (Princeton, NJ)
Generic service model?

Opportunistic Network Framework – get(), set(), put(), rm()

ZigBee  BlueTooth  mDNS/DNS-SD  DHTs?  Gnutella?

Sarnoff 2009 (Princeton, NJ)
Conclusion

• Abandon notion of a clean-slate next-generation Internet
  – that magically fixes all of our problems
• Need for good engineering solutions
  – with user needs, not (just) vendor needs
• Research driven by real, not imagined, problems
  – factor 10 problems: reliability & OpEx
  – more reliability and usability, less sensor networks
• Build a 5-nines network out of unreliable components
• Make network disruptions less visible
• Transition to “self-service” networks
  – support non-technical users, not just NOCs running HP OpenView or Tivoli

Sarnoff 2009 (Princeton, NJ)