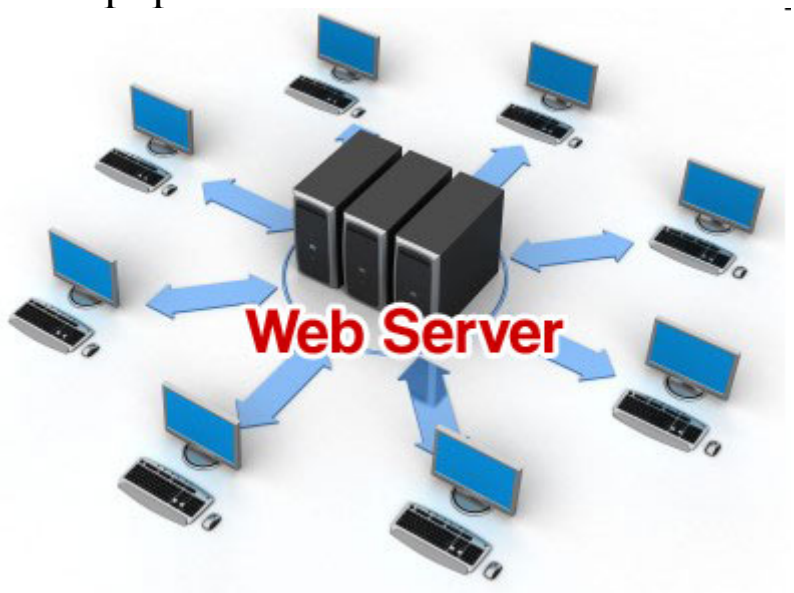
# Architektura klient-server

- Existuje jeden centrální server, se kterým komunikují uživatelé sítě
  - Např. stahují soubor pomocí http
- Ačkoliv server najdeme i v p2p sítích, architektura klient server není p2p síť




http://www.tastyseed.com/website-10/Web-Behavior

# Segmentové stahování

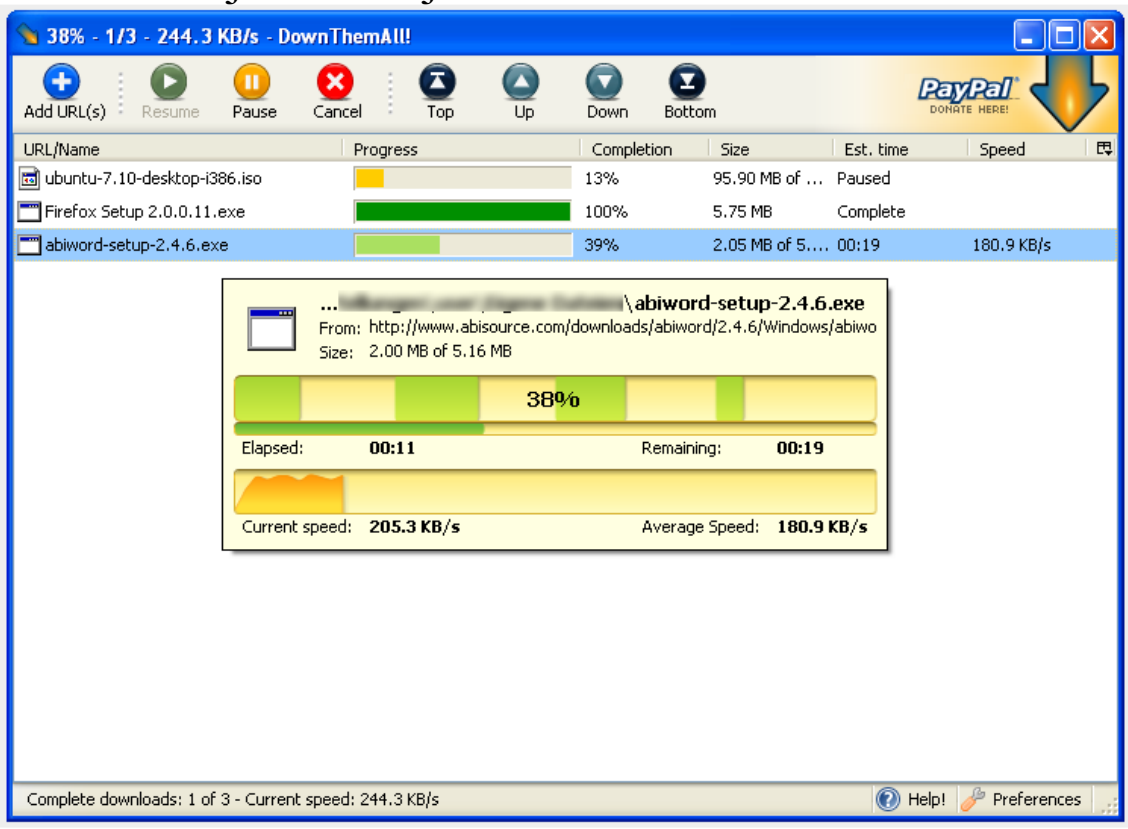
- Některé servery omezují odchozí kapacitu spojení
- V tom případě pomůže rozdělit stahovaný objem dat na jednotlivé segmenty a pro každý segment vytvořit vlastní spojení
  - => dojde k nárůstu rychlosti stahování dat

