

Opdracht 1 - Peer-To-Peer

a) Can you mention some other popular P2P file sharing protocols?

Een aantal andere populaire P2P protocollen zijn:

- *Gnutella*, deze wordt o.a. gebruikt door BearShare en LimeWire);
- *Bitcoin*, bekend van uiteraard Bitcoin;
- *FastTrack*, gebruikt door o.a. Kazaa.

b) What are the stakeholders in a peer-to-peer network like XChange? What are their concerns?

Er zijn een aantal stakeholders betrokken bij het project met verschillende belangen. Ze staan hieronder opgesomd.

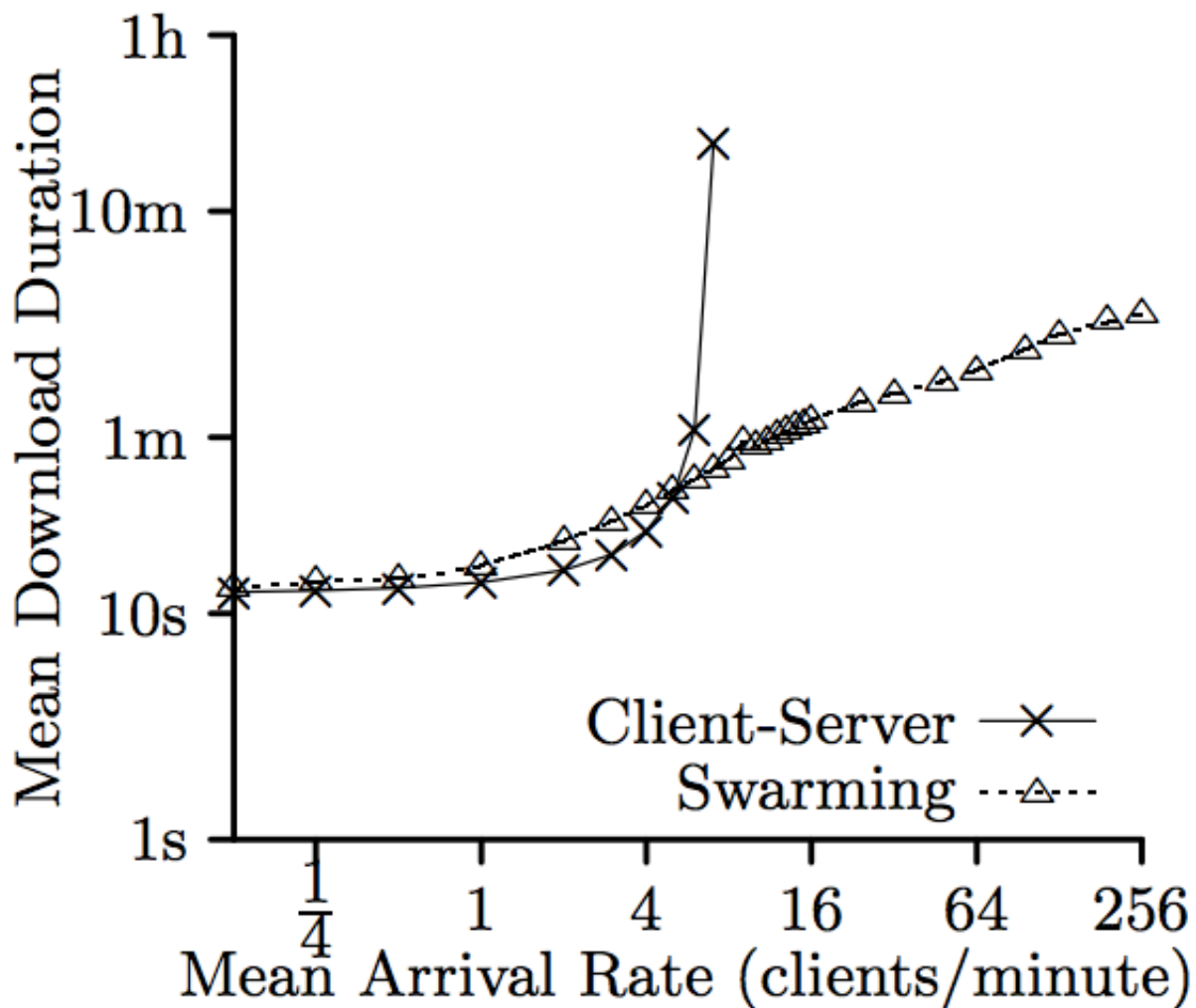
- Opdrachtgever
 - autonomy: de opdrachtgever verwacht dat het systeem zonder tussenkomst van een administrator zal werken;
 - affordability: de opdrachtgever ontwerpt een systeem voor studenten, dus het moet betaalbaar zijn.
 - efficiency: het moet een betaalbaar systeem zijn, dus het zal zo efficiënt mogelijk om moeten gaan met de resources.
- Gebruikers
 - availability: het is belangrijk dat de bestanden altijd beschikbaar zullen zijn;
 - dependability: het moet een systeem worden, waarop te vertrouwen is; een systeem dat niet telkens faalt wanneer het nodig is;
 - usability: voor de gebruiker, en zeker voor studenten, is het belangrijk dat het systeem eenvoudig is in gebruik.

c) What are the advantages and disadvantages of P2P file sharing compared to client/server file sharing? Discuss at least the 4 major perspectives (from the SSA book).

