

Aktuelle Entwicklungen beim Linux-Kernel

%pre

\$ whoami

technischer Redakteur

**schreibe vornehmlich
über Linux-Themen**

**nicht über spielkram wie
Steam, sondern eher die
harten Sachen**

**unter anderem eben
auch den Linux-Kernel**

Thorsten Leemhuis

Kernel-Log

Erste Vorabversion des Linux-Kernels 3.8

Linux 3.8 bringt 3D-Unterstützung für aktuelle GeForce-Chips und ein neues Dateisystem für Flash-Datenträger. Auf Intels 386er, für den Linux derzeit erfunden wurde, wird der im Februar erwartete Kernel allerdings nicht mehr laufen.

Der Kernel-Treiber Nouveau wird bei Linux 3.8 alles mitbringen, damit der zu Mesa 3D 9.0 gehörende Nouveau-OpenGL-Treiber ohne weitere Konfiguration die 3D-Beschleunigung sämtlicher bis dato vertriebener Grafikkarten der GeForce-Serie nutzen kann. So weit haben es die Nouveau-Entwickler, die die zur Treiberprogrammierung nötigen Informationen per Reverse Engineering gewinnen, bisher noch nie geschafft; daher wurden einige Fermi-GPUs und die seit zehn Monaten verkauften Grafikkarten der Kepler-Generation bislang nur eingeschränkt unterstützt. Für viele PCs wird Nvidias proprietärer Grafitreiber aber auch weiter die bessere Wahl bleiben, da Nouveau die schnelleren Betriebsmodi moderner Nvidia-Chips nicht aktivieren kann und daher nur dürftige 3D-Performance liefert; auch bei der Unterstützung für die Lüfterregelung und an anderen Stellen hapert es.

Linux 3.8 bringt erstmals ein einfaches Kernel-Grafitreiber für die Grafikkarte mit, die in Nvidias SoCs (System-on-Chip) der Tegra-Generation 2 und 3 steckt. Der Treiber stammt nicht von Nvidia, sondern von einem Embedded-Entwickler der deutschen Firma Avionic Design. Nvidia hat sich allerdings überraschend in die Entwicklung eingeschaltet und vor einigen Wochen Erweiterungen veröffentlicht, durch die der Treiber die Beschleunigungsfunktionen dieser Grafikkarte verfügbar macht; diese Verbesserungen sind bei 3.8 noch außen vor geblieben. Zur Nutzung der Beschleunigungsfunktionen sind noch Userland-Treiber nötig; derzeit deutet nichts darauf hin, dass Nvidia solche unter einer Open-Source-Lizenz freigeben will.

Bereits Ende Dezember hatte Linus Torvalds noch zahlreiche weitere Änderungen für die im

Februar erwartete Kernel-Version integriert. Der Grafitreiber 915 etwa erhielt Erweiterungen, durch die er die Grafikkarte in Haswell-Prozessoren nun ordentlich unterstützen soll; die Intel in einigen Monaten als Core i-4000 einführen will.

Dateisysteme

Linux 3.8 wird das erst im Oktober von Samsung-Entwicklern präsentierte Dateisystem F2fs (Flash-Friendly File System) unterstützen. Es ist für den Einsatz auf Flash-Datenträgern angelegt, die einen simpleren Flash Translation Layer (FTL) verwenden als SSDs für Desktop-PCs oder Server – dazu gehören USB-Sticks, Speicherkarten sowie die in Kameras, Tablets und Smartphones integrierten Speichermedien. F2fs ist ein Log-structured File System (LSFS) und füllt Datenträger von vorne nach hinten; erst wenn es dort angekommen ist, beginnt es wieder vorne und nutzt zwischenzeitlich freigeordnete Bereiche. Dabei fragmentieren die Daten stark, was bei Flash-Datenträgern aber kein Nachteil ist. Durch Design-Kniffe soll F2fs einige beim LFS-Ansatz bekannte Probleme vermeiden und besser mit simplen FTLs harmonisieren als andere LFS-Dateisysteme oder Ext4.

