Conceptes Avançats de Sistemes Operatius

Facultat d'Informàtica de Barcelona
Dept. d'Arquitectura de Computadors

Curs 2013/14 Q1

Sistemes de fitxers

Departament d'Arquitectura de Computadors

Índex

- Gestió de quotes
- Journaling
- Sistemes de fitxers en xarxa

Índex

- Gestió de quotes
- Journaling
- Sistemes de fitxers en xarxa

Gestió de quotes

- Quota, què és?
  - Habilitat de limitar la quantitat de dades que un usuari (o grup d'usuaris) té en un sistema de fitxers (partició)
- Mecanisme
  - Independent del sistema de fitxers
- Requereix
  - Que el sistema de fitxers les suporti
  - Que el kernel les suporti

Gestió de quotes

- Preparació de la partició
  - Ha de ser muntada amb les opcions 'usrquota' i/o 'grpquota'
    - Es pot usar /etc/fstab
      - /dev/sda9 /home ext4 defaults,usrquota,grpquota 1 1
  - Comanda quotacheck per crear els fitxers de quota
    - quotacheck -vagum
      - verbose, all, group, user, no-remount
    - Crea
      - /aquota.user
      - /aquota.group

Gestió de quotes

- Activació i aturada de les quotes
  - /sbin/quotaon -avug (all, verbose, user, group)
    - Activa el mecanisme de quotes
  - /sbin/quotaoff per desactivar-lo
- Edició de quotes
  - edquota, obre un editor, estil crontab

Disk quotas for user xavim (uid 500):

Filesystem	blocks	soft	hard	inodes	soft	hard
/dev/loop0	32	16	32	2	0	0

- Examinar les quotes: quota -v

Disk quotas for user xavim (uid 500):

Filesystem	blocks	quota	limit	grace	files	quota	limit	grace
/dev/loop0	32*	16	32	6days	2	0	0	

Gestió de quotes

- "Grace period"
  - Temps durant el qual l'usuari pot arribar al límit "hard", només amb warnings per part del sistema
  - Si expira el "grace period", llavors el sistema de quotes ja no deixa passar del "soft" límit

Índex

- Gestió de quotes
- Journaling
- Sistemes de fitxers en xarxa

Journaling

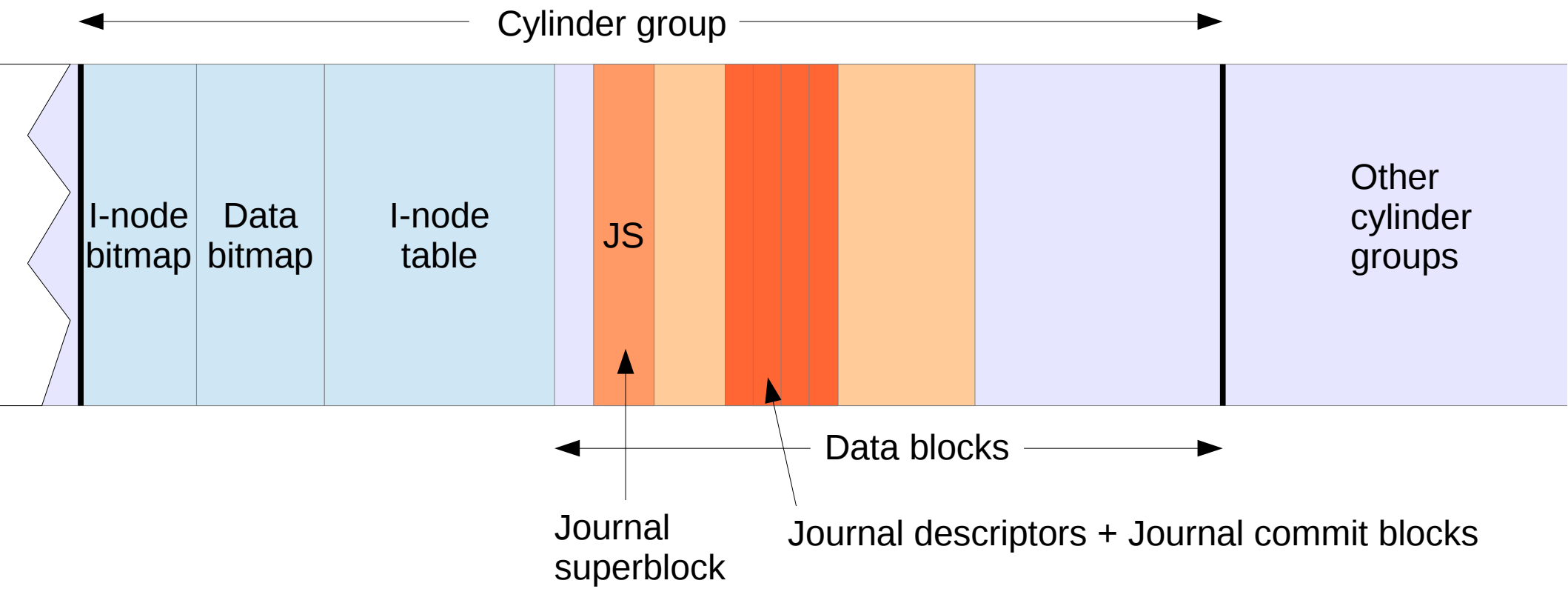
- Habitualment, les operacions sobre fitxers inclouen diverses operacions al disc
  - Exemple: esborrar un fitxer
    - Esborrar l'entrada del directori
    - Marcar l'i-node com a lliure a la taula d'i-nodes
    - Marcar els blocs de dades com a lliures a la taula de blocs
  - Si el sistema s'apaga en mig d'aquests passos...
    - Pot quedar un conjunt de blocs de dades ocupats i sense nom
    - L'entrada del directori pot quedar apuntant a un conjunt de blocs alliberats → poca seguretat!

