This Time

- sample services, cont’d:
  - grid computing
  - XML
  - RSS
  - RFID
  - some security considerations

Grid Computing

- make a collection of distributed resources appear to a user as a single, coherent, computing resource
  - e.g., cpu, storage, visualization & display
  - even if different platforms are involved
- connect resources over a network to make available to user
- may dynamically create such collections
- start with a research-oriented system...

The TeraGrid Project

- NSF stimulated project to create resource for researchers
- consists of:
  - 20 teraflops over 5 distinct locations
  - storage management of 1 Pb (2^50 bytes)
  - tightly integrated components in each cluster
  - connected to other clusters via 40 Gb/s network
  - e.g., linux cluster (Itanium® processors): 1 Tf

TeraGrid
Grid Computing for Mortals

- same goal: creating computing resource from separate distributed resources
- use standard Internet
- need special software to build grid
  - "Legion" from Avaki (Grimshaw at U. Virginia)

Layered Grid Architecture

- Application
- Collective
- Resource
- Connectivity
- Fabric

Grid Software

- need to 'virtualize' resources
- Open Grid Services Architecture (OGSA)
  - common 'view' of a resource and how to use it
- for service ensembles, OGSA supports:
  - creation
  - maintenance
  - application
- see http://www.ogs.org/ogsa

The Emerging Grid

- grid: "...emergence of a new infrastructure upon which first science, and then the whole economy, will be built " (Larry Smarr)
- do for computing what Internet did for documents
- aiming at worldwide governance & standards body analogous to IETF: http://www.gridforum.org

Evolution of Web Services

- limited to HTML
- improve function via ‘active pages’
  - those that require server to run command parser on requested HTML file and perform computing as needed
- but still limited:
  - to what can be embedded in near-HTML pages
- what if could generate pages containing tags not previously defined?
  - i.e., tags specific to particular application needs?

Working the Web

- current web pages written in html
- browser/client renders html
- html tags:
  - understood by both client to have particular meaning
  - who defines?
  - can you have new ones? whenever you want?
Working the Web

- html tags:
  - understood by both client to have particular meaning
  - who defines?
    - W3C
  - can you have new ones? whenever you want?
    - not really
  - html tags only describe layout of document contents

Beyond HTML

- web purchase example: buy a hard disk
  - go to vendor site and
    - navigate through their web pages to find item
      - maybe you don’t like their price
    - go to another vendor site
      - navigate through their web pages to find item
        - you like price, but they’re out of stock
      - go to another vendor site...

Better Than Before...

- better way?
- software agent on your computer:
  - presents a form where you say what you want
  - it returns a summary of vendors that meet criteria, automatically
- cannot (practically) use html to do this
  - how come?

...but need new language

- html only specifies layout
  - to do order form: need tags that apply to content of document
    - e.g., identify it as a vendor query
    - e.g., indicate what part describes sought part
  - need something that extends html capabilities:
    - eXtensible Markup Language, xml

xml

- xml, like html, uses tags mixed in with document content
- xml lets you define and create any tag you want in a document, e.g.,

  `<particular>`
  <type:computer:hard:disk>
  <capacity:60:gb</capacity:
  >speed:7200:rpm</speed:
  </type:
  `particular`>
  - but who will understand these tags?
  - who can you send this to?

xml

- xml tags used to describe content of document
  - not really used to describe layout
  - so how can you see the content of an xml document?
    - some html browsers have simple xml rendering ability
  - need another document that describes how to render the tags
    - the same content can thus be rendered differently depending on the style sheet that describes how to render
  - what do these style sheet documents look like?
**xslt style sheet**

- xslt to convert xml parts query back to plain text
- "attribute:value" format
- `<xsl:stylesheet xmlns:xs="http://www.w3.org/1999/XSL/Transform" version="1.0">`
- `<xsl:output method="text" indent="yes" />`
- `<xsl:template match="/">`
- `<xsl:template match="type">` Part Type: `
- `<xsl:value-of select="." />`

**xml and xslt**

- rendering is a transformation operation
- e.g., from 'abstract' xml to plain text
- xslt can be used for any such transformation
- e.g., from one xml document to a different xml document
- ability to transform depends on having an 'understanding' of the source xml document's structure
- internal working representation: tree constructed from the xml

**xml parsing**

- an xml parser builds such a representation
- e.g.,

```
<Program/>
```

**xml parsers**

- two popular parsers:
- Domain Object Model (DOM):
  - from W3C
  - passes through entire xml document, builds representation in memory (e.g., tree)
- Simple API for XML (SAX):
  - from XML-DEV
  - issues callbacks as tags are encountered during parsing
- freely available parsers of both kinds available for many platforms

**xml: missing link**

- have:
  - arbitrary xml tags describing content of document
  - ability to parse document into structure
  - can use xslt to transform into other structures
- don't have?

