

ADMIN

Netzwerk & Security



Auf DVD:

Das brandneue
Debian 6.0 „Squeeze“
für 32 und 64 Bit

STROM SPAREN

GREEN-IT IN DER TÄGLICHEN PRAXIS

VIRTUALISIERT

Gastsysteme durchleuchtet
Netzwerk-Switch virtuell

PowerTOP
Stromfresser
per Software
aufgespürt



SSD-SPEICHER

So funktionieren die Festplatten-Nachfolger

Samba

Fehlersuche und
Monitoring mit dem
Traffic Analyzer

Func

Parallele Remote-
Shell für Server



Knoppix

Klaus Knoppers
Rettungs- und
System-Tools



Git

Workshop zur
Versionskontrolle

eCryptfs

Verschlüsselung
userspezifisch



www.admin-magazin.de

D EUR 9,80
A EUR 10,80 - BeHeLux EUR 11,25
CH sfr 19,60 - E / J EUR 12,75



Plastik sparen



Liebe Leserinnen und Leser,

in der Nacht auf den 6. Februar um 3:06 war es endlich soweit: Debian-Entwickler und -Sprecher Alexander Reichle-Schmehl erwähnte noch in einem Nebensatz „We’ve release now“, bevor er sich selbst ins Bett begab. Seit diesem Sonntag kann sich die Linux-Welt an der neuesten Debian-Version erfreuen, rund zwei Jahre nach dem letzten Release haben die Debian-Entwickler nun Version 6.0 „Squeeze“ freigegeben.

Wie gewohnt setzen die Debian-Entwickler dabei auf Stabilität, von der nicht nur Endanwender, sondern auch die Entwickler anderer auf Debian basierender Distributionen wie Ubuntu profitieren. So kommt als Kernel die Version 2.6.32 zum Einsatz, den die Linux-Entwickler als lange unterstütztes Long Term Release pflegen. Besonderen Wert legen die Debian-Entwickler darauf, dass die freie Lizenzierung der enthaltenen Software beachtet wird, und haben deshalb einen Großteil „unfreier“ Firmware aus der Distribution entfernt.

Squeeze enthält über 10 000 neue und damit insgesamt beinahe 30 000 Software-Pakete. Über 15 000 Pakete wurden dabei aktualisiert. Die Skriptsprache Python ist in den Versionen 2.6.6 und 3.1 enthalten, Perl als 5.10.1, PHP mit 5.3.3 vertreten. Der GNU-Compiler GCC befindet sich auf dem Stand der Version 4.4.5.

Im Server-Bereich liefert Debian 6.0 die Datenbank MySQL auf dem Stand von 5.1.49 aus, bei Postgresql müssen Anwender noch auf die Features der 9.0-Reihe verzichten und sich mit 8.4.5 begnügen. Der Windows-kompatible Datei- und Druckserver Samba 3.5.5 ist ebenso dabei wie der Mailserver Postfix 2.7.1. Bei den Dateisystemen bietet Debian 6.0 nun Ext4 und Btrfs, Standard bleibt aber weiterhin Ext3. RAIDs und LVM-basierte Storage-Systeme lassen sich nun leichter einrichten, als Bootloader dient Grub 2. Die traditionellen Init-Skripts berücksichtigen beim Booten Abhängigkeiten zwischen den Paketen.

Debian 6.0 existiert für neun verschiedene Rechnerarchitekturen, zusätzlich gibt es eine Debian-Distribution mit FreeBSD-Kernel für zwei Architekturen. Dieser ADMIN-Ausgabe liegt eine Multiarch-DVD bei, auf der sich Debian Linux 6.0 für 32- und 64-Bit-Rechner befindet. Zum ersten Mal verwenden wir dafür die so genannte Ecodisc, die sich mit nur halb soviel Energie- und Materialaufwand wie herkömmliche DVDs produzieren und sogar recyceln lässt.

Viel Spaß damit wünscht

Oliver Frommel



Heft-DVD

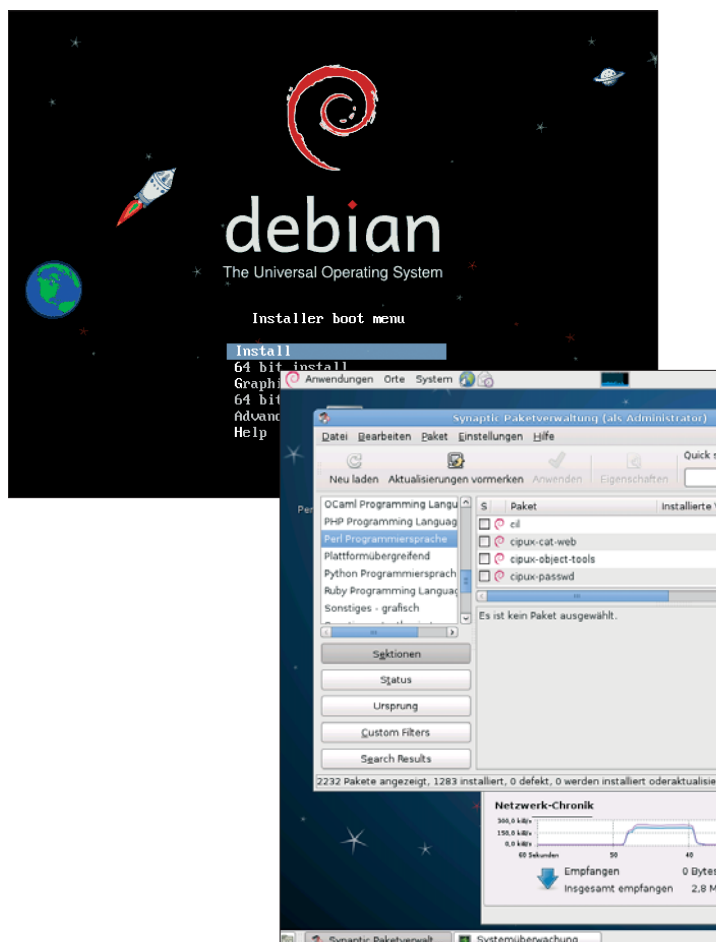
Auf der beiliegenden Heft-DVD finden Sie das aktuelle Debian 6.0 „Squeeze“, das im Februar 2011 nach zweijähriger Entwicklungszeit veröffentlicht wurde:

- **Neuestes Debian-Stable-Release 6.0 „Squeeze“.**
- **Zugriff auf mehr als 29 000 Software-Pakete über das Debian-Software-Repository.**
- **Multi-Arch-DVD für 32- wie auch 64-Bit-Rechner.**
- **Datenträger ist die neuartige Ecodisc, die bei der Produktion nur halb soviel Energie verbraucht wie eine konventionelle DVD.**

Legen Sie einfach nur die DVD ein und starten Sie den Rechner. Möglicherweise müssen Sie noch im BIOS die richtige Boot-Reihenfolge einstellen, damit das DVD-Laufwerk vor der Festplatte an die Reihe kommt.

Info

- [1] Debian-Projekt: [\[http://www.debian.org\]](http://www.debian.org)
- [2] Paket-Datenbank: [\[http://www.debian.org/distrib/packages\]](http://www.debian.org/distrib/packages)
- [3] Ecodisc: [\[http://www.ecodisc.org/\]](http://www.ecodisc.org/)



Bücher über Hochverfügbarkeit und Xen

Vorgelesen

Das Xen-Kochbuch präsentiert direkt anwendbare Kochrezepte zur Virtualisierung. Ein neues Werk zur Linux-Hochverfügbarkeit ist ähnlich praktisch ausgerichtet. Oliver Frommel, Konstantin Agouros

Praxislösungen vorzustellen, ist das erklärte Ziel, wenn man den Untertitel des Buchs „Linux Hochverfügbarkeit“ von Oliver Liebel wörtlich nehmen darf. Und so verschwendet er auch wenig Zeit mit historischen Exkursen und dem beliebten Rekapitulieren von OSI-Stacks und Ähnlichem, sondern kommt recht schnell zur Sache. Bevor er sich mit Hochverfügbarkeit im Cluster-Verbund beschäftigt, erklärt er recht ausführlich das Setup von RAID-Systemen unter Linux und migriert zum Einstieg Schritt für Schritt mit dem Leser zusammen ein konventionelles Setup auf ein RAID-System. Komplexeren Szenarien mit RAIDs und dem Logical Volume Manager (LVM) widmet sich der Rest dieses etwa ein Viertel des Buchs umfassenden Kapitels.

Im Mittelpunkt steht dann das, was sich die meisten Admins wohl unter Hochverfügbarkeit vorstellen: Cluster-Setup mit redundantem Netzwerken und Speichersystemen. Das führt Liebel an den Open-Source-Paketen OpenAIS, Corosync, DRBD und Pacemaker im Detail vor. Die Praxis steht dabei klar im Vordergrund und so nehmen Shell-Kommandos entsprechend großen Raum ein. Er beschreibt und konfiguriert verschiedene Cluster-Typen und geht auch auf spezialisierte Dateisysteme wie OCFS2 und

GFS2 ein, allerdings auf vergleichsweise wenigen Seiten. Auch iSCSI kommt als Storage-System praktisch zum Einsatz. Im letzten Teil des Buchs beschreibt der Autor, wie Virtualisierungstechnologien – konkret Xen und KVM – dabei helfen, hochverfügbare Lösungen aufzubauen. Diesem Thema ist vergleichsweise wenig Raum gewidmet, aber schließlich steht es auch nicht im Mittelpunkt.

Wer mit aktuellen Open-Source-Technologien hochverfügbare Cluster aufbauen will, ist mit dem Buch gut bedient, denn es gibt direkt umsetzbare Anleitungen dafür. Geschmackssache ist der betont lockere Stil des Buchs.

Xen nach Rezept

Kochbücher für die verschiedensten Themen haben bei O'Reilly lange Tradition. In diese reiht sich das Xen-Kochbuch von Hans-Joachim Picht ein, das der Autor laut Vorwort auf seinen Reisen geschrieben hat. Die Arbeit am 2009 erschienenen Buch begann bereits 2007 und das merkt man dem Stand der beschriebenen Software-Releases an.

Das Buch umfasst 20 Kapitel, die beginnend bei einer Einführung in das Thema Virtualisierung bis zu Spezialthemen wie Hochverfügbarkeit oder komplexe-

ren Netzwerkkonfigurationen mit VLANs reichen. Dabei gibt der Autor nach Kochbuchart eine kurze Einführung in das Thema des Kapitels und liefert dann das Rezept zur Lösung des Problems.

Das Buch deckt die Grundinstallation und Konfiguration einer Dom0, der zugehörigen virtuellen oder physischen Festplatten und der Installation der Gast-systeme ab. Dabei geht der Autor auf die gängigen Distributionen (Red Hat, Debian/Ubuntu und Suse; Gentoo fehlt) ein, und beschreibt auch die Installation aus dem Quellcode. Hier ist die Beschreibung allerdings an einigen Stellen zu kurz gehalten, zum Beispiel der Bau von Xen-Kerneln. Neben Xen behandelt der Autor auch ergänzende Themen in Kochrezepten, wie etwa LVM- oder DRBD-Konfigurationen.

Der Autor schwankt in der Ansprache seiner Zielgruppe zwischen erfahrenen Administratoren und Anfängern, was die Lektüre an einigen Stellen etwas anstrengend macht. Liest man das Buch nicht nur auf der Suche nach einem einzelnen Rezept, sondern im Zusammenhang, so kommt sich der Leser an einigen Stellen etwas überflutet vor, ohne dass dabei genügend Struktur mitgeliefert wird.

Wer klare Handlungsanweisungen für Xen-Fragen wie „Wie konfiguriere ich in Xen den Zugriff auf VLANs?“ sucht, ist mit dem Xen-Kochbuch gut beraten. Eine aktualisierte Auflage, die die Änderungen der jüngsten Zeit berücksichtigt, wäre wünschenswert. ■

Der Autor

Konstantin Agouros arbeitet bei der N.runs AG als Berater für Netzwerksicherheit. Dabei liegt sein Schwerpunkt im Bereich der Mobilfunknetze. Sein Buch „DNS/DHCP“ ist bei OpenSource Press erschienen.

Linux Hochverfügbarkeit

Linux Hochverfügbarkeit
Oliver Liebel
Galileo Computing 2011
450 Seiten
50 Euro
ISBN 978-3-8362-1339-4



Xen-Kochbuch

Xen-Kochbuch
Hans-Joachim Picht
O'Reilly 2009
488 Seiten
39,90 Euro
ISBN 978-3-89721-729-4



Neue Software und Produkte

Branchen-News

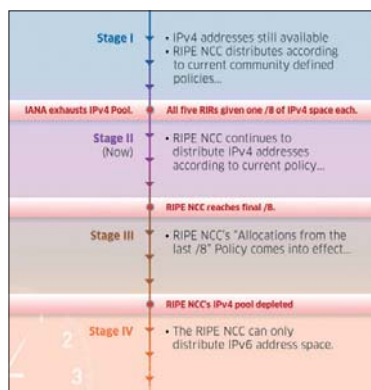
Enterprise Open Source Day

In Nürnberg-Langwasser trafen sich am 1. Februar Vertreter der deutschen Open-Source-Welt zum ersten Enterprise Open Source Day (EOSD). Eingeladen hatte das Nürnberger Systemhaus Ancud. Einige der Aussteller hatten interessante Neuheiten im Gepäck. Univention profilierte sich mit der ersten Referenzimplementierung von Samba 4 als vollständiger Active-Directory-Server, der wohl schon vor Ort bei einigen Kunden als „Drop-in-Replacement“ zahlreiche Windows-Systeme im Griff hat. BI-Spezialist Jaspersoft hatte vor Kurzem die neue Version 4 seiner Management- und Prozessoptimierungssuite veröffentlicht, die EOSD-Besucher jetzt am Stand und in einem mehrstündigen Workshop einer ersten Prüfung unterziehen konnten. Und Helmuth Neuberger von Zarafa erzählte am Rande der Veranstaltung, dass der Groupware-Hersteller sein IMAP/MAPI-Gateway optimiert habe und pünktlich zur Fosdem 2011 die Datenbank-Performance des MySQL-Backends verdoppelt haben will.

In den drei gut besuchten Tracks standen Anwendungen der freien Software-Welt für den Unternehmenseinsatz im Mittelpunkt. Allgegenwärtiges Thema war Cloud Computing nebst Webanwendungen von Sugar und Helium V. Der Open-Source-Professor Dirk Riehle von der Friedrich-Alexander-Universität (FAU) Erlangen-Nürnberg stellte die wissenschaftliche Sichtweise auf das „Single-Vendor Commercial Open Source Business Modell“ vor, gepaart mit zahlreichen Analysen aus dem Bereich der Innovationsforschung und Untersuchungen zum Beispiel der Linux-Kernel-Entwicklung. Der Enterprise Open Source Day 2011 war der erste, und Konstantin Böhm vom Veranstalter Ancud IT zeigte sich sehr zufrieden. Das Konzept sei aufgegangen, es seien mehr Besucher als erwartet gekommen, und man werde die Veranstaltung auf jeden Fall im nächsten Jahr fortführen. Ziel sei es, in Nürnberg mit seinen zahlreichen Open-Source-Firmen ein deutschlandweit bedeutsames Linux-Enterprise-Event fest zu etablieren.

IANA gibt letzte Runde an IPv4-Adressen aus

Der oberste Verwalter des Internet-Zahlenwerks gibt die letzten Blocks von IPv4-Adressen an die regionalen Registries aus. Die Internet Assigned Numbers Authority (IANA) hat zwei /8-Blocks verbleibende IPv4-Adressen an das Asia-Pacific Network Information Centre (APNIC) ausgegeben. Ein solcher /8-Block umfasst etwa 16 Millionen IPv4-Adressen. Die Vergabe der beiden Blocks reduziert die bei der IANA verbleibenden /8-Blocks auf fünf, und läutet damit die letzte Phase ein, in der die IANA diese Blocks an die regionalen Registries verteilt. Wie das APNIC in einer Mail [<https://puck.nether.net/pipermail/cisco-nsp/2011-January/076689.html>] mitteilt, setzt es die normale



Zeitplan der europäischen Registry RIPE zum nahenden Ende der IPv4-Adressen.

Vergabe der alten IP-Adressen noch einige Monate fort, bevor eine neue Phase des Übergangs auf IPv6 beginnt. In dieser Übergangsphase sollen noch einige Jahre lang IPv4-Adressen erhältlich sein. Die für den europäischen Raum zuständige lokale Registry RIPE sieht einen ähnlichen zeitlichen Ablauf vor.

Yahoo entwickelt Hadoop bei Apache weiter

Eric Baldeschwieler, bei Yahoo in führender Position für die Hadoop-Entwicklung zuständig, hat bekanntgegeben, dass das Unternehmen künftig seine Kräfte beim Hadoop-Projekt von Apache einbringen werde.

Die hausinterne Yahoo-Distribution von Hadoop werde hingegen eingestellt, schreibt Baldeschwieler im Yahoo-Entwicklerblog [<http://developer.yahoo.com/blogs/hadoop/>]. Im Zuge der Neuausrichtung auf Apache verschwinden alle Referenzen zu Hadoop von den Yahoo-Seiten, wie „developer.yahoo.com/hadoop“. Zudem werden die Versionsverwaltungen abgeklemmt („yahoo.github.com/hadoop-common“).

Yahoo werde sich bei Apache an der Hadoop-Entwicklung dergestalt beteiligen, dass gehärtete Binary-Releases für Yahoo und andere Hadoop-Nutzer für den Einsatz in ihren Clustern bereitstehen. Aus Hadoop soll durch die Mithilfe Yahoos unter dem Dach von Apache „die“ Open-Source-Plattform für große Datenmengen entstehen. Dazu müssen aber erst die beiden unterschiedlichen Entwicklungszweige von Yahoo und Apache abgeglichen werden.

Yahoo will seine Version „hadoop-0.20-sustaining“, die mit Patches und Fixes bei den rund 40.000 Yahoo-Nodes im Einsatz sei, bei Apache unter „hadoop/common/branches/branch-0.20-security“ einstellen. Die Übertragung habe bereits begonnen, schreibt Baldeschwieler. Yahoo schlägt vor, diese Version Hadoop 20.100 zu nennen und die Community darüber abstimmen zu lassen, ob daraus ein Apache-Release werden könne.

Virtuelle Server Top-Performance zum Tiefpreis!



Debian lädt zur Distributionszählung

Der Derivatives Front Desk von Debian will die Verbindung zu den Derivaten der Linux-Distribution stärken und lädt ein, sich an einem Zensus im Sinne einer regelmäßigen Zählung zu beteiligen. Für den Zensus steht eine neue Wikiseite [<http://wiki.debian.org/Derivatives/Census>] bereit, auf der sich auf Debian aufbauende Linux-Projekte steckbriefartig vorstellen. Wer sich daran beteiligt, verschafft seinem Debian-Derivat innerhalb des Debian-Projekts mehr Aufmerksamkeit, wirbt das Projekt. Durch die Angabe eines Repräsentanten werde die Zusammenarbeit einfacher, und im besten Fall steuern diese Repräsentanten Ideen und sogar Patches oder Bug-Reports bei, die sich auf ihr jeweiliges Derivat beziehen und für Debians Software-Basis relevant sind. Auch der Erfahrungsaustausch zwischen Derivaten soll gefördert werden, indem die Derivate überhaupt sichtbar werden. Die Zensus-Seite soll in Zukunft als Informationssammlung dienen, aus der Debian-Entwickler und Derivat-Projekte Ressourcen für die Programmierung ziehen und die sich in die Debian-Infrastruktur integriert. Ubuntu-Derivate besitzen derzeit noch eine eigene Seite.

Der Derivatives Front Desk existiert seit Juni 2010. Auf der damals eingerichteten Mailingliste „debian-derivatives“ wurde seitdem diskutiert, wie die Zusammenarbeit der Derivate und Debian weitergehen könne. Zum Beispiel ist eine Wikiseite hinzugekommen, die Leitlinien zum Erstellen von Debian-Derivaten sammelt. Für den Austausch gibt es neben der Mailingliste auch einen IRC-Channel (irc.debian.org/debian-derivatives).

Rackspace kauft Openstack-Entwickler ein

Der Internet-Hosting-Provider Rackspace übernimmt die Entwicklungs-Firma Anso Labs, die hinter großen Teilen der Cloud-API Openstack steht. Damit gewinnt Rackspace nun größeres Gewicht in der Entwicklung des Openstack-Projekts, das sich in die beiden Teilprojekte „Compute“ und „Storage“ gliedert. Rackspace hatte bisher am Storage-Teil gearbeitet, während Compute unter dem Namen Nebula von Anso Labs für den Projektpartner Nasa entwickelt wurde.

Mit der Übernahme von Anso Labs steuert Rackspace nun beide Projektteile, was zu Befürchtungen in der Openstack-Community Anlass gibt, denn nun besitzt Rackspace in den Leitungsgremien des offenen Openstack-Projekts deutliche Mehrheiten. Außer der Nasa, Rackspace und Anso Labs unterstützen auch die Firmen Cisco, Citrix, Canonical und Microsoft das Openstack-Projekt.

In der neuesten Version mit dem Codenamen „Bexar“ unterstützt das Cloud-Computing-Framework Openstack erstmals das IPv6-Protokoll. Als Hypervisor für die virtualisierten Cloud-Knoten kann Linux-KVM, Xen und nun auch Microsoft Hyper-V zum Einsatz kommen. VMware ESX soll im kommenden Release „Cactus“ folgen, genauso wie der Support für Linux-Container und Open VZ. Die Openstack-API, die sich derzeit noch in der Entwicklungsphase befindet, wurde um einige Administrations-Features erweitert.

Virtuelle Server von netclub

- bis zu 3 CPU-Kerne und 8 GB RAM
- bis zu 95 GB Festplatte (RAID 10)
- 5 TB Traffic inklusive
- SSL-Zertifikat inklusive
- voller Root-Zugriff (SSH)
- 100 % Backup-Speicher
- 99,9 % garantierte Verfügbarkeit
- auch als Managed Server erhältlich
- viele 64-Bit-Betriebssysteme nach Wahl

**6 Monate
kostenlos**
danach ab 12,99 €*
*

Jetzt kostenlos informieren unter:

0800 638 2587
www.netclub.de/linux

* Aktion „6 Monate kostenlos“ nur gültig bis 30.04.2011. Nach 6 Monaten regulärer monatlicher Grundpreis: VPS L 12,99 €, VPS XL 16,99 €, VPS XXL 29,99 €. Die Mindestvertragslaufzeit beträgt wahlweise 12 Monate (Aktion 6 Monate kostenlos entfällt) oder 24 Monate (6 Monate kostenlos). Abrechnung vierteljährlich. Einmalige Einrichtungsgebühr 9,99 €. Preise inkl. MwSt. Preisänderungen und Irrtümer vorbehalten.

Schluss mit Ruby on Rails in Netbeans

Die Netbeans-Entwickler sollen sich in Zukunft wieder auf ihr Kerngeschäft konzentrieren. In einer Mitteilung an die Community kündigt das Team um die Netbeans-Entwicklungsumgebung an, in Zukunft Ruby on Rails nicht mehr zu unterstützen. Als Grund gibt es an, sich in Zukunft auf den Support des bevorstehenden Java SE 7 und Java Development Kit 7 (JDK 7) zu konzentrieren. Herausragender Java-Support habe immer im Fokus der Netbeans-Entwicklung gestanden, und darauf wolle man sich nun erst einmal konzentrieren. Ein weiterer Grund für die Aufgabe des Ruby-on-Rails-Supports seien die geringen Anwenderzahlen in der Nutzungsstatistik.

In Zukunft sei der Code für Ruby on Rails nicht mehr Teil der Netbeans-Entwicklungs-Builds. Weiterhin an diesem Modul Interessierte werden auf die Netbeans-Entwicklerseite für den Ruby-Support verwiesen und aufgerufen, das Projekt selbst weiterzuführen. [<http://wiki.netbeans.org/RubySupport>]

Google startet Summer of Code 2011

Der Startschuss des von Google finanzierten Summer of Code (SoC) zur Unterstützung von freien Software-Projekten fällt bei der Linuxcon Australien.

Google eröffnet bei der – trotz Überflutungen – in Australien stattfindenden Linux.conf.au den SoC 2011. Auf der Website des SoC 2011 [<http://www.google-melange.com/>] ist der Zeitplan einzusehen. Demnach können Organisationen und Projekte vom 28. Februar bis 11. März ihre zu bearbeitenden Anwendungen einreichen. Am 18. März veröffentlichen die Programm-Administratoren von Google die Liste der akzeptierten Programme. Beim Summer of Code bringt der Suchmaschinenkonzern studentische Entwickler mit Projekten und Organisationen zusammen, die Anwendungen weiterentwickeln wollen. Die Zusammenarbeit der Helfer und der Projekte erfolgt über sogenannte Mentoren aus dem Kreis der Projekte. Google unterstützt und entlohnt die teilnehmenden Entwickler.

Oracle Linux 6 veröffentlicht

Wie der Datenbankhersteller Oracle mitteilt, ist die neueste Version seiner Linux-Distribution verfügbar. Im Wesentlichen handelt es sich um einen Klon des kürzlich erschienenen Red Hat Enterprise Linux 6, mit dem es deshalb die grundlegenden Neuerungen wie das Ext4-Datei-

system, Verbesserungen bei XFS, Performance Counter, FTrace, Latencytop, Tickless Kernel und Control Groups teilt. Regelmäßige Updates bietet Oracle über sein Unbreakable Linux Network.

Als eigenständiges Feature bietet Oracle Linux optional einen eigenen Kernel un-

ter dem Namen Unbreakable Enterprise Kernel an, der in Bezug auf die Interrupt-Verteilung, Locks sowie Netzwerk- und Virtuelle-Speicher-Performance verbessert wurde. Per Default wird aber der Red-Hat-kompatible Kernel mit der Versionsnummer 2.6.32-71.el6 installiert.

IDC: Smartphone-Markt 2010 stark in Westeuropa

In Westeuropa trugen neben Apple mit dem iPhone auch HTC, Nokia und Samsung zum positiven Trend im Smartphone-Markt bei. Weltweit stieg 2010 die Zahl der insgesamt verkauften Handys um fast 20 Prozent auf 1,4 Milliarden. Speziell im westeuropäischen Markt sehen die Marktbeobachter, wie sich immer



Jose Luis Martinez, Vice President, Mobile Computing Solution Marketing mit dem N8.

mehr Kunden von Smartphones überzeugen lassen, und zählen I-Phone 4, BlackBerry 8520, das HTC Desire, das Nokia N8 und das Samsung Galaxy S als Zugpferde auf. Apple findet in Westeuropa sowie in den USA nach wie vor den größten Absatzmarkt und legte mit dem I-Phone 4 im vierten Quartal deutlich zu.

Bei den fünf weltweit führenden Handy-Herstellern kann sich Nokia an der Spitze behaupten: 453 Millionen verkaufte Geräte, fast 33 Prozent Marktanteil bei fünf Prozent Wachstum gegenüber 2009 sind zu verzeichnen. Mit Geräten wie dem C7, C6-01 und C3 hat es Nokia im letzten Quartal gegenüber dem Vorjahresquartal auf 38 Prozent Wachstum im Smartphone-Bereich gebracht. Auf dem zweiten Platz liegt Samsung mit 280 Millionen verkauften Geräten, gut 20 Prozent Marktanteil und 23 Prozent Wachstum. Der Hersteller hat laut IDC erstmals in der Unternehmensgeschichte

allein im vierten Quartal mehr als 80 Millionen Geräte verkauft, von denen zehn Millionen auf Galaxy-S-Telefone abfallen. Abgeschlagener Dritter mit 8,4 Prozent weltweitem Marktanteil und rund 117 Millionen verkauften Geräten ist LG. An fünfter Stelle steht RIM, das weltweit stolze 41 Prozent Wachstum verzeichnete und nunmehr einen Marktanteil von 3,5 Prozent hält. Gegen Ende des Jahres schob sich erstmals ein chinesischer Hersteller in die Top-Five-Liste, hebt IDC hervor, und zwar mit insgesamt 3,7 Prozent Marktanteil auf Platz vier: ZTE, gegründet 1985 im chinesischen Shenzhen, hat von dem weltweit wachsenden Mobiltelefonmarkt profitiert. Speziell auf dem amerikanischen Markt halten mit Dell, Huawei, Kyocera und Sanyo ebenfalls neue Namen Einzug. Den Mobiltelefonmarkt sieht der Marktbeobachter wegen des kräftigen Smartphone-Segments bis in das Jahr 2014 auf Wachstumskurs.

Puppet als Enterprise-Variante erhältlich

Der US-amerikanische Software-Hersteller Puppet Labs hat mit Puppet Enterprise eine kommerzielle Version der sonst freien Puppet-Software für das Konfigurationsmanagement von Rechnern im Netzwerk vorgestellt. Die Enterprise-Version bündelt alle der zahlreichen Puppet-Komponenten sowie die Software-Voraussetzungen wie Apache, Ruby on Rails, Phusion Passenger und mehrere Ruby-Module in einem einzigen Installationsprogramm. Darüber hinaus sei Puppet Enterprise einem umfangreichen Qualitätssicherungsprozess unterzogen worden, verspricht der Hersteller. Im Paket ist außerdem grundlegender Support per E-Mail enthalten.

Das kommerzielle Angebot von Puppet Labs beginnt bei 2500 US-Dollar. Die Preise richten sich nach der Anzahl verwalteter Rechner und der Art des Supports. Mit rund um die Uhr verfügbarem Telefonsupport kostet das Paket beispielsweise gleich das Zehnfache. Eine kostenlose Testversion, mit der sich zwei Rechner managen lassen, kann unter [\[http://info.puppetlabs.com/puppet-enterprise\]](http://info.puppetlabs.com/puppet-enterprise) heruntergeladen werden. Unterstützte Betriebssysteme sind Red Hat Enterprise Linux 5.5, Cent OS 5.5, Ubuntu LTS 10.04, Debian Lenny 5.0.7 und Oracle Enterprise Linux 5 Update 5.

VMware veröffentlicht Groupware Zimbra 7

Zimbra Collaboration Server 7, VMware Zimbra Desktop 7 und VMware Zimbra Appliance 7 heißen die Produkte im Detail, wobei Appliance und Desktop derzeit noch im Betastadium stecken. VMware verspricht neue Funktionen wie den cloud-basierten Dateienaustausch über die sogenannte Briefcase-Funktion von Zimbra. Diverse Optionen für das Dokumentenmanagement sind für die gemeinsam genutzten Dateien in der Briefcase möglich.

Die Weitergabe von Kontakten als V-Card und die „People-Search“ zum simultanen Durchforsten von Kontaktlisten zählen ebenfalls zu den Neuerungen. Ein rollenbasiertes Administrationsmodell soll die Verwaltung erleichtern. Der Nutzer authentifiziert sich in der neuen Version über die standardmäßigen Protokolle und Verfahren wie Single-Sign-On (SSO) über OpenID, Security Assertion Markup Language (SAML) und Kerberos. Zimbra-Server und die Beta-Versionen von Zimbra Appliance und -Desktop gibt es zum Herunterladen. Das Komplettpaket ist kostenpflichtig. Eine Open-Source-Ausgabe zum Testen ist ebenfalls vorhanden, die Zimbra Collaboration Suite Open Source Edition steht unter Zimbra Public License (ZPL) und ist als Quellcode und in Form von Binaries erhältlich.

„Es gibt drei Möglichkeiten, eine Firma zu ruinieren:
mit **Frauen**, das ist das Angenehmste;
mit **Spiele**n, das ist das Schnellste;
mit **Computern**, das ist das Sicherste.“

Oswald Dreyer-Eimbcke, deutscher Unternehmer



Es gibt allerdings auch eine ganze Reihe von Möglichkeiten, mit IT zum Unternehmenserfolg beizutragen. Und genau damit beschäftigen wir uns seit dem Jahr 1989. In weit über 100.000 Stunden IT-Training, technischem Consulting und Software-Entwicklung haben wir eine ganze Menge Know-how entwickelt. Dieses umfassende Wissen kommt Ihnen heute zugute, wenn wir Sie in allen IT-Fragen unterstützen.

Wie Sie von unserem Know-how profitieren, erfahren Sie unter www.lamarc.com.
Rufen Sie uns an +49 611 26 00 23 oder schicken eine E-Mail an info@lamarc.com

Seminare OnSite Coaching Netzwerk Consulting System Consulting Developer Support Software-Entwicklung



Nikita Sobolov, 123RF

VMware-Server verweigert den Dienst

Lost in Update

Ein Administrator gerät bei einer Linux-Distribution mit Paket-Updates in eine Falle, aus der er sich mit Geschick wieder befreien kann. Ein Einblick in einen Unfall und seine praktische Lösung. Dr. Udo Seidel

Gerade im heimischen Umfeld sind Linux-Distributionen mit tagesaktuellem Update-Stand keine Seltenheit. Die Updates sind höchstwahrscheinlich sogar automatisiert und der Benutzer bemerkt davon nichts – es sei denn, eine Applikation verweigert anschließend ihren Dienst.

Prinzipiell gibt es dann verschiedene Ansätze, um das Problem zu lösen. Der vorliegende Artikel erläutert diese am konkreten Beispiel – hier versagte der VMware-Server 2.0 seinen Dienst, nachdem der Administrator das darunterlie-

gende OS von Fedora 13 auf Fedora 14 aktualisiert hatte.

Vorgeschichte

Ausgangspunkt war ein funktionierender VMware-Server [1] unter einem 64-Bit-Fedora 13 (inklusive aller vorhandenen Updates) [2]. Neben den verschiedensten Gästen für Test-Zwecke laufen dort zwei quasi-produktive virtuelle Maschinen, auf die der Autor nur für einen begrenzten Zeitraum verzichten kann. Fedora 14 war gerade frisch erschienen und die ers-

ten Erfahrungen auf anderen Rechnern stimmten positiv auf ein vollständiges Upgrade der heimischen Linux-Welt. Also flugs die neuen Paket-Repositories einbinden und den Paketverwalter Yum [3] anwerfen.

Dank eines schmalen OS-Images und schneller Internet-Anbindung schmückte den Server einige Minuten und einen Reboot später ein taufrisches Fedora 14. Nur irgendwie funktionierte die Verbindung zur VMware-Anwendung nicht mehr. Ein Blick in die Datei »/var/log/messages« offenbarte das zugrunde liegende Problem (**Listing 1**) – der Prozess »vmware-hostd« hatte sich mit einem sogenannten Segmentation Fault verabschiedet, er hat also versucht, auf einen

Listing 1: Segmentation Fault bei »vmware-hostd«

```
01 Dec 6 13:30:08 virtual kernel: [ 175.894212] vmware-hostd[3870]: segfault at 2100001c4f ip
0000003c0cb32ad0 sp 00007f3889e9cb88 error 4 in libc-2.12.90.so[3c0ca00000+19a000]
```

Speicherbereich zuzugreifen, der vor einem solchen Zugriff geschützt ist.

Ein erste Analyse zeigte, dass sich die VMware-Anwendung nicht mit der neueren Version der glibc von Fedora 14 verträgt. Gibt es eine neuere Version des VMware-Servers? Nein, denn VMware entwickelt dieses Produkt nicht mehr weiter [4]. Die Alternative wäre das ESXi-Produkt [5]. Leider ist die vorhandene Hardware nicht kompatibel, also scheidet dieser Ansatz als kurzfristige Lösung aus. Ein Blick auf den Fedora-Update-Server zeigt, dass es auch keine neuere Version des Glibc-Paketes gibt. Ein OS-Update als Lösungsmöglichkeit entfällt somit auch.

Bleiben nur noch zwei Wege, um des Problems Herr zu werden: ein Fallback der Änderung am System oder ein eigener Fix oder Workaround. Der Fallback entspricht praktisch dem Downgrade von Fedora 14 auf Fedora 13. Ohne entsprechende Vorbereitungen wie Dateisystem- oder Volume-Snapshots ist dies keine triviale Angelegenheit. Außerdem ist das Problem der Unverträglichkeit von Applikationen mit neueren Glibc-Version nicht unbekannt und für bestimmte Anwendungen recht einfach zu beheben.

Notbehelf

Der Workaround in diesem Fall ist die Parallel-Installation einer älteren Variante der Bibliothek und die anschließende Mo-

difikation der Applikations-Umgebung, um die „richtigen“ Bibliotheken zu verwenden. Was sich einfach anhört, kann im konkreten Fall ziemlich aufwendig sein und hier führte der erste Versuch leider gar nicht zum Erfolg. Inzwischen war ein ganzer Tag vergangen und der Verlust der zwei oben erwähnten virtuellen Server begann so richtig zu schmerzen. Eine Lösung musste her – also doch der Downgrade (Listing 2).

Hoch und nieder

Prinzipiell könnte man versuchen, nur einzelne Pakete durch die jeweils ältere Version zu ersetzen – angefangen natürlich bei der Glibc. Dies setzt aber voraus, dass der Admin genau weiß, welche Pakete er austauschen muss. Außerdem ist der Mischbetrieb mit Paketen aus zwei verschiedenen Distributions-Versionen nicht gerade empfehlenswert. Unerwünschte Seiteneffekte bis hin zur Instabilität sind die Folge, insbesondere, wenn eine zentrale Komponente wie die Glibc beteiligt ist.

Glücklicherweise beherrscht Yum seit Version 3.2.27 die Option »downgrade« und erleichtert den Schritt von Fedora 14 zurück zu Fedora 13 erheblich. Diese Funktion lässt sich allerdings nur auf einzelne Pakete anwenden. Der Administrator muss somit zuerst eine entsprechende Paket-Liste generieren. Ohne

Verwendung der History-Option von Yum führt der einfachste Weg über »rpm -qa --queryformat '%{NAME}\n' |xargs yum downgrade«. Leider war auch dieses Unternehmen in der Realität nicht von Erfolg gekrönt (siehe Abbildung 1).

Das prinzipielle Problem besteht darin, dass Fedora 14 einzelne Software-Pakete

Listing 2: Handarbeit vor dem Downgrade

```
01 rpm -e libmount --nodeps
02 yum downgrade util-linux-ng libblkid libuuid
03 rpm -e upstart-sysvinit --nodeps
04 yum downgrade upstart
05 rpm -Uvh gdbm-1.8.0-33.fc12.i686.rpm --nodeps --force
06 yum downgrade perl\*
07 rpm -e man-db
08 rpm -e pam_ldap --nodeps
09 yum downgrade nss_ldap
10 ...
```

Listing 3: Bibliotheken von »vmware-hostd«

```
01 $ ldd /usr/lib/vmware/bin/vmware-hostd
02     linux-vdso.so.1 => (0x00007fff75fff000)
03     libz.so.1 => /lib64/libz.so.1
04         (0x0000003c0e600000)
05     libvmomi.so.1.0 => not found
06     libvmacore.so.1.0 => not found
07     libexpat.so.0 => not found
08     libcrypt.so.1 => /lib64/libcrypt.so.1
09         (0x0000003c0da00000)
10     libxml2.so.2 => /usr/lib64/libxml2.so.2
11         (0x0000003c14200000)
12     libstdc++.so.6 => /usr/lib64/libstdc++.so.6
13         (0x0000003c0e200000)
14     libpthread.so.0 => /lib64/libpthread.so.0
15         (0x0000003c0d200000)
16     libgcc_s.so.1 => /lib64/libgcc_s.so.1
17         (0x00007f086f4d5000)
18     libc.so.6 => /lib64/libc.so.6
19         (0x0000003c0ca00000)
20     /lib64/ld-linux-x86-64.so.2
21         (0x0000003c0c600000)
22     libdl.so.2 => /lib64/libdl.so.2
23         (0x0000003c0ce00000)
24     libm.so.6 => /lib64/libm.so.6
25         (0x0000003c0de00000)
26     libfreebl3.so => /lib64/libfreebl3.so
27         (0x0000003c0ee00000)
28 $
29 $ ldd /usr/lib/vmware/bin/vmware-hostd |awk '{print
30     $3}'|grep lib | \
31     xargs rpm --queryformat '%{NAME}\n' -qf |sort -u
32 glibc
33 libgcc
34 libstdc++
35 libxml2
36 nss-softokn-freebl
37 zlib
38 $
```

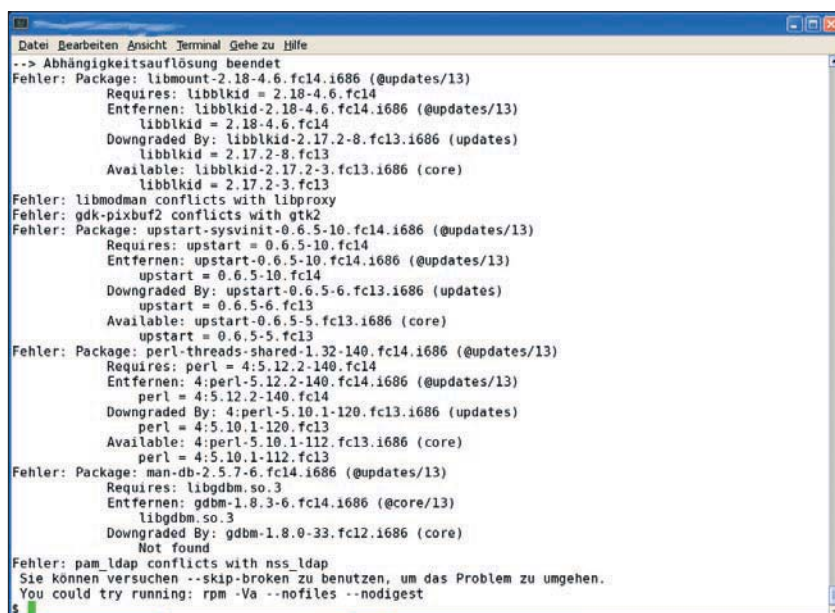


Abbildung 1: Yum erleichtert den Weg von Fedora 14 zu Fedora 13 – erfordert dabei aber immer noch etwas Handarbeit.

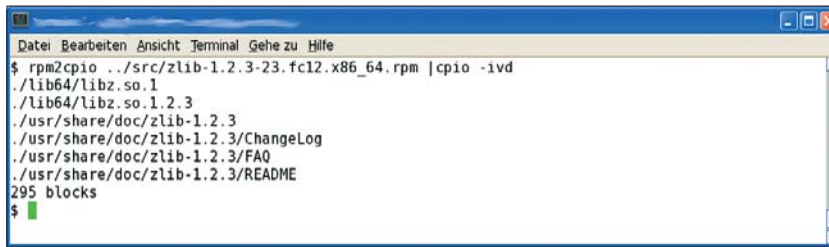


Abbildung 2: RPM-Pakete lassen sich mit »rpm2cpio« und »cpio« einfach extrahieren.

anders auf RPMs aufteilt als Fedora 13. In den meisten Fällen hilft es, das konfliktzeugende Paket zu über »rpm -e Paket1 --nodeps« zu entfernen und anschließend das jeweils andere Paket durch »yum downgrade Paket2« auf die ältere Version zu bringen. Nach ein paar Iterationen dieser Prozedur sollten alle Konflikte aufgelöst sein und der Admin kann das Gesamtsystem auf die ältere Version zurücksetzen. Ein finaler Reboot aktiviert den alten Kernel und die System-Bibliotheken und siehe da, auch der VMware-Server ist wieder zufrieden und tut seinen Dienst.

Ein neuer Versuch

Der Fallback ist allerdings nur eine kurzfristige Lösung. Früher oder später muss entweder ein aktuelleres Betriebssystem oder eine andere Variante der Virtualisierungssoftware her. Für den Moment ist die Lage aber entspannt und der Admin kann sich in Ruhe Gedanken über die endgültige Lösung machen. Der vorliegende Fall der Unverträglichkeit von Glibc und VMware-Server war nicht wirklich unbekannt in der Linux-Gemeinde. Die im Netz vorgestellten Lösungen verwendeten allesamt den oben beschriebenen Ansatz der Parallel-Installation einer älteren glibc-Version. Sollte dies doch der richtige Weg sein? Vielleicht gibt es noch versteckte Abhängigkeiten, beispielsweise weitere Bibliotheken, die der Admin in einer älteren Version parallel installieren muss. Ein Blick auf das Executable »usr/

sbin/vmware-host« offenbart, dass dies nur ein Shell-Skript ist. Die eigentliche Binär-Datei liegt im Verzeichnis »usr/lib/vmware/bin/. Ein »ldd /usr/lib/vmware/bin/vmware-host« zeigt tatsächlich, dass noch weitere Objekte relevant sind, nämlich insgesamt sechs Pakete (siehe Listing 3).

Ausweg

Schließlich bringt das Paket »zlib« die Lösung für das Problem. Der Rest ist dann nur noch Formsache. Zuerst erstellt der Admin ein Verzeichnis für die Parallel-installation: »mkdir -p /usr/lib/vmware/lib/fc13lib64«. Danach besorgt er sich die RPMs von »glibc« und »zlib« für Fedora 13 und packt sie aus. Für die Parallel-installation kann der Admin »rpm« mit einem alternativen Root-Verzeichnis verwenden [6][7]. Die bessere Wahl ist aber, die RPMs in CPIO-Archive umzuwandeln und diese dann auszupacken [7][8]. So umgeht man auf einfache Weise die Ausführung etwaiger Pre- oder Post-Installationsskripte und die RPM-Datenbank des Systems bleibt auch sauber (siehe Abbildung 2).

Untergeschoben

Sind die Bibliotheken am richtigen Platz, muss der Admin dem Binär-Programm »vmware-host« noch beibringen, die parallel installierten Objekte zu verwenden. Dazu setzt er die Shell-Variable »LD_LIBRARY_PATH« entsprechend, bevor aus

dieser Shell das Binär-Programm startet. Das ist recht einfach, da VMware selbst auf die gleiche Weise eigene Bibliotheken integriert. Ist alles erledigt, kann der Admin das Konstrukt testen. Dies kann sogar schon unter Fedora 13 geschehen. Der finale Test kann natürlich nur unter dem neuen OS erfolgen. Vor dem Upgrade sollte der Admin aber diesmal geeignete Maßnahmen ergreifen, die einen Fallback einfacher gestalten (Listing 4). Glücklicherweise bleiben hier weitere negative Überraschungen aus und der VMware-Server verrichtet nun auch unter Fedora 14 seine Dienste.

Fazit

Die hier vorgestellte Lösung funktioniert prinzipiell bei anderen Linux-Distributionen. Je nach verwendetem Paket-Manager muss der Admin das Extrahieren der »alten« Bibliotheken anders bewerkstelligen. Außerdem hat sich wieder bewährt, dass man vor großen Änderungen am System unbedingt einen Plan B haben sollte, der greift, wenn etwas Unerwartetes passiert. (ofr)

Infos

- [1] VMware: [\[http://www.vmware.com/products/server/\]](http://www.vmware.com/products/server/)
- [2] Fedora: [\[http://fedoraproject.org/\]](http://fedoraproject.org/)
- [3] Yum: [\[http://yum.baseurl.org/\]](http://yum.baseurl.org/)
- [4] VMware Life Cycle Policies: [\[http://www.vmware.com/support/policies/lifecycle/general/\]](http://www.vmware.com/support/policies/lifecycle/general/)
- [5] VMware vSphere Hypervisor: [\[http://www.vmware.com/products/vsphere-hypervisor/\]](http://www.vmware.com/products/vsphere-hypervisor/)
- [6] VMware Community, „Fedora 14 segfault“: [\[http://communities.vmware.com/message/1641223\]](http://communities.vmware.com/message/1641223)
- [7] RPM: [\[http://www.rpm.org/\]](http://www.rpm.org/)
- [8] GNU CPIO: [\[http://www.gnu.org/software/cpio/\]](http://www.gnu.org/software/cpio/)

Der Autor

Dr. Udo Seidel ist eigentlich Mathe-Physik-Lehrer und seit 1996 Linux-Fan. Nach seiner Promotion hat er als Linux/Unix-Trainer, System-administrator und Senior Solution Engineer gearbeitet. Heute ist er Leiter eines Linux/Unix-Teams bei der Amadeus Data Processing GmbH in Erding.

Listing 4: »LD_LIBRARY_PATH« für zusätzliche Bibliotheken

```
01 $ tail /usr/sbin/vmware-host                vmware/lib/fc13lib64
02 if [ ! "@VMWARE_NO_MALLOC_CHECK@" = 1 ]; then 08 export LD_LIBRARY_PATH
03     export MALLOC_CHECK_=2                    09 #####
04 fi                                           10
05                                           11 eval exec "$DEBUG_CMD" "$binary" "$@"
06 #####                                       12 $
07 LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/usr/lib/
```



Agiles System-Management in großen Systemlandschaften

Jetzt funkt's

Eine Handvoll Kommandos auf einer großen Anzahl von Systemen abzusetzen, ist ein klassisches Administrations-Problem. Statt mit Schleifen und einer Vielzahl von SSH-Verbindungen zu hantieren, bietet sich in solchen Fällen der Einsatz von Func an. Thorsten Scherf

Schnell mal eben ein Update des Apache-Pakets durchführen oder bestimmte Informationen des Systems einholen. Das ist oft leichter gesagt als getan, gerade dann, wenn Sie diese Aufgabe auf einer ganzen Reihe von Servern durchführen möchten. Ein schneller Workaround besteht im Einsatz einer For-Schleife auf der Kommandozeile. So lässt sich jeweils eine SSH-Verbindung zu den beteiligten Systemen aufbauen und der gewünschte Befehl dort ausführen:

```
for i in `seq 1 10`; do ssh www$i update &
httpd; done
```

Liegt der SSH-Schlüssel des Administrators auf den Systemen, so funktioniert der Aufruf zwar ohne manuelle Authentifizierung, trotzdem wäre eine elegantere Lösung für das Problem wünschenswert.

Abgesichert

Die Software Func [1] (Fedora Unified Network Controller) bietet für das Problem einen dedizierten, authentifizierten sowie verschlüsselten Kommunikationskanal an. Dieser wird zwischen einem zentralen Management-System (dem

Overlord) und allen Client-Systemen (den Minions, Listing 1) aufgebaut. Das Tool »certmaster« dient hierbei als zentrale PKI-Instanz, die sich um das Zertifikatsmanagement kümmert.

Func basiert auf einer Vielzahl von Modulen, die auf jedem Client-System zur Verfügung stehen. Jedes dieser Module bietet bestimmte Methoden an, die sich vom zentralen Management-System oder auch von anderen Client-Systemen aufrufen lassen. Die Module umfassen Methoden zum Ausführen von einzelnen Programmen, der Abfrage von System-Informationen bis zum Reboot der Sys-

teme. Dank der offenen API lässt sich Func sehr leicht durch eigene Module erweitern, das „Hack-Potenzial“ ist somit nahezu unbegrenzt.

Installation und Setup

Func ist Teil der Fedora- und EPEL-Software-Repositories. Für andere Distributionen stehen unter [2] ebenfalls die Sources zum Download bereit. Auf dem Management-System gelingt die Installation einfach mittels Yum:

```
yum install func
```

Im Anschluss ist mittels »service certmaster start« die PKI-Instanz zu starten. Über die Config-Datei »/etc/certmaster/certmaster.conf« bestimmen Sie, ob Zertifikatsanfragen der Client-Systeme automatisch bearbeitet werden (»autosign = yes«) oder ob diese Aufgabe manuell zu bewerkstelligen ist:

```
certmaster-ca --sign www1.virt.tuxgeek.de
```

Mit »certmaster-ca --list« zeigt das Tool alle noch offenen Zertifikatsanfragen an. »certmaster-ca --list-signed« hingegen listet alle bereits ausgestellten Zertifikate auf. Das CA-Zertifikat, das zum Signieren der Anfragen zum Einsatz kommt, liegt dabei üblicherweise als PEM-Datei im Verzeichnis »/etc/pki/certmaster/ca/«. Auf den Client-Systemen ist ebenfalls das Paket »func« zu installieren. Über die Konfigurationsdatei »/etc/certmaster/minion.conf« bekommt das System die Information, wie der Name des Management-Servers lautet. Das ist wichtig, da nach dem Start des Services »funcd« automatisch eine Zertifikatsanfrage an diesen gesendet wird. Diese kommt dann, je nach Konfiguration des Overlords, entweder unmittelbar oder nach dem manuellen Signieren zurück und landet im Verzeichnis »/etc/pki/certmaster/ca/«. Von nun an findet sämtliche Kommunikation zwischen dem Overlord und den Minions verschlüsselt statt. Es fließen also keine sensiblen Daten im Klartext über das Netzwerk. Die korrekte Funktionsweise des Clients ist recht leicht mittels Func zu testen:

Damit der Aufruf von Func auch als unprivilegiertes Benutzer möglich ist, bietet es sich an, für bestimmte Verzeichnisse POSIX-ACLs zu setzen (Listing 2).

Wer Func gerne automatisch auf all seinen Systemen installieren möchte, kann dies über Kickstart-Profile seines Cobbler-Systems [4] automatisieren. In der Cobbler-Konfigurationsdatei existieren hierfür extra zwei Einstellungen für zukünftige Func-Clients:

```
# grep '^func' /etc/cobbler/settings
func_auto_setup: 1
func_master: grobi.tuxgeek.de
```

In der »%packages«-Sektion des Kickstart-Profiles, das zur Installation von neuen Maschinen zum Einsatz kommt, können Sie nun das Cobbler-Template »\$func_install_if_enabled« hinzufügen. Dieses installiert das Func-RPM nur dann, wenn zuvor in der Cobbler-Konfigurationsdatei die Anweisung »func_auto_setup« auf 1 gesetzt wurde. In der »%post«-Sektion sorgt schließlich die Zeile »SNIPPET:func_register_if_enabled« dafür, dass sich das Client-System automatisch beim Func-Master registriert. Ist dieser dann noch so konfiguriert, dass er automatisch ein Client-Zertifikat ausstellt, ist das System sofort als Func-Minion einsatzbereit, eine manuelle Konfiguration ist nicht mehr notwendig. Einfacher geht es kaum noch. Es sei noch erwähnt, dass beim Einsatz von Cobbler das Fedora-Repository »Everything« als Distribution vorhanden sein muss, damit die Installation von Func auf Fedora Systemen erfolgreich ist. Für Red Hat Enterprise Linux oder Cent OS ist entsprechend das EPEL-Repository einzurichten:

```
cobbler repo add --name=e1-5-x86_64-epel
--mirror=http://download.fedora.redhat.com/pub/epel/5/x86_64
```

In den gewünschten Kickstart-Profilen können Sie dieses Repository dann bekanntmachen, sodass Pakete aus diesem Repository installiert werden können (hier das Template »\$func_install_if_enabled«, aus der »%packages«-Sektion):

```
cobbler profile edit --name=e15-ksprofile
--repos="e1-5-x86_64-epel"
```

Nachdem sichergestellt ist, dass Management-System und Clients miteinander gesichert kommunizieren können, bietet es sich erst einmal an, einen Überblick über die vorhandenen Module eines bestimmten Clients zu bekommen. Hierzu hilft der folgende Befehl weiter:

Listing 1: Korrekte Konfiguration eines Minions

```
01 # func "*" check --minion
02 SCAN RESULTS:
03 * FQDN is detected as www1.virt.tuxgeek.de, verify
   that is correct
04 * this minion is configured in /etc/certmaster/
   minion.conf
05 to talk to host 'tiffy' on port 51235 for certs,
   verify that is correct
06 End of Report.
```

Listing 2: Rechte für nicht privilegierte Benutzer

```
01 setfacl -d -R -m 'u:foobar:rX' /etc/pki/certmaster/
02 setfacl -R -m 'u:foobar:rX' /etc/pki/certmaster/
03 setfacl -d -R -m 'u:foobar:rX' /var/lib/certmaster
04 setfacl -R -m 'u:foobar:rX' /var/lib/certmaster
05 setfacl -d -R -m 'u:foobar:rX' /var/lib/certmaster/
   certmaster
06 setfacl -R -m 'u:foobar:rX' /var/lib/certmaster/
   certmaster
07 setfacl -d -R -m 'u:foobar:rX' /var/lib/certmaster/
   certmaster/certs
08 setfacl -R -m 'u:foobar:rX' /var/lib/certmaster/
   certmaster/certs
09 setfacl -d -R -m 'u:foobar:rX' /var/lib/certmaster/
   peers
10 setfacl -R -m 'u:foobar:rX' /var/lib/certmaster/peers
11 setfacl -d -R -m 'u:foobar:rwX' /var/lib/func
12 setfacl -R -m 'u:foobar:rwX' /var/lib/func
13 setfacl -d -R -m 'u:foobar:rwX' /var/log/func/
14 setfacl -R -m 'u:foobar:rwX' /var/log/func/
```

```
# func www1.virt.tuxgeek.de call system
list_modules
```

Die Liste sollte auf allen Systemen zumindest die Standard-Module beinhalten. Zu diesen gehören beispielsweise »yumcmd«

OpenSource Training Ralf Spenneberg
Schulungen direkt vom Autor






Weitere Schulungen:
 XEN, KVM, Modsecurity, Nagios,
 Hacking Webapplications, SELinux,
 AppArmor, Firewall, Postfix

Freie Distributionswahl:
 SuSE, Fedora, Debian, CentOS oder
 Ubuntu (NEU)

Ergonomische Arbeitsplätze
 Echte Schulungsunterlagen mit Übungen
 Bücher als Begleitmaterial

Am Bahnhof 3-5
 48565 Steinfurt
 Tel.: 02552 638755
 Fax: 02552 638757



Weitere Informationen unter www.os-t.de

oder auch »command«. Eine Liste der vorhandenen Methoden zu einem Modul erhält man wie in **Listing 3** dargestellt. Ein bestimmtes Paket aktualisiert der folgende Befehl:

```
# func www1.virt.tuxgeek.de call yumcmd 2
update httpd
```

Entsprechend funktioniert ebenfalls der Zugriff auf Methoden anderer Module.

