CPSC 317 COMPUTER NETWORKING

Module 3: Application Layer Protocols – Day 5 – peer-to-peer

ADMINISTRATION

- Quiz 1 starting today
- PA1 due yesterday (but also 96 hours extra hours)
- PA2 starting today
- iClicker today

READING

- Reading: 2.5

LEARNING GOALS – PEER-TO-PEER

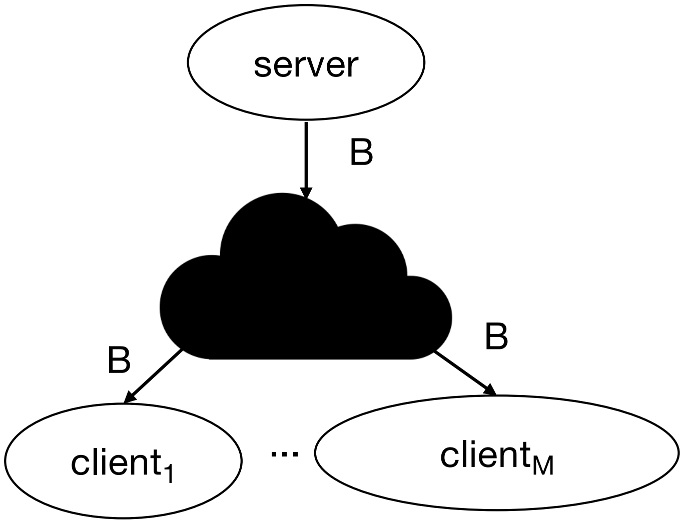
- Describe the architecture of a peer-to-peer application
- Describe the design goals for the bit-torrent protocol
  - Explain the performance benefit of peer-to-peer file sharing over a single server
- Describe the design goals for block-chain protocols
  - Understand the costs

CASE STUDY: BIT-TORRENT

- Purpose: file sharing
- Initially designed in 2001
- Protocol v2 in 2017
  - Mostly upgrading the hash function
- Each node functions as both a consumer and provider of data

FILE SHARING SCENARIO

- Suppose some N machines have a file (N might be just 1)
- Suppose some other M machines want the file (M might be very large)
- If the N machines share the file with the M machines, it might be very slow
  - Limited to the throughput possible by those N machines
  - Imagine if N is 1 and M is 1000



FILE SHARING EXAMPLES IN REAL WORLD

- Gaming
- Software updates
  - e.g., Android updates to billions of mobile phones

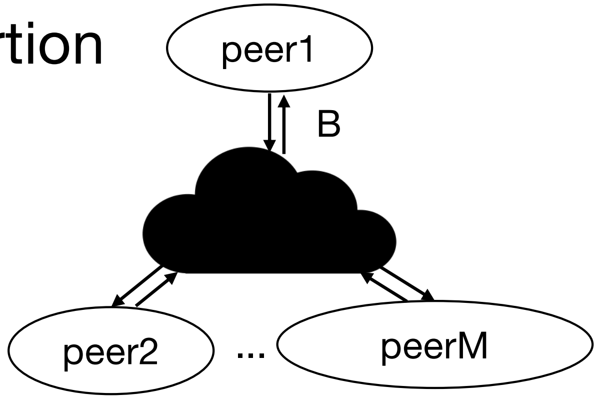
CLICKER QUESTION

Suppose there is one server with a 1Gbyte file and a 1Gbps network connection. Suppose 1000 clients want a copy of the file. How long will it take the server to deliver the file to the 1000 clients? (Choose the closest answer)

A. 2 days
B. 2 hours
C. 2 minutes
D. 2 seconds

BIT-TORRENT

- All N+M machines participate as both sources and sinks of data
- The N hosts are called “seeds”
- All hosts are called “peers”
- As soon as one of the M peers has a portion of the file it can share it with other peers



CLICKER QUESTION

Assume that all peers have a 1Gbps network connection. Suppose there is 1 peer (seed) with a 1Gbyte file. Suppose 1000 peers want a copy of the file. How long will it take the seed to deliver the file to the 1000 peers (assuming the seed is clever)? (Choose the closest answer)