- Některé soubory jsou dostupné na několika serverech z důvodů rozložení zátěže, či poskytnutí redundance
  - A tím odolnosti proti výpadkům, když se jeden ze serverů stane nedostupným

- Nenaváže se tedy několik spojení s jedním serverem, ale několik spojení s několika servery

- Oba přístupy lze zkombinovat

- Používá se v p2p sítích, ale samo o sobě není p2p sítí
- V p2p síti její uživatelé spolu komunikují jako rovný s rovným
  - Peer to peer
  - zde je to stále jenom klient se serverem




http://www.downthemall.net/howto/screenshots/

- Zelené bloky (38%) představují stav současného stahování několika (čtyř) segmentů

- Další text jsou už p2p sítě
  - Jsou uvedeny jejich základní/hlavní vlastnosti
  - Vybrané rysy jedné se časem objevují v dalších

# DC++

- Centrální server, tzv. hub, kam se uživatel musí přihlásit, aby síť mohl používat
  - Serverů je pochopitelně několik, ale nespolupracují spolu
  - Uživatelé sítě se však mohou najednou připojit k několika hubům

- Soubory se nestahují z hubu, ale od uživatelů (peers)

- Chce-li uživatel najít požadovaný soubor,
  - Buď zadá dotaz serveru a ten ho provede
  - Anebo si řekne o seznam všech souborů konkrétnímu uživateli, kterého dohledá v seznamu hubu

  - Soubor se vyhledává buď zadáním jména
  - Anebo zadáním jeho otisku (hash)
    - Viz TTH

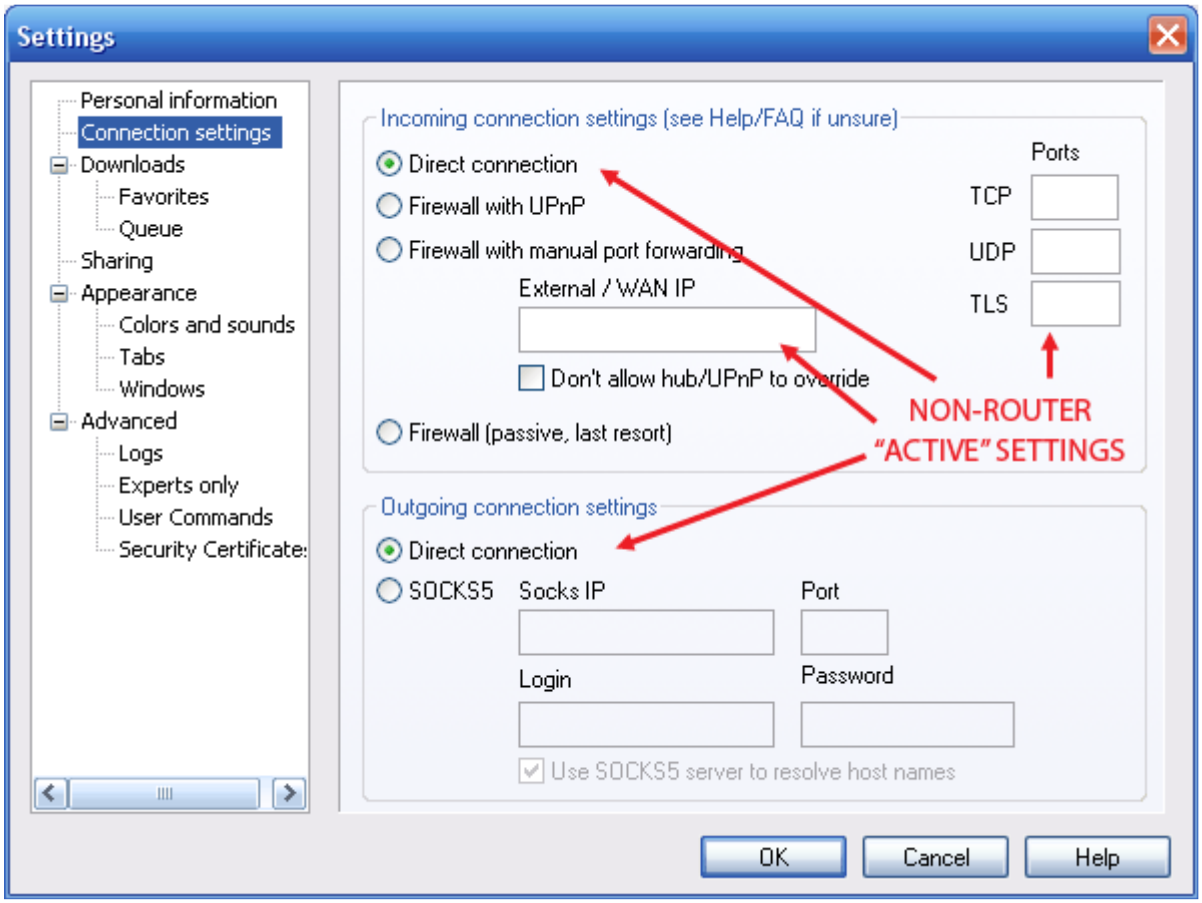
- Uživatel je buď
  - Aktivní
    - Buď má jeho počítač přímo veřejnou IP
    - Anebo je schopný nastavit forwardování portů na posledním uzlu s veřejnou IP, kterou hub vidí na cestě od něj
      - Např. domácí router, který má od ISP přidělenu veřejnou IP
  - Pasivní
    - Lze ho nastavit, i když by byly splněny podmínky pro „Aktivní“

