Announcements

- Reading
  - 7.5-7.6
- Project #5 was due on Monday
  - Demos are on Wed 12/12
- Signups for demos was circulated

News

- Large Collection of newsgroups
  - currently a hierarchalnamespace (used to be rather flat)
  - can be moderated: must be approved before being posted
- Messages
  - have a unique id
  - are associated with one or more newsgroups
  - contain a superset of RFC822 fields
- Transport of news
  - a site a list of one or more sites it gets is newsfeed from
    - a site periodically polls its newsfeeds for news
    - newsfeeds can also push new news out
  - UUCP: Unix-to-Unix CoPy
    - historical path using dialup modems
  - NNTP: Net News Transfer Protocol (TCPport 119)

NNTP

- Provides end-users with news (ala POP for email)
  - LIST and NEWSGROUPS commands
  - GROUP: list all articles in a group
  - NEWNEWS: all news since a specific time
  - ARTICLE: give me a specific article
    - note that this is done by id, not local number
  - POST: post a news article
- Supports moving news between servers
  - all commands from above plus..
  - IHAVE: supports flooding articles around the net
- Performance Problem
  - Well over a gigabyte per day of news
  - protocol is stop and wait (limited for round-trip latency)

Naming in the World Wide Web

- Web history
  - Httpd 0.1 release 6/91; w3 mailing list 10/31/91; first ref to xmosiac (in my email) 2/93.
- Uniform resource locator
  - a single namespace for all objects on the Internet
- <protocol>:<port>//<dns name>/<page reference>
  - protocol: http, ftp, gopher, wais, file, news
  - port: optional - defaults to well known port for protocol
  - dns name: often an alias for a host name
    - also can be an IP address
  - page reference: usually a file path name
    - servers can define default translations
      - default file suffix (e.g. .html)
      - default page for a directory: index.html
      - ~ to expand to a user's home page

WWW (cont.)

- HyperText Markup Language
  - based on SGML
    - font changes, text placement
    - includes support for images
  - supports references to other document (links)
  - supports alternatives to display if browsers can't support a format
- HyperText Transport Protocol
  - used to move HTML from server to client
  - Basic protocol
    - GET: get a page
    - PUT: store a page
    - POST: append to a page

Interactive Web Pages

- Forms
  - HTML can describe fields which permit users to enter data
    - textboxs, checkboxes, lists, etc.
  - contain an action
    - a URL to POST the completed form
- Common Gateway Interface (CGI)
  - Servers can be told that some pages are really programs
    - could be executable binaries, perl programs, etc.
  - An attempt to GET or POST to a CGI script runs it
    - the form data is taken as input
    - CGI script returns an HTML page as output
      - output can be a function of the input
  - common examples:
    - perl scripts
    - interfaces to database systems

CGI Environment Variables

- Basic Fields
  - SERVER_PROTOCOL
    - The name and revision of the information protcol this request came in with. Format: protocol/revision
  - SERVER_PORT
    - The port number to which the request was sent.
  - REQUEST_METHOD
    - The method with which the request was made. For HTTP, this is "GET", "HEAD", "POST", etc.
  - QUERY_STRING
    - The information which follows the ? in the URL which referenced this script. This is the query information.
  - REMOTE_HOST
    - The hostname making the request. If the server does not have this information, it should set REMOTE_ADDR and leave this unset.
  - REMOTE_ADDR
    - The IP address of the remote host making the request.

CGI Environment (cont.)

- Authentication Information
  - AUTH_TYPE
    - If the server supports user authentication, and the script is protected, this is the protocol-specific authentication method used to validate the user.
  - REMOTE_USER
    - If the server supports user authentication, and the script is protected, this is the username they have authenticated as.
- Post Related info (actual data sent on stdin)
  - CONTENT_TYPE
    - For queries which have attached information, such as HTTP POST and PUT, this is the content type of the data.
  - CONTENT_LENGTH
    - The length of the said content as given by the client.

Project Goals

- Background
  - Enterprise application servers huge, and growing
  - Current systems are heavily optimized, but
    - single-purpose systems
    - tuning is complex
- The Approach
  - vertical integration: apps, middleware, OS
  - *but need realistic infrastructure*

Project Components

- Applications
  - 3D volume reconstruction (Davis)
  - Exploration, visualization and processing of remote sensor, microscopy simulation data (JaJa, Sussman)
  - Streaming video (Golubchik)
- Middleware
  - Java in the enterprise (Pugh, Chawathe)
  - Databases (Roussopolous, Chawathe)
  - Harmony (Hollingsworth, Keleher)
  - Compilation (Tseng)
- Systems
  - Active disks (Sussman)
  - Active routers (Bhattacharjee)

