P2P Computing

1. P2P Background
   a) Historical P2P
   b) Modern P2P
   c) Why P2P?

2. P2P Environment
   a) Binding of Peers (Early + Late)
   b) NAT Translation Systems
   c) Firewalls
   d) True P2P

3. P2P Examples
   a) SETI@Home
   b) Napster (expanded)
   c) ICQ
   d) KaZaA
   e) - Gnutella Scenario – Gnutella Lecture…

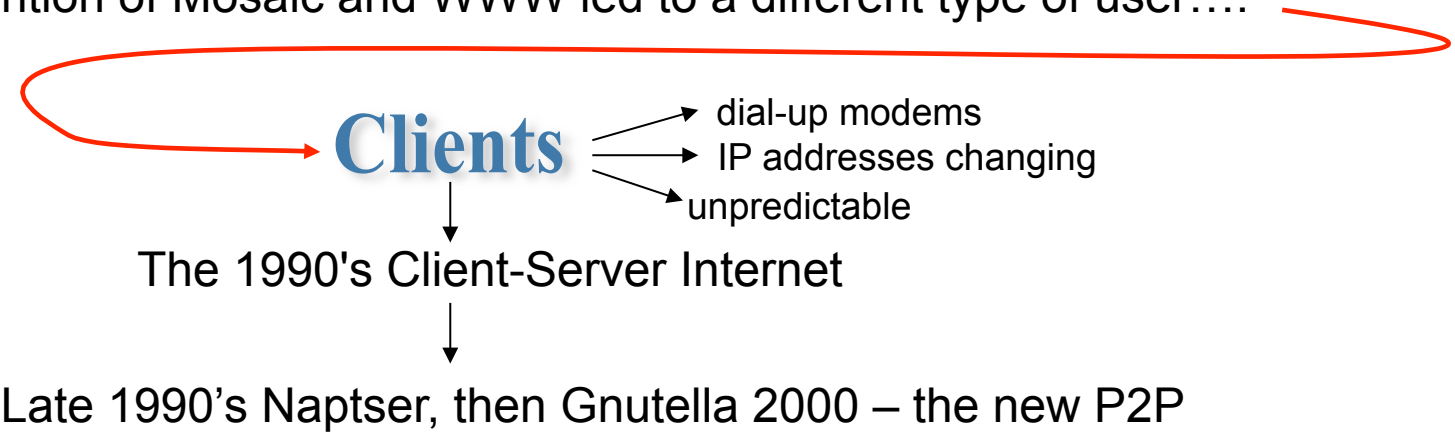
Historical P2P

• Peer to Peer (P2P) - originally used to describe the communication of two peers.

• The Internet started as peer to peer system e.g. ARPANET, 1969
  • goal - to share computing resources around the USA using different networks
  • UCLA, Stanford, Utah and Santa Barbara
  • all had equal status – P2P

• From late 1960's until 1994, machines were assumed to be switch on, connected and had an IP address assigned

• Then, invention of Mosaic and WWW led to a different type of user….