- Chce-li uživatel stáhnou požadovaný soubor
  - Pasivní uživatel může stahovat pouze od aktivních
  - Aktivní uživatel může stahovat od všech

  - Aby byla zajištěna „jakási férovost“, uživatel musí sdílet nějaké minimální množství dat
    - Aby se vůbec dostal na hub
    - A také musí mít hubu vyhovující poměr slotů pro upload a download
      - Kolik lidí od něj může najednou stahovat
      - Od kolika lidí může najednou stahovat on

    - Nejsou-li volné sloty, uživatel je zařazen do fronty (spravuje ji aktivní klient) a čeká
      - A stále čeká… :-)
- TTH (až s Advanced DC protokolem)
  - Tiger Tree Hash
  - Metoda ověřování obsahu souborů pomocí jejich otisků
  - S ověřením se nemusí čekat, až se stáhne celý soubor, ale strom otisků umožní postupně ověřovat části stahovaného souboru, které jsou už stažené

- Kdokoliv v aktivním režimu může od hubu získat seznam všech aktuálně připojených uživatelů
- Od každého uživatele si může vyžádat seznam všech jeho sdílených souborů
- A ke každému souboru si může vyžádat jeho otisk
- A tak se může dozvědět, co vlastně uživatelé sdílí
  - Ačkoliv, se 100% jistotou to může říci teprve tehdy, až si stáhne celý soubor
    - Nebo jeho dostatečně velkou část




http://filesharefreak.com/2007/11/05/direct-connectdc/

# eDonkey

- Síť navržená k distribuování a uchovávání rozsáhlých souborů dat po velmi dlouhou dobu
- V síti existuje několik serverů, kam se uživatelé přihlašují, aby ji mohli používat
- Uživatel se hlásí jenom k jednomu serveru, protože servery spolupracují

- Server pro uživatele vyhledává soubory
  - Buď podle jejich názvu
  - Nebo podle otisku

- Uživatel má buď
  - High ID
  - Nebo Low ID
  - Viz aktivní a pasivní uživatel DC++

- Když chce uživatel stáhnout soubor
  - Získá ID uživatelů, kteří ho mají
    - Od serveru z výsledků vyhledávání
    - Nebo přes SourceExchange
      - Klienti si kromě dat sdílených souborů mohou vyměňovat i seznam zdrojů (uživatelů sdílejících daný soubor)

  - Soubor se rozdělí na několik částí a ty se postupně začnou stahovat od jednotlivých uživatelů
    - Segmentové stahování
    - Počet segmentů a počet uživatelů nemusí (a v praxi ani nebývá) stejný

  - Od klienta s High ID
    - Klient s High ID si sám spravuje frontu požadavků na stahování souborů jinými klienty
    - Pro klienta s Low ID spravuje frontu server

  - Stažené segmenty se ověřují pomocí hierarchického (stromu) systému otisků

- Předem není třeba nic sdílet, ale sdílejí se už segmenty souborů, které se teprve stahují
- V rámci férovosti lze ze sítě stáhnout jenom tolik dat, kolik se jich do ní odeslalo
  - Pochopitelně neplatí úplně na začátku

- U klienta se persistentně uchovává seznam známých zdrojů a souborů
- A to jak
  - pouze známých zdrojů,
  - tak souborů, které se právě stahují,
  - ale i souborů, které už byli staženy

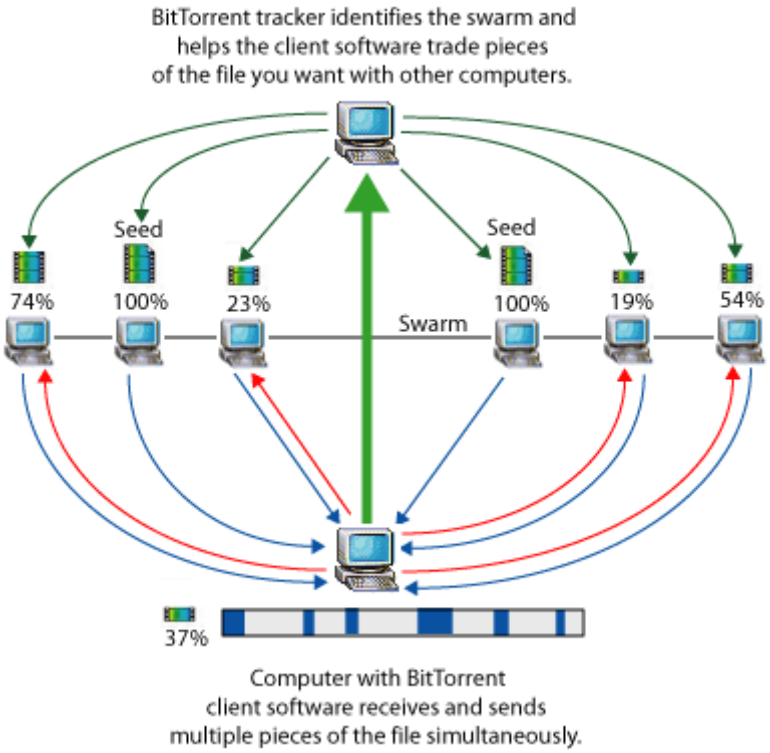
- Vezmou-li se seznamy několika uživatelů, lze z nich začít rekonstruovat, kdo si co stáhnul
  - Úspěch závisí na počtu kopií souboru
  - Na počtu a relevanci použitých seznamů
    - Chceme-li se s 100% jistotou určit, že si někdo mimo seznam stáhnul celý, konkrétní soubor

