CS 43: Computer Networks

07: Concurrency and Non-blocking I/O
Sep 17, 2018
Today

• Under-the-hood look at system calls
  – Data buffering and blocking
• Inter-process communication
• Processes, threads and blocking
Recall Inter-process Communication (IPC)

- Processes must communicate to cooperate
- Must have two mechanisms:
  - Data transfer
  - Synchronization
- On a single machine:
  - Threads (shared memory)
  - Message passing
Message Passing (local)

• Operating system mechanism for IPC
  – send (destination, message_buffer)
  – receive (source, message_buffer)

• Data transfer: in to and out of kernel message buffers

• Synchronization: ?
Where is the synchronization in message passing Inter-process Communication?

A. The OS adds synchronization.

B. Synchronization is determined by the order of sends and receives.

C. The communicating processes exchange synchronization messages (lock/unlock).

D. There is no synchronization mechanism.
Interprocess Communication (non-local)

• Processes must communicate to cooperate

• Must have two mechanisms:
  – Data transfer
  – Synchronization

• Across a network:
  – Threads (shared memory) NOT AN OPTION!
  – Message passing
Message Passing (network)

- Same synchronization
- Data transfer
  - Copy to/from OS socket buffer
  - Extra step across network: hidden from applications
Descriptor Table

For each Process

OS stores a table, per process, of descriptors

Kernel
Descriptors

SOCKET(2) BSD System Calls Manual SOCKET(2)

NAME
socket -- create an endpoint for communication

SYNOPSIS
#include <sys/socket.h>

int
socket(int domain, int type, int protocol);

DESCRIPTION
socket() creates an endpoint for communication and returns a descriptor.

DESCRIPTION

The open() system call opens the file specified by pathname. If the specified file does not exist, it may optionally (if O_CREAT is specified in flags) be created by open().

int open(const char *pathname, int flags);
int open(const char *pathname, int flags, mode_t mode);
Descriptor Table

For each Process

OS stores a table, per process, of descriptors

Kernel

stdin  stdout  stderr
socket()

For each Process

```c
int sock = socket(AF_INET, SOCK_STREAM, 0);
```

- socket() returns a socket descriptor
- Indexes into table

Kernel

0
1
2

stdin  stdout  stderr

0
1
2

...
For each Process

```c
int sock = socket(AF_INET, SOCK_STREAM, 0);
```

OS stores details of the socket, connection, and pointers to buffers

Family: AF_INET, Type: SOCK_STREAM
Local address: NULL, Local port: NULL
Send buffer, Receive buffer
For each Process

```c
int sock = socket(AF_INET, SOCK_STREAM, 0);
```

OS stores details of the socket, connection, and pointers to buffers

Buffer: Temporary data storage location

| Family: AF_INET, Type: SOCK_STREAM |
| Local address: NULL, Local port: NULL |
| Send buffer, Receive buffer |

Kernel
Socket Buffers

For each Process

```c
int sock = socket(AF_INET, SOCK_STREAM, 0);
```

Application buffer / storage space:

Family: AF_INET, Type: SOCK_STREAM
Local address: NULL, Local port: NULL
Send buffer, Receive buffer

Kernel
socket Buffers

For each Process

```c
int sock = socket(AF_INET, SOCK_STREAM, 0);
```

Application buffer / storage space:

```
```

```
```

Family: AF_INET, Type: SOCK_STREAM
Local address: NULL, Local port: NULL
Send buffer □ , Receive buffer □

Kernel

Internet
Socket Buffers

For each Process

```c
int sock = socket(AF_INET, SOCK_STREAM, 0);
```

Application buffer / storage space:

Family: AF_INET, Type: SOCK_STREAM
Local address: NULL, Local port: NULL
Send buffer , Receive buffer

recv(): Move data from socket buffer to process

Kernel

Internet
Socket Buffers

For each Process

```c
int sock = socket(AF_INET, SOCK_STREAM, 0);
```

Application buffer / storage space:

```
Family: AF_INET, Type: SOCK_STREAM
Local address: NULL, Local port: NULL
Send buffer , Receive buffer
```

send(): Move data from process to socket buffer
Socket Buffers

For each Process

```c
int sock = socket(AF_INET, SOCK_STREAM, 0);
```

Application buffer / storage space:

Family: AF_INET, Type: SOCK_STREAM
Local address: NULL, Local port: NULL
Send buffer ☐, Receive buffer ☐

Free space?  Is data here?