Peer to Peer Definition

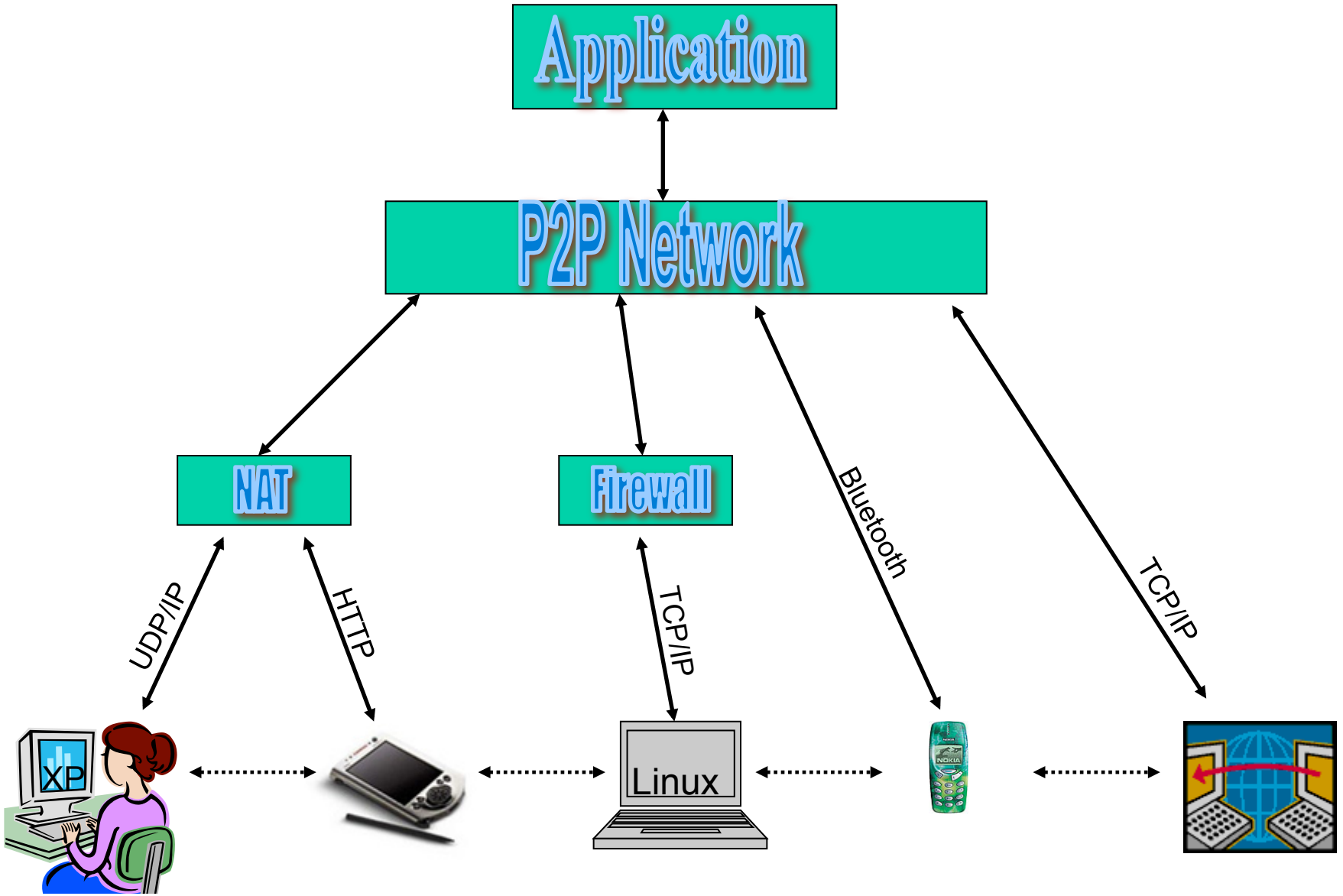
What is an P2P application?

P2P is a class of applications that takes advantage of resources e.g. storage, cycles, content, human presence, available at the edges of the Internet – Clay Shirky

Computers/devices “at the edges of the internet” are those:

- Operating within transient environments - computers come and go frequently
- They can be behind a firewall or NAT systems
- Have to operate outside of DNS
- Often have to deal with differing transport protocols, devices and operating systems

A P2P Network

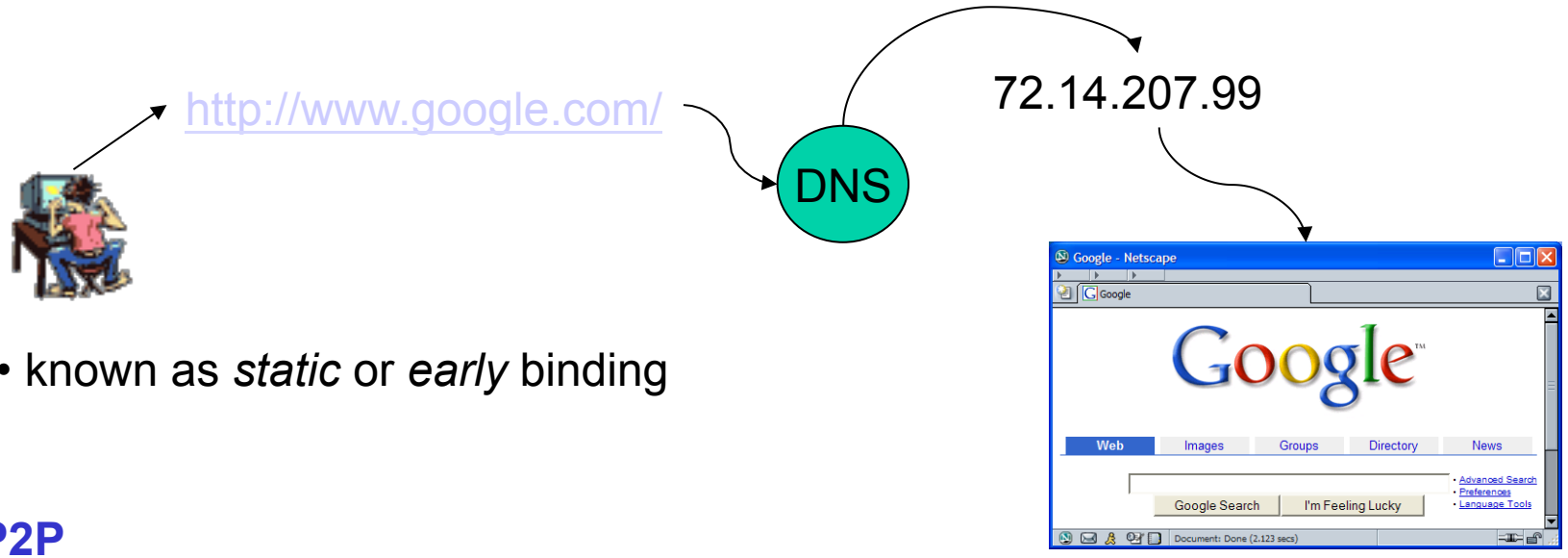


P2P Environments

1. Binding of Peers (Early + Late)
2. Network Translation Systems (NAT)
3. Firewalls
4. True P2P
5. P2P: Summary of Key Points

P2P & the Binding of Peers

Client/Server rely on fixed IP Addresses ...