Viele Hosts

Seine Leistungstärke spielt Func (siehe **Abbildung 1**) vor allem dann aus, wenn es darum geht, eine Vielzahl von Systeme

zu verwalten. Hierzu können Sie auf die gewohnte Shell-Dateinamen-Expansion („Globbing“) zurückgreifen. Es ist allerdings darauf zu achten, das Dateinamensmuster in Hochkommas zu setzen.

Das folgende Beispiel fragt alle LDAP-Systeme aus der Domäne

»virt.tuxgeek.de« nach der eingesetzten Kernel-Version ab, ausgenommen sind Systeme die im Hostnamen den String »slave« enthalten:

```
# func --exclude "*slave*" "*ldap*.virt.2
tuxgeek.de" show hardware kernelVersion
ldap1-ms.virt.tuxgeek.de:
2.6.18-128.el5
ldap2-ms.virt.tuxgeek.de:
2.6.18-128.el5
```

Func bietet ebenfalls die Verwaltung von Maschinen über Gruppen an. Hierfür können Sie entweder die Datei »etc/func/groups« manuell editieren und die gewünschten Gruppen anlegen oder die Arbeit an Func übergeben. Das folgende Beispiel erzeugt die Gruppe LDAP-Server mit den beiden Hosts »ldap1« und »ldap2«, zeigt das Ergebnis an und kopiert schließlich die Datei »etc/hosts« auf alle Systeme der Gruppe (**Listing 4**).

Listing 3: Mehrere Methoden pro Modul

```
01 # func www1.virt.tuxgeek.de call yumcmd list_methods
02 {'www1.virt.tuxgeek.de': ['check_update',
03                             'save_config',
04                             'update',
05                             'module_description',
06                             'module_version',
07                             'get_method_args',
08                             'module_api_version',
09                             'list_methods']}
```

Listing 4: LDAP-Verwaltung

```
01 # func "*" group --ah "@ldap:ldap[1-2].virt.tuxgeek.
de"
02 # func "*" group --la
03 Group : ldap
04 Host : ldap1.virt.tuxgeek.de
05 Host : ldap2.virt.tuxgeek.de
06 # func "@ldap" copyfile --file=/etc/hosts
--remotepath=/etc/hosts
```

Listing 5: Func-API

```
01 #!/usr/bin/python
02
03 import func.overlord.client as fc
04
05 results = fc.Client("@ldap").service.status("dirsrv")
06
07 for (host, returns) in results.iteritems():
08     if returns == 0:
09         print host, "LDAP-Server is UP"
10
11     else:
12         print host, "LDAP-Server is DOWN - starting it
up"
13     result = fc.Client(host).service.start("dirsrv")
14     if result[host] == 0:
15         print host, "LDAP-Server is now UP"
16     else:
17         print host, "LDAP-Server is still down, check
manually..."
```

API für mehrere Sprachen

Am interessantesten wird es, wenn man auf die API des Management-Tools zurückgreift. Grundsätzlich können Sie jede Methode, die Sie auf Kommandozeile aufrufen, ebenfalls als API-Funktion ansprechen. Entsprechende Bindings existieren beispielsweise für die Skriptsprache Python. **Listing 5** zeigt ein sehr einfaches Beispiel, das auf allen Systemen der LDAP-Gruppe überprüft, ob der LDAP-Service dort aktiv ist oder nicht und im Fehlerfall den Service nachstartet:

```
# python check-ldap.py
ldap1.virt.tuxgeek.de LDAP-Server is UP
ldap2.virt.tuxgeek.de LDAP-Server is DOWN 2
- starting it up
ldap2.virt.tuxgeek.de LDAP-Server is now UP
```

Feature Overview

Remote API for systems management

Secure XMLRPC over SSL connections

Python API and CLI interfaces
(Python is great for system administration apps!)

SSL cert management tools, via 'certmaster'



Abbildung 1: Überblick der Func-Features in einer Konferenz-Präsentation.

Wer beim Aufruf seines Editors direkt Spaß bekommen hat und nun eigene Module entwickeln möchte – kein Problem. Unter **[3]** findet sich eine recht ausführliche Anleitung, wie sich dies mithilfe von Python bewerkstelligen lässt.

Fazit

Gerade bei der Administration großer System-Landschaften erweist sich Func als ungemein nützlich. Schon durch den Aufruf auf der Kommandozeile lassen sich viele Arbeiten, die man sonst mühsam mittels SSH-Schleifen durchführen müsste, nun viel eleganter erledigen. Durch die umfangreiche API und die Möglichkeit, eigene Module zu entwickeln, welche eventuell fehlende Funktionen nachrüsten, ist das Hack-Potenzial von Func nahezu unerschöpflich. (ofr) ■

Infos

- [1]** Func-Projektseite: [\[https://fedorahosted.org/func/\]](https://fedorahosted.org/func/)
- [2]** Func-Download: [\[https://fedorahosted.org/func/wiki/GetReleases\]](https://fedorahosted.org/func/wiki/GetReleases)
- [3]** Eigene Module mit Func: [\[https://fedorahosted.org/func/wiki/HowToWriteAnModule\]](https://fedorahosted.org/func/wiki/HowToWriteAnModule)
- [4]** Cobbler: [\[https://fedorahosted.org/cobbler\]](https://fedorahosted.org/cobbler)

Der Autor

Thorsten Scherf arbeitet als Senior Consultant für Red Hat EMEA. Er ist oft als Vortragender auf Konferenzen anzutreffen. Wenn ihm neben der Arbeit und Familie noch Zeit bleibt, nimmt er gerne an Marathonläufen teil.



Ben Heyes, 123RF

Traffic Shaping: Niedrige Latenz bei voller Bandbreite mit Linux

Schnell sein, Held sein

Die Reaktionszeit ausgelasteter Internetzugänge steht und fällt mit dem Vorhandensein und der Art des Traffic Shaping im eigenen Gateway. Beim Zusammentreffen des stadtbekannten Wondershaper und einer selbst entwickelten Lösung kann nur einer gewinnen. Martin Stern

Auch die Besitzer eines schnellen oder gar symmetrischen DSL-Anschlusses kennen den Effekt: Während die Bürokollegen dicke E-Mails verschicken, ISO-Images runterladen und Videos schauen, gerät das eigene VoIP-Telefonat zur von Aussetzern und akustischen Artefakten geprägten Tortur – dagegen hilft nur Traffic Shaping [1]. Damit lassen sich die Latenzzeiten zeitkritischer Übertragungen auf ein Drittel bis ein Sechstel verringern. Das Bandbreitenmanagement hilft auch die Reaktionszeit von interaktiven SSH-, Rdesktop und ähnlichen Verbindung deutlich zu optimieren.

Tabelle 1 vergleicht die Ping-Zeiten auf einer symmetrischen 2-MBit/s-Internetstandleitung zwischen einem Debian Lenny ohne Traffic Shaping und der später in diesem Artikel vorgestellten Shaping-Lösung Trafficcontrol-inetgw. Die Nutzdatenrate lag während der Messung jeweils bei knapp 1,9 MBit/s, Gegenstelle war www.google.de.

Der beliebte Wondershaper [2] dient hier funktionell und in Sachen Performance als Referenz. Er verwendet für den Paketversand eine CBQ-QDisc [3] mit drei War-

teschlangen. Dies erfüllt zwei Aufgaben:

- Es verhindert einen Stau im Modem, der normalerweise deshalb entsteht, weil DSL-Router und Kabelmodems Datenpakete wesentlich schneller aus dem LAN empfangen als diese sich ins Internet versenden lassen. Der Router verwaltet daher eine manchmal mehrere Sekunden lange Warteschlange, die sich der Kontrolle des Linux-Kernels entzieht (**Abbildung 1**). Die CBQ-QDisc dagegen gestattet den Paketversand in Richtung Router nur, solange dort eine bestimmte Bandbreite unterschritten ist. Die hat der Administrator so zu konfigurieren, dass sich gerade keine Warteschlange aufbaut. (Der **Kasten „Glossar“** und [4] helfen beim Navigieren durch die etwas komplizierte Begriffswelt.)

- Zum anderen kann der Kernel Pakete auf die drei Warteschlangen den gewünschten Antwortzeiten entsprechend verteilen. Die CBQ-QDisc arbeitet die Warteschlangen nach ihrer Priorität ab und bevorzugt Latenz-empfindliche Pakete damit deutlich. Das hat mehrere Vorteile. So kommen ACK-Pakete

schneller auf den Weg, wodurch ein Upload einen Download weniger ausbremst. Auch SSH- und ICMP-Pakete erfahren bevorzugte Abfertigung, weitere Protokolle lassen sich leicht ergänzen. Sinnvoll ist das natürlich nur, solange keine zusätzliche Warteschlange im Modem existiert.

Für eingehende Datenpakete bietet der Kernel leider deutlich weniger Möglichkeiten als für den Versand – warum auch, schließlich ergibt es keinen Sinn, ein gerade eingetroffenes Paket erst mal in einer Warteschlange zu deponieren. Eine Warteschlange bildet sich im Downstream trotzdem: beim Provider. Während eines Downloads treffen dort Pakete nämlich deutlich schneller ein, als sie sich an den Kunden weiterleiten lassen.

Um den Stau in der Warteschlange beim Provider zu minimieren, bleibt dem Kunden nur, diesen um einen langsameren Versand zu bitten. Dazu bietet der Linux-Kernel die Ingress-QDisc [5]. Sie erlaubt es, nur bestimmte Pakete zu verwerfen. Der Absender einer TCP-Verbindung schickt das Paket daraufhin erneut und drosselt die Transferrate.

Auch der Wondershaper benutzt die Ingress-QDisc, um Pakete zu verwerfen, sobald der eingehende Datenstrom eine vom Administrator gewählte maximale Bandbreite überschreitet. Leider verwirft die Ingress-QDiscs so gelegentlich auch ACK-Pakete eines Uploads oder andere empfindliche Pakete, was die Latenzen für interaktive Anwendungen erhöht.

Einige noch wenig bekannte Optionen im TC-Subsystem des Kernels versprechen jedoch Hilfe, um Pakete differenzierter zu behandeln, als Wondershaper und andere Traffic-Shaping-Software das tun. Das jetzt vorstellte Trafficcontrol-inetgw [6] ist als Komplettlösung zur Latenzoptimierung auf Internetgateways mit dynamischer Overhead-Berechnung und differenziertem Ingress-Policing ausgelegt.

Dynamischer Overhead

Um die Warteschlangen im lokalem Modem und beim Provider zu vermeiden, muss der Linux-Kernel die übertragenen Paketgrößen inklusive aller Overheads kennen. Daraus ergibt sich die Bandbreite, die wirklich über den Internetzugang läuft. Beim Versand eines TCP-Pakets über DSL kommen zu den Nutzdaten die Header für TCP, IP, PPP, PPPoE, Ethernet und ATM. Besonders trickreich ist, dass ein Ethernet-Paket mindestens 64 Bytes groß sein muss und eine ATM-Zelle immer genau 53 Bytes umfasst. Bei ATM gibt es noch eine Besonderheit. Die letzte für ein Ethernet-Paket verwendete Zelle fasst nur 40 Bytes Nutzdaten, alle vorher 48. Das Modem füllt die fehlenden Bytes mit Nullen auf. **Tabelle 2** verdeutlicht die quantitativen Zusammenhänge.

Der Overhead eines Datenpakets ist also nicht konstant, sondern abhängig von der Nutzdatenmenge. Erst neuere Linux-Kernel zusammen mit neueren IProute-Versionen können die Rohdatenmenge im ATM-Netz korrekt berechnen ([8], [9]) – vorausgesetzt man verwendet eine QDisc, die das unterstützt, etwa Hierarchical Token Bucket (HTB, [10]). Dort gibt es den neuen Parameter »linklay«, der die ATM-Rohdatenmenge korrekt berechnet. Für die Root-HTB-Klasse zum Beispiel sieht das so aus:

```
tc class add dev ppp0 parent 1: classid 1:17
htb rate 205kbit ceil 205kbit linklay atm
```

Tabelle 1: Antwortzeiten einer 2-MBit-Leitung

Bedingungen	Ohne Traffic Shaping	Mit Trafficcontrol-inetgw
Leitung frei	30 ms	30 ms
Downlink ausgelastet	280 ms	90 ms
Up- und Downlink ausgelastet	550 ms	90 ms

Auch beim Ingress-Policing (Empfang) ist der Parameter erlaubt:

```
tc filter add dev ppp0 parent ffff:
protocol ip prio 110 u32 match ip src
0.0.0.0/0 flowid :1 police rate 1710kbit
burst 6k linklay atm conform-exceed
continue/ok
```

Außer ATM unterstützt »linklay« zurzeit Ethernet als Übertragungsschicht.

Auch mit betagtem Kernel

Wenn die gewünschte QDisc »linklay« nicht unterstützt, der Kernel zu alt ist oder gar eine exotische Übertragungsschicht im Spiel ist, gelingt es mit zwei Parametern dem Kernel oft, trotzdem die Rohdatenmenge zu berechnen:

- »mpu« gibt die minimale Größe in Bytes an, die ein IP-Paket beim Übertragen hat. Für einen DSL-Anschluss sind das 106 (zwei ATM-Zellen).

- »overhead« gibt die Byte-Anzahl an, die zur Paketgröße des verwendeten Linux-Netzwerkdevice zu addieren ist. Für den PPP- und Ethernet-Header müsste der Wert 22 Bytes sein. Der Autor hat jedoch experimentell ermittelt, dass auf einem ADSL-Anschluss »14« eine kürzere Antwortzeit erzielt.

Hier ein Beispiel für eine Root-HTB-Klasse mit beiden Parametern:

Glossar

Ingress: Eingehender Datenstrom.

Egress: Ausgehender Datenstrom.

Classifying: Pakete mit Filtern priorisieren und einer Warteschlange zuweisen [7].

Shaping: Hält Pakete in einer Warteschlange zurück, sobald der ausgehende Datenstrom (Egress) eine bestimmte Bandbreite überschreitet. Läuft die Warteschlange voll, werden neue Pakete verworfen.

QDisc: Im Linux-Kernel gibt es verschiedene Queueing Disciplines, also Regeln, nach denen ausgehende Pakete Schlange stehen. Einige QDiscs sind hierarchisch aufgebaut und dürfen weitere Klassen und QDiscs enthalten. Man spricht dann von „classful“.

CBQ: Das Class Based Queueing ist einer von mehreren Mechanismen, um priorisierte Pake-

```
tc class add dev ppp0 parent 1: classid 1:17
htb rate 205kbit ceil 205kbit mpu 106
overhead 14
```

Versuche zeigten, dass »mpu« und »overhead« mit »linklay atm« zu kombinieren die Reaktionszeit tendenziell erhöht.

Performance-Vergleiche

Wie in der oberen Hälfte von **Tabelle 3** zu sehen ist, erreicht das Skript »trafficcontrol-inetgw« mit HTB als QDisc und »linklay atm« kürzere Reaktionszeiten bei leicht höherer Bandbreite als der Wondershaper. Die Messungen liefen mit 100 Ping-Paketen bei einem Paket pro Sekunde zu einem möglichst nahen Router im Internet. Deutlich weicht die erzielte Nutzdatenbandbreite von der erlaubten Rohdatenbandbreite ab.

Bei separaten Up- oder Downloads leidet der Wondershaper noch nicht an der fehlenden Overhead-Berechnung, da die durch Shaping begrenzte Bandbreite eine ausreichende Lücke zur Rohbandbreite lässt. Die Differenz entspricht dem Overhead. Die kleinen ACK-Pakete lasten die andere Übertragungsrichtung nicht aus. Gleichzeitige Up- und Downloads machen jedoch den Vorteil einer dynamischen Overhead-Berechnung deutlich. Jetzt kommt der prozentual sehr große Over-

te abzuarbeiten. Dazu gibt es mehrere unterschiedlich priorisierte Queues, die ein Weighted-Round-Robin-Prozess (WRR) abklappert. Zum Beispiel versendet er zehn Pakete von Queue 1, anschließend fünf Pakete von Queue 2, eines von Queue 3 und dann wieder zehn von Queue 1 und so weiter [3].

HTB: Die seit Kernel 2.4.20 fest in den Kernel integriert Discipline „Hierarchical Token Bucket“ funktioniert ähnlich wie Class Based Queueing, aber mit einem Token-Bucket-Filter: Wie CBQ teilt HTB die verfügbare Bandbreite auf mehrere Klassen auf und garantiert jeder Klasse auch bei starker Auslastung eine gewisse Bandbreite. Überschüssige Bandbreiten verteilt HTB auf die anderen Klassen [10].

Policing: Verwirft bestimmte Ingress-Pakete.

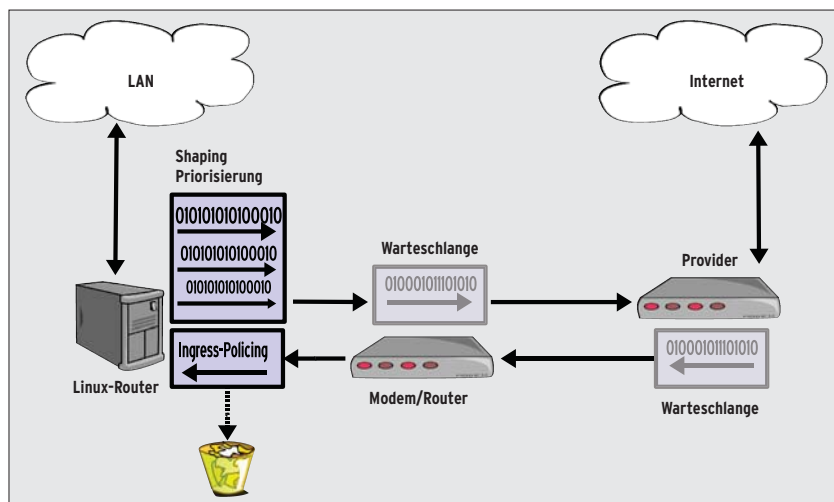


Abbildung 1: Beim Up- und Download bilden sich an mehreren Stellen Warteschlangen. Wer die Reaktionszeit optimieren will, sollten ihr Entstehen beim Provider und im eigenen DSL-Modem oder -Router vermeiden. Der Linux-Kernel hilft mit Shaping und Policing dabei.

head der ACK-Pakete zum Tragen. Ohne dynamische Berechnung des Overheads will das Gateway mehr Daten übertragen, als der Anschluss bewältigen kann, und es bilden sich Warteschlangen.

Die untere Hälfte von **Tabelle 3** zeigt das Ergebnis für einen anderen Internetzugang. Dort ist zu sehen, dass einige ADSL-Zugänge empfindlich auf gleichzeitige Up- und Downloads reagieren und stark schwankende Ergebnisse liefern. Leider hat der Linux-Kernel keine Möglichkeit, sich diesem Verhalten anzupassen.

Die Zeilen 61 bis 101 von **Listing 1** richten Egress-Shaping für den Upload ein. Initial berechnet der Algorithmus ab Zeile 61 ein minimales Quantum für die HTB-Klas-

sen. HTB-Klassen leihen sich gegenseitig Bandbreite, wenn eine Klasse weniger benötigt, als ihr »rate« zuweist, und eine andere gerade Bedarf hat. Quantum gibt dabei die minimale Anzahl auf einmal verleihbarer Bytes an. Es muss größer als die MTU sein, aber so klein wie möglich. Leider berechnet HTB bei den kleinen Upstream-Bandbreiten einiger ADSL-Anschlüsse ein zu kleines Quantum. Das führt zu einer Fehleinschätzung der genutzten Rohdatenbandbreite und damit zu einer Warteschlange im Modem.

Zeile 76 legt die Root-HTB-Klasse an. Sie hat die Aufgabe, die genutzte Upstream-Rohdatenbandbreite per »rate« und »ceil« auf das vom Administrator konfigurierte

Maß zu beschränken. Die Zeilen 82 bis 95 legen drei priorisierte HTB-Klassen für die einzelnen Verbindungstypen an. Die Klasse 1:10 „Low Latency“ hat die höchste Priorität und 1:12 „Second-Order Bandwidth“ die niedrigste. Der Parameter »ceil« bestimmt dabei die Bandbreite, die eine Klasse maximal nutzen darf. Dagegen beschreibt »rate« die für sie reservierte Bandbreite.

Jede Klasse darf immer so viel Bandbreite nutzen, wie ihr »rate« beschreibt. Benötigt sie mehr, kann sie bis zu »ceil« Bandbreite von anderen Klassen leihen – vorausgesetzt dort sind Kapazitäten frei. Wollen mehrere Klassen Bandbreite leihen, erfährt die mit der niedrigeren Priorität eine Bevorzugung.

Ab Zeile 97 legen »tc«-Aufrufe in jeder der drei HTB-Klassen eine Stochastic Fairness Qdisc (SFQ) an **[11]**, was die verschiedenen Verbindungen innerhalb einer Klasse gleichberechtigt behandelt – normalerweise sind Verbindungen mit hohem Bandbreitenbedarf nämlich implizit bevorzugt.

Feingranulare Filter

Die nun folgende Filter-Sektion weist ausgehenden Pakete einer der drei HTB-Klassen zu. Klasse 1:10 „Low Latency“ erhält alle, die eine niedrige Latenz benötigen und voraussichtlich wenig Bandbreite belegen. Das sind kleine TCP-ACK- und alle ICMP-Pakete (Zeilen 106 und 129). Dazu kommen solche, deren IP-Header im TOS-Feld das „Low Latency“-Bit enthält (Zeile 113) oder denen IPtables »10« als Marke verpasst hat (Zeile 117).

Die »u8«- und »u16«-Filterausdrücke ab Zeile 109 prüfen, ob bestimmte Bits im IP-Header gesetzt sind **[12]**. Das erfasst alle Pakete mit weniger als 64 Bytes, in der Regel ACK-Pakete. Da sich die bis zu 80 Bytes großen SACK-Pakete nur aufwändig herausfiltern ließen, erhalten sie in **Listing 2** ab Zeile 22 einfach die Marke »10« aufgedrückt. Die Klasse 1:11 „Bulk and Default“ ist für Datenpakete gedacht, die entweder keine besonders niedrige Antwortzeit benötigen oder für „Low Latency“ zu viel Bandbreite verbrauchen. HTTP-Pakete, die beim interaktiven Surfen entstehen, gehören hierhin. Zeile 133 erfasst alle Pakete, auf die kein anderer Filter angesprochen hat.

Tabelle 2: Variabler Overhead

	TCP-ACK	Ungünstiges TCP-Paket	Optimales TCP-Paket
Linux-PPP-Device			
Nutzdaten	0 Byte	1371 Byte	1418 Byte
TCP-Header	20 Byte	20 Byte	20 Byte
IPv4-Header	20 Byte	20 Byte	20 Byte
Zwischensumme	40 Byte	1411 Byte	1458 Byte
Header			
PPP/PPPoE	8 Byte	8 Byte	8 Byte
Ethernet	14 Byte	14 Byte	14 Byte
Summe	62 Byte	1433 Byte	1480 Byte
Mindestgröße Ethernet-Paket	64 Byte	-	-
ATM			
Anzahl ATM-Zellen	2 Zellen	31 Zellen	31 Zellen
Größe ATM	106 Byte	1643 Byte	1643 Byte
Overhead			
ab PPP-Device	66 Byte	232 Byte	185 Byte

Für Pakete, die ruhig ein paar Millisekunden warten dürfen, gibt es die Klasse 1:12 „Second-Order Bandwidth“, die Zeile 125 erfasst. Die Klasse für Pakete mit der Marke »12« ist vor allem für große Uploads gedacht, die nicht den oft knappen Upstream verstopfen sollen. Um große Uploads vom Surfen zu unterscheiden, markiert Listing 2 ab Zeile 15 Pakete der Typen HTTP und FTP genau dann mit »12«, wenn sie zu einer Verbindung gehören, die schon mehr als 10 MByte Nutzdaten transportiert hat.

In Listing 2 kann der Administrator weitere Protokolle einer der drei HTB-Klassen zuweisen. Wie das Beispiel der großen Uploads zeigt, geht die Granularität von IPtables dabei weit über die von »tc filter« hinaus. Zu beachten ist, dass beim Markieren im IPtables-Skript die letzte Regel greift, während bei »tc filter« in Listing 1 die erste Regel signifikant ist.

Policing

Eine Besonderheit von Trafficcontrol-inetgw ist das differenzierte Ingress-Policing: Statt lediglich beliebige Pakete zu verwerfen, sobald der Downstream die Rohdatenbandbreite überschreitet, analysiert Trafficcontrol-inetgw den Typ des Pakets und verwirft nur solche, auf die es weniger ankommt. Wegen der reduzierten Möglichkeiten des Linux-Kernels beim Ingress, arbeitet Trafficcontrol-inetgw anders als bei Egress ohne Warteschlangen. Jedes Paket durchläuft eine Reihe von Filtern, danach entscheidet Trafficcontrol-inetgw über Annahme oder Verweigerung.

Zeile 143 von Listing 1 prüft für jedes eingehende Paket, ob insgesamt 98 Prozent der vom Administrator konfigurierten Rohdatenbandbreite überschritten sind. Ist das nicht der Fall, nimmt das System das Paket ohne weitere Prüfung an. Die Anweisung »conform-exceed continue/ok« beschreibt, was zu tun ist, wenn der Filter zutrifft (»continue«: weiter mit dem nächsten Filter) oder nicht zutrifft (»ok«: Paket annehmen).

Sind 98 Prozent der Rohdatenbandbreite überschritten, bekommen bis zu 2 Prozent der Pakete ab Zeile 150

die IPtables-Marke »20« aufgedrückt. Diese Pakete verarbeitet dann das Skript »iptables-ingress« (Listing 3) weiter. Es verwirft einzelne Pakete, wenn sie zu einem HTTP- oder FTP-Download gehören, der bereits mehr als 10 MByte Nutzdaten transportiert hat. Die Strategie: Besser schon bei 98 Prozent große Downloads etwas abbremsen, als später das ACK-Paket einer ausgehenden Verbindung verwerfen müssen.

Bei dem Filter ab Zeile 156 kommen nur die Pakete an, die den Filter in Zeile 143 passiert haben. Hier klärt sich, ob diese

Pakete 1 Prozent der Rohdatenbandbreite überschreiten, der Downstream also zu 99 Prozent ausgelastet ist. Wenn ja, gehen ab Zeile 163 einzelne, mindestens 1024 Bytes große Pakete über den Jordan. Das begünstigt ACK-, SSH- und ähnliche Pakete auch dann, wenn die IPtables-Regel in Listing 3 nicht greift, weil keine großen Downloads vorliegen. Erst wenn der Downlink trotzdem zu 100 Prozent ausgelastet ist, verwirft ein Algorithmus ab Zeile 173 auch beliebige Pakete. Das in Listing 1 abgedruckte »trafficcontrol-inetgw« fußt auf Wondershaper, er-

Sicheres und komfortables Login mit Chipkarte statt Benutzername/Passwort.

Sie möchten den neuen Personalausweis für das Login in Ihrer Webanwendung nutzen?

OWOK hilft Ihnen bei der Realisierung.

Mit OWOK steht eine nutzerfreundliche Komplettlösung für das Login per Chipkarte zur Verfügung, die mit geringem Aufwand schnell in beliebige Web-Applikationen eingebunden werden kann.

- Freeware Softwarekomponenten
- Authentifizierung durch Smartcard und PIN
- Maximaler Datenschutz
- Hochsichere Verschlüsselung



Sichere Authentifizierung für Web-Applikationen mit neuem Personalausweis und/oder OWOK-Karte.

Jetzt informieren:

www.reiner-sct.com/owok



OWOK Vortrag
04.03.2011, 12:30 h
iX CeBIT Forum 2011
Halle 3, Stand E08

Listing 1: »trafficcontrol-inetgw«

```

001 #!/bin/bash
002 #Latenzoptimierung für Internetgateways
003 #Martin Stern, 2010 Lizenz: GPL
004 PATH=/usr/sbin:/usr/bin:/sbin:/bin
005
006 #IPTables Bibliotheksverzeichnis gegebenen-
007 #falls für tc überschreiben.
008 #export IPTABLES_LIB_DIR=/usr/local/lib/...
009
010 # Hilfe ausgeben
011 if [ $# -eq 0 -o -gt 3 ]; then
012     echo -e "Usage:"
013     echo -e "$0 [device]"
014     echo -e "  Statistiken anzeigen\n"
015     echo -e "$0 [device] clear"
016     echo -e "  Trafficcontrol deaktivieren\n"
017     echo -e "$0 [device] [downlink] [uplink]"
018     echo -e "  Trafficcontrol einrichten."
019     echo -e "  Downlink/Uplink in Kilobit/s,"
020     exit 1
021 fi
022
023 # Statistic ausgeben
024 if [ $# -eq 1 ]; then
025     echo -e "\nQDisc Statistik:"
026     echo -e "-----"
027     tc -s qdisc ls dev $1
028     echo -e "\nClass Statistik:"
029     echo -e "-----"
030     tc -s class ls dev $1
031     echo -e "\nFilter Statistik:"
032     echo -e "-----"
033     tc -s filter ls dev $1
034     echo -e "\nIngress Filter Statistik:"
035     echo -e "-----"
036     tc -s filter ls parent ffff: dev $1
037     exit
038 fi
039
040 # Konfiguration zurücksetzen
041 if [ $# -eq 2 ]; then
042     tc qdisc del dev $1 root
043     tc qdisc del dev $1 ingress
044     exit
045 fi
046
047 #####
048 #Neue Konfiguration erstellen
049 DEV=$1
050 DOWN=$2
051 UP=$3
052
053 #Alte Konfiguration löschen
054 tc qdisc del dev $DEV root &> /dev/null
055 tc qdisc del dev $DEV ingress &> /dev/null
056
057
058 #####
059 #Uplink konfigurieren (egress)
060
061 #Minimales Quantum festlegen
062 SmallQ=""; BigQ=""
063 z=$(( ${UP}/10*128/3 ))
064 if [ $z -lt 1500 ]; then
065     SmallQ="quantum 1500"
066 fi
067 z=$(( ${UP}/10*128 ))
068 if [ $z -lt 1500 ]; then
069     BigQ="quantum 1500"
070 fi
071
072 #Root HTB QDisc
073 tc qdisc add dev $DEV root handle 1: htb\
074     default 11
075
076 #Root HTB Klasse
077 tcc="tc class add dev $DEV parent"
078 $tcc 1: classid 1:1 htb \
079     rate ${UP}kbit ceil ${UP}kbit \
080     linklay atm $BigQ
081
082 #HTB Klasse 1:10 "Low Latency"
083 $tcc 1:1 classid 1:10 htb \
084     rate $(( ${UP}/3 ))kbit ceil ${UP}kbit \
085     prio 0 linklay atm $SmallQ
086
087 #HTB Klasse 1:11 "Bulk and Default"
088 $tcc 1:1 classid 1:11 htb \
089     rate $(( ${UP}/3 ))kbit ceil ${UP}kbit \
090     prio 1 linklay atm $SmallQ
091
092 #HTB Klasse 1:12 "Second-Order Bandwidth"
093 $tcc 1:1 classid 1:12 htb \
094     rate $(( ${UP}/3 ))kbit ceil ${UP}kbit \
095     prio 2 linklay atm $SmallQ
096
097 #Stochastic Fairness Qdisc für alle drei
098 tcq="tc qdisc add dev $DEV parent"
099 $tcq 1:10 handle 10: sfq perturb 10
100 $tcq 1:11 handle 11: sfq perturb 10
101 $tcq 1:12 handle 12: sfq perturb 10
102
103 #Pakete den HTB Klassen zuweisen.
104 #Der erste passende Filter gilt.
105 #ACK -> "Low Latency"
106 tcf="tc filter add dev $DEV parent"
107 $tcf 1: protocol ip prio 22 u32 \
108     match ip protocol 6 0xff \
109     match u8 0x05 0xf at 0 \
110     match u16 0x0000 0xffc0 at 2 \
111     flowid 1:10
112
113 #TOS Minimum Delay -> "Low Latency"
114 $tcf 1: protocol ip prio 10 u32 \
115     match ip tos 0x10 0xff flowid 1:10
116
117 #IPTables fwmark 10 -> "Low Latency"
118 $tcf 1: protocol ip prio 16 \
119     handle 10 fw flowid 1:10
120
121 #IPTables fwmark 11 -> "Bulk and Default"
122 $tcf 1: protocol ip prio 18 \
123     handle 11 fw flowid 1:11
124
125 #IPTables fwmark 12 -> "Second-Order B.w."
126 $tcf 1: protocol ip prio 20 \
127     handle 12 fw flowid 1:12
128
129 #ICMP -> "Low Latency"
130 $tcf 1: protocol ip prio 24 u32 \
131     match ip protocol 1 0xff flowid 1:10
132
133 #DEFAULT -> "Bulk and Default"
134 $tcf 1: protocol ip prio 50 u32 \
135     match ip dst 0.0.0.0/0 flowid 1:11
136
137 #####
138 #Downlink konfigurieren (ingress)
139
140 #Ingress Root Qdisc anlegen.
141 tc qdisc add dev $DEV handle ffff: ingress
142
143 #Wenn über 98% Auslastung ...
144 $tcf ffff: protocol ip prio 110 u32 \
145     match ip src 0.0.0.0/0 flowid :1 \
146     police rate $(( ${DOWN}*98/100 ))kbit \
147     burst 6k linklay atm \
148     conform-exceed continue/ok
149
150 #... dann überzählige Pakete markieren.
151 $tcf ffff: protocol ip prio 111 u32 \
152     match u32 0 0 flowid :1 \
153     action ipt -j MARK --set-mark 20 \
154     action continue > /dev/null
155
156 #Wenn mehr als 1 weiteres % Auslastung ...
157 $tcf ffff: protocol ip prio 120 u32 \
158     match ip src 0.0.0.0/0 flowid :1 \
159     police rate $(( ${DOWN}*1/100 ))kbit \
160     burst 8k linklay atm \
161     conform-exceed continue/ok
162
163 #... dann große Pakete verwerfen.
164 $tcf ffff: protocol ip prio 121 u32 \
165     match u16 0x0400 0x0400 at 2 flowid :1 \
166     police rate 0kbit burst 2k linklay atm \
167     conform-exceed drop/drop
168
169 $tcf ffff: protocol ip prio 122 u32 \
170     match u16 0x0800 0x0800 at 2 flowid :1 \
171     police rate 0kbit burst 2k linklay atm \
172     conform-exceed drop/drop
173
174 #Wenn noch 1% Auslastung => verwerfen.
175 $tcf ffff: protocol ip prio 130 u32 \
176     match ip src 0.0.0.0/0 flowid :1 \
177     police rate $(( ${DOWN}*1/100 ))kbit \
178     burst 10k linklay atm \
179     conform-exceed drop/drop

```

weitere ihn aber in einigen Punkten. Das Skript benötigt mindestens Kernel 2.6.24 [8] und IProute 20080725. Zur Installation lädt der Administrator die vier in **Tabelle 4** genannten Dateien von [6] herunter und kopiert sie nach »/usr/local/bin/«. Dann erweitert er seine Firewall um die Regeln aus **Listing 2**. Liegt sie als Bash-Skript vor, genügt die folgende Ergänzung:

```
DEV="Device"
source /usr/local/bin/iptables-egress
```

Danach empfiehlt sich ein Testlauf mit »trafficcontrol-inetgw Device geschätzte_Uplink-Rohdatenbandbreite geschätzte_Downlink-Rohdatenbandbreite«. Kommt keine Fehlermeldung, hat alles geklappt. Dann kann man mit »trafficcontrol-inetgw Device« Statistiken abrufen und das System mit »trafficcontrol-inetgw Device clear« wieder dekonfigurieren.

Mit einigen Versionen von IPtables, zum Beispiel 1.4.2 in Debian Lenny, tritt das Problem auf, dass Tc aus dem IProute-Paket die IPtables-Bibliotheken nicht findet, da sie nach »/lib/xtables« verschoben und umbenannt sind. Wer »trafficcontrol-inetgw« aufruft, erhält dann bei Zeile 153 die Meldung »/lib/iptables/libipt_mark.so: cannot open shared object file: No such file or directory failed to find target MARK«. Als Workaround kann der Administrator das »/lib/iptables«-Verzeichnis einer älteren Version nach »/usr/local/lib/oldiptables« kopieren und das »tc« in Zeile 8 mit der Umgebungsvariablen »IPTABLES_LIB_DIR« bekanntgeben.

Die Suche nach einem geeigneten Partner

Zum Testen des Setups benötigt der Administrator einen Host zum Anpingen, der nur wenige Hops entfernt im Internet steht. Die Antwortzeit der Ping-Pakete soll die Latenz des Internetzugangs messen und möglichst wenig vom sonstigem Verkehr im Internet beeinflusst sein. Wenn der Point-to-Point-Partner beim eigenen Provider nicht auf Ping reagiert, lässt der Admin sich mit »mtr -nrc3 www.google.de« alle Router auf dem Weg zu Google anzeigen. Anschließend probiert er nacheinander, ob sie auf Ping reagieren. Der erste Router, der ohne Paketverluste auf »ping IP-Nummer« antwortet, empfiehlt sich als Gegenstelle.

Das Skript »trafficcontrol-searchbw« (hier nicht abgedruckt, aber bei [6] zu finden) unterstützt den Administrator beim Suchen der Rohdatenbandbreite. Statt vieler Testläufe mit »trafficcontrol-inetgw« und anschließendem Ping reicht ein Aufruf zum Untersuchen eines Bandbreitenintervalls. Zuerst ergänzt er in Zeile 18 des Skripts die gerade ermittelte IP-Nummer in der Variablen »PingHost«. Nun startet er einige große Downloads, um den Downlink konstant auszulasten. Mit dem Aufruf von

```
trafficcontrol-searchbw Device DOWN
Untergrenze-Obergrenze Schrittwerte
Ping-Anzahl Upstream
```

startet eine erste Übersichtsmessung. Die einzelnen Parameter in korrekter Reihenfolge erklärt **Tabelle 5**. Ein Beispiel: »trafficcontrol-searchbw ppp0 DOWN 1300-1800 50 5 210«. Das Intervall von 1300 bis 1800 KBit/s Downlink-Rohdatenbandbreite entspricht dabei einer ersten Schätzung des Anwenders. Die 210 KBit/s für den Upstream sind ebenfalls geraten, aber bei dieser Messung nicht wichtig. **Abbildung 2** zeigt das Ergebnis.

Nach der Übersichtsmessung startet der Administrator eine Detailmessung des interessanten Intervalls. Wie in **Abbildung 3** zu sehen, reduziert er dabei die Schrittwerte und erhöht die Anzahl der Pings:

Tabelle 3: Antwortzeiten zweier ADSL-Leitungen

		Default	Wondershaper	Trafficcontrol-inetgw
Internetzugang mit 1,4 MBit/s Downstream				
Leitung frei	Antwortzeit	ø 49 ms	ø 49 ms	ø 49 ms
Erlaubte Rohdatenbandbreite	Download	-	1490 KBit/s	1710 KBit/s
	Upload	-	190 KBit/s	205 KBit/s
Auslastung des Downlinks	Download	1470 KBit/s	1470 KBit/s	1470 KBit/s
	Antwortzeit	470 ms	86 ms	88 ms
	Paketverlust	0 %	0 %	0 %
Auslastung des Uplinks	Upload	200 KBit/s	186 KBit/s	172 KBit/s
	Antwortzeit	1350 ms	100 ms	92 ms
	Paketverlust	2 bis 5 %	0 %	0 %
Gleichzeitige Auslastung des Up- und Downlinks	Download	1455 KBit/s	1235 KBit/s	1450 KBit/s
	Upload	135 KBit/s	145 KBit/s	150 KBit/s
	Antwortzeit	450 ms	420 ms	165 ms
	Paketverlust	3 %	4 %	0 %
Internetzugang mit 2 MBit/s Downstream				
Leitung frei	Antwortzeit	ø 19 ms	ø 19 ms	ø 19 ms
Erlaubte Rohdatenbandbreite	Download	-	2060 KBit/s	2360 KBit/s
	Upload	-	160 KBit/s	175 KBit/s
Auslastung des Downlinks	Download	2040 KBit/s	1990 KBit/s	1990 KBit/s
	Antwortzeit	920 ms	95 ms	85 ms
	Paketverlust	0 %	3 %	0 %
Auslastung des Uplinks	Upload	170 KBit/s	154 KBit/s	152 KBit/s
	Antwortzeit	8700 ms	88 ms	62 ms
	Paketverlust	0 %	0 %	0 %
Gleichzeitige Auslastung des Up- und Downlinks	Download	745 bis 1800 KBit/s	900 KBit/s	1200 bis 1690 KBit/s
	Upload	135 bis 160 KBit/s	150 KBit/s	125 bis 130 KBit/s
	Antwortzeit	1140 bis 3440 ms	3200 ms	520 bis 1230 ms
	Paketverlust	1 bis 4 %	0 bis 3 %	0 %

Tabelle 4: Benötigte Skripte

Dateiname	Funktion
trafficcontrol-inetgw	Konfiguriert des Bandbreitenmanagement im Linux-Kernel
iptables-egress	IPtables-Regeln für die Egress-Prioritisierung
iptables-ingress	IPtables-Regeln für das Ingress-Policing
trafficcontrol-searchbw	Sucht die optimale Rohdatenbandbreiten eines Internetzugangs

```
fwl:~# trafficcontrol-searchbw ppp0 DOWN 1300-1800 50 5 210
Allowed Bandw.: Erlaubte Rohdatenbandbreite (ATM)
Used Bandw.: Verwendete Nutzdatenbandbreite (IP)
```

Allowed (Used)	Allowed (Used)	Avg. RTT
Downstream Bandw.	Upstream Bandw.	
[kBit/s]	[kBit/s]	[ms]
Inaktiv (1345)	Inaktiv (32)	301
1300 (1143)	210 (34)	60
1350 (1218)	210 (41)	58
1400 (1218)	210 (37)	56
1450 (1294)	210 (41)	68
1500 (1103)	210 (34)	64
1550 (1390)	210 (43)	60
1600 (1443)	210 (39)	64
1650 (1481)	210 (41)	64
1700 (1519)	210 (39)	93
1750 (1376)	210 (39)	412
1800 (1369)	210 (26)	499

```
fwl:~#
```

Abbildung 2: Eine Übersichtsmessreihe von erzielten Nutzbandbreiten und Ping-Zeiten. Deutlich zu sehen ist der Latenzsprung zwischen 1700 und 1750 KBit/s – in diesem Bereich liegt die optimale Bandbreite.

»trafficcontrol-searchbw ppp0 DOWN 1650-1750 20 100 210«.

Da diese Messung recht lange läuft, muss der Admin darauf achten, dass die Downloads stabil bleiben. Die jeweils gemessene Nutzdatenbandbreite sieht er in der Ausgabe. Bricht sie ein, muss er die Messung wiederholen. Außerdem darf der Uplink nicht nennenswert anderweitig genutzt sein. Dies erhöht die Latenz und verfälscht damit das Ergebnis. Auch hier sollte er in der Ausgabe die Upstream-Bandbreite prüfen und die Messung gegebenenfalls wiederholen.

Nun kann der Administrator seine optimale Rohdatenbandbreite für den Down-

link ablesen. Es ist die letzte Zeile, die noch eine akzeptable Antwortzeit unter 100 Millisekunden erzielt. Anschließend wiederholt er die Messungen für den Uplink. Steht kein FTP-Server zum Hochladen der Daten ins Internet bereit, kann er sich selbst E-Mails mit sehr großen Attachments schicken. Die Messung muss aber abgeschlossen sein, bevor das System die E-Mails wieder herunterlädt. Ein Beispielkommando für eine Uplink-Übersichtsmessung lautet: »trafficcontrol-searchbw ppp0 UP 150-250 10 5 1710«. Um das System beim Booten zu konfigurieren, fügt der Administrator das Kommando »trafficcontrol-inetgw« zu einem

```
fwl:~# trafficcontrol-searchbw ppp0 DOWN 1650-1750 20 100 210
Allowed Bandw.: Erlaubte Rohdatenbandbreite (ATM)
Used Bandw.: Verwendete Nutzdatenbandbreite (IP)
```

Allowed (Used)	Allowed (Used)	Avg. RTT
Downstream Bandw.	Upstream Bandw.	
[kBit/s]	[kBit/s]	[ms]
Inaktiv (1481)	Inaktiv (33)	486
1650 (1423)	210 (38)	75
1670 (1451)	210 (38)	76
1690 (1467)	210 (38)	79
1710 (1471)	210 (38)	85
1730 (1478)	210 (36)	180
1750 (1482)	210 (34)	482

```
fwl:~#
```

Abbildung 3: Detailmessreihe für das in Abbildung 2 ermittelte Rohbandbreiten-Intervall. Die optimale Rohbandbreite liegt bei 1710 KBit/s, da hier die Latenz bei ausgelastetem Downlink noch unter 100 ms liegt. Damit liefert der Router 1471 KBit/s Nutzdaten. Die Ergebnisse sind genauer als in Abbildung 2, da hier wesentlich länger gemessen wurde.

seiner Startskripte hinzu. Bei einem DSL-Zugang zum Beispiel empfiehlt es sich

```
#!/bin/sh
case "$PPP_IFACE" in
    Device) /usr/local/bin/trafficcontrol
    -inetgw.sh
Device Up Down
;;
esac
```

unter »/etc/ppp/ip-up.d/trafficcontrol« abzulegen. Nun kann er auch seine Firewall um die Regeln aus IPTables-ingress ergänzen. Diese müssen vor den üblichen Regeln für das Connection Tracking stehen, damit IPTables-ingress alle Pakete einer Verbindung auswerten kann. Dazu ergänzt der Admin »source /usr/local/bin/iptables-ingress« in seinem Firewallskript. Diese Regeln werden erst nach der Messung aktiv, da sie Downloads ab dem 10. MByte um zirka zwei Prozent ausbremsen. Dies hätte die maximale Bandbreite verfälscht.

Kurze Latenzen bei beidseitiger Last

Beim Traffic Shaping auf einem Linux-Gateway ist es kein in Stein gemeißeltes Gesetz, dass bei gleichzeitig ausgelastetem Up- und Downstream die Antwortzeiten in den Keller gehen. Denn neue Features des Linux-Kernels und des IP-route-Pakets erlauben die Rohdatenbandbreite dynamisch zu berechnen.

Auch beim oft vernachlässigten Ingress-Policing gelingt es, zwischen Verbindungstypen zu differenzieren. Da der Kernel auf diese Weise keine ACK-Pakete

Listing 2: »iptables-egress«

```
01 #IPTables-Regeln (egress)
02 #Letzte passende Regel gilt.
03
04 /sbin/iptables -t mangle -F
05 iptl="/sbin/iptables -A POSTROUTING "
06 ipt="$iptl -t mangle -o $DEV -p tcp"
07
08 $ipt -m multiport --port 3389 \
09 -j MARK --set-mark 10 #RDesktop
10 $ipt -m multiport --port 5900:5902 \
11 -j MARK --set-mark 10 #VNC
12 $ipt -m multiport --port smtp,ssmtp \
13 -j MARK --set-mark 12 #Mailversand
14
15 #Uploads > 10MB => "Second-Order Bandwidth"
16 $ipt -m connbytes --connbytes-dir both \
17 --connbytes-mode bytes \
18 --connbytes 10485760: \
19 -m multiport --port www,https,ftp-data \
20 -j MARK --set-mark 12
21
22 #Große TCP-ACK-Pakete => "Low Latency"
23 $ipt -m conntrack --ctstate ESTABLISHED \
24 -m length --length 0:128 \
25 --tcp-flags ACK ACK \
26 -j MARK --set-mark 10
```

Listing 3: »iptables-ingress«

```
01 # IPTables Regeln (ingress)
02 for CHAIN in FORWARD INPUT; do
03 /sbin/iptables -A $CHAIN -i $DEV -p tcp \
04 -m connbytes --connbytes-dir both \
05 --connbytes-mode bytes \
06 --connbytes 10485760: \
07 -m multiport --port www,https,ftp-data\
08 -m mark --mark=20 \
09 -m length --length 500: \
10 -j DROP
11 done
```

von Uploads mehr zu verwerfen braucht, beeinflussen sich die beiden Übertragungsrichtungen weniger. Lange Downloads lassen sich um wenige Prozent abbremser, was anderen Verbindungen genug Spielraum eröffnet.

Das hier vorgestellte Trafficcontrol-inetgw macht sich diese Mechanismen erfolgreich zu eigen. Wie Wondershaper benutzt es ein klassenbasiertes Queueing, jedoch keine CBQ-QDisc, sondern HTB-QDisc, die einfacher zu verstehen ist. (jk/ofr) ■

Infos

- [1] Schwerpunkt über Traffic Shaping, Linux-Magazin 02/05
- [2] Wondershaper: [\[http://lartc.org/wondershaper/\]](http://lartc.org/wondershaper/)
- [3] CBQ-QDisc: [\[http://lartc.org/howto/lartc.qdisc.classful.html#AEN939\]](http://lartc.org/howto/lartc.qdisc.classful.html#AEN939)
- [4] Begriffsbestimmung: [\[http://lartc.org/howto/lartc.qdisc.terminology.html\]](http://lartc.org/howto/lartc.qdisc.terminology.html)
- [5] Ingress-QDisc: [\[http://lartc.org/howto/lartc.adv-qdisc.ingress.html\]](http://lartc.org/howto/lartc.adv-qdisc.ingress.html)

- [6] Trafficcontrol-Skripte: [\[ftp://www.linux-magazin.de/pub/listings/magazin/2010/12/TC\]](ftp://www.linux-magazin.de/pub/listings/magazin/2010/12/TC)
- [7] „Classifying packets with filters“: [\[http://lartc.org/howto/lartc.qdisc.filters.html\]](http://lartc.org/howto/lartc.qdisc.filters.html)
- [8] „Final ADSL-optimizer patch series“: [\[http://kerneltrap.org/mailarchive/linux-netdev/2008/4/9/1386524\]](http://kerneltrap.org/mailarchive/linux-netdev/2008/4/9/1386524)
- [9] „Optimization of TCP/IP Traffic Across ADSL“: [\[http://www.adsl-optimizer.dk\]](http://www.adsl-optimizer.dk)
- [10] HTB: [\[http://luxik.cdi.cz/~devik/qos/htb/\]](http://luxik.cdi.cz/~devik/qos/htb/)
- [11] SFQ-Manpage: [\[http://lartc.org/manpages/tc-sfq.html\]](http://lartc.org/manpages/tc-sfq.html)
- [12] U32-Classifer: [\[http://b42.cz/notes/u32_classifier/\]](http://b42.cz/notes/u32_classifier/) und [\[http://lartc.org/lartc.html#LARTC.ADV-FILTER.U32\]](http://lartc.org/lartc.html#LARTC.ADV-FILTER.U32)

Tabelle 5: Trafficcontrol-searchbw

Position	Parameter	Erklärung
1	Device	Netzwerkdevice des Internetzuganges
2	UP/DOWN	Intervall über Up- oder Downstream
3	Untergrenze, Obergrenze	Untergrenze und Obergrenze der zu messenden Rohbandbreite in KBit/s
4	Schrittweite	Schrittweite zwischen den Messungen in KBit/s. Empfohlen: 5 bis 50
5	Pinganzahl	Anzahl Pings pro Messung. Empfohlen: 5 (Übersicht), 100 (Details)
6	Andere Bandbreite	Roh-Bandbreite der anderen Übertragungsrichtung in KBit/s

Der Autor

Martin Stern ist selbstständiger Netzwerk-Administrator. Zwei seiner Schwerpunkte sind Open-VPN-Systeme zur Filialvernetzung mit hochverfügbaren Knoten und Nagios-Monitoringsysteme. Gerne arbeitet er aber auch mit anderen Servern, Routern und Firewalls unter Debian GNU/Linux.

1. Lernen Sie!

Ja, „training-on-the-job“, oft praktiziert, aber nicht überzeugend. Denn die Kollegen haben nie Zeit für echte Erklärungen, außerdem werden „Neue“ sofort von dem vereinnahmt, was im Unternehmen schon seit Ewigkeiten tradiert wird. Warum gibt's seit 2000 Jahren Schulen und Universitäten? „LERNEN“ ist eine vollwertige Tätigkeit, auf die man sich konzentrieren muß, die man nicht 'mal eben so nebenbei tun kann, und die immer auch eine Prise „Erneuerung“ beinhalten sollte!

2. Ineffiziente Arbeit nicht akzeptieren!

Je spezialisierter Sie arbeiten, desto weniger echte, fachliche Kollegen haben Sie in Ihrem eigenen Unternehmen. Wir stellen deshalb Gruppen zusammen, in denen Sie neben hilfsbereiten Kollegen mit ähnlichen Kenntnissen an IHREM Projekt arbeiten. Und ständig ist ein fachlicher Berater anwesend.

„Guided Coworking“ nennen wir das, und es könnte DIE Lösung für so manches Projekt sein, das in Ihrer Firma „hakt“.

3. Hintergrund

Wer den riesigen OpenSource-Baukasten schnell beherrschen muß, geht zu einer unserer über 100 Schulungen. Wer das bereits kann, aber schneller mit seinen Projekten vorankommen will, der kommt mit seiner Arbeit zum Guided Coworking.

Wir sind eine der erfolgreichsten Schulungseinrichtungen im gesamten Bereich „OpenSource“ - sowohl für Admins, als auch für Entwickler.

