Whither WWW? — Status, Problems, Directions

Henning Schulzrinne
GMD Fokus, Berlin
hgs@fokus.gmd.de

Supercomputer 1995

June 27, 1995

Overview

- current deficiencies − emerging technologies = long-term problems
- Internet = $\boxed{\text{WWW}}$ + email + MBONE + IRC + …
- WWW = URL + HTTP + HTML + (older protocols)

Overview

- Internet – in Germany
- network issues (HTTP)
- the WWW lingua franca: HTML
- browsers
- information access: URNs, access control
- legal/political issues

The Internet

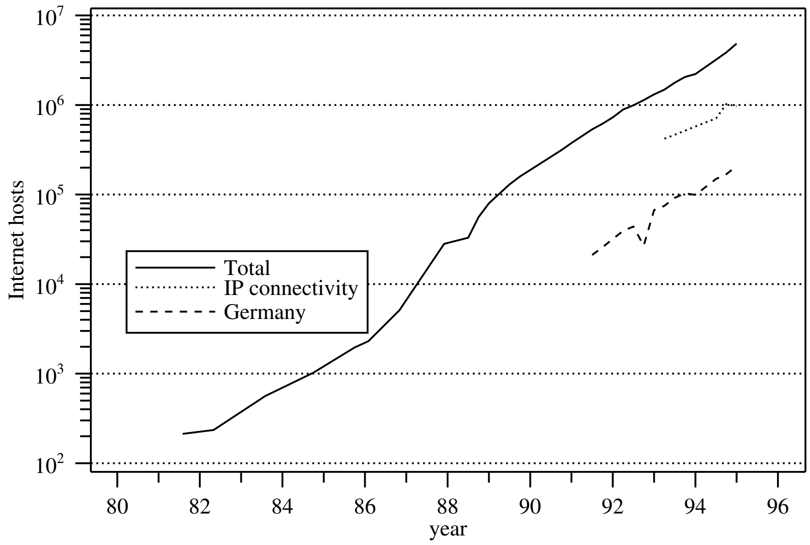
The Internet Hype

The Internet (and WWW) …

- won't cure the common cold
- won't create world peace
- might or might not add jobs
- may replace phone orders
- partial replacement: fax, letters, phone calls, mass mailings, …
- may or may not make you happier (did the telephone?)
- only consultants are making money right now…

On-line services can pay a lot of small providers out of their monthly fees. The Internet would require individual contracts/payment interactions with each information supplier. Services are less likely to be used if I need to pay a little for each use.

Internet in Germany



Internet in Germany

	USA	Germany
highest access speed	45 Mb/s	2 Mb/s
high-speed access	$12,000/a (1.5 Mb/s)	$253,000/a (2 Mb/s)
ATM 155 Mb/s fixed	$7899/month	$26700/month
ATM traffic	free	$1834/hour
local (ISDN) calls	free	$0.80/h
data networks	leased lines	X.25
Internet use	upper-income males	CS students
service offerings	White House, FTD	universities, Spiegel
student use	sociology freshmen	CS PhDs
on-line users	O(5 mio.)	100,000 (700,000 Btx)
president on line	yes	no
Post Office on line	yes	Btx?

"While online enthusiasts will be cheered by this news, Germany is still relatively undeveloped terrain regarding the Internet or systems like CompuServe. These two networks, the Frankfurter Rundschau reported recently, together have a total of about 100,000 private subscribers. The German BTX system, which is offered by Telecom and is available only in Germany, is much better known. BTX currently has 700,000 users." (*This Week in Germany*,

`http://langlab.uta.edu/pub/gic/twig/95mar/03marwig.txt`)

"When it comes to movies-on-demand, only 11% of Americans use pay-per-view (PPV), the study shows, and that number only uses it 14 times a year. Also, 91% of people would rather rent a $1 video than pay for a $5 PPV movie. In the home shopping arena, only eight percent of those surveyed say they watch services such as QVC and Home Shopping Network.

Yet only six percent of Americans connect to an online service each week, and only 38 percent of non-users feel they will do so in the near future. (March 31, 1995)"

Internet Access

In the U.S.:

- cable TV systems (10 Mb/s channel)
- kiosks in libraries, post offices, municipial buildings, ...
- FreeNets (public access dial-in sites)
- accounts through local university for families of employees, high-school students

Not just consumer, but also producers:

- on-line services offer personalized web pages
- small-scale publishing (local event listings, dog breeding club)
- cost to provider is low

GMD fokus

WWW: Future Commercial Uses

- free information
  - advertising (customer expects)
  - customer support
  - public relations
  - non-profits
  - non-commercial content as hook
  - “vanity press”
- with payment
  - software/audio distribution
  - specialized information services
  - mail-order for “no-touch” goods

GMD fokus

WWW and Networks: HTTP, Multimedia, . . .