Recuperació costosa

- Per arreglar aquests errors, cal recórrer completament
  - l'arbre de directoris i fitxers
  - les estructures d'i-nodes i mapes de blocs de dades
- Temps de test i recuperació
- Solució: journal

Journaling

- Inclou una estructura de dades de suport a la recuperació del sistema de fitxers



Analysis and Evolution of Journaling File Systems
Vijayan Prabhakaran, Andrea C. Arpaci-Dusseau, and Remzi H. Arpaci-Dusseau
http://www.cs.wisc.edu/wind/Publications/sba-usenix05.pdf

Journaling

- Ext4
  - Basat en ext3, sense afegir incompatibilitats
    - Limit ext3: 8TB punters a blocs de 32 bits
  - Suport per sistemes de fitxers grans
    - 48 bits d'adreça a bloc → 1 EB (2^{60} bytes)
    - Extents: milloren el tractament de blocs contigus en el fitxer i en disc
  - Tamany de block gran: 4KB – 1MB
    - **Utilitats, sistema**
  - Temps en alta resolució: nanosegons
  - Incorpora suport per quotes en el propi sistema

https://ext4.wiki.kernel.org/index.php/Ext4_Design

Journaling

- Garanteix consistència del sistema de fitxers
  - Fallades de corrent elèctric
  - Fallades del sistema
- Pot guardar-se en un disc diferent
  - Per minimitzar contenció al disc
    - Lectures i escriptures de dades i journal al mateix temps
- Les escriptures al journal es fan **asíncrones** per reduir l'impacte en el rendiment

Gestió del journal

- Al journal cal escriure per avançat, respecte a la resta del disc
  - Introduir dependències entre operacions
- Els canvis escrits en el journal són atòmics:
  - En recuperació, mai es repetirà una seqüència d'operacions que no estigui sencera en el journal
  - Cada seqüència inclou una suma de comprovació
    - Si la suma és incorrecte, no es reproduirà durant la recuperació