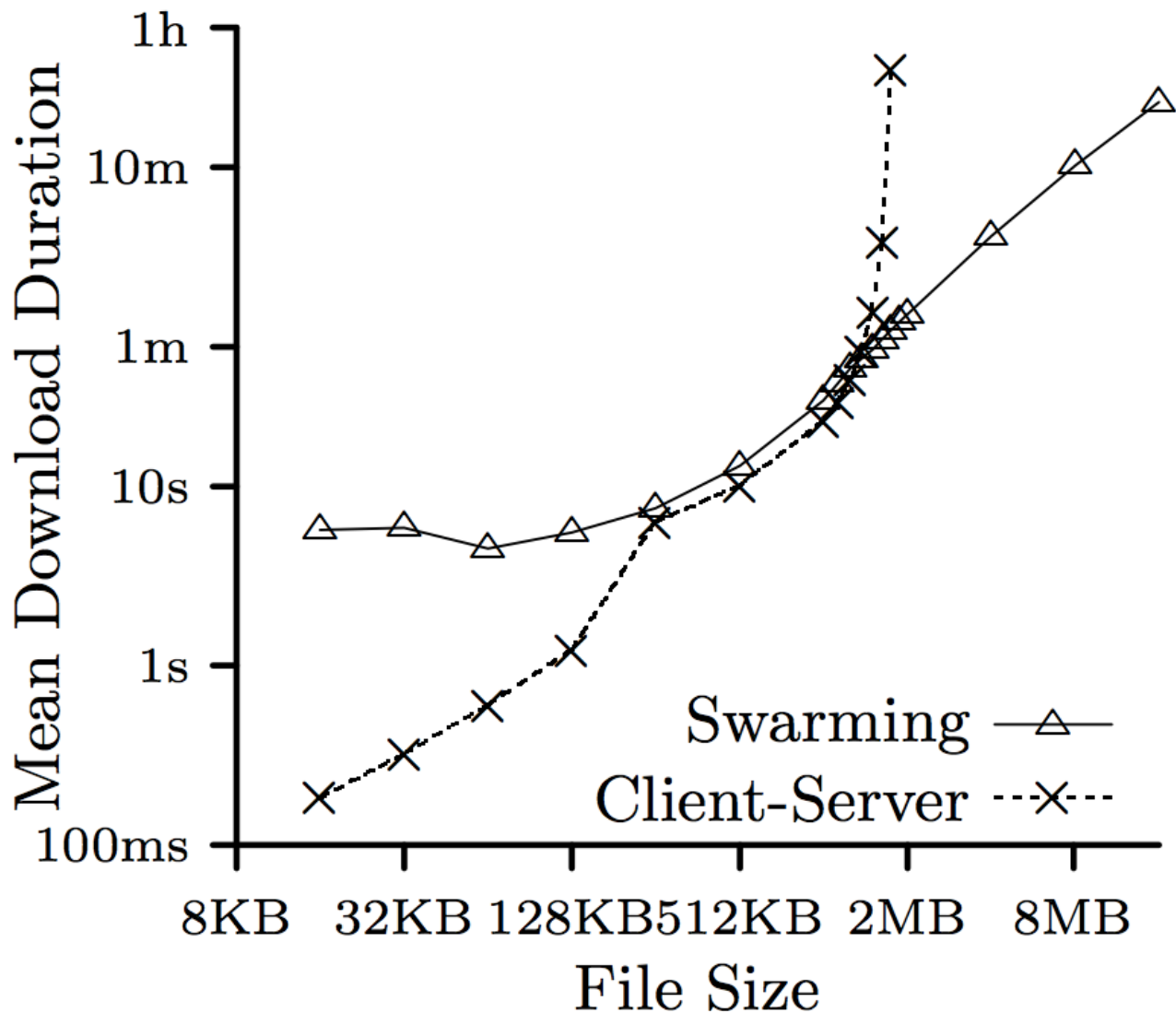
- Security perspective: Er vindt geen scan op malicious software plaats bij P2P bestandsdeling. Bij client/server bestandsdeling wordt er vaak wel gescand op dergelijke software. Daarnaast is het zo, dat wanneer er zich een virus bevindt in een bestand dat gedeeld wordt, dat dit in een P2P netwerk veel sneller verspreidt dan in een client/server model. Het virus is in het P2P model ook veel lastiger uit te roeien, omdat de bestanden met het virus constant verspreidt blijft worden. Bij het client/server model is er maar 1 bestand dat verspreidt wordt, namelijk die op de server. Wanneer dit bestand dus vrij gemaakt wordt van virussen, zijn alle verspreidingen van het bestand in de toekomst dus ook vrij van dat virus.
- Performance and scalability perspective: Zie antwoord op vraag 1.d.
- Availability and resilience perspective: Zie wat betreft de beschikbaarheid vraag 1.g. Hierin wordt niet zo zeer ingegaan op resilience. Resilience houdt in dat het systeem zich kan herstellen na fouten die zich hebben voorgedaan. Wanneer bij het client/server model zich een fout voordoet, dan resulteert dat in veel gevallen in de situatie dat het bestand niet beschikbaar is. Bij het P2P model is dat niet het geval. Wanneer door een failure een bestand corrupt raakt, is dat in de meeste gevallen geen probleem. Wanneer het bestand niet volledig beschikbaar is, kan dit opgevangen worden door het feit dat iedere client het bestand aan de hand van bestandsdelen weer opnieuw kan opbouwen. In weze kan het P2P netwerk, wat betreft de beschikbaarheid van bestanden, zichzelf na een failure weer herstellen.
- Evolution perspective: Het client/server model kan beter omgaan met veranderingen. Wanneer er bijvoorbeeld nieuwe software geïnstalleerd moet worden, hoeft dat bij het client/server model maar op 1 plaats de software geupdatet te worden. Hierbij gaan we er vanuit dat de software van de server geupdatet moet worden en niet de software van de client. Het nadeel hiervan is dat de server voor alle clients tijdelijk onbereikbaar is. Bij het P2P model, draait de software niet op een centrale server, maar bij alle peers. Wanneer het protocol bij P2P geupdatet zou worden, moet dit protocol dus bij alle peers geupdatet worden. Dat heeft heel wat meer voet in aarde dan, zoals bij het client/server model, de software op 1 locatie updaten.

