What’s in a name?

- Any problem in Computer Science can be solved by another layer of indirection
  
  — David Wheeler
  (Chief EDSAC Programmer)

- We name objects in order to:
  - Abstract away details of location, access, user interface
  - Interpose another layer of control, to allow relocation, e.g.

- Naming is a choice
  - To share a common name is to share trust in its meaning
What’s Internet Scale About, Anyway?

This workshop series is dedicated to the proposition that successful Internet applications require more than scalable algorithms:

- They must scale across time — longevity
- They must scale across space — latency
- They must scale across organizations — liability
- Economic, Political, and Social criteria are just as critical

Internet Scale is about more than large numbers...
Powers of Ten illustrates the different rules governing different scales of existence.

- Meteorology, Biology, Chemistry, Quantum Mechanics...
- Geology, Astronomy, Cosmology...

Yet the same rules apply, too!

- Physics is scale-invariant

Let’s try ‘zooming in’ on an Internet-scale name...

A film from the office of Charles & Ray Eames, 1977, running time 8:47
<table>
<thead>
<tr>
<th>URI</th>
<th>Uniform Resource Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolved by</td>
<td>Web Browser</td>
</tr>
<tr>
<td>Hierarchical</td>
<td>Left-to-Right</td>
</tr>
<tr>
<td>Format by</td>
<td>IETF RFC 1630 (6/94)</td>
</tr>
<tr>
<td></td>
<td>IETF RFC 2396 (8/98)</td>
</tr>
<tr>
<td>Entries by</td>
<td>Server Administrator</td>
</tr>
<tr>
<td>Internationalization</td>
<td>US-ASCII (UTF-8)</td>
</tr>
<tr>
<td>Number</td>
<td>$10^{10} +$</td>
</tr>
<tr>
<td>Lifetime</td>
<td>$10^1 - 10^8$ sec</td>
</tr>
</tbody>
</table>

- Replaced complex recipes for fetching network information with a single string
- Composed from four namespaces
  - Scheme, domain, port, path
- Can also have username, password

Browsers resolve URIs to Web Pages

![Web browser interface with United Airlines itinerary](image)

**United Connection on the Web**

- Main Menu
- User Profile
- Help
- Feedback
- What's New
- FAQ

### United Airlines 1720
- **Boeing 767-200**
- **Aug 19 07:41 departs SNA**
- **Aug 19 13:40 arrives ORD**

- **Choose Your Seats**

- **Total Airfare (including taxes)** USD 1019.00

- **Fares are not guaranteed until ticketed.**

- **Please read these hints if this airfare seems too high.**

---

19 August 1999

Internet Scale Namespaces: A Survey
<table>
<thead>
<tr>
<th>URI Scheme</th>
<th>Resolved by</th>
<th>Atomic</th>
<th>Format by</th>
<th>Entries by</th>
<th>Internationalization</th>
<th>Number</th>
<th>Lifetime</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Web Browser</td>
<td>ASCII</td>
<td>IETF RFC 1738</td>
<td>IANA Registry</td>
<td>none</td>
<td>$10^{1+}$</td>
<td>$10^8$-$10^9$ sec</td>
</tr>
<tr>
<td></td>
<td></td>
<td>string</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Quickly identifies information-access system which can resolve the URI path
- Resolves to IANA assigned port numbers
  - Not injective: HTTP and IPP both at 80
- Can be an address, too, as with data:

