Future Internet

Revisiting the Internet architecture?

Prof. Anja Feldmann, Ph.D.
Deutsche Telekom Laboratories
TU Berlin
The "Internet"

From Computer Desktop Encyclopedia
Reproduced with permission.
© 2000 The Computer Museum History Center
The “Internet”: What is it

- Visualization (2002)
  - ~ 535,000 Nodes
  - > 600,000 Links

- Social phenomena
  - Cyperspace
  - Changing/redefining communication
    - Human to human, human to computer, ....
The “Future Internet”?

- Information and Media Society
  - Everyone generates content
  - Sensors and cameras everywhere
  - New distribution channels

- Challenges
  - Verification
  - Fusion/filtering of information
  - Situation adaptation
  - Incentives, trust, privacy
What makes the Internet so sexy

- Applications can be deployed by anybody that is connected to the Internet
  (Fundamentally different to the Telephone world)
- Multi-service network:
  Everything over the Internet
TCP/IP protocol structure

Application (Processes)

Transport (Hosts)

Network

Link

TCP
UDP
IP

Telnet
FTP
HTTP
SMTP
DNS

FDDI
ATM
Tokenring
Ethernet
What makes the Internet so sexy

- Applications can be deployed by anybody that is connected to the Internet (Fundamentally different to the Telephone world)

- Multi-service network:
  - Everything over the Internet

  - Every application protocol over IP
  - IP over any network technology
Internet design goals

(in decreasing order of importance)

- Connect existing networks
- Survivability
- Support multiple types of services
- Must accommodate a variety of networks
- Allow distributed management
- Allow host attachment with a low level of effort
- Be cost effective
- Allow resource accountability
Today’s Internet

Data plane

- HTTP
- SMTP
- POP3
- IMAP
- XMPP
- NTP
- Telnet
- FTP
- IRC
- XDR
- ASN.1
- XML
- SMB
- RPC
- ASP
- RTP
- TCP
- UDP
- IPv4
- MPLS

802.3
802.11
802.16
ATM
802.3x
802.11a
802.16e
SDH
PDH
xDSL
ISDN
FDDI
Token Ring
Today’s Internet: Architectural limits

- Trust assumptions
  - Internet assumes cooperation

- Competition
  - Original Internet assumed no commercial considerations

- Edge diversity
  - Original Internet is host-centric
  - Ignores mobility, sensors, ...

- Network services
  - Original Internet exposes limited information
  - Limits new services
  - Limits network management
  - Almost no changes in the network core

- Designed to be a open, cooperating system
- Focus on data plane
Today’s Internet: Challenges

- Heterogeneity any which way you look
  - Users, applications, hardware, traffic
- An immense moving target
- Highly interacting systems
  - Temporal: between users, hosts and networks
  - Spatial: among different components
  - Vertical: across different networking layers
Why rethink the Internet architecture

- **Reliability and availability**
  - E-Commerce increasingly depends on fragile Internet
  - Debuggability

- **Security**
  - Known vulnerabilities lurking in the Internet
  - Addressing security has a significant cost

- **Scale & Diversity**
  - Cyberspace (everything is networked)

- **Support for new applications/services**
  - Mobility / Quality of service
  - High speed connections to the home

- **Economics**
  - Cost-effectively
  - Business models

❖ All are control plane issues!
Rethinking the Internet architecture

- Explore alternative architectures

- Approach
  - Incremental
    - Apply point-solutions to the current architecture
  - Clean slate design (CSD)
    - Start from scratch

- Advantage CSD
  - No limitations: enables rethinking of the network and service architecture
  - Architecture not intrinsic
  - Experiments and failures are possible
How to get there?

- How to determine that one has a good new architecture?
  - Paperware? No
  - Built, evaluated, used? Yes

- Approach:
  - Experimental facility
  - Research into new architectures

- Benefit:
  - Intellectual challenge: uncover otherwise ignored system aspects
  - Research how to build/operate an experimental facility

☞ Go beyond point solutions
Clean slate design: Drivers

- Technical
  - Virtualization techniques
  - Cloud computing / networking
  - Significant computational resources in the network
  - Fast packet forwarding hardware, e.g., OpenFlow

- Starting points
  - PlanetLab / OneLab
  - Geant2/Internet2
  - Emulab
  - Vini
  - ...
Future Internet: Sample initiatives and projects

- Future Internet network design – FIND [http://find.isi.edu/](http://find.isi.edu/)
- Clean Slate Program (Stanford University)
- Groupe de Reflexion Internet du Futur (France)
- Super Janet funded by EPSRC (UK)
- Internet del Futuro (SP)
- ANR (France)
- G-LAB funded by BMBF (Germany)
- AKARI Project (Japan)
- it839/u-it839 (Korea)
- NICTA (Australia)
- CNGI Project (China)
Revisiting traffic characteristics: Why?