**xml: missing link**

- have:
  - arbitrary xml tags describing content of document
  - ability to parse document into structure
  - can use xslt to transform into other structures
- don't have? anything that knows about
  - data types
  - legal/illegal values for tags we've invented
  - what if we name our xml tags the same as someone else's in a different document
  - where they don’t mean the same thing?
xml namespace

- schema:
  - provides info about structure and content of data and tags
  - allows for tags to be qualified by namespace
    - so tag ‘product_query’ can appear in two different xml documents and be different in structure and content
    - qualified by, e.g.,
      - acme.com/product_query
      - ozime.com/product_query

Other Web Services Components

- how do you know... destination’s capabilities?
- use Web Services Description Language (WSDL):
  - provides standardized way for a site to make known formats & protocols its service accepts

Other Web Services Components

- how do you know...
- use Web Services Description Language (WSDL):
  - provides standardized way for a site to make known formats & protocols its service accepts
**Other Web Services Components**

- how do you get your document there?
- use Simple Object Access Protocol (SOAP): defines "envelope" for web services communication
- envelope mappable to http and other transport protocols
- one-way msg protocol, allows for intermediaries to process or add to msg

**Getting The Word Out**

- suppose you have a special interest in x
  - e.g., news, sports, technical updates, AV equipment
- how do you stay current?
  - discover/learn which web sites have content
  - visit them often
- sites might prefer to be able to push their wares?
  - member subscription; they send email
  - specialized software ‘agent’ goes mining web site

**A Newer Model**

web site D ‘knows’ a little about content on A, B, and C; you can see summaries and entire articles from A-D just by visiting D

**How Build This?**

- use specialized content agents
  - like crawlers, but targeted
  - content-providing sites provide API for content access
How Build This?

- use specialized content agents
- like crawlers, but targeted
- content-providing sites provide API for content access
- content-providing sites provide data dumps from their content

A Better Way To Build This

- use standardized ‘blurb’ format embodied in a technology called RSS
- Rich Site Summary
- a.k.a. Really Simple Syndication
- like syndicating: publish material to some number of locations
  - e.g., like comics in newspapers, some TV shows
- RSS uses XML

An RSS Item

- RSS provides a set of items within a channel to interested readers
- an item looks like:

```xml
<title>RSS Resources</title>
<link>http://www.webreference.com/authoring/languages/xml/rss/</link>
<description>Defined in XML, the Rich Site Summary (RSS) format has quietly become a dominant format for distributing headlines to the Web. Our list of links gives you the tools, tips and tutorials you need to get started using RSS.</description>
</item>
```

RSS Channel

- channel provides set of items in some way related
  - e.g., most recent, same topic
  - up to 15 items per channel
- RSS element may contain at most 1 channel
- each channel must contain tags:
  - title channel’s title
  - description brief text description of channel
  - link an HTML URL to channel’s web site
  - language language encoding for channel (e.g., en-us)
  - item from 1 to 15 items

RSS Channel

- each channel may contain additional tags, including:
  - copyright designates content as copyrighted, names holder
  - pubDate date this channel was published
  - lastBuildDate time of last update to channel
  - image some graphic for channel image
  - seems like a good idea, in general

Too Good...

- RSS developed by Netscape for a service it wanted to provide: version 0.90
  - then UserLand Software did work to simplify original: version 0.91
  - further UserLand refinements: 0.92, 0.93, 0.94
  - RSS–DEV adopted 0.90 and evolved it into version 1.0
    - based on RDF
  - UserLand most recently produced 2.0
  - most versions mutually incompatible
Using RSS: source-side

- to provide RSS feed from your site to others:
  - need xml file defining your RSS channel
  - need your httpd server to know about this file
  - so can be served on demand
- need to keep RSS file up-to-date as content changes on your site
  - can do by hand
  - many tools to semi- or fully automate

Using RSS: client-side

- RSS-clients
  - receive and render XML of the feed
  - maintain local 'tracking' info so can know which feeds user subscribes to
    - and (perhaps) also status of feeds
  - may integrate into browser
    - many current browsers will render the xml file as plain text; ugly, but readable, sort of

