Kernel-Log

Linux 4.11 verlängert Akkulaufzeit

Die neue Kernel-Version beherrscht moderne Stromspartechniken besser. Neu dabei ist Support für selbstverschlüsselnde SSDs. Die Entwickler haben zudem Grundlagen geschaffen, um auf x86-64-Systemen bald bis zu 1 Petabyte Arbeitsspeicher ansprechen zu können.

Von Thorsten Leemhuis

Kurz nach Erscheinen dieser c't dürfte Linus Torvalds die Linux-Version 4.11 freigeben. Sie enthält gleich mehrere Änderungen, um die Stromspartechniken moderner Hardware besser zu nutzen und so die Akku-Laufzeit von Notebooks zu steigern. Das gilt insbesondere für Geräte mit NVMe-SSDs, denn der Kernel beherrscht nun APST (Autonomous Power State Transitions). Durch diese Stromspartechnik gehen NVMe-Datenträger eigenständig schlafen, wenn es gerade nichts zu tun gibt. Das spart oft zwischen 0,5 und 1,0 Watt, was die Leerlauf-Leistungsaufnahme sparsamer Notebooks um 20 Prozent oder mehr reduziert und so die Akkulaufzeit um zwei oder mehr Stunden verlängern kann.

Die Laufzeit mancher Notebooks dürfte auch steigen, weil der i915-Treiber nun auch bei den neuesten Intel-GPUs automatisch Framebuffer Compression (FBC) aktiviert. Theoretisch ist bei Intel-Notebooks aber noch mehr drin, denn nach wie vor lässt der Kernel einige Stromsparfunktionen moderner Intel-GPUs standardmäßig links liegen, weil sie auf einzelnen oder vielen Geräten zu Schwierigkeiten führen.

Manche per PCI Express (PCIe) angebundene Bausteine moderner PCs können jetzt tiefer schlafen gehen, wenn sie gerade untätig sind. Das ist dem Support für „ASPM L1 Substates“ zu verdanken. Diese Stromspartechnik wurde bei PCIe 3.1 spezifiziert und wird von einigen Chips bereits implementiert.

Genug RAM für jeden

Linux 4.11 legt das Fundament für Änderungen, durch die 4.12 auf x86-64-Systemen bis zu 4 Petabyte (4096 Terabyte) ansprechen können soll – „genug für jeden“, scherzt der Entwickler in Anspielung auf das vermeintliche Bill-Gates-Zitat „640 kB sollten eigentlich genug für jeden sein“. Bislang ist auf x86-64-Systemen bei 64 Terabyte Arbeitsspeicher Schluss; eine Grenze, an die erste Hardware-Hersteller gerade stoßen.

Einige Umbauten am Speichermanagement-Code versprechen, die Swap-Performance leistungsstarker Systeme zu verbessern. Die Änderungen entstanden vornehmlich für Mehrprozessor-Systeme mit besonders schnellen Datenträgern, wie sie häufiger in leistungsfähigen Cloud-Servern zum Einsatz kommen. Durch schnelle PCIe-SSDs oder persistente Memory/NVDIMMs ist der Performance-Einbruch durch Swapping hier längst nicht so krass, wie man es von PCs kennt, die Arbeitsspeicher auf Festplatten auslagern.

Selbstverschlüsselnde SSDs

Über Linux 4.11 lässt sich die Verschlüsselungsfunktion von SSDs nutzen, welche die Opal Storage Specification der Trusted Computing Group (TCG) implementieren. Bei solchen Self-Encrypting Drives (SED) verschlüsselt die SSD die Daten selbst; Microsoft nennt so konfigurierte Datenträger „Encrypted Hard Drive“ (eDrive). Der Ansatz ist für PCs interessant, die verschiedene Betriebssysteme nutzen; außerdem kann er Lebensdauer und Performance steigern, weil Mechanismen wie Wear Leveling und Trim besser greifen. Um Speicherbänder zu entsperren und Opal-SEDs anderweitig zu administrieren, sind aber noch Userspace-Werkzeuge nötig. Es ist ungewiss, wann Linux-Distributionen diese Programme mitbringen und idealerweise so integrieren, dass Anwender sich nicht mit Interna auseinandersetzen müssen.