- Constantly changing
- Basis of most architectural changes

- Residential broadband access popular/widespread
- Differs from well-studied campus and enterprise traffic
  - Not subject to acceptable-use policies
Usage changes: „Killer application?“

- WWW and the Internet
  - 1993: ... Hardly any WWW traffic on the Internet
  - 1994: ... About 10% of total Internet traffic is WWW
  - 95/96: ... Up to 60-70% of overall Internet traffic is WWW
  - ...????????...
Application mix?
However: MWN traffic by port
(24 hours of traffic to/from MWN clients in 2006)

<table>
<thead>
<tr>
<th>Port</th>
<th>% Conns</th>
<th>% Success</th>
<th>% Payload</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web</td>
<td>80</td>
<td>70.82%</td>
<td>68.13%</td>
</tr>
<tr>
<td>445</td>
<td>3.53%</td>
<td>0.01%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Web</td>
<td>443</td>
<td>2.34%</td>
<td>2.08%</td>
</tr>
<tr>
<td>SSH</td>
<td>22</td>
<td>2.12%</td>
<td>1.75%</td>
</tr>
<tr>
<td>Mail</td>
<td>25</td>
<td>1.85%</td>
<td>1.05%</td>
</tr>
<tr>
<td>1042</td>
<td>1.66%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>1433</td>
<td>1.06%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>135</td>
<td>1.04%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>&lt; 1024</td>
<td>83.68%</td>
<td>73.73%</td>
<td>79.05%</td>
</tr>
<tr>
<td>&gt; 1024</td>
<td>16.32%</td>
<td>4.08%</td>
<td>20.95%</td>
</tr>
</tbody>
</table>
Application mix - today?

*Based on limited dataset using payload inspection
Outline

- Revisiting traffic characteristics
  - Data sets
  - Dominant characteristics
    - Application usage
    - HTTP usage
    - NNTP usage
    - Performance/path characteristics
- Revisiting ISP – application relationship
- Revisiting Routing
- Revisiting Network structure
- Revisiting Splitting control and forwarding
Outline

- Revisiting traffic characteristics
  - Data sets
  - Dominant characteristics
    - Application usage
    - HTTP usage
    - NNTP usage
    - Performance/path characteristics
- Revisiting ISP – application relationship
- Revisiting Routing
- Revisiting Network structure
- Revisiting Splitting control and forwarding
Data sets (1)

- Large European ISP (>10M customers in total)
- Anonymized packet level traces
  - Covering >20,000 DSL customers
  - One urban area
- Overview of packet level traces
  - 14 x 90min; twice per day over 1 week in Aug 2008
  - 24hr in Sep 2008 (>4TB)
  - 24hr in Apr 2009 (>4TB)
- Bro Intrusion Detection System for analysis
Data sets (2)

- Anonymized DSL session traces
  - DSL connect / disconnect times
  - Anonymized line-card ID
  - Access bandwidth
  - Augments packet data

- Overview of session traces
  - One for each packet trace
  - 10 day in Feb 2009
Outline

- Revisiting traffic characteristics
  - Data sets
  - Dominant characteristics
    - Application usage
    - HTTP usage
    - NNTP usage
    - Performance/path characteristics
- Revisiting ISP – application relationship
- Revisiting Routing
- Revisiting Network structure
- Revisiting Splitting control and forwarding
Methodology

- Using Bro's Dynamic Protocol Detection (DPD)
  - Protocol semantics and/or
  - Signatures
- 85% of bytes classified
- Another 3.6% on well-known ports
- No dominant day-of-week effects

Verification
- With NetFlow data (port based)
- With commercial Deep Packet Inspection (DPI) system at different location
Application usage per hour
Application usage per b/w

percent

access bandwidth [Kbps]

1200  2300  3500  6500  17000

unclassified
well-known
other DPD
NNTP
eDonkey
BitTorrent
HTTP
Key results

- HTTP dominates?: 57% of bytes
- P2P less than 14%
- Unclassified: 11%
- Other significant protocols
  - NNTP 2–5%
  - Streaming (non-HTTP) 5%
  - Voice-over-IP 1.3%
- Port based classification works well for non-P2P protocols