Sample RSS 'page'

```xml
<channel>
    <title>NPR News: Top Stories</title>
    <lastBuildDate>Sun, 28 Nov 2004 18:44:37 EDT</lastBuildDate>
    <language>en-us</language>
    <link>http://www.npr.org/topics/topic.php?topicId=2&sourceCode=RSS</link>
</channel>
```

Better RSS XML Handling

- RSS 'viewer'
  - top picks as recommended by blogspace.com:
    - for Macintosh:.NetNewsWire
    - for Windows: SharpReader
    - for Linux: Straw
  - for web: Bloglines
  - common to use Python as part of receiving/rendering process
    - a scripting language similar in some ways to Perl and Tcl

Aggregation

- RSS on server does publishing or syndication
- on client-side, collect feeds from multiple sources
  - ...aggregators collect news, weblog and other feeds over the web and aggregate them so the news items are readable from a single place, regardless of their source.
    - http://www.nongnu.org/straw
- a site performing aggregation may issue the aggregate as its own RSS feed

Some RSS Feeds...

- recent earthquake information:
  http://www.reuters.co.uk/newsrss.jhtml
- tech news: http://slashdot.org/index.rss
  very widespread use with blogs
RSS Standard?

- current RSS users should provide support for 1.0 and 2.0
- what about a single standard?
  - e.g., what’s the IETF say?

Atom

- from “draft-ietf-atompub-format-03:”

Extending Use of Wireless

- pervasive computing “kitchen helper” application:
  - can present recipes (from local or remote db)
  - can show videos of recipes, techniques
  - focused or general
  - can report which local grocery stores have ingredients
  - how many, prices
  - possibly purchase items at your request

Wireless to the Rescue

- solution:
  - db program needs to know
  - count of items
  - location of items
  - allow items to respond to an application query
  - could place small, active transmitter in/on each item
  - continually transmit an identifying string
central receiver collects signals and relays to application software
No Bunnies

- even the best battery eventually goes dead
- so transmitter would stop
- don't use batteries
- use passive circuit that:
  - is normally off
  - gets power from an RF pulse
  - when powers-up, transmits a unique id number
    - factory programmed, typically ≤64-bits
  - and RF–generating receiver
  - to generate the RF pulse
  - listen for reply from passive circuit

Radio–Frequency IDentifiers

- the passive circuits often called tags
- two frequency ranges:
  - low: 100s of kHz
  - high: 13.56 MHz (ISO 15693)
- work over small distances
  - from < 1 m to few hundred metres
- typical usage setting:

RFID tags

- the tag (a.k.a. transponder) devices can take many forms:
  - 3.85 X 23mm glass transponder
  - 64 bits memory
  - 134.2 kHz (low frequency), HDX
  - ≤ 60 cm reader distance
  - 70 ms to read value (typical)
  - -40 to +85 °C operating range

RFID readers

- perform wireless exchange with RFID tags
- perform exchange with host application
  - relays to app info from tag
  - can send to tag (if its writeable) info from host
  - often wired, often fixed in location
  - serial communication to host
  - 600 - 57600 Baud
  - can store ≤ 909 read ID values
  - can be synchronized to 'avoid' other nearby readers

The RFID tag

- may be read–only
- may be read–write
  - allow user to store other information about the tagged object
  - may be write–once
- may also respond to commands
  - not simply emit id on power-up
  - e.g., "stay quiet"
Application 1: textile manufacturing

- Malden Mills uses RFID in strips of material it manufactures to mark flaws
- Mark flaw start and end with pair of tags
- Slitting machine automatically adjusts action when encounters first tag, re-adjusts on second
- Uses (and re-uses) simple read-only RFID tags

Application 2: airport luggage

- McCarran International Airport, Las Vegas:
  - From check-in desk to pick up at destination: use standard luggage tag with embedded RFID tag
  - Uses 900 MHz RFID tags
  - Use receivers at critical points along handling path, e.g., explosives scan station
  - Critical issue: cost
  - How much does use of an RFID tag increase cost of handling?
  - Supplier quotes $0.25 per tag

Application 3: supply chain

- Complete supply-chain tracking
  - See, e.g., Scientific American Jan 2004 pp. 56 – 65
  - Spaghetti sauce jars tagged as filled/labelled/lidded
  - Boxes of n jars have their own tags, so too, pallets
  - Passively scanned as:
    - Leave food factory
    - Arrive at distribution centre
    - Leave distribution centre
    - Arrive in grocery store

Application 3: supply chain

- When leave grocery store:
  - Shopper carries out past check-out scanner
  - Shopper has store card in wallet
  - Total ‘shopping cart’ scanned and shopper’s account charged
    - And other shopper records updated
  - At home, scanner can tell when jar is removed from pantry: time to order more
  - Tag on container can be used at recycling plant to separate material
  - Or to catch cheaters who put jar in trash?

