Kernel-Log

Linux 4.12: Vega-Treiber und mehr Live Patching

Die neue Kernel-Version unterstützt AMDs neue High-End-Grafikchips – eine der wichtigsten Grafiktreiber-Funktion fehlt aber vorerst. Zwei neue I/O Scheduler versprechen die Reaktionsfreude von Desktops und Servern zu steigern. Die Schnellstraße durch den Netzwerkstack funktioniert jetzt universell.

Von Thorsten Leemhuis

Das Anfang Juli veröffentlichte Linux 4.12 ist zwar kein Rekordhalter, aber wie Linus Torvalds anmerkte, eine der Kernel-Versionen mit den meisten und umfassendsten Änderungen. Zu denen gehören Umbauten an der Infrastruktur zum Kernel Live Patching (KLP), durch die sich in Zukunft deutlich mehr Sicherheitslücken des Linux-Kernels im Betrieb beheben lassen sollen. Bislang kann man via KLP nur zirka 90 Prozent der typischen Sicherheitslücken stopfen, während die indirekten Vorläufer Kpatch und Kgraft zuvor schon 95 Prozent meisterten. Mit dem jetzt integrierten „Per-Task Consistency Model“ soll KLP nicht nur endlich aufholen, sondern überholen: Das Konsistenzmodell legt Grundlagen, um mit KLP letztlich alle Lücken zur Laufzeit beheben zu können. KLP kann jetzt nämlich zuverlässig sicherstellen, dass es keinen Codeabschnitt modifiziert, der gerade irgendwo ausgeführt wird; außerdem lassen sich damit auch Lücken stopfen, bei denen ein Live Patch im Speicher liegende Funktions- oder Daten-Semantiken modifizieren muss. Das Ganze klappt vorerst aber nur auf x86-64-Systemen.

Vega-Treiber

Die neue Kernel-Version bringt Basis-Support für AMDs Grafikprozessoren der „Vega“-Generation (siehe S. 29). Eine wichtige Funktion fehlt dem Amdgpu-Treiber allerdings noch: Er kann mit Vega-Karten bislang keine Monitore ansteuern, daher gelingt letztlich nur ein zum Rechnen mit Grafikkarten interessanter Headless-Betrieb. Patches, die dieses Manko beseitigen, sind Teil der schon länger als „DC“ (Display Core) entwickelten Patch-Sammlung. AMDs Entwickler bauen diese zuvor DAL (Display Abstraction Layer) genannten Umstrukturierung gerade um, damit sie die Qualitätsansprüche der Kernel-Entwickler erfüllen und in Linux einfließen können. Vielleicht dauert das nur ein paar Monate, vielleicht aber auch noch ein Jahr oder mehr.

Der Nouveau-Treiber beherrscht jetzt 3D-Beschleunigung bei Nvidias Pascal-Grafikchip der GeForce-1000er-Serie; außerdem unterstützt der Treiber jetzt auch GeForce-1050-Modelle. Zusammen mit dem zugehörigen OpenGL-Treiber der neuesten Mesa-Version erzielen Nvidias aktuelle Grafiklösungen so eine 3D-Performance, die für einfache Spiele und Desktop-Oberflächen wie Gnome oder KDEs Plasma typischerweise ausreicht. Nvidias proprietärer Grafiktreiber entlockt diesen GPUs aber deutlich mehr Leistung.

Reaktionsfreude

Der neue Storage-I/O Scheduler „Budget Fair Queueing“ (BFQ) soll manche PCs reaktionsschneller machen. Ob die Performance zulegt, hängt allerdings stark vom eingesetzten Datenträger und den Zugriffsmustern der Programme ab. Vorteile dürfte BFQ am ehesten bei klassischen Magnetfestplatten zeigen, nicht aber bei besonders schnellen SSDs.

BFQ ist auf Desktop-Systeme ausgerichtet und arbeitet mit zahlreichen Heuristiken. Der mit I/O-Budgets arbeitende Scheduler bevorzugt beispielsweise die I/O-Operationen von Programmen, mit denen der Anwender gerade zu interagieren scheint. Außerdem räumt BFQ Leseoperationen eine höhere Priorität ein; ebenso ist es bei Operationen von als Echtzeit markierten Prozessen. Mit diesen und weiteren Tricks hat sich BFQ in Tuning-Kreisen einen guten Ruf erarbeitet, obwohl man ihn bislang meist umständlich nachrüsten musste. Zum Einsatz von BFQ wird man auf vielen Systemen aber weiterhin Hand anlegen müssen, denn die in den Kernel integrierte BFQ-Variante funktioniert nur mit dem bei Linux 3.17 eingeführten Multi-Queue Block IO Queueing Mechanism (Blk-Mq). Den nutzt der typischerweise für SATA-Festplatten verwendete AHCI-Treiber allerdings nur, wenn man den Kernel mit dem Parameter `scsi_mod.use_blk_mq=y` startet oder beim Erstellen eines eigenen Kernels die Konfigurations-Option SCSI_MQ_DEFAULT setzt. Außerdem muss man BFQ über Dateien wie `/sys/block/sda/queue/scheduler` explizit für jeden einzelnen Datenträger aktivieren.

An Admins von High-End-Servern richtet sich der zweite neue I/O Scheduler: der maßgeblich von Facebook-Mitarbeitern entwickelte Kyber. Die haben ihn auf besonders schnellen Datenträger abgestimmt, die mit mehreren Warteschlangen arbeiten – beispielsweise per PCIe angebundene NVMe-SSDs. Mit Hilfe der verschiedenen Queues arbeitet Kyber darauf hin, Leseoperationen bevorzugt abzuarbeiten, weil Nutzer häufig auf deren Ergebnis warten.