Ext4 wird bei 3.8 Inline Data Support beherrschen; so kann es wenige Byte kleine Dateien zusammen mit dem Inode-Eintrag speichern, was Speicherplatz spart und den Zugriff beschleunigt. Anhand von Checksummen soll XFS Metadaten-Verfälschungen durch Schreib- oder Lesefehler erkennen können. Das weiterhin experimentelle Btrfs enthält eine Replace-Funktion, die Daten schneller als zuvor von einem Laufwerk auf ein anderes transferiert – etwa beim Datenträgeraustausch im laufenden Betrieb.

Über den bei 3.8 neuen Hyper-V Balloon Driver kann Microsofts Hypervisor zur Laufzeit die Menge des Speichers reduzieren, die ein Linux-Gastsystem nutzt. Der vor allem mit KVM und Xen eingesetzte Netzwerktreiber virtio_net soll bessere Performance liefern, da er nun mehrere Queues einrichten kann; die beherrscht nun auch der Tun/Tap-Treiber, der auch, aber nicht ausschließlich bei der Systemvirtualisierung zur Emulation von Netzwerk-Hardware genutzt wird. Ebenfalls neu ist der Treiber rtl8723ae für den PCIe-WLAN-Chip Realtek RTL8723AE.

Ressourcenfragen

Linux 3.8 soll größere, in den letzten Monaten vorgenommene Umbauarbeiten zur Unterstützung für User Namespaces weitgehend abschließen. Durch einige der Änderungen sollen in

lässt, den der Kernel selbst zur Verwaltung von Prozessen nutzt. Bei passend konfigurierten Limits soll der Kernel dadurch eine mit der Shell ausgeführte „Fork-Bombe“ auch in Containern im Zaum halten können, in denen Anwender Root-Rechte haben.

Bei Kernen mit Unterstützung für Transparent Huge Pages kann Linux 3.8 eine „Huge Zero Page“ vorhalten, die als gemeinsamer, mit COW/Copy-on-Write genutzter Ausgangspunkt für neue Huge Pages (große Speicherseiten) dient. Das soll den Speicherverbrauch erheblich reduzieren, wenn diese Speicherbereiche nur angefordert, aber gar nicht genutzt werden.

Nach längen, teilweise heftigen Streitigkeiten zwischen verschiedenen Kernel-Entwicklern wurden einige unter dem Schlagwort „Balancing“ entwickelte Funktionen für Linux 3.8 aufgenommen. Im Idealfall kann der Kernel durch sie automatisch Prozessor und Arbeitsspeicher zusammenhalten, die ein Prozess verwendet. Solch eine Platzierung ist für optimale Performance auf den heute verbreiteten Multiprozessor-Systemen mit NUMA (Non-Uniform Memory Access) wichtig, denn ein Prozess erreicht dort am schnellsten den Arbeitsspeicher des Prozessors, auf dem er läuft (siehe S. 166). Bei Linux 3.9 und späteren Versionen sollen verbesserte Automatismen folgen, die auf den jetzt integrierten Grundlagen aufbauen.

Bye-bye, 386er

Das Kernel 3.8 beiliegende Diagnosewerkzeug Turbostat kann bei neuem Intel-Prozessoren auslesen, wie viel Watt die einzelnen Grafik- und Prozessorkerne und der Prozessor als Ganzes ungefähr verbrauchen. Um Wartung und Weiterentwicklung von Linux zu erleichtern, haben die Kernel-Entwickler die Unterstützung für Intels 386er und andere CPUs entfernt, die diese vor über 20 Jahren verbreitete Prozessor-Architektur verwenden; der 486er und modernere x86-32-Prozessoren werden weiterhin unterstützt. Auf einem 386er hatte Torvalds damals die Linux-Entwicklung begonnen. Er weint dem Code aber keine Träne nach und schrieb beim Entfernen: „Ich bin nicht sentimental. Gut, dass wir das los sind.“ (th)





heise open > Kernel-Log > Kernel-Log – Was 3.8 bringt (1): Dateisysteme und Storage

Kernel-Log 21.01.2013 14:59

Thorsten Leemhuis

Kernel-Log – Was 3.8 bringt (1): Dateisysteme und Storage

Linux unterstützt jetzt das speziell auf Datenträger mit Flash-Chips zugeschnittene Dateisystem F2fs. Btrfs soll bestimmte Aufgaben flotter erledigen und Ext4 kann kleine Dateien nun effizienter handhaben.