Not Quite There Yet

- Problems remain, largely:
  - Volume of responses to RF pulse
  - Who already answered?
    - At least wastes CPU time
  - Cost of tags needs to be lower
  - Acceptance issues:
    - Privacy concerns
    - Labour worries

Privacy Problem

- Benetton wanted to tag clothes for inventory tracking worldwide (warehouses, stores)
  - Philips claimed tags readable only in-store
  - Wal Mart and Gillette wanted to tag products
  - Gillette agreed to tag the packaging, not the product
**Proposed Guidelines for Tag Use**

- four suggested guidelines:
  1. consumers must be warned that a product has an RFID tag
  2. all RFID tags must be recognizable and removable by customer
  3. tag dies after leaving checkout counter
  4. tag should be on packaging, not product itself

**Security Notions**

- confidentiality:
  - information can only be seen by those we wish to see it
  - involves OS, network, as well as “common sense”

**Security Notions**

- confidentiality
- integrity:
  - resources cannot be modified except by authorized users
  - modification can be detected

**Security Notions**

- confidentiality
- integrity
- authentication:
  - ensuring communicating parties are really who they say they are

**Security Notions**

- confidentiality
- integrity
- authentication
- availability
- non-repudiation:
  - sender cannot later claim not to have sent msg
Security Notions

- confidentiality
- integrity
- authentication
- availability
- non-repudiation
- key distribution:
  - establish session keys among communicating parties

Classes of Threat

- view computer as repository of information
  - data, reports, intelligence, ...
  - programs to process information
- normal use:
  ![Diagram of normal use]
  - information source
  - information destination
- user initiates request for info, and receives it

Classes of Threat

- interruption: information fails to arrive at destination
  ![Diagram of interruption]
  - information source
  - information destination

Classes of Threat

- interception: information arrives at its destination and also at another, unintended, destination
  ![Diagram of interception]
  - information source
  - information destination
  - “bad guy”

Classes of Threat

- modification: information arrives at its destination but only after being modified by a third party
  ![Diagram of modification]
  - information source
  - information destination
  - “bad guy”

Classes of Threat

- fabrication: information arrives at its destination but not from presumed source
  ![Diagram of fabrication]
  - information source
  - information destination
  - “bad guy”
Security Threats

- what threat is shown here? how countermanded?

Assets

- hardware:
  - can be stolen (entirely or partially)
  - can be damaged
  - hard to automate protection schemes

Assets

- hardware
- software
- data:
  - may be destroyed accidentally or maliciously
  - may be modified accidentally or maliciously
  - may be unintentionally released

Security Tools

- physical
  - locked doors, access control to physical resources

Security Tools

- physical
  - operational:
    - locked doors, access control to physical resources
    - monitoring
    - backups (RAID, off-site storage)
    - disaster testing/readiness exercises
Security Tools

- physical
- operational
- integrity:
  - use of security hashes, msg digests, checksums

MDS Checksum

- perform operation on file that results in a 128-bit value that only that file contents can produce
  - change one bit of file, MDS sum is different
- users perform mdssum computation on data they receive, compare with sum advertised by provider
  - if sums don’t match, file has been adulterated
  - e.g., MDS sum for lecture 11’s 2-up pdf handouts is: 954dac5ba7e8bdc5c09699999545ef
- other popular techniques: SHA-1
  - produces 160-bit result
  - e.g., ab7d4bb1ebec49886122ed12338b14dc03d31b

Confidentiality Security Tools

- encryption:
  - ‘scrambling’ of bits so that message content is unintelligible to third party
  - un-scrambling operation = decryption
  - requires use of ≥ 1 keys
    - keys involved in both encryption and decryption
    - key values must not be known to third parties
  - key exchange: secure way to communicate key values between communicating parties

Encryption Systems

- receive as input plain text, P
  - i.e., original, “in the clear” message
  - using a key, apply a sequence of well-defined operations to plain text to generate ciphertext C
  - 2 major kinds of system:
    - use one key = symmetric key encryption
    - use two keys = public key encryption (PKE)
Encryption Systems: Keys

- key value must remain secret or any third party with key can decrypt message
- how deliver key to recipient?
  - by hand
  - using another encryption scheme
    - e.g., Diffie-Hellman key exchange