Siehe www.linuxhotel.de



SMB Traffic Analyzer SMBTA

Tanzlehrer

Der SMB Traffic Analyzer ist ein VFS-Modul für Samba 3.6, das eine Echtzeitanalyse des Datendurchsatzes im SMB-Netzwerk ermöglicht. Es bringt eigene Analyse-Tools mit oder bereitet die gewonnenen Daten mit RRD-Tool grafisch auf. Thomas Drilling

Der SMB Traffic Analyzer ist ein als VFS-Modul realisiertes Tool, das einen Samba-/CIFS-Server in die Lage versetzt, Traffic-Statistiken im Samba-Netzwerk aufzuzeichnen. Der SMBTA-Daemon (SMBTAD) speichert diese in einer Datenbank und macht sie unter anderem via SQL verfügbar. Die Architektur ist übersichtlich und besteht aus einem Modul im Samba-VFS (Virtual File System), einem Daemon sowie einem Set von Client-Tools (»smbttools«) zum Auswerten und Visualisieren. SMBTA nutzt eine vorhandene SQL-Datenbank (»sqlite3«) zum Speichern der Traffic-Daten. SMBTA-Entwickler und Erfinder Holger Hettrich von Novell arbeitet seit der

SambaXP-Conference 2007 in Göttingen an seinem Samba Traffic Analyzer und konnte bisher in Fachkreisen, etwa beim Samba-Team, viel Aufmerksamkeit erregen, was ihm eine Reihe von gut besuchten Vorträgen etwa auf der Samba Xperience und bei anderen Gelegenheiten in Deutschland und den USA bescherte. Inzwischen engagiert sich sogar Arbeitgeber Novell insofern, dass Hettrich einen Teil seiner Arbeitszeit auf SMBTA verwenden kann.

Leider fehlt es in der Öffentlichkeit noch am breit angelegten Interesse, was aber bisher primär daran lag, dass jegliches Nutzer-Feedback direkt von und über Holger Hettrich abgewickelt wurde und nicht über die passenden Mailinglisten oder sonstige Kanäle, was schlicht eine

Frage gekonnter Öffentlichkeitsarbeit ist. Deshalb ist auch die Anzahl an Referenz-Nutzern derzeit noch überschaubar. Das dürfte sich aber mit der kommenden Samba-Version 3.6 ändern, die den SMB Traffic Analyzer als offiziellen Bestandteil von Samba enthalten wird [\[1\]](#).

Spezialisiert

Zwar könnten versierte Admins auch mit einem gewöhnlichen Portsniffer ausgewählte Schnittstellen etwa auf typischen Netbios-Traffic abhören, SMBTA konzentriert sich aber ausschließlich auf Samba-Traffic und ermöglicht das Erstellen umfangreicher und aussagekräftiger Statistiken, weil das Tool quasi echtes Datamining betreibt, in dessen Zentrum die Sqlite-Datenbank steht. So ist es etwa mit SMBTA möglich, Analysen gezielt auf



einzelne Benutzer, Shares oder Dateien zu beschränken. Auf Client-Seite stehen im Prinzip drei Tools zum Auswerten zur Verfügung, die Bestandteil des Pakets »smbttools« sind. Zum einem ermöglicht ein RRD-Treiber mithilfe von RRD-Tool (Round Robin Database [2]), sich in Echtzeit an den SMBTA-Daemon zu hängen, was die grafische Aufbereitung der Traffic-Daten oder deren Weiterverarbeitung etwa in Perl-Skripten ermöglicht, wahlweise via IP oder Unix-Domain-Sockets. Das Werkzeug »smbtaquery« kann die Datenbank via XML auslesen. Wer eher auf visuelle Effekte steht, kann mit dem »smbtmonitor« auch ganz ohne SQL-Kenntnisse auf die gespeicherten Daten zugreifen.

Bezugsquellen

Wer SMBTA selbst einsetzen möchte, kann entweder die Sourcen der aktuellen Version 1.1.2 von [3] herunterladen oder sich von Holger Hettrichs Wordpress-Blog Binaries für Open Suse besorgen [4]. Allgemeine Informationen zu SMBTA finden sich unter [5]. Aufgrund der Tatsache, dass das Installieren von SMBTA Backports auf Samba 3.6 erfordert, fahren Anwender mit dem RPM-Binary oder dem One-Click-Installer für Open Suse 11.3 am besten. Selbstverständlich ist auch das Auschecken von SMBTA per Git möglich [6].

Last but not least können Suse-Nutzer auch die Paketquelle [7] in Yast einbinden und SMBTA mit dem Paketmanager installieren (Abbildung 1 und 2). Noch stressfreier gelingt das Ausprobieren von SMBTA mit der von Hettrich geschnürten Appliance »Stresstest«, die aktuell

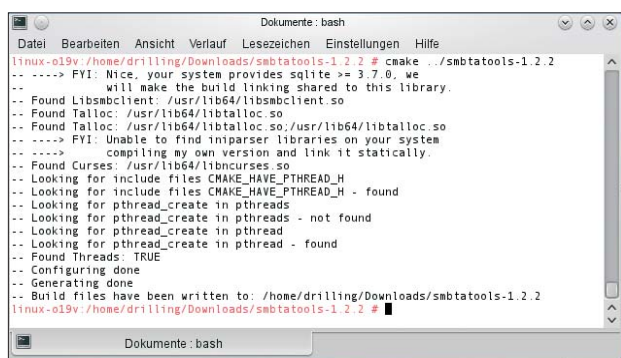


Abbildung 1: Der SMBTA-Daemon lässt sich relativ einfach aus den Quellen übersetzen oder über das angegebene Repository als Binary mit Yast installieren.

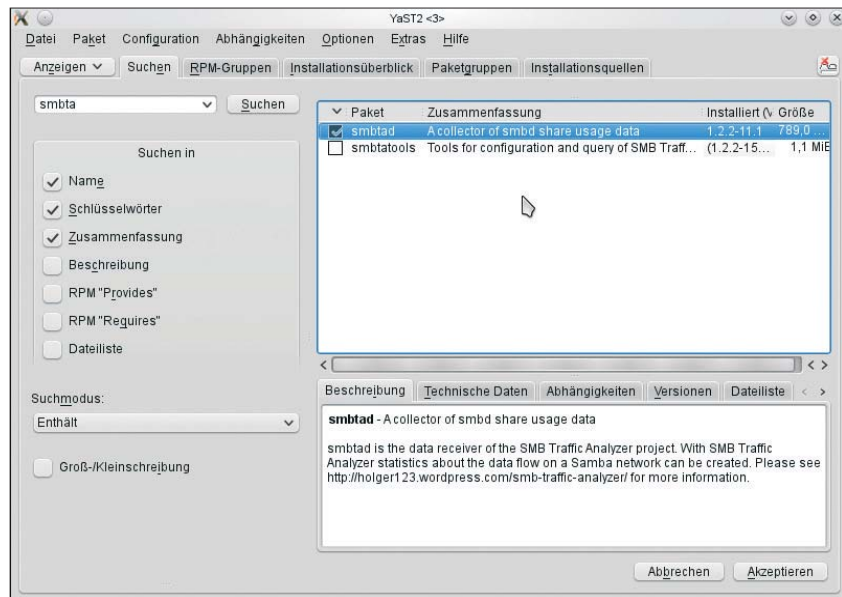


Abbildung 2: Open-Suse-Nutzer können sowohl SMBTA als auch die SMBTA-Tools direkt aus den angegebenen Repositories installieren.

in der Version 0.0.2 vorliegt [8] (siehe Kasten).

Wer SMBTA schnellstmöglich ausprobieren, dazu aber keinen dedizierten Samba-Server aufsetzen möchte, muss SMBTA und die »smbttools« aus den Quellen bauen. Damit das klappt, müssen »cmake«, »libsmbclient-devel«, »libtalloc-devel« und »ncurses-devel« installiert sein. Außerdem müssen die Datenbankumgebung Sqlite3 und die zugehörigen Devel-Pakete installiert sein. Schließlich ist zu prüfen, ob »libxslt« installiert ist, was bei Open Suse per Default der Fall sein sollte. Nun entpackt man zunächst die Sourcen »smbttools-1.2.2.tar.bz2« und wechselt in das entstandene Verzeichnis. Der passende Aufruf von »cmake« ausgehend vom Build-Verzeichnis konfiguriert das Paket für die Übersetzung:

```
cmake ../?
smbttools-1.2.2
```

»make« und »make install« kompilieren das Paket und kopieren die Programme in den passenden Ort. Zum Starten des Daemons genügt »smbtd -u -n«, womit Daemon (»u«) und Client (»n«)

über Unix Domain Sockets kommunizieren. Der SMBTA-Daemon hat primär die Aufgabe, die SQL-Datenbank mit den Daten zu füttern, die er vom VFS-Modul empfängt. Außerdem ist er für das Abhandeln sämtlicher Client-Anfragen an die Datenbank verantwortlich. Möchte der Admin nun jeglichen Traffic auf einer ausgewählten Freigabe aufzeichnen, muss er lediglich in der betreffenden Share-Definition das VFS-Modul wie folgt laden.

```

vfs objects = smb_traffic_analyzer
smb_traffic_analyzer:protocol_version = v2
smb_traffic_analyzer:mode = unix_domain_socket
  
```

wobei sich der letzte Parameter vom Administrator an die eigenen Bedürfnisse anpassen lässt (siehe auch Abbildung 3). Soll die Kommunikation zum Beispiel über TCP/IP laufen, sähe eine entsprechende Share-Definition etwa so aus:

```

vfs objects = smb_traffic_analyzer
smb_traffic_analyzer:protocol_version = v2
smb_traffic_analyzer:host = localhost
smb_traffic_analyzer:port = 3490
  
```

Der SMBTAD-Daemon wäre in diesem Fall mit

```
smbtd -i 3490 -p 3491
```

zu starten und wartet damit auf Anfragen via Port 3490 an das VFS-Modul und behandelt Client-Anfragen auf Port 3491.

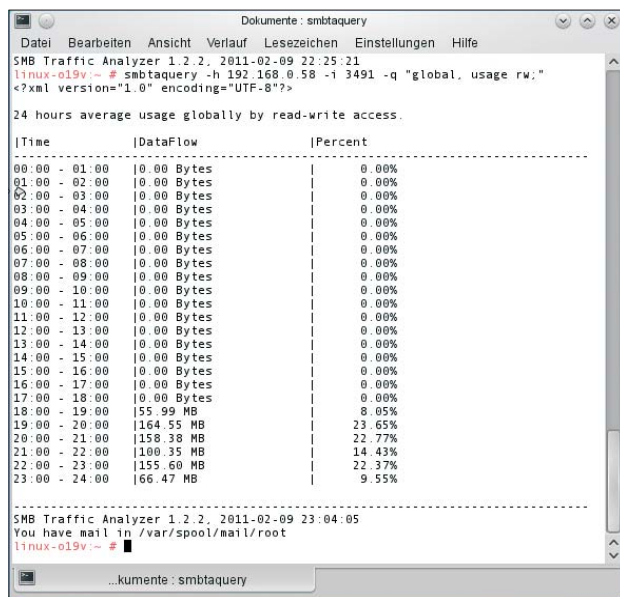


Abbildung 3: Der Usage-Parameter ermöglicht ein zeitlich gewichtetes Auswerten der Aktivität etwa auf einer bestimmten Freigabe.

Per Default legt »smbtd« seine SQLite-Datenbank unter »\$HOME/.smbtd/stadddb« an, es sei denn, die Datenbank existiert bereits. Eine ganze Reihe weiterer Parameter lassen sich mit »smbtd -help« in Erfahrung bringen. Eine ausführliche Erläuterung sämtlicher Parameter findet sich darüber hinaus in der hervorragenden Dokumentation. Selbstverständlich ist es auch möglich, alle benötigten Konfigurationsparameter in einer Konfigurationsdatei »/etc/smbtd.conf« zusammenzufassen, im typischen Ini-Datei-Format, mit »#« als Kommentarzeichen.

Abfrage-Tool

Wie erwähnt, stehen an Client-Tools ein RRD-Treiber, das Befehlszeilenwerkzeug »smbtaquery« sowie »smbtamonitor« zur Verfügung. Selbstverständlich können versierte Admins der Traffic-Datenbank auch mit beliebigen SQL-Werkzeugen ins Innerste schauen, komfortabler ist aber »smbtaquery«, das speziell auf das Datenbank-Setup von SMBTA ausgerichtet ist und eine Reihe vorkonfigurierter Abfragen zum Erzeugen statistisch verwertbarer Daten mitbringt. Smbtaquery erzeugt dazu XML-Output, den der Admin bei installiertem XSLT-Prozessor in sein favorisiertes Format überführen kann. Der XSLT-Prozessor erhält dazu passende Stylesheet-Informationen von »smbtaquery«. Um das Abfragen der

Datenbank zu vereinfachen, enthält »smbtaquery« einen einfachen Interpreter, der auf die Zusammenarbeit mit SMBTA spezialisiert ist.

Abfragedatei

Es gibt grundsätzlich zwei Möglichkeiten, den eingebauten Interpreter zu benutzen. Die eine besteht darin, eine Datei zu übergeben, die bereits sämtliche Abfrage-Befehle enthält. Der Datei-

name wird dabei durch den Parameter »-f« (File) gekennzeichnet.

```
smbtaquery -h Host -i 3491 -f ?
befehlsdatei.txt
```

Sowohl in der Datei als auch im Interpreter-Modus ist jedes Kommando durch ein Komma zu separieren, Parameter durch ein Leerzeichen. Jede Zeile endet mit einem Semikolon. In der Konfigurationsdatei werden Kommentare wie üblich mit »#« gekennzeichnet.

Interaktive Queries

Als zweite Möglichkeit kann der Admin auch direkt den eingebauten Interpreter benutzen, was er »smbtaquery« mit dem Parameter »-q« (query) signalisiert.

```
smbtaquery -h Host -i 3491 -q 'Abfrage'
```

Die Parameter »-h« und »-i« erschließen sich aus den bisherigen Erläuterungen. Hinter »-q« für „query“ folgt die eigentliche Abfragesyntax, zum Beispiel:

```
smbtaquery -h Host -i?
3491 -q 'global,?
usage rw;'
```

womit »smbta-

query« den globalen Traffic im kompletten Samba-Netzwerk zählt. Selbstverständlich lässt sich der zu untersuchende Traffic auch auf einen Benutzer oder einen Share beschränken:

```
smbtaquery -h Host -i 3491 ?
-q 'user drilling, total w;'
```

Smbtaquery kann sich selbstverständlich nicht nur via Hostname und TCP-Port verbinden, sondern wie erwähnt auch Unix Domain Sockets verwenden.

```
smbtaquery -u -q 'global, usage rw;'
```

Übrigens gibt »smbtaquery« per Default sämtliche Ausgaben auf dem Terminal aus, vom dem es gestartet wurde. Für das Umleiten der Ausgabe in eine Datei kann der Admin gewöhnliche Unix-Operatoren benutzen, wie »> ausgabe.txt«. Es ist aber auch möglich, mithilfe des Parameters »-o« beispielsweise die Ausgabe im HTML-Format zu erzeugen:

```
smbtaquery -u -q 'global, usage rw;'?
-o html > ausgabe.html
```

Abfrage-Beispiele können der hervorragenden Dokumentation entnommen werden. So ermittelt beispielsweise

```
'user drilling, total r;'
```

die absolute Anzahl (»total«) an Bytes, die vom Benutzer »user drilling« über das Samba-Netzwerk gelesen (»r«) wurden. Ein weiteres Beispiel wäre:

```
'share USB-Fritzbox, usage rw;'
```

Hierbei zeigt die Usage-Funktion die zeitliche Verteilung der Aktivität eines Objektes für einen „virtuellen“ Tag im 24-Stunden-Raster, im Beispiel für Lese-Schreib-Zugriffe auf der Freigabe »USB-Fritzbox«,

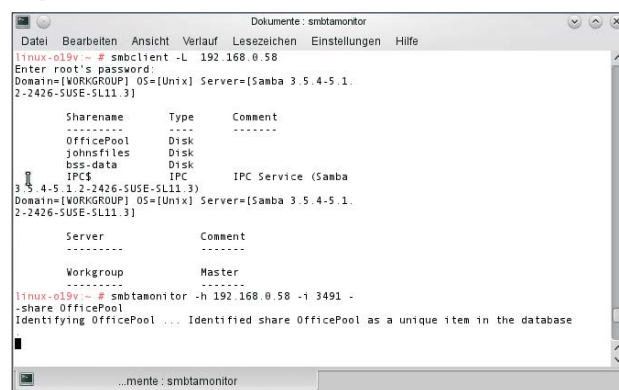


Abbildung 4: Smbtamonitor bindet sich immer explizit an ein Objekt, also entweder einen Share, einen User oder einen Dateinamen.

einer am Router angeschlossenen externen Festplatte (**Abbildung 3**). Alternativ dazu zeigt der Ausdruck

```
'share USB-Fritzbox, total w;'
```

die absolute Anzahl an Bytes, die auf die Freigabe „USB-Fritzbox“ geschrieben wurden. Die »total«-Funktion ermittelt und zeigt stets die totale Summe an Bytes, die vom angegebenen Objekt (Share, User, gesamtes Netz) gelesen und/oder geschrieben wurden. Weitere leistungsfähige und interessante Parameter, wie »top«, »list« oder »last_activity« lassen sich ebenfalls der Dokumentation entnehmen.

Smbtamonitor

Das Tool »smbtamonitor« ermöglicht dem Administrator, den kompletten Samba-Traffic in Echtzeit zu überwachen. Der Client verbindet sich dazu direkt mit dem Daemon »smbtd«, anstatt gespeicherte

Traffic-Informationen aus der Datenbank zu holen. Der Daemon sendet alle Daten-Pakete, die er vom VFS-Modul empfängt, an den »smbtamonitor«.

Der wiederum horcht permanent am SMBTA-Socket und visualisiert alle empfangenen Pakete in einem Curses-Graphen.

Dazu bindet sich jede Smbtamonitor-Instanz stets

an ein Objekt, User, Share oder Datei (**Abbildung 4**). Dabei kann der Admin beliebig viele Smbtamonitor-Instanzen starten. So ist es mit »smbtamonitor« beispielsweise ebenfalls möglich, die ab-

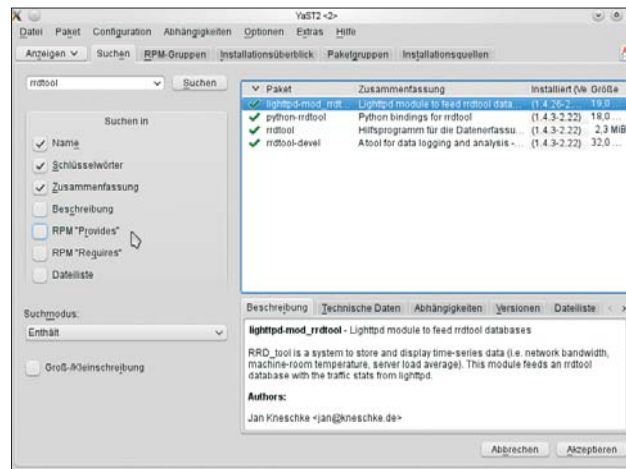
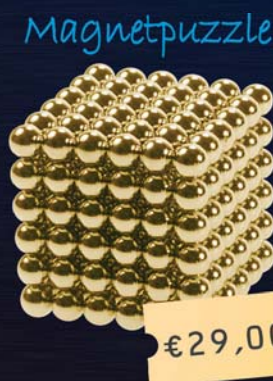


Abbildung 5: Die RRD-Tools lassen sich bei Open Suse mit Yast installieren und stellen ein leistungsstarkes Interface zur Verfügung, um grafische Messdaten zu visualisieren oder in Skripten weiterzuverarbeiten.

solute Anzahl übertragener Bytes und/oder den Datendurchsatz/Sekunde für das angegebene Objekt sichtbar zu machen. Smbtamonitor erfordert entweder die Angabe von Host (»h«) und Portnum-

getDigital.de
YOUR GEEK STUFF SUPPLIER

In unserem Geek Shop findest Du unzählige spannende T-Shirts für IT-Admins und andere Nerds, sowie ausgefallene Gadgets, besondere Tastaturen und mehr für Technikfans.



RFID Schutzhüllen

www.getdigital.de/admin_shop



try me!

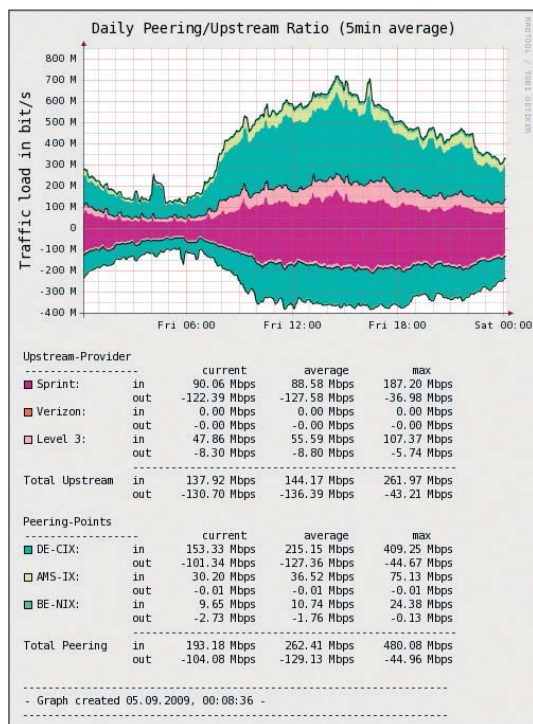


Abbildung 6: Der RRD-Treiber implementiert eine Schnittstelle zwischen SMBTA und RRD-Tool, mit deren Hilfe sich umfangreiche Graphen erstellen lassen.

mer (»i«) oder das Initiieren der Verbindung via Unix-Domain-Socket mithilfe des Parameters »-n«:

```
smbtmonitor -h Host -i 3491 --share 2
Freigabe
```

Einzelne Dateien kontrollieren

Smbtmonitor kann sich auch gezielt an eine Datei binden. So ist es etwa möglich, in Echtzeit zu visualisieren, ob und in welchem Umfang die Nutzer im Netz von der Datei »RELEASENOTE.TXT« Kenntnis nehmen, die auf einem Netzlaufwerk abgelegt ist:

```
smbtmonitor -h Host -i 34912
--file RELEASENOTE.TXT
```

Übrigens lässt sich auch das Tool »smbtmonitor« mit einer Konfigurationsdatei »\$HOME/.smbttools/monitor-config« konfigurieren, in der Hostname und Portnummer hinterlegt sind:

```
[network]
Hostname = SMBTA-Host
Port = 3491
```

Viele weitere nützliche Parameter finden sich ebenfalls in der Dokumentation.

RRD zeichnet Graphen

Das Round Robin Database Tool (RRD-Tool) ist für seine Robustheit und die einfache Handhabung bekannt und ermöglicht eine grafische Visualisierung etwa des Durchsatzes auf einem Samba-Share. RRD-Tool steht unter [2] zur Verfügung, ist aber bei Open Suse auch in den Repositories zu finden und leicht mit dem Yast-Paketmanager zu installieren (Abbildung 5).

SMBTA selbst enthält allerdings lediglich einen Treiber als Schnittstelle zu RRD-Tool, der sich durch Aufruf von »rrddriver« gefolgt von einem oder mehreren Argumenten starten lässt (Abbildung 6). Darunter die Bekannten »-h« (Host), »-i« (TCP-Port), »-n« (Domain Sockets), »-s« (Share), »-u« (User) aber auch Neue, wie »-r«, mit dessen Hilfe der Admin einen RRD-Tool-Setup-String definiert. Default ist

```
DS:readwrite:GAUGE:10:U:U
DS:read:GAUGE:10:U:U
DS:write:GAUGE:10:U:U
```

Ein Beispiel für das Verwenden des RRD-Treibers könnte wie folgt aussehen:

```
rrddriver -b meinrrd -h Host2
-i 3491 -user drilling
```

Es startet den RRD-Treiber für alle Traffic-Informationen, die der Nutzer »drilling« produziert.

Per Default aktualisiert RRD-Tool seine Datenbank alle zehn Sekunden. Wer es genauer haben möchte, kann das Intervall beispielsweise mit »-S 2« auch auf

zwei Sekunden senken. Jetzt kann der Administrator die RRD-Tools mit seinen zahlreichen Möglichkeiten nutzen. Mehr Informationen zu RRD-Tool gibt der gleichnamige Kasten. Auch die Projektseite des RRD-Tool ist eine ergiebige Quelle und stellt umfangreiches Dokumentationsmaterial zur Verfügung. Ein Beispiel für den manuellen Aufruf von RRD-Tool gibt Listing 1. Er erzeugt eine PNG-Grafik »bild-smb-durchsatz.png« mit dem Titel »Datendurchsatz auf Share 'johnsfiles'«, die den Schreib- und Lese-Durchsatz auf der genannten Freigabe zeigt, getriggert im Unix-Time-Format, das der Parameter »-s« spezifiziert. Hinter »DEF« folgt der virtuelle Name

RRD-Tool

Das ursprünglich von Tobias Oetiker entwickelte RRD-Tool hat sich über die Jahre zum Quasi-Standard beim Speichern von Netzwerk-Überwachungsdaten entwickelt. Es ist eine Software, mit der sich zeitbezogene Messdaten speichern und visualisieren lassen. Inzwischen steht RRD-Tool unter der GNU General Public License, sodass heute eine ganze Reihe von Autoren daran mitarbeiten. RRD-Tool ist im Sourcecode wie auch als Binary für eine Reihe von Betriebssystemen verfügbar.

Die Abkürzung RRD steht für Round Robin Database und somit die Art und Weise, wie es seine Daten speichert: Beim Anlegen seiner Datenbank reserviert das System Speicher nur für eine bestimmte Zeitspanne und erweitert die Datenbank nach deren Ablauf nicht etwa, sondern überschreibt die jeweils ältesten Daten. In der englischen Informatik heißen solche Reihum-Methoden auch Round Robin. Die Idee dabei: Stetig eintrudelnde zeitbezogene Messdaten sollen mit anwachsender Uptime nicht die Festplatte vollmüllen, denn bei älteren Daten reicht oft ein grober Überblick, während sich das aktuelle Geschehen in der Regel durch wichtige Details offenbart.

Das Benutzer-Interface von RRD-Tool besteht aus einem Satz von Kommandozeilen-Tools, deren Funktion auf der Projektseite ausführlich erklärt ist. Außerdem sind APIs für viele Programmiersprachen definiert, allen voran C und Perl, damit sich RRD-Tool von anderen Programmen zum Speichern nutzen lässt. RRD-Tool ruft man normalerweise nicht über die Kommandozeile auf, meistens dient es anderen Programmen als Datenquelle- und/oder Speicher, etwa Cacti [9] oder MRTG [10]. Eine umfangreiche Liste findet sich ebenfalls auf der RRD-Tool-Homepage.

Listing 1: RRD-Tool

```
01 rrdtool graph bild-smb-durchsatz.png -s 1290772099 -S 1 --title "Datendurchsatz auf Share
'johnsfiles'" DEF:read_in=testdb:read:AVERAGE DEF:write_in=testdb:write:AVERAGE "AREA:write_
in#AA0000:Write" "STACK:read_in#AA9999:Read"
```

Buntes treiben: USB via Netzwerk



USB-Server können..

.. Distanzen überbrücken

Lagern Sie Ihren USB-Port dorthin aus, wo Sie ihn in Ihrem Netzwerk benötigen. Angeschlossene USB-Geräte verhalten sich, als wären sie lokal an Ihren Windows PC angeschlossen.

.. gemeinsam zugreifen

Geräte wie Scanner, Drucker, Dongles, usw... werden ohne Kabel-/Geräteumstecken von mehreren Anwendern genutzt.

.. aus virtuellen Umgebungen zugreifen

Installieren Sie einen USB-Host-Controller in Ihr virtuelles Windows (Hyper-V, VM-Ware oder VirtualBox).

.. Kertreiber nutzen

Auch ohne Benutzeranmeldung steht der virtuelle USB-Controller Windows-Diensten bereits zur Verfügung.

Weitere Informationen
unter:
USB-Server.de/admin

für den darzustellenden Wert, dahinter die Datenquelle, der auszulesende Wert und die sogenannte Konsolidierungsmethode (hier »AVERAGE«).

Fazit

Datendurchsätze im Netzwerk messen und anzeigen an sich ist keine große Sache. Interessiert sich der Admin aber speziell für Datenverkehr, der vom CIFS-Server Samba verursacht wird, kommt vielen Administratoren ein waschechtes Datenbankbasiertes Datamining-Tool wie SMBTA sicher wie gerufen. Zumindest unter Fachleuten erntete Autor Holger Hettrich bisher so viel Aufmerksamkeit, dass sich das Samba-Team entschlossen hat, den

Der Autor

Thomas Drilling ist seit mehr als zehn Jahren hauptberuflich als freier Journalist und Redakteur für Wissenschafts- und IT-Magazine tätig. Er selbst und das Team seines Redaktionsbüros verfassen regelmäßig Beiträge zu den Themen Open Source, Linux, Server, IT-Administration und Mac OSX. Außerdem arbeitet Thomas Drilling als Buchautor und Verleger, berät als IT-Consultant kleine und mittlere Unternehmen und hält Vorträge zu Linux, Open Source und IT-Sicherheit.

SMB Traffic Analyzer zum festen Bestandteil von Samba 3.6 zu machen. Allerdings besteht bei der derzeitigen Architektur von Samba im Allgemeinen und SMBTA im Speziellen die Gefahr, dass die von SMBTA verwendete Datenbank sehr schnell anwächst. (ofr)

Infos

- [1] SMBTA wird Bestandteil von Samba 3.6 :[<http://samba.org/samba/ftp/pre/WHATSNEW-3-6-Opref1.txt>]
- [2] RRDTool: [<http://www.mrtg.org/rrdtool/>]
- [3] SMBTA Sourcecode: [<http://morelias.org/smbta>]
- [4] SMBTA Binaries für openSUSE 11.3: [<http://holger123.wordpress.com/smb-traffic-analyzer/smb-traffic-analyzer-download>]
- [5] SMBTA Homepage: [<http://holger123.wordpress.com/smb-traffic-analyzer>]
- [6] SMBTA Git: [<https://github.com/hhetter>]
- [7] Suse-Samba-Repository: [<http://download.opensuse.org/repositories/network:/samba:/STABLE/>]
- [8] Stresstest Appliance 0.0.2: [<http://holger123.wordpress.com/2011/01/28/smbta-stresstest-0-0-2-released-built-with-smb-traffic-analyzer-1-2-2>]
- [9] Cacti: [<http://www.cacti.net>]
- [10] MRTG: [<http://oss.oetiker.ch/mrtg/pub/?M=D>]

Stresstest

Bei Stresstest (aktuell: 0.0.2) handelt es sich um eine fix und fertig geschnürte Suse-Appliance mit installiertem Samba-Server inklusive aktiviertem SMBTA-VFS-Modul (Abbildung 8). Stresstest basiert zwar ebenfalls auf SMBTA 1.2.2, enthält aber darüber hinaus eine Reihe von Patches, die nicht in 1.2.2 enthalten sind. Die Stresstest-Appliance ist primär für das Testen von SMBTA gedacht und wird auch von den Entwicklern dazu intensiv genutzt. Wer einen schnellen Blick auf die Analysequalitäten von SMBTA werfen möchte, ohne erst ein komplexes Szenario aufzusetzen, für den ist Stresstest die beste Lösung. Die Appliance im Open Virtualization Format (OVF) enthält dazu mit »smbtator-

turesrv« eine kleine Server-Applikation, die über mehrere Prozess-Instanzen Dateinamen und Pfade über die Testumgebung verteilt.

Bei Stresstest 0.0.2 sind sechs User aktiv, die allesamt die Applikation »smbtorture« benutzen, um den Serverprozess ausreichend zu beschäftigen. Smbtorture ist quasi eine kleine Testsuite für SMBTA und wird bisher hauptsächlich von den Entwicklern selbst für Langzeittests genutzt. Das Tool simuliert ein typisches Load-Verhalten von Office-Anwendungen. Zwischen den einzelnen Verkehrsproduktions-Zyklen legt das Tool jeweils Pausen von einigen Sekunden ein. Außerdem misst es die Zeit, die

es selbst läuft, und kann seine eigene Aktivität aufzeichnen und reproduzieren. Mehrere Smbtorture-Prozesse können problemlos parallel laufen. SMBTA-Stresstest benutzt übrigens Port 3491, über den die Werkzeuge aus smbtatools ihre Abfragen abwickeln, was bei der Firewall/Paketfilter-Konfiguration des Client-Rechners zu berücksichtigen ist. Ansonsten ist die Appliance wie folgt konfiguriert:

Netzwerk: DHCP
Timezone: Europe/Berlin
Language: de_DE.UTF-8

Firewall: disabled

Das Root-Passwort sowie die Passwörter der übrigen sechs User »holger«, »nelson«, »john«, »bjoern« und »bt-ram« sind jeweils »linux«.

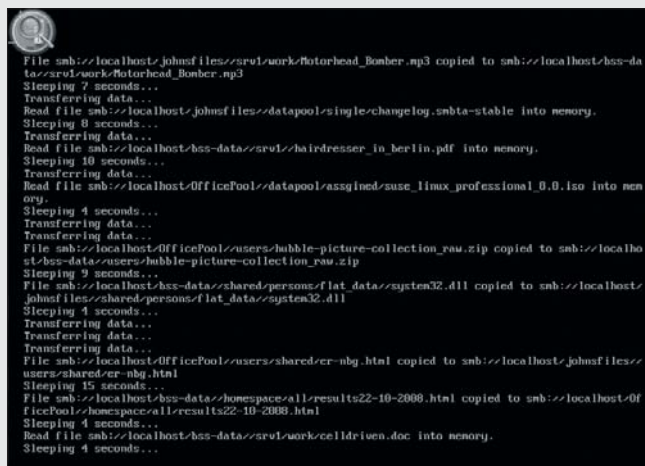


Abbildung 8: Stresstest setzt SMBTAD unter Stress, indem die Appliance laufend Samba-Traffic generiert.



Virtueller Switch mit Openswitch

Schaltstelle

Die Virtualisierung mit VMware, KVM und Xen ist nicht mehr aufzuhalten. Leider fehlte bisher ein virtueller Switch, der komplexe Szenarien unterstützt. Openswitch unterstützt Flows, VLANs und Trunking, Policies und das Bündeln von Netzwerkkarten wie die großen Hardware-Switches. Ralf Spenneberg

Immer mehr Unternehmen stellen ihre gesamte Infrastruktur auf virtuelle Systeme um. Sie virtualisieren zentrale Komponenten wie SAP-Systeme, Oracle-Datenbanken, E-Mail-Systeme und Fileserver und verringern so den Verwaltungsaufwand. Gleichzeitig muss der Administrator für die Wartung der Hardware nicht mehr zwingend die Sys-

teme ausschalten, da er diese zuvor im laufenden Betrieb auf andere virtuelle Hosts verschieben kann.

Ein großer Nachteil der virtuellen Umgebungen war lange Zeit deren einfache Netzwerkstruktur. Während reale Netzwerkeswitches neben VLANs, Trunking, QoS, Port-Aggregation und Firewalling auch Layer-3-Funktionalität bieten, sind

die virtuellen Switches sehr simpel. VMware hat dieses Problem gemeinsam mit Cisco gelöst, das für VMware-Umgebungen den virtuellen Switch Nexus 1000V anbietet. Dieser integriert sich in die VMware-Umgebung und bietet entsprechend fortgeschrittene Funktionen. Für freie Virtualisierungslösungen fehlte bisher ein entsprechendes Produkt. Mit

Openflow

Das Openflow-Projekt der Stanford Universität will die Router/Switch-Welt revolutionieren. Ein klassischer Router oder Switch vereint zwei Funktionen in einem Gerät:

- Schnelle Weiterleitung der Pakete (Data Path)
- Entscheidung wie und wohin die Pakete weiterzuleiten sind (Control Path)

Üblicherweise arbeiten diese beiden Systeme unabhängig voneinander auf demselben Gerät. Nur wenn der Data-Path-Anteil nicht weiß, wie und wohin ein Paket zu senden ist, fragt er den Control Path. Dieser ermittelt den Pfad/Route und speichert diese in der Flow-Tabelle. Alle wei-

teren Pakete desselben Flows kann die Data Path Engine schnell weiterleiten.

Der Openflow-Ansatz verschiebt nun den Control Path auf einen separaten Controller. Hierbei kann es sich um einen einfachen Server handeln. Der Openflow-Switch (Data Path) und der Controller kommunizieren dann über einen sicheren Kanal.

Der Openflow-Switch speichert die Flow-Tabelle. In dieser speichert der Controller die einzelnen Flows. Jeder Flow beschreibt die Eigenschaften der Pakete, die Bestandteil des Flows sind, und wie der Switch diese Pakete behandeln soll (Drop, Send-out-Port und so weiter). Sobald der

Switch ein Paket erhält, für das kein passender Eintrag in der Tabelle existiert, sendet er das Paket an den Controller, der das Paket analysiert, eine Entscheidung trifft und diese in der Flow-Tabelle speichert.

Durch die Zusammenarbeit mit mehreren Herstellern konnten die Entwickler die Unterstützung für Openflow bereits in einigen kommerziellen Netzwerkgeräten erreichen. So existieren angepasste Firmwares für einige Switches von Hewlett-Packard, NEC, Toroki und Pronto [3]. Openswitch ist eine Software-Implementierung, die beide Funktionalitäten (Data Path und Controller) zur Verfügung stellt.

Openvswitch [2] entsteht nun eine OpenSource-Lösung für dieses Problem. Dabei unterstützt Openvswitch neben Xen, KVM und Virtual Box auch den XenServer. Auch Citrix wechselt in der kommenden Version auf Openvswitch, das auf dem Openflow-Projekt [1] der Universität Stanford basiert. Openflow ist ein neuer offener Standard, der die Verwaltung von Switchen und Routern mit beliebiger Software erlauben soll (siehe **Kasten „Openflow“**)

Reichhaltig

Mit Openvswitch kann der Administrator auf einem Linux-System die folgenden Funktionen nutzen:

- Voll-funktionaler Layer-2-Switch
- Unterstützung für Netflow, sFlow(R), SPAN, and RSPAN
- 802.1Q VLANs mit Trunking
- Quality of Service (QoS)
- Port Aggregation
- GRE Tunneling
- Kompatibilität zum Linux-Bridge-Code (»brctl«)
- Switch-Implementierungen im Kernel und im Userspace

Bevor der Administrator jedoch loslegen kann, muss er Openvswitch installieren. Für Debian Sid (Unstable) gibt es fertige Pakete. Für Fedora/RedHat hat der Autor auf seiner Seite [4] Pakete zur Verfügung gestellt. Die Installation kann aber auch einfach aus den Quellen erfolgen (siehe **Kasten „Installation“**).

Während die Pakete entsprechende Start-Skripte für die einfache Nutzung liefern,

Installation

Nach dem Auspacken der Quellen übersetzt und installiert der Administrator die Openvswitch-Software mit den üblichen Befehlen:

```
./configure --with-l26=/lib/modules/$(uname -r)/build
make
sudo make install
```

Um das Kernel-Modul zu übersetzen, müssen die Kernel-Header installiert sein. Diese sind bei den meisten Distributionen in einem Paket wie »kernel-devel« oder »kernel-headers« untergebracht.

Anschließend sollte der Admin die Installation prüfen und die Software das erste Mal starten. Hierzu lädt er manuell das passende Kernel-Modul:

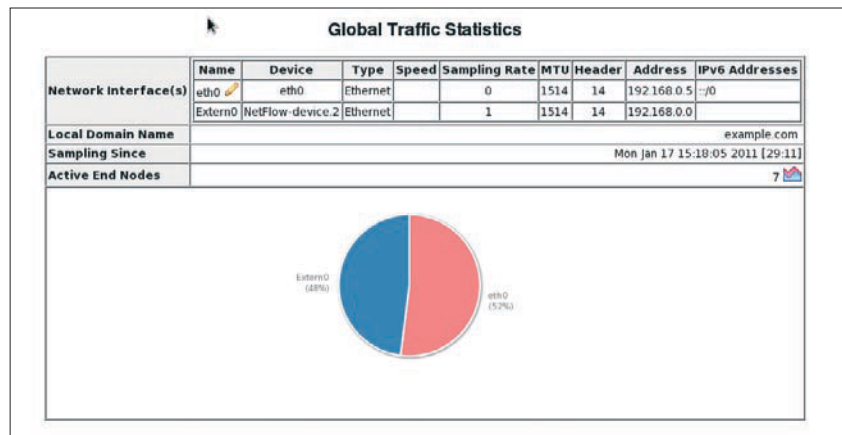


Abbildung 1: Ntop zeigt die Flows der Bridge an.

muss der Admin bei manueller Installation die Dienste auch manuell starten oder ein eigenes Startskript erstellen. Die Konfigurationsdatenbank übernimmt die Verwaltung des Switches (Listing 1). Anschließend startet der Admin den Openvswitch-Dienst:

```
ovs-vswitchd unix:/usr/local/var/run/
openvswitch/db.sock
```

Nun lassen sich mit dem Befehl »ovs-vsctl« bereits neue Switches erzeugen, Ports hinzufügen und konfigurieren. Da die meisten Skripte für Xen und KVM aber den Befehl »brctl« für die Administration der Bridge nutzen, sollte der Admin auch den Bridge-Kompatibilitäts-Daemon starten. Hierzu lädt er zunächst das Kernel-Modul und startet dann den Dienst:

```
modprobe datapath/linux-2.6/brcompat_mod.ko
ovs-brcomptd --pidfile --detach -vANY:
console:EMER unix:/usr/local/var/run/
openvswitch/db.sock
```

```
modprobe datapath/linux-2.6/openvswitch_
mod.ko
```

Schlägt dieser Befehl fehl, so muss der Admin meist ein bereits geladenes Bridge-Modul zunächst entfernen: »rmmod bridge«. Möglicherweise passt die Version des Kernel-Moduls auch nicht zum aktuellen Kernel. Dies kann besonders bei dem Einsatz fertiger Pakete ein Problem darstellen. Dann muss der Admin das Modul erneut übersetzen. Anschließend initialisiert er noch die Konfigurationsdatenbank von Openvswitch:

```
ovsdb-tool create /usr/local/etc/ovs-
vswitchd.conf.db vswitchd/vswitch.ovsschema
```

Bei weiteren Problemen gibt die Datei »INSTALL.Linux« Tipps zur Fehlersuche.

Nun kann er Openvswitch auch mit den Bridge-Utilities verwalten:

```
brctl addbr extern0
brctl addif extern0 eth0
```

Auch alle Skripte der Distributionen für die Erzeugung der Bridges funktionieren nun wie gewohnt. Natürlich kann der Admin auch den Befehl »ovs-vsctl« nutzen, um die Bridge zu verwalten. Beide Befehle lassen sich nun gleichzeitig nutzen (Listing 2).

Falls der Befehl »brctl show« meldet, dass bestimmte Dateien in dem Verzeichnis »/sys/« nicht gefunden werden, sind die Bridge-Utilities (zum Beispiel unter RHEL

Listing 1: Konfiguration

```
01 ovsdb-server /usr/local/etc/ovs-vswitchd.conf.db \
02 --remote=punix:/usr/local/var/
run/openvswitch/db.sock \
03 --remote=db:Open_
vSwitch,managers \
04 --private-key=db:SSL,private_
key \
05 --certificate=db:SSL,certific
ate \
06 --bootstrap-ca-cert=db:SSL,ca_
cert
```

Listing 2: Kontrolle der Bridge

```
01 [root@kvm1 ~]# brctl show
02 bridge name      bridge id                STP enabled
   interfaces
03 extern0          0000.00304879668c       no
   eth0
04 vnet0
05 [root@kvm1 ~]# ovs-vsctl list-ports extern0
06 eth0
07 vnet0
```

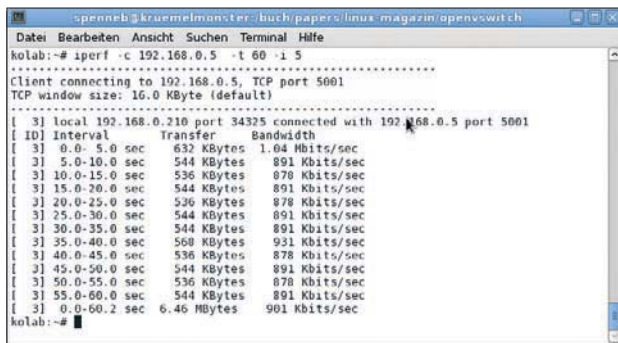


Abbildung 2: Iperf beweist die Beschränkung der Sendeleistung.

6) zu aktuell. Hier empfiehlt sich ein Downgrade auf die letzte Version von RHEL 5.

Bis hierher verhält sich Openvswitch ganz genauso wie eine Bridge, die mit den Bridge-Utilities aufgebaut wurde. Um die fortgeschrittenen Funktionen zu nutzen, muss der Admin nun weitere Konfigurationen vornehmen. Sämtliche Einstellungen der Openvswitch-Konfigurationsdatenbank lassen sich mit dem Befehl »ovs-vsctl« verwalten.

Netflow

Openvswitch kann die Netflows innerhalb des Switches exportieren. Hierzu muss der Admin zunächst eine neue Netflow-Probe anlegen.

```
# ovs-vsctl create netflow target="192.168.0.5:\:5000"
75545802-675f-45b2-814e-0875921e7ede
```

Nun verbindet er die Probe mit der Bridge »extern0«:

```
# ovs-vsctl add bridge extern0 netflow 7
75545802-675f-45b2-814e-0875921e7ede
```

Hat der Admin vorher auf dem System

mit »ovs-vsctl list bridge« und »ovs-vsctl list netflow« kontrollieren und mit »ovs-vsctl destroy ...« wieder entfernen.

QoS

Häufig möchte der Admin die Bandbreite einzelner virtueller Gäste kontrollieren und einschränken. Dies ist insbesondere der Fall, wenn unterschiedliche Kunden in einer virtuellen Umgebung bedient werden. Entsprechend den ausgehandelten Service Level Agreements erhalten dann die unterschiedlichen Gäste die bezahlte Leistung.

Openvswitch kann recht einfach die maximale Sendeleistung einzelner Gäste beschränken. Um dies zu testen, sollte zunächst der normale Durchsatz gemessen werden. Hierzu startet der Admin auf einem System Iperf als Server und auf einem virtuellen Gast Iperf als Client (Listing 3).

Nun kann der Admin die Sendeleistung beschränken. Der Befehl erwartet die Sendeleistung in Kbit/s. Neben der reinen Sendeleistung sollte auch der Burst angegeben werden, der etwa ein Zehntel der Sendeleistung betragen sollte. Die Schnittstelle »vnet0« ist in diesem Beispiel der Switchport, an dem der virtuelle Gast angeschlossen ist.

```
# ovs-vsctl set Interface vnet0 ingress
_policing_rate=1000
# ovs-vsctl set Interface vnet0 ingress
_policing_burst=100
```

Das Ergebnis kann direkt mit Iperf wieder getestet werden. Der Versuch zeigt, dass die Beschränkung funktioniert (Abbildung 2).

Kennt sich der Admin mit dem »tc«-Kommando und dem Class-Based-QoS von Linux mit unterschiedlichen Queuing

192.168.0.5 einen Netflow Collector auf Port 5000 gestartet (zum Beispiel Ntop), so kann er nun die Daten anzeigen (siehe Abbildung 1).

Die Konfigurationseinstellungen in der Datenbank kann der Admin

Disciplines aus, kann er auch diese mit Openvswitch verwenden. Die Manpage liefert hierzu weitere Beispiele.

Mirroring

Um ein Intrusion Detection System zu betreiben, benötigt man einen Mirror-Port auf dem Switch. Auch diesen stellt Openvswitch zur Verfügung. Hierzu muss der Admin zunächst den Mirror-Port erzeugen und anschließend zum richtigen Switch hinzufügen. Um einen Mirror-Port zu erzeugen, der den Verkehr sämtlicher anderer Ports empfängt und auf »vnet0« spiegelt, verwendet man den folgenden Befehl:

```
ovs-vsctl create mirror name=mirror select_2
all=1 output_port=e46e7d4a-2316-407f-ab11-7
4d248cd8fa94
```

Der Befehl »ovs-vsctl list port vnet0« ermittelt die dazu benötigte ID des Output-Ports. Nun muss der Admin den entstandenen Mirror Port zur Bridge hinzufügen:

```
# ovs-vsctl add bridge extern0 mirrors 2
716462a4-8aac-4b9c-aa20-a0844d86f9ef
```

VLANs

Auch VLANs lassen sich mit Openvswitch realisieren. Hierfür bietet Openvswitch zwei verschiedene Möglichkeiten an. Zunächst ist jeder Openvswitch VLAN-fähig. Fügt der Admin einen Port zum virtuellen Switch hinzu, so handelt es sich immer um einen VLAN-Trunk-Port, der alle VLANs getaggt transportiert. Um einen Access-Port zu erzeugen, der ein VLAN nativ und ohne Tags transportiert,

Listing 3: Performance-Messung

```
01 ## Server:
02 ## iperf -s
03 ## Client:
04 # iperf -c 192.168.0.5 -t 60
05 -----
06 Client connecting to 192.168.0.5, TCP port 5001
07 TCP window size: 16.0 KByte (default)
08 -----
09 [ 3] local 192.168.0.210 port 60654 connected with
10 192.168.0.5 port 5001
11 [ ID] Interval      Transfer      Bandwidth
12 [ 3] 0.0-60.0 sec  5.80 GBytes   830 Mbits/sec
```

Der Autor

Ralf Spenneberg arbeitet als freier Unix/Linux-



Trainer, Berater und Autor. Mit seinem Unternehmen OpenSource Training Ralf Spenneberg führt er Schulungen und Beratungen durch. Er veröffentlichte bereits mehrere Bücher

zu den Themen Intrusion Detection, SELinux, Firewalling und Virtuelle Private Netzwerke. Vor wenigen Monaten ist die zweite Auflage seines Buches „VPN mit Linux“ erschienen.

kann der Admin den folgenden Befehl nutzen:

```
ovs-vsctl add-port extern0 vnet1 tag=1
```

Mit »brctl« kann der Admin nicht direkt einen derartigen Port erzeugen. Hierzu muss er einen Umweg über eine Fake-Bridge gehen. Openvswitch unterstützt Fake-Bridges, die man dann einzelnen VLANs zuordnet. Jeder Port an einer Fake-Bridge ist dann ein Access-Port in dem VLAN.

Hierzu erzeugt der Admin zunächst die Fake-Bridge. Diese legt er immer als Tochter-Bridge einer übergeordneten Bridge an. In dem Beispiel dieses Artikels wäre das:

```
# ovs-vsctl add-br VLAN1 extern0 1
```

Die neue Fake-Bridge hat nun den Namen VLAN1 und transportiert das VLAN mit dem Tag 1. Diese muss der Administrator noch aktivieren und kann dabei gleichzeitig auch eine IP-Adresse vergeben.

```
# ifconfig VLAN1 192.168.1.1 up
```

Jeder Port, der nun an dieser Bridge erzeugt wird, ist ein Access-Port in dem VLAN 1. Damit kann nun auch wieder der Befehl »brctl« genutzt werden.

Weitere Funktionen

Openvswitch bietet darüber hinaus noch viele weitere Funktionen. So kann er GRE-Tunnel zwischen mehreren Systemen aufbauen und darüber auch ein VLAN-Trunking durchführen. Damit ist es auch möglich, virtuelle Maschinen auf andere Rechner zu verschieben, die sich nicht im selben lokalen Netz befinden. Die Kommunikation erfolgt dann über den GRE-Tunnel.

Außerdem kann Openvswitch Ports aggregieren. Die Linux-Kernel-Entwickler sprechen hier von Bündelung, Cisco nennt diese Funktion Etherchannel. Damit kann der Admin mehrere physische

Ports als einen logischen Port zusammenfassen und diese zur Lastverteilung und Hochverfügbarkeit nutzen.

Openvswitch ist ein sehr interessantes Projekt, das im Moment in erster Linie unter seiner fehlenden Bekanntheit und der geringen Dokumentation leidet. Hoffentlich werden die Distributionen dieses Projekt aufnehmen und in ihre Werkzeuge, zum Beispiel Libvirt nativ integrieren. Dann ist auch die Nutzung des Bridge-Kompatibilitäts-Daemon nicht mehr erforderlich. (ofr) ■

Infos

- [1] Openflow: [\[http://www.openflow.org\]](http://www.openflow.org)
- [2] Openvswitch: [\[http://www.openvswitch.org\]](http://www.openvswitch.org)
- [3] Openflow Switches: [\[http://openflow.org/foswiki/bin/view/OpenFlow/Deployment/Components\]](http://openflow.org/foswiki/bin/view/OpenFlow/Deployment/Components)
- [4] RPM-Pakete von Openvswitch: [\[http://www.spenneberg.org/Openvswitch/\]](http://www.spenneberg.org/Openvswitch/)

German Unix User Group

Die Fachkonferenz
für professionelle
Systemadministratoren
und Spezialisten im
Bereich Unix, Netze
und IT-Sicherheit

guug

Frühjahrsfach-
gespräch

22. – 25.3.2011

Bauhaus
Universität
Weimar

<http://guug.de/ffg>

Keynote von
Gunter Dueck
CTO IBM

Vorträge und
Tutorien zu
Netzwerksicherheit,
Virtualisierung, Mail,
Backup, SAP,
Monitoring, MySQL,
Konfigurations-
management, ...



So funktionieren SSD-Speicher

Frische Chips

Elena Moiseeva, 123RF

Solid State Drives werden von Tag zu Tag populärer. Die kleinen Performance-Wunder versprechen eine deutlich höhere I/O-Performance als herkömmliche Festplatten - und das bei niedrigerem Stromverbrauch. Wie SSDs funktionieren und wann sich deren Einsatz lohnt, verrät dieser Artikel. Werner Fischer

SSDs haben mit normalen Festplatten nur eine Gemeinsamkeit: Beide Medien speichern Daten. Intern arbeiten SSDs allerdings vollkommen anders als Festplatten. Während Festplatten die Daten auf mehreren Magnetscheiben speichern, nutzen SSDs mehrere Flash-Chips – meist bis zu zehn Stück – für die Datenspeicherung. Da bei SSDs keine mechanischen Teile wie bei Festplatten bewegt werden, sind vor allem zufällig verteilte Datenzugriffe deutlich schneller.

Festplatte im Vergleich

Um die Performance von SSDs mit Festplatten zu vergleichen, lohnt zuvor ein kurzer Blick auf den Aufbau herkömmlicher Festplatten. Diese speichern die Daten auf mehreren rotierenden Magnetscheiben. Die Daten werden dabei mit Schreib-/Leseköpfen auf die Scheiben geschrieben beziehungsweise von diesen gelesen.

Soll nun ein zufälliger Sektor von der Festplatte gelesen werden, muss die Festplatte den Schreib-/Lesekopf an die richtige Stelle bewegen (Seek Time) und dann darauf warten, dass der gewünschte Sektor auf der rotierenden Scheibe unter dem Schreib-/Lesekopf vorbeifährt (La-

tency Time). Die Latency Time beträgt dabei im Mittel die Dauer, welche für eine halbe Umdrehung notwendig ist. Die Summe aus Seek Time und Latency Time ergibt die Random Access Time (mittlere Zugriffszeit). Diese beträgt je nach Festplatte und deren Drehzahl zwischen fünf und 15 Millisekunden, was 200 respektive 66 IOPS (I/O-Operationen pro Sekunde) entspricht.

SSDs erreichen im Vergleich dazu mehrere tausend zufällig verteilte IOPS. Bei kontinuierlichen Datentransfers kann die Festplatte allerdings deutlich besser mithalten. Sie liefert dabei je nach Typ zwischen 60 und 150 MByte/s. SSDs können zwar zu Beginn über 250 MByte/s liefern. Mit höherem Füllstand der SSD sinkt diese Rate jedoch. Wie sehr die Datenrate zurückgeht, hängt dabei von den Algorithmen im SSD-Controller ab.