Linus Torvalds hat am Freitag die vierte Vorabversion von Linux 3.8 veröffentlicht. Er rief dabei zum Testen auf und erwähnte erfreut, die Entwicklung komme seiner Ansicht nach zur Ruhe. Alle größeren Neuerungen für Linux 3.8 haben Torvalds und seine Mitstreiter wie üblich aufgenommen, als sie mit der Entwicklung dieser Version begonnen haben. Größere Änderungen im jetzigen Abschnitt der Stabilisierungsphase sind selten, daher kann das [Kernel-Log](#) bereits jetzt einen umfassenden Überblick über die wichtigsten Neuerungen der Mitte Februar erwarteten Linux-Version geben.



Artikelanfang

Die kleinen Perlen

Seite « 1 2 »

[English version](#)

Anzeige

Erfahren Sie, wie wir Trinity Health geholfen haben, das Betriebsmodell umzustellen.

► [Lesen Sie hier die ganze Story](#)

Meistgelesen

Meistkommentiert

CeBIT 2013: Linux und Open Source
Hands on: Ubuntu für Smartphones
[Dells Ubuntu-Ultrabook jetzt bestellbar](#)
Ubuntu für Tablets und Smartphones zum Ausprobieren
[Ubuntu Touch jetzt mit Daily Builds](#)

**Früher ein sehr aktiver
Fedora-Entwickler**

page

discussion

edit

history

move

unwatch

WIKI

- [Fedora Project](#)
- [Wiki](#)
- [News](#)
- [Events](#)
- [Features](#)
- [Recent changes](#)
- [Random page](#)
- [Help](#)

NAVIGATION

- [Home](#)
- [Get Fedora](#)
- [Join Fedora](#)

SUB-PROJECTS

- [Ambassadors](#)
- [Bug Zappers](#)
- [Design](#)
- [Documentation](#)
- [EPEL](#)
- [Infrastructure](#)
- [Internationalization](#)
- [Localization](#)
- [Marketing](#)
- [Package Maintainers](#)

Kernel Vanilla Repositories

Contents [\[show\]](#)

Linux vanilla kernels for Fedora [\[edit\]](#)

Overview [\[edit\]](#)

This page contains information about a [set of package repositories](#) which contain RPMs of Linux vanilla kernels built for Fedora. Vanilla in this scope means 'unmodified', hence these packages do not contain any of those enhancements the Fedora developers integrate into the Linux kernel packages that Fedora normally uses.

Currently there are these repositories:

- [kernel-vanilla-mainline](#) for F19 and F20 [\[edit\]](#)
- [kernel-vanilla-stable](#) for F19 and F20 [\[edit\]](#)
- [kernel-vanilla-stable-testing](#) for F19 and F20 [\[edit\]](#)

To install the kernels from these repositories download the repo definitions for [kernel-vanilla-mainline](#), [kernel-vanilla-stable](#) or [kernel-vanilla-stable](#) and put it in `/etc/yum.repos.d/`. You can do this with commands like these:

```
# for the mainline repo
curl -s http://repos.fedorapeople.org/repos/thl/kernel-vanilla-mainline/fedora-kernel-vanilla-mainline.repo

# for the stable repo
curl -s http://repos.fedorapeople.org/repos/thl/kernel-vanilla-stable/fedora-kernel-vanilla-stable.repo

# for the stable-testing repo
curl -s http://repos.fedorapeople.org/repos/thl/kernel-vanilla-stable-testing/fedora-kernel-vanilla-stable-testing.repo
curl -s http://repos.fedorapeople.org/repos/thl/kernel-vanilla-stable-testing/fedora-kernel-vanilla-stable-testing.repo
```

**Kein
Kernel-Entwickler**

**guten Überblick über
das große Ganze**

und genau der folgt!

**schnelle Themen-
sprünge, also notiert
euch Fragen!**

== Vortragsabschnitte ==

- 1. jüngste Verbesserungen**
- 2. wichtige Entwicklungen**
- 3. woran gearbeitet wird**

= 1. Abschnitt =

**Jüngst vorgenommene
Verbesserung**

Aktuelle Version und dessen Nachfolger

Linux 3.19 und ~~3.20~~ 4.0

**4.0 statt 3.20:
Linus wurde die zweite
Zahl zu groß**



Linus Torvalds

Shared publicly - Feb 13, 2015

So, I made noises some time ago about how I **don't** want another 2.6.39 where the numbers are big enough that you can't really distinguish them.