Beim Multi-Queue Block IO Queueing Mechanism (Blk-Mq) können sich nun I/O-Scheduler einklinken. Solche können die gerade anstehenden Lese- und Schreiboperationen umsortieren, um die Performance zu verbessern – beispielsweise indem sie die zeitraubenden Wege reduzieren, die Schreib-/Leseköpfe von Festplatten nehmen müssen. Bei 4.12 soll eine auf die neue Scheduler-Infrastruktur angepasste Variante des Budget Fair Queueing (BFQ) Storage-I/O Schedulers folgen, der in Tuning-Kreisen schon länger als leistungssteigernd empfohlen wird.

An Programmierer richtet sich der neue System-Funktionsaufruf `statx()`, der eine umfassendere und effizientere Abfrage von Datei- oder Verzeichniseigenschaften ermöglicht. Der neue Syscall versucht `stat()` und `fstat()` zu beerben, über die Anwendungen bislang Informationen zu Dateisystemeinträgen abfragen – etwa Größe, Berechtigungen oder erweiterte Attribute. Der neue Funktionsaufruf ist flexibler, mächtiger und schneller, denn Programme können bei der Abfrage gezielt festlegen, welche Informationen sie wollen; Dateisysteme brauchen dadurch keine Daten zusammenzusuchen und weiterzugeben, die das Programm ohnehin nicht interessieren.

Sicherheit für Snap-Pakete

Nachdem die Entwicklung am Kernel-Code von AppArmor 2014 und 2015 schon eingeschlafen zu sein schien, hat diese im letzten Jahr wieder Fahrt aufgenommen. Das Tempo hat jetzt noch mal zugelegt, denn bei 4.11 gab es über sechzig Änderungen an der Sicherheitslösung. Sie unterstützt nun Policy Namespaces und einige andere Funktionen, mit denen die Paketmanagement-Lösung „Snap“ die mit ihr ausgeführte Software abschirmt. Weitere dafür nötige Änderungen sollen in die beiden nächsten Kernel-Versionen einfließen. Der Schutz, der zu einer der Haupteigenschaften des mit Flatpak kon-

kurrierenden Paketformats zählt, dürfte damit mittelfristig auch bei anderen Distributionen verwendbar sein; bislang bekommt man ihn nur bei Ubuntu.

Linux 4.11 zeigt in den von `dmesg` ausgegebenen Log-Meldungen jetzt an, ob UEFI Secure Boot auf dem jeweiligen System aktiv ist. Das ist einigen Patches zu verdanken, die in ähnlicher Form schon länger in den Kerneln großer Linux-Distributionen stecken. Dort finden sich auch Anpassungen, die zu einigen Einschränkungen im Betrieb führen und beispielsweise das Laden unsignierter Kernel-Module blockieren. Es gibt Bestrebungen, auch diese Patches in den offiziellen Linux-Kernel zu integrieren; noch ist aber ungewiss, ob das passieren wird.

Über /sys/kernel/security/lsm lässt sich nun auslesen, ob AppArmor, Capabilities, Yama, SELinux oder andere über Linux Security Modules (LSM) andockende Schutztechniken aktiv sind. Wie zuletzt üblich gab es viel mehr Detailänderungen zum Verbessern der Betriebssicherheit, als es vor zwei und mehr Jahren die Regel war. Neu dabei ist beispielsweise das GCC-Plug-in „structleak“, das für eine explizite Initialisierung von Datenstrukturen sorgt, die der Kernel mit dem Userspace teilt; das soll Informationslecks vermeiden, die Angreifern bei der Übernahme eines Systems dienlich sein könnten.

Netzwerk-Tuning

Die bei Linux 3.5 integrierte Unterstützung für das in RFC5827 definierte TCP Early Retransmit haben die Entwickler wieder entfernt: Dank Verbesserungen am bei Linux 4.4 eingeführten Package-Loss-Algorithmus RACK (Recently ACK) sei die Funktion jetzt unnötig. Beide Techniken versprechen einen Geschwindigkeitszuwachs bei TCP-Verbindungen, die häufiger Netzwerkpakete verlieren.

TCP Fast Open (TFO), mit dem Programme seit Linux 3.13 den Aufbau von HTTP-Verbindungen beschleunigen können, lässt sich jetzt über ein weiteres API nutzen. Das soll Vorteile für Anwendungen bieten, die gleich nach dem Verbindungsaufbau Daten per `write()` senden.

