

LUENTO 24

Kertaus

Koealue: (Koe to 13.12 klo 16.00 A111)

- Kirjan luvut 9-16 (ei 13) + Appendix B.4
- Luennot 11-24, harjoitukset 7-12
- Opintopiiritehtävä 3

1

Sisältöä

- Vuorotus
- I/O
- Tiedostojärjestelmä
- Hajautettu prosessi
- Tietoturva

2

Vuorotus

3

Milloin?

systemiin?

- Long-term
 - otetaanko uusi prosessi suoritettavaksi?
 - mahtuuko muistiin? riittäkö swap-tila?
- Medium-term
 - milloin (heitto) vaihdettu prosessi muistiin?
 - vapautta muistia?
 - monijakoaste?
- Short-term
 - mille prosessille annetaan CPU?

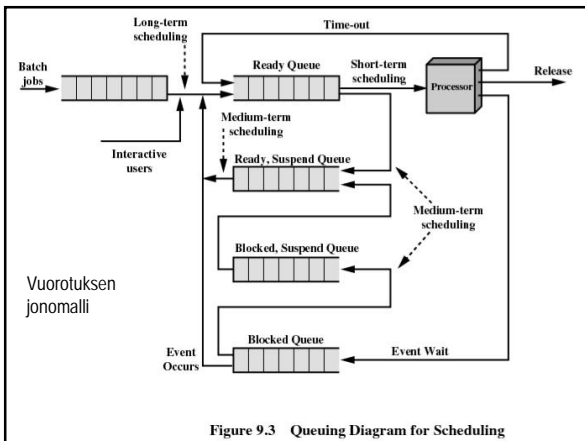
muistiin?

suorittimelle?

I/O-laitteelle?

- I/O
- minkä prosessin I/O pyyntö palvellaan ensin?

Figure 9.2 Levels of Scheduling



Kiinteä ja vaihteleva prioriteetti

	high pri		
	0	missile avoidance process	
	1	flight surface control	
	2	navigation	kiinteä pri
	3		
	...		
	63		
	64	load control	
	65	swapper	
	66	cleaner	kiinteä pri
	...		
	127		
	128	odottanut kauan RR jonossa	
	129	calculator (128-140)	vaihteleva pri
	130		
	131	käyttänyt paljon CPU-aikaa (esim. koko aikaviipale)	
	...		
	192		
	low pri		

6

Algoritmit

- First-Come-First-Served FCFS
- Round Robin RR
- Virtual Round Robin VRR
- Shortest Process Next SPN
- Shortest Remaining Time SRT
- Highest Response Ratio Next HRRN
- Multilevel Feedback feedback
- Fair Share Scheduling FSS

7

Feedback $q=1$

(Fig 9.5 [Stal05]) keskim. 10.0

- Jonot RQ0, RQ1, RQ2, ...
 - RQ0: pura alkaviipaleittain, FCFS, siirrä seuraavaan jonoon
 - RQ2..RQn-1: pura alkaviipaleittain, FCFS, siirrä seuraavaan jonoon
 - RQn: pura alkaviipaleittain, RR, pidä samassa jonoissa
- Pitkät työt voivat kestää kauan, nälkiintymisvaara

8

EDF – Earliest Deadline First

Kaksi jaksollista (periodic) työtä: Fig 10.5 [Stal05]

- Saapumisajat A 20 ms, B 50 ms välein
- Suoritusajat A 10 ms, B 25 ms
- Vuorottaminen 10 ms:n välein
- Valm. takaraja A 20 ms, B 50 ms saapumisesta

9

Rate Monotonic Scheduling

- Vain jaksollisille (periodisille) töille (Fig 10.7 [Stal05])
- Sama työ tasaisella tahdilla
 - käyttöaste yhden työn osalta on $U = C/T$
- Jakson loppu = Hard deadline
- Pienin jakso (T) = suurin prioriteetti

Rate: määrä yksikköä kohden, taajuus

10

Ajoitettavuuden arviointi käyttöasteen avulla

Tbl 10.4 [Stal05]

- RMS: Selvästi

$$U_i = C_1/T_1 + C_2/T_2 + \dots + C_n/T_n \leq 1$$

- Riittävä ehto sille, että työt voidaan ajoittaa RMS-algoritilla on

$$C_1/T_1 + C_2/T_2 + \dots + C_n/T_n \leq n(2^{1/n} - 1)$$

- Arvon n kasvaessa, RMS:lle yläraja lähenee

$$\ln 2 \sim 0.693 \text{ eli rajatapaus } U_i < 0.693$$

- EDF: tehokkaampi, sille riittää (riittävä ja välttävä ehto)

$$U_i = C_1/T_1 + C_2/T_2 + \dots + C_n/T_n \leq 1$$

12

Prioriteetin kääntyminen (Priority inversion)