Manipulationserkennung

Linux kann jetzt Manipulationen an verschlüsselten Volumes erkennen. Ein Baustein dafür ist das neue Device-Mapper-Target „dm-integrity“, das ein Block-Device emuliert, das zu jedem gespeicherten Sektor einige Metadaten für spätere Integritätsprüfungen speichert. In Kombination mit der ebenfalls neuen Cryptographic Data Integrity Protection gelingt Authenticated Encryption (AE); dadurch kann der Kernel erkennen und warnen, wenn Blöcke eines verschlüsselten Volumes ohne den passenden Schlüssel modifiziert wurden.

Einige Detailänderungen am MD-Subsystem versprechen die RAID-5-Recovery zu beschleunigen und die Perfor-

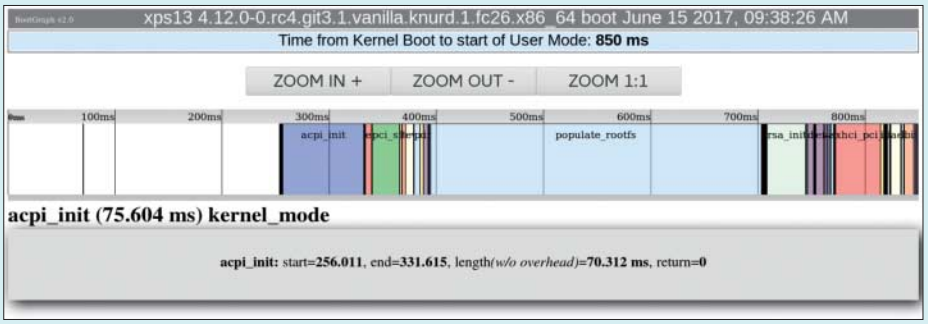
mance von Festplatten-Verbünden der Level 5 oder 6 zu steigern. Einige Anpassungen bei Btrfs beseitigen bekannte Probleme und Funktionslücken bei der Dateisystem-eigenen Implementierung von RAID 5 und 6; diese Funktion von Btrfs gilt aber nach wie vor als nicht praxisreif. Umbauten am Ext4-Dateisystem versprechen, die Performance großer Dateisysteme zu verbessern; auch Workloads mit verteilten Schreiboperationen (random write) sollen profitieren.

Schnellstraße für alle

Der bei Linux 4.8 eingeführte Express Data Path (XDP) lässt sich Dank „Generic XDP“ jetzt mit beliebigen Netzwerkschnittstellen nutzen. XDP ist grob gesagt ein Schnellverarbeitungsweg für bestimmte Netzwerk-Pakete, der etwa effizienteres Forwarding und eine bessere Abwehr von DDoS-Attacken ermöglicht. Damit XDP seine Vorteile ausspielt, ist aber nach wie vor XDP-Support im Treiber nötig: Die neue XDP-Variante dient unter anderem als Referenzimplementierung und soll Entwicklern zugleich ermöglichen, XDP-Programme mit beliebigen Netzwerkchips zu testen.

Anfahren

Das jetzt in den Kernel-Quellen enthaltene Diagnosewerkzeug AnalyzeBoot v2.0 erzeugt eine HTML-Datei mit Details zum Boot-Prozess. Sie zeigt mit einer interaktiven Balkengrafik, wie lange welches Kernel-Subsystem zur Initialisierung benötigt. Der Beobachtungszeitraum endet allerdings bereits beim Aufruf des Init-Prozesses. Das Tool richtet sich daher vornehmlich an Entwickler, die beim Start trödelnden Code finden und beseitigen wollen. Für Endanwender sind Werkzeuge wie Systemd-Bootchart oft die bessere Wahl, denn sie beobachten die gesamte Systeminitialisierung.



Zur Startzeitoptimierung liegt den Linux-Quellen jetzt das Werkzeug „AnalyzeBoot“ bei. Seine interaktive Balkengrafik zeigt, wie lange die verschiedenen Kernel-Subsysteme zur Initialisierung benötigen.

Durch neue und verbesserte Treiber unterstützt Linux 4.12 über 750 Geräte mehr als sein Vorgänger. Darunter sind allein rund hundert USB- oder PCI/PCIe-Devices. Der Treiber für die Grafikkerne des Raspberry Pi kann per HDMI nun auch Audio-Signale ausgeben. Neu dabei ist auch Support für einen weiteren der beim Raspi eingesetzten Controller für SD-Karten.

Zerhacken

Die Standard-Konfiguration schaltet Kernel Address Space Layout Randomization (KASLR) nun ein und auch der Konfigurations-Hilfetext rät zum Aktivieren. Linux kann durch die Sicherheitstechnik einige seiner Kernel-intern verwendeten Speicherbereiche seit einer Weile verstreut ablegen, was Angreifern einen Einbruch erschwert.

Die bei Linux 4.11 begonnen Umbauten zur Unterstützung von 64 TByte Arbeitsspeicher auf x86-64-Systemen wurden fortgesetzt, aber anders als erwartet noch nicht bei 4.12 abgeschlossen. Das soll jetzt mit 4.13 passieren. Diese Version dürfte Anfang September erscheinen, wenn Torvalds und seine Mitstreiter im gewohnten Tempo arbeiten.

(thl@ct.de) c't