?Erman et al. found very similar results in cotemporaneous work presented at WWW '09
Outline

- Revisiting traffic characteristics
  - Data sets
  - Dominant characteristics
    - Application usage
    - HTTP usage
    - NNTP usage
    - Performance/path characteristics
- Revisiting ISP – application relationship
- Revisiting Routing
- Revisiting Network structure
- Revisiting Splitting control and forwarding
Motivation & methodology

- Why is HTTP so popular (again)?
  - HTTP offers popular high-volume content?
  - HTTP as transport protocol for other applications?
- Anonymized HTTP headers extracted via Bro
- Determine content-type
  - Content-type header
  - Libmagic
- Second level domain (from Host header)
- User-Agent header
Key results

- Popular content-types by volume
  
  ![Bar chart showing content types](chart.png)

  - Flash-video: 25.2%
  - RAR: 14.7%
  - Image: 11.5%
  - Video: 7.6%
  - Other: 23.4%
  - Unclass.: 17.6%

- Domain popularity:
  - One-click-hoster is top domain: 15% of HTTP bytes
  - Video portals (using flash-video) follow

- No significant hiding / tunneling via HTTP

- HTTP dominance due to popular high-volume content
Outline

- Revisiting traffic characteristics
  - Data sets
  - Dominant characteristics
    - Application usage
    - HTTP usage
    - NNTP usage
    - Performance/path characteristics
- Revisiting ISP – application relationship
- Revisiting Routing
- Revisiting Network structure
- Revisiting Splitting control and forwarding
Usenet – NNTP

- Exchange of news/messages
  - Subsumed by forums, wikis and blogs
  - Said to be outdated and only used by “geeks”
  - Most servers do not allow binary content and have short retention times.

- What has changed?
  - Fee-based NNTP server operators, e.g., UseNeXT or GigaNews
  - 99 % NNTP volume is binary
  - Competes with One-Click-Hosters as client/server based alternative for file-sharing
Outline

- Revisiting traffic characteristics
  - Data sets
  - Dominant characteristics
    - Application usage
    - HTTP usage
    - NNTP usage
    - Performance/path characteristics
- Revisiting ISP – application relationship
- Revisiting Routing
- Revisiting Network structure
- Revisiting Splitting control and forwarding
TCP options / performance

- Window scaling
  - Approx. 50% of host advertise scaling; most non-zero
  - Maximal advertised window often 64KB
- Loss/reordering in 10% of connections
- Bandwidth-delay product > max. rcv. window
  - Affects 44% of connections with >50KB volume (downstream)
- Most lines only use small fraction of bandwidth
Round-trip-times (RTT)

- Assessed during TCP handshake

- Local component dominates (DSL interleaving)
- Median: 74ms
- 99th perc: 1328ms
- Wireless equipment can cause significant delays
Achieved flow throughput
Achieved throughput

- Most lines only use small fraction of bandwidth
- Throughput by application and flow
  - HTTP, NNTP have order of magnitude higher throughput than P2P
- Mean number of parallel flows
  - P2P has 5 times as many as HTTP
  - P2P and NNTP similar
Summary – Residential traffic study

- High IP address churn (4% assigned > 10 times)
- HTTP dominates traffic: >57%
  - P2P only 14%
  - NNTP noticeable
- Flash-video (video portals) most popular in HTTP: >25%
  - RAR-archives (One-click-hosters): >14%
- Performance
  - DSL bandwidth in general not fully utilized
  - Window advertisements might limit performance
  - Local RTT component dominates (DSL interleaving)
Outline

- Revisiting traffic characteristics
  - Data sets
  - Dominant characteristics
    - Application usage
    - HTTP usage
    - NNTP usage
    - Performance/path characteristics

- Revisiting ISP – application relationship
- Revisiting Routing
- Revisiting Network structure
- Revisiting Splitting control and forwarding
Internet and traffic engineering

