

Lecture 02

WIS Introduction

By A/Prof Heng Tao SHEN
School of Information Technology and Electrical Engineering
The University Of Queensland

P. 2

Outline

- What is a Web Information System?
- About the Internet and the Web
 - Internet, WWW, servers, clients, browsers, URLs, HTTP, ...
- Modern WIS development technologies
 - HTML, Applets, JavaScripts, AJAX, Servlets, JSP, ASP, ...
 - J2EE and .NET
 - XML, Web Services, SaaS
 - Security
 - ...

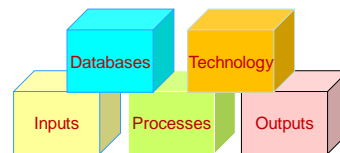
Information and System

- What is **Information**?
 - Data put into meaningful and useful context and communicated to recipients who use it to make decisions
 - Quality of information
 - Accuracy, timeliness and relevance
- What is a **System**?
 - A set of components that interact to accomplish some purpose
 - legal system, banking system, education system ...

P. 3

Information Systems

- What is an **Information System**?
 - interaction between people, processes, data and technology
 - building blocks



P. 4

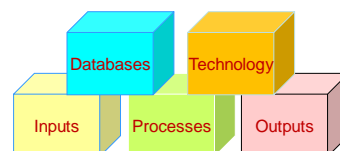
Web Information Systems

- What is a **Web Information System**?
 - An information system using WWW as part of the deployment and implementation technology
 - Deployment: Web as communication mechanism
 - Implementation: Web as distributed computing platform
- Different Types of WIS
 - Informational
 - Disseminate information (static or dynamic)
 - Share information (among visitors or WISs)
 - Transactional
 - Conduct transactional business (B2C, C2C, B2B...)
 - Perform tasks (simple, complex and collaborative)

P. 5

WIS Components

- Still the same components, but with new meanings now



P. 6

Database Component

- Database
 - DB is a structured collection of related facts
 - DBMS is a collection of programs that enables users to create and maintain a DB
 - The three components of a database system
 - DB, DBMS and applications
- New dimensions
 - DB for the Web
 - Massive amount, new data types, online performance...
 - XML database, Web interfaces and APIs...
 - Web as DB
 - The Web is a giant distributed database (or, is it?)

P.7

Technology Component

- Hardware devices
 - CPU, storage, input devices, printers ...
- Communications
 - Transmission media (fibre optic cables, radio or satellite transmission, microwave transmission ...)
 - Topology (Star, Ring, Bus ...)
- Software
 - Operating Systems (Unix, Windows ...)
 - Network Protocols (TCP/IP, Internet ...)
 - Programming languages (C, C#, Java, Perl, scripting languages...)
 - Others (various algorithms and services, data exchange protocols...)

P.8

The Internet

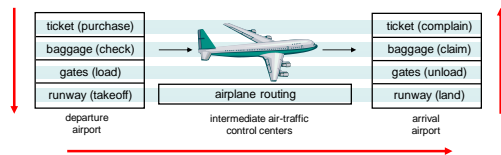
- The Internet is a huge collection of interconnected computers or other devices, since 1995
- A single protocol
 - TCP/IP (transmission control protocol/ internet protocol)
 - Application (process-to-process) layer: HTTP, SOAP, IMAP, ...
 - creates and communicates data to other processes or applications
 - Transport (host-to-host) layer: TCP, UDP, DCCP, ...
 - deals with opening and maintaining connections
 - Network layer: IPv4 (32-bit), IPv6 (128-bit), ...
 - moves data packets from node to node
 - Link layer: ARP, RARP, MAC, ...
 - describes local network topology and the interfaces
 - Full names?
 - Hypertext Transfer Protocol (HTTP); Internet message application protocol (IMAP); Simple Object Access Protocol (SOAP); User Datagram Protocol (UDP); Datagram Congestion Control Protocol (DCCP); Address Resolution Protocol (ARP); Media Access Control (MAC); ...

Application
Transport
Network
Link

P.9

An Airplane Trip

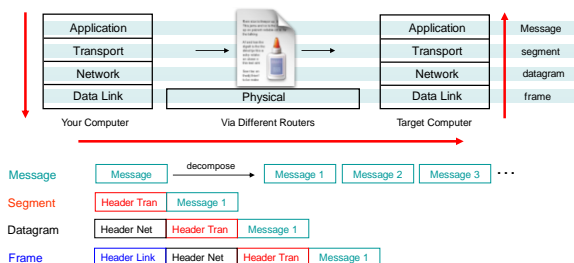
- Before we talk about how the data are sent via the network...
 - Let's consider: How we take an airplane trip?



