Distributed Systems

03. Remote Procedure Calls

Paul Krzyzanowski
Rutgers University
Fall 2014
Networking Recap
Multiple Access Protocols

- Share a communication medium
- Statistical multiplexing
  - No timeslots
  - Anyone can transmit when ready
  - But be prepared for collisions or dropped packets
The Internet

- Internet
  - Logical network on top of multiple physical networks
  - Packets routed through routers: store & forward
  - Fault tolerant
    - No central point of control
    - Packets can take alternate routes
    - Best effort delivery: no guarantees

- End-to-end principle → smart endpoints, dumb network
  - Only network endpoints (hosts) should store state
  - Any application-specific functions should be implemented at the hosts
    - end-to-end vs. hop-by-hop
  - E.g., TCP state is preserved even if traffic moves to different routers
IP vs. OSI protocol stack

Internet protocol stack

1. Physical
2. Data Link
3. Network
4. Transport
5. Session
6. Presentation
7. Application

Middleware

OSI protocol stack

1. Physical
2. Data Link
3. Network
4. Transport
5. Session
6. Presentation
7. Application

TCP, UDP
IP, IPv6
Ethernet MAC

September 15, 2014
© 2014 Paul Krzyzanowski
IP transport layer protocols

Enable applications to communicate

– **TCP: Transmission Control Protocol**
  - Connection-oriented service – operating system keeps state
  - Full-duplex connection: both sides can send messages over the same link
  - Reliable data transfer: the protocol handles retransmission
  - In-order data transfer: the protocol keeps track of sequence numbers
  - Flow control: receiver stops sender from sending too much data
  - Congestion control: “plays nice” on the network

– **UDP: User Datagram Protocol**
  - Connectionless service: lightweight transport layer over IP
  - Data may be lost
  - Data may arrive out of sequence
  - Checksum for corrupt data: operating system drops bad packets
  - Sender can keep blasting data to the receiver even if it can’t keep up
Point-to-Point Communication
Socket-based communication

• Socket API: all we get from the OS to access the network
• Socket = distinct end-to-end communication channels
• Read/write model
  – Send a bunch of bytes
  – Read a bunch of bytes
  – Send a bunch of bytes
  – Read a bunch of bytes
  – …
• Application implements its protocol
• Line-oriented, text-based protocols common
  – Not always efficient but easy to debug & use
Connection-Oriented (TCP) socket operations

Client

1. Create a socket
2. Name the socket (assign local address, port)
3. Connect to the other side
4. read / write byte streams
5. close the socket

Server

1. Create a socket
2. Name the socket (assign local address, port)
3. Set the socket for listening
4. Wait for and accept a connection; get a socket for the connection
5. read / write byte streams
6. close the socket
7. close the listening socket

socket
bind
connect
read/write
close
socket
bind
listen
accept
read/write
close
close
Connectionless (UDP) socket operations

Client

- Create a socket
- Name the socket (assign local address, port)
- Send a message
- Receive a message
- close the socket

Server

- Create a socket
- Name the socket (assign local address, port)
- Receive a message
- Send a message
- close the socket
Java provides shortcuts that combine calls

Example

Java

```
Socket s = new Socket("www.rutgers.edu", 2211)
```

C

```
int s = socket(AF_INET, SOCK_STREAM, 0);

struct sockaddr_in myaddr; /* initialize address structure */
myaddr.sin_family = AF_INET;
myaddr.sin_addr.s_addr = htonl(INADDR_ANY);
myaddr.sin_port = htons(0);
bind(s, (struct sockaddr *)&myaddr, sizeof(myaddr));

/* look up the server's address */
struct hostent *hp;   /* host information */
struct sockaddr_in servaddr; /* server address */
memset((char *)&servaddr, 0, sizeof(servaddr));
servaddr.sin_family = AF_INET;
servaddr.sin_port = htons(2211);
hp = gethostbyname("www.rutgers.edu");

if (connect(fd, (struct sockaddr *)&servaddr, sizeof(servaddr)) < 0) { 
    /* connect failed */
}
```
Sample SMTP Interaction

$ telnet cs.rutgers.edu 25
Trying 128.6.4.2...
Connected to cs.rutgers.edu.
Escape character is '^]'.
220 aramis.rutgers.edu ESMTP Sendmail 8.11.7p3+Sun/8.8.8; Mon, 15 Sep 2014 12:12:01 -0400 (EDT)
HELO pk.org
250 aramis.rutgers.edu Hello aramis.rutgers.edu [128.6.4.2], pleased to meet you
MAIL FROM: <pxk@cs.rutgers.edu>
250 2.1.0 <pxk@cs.rutgers.edu>... Sender ok
RCPT TO: <p@pk.org>
250 2.1.5 <p@pk.org>... Recipient ok
DATA
354 Enter mail, end with "." on a line by itself
From: Paul Krzyzanowski <pxk@cs.rutgers.edu>
Subject: test message
Date: Mon, 11 Feb 2013 17:00:16 -0500
To: Whomever <testuser@pk.org>