Aufbau

Die kleinste Einheit in einem Flash-Chip einer SSD ist die Speicherzelle. Je nach Typ speichert eine einzelne Speicherzelle ein oder mehrere Bits:

- SLC (Single Level Cell) 1 Bit
- MLC (Multi Level Cell) 2 Bits
- TLC (Triple Level Cell) 3 Bits

Eine SLC unterscheidet zwei Ladungszustände, eine MLC vier und eine TLC acht. Die angelegte Spannung beim Schreiben auf eine Speicherzelle ist allerdings fix. Daher muss bei einer MLC für einen Schreibvorgang bis zu vier Mal Spannung angelegt werden, um in den höchsten Ladungszustand zu kommen. Bei einer SLC reichen maximal zwei solcher Vorgänge – die Schreibperformance ist damit höher. TLCs erfordern bis zu acht Mal Spannung und werden daher für SSDs nicht verwendet. Aufgrund ihrer hohen Datendichte und höheren Kapazität werden TLCs aber für USB-Sticks und SD-Karten genutzt. Dazu reicht ihre Performance aus.

Bei jedem Anlegen von Spannung kommt es zu einer kleinen Abnutzung der Isolationsschicht der Speicherzelle. Da für einen einzelnen Schreibvorgang auf eine MLC öfter Spannung angelegt wird als bei einer SLC, verträgt die MLC weniger Schreibvorgänge (sogenannte Program/Erase Cycles oder p/e-Zyklen). MLCs haben meist eine Lebensdauer zwischen 10 000 und 30 000 p/e-Zyklen, SLCs rund 100 000 p/e-Zyklen.

Mehrere Speicherzellen bilden eine Page. Sie ist die kleinste Struktur, die vom SSD-Controller gelesen oder beschrie-

ben werden kann. Allerdings kann der SSD-Controller den Inhalt einer bereits beschriebenen Page nicht löschen oder verändern. Eine Page ist in der Regel vier Kibibyte (KiB) groß (= 4096 Bytes). Mit MLCs entspricht das 16 384 Speicherzellen. Künftige Flash Chips mit einer Fertigungsstrukturbreite von 25 nm haben Pages mit acht KiB Größe.

Mehrere Pages sind in Blöcken zusammengefasst. Aktuell besteht ein solcher Block aus 128 Pages und fasst damit 512 KiB an Daten. Blöcke der künftigen 25nm-Fertigung nutzen 256 Pages à acht KiB, in Summe also 2 MiB. Ein Block ist die kleinste Einheit, die der SSD-Controller löschen kann, indem er eine Löschspannung am gesamten Block anlegt. Erst nach dem Löschen eines Blocks kann der SSD-Controller die einzelnen Pages wieder neu beschreiben.

1024 Blöcke bilden eine Plane. Vier Planes sind wiederum auf einem Die untergebracht, wie **Abbildung 1** zeigt. Ein Die ist dabei etwa so groß wie ein Fingernagel (167 mm²). In der Produktion werden circa 200 bis 300 Dies aus einem Wafer gewonnen. Je nach gewünschter Kapazität landen ein bis acht Dies in einem Thin small-outline package (TSOP) – jene schwarzen Chips, wie sie auch auf Speichermodulen vorhanden sind. Eine SSD besteht schließlich aus bis zu zehn TSOPs und einem SSD-Controller.

Schreibtechniken

Die geschilderte interne Funktionsweise einer SSD erfordert spezielle Schreibtechniken um sowohl eine hohe Performance als auch eine lange Lebensdauer der SSD zu erreichen. Das erstmalige Beschreiben einer neuen SSD ist dabei noch einfach. Alle Blöcke (und somit auch alle Pages) sind gelöscht. Neue Daten schreibt der SSD-Controller direkt auf die entsprechenden Pages.

Schwieriger wird es, sobald bereits vorhandene Daten verändert werden. Ohne zusätzliche Mechanismen müsste der SSD-Controller sämtliche Pages eines Blocks zuerst in einen Cache einlesen, anschließend den gesamten Block löschen, um abschließend die zwischengespeicherten Pages versehen mit den Änderungen wieder in den Block zu schreiben. Die Änderung eines einzel-

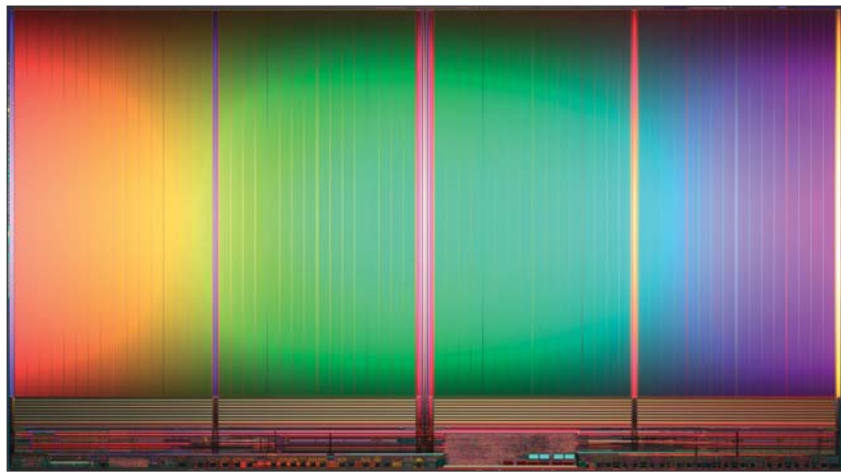


Abbildung 1: Die eines Flash-Chips mit vier Planes. (Quelle: Intel)

nen Bits würde so das Lesen, Löschen und Neuschreiben von 512 KiB an Daten verursachen.

Das vermeidet der SSD-Controller mit einem einfachen Trick. Wenn er Daten einer Page verändert, schreibt er dazu einfach die neuen Daten in eine andere Page, die noch gelöscht ist. In einer internen Zuordnungstabelle vermerkt der Controller dabei, dass die Daten der logischen LBA-Adresse A nun nicht mehr in Page X, sondern in Page Y zu finden sind. Die ursprüngliche Page X markiert er als ungültig. Damit diese Methode funktioniert, benötigt der SSD-Controller zusätzliche Speicherkapazität. Ansonsten könnte er auf einer vollgeschriebenen SSD keine Daten mehr verändern. Daher hat jede SSD eine sogenannte Spare Area. Typischerweise beträgt die Größe dieser Spare Area zwischen sieben und 27 Prozent der Nennkapazität einer SSD. Eine 160 GB SSD kann etwa tatsächlich über 172 GByte verfügen, 12 GByte versteckt die SSD aber vor dem Betriebssystem und nutzt diese Datenmenge als Spare Area. Mit fortschreitender Nutzungsdauer verändern sich immer mehr vorhandene Daten. Mehr und mehr Pages markiert der SSD-Controller als ungültig. Einzelne Blöcke enthalten nur mehr 40 bis 60 Prozent echte Daten, die restlichen Pages darin sind ungültig und können vorerst nicht weiter genutzt werden. Bevor nun die Spare Area zur Neige geht, räumt der SSD-Controller auf. Bei dieser Garbage Collection kopiert der SSD-Controller die noch gültigen Pages eines Blocks in einen freien Block, der damit nur teilweise gefüllt ist. Die restlichen Pages dieses neuen

Blocks bleiben noch unbeschrieben und können für weitere Schreiboperationen genutzt werden. Den ursprünglichen Block löscht der SSD-Controller danach. Somit sind alle Pages dieses gelöschten Blocks wieder beschreibbar. **Abbildung 2** zeigt ein solches Beispiel.

Wie eingangs erwähnt, ist die Anzahl der möglichen Schreibvorgänge (p/e-Zyklen) pro Speicherzelle begrenzt. Sollen nun neue Daten gespeichert werden, schreibt der SSD-Controller daher immer zuerst auf Pages, die noch wenig abgenutzt sind. Dieser Vorgang wird als Dynamic Wear Leveling (dynamische Abnutzungsverteilung) bezeichnet. Dynamisch deshalb, weil nur neue oder veränderte Daten verteilt werden.

Diese Methode erhöht die Lebensdauer der SSD. Daten, die nur einmal geschrieben werden und sich danach nicht mehr ändern, bleiben damit aber auf ihren Pages. Static Wear Leveling geht daher einen Schritt weiter und verschiebt auch solche Daten periodisch auf andere Pages, die schon mehr abgenutzt sind. Damit werden tatsächlich alle Pages einer SSD gleichmäßig abgenutzt und die Lebensdauer der SSD steigt weiter.

Write Amplification

Durch die geschilderten Algorithmen werden einmal geschriebene Daten, selbst wenn sie nicht verändert werden, in andere Pages kopiert (etwa durch Garbage Collection oder Static Wear Leveling). Ein geschriebenes Byte kann also mit der Zeit vom SSD-Controller durchaus mehrfach kopiert werden und somit zu

mehreren Schreibvorgängen führen. Der Faktor, wie oft es statistisch zu einem solchen Kopiervorgang kommt, wird als Write Amplification bezeichnet.

Eine hohe Write Amplification führt naturgemäß zu einer höheren Abnutzung und damit geringeren Lebensdauer. Manche SSD-Controller (etwa von Sandforce) versuchen, durch eine Komprimierung der Daten im SSD-Controller die nötigen Schreibvorgänge zu reduzieren, um damit die Abnutzung zu reduzieren. Eine weitere Möglichkeit ist die Vergrößerung der Spare Area. Dadurch sinkt zwar die nutzbare Kapazität, da aber seltener eine Garbage Collection erforderlich ist, steigt dennoch die Lebensdauer der SSD.

Lebensdauer

Die Lebensdauer von SSDs ist durch die Lebensdauer der Speicherzellen begrenzt. Mit fortschreitender Nutzung können vermehrt Blöcke ausfallen. Wobei ein „Ausfall“ dabei relativ ist. Je mehr p/e-Zyklen stattfinden, umso länger dauert das Löschen eines Blocks. Überschreitet diese Zeit zum Löschen einen bestimmten Schwellwert, markiert der SSD-Controller diesen Block als „Bad Block“ und nutzt stattdessen einen Spare Block aus der Spare Area. Es kommt dabei also zu keinem Datenverlust. Es sinkt nur die Anzahl der Spare-Blöcke. Einzelne Bitfehler können zwar auftreten (raw bit error rate, RBER), werden aber durch ECC-Mechanismen korrigiert. Erst

bei zu vielen Bitfehlern greift ECC nicht mehr, und es kommt zu einem unkorrigierbaren Fehler (uncorrectable bit error rate, UBER).

Damit künftig die Lebensdauer von unterschiedlichen SSDs einfach vergleichbar sind, hat die JEDEC Solid State Technology Association die beiden Standards JESD218 (SSD Requirements and Endurance Test Method) und JESD219 (SSD Endurance Workloads) verabschiedet. Damit können Hersteller die Lebensdauer ihrer SSDs in TBW (Terabytes written, geschriebene Terabytes) angeben. Einige SSDs geben auch schon jetzt per SMART Auskunft über die verbleibende Lebensdauer (Media Wearout Indicator) [1]. Generell ist die zu erwartende Lebensdauer von SSDs bei normalen Schreibmengen durchaus mit jenen von Festplatten vergleichbar, oft sogar höher, da bei es bei SSDs zu keinen mechanischen Ausfällen (wie etwa Head-crashes bei Festplatten) kommt.

Einsatzgebiete

SSDs eignen sich aufgrund ihrer Eigenschaften für zahlreiche Einsatzgebiete. Die folgende Auflistung ist nach aufsteigenden Investitionskosten gegliedert.

Bereits eine SSD mit geringerer Kapazität zwischen 40 und 80 GByte bringt bei einem Einzelplatz-PC deutliche Produktivitätsvorteile. Bei geringen Investitionskosten dient sie als Installationsmedium für das Betriebssystem und die Anwendungsprogramme neben einer normalen Festplatte, die Benutzerdaten enthält. Das Hochfahren des Rechners geht damit deutlich schneller. Auch die einzelnen Programme starten rascher, das Arbeiten am PC ist flüssiger.

SSDs mit höheren Kapazitäten ab 160 GByte können die bisherige Festplatte in einem Rechner vollständig ersetzen. Das erhöht zwar die Kosten, beschleunigt aber dafür auch den Zugriff auf die Daten des Benutzers. Weitere Vorteile durch den Wegfall einer herkömmlichen Festplatte sind der Wegfall des Betriebsgeräusches der Festplatte und der deutlich geringere Stromverbrauch. Bei Notebooks lässt sich damit die Akkulaufzeit um 20 bis 30 Mi-

nuten erhöhen. Auch im Serverumfeld lassen sich SSDs sinnvoll einsetzen. Vor allem bei Datenbanken mit vielen zufälligen I/O-Zugriffen spielen SSDs ihre Vorteile aus. Eine Spiegelung von SSDs mit einem RAID 1 schützt vor Datenverlust beim Ausfall einer einzelnen SSD. Manche RAID-Controller unterstützen dabei sogar ein RAID 1 bestehend aus einer SSD und einer herkömmlichen Festplatte, wobei der RAID-Controller dabei Lesezugriffe im Normalbetrieb ausschließlich an die SSD schickt.

SSDs können darüber hinaus auch als beschleunigender Lesecache in einem RAID-Verbund mit normalen Festplatten genutzt werden. Adaptec-Controller mit Maxcache-Unterstützung legen häufig gelesene Datenbereiche zusätzlich auf SSD ab [2]. Um diese Daten später erneut zu lesen, wird direkt von den SSD gelesen und nicht von den RAID-Festplatten. Im Gegensatz zum normalen Cache des RAID-Controllers hat dieser SSD-Cache eine deutlich höhere Kapazität und bleibt auch bei einem Reboot erhalten.

Fazit

SSDs sind zwar nach wie vor mit höheren Einstiegskosten verbunden. Sie zahlen sich aber durch ihre hohe I/O Performance oft sehr schnell aus. Entscheidend für eine gute Performance ist aber die Qualität, wie gut der Hersteller die Algorithmen im SSD-Controller implementiert hat.

Im nächsten ADMIN zeigt der zweite Teil dieser Artikelserie, wie sich die SSD-Performance mit AHCI, ATA TRIM, einer vergrößerten Spare Area und durch ein korrektes Partition Alignment optimieren lässt. (ofr)

Infos:

- [1] SSD-SMART-Analyse: <http://www.thomas-krenn.com/SSD-SMART-Analyse>
- [2] Maxcache-Support: <http://www.thomas-krenn.com/maxCache>

Der Autor

Werner Fischer ist seit 2005 Technology Specialist bei der Thomas-Krenn.AG und Chefredakteur des Thomas Krenn Wikis. Seine Arbeitsschwerpunkte liegen in den Bereichen Hardware-Monitoring, Virtualisierung, I/O Performance und Hochverfügbarkeit.

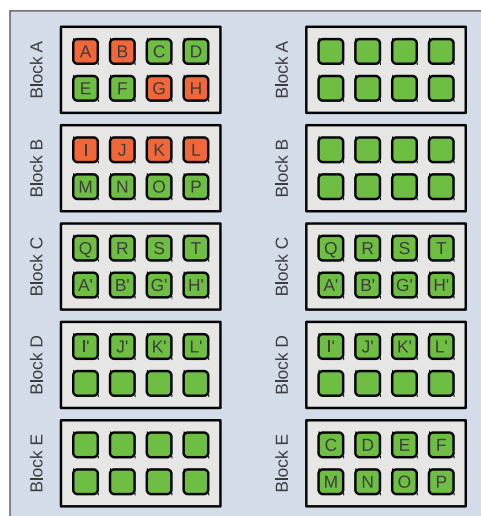


Abbildung 2: Viele Pages sind ungültig, weil deren Daten verändert und woanders abgespeichert wurden. Die Garbage Collection macht die ungültigen Bereiche wieder nutzbar.

Ohne Mythen: Green-IT in Zahlen

Grüne Verheißung

Auch Rechner kommen in die Jahre und leiden dann nicht nur unter Zipperlein, sondern halten auch nicht mehr mit der Software Schritt, die sich mit jedem Release verjüngt und anspruchsvoller wird. Was tun? Jens-Christoph Brendel

Sitzen drei Admins beim Bier, braucht man nicht lange zu warten, bis die Rede auf alte Zeiten kommt und darauf, dass sich heute keiner mehr vorstellen kann, was doch noch vor zehn Jahren Stand der Technik war: Kleiderschrankgroße Ungetüme mit der Speicherkapazität einer heutigen Festplatte im Zigarrenkasten-Format oder sündhafte teure Server mit einer Rechenleistung, die man aktuell in Consumer-Elektronik einbettet. IT ist ein so schnelllebiges Geschäft, dass ihr Tempo selbst die Akteure zuweilen verblüfft. Wäre es unter diesen Umständen nicht konsequent, etwa alle drei Jahre die abgeschriebenen PCs mit erloschener Garantie gleich ganz zu ersetzen? Dann wäre man immer up-to-date und täte vielleicht sogar der Umwelt etwas Gutes, käme doch so immer die stromsparendste Technik zum Einsatz. Oder ist es besser, dem alten PC regelmäßig eine Frischzellenkur zu gönnen – was nicht nur billiger käme, sondern vielleicht auch umweltschonender ist, weil kein neuer

PC zu produzieren und kein Alter zu entsorgen wäre?

Studieren geht über Probieren

Die Studie „Total Cost of Ownership & Green IT“ des Innsbrucker Kompetenzzentrum IT, der Fachhochschule Kufstein und des IZES Institut für Zukunftssysteme GmbH, Saarbrücken hat genau das untersucht. Ihre Fragestellung lautete: Wie ökonomisch und wie ökologisch wären Austausch oder Nachrüstung der kompletten Hardware nach jeweils drei Jahren, wenn man einen Zeitraum von neun Jahren ins Auge fasst?

„Die Beurteilung der wirtschaftlichen Situation bezieht sich hauptsächlich auf die Berechnung der Total Cost of Ownership.“, schreiben die Autoren. Ihre TCO-Berechnung stützte sich dabei auf elf einzelne Faktoren und beinhaltete alle Kosten, die im Lauf der Anschaffung, Nutzung und Entsorgung der Geräte für

die Unternehmen anfallen, einschließlich der Entsorgung des Elektroschrott. Die Rechnung bezieht auch indirekt zurechenbare Kosten wie Installations-, Wartungs-, Ausfalls- oder Risikokosten ein.

„Die Bewertung der Umweltkosten“, heißt es in der Studie weiter, „erfolgt hauptsächlich auf Basis einer CO₂-Berechnung.“ Die Übernahme das IZES (Institut für Zukunftssysteme), Saarbrücken. Dabei haben die Forscher versucht, nicht nur den CO₂-Ausstoß von diversen IT-Geräten während ihrer Nutzungsdauer zu betrachten, sondern auch das CO₂, das bei der Produktion und der Entsorgung der jeweiligen Geräte anfällt (**Abbildung 1**).

„Der zweite Hauptgesichtspunkt bei der Beurteilung der Auswirkungen auf die Umwelt ist die Umweltbilanz. Sie enthält Kosten, die bei der Produktion, Nutzung und Entsorgung von IT-Infrastruktur entstehen und die Umwelt schädigen, wie zum Beispiel auch die Kosten für nicht erneuerbare Rohstoffe. Auch die Entsor-

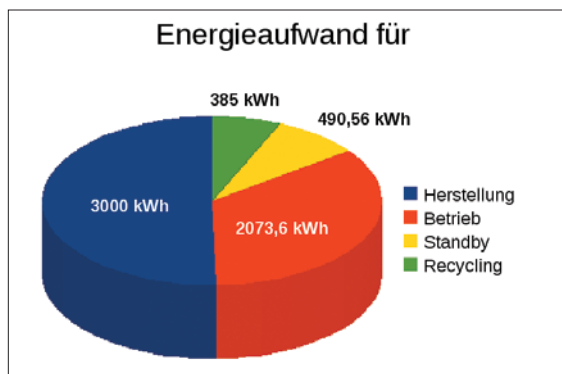


Abbildung 1: In die Bilanz der Forscher ging der CO₂-Austausch während des gesamten Lebenszyklus ein: Von der Herstellung über den Betrieb eines 120-Watt-Rechners, acht Stunden täglich über sechs Jahre hinweg, über das Stand-by (14 Watt, 16 Stunden am Tag) bis zum Recycling.

gung von Elektroschrott beziehungsweise giftigen Bestandteilen von Komponenten belastet die Umwelt und bringt Kosten mit sich, die zu berücksichtigen sind.“

Untersuchungsszenarien

Verständlicherweise konnte die Studie keine potenziell unbegrenzte Zahl von Einzelfällen untersuchen, sondern musste sich auf typische Szenarien beschränken. Dabei konzentrierten sich die Ersteller der Studie auf vier Größenklassen von Unternehmen, für die sie die in der **Tabelle 1** aufgeführte durchschnittliche Ausstattung annahmen. Bei der Ich-AG gingen sie davon aus, dass neben einem stationären PC auch ein Notebook im Einsatz ist (*). Für jede Komponente der Ausstattung legten die Analysten einen Stückpreis zugrunde, so dass sich Gesamtsummen für den Wert der IT zwischen 3400 (Einzelunternehmer) und knapp sechs Millio-

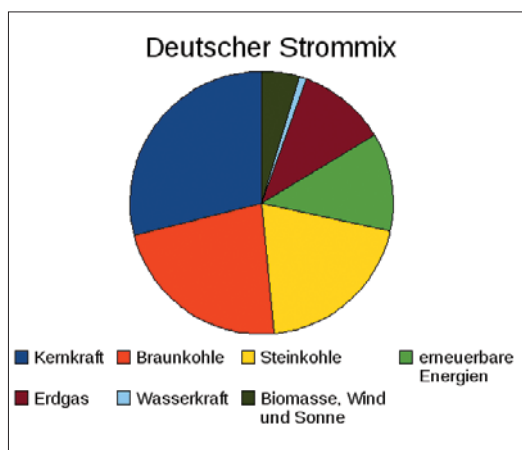


Abbildung 2: Aus diesen Quellen stammt der deutsche Strom. Seine Erzeugung kostet im Mittel 550 Gram CO₂ pro Kilowattstunde.

nen (Großunternehmen) Euro ergaben. Den gesamten Untersuchungszeitraum untergliederten sie in dreijährige Intervalle, weil nach dieser Frist häufig die Herstellergarantie auf Hardware ausläuft und die PCs als Hauptkomponente abgeschrieben sind. Weiter betrachtete die Studie pro Unternehmensklasse die beiden grundsätzlichen Handlungsalternativen Aufrüstung oder Neuanschaffung. Bei der Neu-

anschaffung fand auch die vermutliche Preisentwicklung Berücksichtigung. Die Variante Aufrüstung erwies sich als nicht ganz einfach zu berechnen, da die einzelnen Komponenten oft nicht unabhängig voneinander austauschbar sind. Für die TCO-Berechnungen der Ein-Mann-Firma, des Kleinunternehmens und des mittelständischen Unternehmens arbeiteten die Untersucher daher mit einem konkreten Aufrüstungsszenario, in das ein Austausch des Mainboards mit Onboard-Grafik- und -Soundkarte, des Prozessors, des Arbeitsspeichers (1024 MByte) und der Festplatte (320 GByte) eingerechnet war. Nach einer Preisrecherche kristallisierte sich heraus, dass die Hardware für die Aufrüstung etwa halb so teuer wie für eine komplette Neuanschaffung sein würde. Bei den Großunternehmen wendete man die Aufrüstung einzelner Komponenten nur auf die Server an – anstelle der Clientrechner kamen dort im Zuge der Aufrüstung Thin Clients mit einem Stückpreis von 300 Euro zum Einsatz.

TCO-Berechnungen

In die TCO-Berechnungen flossen direkte wie indirekte Kosten ein. Zu den direkten Kosten rechnen die Anschaffungskosten für neue Hardware beziehungsweise die Hälfte davon als Aufrüstungskosten. Weiter die Aufwendungen für Installation und Integration der neuen Technik und ebenso die für Abholung und Entsorgung der alten. Da ein

ESET-Virenschutz für die gesamte Firmen-IT

Damit Ihr Unternehmen kein Pflegefall wird



- Proaktive Erkennungstechnologien schützen auch vor neuen, noch unbekannten Bedrohungen
- Ausgezeichnete Performance, die Ihre Rechnerleistung nicht einschränkt
- Hohe Scangeschwindigkeit, kleine Updates.
- Intuitive Bedienbarkeit, zentrale Installation und Administration
- Zügiger, professioneller, deutschsprachiger Support

www.eset.de/partner



eset

www.eset.de

Anzeige wird
separat angeliefert

Anzeige wird
separat angeliefert

Teil der alten Geräte noch verwendbar sein dürfte und verkaufbar ist, sieht das Rechenschema auch Aufwendungen wie Erlöse für einen solchen Verkauf vor.

Indirekte Kosten entstehen etwa durch die notwendige Recherche im Zuge der Investitionsvorbereitung, gegebenenfalls durch externe Beratung, weiter durch die nötige Schulung von Mitarbeitern und Administratoren. Hinzukommen Betriebskosten und auch die Ausfallkosten, die eine Umrüstung der Hardware mit sich bringen könnte. Die Studie veranschlagt sie allerdings nie, weil man sie auch in der Realität unter allen Umständen vermeiden würde.

Umweltfreundlichkeit

Wie beurteilten die Forscher nun die Umweltfreundlichkeit jeder Lösung? Sie

schreiben: „In Bezug auf die Umweltfreundlichkeit sollen die Computerbestandteile während eines kompletten Lebenszyklus eines Computers auf ihren Energieverbrauch und ihre CO₂-Emissionen untersucht werden. Der komplette Lebenszyklus umfasst die Gewinnung der Rohmaterialien beziehungsweise die Bereitstellung der verwendeten Materialien, die Produktion der Einzelkomponenten und den Zusammenbau dieser Komponenten zu einem vollständigen PC, den Gebrauch des Gerätes und nach seiner Nutzung dessen Entsorgung mit Recycling und Deponierung der nicht verwertbaren Bestandteile.“

Die Rechnung setzt dabei voraus, dass neue Technik grundsätzlich energiesparender arbeitet. Daraus leitet sich ein CO₂-Einspareffekt ab. Legt man den deutschen Strommix zugrunde ([Abbildung](#)

2), dann ergeben sich einer Greenpeace-Studie zufolge CO₂-Emissionen von 550 Gramm pro Kilowattstunde. Von diesem Wert gehen die Berechnungen der Studie aus.

Methodisch stützt sich die Studie sowohl auf die direkte Erhebung von Daten zu durchschnittlicher Ausrüstung, Stromverbrauch, Alter, Lebensdauer und Betriebsstunden der technischen Ausrüstung jeweils eines konkreten Unternehmens pro untersuchter Größenklasse, als auch auf Literatur- und Internet-Recherche. Dabei verglichen die Wissenschaftler und Studenten eine ganze Reihe verschiedener Quellen. Nicht selten standen dabei eigene Messergebnisse im Widerspruch zu Herstellerangaben.

TCO-Berechnungen

Bei den Energieverbrauchsmessungen an PCs diente ein Acer Aspire M3610A als Referenzmodell, nachdem er sich mit 75 Watt im Betrieb und nur 6 Watt im Stand-by-Modus als das energiesparendste Modell erwiesen hatte. Die Verbrauchswerte aller im Kleinunternehmen gemessenen Modelle lagen zwischen 91 und 170 Watt im Betrieb. Sie unterschieden sich auch im Stand-by-Modus deutlich, in dem immerhin bis zu 9 Watt anfielen, was sich merklich in der Stromrechnung niederschlagen dürfte. Erwartungsgemäß waren neuere Modelle effizienter, insbesondere der Arbeitsspeicher hatte einen höheren Wirkungsgrad. Da die Rechner des Mittelständlers im Vergleich mit dem Kleinunternehmen neuer waren, wiesen sie auch eine bessere Energiebilanz auf (Leistungsaufnahme 74 bis 96 Watt bei durchschnittlicher Last, unter Maximallast bis 140 Watt).

Wesentlich größere Unterschiede ergaben sich bei den Servern, wo die Differenz zwischen geringstem und höchstem Stromverbrauch im Schnitt bei 251 Watt, unter Volllast sogar bei 448 Watt lag. Im Fall des Großunternehmens wurde zusätzlich die Alternative Thin Client betrachtet, bei der nur ein Viertel des Energieverbrauchs eines durchschnittlichen PC anfällt.

Die TCO-Berechnungen nahmen für alle vier Größenklassen an, dass das Unternehmen entweder alle drei Jahre die alte gegen neue Hardware tauschte oder nur

Tabelle 1: IT-Ausstattung pro Größenklasse

	Konzern	Mittelständler	Kleinunternehmen	Ich-AG
Arbeitsstation (mit Monitor)	10	50	10	1*
Arbeitsplatzdrucker	1 300	15	4	1
Scanner	50	1	1	0
Sicherung	50	1	1	1
Netzwerkdrucker	100	1	0	0
Fileserver	50	1	1	0
Messaging Server	10	1	0	0
Terminal-Server	100	0	0	0

Tabelle 2: TCO für den Alleinunternehmer

	Neuanschaffung	Aufrüstung
Direkte Kosten		
Hardware und Installation		
Hardware	6000	2250
Installation und Integration	840	945
Entsorgung		
Abholung und Entsorgung	24	1
Verkaufseinnahmen	0	0
Verkaufsaufwand	0	0
Indirekte Kosten		
Informationsrecherche		
interner Aufwand	0	0
externe Beratung	420	210
Schulungskosten	0	0
Mitarbeiterschulungen	405	0
Administratorschulungen	0	0
Betriebskosten		
Stromkosten	157	303
Produktionsausfallkosten	0	0
TCO	7 847	3 709

aufrüstete. Für den Alleinunternehmer hätte das Aufrüsten nach neun Jahren eine Ersparnis von insgesamt über 4000 Euro ergeben, wobei die Forscher unterstellten, dass er sich bei der Neuanschaffung hätte beraten und schulen lassen müssen (Tabelle 2).

Auch das Kleinunternehmen hat wie die Ich-AG keine eigene IT-Abteilung, muss aber mindestens einen Admin schulen. Die Schulungskosten bei Neuanschaffung liegen dabei in derselben Größenordnung wie die mit der moderneren Hardware eingesparten Stromkosten. Bleiben allerdings die viel höheren Hardwarekosten, die eine regelmäßige Neuanschaffung um mehr als 17 000 Euro gegenüber der Aufrüstung verteuern (Tabelle 3).

Der Mittelständler in der Studie verfügt über eine eigene IT-Abteilung, die auch die nötige Recherche und Beratung übernehmen kann, für die in diesem Fall ein interner Aufwand entsteht. Mehr neue Rechner sparen natürlich auch mehr Strom, so dass dieser Faktor hier erstmals mit fast 1 000 Euro im Jahr ins Gewicht fällt. Trotzdem ist auch hier die Version mit Aufrüstung statt Neukauf deutlich günstiger (Tabelle 4).

Der Großkonzern steigt im Modell der Studie auf Thin Clients um. Weil die nur ein Viertel des Stroms brauchen, den ein PC benötigt, lassen sich in diesem Maßstab fast 100 000 Euro Stromkosten einsparen, aber auch das reicht lange nicht, um die Mehrkosten der Neuanschaffung zu kompensieren (Tabelle 5).

Teurer, aber ökologischer?

Wie sich zeigt, ist in jeder Größenklasse die komplette Neuausrüstung im Drei-Jahres-Takt immer der teurere Weg ist (Tabelle 6). Das heißt ja aber noch nicht, dass damit nicht vielleicht so gravierende ökologische Nutzeffekte verbunden sein könnten, dass diese Variante dennoch gerechtfertigt scheint. Ob dem so ist, sollte eine Untersuchung der ökologischen Auswirkungen der beiden untersuchten Strategien zeigen.

Der Analyse der ökologische Effekte diene im Rahmen der Studie eine Untersuchung des kompletten PC Lebenszyklus. „Das Ziel dieser Lebenszyklusanalyse“, schreiben die Autoren, „ist die Identifikation der Computerbestandteile

und Analyse des Energieverbrauchs und der CO₂-Emissionen während eines kompletten Lebenszyklus eines Computers. Diese Betrachtung beschränkt sich auf die Zusammenstellung des Energieaufwandes und der entstehenden CO₂-Emissionen in jedem Lebensabschnitt.“

Der gesamte Herstellungsprozess eines neuen Computers verbraucht 2136 MJ

Energie und produziert rund 147 kg CO₂-Emissionen. Der größte Teil davon (1447 MJ Energie und 98,92 kg CO₂) entfällt dabei auf die Herstellung der Ausgangsmaterialien, die eigentliche Produktion des Endprodukts benötigt nur noch 286 MJ Energie und setzt 17,03 kg CO₂ frei, der Rest ist der Distributionsphase zuzurechnen.

Tabelle 3: TCO für den Kleinbetrieb

	Neuanschaffung	Aufrüstung
Direkte Kosten		
Hardware und Installation		
Hardware	32 700	13 800
Installation und Integration	4200	5775
Entsorgung		
Abholung und Entsorgung	102	16
Verkaufseinnahmen	0	0
Verkaufsaufwand	0	0
Indirekte Kosten		
Informationsrecherche		
interner Aufwand	0	0
externe Beratung	420	336
Schulungskosten		
Mitarbeiterschulungen	540	0
Administratorschulungen	270	0
Betriebskosten		
Stromkosten	3454	4468
Produktionsausfallkosten	0	0
TCO	41 686	24 395

Tabelle 4: TCO für den Mittelständler

	Neuanschaffung	Aufrüstung
Direkte Kosten		
Hardware und Installation		
Hardware	80 100	58 800
Installation und Integration	16 800	22 260
Entsorgung		
Abholung und Entsorgung	420	78
Verkaufseinnahmen	0	0
Verkaufsaufwand	0	0
Indirekte Kosten		
Informationsrecherche		
interner Aufwand	1050	1365
externe Beratung	0	0
Schulungskosten		
Mitarbeiterschulungen	675	0
Administratorschulungen	540	0
Betriebskosten		
Stromkosten	11 613	20 126
Produktionsausfallkosten		
TCO	111 198	102 629

Diese Werte lassen sich auf einzelne Komponenten herunterbrechen, wie **Tabelle 7** zeigt. Den größten Anteil am Energieverbrauch wie am CO₂-Ausstoß hat das Mainboard, gefolgt vom Netzteil des Rechners.

Die Gebrauchsphase beurteilten die Untersucher anhand der oben bereits erwähnten Energieverbrauchsmessungen im Betrieb. Dabei ergaben sich als Mittelwerte für die Leistungsaufnahme pro Stunde bei einem PC 81 Watt, bei einem Server 188 Watt, für einen Bildschirm 37 Watt, für einen Drucker 300 Watt oder 46 Watt für einen Laptop. In dem untersuchten mittelständischen Unternehmen summierten sich die Werte aller EDV-Geräte zu einem Stromverbrauch von 1 319 Watt pro Stunde, was 87 Prozent des Stromverbrauchs der gesamten Firma entsprach. Einsparpotenziale liegen, so die Autoren, vor allem im Stand-by-Betrieb.

Was bedeutet dies nun im Hinblick auf die Entscheidung zwischen Neukauf und Ausrüstung. Die Autoren der Studie ziehen nach umfangreichen Berechnungen folgendes Fazit: „Aus Sicht der Umwelt wirkt sich die Neuanschaffung eines energieeffizienten Computers im Vergleich zu einer Aufrüstung nur in den ersten drei Jahren und sieben Monaten negativ aus. Nach diesem Zeitraum macht sich die Energieeffizienz des neuen Rechners bemerkbar. Demnach ist eine Entscheidung zwischen Neuanschaffung und Aufrüstung aus ökologischer Sicht in den definierten Betrachtungszyklen von 3 Jahren nicht eindeutig möglich. Nach den ersten drei Jahren soll aufgerüstet und nach den zweiten 3 Jahren neu angeschafft werden. Damit sollten sich die beiden Varianten, Neuanschaffung und Aufrüstung, in Hinsicht auf ein ökologisches Optimum abwechseln.“

Die Autoren warnen davor, in die Falle zu tappen: „Weitere Überlegungen führen jedoch dazu, dass die Entscheidung zur Aufrüstung nach 3 Jahren nicht die ökologischste Variante darstellt. Dies liegt in der Tatsache begründet, dass die nächsten 3 Jahre nur zu einem kleinen Teil, nämlich 7 Monate, noch als energieeffizient zu bewerten sind. In den restlichen 2 Jahren und 5 Monaten befindet sich der jeweilige Rechner aus ökologischer Sicht in einem ineffizienten Status. Aus diesem Grund kann aus der vorliegenden

Studie geschlossen werden, dass im ökologischen Sinne eine Neuanschaffung alle 3 Jahre über einen Betrachtungszeitraum von 9 Jahren gegenüber der Aufrüstung zu bevorzugen ist.“ Das Optimum, so die Studie, läge bei einem Neuanschaffungszyklus von 3 Jahren und 7 Monaten.

Fazit

Damit endet der Vergleich in einem Patt: Ökonomisch schneidet das Aufrüsten besser ab, es ist in jedem Fall günstiger.

Der Umwelt kommt der Neukauf aller Rechner eher entgegen. Zwar resultiert aus der Energieeffizienz neuer Hardware kein ökonomisches Plus, wohl aber ein ökologisches.

Eine ganze Reihe von Faktoren konnten aus unterschiedlichen Gründen in der Studie keine Berücksichtigung finden. Wohl aber lassen sich weitere Kriterien in individuelle Untersuchungen im Rahmen konkreter Projekte aufnehmen. Dabei kann die Methodik der vorliegenden Arbeit als Vorlage dienen. (jcb/ofr) ■

Tabelle 5: TCO für den Konzern

	Neuanschaffung	Aufrüstung
Direkte Kosten		
Hardware und Installation		
Hardware	17 787 000	9 880 000
Installation und Integration	1 011 500	1 080 100
Entsorgung		
Abholung und Entsorgung	69 960	54 190
Verkaufseinnahmen	0	0
Verkaufsaufwand	0	0
Indirekte Kosten		
Informationsrecherche		
interner Aufwand	136 500	115 500
externe Beratung	0	0
Schulungskosten		
Mitarbeiterschulungen	162 000	54 000
Administratorschulungen	40 500	13 500
Betriebskosten		
Stromkosten	571 428	666 862
Produktionsausfallkosten	0	0
TCO	19 778 888	11 864 152

Tabelle 6: Kostennachteil bei Neuanschaffung

	Neuanschaffungskosten vs. Aufrüstung in Prozent
Ich-AG	211,48
Kleinbetrieb	170,87
Mittelständler	108,35
Konzern	166,71

Tabelle 7: PC-Komponenten

Komponente	Energieaufwand (MJ)	CO ₂ -Emission (kg)
Mainboard	639	44
Power Supply	351	24
CD/DVD	158	10
Festplatte	117	9
Grafikkarte	181	13
Floppy	58	4
Verpackung	28	1
LCD-Monitor	149	9

ADMIN

Netzwerk & Security

Alle Artikel des
Jahres 2010

JAHRES-DVD 2010



- Artikel zu Storage, Backup, Netzwerk, Monitoring, Virtualisierung u.v.m.
- PDFs zum Lesen am Bildschirm oder Ausdrucken
- Search Engine für Artikel-Volltext-Suche
- Auf der zweiten Seite der DVD: Release-Kandidat von Debian 6.0 „Squeeze“

 **debian 6.0** -BETA-

JETZT BESTELLEN:

• www.admin-magazin.de/DVD2010

• Telefon 07131 / 2707 274 • Fax 07131 / 2707 78 601
• E-Mail: info@admin-magazin.de • Shop: Shop.LinuxNewMedia.de

**NUR
€14,95**



Energiefresser unter Linux mit PowerTOP enttarnen

Auf frischer Tat ertappt

Zahlreiche Anwendungen stören den Prozessor bei seinen Strom sparenden Nickerchen. PowerTOP nennt diese Übeltäter nicht nur beim Namen, sondern liefert auch Hinweise auf ihre Entsorgung. Die damit eingesparten Energiekosten dürften Notebook-Nutzer und Server-Betreiber freuen. Tim Schürmann

Sobald es für einen Prozessor nichts mehr zu tun gibt, versetzt er sich automatisch in einen von mehreren Schlafzuständen. Auf modernen Prozessoren passiert das gleich mehrfach pro Sekunde. Je tiefer die CPU dabei einschläft, desto weniger Strom benötigt sie. Folglich sollte man sie nur dann wecken, wenn es auch wirklich notwendig ist. Dummerweise rütteln einige selbstsüchtige Prozesse und Gerätetreiber den Prozessor immer wieder aus seinem Schlaf und treiben so die Stromkosten unnötig nach oben.

Detektei

Genau diese Störenfriede ermittelt das kleine Werkzeug PowerTOP auf Linux-Systemen. Es beobachtet eine Weile lang alle laufenden Prozesse und Gerätetreiber, schätzt den durch ihren Einsatz verursachten Energieverbrauch und präsentiert dann die mutmaßlich größten Stromverschwender. Obendrauf liefert die eingebaute Datenbank weitere Ener-

giespartipps, die PowerTOP mit einem Tastendruck sogar direkt anwendet. Damit erhöhen insbesondere Notebook-Besitzer die Batterielaufzeit, während Server-Betreiber ihre Energiekosten senken – gerade bei Server-Farmen ein Hauptkostenfaktor. Darüber hinaus können Software-Entwickler schon frühzeitig prüfen, ob ihre Software den Prozessor zu oft unnötig drangsaliert.

PowerTOP wurde 2007 von Intel ins Leben gerufen, läuft mit kleineren Einschränkungen aber auch auf AMD-, ARM- und UltraSPARC-Prozessoren. Das Werkzeug steht unter der freien GPL-v2-Lizenz und liegt mittlerweile jeder größeren Linux-Distribution bei. In den Repositories lagert allerdings häufig nur eine ältere Version. Da neuere grundsätzlich mehr Energiespartipps kennen, sollte man sich im Zweifel das aktuelle Quellcode-Archiv von der PowerTOP-Homepage angeln [1] und selbst übersetzen. Dazu benötigt man lediglich den C-Compiler, »make«, »gettext« sowie die Entwicklerpakete zur

»libncursesw«. Das »w« ist übrigens kein Tippfehler: Es handelt sich hierbei um die Wide-Character-Version der bekannten »ncurses«-Bibliothek. Unter Ubuntu benötigt man deshalb neben dem Paket »libncurses5-dev« auch noch seinen Kollegen »libncursesw5-dev«, während Open Suse alles Notwendige im »ncurses-devel«-Paket zusammenschnürt. Ein schlichter Aufruf von PowerTOP übersetzt und installiert schließlich »make«, gefolgt von »sudo make install«. Bevor man das fertige Werkzeug starten kann, müssen allerdings noch ein paar Voraussetzungen erfüllt sein.

Taktstock

Ältere Linux-Kernel nutzten bis zur Version 2.6.20 einen festen Herzschlag mit 1000 Hz. Mit jedem Schlag (Tick) weckte der Kernel den Prozessor und prüfte, ob irgendwelche Aufgaben anstanden. Energiesparen war so nur eingeschränkt möglich. Ab dem Kernel 2.6.21 gibt es

```

ubuntu@ubuntu: ~
Datei Bearbeiten Ansicht Suchen Terminal Hilfe
PowerTOP version 1.13 (C) 2007 Intel Corporation

Cn      Verweildauer      P-States (Frequenzen)
C0 (Prozessor läuft) ( 2,6%)      2,00 GHz      2,0%
zyklisches AbfräC1 mwait      0,0m      1,87 GHz      0,0%
C1 mwait      0,0ms ( 0,0%)      1,60 GHz      98,0%
C2 mwait      0,1ms ( 0,2%)
C4 mwait      16,2ms (97,2%)

Aufwachen pro Sekunde : 84,6      Intervall: 15,0s
Stromverbrauch (ACPI-Schätzung): 16,7W (1,6 Std.)

Häufigste Ursachen für das Aufwachen:
45,6% ( 34,5) [kernel scheduler] Load balancing tick
13,6% ( 10,3) [extra timer interrupt]
6,6% ( 5,0)  syndaemon
4,5% ( 3,4)  [ohci_hcd:usb3] <interrupt>
0,0% ( 0,0)  gvfsd-metadata
4,0% ( 3,1)  USB Gerät 3-1 : Apple Optical USB Mouse (Mitsumi Electric)

Vorschlag: Deaktivieren Sie die ungenutzte Bluetooth-Schnittstelle mit
folgendem Befehl: hciconfig hci0 down ; rmmod hci_usb
Bluetooth verbraucht zum Senden Strom und belastet die USB-Schnittstelle.
Q - Beenden      R - Aktualisieren      B - Bluetooth ausschalten

```

Abbildung 1: Auf einem Macbook unter Ubuntu 10.10 empfiehlt PowerTOP, die (ungenutzte) Bluetooth-Schnittstelle zu deaktivieren und nennt auch gleich noch den passenden Befehl dafür. Je größer das Terminalfenster ist, desto mehr Übeltäter zeigt PowerTOP an.

diese regelmäßigen Ticks nicht mehr. Stattdessen klärt Linux kurz, wann in der Zukunft neue Aufgaben anstehen und stellt dann einfach den in modernen PCs verbauten Hardware-Wecker (in Form des High Precision Event Timers, kurz HPET). Der Prozessor kann sich dank dieser sogenannten „dynamischen Ticks“ (kurz Dyticks) endlich für längere Perioden schlafen legen.

PowerTOP setzt deshalb einen solchen „tickless“-Kernel voraus. Ein 32-Bit-Kernel muss folglich mindestens Version 2.6.21 tragen, ein 64-Bit-Kernel hingegen die Versionsnummer 2.6.23. Ob das der Fall ist, verrät etwa »uname -r«. Auf der sicheren Seite ist man ab Open Suse 11.0 und Ubuntu 8.04. Zusätzlich muss im Kernel die Unterstützung für die dynamischen Ticks aktiviert sein. Bei allen großen Distributionen ist dies durchweg der Fall, die dabei vorgenommenen Einstellungen verrät **Kasten „Einstellungssache“**.

Top Sache

Sind alle Voraussetzungen erfüllt, startet PowerTOP ein simples

```
sudo powertop
```

Man kann das Werkzeug auch als normaler Nutzer aufrufen, sieht dann aber nur einen Bruchteil der Informationen und darf zudem keine Verbesserungsvorschläge anwenden.

PowerTOP sammelt jetzt für die Dauer von fünf Sekunden Informationen und

```

tim@ubuntu: ~
Datei Bearbeiten Ansicht Suchen Terminal Hilfe
PowerTOP version 1.13 (C) 2007 Intel Corporation

P-States (Frequenzen)
2,13 GHz      10,4%
1,60 GHz      89,6%

Aufwachen pro Sekunde : 52,3      Intervall: 3,0s
Keine ACPI Stromverbrauch-Schätzung verfügbar

Häufigste Ursachen für das Aufwachen:
25,3% ( 27,7) [ehci_hcd:usb2, uhci_hcd:usb5, EMU10K1] <interrupt>
22,9% ( 25,0) USB Gerät 2-5 : USB Mass Storage Device (USBest Technology)
12,8% ( 14,0) [kernel scheduler] Load balancing tick
10,7% ( 11,7) [kernel core] hrtimer_start (tick_sched_timer)
6,1% ( 6,7)  [uhci_hcd:usb4, pata_jmicron, hda_intel] <interrupt>
3,7% ( 4,0)  [uhci_hcd:usb3, ahci, nvidia] <interrupt>

A SATA device is active 100,0% of the time:
SAMSUNG SP2504C

Q - Beenden      R - Aktualisieren      S - SATA Link Stromsparfunktionen einsc

```

Abbildung 2: Auf Desktops und Servern fehlen die Angaben links oben zu den C-States, bei Kabelbetrieb zusätzlich noch die **Stromverbrauch-Schätzung**.

meldet sich dann mit dem etwas

kryptischen Ergebnis aus **Abbildung 1**. Auf einem Notebook – und nur dort – zeigt PowerTOP links oben an, wie lange der Prozessor während der fünf Sekunden durchschnittlich wie tief geschlafen hat. Das »C« gefolgt von einer Zahl steht dabei jeweils für einen von mehreren Stromsparszuständen, den sogenannten C-States. Je höher die Nummer hinter dem »C«, desto weniger Energie verbraucht der Prozessor, aber desto länger benötigt er auch, um wieder aufzuwachen. Im Zustand C0 rennt der Prozessor mit voller Kraft und führt Anweisungen aus. Im Zustand C1 befindet er sich im Leerlauf, in den übrigen legt er nach und nach immer mehr Bereiche und Rechenwerke still. Die meiste Energie spart der Prozessor in den Zuständen C3 und C4, folglich sollte er darin auch die meiste Zeit verbracht haben. Im Idealfall schläft die CPU 95 Prozent ihrer Zeit in einem dieser beiden

Zustände. Welche der C-States PowerTOP anzeigt, hängt übrigens vom BIOS ab. Einige verstecken beispielsweise den C3- und C4-Zustand, sobald das Notebook am Stromkabel hängt.

Moderne Prozessoren können zusätzlich noch ihre Taktfrequenz automatisch schrittweise senken beziehungsweise erhöhen. Intel nennt dieses Technik Speedstep. Welche Frequenzen das sind und wie lange der Prozessor mit ihnen gerechnet hat, verraten rechts oben die sogenannten Performance-States, kurz »P-States«. Meist, aber nicht grundsätzlich, gilt: Je langsamer die CPU taktet, desto weniger Energie verbraucht sie.

Nervende Bande

Unter den Zahlen markiert ein farbiger Balken, wie oft der Prozessor in einer Sekunde durch irgendetwas oder irgendetwas geweckt wird (*Wakeups per second*). Je kleiner die Zahl ist, desto besser. Mit einem Gnome-Desktop sollen im besten Fall drei Weckrufe pro Sekunde erreichbar sein, gute Werte liegen in der Praxis unter 20. Die Balkenfarbe visualisiert übrigens noch einmal gute und schlechte Werte: Bei einem knallroten Balken herrscht dringender Handlungsbedarf, ein gelber Ton deutet auf eine akzeptable, energie-sparendere Situation hin.

Wenn PowerTOP auf einem mobilen Gerät läuft und dieses gerade seine Batterie leersaugt, erfährt der Anwender direkt unter dem Balken, wie viel Watt der Computer gerade verbraucht und wie lange der Strom wahrscheinlich noch ausreicht.

Einstellungssache

Wer einen neuen Kernel übersetzen möchte, sollte laut PowerTOP-FAQ folgende Einstellungen aktivieren:

```
CONFIG_NO_HZ
CONFIG_HIGH_RES_TIMERS
CONFIG_HPET_TIMER
CONFIG_CPU_FREQ_GOV_ONDEMAND
CONFIG_USB_SUSPEND
CONFIG_SND_AC97_POWER_SAVE
CONFIG_TIMER_STATS
```

und die folgenden (sofern vorhanden) deaktivieren:

```
CONFIG_IRQBALANCE
CONFIG_ACPI_DEBUG
```

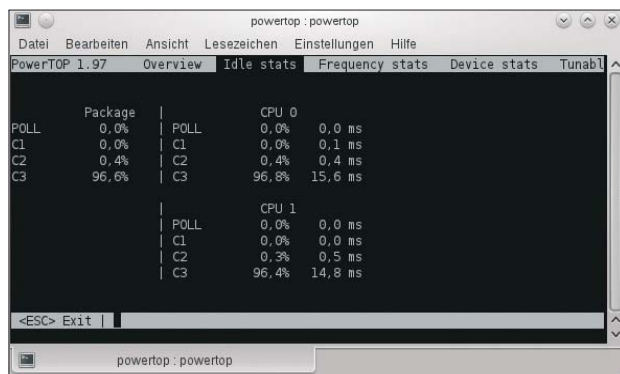


Abbildung 3: PowerTOP 2.0 nennt die C-States für jeden vorhandenen CPU-Kern.

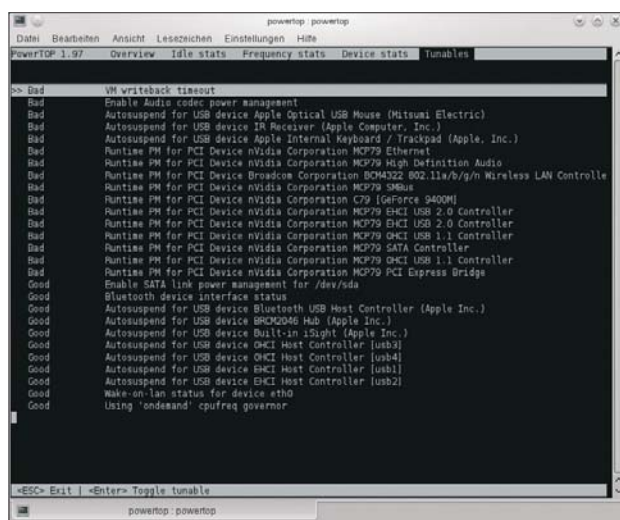


Abbildung 4: Unter **Tunables** findet man unter PowerTOP 2.0 jetzt alle Stromspartipps gesammelt. Das »VM« im ersten Tipp steht übrigens für Virtual Memory und nicht etwa für Virtual Machine.

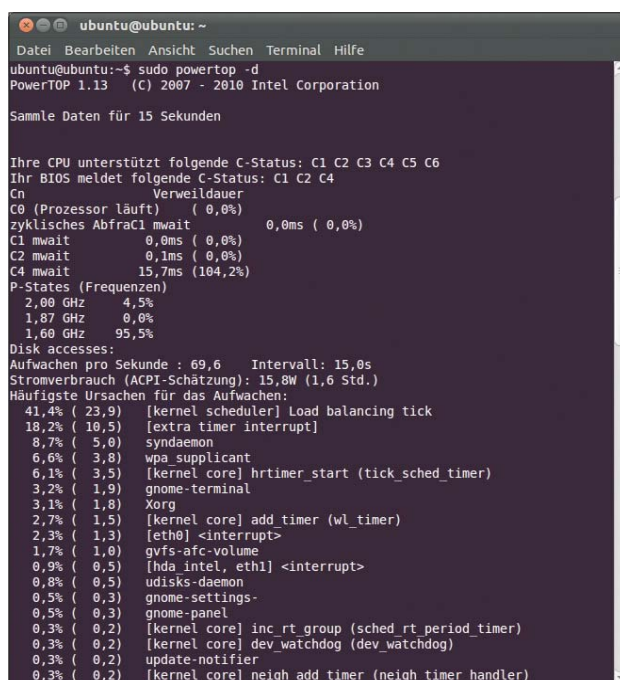


Abbildung 5: Mit dem Parameter »-d« liefert PowerTOP einen ausführlichen Bericht ...

Wer oder was das System in den fünf Sekunden ordentlich auf Trab gehalten hat, steht unter den »Häufigsten Ursachen für das Aufwachen« (in der englischen Version »Top causes for wakeups«). Die fettgedruckten Verursacher hält PowerTOP für die größten Stromfresser. Neben den Prozessen tauchen in der Liste auch Geräte und Kernel-Funktionen auf. Die Interpretation ist allerdings nicht immer ganz einfach. Beispielsweise steckt hinter dem ominösen »i8042« ein Kernel-Treiber für Tastatur und Maus. Im Zweifelsfall hilft eine kurze Google-Anfrage beim Verständnis.

CPU-belastende Geräte sind in Notebooks vor allem die WLAN- und Bluetooth-Komponenten. Sie erkennt man in der Liste schnell am vorangestellten Schnittstellennamen. Um sie abzuschalten, genügt es häufig schon, das entsprechende Kernel-Modul per »sudo modprobe -r modulname« zu entfernen und eventuell einen zugehörigen (Bluetooth-)Daemon herunterzufahren. Einige Notebooks besitzen spezielle Tasten, die WLAN und Bluetooth deaktivieren. Ob es geklappt hat, erfährt man wieder über PowerTOP. Besonders fies sind zudem USB-Adapter, da sie auch gleich noch das USB-Subsystem unter Strom halten.