Keys: Size Matters

<table>
<thead>
<tr>
<th>Key Size</th>
<th>1/µs</th>
<th>10¹⁰/µs</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>35.8 min</td>
<td>2.13 ms</td>
</tr>
<tr>
<td>56</td>
<td>1142 yrs</td>
<td>10.01 hrs</td>
</tr>
<tr>
<td>80</td>
<td>19,154 yrs</td>
<td>7 days</td>
</tr>
<tr>
<td>128</td>
<td>5.4 x 10⁷ yrs</td>
<td>5.4 x 10⁷ yrs</td>
</tr>
<tr>
<td>168</td>
<td>5.9 x 10⁸ yrs</td>
<td>5.9 x 10⁸ yrs</td>
</tr>
</tbody>
</table>

- no one really takes under 80 bits very seriously today
- usually want 128 or better

Breaking Encryption Schemes

- bad guys listen to encrypted message and try to decrypt it without key
- need to ‘guess’ key
  - keep trying key values until success
  - worst case: have to try them all
  - expected value: typically need to try half of all key values
- jackpot: be able to determine key for any message ‘easily’
  - break algorithm, not just individual message

Public Key Encryption

- encryption and decryption algorithms are well known
- encryption key is made public but decryption key is kept secret
- PK algorithms based on easy to compute functions whose inverses are very hard to find
  - e.g., base on problem of factoring two large prime numbers (RSA algorithm)

DES Iterations

- DES is symmetric–key block cipher
  - originally from IBM
  - developed further by NIST
- was broken by EFF in 1998
- now use triple–DES (TDEA, 3DES)
  - effective key length 168 bits (3 x 56)

RSA

- key pairs
- encryption
- CIPHER TEXT
- decryption
- PLAIN TEXT
- PLAIN TEXT
- private key
- public key
- encryption
Who Can You Trust?

- suppose Bob encrypts a session key with his private key then Alice’s public key
- to be sent to Alice
- how does Bob know that the key he got as Alice’s public key really is Alice’s public key?
- because she says it is?

Signing Authorities

- Alice can register her public key with a trusted signing authority.
- she sends her public key
- they (somehow) ensure the key is really hers
- they ‘sign’ her key attesting its genuineness
- generate MD5 hash of this message
- encrypt result with their private key
- result is a certificate
- now anyone can verify key using signing authority’s public key

Certificates

- X.509 (v3) standard certificate:
  - see, e.g., RFC1991, RFC2440
  - X.509 Simple Public Key Infrastructure

- some other types of certificates:
  - PGP: Pretty Good Privacy
  - SKI: Simple Key Infrastructure

Implementations

- Secure Sockets Layer (SSL):
  - provides certificate-based authentication
  - server – client
  - client – server
  - provides key-exchange
  - provides encryption
  - open-source version available (for all platforms):
    OpenSSL [http://www.openssl.org](http://www.openssl.org)

Stunnel Session Log

```plaintext
$9900 started
$9900 connected from 127.0.0.1:9900
sslverizon: mk
SLL state (connect): before/connect initialization
SLL state (connect): SSLv3 write client hello A
SSL state (connect): SSLv3 read server hello B
SSL state (connect): SSLv3 write client key exchange A
SSL state (connect): SSLv3 write client cert A
SSL state (connect): SSLv3 write server key exchange A
SSL state (connect): SSLv3 write certificate verify A
SSL state (connect): SSLv3 write finished A
SSL state (connect): SSLv3 write finished A
SSL state (connect): SSLv3 write finished A
SSL state (connect): SSLv3 write finished A
SSL state (connect): SSLv3 write finished A
SSL state (connect): SSLv3 read finished A
Protocol: [SSLv3 Kea=MD
Authentication: [MD5-SHA1=SSLv3 hmac]
```
**Stunnel Session Log**

SSL alert (read): warning: close notify
SSL closed on SSL_read
SSL write shutdown (output buffer empty)
Socket write shutdown (output buffer empty)
SSL alert (write): warning: close notify
SSL_handshake finished; sent close_notify
Connection closed; 8 bytes sent to SSL; 805 bytes sent to socket
SSL Finished (0 bytes)

**Privacy and NEW**

![Diagram](image)

**Escrowed Keys**

- In 1990s, law enforcement drove idea of:
  - Require (by law) everyone to use an NSA developed encryption algorithm (Skipjack)
  - 80-bit encryption
  - But has ability to deliver 'master-key' to let any msg so encrypted be decrypted by law enforcement
- Needed pair of keys to make up 'master'
  - One held by NIST
  - Other held by Dept. of the Treasury
- Not put into force; algorithm now openly avbl