Pachyderm

A Strategy for Working on the Web

Andrew Birrell
Mike Schroeder
Raymie Stata
Ted Wobber

October 1997
The System

- Mail Service
- POP3 & SMTP
- Pachyderm Server
- Web Server
- HTTP
- Web Browser
- Msg Store & Index
- Perhaps a Firewall
Personal Computing in the Late 1990’s

- Communicate with lots of different people
- Access lots of sources of information
- Move around a lot
- Lots of accumulated data
- Lots of valuable data
A Solution

- The platform is the web
- Location-independence
- Bandwidth tolerance
- Indexing for Information Structure and Retrieval
The Platform Is The Web

- Client hosts communicate *only* through HTTP and web servers

- Client is a web browser — no software to install or manage

- Build the user interface from HTML and Java

- Use Java to adjust where the work gets done
Location-Independent Computing

- Any web browser-equipped computer will do
- Any location will do
- Don’t lock state in personal machines or behind firewalls
- Security must not be compromised by location
Bandwidth Tolerance

- Acceptable performance even with 28.8K dial-up
- Ship data only as needed by client
- Use concurrency to hide latency
- Use Java to move latency-critical work into client
Information Retrieval

• Manual classification schemes alone are too difficult

• Fixed classification schemes:
  — don’t scale
  — don’t adapt to change
  — need too much expertise

• Use full-text index/query (as in AltaVista.digital.com)

• Allow user-created labels to add structure
Pachyderm: System Structure

Mail Server

Index Server

Web Server

Browser

Message Store & NI2 SDK

Web Server

CGI program

MIME filter

Java Applet

Msg Store & Index
Pachyderm Design Details

• Repository designed for lots of messages (e.g. GByte/user)

• Index is NI2 (same as AltaVista) so it scales well
  — can index attachments too (e.g. PowerPoint)

• Supports location-independence — just walk away from one client and start another.

• Adequate performance over low-bandwidth high-latency lines:
  — Java for interactive UI
  — Hot-links for attachments reduce message transfer size
  — Concurrent write-behind to reduce update delays

• Mutable labels to apply structure to index
Mutable Labels for a Full-Text Index

- Can attach any number of arbitrary strings to any message
- Easy and efficient to modify the set of attached labels
- Queries can include label terms
- User can use labels instead of folders
- System can use labels for manipulating state:
  - pending new mail
  - read/unread flag
  - hidden/deleted flag
Named Queries

- Set of queries associated with each user account
- Frequently used queries
- Inbox filter (messages to be kept, but omitted from inbox)
- Mailing list subscriptions
- Newsgroup subscriptions
Message Filter Details

- Presents RFC822 messages as HTML

- Generates hot-links for:
  - MIME attachments
  - documents embedded as text, e.g. uuencode, base64, postscript
  - strings that look like URLs
  - strings that look like email addresses
  - raw message text

- heuristic recognition
Secure Web Tunnel: The Artifact

Internal Web

Secure Web Tunnel
- Authenticator
- Redirector
- SSL/HTTP proxy

Netscape or IE

HTTP
SSL
SSL
HTTP

Logging & Auditing

Firewall
Secure Web Tunnel: Details

• Several choices on authentication:
  — cryptokey handshake (stores token as cookie)
  — name and password (stores token as cookie)
  — X.509 certificate + secret in user’s smartcard (or laptop)

• Several choices on security options:
  — logging and traffic analysis
  — restricted access based on user, location, etc.
Status: Pachyderm

- Prototype working, several users at SRC and rest of DEC
- Transfer in progress to product team
- Implementing complete UI with Java applet really works
- Location independence (even at low-bandwidth) is wonderful
- Indexed email is wonderful
- Several patents filed
Status: Secure Web Tunnel

- Prototype working
- Transfer in progress to product team
- Paper written
- Patent filed
What’s Next?

• Make web tunnel really usable

• Add Pachyderm access to newsgroups and shared documents

• Provide disconnected operation for Pachyderm

• Get products to market and/or sell the technology

• Choose the next step along the vector ...
For a description of Pachyderm:

http://www.research.digital.com/SRC/pachyderm/