Challenge: Your process does NOT know what is stored here!
int sock = socket(AF_INET, SOCK_STREAM, 0);  
(assume we issued a connect() here...) 
int recv_val = recv(sock, r_buf, 200, 0); 

Family: AF_INET, Type: SOCK_STREAM 
Local address: ..., Local port: ...
Send buffer, Receive buffer

Is data here?
What should we do if the receive socket buffer is empty? If it has 100 bytes?

For each Process

```c
int sock = socket(AF_INET, SOCK_STREAM, 0);  // r_buf (size 200)
(assume we connect()ed here...)
int recv_val = recv(sock, r_buf, 200, 0);
```

Two Scenarios:

- **Socket buffer**
  - Empty
  - 100 bytes

Kernel
What should we do if the receive socket buffer is empty? If it has 100 bytes?

For each Process

```c
int sock = socket(AF_INET, SOCK_STREAM, 0);  // r_buf (size 200)
(assume we connect()ed here...)
int recv_val = recv(sock, r_buf, 200, 0);
```

Two Scenarios:

<table>
<thead>
<tr>
<th></th>
<th>Empty</th>
<th>100 Bytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Block</td>
<td>Block</td>
</tr>
<tr>
<td>B</td>
<td>Block</td>
<td>Copy 100 bytes</td>
</tr>
<tr>
<td>C</td>
<td>Copy 0 bytes</td>
<td>Block</td>
</tr>
<tr>
<td>D</td>
<td>Copy 0 bytes</td>
<td>Copy 100 bytes</td>
</tr>
<tr>
<td>E</td>
<td>Something else</td>
<td></td>
</tr>
</tbody>
</table>

Socket buffer

- Empty
- 100 bytes

Kernel
What should we do if the send socket buffer is full? If it has 100 bytes?

For each Process

```
int sock = socket(AF_INET, SOCK_STREAM, 0);   // s_buf (size 200)
(assume we connect()ed here...)
int recv_val = recv(sock, r_buf, 200, 0);
```

Two Scenarios:

Socket buffer

```
Full
```

```
100 bytes
```
What should we do if the send socket buffer is full? 
If it has 100 bytes?

For each Process:

```
int sock = socket(AF_INET, SOCK_STREAM, 0);  
s_buf (size 200)
(assume we connect()ed here...)
int recv_val = recv(sock, r_buf, 200, 0);
```

Two Scenarios:

<table>
<thead>
<tr>
<th></th>
<th>Full</th>
<th>100 Bytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Return 0</td>
<td>Copy 100 bytes</td>
</tr>
<tr>
<td>B</td>
<td>Block</td>
<td>Copy 100 bytes</td>
</tr>
<tr>
<td>C</td>
<td>Return 0</td>
<td>Block</td>
</tr>
<tr>
<td>D</td>
<td>Block</td>
<td>Block</td>
</tr>
<tr>
<td>E</td>
<td>Something else</td>
<td></td>
</tr>
</tbody>
</table>

Kernel

Socket buffer

Full

100 bytes
Blocking Implications

send()
- **Do not** assume that you will `recv()` all of the bytes that you ask for.
- **Do not** assume that you are done receiving.
- **Always** receive in a loop!

recv()
- **Do not** assume that you will `send()` all of the data you ask the kernel to copy.
- Keep track of where you are in the data you want to send.
- **Always** send in a loop!

* Unless you’re dealing with a single byte, which is rare.
ALWAYS check send() return value!

• When send() return value is less than the data size, you are responsible for sending the rest.

Data sent: 0
Data to send: 130

send(sock, data, 130, 0);
ALWAYS check send() return value!

• When send() return value is less than the data size, you are responsible for sending the rest.

Data:
Data sent: 0
Data to send: 130
send(sock, data, 130, 0);

Data:
Data sent: 60
Data to send: 130
**ALWAYS check send() return value!**

- When send() return value is less than the data size, **you are responsible for sending the rest.**

```c
send(sock, data, 130, 0);
```

Data sent: 0  
Data to send: 130

Data sent: 60  
Data to send: 130

// Copy the 70 bytes starting from offset 60.
```
send(sock, data + 60, 130 - 60, 0);
```
ALWAYS check send() return value!

- When send() return value is less than the data size, you are responsible for sending the rest.

Repeat until all bytes are sent. (data_sent == data_to_send)...