GMD fokus

Impact of WWW on Networks

Worst of data and interactive services:

- low latency requirement: a human is waiting
- high data rate, very bursty
- needs to "feel" fast
- short documents ➠ high impact of latency, signaling
- little spatial correlation

GMD fokus

HTTP Problems

- extremely simple to use (ASCII, can use telnet)
- inefficient:
  - new connection per request
  - typical: text + lots of small images
  - separate transfer for each page, image
  - TCP open/close + slow start for each transfer
  - slow links + TCP timeout (4 min.) ➠ lots of protocol state
  - even slower with ATM connection setup
  - textual (ASCII) headers: slow processing?

➠ try to fix latency problems (they won't go away with faster links, machines)

GMD fokus

HTTP is mostly an 8-bit protocol, so it avoids the kludges necessary for Internet email. However, the encoding of forms content in URLs and as x-www-url-encoded is still rather cumbersome, particularly for binary data.

Faster HTTP

Incremental improvement over HTTP:

- client keeps TCP connection open
  ➠ fewer connections, lower latency
- pipelining
  - GETALL request: all images in page
  - GETLIST request: enumerate URLs
- with lots of images, latency halved
- mostly backward-compatible

See

`http://www.ncsa.uiuc.edu/SDG/IT94/Proceedings/DDay/mogul/HTTPLatency.html`

for paper.

GETALL generates unnecessary net traffic if a page changes text, but icons remain the same and are still cached.

HTTP-NG

- proposal within IETF
- single stream for several requests
- requests and responses are asynchronous
- binary encoding (packed ASN.1)
- interoperate through proxies

GMD fokus

`http://sunsite.unc.edu/ses/ng-notes.txt`

Given that most HTTP problems can be fixed within the existing protocol, it appears unlikely that HTTP-NG will succeed.

HTTP Negotiation

- content type negotiation not implemented
- language negotiation: "German first, English if you must"
- character set (Japanese)
- max. price for information?

▼ mostly not implemented

GMD fokus

Cacheing

- cacheing
  - browser: store recently used images, text locally
  - proxy (e.g., on firewall): returns locally cached version after checking expiration date
- needed for scaling (on-line services)

But…

- proxies add delay, another failure point ➠ not very popular
- access authorization ➠ similar to site licensing
- for provider, difficult to measure "hit count" (specially equipped browsers)

Even with cacheing, world-wide web access is still easier to measure than, say, TV or print-ad reach. Eventually, there will probably be the equivalent of Nielsen households, e.g., browsers that log accesses.

New Interaction Modes

- currently: user pull only, stateless
- need:
  - continuous or periodic updates (Netscape)
  - state (identity) from one click to next
  - "what's new since your last visit"
  - local agent collects overnight news into morning paper
  - email notification based on personal profile
    * aggregated
    * per-service

GMD fokus

HotWired (`http://www.wired.com`) has a personalized "what's new" page. Agents checking WWW pages and email notification accomplish roughly the same, just driven by own computer or external information aggregation service, respectively. An example of a service that polls sites is

`http://www.netmind.com/URL-minder/URL-minder.html`

.

See

`http://home.netscape.com/assist/net_sites/pushpull.html`

for dynamic documents. See

`http://www.netcreations.com/gadget/`

for a demonstration.

`http://home.netscape.com/assist/net_sites/bg/index.html` describes backgrounds. Netscape also implements "cookies" for maintaining state, as described in

`http://home.netscape.com/newsref/std/cookie_spec.html`

This allows state maintenance in the client using HTTP headers. This is useful for keeping track of a single customer in an on-line shop or to avoid unnecessary reauthentication.

Continuous Media

= audio, video, games, …

- delivery: HTTP over TCP is not suitable (other protocols: RTP)
- click, wait for download, play: not tenable
- integration: video window in browser
- VCR control: currently too cumbersome
- want to click on items in live video
- "twitch" (arcade) games
  - probably too slow (network latency)
  - need more than clicking
  - local execution with communication

GMD fokus

For audio, RealAudio offers a proprietary encoding and server for interactive, low quality audio through modem lines:

`http://www.realaudio.com/`

Highly interactive games, video arcade style, are probably not suited for current WWW technology.

HTML

HTML Advantages

- simple to understand & generate from applications (e.g., databases)
- bandwidth efficient
- adaptation to local environment (ASCII, different screen resolutions)
- client-side text searches
- easy conversion at browser: wordprocessor, Braille, speech, …

HTML Generations

version	browser	features
1.0	historic	headers, emphasis, lists
2.0	Mosaic	inline images, forms
“2.1”	Netscape	alignment, tables, appearance
3.0	Arena	math, figures, stylesheets

▼ no two browsers support exactly the same features
▲ but usually no damage

HTML 2.0 Limitations

- H1, H2, … for font changes, not structure
- traditional article hierarchical structure appropriate?
- everything as imagemap GIFs?
- …or Acrobat PDF?
- no tables, math
- stylesheets: map HTML to presentation (e.g., <H1> maps to 18 pt Helvetica Bold, centered)
- client-side highlighting for imagemaps

22-1

See

`http://www-swiss.ai.mit.edu/philg/research/shame-and-war.html`

for another HTML critique.