Source: Arbor Networks
2009
Internet and traffic engineering

Traffic Engineering:
Adjust routing or peering, dimension the network

Offline Process

Source: Arbor Networks
2009
The new Internet

→ New core of interconnected content and consumer networks

Source: Arbor Networks 2009
The new Internet

→ New core of interconnected content and consumer networks

Source: Arbor Networks 2009
The new Internet

Moving Target I: Popular Applications

New core of interconnected content and consumer networks

Source: Arbor Networks 2009
The new Internet

Moving Target I: Popular Applications
Moving Target II: Bottlenecks

→ New core of interconnected content and consumer networks

Source: Arbor Networks 2009
The new Internet

- New core of interconnected content and consumer networks
- ISPs lost control of their traffic

Source: Arbor Networks 2009
The new Internet

Global Internet Core

Regional / Tier2 Providers

Moving Target I
Popular Application

→ New cont

→ ISPs
Challenge

Content-aware Traffic Engineering

ISPs re-gain control of their traffic by biasing host selection
Grand challenge

Content-aware Traffic Engineering

ISPs re-gain control of their traffic by biasing host selection

→ Online
→ No routing re-configuration
→ No additional investments
→ Possible reduction of operational cost
→ Potential negotiation tool
Roadmap

Measurements

System Design & Deployment

Field Test

ISP-Application Collaboration
Residential traffic

→ Recall: HTTP is responsible for around 60% of total traffic
→ This trend should continue (flash video, cloud applications, datacenters)

Source: Maier et al, IMC'09
Content Distribution and DNS

1. Client queries DNS.
2. DNS query forwarded to Internet Service Provider (ISP).
3. ISP returns DNS reply to Provider DNS.
4. Provider DNS returns DNS reply to Client.
5. Client receives DNS reply.

External DNS
Provider DNS
Internet Service Provider (ISP)
DNS Reply Aggregator
Host
Reply anatomy

→ Requesting a photo from Facebook

$ dig photos-h.ak.fbcdn.net
<<>> DiG 9.7.0-P1 <<>> photos-h.ak.fbcdn.net

;; QUESTION SECTION:
photos-h.ak.fbcdn.net. IN A

;; ANSWER SECTION:
photos-h.ak.fbcdn.net. 6099 IN CNAME photos-d.ak.facebook.com.edgesuite.net.
photos-d.ak.facebook.com.edgesuite.net. 20492 IN CNAME a998.mm1.akamai.net.
a998.mm1.akamai.net. 7 IN A 62.41.85.74
a998.mm1.akamai.net. 7 IN A 62.41.85.90
...

2nd Level Domain → Application

Redirection → Content Provider
Consolidation of content

Top-10 applications or content providers are responsible for around 50% of the HTTP traffic.
Diversity of paths

More than 60% of the HTTP traffic can be downloaded from at least 3 different locations.
PaDIS

Provider-aided Distance Information System
PaDfS

External DNS

Internet Service Provider (ISP)

Provider DNS

PaDfS

Client

Host

Full View of the ISP Network
PaDIS

External DNS

Internet Service Provider (ISP)

Provider DNS

PaDIS

Client

Host

Content can be downloaded from any eligible host!
PaDIS

External DNS

Provider DNS

Service Provider (ISP)

Full View of the ISP Network

Client

Host1
Host2
Host3
Host4

PaDIS
PaDIS

External DNS

Host1
Host2
Host3
Host4

Provider DNS

Service Provider (ISP)

Full View of the ISP Network

Client

Host

PaDIS

1 2 3 4 5 6
Network load balancing
Network load balancing

Host A

Host B

Host C

Client
Network load balancing
Path diversity

- Rebalance of traffic to less congested links
- Top-10 content providers and applications
- Reduction up to 30% on congested link
- 5-10% reduction in total traffic
- No increase in path length
- Improve locality of HTTP traffic from 25% to 50%
Roadmap
Improving content access time

Case study: CDN
Improving content access time
Case study: One-Click Hosters
Roadmap

- Measurements
- System Design & Deployment
- Field Test
- ISP-Application Collaboration
ISP applications collaboration

1. Client
2. Internet Service Provider (ISP)
3. PaDIS
4. Provider DNS
5. External DNS
6. ISP applications collaboration
7. Host
ISP-applications collaboration

External DNS

Provider DNS

Internet Service Provider (ISP)

PaDIS

Client

Host
ISP-applications cooperation

Client

External DNS

Internet Service Provider (ISP)

Provider DNS

PaDIS

Host1
Host2
Host3
Host4

Host2
Host3
Host4

Host1
Host4
Host3
Host2

Host

1
2
3
4
5
6
7
7
Summary

- Alternative traffic engineering
  - Do not change the routing
  - Change the traffic matrix!