12

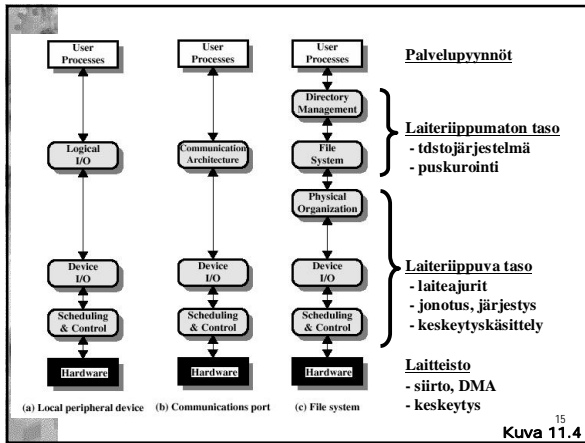
Prioriteetin kääntymisen välttäminen

- Keskeyttämättömät kriittiset alueet
 - Luovat tarpeeton odotusta.
 - Käyttökelpoisia vain lyhyille kriittisille alueille.
- Sisääntuloprotokolla kriittiselle alueelle
 - Prioriteetin perintä (Priority Inheritance Protocol).
 - Prioriteetin kattomenetelmä (Priority Ceiling Protocol).

13

I/O

14



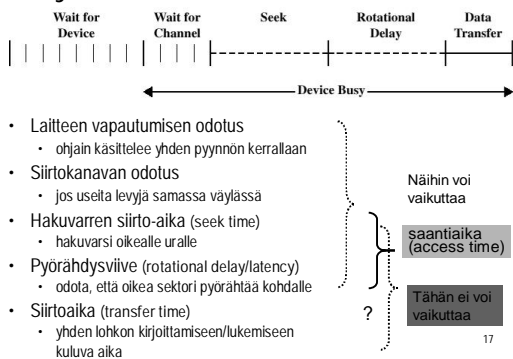
Laiteajurit

- Erityyppisille laitteille omat ajurinsa
- Etsi ajuri laitenumeron perusteella laitekuvaajalistasta
 - siirtoa käynnistettäessä
 - siirron päättyessä (keskeytys!)
- Laitekuvaaja
 - laitteen tunnistus, device id
 - tilatietoa, kenelle laite varattu
 - mitä ajuria käyttää
 - mitä ajurin funktiota (handler) kutsuttava missäkin tilanteessa
 - open(), read(), write(), close() ..., keskeytys
 - jono pyynnöistä parametreineen
 - mm. linkki pyynnön tehneen prosessin PCB:hen

16

Levyhaku

(Fig 11.6 [Stal05])



Algoritmeja

- Hakuvarren siirtoaika pisin
 - kannattaa minimoida siirrot
- Random? FIFO? PRI? LIFO?
 - huonoja, eivät huomioi hakuvarren nykyistä positiota
- Ota huomioon hakuvarren sijainti
 - SSTF
 - SCAN
 - C-SCAN
 - N-step-SCAN ja FSCAN

18

Levyn vuorotus algoritmeja

Table 11.2 Comparison of Disk Scheduling Algorithms

(a) FIFO (starting at track 100)		(b) SSTF (starting at track 100)		(c) SCAN (starting at track 100, in the direction of increasing track number)		(d) C-SCAN (starting at track 100, in the direction of increasing track number)	
Next track accessed	Number of tracks traversed	Next track accessed	Number of tracks traversed	Next track accessed	Number of tracks traversed	Next track accessed	Number of tracks traversed
55	45	90	10	150	50	150	50
58	3	58	32	160	10	160	10
39	19	55	3	184	24	184	24
18	21	39	16	90	94	18	166
90	72	38	1	58	32	38	20
160	70	18	20	55	3	39	1
150	10	150	132	39	16	55	16
38	112	160	10	38	1	58	3
184	146	184	24	18	20	90	32
Average seek length	55.3	Average seek length	27.5	Average seek length	27.8	Average seek length	35.8

19

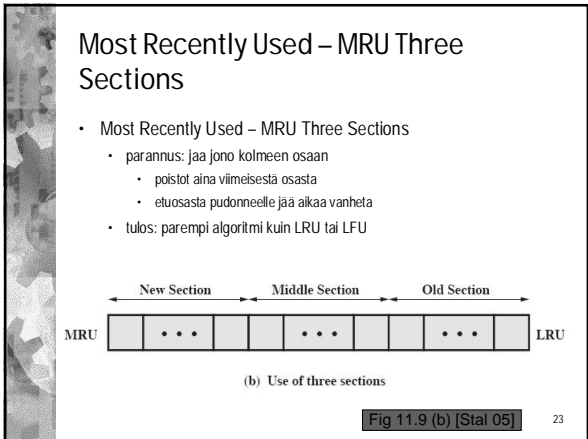
- ### RAID - Redundant Array of Independent Disks
- RAID 0 (ei redundanssia, ei toisintoja)
 - RAID 1 (mirror, kahdennettu)
 - RAID 2 (Hamming)
 - RAID 3 (pariteettibitti)
 - RAID 4 (pariteettilohko)
 - RAID 5 (hajautettu pariteettilohko)
 - RAID 6 (2 haj. pariteettilohkoa)
- 20

Lohkopuskurit

Ch 11.7 [Stal 05]