d) Why do P2P networks have excellent scalability? Give numerical examples and graphs to support your answer.

Bij een client/server netwerk is de grootste bottleneck de capaciteit van de server. Wanneer het aantal clients verdubbeld, moet de server zijn resources verdelen over het dubbele aantal clients. Dat betekent dat iedere client in plaats van, bijvoorbeeld, 1024 MB/s tot zijn beschikking heeft, dat het maar 512 MB/s tot zijn beschikking heeft. Dat is bij P2P niet het geval. Bij een P2P netwerk wordt gecontroleerd wie het bestand aanbiedt. Degene die het bestand het snelste kan leveren, zal het bestand vervolgens leveren. Bij het P2P netwerk geldt dus: wanneer het aantal peers verhoogd, zal de beschikbaarheid van bestanden verhoogd worden en de snelheid waarmee een bestand gedownload wordt, minder snel oplopen dan bij een client/server bestandsdeling. Dat is ook terug te zien in onderstaande grafiek.



Daarnaast is de bestandsgrootte van invloed op de downloadsnelheid. Bij een client/server bestandsdeling is de impact van de bestandsgrootte echter veel groter dan bij een P2P bestandsdeling. Uiteraard is de benodigde tijd voor het downloaden van een groter bestand meer dan bij een kleiner bestand. Maar bij een client/server bestandsdeling stijgt deze tijd exponentieel, terwijl deze bij P2P lineair stijgt. Zie hiervoor ook onderstaande grafiek.



e) Explain in what way the integrity of (shared) files is preserved.

Bij peer-to-peer is het niet noodzakelijk dat iedere peer over het volledige bestand beschikt. Het kan voorkomen dat een peer maar de helft van een bestand heeft, omdat de andere helft corrupt is. De integriteit van het bestand kan dan nog steeds gewaarborgd worden. Een andere peer kan namelijk over de tweede helft van het bestand beschikken. Een derde peer kan dan bij beide andere peers de benodigde delen van het bestand ophalen en het bestand reconstrueren.

Bij een client/server is dit niet mogelijk. Wanneer een bestand corrupt is op de server, kan het bestand niet gereconstrueerd worden, omdat er maar 1 versie van beschikbaar is.

f) Explain what is meant by torrent index and tracker. And why is it difficult to forbid torrent indexes and trackers?

Een tracker is een HTTP(S) service die gebruikt wordt om te managen waar de peers zich bevinden. Peers weten in principe niet van elkaars bestaan af. Wanneer een peer zich aanmeldt bij de tracker, communiceert het over welke bestanden/bestandsdelen het beschikt. Wanneer een andere peer om een bestand/bestandsdeel vraagt bij de tracker, koppelt de tracker deze twee peers met elkaar, zodat de bestandsoverdracht kan plaatsvinden.

Een torrent index is een index van alle torrents die te downloaden zijn. Binnen een torrentbestand ligt vervolgens opgeslagen welke trackers beschikbaar zijn voor dit bestand.

Het is niet verboden om de locatie van een bestand op te slaan. Daarnaast is het ook niet verboden om, wanneer een gebruiker daar toestemming voor gegeven heeft, op te slaan waar deze zich bevindt. Zeker niet als dit maar tijdelijk opgeslagen wordt. Het is dus heel moeilijk om torrent indici en trackers te verbieden.

g) Evaluating availability, what are in your opinion the weak points in the architecture of the protocol?

De peers in het peer-to-peer netwerk zijn heel vluchtig. Het komt vaak voor dat wanneer een peer een bestand heeft gedownload, dat het de verbinding sluit. Daardoor daalt de beschikbaarheid voor andere peers in het netwerk. Wanneer het totale aantal peers in een netwerk hoog genoeg is, is de fluctuatie in het aantal peers niet zo sterk merkbaar. Maar wanneer het aantal peers in een P2P-netwerk wat lager is, neemt de beschikbaarheid van een bestand dus ook af.

h) Can you think of a design pattern that is comparable to the role of a tracker in the bittorrent protocol?

