

Praktikum angewandte Systemsoftwaretechnik (PASST)

Dateisysteme / Aufgabe 6

10. Januar 2019

Tobias Langer, Stefan Reif, Michael Eischer,
Bernhard Heinloth und Florian Schmaus

Lehrstuhl für Informatik 4
Friedrich-Alexander-Universität Erlangen-Nürnberg



Lehrstuhl für Verteilte Systeme
und Betriebssysteme



FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

TECHNISCHE FAKULTÄT

Dateisysteme

Was ist ein Dateisystem?

- Beispiele: ext{2,3,4}, ntfs, ...
 - Verwaltet persistenten Speicher
 - Einheitliche Schnittstelle
 - Implementierung von `open()`, `read()`, `write()`, ...
- Pseudo-Dateisysteme: procfs, sysfs
 - Exportieren Kernel-Datenstrukturen
 - Zugriff über Dateischnittstelle

Was ist ein Dateisystem?

- Beispiele: ext{2,3,4}, ntfs, ...
 - Verwaltet persistenten Speicher
 - Einheitliche Schnittstelle
 - Implementierung von `open()`, `read()`, `write()`, ...
- Pseudo-Dateisysteme: `procfs`, `sysfs`
 - Exportieren Kernel-Datenstrukturen
 - Zugriff über Dateischnittstelle

Aufgabe 6

Implementierung eines Dateisystems im Linux-Kernel

- Dateien sind als Baum organisiert
 - „hierarchisches“ Dateisystem
- Wurzelverzeichnis: /
- Manche Einträge sind „Mountpoints“
 - Dahinter verbirgt sich eine weitere Dateisysteminstanz
 - Operationen können dort eine vollständig andere Semantik haben

Bestandteile eines einfachen Dateisystems

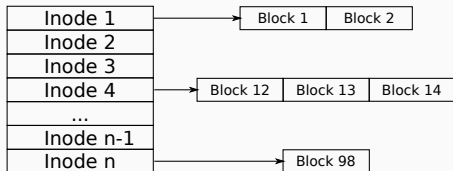
Prinzipieller Aufbau eines Unix-Dateisystems:

■ Inodes

- Metadaten der Dateien
- Verweise auf Dateiblöcke

■ Blöcke

- Nutzdaten
- Verzeichnisinhalte



Bestandteile eines einfachen Dateisystems

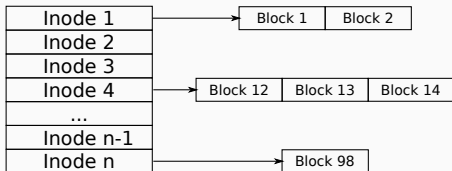
Prinzipieller Aufbau eines Unix-Dateisystems:

■ Inodes

- Metadaten der Dateien
- Verweise auf Dateiblöcke

■ Blöcke

- Nutzdaten
- Verzeichnisinhalte



Blockgröße \neq Blockgröße

- Blockgröße wird durch Dateisystem vorgegeben
- Entspricht nicht phys. Blockgröße des Datenträgers

Inode

Nummer	Type	Größe	Besitzer	Verweiszähler	...	Verweis auf Datenblöcke
--------	------	-------	----------	---------------	-----	-------------------------

- Eine Index node beschreibt Objekte im Dateisystem
 - Eindeutige ID
 - Typ & Größe der Datei
 - Besitzer & Zugriffsrechte
 - Verweis auf die Nutzdaten (wenn vorhanden)
 - Verweiszähler
- Verzeichnisse weisen Inodes Namen zu
 - Mehrere Namen für eine Inode möglich

Fallstudie: Ein einfaches Dateisystem (1/2)

Dateisystem

Superblock	Inode Freiliste	Block Freiliste	Inodes	Blöcke
------------	--------------------	--------------------	--------	--------

- Blockgröße: 512 Byte
 - Max. Inodes: 4096
 - Max. Blöcke: $16384 = 8 \text{ MiB}$
-
- Feste Zahl an Inodes & Blöcken
 - Block- & Inodefreilisten als Bitmaps
 - Superblock enthält statische Informationen
 - Dateien werden als Blockintervall gespeichert

Fallstudie: Ein einfaches Dateisystem (2/2)

Superblock

Magic Number	Blockgröße	Offset Inode Freiliste	Offset Block Freiliste	Offset Inodes	Offset Blöcke	Reserviert
--------------	------------	------------------------	------------------------	---------------	---------------	------------

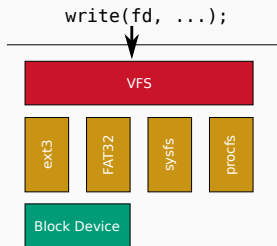
- Magic Number: 0x53462d5453534150
- Blockgröße (idR. 512 Byte)
- Absolute Offsets der Dateisystemsegmente (in Blöcken)

Frage

- Welche Vor- / Nachteile hat das Layout?