21

- ### Lohkopuskurit, levypuskurit
- KJ:n data-alueella oleva puskurit muistiinluettuja levylohkoja varten
 - Paikallisuusperiaate
 - Ennaltanouto / viivästetty kirjoitus
 - Poistoalgoritmit
 - LRU: Least Recently Used
 - LFU: Least Frequently Used
 - Most Recently Used – MRU FIFO (Frequency Based Replacement)
- 22

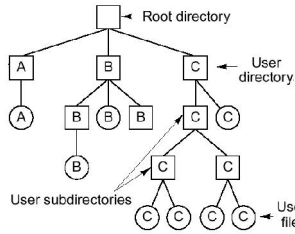


Tiedostojärjestelmä

24

Hakemisto: Hierarkinen puurakenne

- Juurihsto, kotihakemistot, alihakemistot
- Myös prosessit voivat luoda alihakemistoja
- Juurihstolla kiinteä paikka levyllä



25

Tiedostojen yhteiskäyttö

- Sama tiedosto käytössä monta kertaa
- Hard link
 - monia omistajaa (vai monella omistajan oikeudet?)
 - kaikilla samat oikeudet
 - omistaja poistaa → muilla silti käytössä normaalisti
- Soft link eli symbolinen linkki
 - tiedoston tyyppi: symbolinen linkki
 - tiedoston sisältö: merkkijono, joka indikoi varsinaisen tiedoston
 - vain yksi omistaja
 - omistaja poistaa → muiden linkit epäkelvoja

26

Hard link

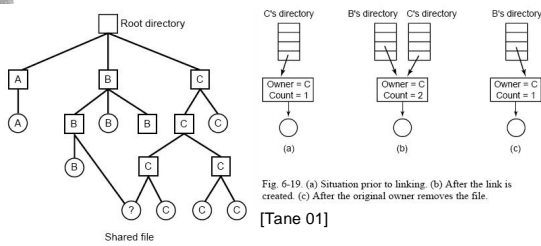


Fig. 6-19. (a) Situation prior to linking. (b) After the link is created. (c) After the original owner removes the file.

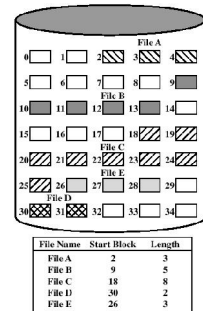
[Tane 01]

Fig. 6-18. File system containing a shared file.

27

Koko tdsto yhdelle alueelle

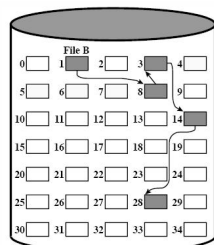
- Hstoalkiossa vain ens. lohkon numero sekä tdston koko (pituus)
- Koon muuttaminen vaikeaa
 - arvioitava varausta tehtäessä
 - saatetaan joutua kopiaimaan uudelle alueelle



28

Lohkojen ketjutus

- Varaus lohko kerrallaan vasta tarvittaessa
- Hstoalkiossa viite tdston ensimm. lohkonumeroon sekä tdston koko



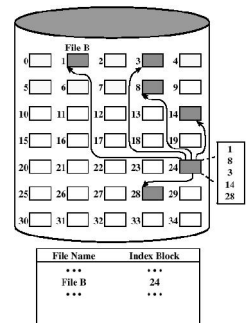
File Name	Start Block	Length
...
File B	1	5
...

Fig 12.9 [Stal05]