Tipus de journals

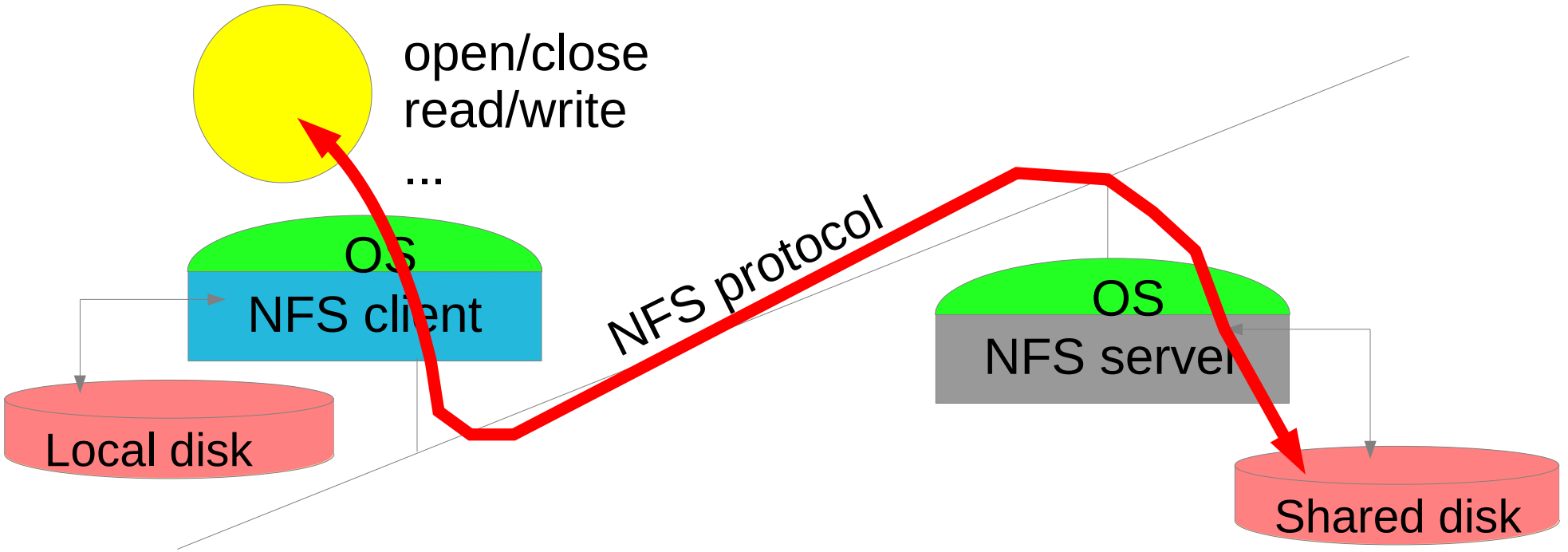
- Físic
  - Grava una còpia de cada bloc!
- Lògic
  - Grava només els canvis a les metadades del sistema de fitxers
  - Poden tenir corrupció de dades
    - Però no de l'estructura del sistema de fitxers

Índex

- Gestió de quotes
- Journaling
- Sistemes de fitxers en xarxa

Sistemes de fitxers en xarxa

- Network File System (NFS)
  - Tranparent als usuaris
  - Implementat sobre Remote Proc. Calls (RPCs)
  - Centralitzat en un servidor



Sistemes de fitxers en xarxa

- AFS, Andrew File System
  - Distribuït en diferents servidors
    - Presenta una visió homogènia dels fitxers independent de la localització de l'usuari
  - OpenAFS – Linux, MAC, Windows

Object Exchange (OBEX)

- Protocol d'intercanvi de dades amb dispositius mòbils
  - fitxers de tot tipus
  - Entrades de calendari
  - Tarjetes de visita

http://openobex.sourceforge.net/about.html

Object Exchange (OBEX)

- Xarxes
  - USB
  - Infrared (IrDA, IrLAN)
  - Bluetooth
  - Serial ports / ttys
- Sistemes de fitxers
  - FUSE – Filesystem in User Space

http://fuse.sourceforge.net/

FUSE

- Configuració i mòdul en el kernel
  - CONFIG_FUSE_FS=m fuse.ko
- Llibreries
  - /lib64/libfuse.so
- Comanda
  - /bin/fusermount – setuid a root
- Interfície
  - fuse_main(argc, argv, &operations, NULL);

