IS 651: Distributed Systems
Chapter 1: Distributed Systems
Introduction

Jianwu Wang

Spring 2016
What is this course about?

• Overview of the concepts, systems and techniques of distributed systems
  ▪ Basic concepts and principles of distributed systems, which are useful in many real-world applications/projects
  ▪ Lectures/readings plus extensive hands-on exercises/homework
  ▪ By the end of this course, you will have a good technical understanding of many distributed system related technologies

• NOT a programming class
  ▪ We will use several languages (XML, JavaScript, Xquery, etc.), only the basics for exercises/homework
  ▪ No real programming. Only need to understand programs and make some changes

• More technical than MIS, more application-oriented than CS
Why this course is important?

• People are using distributed systems everyday
  ▪ Web sites: Gmail, Facebook, ...
  ▪ Distributed databases: MySQL Cluster, Hbase, ...
  ▪ Distributed file systems: NSF, AFS, ...
  ▪ Distributed scientific software: MPI, OpenMP, ...

• Knowing the knowledge might help your future career
  ▪ One of most actively evolving topic: Cloud, Big Data, Mobile, GPU, ...
  ▪ Design/implement a new distributed system
  ▪ Running data analytics on a distributed system
  ▪ Managing a distributed database
  ▪ ...
How to Study Well?

• Start exercise and homework early
• Read the requirements carefully
• Read book chapters, especially before exams
• If needed, ask for help early
  ▪ If you are experiencing any problems that affect your performance in this class, please contact us immediately
• Study hard from the beginning of the semester
Learning Outcomes

• After learning chapter 1, you should be able to
  ▪ Understand the basics of distributed systems and service oriented architecture (SOA)
  ▪ Understand the layered architecture of network
  ▪ Can access and navigate Linux server (gl.umbc.edu)
Distributed Systems

• A distributed system consists of multiple autonomous computers that communicate through a computer network. The computers interact with each other in order to achieve a common goal.

• Definition: A distributed system is a collection of independent computers that appears to its users as a single coherent system.
Service-Oriented Architecture (SOA)

• SOA is an evolution of distributed computing
• SOA defines how two computing entities, such as programs, interact in such a way as to enable one entity to perform a unit of work on behalf of another entity.
• Service interactions are defined using a description language. Each interaction is self-contained and loosely coupled, so that each interaction is independent of any other interaction.
• This is an abstract architectural concept and no specific technology is assumed
SOA Concepts

• Top-down
• Bottom-up
• Loosely-coupled
• Interface, not application
XML Web Service

• XML Web Services are the current technology most associated with SOA

• A Web Service (XML Web Service) is a unit of code that can be activated using HTTP requests. Stated another way, a Web Service is an application component (in any language) that can be remotely callable using standard Internet Protocols such as HTTP and XML.

• Web Services came into existence to deliver distributed computing over the Internet.

• A major advantage of the Web services architecture is, it allows programs written in different languages on different platforms to communicate with each other in a standards-based way.
Traditional B2B Scenario

• To support the company's inventory query and supply ordering with no web services, the enterprise architecture requires that four systems be connected using three proprietary interfaces.

B2B Scenario with SOA

• SOA opens up a number of new possibilities for conducting B2B commerce without significant reworking of the underlying systems.

• In addition to eliminating the proprietary interfaces, the SOA makes it easily possible for the first plant to check directly with the second plant and place orders without going through the HQ computer.
B2B Scenario with More SOA

- The manufacturer now wants to institute an electronic competitive bidding system for its orders.
- The suppliers who want to bid on the opportunity to win business from the manufacturer can connect to the bidding system through a Web service.
Course Prerequisites

• Programming with systems analysis and design
• Networking
• Databases
Networking - OSI Seven Layer Model

<table>
<thead>
<tr>
<th>Media/Host</th>
<th>Data Unit</th>
<th>Layer</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Host</td>
<td>Data</td>
<td>7. Application</td>
<td>HTTP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6. Presentation</td>
<td>Representation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5. Session</td>
<td>Dialogue</td>
</tr>
<tr>
<td>Segment</td>
<td></td>
<td>4. Transport</td>
<td>End-to-end</td>
</tr>
<tr>
<td>Media</td>
<td>Datagram</td>
<td>3. Network</td>
<td>Routing</td>
</tr>
<tr>
<td></td>
<td>Frame</td>
<td>2. Data Link</td>
<td>MAC address</td>
</tr>
<tr>
<td></td>
<td>Bit</td>
<td>1. Physical</td>
<td>Signals</td>
</tr>
</tbody>
</table>
Networking - Peer-to-Peer (P2P) Communication

- Peer-to-peer communications through protocol data unit (PDU) headers
- The data unit of one layer is part of data unit of its underlying layer
Networking - TCP/IP

From http://www.whatisnetworking.net/tag/advantages-and-disadvantages-of-tcpip-model/
Distributed Systems Topics

• **Architecture**: how distributed systems are put together to provide a single system abstraction from many separate parts.

• **Fault-tolerance**: how distributed systems can continue to provide service when some parts have failed to provide availability until they are recovered.

• **Consistency**: how distributed systems can maintain logical coherency when data is distributed.

• **Scalability**: how a distributed system can grow to meet demand in an efficient and effective way.

• **Performance**: how to optimize response time when components are distributed.

• **Security**: how to ensure data integrity and confidentiality is a distributed environment.
Eight Fallacies for Distributed Systems

• The network is reliable.
• Latency is zero.
• Bandwidth is infinite.
• The network is secure.
• Topology doesn't change.
• There is one administrator.
• Transport cost is zero.
• The network is homogeneous.
Unix/Linux Command-line

• Use of an SSH client
  • Mac: Terminal command line
  • Windows: Putty, Bitvise SSH Client

• Use of an SCP client
  • Mac: Terminal command line, cyberduck
  • Windows: cyberduck, Bitvise SSH Client, WinSCP

• Use of a small number of unix commands
  • ls, mkdir, cd, pwd, man, more, cat, ...

• Edit a text file on a unix machine
  • pico, vi, nano, emacs, ...

• Navigation of your student account on gl.umbc.edu
Exercise/Homework Notes

• You can help each other for exercises, but not homework
  ▪ No communication before homework submission

• Same exercise/homework code from multiple students will be investigated
  ▪ No cheating

• Make sure you submit your exercise/homework by the deadline
  ▪ Grader will grade them the next morning

• Any time an exercise/homework asks you to put information into a file with a .txt extension, it must be a plain text file
  ▪ Never use a word processor format