Data sent: 0
Data to send: 130
send(sock, data, 130, 0);

Data sent: 60
Data to send: 130
// Copy the 70 bytes starting from offset 60.
send(sock, data + 60, 130 - 60, 0);
Blocking Summary

**send()**

- Blocks when socket buffer for sending is full
- Returns less than requested size when buffer cannot hold full size

**recv()**

- Blocks when socket buffer for receiving is empty
- Returns less than requested size when buffer has less than full size

Always check the return value!
Blocking Summary

send()
• Blocks when socket buffer for sending is full
• Returns less than requested size when buffer cannot hold full size

recv()
• Blocks when socket buffer for receiving is empty
• Returns less than requested size when buffer has less than full size

Always check the return value!
Concurrency

- Think you’re the only one talking to that server?
Without Concurrency

- Think you’re the only one talking to that server?

```
Web Server
recv()
request
```

Client
Without Concurrency

- Think you’re the only one talking to that server?

Client taking its time...

Web Server

Server Process Blocked!

recv() request

Ready to send, but server still blocked on first client.

If only we could handle these connections separately...
Multiple processes

Web Server

Server fork()s

Web Server

Child process recv()s

Web Server

Server fork()s

Services the new client request

Client

Client
Processes/Threads vs. Parent
(More details in an OS class…)

**Spawned Process**
- Inherits descriptor table
- Does not share memory
  - New memory address space
- Scheduled independently
  - Separate execution context
  - Can block independently

**Spawned Thread**
- Shares descriptor table
- Shares memory
  - Uses parent’s address space
- Scheduled independently
  - Separate execution context
  - Can block independently
Processes/Threads vs. Parent
(More details in an OS class...)

Spawned Process
- Inherits descriptor table
- Does not share memory
  - New memory address space
- Scheduled independently
  - Separate execution context
  - Can block independently

Spawned Thread
- Shares descriptor table
- Shares memory
  - Uses parent’s address space
- Scheduled independently
  - Separate execution context
  - Can block independently

Often, we don’t need the extra isolation of a separate address space. Faster to skip creating it and share with parent – threading.
Threads & Sharing

• Global variables and static objects are shared
  – Stored in the static data segment, accessible by any thread

• Dynamic objects and other heap objects are shared
  – Allocated from heap with malloc/free or new/delete

• Local variables are not shared
  – Refer to data on the stack
  – Each thread has its own stack
  – Never pass/share/store a pointer to a local variable on another thread’s stack
Both processes and threads:

- Several benefits
  - Modularizes code: one piece accepts connections, another services them
  - Each can be scheduled on a separate CPU
  - Blocking I/O can be overlapped
Which benefit is most critical?

A. Modular code/separation of concerns.

B. Multiple CPU/core parallelism.

C. I/O overlapping.

D. Some other benefit.
Both processes and threads

• Several benefits
  – Modularizes code: one piece accepts connections, another services them
  – Each can be scheduled on a separate CPU
  – Blocking I/O can be overlapped
• Still not maximum efficiency...
  – Creating/destroying threads still takes time
  – Requires memory to store thread execution state
  – Lots of context switching overhead
Non-blocking I/O

• One operation: add a flag to send/receive
• Permanently, for socket: `fcntl()` – “file control”
  – Allows setting options on file/socket descriptors

```c
int sock, result, flags = 0;
sock = socket(AF_INET, SOCK_STREAM, 0);
result = fcntl(sock, F_SETFL, flags | O_NONBLOCK)
check result – 0 on success
```
Non-blocking I/O

• With O_NONBLOCK set on a socket
  – No operations will block!

• On recv(), if socket buffer is empty:
  – returns -1, errno is EAGAIN or EWOULDBLOCK

• On send(), if socket buffer is full:
  – returns -1, errno is EAGAIN or EWOULDBLOCK
Will this work?

server_socket = socket(), bind(), listen()
connections = []
while (1)
    new_connection = accept(server_socket)
    if new_connection != -1,
        add it to connections
    for connection in connections:
        recv(connection, ...) // Try to receive
        send(connection, ...) // Try to send, if needed
Will this work?

A. Yes, this will work.
B. No, this will execute too slowly.
C. No, this will use too many resources.
D. No, this will still block.

server_socket = socket(), bind(), listen()
connections = []
while (1)
    new_connection = accept(server_socket)
    if new_connection != -1,
        add it to connections
    for connection in connections:
        recv(connection, ...) // Try to receive
        send(connection, ...) // Try to send, if needed
Event-based concurrency

• Rather than checking over and over, let the OS tell us when data can be read/written
• Create set of file/socket descriptors we want to read and write
• Tell system to block until at least one of those is ready for us to use. The OS worries about selecting which one(s).

select()
Event-based concurrency

• Rather than checking over and over, let the OS tell us when data can be read/written

• Tell system to block until at least one of those is ready for us to use. The OS worries about selecting which one(s).

• Only one process/thread (or one per core)
  – No time wasted on context switching
  – No memory overhead for many processes/threads