<table>
<thead>
<tr>
<th>Scheme Protocol</th>
<th>RFC</th>
<th>Port</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTP File Transfer Protocol</td>
<td>1738</td>
<td>21</td>
</tr>
<tr>
<td>Telnet Interactive Sessions</td>
<td>1738</td>
<td>23</td>
</tr>
<tr>
<td>Gopher The Gopher Protocol</td>
<td>1436</td>
<td>70</td>
</tr>
<tr>
<td>HTTP Hypertext Transfer</td>
<td>2616</td>
<td>80</td>
</tr>
<tr>
<td>NNTP Netnews Transfer</td>
<td>977</td>
<td>119</td>
</tr>
<tr>
<td>WAIS Wide Area Inf. Svc</td>
<td>1625</td>
<td>210</td>
</tr>
<tr>
<td>Z39.50s Z39.50 Session</td>
<td>ANSI</td>
<td>210</td>
</tr>
<tr>
<td>Mailto Invoke mailer</td>
<td>821</td>
<td>25</td>
</tr>
<tr>
<td>Https (443), snews (563), ftps (990)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
  - Single-bit security flag
DNS
Resolved by DNS Protocol
Hierarchical Right-to-Left
Format by IETF RFC 883 (11/83)
Entries by ICANN-delegated registrar
Internationalization [A-Z][a-z][0-9]-
Number $10^8+$ (63/254 char limit)
Lifetime $10^7 - 10^8$ sec

Composed of hierarchical namespaces
- com (ICANN), united (NSI), www (United)
Uniqueness requirement forces political solutions: United Van Lines or Air Lines?
- Actually, neither: this domain is disputed
Resolved by an 13-rooted planetary tree

DNS Resolvers resolve Hostnames into Internet Addresses
National TLDs ISO-3166 two-letter codes
- Iceland this.is/keyword
- Montserrat linux.versus.ms

Original TLD intentions:
- .gov US Federal Gov’t
- .net Network service providers
- .int International treaty orgs

‘Localhost’ is a reserved name
Reverse lookups
- 213.21.195.128.arpa.in-addr
Competing global trademark registries
- RealNames, WHOIS (RFC 2345)

Urgent need to expand number of roots
- Allow several to masquerade as one
IP (Internet Protocol Address)
- Resolved by TCP/IP Stack
- Hierarchical Right-to-Left
- Format by IETF RFC 791 (9/81)
- Entries by IANA-delegated IP registry
- Internationalization none
- Number
- Lifetime

TCP/IP Stacks resolve Internet Addresses to MAC (physical) Addresses or next-hop Internet Addresses

Regional IP numbering registries
- Europe - RIPE, Asia- APNIC, US- ARIN
- Allocation Policy set by RFC 2050

Reserved ranges
- This network: 0.x.x.x
- Broadcast: 255.255.x.x (Class B)
- Multicast: 224.0.0.0 and up
- Reserved loopback address: 127.0.0.1

Sample Netnumbers circa 1981 (RFC 790)

<table>
<thead>
<tr>
<th>Internet Address</th>
<th>Name</th>
<th>Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>001.rrr.rrr.rrr</td>
<td>BBN-PR</td>
<td>BBN Packet Radio Network</td>
</tr>
<tr>
<td>002.rrr.rrr.rrr</td>
<td>SF-PR-1</td>
<td>SF Packet Radio Network</td>
</tr>
<tr>
<td>003.rrr.rrr.rrr</td>
<td>BBN-RCC</td>
<td>BBN RCC Network</td>
</tr>
<tr>
<td>004.rrr.rrr.rrr</td>
<td>SATNET</td>
<td>Atlantic Satellite Net</td>
</tr>
<tr>
<td>005.rrr.rrr.rrr</td>
<td>SILL-PR</td>
<td>Ft. Sill Packet Radio</td>
</tr>
<tr>
<td>007.rrr.rrr.rrr</td>
<td>CHAOS</td>
<td>MIT CHAOS Network</td>
</tr>
</tbody>
</table>

Composed of subnet and link numbers
- Class A, B, C and CIDR net mask prefixes
- Topological consistency of net ranges

Also demuxed by 16-bit TCP port number

Network Address Translators (NATs)
- fudge injectivity – address collisions poss.
MAC Media/Multiple Access Control
Resolved by LAN Address Res. Protocol
Hierarchical Org. Unique ID + device ID
Format by IEEE 802.3 &c
Entries by IEEE Registration Authority
Internationalization none
Number 2^48 = 10^{14}
Lifetime 10^8 - 10^{10} sec

Maps onto individual link endpoints (network stations)
Absolutely must be link-unique
  Analogous MACs for ATM, Token Ring
ARP uses a simple lookup table

ARP & RARP resolve Ethernet Addresses to/from Internet Addresses
Blocks of 4,096 are sold to Ethernet adapter manufacturers for $500, after a $1,250 “initiation” fee
What portions of this space are reserved?