- Benefits
  - ISPs: Regain control of network traffic
  - User: Performance improvements
    - Win-win situation for ISPs and end-users
    - ISPs can share benefits with content and application providers

- PADIS
  - Simple and easy to implement
  - Prototype running
Outline

- Revisiting traffic characteristics
  - Data sets
  - Dominant characteristics
    - Application usage
    - HTTP usage
    - NNTP usage
    - Performance/path characteristics
- Revisiting ISP – application relationship
- Revisiting Routing
- Revisiting Network structure
- Revisiting Splitting control and forwarding
Exploring Alternative Architectures

HAIR: Hierarchical Architecture for Internet Routing
Routing scalability: Problems

- Routing table size growth
  - Multi-homing
  - Traffic engineering
  - Prefix disaggregation

- IP addresses usage
  - Locator within the Internet
  - Identifier for applications

Graph showing the growth of routing table entries from 1989 to 2008.
Routing scalability: Problems

- Churn: High update rates
  - Due to mobility
  - Due to global visibility
  - Due to „overuse“ of policy
Routing scalability: Current workarounds

Scalability issues
- large RT
  - static
  - dynamic
- high upd rate
  - static
  - dynamic

Consequences
- expensive TCAM
  - data plane
- high workload to maintain RT
  - control plane

Workarounds
- massive filtering
  - static
  - dynamic
- dampening
  - dynamic

Problems
- limited TE
- limited mobility
Approach

Key ideas

- Separation of locator/identifier function of IP address
  => separation of routing and location mapping

130.149.220.23
TU-Berlin
Approach

Key ideas

- Separation of locator/identifier function of IP address
  => separation of routing and location mapping

- **Hierarchy** for routing and location mapping
Approach

- Key ideas
  - Separation of locator/identifier function of IP address
    - => separation of routing and location mapping
  - **Hierarchy** for routing and location mapping

- Two components
  - Routing system based on locator
  - Mapping system to map an identifier to a locator
Hierarchical routing

- Network is organized in multiple levels
- Levels are separated by separators
- Routers only know the details about their level
Hierarchical routing: Internet

- Where do we have small separators?
- Internet structure
  - Core
    - Set of interconnected autonomous systems (ASs)
    - Tier-1, tier-2 ASs, ...
    - Transit ASs
- AS core
  - ~5000 ASs
- AS edge
  - ~30000 AS
- **AS core**
  - ~5000 ASs
- **AS edge**
  - ~30000 AS

Potential large separator

Potential small separator

Core

Enterprise Network

ISP1

ISP2

ISP3

Transit AS 1

Transit AS 2

Access Provider

Stub AS
Hierarchical routing: Internet

- Where do we have small separators?
- Internet structure
  - Core
    - Set of interconnected autonomous systems (ASs)
    - Tier-1, tier-2 ASs, ...
    - Transit ASs
  - Intermediate
    - Stub ASs, e.g., metropolitan area networks
    - Enterprise networks
    - Content distribution networks
  - Edge
    - Local area networks
Hierarchical routing: Internet

- Separator size
  - Core -> Intermediate
    - Stub ASs, e.g., metropolitan area networks: < 10 links
    - Enterprise networks: < 10 links
    - Content distribution networks: < 1000 links
  - Intermediate -> Edge
    - Local area networks: < 10 links

- Terminology
  - Core / WAN
  - Intermediate / MAN
  - Edge / LAN
  - Separator / Attachment point (AP)
Hierarchical network

- Example: Three levels of hierarchy
  - Routing via intermediate points - the separators
    - specify attachment points
  - WAN APs: WAP
    - Provider access links
  - MAN APs: MAP
    - Firewalls
Sending a packet

- Routing via intermediate access points
  - Mapping service: resolve identifier to locator
  - 3 locator parts: WAP|MAP|ID
Routing scalability

- **Core**
  - Routing based on WAPs
  - Stable business relationships
  - Almost no churn
  - Aggregatable addresses
  - Common routing protocol (e.g., BGP)

- **Intermediate (smaller ISPs/enterprises)**
  - Routing based on MAPs
  - Separate addresses and routing
  - Local changes → local impact

- **Edge (e.g., Ethernet LAN)**
  - Standard L2 switching
Mapping system

- Design requirements
  - Scales with number of hosts
  - Fast response times
  - Easy to update

- Approach
  - Clients are responsible
  - Hierarchical design
    - Global DHT or DNS like system
      - For each identifier: pointer to MMS
      - WANs contribute resources
    - MAN mapping service (MMS)
      - Stores locators for attached nodes
      - Provided by MAN(s)
Mapping identifiers to locators

Steps
- Client queries
  - Global DHT
  - MMS

To avoid lookups
- Use caching
- Include source locators in packet
- ...