Het Broker Pattern is enigszins vergelijkbaar met de rol van de tracker in het bittorrent protocol. Het Broker Pattern wordt gebruikt bij gedistribueerde applicaties. Het broker component is, net als de tracker, verantwoordelijk voor het coördineren van de communicatie.

i) Can you imagine how the architecture of bittorrent could be applied to a complex application or system? If so, give an example.

Ja, het kan gebruikt worden in applicaties/systemen met een grote hoeveelheid data verdeeld over meerdere servers. Je zou daarbij kunnen denken aan een systeem op de schaal van Facebook. Facebook beschikt namelijk over een zeer grote hoeveelheid data die over meerdere servers gedistribueerd moet worden. Wanneer een post of update op één van de servers opgeslagen wordt, moet daar een backup van gemaakt worden op alle servers. Dit kan gedaan worden door vanuit die ene server, waar de post/update op is opgeslagen, de post/update direct naar alle andere servers te distribueren. Dit heeft een aantal nadelen:

1. Het kost de server veel dataverkeer om 1 post naar alle servers te backupper;
2. Het kost veel tijd voor die ene server om een update die in China is gedaan, op te slaan op een server in Amerika;
3. Een server die veel gebruikt wordt, is ook veel tijd kwijt aan het synchroniseren van de gegevens naar alle andere servers.

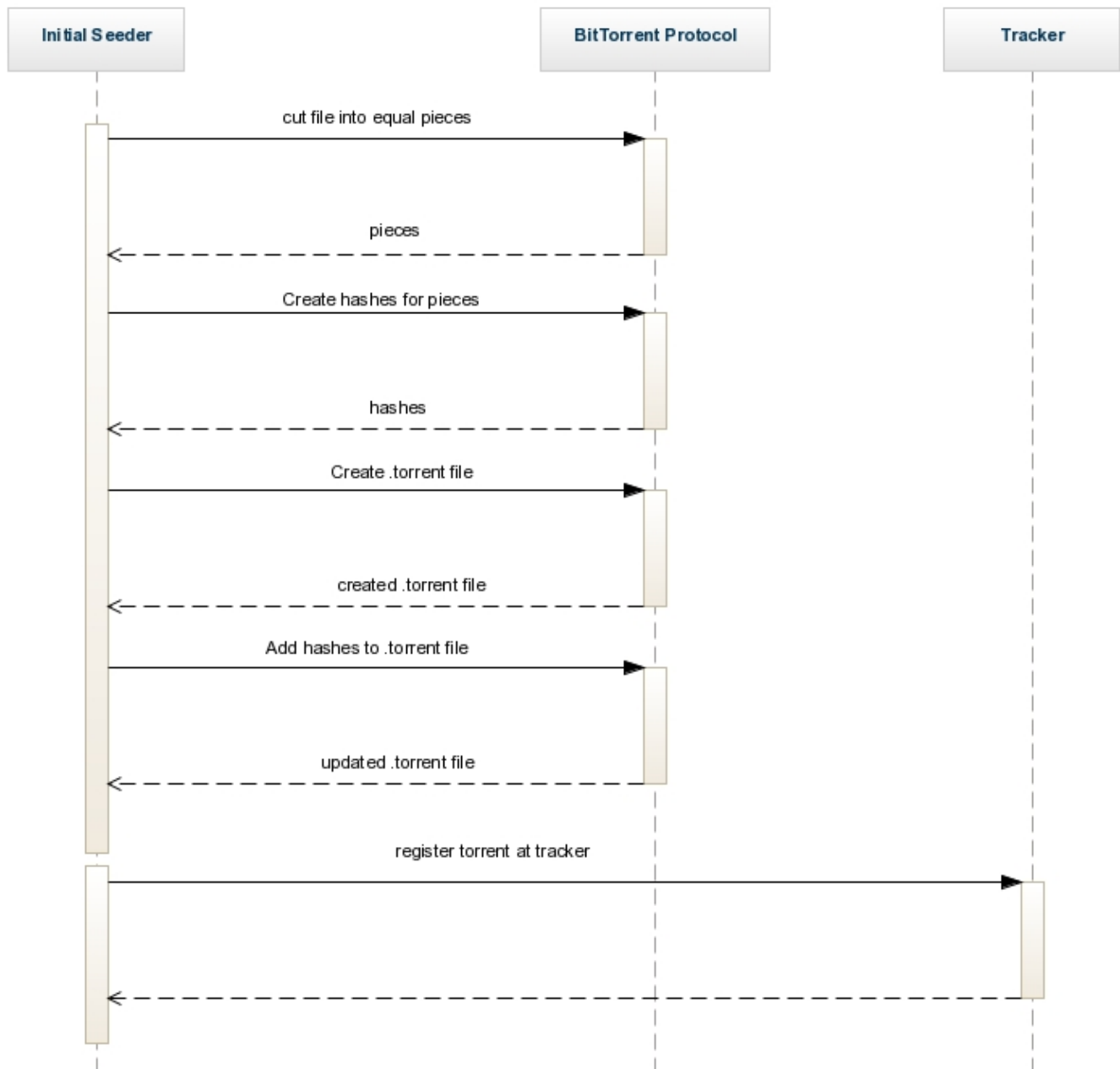
Dit kan opgeslost worden door het via torrents te distribueren. Door de updates via torrents te distribueren, wordt de load over alle servers (ongeveer) gelijk verdeeld. Daarmee voorkom je overbelasting van een specifieke server en is de post/update ook sneller verdeeld over alle andere servers.