Address Resolution Protocol, RFC 826, November 1982
  Announce own IP, request target IP’s MAC
Reverse ARP, RFC 903, June 1984
  Broadcast a request to get an IP address
PPP Address
- Resolved by Point-to-Point Prot. + modem
- Hierarchical
- Format by ITU E.164 (Bell, 1947)
- Entries by N. Am. Numbering Plan
- Internationalization

Phone number
- Lifetime  $10^{10}$
- $10^5 - 10^9$ sec

PPP Link driver itself operates over a phone circuit

Phone numbers hierarchically assigned
- Nation, Area, Exchange, Subscriber
- Absolutely must be world-unique

Indexed by Yellow and White Pages

Telephones resolve phone numbers into circuits

Networks can be countries, too: Iridium satellite phone subscribers get +8816

Phone numbers are represented in many common forms
- ITU form: +1-(626)-806-7574
- DNS form: 4.7.5.6.0.8.6.2.6.1.tpc.int
- URI form: phone://16268067574/

Tel:, fax: and modem: proposals, too

Geocoded: Madison 437 (archaic)

Reserved portions
- 555 testing & information services
- 800-855-xxx Teletype toll-free info (TDD)

Caller-ID (ANI) reveals source address

Reverse lookup possible too
Maps onto individual resources
- But representation on the wire may still depend on content language, media type, authorization, &c.
- Must be server-unique; may be aliased
- No versioning by default; can rot

Web Servers resolvePathnames into HTTP Representations (replies)
- Becomes BASE for resolving relative URLs
  - This resource identifier resolves to the HTML outline of a page that is completed with several subsidiary resources (graphics, sounds, style sheets)
- Can be a collection resource (DAV)
  - Supports enumeration, searching of directories
- Can have properties (DAV)
  - Such as Author, Words, Cost...
    - Which come from yet other property namespaces...
Filename
Resolved by
Hierarchical
Format by
Entries by
Internationalization
Number
Lifetime

Web Servers resolve path components into filenames
Operating Systems resolve filenames into inodes
Disk Drivers resolve inode into track+sector addresses
Disk Controllers resolve track+sector addresses into data blocks

Maps onto individual files or processes
Server typically rewrites the URL by substituting root, user directory, extension
Security and accounting controlled by OS, not necc. the web server’s control
PNR = Passenger Name Record
Resolved by Airline Distribution System
Atomic Alphanumeric picture string
Format by Length, pattern vary by GDS
Entries by Airline
Internationalization none
Number $10^8$ +
Lifetime $10^1 - 10^7$ sec
Maps onto individual reservations
  - Every booking and confirmation is kept until flight time
  - Resolves to an IATA + Airline ticket number
    - Permanent identifier lasts for years
Must be unique over its design lifetime

A Reservation Database Process resolves PNR keys into Reservation records
Talking to a process, not a bag of bits.
Interoperability standards are crucial for interline ticketing, but still fragmented by each GDS (Sabre, Apollo, Amadeus, etc).