Hi,
This is a test
.
250 2.0.0 s8FGE2S13883 Message accepted for delivery
quit
221 2.0.0 aramis.rutgers.edu closing connection

This is the message body. Headers may define the structure of the message but are ignored for delivery.
TCP Sockets: Disambiguating Messages

host 1.2.3.4

P₀  port 2036
P₁  port 2037

host 2.2.2.2

P₃  port 25

on host 2.2.2.2

<table>
<thead>
<tr>
<th>Local</th>
<th>Port</th>
<th>Source Host</th>
<th>Source Port</th>
<th>Listen?</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>2.2.2.2</td>
<td>25</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>s₁</td>
<td>2.2.2.2</td>
<td>25</td>
<td>1.2.3.4</td>
<td>2036</td>
</tr>
<tr>
<td>s₂</td>
<td>2.2.2.2</td>
<td>25</td>
<td>1.2.3.4</td>
<td>2037</td>
</tr>
<tr>
<td>s₃</td>
<td>2.2.2.2</td>
<td>25</td>
<td>1.2.3.5</td>
<td>2036</td>
</tr>
</tbody>
</table>

on host 2.2.2.2

host 1.2.3.5

P₂  port 2036
Remote Procedure Calls
Problems with the sockets API

The sockets interface forces a read/write mechanism

Programming is often easier with a functional interface

To make distributed computing look more like centralized computing, I/O (read/write) is not the way to go
1984: Birrell & Nelson
– Mechanism to call procedures on other machines

Remote Procedure Call
You write:

\[ x = f(a, \text{"test"}, 5); \]

The compiler parses this and generates code to:

a. Push the value 5 on the stack
b. Push the address of the string “test” on the stack
c. Push the current value of a on the stack
d. Generate a call to the function f

In compiling \( f \), the compiler generates code to:

a. Push registers that will be clobbered on the stack to save the values
b. Adjust the stack to make room for local and temporary variables
c. Before a return, unadjust the stack, put the return data in a register, and issue a return instruction
Implementing RPC

No architectural support for remote procedure calls

*Simulate it* with tools we have
(local procedure calls)

Simulation makes RPC a language-level construct

instead of an operating system construct

The compiler creates code to send messages to invoke remote functions

The OS gives us sockets
The trick:

Create **stub functions**
to make it appear to the user that the call is local

On the client
The stub function has the function’s interface
Packages parameters and calls the server

On the server
The stub function (skeleton) receives the request and calls the local function
1. Client calls stub (params on stack)

- Client functions
- Client stub (proxy)
- Network routines
- Server functions
- Server stub (skeleton)
- Network routines

OS

OS
2. Stub **marshals** params to net message

Marshalling = put data in a form suitable for transmission over a network (serialized)
3. Network message sent to server

- **Client functions**
  - Client stub (proxy)
  - Network routines

- **Server functions**
  - Server stub (skeleton)
  - Network routines

OS

© 2014 Paul Krzyzanowski
4. Receive message: send it to server stub

- **Client functions**
  - Client stub (proxy)
  - Network routines

- **Server functions**
  - Server stub (skeleton)
  - Network routines
5. Unmarshal parameters, call server function

- Client functions
- Client stub (proxy)
- Network routines
- Server stub (skeleton)
- Server functions
- Network routines
6. Return from server function
7. Marshal return value and send message
8. Transfer message over network

- **Client functions**
  - Client stub (proxy)
  - Network routines

- **Server functions**
  - Server stub (skeleton)
  - Network routines

**OS**
9. Receive message: client stub is receiver
Stub functions

10. Unmarshal return value, return to client code
A proxy looks like the remote function

• Client stub has the same interface as the remote function

• Looks & feels like the remote function to the programmer
  – But its function is to
    • Marshal parameters
    • Send the message
    • Wait for a response from the server
    • Unmarshal the response & return the appropriate data
    • Generate exceptions if problems arise
A skeleton is really two parts

- **Dispatcher**
  - Receives client requests
  - Identifies appropriate function (method)

- **Skeleton**
  - Unmarshals parameters
  - Calls the local server procedure
  - Marshals the response & sends it back to the dispatcher

- **Invisible to the programmer**
  - The programmer doesn’t deal with any of this
  - Dispatcher + Skeleton may be integrated
    - Depends on implementation
RPC Benefits

• RPC gives us a procedure call interface

• Writing applications is simplified
  – RPC hides all network code into stub functions
  – Application programmers don’t have to worry about details
    • Sockets, port numbers, byte ordering

• Where is RPC in the OSI model?
  – Layer 5: Session layer: Connection management
  – Layer 6: Presentation: Marshaling/data representation
  – Uses the transport layer (4) for communication (TCP/UDP)
RPC has challenges
Parameter passing

Pass by value
  - Easy: just copy data to network message

Pass by reference
  - Makes no sense without shared memory
Pass by reference?