We're slowly getting up there again, with 3.20 being imminent, and I'm once more close to running out of fingers and toes.

I was making noises about just moving to 4.0 some time ago. But let's see what people think.

So - continue with v3.20, because bigger numbers are sexy, or just move to v4.0 and reset the numbers to something smaller?

30,328 votes

✓ I like big versions, and I cannot lie

44%

v4.0, 'cause I get confused easily

56%

**wie beim Sprung von
2.6.39 auf 3.0**

**wie damals: Änderung
im gewohnten Umfang**

- 3.19 -

**3.19 ist Anfang
Februar erschienen**

**Update bei Fedora 21,
Ubuntu 15.04, ...**

3.19: offloading für routing und switching

**IOW: irgendwann laufen
Debian, RHEL und SLES
vielleicht auf Switches**

**erleichtert Management
sowie NFV und SDN**

3.19: Treiber für HSA-HQ bei AMD

**IOW: Es wird einfacher,
GPUs für Berechnungen
zu nutzen**

auch in 3.19:

**Intel Memory Protection
Extensions (MPX)
Elantech-Treiber**

- 4.0 -

**4.0 erscheint vermutlich
Mitte April**

4.0: Kernel Live Patching (KLP)

**IOW: Sicherheitslücken
des Kernels ohne
Neustart korrigieren**

Wofür?

**monatelange
Simulationen:
Neustarts unmöglich**

**hochverfügbare Server:
Neustarts unbequem**

**stopft ~90% der
typischen Lücken**

**direkte Vorläufer
schaffen ~95%**

daher:

"Live Patching Basics"

**Live Patches
sind Kernel-Module**

**daher: keine neue
Sicherheitsgefahr**

**Tools erzeugen
Patch-Module aus
Quellcode-Patches**

**Quellcode-Patch darf
keine Datenstrukturen
verändern**

**Review durch
Programmierer
erforderlich!**

**Wie Hotplug für PCIe:
cool, aber zu Hause hat
man es nicht**

**Seife erläutert Details
um 17:15 in diesem
Raum**

4.0: nfsd beherrscht Parallel NFS (pNFS)

**IOW: NFS-Direktzugriff
auf Nutzdaten inkl.
RAID-0-artiger Funktion**

noch mehr in 4.0:

Ext4 für NVDIMMs

lazytime

**Congestion-Control-
Algorithms pro Route**

= 2. Abschnitt =

**Wichtige Entwicklungen
der letzten Monate**

- btrfs -

**"Next generation Linux
file system"**

**Mittlerweile über
7 Jahre alt...**

**Features machen es
Interessant**

COW

Volume Management inklusive RAID

Subvolumes

Snapshots

Checksummen & Scrubbing

Kompression

Send/Receive

Deduplication



**Kernfunktionen laut
Hauptentwickler stabil**

**RAID 5&6 nach wie
vor experimentell**

**RAID 5&6: scrubbing
und device replace erst
seit 3.19**

Standard-Root-FS bei openSUSE 13.2 & SLE12

**Hauptmotivation:
Snapshots mit Rollback**

**Facebook und Google
setzen es ein**

**IOW: btrfs wird langsam
alltagstauglich**

reichlich Stolpersteine

balance

**Reparatur-Werkzeuge
schwer zu
druckschauen**

**Performance dürftig bei
Datenbanken und VMs
(COW bedingt)**

**Einsatz gut
durchdenken!**



January 14, 2015

Why you should consider using btrfs, real COW snapshots and file level incremental server OS upgrades like Google does.

Marc MERLIN



NEXT DAY VIDEO

Tim Videos

0:02 / 46:51

Why you should consider using btrfs ... like Google does.