Global DHT/MMS
- Can store multiple alternatives

Failure recovery
- Via multiple alternatives
Discussion (1)

- **Scalability**
  - Hierarchical routing AND mapping system
  - Updates are localized => low update rates
  - No manual configuration

- **Mobility: local visibility of changes**
  - Intra-MAN mobility: frequent
    - Updates restricted to MMS
  - Inter-MAN mobility: less frequent
    - Update global DHT (fast)
    - Move locators to new MMS
Discussion (2)

- Multihoming
  - Inherent support: APs exposed to routing system

- Multipath
  - Use multiple locators in parallel

- Inbound traffic engineering
  - Per-host basis
  - MANs/MMS have control

- Migration path
  - To support legacy hosts
Migration via NATs/Firewalls: Sending

- Firewalls/NAT act as MAPs
- Legacy packet arrives from LAN
  - Treat dst address as dst ID
  - Resolves locator for ID
  - Add source locator to packet header
  - Encapsulate original packet and sends it
Migration: Receiving

- WAP strips encapsulation
- MAP/NAT strips the second layer
  - May get the mapping for the source locator
- Packet is routed onward
What’s different here

- Routing hierarchy based on structure of the Internet
  - Smaller table sizes
  - Lower update rates
- Mapping service is hierarchical
  - With local control and responsibility
- Hosts are responsible for obtaining mapping
- Incremental deployment possible
Lessons learned

- **Main goals**
  - Scalability
  - Support for multi-homing, TE, mobility, etc.
  - Smooth migration, support for legacy hosts

- **Key ideas**
  - Separation of locator/identifier function of IP address
  - Hierarchical routing and location mapping scheme

- **Two components**
  - Routing system based on locator
  - Mapping system to map an identifier to a locater
Outline

- Revisiting traffic characteristics
  - Data sets
  - Dominant characteristics
    - Application usage
    - HTTP usage
    - NNTP usage
    - Performance/path characteristics
- Revisiting ISP – application relationship
- Revisiting Routing
- Revisiting Network structure
- Revisiting Splitting control and forwarding
Enabling Alternative Architectures

Network Virtualization Architecture: Proposal and Initial Prototype
Network virtualization scenarios

- Virtual network
  - Resource isolation
  - Different architecture/protocol per virtual network
    - Does not have to be IP protocol
    - Some with some QoS and security
  - Expose network components to applications and services
    - Overcome Internet impassé
  - Dynamic
    - New ones will come and old ones will go
    - Migration / Expansion / Contraction
  - Multiple networks in parallel == diversity

- Simplify network management and service offerings

- Virtual networks != VPN – VPN is just a service!
  Virtual networks != P2P network – P2P is just an overlay
Virtualization: Vision

Virtualization Management
Provisioning of Virtual Networks
(on-demand instantiation of virtual networks)

Infrastructure

Virtualized Substrate

Virtualization of Resources
(partitioning of physical infrastructure into “slices”)

Virtualization Management

Virtual Network

Virtual Network
Benefits of virtualization

- Overcome *ossification* of network core
  - Isolation as enabler for new technologies
    - Traditional: IPv6, multi-cast, ...
    - CSD: novel network architectures
  - Deployment of *innovative* products
  - Network diagnosis

- Efficient utilization of resources
  - Migration of devices (such as routers)
    - similar to server virtualization
  - Traffic load balancing (“migration” of links)

- New business opportunities
  - Sharing of physical resources
    (e.g., T-Mobile UK and 3 UK)
Virtual network – terminology

- Physical node
- Virtual links
- Infrastructure provider link
- Substrate nodes
- Virtual nodes
- Infrastructure provider A
- Infrastructure provider B
- Legacy nodes
Virtual node – terminology

- VNet User
- Physical Resources
- Substrate node control
- Physical Link
- Infrastructure Provider
Roles in the Internet

Traditional roles:
- Service providers (SP)
  - Google, World of Warcraft, ...
- Internet Service Providers (ISPs)
  - Deutsche Telekom, AT&T, ...