Meer informatie hierover kan gevonden worden op [TorrentFreak](#) en [Wikipedia](#).

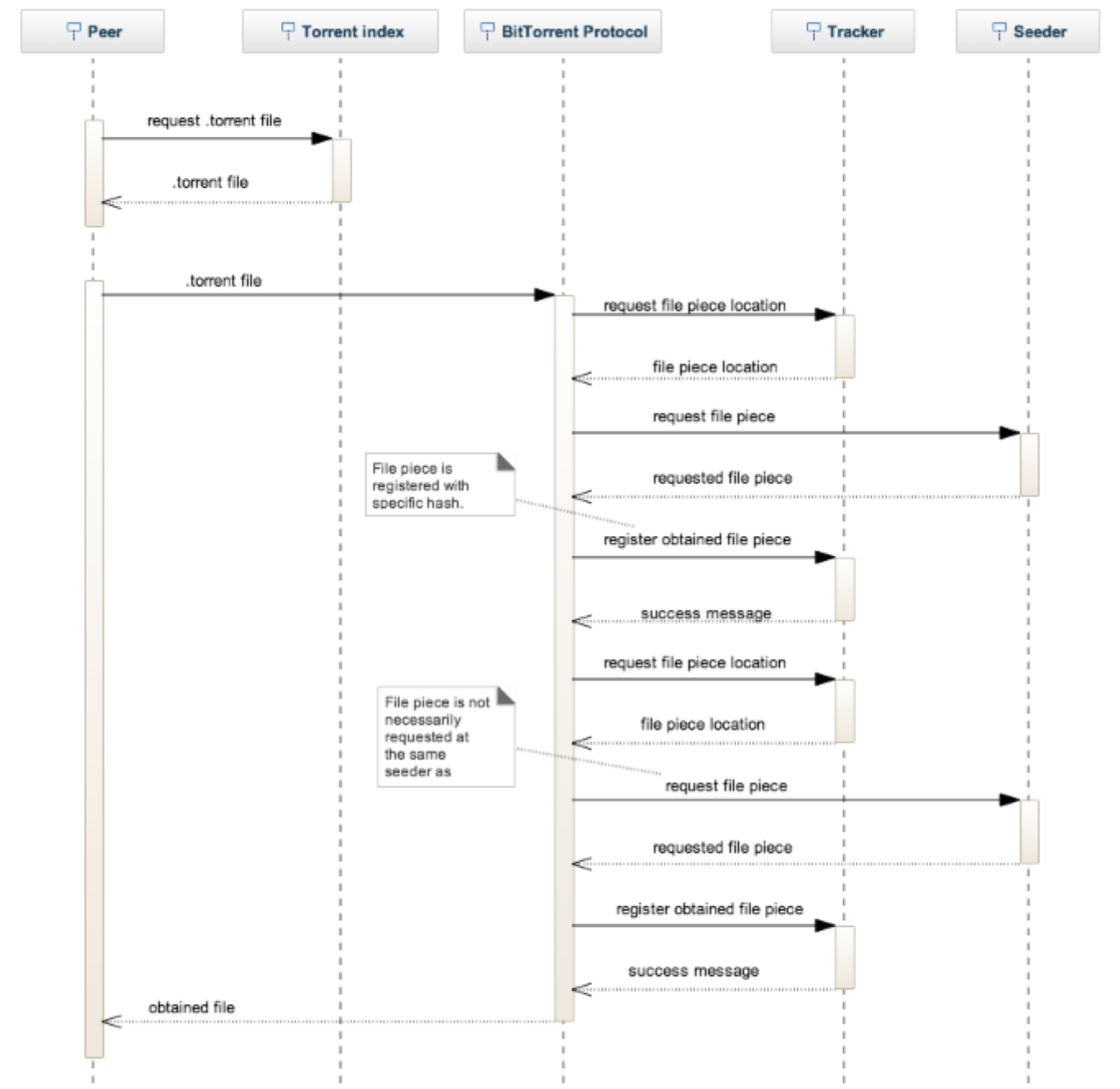
j) Explain in your own words the working of a bittorrent peer-to-peer system by drawing deployment and sequence diagrams. Supplement the drawings with text. Go into depth about which data is on which node, hashes, etc.

De vraag is dubbelzinnig gesteld: er wordt gevraagd om in eigen *woorden* uit te leggen hoe een bittorrent peer-to-peer systeem werkt, door diagrammen te *tekenen*. Aangezien de diagrammen door onszelf (dus met eigen "woorden") gemaakt zijn, gaan we er vanuit dat de tekeningen op zich volstaan om deze vraag te beantwoorden.