Schuldfrage

Taucht der X-Server (»Xorg«) in der Liste auf, muss nicht immer er selbst der Schuldige sein. Für gewöhnlich stellt hier ein anderes Programm ständig Anfragen an den X-Server, der dann seinerseits zwangsweise den Prozessor weckt. Auslöser kann hier schon ein blinkender Cursor in einem Terminal sein. Bei einem störenden Interrupt kommt man dem Übeltäter am schnellsten auf die Spur, indem man nach und nach die von ihm genutzten Kernel-Module entlädt. Nach jeder Aktion kontrolliert man in PowerTOP erneut den Status.

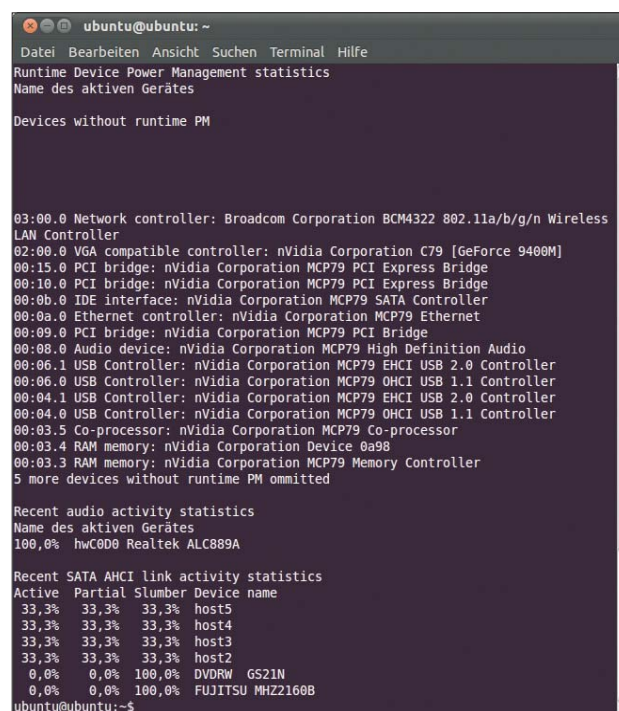


Abbildung 6: ... der zahlreiche zusätzliche Informationen enthält.

SELBST IST DER ADMIN!

Mit der **MobyDick Telefonanlage** haben Sie alle Fäden selbst in der Hand. Außerdem verbinden Sie die Features modernster **Telefonie** mit den Vorteilen von **Voice over IP**.

NEU
Kostenlose
Community
Version
erhältlich

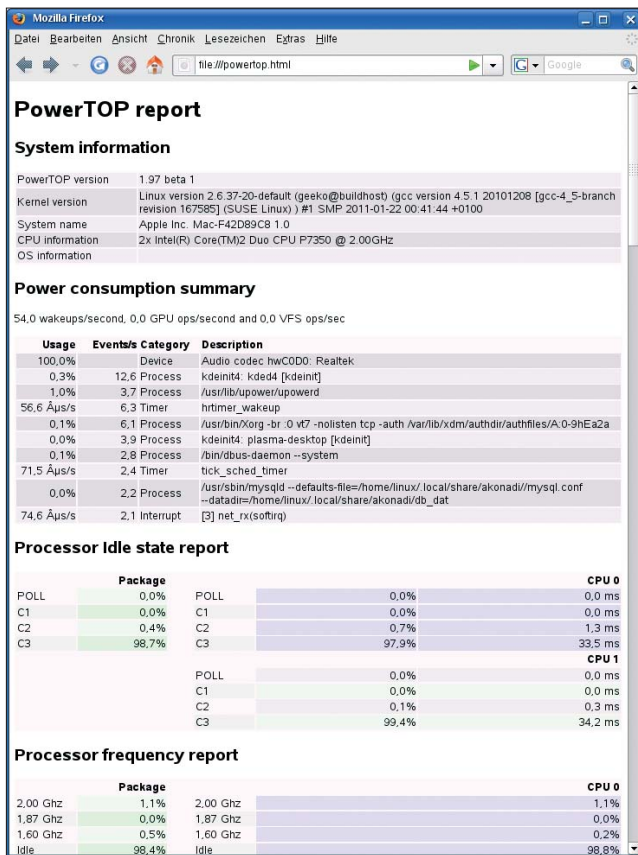


Abbildung 7: PowerTOP 2.0 liefert seine Berichte übersichtlich als HTML-Dokument.

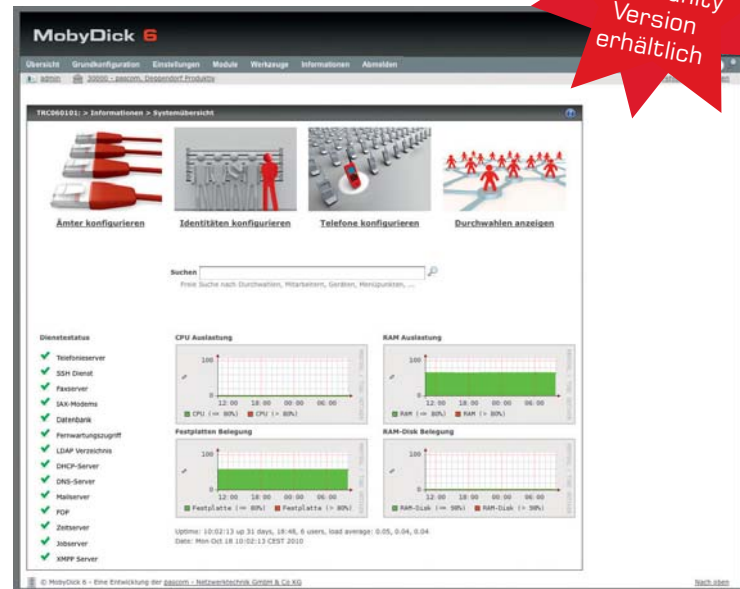
Bevor man einen identifizierten Prozess oder ein Kernel-Modul in der Liste endgültig abschießt, sollte man sich allerdings gut überlegen, ob er wirklich nicht mehr gebraucht wird oder ob es nicht vielleicht schon eine neue Version des Programms gibt, die schonender mit den Ressourcen umgeht.

Weiser Mann

Ganz am unteren Rand gibt PowerTOP noch einen Stromspartipp. Ihn setzt man direkt über die angezeigte Taste rechts unten um. Beispielsweise aktiviert [S] den SATA-Stromsparmodus, [B] knipst der Bluetooth-Schnittstelle das Licht aus, [A] schaltet die HD-Audio Stromsparfunktionen ein, während [K] die wiederholten Abfragen an das CD-Laufwerk unterbindet.

Allerdings sollte man dabei seinen Verstand nicht ab- und einfach alle Vorschläge anschalten. Bildet auf dem Notebook das WLAN den (derzeit) einzigen Weg nach draußen, sollte man es vielleicht doch besser aktiviert lassen, während wiederum auf einem Server das Einschalten des Festplatten-Powermanagements bremsend und somit kontraproduktiv sein kann (Abbildung 2).

Einige dieser Maßnahmen bleiben dauerhaft aktiv, wie etwa Udev-Regeln, andere nur bis zum nächsten Systemstart. Um Letzteres zu verhindern, nimmt man den vorgeschlagenen Befehl in die Startskripte auf. Für gewöhnlich bietet sich dazu »/etc/rc.local« an. Auf Notebooks ist es eine Überlegung wert, die Befehle in einem Skript zu sammeln und dieses nur bei Bedarf aufzurufen – beispielsweise wenn man von Kabel auf Batteriebetrieb wechselt.



DAS BESTE AUF DEN PUNKT GEBRACHT.

- Einfache Administration
- Intuitive Bedienbarkeit
- Hoch skalierbar von 1 bis 1000 User pro Server
- Standortvernetzung und Home Office Integration
- Quelloffen anpassbar
- auf OpenSource basierend
- Hochverfügbar
- Günstiger als Sie glauben
- Unified Communications:

- Telefon
- Video
- VoiceMail
- Präsenzmanager
- Instantmessaging
- FaxEmail Gateway
- PrintFax Gateway
- Conferencing

Mehr Informationen finden Sie unter:

<http://www.pascom.net>

<http://community.pascom.net>



pascom

Netzwerktechnik GmbH & Co. KG

Berger Straße 42

94469 Deggendorf

Tel.: +49 991 27006 - 0

PowerTOP läuft kontinuierlich weiter und aktualisiert alle fünf Sekunden seine Anzeigen. Damit ändert sich unter Umständen auch der Tipp am unteren Ende. Per [R] kann man diese Aktualisierung erzwingen, [Q] oder [Strg] + [C] beendet das Werkzeug.

Next Generation

Mit dem Erscheinen dieser Ausgabe dürfte bereits die komplett neu programmierte PowerTOP-Version 2.0 erhältlich sein. Sie bringt eine runderneuerte Benutzeroberfläche mit und unterstützt die neuen Stromsparfunktionen des Kernel 2.6.37. Ihn nutzen unter anderem die kommenden Distributionen Ubuntu 11.04 und Open Suse 10.4. Des Weiteren greift PowerTOP 2.0 auf die „perf“-Infrastruktur des Kernels zurück, um genauere Angaben zum Energieverbrauch liefern zu können. Mit genügend gesammelten Daten sagt die Neuauflage zudem den Stromkonsum jedes einzelnen Geräts voraus, wenngleich Chefentwickler Arjan van de Ven diese Funktion noch als experimentell bezeichnet. Um PowerTOP 2.0 sinnvoll nutzen zu können, benötigt man allerdings mindestens einen Kernel 2.6.36, besser den Nachfolger 2.6.37. Wer einen älteren Kernel betreibt, sollte beim alten PowerTOP 1.x bleiben.

Bis zum Redaktionsschluss musste man sich den Quellcode der BETA-Version von PowerTOP 2.0 noch unter [2] herunterladen. Um ihn zu übersetzen, sind der GNU-C++-Compiler sowie

die Entwicklerpakete zu »libstdc++«, »pciutils«, »ncurses«, »zlib« und »libnl« notwendig. Unter Open Suse stecken sie in den Paketen »gcc-c++«, »pciutils-devel«, »ncurses-devel«, »zlib-devel« und »libnl-devel«, Ubuntu-Nutzer installieren hingegen »g++«, »libpci-dev«, »libncurses5-dev«, »libncursesw5-dev«, »zlib1g-dev« und »libnl-dev«. Um PowerTOP zu übersetzen und einzuspielen, reicht wieder ein »make«, gefolgt von »sudo make install«.

Wachposten

Wie seinen Vorgänger startet man PowerTOP 2.0 via »sudo powertop«, woraufhin die etwas aufgeräumtere Oberfläche aus **Abbildung 3** erscheint. Sie verteilt alle Informationen auf mehrere Register, zwischen denen man mit den Pfeiltasten nach links und rechts umschaltet. Der Karteireiter »Overview« führt zunächst wieder die Anzahl der Weckrufe pro Sekunde auf, darunter alle stromfressenden Quälgeister. »Idle Stats« listet die C-States, »Frequency Stats« die P-States auf – diesmal jedoch wesentlich detaillierter und für jeden CPU-Kern einzeln (**Abbildung 3**). Unter »Device stats« findet man alle nervenden Geräte, wie etwa Netzwerkkarten und USB-Adapter. Bei den Netzwerkschnittstellen verrät PowerTOP, wie viele Datenpakete sie pro Sekunde durch das Netz geschleust haben. Auf dem Register »Tunables« listet PowerTOP schließlich alle möglichen Stromspartipps auf. Ein »Bad« zeigt Hand-

lungsbedarf an, während »Good« bereits umgesetzte Maßnahmen kennzeichnet. Mit den Pfeiltasten steuert man einen Spartipp an und schaltet ihn mit der Eingabetaste frei (**Abbildung 4**).

Fazit

Je nach Linux-System lassen sich mit der Hilfe von PowerTOP im Schnitt bis zu 10 Watt einsparen. Notebooks danken dies mit längeren Batterielaufzeiten, in Server-Farmen reduzieren sich die Stromkosten. Es lohnt sich folglich, sein System einmal kurz mit dem kleinen Werkzeug auf CPU weckende Programme hin zu untersuchen und unnötige Stromfresser zu stoppen. Allerdings ist auch PowerTOP kein Allheilmittel. Wer konsequent Strom sparen möchte, sollte alle nicht genutzten Geräte und Schnittstellen abschalten und konsequent auf energieeffiziente Hardware setzen. Nützliche, von PowerTOP unabhängige Stromsparhinweise finden sich beispielsweise in der Ausgabe 03/2010 unseres Schwestermagazins Linux-User [3]. (ofr)

Infos

- [1] PowerTOP-Homepage: [<http://www.lesswatts.org/projects/powertop/>]
- [2] Quellcode von PowerTOP 2.0 Beta: [<http://www.kernel.org/pub/linux/status/powertop/powertop-1.97.tar.bz2>]
- [3] Marcel Hilzinger, Stehvermögen, Stromverbrauch bei Notebooks vermindern, Linux-User 3/2010

Auf Kommando

PowerTOP kennt in Version 1.13 noch ein paar interessante Kommandozeilenparameter. So blendet »powertop -p« die Prozess-IDs in der Liste der häufigsten Aufwecker ein. Die Nummern stehen dann in rechteckigen Klammern vor den Prozessnamen. Mit dem Parameter »-t« erhöht man das Zeitintervall, in dem PowerTOP die Daten sammelt. Die Version 1.13 enthält hier allerdings offenbar einen Fehler und ignoriert diesen Parameter konsequent. Abschließend läuft PowerTOP mit »-d« genau einmal 15 Sekunden, druckt seine Ergebnisse in die Standardausgabe und schiebt als Bonus noch alle dabei gesammelten Daten hinterher (**Abbildung 5,6**). Einen solchen Bericht generiert auch das neue PowerTOP 2.0 über den Befehl »sudo powertop --html«, speichert die Informationen aber in der HTML-Datei »powertop.html« (**Abbildung**

7). Sofern PowerTOP 2.0 auf einem Notebook im Batteriebetrieb läuft, sagt es den Energieverbrauch verschiedener Geräte beziehungsweise Aktivitäten voraus. Diese Angaben sind genauer, wenn man das Werkzeug über den Befehl »sudo powertop --calibrate« startet. PowerTOP 2.0 führt dann mehrere Kalibrierungsläufe durch, in denen es unter anderem verschiedene Bildschirmhelligkeiten durchprobiert. PowerTOP kennt in Version 1.13 noch ein paar interessante Kommandozeilenparameter. So blendet »powertop -p« die Prozess-IDs in der Liste der häufigsten Aufwecker ein. Die Nummern stehen dann in rechteckigen Klammern vor den Prozessnamen. Mit dem Parameter »-t« erhöht man das Zeitintervall, in dem PowerTOP die Daten sammelt. Die Version 1.13 enthält hier allerdings offenbar einen Fehler und ignoriert

diesen Parameter konsequent. Abschließend läuft PowerTOP mit »-d« genau einmal 15 Sekunden, druckt seine Ergebnisse in die Standardausgabe und schiebt als Bonus noch alle dabei gesammelten Daten hinterher (**Abbildung 5**). Einen solchen Bericht generiert auch das neue PowerTOP 2.0 über den Befehl »sudo powertop --html«, speichert die Informationen aber in der HTML-Datei »powertop.html« (**Abbildung 7**). Sofern PowerTOP 2.0 auf einem Notebook im Batteriebetrieb läuft, sagt es den Energieverbrauch verschiedener Geräte beziehungsweise Aktivitäten voraus. Diese Angaben sind genauer, wenn man das Werkzeug über den Befehl »sudo powertop --calibrate« startet. PowerTOP 2.0 führt dann mehrere Kalibrierungsläufe durch, in denen es unter anderem verschiedene Bildschirmhelligkeiten durchprobiert.

Verschlüsselte Festplatten per eCryptfs

Sichtschutz

Viele Linux-Distributionen bieten mittlerweile verschlüsselte Partitionen an, die aber den Zugriff auf einen Anwender beschränken. Feiner regulierte Verschlüsselung bietet eCryptfs. *Juliet Kemp*

eCryptfs [1] bietet Verschlüsselung auf Dateiebene nach PGP-Art. Speichert der Anwender eine Datei in einem per eCryptfs verschlüsselten Verzeichnis, verschlüsselt eCryptfs sie transparent. Wird eine solche Datei gelesen, geschieht auch die Entschlüsselung automatisch. Das alles passiert hinter den Kulissen, wobei mit den Default-Einstellungen der Login-Schlüssel verwendet wird, was den ganzen Ablauf ohne weiteren Eingriff automatisiert.

Datei statt Block

Einer der Vorteile von dateibasierter gegenüber blockbasierter Verschlüsselung liegt darin, dass mehrere Schlüssel auf einem System möglich sind. Das bedeutet, dass Benutzer auf dem gleichen System ihre Dateien oder Verzeichnisse unabhängig voneinander verschlüsseln können. Verschiedene Benutzer können selbst im gleichen Verzeichnis ihre Dateien mit unterschiedlichen Schlüsseln ablegen. Mit Ubuntu hat der Anwender die Möglichkeit, bei der Installation sein Home-Verzeichnis zu verschlüsseln. Wer einen neuen Benutzer anlegt, kann das ebenso mit einem verschlüsselten Home-Ver-

zeichnis tun, wenn die »ecryptfs-utils« installiert sind:

```
sudo adduser --encrypt-home Benutzer
```

Leider funktioniert dieser Befehl nur unter Ubuntu. Debian unterstützt zwar eCryptfs, aber sein Adduser-Tool kennt nicht die Option »--encrypt-home«. Alternativ können Sie natürlich auch eCryptfs von Hand installieren und einrichten:

```
ecryptfs-setup-private
```

Dieser Aufruf legt ein Verzeichnis »~/Private« an und richtet es als verschlüsseltes Verzeichnis ein (**Abbildung 1**). Wenn Sie sich ausloggen, wird es unter dem Namen »~/Private« gespeichert, beim Neueinloggen wird es automatisch als »~/Private« gemountet. Das Login-Passwort dient dazu, die Passphrase zum Verschlüsseln zu speichern.

Am sichersten ist es, eCryptfs selbst eine zufällige Passphrase erzeugen zu lassen, denn ein zufälliges Passwort ist meist resistenter gegen Brute-Force-Attacken als ein selbst ausgedachtes. Sie sollten die Passphrase an einem sicheren Ort speichern, falls ein Problem auftritt und Sie die Dateien wieder von Hand entschlüsseln müssen. Der folgende Befehl gibt die

gespeicherte Passphrase aus:

```
ecryptfs-unwrap-passphrase .ecryptfs/wrapped-  
-passphrase
```

Wenn Sie sich entschieden haben, welche Verzeichnisse Sie verschlüsseln möchten, verschieben Sie sie nach »~/Private« und legen einen symbolischen Link zu Ihrem alten Speicherort an. Um zum Beispiel das Verzeichnis Ihrer Bankdaten zu verschlüsseln, verwenden Sie diese Befehle:

```
cd  
mv banking Private/  
ln -s Private/banking banking
```

Wenn Sie sich ausloggen oder »Private« unmounten, werden sowohl der Inhalt dieses Verzeichnisses als auch die Dateinamen verschlüsselt. Andere Benutzer können also bei einem Blick in das Verzeichnis nicht feststellen, was dort gespeichert ist. Ein anderes Verzeichnis dorthin zu verschieben, führt dazu, dass es automatisch verschlüsselt wird.

Wenn Sie nicht wollen, dass »Private« beim Einloggen automatisch gemountet wird, löschen Sie »~/ecryptfs/auto-mount« (und »~/ecryptfs/auto-umount«). Bei beiden handelt es sich um leere Dateien,

Sie können sie also einfach mit »touch« wieder anlegen, um das Automounten wieder zu aktivieren. Wenn Sie wichtige Dateien nach »Private« verschoben und Automount aktiviert haben, werden Sie vielleicht mit Fehlermeldungen konfrontiert, zum Beispiel bei wichtigen Konfigurationsdateien für den Desktop.

Handverschlüsselt

Wenn Sie ein anderes Verzeichnis als »Private« verwenden möchten oder ein zweites verschlüsseltes Verzeichnis (zum Beispiel mit einem anderen Schlüssel), aktivieren Sie eCryptfs folgendermaßen:

```
mkdir ~/secret
chmod 700 ~/secret
sudo mount -t ecryptfs secret secret
```

Das Programm fragt Sie dann nach dem gewünschten Verschlüsselungsverfahren und der Schlüssellänge – verwenden Sie einfach die Defaults, wenn Sie sich nicht sicher sind. Außerdem werden Sie gefragt, ob Sie Plaintext-Passthrough verwenden möchten. Dieses Feature erlaubt es, innerhalb eines eCryptfs-Verzeichnisses unverschlüsselte Dateien abzulegen. Das mag zwar in manchen Situationen nützlich sein, aber man verliert leicht den Überblick darüber, was nun verschlüsselt ist und was nicht. Meistens ist die Default-Einstellung, es nicht zu erlauben, am besten.

Filename-Encryption bedeutet, dass nicht nur Dateiinhalte, sondern auch Dateinamen verschlüsselt werden. Dann können nur Sie die Dateinamen im verschlüsselten Verzeichnis sehen, wenn es mit eCryptfs gemountet ist. Schließlich werden Sie noch nach der Passphrase gefragt. Weil es das erste Mal ist, dass Sie diesen Mountpoint verwenden, erhalten Sie eine Warnmeldung darüber, dass Sie die Passphrase zum ersten Mal verwenden. Außerdem möchte das Tool wissen, ob Sie die Passphrase im Keyring speichern wollen, um künftige Warnmeldungen zu vermeiden.

In den Mount-Parametern steht die erste Pfadangabe für das Verzeichnis, das Sie per eCryptfs mounten wollen, die zweite für den Mountpoint. Diese beiden Pfade können sich unterscheiden, wenn sie aber gleich sind, bedeutet das, dass Sie auf Dateien im verschlüsselten Verzeich-

nis immer über eCryptfs zugreifen müssen. Dadurch wird der Zugriff auf diese transparent.

Wenn Sie die Arbeit mit dem Verzeichnis beenden und unmounten, sehen Sie statt der Dateinamen nur noch die verschlüsselten Versionen (jedenfalls, wenn Sie die Verschlüsselung von Dateinamen aktiviert haben). Wenn Sie einen Blick in die Dateien werfen, sehen Sie nur zufällig erscheinende Binärdaten.

Aufwand

Leider bringt die Verschlüsselung per eCryptfs auch einige Umstände mit sich. So müssen Sie jedes Mal die Optionen angeben, wenn Sie das Verzeichnis mounten wollen:

```
sudo mount -t ecryptfs secret/ secret/ -o 2
ecryptfs_cipher=aes,ecryptfs_key_bytes=16,2
ecryptfs_passthrough=n
```

Dummerweise gibt es für die Verschlüsselung von Dateinamen keine Kommandozeilenoption. Sie können ein Verzeichnis aber ohne Probleme einmal mit und einmal ohne diese Option mounten.

Der andere Nachteil ist, dass das obige Mount-Kommando Root-Rechte erfordert. Um das zu ändern, binden Sie als Administrator das Verzeichnis ein und sehen sich dann die Datei »/etc/mtab« an. Tragen Sie die entsprechende Zeile in »/etc/fstab« ein und fügen Sie noch die Option »user« hinzu. Das Ergebnis sieht dann etwa aus wie [Listing 1](#).

Fügen Sie Ihren Schlüssel mit »ecryptfs-manger« dem User-Session-Keyring hinzu. Wählen Sie Option 1 (»add passphrase to keyring«) und tippen Sie dann Ihre Passphrase ein. Mit Option 4 verlassen Sie das Programm. Wenn Sie nun »mount -i newsecret« eingeben, sollten Sie das verschlüsselte Verzeichnis verwenden können (»-i« verhindert den Aufruf des Mount-Hilfsprogramms). Unmounten Sie es wieder und löschen Sie mit »keyctl clear @u« den Session-Keyring.

Zwei weitere Schritte automatisieren diesen Vorgang für das Login. Fügen Sie als Erstes die Zeile »mount -i secret« der »~/.bashrc« hinzu. Tragen Sie dann noch als Administrator die folgende Zeile in »/etc/pam.d/login« ein:

```
auth required pam_ecryptfs.so
```

vollmar net

stark · schnell · innovativ



🔌 **Xen-Server**

🔌 **IP-Tunnel**

ab 4,95 mtl.

🔌 **Dedicated
Server Special**

- 🔌 AMD64 x2 ab 3,0 Ghz, 4096 MB RAM
- 🔌 2 x 500GB HD SATA II, 2000GB Traffic incl.

39 €* mtl.

vollmar.net e.K.
Deutschherrnstr. 15-19
90429 Nürnberg

<http://vollmar.net>
info@vollmar.net
Tel: 08000 - 87 87 76 76 (kostenfrei)

full support

IPv6
running

```

oliver@debian: ~
Datei Bearbeiten Ansicht Terminal Hilfe
oliver@debian:~$ /usr/bin/ecryptfs-setup-private
Enter your login passphrase:
Enter your mount passphrase [leave blank to generate one]:
Enter your mount passphrase (again):

*****
YOU SHOULD RECORD YOUR MOUNT PASSPHRASE AND STORE IT IN A SAFE LOCATION.
ecryptfs-unwrap-passphrase ~/.ecryptfs/wrapped-passphrase
THIS WILL BE REQUIRED IF YOU NEED TO RECOVER YOUR DATA AT A LATER TIME.
*****

Done configuring.

Testing mount/write/umount/read...
Testing succeeded.

Logout, and log back in to begin using your encrypted directory.

oliver@debian:~$ ls Private/
Access-Your-Private-Data.desktop  README.txt
oliver@debian:~$

```

Abbildung 1: Ein privates Verzeichnis für verschlüsselte Dateien.

Loggen Sie sich aus und wieder ein, und »secret/« sollte automatisch gemountet werden. Das Gleiche können Sie mit weiteren Verzeichnisse machen, die Sie verschlüsseln möchten. Sie können auch das komplette Home-Verzeichnis mounten statt nur einzelner Verzeichnisse, aber das ist relativ kompliziert. Wie das geht, verrät ein Eintrag im Entwickler-Blog [2].

Die Passphrase zum Mounten von eCryptfs-Verzeichnissen ist in »~/.ecryptfs/wrapped-passphrase« gespeichert und symmetrisch mit dem Login-Passwort verschlüsselt. Per Default ist die Mount-Passphrase ein zufällig generierter 128-Bit-Schlüssel, der schwer zu merken, aber auch schwer zu knacken ist. Wenn Sie ein eCryptfs-Verzeichnis mounten, wird die Passphrase im Kernel-Keyring gespeichert.

Wenn Sie auch Dateinamen verschlüsseln, kommt ein zweiter Schlüssel zum Einsatz, der ebenfalls in den Kernel-Keyring geladen wird. Diese beiden Schlüssel werden dann immer verwendet, wenn Sie auf eine Datei im verschlüsselten Verzeichnis zugreifen.

Um die Sicherheit zu erhöhen, können Sie die Passphrase auf einem portablen Speichermedium wie einem USB-Stick speichern, das Sie an einem sicheren Ort verwahren, wenn Sie nicht eingeloggt sind. Legen Sie dann einen symbolischen Link von »~/.ecryptfs/wrapped-passphrase« zu Schlüsseldatei auf dem Stick an. Da-

mit das auf Dauer zuverlässig funktioniert, müssen Sie den Stick noch in »/etc/fstab« eintragen, damit er immer im selben Verzeichnis eingebunden wird:

```
/dev/sdb1 /media/usbkey2
ext3 defaults 0 0
```

Besonders sauber erledigen Sie das Gleiche bei modernen Linux-Systemen

über die Udev-Konfiguration, in der Sie beispielsweise abhängig von den USB-Vendor- und -Produkt-IDs den Stick ins passende Verzeichnis mounten. Wenn Sie den Schlüssel auf einem USB-Stick speichern, sollten Sie noch dringender an ein Backup denken, denn die kleinen Geräte gehen schnell einmal verloren.

Sie können auch die Option »-w« von »ecryptfs-setup-private« verwenden, die statt des Login-Passworts noch eine zweite Passphrase verwendet. Damit können Sie das Security-Level noch einmal steigern, aber das verschlüsselte Verzeichnis kann nicht mehr automatisch beim Einloggen gemountet werden.

Systemverzeichnisse

Denken Sie bei Ihren Sicherheitserwägungen auch daran, dass sensible Daten auch an anderen Stellen landen können, zum Beispiel im tmp-Verzeichnis. Deshalb wollen Sie vielleicht auch andere Verzeichnisse mit eCryptfs schützen. Das können Sie genauso machen wie weiter oben beschrieben, denken Sie nur daran, das entsprechende Verzeichnis in »/etc/fstab« einzutragen, damit Nicht-Root-Benutzer es auch mounten dürfen.

Wenn das fragliche Verzeichnis schon Daten enthält, verkompliziert sich die Prozedur etwas. Machen Sie als Erstes ein Backup der Daten und gehen Sie dann die folgenden Schritte durch:

- Legen Sie ein neues Verzeichnis an und richten Sie es wie beschrieben als eCryptfs-Verzeichnis ein.
- Mounten Sie es und kopieren Sie alle Dateien aus dem alten Verzeichnis hinein. Denken Sie dabei auch an die mit einem Punkt beginnenden Dot-Files.

■ Benennen Sie das alte Verzeichnis um und unmounten Sie das eCryptfs-Verzeichnis.

■ Geben Sie dem eCryptfs-Verzeichnis den vorigen Namen des alten Verzeichnisses, tragen Sie es in »/etc/fstab« ein und mounten Sie es wieder. Ihre Daten sind nun verschlüsselt.

Im Fall des tmp-Verzeichnisses ist es am besten, alle temporären Daten zu löschen und es als eCryptfs-Verzeichnis neu anzulegen.

Selbst die Swap-Partition lässt sich verschlüsseln, allerdings funktionieren dann die Schlaf-Modi (Hibernation/Sleep) nicht mehr. Damit ist jedenfalls sichergestellt, dass keine entschlüsselten Dateien im Swap-Bereich liegen. Um die Swap-Partition zu verschlüsseln, installieren Sie »cryptsetup« und führen dann als Administrator »ecryptfs-setup-swap« aus. Das Tool stoppt dann den Swap-Vorgang, verschlüsselt die Partition und startet das Swappen wieder. Zwar wird automatisch ein neuer Eintrag in »/etc/fstab« hinzugefügt, aber der alte bleibt dummerweise erhalten. Sie müssen also die Datei von Hand editieren und den alten Eintrag entfernen. Die verschlüsselte Swap-Partition heißt dann »/dev/mapper/cryptswap«.

Mehr Infos

Für aktuelle Informationen zu eCryptfs und einer Reihe hilfreicher Tutorials werfen Sie einen Blick in das Blog des Entwicklers Dustin Kirkland [3]. (ofr) ■

Infos

- [1] eCryptfs: <https://launchpad.net/ecryptfs>
- [2] Migrating to an Encrypted Home Directory: <http://blog.dustinkirkland.com/2009/06/migrating-to-encrypted-home-directory.html>
- [3] Dustin Kirklands Blog: <http://blog.dustinkirkland.com>

Die Autorin

Juliet Kemp beschäftigt sich mit Linux, seit sie gemerkt hat, dass es mehr Spaß macht als Abschlussprüfungen. Mittlerweile arbeitet sie schon über zehn Jahre als Systemadministratorin. Sie ist die Autorin des Buchs „Linux Systems Administration Recipes: A Problem-Solution Approach.“

Listing 1: »/etc/fstab«

```
01 /home/juliet/secret /home/juliet/secret ecryptfs
user,rw,ecryptfs_sig
=9ffdc5087c0c049,ecryptfs_fnek_
sig=9ffdc5087c0c049,ecryptfs_unlink
_sigs,ecryptfs_cipher=aes,ecryptfs_key_bytes=16 0 0
```



Projekt Honey Pot kämpft gegen Spam

Angelockt

Brux@bk.ru Brubov, 123RF

Einzelkämpfer haben im Kampf gegen Spam wenig Chancen. Das Projekt Honey Pot versucht Kräfte zu bündeln, und über verteilte Honeypots möglichst viele Spam-Versender zu identifizieren. James Mohr

Schon lange hat Spam aufgehört, ein kleines Ärgernis zu sein. Irgendwann einmal waren die zwei bis drei Spam-Mails am Tag kaum der Rede wert, aber jetzt treffen an einem guten Tag „nur“ 200 ungebetene Nachrichten in meiner Inbox ein. Auf diesem Niveau stellt Spam eine richtig hartnäckige Störung dar.

Jede Spam-Mail, die bei Ihrem Server ankommt, bedeutet einen Aufwand an Netzwerk-Kapazität, CPU-Zeit und Speicherplatz, was wiederum auf bares Geld hinausläuft. Und diese Kosten fallen sogar dann an, wenn eine Spam-Mail nie das Postfach eines Benutzers erreicht. Die beste Lösung ist also, Spam bereits zu stoppen, bevor sie den eigenen Server erreicht.

Stand der Technik

Viele Open-Source-Tools wie Spamassassin versuchen wie kommerzielle Software, Spam nach dem Eintreffen zu erkennen und auszusortieren, statt sie den Anwendern zuzustellen. Die unerwünschten Mails sind dann aber schon auf dem Server und verursachen zum Teil einigen Aufwand an Rechenzeit.

Unspam Technologies [1] hat deshalb das Projekt Honey Pot [2] ins Leben gerufen, das erste verteilte System, um

Spammer im Internet zu identifizieren. Die Firma selbst bietet Antispam-Services für Enduser genauso wie für Firmen und Regierungsstellen. Das Honey-Pot-Projekt besteht im Wesentlichen aus einzelnen sogenannten Honeypot-Maschinen, die im Internet verteilt sind, und Spammer anlocken sollen.

Historisch waren Honeypots meist einzelne Server, die speziell dafür eingerichtet wurden, potenzielle Einbrecher anzulocken. Sie waren oft speziell mit Sicherheitslücken ausgestattet, um das Interesse böswilliger Hacker zu wecken, diese zu einem Einbruchversuch zu motivieren und von den eigentlich sensiblen Servern abzulenken.

Diese Honeypots sollten nicht nur zur Beschäftigungstherapie der Cracker dienen. Stattdessen konnten Administratoren dabei versuchen, die Herkunft der Einbrecher nachzuvollziehen, ihre Methoden studieren, Gegenmaßnahmen ergreifen und die anderen Server dadurch besser absichern.

Das gleiche Prinzip lässt sich auch auf Spammer anwenden. Aus der Beobachtung ihrer Methoden kann man Techniken zu ihrer Abwehr entwickeln. Im einfachsten Fall geht das über eine Analyse von Spam-Mail, um ein Muster zu finden, welche Mails Spam enthalten und wel-

che nicht. Solche Muster lassen sich an Tools wie Spamassassin füttern, um die Mail-Nachrichten nach dem Eintreffen zu filtern. Aber das ist, wie bereits angesprochen, nicht der beste Weg.

Erntezeit

Das schmutzige Handwerk der Spammer vollzieht sich in mehreren Schritten. Am Anfang steht das sogenannte Harvesting (Ernten), bei dem sie auf Websites nach E-Mail-Adressen suchen. Schon in den Urzeiten des World Wide Web haben Anwender arglos ihre E-Mail-Adressen auf Webseiten geschrieben. Von dort können Spammer sie mit einfachsten Mitteln sammeln und in ihren Datenbanken speichern.

Wenn ein Harvester eine Webseite aufruft, hinterlässt er dabei die typischen Spuren in den Logdateien des Servers, zum Beispiel die IP-Adresse. Genauso funktioniert es beim Projekt Honey Pot: Ruft jemand die entsprechende Webseite mit einer E-Mail-Adresse ab, speichert der Honeypot die IP-Adresse des Clients. Erhält die E-Mail-Adresse später Spam, schließt sich der Kreis und die IP des Spammers ist identifiziert.

Der direkte Nutzen aus dieser Erkenntnis besteht darin, die Harvester auszu-

sperren. Wer eine IP-Adresse als Spam-Schleuder identifiziert hat, kann diese in der Firewall sperren und damit vom weiteren Ernten abhalten. Das bedeutet weniger Spammer mit gültigen E-Mail-Adressen, löst aber nur einen Teil des Problems. Schließlich will man auch Spammer stoppen, die die eigene Adresse schon gespeichert haben. Deshalb gilt es auch die Rechner zu finden, die Spam verschicken.

Meistens gehören diese Spam-Server gar nicht den Spammern selbst, denn es ist nicht sehr schwer, die Absender-IP zu ermitteln. Weil die meisten mit Spam verbundenen Tätigkeiten illegal sind, wären die Spammer leicht zu identifizieren und juristisch zur Verantwortung zu ziehen. Deshalb verschleiern sie ihre Spuren zum Beispiel, indem sie die Werbe-Mails von den Rechnern argloser Anwender verschleiern. Solche sogenannten Zombie-Rechner wurden meist mithilfe einer Malware gekapert, die im Hintergrund schlummert, wenn der legale Besitzer im Internet surft. Der Spammer kann seine eingeschleuste Software dann zum Leben erwecken und für seine niederen Dienste einspannen.

Wurde ein Rechner als Spamschleuder identifiziert, trägt Projekt Honey Pot ihn in eine Liste ein. Administratoren müssen darüber hinaus seine Adresse in eine lokale Blacklist des Mailservers eintragen.

Kein Kommentar

Immer mehr greift die Unsitte des Kommentar-Spams um sich. In meinem eigenen Blog landen zum Beispiel jede Woche mehrere etwas kryptische Kommentare. Meistens handelt es sich dabei um eine mehr oder weniger sinnvolle Sammlung von Wörtern, zusammen mit einem Link auf eine Website.

Projekt Honey Pot analysiert den Inhalt solcher Spam-Kommentare und erzeugt eine Liste der meistverwendeten Spam-Wörter. Viele Content-Management- und Blog-Systeme akzeptieren eine Liste von Wörtern, aufgrund derer sie einen Kommentar als Spam klassifizieren. Die Listen von Honeypot bieten eine gute Grundlage dafür, auf diesem Weg die Erkennungsquote dramatisch zu steigern. Das Projekt publiziert außerdem eine Liste von Domain-Namen und Adressen, auf die

Spammer in Kommentaren besonders häufig verweisen.

Ein beliebtes Einfallstor für Spammer entsteht, wenn Administratoren ihren Mailserver nicht richtig konfigurieren. Dann wird aus ihm schnell ein sogenanntes offenes Relay, was bedeutet, dass jeder über den Server Mail verschicken darf, egal welche Absender- und Empfängeradresse er verwendet. Ein besondere Gefahr entsteht hieraus, weil die Spam-Nachrichten in diesem Fall von einem ansonsten legitimen Absender stammen. Schnell landet dieser dann auf einer Blacklist und sperrt damit seine Anwender von der E-Mail-Nutzung aus – auch wenn diese Blacklist-Praxis nicht unumstritten ist [3].

Cleverer Spammer

Die Spammer sind aber auch nicht dumm und wollen natürlich wissen, wenn der von ihnen verwendete Server blockiert wird. Deshalb testen sie offene Relays, indem sie Mails an sich selbst verschicken. Nur wenn eine Mail auch ankommt, handelt es sich um einen für ihre Zwecke nutzbaren Server. Auch daran haben die Macher des Projekts Honey Pot gedacht. Sie versuchen zu identifizieren, welche Adresse die Spammer für solche Tests verwenden. Sie lässt der Honey Pot dann im Folgenden durch, verwirft aber alle anderen Nachrichten.

Jeder Administrator kann sich beim Honey-Pot-Projekt einbringen, indem er einen eigenen Honeypot installiert. Dazu muss er eine Vereinbarung unterzeichnen, die unter anderem besagt, dass er eine Seite mit Nutzungsvereinbarungen in seine Website aufnimmt, die das Harvesting von E-Mail-Adressen untersagen. Diese neue Seite enthält auch die E-Mail-Adresse des eigentlichen Honeypot. Diese Adresse gibt es im Projekt nur einmal, sodass sie eindeutig zugeordnet werden kann. Weil sie nur über die neue Webseite zu finden ist, kann niemand sich darauf berufen, die Nutzungsvereinbarungen nicht zu kennen. Honeypot speichert wie beschrieben die IP-Adresse des Harvesters. Ein solcher Honeypot lässt sich zu jeder Website hinzufügen, die die Ausführung eigener Skripts erlaubt, zum Beispiel mit PHP, Mod-Perl, ASP, Python und so weiter.

Je nachdem, ob Sie es auf Harvester oder Spammer abgesehen haben, kann die Webseite auch Links oder Formulare enthalten. Webformulare besitzen meist einen ähnlichen Aufbau, weshalb sie von Spammern einfach auszufüllen sind. Der Prozess lässt sich leicht automatisieren, dennoch beschäftigen viele Spammer dafür auch Menschen mit niedrigem Stundenlohn. Honeypot wertet die Eingaben solcher Formulare aus, um verdächtige Muster zur Spam-Erkennung zusammenzustellen.

Lockstoff

Üblicherweise ist die auf der Webseite enthaltene E-Mail-Adresse vor menschlichen Besuchern versteckt. Auf diese Weise soll sichergestellt werden, dass nur automatisierte Harvester sie sammeln. Der Text für die Nutzungsbedingungen unterscheidet sich von Installation zu Installation, was die automatisierte Erkennung des Honeypots durch die Spammer erschwert. Jedes Mal, wenn ein Harvester auf den Honeypot trifft, wird eine neue Adresse erzeugt. So kann das Projekt Honey Pot den Harvester nicht nur über die E-Mail-Adresse identifizieren, sondern auch über den genauen Zeitpunkt des Abrufs.

Projekt Honey Pot lädt Unternehmen und andere Organisationen zum Mitmachen ein. Wegen der juristischen Bedeutung der Nutzungsvereinbarung schrecken manche Firmen davor zurück, einen Honeypot auf ihrer Website einzurichten. Außerdem legen die meisten Wert auf ein einheitliches Erscheinungsbild. Das stellt allerdings ein kleineres Problem dar, denn die Firma hinter Projekt Honey Pot arbeitet mit Firmen zusammen, um die Honeypots soweit möglich anzupassen. Weitverbreitet ist die Angst, dass ein Honeypot die Menge an Spam in die Höhe treibt. Aber weil die auf der Website veröffentlichte E-Mail-Adresse unter der Kontrolle des Honeypot-Projekts steht, beschränkt sich aller Mail-Spam darauf. Einige Admins befürchten auch, dass der eigene Server auf den Blacklisten des Projekts landet. Das passiert typischerweise mit dynamischen IP-Adressen, wenn beispielsweise eine Zombie-Maschine Spam verschickt und später ein anderer Rechner deren IP-Adresse zugewiesen bekommt. Für solche Fälle bietet das

Directory of Malicious IPs

This page displays the top IPs by different categories. You may sort or limit this list by selecting from the menus below.

The list below is comprised of Malicious IPs (limited to the top 50) that are:

- Arranged by their Last Bad Event

Global Statistics
Last Bad Event
Any IP
From All Countries

See [comment spammers](#), [dictionary attackers](#), or [mail servers](#) from the same region.

You may also [lookup information](#) on a specific IP address.

To track IPs targeting a website not currently listed in the pulldown, simply [install a honey pot](#) on it.

An [RSS feed](#) for this page is available.

WHO IS...
Updating Facebook During Work?

Malicious IP	Event	Total	First	Last
109.230.217.105 C	Bad Event	26	2010-12-03	2010-12-03
213.108.217.82 C	Bad Event	51	2010-11-07	2010-12-03
93.116.183.14 C	Bad Event	1,717	2010-10-28	2010-12-03
213.5.66.16 C	Bad Event	3,423	2008-08-20	2010-12-03
193.105.210.13 C	Bad Event	277	2010-10-21	2010-12-03
41.190.16.17 C	Bad Event	1,881	2010-05-16	2010-12-03
109.230.221.166 C	Bad Event	18	2010-12-03	2010-12-03
91.201.66.121 C	Bad Event	201	2010-08-12	2010-12-03
95.211.8.146 C	Bad Event	771	2010-10-18	2010-12-03
210.187.51.43 SC	Bad Event	143	2009-03-05	2010-12-03
188.163.72.235 C	Bad Event	4	2010-12-03	2010-12-03
213.5.71.163 CR	Bad Event	2,212	2010-09-30	2010-12-03
93.174.93.58 C	Bad Event	175,297	2009-03-13	2010-12-03
91.201.66.86 C	Bad Event	3,245	2010-08-03	2010-12-03
91.201.66.103 C	Bad Event	63	2010-08-17	2010-12-03
190.90.5.83 C	Bad Event	86	2010-10-02	2010-12-03
208.91.190.13 C	Bad Event	9	2010-11-27	2010-12-03

Abbildung 1: Im Webinterface des Projekt Honey Pot sind beispielsweise die IP-Adressen von Spam-Schleudern aufgelistet.

Projekt Honey Pot eine Möglichkeit, sich selbst von einer Blacklist zu löschen. Gerät man jedoch wieder darauf, wird das Entfernen schwieriger. Schließlich sollen sich notorische Spammer nicht jedes Mal wieder selbst von einer Blacklist löschen können.

Um die Erfolgsquote von Honey Pot zu steigern, sind viele kleine Mailserver gefragt mitzumachen. Die Spammer kennen das Projekt mittlerweile und versuchen, Honeypot-Rechner zu identifizieren. Mehr Server bedeuten für die Spammer mehr Arbeit dafür, ein Muster zu erkennen und Honeypot-Rechner zu finden. Man kann das Projekt auch unterstützen, indem man einen MX-Eintrag zur Verfügung stellt. Dazu tragen Sie einen MX-Eintrag in der DNS-Konfiguration einer eigenen Subdomain (zum Beispiel »mail2.domain.com«) ein, der auf einen Mailserver von Honey Pot zeigt. Das Projekt empfängt dann alle E-Mails, die an eine Adresse in der entsprechenden Domain gehen (»user@mail2.domain.com«). Je mehr Hosts hierbei mitmachen, umso effizienter funktioniert auch die Spam-Erkennung.

Man soll allerdings keine Subdomains zur Verfügung stellen, die schon einmal Spam empfangen haben. Sonst wird es zu schwierig, zwischen Spam-Mails zu unterscheiden, die vom Honeypot stammen und solchen, die aus früheren Sammelaktionen stammen. Eine neue Subdomain anzulegen, ist normalerweise immer möglich. Außerdem sollte

über diese Domain auch keine normale Mail empfangen worden sein. Honey Pot nimmt nämlich an, dass jede empfangene Mail Spam ist. Schickt jemand eine echte Mail an die Adresse, verkompliziert das die Analyse.

Wenn Sie sich dafür entscheiden, dem Projekt einen MX-Record zu spendieren, können Sie ihn als private oder public markieren. Ist er public, kann jeder Honeypot E-Mails an den vom MX-Record referenzierten Host senden. Anderenfalls schicken nur Honeypots vom gleichen Account – also Ihre eigenen – Mails an diesen Server. Sie können Honey Pot auch anweisen, E-Mail-Adressen nur innerhalb der Subdomain zu erzeugen, die Sie gespendet haben. Mehr Details zu diesen Zusammenhängen sind auf der Honey-Pot-Website zu finden.

Andere Dienste

Wie schon angedeutet betreibt das Projekt Honey Pot auch einen HTTP-Blacklist-Dienst (http:BL), über den die gesammelten Informationen für User verfügbar sind. Mit diesem Dienst können Sie eine DNS-Abfrage für einen bestimmten Rechner durchführen und http:BL sagt Ihnen dann, ob Sie diese Maschine besser blockieren oder Ihre Daten durchlassen. Die Details hierzu finden sich ebenfalls auf der Projekt-Website.

Das Projekt bietet außerdem verschiedene Statistiken und Listen, wie zum Beispiel die aktivsten Harvester und Spammer, die

am häufigsten verwendeten Spam-Wörter und so weiter. Auch Statistiken über verschiedene Spam-Fallen sind verfügbar, genauso wie Berichte von überführten Spammern. Zwei Beispiele dazu: Die längste Zeitspanne zwischen dem Harvesten einer Adresse und dem Empfang der dazugehörigen Spam-Mail war über drei Jahre, die kürzeste dagegen nur eine Sekunde. Die Statistiken sind übersichtlich aufbereitet und in verschiedenen Ansichten konfigurierbar (Abbildung 1). Die im Lauf der Zeit gesammelten Daten stellt das Projekt auch Entwicklern von Antispam-Software oder -Forschern zur Verfügung.

Fazit

Seit der Aufnahme des Betriebs im Jahr 2004 hat Projekt Honey Pot mehr als 80 Millionen IP-Adressen mit 50 Millionen Spam-Fallen und 1,2 Milliarden Spam-Mails überwacht. Mit ihrer derzeitigen Kapazität könnte es diese Zahl bis auf etwa 500 Milliarden Fallen steigern. Mit bis zu 20 Millionen Nachrichten pro Woche, die es durch Honeypots empfängt, entstehen derzeit etwa ein Terabyte an Daten.

Auch wenn das Projekt Honey Pot das Spam-Problem nicht völlig aus der Welt schaffen wird, kann es den Spammern dennoch recht große Brocken in den Weg legen. Sich in das Projekt einzubringen, dauert nur ein paar Minuten, kann aber eine große Wirkung haben. Spam funktioniert, weil die Spammer Daten in so großem Stil verschicken. Honey Pot versucht das gleiche Prinzip bei der Abwehr zu verwenden. (ofr)

Infos

- [1] Unspam Technologies Inc.: [\[http://www.unspam.com\]](http://www.unspam.com)
- [2] Project Honey Pot: [\[http://www.projecthoneypot.org\]](http://www.projecthoneypot.org)
- [3] Wolf-Dieter Mergenthaler, Dunkle Mächte, ADMIN 06/2010, S. 89

Der Autor

James Mohr ist verantwortlich für das Datacenter-Monitoring bei einem Business Solutions Provider in Coburg. Er betreibt die Linux-Tutorial-Website [\[http://www.linux-tutorial.info\]](http://www.linux-tutorial.info). James ist Autor mehrerer Bücher und vieler Artikel zu IT-Themen.



Eine Tour durch die Debugging- und Rettungs-Tools in Knoppix

Zu Hilfe!

Geht bei einem System-Crash das Dateisystem kaputt, ist guter Rat teuer. Zumindest, wenn man Recovery-Experten zurate zieht. Der erfahrene Admin weiß sich selbst zu helfen, zum Beispiel mit den Tools auf einer Knoppix-DVD. Klaus Knopper

Viele der angeblich automatisch funktionierenden Reparatur-Tools richten mehr Schaden an, indem sie krampfhaft versuchen, teilweise zerstörte Daten wiederherzustellen. Weil solche Programme nichts darüber wissen, wie es zum ursprünglichen Problem gekommen ist, gehen sie leicht von falschen Voraussetzungen aus, wenn sie an die Arbeit gehen.

Deshalb ziehen erfahrene Administratoren es vor, zur Datenrettung einfachere Tools zu verwenden, deren Auswirkungen sie besser überblicken können. Auf jeder Knoppix-DVD befinden sich eine ganze Reihe solcher Programme, die Ad-

mins in ihrer täglichen Arbeit verwenden können, um Probleme zu diagnostizieren und Daten zu retten. Weil es sich bei Knoppix um eine Live-DVD handelt, erfüllt es bereits mehrere wesentliche Voraussetzungen für ein Rettungssystem.

Zum einen erlaubt es überhaupt ein System zu booten, selbst wenn die Boot-Festplatte es nicht mehr tut. Zum anderen verhindert der Einsatz von Knoppix als laufendes Betriebssystem, dass weitere Daten auf eine beschädigte Festplatte geschrieben werden, womit das Rettungsvorhaben sehr schnell zum Scheitern verurteilt wäre. Weil Knoppix auch das

Windows-Dateisystem NTFS unterstützt, eignet es sich gleichermaßen zur Rettung von Windows- wie Linux-Systemen.

Das virtuelle Procs-Dateisystem, das normalerweise unter »/proc« gemountet ist, enthält nicht nur Informationen über alle laufenden Prozesse (daher sein Name), sondern auch über Kernel-Subsysteme, Netzwerke und Dateisysteme. Schon mit den einfachsten Unix-Tools wie »cat« und »echo« lassen sich Werte auslesen und setzen. **Tabelle 1** zeigt einige Beispiele für im Procs enthaltene Informationen. **Listing 1** zeigt, wie man die Interrupts eines Systems ausliest.

Auch das Sysfs, das üblicherweise unter »/sys« gemountet ist, führt Laufzeitinformationen des Kernels, in diesem Fall der Hardware-Konfiguration eines Systems. Der folgende Befehl liest aus dem Sysfs die eingebauten Block-Devices (Festplatten) aus:

```
cd /sys/block; for i in sd*; do echo -n "
$i: "; cat $i/device/model; done
```

Über das Sysfs lässt sich beispielsweise die LED ausschalten, die anzeigt, ob das Touchpad aktiv ist:

```
echo 0 > /sys/devices/platform/eeepc/leds/
eeepc::touchpad/brightness
```

Das Sysfs kann genauso durchsucht werden wie ein konventionelles Dateisystem.

Knoppers Rettungstipps

Meine persönlichen Faustregeln bei der Datenrettung sind:

- Sie sollten das System kennen.
- Lesen und schreiben Sie nur so oft wie wirklich nötig, vor allem, wenn der Datenträger beschädigt ist. Wer immer und immer wieder versucht, einen defekten Sektor zu lesen, macht die Sache möglicherweise noch schlimmer.
- Kopieren Sie die komplette Festplatte, nicht nur die mutmaßlich interessanten Teile – zumindest wenn es die Zeit erlaubt. Vielleicht finden sich an anderen Stellen auf der Platte Anhaltspunkte, um Daten zu retten.
- Arbeiten Sie immer nur mit einer Kopie, wenn Sie Änderungen schreiben. Ich verwende Disk-Images statt direkt mit einer

Festplatte zu arbeiten. Natürlich muss das Dateisystem groß genug sein, um eine entsprechend große Image-Datei speichern zu können. Ext4 sollte auf jeden Fall genügen. Möglicherweise müssen Sie dazu sogar eine extra Festplatte besorgen.

- Geduld ist gefragt. Datentransfer über eine USB2-Schnittstelle ist nicht besonders schnell, nur etwa 5 MByte/s. Überlegen Sie sich, ob Sie genug Zeit haben, um alle Daten zu kopieren.
- Prüfen Sie vor jedem Befehl noch einmal die Geräte- und Partitionsbezeichnungen. Sie arbeiten meistens mit Root-Rechten, da ist eine Partition schnell einmal überschrieben, wenn Sie nur eine falsche Ziffer verwenden!

Wer nach einem spezifischen Feature suchen möchte, dessen Namen bekannt ist, kann dazu einfach den Find-Befehl verwenden, zum Beispiel für den Schalter »rfkill«, der die Wireless-Antenne abschaltet:

```
find /sys -iname \*rfkill\*
```

Einige – aber nicht alle – WLAN-Chipsets lassen sich aktivieren, indem man den Wert 1 in die entsprechende Rfkill-Datei schreibt. Umgekehrt schaltet eine 0 sie aus. Das kann sich als praktisch erweisen, wenn die dazu gehörige Funktionstaste nicht mehr funktioniert. Geladene Module und ihre Parameter finden sich in »/sys/module/modulename«, wo Sie gegebenenfalls die Parameter überprüfen können, die beim automatischen Laden eines Moduls verwendet wurden.

Aus dem Procs und dem Sysfs zu lesen erfordert normalerweise keine Root-Rechte (außer es geht um sensible Daten wie Krypto-Schlüssel und Ähnliches), zum Schreiben sind sie allerdings nötig. Ein Befehl wie »sudo echo -n 1 > /sys/devices/platform/eeepc/cdr« (der den internen Card-Reader aktiviert) wird wegen der Umleitung durch »>« nicht funktionieren. Verwenden Sie stattdessen besser »echo -n 1 | sudo tee /sys/devices/platform/eeepc/cdr«, was die Ausgabe des Echo-Befehls dem Tee-Kommando übergibt, das mit Root-Rechten läuft.