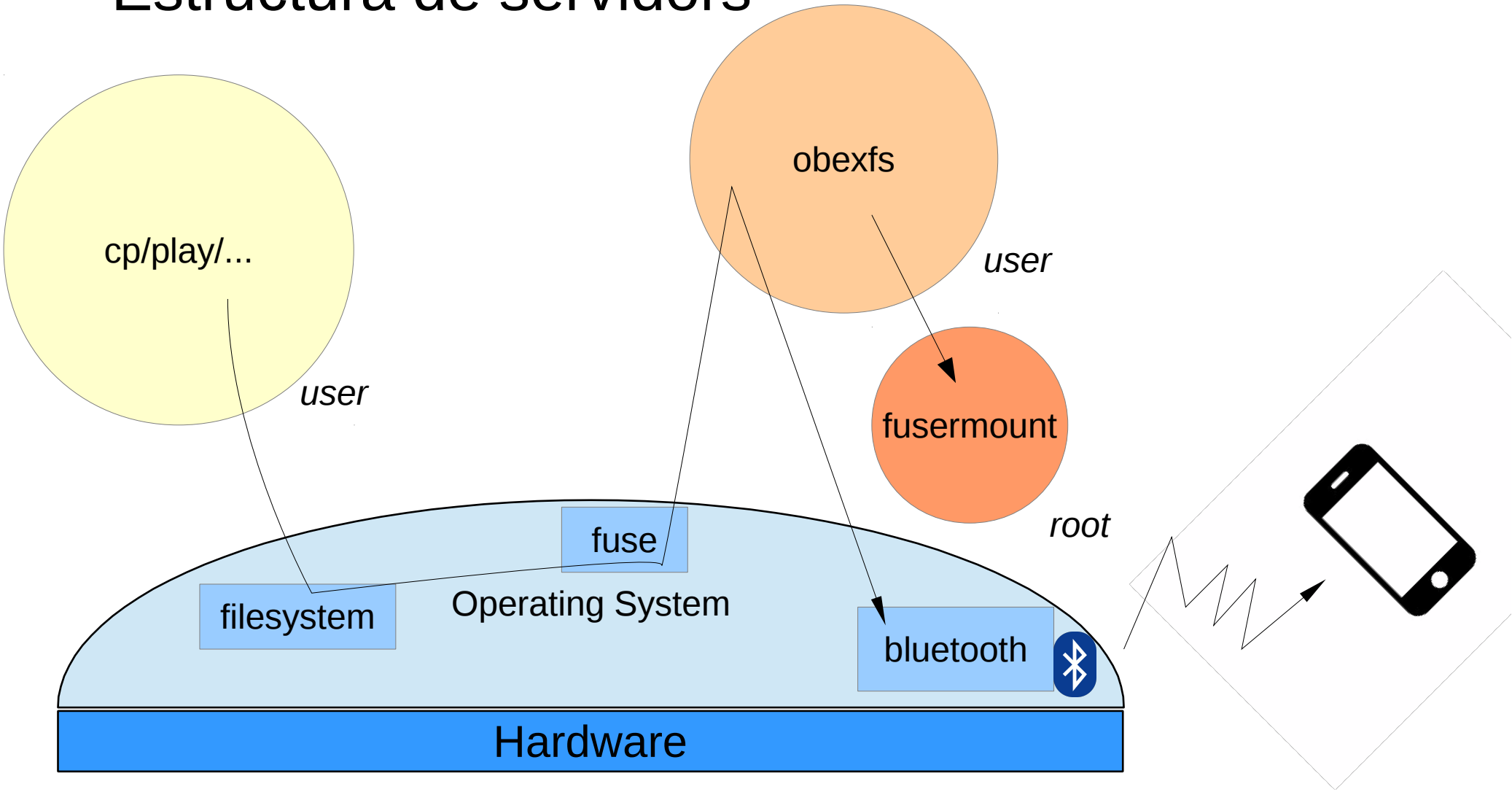
FUSE

- Interfície
  - open / create / read / write / fsync / flush / release
  - get file attributes / stat / statfs / access
  - symlink / readlink / link / rename
  - mknod / mkdir
  - unlink / rmdir
  - opendir / readdir / releasedir / fsyncdir
  - chmod / chown
  - truncate
  - +++

Exemple
https://github.com/zuckschwerdt/obexfs/blob/master/fuse/obexfs.c

OBEX + FUSE

- Estructura de servidors



OBEX + FUSE (exemples)

- Llistar el contingut d'un directori
  - obexftp -b DC:tt:zz:aa:xx:yy -c <path> -l
- Transferir fitxers al dispositiu
  - obexftp -b DC:tt:zz:aa:xx:yy -c <path> --put <file>
- Transferir fitxers des del dispositiu
  - obexftp -b DC:tt:zz:aa:xx:yy -c <path> --get <file>
- Muntar el dispositiu a <mountpoint>
  - obexfs -b DC:tt:zz:aa:xx:yy – – <mountpoint>

Activitat de Bluetooth

- tcpdump permet veure l'activitat del Bluetooth
  - Cal determinar el dispositiu
    - lsusb -v | grep -i bluetooth →
      - Bus 004 Device 004: ID 0a5c:2145 Broadcom Corp. Bluetooth …
    - tcpdump -D →

      1.eth0

      ...

      7.usbmon4 (USB bus number 4)

      …

      12.any (Pseudo-device that captures on all interfaces)

      ...
  - tcpdump -i 7 -w - | od -c

Activitat

- Documentar-se sobre "disk scheduling"
  - http://en.wikipedia.org/wiki/Category:Disk_scheduling_algorithms
- Entregueu:
  - Diferències entre els algorismes
    - FSCAN
    - N-Step-SCAN