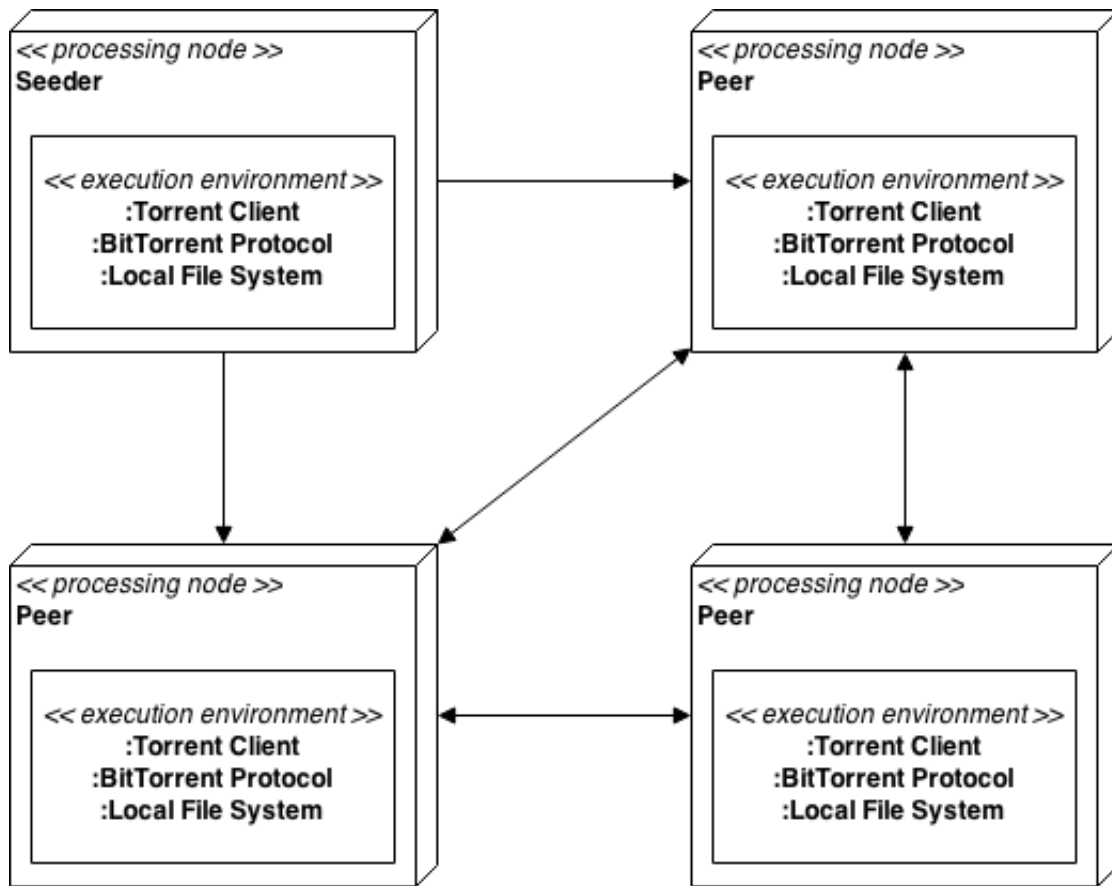
In onderstaand diagram wordt uitgelegd hoe een seeder ontstaat. Deze beschikt over een bestand en meldt zich aan bij de tracker, zodat het bestand met andere peers gedeeld kan worden. Door een compileerfout ontbreekt de return message van de tracker naar de seeder. Hier had *success message* moeten staan.



In onderstaand diagram wordt uitgelegd hoe het delen van een bestand via het bittorrent protocol werkt.



Onderstaand diagram is een deployment diagram. Hierin is terug te vinden welke software op de verschillende componenten binnen het netwerk draait. Om het overzicht te behouden is onderstaand diagram een vereenvoudigd model.



Opdracht 2

Deel 1

Deze antwoorden zijn gebaseerd op [Enterprise Architecture for Systems Engineers by IBM](#).

a) Explain in your own words what "stovepipe systems" are (p.1).

De term "stovepipe systems" (kachelpijp systeem) refereert naar een antipatroon in softwareontwikkeling, waarbij er een deel zichtbaar is voor de buitenwereld, maar het echte werk binnen gebeurt. Hiermee is er maar een beperkt deel van de functionaliteit zichtbaar/buikbaar voor anderen en zijn dingen zoals (data)opslag volledig losgekoppeld - en kan dit niet makkelijk worden gedeeld met anderen. Een voorbeeld is een systeem waarbij elk component een eigen database gebruikt voor gebruikersauthorisatie. Waardoor er nieuwe problemen, zoals redundantie van gebruikersaccounts, ontstaan.

Een voordeel van een kachelpijp systeem is zelfstandigheid, wat inhoudt dat er geen afhankelijkheden zijn met andere componenten. Waardoor de zogeheten "dependency hell" voorkomen kan worden.

b) Explain in your own words what "business/IT alignment" is (p.1).