Das Linux Virtual Filesystem (VFS) (1/2)

- Abstraktion für Dateisysteme
 - Generische Schnittstelle
 - Implementiert `write()`, `open()`, ...



- Untergliedert entsprechend Dateisystemaufbau
 - Dateisystemtyp
 - Superblock
 - Inode
 - Datei

- „Objektorientierung“ im Kernel
 - Struktur für Objekt-Instanz
 - Zusätzliche Struktur für „Methoden“ (= Callbacks)
- VFS stellt generische Implementierung bereit
- Spezialisierung durch Treiber
 - Implementieren von Initialisierern (\approx Konstruktor)
 - Implementieren & Setzen von Callbacks (\approx Vererbung)

Der Dateisystemtyp im VFS

Dateisystem				
Superblock	Inode Freiliste	Block Freiliste	Inodes	Blöcke

Dateisystemtyp

- Repräsentiert durch `struct file_system_type`
- Erzeugt Dateisysteminstanzen

Dafür benötigt der Treiber ...

- Attribute des Dateisystems (Namen, Statusflags, ...)
- `fill_super`-Callback

Der Superblock im VFS

Dateisystem

Superblock	Inode Freiliste	Block Freiliste	Inodes	Blöcke
------------	-----------------	-----------------	--------	--------

Superblock - Instanz des Dateisystems

- Repräsentiert durch `struct super_block`
- Generische Kernelschnittstelle

Dafür benötigt der Treiber ...

- `struct super_operations`
- Implementierung der unterstützten Callbacks

Ergebnis im Userspace

- Mounten eines Dateisystems

Inodes im VFS

Dateisystem

Superblock	Inode Freiliste	Block Freiliste	Inodes	Blöcke
------------	-----------------	-----------------	--------	--------

Inode

- Repräsentiert durch `struct inode`
- Generische Kernelschnittstelle

Dafür benötigt der Treiber ...

- `struct inode_operations`
- Implementierung der unterstützten Callbacks

Ergebnis im Userspace

- `open`, `link`, `rename`, ...Syscalls funktionieren

Dateien im VFS

Dateisystem

Superblock	Inode Freiliste	Block Freiliste	Inodes	Blöcke
------------	-----------------	-----------------	--------	--------

Geöffnete Dateien

- Repräsentiert durch `struct file`

Dafür benötigt der Treiber ...

- `struct file_operation`
- Implementierung der unterstützten Callbacks

Ergebnis im Userspace

- Dateien können gelesen & geschrieben werden
- Verzeichnisinhalte können gelesen werden

- Dentries - Direcory entries
 - Repräsentiert einen Verzeichniseintrag
 - Verweis auf Inode und Eltern-Dentry
 - Werden für schnelleren Zugriff gecached
 - **Werden nicht auf der Festplatte gespeichert**

- Adressraum-Operationen
 - Dateioperationen interagieren indirekt mit Blockdevice
 - Stattdessen: Adressraum-Operationen
 - Implementierung des `get_block` Callbacks nötig
 - Mapping Dateisystemblöcke auf Blockdeviceblöck

Und jetzt...?

- Schrittweises Vorgehen:
 1. Mount
 2. Anzeigen von Ordnerinhalten
 3. Lesen von Dateien
 4. Schreiben von Dateien
- Dateisystemformat ist bei Weitem nicht ausgereift
 - Fragmentierung bei Verkleinerung / Löschen von Dateien
 - Kopieraufwand bei Vergrößerung von Dateien
 - ...

Aufgabe 6

- Einarbeiten in die benötigten APIs im Linux-Kern
 - Dokumentation, Codebeispiele
 - Empfohlenes Dateisystem zum Verständnis: bfs, minix
- Lesen & Verstehen der Spezifikation
- Programmieren des Dateisystems
- Testen des bereitgestellten Images

Abgabe: Bis 29. Januar 2019 durch Vorführung in einer Rechnerübung

Fragen?