# BitTorrent

- Popis souboru ke stažení, .torrent, obsahuje URI serveru, kde lze získat seznam uživatelů, od kterých je možné soubor stahovat a otisk souboru (SHA-1)
- Server se nazývá tracker

- Soubor lze stahovat pouze od uživatele
  - Který by v síti DC++ byl aktivní
  - Který by v síti eDonkey měl High ID
  - A zároveň je k dispozici funkční tracker
    - Novější verze klientů už po vzoru ostatních sítí podporují
      - Source Exchange
      - Distributed Hash Table
      - A tak se mohou obejít bez trackeru
        - Proč tomu pak ale říkají BitTorrent?
        - eMule s Kademlií to už uměla dříve

- Leecher – uživatel, který soubor pouze stahuje, ale nesdílí
- Seder – uživatel, který soubor sdílí
- Swarm – skupina klientů, kteří zároveň sdílejí a stahují ten samý soubor




http://computer.howstuffworks.com/bittorrent.htm/printable

- Zjistit, kdo sdílí daný soubor, na to se stačí zeptat trackeru – je to standardní funkce protokolu
  - Uživatel se proti takovému dotazu nemůže nijak bránit – dáno návrhem protokolu
  - Server navíc po nějakou dobu uchovává historii

- Stejně jako u DC++ férovost sdílení vynucuje hub, i tady se musí férovost vynutit mimo protokol
- Seeder nepustí poslední díl souboru, dokud neexistuje dostatek jeho téměř kompletních kopií (až na ten díl)
- Spoléhá se na férovost jednání uživatele

# DHT

- Distributed Hash Table
  - Není název protokolu/sítě, ale know-how
  - DHT je např. Kademlia
  - Např. Gnutella je také decentralizovaná síť, ale nepoužívá DHT
    - Není škálovatelná a jako řešení se ujala DHT

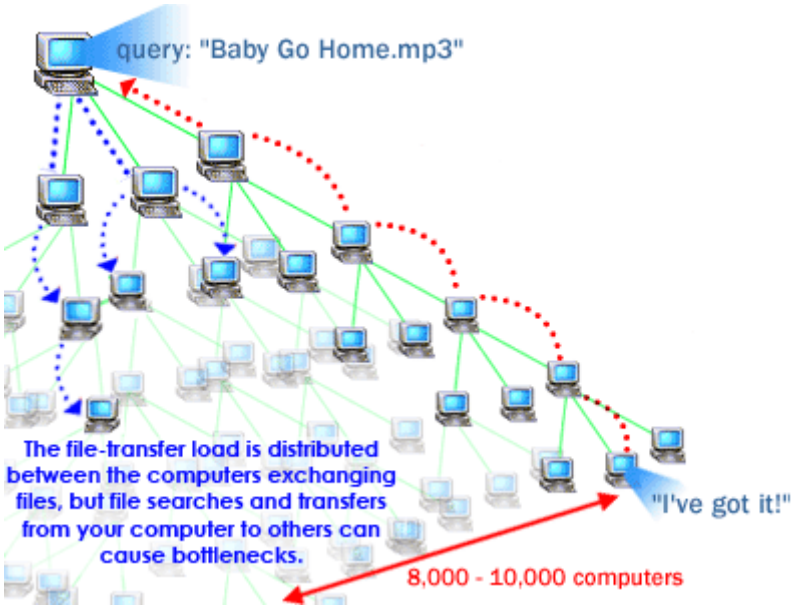
- V síti neexistuje žádný server, který by bylo možné vypnout a tím vyřadit, nebo alespoň výrazně ochromit celou síť
- Tj. celá síť je decentralizovaná a jsou v ní pouze klienti
  - Nápadně to připomíná požadavek na decentralizovanou komunikaci, kterou požadoval Pentagon a dostal ji v podobě IP
    - => Digital Battlefield
      - Vojenský komunikační systém
        - Pochopitelně mimo rozsah KIV/ZPS

- Další výhodou je škálovatelnost sítě
  - Jedno, kolik má uzlů, výkonnost roste
  - Jakýkoliv centrální server má horní hranici, kolik klientů zvládne obsloužit
    - => limit

- Pokud vypadne některý z uzlů, nic se neděje
  - => odolnost proti chybám
    - Fault-Tolerance