Naarmate de IT wereld groeide is het steeds belangrijker geworden dat de IT architectuur aansluit op de doelstellingen van het bedrijf, waardoor uitgaven aan het systeem verantwoord kunnen worden door bijbehorende zakelijke voordelen. Business/IT alignment is het afstemmen van deze zakelijke doelstellingen. Dit wordt gedaan door de system engineers eerder te betrekken in het ontwikkelproces en wijzigingen gedreven te laten worden door metingen (analyse) en feedback.

Deze manier zorgt voor verbeterde transparantie en tracht de system engineers en de mensen van het management op één lijn te brengen. Dit door het niveau van betrokkenheid en communicatie te verhogen, waardoor er minder met de vinger wordt gewezen maar een ieder zijn of haar verantwoordelijkheid heeft in de gemaakte keuzes.

c) Explain in your own words what "governance" is (p.1).

Governance is bestuur en in deze context het technisch bestuur omtrent het systeem. Denk hierbij aan het sturen in het maken van keuzes vanuit een technisch perspectief. De system engineers kunnen hierin bijdragen door aan het management te laten zien welke impact bepaalde keuzes hebben op zowel projectniveau als op het niveau van de enterprise architectuur.

Door elke stap van het ontwikkelproces (het ontwerpen, bouwen, en eventueel aanpassen of uitbreiden van het systeem) ook vanuit een technisch oogpunt te bekijken, kunnen er bij voorhand problemen worden voorzien. Doordat bepaalde keuzes technische implicaties kunnen hebben en hier vroegtijdig een oplossing voor gevonden kan worden, is de impact van deze aanpassingen laag en daarmee ook goedkoop ("cheap").

d) Explain in your own words the relationship between programs and projects (p.3).

Zoals in het artikel beschreven is een program (programma) een initiatief dat wordt ondernomen om de staat van de enterprise te wijzigen door het bieden van een nieuwe of verbeterde capaciteit. En zoals beschreven is een project een ontwikkelactiviteit met een specifieke focus, begin, en eind, met als doel om een meetbaar resultaat te leveren dat bijdraagt aan de capaciteit van het programma.

De relatie is goed te zien in *Figure 2: How programs and projects affect the enterprise*. Waarin te zien is dat een programma meerdere projecten kan overkoepelen. Het programma is het einddoel, waarbij de projecten het middel zijn om dit doel te realiseren. Een programma kan in scope variëren van het aanpassen van een enkel aspect tot het wijzigen van de gehele business. Projecten implementeren het programma.

De artefacten die programma's leveren verschillen van die van de projecten. Beide leveren deltas omtrent requirements, architectuur, en testen plus nieuw gevonden defecten. Echter is dat bij een programma op enterprise niveau en bij een project op systeemniveau.

Deel 2

Intro: *The RDW ("Dienst Wegverkeer", www.rdw.nl) is a Dutch governmental organization which registers cars, drivers licenses, etc. They currently have a large legacy base on mainframe/Cobol (both databases and applications). One of their programs, which is called Platform Independence, is to migrate from the mainframe to a Microsoft platform.*

De volgende presentatie is gebruikt bij het antwoorden van de vragen: <https://www.ngi-ngn.nl/Regios/Noord/Verslagen/Lezing-RDW-succesvol/Bijlagen/1-juli-14-presentatie-over-Westpoort-datacenter-verhuizing-public.pdf>.

e) Give three reasons why the RDW would want to migrate away from the mainframe.

1. Voor een modern, geïntegreerd, multi-platform, en op PC's gebaseerde ontwikkelomgeving;
2. Om alle ontwikkeltools en mutli-platform processen te integreren in een gecontroleerde omgeving;
3. Om de fundering te leggen voor een toekomstige modernisering van de IT binnen het bedrijf.

Bron: <http://www.d-s-t-g.com/en/customers/references/trade-and-public-administration/rdw.html#3>

f) Give two reasons why the RDW would *not* want to migrate away from the mainframe.

1. De risico's die zijn verbonden aan het migratietraject. Zo kan er ongemak ontstaan voor de klanten, door bijvoorbeeld verlies van data of incorrect werkende functionaliteit (bijv. door compatibiliteitsproblemen). Dit kan zowel tijdens als na de migratie voorkomen;

2. Ook kunnen er aannames en veronderstellingen worden gedaan zonder deze correct te toetsen/testen. Dit kan ontstaan door bijvoorbeeld ontbrekende specifieke kennis van interne medewerkers. Ook kunnen de prestaties tegenvallen (bijv. energieverbruik ten opzichte van verwerkingssnelheden) en zal ook hiervoor een oplossing gevonden moeten worden. Tot slot kan het migratietraject meer tijd/geld kosten dan verwacht. Waardoor het resultaat mogelijk niet opweegt tegen de kosten.

g) Which one of the strategic goals on

<https://www.rdw.nl/overrdw/Paginas/Strategische-doelstellingen.aspx> is the main driver for this program?