P.10

Internet Communication

- Internet as a layered architecture

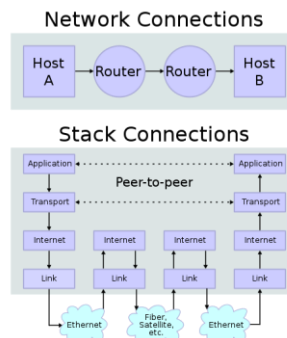


Question 1: Why we have to decompose the original message?

Question 2: Why we have to add "header"

P.11

Another view with two routers



P.12

How the Layers Communicate?

- Using Protocol
 - Set of rules/agreements in telecommunication
 - So that we can understand (encode/decode) the data by following some standards
- Note
 - Same layer in different computers is communicated by using some standard protocols
 - Different layers do not need to communicate
 - Layer A only needs to route the data to Layer B. Layer B does not need to *understand* the data
 - Difficult to understand? The immigration department does not care where have you bought the ticket.

P. 13

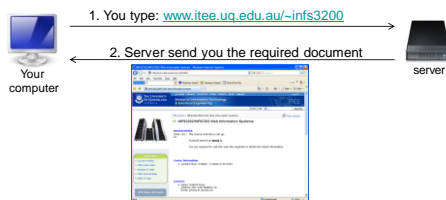
Internet Protocol Suite (1/6)

- Strictly speaking:
 - When you try to exchange data via Internet Protocol Suite, you need to specify protocol and port number in the application.
 - A computer just like an airport.
 - A port just like a gate in the airport.
 - There are totally 65535 ports
 - E.g. Access web page – 80; Send email – 25; ...
 - For example, the format of URL (Uniform Resource Locator) is:
 - Format:
 - Protocol :// Host : Port / File
 - Example:
 - http :// www.google.com.hk : 80 / index.html
 - try with different port number, e.g. 25
 - Do you know why we never need to specify 80?
 - You do not need to specify them because the application already "completed" for you

P. 14

Internet Protocol Suite (2/6)

- Example:
 - Request a web page
 - www.itee.uq.edu.au/~infs3202
 - From a user point of view:



- What is the underlying mechanism?

P. 15

Internet Protocol Suite (3/6)

- The basic idea:

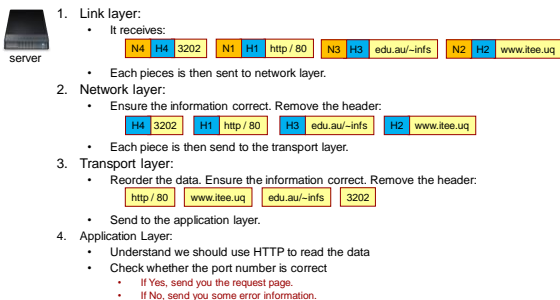


- After you typed: www.itee.uq.edu.au/~infs3202
 - Application layer determine your protocol and port
 - For web searching, by default, it is "http" (protocol) and "80" (port)
 - Application layer decompose your message into several smaller ones:
 - http / 80 www.itee.uq edu.au/~infs 3202
 - Each piece is then send to the transport layer
- Transport layer:
 - TCP (Transmission Control Protocol) add some information.
 - H1 http / 80 H2 www.itee.uq H3 edu.au/~infs H4 3202
 - Each piece is then send to the network layer
- Network layer:
 - IP (Internet Protocol) add some information.
 - N1 H1 http / 80 N2 H2 www.itee.uq N3 H3 edu.au/~infs N4 H4 3202
 - Each piece is then send to the link layer
- Link layer:
 - Transmission begins. Each piece of data is regarded as a frame
 - Header information should be added, but is omitted in this example for clarity

P. 16

Internet Protocol Suite (4/6)

- Due to a number of reasons, the frames received at the sever side may be out of order.



P. 17

Internet Protocol Suite (5/6)

- If a frame is lost...



- Link layer:
 - It receives:
 - N4 H4 3202 N1 H1 http / 80 N2 H2 www.itee.uq
- Network layer:
 - Ensure the information correct. Remove the header:
 - H4 3202 H1 http / 80 H2 www.itee.uq
- Transport layer:
 - Reorder the data
 - H1 http / 80 H2 www.itee.uq H4 3202
 - Identify a data is missing (H3). Request the missing data.
 - Send information to network layer:
 - Request H3
- Network layer
 - Add some information and send to link layer:
 - N Request H3
- Link layer:
 - Route back to your computer to request the missing information

P. 18

Internet Protocol Suite (6/6)

(Cont'd)



Your computer

1. Link layer:
 - It receives:
 - N** Request H3
2. Network layer:
 - Ensure the information correct. Remove the header:
 - Request H3**
3. Transport layer:
 - Identify it is a request of a missing data (H3). Pass the missing data to the network data.
 - H3** edu.au~infs
4. Network layer
 - Add some information and send to link layer:
 - N3 H3** edu.au~infs
5. Link layer:
 - Route to the destination computer again.