GDS = Global Distribution System
Anatomy of a URI

User
  Application
    Transport
      Network
        Link

URI
  Browser
    Scheme
      Protocol Handler
    Hostname
      DNS Resolver
    Pathname
      HTTP Server

IP Address
  TCP Stack
    MAC Address
      Network Link
    Phone Number
      PPP Link

Port Number
  TCP Stack

Filename
  Server OS

Reservation
  Database Key

Inode
  Disk Driver

Track, Sector
  Disk Controller
A URI resolves to an HTTP Message

- Zooming further in, an HTTP response message uses several more namespaces
- Method Standards-track RFC
- Reply Code RFC; new IANA Registry
- Content-Type IANA Media-Type
- Content-Language ISO language codes
- Character-set IANA ref to ISO Charset
- ETag Uniquely identify the resource
- PICS label URI pointing to schema
  - Here, ‘Good Clean Fun’ specifies its own suds/density/color ratings namespace
- Digital Signature Hash of resource
  - Algorithm identifiers are URIs, too
  - But signing principals are another scale...
Principals resolve to People & Organizations

- Signing principals must use larger-scale names:
  - Lifetime of name validity is \( \gg \) duration of Web transaction
  - Social scope of name is \( \gg \) than just immediate parties
  - Typically also used across multiple applications

- Resolving any identifier onto the range of people and incorporated entities raises non-technical questions
  - Privacy – is the resolver function known to all? Breakable?
  - Trust – such identities are invariably intended to bind legally
  - Economic – injectivity creates scarcity (one-to-one map)
  - Politics – surjection could be compelled (universal IDs)
## X.500 Directory Hierarchy

- **Distinguished Name (DN)**
  - Common Name (`cn`)
  - Address (`street`)
  - Locality / Region (`l`)
  - State / Province (`st`)
  - Organizational Unit (`ou`)
  - Organization (`o`)
  - Country (`c`)

- **Examples**
  - `cn=Rohit Khare, o=4K Associates, c=US`
  - `cn=Rohit Khare, ou=Information and Computer Science, l=Irvine o=University of California, st=CA, c=US`
  - Took 10 years, CCITT vs ISO friction, too

- **X.509 Certificates & Revocation Lists**
  - Resolve DNs into public keys

- **Each component of a DN can be a Certification Authority (CA)**

- **Yields a pyramid-shaped trust structure, with increasingly liable, larger-scope organizations delegating central authority**

- **E.g. all https servers must buy certificates from a small number of roots, such as Verisign**

- **Role/authorization relegated to ‘extended attribute’ fields**
Pretty Good Privacy, by contrast, allows any ASCII string to represent a keyholder
- Typically, eMail Address(es)

In the beginning, there is the self-signed key

After verifying key ‘fingerprints’ offline, correspondents can also sign your key

Names are imported into your keyring only when signed by ‘trusted’ correspondents

SDSI works similarly

Great, if everyone is known to each other... spontaneous messaging requires a bootstrap

Brian LaMacchia’s PGP Keyserver is a centralized cache of people’s signed public keys

PGP tools can interactively query it, attempting to construct a chain of trusted ‘introducers’

Thus, we have a radically decentralized namespace – ’Rohit’ is in the eye of the beholder – but implemented centrally...
XML Namespaces

- Suppose we zoom further into our itinerary web page:
  - `<B>Total: <FARE currency='usd' basis='R'>$6010</FARE>`
- Their XML element for distinguishing fare amounts is an addition to the HTML tag namespace
- XML Namespaces essentially turns tags into URIs:
  - `<HEAD xmlns:u='http://united.com/schemas/fares'>...
    <u:FARE u:currency='usd' u:basis='R'>$6010</u:FARE>`
- But how to compare United’s fares to another’s?
  - XML namespaces are a nifty Internet-scale solution
IScale Properties of XML Namespaces

- Binding an ontology (vocabulary) to a URI allows communities of different scales to share semantics:
  - Over time, it could be ratified to http://iata.int/fareschema
  - In restricted beta-testing of advanced features, it could be delegated to http://dev.united.com/rel3/fares.v1