- known as *static* or *early* binding

P2P

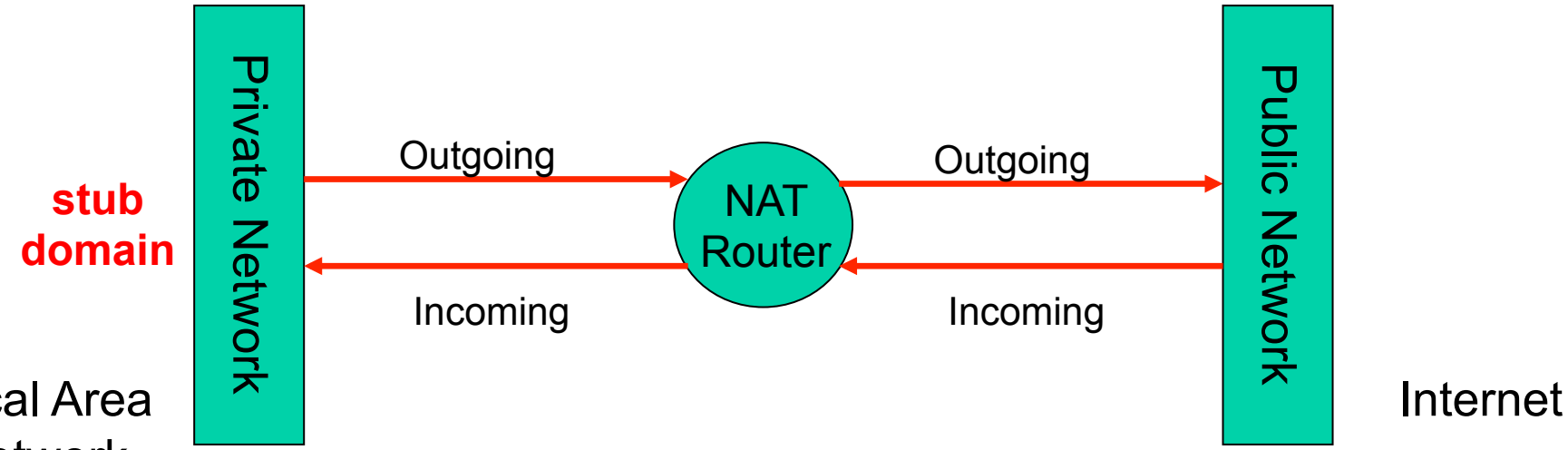
- Computers do not have a fixed address e.g. hidden behind NATs
- Need a *late* binding of their addresses with their identifier
- this is the norm in P2P
- DHCP often used for LANs for dynamic IP address assignment
- Dynamic DNS service providers offer late binding of changing IPs to the DNS system via their Domain Name servers.

NAT Systems

- IP address - unique 32-bit number, identifies computer location computer on network

- A possible 2^{32} but really around 3 billion, why?
  - addresses are set aside for multicasting, testing or other special uses
  - addresses are wasted because of the hierarchical nature of the address.
  - e.g. Google owns the block **209.85.128.0 - 209.85.255.255**