Recently:
- Physical infrastructure provider (PIPs)
- Bit-pipe providers
- Service providers (SP)
Roles with network virtualization

- Service Provider
- VNet Operator
- VNet Provider

Infrastructure provider
Tasks: Birdseye view

- Physical Infrastructure Provider (PIP)
  - Provides Virtual Resources + Resource Control Interface

- VNET Provider (VNP)
  - Assembles virtual networks
  - Intuitively: provides layer of indirection

- VNET Operator (VNO)
  - Operates, controls, manages virtual networks
    (e.g., comparable to NOC)

- Service provider (SP)
  - Offers the service
Physical Infrastructure Provider

- **Services:**
  - Provides Virtual Resource
  - Resource Control Interface

- **Input:**
  Requests for virtualized resources from VNP

- **Task:**
  - Creation of topology (constituents)
  - Pointers to virtual resources
  - Resource Control Interface
    - Virtual Node Bootstrapping
    - Interconnection
VNET Provider (VNP)

- **Service:**
  - Instantiated virtual networks (interconnected virtual nodes with bootstrapping environment)
  - Handles contracts with PIP and VNO.
- **Input:** Abstract request for VNet
- **Task:**
  - Identify appropriate PIPs
  - Negotiate contracts
  - Partition network topology and acquire partial VNETs and Control Interfaces
  - Assemble virtual networks and control interfaces from partial VNETs provided by PIPs
VNET Operator (VNO)

- **Service:**
  - Bootstraps, operates, controls, manages fully instantiated virtual network
  - Operates on virtual resources, identified by Identifiers (not Locators)

- **Input:** Interconnected virtual network
- **Task:** operating, managing of virtual network
Control interfaces
VNET signaling and control
Spec includes policy (e.g. pref IP) price etc.
VNET Operator

NYC 1

London 2

Spec includes policy (e.g. pref IP) price etc.

VNET Provider

Management VNP

Management PIP1

Management PIP2

PIP
VNET operator

NYC 1

London 2

VNET Provider

Spec includes policy (e.g. pref IP) price etc.

Management VP

console1 console2

Management IP2

console1

Management IP1

console1

IP

VM

C

128
Lessons learned

- Isolate tasks => business opportunities
  - E.g.: Magnitude of the investment cost
    - AT&T plans to invest 17–18 Bn $ in 2009 compared to a revenue of 124 Bn $ in 2008
    - Deutsche Telekom plans to invest 8.7 Bn Euro compared to revenues of 62 Bn Euro in 2008
    1% is substantial!

- Don’t forget control interfaces
- Interprovider issues are tricky
- Indirection and resource isolation are great tools
Case study:
Locating performance problems aided by network virtualization
Network debugging: Motivation

How do you test an update to the configuration or software of your system?

- simulation
- scalable
- cheap
- accuracy?

- testbed
- may be more accurate
- but user behavior?
- expensive to run
- on large scale?
- longtime?
Network diagnosis/debugging

Problem:
Implementation/configuration issue surface in large-scale, long-term deployments with real user traffic

Goal:
- Do not change network under test
- Avoid probe effect

Diagnosis methods:
- Instrumentation
- Regression tests
Instrumentation

- Pair production VNet with monitoring VNet
- Copy all/selected packets to monitoring VNet
- Processing is accounted to monitoring VNet
Regression testing - Shadow Vnet

Input dist'ed to Vnet 1.0 and Vnet 1.1

Output of Vnet 1.0 dist'ed to ext entities

Substrate
VNet running V1.0
VNet running V1.1
Control Vnet
Regression testing – Shadow

Vnet

- Run VNet1.0, VNet1.1 monitoring VNet
- Distribute external input to both VNet1.0 and VNet1.1
- Ctrl compares output behavior of VNet1.0 and VNet1.1 for semantic equality
- Only output of VNet1.0 is distributed to external entities
Example: VoIP with background load

- Phase 1: Minimal background traffic
- Phase 2: Background traffic increases
- Phase 3: Start ShadowVNet: VNET B
- Phase 4: Enable QoS in VNET B
- Phase 5: VNET B becomes operational
Example: VoIP with background load

User perceived quality is restored when the ShadowVNet is activated
Lessons learned

- New network debugging features
  - Instrumentation
  - Regression tests
  - Distributed debugger

- Goals
  - To not change network under test
  - Avoid probe effect

- Solution: Network virtualization
  - Isolation
  - Accounting of resources
Outline