The whole process continue until the transport layer at the server side received all of the data, and then pass to the application layer in there.

P. 19

Application Layer and Transport Layer

- Application layer:
 - Support network applications
 - Internet, Email, Instant messaging (MSN, Skype), P2P file transfer (BT), Network Game, Remote desktop, Internet telephony ...
 - Web – HTTP (Hyper Text Transfer Protocol)
 - Email – SMTP (Simple Mail Transfer Protocol)
 - File Transfer – FTP (File Transfer Protocol)
 - ...
- Transport layer
 - Manage the received data
 - TCP - provides reliable, ordered delivery
 - UDP - provides unreliable service and datagrams may arrive out of order, appear duplicated, or go missing without notice; stateless nature is also useful for servers answering small queries from huge numbers of clients, such as IPTV, online games, etc

P. 20

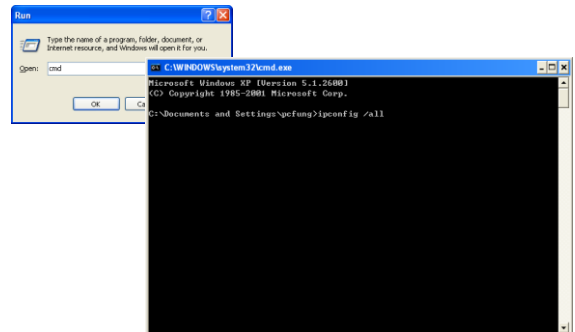
Network Layer (1/2)

- Network layer
 - Route the data
 - IP – Internet Protocol
 - In order to route the data, we need two information:
 - IP Address (like your home address)
 - Logical address
 - In the following form: xxx.xxx.xxx.xxx
 - Identify the physical location of the computer
 - » In UQ domain, all computers are: 130.102.xxx.xxx
 - Media Access Control (MAC) Address (like your passport)
 - Physical address
 - The identity of the network card
 - » of coz, without a network card, you cannot online...
 - ALL network card in the world have DIFFERENT MAC Address!!!
 - » IEEE standard – 2²⁴ combinations, i.e. 281474976710656

P. 21

Network Layer (2/2)

- Find your IP address and MAC address:



Link and Physical Layer

- Link and Physical Layer:
 - Data transfer between neighbouring network elements
 - Bits on the wire

P. 23

The Web

- A hypermedia-based system that provides a simple 'point and click' means of browsing information on the Internet using hyperlinks
 - Web servers and clients
 - Documents are written in HTML
 - Servers and clients communicate via HTTP
 - Documents and locations are identified using URL

P. 24

Web Servers

- Web servers are programs that provide documents to browsers
 - Documents → resources (thus, dynamic results)
 - Typically a request is independent of each other, even when they are from the same client
- Some popular Web server software
 - Apache
 - Microsoft Internet Information Services (IIS) and Personal Web Server (PWS)
- The number of Web servers is huge

P. 25

Web Browsers

- A client side software to request, receive and process web pages from a Web server
 - IE, Netscape, Mozilla, Opera...
 - All can handle HTML and HTTP
 - Very sophisticated now
 - HTML/XHTML, CSS, various scripts, XML...
 - Which browser you client uses?
 - In theory, it does not matter as they should all be standard-compliant
 - In reality, they take "suggestions", not "instructions"
 - In practice, you have to deal with this heterogeneity
 - Browser hosts are diversified too
 - Hand-held, wireless devices and appliance-based

P. 26

HTTP

- HTTP is the **protocol** used to transfer Web pages through the Internet, at application layer
 - One protocol for the entire Web
- Based on request-response paradigm
 - **Stateless connection**: the server has no memory of previous ones (i.e., a request is closed once responded)

P. 27

HTML/XHTML

- HTML is the document formatting language used to design most Web pages
 - HyperText Markup Language
- An HTML document is a mixture of content and controls
 - Regular text, and tags to specify their appearance
- XHTML 1.0 becomes the standard in 2000
 - Supporting XML, more rigid and richer controls, separation of presentation and structure of a document...

P. 28

URL

- URL is a string of alphanumeric and characters that represents the location or address of a resource on the Internet and how that resource should be accessed
 - Uniform Resource Locators
 - The basis for HTTP and the Web
- Syntax
 - `<protocol>://<host>[:<port>]/<path_to_document[? Arguments]`
 - EG, `http://www.itee.uq.edu.au/~infos3202/profile.pdf`
 - Protocols: http, file, mailto, ftp,
 - Path: complete or partial, depending on web server configuration
 - Many characters, such as space, '+', ':', '&' cannot be used