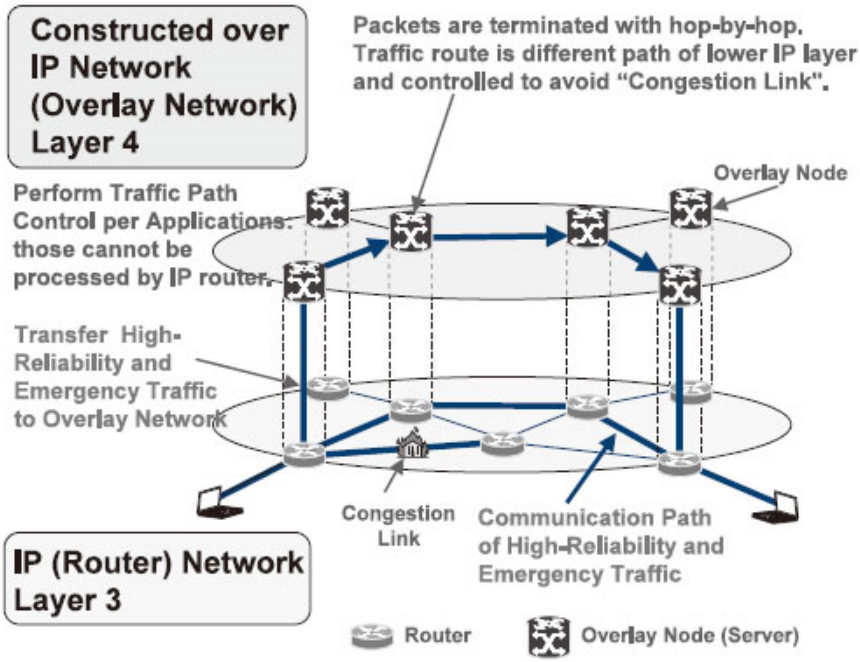
- K připojení do sítě stačí znát IP adresu jednoho jediného, libovolného klienta

- Od něj se potom získá seznam ostatních uzlů
- Distribuované systémy jako Gnutella vyhledávaly pomocí broadcastu
    - Což sice nevyžadovalo server, ale bylo to méně výkonné než se serverem



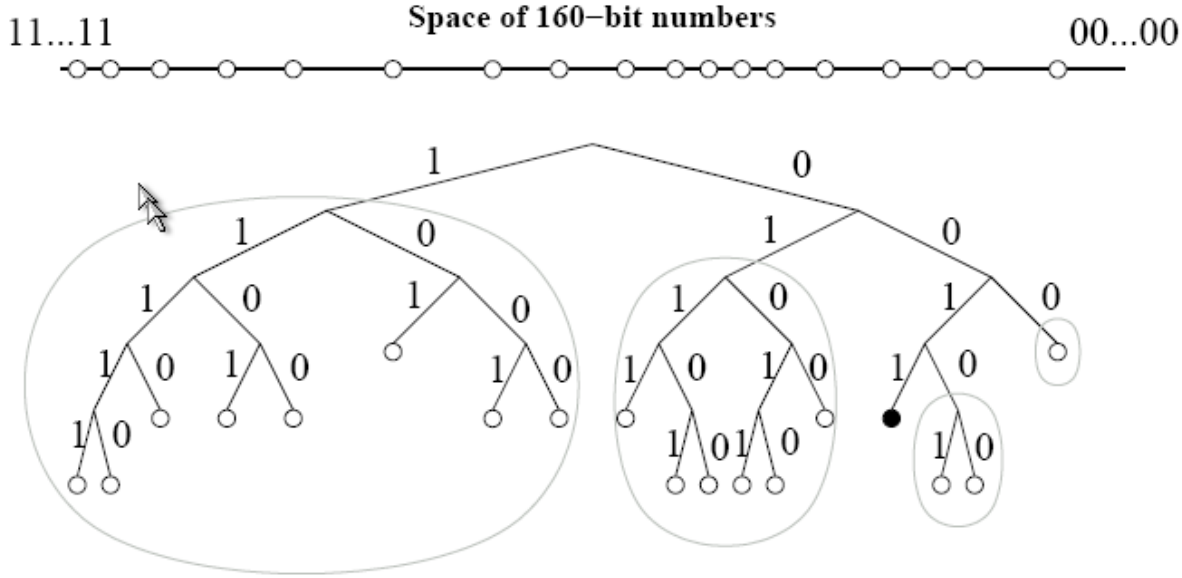

http://computer.howstuffworks.com/bittorrent.htm/printable

- DHT systém vytváří virtuální síť nad již existující sítí
    - Overlay Network nad IP



http://www.nec.co.jp/techrep/en/journal/g07/n02/070222-6.html

- Uzel tak v DHT není adresován jeho IP adresou, ale pomocí jeho ID
  - Např. 160 bitů u Kademlie
    - Ke každému bitu je přiřazen seznam uzlů, v jejichž ID je nastaven ten samý bit




http://pdos.csail.mit.edu/~petar/papers/maymounkov-kademlia-lncs.pdf

    - Najít uzel tak znamená porovnat hledanou adresu se seznamem a vybrat ty, kde je nejmenší rozdíl
      - XOR

- Každý uzel ukládá dvojice <klíč, hodnota>
  - Klíč může být otisk toho, co hledáme; velký jako ID
  - Hodnota pak může být ID uzlu, kde to najdeme
  - DHT se chová jako asociativní paměť
  - Uzel replikuje dvojice na ostatní uzly sítě
    - Dvojice mají omezenou životnost