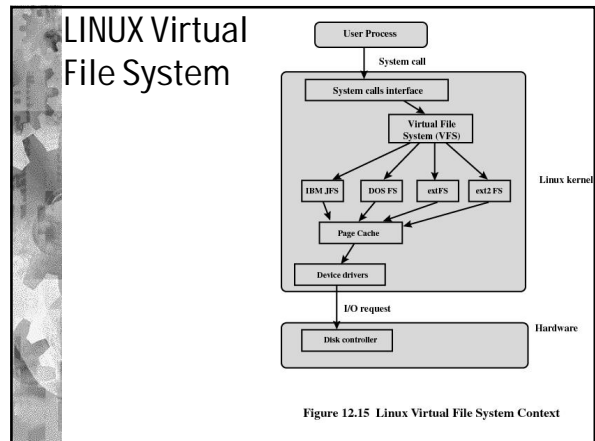
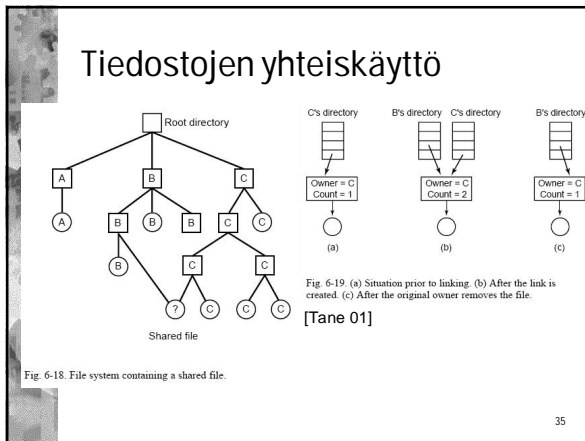
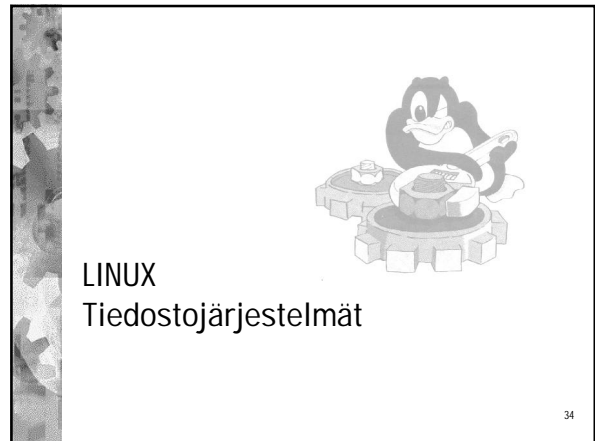
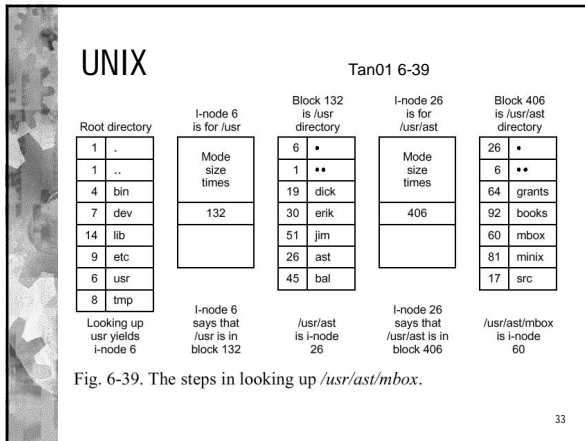
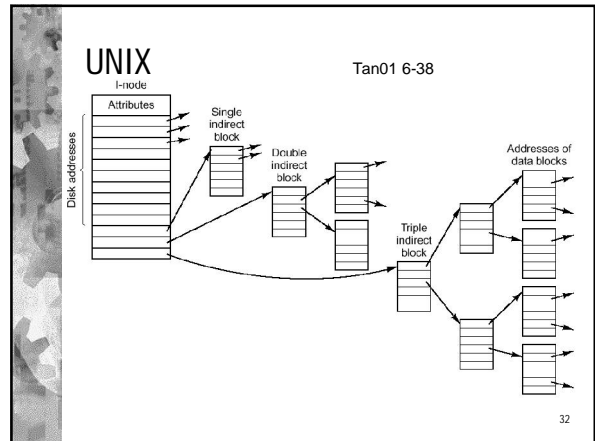
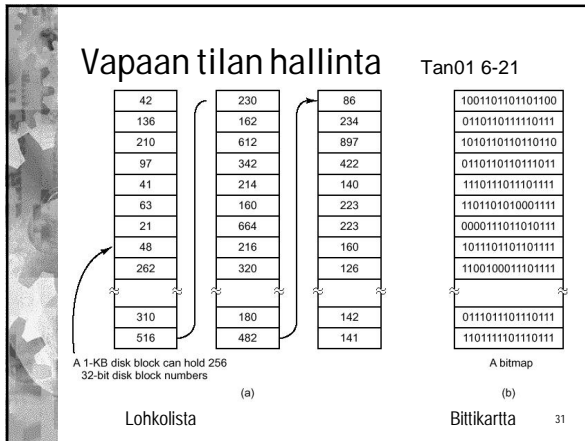
29

Lohkohakemisto

- Erillinen hakemisto tdstolle varatuista lohkoista
- Usein erillään omassa lohkoissa
 - hstoalkiossa vain hstolohkon numero



30



Linux tiedostojärjestelmät

- ext2fs (second extended file system)
 - Linuxia varten kehitetty tiedostojärjestelmä
 - esikuvana BSD Fast File System (FFS)
 - lohkoryhmit
 - tehokkuus, luotettavuus
- /proc
 - erikoistiedostot, luodaan 'lennosta'
 - esim. ytimen parametrien kysely/asettaminen
 - KJ-palvelut pilotoitu tiedostojärjestelmän käytöksi
 - käytön valvonta tiedostojärjestelmän suojauksen avulla
- ext3fs
 - journaling file system, log-structured file system (LFS)
 - Red Hat Linux'issa

37

Fig. 6-11. A possible file system layout. [Tane01]

Fig. 10-35. Layout of the Linux Ext2 file system.

38

Linux ext2fs levy

- Lohkoryhmit (block groups)
 - yhtenäisesti levyllä allokoitu alue
 - datalohkot ja i-nodet fyysisesti lähellä toisiaan
 - säästä hakuvarren silloissa
- Kaikki lohkot samankokoisia (1 KB)
- Kaikki i-nodet 128B (tavallinen UNIX 64B)

Fig. 10-35. Layout of the Linux Ext2 file system.