Infrastructure

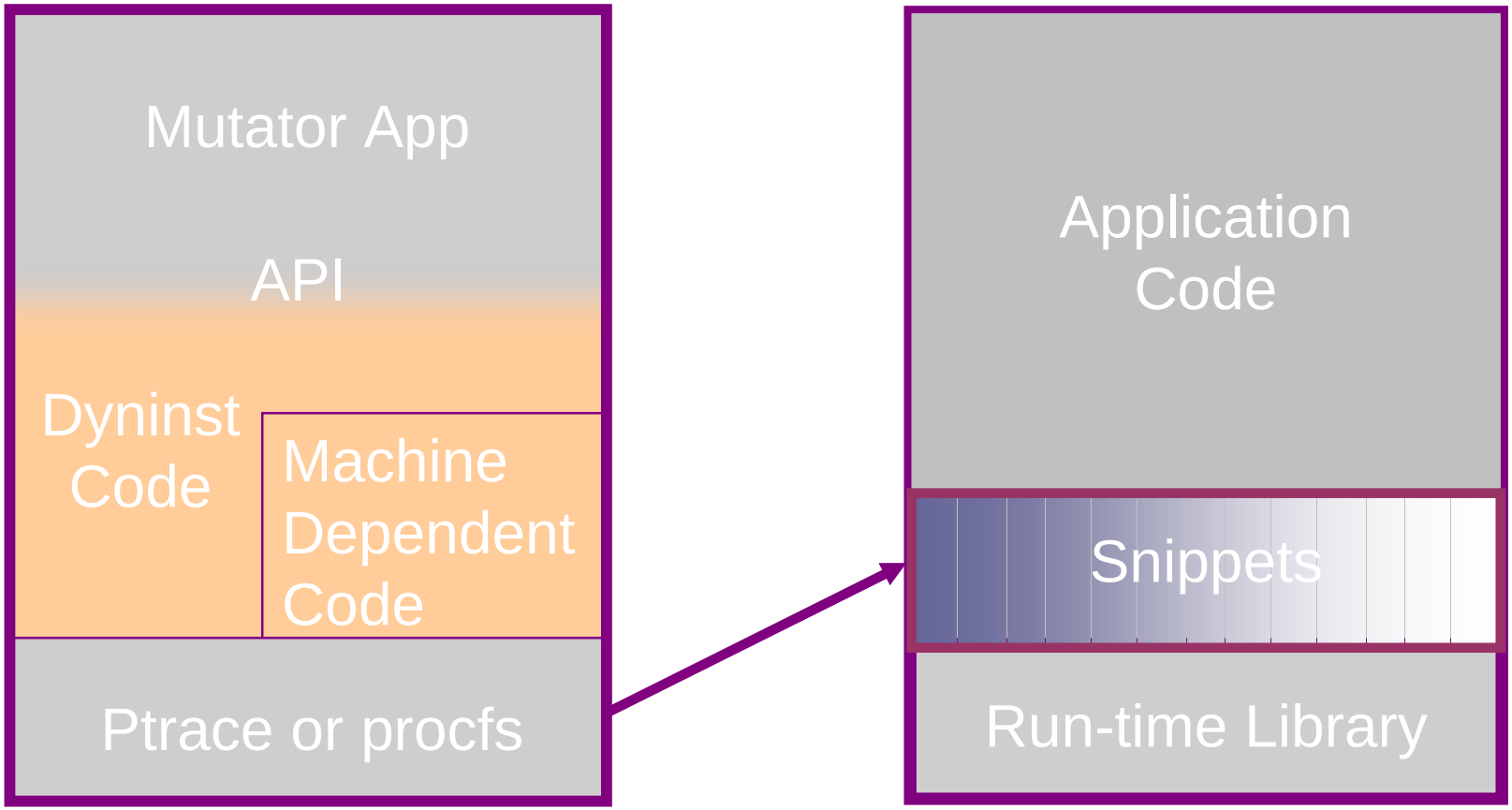
- Sun Starfire
  - 24 processors (Ultra-Sparc III)
  - 24 Gigabytes of memory
  - 0.8 terra-bytes of storage

- Storage Cluster
  - 50 Netfinity nodes (650 Mhz, 128 MB)
  - 9 terra bytes of storage
    - 2-4 disks per node
  - Each node has 3 100 Mbps Ethernet cards
    - 2 cards used for I/O
    - 1 card used for system management

Applications of Runtime Code Patching

- Performance measurement
  - Recording application behavior
- Correctness debugging
  - Fast conditional breakpoints
  - Data breakpoints
- Execution driven simulation
  - Architecture studies
- Testing
  - Code coverage testing
  - Forcing hard to execute paths to be taken

Structure of the Dyninst Library



API Library

- Provides
  - Functions for control of mutatee
  - Runtime code generation
  - Information about mutatee
- A set of C++ classes
  - Machine independent representation of a program
  - Processes and threads
  - Representation of new code to patch into program

Representing Code Snippets

- Platform Independent Representation
  - Same code can be inserted into apps on any system
- Simple Abstract Syntax Tree
  - Can refer to application state (variables & params)
  - Includes simple looping construct
  - Permits calls to application subroutines
- Type Checking
  - Ensures that snippets are type compatible
  - Based on structural equivalence
    - allows flexibility when adding new code