Netscape HTML Enhancements

"HTML 2.1"

- backgrounds ("wallpaper")
- figure placement: text alongside figures
- different font sizes
- tables
- blinking text (avoid!)

HTML 3.0

- tables
- notes, warnings (with locally meaningful symbols)
- stylesheets
- client-side hotzones
- scripts for forms (checking, calculated fields) – *unspecified*
- mathematics (subscripts, superscripts, fractions, ...)

Alternative: HyperTeX, PDF with links.

Math features are not likely to be high on the priority list of commercial browser developers, but ARENA supports it. HTML 3.0 math features are based on, but still not quite as rich as LaTeX. Arena can be found at

`http://www.w3.org/hypertext/WWW/Arena/`.

HyperTeX is described at `ttp://xxx.lanl.gov/hypertex`. It is not quite clear what major advantages full SGML would bring for most information, but it is useful to gain direct access to the large body of SGML material (TEI for critical texts, CALS for military, etc.).

HTML Limitations

Even with HTML 3.0:

- layout features will never equal PostScript
- automatic conversion from branched scroll to linear printed page?
- other layout (two-column, fancy figure wrapping)
- full SGML?
- no multimedia capabilities ➠ HyTime?

GMD fokus

WWW as the Universal Network GUI?

User interaction:

- fill out form, press $\boxed{\text{Submit}}$
- click on imagemap
- like old mainframe terminals

Can't do:

- rapidly changing UI
- per-field verification
- drag-and-drop
- sliders, dials, ...

Partial solution: client-side scripts (SafeTcl, Java)

Virtual Reality

- low-bandwidth description of environments (VRML)
- allow navigation with keyboard, mouse or headgear
- `http://www.sgi.com/`

WWW Browsers

Advanced Browsers: Java

developed by Sun Microsystems

- client-side program execution, based on OO language
- compiled scripts downloaded and executed by client *virtual machine*
- need restricted execution environment
- allows more sophisticated user interface
- allows local forms processing

alternative: SafeTcl

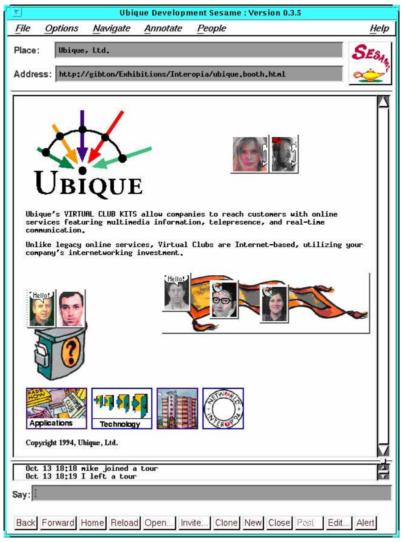
[Demonstration of Java: moving graphics objects; games]
`http://java.sun.com/` and (in German) `http://www.iway.de/iw95-22/java.html`

Browser Extensions: Virtual Places

- communicate with visitors on same page by typing, audio
- take visitors on WWW tour ("bus")
- requires custom browser or browser enhancement
- applications: customer support, trade shows, education

`http://www.ubique.com`

Virtual Places



Client-Side Interface

Problem: browsers cannot communicate with local applications

- external applications control browsers:
  - URLs in email trigger browser
  - guided tours, slide shows
- Netscape: remote control using X properties
  - get URL
  - all commands reachable by menus, buttons
- Common Client Interface (NCSA Mosaic):
  - control operations
  - use browser as URL resolver to retrieve data

For Netscape and Unix X11, see

`http://home.netscape.com/newsref/std/x-remote.html`

For Mosaic CCI, see

`http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/CCI/cci-spec.html`

The protocol is textual, using a TCP connection between application and browser.

The Future of Browsers

- better bookmark management (by topic)
- integrated with email, news reader, file system
- every application understands URLs (word processor, editor, data base, …) ➠ extension of object linking (OLE) idea
- ➠ browsers need to "borrow" windows
- applications need to borrow browser windows (live video!)

If bookmarks become files, part of the normal directory structure, we really only need display applications that are URL-aware and a stack that can handle URN resolution, the HTTP transfer, etc. This would leave a browser as an HTML viewer.

The Future of Browsers

- browser as local file system manager?
- local file system as hypertext instead of directory?
- universal network user interface (high-level X?)
  - not interactive enough
  - but: scripts
- do we need separate browsers?