39

ext2fs superlohko (superblock)

0	Number of i-nodes	Number of blocks		
8	Number of reserved blocks	Number of free blocks		
16	Number of free i-nodes	First data block		
24	Block size	Fragment size		
32	Blocks per group	Fragments per group		
40	i-nodes per group	Time of mounting		
48	Time of last write	Status	Max. mnt cnt	
56	Ext2signal	Status	Error behav.	Pad word
64	Time of last test	Max test interval		
72	Operating system	File system revision		
80	RESUID	RESGID	Pad word	
Pad words				

- 1 lohko
- Kuuaa koko ext2fs-partition rakenteen
- Kopio jokaisen lohkoryhman alussa
- luotettavuus, virheestä toipuminen

blocksize

40

ext2fs i-node

0	Mode	Uid	File size
8	Access time	Time of creation	
16	Time of modification	Time of deletion	
24	Gid	Link counter	No. of blocks
32	File attributes		Reserved (OS-dependent)
40	12 direct blocks		
88	One-stage indirect block	Two-stage indirect block	
96	Three-stage indirect block	File version	
104	File ACL	Directory ACL	
112	Fragment address	Reserved (OS-dependent)	
120	Reserved (OS-dependent)		

Access Control List

41

Kirjaava tiedostojärjestelmä

- Kirjataan kaikki muutokset (journalointi?)
- Pitää tiedostojärjestelmän eheänä
- Kirjataan
 - Vain metatieto muutoksista – journal
 - Sekä metatieto että itse data – loki
- Tarve:
 - Tiedostojärjestelmän tarkistus (check) kestää liian kauan, jos epänormaali 'kaatuminen'
 - Valtaosa levyoperaatioista on kirjoituksia, lukuoperaatiot tehdään puskureista
- Useimmat kirjoitukset pieniä päivityksiä
 - levyn hakuvarsi liikkuu paljon, vähän dataa siirtyä


42

Perusidea

- Ongelma tavallisen tiedostojärjestelmän uuden tiedoston X luomisessa:
 - kirjoita hakemiston i-node, hakemisto, tiedoston i-node ja lopulta tiedosto
 - virta poikki (tms vika) kesken kaiken? Oooops.
- Ratkaisu: tapahtumaloki, joka takaa tiedostojärjestelmän eheyden - vrt tietokantojen loki
- Esim: Microsoft NTFS, Red Hat Linux ext3fs

43

NTFS: Piirteitä



- Kaatumisista ja levyvirheistä toipuminen
 - LFS lokitiedoston avulla
- Käyttöoikeudet
 - pääsyylistat (security descriptor)
- Sallii suuret levyt ja tiedostot
 - FAT32:ssa vain 2^{32} lohkoa, suuri allokointitaulu
- Tiedosto-oliot ovat (*arvo*, *attribuutti*) -pareja
- Mahdollisuus indeksointiin tiedoston käsittelyn nopeuttamiseksi
- Lohko, cluster
 - yksi tai useampi peräkkäinen sektori (esim. 512 B - 4 KB)
 - 32 GB levyllä 128 sektoria/lohko (→ lohko 64-512 KB)
 - varauksen ja kirjanpidon perusyksikkö
- Partitio, volume
 - fyysinen levyn looginen osa, jolla oma tiedostojärjestelmä

44

NTFS-partitio

(Fig. 12.17 [Stal 05])

partition boot sector	Master File Table	System Files	File Area
-----------------------	-------------------	--------------	-----------

- **Boottilohko**
 - partition ja tiedostojärj. rakenne, boottitietue ja -koodi
 - MFT:n sijainti
- **MFT**
 - tietoa tiedostoista, hakemistoista (folders) ja vapaasta tilasta
- **System Files (- 1MB)**
 - kopio MFT:n alkuosasta
 - virheistätoipumislöki, bittikartta vapaat/varatut lohkot, attribuuttien kuvaustaulu
- **File Area** - tiedostojen lohkoille

45

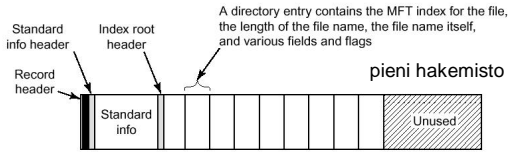
NTFS – MFT (Master File Table)

- 1 KB:n kokoisia MFT-tietueita
 - jokainen kuvaa yhden talliolla olevan tiedoston
 - myös hakemisto on tiedosto
- vaihtelevanmittainen osa käytössä
 - (attribuutti, arvo) pareja (ei paikkasidonnainen!)
 - data attribuutti, "arvo" = lohkojen sijainti
- **16 ensimmäistä** tietuetta varattu ns. metadatalle
 - 16 \$-alkuisia tiedostoa
- Jos pieni tiedosto, tietue **sisältää myös datan**
- Jos iso tiedosto, data erillisellä tallealueella
 - MFT-tietueessa lohkonumeroita
 - kuvaus voi jatkua useampaan MFT-tietueeseen

46

Hakemiston MFT-tietue

(Fig. 11-38 [Tane01])

