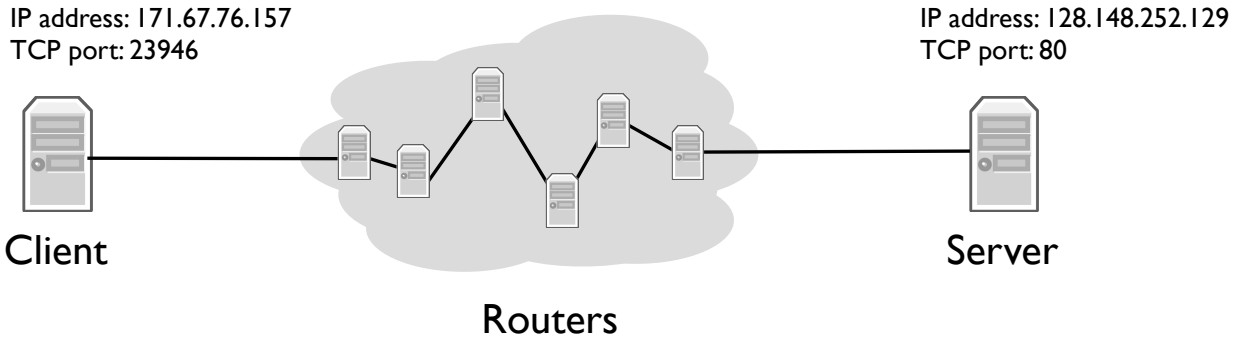
Longest Prefix Match

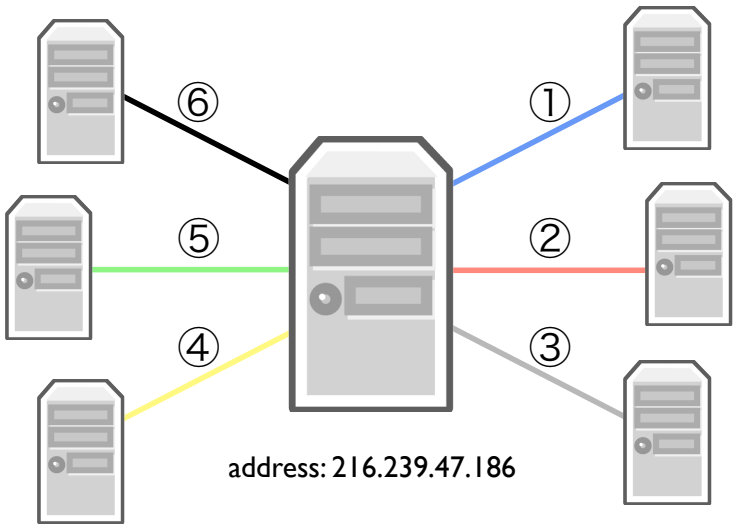
Internet routers can have many links. They have many options for which direction to forward a received packet. To select which link to forward a packet over, routers today typically use an algorithm called Longest Prefix Match.

Inside the Stream



In this example a client wants to open a TCP connection to a server on port 80, the typical port for web servers. The packets to set up the connection and transfer data take many hops between the client and server. On each hop of each packet, a router decides which link to forward the packet over.

Inside Each Hop



dest	link
default	1
171.33.x.x	5
23.x.x.x	2
28.33.5.x	4
171.32.x.x	2
67.x.x.x	6
216.x.x.x	1

How does a router make this decision? It does so through something called a forwarding table, shown here on the right. A forwarding table consists of a set of partial IP addresses. The x's show that the addresses are partial. The x's represent wildcards. For example, the second entry, reading 171.33.x.x means "any IP address whose first byte is 171 and whose second byte is 33." This particular entry, for example, includes 171.33.5.245 as well as 171.33.1.1.

When a packet arrives, the router checks which forwarding table entry best matches the packet and forwards the packet along the link associated with that forwarding table entry. By "best", I mean most specific. The default route is effectively all wildcards -- it matches every IP address. If, when a packet arrives, there isn't a more specific route than the default route, the router will just use the default one.

Longest Prefix Match

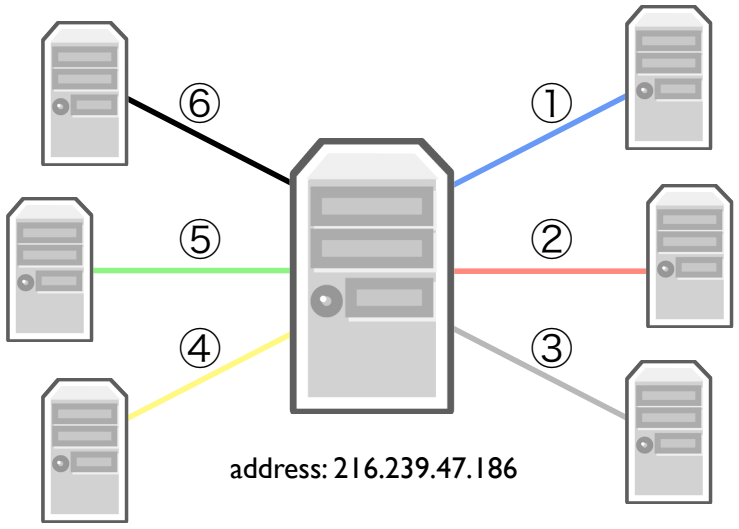
- Algorithm IP routers use to chose matching entry from forwarding table
- Forwarding table is a set of CIDR entries
  - An address might match multiple entries
  - E.g., 171.33.0.1 matches both entries on right
- Algorithm: use forwarding entry with the longest matching prefix
  - Longest prefix match will chose link 5 for 171.33.0.1