- Chceme-li najít soubor
  - Buď zadáme přímo otisk
  - Nebo hledaný řetězec (např. název souboru) převedeme na klíčová slova a hledáme jejich otisky

- Chceme-li stáhnout soubor
  - Nejprve najdeme ID uzlů, které daný soubor vlastní
    - Nebo alespoň jeho část
  - Pak s nimi navážeme spojení pomocí IP adres
  - A soubor přeneseme

- Na rozdíl od uvedených předchozích sítí
  - Je stále možné dotazem do sítě zjistit, kdo sdílí daný soubor
  - Ale v DHT nedostaneme jeho IP adresu, nýbrž pouze jeho ID

- Chceme-li získat IP adresu konkrétního uzlu, musíme se ho zeptat
  - Ovšem, i my se ho ptáme z nějaké IP adresy
  - Pokud se však správce rozhodne filtrovat příchozí požadavky podle IP adresy, ze které přicházejí
    - A naši dá na černou listinu
      - Tak se ji tímto způsobem nedozvíme

  - Další možností je získat záznamy některého jiného uzlu, který už přímo komunikoval s uzlem, jehož IP chceme zjistit
    - Tomu už majitel uzlu, jeho IP chceme zjistit, nedokáže zabránit

- Možností je použít anonymizační vrstvu, např. TOR, a tak skrýt svou IP adresu
  - Někteří klienti DHT sítí však používají UDP, ne TCP => možný problém

- I tato situace vedla k vytvoření anonymních P2P sítí

# Anonymní p2p

- Příkladem
  - Freenet – distribuované datové úložiště
  - I2P – komunikační vrstva

- Motivací je
  - Každý režim není považován za demokratický

  - Svoboda projevu
    - Na všechny témata nelze promluvit, aniž by se mluvčí nevystavoval riziku postihu
    - Typicky webová fóra umožňují cenzurovat jim nepohodlné názory
  - Možnost svobodně číst různé informace
    - Aniž by tím na sebe čtenář neupoutával pozornost a nevznikalo neodůvodněné podezření
    - Protože by se čtenář mohl obávat postihu za vyhledávání určitých informací, mohl by se raději rozhodnout je nečíst
    - Ve svém důsledku by to vedlo k umlčování informačního zdroje

- Vývoj anonymních p2p sítí je podporován i na úrovni vlády
  - TOR a FreeHaven obdržely granty od US Navy
  - Finanční podporu projektu TOR si je v USA možné odečíst z daní

- Bohužel se i tak najdou lidé, kteří poskytovanou anonymitu zneužijí
- Primárním cílem je odstranit informaci, kdo si s kým vyměňuje data
  - Blog, fórum či e-mail jsou také datové soubory

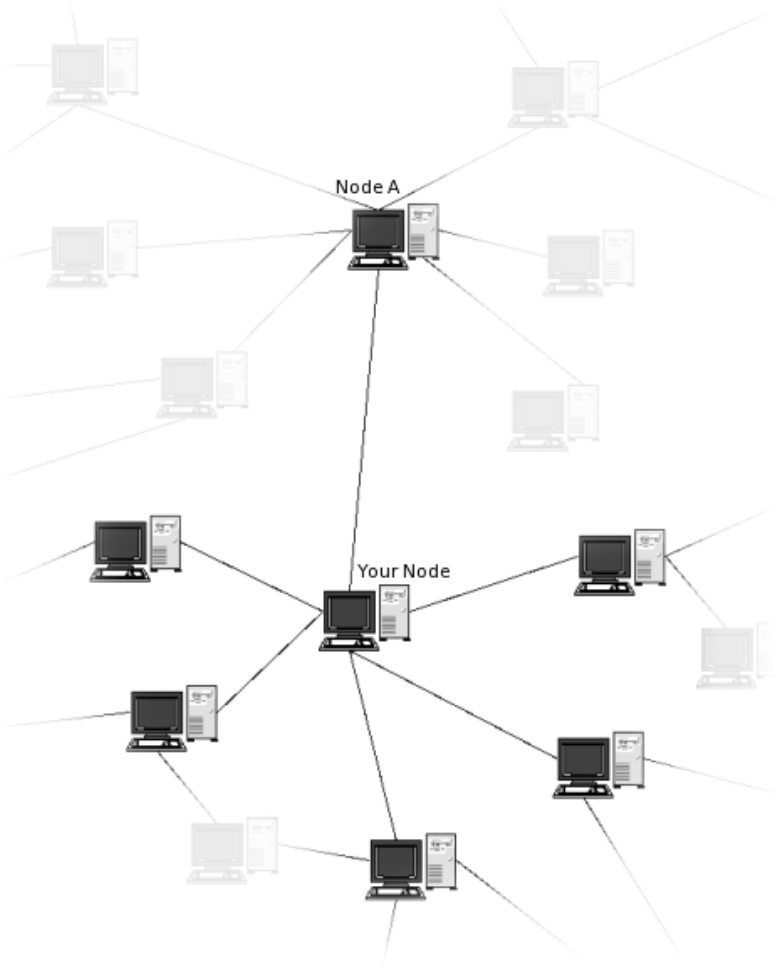
- Protokol nezná zprávu ve stylu, kdo komu něco poslal
- Uzel přijme zprávu a pošle ji dál bez ohledu na to,
  - zda byl její příjemce,
  - nebo jestli je přímý soused příjemce
  - else //všechny ostatní možnosti