- Revisiting traffic characteristics
  - Data sets
  - Dominant characteristics
    - Application usage
    - HTTP usage
    - NNTP usage
    - Performance/path characteristics
- Revisiting ISP – application relationship
- Revisiting Routing
- Revisiting Network structure
- Revisiting Splitting control and forwarding
Case Study: OpenFlow

An exciting new technology for separating hardware and software
OpenFlow - An enabler for open control of the network

- OpenFlow moves control path in each switch/router (middlebox) to an external controller, which makes policy decisions at the flow level (flow is defined by Layer 2, Layer 3 and/or Layer 4 header fields).
- With OpenFlow, each switch speaks a separate control protocol with an external controller over SSL.
Example service using OpenFlow
Per-flow policy and QoS on demand

- Interface between the OpenFlow network and the client enables, e.g., on demand
  - Constraints on route
  - Temporary upgrade in service/QoS

QoS boost: Increase bandwidth for this specific flow.
OpenFlow is already supported on routers/switches.

- Juniper MX-series
- NEC IP8800
- WiMax (NEC)
- HP Procurve 5400
- Cisco Catalyst 6k
- PC Engines
- Quanta LB4G
An OpenFlow based Router

Taking advantage of
+ OpenSource Routing Software
+ Inexpensive Switch Hardware
OpenFlow based router: FIBIUM. Evaluation of open-source routers.

Levers and current situation

- ISPs spend significant annual CAPEX for routers etc.
- Currently a quasi-monopoly vendor market for backbone routers and routers in huge data centers.
- Modularization/standardization of router components/open-source software may open the market
  - Similar to Linux
  - Industry standards for blade servers

Our Approach

- Understand if low cost switches with open-source software can be suitable replacement for high-cost routers.
- Build and evaluate prototype router using low-cost components and open-source software.

The OpenFlow might be a low cost, flexible alternative
OpenFlow based router: FIBIUM.

Today’s inflexible routers.

- Proprietary software
- Limited features customization
- No access to datapath
- Optimized but not programmable hardware

Current IP routers

- Carrier-grade open-source based routers, e.g., Vyatta, IPInfusion, Quagga
- High-performance and inexpensive Ethernet layer-3 switches with OpenFlow support, e.g. HP, NEC, Cisco

Observations
OpenFlow based router: FIBIUM.
Divide and conquer.

**Principles**
- Decouple routing and datapath
- Combine inexpensive OpenFlow-enabled switch with open-source routing software on commodity hardware

**Observations**
- Current switches have sufficient datapath performance
- Switch control logic is limited
- Current commodity hardware better than route controllers on routers
OpenFlow based router: FIBIUM.
From concept to reality.

- Leverages OpenFlow interface of switch:
  - RouteVisor programs the switch
  - Route cache management ensures good fast path performance despite limited switch control logic
  - Slow path handled by PC

FIBIUM

RouteVisor

- Ensures that switch and PC combination appears as a router to the outside world
  - Interface between route control logic on PC and switch
  - Collects traffic statistics from switch and updates data path on switch
OpenFlow based router: FIBIUM. Prototype.

Test lab
- Test: 2 HP switches, commercial routers (Cisco and Juniper) and Linux routers
- RouteVisor tasks:
  - Switch configuration
  - Handle OSPF and BGP messages
  - Route Cache management

Prelim. performance evaluation
- Control plane: 100K BGP updates per minute
- Communication channel between PC and switch:
  - Traffic statistics: 100K per second
  - Switch data path updates: 1000 modifications per second
Lessons learned

- Open interface == new opportunities
- Flexibility of software wins over closed systems
- Example: FIBIUM
  - Control: Open source routing software
  - Data path: Hardware
Rethinking the Internet architecture

- Explore alternative architectures

- Approach
  - Incremental
    - Apply point-solutions to the current architecture
  - Clean slate design (CSD)
    - Start from scratch

- Advantage CSD
  - No limitations: enables rethinking of the network and service architecture
  - Architecture not intrinsic
  - Experiments and failures are possible
CSD: Reshaping the Internet

Impact on users:
- Ease of access to relevant information
- New control plane with new capabilities
- Easy to introduce new applications with new features
  - Security, mobility, quality of service

Impact of new economic models:
- New interfaces between providers (network/service)
- New value-chain and new roles for providers
- Open interfaces may enable new ecosystems of business alliances

Impact on society:
- Information society

Impact on operators:
- Easier network management