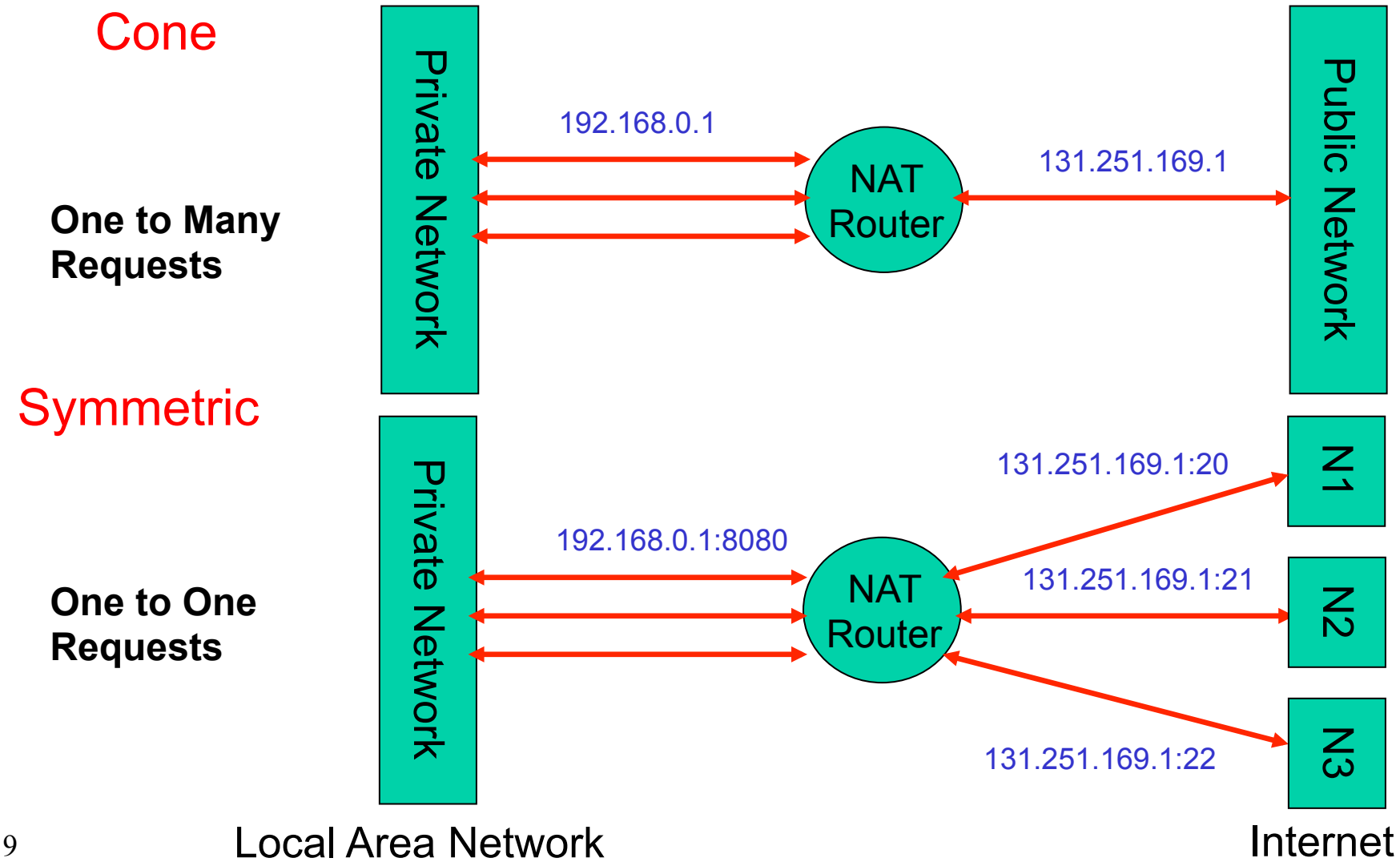
- Explosion of the Internet - available IP addresses is not enough! Solutions:
  - redesign the address format to allow more addresses ?
    - Yes, IPv6 (requires modification to the entire infrastructure of the Internet – but it is happening) supports 2^{128}
  - or use NAT Systems



NAT System Types

- Full Cone
  - Once a mapping between an internal and external address is made, it remains.
  - External messages directed at the external address will get routed to the mapped internal address.
- Restricted Cone
  - External messages will only be routed to the mapped internal IP address if an initiating message first came from inside the NAT.
- Port Restricted Cone
  - Same as a restricted cone, but the port number is also taken into account.
- Symmetric NAT
  - Every outgoing message from a particular address and port is mapped to a unique external address.
  - Two separate messages from the NAT-bound node will have distinct external mappings.
  - If Restricted: External messages will only be routed to the mapped internal IP address if an initiating message first came from inside the NAT.

NAT System Types



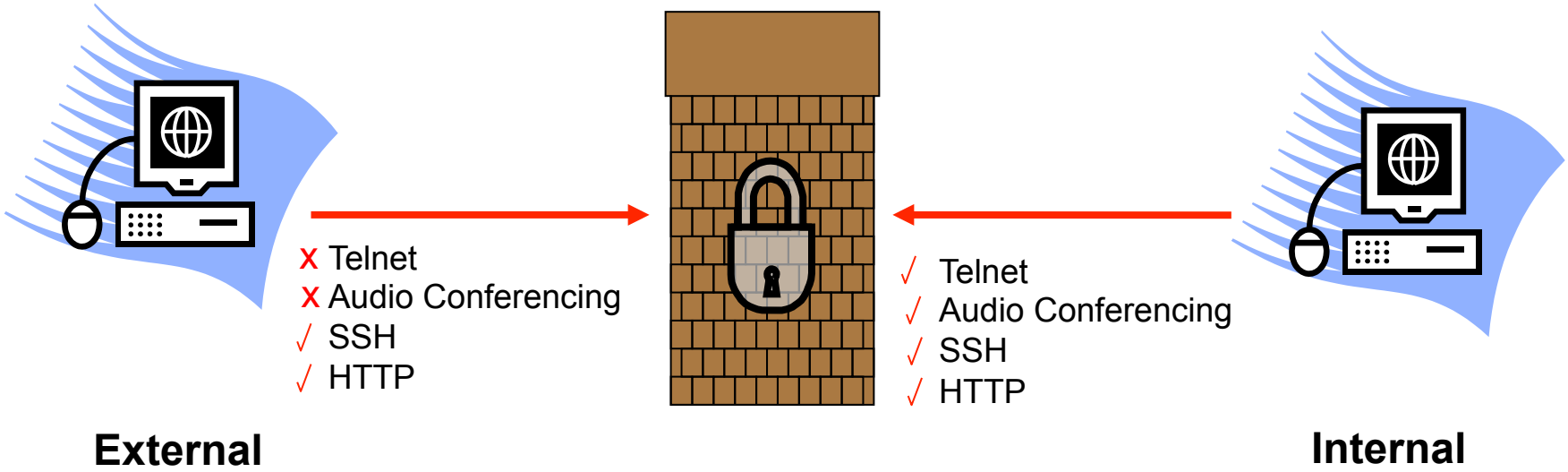
How do you get around a NAT?