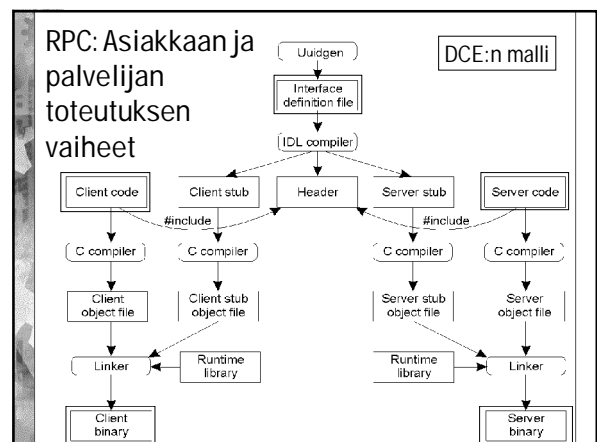
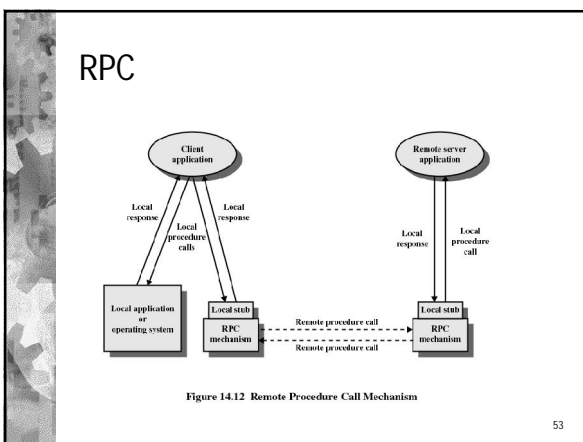
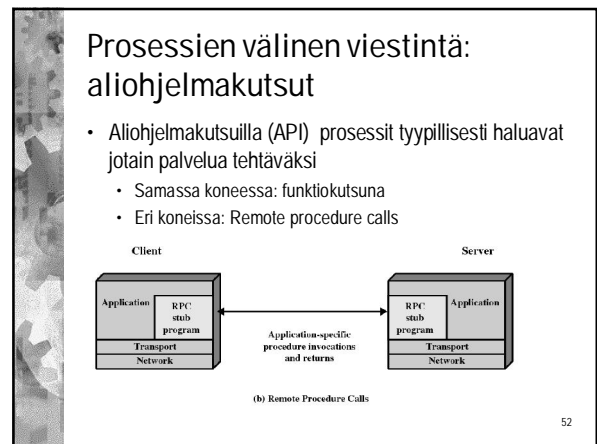
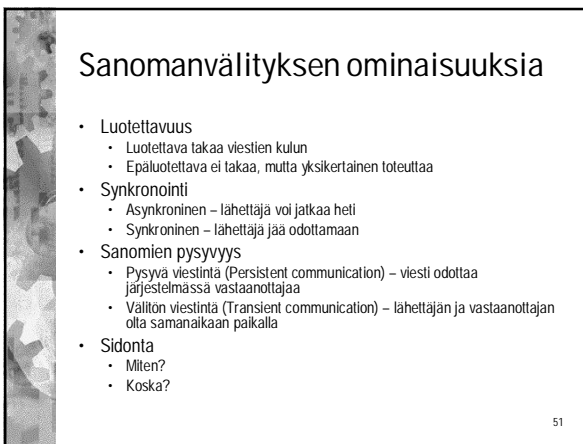
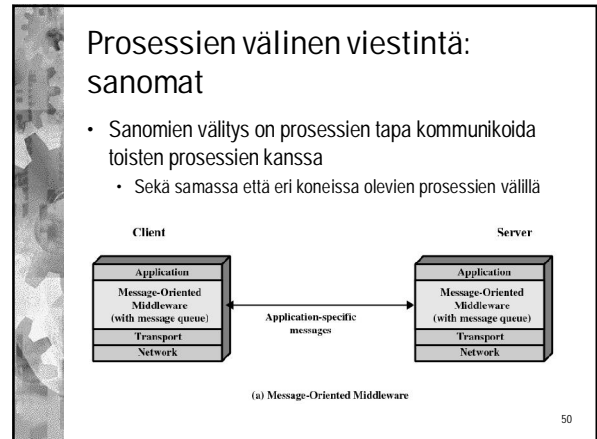
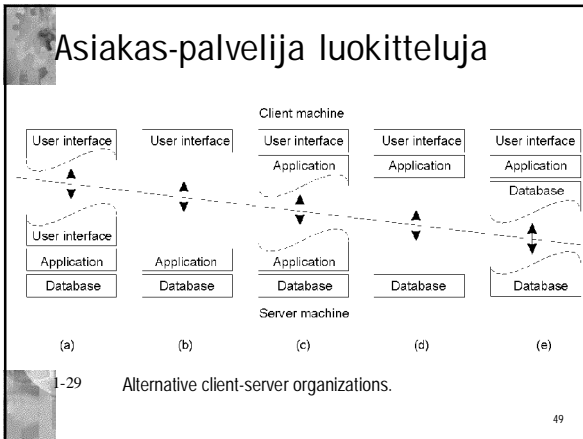