Access: URNs, Access Control

Location Independence: URN

- URL (uniform resource locator) ➠ fixed host and location in server file system
- URC (uniform resource citation) ➠ description
- URN (uniform resource name) ➠ map permanent name to URLs
- how generic? ➠ URN for stock quote, with qualifiers for cost, timeliness, etc.
- different, permanent naming authorities (ISBN, institutions, ...)
- suitable for load balancing
- need fast, reliable directory service

finally: need pronouncable URLs...

Currently, limited machine independence and load balancing can be accomplished with DNS (domain name service) aliases. X.500 directories are probably too slow; there are proposals within the IETF to use DNS also for URN lookup, with the domain name as the most significant part of the resource identifier.

Parental Control

"Just one mouse click away from pornography/bomb instructions/gambling/…" ➠ need to control content for

- parents
- public libraries
- offices

Approaches:

- control authors ➠ censorship ("prior restraint"), international jurisdiction
- control online services ➠ impractical due to volume; what standard? how to verify age?

Parental Control: Client Side

Alternative: let client control content:

- filter on browser ➠ positive/negative URL list
- document header contains rating from different agencies (Catholic Church, Playboy, …)
- browser configured to display only those containing appropriate rating
- URL contains age restrictions (KidCode)

GMD foKus

Particularly in the U.S., access control has assumed great significance in the last few months, particularly since private and primary school net use are much more widespread than in Europe. There have been legislative proposals to make authorship and transmission of offensive material a criminal offense. One product based on a negative list is `http://www.surfwatch.com/`. KidCode inserts special character sequences into the URL like

`http://bookstore.com/KidCode.12.profanity/`.

It is described in

`ftp://ds.internic.net/draft-borenstein-kidcode-00.txt`.

Sites could easily bypass subscription schemes, but are unlikely to do that, as they are not likely to be interested in being sued by a parent. An organization working on classification is SafeSurf (`http://www.safesurf.com/wave/`).

Legal Issues

Legal Issues

Success of commercial use depends on legal infrastructure:

- encryption for privacy ↔ law enforcement, export
- authentication ➠ avoid having different passwords/ids for each service ➠ certificates ➠ certification authorities
- electronic cash ↔ central banks, tax authority
- copyright: you can't "loan" bits
- gambling, pornography, copyright: which law applies? who enforces it?

In the U.S., the Postal Service (`http://www.usps.gov`) apparently is considering acting as certification authority (CA), other candidates are banks and federal/state/local institutions. In Germany, the *Einwohnermeldeamt* would be a natural CA. It is likely that persons may have several identities, just like different identity cards today. Also, functional identities ("authorized to sign for Joe's Bakery").

Off-shore gambling seems to attract a lot of attention (see *TIME Magazine*, June 12, 1995).

Casual copyright infringement using the WWW is actually somewhat less likely than with disks and ftp, as you have to assemble bits and pieces and explicitly store them locally. There are efforts to mark electronic copies so that printed versions can be traced to the original purchaser.

WWW: The Competition

WWW Competition

- BTX, Minitel: slow, low-res graphics, geared to dumb terminals
- BTX+: proprietary, Germany only
- VOD: special purpose; navigation?
- MHEG (French, RACE/ACTS project)
  - allows multimedia synchronization
  - hides program "source"
  - much more complicated than WWW
  - not well suited for diverse delivery
  - set-top boxes?

WWW Competition

- Microsoft Network (MSN)
  - better integration with OS?
  - possibly superior document language
  - content?
  - not technical issue: bundled ↔ "it's Microsoft"

WWW vs. Video on Demand

- enthusiasm for VoD (video on demand) is shrinking (at least in the US):
  - high cost – pay-per-view has been niche market
  - little additional TV time available
  - cheap competition from video stores/purchases
- WWW as competition to video shopping:
  - potentially high information content for WWW
  - video footage helpful?
- VoD: hot lists?, no local processing
- WWW could subsume video-on-demand
- remote-control version of WWW ➠ few links, no forms entry, …

GMD fokus

WWW is considered too complicated for non-computer users, but video-on-demand is simple only because (usually a single) service provider severly restricts choices. It would be just as easy to build specialized link pages.

Video footage, as envisioned for home shopping, is mostly useful for items that many people prefer to experience first-hand (cars, clothes, …). For pure entertainment, the current shopping channel will do fine.

Summary: The Major Problems

Covered a lot of technical detail ➠

- Internet business model?
- WWW suitability for video-on-demand? games?
- WWW in a set-top-box?
- access controls?
- security: mostly a *legal*, not a *technical* problem
- standards (too many players?)

GMD fokus

Conclusion

- WWW is not the only Internet application
- …but likely the driver for home market (except cheap telephony)
- lots of kludges, but does 80% of what's needed
- grass-roots ➠ will it be "allowed" to catch on in Europe?

GMD fokus