P. 29

Client Side Technologies

- Embedded in HTML, and interpreted by the Web browser
- Client-side processing
 - For animations, user interactions and validations
 - First introduced by Java Applet
 - Now dominated by HTML/CSS/DOM/JavaScript/Ajax
- Motivation
 - Utilisation of client-side resources to save bandwidth and server-side load
 - Key consideration: client-side computing power

P. 30

Server Side Technologies

- Generated by some scripts that convert live data to HTML
 - These scripts need to understand the request from the clients and the results generated by applications that use databases
- Technologies
 - Many competing technologies, with identical or very similar features, capabilities and architectures
 - CGI (common gateway interface), Server-Side Includes, HTTP Cookies, Java (and Applets and Servlets), Scripting Languages (Perl, PHP, JavaScript, VBScript...), XML...
 - Dominated by JSP and ASP, both with database and XML support
 - HTML template with embedded scripts
 - Runtime execution to translate into pure HTML docs

P. 31

J2EE vs .NET

- J2EE
 - One language, any platform
- .NET
 - Any language, one platform
- They offer pretty much the same list of features, albeit in different ways

P. 32

XML

- eXtensible Markup Language
 - One of the most important technologies to support WWW
 - GML (1969), SGML (1974-1986), HTML (1990-1993), XML (1996-1998), XHTML(2000)
- HTML: how a document looks like
 - A fixed set of tags whose meaning is fixed
- XML: what a document means
 - Allows the user to define tags
 - Separation of data and its presentation
 - CSS and XSL (eXtensible Style Language)
 - Essential for data exchange and customization

P. 33

A Meta-language

- A Language used to define other languages
 - XML itself has no tags
 - But it can be used to define a set of tags
 - Eg, XHTML
 - Language here means a set of tags
 - Vocabulary: elements and attributes
 - Grammar: where a tag can appear and how they can appear
 - From this sense, a table in a database is a 'language'

P. 34

Web Services

- A client of a website can be an application!
 - This is the future envisioned by B2B
 - Automatic production and consumption of Web data via Internet protocols such as HTTP
 - The need of self-describing data
 - The publication of remote Web functions
 - Abstraction and integration of distributed Web functionality
- A Web service is a remote component containing functionality that can be invoked using Internet protocols
- Web Services technology
 - XML
 - The following 3 technologies are all specified in XML
 - SOAP (Simple Object Access Protocol)
 - WSDL (Web Service Description Language)
 - UDDI (Universal Description, Discovery and Integration)

P. 35

SaaS

- **Software as a service**, a model of software deployment whereby a provider licenses an application to customers for use as a service on demand
 - SaaS software vendors (such as CyberShift or SalesForce.com) may host the application on their own web servers or upload the application to the consumer device, disabling it after use or after the on-demand contract expires
- Key Features:
 - Network-based access to commercially available software
 - Activities managed from central locations
 - Single instance to multi-tenant
 - Frequent integration into a larger network - either as part of a mashup or as a plugin to a platform as a service
- What about **Service Oriented Architecture**?

P. 36

Web Security

- Transmit information over the Internet such that:
 - It is accessible to the sender and the receiver only (**privacy**)
 - It cannot be changed during transmission (**integrity**)
 - The receiver can be sure it came from the sender (**authenticity**)
 - The sender can be sure the receiver is genuine (**non-fabrication**)
 - The sender cannot deny s/he sent it (**non-repudiation**)
- Security is beyond transmission
- Protection required for clients as well as the servers
 - Various malicious attacks
- Many technologies
 - Firewall, digital certificates and signatures, SSL (Secure Socket layer), ...

P. 37

WIS Requirements

- Security and integrity
 - Access corporate data in a secure manner
 - Conduct business online in a secure manner
 - Identification and authentication (session- and application-based) etc
 - Support multiple HTTP request transactions
- Future-proof
 - Data and vendor independent connectivity
 - Web browser independent interface
 - Open-architecture for interoperability
- Performance
 - Fully fledged database connectivity
 - Efficient and scalable
 - High-level productivity tools for development, maintenance and deployment
- However, it is hard to satisfy all...

P. 38

WIS Architectures

- Two-tier architecture
 - Client (user interface + main business logic)
 - Server (some business logic + database management)
 - Thin and fat client
- Three-tier architecture
 - Client (user interface)
 - Application server (business logic)
 - Database server (data management)
 - The three-tier architecture has many advantages, one of them is a natural mapping to the Web environment

P. 39

Summary

- An introduction to Web Information Systems
- Some important concepts and technologies
- Next week:
 - Client-side Development I - HTML, XHTML, CSS

P. 40

References

- Robert W. Sebesta. Programming the World Wide Web. 3rd Edition, Chapter 1, 2005.
- Behrouz A. Forouzan. Data Communications and Networking. McGraw Hill. 4th Edition. 2007.
- Internet Protocol Suite:
http://en.wikipedia.org/wiki/Internet_Protocol_Suite

P. 41