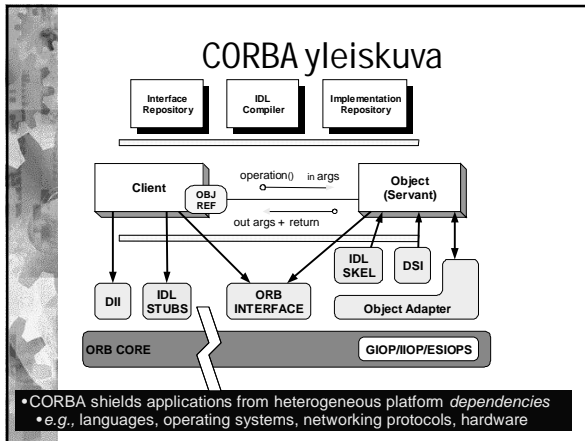
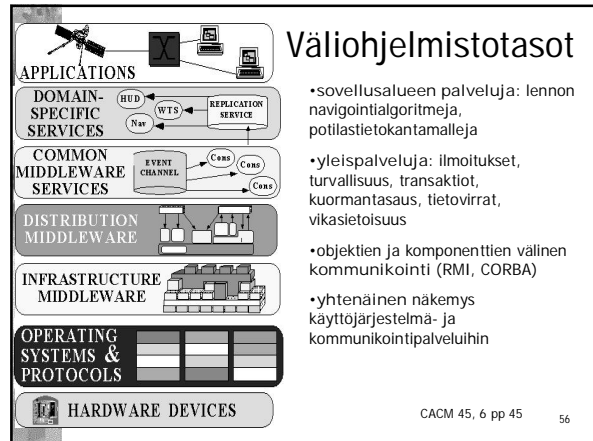
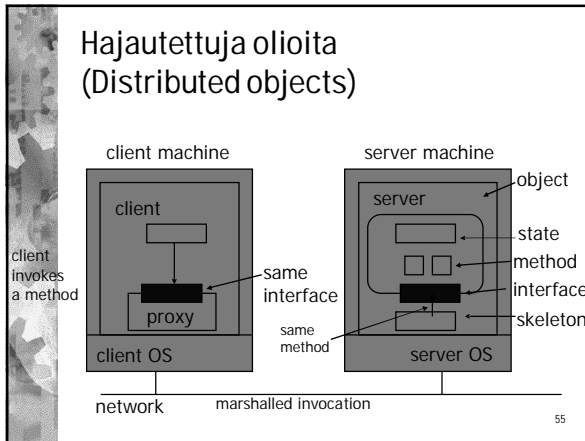
- Pienissä hakemistoissa MFT-tietueet peräkkäisjärjestyksessä
- Isoissa hakemistoissa MFT-tietueessa B-puun (B-tree) indeksirakenne
 - nimen etsintä ei ole peräkkäishakua

47

Hajautettu prosessi

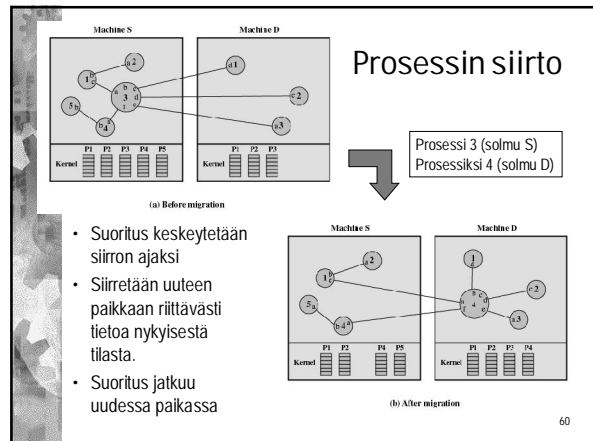
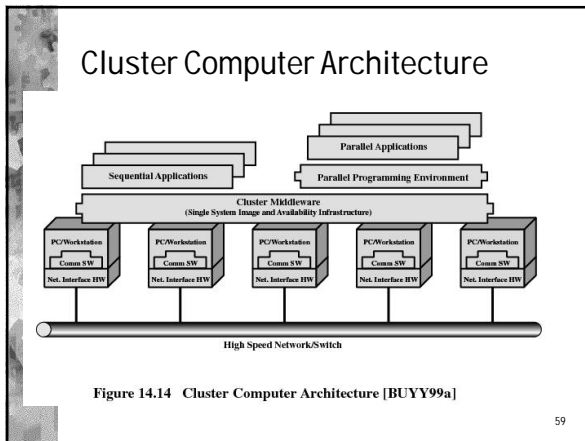
48





Klusterit: luokittelua

Clustering Method	Description	Benefits	Limitations
Passive Standby	A secondary server takes over in case of primary server failure.	Easy to implement.	High cost because the secondary server is unavailable for other processing tasks.
Active Secondary	The secondary server is also used for processing tasks.	Reduced cost because secondary servers can be used for processing.	Increased complexity.
Separate Servers	Separate servers have their own disks. Data is continuously copied from primary to secondary server.	High availability.	High network and server overhead due to copying operations.
Servers Connected to Disks	Servers are cabled to the same disks, but each server owns its disks. If one server fails, its disks are taken over by the other server.	Reduced network and server overhead due to elimination of copying operations.	Usually requires disk mirroring or RAID technology to compensate for risk of disk failure.
Servers Share Disks	Multiple servers simultaneously share access to disks.	Low network and server overhead. Reduced risk of downtime caused by disk failure.	Requires lock manager software. Usually used with disk mirroring or RAID technology.