Linux.conf.au 2015 -- Auckland, New Zealand

Subscribe 1,067

5,230

- Ext4 & XFS -

**für manche Einsatz-
Zwecke einfach
besser als Btrfs**

**immer wieder
weiterentwickelt**

**XFS default bei RHEL7
(überall) und SLE12
(Daten-Partitionen)**

- Overlayfs -

**Linux kann seit 3.18
Dateisysteme
übereinander schichten**

ab 4.0 sogar mehrere

**alles geschriebene geht
in die oberste Schicht**

**wichtig für Live-Medien,
Docker, Embedded, ...**

hat 15+ Jahre gedauert

aufs und Co.

**keine echte Option – zu
komplex, zu viele Bugs**

- Grafiktreiber -

**Intel: Treiber nicht
perfekt, aber richtig gut**

**AMD: OSS-Treiber
hat ordentlich aufgeholt**

**AMDs Kernel-Treiber
bald Unterbau für
proprietären 3D-Treiber**

**Nvidia: Nouveau lässt
nach wie vor einiges zu
wünschen übrig**

**größtes Problem:
Taktregelung**

**(und dadurch
Performance und Power
Management)**

**Optimus: automatisches
Ausschalten funktioniert**

**GPU verwendbar, aber
so eine Sache für sich**

**- Grafiktreiber,
allgemein -**

**4k-Displays und MST
noch ein Problem**

Wayland kommt jetzt

Fedora 22: Log-In- Manager nutzt Wayland

**X-Wayland sorgt für
Abwärtskompatibilität
(auch Remote!)**

- Netzwerk -

3.18: Transmission Queue Batching

**IOW: beim Versenden
gleich Daten mitnehmen,
die ohnehin bereit liegen**

**Sende-Performance
nahe am theoretischen
Maximum**

3.19:

SO_INCOMING_CPU

IOW: Lokalität waren

3.19: Ipvlan

IOW: Container verbinden

**ähnlich macvlan, aber
multiplexing auf
höherem Layer**

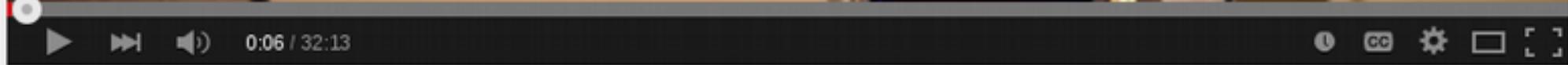
3.18: Data Center TCP (DCTCP)

**IOW: Datenfluss in
Rechenzentren
optimieren (nutzt ECN)**

3.18: foo-over-udp

**IOW: beliebige
Protokolle via
UDP tunneln**

**In the end:
Netzwerkstack-
Entwickler finden nach
wie vor enorm viel zum
Verbessern**



Netdev 0.1 - Keynote



Netdev 0.1

 **Subscribe** 39

500 views

**- tracing & performance
monitoring -**

**eBPF schafft
Grundlagen für Dtrace-
ähnliche Möglichkeiten**

**perf, ftrace und Co. sind
mächtig geworden**



Linux Performance Tools, Brendan Gregg, LinuxCon Europe 2014



DiaMonWorkgroup

Subscribe 57

3,818

+ Add to Share ... More

36 0

Linux Performance

This page links to various Linux performance material I've created, including the tools maps on the right. These show: [Linux observability tools](#), [Linux benchmarking tools](#), [Linux tuning tools](#), and [Linux sar](#). For more diagrams, see my slide decks below.