- Manually configure NAT – not fun, not always possible
- Initiate an outward-bound connection, and KEEP IT OPEN
- Internet Gateway Device (IGD) Protocol via UPnP
  - Some NATs expose themselves using IGD
    - allows UPnP enabled nodes to discover external IP addresses and port mappings.
  - STUN (Simple Traversal of UDP over NATs - now Session Traversal Utilities over NAT)
    - Requires rendezvous nodes which accepts incoming messages from NAT-bound nodes. They return the address they see to the NAT-bound node.
  - TURN (Traversal Using Relay NAT)
    - Requires a dedicated relay server outside the NAT that receives data on behalf of the NAT bound node and forwards it.

Firewalls

- a system designed to prevent unauthorized access to or from a private network
- Allows internal/external connections through **specific ports** – and hence **protocols**
- Can restrict based on local and remote IP address as well.
- firewall examines each message and blocks those that do not meet the specified security criteria

Example



How to get Around Firewalls?

- Firewall dependent
- Open an initiating connection and KEEP IT OPEN
- Tunnelling
  - wrap the payload of the restricted protocol in a protocol that is not restricted
  - HTTP is commonly allowed through firewalls
    - employees can argue they could not be productive with access to the Web.
- Relay nodes
  - act as ‘mailboxes’ that a restricted node can pull messages from.

P2P Examples

1. P2P Examples
   a) SETI@Home
   b) Napster (expanded)
   c) ICQ
   d) KaZaA
   e) - Gnutella Scenario – later…

True P2P ?

Three main categories of systems

- Centralized systems: peer connects to server which coordinates and manages communication. e.g. SETI@home

- Brokered systems: peers connect to a server to discover other peers, but then manage the communication themselves (e.g. Napster). This is also called *Brokered P2P*.

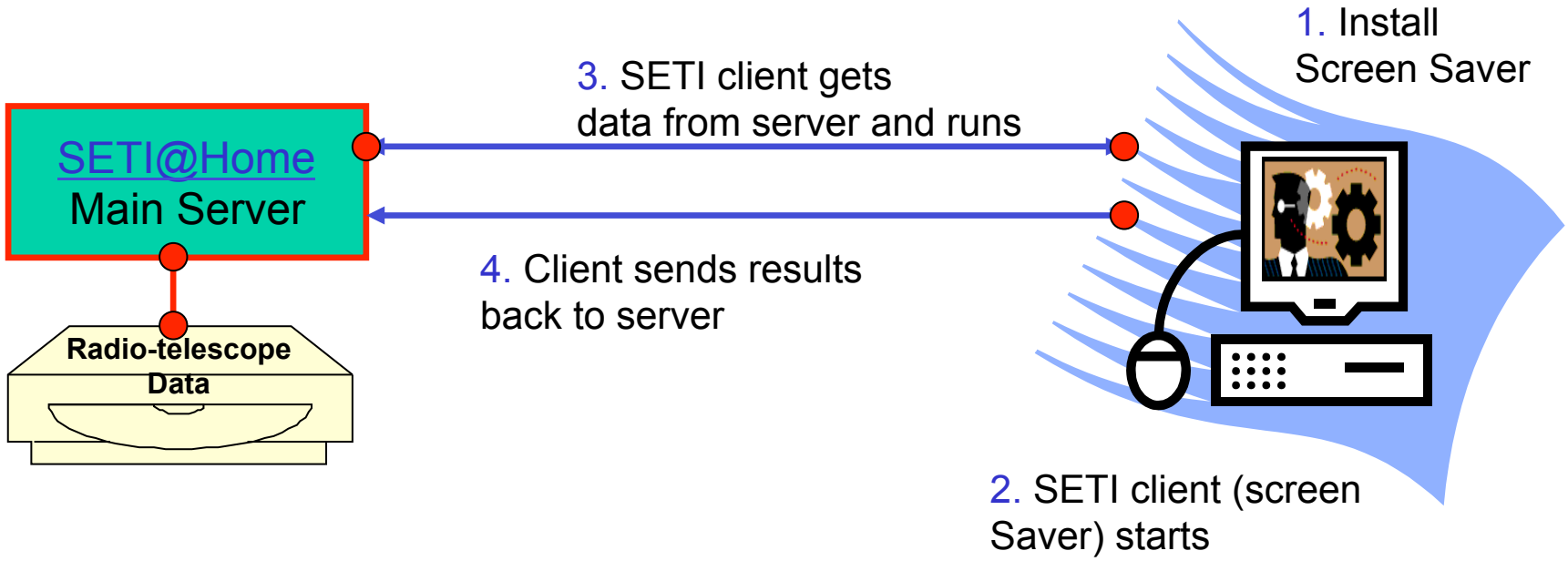
- Decentralized systems: peers run independently with no central services. Discovery is decentralized and communication takes place between the peers. e.g. Gnutella