dest	link
0.0.0.0/0	1
171.33.0.0/16	5

Longest prefix match, or LPM, is the algorithm IP routers use to decide how to forward a packet. Every router has a forwarding table. Entries in this forwarding table have two parts: a CIDR entry describing a block of addresses, and a next hop for packets that match that CIDR entry. An address might belong to multiple CIDR entries.

For example, in this routing table on the right, there are two entries, one for the default route, which has a prefix of length 0, and one for 171.33.0.0/16. By default, all packets will go over link 1. However, if the first 16 bits, two octets, of a packet destination address matches 171.33, the router will send it over link 5. This is because a 16 bit prefix is a longer prefix than 0 bits, it's more specific.

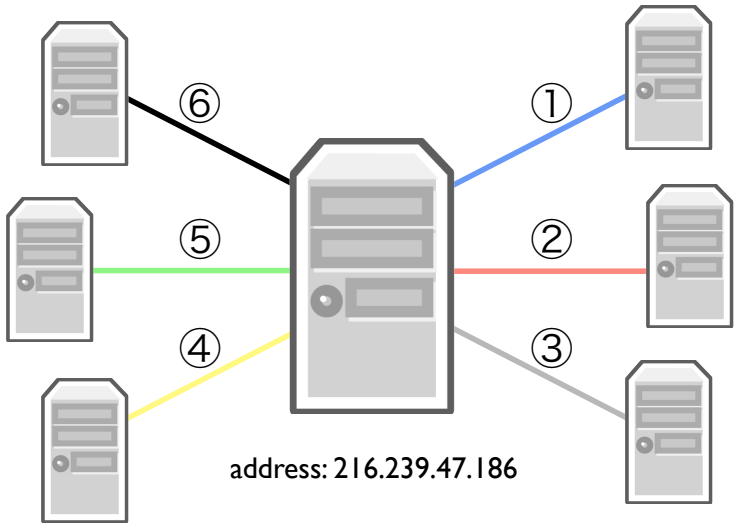
Inside Each Hop



dest	link
default	1
171.33.x.x	5
23.x.x.x	2
28.33.5.x	4
171.32.x.x	2
67.x.x.x	6
216.x.x.x	1

So let's go back to our earlier example, where we showed a forwarding table with Xs denoting wildcards. Here's the router and its forwarding table.

Inside Each Hop (for real)



dest	link
default	1
171.33.x.x	5
23.x.x.x	2
28.33.5.x	4
171.32.x.x	2
67.x.x.x	6
216.x.x.x	1

dest	link
0.0.0.0/0	1
171.33.0.0/16	5
23.0.0.0/8	2
28.33.5.0/24	4
171.32.0.0/16	2
67.0.0.0/8	6
216.0.0.0/8	1

If we represent this forwarding table as CIDR entries, this is what it looks like. Since in this simple example all of the prefixes are in terms of bytes, all of the prefixes have length 0, 8, 16, or 24 bits.

Quiz

With the forwarding table on the right, over which link will a router using longest prefix match send packets with the following IP destination address?

A. 63.19.5.3

B. 171.15.15.0

C. 63.19.5.32

D. 44.199.230.1

E. 171.128.16.0

dest	link
0.0.0.0/0	1
18.0.0.0/8	5
171.0.0.0/8	2
171.0.0.0/10	4
171.0.15.0/24	1
55.128.0.0/10	6
63.19.5.0/30	3

With the forwarding table on the right, over which link will a router using longest prefix match send packets with the following IP destination address?

The answer for A, 63.19.5.3, is link 3. 63.19.5.3 matches two prefixes: the default route and prefix 63.19.5.0/30. The prefix is 30 bits long and 63.19.5.3 differs in only the last two bits. /30 is a longer prefix than /0 so the router will pick link 3.

The answer for B,171.15.15.0, is link 4. 171.15.15.0 matches three entries. It matches the default route, 171.0.0.0/8 and 171.0.0.0/10. It does not match 171.0.15.0/24 because B's second octet is 15, not 0. The third match, 171.0.0.0/10, is the longest prefix, so the router sends the packet along link 4.

The answer for C, 63.19.5.32, is link 1. The longest prefix match is the default route. It does not match 63.19.5.0/30 because it differs in the 26th bit.

The answer for D, 44.199.230.1, is link 1. The longest prefix match is the default route.

The answer for E, 171.128.16.0, is link 2. This address matches two prefixes, the default route and 171.0.0.0/8. It does not match 171.0.0.0/10 because it differs on the 9th bit. 171.0.0.0/8 is the longest prefix, so the router will forward the packet on link 2.