What makes the world wide web work?

Karen Reid
Simple Web Request

MarkUs

MarkUs is a web application for giving high quality feedback to students. The administrative
The Request

• How do we tell the web server what we want?
• How do we even find the web server?
• How do the web server and browser talk to each other?
HTTP Request

**Request:**

GET / HTTP/1.1
Host: markusproject.org
...

**Reply:**

HTTP/1.1 200 OK
Date: Tue, 13 Mar 2017
Server: Apache/2.2.22(Debian)
Content-Type: text/html
How do we find the server?

• Every computer on the Internet has an Internet address.
• Called an IP address (Internet Protocol)
• An IP address is 4 numbers separated by dots.

markusproject.org = 69.164.221.145
Domain Name Servers

- browser
- markusproject.org?
- root name server
- local name server
- 69.164.221.145
- delegate server
- "org" name server
This is getting complicated!

Number of messages? 8 (or so)
Now what?

• Okay, we have the address.
• What do we do with it?
• Let’s look at how two computers communicate.
• HTTP is a high-level protocol
• HTTP is specific to the web.
• Computers communicate for many reasons.
Protocols

• Computers use several layers of general protocols to communicate.
• To understand why these layers are important, think about how a company sends you an invoice for a purchase.
Protocols

Invoice:
Customer: Karen Reid
Order No: 5379

<table>
<thead>
<tr>
<th>Qty</th>
<th>Item</th>
<th>Unit Price</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Athalon</td>
<td>219.00</td>
<td>219.00</td>
</tr>
<tr>
<td>2</td>
<td>128 MB</td>
<td>149.95</td>
<td>299.90</td>
</tr>
</tbody>
</table>

Subtotal: 518.90
Tax: 77.84
TOTAL: 596.74

Payable to: CPUS are us $596.74
Five hundred ninety six 74/100

CPUS are us
Karen Reid
Dept. of Computer Science
University of Toronto

We deliver!

Courier
TCP/IP

- Transmission Control Protocol.
- Tells us how to package up the data.

<table>
<thead>
<tr>
<th>source address</th>
<th>dest. address</th>
</tr>
</thead>
<tbody>
<tr>
<td>bytes</td>
<td>ack</td>
</tr>
<tr>
<td>port</td>
<td></td>
</tr>
</tbody>
</table>

data
TCP Connection

3-way handshake
SYN

Hi 207.99.53.46
Connection port 80?

Hi 192.168.1.1
Let’s talk

okay

Send me a file

Got it

Here’s some data

ack

Here’s some more

Got it

Got it

I’m done

ack

I’m done too

ack

I’m done too

fin

fin
Packaging up the data

- make packets
- put in an IP envelope with another header

Each TCP packet is given a header:
- sequence number
- checksum

To 207.99.53.46
The Big Picture

- **Client-Server model**: a client process wants to talk to a server process
- Client must find server - **DNS lookup**
- Client must find process on server - **ports**
- Finally **establish a connection** so two processes can talk
7 cities, 5 states/prov, 2 countries

browser

- home network
- Rogers cable
  - Toronto
  - Tempe
  - Marina Del Rey
  - Parsippany
  - Newark
  - Greendell

markusproject.org
Putting it together
How many messages?

- It depends on the size of the web page.
- The web page that appears for markusproject.org is less than 30 Kbytes.
- If the web page is 30 Kbytes (small!) it will likely be broken up into ~20 IP packets.

\[8 \text{ (DNS)} + 20 \times 15 \text{ hops} = 308 \text{ messages}\]
When something goes wrong

• A packet might not arrive
  – traffic overload
  – bit corruption
• Receiver asks for missing packets to be resent.
• Want to send data as fast as possible.
• But sending too fast wastes resources.
TCP Congestion Control

• Window-based:
  – some number of packets allowed to be sent and not ack’d
  – as successful ack’s arrive, grow window
  – if packet loss is detected, cut window size
TCP Congestion Control

- Time
- Window size

packet losses
All we did was click on a link...
Take aways

• The web today is made up of complex layers of software
• No one person, organization, or company could have created it in isolation
• We can understand it because we can study one layer at a time
• We can create new things by building on top of existing layers