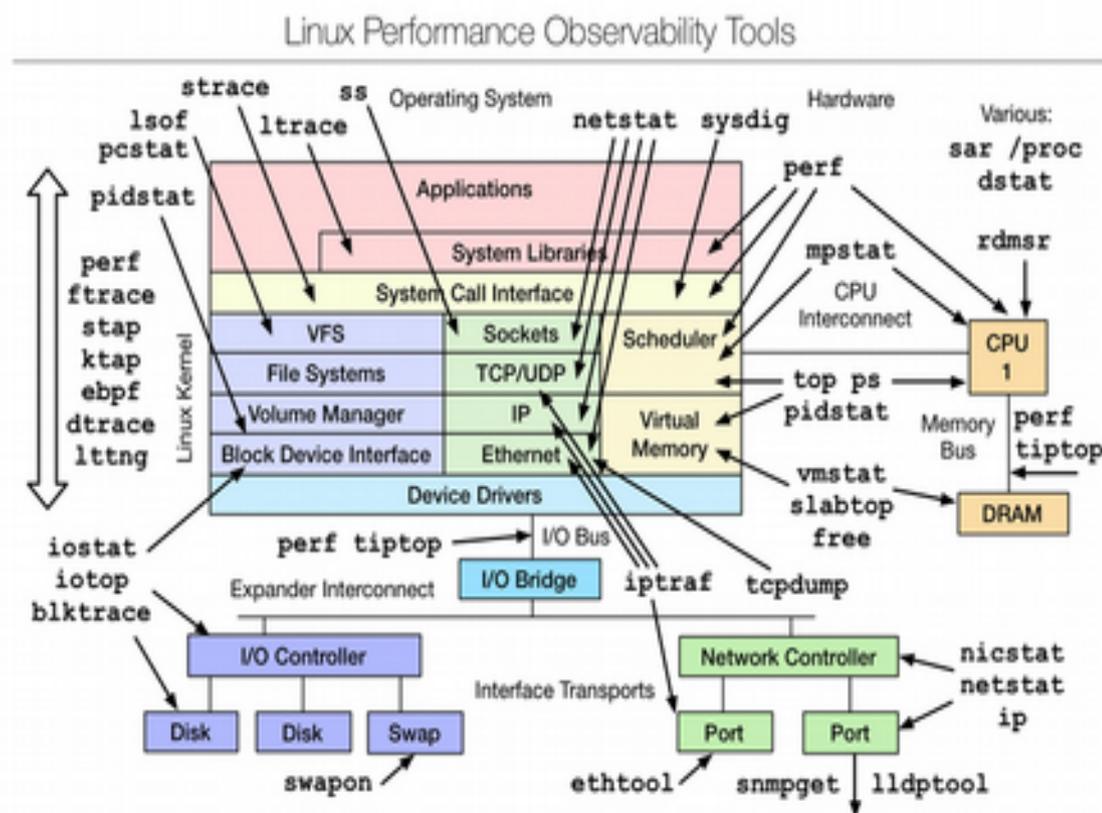
Tools

- [perf_events](#): perf one-liners, examples, visualizations.
- [perf-tools](#): perf analysis tools using perf_events and ftrace (github).
- [ktap](#): one-liners, examples, and scripts.
- [Flame Graphs](#): using [perf](#), [SystemTap](#), and [ktap](#).

Documentation

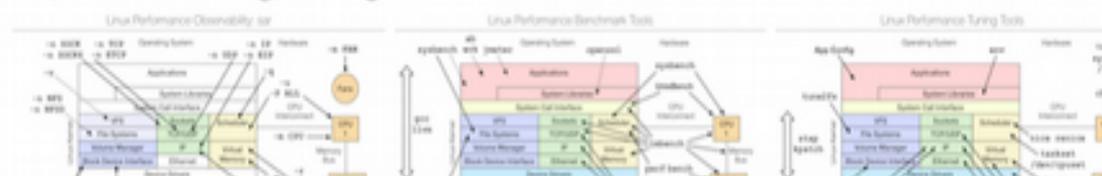
- My post [Performance Tuning Linux Instances on EC2](#) includes the tunables we're using at Netflix (2015).
- My [lwn.net](#) article [Ftrace: The Hidden Light Switch](#)

Observability:



<http://www.brendangregg.com/linuxperf.html> 2014

sar, benchmarking, tuning:



google "Brendan Gregg"

**Daniel Kobras, morgen
11:30: "Ohne
Tempolimit: I/O-
Performanceproblemen
in Linux-Systemen auf
der Spur"**

- was noch? -

Multiqueue block layer (blkmq)