Das CIFS-Dateisystem beherrscht nun Freigaben-spezifische Verschlüsselung (per-share encryption), die SMB3 ermöglicht, das aktuelle Samba- und Windows-Versionen beherrschen.

Treibereien

Zur mit 4.11 erstmals unterstützten Hardware gehören die HD-Audio-Codecs ALC299 und ALC1220 von Realtek; Letzterer steckt auf einigen der Mainboards für die aktuellen Prozessorserien von AMD (Ryzen) und Intel (Kaby Lake/Core-i 7xxx). Ferner haben die Entwickler die Unterstützung für TPM 2.0 verbessert. Multi-Queue Support im HyperV-Storage-Treiber verspricht bessere I/O-Performance, wenn Linux unter einem modernen Hypervisor von Microsoft läuft.

Der für die verschiedenen Raspberry-Pi-Modelle zuständige Grafiktreiber VC4 kann nun auch Bildschirm-Panels ansteuern, die via Display Serial Interface (DSI) angebunden werden. Über einen neuen Sound-Treiber funktioniert nun auch HDMI Audio bei Raspis. Er hat aber eine Reihe von Qualitätsmängeln und liegt daher im Staging-Bereich. Das gilt auch für einen neuen Treiber, der das Kamera-Interface von Raspis unterstützt.

GPU-Virtualisierung

Bei den Treibern für AMD-GPUs gab es einige Detailänderungen, die die 3D-Leistung steigern und die Unterstützung von Stromsparfunktionen und Mehrschirmbetrieb verbessern. Außerdem unterstützt der AMD-Treiber nun noch mehr Polaris12-Chips, die auf der Radeon Rx 550 stecken (siehe S. 24). Ferner enthält der Grafiktreiber Amdgpu nun eine rudimentäre und noch nicht alltagstaugliche Infrastruktur zur GPU-Virtualisierung. Damit kann er Teile des Grafikchips als virtuelle Geräte bereitstellen, die sich an virtuelle Maschinen (VMs) überstellen lassen. Die darin laufenden Betriebssysteme und Anwendungen können die GPU darüber ohne sonderlichen Geschwindigkeitsverlust nutzen, ohne den Host zu gefährden. AMD überstellt Teile der GPU allerdings mit Hilfe von SR-IOV (Single-root Input/Output Virtualization), das bislang nur AMDs professionelle Grafikkarten bieten; Mainstream-Grafikkarten sollen die Funktion aber auch bald erhalten.

Beim i915-Treiber für Intel-GPUs funktioniert das Weiterleiten von Audio via DisplayPort jetzt auch bei Bildschirmen, die der Grafikchip per Multi-Stream Transport (MST) anspricht. Die Grafik- und Audio-Treiber von Intel unterstützen schon jetzt Prozessoren der Geminilake-Platform, die zur Atom-Klasse gehörende CPU-Kerne enthalten und in einigen Monaten auf den Markt kommen sollen.

Treiberbremse

Der Nouveau-Treiber kann durch einige Umbauten nun die Firmware-Dateien handhaben, die Nvidia kürzlich für die GP1xx-GPUs von GeForce-1000-Karten veröffentlicht hat. Mit dieser Firmware und einigen für Linux 4.12 vorgesehenen Änderungen sollen Linux-Distributionen bald die 3D-Beschleunigung dieser Grafik-Chips nutzen können, ohne dass der Anwender manuell Treiber installieren muss.

Wie bei den GPUs der GeForce-900er-Serie ermöglicht Nvidias Firmware auch bei den aktuellen Grafikchips keinen Wechsel der Geschwindigkeitsstufe. Nouveau kann die schnellsten Betriebszustände moderner GeForce-Hardware daher ebenso wenig ansteuern wie die stromsparendsten; vielmehr laufen Grafikprozessor und Speicher mit einem mittleren Standardtakt, der beim Einschalten gesetzt wird. Damit hält sich Nvidia die Konkurrenz vom Leib: Die quelloffenen Treiber können so unmöglich das Performance-Niveau erreichen, das Nvidias proprietäre Grafiktreiber erzielen. *(thl@ct.de)* c't