  - zpráva tak sice má nějaký počáteční bod a alespoň jednoho příjemce, ale tyto informace se neuchovávají

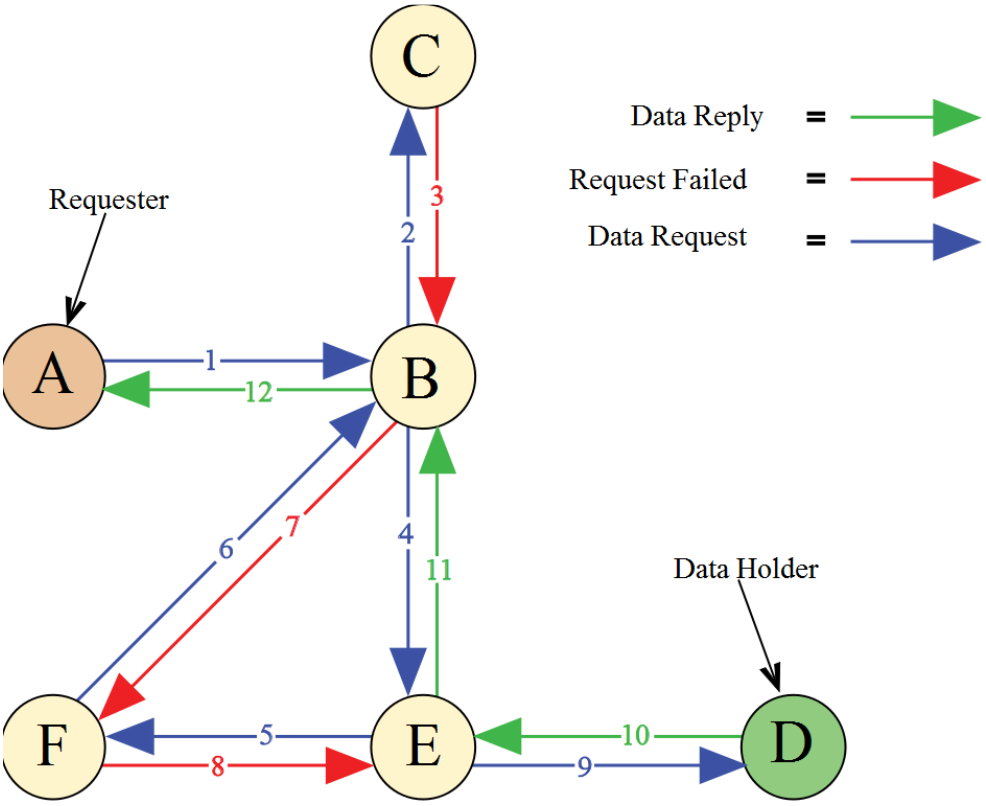
- Zároveň už tak neexistuje přímé spojení dvou klientů podle IP adres
  - IP adresu znají pouze sousední klienti
    - Ve smyslu překryvné sítě
    - A – B – C
      - B zná IP adresu A, C ji však nezná
  - V uvolněném režimu může být sousedem kdokoliv a síť je flexibilnější, než když
  - sousedem může být pouze prověřený známý, což ale zaručuje větší bezpečí konkrétnímu uživateli

- Distribuce souborů v síti je zaručena tím, že uzel uchovává, cachuje, data, která přes něj procházejí
  - Systém se tak chová jako distribuovaná cache

- Nejenom, že jsou spojení mezi uzly šifrované, šifrované jsou i uložené soubory
  - Uživatel tak ani neví, jaká data vlastně uchovává




http://freenetproject.org/connect.html




http://en.wikipedia.org/wiki/Freenet

# Dění okolo p2p

- Pro zajímavost, mimo rozsah KIV/ZPS

- K 18. 7. 2008 autorovi přednášek není znám zákon, který by výslovně zakazoval p2p sítě
- P2P sítě a distribuované systémy samy o sobě nic nelegálního nedělají, k nelegálním aktivitám je používají někteří jejich uživatelé – tj. ne všichni
    - Např. Sciencenet, P2PTV, Digital Battlefield, groupware a FreeHaven (Freedom of Speech)

- Nicméně:

*The music industry made the following claims against Napster:*

*(1) That its users were directly infringing the plaintiff's copyright; (2) That Napster was liable for contributory infringement of the plaintiff's copyright; and (3) That Napster was liable for vicarious infringement of the plaintiff's copyright.*

*The court found Napster guilty on all three claims.*

http://en.wikipedia.org/wiki/Napster
A&M Records, Inc. v. Napster, Inc., 239 F.3d 1004, 1013, 1020 (9th Cir. 2001)

- Následně se pak můžeme ptát, proč musíme platit autorský poplatek za jakýkoliv datový nosič, ačkoliv na něj umístíme např. výhradně vlastní fotky a video z dovolené
    - Nebo proč ho zaplatí i student, který si byl okopírovat volně šířené texty přednášek