**IOW: Durchsatz steigern,
Latenzen und CPU-Last
reduzieren**

3.19: Nvme-Treiber auf blkmg portiert

SCSI-Subsys.: optional
SCSI-Treiber: WIP

**Und wie immer: Bessere
Hardware-Unterstützung**

Meta-Ebene

**Kernel-Entwicklung
wird schneller**

**Neue Versionen
alle 2 Monate
(9 Wochen +/- 1 Woche)**

Linux-Version	Anzahl Dateien ¹	Zeilen Quelltext (Ohne Doku) ²	Entwicklungszeitraum	Anzahl Commits ³	Diffstat ⁴
<u>3.15</u>	46780	18632574 (17323446)	53 Tage	13722	11427 files changed, 932110 insertions(+), 571520 deletions(-)
<u>3.16</u>	47425	18879129 (17527507)	56 Tage	12804	9807 files changed, 513830 insertions(+), 267276 deletions(-)
<u>3.17</u>	47490	18864388 (17484200)	63 Tage	12354	10605 files changed, 651466 insertions(+), 666200 deletions(-)
<u>3.18</u>	47971	18994096 (17586160)	63 Tage	11379	9303 files changed, 485509 insertions(+), 355800 deletions(-)
3.19	48424	19130604 (17692109)	63 Tage	12617	10739 files changed, 483355 insertions(+), 346843 deletions(-)

**Typischerweise
~12.000 Änderungen**

**~80% der Änderungen
von bezahlten
Entwicklern**

Linux Kernel Development

*How Fast is it Going, Who is Doing It, What
Are They Doing and Who is Sponsoring the Work*

A Linux Foundation publication
February 2015

AUTHORS

Jonathan Corbet, LWN.net

Greg Kroah-Hartman, The Linux Foundation

Amanda McPherson, Linux Foundation

www.linuxfoundation.org

= 3. Abschnitt =

Woran gearbeitet wird

- kdbus -

IPC-Dienst im Kernel

vermutlich in 4.1

**designierter Nachfolger
des D-Bus-Daemon**

**also dem Transportweg
von D-Bus**

warum was neues?

**schneller – ermöglicht
neue Einsatzgebiete**

**bessere Abschirmung,
Container-tauglich**

**schon früh im Boot-
Prozess nutzbar**

**IOW: Performance,
Sicherheit, Early-Boot**

**wichtig u.a. für "Linux
Desktop Apps"**

Get on the Bus!

devconf.cz 2015, Brno

February 2015

Get on the Bus!



5:25 / 46:58



Lennart Poettering - Get On The Bus!



RedHatCzech



Subscribe

795

2,026

What's new in systemd, 2015 Edition
What's new in the systemd world, and what's coming next
Lennart Poettering



FOSDEM 15 .org

Copyright © 2015 FOSDEM VZW
This work is licensed under the Creative Commons Attribution 2.0 Belgium Licence.

Video player controls: play/pause, next, volume, 1:33 / 56:06, full screen, and other icons.

FOSDEM 2015 - What's New in Systemd - Lennart Poettering



Isdowdle

Subscribe 28

1,126

+ Add to Share ... More

1 1

- Festplatten mit SMR -

**SMR = Shingled
Magnetic Recording
(SMR)**

**Speicherbereiche
überlappen sich wie bei
Dachschindeln**

**kleinen Änderung ->
riesigen Bereich neu
schreiben**

**Spezialbehandlung für
optimale Performance
oder Einsatz wichtig**

Lösungen in der Diskussion

- Schnelle VM-Grafik -

**virtio-gpu KMS driver:
effizienter mit Qemu
interagieren**

**Virgil 3D: 3D-Befehle aus
VM nach oben reichen**

**XenGT/KvmGT:
Teilfunktionen der GPU
in VM hinein reichen**

- was noch? -

Y2K38

Kernel Tinification

<http://tiny.wiki.kernel.org>

Multipath TCP

LLVMLinux

RT

SSHDS

Persistent Memory

Atomic I/O operations

%post

**Präsentationsfolien,
Artikel und Videos zu
allem Angesprochenem
im Netz**

**zwei Sachen muss
ich noch loswerden**

(1) hilft beim Testen!

**insbesondere bei
Hardware jenseits des
Mainstreams**

**denn wenn ihr die nicht
testet, tut es keiner**

**Fehler bleiben lange
unentdeckt...**

**...uns dann umso
schwerer zu beheben**

(2) feedback bitte!

sprecht mich an

sagt mir, was ihr denkt

**so, jetzt war
es das aber**

**mit etwas Glück ist jetzt
noch Zeit für Fragen**

**denn mit Slide 195
endet der Vortrag jetzt**

Thorsten Leemhuis
<linux@leemhuis.info>

**[http://plus.google.com/
+ThorstenLeemhuis](http://plus.google.com/+ThorstenLeemhuis)**

#EOF