Het achterliggende strategische doel voor de migratie is: *Nieuwe taken uitvoeren die het publieke belang dienen, zowel in de mobiliteitsketen als op het gebied van veiligheid en milieu.*

De migratie naar een modern Windows platform vereenvoudigd het uitvoeren van nieuwe taken en dient in het publieke belang vanwege de volgende aspecten:

- Stabiliteit van het systeem - het mag niet zo zijn dat er bij een wijziging of toevoeging een deel van het systeem stopt met functioneren. Bij het opnieuw inrichten van een infrastructuur kunnen er nieuwere, en mogelijk stabielere versies van, softwarepakketten worden gebruikt. Ook kunnen hierdoor oude beveiligingslekken worden gedicht;
- Eenvoud voor ontwikkelaars - het implementeren van nieuwe functionaliteit mag geen moeilijke taak zijn. Hierbij geldt in het geval van het RDW dat er steeds minder (goede) Cobol ontwikkelaars zijn. En dat deze uiteraard steeds duurder worden vanwege hun expertise in een niche. Door een te complex ontwikkelproces kan vooruitgang stagneren. [C# van .NET wint daarentegen steeds meer aan populariteit](#);
- Mobiliteit en schaalbaarheid - waarbij de stijgende groei in gebruik (zie voorgaande link naar presentatie) geen problemen mag veroorzaken, wat kan worden gegarandeerd door een moderne en schaalbare infrastructuur.

Bij de migratie wordt er ook rekening gehouden met de CO2 footprint van het RDW. Het RDW is zich goed bewust van de impact die ze hebben op het milieu. De onderstaande grafiek beschrijft de huidige CO2 footprint:

Emissiebron	2013	2012	2011	Vershil 2012-2013
Aardgas	1.844	1.618	1.527	14,0%
Elektriciteit	2.803	2.501	2.073	12,1%
Transportbrandstof	2.051	2.028	1.864	1,2%
OV-kilometers	30	33	15	-7,7%
Koelvloeistof	-	-	-	-
Afval	224	188	181	19,4%
Kerosine	261	351	329	-25,5%
Papier	285	308	345	-7,2%
Totaal	7.498	7.025	6.334	6,7%


De volgende afbeelding komt uit de hiervoor genoemde presentatie:

It consequenties en uitdagingen - verantwoord

Datacenters constructed in the last 10 years are functionally obsolete relative to power and cooling capacity required for new computer equipment.

- **Verantwoord ondernemen (stroom en koeling):**
 - **Nieuwe datacenters hebben een beter efficiëntie**
 - **Gedeelde datacenters zijn vaak efficiënter dan kleine opzichtstaande locaties.**
 - **Gedeelde datacenters zullen “groene” oplossingen sneller realiseren**
 - **Introducieren van nieuwe technologie zal de power consumptie drastisch verlagen.**

The heat/power problem is a mid-term problem.

 RDW

Het migratietraject zal een infrastructuur opleveren die zuiniger is in energieverbruik, wat het milieu direct ten goede komt. Het moge duidelijk zijn dat het RWD verantwoord wil ondernemen en hier ook rekening mee houdt bij het maken van keuzes.

h) Think of three possible projects that could be part of the program Platform Independence.

In de eerder genoemde presentatie wordt het migratietraject beschreven (slide 44). De individuele acties kunnen worden opgedeeld in de drie fases:

- Voorbereiding
 - Samenwerkingsverband opgericht met Noordelijke (semi-)overheden
 - Gezamenlijke aanbesteding om de bouw van het nieuwe Westpoort datacenter in Groningen te realiseren
 - Op vrijwillige basis participeren in het landelijke datacenter consolidatie programma van de centrale overheid

- Westpoort computervloer gereed maken voor het inhuizen van de RDW systemen
- Westpoort locatie redundant koppelen aan het RDW netwerk
- Verhuis- en migratieplannen maken / Voorbereiden op verhuizing
- Migratie
- Afronding
 - Rozenburg computervloer ontmantelen (ICT technisch niet EW installatie)
 - Realisatie beperkte computerruimte voor Test/Winframe witness/Tape/PKI
 - Procedures aanpassen aan Westpoort <> Zernike combinatie

Omdat de, hiervoor te ondernemen, acties op een dusdanig grote schaal worden uitgevoerd, kan elke actie als project worden beschouwd. Een drietal projecten zijn:

1. Westpoort computervloer gereed maken voor het inhuizen van de RDW systemen (1200 servers, 136 actieve netwerkcomponenten) en het redundant koppelen van deze locatie aan het RDW netwerk. Het meetbare resultaat van dit project is een fysieke basisinfrastructuur voor de uiteindelijke verhuizing/migratie;
2. Realisatie beperkte computerruimte voor Test/Winframe witness/Tape/PKI. Het resultaat is de gerealiseerde infrastructuur voor het uitvoeren van de testen en backups, waarbij er veel rekening moet worden gehouden met beschikbaarheid (plannen voor terugvallen bij uitval);
3. Procedures aanpassen aan combinatie van Westpoort en Zernike. De combinatie van nieuwe afspraken en de nieuwe locatie vereist dat bepaalde procedures worden aangepast, hiervoor moeten de veranderingen en hun impact in kaart worden gebracht. Het resultaat hiervan zijn de bijgewerkte artefacten.