Die beiden bekannten Befehle »ps« und »lspci« stellen bequeme textuelle Benutzer-Interfaces zur Verfügung, um die Informationen aus Procs und Sysfs anzuzeigen. So zeigt

```
lspci -v -k
```

die vorhandene Hardware und die dazu gehörigen Kernel-Module an. Der Befehl »dmesg« verrät die Reihenfolge, in der die Hardware vom Kernel erkannt und initialisiert wurde. Die Ausgabe enthält möglicherweise auch Fehlermeldungen, die ansonsten

einfach untergehen. So sind manche Festplatten-Fehler zuerst mit Dmesg zu erkennen, wie das Beispiel in Listing 2 zeigt. Der Lesefehler bei der Datei »test.data« entpuppt sich bei »dmesg« als physischer Lesefehler der Festplatte.

Synchron

Das Rsync-Kommando ist eine gute Wahl, wenn es darum geht Backups anzulegen, komplexe Verzeichnisbäume zu kopieren und dabei noch übers Netzwerk sichere SSH-Verbindungen zu verwenden. Der Befehl

```
rsync -HavP Original Kopie
```

legt eine identische Kopie einer Datei oder eines Verzeichnisses an und schließt dabei dank »-a« Dateitypen, Änderungsdatum und Zugriffsrechte ein.

Die Option »-H« sorgt dafür, dass Hard-Links beim Kopieren erhalten bleiben, »-v« und »-P« zeigen den Fortschritt an. Bricht der Vorgang ab, macht Rsync beim nächsten Aufruf an der richtigen Stelle weiter und verarbeitet nur die bisher nicht kopierten Daten. Mit einem Aufruf der Art »Benutzer@Rechner:Verzeichnis« für Quelle oder Ziel verwendet Rsync automatisch SSH zur Übertragung.

Blockweise

Sowohl »dd« (das Standard-Unix-Kommando) wie auch »dd_rescue« (ein spezialisiertes Derivat des Ahnen) kopieren Daten blockweise. Der Hauptunterschied liegt, abgesehen von der unterschiedlichen Aufrufsyntax, darin, dass »dd_res-

Listing 1: »cat /proc/interrupts«

```
01          CPU0          CPU1
02  0:         1021         0  IO-APIC-edge    timer
03  1:        150241         0  IO-APIC-edge    i8042
04  9:        4024535         0  IO-APIC-fastestoi    acpi
05 12:       13426923         0  IO-APIC-edge    i8042
06 14:       1386145         0  IO-APIC-edge    pata_sch
07 15:           0         0  IO-APIC-edge    pata_sch
08 16:           9         0  IO-APIC-fastestoi    pciehp
09 17:       3398067         0  IO-APIC-fastestoi    pciehp, ath9k
10 18:           0         0  IO-APIC-fastestoi    uhci_hcd:usb4
11 19:       2384012         0  IO-APIC-fastestoi    ehci_hcd:usb1
12 20:       199725         0  IO-APIC-fastestoi    uhci_hcd:usb2
13 21:       1487680         0  IO-APIC-fastestoi    uhci_hcd:usb3
14 22:       22998855         0  IO-APIC-fastestoi    psb@pci:0000:00:02.0
15 23:       3084145         0  IO-APIC-fastestoi    hda_intel
16 24:           1         0  PCI-MSI-edge    eth0
17 ...
```

Listing 2: Lesefehler in »dmesg«

```
01 knopper:~# md5sum test.data
02 md5sum: test.data: Input/output error
03
04 knopper:~# dmesg|tail
05 ide: failed opcode was: unknown
06 end_request: I/O error, dev hda, sector 119422514
07 hda: dma_intr: status=0x51 { DriveReady SeekComplete Error }
08 hda: dma_intr: error=0x40 { UncorrectableError }, LBAsect=119422536,
    high=7, low=1982024, s
09 ector=119422522
```

cue« auftretende Lesefehler überspringen kann und die nicht lesbaren Daten mit Nullen ersetzt, sodass die Länge kaputter Dateien erhalten bleibt. Um eine fehlerfreie Festplatte in eine Image-Datei zu kopieren, verwenden Sie den Befehl:

```
dd if=/dev/sda of=sda.img bs=1M
```

Normalerweise beendet sich »dd« bei Fehlern oder nimmt jedenfalls mit »conf=noerror« Änderungen an den Daten vor. Besser geht das deshalb mit »dd_rescue«, wie Listing 3 zeigt.

Die resultierende Image-Datei »sdb.img« enthält alle von »/dev/sdb« noch lesbaren Informationen. Die fehlerhaften Bytes ersetzt das Tool durch Nullen, die Image-Datei behält so die Originalgröße bei und lässt sich im besten Fall auch per Loopback mounten.

In manchen Fällen empfiehlt sich die Option »-r«, die »dd_rescue« anweist, die Festplatten von hinten zu lesen, also mit dem letzten Block zu beginnen. Wenn die Partitionstabelle noch in Ordnung ist und die vorhandenen Dateisysteme richtig beschreibt, können Sie auch statt der kom-

Tabelle 1: Informationen im Procs

»cat /proc/cpuinfo«	Read CPU information
»cat /proc/meminfo«	Read memory usage information
»cat /proc/interrupts«	Read interrupt allocation
»cat /proc/partitions«	Read detected disk partitions
»echo 1 > /proc/sys/net/ipv4/ip_forward«	Enable IP forwarding:

Listing 3: Kopieren einer fehlerhaften Festplatte

```
01 dd_rescue -A /dev/sdb sdb.img
02
03 dd_rescue: (info): ipos: 1667072.0k, opos: 1667072.0k, xferd: 1667072.0k
04 errs: 0, errxfer: 0.0k, succxfer: 1667072.0k
05 +curr.rate: 5270kB/s, avg.rate: 5808kB/s, avg.load: 8.1%
```

pletten Disk auch einfach eine Partition kopieren. Dazu ersetzen Sie im Aufruf von **Listing 3** »/dev/sdb« beispielsweise durch »/dev/sdb3«. Wollen Sie der ohnehin schon kaputten Platte überflüssige Bewegungen ersparen, können Sie mit »-s Startposition« das Tool anweisen, an einer bestimmten Position zu beginnen.

Testdisk

Ein häufiger Fehlerfall ist eine fehlende oder kaputte Partitionstabelle. Dann kann beispielsweise das BIOS keine Partition zum Booten finden oder Linux findet beim Start die einzubindenden Partitionen nicht. Das kann das Werk eines Windows-Virus sein, der sich im Master Boot Record einnistet oder eine versteckte Partition anlegt. Natürlich kann so ein Fehler auch durch fehlgeschlagene Versuche entstehen, die Platte zu partitionieren, oder einfach durch einen Festplattenfehler in den ersten 512 Bytes. Mindestens im letzten Fall ist das Umkopieren der kompletten Festplatte in eine Image-Datei ein Muss.

Beim herkömmlichen PC-Partitionsschema sind die ersten vier Partitionen direkt in den ersten 512 Bytes der Platte definiert und heißen primäre Partitionen. Die Definition der sogenannten erweiterten (extended) Partitionen befindet sich in weiteren Sektoren.

Ein Backup der primären Partitionstabelle lässt sich leicht mit Unix-Standard-Tools anlegen:

```
dd if=/dev/sda of=sda.mbr bs=512 count=1
```

Um ein solches Backup wiederherzustellen, ohne den vorhandenen Boot Record zu überschreiben, verwenden Sie das folgende Kommando:

```
dd if=sda.mbr of=/dev/sda bs=1 count=642
skip=446 seek=446
```

Ist die Partitionstabelle futsch, ist dennoch nicht alles verloren. Die beiden Tools »testdisk« und »gpart« versuchen Anfang und Ende vorhandener Partitionen zu finden. Testdisk ist für den interaktiven Gebrauch gedacht, Gpart lässt sich mit Kommandozeilenoptionen steuern. Testdisk zeigt in seinem Benutzer-Interface auch Dateien und Verzeichnisse auf gefundenen Partitionen an, womit Sie überprüfen können, ob das Tool richtig geraten hat. Leider ist dieses Feature nicht sehr stabil und kann zum Absturz von Testdisk führen, wenn die Daten einer Partition fehlerhaft sind.

Wer statt mit den Originaldaten lieber mit einer Kopie arbeiten möchte, kann »losetup« zusammen mit einem der Loopback-Geräte »/dev/loop*« verwenden, wie es dieser Artikel weiter unten genauer beschreibt. Es empfiehlt sich, die von Testdisk oder Gpart gefundenen Werte zu notieren und die Festplatte mit »sfdisk« oder »fdisk« zu partitionieren, nachdem man die Blocks der Partitionstabelle gesichert hat.

Fdisk

Das kleine Programm sfdisk kommt häufig in Shell-Skripts zum Einsatz, um eine Festplatte automatisch zu überprüfen und neu zu partitionieren. Es lässt sich aber auch gut dafür verwenden, lesbare Backups von primären und erweiterten Partitionen anzulegen. Der Aufruf

```
sfdisk -uS -d /dev/sda > sda.sfdisk
```

gibt eine Text-Datei aus, wie sie in **Listing 4** zu sehen ist. Die Tabelle zeigt die Partitionsgrenzen in Sektoren an. Ge-

nauso leicht ist es, aus der Textdatei eine Partitionstabelle wiederherzustellen:

```
sfdisk < sda.sfdisk
```

Der Befehl »fdisk« war lange Zeit das Standardprogramm, um auf der Kommandozeile interaktiv Partitionen anzulegen. Seine Ausgabe ist leichter lesbar als diejenige von »sfdisk«, wie **Listing 5** zeigt. Im interaktiven Modus präsentiert es dem Anwender ein Menü, über das er Partitionen anlegen, löschen, Disk-Parameter und anderes ändern kann. Mit »w« schreibt der Anwender die Änderungen auf die Festplatte. Verlässt er das Programm mit »q«, gehen sie verloren, und die Platte bleibt unangetastet. Es empfiehlt sich also, vor dem Schreiben noch einmal mit »p« das Partitionsschema genau anzusehen.

Beide Programme können auch mit Image-Dateien umgehen, allerdings müssen Sie »fdisk« dabei die „Disk-Geometrie“ mitteilen, die es bei einer echten Festplatte automatisch ermittelt.

Loopback

Mit Image-Dateien zu arbeiten, funktioniert viel besser, wenn Sie Linux dazu bringen, sie wie richtige Block-Geräte zu behandeln, einschließlich der I/O-Aufrufe, die die Geometrie und Partitionen ermitteln. Dafür bietet sich der Einsatz eines sogenannten Loopback-Device an, das der Linux-Kernel zur Verfügung stellt. Es bildet eine Datei auf ein oder mehrere Block-Geräte »/dev/loop*« ab. Im System tauchen so weitere Festplatten und Partitionen auf, sobald die Image-Datei mit »losetup« aktiviert ist. So bindet

```
losetup /dev/loop0 sda.img
```

die Datei »sda.img« über die Gerätedatei »/dev/loop0« ein, wenn diese noch zur Verfügung steht. Anschließend lässt sich dieses neue Gerät genauso mit »testdisk« oder »fdisk« bearbeiten wie eine echte Festplatte.

Um auf Partitionen zuzugreifen, die sich in einem Image befinden oder auf einem Blockgerät ohne Partitionseintrag, geben Sie hinter dem Offset-Parameter den Wert in Bytes an:

```
losetup /dev/loop0 sda.img -o Offset
```

Die mit »sfdisk -uS« oder »fdisk -lu« er-

Listing 4: Ausgabe von »sfdisk«

```
01 # partition table of /dev/sda
02 unit: sectors
03
04 /dev/sda1 : start=      63, size=160649937, Id=83,
    bootable
05 /dev/sda2 : start=160650000, size= 8032500, Id=82
06 /dev/sda3 : start=168682500, size=319709565, Id=83
07 /dev/sda4 : start=      0, size=      0, Id= 0
08
09 ---
```

Anzeige wird
separat angeliefert

haltenen Werte müssen Sie dazu mit 512 multiplizieren.

In **Listing 6** beginnt die erste Partition beim Sektor 8192, »losetup« weist sie also mit dem folgenden Befehl richtig zu:

```
losetup /dev/loop0 sda.img -o $((512 * 8192))
```

Hierbei multipliziert das Shell-Konstrukt »\$(())« die beiden Werte. Jetzt sollte es möglich sein, die Partition zu mounten. Das Dateisystem weiß selbst, wo es beginnt und endet, also macht es nichts aus, dass die per »losetup« gemappte Partition den Rest der Festplatte umfasst. Um das Tool selbst das verwendete Dateisystem erkennen zu lassen, rufen Sie

```
mount /dev/loop0 /mnt
```

auf. Alternativ geben Sie den Dateisystemtyp hinter »-t« an:

```
mount -t Dateisystem /dev/loop0 /mnt
```

Listing 5: Ausgabe von »fdisk«

```
01 fdisk -l /dev/sda
02
03 ---
04
05 Disk /dev/sda: 250.1 GB, 250059350016 bytes
06 255 heads, 63 sectors/track, 30401 cylinders
07 Units = cylinders of 16065 * 512 = 8225280 bytes
08 Disk identifier: 0xabf319e9
09
10 Device Boot      Start       End      Blocks    Id
11 /dev/sda1  *           1       10000     80324968+
12    83  Linux
13 /dev/sda2           10001     10500      4016250
14    82  Linux swap / Solaris
15 /dev/sda3           10501     30401     159854782+
16    83  Linux
17 ---
```

Listing 6: Partitionstabelle eines Image

```
01 sfdisk -uS -d sda.img
02
03 ---
04
05 # partition table of sda.img
06 unit: sectors
07
08 sda.img1 : start=      8192, size=  7736320, Id= b
09 sda.img2 : start=        0, size=        0, Id= 0
10 ...
11
12 ---
```

Weil der Mount-Befehl über Loopback-Geräte Bescheid weiß, können Sie beide Schritte auch in einem erledigen, indem Sie »mount« gleich die richtigen Parameter mitgeben:

```
mount -o loop,offset=$((512 * 8192)) sda.
img /mnt
```

Wenn Mount selbst kein freies Loop-Device findet, geben Sie mit »loop=/dev/loop1« eines an. Jetzt können Sie von der gemounteten Partition nach Belieben Dateien kopieren und auch ändern – wenn Sie sie nicht mit »-r« nur-lesbar gemountet haben.

Kompression und Kryptographie

Dieser Ansatz funktioniert auch mit komprimierten Image-Dateien wie bei Knoppix, wenn Sie die entsprechenden Gerätedateien »/dev/cloop*« verwenden. Je nach verwendeter Version von »losetup« und »mount« ist es dabei obligatorisch, die Datei read-only zu mounten.

»losetup« unterstützt auch die Cryptoloop-Erweiterung des Kernels mit den folgenden Optionen:

- »-e Verschlüsselungstyp«
 - »-k Schlüssellänge«
 - »-N«: verwende Passphrase anstatt eines Hash-Algorithmus als Schlüssel
- Ein Beispiel eines Aufrufs mit AES-Verschlüsselung und 256 Bits langem Key:

```
losetup -e aes -k 256 -N /dev/loop0
partition.img
```

Wenn Ihre Distribution eine spezielle Version von »losetup« aus dem Loop-AES-Paket einsetzt, besitzt der Befehl andere Optionen:

- »-e Verschlüsselungstyp«
- »-H Hash-Algorithmus«.

Der Hash-Algorithmus wird dazu verwendet, die Schlüssel unkenntlich zu machen. Das obige Beispiel sieht damit so aus:

```
losetup -e AES256 -H unhashed2 /dev/loop0
partition.img
```

Mounten lässt sich eine solche Partition wie oben beschrieben, nur die Daten werden verschlüsselt gespeichert. Wenn »losetup« merkwürdige Fehlermeldungen ausgibt, müssen Sie möglicherweise erst die entsprechenden Verschlüsselungsmo-

dule in den Kernel laden wie im folgenden Listing:

```
for m in loop cryptoloop aes_generic
aes_i586 cbc; do
[ -d /sys/module/"$m" ] || modprobe "$m"
done
```

Vergessen Sie nicht, nach dem Unmounten die Datei wieder vom Loopback-Device zu trennen. Sonst bleibt es für andere Anwendungen blockiert. Wenn sie Unmount mit der Option »-d« aufrufen, geschieht dies automatisch. Haben Sie es vergessen, können Sie es mit »losetup -d /dev/loop0« nachholen.

Forensik

Einige Rettungsvorhaben wie die Wiederherstellung einzelner Dateien lassen sich auch unternehmen, ohne sich mit Partitionen und Dateisystemen aufzuhalten. Das Tool »foremost« untersucht die rohen Daten direkt nach bekannten Dateitypen und legt für seine Fundstücke neue Dateien an. So durchsucht

```
foremost sda.img
```

die Image-Datei »sda.img« nach bekannten Dateien wie PNG-Bildern, Filmen und Office-Dokumenten – die alle einen Header besitzen, der die Dateigröße verzeichnet. Gefundene Dateien legt Foremost per Default im Verzeichnis »output« ab.

Bei Unix-Dateisystemen, die Blocks mit Verweisen auf andere Blocks enthalten (siehe dazu **[1]**), kann die Option »-d« hilfreich sein. Wenn sich einige Dateien nicht komplett wiederherstellen lassen, weil sie teilweise überschrieben wurden, probieren Sie die Foremost-Option »-a« aus, die das Tool auch Fragmente wiederherstellen lässt. Der Einsatz dieser Option lässt die Menge an Daten explodieren, weil sich überlappende Teile mehrfach wiederhergestellt werden.

Wie Foremost stellt auch Photorec Bilder, andere Multimedia- und Office-Dateien wieder her, macht aber viel mehr als sein Name vermuten lässt. Das Tool ist interaktiv und erinnert an das oben vorgestellte Testdisk, vermutlich weil es vom selben Autor stammt. Das ist auch der Grund, warum es bei Debian-Distributionen im gleichen Paket steckt wie Testdisk. Wenn Sie nur Dateien von einer versehentlich formatierten Flash-Disk retten

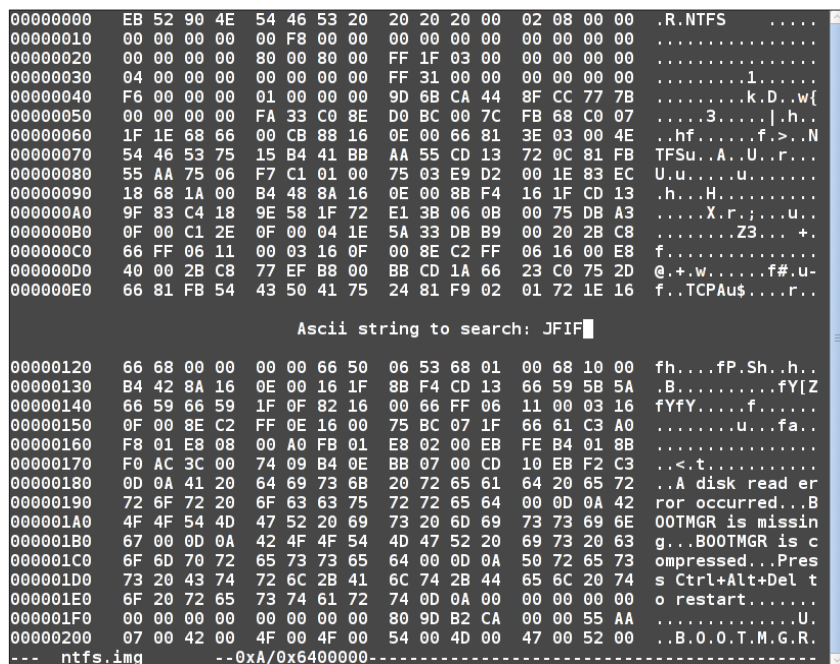


Abbildung 1: Wenn nichts mehr hilft, hilft vielleicht noch Hexedit: byteweises Durchsuchen einer Festplatte.

wollen (wobei das bei den meisten Kameras nur bedeutet, dass der Dateindex gelöscht wurde), sollten Sie als Erstes Foremost oder Photorec ausprobieren.

Hexadezimal

In manchen Fällen hilft keines der hier vorgestellten Werkzeuge weiter, zum Beispiel weil große Teile einer Partition überschrieben wurden oder unlesbar sind. Dann ist es aber immer noch möglich, einzelne Teile von wichtigen Dateien zu retten, die Sie auf keinen Fall verlieren möchten. Hexedit ist ein Tool, mit dem Sie nach einzelnen Bytes oder Zeichenketten suchen können, um festzustellen, ob sich weitere Rettungsversuche lohnen.

Hexedit funktioniert mit echten Festplatten und Partitionen genauso wie mit Image-Dateien. Sie verwenden es interaktiv und können zwischen der Binär-Ansicht (Hexadezimal) links und der Textansicht rechts mit der Tab-Taste umschalten. Das ist besonders wichtig, wenn Sie mit der Taste [/] nach Textmustern suchen. Dazu müssen Sie erst auf die rechte Seite navigieren, denn sonst suchen Sie nach dem entsprechenden Hex-Code.

Abbildung 1 zeigt einen Ausschnitt einer NTFS-Partition in Hexedit. In diesem Beispiel wird die Ascii-Suche nach »JFIF« dazu verwendet, um Jpeg-Bilder zu fin-

den. Die Suche nach Textmustern hilft dabei, Textdateien oder ältere Backups zu finden. Journalling-Dateisysteme unter Linux überschreiben üblicherweise nicht die alte Version einer Datei, wenn eine neue gespeichert wird. So finden sich oft noch zahlreiche Versionen einer Datei, selbst wenn sie mittlerweile gelöscht wurde.

Wenn Sie ein Dateifragment gefunden haben, markieren Sie Beginn und Ende der Daten und speichern es in eine neue Datei. Wenn Sie nicht genau erkennen können, wo eine Datei beginnt und endet, bemessen Sie die Grenzen etwas großzügiger und bearbeiten die Datei später. Hexedit lässt sich auch dazu verwenden, beispielsweise Partitionstabellen direkt zu editieren, und wenn Sie den passenden architekturenspezifischen Maschinencode kennen, können Sie das Verhalten eines Bootloaders ändern, indem Sie den Code direkt ändern und wieder speichern.

NTFS

Seit es NTFS-3G gibt, die Userspace-Implementierung des NTFS-Dateisystems, ist NTFS zu einer guten Wahl geworden, wenn es um den Datenaustausch mit Windows-Anwendern geht. NTFS kann im Gegensatz zu FAT32 mit Dateien größer als 4 GByte umgehen und bringt sogar einige von Linux bekannte

Features mit wie Extended Attributes und Zugriffsrechte. Wenn man das NTFS-3G-Paket installiert, wandern auch die passenden Mount-Helfer auf die Festplatte und der Mount-Befehl lässt sich dann mit »-t ntfs-3g« aufrufen. Einige Distributionen haben den NTFS-Typ auch auf NTFS-3G umgebogen. Im Gegensatz zum Kernel-Treiber beherrscht NTFS-3G auch das Schreiben von NTFS-Partitionen. Eine NTFS-Partition mounten Sie mit:

```
ntfs-3g /dev/partition /mountpoint
```

oder

```
mount -t ntfs-3g /dev/partition /mountpoint
```

Befindet sich die Partition in einem inkonsistenten Zustand, versucht NTFS-3G sie selbstständig zu reparieren, zumindest soweit, dass sie sich mounten lässt. Das kleine Werkzeug »ntfsfix« verbirgt sich im Paket »ntfsprogs«. Es macht allerdings nicht mehr, als das NTFS-Journal zurückzusetzen und das Dateisystem für den nächsten Windows-Boot zum Check zu markieren. Dieser Schritt kann Windows helfen, ein Dateisystem zu reparieren. Um ein beschädigtes Dateisystem mit NTFS-3G zu mounten, ist es meistens überflüssig.

Der Befehl »ntfsclone« erzeugt eine exakte Kopie eines NTFS-Dateisystems und schreibt sie in ein sogenanntes Sparse-File. Darin sind nur die wirklich beschriebenen Blocks enthalten. Nicht belegte Sektoren überspringt das Tool beim Kopieren. Zwar können auch »rsync« und »cp« Kopien eines gemounteten Dateisystems anfertigen, aber nur »ntfsclone« erhält die NTFS-spezifischen Extended Attributes.

Fazit

Das war eine kurze Tour durch die nützlichsten in Knoppix enthaltenen Rettungs-Tools. Ich hoffe, dass Sie von ihnen irgendwann produktiv Gebrauch machen können – oder am besten, dass Sie sie nie brauchen. (ofr)

Infos

[1] Ben Martin, Wie funktionieren Linux-Dateisysteme?, ADMIN 05/2010, <http://www.admin-magazin.de/content/wie-funktionieren-linux-dateisysteme/>



Wiki-Engines im Praxistest

Webstuhl fürs Web

Sie sind duldsame Datenspeicher, Sammelstellen für Ideen oder Dokumentationen und wahre Fundgruben für Wissen aller Art. Wiki-Programme bieten faszinierende Werkzeugkästen und sind - dank Wikipedia - aus dem Internet nicht mehr wegzudenken. Sechs dieser Engines beweisen sich im Praxistest. Mela Eckenfels

Wikis (Hawaiisch für „schnell“) sind schon lange Bestandteil des privaten und beruflichen Alltags. Familien und Freunde sammeln ihre leckersten Rezepte, Fans tauschen sich über Fernsehserien aus, Linuxer schreiben gemeinschaftlich Artikel zu ihrer Lieblingsdistribution, und auch in vielen Unternehmen ersetzen sie teure Knowledgebase-Software. Wikis eignen sich ideal, um die eigenen Gedanken und Worte zu weben, und erleichtern das Texten im Team.

Das Handwerkszeug besteht aus einem Browser auf Anwenderseite und einer Wiki-Software auf dem Server. Die Zahl dieser Engines ist inzwischen ins Dreistellige gewachsen, und so ist es an der Zeit für einen Test, wer die schönsten Seiten webt.

Der Stoff, aus dem die Träume sind

Wikis sind fast so alt wie das Web selbst. Bereits 1994 erkannte Ward Cunningham die Notwendigkeit, den Inhalt vom HTML-Code einer Webseite zu trennen und gleichzeitig den Nutzern eine vereinfachte Textauszeichnung an die Hand

zu geben, damit sie gemeinsam auf einer Plattform arbeiten und Wissen ohne lästige Zeitverzögerung verwalten können. Revolutionär in seinem Wikiwikiweb [1] war sicherlich auch die Idee der „Wiki-Wörter“ (CamelCase), die automatische Verlinkung durch Wörter mit Binnenmajuskel zu den gleichnamigen Wiki-Seiten.

Bis Ende der 1990er fristeten Wikis eher ein Schattendasein und kamen hauptsächlich bei Software-Entwicklern zum Einsatz. Erst mit dem inzwischen wohl bekanntesten Wiki-Projekt, der Wikipedia, traten sie mehr und mehr in den Fokus der Öffentlichkeit. Zu diesem Zeitpunkt war das Ur-Wiki schon wieder in der Bedeutungslosigkeit verschwunden. Ein Grund lag sicherlich im eingeschränkten Markup, aber auch die offene Struktur früher Wiki-Engines ist inzwischen überholt.

Heutzutage ist Wiki-Software entweder gegen Internetphänomene der zerstörerischen Art gerüstet oder sie stirbt aus. Eine Übersicht der letzten Änderungen samt Versionskontrolle, Benutzerverwaltung und Spamschutz sind inzwischen ein Muss. Waren die ersten Wikis noch

in Perl geschrieben, so ist laut Wikimatrix [2] die beliebteste Programmiersprache jetzt PHP, dicht gefolgt von Java und C. Etwa die Hälfte aller Wiki-Engines speichert Inhalte in einer Datenbank, viele andere nutzen das Dateisystem zur Ablage, nur eine Minderheit verwendet für diesen Zweck ein Versionskontrollsystem.

Im Test treten Mediawiki, Moinmoin, Dokuwiki, Twiki, Tikiwiki und Tiddlywiki gegeneinander an und zeigen, was sie in puncto Usability und Administration zu bieten haben. Die letzten beiden Kandidaten laufen außer Konkurrenz – Tikiwiki ist weit mehr als nur eine Wiki-Engine, sondern eher eine ausgereifte Community-Plattform, und Tiddlywiki eignet sich als „Wiki on a stick“ nicht für große Datensammlungen und viele Benutzer.

► Mediawiki

Am Platzhirsch kommt keiner vorbei: Es dürfte kaum jemanden geben, der nicht eines der unzähligen auf Mediawiki [3] basierenden Wiki-Projekte nutzt – selbst wenn er die Wikipedia vermeidet. Hobbyköche tauschen sich im Rezepte-Wiki [4]

aus und Trekkies befragen die Wissensdatenbank Memory Alpha [5]. Wer kurz nachschlagen möchte, welche Sherlock-Holmes-Anspielung in der letzten Folge von „Dr. House“ vorkam, landet garantiert beim Dienst Wikia [6], der unter anderem zu vielen Fan-Wikis beliebter Fernsehserien verlinkt.

Mediawiki entstand 2002 und wurde gemeinsam mit der Wikipedia groß. Die Engine ersetzt das in Perl geschriebene Usemod [7], auf dem das Wikipedia-Projekt von 2000 bis 2002 aufbaute. Für Usemod bedeutete der Umstieg das Ende, die Entwickler stellten ihre Arbeit im Jahr 2003 ein. Mediawiki ist in PHP implementiert, als Datenbank-Backend dienen wahlweise MySQL, PostgreSQL oder SQLite.

Die aktuelle Version 1.16.0 zeigt sich deutlich performanter als ihre Vorgänger. Ein Administrator mit sehr guten MySQL-Kenntnissen ist allerdings Voraussetzung für den Betrieb dieser Engine. Soll das Wiki öffentlich zugänglich sein, ist nicht nur die Pflege der Datenbank aufwändig. Mit schöner Regelmäßigkeit tauchen Sicherheitslücken in Mediawiki auf. Diese schließt der Admin besser zügig, da unter anderem Botnetze in Wiki-Installationen nach diesen Schwachstellen suchen. Mediawiki leidet also ein bisschen an der eigenen Popularität.

Die Installation ist unproblematisch. Pakete für die gängigen Betriebssysteme stehen bereit. Alternativ reicht es, die

Quellen von der Projekt-Homepage auf dem eigenen Webpace zu entpacken und das Verzeichnis im Browser aufzurufen. Dort macht der Admin im Prinzip nur noch Angaben zur Datenbank, um die Einrichtung abzuschließen.

Dass der Wiki-Betreiber die Hauptkonfigurationsdatei »LocalSettings.php« danach von Hand umkopieren muss, mag begründet sein – lästig ist es dennoch. Unschön ist ebenfalls, dass saubere URLs (also solche, die nichts über die dahintersteckende Technologie preisgeben, [8]) ohne Apache-Kenntnisse nicht möglich sind. Andere Softwareprojekte machen seit Jahren vor, dass es auch einfacher und benutzerfreundlicher geht, während bei Mediawiki zumindest in diesem Punkt Stillstand herrscht.

Archaisch

Wenig Bewegung gibt es auch in Sachen Markup (Listing 1). Zwar bietet kaum eine andere Wiki-Engine so viele Möglichkeiten, wenn es darum geht, komplexe Darstellungsvorlagen für ganze Seiten oder Seitenteile zu erstellen. Demgegenüber steht allerdings eine stellenweise unintuitive und unübersichtliche Basis-Syntax. Unsinnigerweise unterscheiden sich interne und externe Verlinkungen immer noch im Markup, selbst einfache Tabellen ohne Sonderwünsche für Zeilen- und Spaltenbreite sind nicht ohne Blick ins Handbuch zu erstellen.

Plugins für Editoren und Textverarbeitungen oder Firefox-Addons helfen über die Stolpersteine hinweg, eine bessere Dokumentation wäre aber schön. Das Handbuch wirkt oft an der Zielgruppe vorbeigeschrieben – die Mediawiki-/Wikipedia-Community schafft es nicht, über den Tellerrand der eigenen Begrifflichkeiten hinwegzuschauen.

Neueinsteiger verwöhnt Mediawiki also nicht – Poweruser dagegen bekommen mit Templates, Parsing Instructions, Includes und Variablen einen Werkzeugkasten mit mächtigen Tools an die Hand. Look & Feel des eigenen Wiki passt der Admin mit Skins an. Das Netz ist eine wahre Fundgrube für diese Oberflächen. Dennoch nutzen viele Mediawiki-Installationen den Standard namens Monobook (siehe Abbildung 1), was allerdings schnell zu Verwechslungen mit der Wikipedia führt.

Bei aller Kritik bleibt Mediawiki eine gute Engine. Fast alle Wiki-Fans haben schon eine Mediawiki-Installation genutzt oder zu einem mit Mediawiki betriebenen Projekt beigetragen und sich dabei an Look & Feel und auch an das unhandliche Markup gewöhnt.

► Moinmoin

Zu Unrecht hat das Wiki mit dem plattdeutschen Namen [9] den Ruf, ein selten genutzter Exot oder eine rein deutsche Entwicklung zu sein. Zahlreiche Open-Source-Projekte und Firmen auf dem ganzen Globus setzen auf diese Engine im Intra- oder Internet, darunter Dropbox, Open Office und Edubuntu.

Moinmoin ist in Python geschrieben und legt die Wiki-Inhalt als normale Textdateien ab. Die Engine speichert Revisionen nicht inkrementell, sondern erstellt für

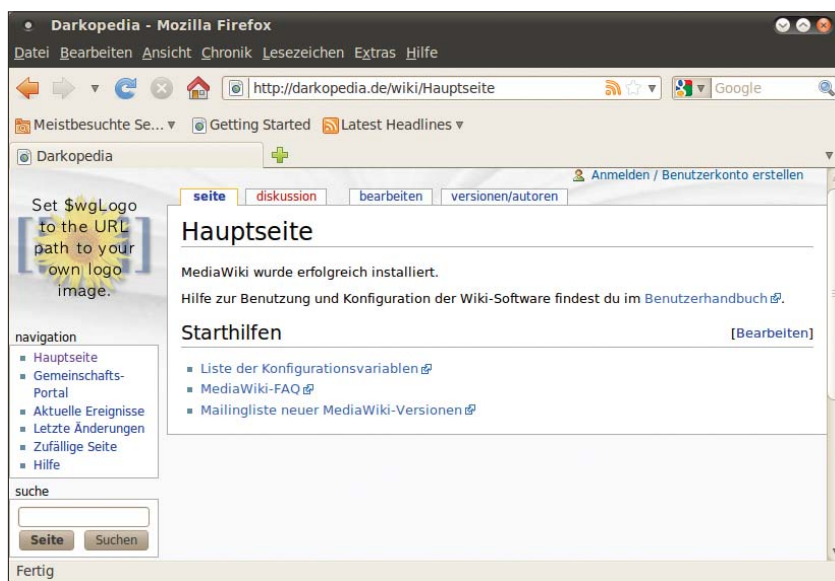


Abbildung 1: Verwechslungsgefahr: Es sieht aus wie Wikipedia, ist aber eine frische Mediawiki-Installation mit der Standard-Skin Monobook.

Listing 1: Mediawiki-Markup

```
01 ==Hallo Welt==
02 Wikis sind eine '''fette Sache''', aber manchmal auch
   ganz schön '''schräg'''.
03 [[Intern|Drinnen]] gibts viel zu sehen, [http://www.
   linux-magazin.de/ Draussen] aber auch.
04 {
05 |Übersichtlichkeit
06 |in Tabellen
07 |-
08 |ist nicht immer
09 |gegeben.
10 }
```

jede Änderung eine neue Datei. So kann eine über Jahre ungepflegte Moinmoin-Installation schnell zum Platzfresser werden. Die Wiki-Software liefert daher mit dem Kommandozeilentool Moin [10] einen zuverlässigen Administrations-Helfer aus, der solche und andere Missstände beseitigt.

Moinmoin bietet zwei Installations-Szenarien an: eine Quick-Install- und eine Server-Variante. Erstere setzt lediglich den Download des Tarball voraus, bietet einen eingebauten Webserver, der anstelle oder parallel zu einem laufenden arbeitet, und eignet sich daher ideal für das Einrichten im eigenen Homeverzeichnis. Die zweite Variante setzt Root-Rechte voraus; eine Anleitung zur Modifikation der entsprechenden Konfigurationsdateien findet sich im Handbuch der Wiki-Software.

Eine bequeme Einrichtung über den Browser ist auch in der aktuellen Version 1.9 nicht möglich, sodass die Konfiguration zur Frickelei ausartet, wenn kein Shellzugriff und nur ein FTP-Zugang zum Server besteht. Moinmoin entschädigt den Anwender mit einer übersichtlichen Konfigurationsdatei »wikiconfig.py«.

Um das Wiki zu erweitern, kann der Nutzer Makros in normale Seiten einbinden, global zusätzliche Funktionen mit Actions bereitstellen oder auf Themes zurückgreifen. Im Angebot der Zusatzsoftware sind Parser, Formatter und XML-RPC-Funktionen. Alle Addons richtet der Admin auf der Shell ein, eine Installationsmöglichkeit über ein Webinterface sucht er auch hier vergebens.

Unscheinbarer Star?

Verglichen mit der Fülle an Erweiterungen für Mediawiki ist das Angebot von Moinmoin eher überschaubar. Mit ein wenig technischem Sachverstand und minimalen Programmierkenntnissen sollte es aber problemlos möglich sein, beste-

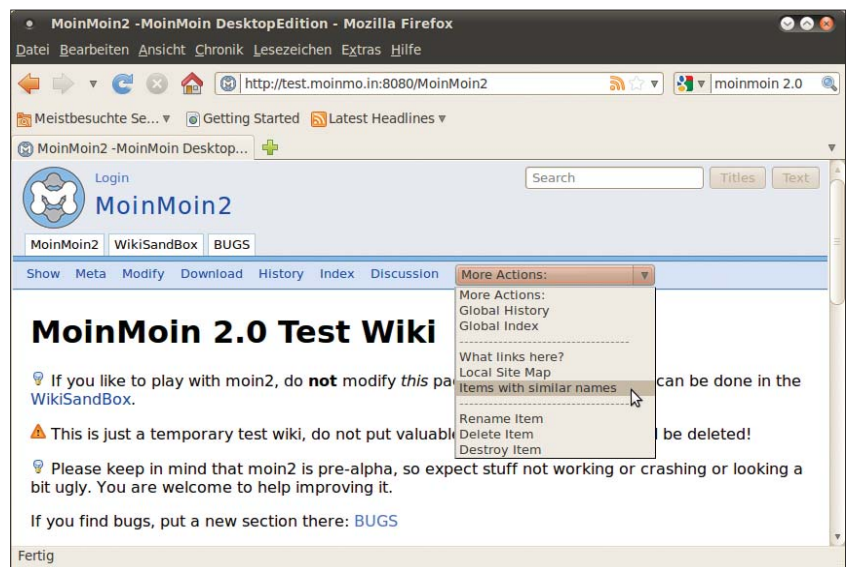


Abbildung 2: Die Moinmoin-Entwickler versprechen zahlreiche Neuerungen für die nächste Release. Einen Vorgeschmack auf Version 2.0 gibt's bereits auf der Projekt-Webseite.

hende Addons zu patchen und an eigene Bedürfnisse anzupassen.

Wer mit Moinmoin mehrere Wiki-Instanzen auf einem Server betreiben möchte, installiert die Software ein Mal und vervielfältigt das Verzeichnis, auf das die User übers Web zugreifen. Dazu ist nicht mehr als ein halbes Dutzend Handgriffe nötig, die schnell in ein Skript verpackt sind. Moinmoin leistet darüber hinaus auch gute Dienste als Contentmanagement-System. Access Control Lists erlauben eine gezielte Rechtevergabe für die gesamte Installation oder für einzelne Unterseiten.

Die Moinmoin-Entwicklung schreitet extrem schnell voran. Allein im Jahr 2008 gab es drei große Versionssprünge. Die Programmierer arbeiten momentan auf Hochtouren an Version 2.0, die viele neue Features mitbringen soll [12]. Ein Testwiki mit Sandbox lädt zum Experimentieren ein (siehe **Abbildung 2**). Eine Überarbeitung großer Teile des Code ist vorgesehen. Anstelle von Seiten und Attachments kennt die Engine dann nur noch Items verschiedenen Typs. Außerdem integriert sie über ein Backup-API verschiedene Datenspeicher. Auch wollen die Moinmoin-Mitarbeiter stärker auf den Code anderer Projekte zurückgreifen, zum Beispiel auf Jinja2 [13] als Template-Engine.

Auf den ersten Blick scheint Moinmoin nur eine durchschnittliche Wiki-Engine ohne große Besonderheiten zu sein.

Schaut man etwas genauer hin, stellt sich die Software jedoch als flexibel und leistungsstark heraus. Moinmoin bleibt selbst bei hohen Zugriffszahlen performant, durch die Theme-Schnittstelle passt der Admin das Wiki mit verhältnismäßig geringen Aufwand an eigene Bedürfnisse an und die Hilfeseiten sind tatsächlich hilfreich.

Bereits seit Version 1.6 ist mit Hilfe des Creole-Projekts [14] ein grundlegend überarbeitetes und vereinfachtes Markup entstanden (siehe **Listing 2**). Die Engine punktet mit einem guten Spamschutz, nachdem in der Vergangenheit Spambots viele der mit Moinmoin betriebenen Wiki-Projekte attackierten und unbenutzbar machten.

► Dokuwiki

Von allen getesteten Engines gewinnt Dokuwiki [15] den Preis für die schnellste und problemärmste Installation. Das System eignet sich ideal für Anwender, die nicht für jede kleine Konfigurationsänderung auf die Shell wechseln können oder wollen. Das in PHP geschriebene Wiki speichert seine Inhalte genau wie Moinmoin in Textdateien und bewahrt alte Versionen ebenfalls vollständig auf – allerdings in gepackter Form.

Die Einrichtung über ein Webinterface ist in Sekundenschnelle erledigt, über diese Schnittstelle nimmt der Admin auch Feineinstellungen wie die Einrichtung von

Listing 2: Moinmoin-Markup

```
01 = Hallo Welt =
02 Wikis sind eine '''fette Sache''', aber manchmal auch
   ganz schön ''schräg''.
03 [[Intern|Dinnen]] gibts viel zu sehen, [[http://www.
   linux-magazin.de/|Draussen]] aber auch.
04 || Übersichtlichkeit || in Tabellen ||
05 || ist nicht immer || gegeben. ||
```

ACLs und Plugins vor (Abbildung 3). Lediglich zur Installation von Templates ist ein Shellzugriff nötig. Diese lädt der Wiki-Betreiber von Hand auf den Server und entpackt sie im Verzeichnis »lib/tpl/«.

Auf den ersten Blick scheint Dokuwiki sogar in der Lage zu sein, mit einem einzigen Klick saubere URLs zu aktivieren. Von dieser Option sollten Anwender jedoch tunlichst die Finger lassen oder zumindest vorher gründlich die Anweisungen der Rewrite-Dokumentation [16] gelesen haben. Andernfalls ist die Gefahr groß, mit einer unbenutzbaren Installation zu enden.

Besseres Markup

Das Dokuwiki-Markup ist angenehm durchdacht und ähnelt dem von Moinmoin (Listing 3). Beide Engines sind spürbar vom Creole-Markup beeinflusst, das sie zusätzlich zur hauseigenen Auszeichnung unterstützen. Dokuwiki benötigt dazu ein entsprechendes Plugin.

Außer mit der einfachen Administration trumpft die Engine vor allem durch ihr Template-System auf. Eine Suche im Netz führt den User zu vielen schicken Looks, auch die Anleitung zur Erstellung ist gut verständlich. Während fast alle Mediawiki-Installationen gleich aussehen und direkt als Wiki identifizierbar sind, steckt hinter manch aufwändig gestalteter Web-

seite kein sündhaft teures CMS, sondern Dokuwiki [17].

► Twiki

Verglichen mit der reibungslosen Installation von Dokuwiki gerät die Einrichtung von Twiki [18] zu einer Bergwanderung ohne Geländer. Böse Zungen behaupten, das sei man ja von Perl und seinen Entwicklern gewohnt. Auch die Dokumentation ist für Normalsterbliche kaum zu erfassen, ohne dass das Gehirn automatisch in den Bildschirmschoner-Modus schaltet. Aber was ein echter Programmierer ist, der liest keine Anleitungen, der weiß einfach, wie es geht.

Trotzdem – oder möglicherweise genau aus diesem Grund – erfreut sich Twiki der vielleicht größten Nutzergemeinde neben Mediawiki. Es ist die älteste der getesteten Wiki-Engines und hat über die Jahre die Businesswelt für sich eingenommen. Unternehmen wie Motorola oder Disney nutzen Twiki als interne Wissensdatenbank oder Dokumentenverwaltung und setzen es sogar als Projektmanagement-Tool ein.

Bodenständig stabil

Twiki benötigt neben Perl (5.6.1 oder neuer, Version 5.8.4 oder neuer wird empfohlen) eine Reihe von CPAN-Modu-

len, welche die Dokumentation auflistet. Darüber hinaus verwendet die Engine einige GNU-Tools wie Diff, Patch und Egrep/Fgrep. Cron und ein Webserver (Empfehlung der Entwickler: Apache) sind ebenfalls Voraussetzung. Ist Twiki erst einmal installiert, dann läuft und läuft es und erfreut den Verwalter mit zahlreichen Konfigurationsmöglichkeiten innerhalb des Wiki (siehe Abbildung 4). Aber bis dahin geht nichts ohne Perl-kundige Hände. ►

Twiki gehört zu den Engines im Test, die auf eine Datenbank verzichten und die Inhalte im Dateisystem des Servers ablegen. Die Versionskontrolle ist in Perl im Wiki selbst implementiert, optional kann der Admin aber auch RCS [19] verwenden.

Ähnlich wie bei Dokuwiki existieren zahlreiche Erweiterungen für die Engine. Zu den interessantesten Plugins gehört sicherlich Twiki Draw [20], mit dem der Benutzer gleichzeitig Text, Grafiken und Diagramme innerhalb des Wiki bearbeiten kann. Auch wenn das Markup kein Creole unterstützt, ist es doch relativ intuitiv (siehe Listing 4).

Twiki verwaltet die Metadaten einer Seite mit Formularen, während Mediawiki auf Kategorien setzt. Dient ein Wiki beispielsweise als Projektmanagement-Tool, kann der Nutzer jeder Task innerhalb eines Projekts auch eine eigene Seite zuordnen und über ein eingebettetes Formular den Status der Aufgabe verändern.

Twiki erlaubt darüber hinaus eine schnelle Suche über diese Meta-Informationen

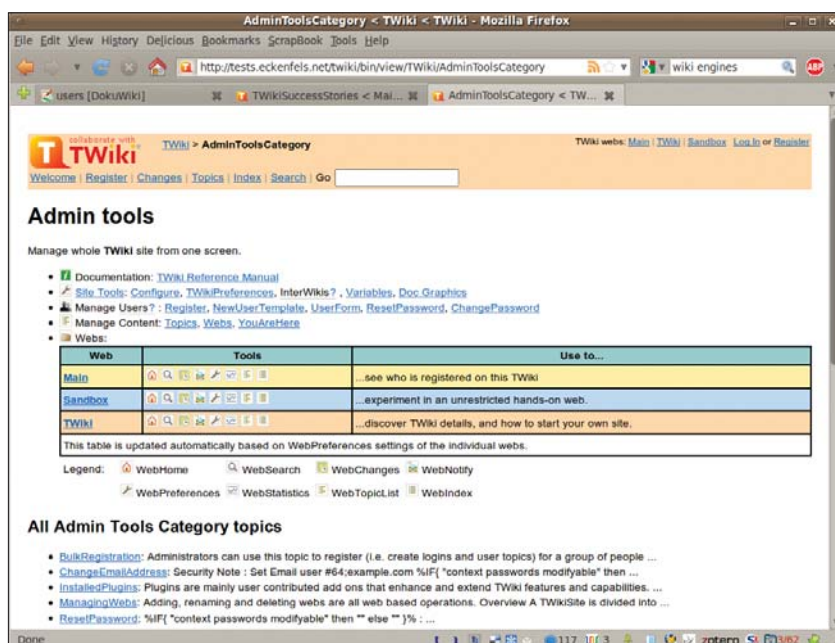


Abbildung 4: Vor die Twiki-Engine haben die Entwickler Perl gesetzt. Danach überzeugt das System aber durch komfortable Einrichtungsmöglichkeiten, die alle auf einer Seite versammelt sind.

Listing 3: Dokuwiki-Markup

```
01 ===== Hallo Welt =====
02 Wikis sind eine *fette Sache*, aber manchmal auch
   ganz schön //schräg//.
03 [[Intern|Drinnen]] gibts viel zu sehen, [[http://www.
   linux-magazin.de/|Draussen]] aber auch.
04 | Übersichtlichkeit | in Tabellen |
05 | ist nicht immer | gegeben. |
```

Listing 4: Twiki-Markup

```
01 ---+ Hallo Welt
02 Wikis sind eine *fette Sache*, aber manchmal auch
   ganz schön _schräg_.
03 [[InTern|Drinnen]] gibts viel zu sehen, [[http://
   www.linux-magazin.de/][Draussen]] aber auch.
04 | Übersichtlichkeit | in Tabellen |
05 | ist nicht immer | gegeben. |
```

Tabelle 1: Feature-Übersicht der getesteten Wiki-Engines

Wiki-Engine	Mediawiki	Moinmoin	Dokuwiki	Twiki	Tikiwiki	Tiddlywiki
Lizenz	GPLv2	GPLv2	GPLv2	GPLv2	LGPL 2.1	BSD
Sprache	PHP	Python	PHP	Perl	PHP	Javascript
Backend	MySQL oder PostgreSQL	Textdateien	Textdateien	Textdateien	MySQL	einzelne HTML-Datei
ACL-Unterstützung	nein	ja	ja	ja	ja	-
Creole-Markup-Unterstützung	geplant	ja	ja	nein	nein	ja
Integrierte Themes	manuelle Inst.	manuelle Inst.	manuelle Inst.	manuelle Inst.	ja	ja
Mehrere Instanzen (Farming)	Plugin	ja	Plugin	Plugin	ja	-

und gibt auf Knopfdruck beispielsweise alle Tasks aus, deren Status auf »erledigt« gesetzt ist.

► Tikiwiki

Tikiwiki [21] präsentiert sich als eierlegende Wollmilchsau und versammelt alles unter einer Oberfläche: Wiki, Blog, Forum, Terminverwaltung, Bildergalerie, CMS, DMS, Groupware, Community – Lebensmittelpunkt. Vom Look & Feel her erinnert Tikiwiki an PHP Bulletin Board [22]. Mit einem normalen Wiki hat die Engine eigentlich nichts mehr gemein und die Benutzerfreundlichkeit leidet schwer unter der „Featuritis im Endstadium“.

Dafür gibt es an der Installationsprozedur wenig zu meckern. Wie die meisten Community-Tools ist Tikiwiki in PHP geschrieben und fordert als Backend eine MySQL-Datenbank. Das Einrichten per Webbrowser läuft tadellos – von Grundfunktionen über Plugins bis hin zur Auswahl verschiedener Themes ist alles drin

(siehe Abbildung 5). Auch das Markup ist relativ einfach zu erlernen, auch wenn es sich durchaus von der Auszeichnung der anderen Wikis unterscheidet (siehe Listing 5).

Tikiwiki eignet sich ideal, um schnell eine kleine Community-Plattform zusammenzuzimmern. Sind allerdings vor allem Wiki-Grundfunktionen gewünscht, sollte der Admin lieber zu einer anderen Engine greifen, da die Fülle der Möglichkeiten schnell überfordert und die Wiki-Funktionalität nur Beiwerk ist.

► Tiddlywiki

Ebenfalls außer Konkurrenz startete Tiddlywiki [23] im Test. Der Grund: Das Wiki für die Hosentasche besteht aus einer einzigen HTML-Datei mit einer eingebetteten Javascript-Anwendung,

welche die Wiki-Funktionalität implementiert. Eine Installation ist daher nicht nötig – der Anwender lädt einfach die Datei »empty.html« herunter, benennt sie entsprechend um und öffnet sie im Browser.

Tiddlywiki ist nicht dafür gemacht, zahlreiche Benutzer oder Änderungen beziehungsweise Seiten zu verwalten. Strenggenommen eignet sich das Wiki nur zum Arbeiten auf dem lokalen Rechner und nicht über eine Internetverbindung hinweg. Es deaktiviert nämlich die Bearbeitungsfunktion, wenn der Nutzer über HTTP darauf zugreift.

Und doch hat das kleine Tool einen Sinn: Auf einem USB-Stick am Schlüsselbund kann man das persönliche Notiz-Wiki mit auf Reisen nehmen und Informationen handlich ablegen sowie durch Tags sortieren (Abbildung 6). Es spricht auch

Listing 5: Tikiwiki-Markup

```
01 ! Hallo Welt
02 Wikis sind eine __fette Sache__, aber manchmal auch
   ganz schön 'schräg'.
03 ((Drinne|Intern)) gibts viel zu sehen, [http://www.
   linux-magazin.de/]Draussen] aber auch.
04 ||Übersichtlichkeit | in Tabellen
05 ist nicht immer | gegeben. ||
```

Listing 6: Tiddlywiki-Markup

```
01 ! Hallo Welt
02 Wikis sind eine 'fette Sache', aber manchmal auch
   ganz schön //schräg//.
03 [[Drinne|Intern]] gibts viel zu sehen,
   [[Draussen|http://www.linux-magazin.de/]] aber auch.
04 | Übersichtlichkeit | in Tabellen |
05 | ist nicht immer | gegeben. |
```

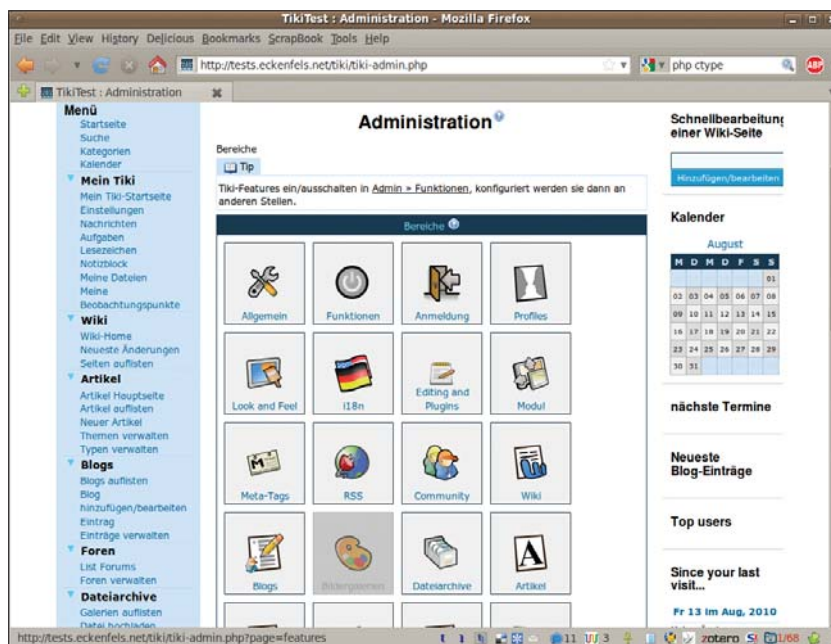


Abbildung 5: Die Administrations-Oberfläche von Tikiwiki sucht ihresgleichen. Aussehen, Sprache, Plugins, RSS-Feeds, Blogs und Foren richtet der Wiki-Betreiber an zentraler Stelle ein.

nichts dagegen, Programmhandbücher oder FAQs mit Tiddlywiki zu erstellen. Die einzelne HTML-Datei können Entwickler problemlos mit einer Installations-CD/-DVD ausliefern. Selbst Passwörter lassen sich in Tiddlywiki mit Hilfe eines Krypto-Plugins verwalten.