A. 2 days

B. 2 hours

C. 2 minutes

D. 2 seconds

A FEW DETAILS

The file is broken into many pieces

- Fixed size (except for the last one)
- Protected by a cryptographic hash (we'll talk about these in module 8)
  - Allows reliable detection of corruption

A summary (torrent file) gives the necessary start up information:

- How many pieces
- The hash of each piece
- Somewhere to start looking for peers (Seeds, Trackers)

BASIC OPERATION

Finding other peers

- Seeds or trackers to start with
- Peer exchange: Each peer tells the other peers it is talking to regarding the peers it knows about
- The group of peers for one particular file is a “swarm”
- Each peer talks to some subset of the “swarm” at any time

Finding pieces

- Each peer shares the identity of the pieces it has with the peers it is talking to
- A peer who doesn’t have a piece asks a peer who has to share it

A FEW POLICY QUESTIONS

Which piece does a peer ask for first?

- Rarest first
- Increases the overall “health” of a file

Which of all the peers in a swarm should a peer send data to?

- The ones that are sending the most data to it (preferred peers)
- Tit-for-tat
- Random “opportunistic unchoking”

IMPLEMENTATION

- Bit-torrent is an open protocol with many implementations
- Most use TCP as the transport mechanism
- Some use μTP – a UDP-based reliable transport protocol

FINAL THOUGHTS

- Popular files can be found and obtained very quickly
- Unpopular files can be hard to find
- Often used to copy copyrighted material
  - The protocol and its implementations aren't illegal in any way
  - Copying material that someone else holds the copyright to is illegal (no matter how it is done)
- But not exclusively
  - Used by Facebook and Twitter to share content between servers

CASE STUDY: BLOCKCHAIN

- Purpose: unmodifiable transaction history
- Initially designed in 2008 (based on earlier work) by someone(s) using the pseudonym Satoshi Nakamoto
- Uses cryptography to ensure that records added to the history can never be changed or removed
- Foundation of bitcoin and many other cryptocurrencies

CENTRALIZED OR DE-CENTRALIZED

- Blockchain can be implemented both ways
- The primary value of the de-centralized approach is that it doesn't require a trusted agent
- The group of peers collaborate to decide which next "link" in the chain is accepted

MECHANISM

- All changes are broadcast (using gossip) to every other peer
- Changes are grouped into blocks
- When a block fills up, it is added to the chain
- There is no central repository of the “truth”
  - Every peer holds all of the history
  - This makes it very hard to forget the history since it is replicated very, very heavily

HISTORY "FORKS"

- There can be some disagreement over whether newly added blocks are actually in the chain
  - Concurrent modification
- The group of peers eventually agree on the history
  - Blocks are "scored" somehow, with the group of peers choosing the highest scoring history
- The likelihood of a block being removed from the history gets progressively (exponentially) smaller as it gets older

BLOCKCHAIN CASE STUDY: BITCOIN

- There are about 10,000 active computers in the bitcoin peer-to-peer network (2019)
- Anyone can join
  - Initial peers found via DNS
- Communication is built on top of TCP
  - Gossip-based
  - Share what you know that your neighbours don't

COSTS – STORAGE

- Bitcoin's block chain is growing rapidly

Year	Size
2014	20GB
2015	30GB
2016	50GB
2017	100GB
2020	>200GB

COSTS – ENERGY

- “Proof of work”
  - Each peer must demonstrate how much they like a particular version of the history by doing lots of work to support that version
  - Peers are rewarded with bitcoin for doing this (mining)
- Bitcoin is estimated to consume as much electrical energy (121 terawatt hours per year) as Argentina and more than the Netherlands (109) (22% of Canada’s 545)
- Each bitcoin transaction costs 708 kilowatt-hours
  - Electricity in BC (residential) costs $0.126 per kwh
  - 708 * 0.126 = $89.20 CAD

IN-CLASS ACTIVITY

- ICA35