1. Copy items referenced to message buffer
2. Ship them over
3. Unmarshal data at server
4. Pass *local* pointer to server stub function
5. Send new values back

To support complex structures
- Copy structure into *pointerless* representation
- Transmit
- Reconstruct structure with local pointers on server
Representing data

No such thing as *incompatibility problems* on local system

Remote machine may have:

– Different byte ordering
– Different sizes of integers and other types
– Different floating point representations
– Different character sets
– Alignment requirements
Representing data

IP (headers) forced all to use **big endian** byte ordering for 16- and 32-bit values

**Big endian:** Most significant byte in low memory
- SPARC < V9, Motorola 680x0, older PowerPC

**Little endian:** Most significant byte in high memory
- Intel/AMD IA-32, x64

**Bi-endian:** Processor may operate in either mode
- ARM, PowerPC, MIPS, SPARC V9, IA-64 (Intel Itanium)

```c
main() {
    unsigned int n;
    char *a = (char *)&n;

    n = 0x11223344;
    printf("%02x, %02x, %02x, %02x\n",
           a[0], a[1], a[2], a[3]);
}
```

Output on an Intel: 44, 33, 22, 11

Output on a PowerPC: 11, 22, 33, 44
Representing data: serialization

Need standard encoding to enable communication between heterogeneous systems

• Serialization
  – Convert data into a pointerless format: *an array of bytes*

• Examples
  – XDR (eXternal Data Representation), used by ONC RPC
  – JSON (JavaScript Object Notation)
  – W3C XML Schema Language
  – ASN.1 (ISO Abstract Syntax Notation)
  – Google Protocol Buffers
Representing data

Implicit typing
– only values are transmitted, not data types or parameter info
– e.g., ONC XDR (RFC 4506)

Explicit typing
– Type is transmitted with each value
– e.g., ISO’s ASN.1, XML, protocol buffers, JSON
Where to bind?

Need to locate host and correct server process
Where to bind? – Solution 1

Maintain a centralized DB that can locate a host that provides a particular service

(Birrell & Nelson’s 1984 proposal)

Challenges:

– Who administers this?
– What is the scope of administration?
– What if the same services run on different machines (e.g., file systems)?
Where to bind? – Solution 2

A server on each host maintains a DB of *locally* provided services
TCP or UDP? Which one should we use?

- Some implementations may offer only one (e.g. TCP)

- Most support several
  - Allow programmer (or end user) to choose at runtime
When things go wrong

• Local procedure calls do not fail
  – If they core dump, entire process dies

• More opportunities for error with RPC

• Transparency breaks here
  – Applications should be prepared to deal with RPC failure
When things go wrong

• **Semantics of remote procedure calls**
  – Local procedure call: *exactly once*

• A remote procedure call may be called:
  – 0 times:
    server crashed or server process died before executing server code
  – 1 time:
    everything worked well, as expected
  – 1 or more times: excess latency or lost reply from server and client retransmission
RPC semantics

• Most RPC systems will offer either:
  – **at least once** semantics
  – or **at most once** semantics

• Understand application:
  – **idempotent** functions: may be run any number of times without harm
  – **non-idempotent** functions: those with side-effects

• Try to design your application to be idempotent
  – Not always easy!
  – Store transaction IDs, previous return data, etc.
More issues

Performance
   – RPC is slower … a lot slower (why?)

Security
   – messages may be visible over network – do we need to hide them?
   – Authenticate client?
   – Authenticate server?
Programming with RPC

Language support

- Many programming languages have no language-level concept of remote procedure calls (C, C++, Java <J2SE 5.0, …)
  - These compilers will not automatically generate client and server stubs

- Some languages have support that enables RPC (Java, Python, Haskell, Go, Erlang)
  - But we may need to deal with heterogeneous environments (e.g., Java communicating via XML)

Common solution

- Interface Definition Language (IDL): describes remote procedures

- Separate compiler that generate stubs (pre-compiler)
Interface Definition Language (IDL)

• Allow programmer to specify remote procedure interfaces (names, parameters, return values)

• Pre-compiler can use this to generate client and server stubs
  – Marshaling code
  – Unmarshaling code
  – Network transport routines
  – Conform to defined interface

• An IDL looks similar to function prototypes
RPC compiler

IDL → RPC compiler → IDL

client code (main) → client stub → data conv. → headers → data conv. → server skeleton → server functions

compiler → client → compiler

compiler → server

Code you write

Code RPC compiler generates
Writing the program

• Client code has to be modified
  – Initialize RPC-related options
    • Identify transport type
    • Locate server/service
  – Handle failure of remote procedure calls

• Server functions
  – Generally need little or no modification
What kind of services does an RPC system need?

• **Name service operations**
  – Export/lookup of binding information (ports, machines)
  – Support dynamic ports

• **Binding operations**
  – Establish client/server communications using appropriate protocol (establish endpoints)

• **Endpoint operations**
  – Listen for requests, export endpoint to name server (often the main program on the server)
What kind of services does an RPC system need?

- **Security operations**
  - Authenticate client/server

- **Internationalization operations (maybe)**

- **Marshaling/data conversion operations**

- **Stub memory management**
  - Dealing with “reference” data, temporary buffers

- **Program ID operations**
  - Allow applications to access IDs of RPC interfaces
  - Can you pass references to remote functions to other processes?
The End