True P2P

Example1: SETI@HOME (Client/Server)

Client/Server
P2P

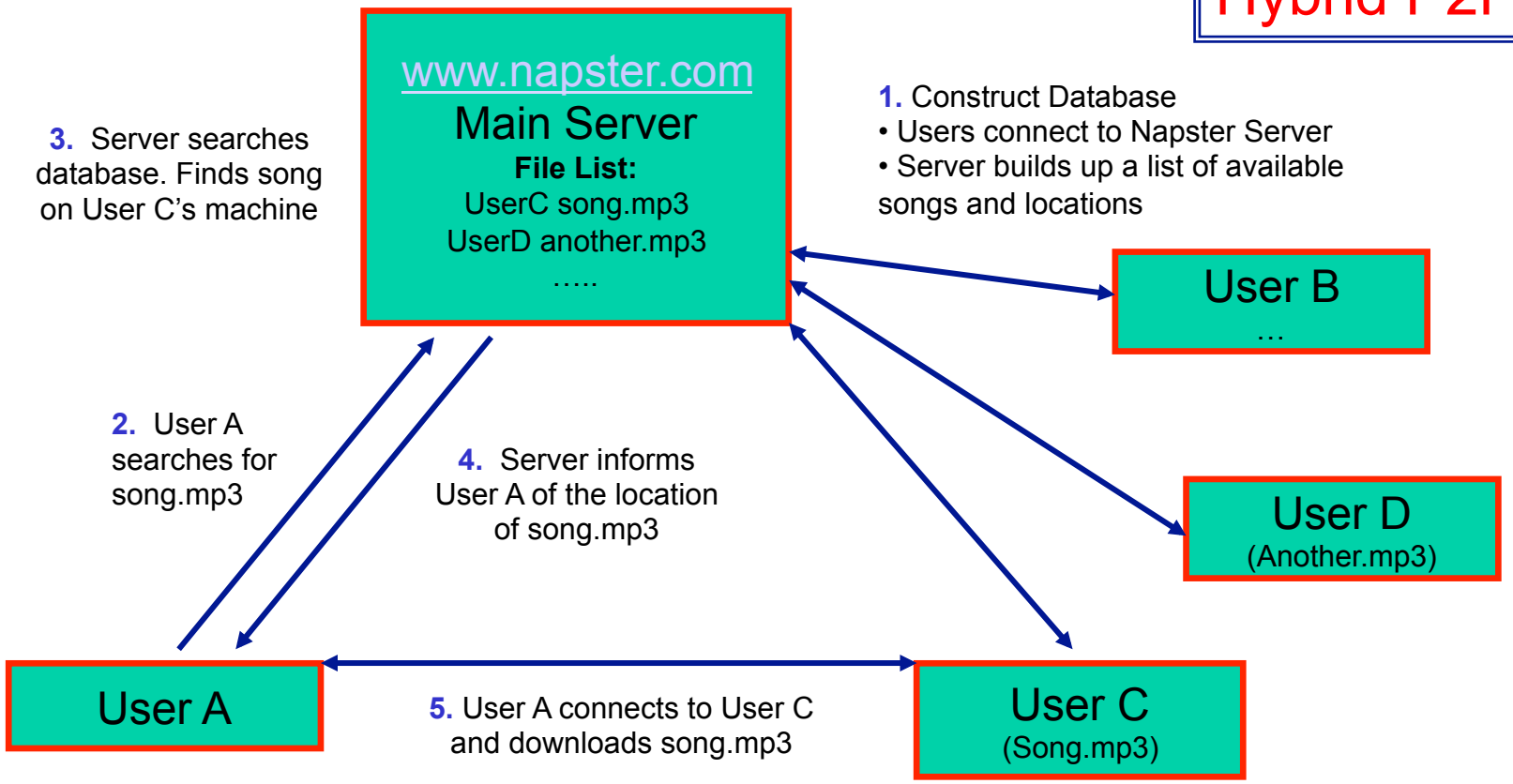
- Launched In 1996
- Scientific experiment - uses Internet-connected computers in the Search for Extraterrestrial Intelligence (SETI)
- Distributes a screen saver–based application to users
- Applies signal analysis algorithms to different data sets to process radio-telescope data.
- Has more than 3 million users - used over a million years of CPU time to date