Tiddlywiki kooperiert mit den meisten modernen Browsern. Lediglich der Konqueror sperrt sich, wenn es darum geht, die HTML-Datei zu verändern. Das Markup ist relativ überschaubar (siehe [Listing 6](#)), vom Creole-Projekt beeinflusst und unterstützt diese Auszeichnungsform auch.

Fazit

Ein vollständiger Überblick über alle Wiki-Engines würde Bücher füllen – zu groß ist das Angebot und zu unterschiedlich sind die Anforderungen der Nutzer. Alle Testkandidaten waren mehr oder weniger komfortabel im Hinblick auf Installation und Einrichtung. Die Engines, die ins Dateisystem schreiben, zeigten sich im Test zunächst performanter. Das Bild wandelt sich allerdings, wenn der Datenbestand wächst.

Die größten Unterschiede zwischen den Wikis bestehen nach wie vor im Markup.

Die Entscheidung für oder wider eine Software fällt selten nach technischen Gesichtspunkten – ausschlaggebend ist vielmehr, mit welchem System die Nutzerzielgruppe umgehen kann und will. Nicht-übertragbares Markup macht einen späteren Umstieg auf ein anderes Wiki unattraktiv und arbeitsaufwändig. Das Creole-Markup [\[14\]](#) kann dabei helfen, Inhalte ohne Anpassung der Auszeichnungssprache zwischen Wikis zu übertragen.

Unter einem Manko leiden alle getesteten Programme gleichermaßen: Sie wurden von Techies für Techies entwickelt. Wysiwyg-Editoren oder Export-Plugins für Fremdformate verschleiern die Markup-Probleme nicht. Solange sich Anwender mit einer Fülle von Symbolen, Includes und Variablen herumschlagen müssen, zwischen denen sich der eigentliche Text versteckt, bleiben Wikis das Werkzeug der IT-Abteilungen und technikaffiner Datensammler. Es fehlt an Oberflächen, die wirklich intuitiv zu bedienen sind. (*hej/ofr*) ■

Infos

- [\[1\]](#) Wikiwikiweb: [\[http://c2.com/cgi/wiki?WelcomeVisitors\]](http://c2.com/cgi/wiki?WelcomeVisitors)

- [\[2\]](#) Vergleich von Wiki-Software: [\[http://www.wikimatrix.org\]](http://www.wikimatrix.org)
[\[3\]](#) Mediawiki: [\[http://www.mediawiki.org\]](http://www.mediawiki.org)
[\[4\]](#) Rezepte-Wiki: [\[http://www.rezeptewiki.org\]](http://www.rezeptewiki.org)
[\[5\]](#) Star-Trek-Enzyklopädie: [\[http://memory-alpha.org\]](http://memory-alpha.org)
[\[6\]](#) Wikia: [\[http://www.wikia.com\]](http://www.wikia.com)
[\[7\]](#) Usemod: [\[http://www.usemod.com\]](http://www.usemod.com)
[\[8\]](#) Saubere URLs in Mediawiki: [\[http://www.mediawiki.org/wiki/Manual:Short_URL\]](http://www.mediawiki.org/wiki/Manual:Short_URL)
[\[9\]](#) Moinmoin: [\[http://moinmo.in\]](http://moinmo.in)
[\[10\]](#) Kommandozeilentool Moin: [\[http://moinmo.in/HelpOnMoinCommand\]](http://moinmo.in/HelpOnMoinCommand)
[\[11\]](#) Moinmoin 2.0: [\[http://moinmo.in/MoinMoin2.0\]](http://moinmo.in/MoinMoin2.0)
[\[12\]](#) Ankündigung der Moinmoin-2.0-Features auf der EuroPython 2010: [\[http://europepythontalks.blip.tv/file/3980887/\]](http://europepythontalks.blip.tv/file/3980887/)
[\[13\]](#) Jinja2: [\[http://jinja.pocoo.org/2/\]](http://jinja.pocoo.org/2/)
[\[14\]](#) Creole-Projekt: [\[http://www.wikicreole.org\]](http://www.wikicreole.org)
[\[15\]](#) Dokuwiki: [\[http://www.dokuwiki.org\]](http://www.dokuwiki.org)
[\[16\]](#) Saubere URLs in Dokuwiki: [\[http://www.dokuwiki.org/rewrite\]](http://www.dokuwiki.org/rewrite)
[\[17\]](#) Liste von Dokuwiki-Websites: [\[http://www.dokuwiki.org/users\]](http://www.dokuwiki.org/users)
[\[18\]](#) Twiki: [\[http://twiki.org\]](http://twiki.org)
[\[19\]](#) Revision Control System: [\[http://www.gnu.org/software/rcs/rcs.html\]](http://www.gnu.org/software/rcs/rcs.html)
[\[20\]](#) Twiki Draw: [\[http://twiki.org/cgi-bin/view/Plugins/TWikiDrawPlugin\]](http://twiki.org/cgi-bin/view/Plugins/TWikiDrawPlugin)
[\[21\]](#) Tikiwiki: [\[http://www.tikiwiki.org\]](http://www.tikiwiki.org)
[\[22\]](#) PHP BB: [\[http://www.phpbb.com\]](http://www.phpbb.com)
[\[23\]](#) Tiddlywiki: [\[http://www.tiddlywiki.com\]](http://www.tiddlywiki.com)

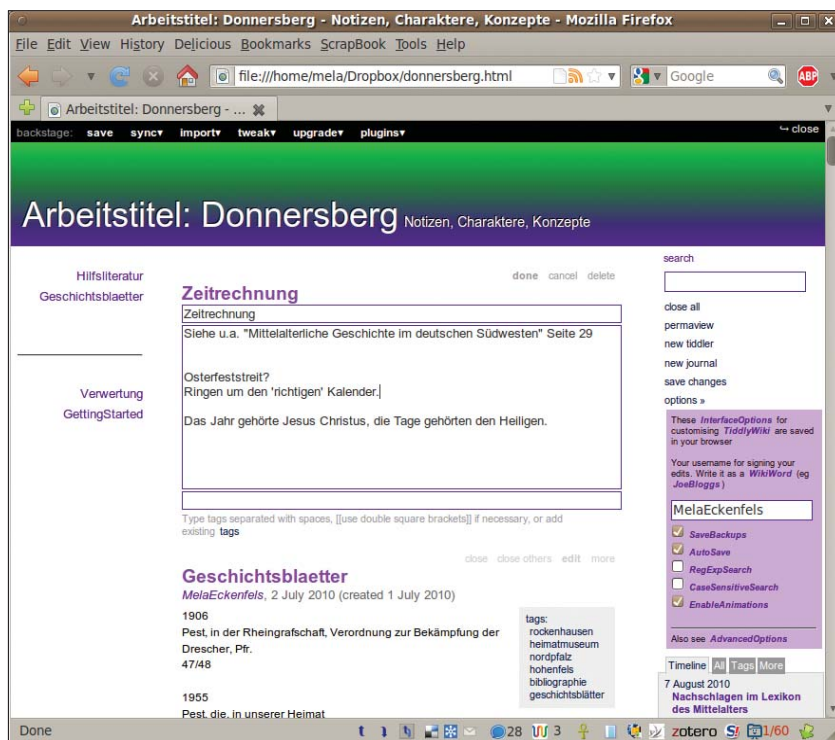


Abbildung 6: Tiddlywiki ist immer bereit und braucht nicht mal eine Internetverbindung. Damit eignet sich die Software ideal für Offline-Ideensammlungen.

Fernstudium IT-Sicherheit

staatlich geprüft und
anerkannt

Aus- und Weiterbildung zur Fachkraft für IT-Sicherheit. Ein Beruf mit Zukunft. Kostengünstiges und praxisgerechtes Studium ohne Vorkenntnisse. Beginn jederzeit.

NEU: PC-Techniker, Netzwerk-Techniker, Linux-Administrator LPI, Webmaster

Teststudium ohne Risiko.
GRATIS-Infomappe gleich anfordern!

FERNSCHULE WEBER
 -seit 1959-
 Postfach 21 61
 Abt. C25
 26192 Großenkneten
 Tel. 0 44 87 / 263
 Fax 0 44 87 / 264

www.fernschule-weber.de

RAID-Systeme unter Linux optimal konfigurieren

Eingestellt

Wer die Parameter seines RAID-Systems unter Linux richtig konfiguriert, kann das Speichersystem um bis zu 15 Prozent beschleunigen. Ben Martin

Unter Linux ein RAID-System zu konfigurieren, ist heutzutage keine große Sache mehr. Ein Aufruf von »mdadm« und »pvcreate« genügt, und schon steht ein RAID-5- oder RAID-10-System Gewehr bei Fuß. Tatsächlich ist die Einrichtung eines RAID so einfach, dass sich viele Administratoren einfach durch die Kommandozeile oder GUI hangeln und keine großen Gedanken auf die vielen möglichen Einstellungen verschwenden, die die Leistung beeinflussen.

Täuschend

Bei RAID-Systemen garantieren die Default-Einstellungen nicht unbedingt optimale Performance. Dieser Artikel zeigt, dass die Abstimmung der Parameter von Dateisystem und RAID einen Unterschied von bis zu 15 Prozent ausmachen.

Ein Dateisystem ist eine nur scheinbar einfache Sache. Auf den ersten Blick scheint nicht viel dahinterzustecken. Eine Datei abspeichern und später wieder lesen, was kann daran schon schwierig sein? Wie so oft steckt der Teufel im Detail. Schon das Speichern einer kleinen Datei ist ein relativ komplexer Vorgang, denn beispielsweise sind auf dem Weg bis zur Festplatte einige Caches im Spiel: Der Kernel hat einen eigenen Cache, der Disk-Controller möglicherweise auch, und schließlich besitzt auch noch die Festplatten-Hardware einen kleinen Zwischenspeicher. Wenn ein Dateisystem also sicherstellen will, dass die Daten wirklich auf die Platte geschrieben wurden, wird daraus ein recht komplexer Vorgang, der auch Zeit kostet. Wenn Sie mit vier Festplatten ein RAID-5 (siehe Kasten) anlegen, schreibt der Li-

nux-Kernel drei Blocks mit Nutzdaten auf drei Platten und legt die Parity-Informationen der drei Blocks auf der vierten Platte ab. Auf diese Art lassen sich die Daten wiederherstellen, selbst wenn eine der vier Platten kaputtgeht. Weil man bei Dateisystemen häufig von Blocks spricht, bezeichnen RAID-Experten die drei Datenblocks meist als „Chunks“, um Verwirrung zu vermeiden.

Bei vier Festplatten gibt es drei Daten-Chunks und ein Parity-Chunk. Die Stripe-Größe ist also drei Chunks und die Parity-Stripe-Größe ein Chunk. Die Stripe-Größe ist sehr wichtig, denn ein Dateisystem sollte versuchen, alle Chunks eines Stripe gleichzeitig zu schreiben, damit der Parity-Chunk aus den drei Chunks im Hauptspeicher berechnet und dann auf die Platte geschrieben werden kann. Der zu schreibende Chunk soll im Folgen-

RAID-5

Dieser Artikel geht davon aus, dass Sie mit den Konzepten und der Konfiguration von RAID-Systemen vertraut sind. Fehlt Ihnen dieses Wissen, finden Sie einige Dokumentation dazu online unter [1]. Das Redundant Array of Inexpensive Disks (RAID) steht für eine Sammlung von Techniken, um fehlertolerante Speichersysteme aufzubauen. Der Begriff stammt aus dem Artikel „A Case for Redundant Arrays of Inexpensive Disks“ von David Patterson, Garth Gibson und Randy Katz [2]. Fehlertolerante Storage-Systeme sind ein Mittel, Datensicherheit zu garantieren, wenn beispiels-

weise eine Festplatte versagt. Der einfachste Weg dazu führt über die Duplizierung der Daten auf zwei Festplatten. Das entspricht mehr oder weniger der bekannten Mirroring-Methode, die als RAID-1 bekannt ist. Auch wenn Disk-Mirroring sicher einige der Probleme löst, ist es weder besonders elegant noch besonders effektiv. 50 Prozent des verfügbaren Speicherplatzes werden darauf verwendet, redundante Informationen zu speichern. Der ursprüngliche Artikel der RAID-Autoren untersucht alternative Wege, um auf effizientere Art Fehlertoleranz zu gewährleisten. Ein in der IT-Welt

weitverbreiteter Ansatz ist das Disk-Striping mit Parity, das auch als RAID-5 bekannt ist. RAID-5 erfordert drei oder mehr Festplatten. In einem RAID-Array mit N Festplatten landen die Daten auf N-1 Platten, während die verbleibende Disk die Parity-Informationen speichert, die zur Wiederherstellung der Originaldaten dient, wenn eine Datenplatte ausfällt. Die Anzahl von Festplatten, die zur Speicherung redundanter Daten abgestellt sind, ist deshalb 1/N. Je mehr Festplatten also im RAID-Verbund stecken, desto geringer ist die relative Verschwendung von Plattenplatz.

den Chunk-X heißen, der Parity-Chunk Chunk-P. Eine Möglichkeit ist es, die anderen Daten-Chunks des Stripe zu lesen, Chunk-X zu schreiben, Chunk-P neu zu berechnen und dann zu schreiben. Alternativ kann das RAID Chunk-P und den alten Chunk-X lesen und dann Chunk-P so ändern, dass er zum alten Chunk-X passt.

Wie man sieht, können sich RAID-Algorithmen ziemlich verkomplizieren, wenn man nicht nur zusammenhängende Blocks nacheinander auf das System schreibt. Noch komplexer wird es, wenn man das Dateisystem miteinbezieht, wie etwa die Metadaten, die für jede Datei und jedes Verzeichnis existieren. Wenn der Anwender zum Beispiel eine Datei anlegt oder löscht, ändern sich die Metadaten, das heißt einige wenige Daten müssen auf die Festplatte geschrieben werden.

Abhängig vom Aufbau eines Dateisystems muss es auch beim Lesen eines Verzeichnisses viele kleine Datenssegmente von der Platte lesen. Wenn das Dateisystem über Chunk- und Stripe-Größe Bescheid weiß, kann es die Metadaten in Chunks ablegen, was wiederum den Zugriff vereinfacht und beschleunigt.

Am Anfang der optimalen Abstimmung der Parameter von Dateisystem und RAID steht, die Stripe- und Chunk-Größe beim Anlegen des Dateisystems richtig festzulegen. Wer XFS verwendet und das Dateisystem direkt auf dem RAID anlegt, hat es leicht, denn »mkfs.xfs« wählt automatisch die richtigen Werte. Wenn allerdings der Logical Volume Manager (LVM) ein RAID verwendet und man ein XFS-System auf einem Logischen LVM-

Details

Dieser Artikel konzentriert sich auf die Ext3- und XFS-Dateisysteme. Beim Test kam ein RAID-5 mit vier 500-GByte-Festplatten zum Einsatz. Im Einzelnen waren es drei Samsung HD501LJ mit 500 GByte und eine Samsung HD642JJ mit 640 GByte. Ein Vorteil von Linux besteht darin, dass man sein RAID-System auch über Platten unterschiedlicher Größe verteilen kann, solange die Partitionen gleich groß sind. Das im Test eingesetzte RAID belegte 343 525 896 Blocks auf jeder Festplatte, was in einem Dateisystem von etwas weniger als 1 TByte resultiert. Das verwendete Mainboard war ein Gigabyte P35.

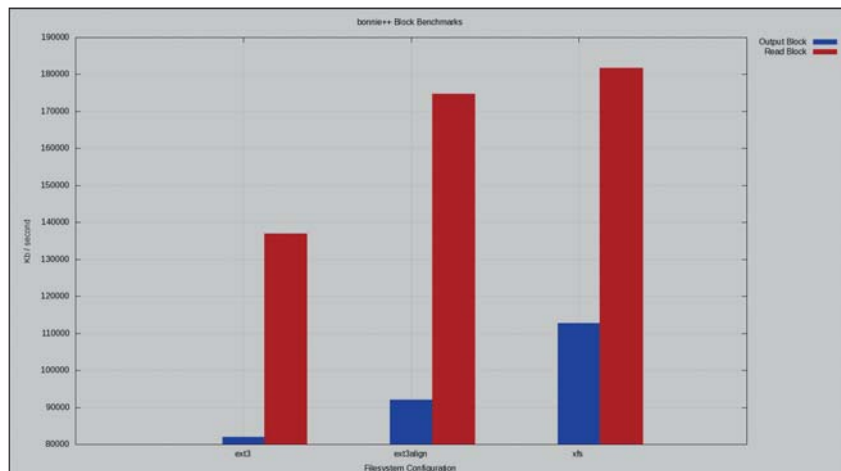


Abbildung 1: Bonnie-Benchmark-Ergebnisse für Lesen und Schreiben. Die Performances zwischen Standard-Ext3 und für RAID optimiertem Ext3 sind bereits enorm.

Volume anlegt, optimiert »mkfs.xfs« die Parameter fürs Anlegen des Dateisystems nicht.

Zur Tat

Der Befehl in Listing 1 legt ein RAID-System an. Die Default-Chunk-Größe ist dabei 64 KByte in einem Verbund von vier Festplatten. Wenn man die Festplatten/Partitions-IDs in »/dev/disk/by-id« verwendet, läuft man weniger Gefahr, aus Versehen die falsche Festplatte einzubinden.

Wir haben für Ext3 und XFS jeweils zwei Tests durchgeführt: einen mit einem Dateisystem, das mit »mkfs« und Standardeinstellungen angelegt wurde, den anderen mit auf das RAID abgestimmten Parametern. Es werden also vier Dateisysteme verglichen, nämlich Ext3-Standard (Ext3), Ext3-Ausgerichtet (Ext3-Aligned), XFS-Standard (XFS) und XFS-Ausgerichtet (XFS-Aligned). Die einzige Zusatzoption beim Standard-XFS ist »lazy-count=1«, das bei hoher Last den konkurrierenden Zugriff auf den Superblock verbessert und als Standard-Option empfehlenswert ist. Mehr Informationen zur Test-Hardware finden Sie im Kasten „Details“.

Der Aufruf von »mkfs« ist in Listing 2 zu sehen. Um die Funktion der Parameter zu verdeutlichen, sind die Werte in aussagekräftigen Variablen gespeichert. Der Stripe-Parameter teilt Ext3 mit, wie groß jeder RAID-Chunk in 4 KByte großen Disk-Blocks ist. Der Stripe-Width-Parameter legt fest, wie groß ein einzelnes

Stripe in Daten-Blocks ist. Effektiv entspricht es bei der RAID-5-Konfiguration mit vier Platten dreimal dem Wert von Stripe.

Den entsprechenden Aufruf für XFS zeigt das folgende Listing. Die Parameter »sunit« und »swidth« ähneln »stride« und »stripe-width« bei Ext3, allerdings werden die Werte in 512-Byte-Blocks angegeben:

```
mkfs.xfs -f -l lazy-count=1 -d sunit=2
$((($CHUNK_SZ_KB*2)),swidth=$((($CHUNK_SZ_KB*2)
*$NON_PARITY_DRIVE_COUNT)) $RAID_DEVICE
```

Für die Benchmarks kamen Bonnie++ und Iozone zum Einsatz. Bonnie++ ist als Dateisystem-Benchmark recht bekannt und testet zeichen- sowie blockweises Lesen und Schreiben, Rewrite, Seek und Metadatei-Operationen wie

Listing 1: RAID anlegen

```
01 cd /dev/disk/by-id
02 mdadm --create --auto=md --verbose --chunk=64 \
03 --level=5 --raid-devices=4 /dev/md/md-alignment-test
\
04 ata-SAMSUNG_HD501LJ_S0MUxxx-part4 ata-SAMSUNG_
HD501LJ_S0MUxxx-part4 \
05 ata-SAMSUNG_HD501LJ_S0MUxxx-part4 ata-SAMSUNG_
HD642JJ_S1AFxxx-part7
```

Listing 2: Ext3 optimiert

```
01 export RAID_DEVICE=/dev/md/md-alignment-test
02 export CHUNK_SZ_KB=64
03 export PARITY_DRIVE_COUNT=1
04 export NON_PARITY_DRIVE_COUNT=3
05
06 mkfs.ext3 -E stripe-width=$((NON_PARITY_DRIVE_
COUNT*CHUNK_SZ_KB/4)),\
07 stride=$((CHUNK_SZ_KB/4)) $RAID_DEVICE
```

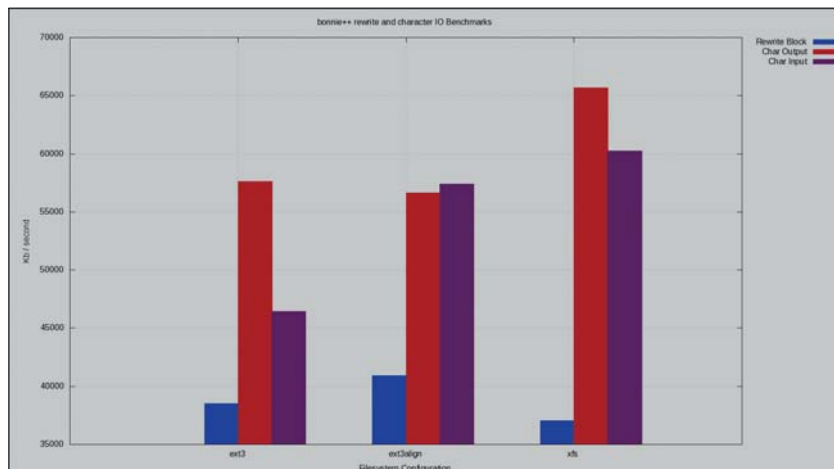


Abbildung 2: Bonnie-Ergebnisse für Rewrite und zeichenweises Lesen und Schreiben. Auch hier sind bei korrekter Ausrichtung des Ext3-Systems Performance-Gewinne zu verzeichnen.

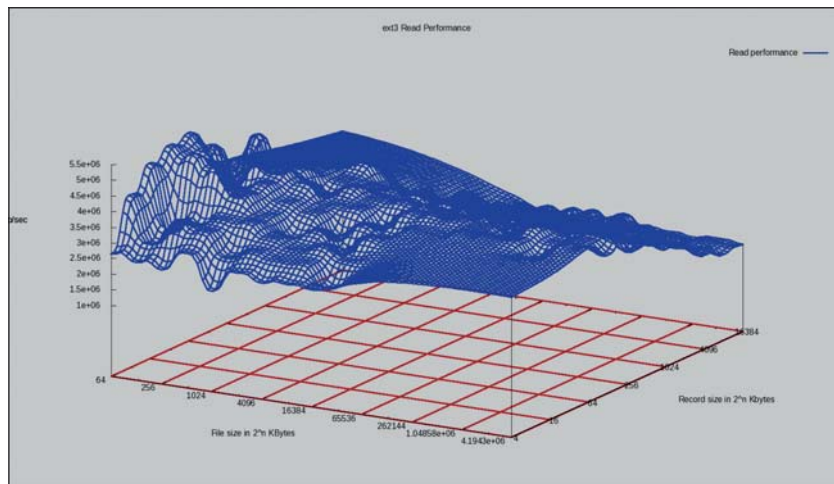


Abbildung 3: Izone-Ergebnisse fürs Lesen mit Ext3.

Anlegen und Löschen. Izone führt eine ganze Reihe an Lese- und Schreibtests mit unterschiedlichen Größen durch. Das Ergebnis lässt sich abhängig von den Größen in einem dreidimensionalen Graphen darstellen.

Abbildung 1 zeigt das erste Bonnie-Ergebnis beim Lesen und Schreiben. Weil »mkfs.xfs« wie erwähnt die RAID-Konfiguration erkennt und die Parameter anpasst, unterscheiden sich die Werte von XFS und XFS-Aligned nicht, deshalb ist im Diagramm nur XFS zu sehen. Bei Ext3 ist bereits ein sehr großer Unterschied zwischen Standard- und optimiertem Dateisystem zu erkennen. Auf eine um 10 MByte/s verbesserte Schreibperformance wird man nur ungern verzichten wollen!

Die von Bonnie gemessene Rewrite-Leistung ist ein aussagekräftiger Wert, wenn man beispielsweise auf einem Dateisys-

tem eine Datenbank betreiben möchte. Zwar ersetzen auch typische Desktop-Anwendungen wie das Speichern einer Datei, eine komplette Datei in einer Ope-

ration, aber Datenbanken überschreiben häufig Daten an ihrem altem Speicherort.

Die Rewrite-Leistung ist in **Abbildung 2** zu sehen. Auch wenn zeichenbasierte Ein-/Ausgabe weniger wichtig ist als blockbasierte, beläuft sich der Gewinn bei einem optimal ausgerichteten Ext3-Dateisystem auf einen Unterschied von 2 MByte/s, also etwa 6 Prozent mehr Leistung.

Die Izone-Ergebnisse für Ext3 und Ext3-Aligned sind sehr ähnlich, zeigen aber einige Anomalien, auf die es wert ist hinzuweisen. Die Lese-Performance ist in den **Abbildungen 3 und 4** zu sehen. Man sieht, dass Ext3-Aligned beim Lesen kleiner Datenmengen schneller ist als Ext3. Eigenartigerweise fällt die Performance aber mit Ext3-Aligned bei Dateien größer 16 MByte ab.

LVM

Wie erwähnt kann XFS selbst die optimalen Parameter wählen, wenn es sich um ein Software-RAID ohne LVM handelt. Bei einem Hardware-RAID oder der Verwendung von LVM muss der Administrator die besten Werte selbst herausfinden und beim Anlegen des Dateisystems angeben.

Wir haben zum Test ein manuell ausgerichtetes XFS und ein XFS mit Standard-Parametern angelegt, um zu untersuchen, welche Auswirkungen das in der Praxis hat. Viele Admins verwenden LVM über einem RAID, um in den Genuss von mehr Flexibilität und einfacherer Verwaltung

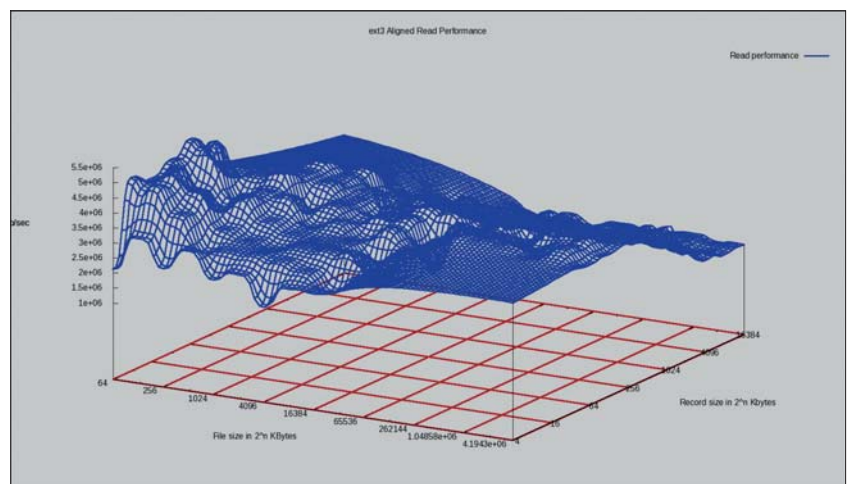


Abbildung 4: Izone-Ergebnisse fürs Lesen mit dem ausgerichtetem Ext3. Bemerkenswert ist die Mulde um eine Dateigröße von 16 386 Bytes herum.

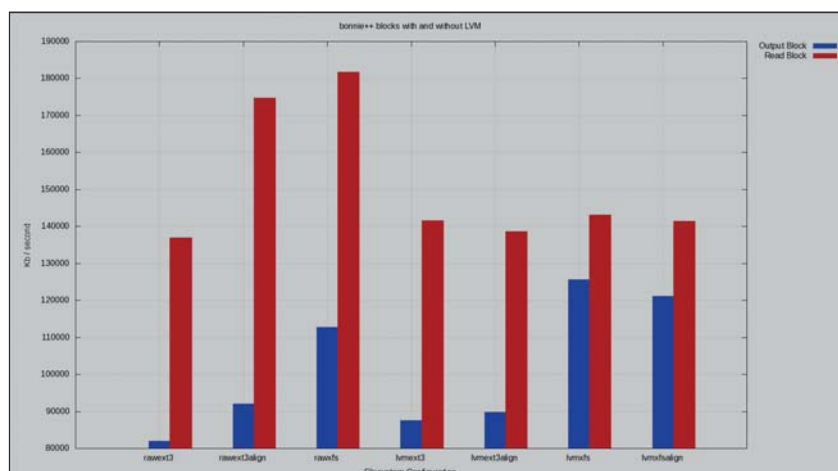


Abbildung 5: Bonnie-Ergebnisse für Lesen und Schreiben, mit und ohne LVM.

zu kommen. Deshalb haben wir die Ext3-Konfigurationen zusammen mit LVM getestet, um zu sehen, wie viel Performance die LVM-Schicht kostet.

Abbildung 5 zeigt die Ergebnisse für Ext3 und XFS auf demselben RAID mit und ohne LVM. Bei der mit „raw“ benannten Konfiguration wurde das Dateisystem direkt auf dem RAID angelegt. Bei der Konfiguration mit dem Namen „lvm“ wurde ein physisches LVM-Volume auf dem RAID angelegt und darin noch einmal ein logisches LVM-Volume, auf dem sich schließlich das Dateisystem befindet.

Man sieht, dass sich der Lese-Performance-Gewinn bei der Ausrichtung von Ext3 in Luft aufgelöst hat und die Lese-Performance über alle Dateisysteme hinweg identisch ist. Das Stripe auf dem XFS-Dateisystem auszurichten, hat die Schreib-Performance sogar noch verschlechtert. LVM scheint die Block-Lese-Geschwindigkeit zu begrenzen, sodass alle Versuche durch das Ausrichten der Blocks keinen Vorteil bringen.

Interessanterweise hat LVM bei XFS zu einer Steigerung der Schreibleistung geführt. XFS versucht das Schreiben so lange wie möglich zu verzögern (delayed write). Vielleicht ist das der Grund, warum XFS beim Schreiben mit LVM schneller ist. Wenn LVM einen eigenen Cache pflegt, können Daten möglicherweise effizienter geschrieben werden, wenn sie in großen Blöcken vom Dateisystem kommen. Auf der anderen Seite bricht die Lesegeschwindigkeit mit LVM dramatisch ein.

In Abbildung 6 sind die Ergebnisse des Bonnie-Rewrite-Benchmarks zu sehen.

Das Diagramm zeigt, dass die Leistungssteigerung durch das Ausrichten des Ext3-Systems zum Teil erhalten blieben. Das Ext3-Aligned ist auch mit LVM immerhin noch fast 1 MByte/s schneller

als das Standard-Ext3. Der Unterschied zwischen den beiden ist mit LVM jedoch kleiner als ohne.

Auch wenn ein Unterschied von 1 MByte/s wenig erscheinen mag, macht er zwischen den Extremen Ext3 auf LVM und Ext3-Aligned direkt auf dem RAID dennoch 15 Prozent aus! Diese Leistungssteigerung kommt auf der gleichen Hardware ohne zusätzliche Kosten zustande.

Weil es schwierig ist, zwei dreidimensionale Diagramme nur per Augenschein zu vergleichen, haben wir im Weiteren den Unterschied zwischen den ausgerichteten und den Standard-Dateisystemen visualisiert. In Abbildung 7 ist die mit Iozone gemessene zufällige Lese-Performance von Ext3-Aligned und Ext3 zu sehen. Es ergibt sich kein komplett eindeutiges Bild, aber in einzelnen Fällen ist Ext3-Aligned schneller. Abbildung 8 visualisiert den Unterschied in

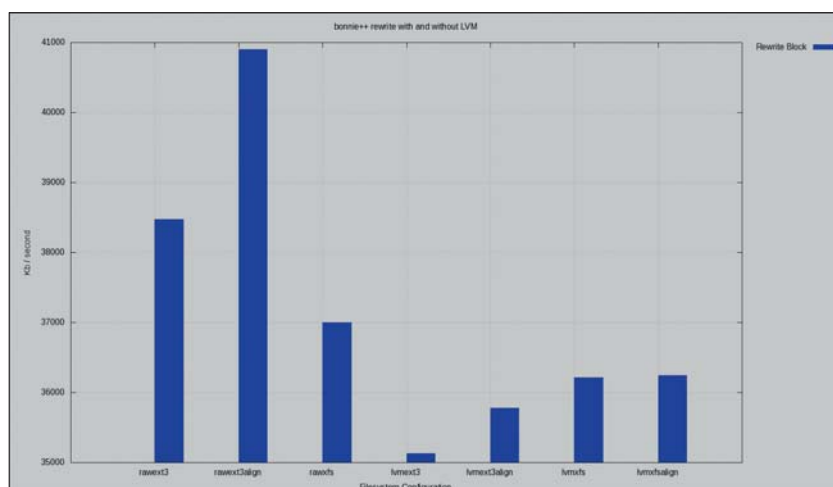


Abbildung 6: Bonnie-Ergebnisse für Rewrite mit und ohne LVM. Das Dateisystem direkt auf dem RAID ist der LVM-Version wieder überlegen.

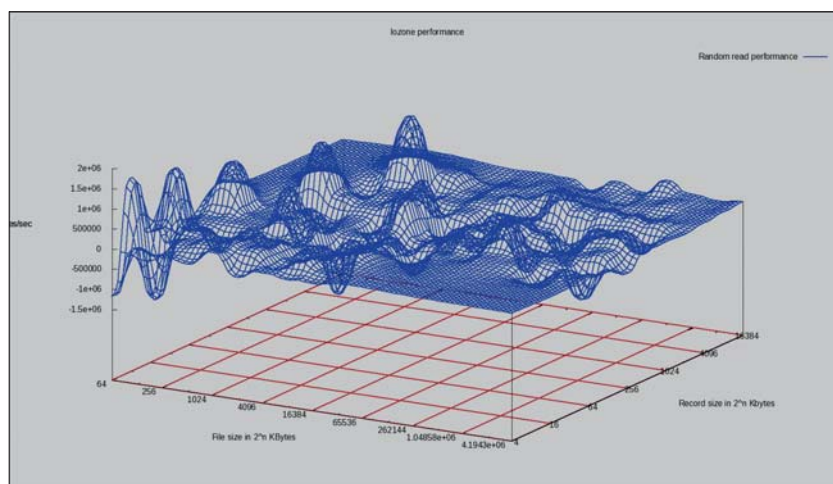


Abbildung 7: Zufälliges Lesen von Ext3 mit und ohne Ausrichtung auf LVM.

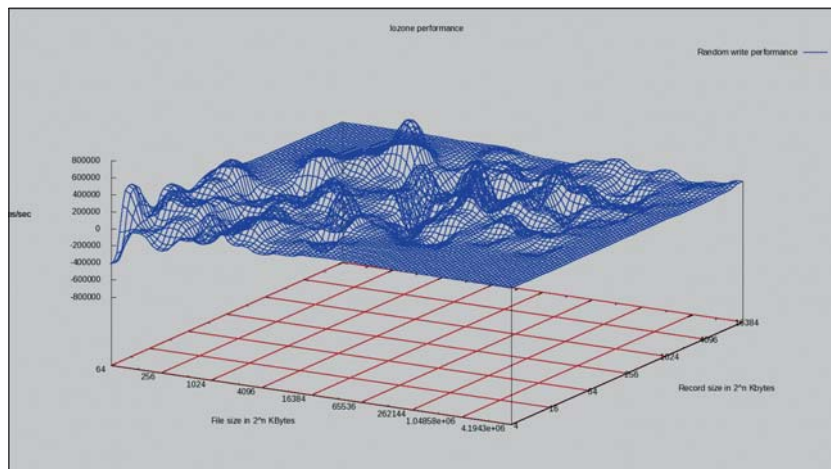


Abbildung 8: Unterschiede beim I/Ozone-Benchmark für zufälliges Schreiben mit Ext3.

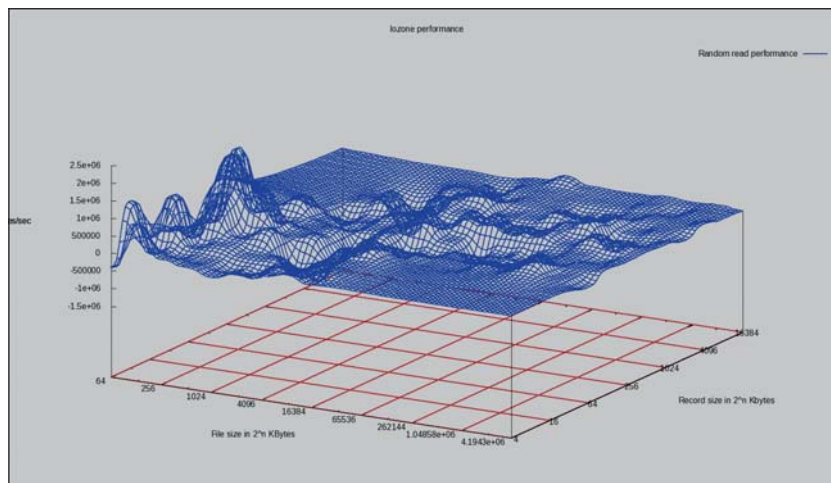


Abbildung 9: Zufälliges Lesen vom XFS-Dateisystem mit und ohne Ausrichtung auf LVM.

der zufälligen Schreib-Performance von Ext3 auf LVM in der Standard- und der ausgerichteten Variante. Die Senken und Erhebungen befinden sich an den gleichen Stellen, sind aber nicht ganz so ausgeprägt wie bei der Lese-Performance. Insgesamt zeigt das Ext3-Aligned weniger Leistungseinbrüche als beim Lesen.

Schließlich ist in **Abbildung 9** noch der Vergleich von XFS auf LVM mit und ohne

Alignment beim zufälligen Lesen zu sehen. Verglichen mit Ext3 ist der Graph relativ flach. Zufällige Leseoperationen kleiner Dateien sind auf einem ausgerichteten XFS viel schneller, wie die Gipfel auf der linken Seite des Diagramms zeigen. Bei kleineren Schreiboperationen auf 4 MByte große Dateien ist das ausgerichtete Dateisystem etwas langsamer als eines mit Standard-Optionen. Bei größeren Schreiboperationen auf Dateien zwischen

Tiefer graben

Seit Version 2.6.28 unterstützt der Linux-Kernel das Ext4-Dateisystem, mittlerweile hat es auch in den meisten Distributionen Einzug gehalten. Gegenüber seinem Vorgänger Ext3 bringt es vor allem bessere Performance und den Support größerer Partitionen.

Ein Nachteil davon, die RAID-Stripe-Größe beim Anlegen eines Dateisystems zu berücksichtigen, ist dass nachträgliche Änderungen am RAID zu Unstimmigkeiten führen.

Wenn Sie auf einmal vier statt drei Daten-Disks in Ihrem RAID-5 haben, ist die Stripe-Ausrichtung des Dateisystems nicht mehr korrekt. Ext4 besitzt aber eine Option, die es erlaubt, die Stripe-Größe beim Mounten anzugeben. Wenn Sie also dem RAID eine neue Platte hinzufügen oder ein LVM-Volume erweitern, teilen Sie das dem Ext4-System einfach mit, und es kann seine Operationen in Zukunft darauf abstimmen.

16 und 256 MByte liegt das ausgerichtete XFS wieder vorne.

Fazit

Wie dieser Artikel zeigt, lassen sich zwischen 10 und 15 Prozent an Performance gewinnen, wenn man beim Anlegen eines Ext3-Dateisystems auf die Ausrichtung an RAID-Chunk- und -Stripe-Größe achtet. Auch wenn die Unterschiede bei XFS nicht so groß sind, kann die Ausrichtung beim Anlegen und Löschen großer Dateien einen Unterschied machen. Generell lassen sich folgende Faustregeln angeben:

- Wer Ext3 oder Ext4 verwendet, sollte mit »-E stripe-width = x, stride = y« immer dem Filesystem die RAID-Konfiguration mitteilen.
- Wenn man ein XFS-Dateisystem direkt auf einem RAID anlegt, sollte das System selbst die richtigen Parameter verwenden.
- Wer XFS in einer LVM-Partition auf einem RAID verwendet, sollte mit »-d sunit = x, swidth = y« an »mkfs.xfs« die passenden Werte übergeben.
- Bei anderen Dateien sollten Sie überprüfen, ob »mkfs« die richtigen Werte für das konfigurierte RAID ermitteln konnte.

Der mit diesen Tipps erzielte Leistungsgewinn wird je nach Konfiguration anders ausfallen, aber unabhängig von den konkreten Zahlen, ist es eine gute Strategie, das Dateisystem nach den RAID-Parametern auszurichten, vor allem, wenn man Ext3 oder Ext4 verwendet. (ofr) ■

Infos

- [1] OSDL Linux RAID: [http://linux-raid.osdl.org/index.php/Linux_Raid]
- [2] David A. Patterson, Garth Gibson, and Randy H. Katz, A Case For Redundant Arrays of Inexpensive Disks (RAID): [<http://www-2.cs.cmu.edu/~garth/RAIDpaper/Patterson88.pdf>]

Der Autor

Ben Martin arbeitet seit mehr als zehn Jahren an Dateisystemen. Er bietet Consulting-Leistungen zu Libferris, Dateisystemen und Qt/C++-Entwicklung an. Er schreibt genauso gerne über Open Source, wie er Code beisteuert.

Zugriff auf virtuelle Disk-Images mit Libguestfs

Gastfreundlich

Die Libvirt vereinheitlicht den Zugriff auf virtuelle Maschinen unterschiedlicher Couleur zur Laufzeit. Mit der Libguestfs können Administratoren und Programmierer die dazu gehörigen Disk-Images bearbeiten. Richard W.M. Jones

In der Vergangenheit hatten Administratoren es mit virtuellen Disk-Images nicht leicht. Ganz am Anfang mussten sie sogar noch die Images für virtuelle Maschinen von Hand anlegen, aber diese Zeiten sind unter anderem dank des Storage-Management-Codes der Libvirt vorbei. Wer aber unabhängig von einer laufenden virtuellen Maschine ein Disk-Image bearbeiten wollte, musste sich immer noch mit Tools wie Kpartx begnügen. Das erforderte Root-Rechte und barg somit sogar die Gefahr für Root-Exploits durch manipulierte Disk-Images.

Ich fand die Situation völlig inakzeptabel. Schließlich sind Disk-Images auch nur Dateien, und es ist nicht einzusehen, wieso man dafür Root-Rechte brauchen sollte. Man muss auch nicht Superuser werden, wenn man ein Textdokument öffnet oder spezielle Device-Nodes anlegen, um mit Gimp zu arbeiten.

Problem erkannt

Die Libguestfs [1] löst nun alle diese Probleme. Es ist eine skriptbare, sichere Bibliothek, über die ein Anwender Disk-

Images unterschiedlicher virtueller Maschinen ohne Root-Rechte bearbeiten kann. Das Paket bringt eine ganze Reihe nützlicher Kommandozeilentools mit und sogar eine interaktive Shell, um innerhalb des Disk-Image zu navigieren.

Anwender von Fedora, Red Hat Enterprise Linux und Cent OS haben es leicht. Fedora-User beispielsweise installieren die Bibliothek und die Tools mit

```
yum install \*guestf\*
```

Wer RHEL oder Cent OS einsetzt, muss vorher noch das EPEL-Repository einbinden und kann dann Libguestfs mit dem



Tabelle 1: Libguestfs-Tools

»virt-cat«	»virt-install«	»virt-rescue«
»virt-clone«	»virt-list-filesystems«	»virt-resize«
»virt-convert«	»virt-list-partitions«	»virt-tar«
»virt-df«	»virt-ls«	»virt-top«
»virt-edit«	»virt-make-fs«	»virt-viewer«
»virt-image«	»virt-manager«	»virt-win-reg«
»virt-inspector«	»virt-pki-validate«	»virt-xml-validate«

gleichen Befehl installieren. Debian- und Ubuntu-User finden einen passenden Link in den Libguestfs-FAQ [2]. Hilfe versprechen ein freundlicher IRC-Channel namens #libguestfs im Freenode-Netz und eine Mailing-Liste, die Sie über die Website abonnieren können.

Libguestfs kann von Hardware-Beschleunigung Gebrauch machen, dazu muss das Device »/dev/kvm« Lese- und Schreibrechte für den jeweiligen Benutzer besitzen. Wenn die Tools installiert sind, verschaffen Sie sich einen Überblick, indem Sie in der Shell »virt-« eingeben und zweimal die Tab-Taste drücken. **Tabelle 1** listet die wichtigsten Befehle auf.

Für jedes Tool gibt es eine eigene Manpage (zum Beispiel »man virt-df«). Guestfish erlaubt den vollständigen Zugriff auf die Libguestfs-API zum Beispiel für Shellskripts, mit Guestmount können Sie ein Disk-Image in einem Verzeichnis mounten (**Abbildung 1**). Die API ist außerdem für die Programmiersprachen C, C++, Perl, Python, Ruby, Ocaml, Java, Haskell und C# (Mono) verfügbar.

Wie der Name schon andeutet, stellt »virt-df« das Pendant zum gewohnten »df«-Kommando dar. Ohne weitere Optionen aufgerufen zeigt es den freien und belegten Speicherplatz jeder virtuellen Disk an, die von Libvirt verwaltet wird.

Warnung

Verwenden Sie niemals Libguestfs-Tools im Lese-/Schreibmodus mit Images einer laufenden virtuellen Maschine. Das würde höchstwahrscheinlich zu einem unbrauchbaren Image führen. Meistens überprüfen die Tools das selbst, aber es gibt Fälle, in denen Sie es nicht merken. Die meisten Programme kennen die Option »--ro«, um Änderungen zu verhindern. Damit lassen sich die Tools gefahrlos selbst mit laufenden VMs verwenden. Manche Tools lesen von Haus aus nur, und kennen deshalb diese Option nicht. Werfen Sie im Zweifelsfall einen Blick in die Manpage.

Beachten Sie, dass per Default nur Root die Images lesen darf, die beispielsweise in »/var/lib/libvirt/images« liegen. Damit Sie als gewöhnlicher User die Images lesen dürfen, müssen Sie die Rechte also entsprechend anpassen. Das Beispiel einer Ausgabe von »virt-df« zeigt **Listing 1**. Alternativ rufen Sie virt-df mit einem Disk-Image als Parameter auf (**Listing 2**).

Virt-df eignet sich hervorragend dafür, kontinuierlich den Speicherplatz zu überprüfen und rechtzeitig neue Kapazitäten einzuplanen. So können Sie es als Cronjob ausführen und die kommaseparierte Ausgabe (»--csv«) beispielsweise in Datenbanken oder eine Tabellenkalkulation importieren.

Die Größe eines Disk-Image lässt sich mit »virt-resize« ändern (**Listing 3**). Das Tool verändert nicht direkt das Disk-Image, sondern legt ein neues an und kopiert die bestehenden Inhalte dorthin. Damit diese konsistent sind, sollte man keine Images mit einer gerade laufenden Maschine verändern. Fahren Sie diese also vorher herunter.

Wenn es beispielsweise ein Problem mit dem Boot-Manager einer virtuellen Maschine gibt, lässt sich mit »virt-edit« direkt eine Datei auf dem Image editieren, wie etwa die Grub-Konfigurationsdatei »menu.lst«:

```
virt-edit Debian5x32 /boot/grub/menu.lst
```

Alternativ dazu können Sie eine Reparatur mit »virt-rescue« versuchen, das ähnlich wie eine Rescue-CD eine Rettungs-Shell startet. Versuchen Sie auch das besser nicht mit einer laufenden VM (siehe **Kasten „Warnung“**).

Um Dateien eines Image anzuzeigen, gibt es den Befehl »virt-cat«. Damit können Sie beispielsweise die Logdateien einer virtuellen Maschine anzeigen oder nach verdächtigen Spuren suchen. Das Skript

Listing 1: »virt-df«

```
01 # virt-df
02 Filesystem                                1K-blocks
    Used Available Use%
03 Debian5x32:/dev/debian5x32.home.annexia.org/home
04                                         2027920
    35844    1889064    2%
05 Debian5x32:/dev/debian5x32.home.annexia.org/root
06                                         329233
    81728    230507    25%
07 Debian5x32:/dev/debian5x32.home.annexia.org/tmp
08                                         170549
    5663     156080    4%
09 Debian5x32:/dev/debian5x32.home.annexia.org/usr
10                                         2100444
    298988    1694756    15%
11 Debian5x32:/dev/debian5x32.home.annexia.org/var
12                                         955480
    202308    704636    22%
13 Debian5x32:/dev/vda1                      233335
    9546     211341    5%
14 Windows7x32:/dev/vda1                    102396
    24704     77692    25%
15 Windows7x32:/dev/vda2                     8284156
    7229712    1054444    88%
```

in **Listing 4** verwendet »virt-cat«, um nach Backdoor-Accounts in einem Linux-Gastsystem zu suchen.

Auch wenn sich viele Administrationsaufgaben mit den Kommandozeilentools des Pakets erledigen lassen, erschließt sich der volle Leistungsumfang der Libguestfs nur durch Guestfish, die interaktive Shell. Guestfish kennt intern beinahe 300 Befehle, von denen dieser Artikel nur die allerwichtigsten kurz erklärt.

Listing 2: »virt-df« mit Disk-Image

```
01 $ virt-df -h ~/disk.img
02 Filesystem                                Size
    Used Available Use%
03 /home/rjones/disk.img:/dev/vda1          193.7M
    21.6M    162.1M    12%
04 /home/rjones/disk.img:/dev/vg_f12x32/lv_root
05                                         5.2G
    2.3G     2.6G    45%
```

Listing 3: »virt-resize«

```
01 $ truncate -s 10G ~/enlarged.img
02 $ virt-resize ~/disk.img ~/enlarged.img --expand /dev/sda2
03 Summary of changes:
04 /dev/sda1: partition will be left alone
05 /dev/sda2: partition will be resized from 5.8G to 9.8G
06 /dev/sda2: content will be expanded using the 'pvresize' method
```

Listing 4: Backdoor-Suche

```

01 #!/bin/sh -
02 # Get list of guests from libvirt.
03 guests=$(
04   virsh list --all | tail -n+3 | head -n-1 |
05   awk '{print $2}'
06 )
07 for n in $guests; do
08   virt-cat $n /etc/passwd |
09   awk -F: '$1 != "root" && $3 == 0 {
10     print "BACKDOOR ACCOUNT FOUND:", $1
11   }'
12 done

```

Listing 5: Datei-Upload ins Gast-System

```

01 ><fs> mount /dev/sda1 /
02 ><fs> upload foo.rb /foo.rb
03 ><fs> ll /
04 total 18
05 drwxr-xr-x 3 root root 1024 Jan 19 11:59 .
06 drwxr-xr-x 20 root root 4096 Jan 19 11:56 ..
07 -rw-r--r-- 1 root root 1014 Jan 19 11:59 foo.rb
08 drwx----- 2 root root 12288 Jan 19 11:57 lost+found

```

Listing 6: Automatik mit »-i«

```

01 $ guestfish --ro -i Debian5x32
02
03 ><fs> less /boot/grub/menu.lst
04 # menu.lst - See: grub(8), info grub, update-grub(8)
05 #
06 # grub-install(8), grub-floppy(8),
07 # grub-md5-crypt, /usr/share/doc/grub
08 # and /usr/share/doc/grub-legacy-doc/.
09 ...

```

Listing 7: Cloning VMs

```

01 #!/bin/sh -
02
03 template="$1"
04 newimage="$2"
05 nameserver="$3"
06 hostname="$4"
07
08 dd if="$template" of="$newimage" bs=1M
09
10 echo > /tmp/network <<EOF
11 NETWORKING=yes
12 HOSTNAME=$hostname
13 EOF
14
15 guestfish -i "$newimage" <<EOF
16 write-file /etc/resolv.conf "nameserver $nameserver"
17   0
18 upload /tmp/network /etc/sysconfig/network
19 sync
20 EOF
21 rm /tmp/network

```

Sie können Guestfish ebenfalls mit einem existierenden Disk-Image starten oder direkt ein neues anlegen. An Formaten beherrscht Guestfish das Rohformat von Festplatten (Raw), CD-ISOs, virtuelle Floppy Disks (VFD), komprimierte Formate wie Qcow2, SD-Karten und sogar Dateisysteme. Guestfish kann sie alle lesen und die meisten sogar schreiben. Zum Starten des Programms geben Sie in der Shell »guestfish« ein. Ein 100 MByte großes Sparse-Image legen Sie dann folgendermaßen an:

```

><fs> sparse test.img 100M
><fs> run

```

Das dauert ein paar Sekunden. Partitionieren Sie dann die Disk und legen Sie ein Dateisystem an:

```

><fs> part-disk /dev/sda mbr
><fs> mkfs ext2 /dev/sda1

```

Sie können die Partition dann mounten und beispielsweise eine Datei aus dem Host-System in die neue Gast-Partition hochladen (Listing 5). Umgekehrt geben Sie mit dem folgenden Befehl den Inhalt der Datei aus dem Disk-Image aus:

```

guestfish --ro -a test.img -m /dev/sda1 2
cat /foo.rb

```

Der Schalter »-a« fügt das Image hinzu, und »-m« teilt Guestfish mit, wo es das Dateisystem findet. Von außen betrachtet ist ein Disk-Image nur eine Datei, das Mounten eines Dateisystems gehorcht aber den Konventionen des jeweiligen Betriebssystems. So kann »/dev/sda1« auf »/boot« eingebunden werden, »/dev/vg_lv_var« auf »/var« oder »/dev/sda2« als Laufwerk »C:« unter Windows. Woher soll man nun als Anwender wissen, wie man eine Partition einbindet? Libguestfs bringt ein Tool namens »virt-inspector« mit, das mit einigen Regeln und Heuristiken die Zuweisung herausfindet. Über den Schalter »-i« bietet Guestfish eine ähnliche Funktion (Listing 6).

Um beispielsweise herauszufinden, wie viel Platz die Logdateien einnehmen, geben Sie »du /var/log« ein. Mit »grep ^root: /etc/password« listen Sie den

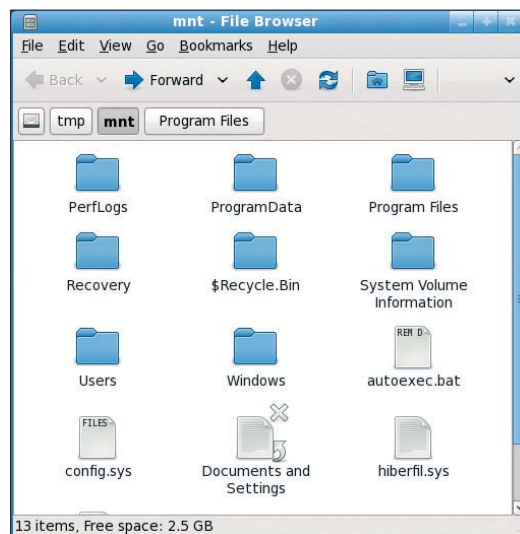


Abbildung 1: Mit dem Befehl »guestmount« kann der Administrator ein virtuelles Image direkt mounten. Hier ein Windows-Dateisystem im Nautilus-Dateibrowser.