- Versioning is a red-herring: new namespace, new URI
- HTTP content negotiation leaves schema format open
- ... but disagreements are still accurately flagged
  - Disambiguates Air, Hotel, and Auto definitions of <DAY>
... and many more IScale namespaces:

- Dublin Core
- Library of Congress classifications
- Yahoo! Categories
- ISBN / ISSN numbers
  - http://isbn.nu/<isbn> - try it!
- UPC product bar codes
- GPS coordinates (?)
- RFCs & Internet-Drafts
- User & Group profiles
- Printer Descriptions (PPDs)
- Video Codecs
- Fonts
- Colorspaces
- Java class files
- GUIDs (globally unique IDs)
- Social Security Numbers
- DUNS business ID number
**Recap: Key Namespace Features**

- Name of the Namespace
- Resolver system accepting such addresses
- Authority governing form of names
- Authority governing entries in namespace
- Internal structure of names, if any
  - Directionality, if hierarchical
- Lifetime of name (domain)
- Lifetime of address (range)
- Density – current size / potential size
- User Interface implications – internationalization
- Give three example entries in the namespace
- What subspaces are reserved, and for which purposes?
- Formally, is the resolver function a bijection (i.e. injective and surjective – having unique addresses, and names for all addresses?)
- What other namespaces map to it?
  - E.g. phone numbers are also represented in the .tpc.int domain
- Context-sensitivity – any additional parameters to the resolver function?
- What’s the resolution algorithm?
Part II: Identifying IScale Issues

Recall the three requirements we set forth:

- Names must scale across time — longevity
  - Human- and machine-readability
  - Security and reliability
- Names must scale across space — latency
  - Scalable, nomadic, decentralized algorithms
  - Geography and other context-dependencies
- Names must scale across organizations — liability
  - Names reflect trust decisions
  - Accommodating anonymity
**IScale: Across Time**

- **Longevity requires readability**
  - Fixed format standards preserve machine-readability
  - Human-readable names for recoverability and usability
    - Internationalized, graphical, and audio “names” exist, too

- **Longevity requires security and reliability**
  - Formats, protocols, and policies must be stable standards
  - Resolution services must be audited and bullet-proof
  - Reliable on-line access can increase fidelity (up-to-date)
  - Mobility, by contrast, calls for agility; rapid updates
Physical scale’s most salient constraint is latency
- Far beyond a LAN’s RTT of 30 ms, past Internet’s 300 ms, all the way to nomadic disconnection for days at a time
- Calls for new mobile, decentralized resolution strategies

Physical scale is also an opportunity
- ‘Geospatial hypertext’ shows the way to content that resolves specifically for a reader’s location
  - E.g. having “united.com” return the nearest ticket office
- Conversely, planetary reach mocks global namespaces
  - E.g. “tollroad.com”, which resolves to a few miles of Hwy 73 at UCI...
Organizational boundaries are trust boundaries

- Thus, multilaterality is a key IScale issue:
  - Explicit delegation of naming authority can reduce contention
  - Explicit levels of commitment: private, experimental, public, and so on

- Paranoia also follows from strong trust boundaries:
  - So decentralization is even more of an IScale issue than distribution

- Liability accrues at those boundaries
  - Drives need to explicitly articulate the namespaces used

- Anonymity and pseudonymity are also solutions!
A Vision: Postmodern Naming

How do human societies handle naming, anyway?
- People are not uniquely named
- Not all people are even uniquely addressable
- No person or organization can enumerate all people
- People arguably manage self-organizing namespaces

Everyone has their own personal namespace, yet we’re all only a few degrees of separation apart

What will the meaning of a name be when computers have to play ‘six degrees of separation’, too?
Take-Home Points

- There are many, many kinds of IScale Namespaces
- There are genuinely Internet Scale issues
  - Decentralized Algorithms: Protocols, Standards
  - Decentralized Policies: Politics, Trust, Economics
- There are genuinely Internet Scale solution patterns engineers need documented