Example 2: File Sharing with Napster

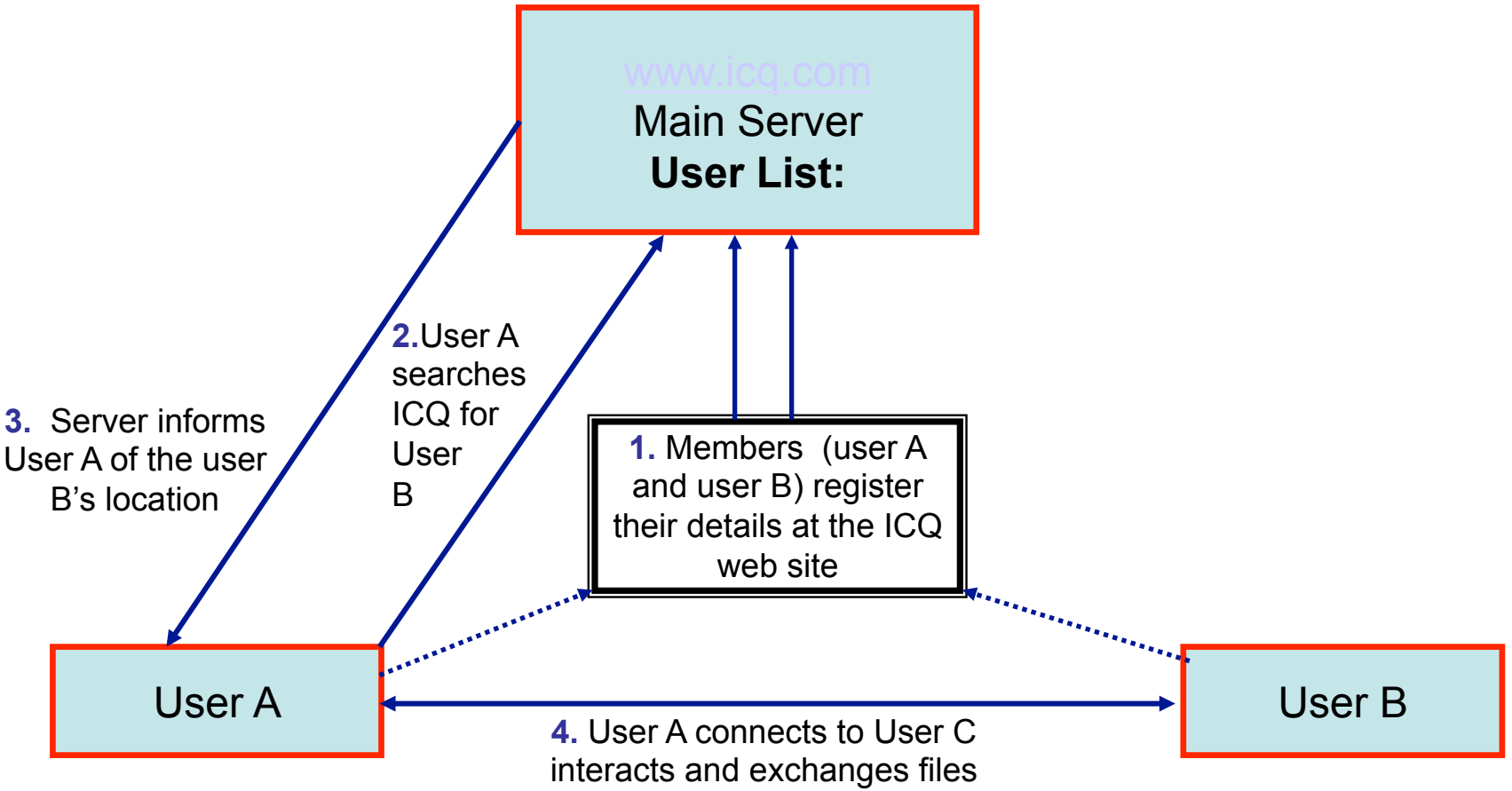
- Launched in May 1999, by Shawn Fanning (19) and Sean Parker (20)
- Allowed Users to download MP3 Files - compression format, good quality but 1/12th original size
- April 2000 – Metallica starts law suit – Huge and long court case
- November 2000 – Napster has 38 Million members
- July 2001 – Napster ordered offline, June 2002 bankrupt