- Je také možné se ptát, kolik lidí by si opravdu koupilo dané dílo, pokud by nebyl jiný způsob, jako ho získat
  - Tj. zda opravdu dochází k finanční ztrátě
  - Příkladem lze argumentovat hypotetickým(?) filmem, který je prakticky ztráta času vidět, ale má výborně provedený trailer

- Takových otázek existuje více a lze se domnívat, že některé odpovědi na ně dávají některým lidem pocit, že už tak platí za něco, co neprovedli – autorský poplatek
  - A tím si zřejmě ospravedlňují překračování zákona
  - Zřejmě to také slouží jako motivace pro tvůrce příslušného software a provozovatele serverů

**Kingston - 8GB DataTraveler Vault Privacy**
USB 2.0 FlashDisk, 256-bit AES hardwarová enkrypce dat


| | |
|---|---|
| **Výrobce** | Kingston |
| **Kód** | |
| **Cena** | 4754 Kč |
| **Cena s DPH** | 5657 Kč |
| V ceně zahrnuto: | |
| recyklační poplatek | 0.00 Kč |
| autorský poplatek | 96.00 Kč |
| **Záruka** | 60 měsíců |
| **ks** | Kup |

**Podrobný popis**
Kingstons DataTraveler Vault Privacy Edition USB Flash drive ochrání i ta nejcitlivější data

256-bit AES hardwarová enkrypce dat
dodatečná komplexní ochrana hesly
neuvěřitelně odolný kovový kryt
výrábí se jen v USA
rychlost čtení až 24MB/s a zápisu až 10MB/s
podpora Windows Vista ReadyBoost
kompatibilní Windows Vista (32-bit only),2000 (SP4) and XP (SP1, SP2)
podrobná specifikace na http://www.kingston.com/flash/DTVault.asp?id=3

http://www.axes.cz/components.php?&show=cats&cats=components&maincat=5&subcat=70&id=3857

- Kroky podniknuté proti p2p sítím a jejich uživatelům deklarativně spadají do boje proti sw pirátství

- Např. v únoru 2006 policie zavřela eDonkey server Razorback 2
  - Kapacita1,3 miliónů uživatelů
  - 170 miliónů indexovaných souborů
  - V tu dobu už však byl rozšířen klient eMule, který podporoval Kademlii
  - A následoval nárůst na 130% původního objemu stahovaných dat

- Dále je možné jít přímo proti jednotlivým uživatelům
  - Což se děje
  - Má to psychologický efekt, protože je to nepříjemné
  - Ale viděno čistě statisticky, uživatelů p2p sítí už je prostě příliš mnoho
  - Jak protiopatření se objevily anonymní p2p sítě

- Dalším pokusem, jak ochromit síť, je kontrola nad softwarem
  - V září 2006 fa MetaMechine, Inc. zaplatila 30 mil. USD RIAA, aby se vyhnula žalobě za porušování autorských práv
    - MetaMachine vyvinula klienta a síť eDonkey
    - Data sdílí uživatelé, ne programátor
    - Viz 2. bod žaloby Napsteru
    - MetaMachine ukončila vývoj sw
  - Následoval vývoj open-source klienta a serveru
    - Klientem byla právě eMule
    - Serverem Lugdunum

    - V březnu 2000 fa Nullsoft uvolnila klienta sítě Gnutella
        - Klient byl uzavřený software
        - Hned druhý den fa AOL (vlastnící Nullsoft) jeho distribuci zastavila a zakázala Nullsoftu další práci na této síti
        - Následně se začali objevovat kompatibilní klienti
            - Open Source
            - A síť se ujala i přes snahu AOL ji zastavit

- Dalším pokusem o odrazení uživatelů od sítě se stalo nahrávání falešných souborů – tzv. fake
- A pokusy zahltit síť např. příliš častými dotazy na vyhledávání
- Případně blokování TCP/UDP portů
    - Odpovědí byla např. uzavřenost zdrojového kódu serveru Lugdunum, ač je jinak poskytován zdarma
    - Dále byla omezena možnost příliš častého, nebo příliš mnoha současných vyhledávání
    - Zavedly se komentáře a hodnocení souborů, takže když někdo stáhnul fake a ohodnotil ho, ostatní ho už nestahovali
        - Díky otiskům bylo úplně jedno, kolik různých jmen soubor měl
    - Zavedlo se „protocol obfuscation“ a možnost konfigurování portů, které se mají použít

- Dále vznikly seznamy IP adres firem pracujících např. pro RIAA a uživatelé je začali blokovat

- Zatím jsme se bavili pouze o tradičních IP sítích

- Programovatelné sítě mají daleko větší potenciál
- Dalším potenciálně možným vývojem je např. použití steganografie
  - Album fotek v němž bude ukryt soubor s jiným obsahem?

- Zdá se, že každé represivní opatření proti uživatelům p2p sítí přineslo z jejich strany větší odhodlání a technologický pokrok
  - Což se dalo předvídat
  - Pokud RIAA, MPAA a další nezmění svůj přístup, zřejmě bojují předem prohraný boj?

http://en.wikipedia.org/wiki/Borg_Queen

*Lower your shields and surrender your rights.*
*We will add your movie and music files to our own.*
*Resistance is futile.*