Siirtopoliitikoja

- Eager (all):
 - Siirrä kaikki (muistialueet+muut)
- Precopy:
 - Prosessia vielä suoritetaan, kun muistialueita jo kopioidaan
- Eager (dirty):
 - Siirrä vain se osa muistiavaruutta, joka on keskusmuistissa ja jota on muutettu
- Copy-on-reference:
 - Siirrä sivua vain viittattaessa
- Flushing:
 - Kopioi prosessin muuttuneet sivut levyille

61

Tietoturva

62

Perinteinen, symmetrinen salaus

Sama avain molemmilla!

Plaintext input (selväkieli) → Encryption algorithm (e.g., DES) (salakirjoitus) → Transmitted ciphertext (salakirjoitetun tekstin siirto) → Decryption algorithm (reverse of encryption algorithm) (salakirjoituksen purku) → Plaintext output (selväkieli)

(Fig 16.14 [Stal 05])

63

Julkisen avaimen salakirjoitusmenetelmä

Salaus

Plaintext input (selväkieli) → Encryption algorithm (e.g., RSA) (salakirjoitus) → Transmitted ciphertext (salakirjoitetun tekstin siirto) → Decryption algorithm (reverse of encryption algorithm) (salakirjoituksen purku) → Plaintext output (selväkieli)

(a) Encryption (Fig 16.15 (a) [Stal 05])

64

Tunnistus julkisen avaimen menetelmällä

Plaintext input (selväkieli) → Encryption algorithm (e.g., RSA) (salakirjoitus) → Transmitted ciphertext (salakirjoitetun tekstin siirto) → Decryption algorithm (reverse of encryption algorithm) (salakirjoituksen purku) → Plaintext output (selväkieli)

(b) Authentication (Fig 16.15 (b) [Stal 05])

65

Passiiviset hyökkäykset (kuuntelu, nuuskinta)

Fig 16.3 [Stal 05]

- Luottamuksellisuus rikkoontuu, eheys ei rikkoudu
- Tietoliikenneyhteydät, -verkko
 - salakuuntelu, tarkkailu, vuotaminen julkisuuteen (release of contents)
 - puhelut, sähköposti, tiedostojensiirto
 - salaus, salakirjoitus
 - silti analysoitavissa (traffic analysis)

66

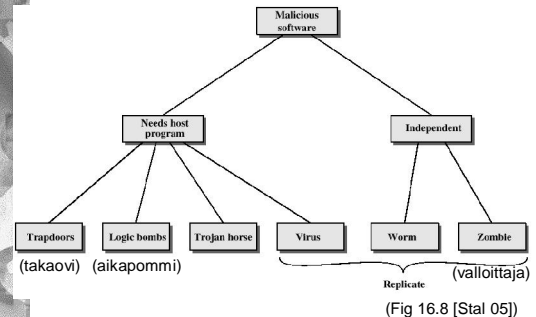
Aktiiviset hyökkäykset

- Tiedon eheys rikkoontuu
- Tietoliikennyhteidet, -verkko
 - lähettäjä teeskentelee olevansa joku muu (masquerade)
 - virheellinen toisto (replay)
 - viivyttäminen, muuttaminen, uudelleenjärjestely (modification of msg contents)
- käytön esto (DoS = denial of service)
 - ylikuormitus, yhteyksien sabotointi
 - yritetään havaita ja toipua nopeasti



67

Luokittelua



(Fig 16.8 [Stal 05])

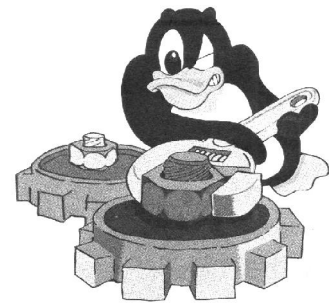
68

Aiemmin kokeessa kysyttyä

- Vuorotus
 - Yhden prosessorin: Multilevel feedback, Fair share, ...
 - Moniprosessori: kimpvuorotus
 - Reaaliaikajärj.: Rate monotonic
- I/O ja tiedostonhallinta
 - Puskurointi, lohkopuskurien allokointi
 - Levyhakuja: SCAN, FIFO, SSTF
 - Ext2fs, NTFS, tiedostojen suojaus
- Tietoturva
- Hajautuksesta:
 - Klusterit
 - RPC, CORBA

69

-- END --



Käyttöjärjestelmät

70