Brokered/ Hybrid P2P



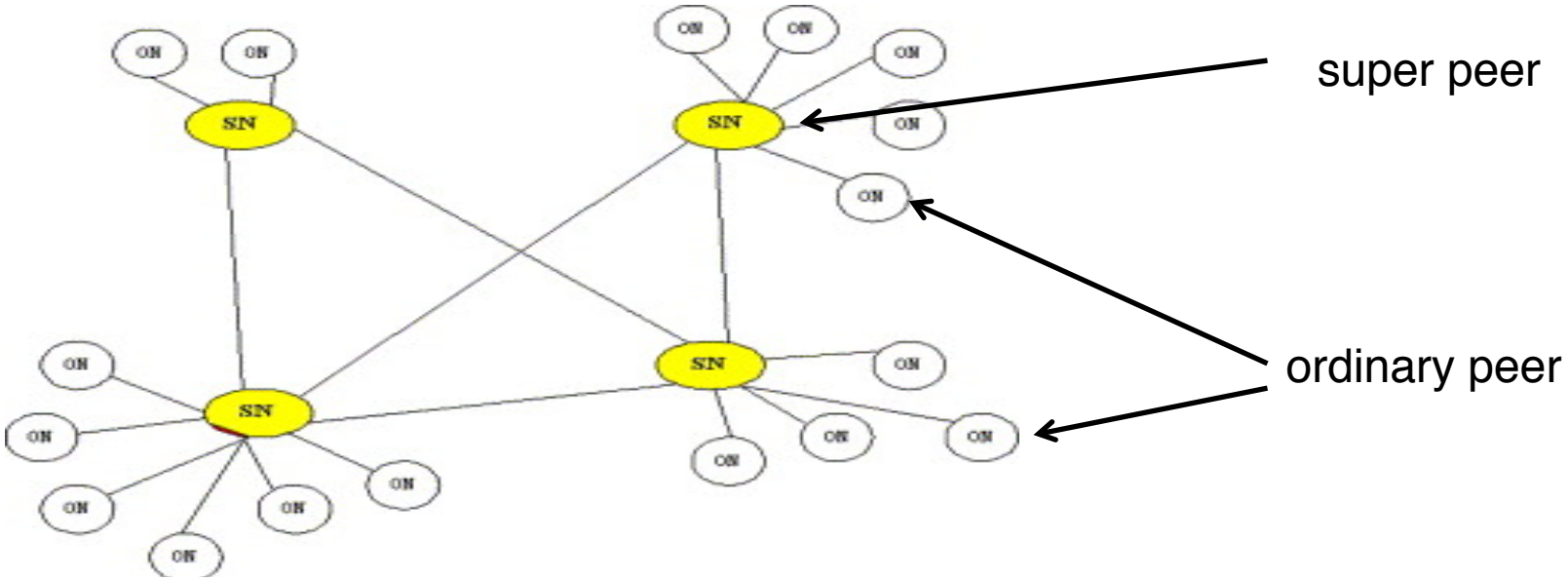
Example 3: ICQ: Instant Messaging

- Released in November 1996
- Allows users to be notified when their friends come online
- Users can send messages to their friends (instant messaging)
- Also allows users to exchange files



Example 4: KaZaA

- Uses the FastTrack network
  - up to 3 million users a day in 2002
  - designed by the Skype guys
  - not an open protocol
- SuperPeer Architecture - more decentralized form of brokered architecture
- maintains a file index that maps file identifiers to the IP addresses.
- Index is distributed across the SuperPeers.
- SP maintains a local index for all of its children.
  - similar to a (mini) Napster hub, but not a dedicated server – belongs to an individual user.



KaZaA cont…

- Superpeers create an overlay with a certain number of connections to other Superpeers.
  - roughly between 30 – 50 neighbours
  - change their connections frequently – every 10 – 20 minutes
- Superpeers service roughly 100 – 160 ordinary nodes.
  - 3 million nodes
  - roughly 30,000 superpeers.
- reduces the flooding of the entire network by queries.
- instead, ordinary nodes are connected to a small number of Superpeers and query only them.
- Queries are selectively flooded amongst the superpeer overlay.
- measurements suggest the network is highly dynamic
  - supernodes look for new neighbours with a low workload.
  - supernodes connect briefly to share metadata
  - periodically reconfigure themselves to provide a good spread of connections across the network providing better query results over time.

- second generation Gnutella has adopted the superpeer architecture.
- we'll look at this again next week.

Examples of P2P Technologies

File sharing/storage programs

Gnutella	LimeWire
Napster	Kazaa
eMule	Freenet

CPU resource-sharing systems:

SETI@HOME
BOINC
United Devices
Entropia

Instant messaging: ICQ
Jabber (http://www.jabber.org/)
became Extensible Messaging and Presence Protocol (XMPP)

Conferencing:

T.120 – Netmeeting, Sun Forum & SGI Meeting
- white-boarding, voice over IP (Skype)

Concluding Remarks

1. P2P, Things to Know
   a) What is P2P: “Connecting devices at the edges of the internet”
   b) Know basics of P2P Environment e.g. transient, hostile, NAT, Firewalls
   c) Know good P2P Techniques - decentralization, equal peers etc.
   d) Examples: Napster + Gnutella + KaZaA – know the differences