Root-Account des Systems auf. Auch Pfad- und Dateivervollständigung per Tab-Taste funktioniert in Guestfish wie von der Bash-Shell gewohnt. Partitionen und logische Volumes listen die folgenden Befehle auf:

```

><fs> list-partitions
/dev/vda1
/dev/vda2
><fs> lvs
/dev/vg_rhel6/lv_root
/dev/vg_rhel6/lv_swap

```

Ihr Typ lässt sich dann mit »vfs-type« und »file« abfragen:

```

><fs> vfs-type /dev/vg_rhel6/lv_root
ext4
><fs> file /dev/vg_rhel6/lv_root
Linux rev 1.0 ext4 filesystem data (needs 2
journal recovery) (extents) (large files) 2
(huge files)

```

Mit den Guestfish-Befehlen »tgz-out« und »tgz-in« packen Sie Dateien aus dem Gast-System in ein Tar-Paket auf dem Host-System oder umgekehrt. Mit dem Augeas-Konfigurations-Editor [3] können Sie die Konfiguration verschiedener Linux-Systeme bearbeiten, zum Beispiel die Apt-Konfigurationsdateien einer Debian-Distribution:

```

><fs> aug-init / 0
><fs> aug-get /files/etc/apt/sources.list/1/2
uri
http://ftp.uk.debian.org/debian/
><fs> aug-get /files/etc/apt/sources.list/1/
distribution
lenny

```

```

rjones@trick:~$ virt-win-reg /dev/vg_trick/Windows7x32 HKLM\SOFTWARE\MICROSOFT\
\WINDOWS NT\CurrentVersion | head
[KEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion]
"BuildGUID"=hex(1):65,00,33,00,33,00,31,00,63,00,65,00,32,00,34,00,2d,00,33,00,3
7,00,37,00,61,00,2d,00,34,00,37,00,62,00,64,00,2d,00,38,00,36,00,64,00,65,00,2d,
00,39,00,32,00,61,00,65,00,31,00,61,00,61,00,31,00,61,00,65,00,36,00,35,00,00,00
"BuildLab"=hex(1):37,00,36,00,30,00,30,00,2e,00,77,00,69,00,6e,00,37,00,5f,00,72
,00,74,00,6d,00,2e,00,30,00,39,00,30,00,37,00,31,00,33,00,2d,00,31,00,32,00,35,0
0,35,00,00,00
"BuildLabEx"=hex(1):37,00,36,00,30,00,30,00,2e,00,31,00,36,00,33,00,38,00,35,00,
2e,00,78,00,38,00,36,00,66,00,72,00,65,00,2e,00,77,00,69,00,6e,00,37,00,5f,00,72
,00,74,00,6d,00,2e,00,30,00,39,00,30,00,37,00,31,00,33,00,2d,00,31,00,32,00,35,0

```

Abbildung 2: Zugriff auf die Windows-Registry mit »virt-win-reg«.

Das hier verwendete Präfix »/files« wird von Augeas verwendet, um Konfigurationsdateien zu kennzeichnen, es hat nichts mit der Libguestfs zu tun.

Kein Stiefkind

Libguestfs bringt auch guten Support für Windows mit. Genauso wie Linux-Gäste lassen sich Windows-Gäste mounten und untersuchen.

```

$ guestfish --ro -i Windows7x32
><fs> ls /Windows/System32/drivers | head -2
1394bus.sys
1394ohci.sys

```

Abbildung 2 zeigt, wie sich mit dem Tool »virt-win-reg« sogar Registry-Einträge bearbeiten lassen.

Sein volles Potenzial entfaltet Guestfish, wenn man es in eigenen Skripten einsetzt. So klonst das Skript in Listing 7 eine virtuelle Maschine. Die folgenden Zeilen verwenden das Skript, um von einem Template eine neue VM zu klonen und dann die Konfiguration zu verändern:

```

# cd /var/lib/libvirt/images
# /tmp/clone.sh oldguest newguest 192.168.1.12
newguest.example.com
# virt-install --import --file newguest

```

Auf diese Art kann jeder Administrator auch ohne GUI recht unkompliziert Management-Skripts für virtuelle Umgebungen schreiben.

Die Zukunft

Die aktuelle Libguestfs-API ist umfassend, ausgereift und gut getestet. Während der Entwicklung hat das Team auf Rückwärtskompatibilität von API und ABI geachtet und nur neue Aufrufe hinzugefügt. Auf der Agenda für kommende Versionen stehen noch mehr bequeme

Administrations-Tools, zum Beispiel um schnell neue Gastsysteme bereitzustellen oder Disk-Images automatisch zu verkleinern. Am Windows-Support wollen die Entwickler genauso arbeiten wie am CIM-Support für Virt-Inspector und Integritäts-Checks für im Gast installierte Software. Auch Updates für »virt-p2v« und »virt-v2v« sind geplant.

Bessere Integration in Management-Tools soll für bequemere Anwendung sorgen. So ist es derzeit möglich, mit Fuse und Guestmount Gastsysteme zu mounten. In Zukunft können Administratoren dann im grafischen Virt-Manager direkt die Images der Gast-Systeme öffnen. Weitere Optionen für die Zukunft sind die Integration von Security-Tools, Rootkit- und Viren-Scannern und Disk-Monitoring-Tools. Mitarbeiter zum Paketieren für Ubuntu, Gentoo und Mac OS X sind herzlich willkommen. Allerdings ist es recht aufwendig, Libguestfs von Hand zu kompilieren, deshalb sollte man genügend Zeit dafür erübrigen können.

Eine Frage taucht immer wieder auf: Was ist mit einer GUI? Die Antwort darauf lautet erst einmal: Es gibt keine. Es ist schwer vorstellbar, wie eine GUI aussehen könnte, die den Funktionsumfang von Guestfish besitzt. Es wäre jedenfalls eine ziemlich komplexe und vermutlich unhandliche GUI. Falls sich aber doch jemand dazu entschließen sollte, kann er auf große Unterstützung durch die Libguestfs-Community rechnen. Einen ersten Vorschlag dazu habe ich meinem Blog gemacht. (ofr)

Infos

- [1] Libguestfs: [<http://libguestfs.org/>]
- [2] Libguestfs FAQ: [<http://libguestfs.org/FAQ.html>]
- [3] Augeas: [<http://augeas.net/>]



Joomla!™

Web Hosting

BEST-PRICE JOOMLA

Joomla fix & fertig installiert
Inklusive 2 .de Domains
800 MB Speicherplatz
100 E-Mail Adressen
1 MySQL Datenbank
2 FTP Konten
u.v.m.

3,95 €
monatlich

JOOMLA STANDARD

Joomla fix & fertig installiert
Inklusive 3 Domains
2000 MB Speicherplatz
500 E-Mail Adressen
3 MySQL Datenbanken
20 Subdomains
10 FTP Konten
1 Joomla Buch
u.v.m.

30% Rabatt
6,95 €
statt 9,95 € / Monat

UNSER SERVICE

Keine Einrichtungsgebühren
Keine versteckten Kosten
Traffic: Flatrate
150 fertige Joomla Designs
Joomla Update Service
Kostenloser Telefonsupport

www.provider4u.de

☎ 0800 - 11 44 1 44

Provider4u.de

Leistungsfähige und einfache Versionsverwaltung

Code im Blick

Seit Linus Torvalds höchstpersönlich das Versionskontrollsystem Git erfand, hat sich dieses unter immer mehr Open-Source-Projekten verbreitet. Dieser Artikel gibt eine Einführung für alle, die Software schreiben, Quellcode aus Repositories laden oder Konfigurationsdateien mit Git verwalten wollen. Juliet Kemp

Mit einer Versionsverwaltungs-Software

verfolgen Sie die Geschichte Ihres Projekts Byte für Byte, und machen jede beliebige Änderung rückgängig, falls Sie sich doch anders entscheiden. Außerdem lassen sich mehrere Versionen oder Zweige innerhalb eines Projekts anlegen und verfolgen.

Projekte in der Software-Entwicklung nutzen immer eine Form von Versionskontrolle, und immer mehr von ihnen verwenden Git [1]. Das Tool wurde ursprünglich von Linus Torvalds für das Entwicklerteam des Linux-Kernels geschrieben, aber wegen seiner Flexibilität, Schnelligkeit und verteilten Struktur verbreitete es sich bei sehr vielen Projekten sehr schnell. Dieser Artikel zeigt Ihnen, wie Sie Git für die Erfassung einzelner Schritte in Ihrem Projekt erfolgreich anwenden.

Git ist auf hochverteilte, sehr schnelle und sehr flexible Leistung ausgelegt. „Hochverteilt“ bedeutet in diesem Sinne, dass Git im Gegensatz zu anderen Versionsverwaltungssystemen, wie etwa CVS oder Subversion, eigentlich gar kein zentrales Repository benötigt. Stattdessen haben alle Repositories den gleichen Status, und

jedes Repository kann mit jedem anderen verglichen und aktualisiert werden. Das prädestiniert Git für Projekte mit vielen kooperierenden Teilnehmern [2].

Chaos im Griff

Eines der Hauptmerkmale von Git ist, dass es für die Unterstützung nicht linearer Entwicklungsarbeiten konzipiert wurde: Es geht davon aus, dass Änderungen immer wieder zusammengeführt werden, sobald sie von den verschiedenen Testern und Entwicklern eingereicht werden. Somit ist es ein Leichtes, Zweige zusammenzuführen oder sogar ganze Bäume, unabhängig davon, ob sie eine gemeinsame Vorversion haben. Im Gegensatz zu vielen anderen Versionskontrollsystemen ist es mit Git auch einfach, nicht versionierte Dateien in einen existierenden Baum einzubinden. Das macht es zum idealen Werkzeug für Projekte mit stark verteilter Entwicklung; aber wegen seiner beispiellosen Flexibilität auch für die heimische Verwendung.

Installieren Sie das Paket »git-core« unter Ubuntu/Debian und openSUSE/Fedora, um die Grundlagen zu erhalten; nützlich

ist auch das Paket »git-doc« mit der Dokumentation. Verschiedene andere Pakete sind noch als Erweiterung erreichbar, wie »git-svn«, das die Subversion-Interoperabilität ermöglicht und »gitweb«, das ein Web-Interface zur Verfügung stellt. Alternativ installieren Sie Git direkt aus dem Quellcode [3].

Haben Sie Git einmal installiert, erstellen Sie im Nu Ihr erstes Repository. Der dezentralisierte Charakter von Git zeigt sich darin, dass jede Arbeitskopie ihr eigenes Repository überall mit sich trägt (im »git«-Unterverzeichnis), anstatt wie bei CVS oder SVN ein zentrales Online-Repo zu pflegen.

Haben Sie ein bestehendes Verzeichnis, das Sie unter Versionskontrolle stellen wollen, ist der Vorgang äußerst einfach:

```
cd my_directory
git init
git add .
git commit
```

Haben Sie Ihr von Git kontrolliertes Verzeichnis eingerichtet (Abbildung 1), legen Sie nun in der Konsole über »touch Datei« eine neue Datei an und tippen »git status« ein. Es erscheint eine Warnung, dass sich

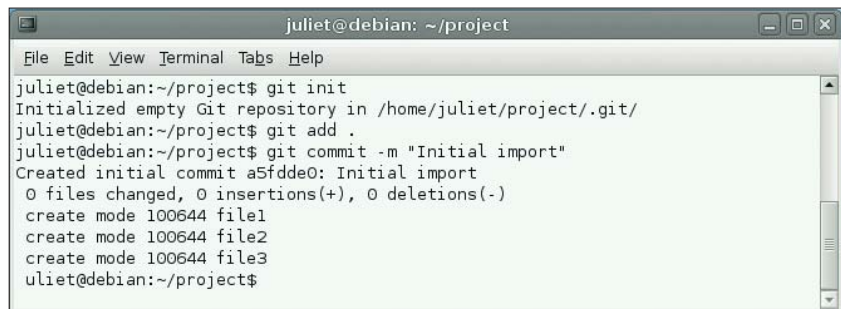
im Verzeichnis eine Datei befindet, deren Version noch nicht verfolgt wird. Sie fügen diese Datei mit »git add *Datei*« Git hinzu, und mit »git commit« übertragen Sie die Datei ins Repository. Daraufhin öffnet sich zuerst ein Texteditor, in dem Sie einen Kommentar zur gemachten Änderung hinterlegen müssen. Möchten Sie den Kommentar gleich mit übertragen, dann benutzen Sie dazu den Befehl »git commit -m "*Kommentar*"«. Sämtliche Anmerkungen speichert Git im Protokoll mit. Gewöhnen Sie es sich deshalb an, gute Kommentare zu schreiben, für den Fall, dass Sie später eine Änderung wieder rückgängig machen wollen. Mit folgenden zwei Befehlen fügen Sie alle bis jetzt noch nicht erfassten Dateien eines Verzeichnisses zu Git hinzu und übertragen Sie diese ins Repository:

```
git add .
git commit -m "Anmerkungen"
```

Bearbeiten Sie nun eine der vorhandenen Dateien, und geben Sie dann den Befehl »git status« ein. Git zeigt dann diese Datei als „Changed but not updated“ („verändert, aber nicht aktualisiert“) an. Im Gegensatz zu einigen anderen Systemen übernimmt Git keine Änderungen, bis Sie es dem Tool nicht ausdrücklich befehlen. Um alle bereits erfassten aber inzwischen geänderten Dateien dem Repository hinzuzufügen, benutzen Sie den Befehl »git commit -am "*Anmerkungen*"«; hierbei werden nur die bereits erfassten und verfolgten Dateien berücksichtigt, für die noch nicht erfassten Dateien müssen Sie immer »git add .« eingeben.

Ihre vorhin erstellte Test-Datei löschen Sie mit »git rm *Datei*« und dann »git commit«. Wurde die Datei aber seit der letzten Übergabe an das Repository geändert, erscheint eine Warnung. Um die Datei dennoch zu löschen, benutzen Sie »git rm -f *Datei*«.

Ob es in Ihrem Verzeichnis Dateien gibt, die von Git noch nicht erfasst wurden, erfahren Sie aus der Ausgabe von »git status«. Allerdings haben Sie möglicherweise einige Dateien in diesem Verzeichnis, die Sie nie hinzufügen möchten (Backup-Dateien oder temporäre Dateien). Normalerweise erscheinen diese auch in der Ausgabe von »git status«, da sie per »git add .« auch hinzugefügt wurden. Um dies zu vermeiden, erstellen Sie eine einfache



```
juliet@debian: ~/project
File Edit View Terminal Tabs Help
juliet@debian:~/project$ git init
Initialized empty Git repository in /home/juliet/project/.git/
juliet@debian:~/project$ git add .
juliet@debian:~/project$ git commit -m "Initial import"
Created initial commit a5fdde0: Initial import
0 files changed, 0 insertions(+), 0 deletions(-)
create mode 100644 file1
create mode 100644 file2
create mode 100644 file3
juliet@debian:~/project$
```

Abbildung 1: Einrichten eines Verzeichnisses als Git-Repository: schmerzlos!

Textdatei mit dem Namen ».gitignore« in Ihrem Arbeitsverzeichnis. Listen Sie darin alle zu ignorierenden Dateien auf. Eine »gitignore« Datei sieht gewöhnlich etwa so aus:

```
.*.swp*
tmp
images
```

Mit diesen Einstellungen werden Dateien mit Namen wie ».myfile.swp« (die temporären Dateien von Vim) sowie der Inhalt der Unterverzeichnisse »tmp/« und »images/« ignoriert.

Möchten Sie nun nach all diesen Experimenten, dass die Versionskontrolle dieses Verzeichnis nicht weiter verwaltet (sprich: es soll nicht länger als Git-Repository funktionieren), entfernen Sie einfach das »git« Unterverzeichnis mit »rm -rf .git«. Geben Sie jetzt den Befehl »git status« ein, dann erscheint eine Nachricht, dass dieses Verzeichnis kein Git-Repository ist.

Dieser Schritt ist so einfach, dass Sie nicht vergessen sollten, regelmäßig Backups des Git-Repositorys anzulegen!

Git konfigurieren

Früher oder später erhalten Sie bei den Arbeiten mit Git eine Warnmeldung, dass Ihr Name und Ihre E-Mail-Adresse fehlt. Sämtliche Commits in Git führen den Namen und die E-Mail-Adresse der Person, die den Code eingereicht hat (Committer). Git selbst versucht diese Informationen selbstständig herauszufinden, aber das geht in den meisten Fällen schief. Am besten nehmen Sie diese Einstellungen einmalig von Hand vor, wie es die folgenden Zeilen demonstrieren:

```
git config --global user.name Juliet Kemp
git config --global user.email juliet@earth.2
11
```

So werden Sie nicht nur die lästige Warnmeldung los, sondern sind auch gleich für die Zukunft gerüstet. Die Angaben gelten für sämtliche Repositories, die Sie unter diesem Benutzernamen auf diesem Rechner vornehmen. Wenn Sie also einen Beitrag zu einem anderen Projekt leisten, das auch Git als Versionskontrolle benutzt, übermittelt Git auch diese Standardangaben. Die Angaben lassen sich für ein bestimmtes Repository auch ändern, wie in Listing 1 gezeigt.

Es sind auch noch zahlreiche andere Einstellungen möglich. Sie können zum Beispiel festlegen, dass Git in seiner »diff«-, »status«- und »branch«-Ausgabe verschiedene Farben verwendet (Abbildung 2):

```
git config --global color.diff auto
git config --global color.status auto
git config --global color.branch auto
```

Für weitere Einstellungen lesen Sie das Handbuch [4] und das Git-Tutorial [5].

Git außer Haus

Bisher haben wir behandelt, wie Sie Git für eigene, lokale Projekte verwenden. Wenn Sie jedoch bei einem externen Projekt mitarbeiten, das ebenfalls Git verwendet, müssen Sie zuerst den Bestand des externen Repositorys importieren. Da Git ein verteiltes Versionskontrollsystem ist, können Sie im Prinzip jedermanns Repository benutzen, solange Ihnen der Zugang erlaubt ist.

Dennoch setzen die meisten Projekte wegen der Benutzerfreundlichkeit und der schnelleren Veröffentlichung der Ausgaben meistens ein zentrales Repository

Listing 1: Git-Konfiguration

```
01 cd /home/juliet/my_anonymous_repository
02 git config user.name Someone Else
03 git config user.email someone.else@gmail.com
```

```

juliet@debian: ~/project
File Edit View Terminal Tabs Help
juliet@debian:~/project$ git config --global color.diff auto
juliet@debian:~/project$ git config --global color.status auto
juliet@debian:~/project$ git config --global color.branch auto
juliet@debian:~/project$ git status
# On branch master
# Changed but not updated:
#   (use "git add <file>..." to update what will be committed)
#
#       modified:   file1
#
# Untracked files:
#   (use "git add <file>..." to include in what will be committed)
#
#       newfile
no changes added to commit (use "git add" and/or "git commit -a")
juliet@debian:~/project$ git branch
* master
juliet@debian:~/project$

```

Abbildung 2: Die drei oberen Aufrufe stellen die farbige Ausgabe von „git status“ und „git branch“ ein, die weiter unten folgt.

ein, das als Master-Version des Projekts funktioniert und aus dem Sie zu Beginn Ihrer Arbeit die Dateien herunterladen müssen. Das Importieren eines bestehenden Projekts ist ebenso einfach wie das Anlegen eines eigenen Repositories. Möchten Sie den Kernel-Zweig von Linus Torvalds herunterladen, geben Sie folgenden Befehl ein: (Achtung: Dieser Baum ist recht groß und es wird eine Weile dauern!)

```
git clone git://git.kernel.org/pub/scm/linux/kernel/git/torvalds/linux-2.6.git linux-2.6
```

Andere Linux-Projekte, die Git verwenden sind unter anderem Gimp, Debian, One Laptop Per Child und Perl (für eine längere Liste siehe die Seite Git auf Wikipedia); auf der Webseite des jeweiligen Projekts finden Sie immer den Link zu seinem zentralen Repository. Sobald Sie eine Kopie des Projekt-Repositories auf Ihrem eigenen Rechner haben, können Sie mit der Arbeit am Code beginnen. Es empfiehlt sich, einen Zweig (Branch) für Ihre eigenen Änderungen anzulegen. Das Aktualisieren aus dem zentralen Repository übernimmt »git pull«. Später möchten Sie vielleicht einen Patch

aus den eigenen Änderungen generieren, um ihn dem Projekt zur Überprüfung und Einbeziehung zu senden. Das Arbeiten mit Zweigen (Branching), die Aktualisierung Ihrer Arbeitskopie (Updating), und wie man einen Patch erstellt, beschreibt der Artikel weiter unten.

Versionsgeschichte

Ähnlich wie alle anderen Versionsverwaltungssysteme, führt Git ein vollständiges Protokoll über die Änderungen jeder einzelnen Übertragung (commit). Das bedeutet, dass Sie verschiedene Versionen miteinander vergleichen, die früheren Schritte im Nachhinein überprüfen und sogar zu einer früheren Version des Projekts zurückkehren können.

Der Befehl »git show« zeigt Details zum letzten Commit. Sie sehen hier die Revision-ID, den Autornamen und seine E-Mail-Adresse, das Datum der Übertragung in das Repository, die Anmerkungen des Autors zu den Änderungen und ein Diff der Änderungen. **Listing 2** zeigt ein Beispiel.

Der Befehl »git log« zeigt kurze Informationen zu jedem Commit in der Versionsgeschichte. Welche Informationen genau dabei aufgelistet werden, erklärt die Manpage. Sehr nützlich ist »git log Dateiname«, der nur die Änderungen der Datei *Dateiname* anzeigt. Eine hübsche Einzeilen-Ausgabe erhalten Sie per »git log pretty=oneline« (andere Optionen sind »short«, »medium«, »full«, »fuller«, »email«, und »raw«).

Die Commits haben lange alphanumerische IDs (wie 351e8cf452b92ed591f19fdbd63023a68475a364 in **Listing 2**). Sie sind das Ergebnis eines interessanten Sicherheitsfeatures von Git: Der Name jeder Übertragung ist ein Hash-Wert, errechnet aus dem Inhalt der Übertragung. Dies bedeutet, dass die Authentizität der Versionsgeschichte in zweierlei Hinsicht gewährleistet ist.

Erstens kann ein Angreifer den Inhalt eines Commits nicht modifizieren, ohne dabei auch den Namen zu ändern. Zweitens ist der Name unverwechselbar (global unique). Wenn Sie also mit anderen zusammenarbeiten, ist es immer eindeutig und in jedem Repository nachvollziehbar, worauf sich Ihre Änderungen beziehen. Die Handhabung von Revisionsnummern wird dadurch zwar etwas umständlicher, aber Git macht eine Autovervollständigung, nachdem Sie die ersten paar Buchstaben eintippen, oder Sie verwenden einfach Copy & Paste.

Der Befehl »git diff« zeigt Ihnen die Unterschiede zwischen der zuletzt übertragenen und Ihrer aktuellen Version. Die Unterschiede zwischen zwei früheren Commits ermitteln Sie mit »git diff commitID1 commitID2«. Einen Commit zu widerrufen, gelingt mit »git revert commitID«. Ihren kompletten Baum bringen Sie mit »git checkout« auf den letzten Versionsstand zurück (Achtung: Dadurch werden alle Änderungen überschrieben!).

Abzweigen

Git ist enorm flexibel, und die Möglichkeiten zur Verzweigung (branches) und Zusammenführung (merge) sind nahezu unbegrenzt. Abzweigen (das Erstellen mehrerer „Kopien“ vom gleichen Repository) bedeutet, dass Sie eine Reihe von Änderungen zeitweilig abtrennen, während Sie mit ihnen experimentieren, oder dass Sie verschiedene Versionen eines Projekts erstellen, um den Baum nicht zu beeinträchtigen. Das Abzweigen in Git ist schnell und einfach, und es ist ebenso einfach, wieder alles in den Baum zurückzuführen. Arbeiten Sie also nach Möglichkeit zunächst in einem separaten Branch. Die grundlegenden Git-Befehle für Branches sind:

■ »git branch« listet die aktuellen Zweige des Projekts auf.

Listing 2: Commit und Diff

```

01 commit 351e8cf452b92ed591f19fdbd63023a68475a364
02 Author: Juliet <juliet@earth.li>
03 Date: Tue Mar 2 11:47:04 2010 +0000
04     Testing delete
05 diff --git a/<I>Datei<I>.txt b/<I>Datei<I>.txt
06 deleted file mode 100644
07 index e69de29..0000000

```

- »git branch Name« erzeugt einen neuen, lokalen Zweig.
- »git branch -d Name« löscht den entsprechenden Zweig.
- »git checkout Name« wechselt in den lokalen Zweig über. Verwenden Sie »git checkout -b Name«, um den Zweig auch gleich zu erstellen, wenn Sie darin wechseln.

Nehmen wir an, Sie wollen einen neuen Zweig für Ihr Projekt erstellen und diesen V1.5 benennen. Zunächst erstellen Sie den neuen Zweig: »git branch V1.5«. Ihr Repository hat nun zwei Zweige, aber das aktuelle Verzeichnis entspricht immer noch dem Master (dem Hauptzweig, der automatisch „Master“ heißt). Der Befehl »git branch« zeigt Ihnen das an: Alle bestehenden Zweige sind aufgelistet, und vor dem aktuellen Arbeitszweig steht ein Stern (**Abbildung 3**).

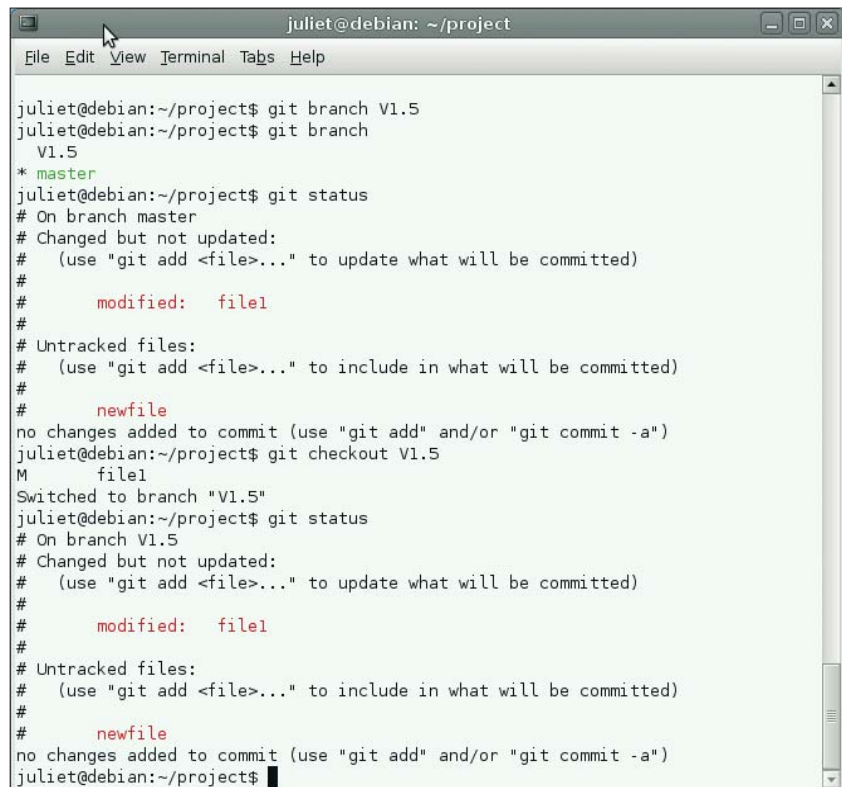
Sie wechseln nun mit »git checkout V1.5« in Ihren neuen, lokalen Zweig. In der Ausgabe von »git status« erscheint nun dieser Zweig mit dem Stern davor. Nun arbeiten Sie in diesem Zweig. Führen Sie die gewünschten Änderungen durch und übertragen (committen) Sie diese wie üblich. Zum Master wechseln Sie wieder mit »git checkout master« zurück. Ihre Änderungen werden hier nicht sichtbar sein. Wechseln Sie wieder zum Zweig V1.5, und da sind sie wieder.

Zusammenführen

Um Zweige wieder zusammenzuführen, verwenden Sie »git merge Name«. Dies wird den Branch »Name« in den aktuellen Zweig (in das aktuelle Arbeitsverzeichnis) einfädeln.

In unserem Beispiel wechseln Sie also zunächst zum Master mit (»git checkout master«), dann führen Sie die Änderungen vom Zweig V1.5 mit »git merge V1.5« in den Master hinein. Der Zweig V1.5 und der Master sind nun zwar identisch, aber »git status« zeigt, dass der zusätzliche Zweig immer noch existiert. Wechseln Sie nun wieder in den Zweig V1.5, um Ihre Entwicklungsarbeit weiterzumachen!

Beachten Sie, dass Git keinen Merge-Vorgang zulässt, solange es noch nicht übertragene Änderungen gibt, da beim Merge das bestehende Arbeitsverzeichnis überschrieben wird. (Als Workaround



```
juliet@debian: ~/project
File Edit View Terminal Tabs Help

juliet@debian:~/project$ git branch V1.5
juliet@debian:~/project$ git branch
  V1.5
* master
juliet@debian:~/project$ git status
# On branch master
# Changed but not updated:
#   (use "git add <file>..." to update what will be committed)
#
#       modified:   file1
#
# Untracked files:
#   (use "git add <file>..." to include in what will be committed)
#
#       newfile
no changes added to commit (use "git add" and/or "git commit -a")
juliet@debian:~/project$ git checkout V1.5
M       file1
Switched to branch "V1.5"
juliet@debian:~/project$ git status
# On branch V1.5
# Changed but not updated:
#   (use "git add <file>..." to update what will be committed)
#
#       modified:   file1
#
# Untracked files:
#   (use "git add <file>..." to include in what will be committed)
#
#       newfile
no changes added to commit (use "git add" and/or "git commit -a")
juliet@debian:~/project$
```

Abbildung 3: Einen lokalen Zweig erzeugen und in den lokalen Zweig wechseln (beachten Sie den Zweignamen am Anfang der Status-Ausgabe).

eignet sich in einem solchen Fall »git stash« – siehe weiter unten in diesem Artikel.)

Das Zusammenfügen von Dateien erledigt Git beim Merge so gut wie möglich automatisch. Aber manchmal gibt es Konflikte (Änderungen, die einander gegenseitig beeinflussen). In einem solchen Fall wird der Merge scheitern, und Git zeigt eine Warnung mit den problematischen Dateien an. Auch »git status« führt diese Dateien als nicht zusammengeführt auf. Öffnen Sie dann die problematischen Dateien in einem Texteditor, um den Konflikt selbst aufzulösen. Die entsprechenden Abschnitte werden gekennzeichnet, wie in **Listing 3**.

Beide Versionen der Datei werden angezeigt, und Sie entscheiden, wie Sie das Problem lösen. Sobald Sie mit der Korrektur der Dateien fertig sind, führen Sie »git add Dateiname« bei allen problematischen Dateien aus, dann vollenden Sie den Merge mit »git commit«. Alternativ können Sie auch »git commit -a« verwenden, um alle Konflikte als gelöst zu markieren. Achten Sie darauf, dass diese wirklich aufgehoben wurden, bevor Sie das tun!

Arbeiten Sie im Team, dann gibt es mehrere Möglichkeiten, um die eigenen Änderungen den anderen Mitgliedern bereitzustellen. Die übliche Vorgehensweise besteht darin, einen Patch zu erzeugen (eine Textdatei, die Ihre Änderungen gegenüber dem Projekt-Baum beschreibt) und diesen per E-Mail zu verschicken. Glücklicherweise macht das Git-Patching-System diese Sache sehr einfach.

Bevor Sie überhaupt mit irgendwelchen Änderungen anfangen, sollten Sie zuerst einen neuen Zweig anlegen:

```
git checkout -b MyBugFix
```

Von nun an arbeiten Sie im Zweig »MyBugFix« weiter. (Lesen Sie die Informationen zum Befehl »git stash« weiter unten in diesem Artikel, falls Sie die Arbeit bereits am Master-Zweig begonnen haben.) Nachdem Sie Ihre Änderungen vollbracht

Listing 3: Konflikt

```
01 test
02 <<<<<< HEAD
03 this is the master branch
04 =====
05 this is the V1.5 branch
06 >>>>>> V1.5
```

und ins lokale Git-Repo übertragen haben, müssen Sie zunächst sicherstellen, dass Ihr lokales Repository auf dem neuesten Stand ist. Sonst würden Sie einen Patch erstellen, der sich auf eine veraltete Version bezieht. Wechseln Sie dazu mit »git checkout master« zurück in den Master-Zweig und aktualisieren Sie diesen mit »git pull« aus dem ursprünglichen Repository.

Wechseln Sie nun wieder mit »git checkout MyBugFix« zu Ihrem Arbeitszweig über, und geben Sie hier den Befehl »git rebase master« ein. Damit übernehmen Sie die Master-Änderungen in Ihren »MyBugFix« Arbeitszweig. (Selbstverständlich müssen Sie die eventuellen Konflikte ebenfalls aufheben.) Git frisst die Versionsgeschichte auf und Ihr Zweig stammt nun von der allerjüngsten Version des Masters ab: ideale Voraussetzungen, um einen Patch zu erstellen:

```
git format-patch master --stdout > 2
mybugfix-patch.diff
```

Dadurch wird der aktuelle Arbeitszweig (hier »MyBugFix«) mit dem Master-Zweig verglichen, und jeder Commit ausgefiltert, der im Master-Zweig nicht vorhanden ist. Je nach Commit wird ein Patch auf stdout ausgegeben und dann in die Datei »mybugfix-patch.diff« weitergeleitet. Alternativ verwenden Sie »git format-patch master«, um im aktuellen Arbeitsverzeichnis einen Patch pro Commit zu generieren oder »git format-patch master -o patchdir«, um diese im Patch-

Verzeichnis abzulegen. Die einzelnen Patches sehen wie eine E-Mail aus.

Einen Patch von jemand anderem wenden Sie mit »git am« an. Zunächst erstellen Sie dazu einen neuen Zweig mit »git checkout -b SarahsPatch«, um den Patch nicht mit eigenen Änderungen zu verwechseln. (Es ist mit Git sehr einfach Zweige zu erstellen und diese wieder zusammenzuführen, nutzen Sie also dieses Feature!) Dann führen Sie die Änderungen mit »git am sarahspatchfile.diff« zusammen, um den Überblick zu haben.

Bei extern verwalteten Projekten möchten Sie wahrscheinlich nicht gleich alle Änderungen in Ihren Master-Zweig einfügen, sondern warten lieber bis diese zentral anerkannt wurden (also bis zum nächsten Update mit »git pull«). Geht es aber um ein kleines Projekt, in dem nur Sarah und Sie beteiligt sind, und sind Sie mit diesem Patch zufrieden, dann führen Sie diesen Zweig mit »git checkout master; git merge SarahsPatch« in den Master-Zweig über.

Beachten Sie, dass »git am« auch mit Patch-Dateien umgehen kann, die in einer E-Mail angekommen sind. Schmeißen Sie dazu einfach all Ihre Patches in eine Mailbox (im Linux Mailbox-Format) und geben Sie dann den Befehl »git am mailboxfile« ein. Der Autorname wird dabei aus der »From:«-Zeile, das Datum aus der »Date:«-Zeile, und die Benennung aus der »Subject:«-Zeile der E-Mail übernommen. Die Beschreibung übernimmt Git aus der »Subject:«-Zeile und dem Text des E-Mail-

Body bis zum Beginn des eigentlichen Patches. Dieses Format benutzt Git auch beim Speichern eines Patches mit »git format-patch«.

Schummeln mit Stash

Ein weiteres nützliches Werkzeug bei der Generierung von Patches und der Arbeit an sehr komplexen Projekten ist »git stash«. Es bietet sich in erster Linie dann an, wenn Sie eine Änderungen aus dem aktuellen Baum entfernen, diese aber nicht verlieren möchten (also nicht endgültig zu einer früheren Version zurückkehren wollen). Der Befehl »git stash save "Arbeit am Projekt Foo"« speichert die bisherigen Änderungen vorübergehend im „stash“ und kehrt zur letzten Version des Projekts zurück. Jetzt können Sie die nötigen kleineren Verbesserungen wie gehabt committen. Benutzen Sie schließlich »git stash pop«, um die gespeicherten Änderungen wieder einzufügen.

Den Befehl »git stash« verwendet man üblicherweise dann, wenn man mitten in einer größeren Arbeit plötzlich einen kleinen separaten Bug entdeckt. In diesem Fall speichert man die Änderungen, behebt den Bug, überträgt die Änderungen, und generiert einen Patch.

Wie Sie bereits gesehen haben, überlappen sich »git stash« und »git branch« teilweise – betrachten Sie also die Stash-Funktion als eine einfachere Art von Branch.

Falls Sie plötzlich merken, dass Sie im falschen Zweig gearbeitet haben, hilft Ihnen ebenfalls »git stash«. In diesem Fall speichern Sie die Änderungen im Stash, wechseln den Zweig und übertragen die Änderungen aus dem Stash in den neuen Zweig.

Taggen mit Git

Der Befehl »git tag« erstellt eine Art Tags oder Labels bei den einzelnen Commits. Tags werden meistens dazu verwendet, Release-Punkte zu markieren, aber Sie können sie auch für andere Zwecke benutzen. Eine Liste der Tags in Ihrem aktuellen Arbeitsverzeichnis in alphabetischer Reihenfolge erhalten Sie über den Befehl »git tag«. Um nach bestimmten Mustern zu suchen, versuchen Sie »git tag v1.*« für Tags, die mit »v1.« anfangen.

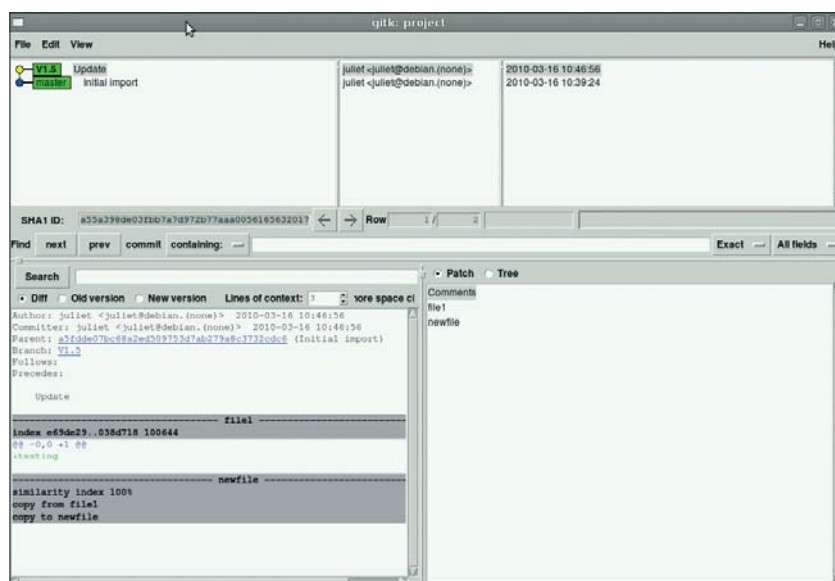


Abbildung 4: Das Git-Fenster mit Versionsgeschichte, Verzweigungen und Diffs.

Eigene Tags zum aktuellen Stand fügen Sie mit »git tag -a« hinzu:

```
git tag -a "April" -m "Neuer Entwickler: Sam"
Sam"
```

Dadurch bekommt der letzte Commit ein Tag namens „April“ mit dem Kommentar „Neuer Entwickler: Sam“ (lassen Sie »-m« weg, erscheint ein Texteditor, in dem Sie Ihren Kommentar verfassen). Der Befehl »git tag« zeigt nun den Namen des Tags an, und »git show April« zeigt den Kommentar zum Tag und die Details des letzten Commits.

GPG-Signaturen

Falls gewünscht können Sie Tags per GPG signieren. Dazu geben Sie einfach »-s Name« statt »-a Name« ein, und das System fragt nach dem Passwort für Ihren GPG-Schlüssel. Möchten Sie eine Signatur überprüfen, benutzen Sie dazu den Befehl »git tag -v Name«.

Einfache Tags ohne Kommentar oder GPG-Signatur erstellen Sie mit »git tag Name«: »git show April« zeigt in diesem Fall nur Informationen zum Commit, keine Tag-Information. Diese Nutzung ist dann sinnvoll, wenn Sie lediglich eine kurze Referenz für eine bestimmte Übertragung brauchen und keinen langfristigen Marker mit vielen Informationen.

Schließlich können Übertragungen auch Tags mit einem Teil der Commit-ID anhängen. Verwenden Sie dazu zuerst »git log«, um die ID zu ermitteln, danach »git tag -a Juni -m "Neue Entwicklerin: Sarah" ad829ce«, um dem Commit mit der ID „ad829ce“ ein Tag anzuhängen.

Weitere nützliche Dinge

Dieser Artikel behandelte bislang die Grundlagen der Arbeit mit Git, damit Sie es bei Ihren eigenen Projekten sinnvoll einsetzen und auch bei größeren Projekten mitarbeiten können. Git kann aber noch viel mehr!

Alle bislang erwähnten Befehle sind sehr flexibel, und es gibt eine breite Palette von Optionen für alle unter ihnen – für mehr Information lesen Sie »man git-[tool]« (zum Beispiel »man git-add«). In den oben genannten Beispielen bezogen sich die Befehle immer auf den letzten Commit, die meisten Befehle sind aber auch

fähig, ganze Bereiche von Commit-IDs zu handzuhaben.

Hier sind noch einige interessante Features, die Sie kennen sollten:

- Installieren Sie das Paket »gitk«. Das »gitk«-Tool zeigt eine überschaubare Versionsgeschichte von Ihrem Projekt (Abbildung 4).

- Git unterstützt sogenannte Hook-Skripte: sie ermöglichen es, Skripte zu setzen, die vor oder nach bestimmten Ereignissen ausgeführt werden. So starten Sie zum Beispiel einen Unit-Test zur Überprüfung von unerwünschten Leerzeichen, bevor ein Commit angenommen wird, oder verschicken nach einem Commit automatisch eine Mail. Einige Beispiele finden Sie im Ordner »git/hooks/« in Ihrem Arbeitsverzeichnis. Um diese auszuprobieren, müssen Sie die »sample«-Erweiterung entfernen.

- Es gibt auch Tools für die Zusammenarbeit mit anderen Versionskontrollsystemen. Schauen Sie sich zum Beispiel »git-svn« gründlicher an, um Git mit einem bestehenden Subversion-Repository zu verwenden [6].

Beachten Sie, dass Git kein hübsches Backup anbietet, wie manche zentralisierte Systeme: Löschen Sie versehentlich das Arbeitsverzeichnis, so geht auch die Versionsgeschichte verloren. In einem verteilten Projekt können Sie natürlich darauf hoffen, dass, andere Team-Mitglieder ihre Backup-Geschichten aufbewahren, ein regelmäßiges Backup ist aber sicher kein Fehler.

Fazit

Git ist schnell, einfach zu bedienen, und unglaublich flexibel. Arbeiten Sie an einem hochgradig verteilten

Projekt ohne festes zentrales Repository, bei dem nur stabile Versionen veröffentlicht werden, eignet sich Git dafür besonders gut. Es ist aber auch ideal für die Arbeit an Ihrem eigenen kleinen Projekt und selbst für Konfigurationsdateien geeignet. (mhi/ofr)

Infos

- [1] Offizielle Webseite von Git: [\[http://git-scm.com/\]](http://git-scm.com/)
- [2] E-Mail von Linus Torvalds über Vorteile eines verteilten Systems: [\[http://lwn.net/Articles/246381/\]](http://lwn.net/Articles/246381/)
- [3] Git-Download: [\[http://www.kernel.org/pub/software/scm/git/\]](http://www.kernel.org/pub/software/scm/git/)
- [4] Git User's Manual: [\[http://www.kernel.org/pub/software/scm/git/docs/user-manual.html\]](http://www.kernel.org/pub/software/scm/git/docs/user-manual.html)
- [5] Offizielles Git-Tutorial: [\[http://www.kernel.org/pub/software/scm/git/docs/gittutorial.html\]](http://www.kernel.org/pub/software/scm/git/docs/gittutorial.html)
- [6] Ein Subversion-to-Git Crash-Course: [\[http://git-scm.com/course/svn.html\]](http://git-scm.com/course/svn.html)

Linux-Server



815 S., 2011, 49,90 €
» www.GalileoComputing.de/2205

Linux Hochverfügbarkeit



454 S., 2011, 49,90 €
» www.GalileoComputing.de/1999

Admin-Know-how

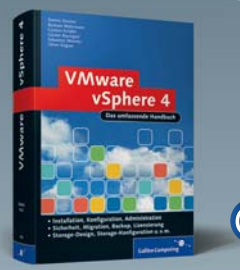
www.GalileoComputing.de

Webserver einrichten und administrieren



497 S., 2. Auflage 2011, mit CD, 39,90 €
» www.GalileoComputing.de/2529

VMware vSphere 4



1.052 S., 2010, 89,90 €
» www.GalileoComputing.de/2179



Galileo Computing
Wissen, wie's geht.



Roman Smirko, 123RF

Umgang mit VHD-Devices

Die Welt ist eine Scheibe

Windows 7 realisiert virtuelle Festplatten durch sogenannte VHD-Container, die eine Reihe von Vorteilen gegenüber physischen Laufwerken bieten. Da die benötigte Virtualisierungs-Technik bereits Teil von Windows 7 ist, lohnt sich für den Admin ein genauerer Blick darauf. *Thomas Drilling*

In rund einem Jahr seiner Existenz hat Windows 7 auch Skeptiker davon überzeugt, dass es weit besser ist als das verunglückte Vista. Auch bei Fachleuten kommt das neue Desktop-Betriebssystem von Microsoft recht gut weg. Traurig für Microsoft ist nur, dass viele Windows-Anwender sich gar nicht so sehr für viele neue Funktionen zu interessieren scheinen. Vielmehr kursieren Anleitungen und Tipps, wie man der Windows-7-Oberfläche wieder das antiquierte Verhalten von XP beibringt, damit der Desktop sich so anfühlt und so aussieht wie eh und je. Eine der vielen Neuerungen besteht darin, dass das Festplatten-Verwaltungsprogramm von Windows 7 mit virtuellen Festplatten umgehen kann. Diese sind bei Microsoft traditionell im VHD-Format gespeichert und nichts anderes als Container-Dateien im Dateisystem des Hosts. Das VHD-Format kommt auch bei Microsofts Virtual PC und bei der Hyper-V-Komponente von Windows Server 2008

zum Einsatz, sodass sich eine mit Windows 7 eingerichtete VHD-Datei auch in Virtual PC oder Hyper-V nutzen lässt. Noch interessanter scheint die Option, von virtuellen Festplatten booten zu können: eine Funktion, die sich Microsoft aber extra bezahlen lässt.

Nutzen-Rechnung

Ein Nutzen von VHD-Dateien auf Windows-7-System-Ebene liegt darin, dass sich virtuelle Festplatten bequem sichern und verwalten lassen, da es sich lediglich um Dateien im Hostsystem handelt. Nachteile gegenüber physischen Festplatten oder Partitionen gibt es kaum. Die Performance beim Schreib- und Lesezugriff liegt nur wenige Prozentpunkte unter der beim physischen Zugriff. Leider funktionieren Ruhezustand und Microsofts Festplattenverschlüsselung BitLocker nicht mit VHD-Dateien. Der eigentliche Clou von VHD-Dateien liegt

darin, dass Windows 7 sogar von virtuellen VHD-Laufwerken bootet, womit sich für den Admin interessante Perspektiven für das unternehmensweite Deployment von Windows-7-Installationen auftun. Windows 7 lässt sich nämlich auch so konfigurieren, dass es ausschließlich von VHD-Dateien bootet, also ohne dass auf dem Arbeitsplatzsystem überhaupt ein Betriebssystem installiert ist.

Der Papierform nach scheint Microsoft mit VHD-Laufwerken eine interessante Technologie in Windows 7 eingebaut zu haben. Gerade wer beispielsweise als Administrator häufig mit Betriebssystemen experimentiert, braucht entweder einen ausgewachsenen Virtualisierungsserver auf Hypervisor-Basis wie VMware ESX oder Citrix XenServer oder viele Festplatten nebst entsprechendem Partitionierungs-Know-how.

Darüber hinaus sind verschiedene Dinge zu berücksichtigen. So braucht Windows 7 zum Beispiel nach wie vor eine Start-

Partition, die auf einer Festplatte mit Master-Boot-Record liegt. Daher arbeitet der ambitionierte Admin heute in der Regel mit virtuellen Maschinen, was allerdings ebenfalls Weitsicht erfordert. Zwar ist ein virtueller PC mit kostenlosen Lösungen wie Virtual Box schnell aufgesetzt, und mit Tools wie dem VMware-Player steht auch dem Zugriff auf eine riesige Auswahl fertiger Images aus dem Netz nichts im Wege. Die spätere Austauschbarkeit solcher virtueller Maschinen scheitert aber gelegentlich an den unterschiedlichen und untereinander inkompatiblen Formaten von Virtual Box, Parallels, KVM oder VMware.

Microsofts ebenfalls kostenlos verfügbares Virtual PC fällt in dieser Rangliste sogar unter den Tisch, weil es keine Nicht-Microsoft-Betriebssysteme unterstützt. Zwar gibt es Konvertierungs-Tools, trotzdem ist der Admin in der Regel auf

die Verfügbarkeit eines Virtual Machine Monitors angewiesen und legt sich damit im Prinzip auch auf ein Format fest.

Eine VHD-Datei ist mithilfe der Datenträgerverwaltung schnell erstellt und eingehängt und steht sofort als neues Laufwerk im Explorer zur Verfügung. Dazu genügt in der Computerverwaltung ein Rechtsklick auf die Datenträgerverwaltung und das Auswählen des Kontextmenü-Eintrags »Virtuelle Festplatte erstellen«.

Es folgt die Angabe des gewünschten Pfades und der Größe der VHD-Datei. Der Administrator kann zwischen einer dynamisch wachsenden virtuellen Festplatte oder solchen mit fester Größe wählen und diese bei »Größe der virtuellen Festplatte« eintragen. Anschließend ist die neue virtuelle Festplatte zu initialisieren, damit der Logical Diskmanager von Windows 7 darauf zugreifen kann (Abbildung 2).

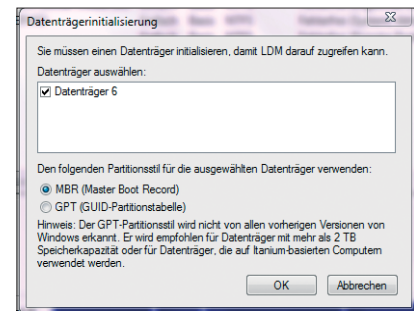


Abbildung 2: Bevor der LDM von Windows 7 auf VHD-Festplatten zugreifen kann, müssen diese initialisiert werden.

Dieser unterstützt übrigens neben dem klassischen MBR-Partitionsstil auch GUID-Partitionstabellen (GPT). Danach lässt sich auf dem Datenträger in der Datenträgerverwaltung ein neues Volume erstellen, dann kann diesem ein Laufwerksbuchstabe zugewiesen und es anschließend formatiert werden, worauf es im Dateimanager zur Benutzung zur Verfügung steht. Dabei lässt sich das virtuelle Laufwerk alternativ auch in einem leeren NTFS-Ordner verfügbar machen. Ambitionierte Admins können dazu das Powershell-Tool »Diskpart« benutzen. Exemplarisch gelingt das Einbinden der VHD-Disk »testplatte.vhd« nebst Zuweisen eines Laufwerksbuchstabens von der Kommandozeile wie folgt:

```
#diskpart
select vdisk file=C:\Pfad-zur-VHD-Datei
attach vdisk
select volume {Volume-Nummer}
assign letter=v
exit
```

Die Volume-Nummer bringt (im Diskpart-Kommando-Modus) der Befehl »list volume« in Erfahrung (Abbildung 3).

Virtuell booten

Wer in das Bootmanagement neuerer Windows-Versionen eingreifen möchte, muss sich seit Windows Vista mit einem neuen Konzept vertraut machen, weil seither die mit einem Texteditor bearbeitbare »boot.ini« von Windows XP ausgedient hat. Das Konzept von Vista und Windows 7 ist flexibler und kompatibel mit EFI-Systemen. Es gliedert sich in drei Komponenten: den Windows Boot Manager, die Startroutine für das Betriebssystem und die Routine zum Fortsetzen von Windows aus dem Ruhezustand.



Abbildung 1: Mit »/?« aufgerufen, gibt »bcdedit« eine ausführliche Hilfeseite aus.

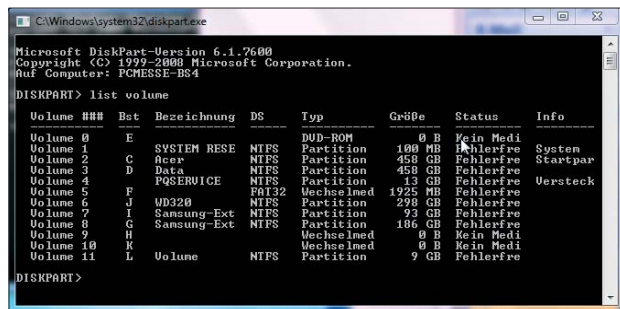


Abbildung 3: Die Zuordnung von Volume-Nummer und Laufwerkbuchstabe fördert »list volume« zutage.

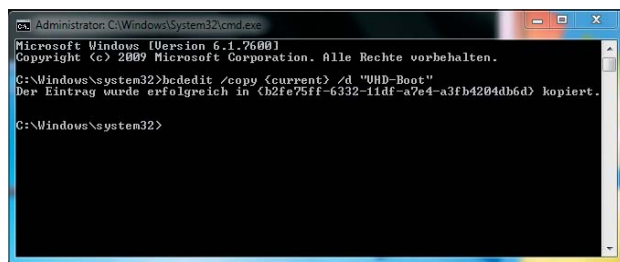


Abbildung 4: Das Kommando kopiert den vorhandenen Bootspeicher-Inhalt und erweitert das Bootmenü um einen neuen Eintrag.

Bei Windows XP und 2000 kommt für alle diese Aufgaben »ntldr« zum Einsatz. Die früher in der »boot.ini« gespeicherten Informationen stecken seit Vista im sogenannten Boot Configuration Data Store (BCD), einer Binärdatei, die je nach System entweder im NVRAM (EFI-Systeme) oder im Verzeichnis »\boot« der Startpartition liegt. Jegliche Änderungen an der Boot-Konfiguration lassen sich nur noch mithilfe des Kommandozeilen-Tools »bcdedit« erledigen.

Die BCD-Hierarchie gliedert sich in drei Ebenen: den Store, die Objekte und die Elemente. Der Store ist ein Container für BCD-Objekte, die ihrerseits wieder BCD-Elemente enthalten. Ein typisches Objekt beschreibt beispielsweise eine Boot-Umgebung, etwa für ein installiertes Windows 7. Alle Elemente enthalten Einstellungen wie den Namen, Debug-Settings oder das Device, auf dem das Betriebssystem zu finden ist. Der Windows Boot Manager ist ebenfalls ein eigenständiges BCD-Objekt, dessen Elemente Parameter wie Timeout, Default-Betriebssystem und Anzeigereihenfolge festlegen, etwa so wie die Einträge des Abschnitts »boot loader« in der »boot.ini«.

Außerdem gibt es im Store meist ein Objekt namens »Windows Boot Loader«, das in der »boot.ini« seine Entsprechung im Abschnitt »operating systems« findet.

Systeme oder für Windows XP und 2000. Letztere aktivieren letztendlich wieder den alten »ntldr«, der seinerseits die »boot.ini« ausliest. Wer Windows 7 oder Vista zusätzlich zu einem bestehenden XP installiert hat, bei dem sollten die hierfür erforderlichen Einträge im BCD-Store bereits existieren. Falls nicht, lassen sie sich mit Administratorrechten durch Eingabe von

```

bcdedit /create {legacy} /d "Windows XP"
bcdedit /set {legacy} device partition=E:
bcdedit /set {legacy} osdevice partition=E:
bcdedit /set {legacy} path \ntldr
bcdedit /displayorder {legacy} /addlast

```

erzeugen. Windows 7 legt dann einen neuen Eintrag im BCD-Store an, der sich im Bootmenü mit »Windows XP« zu erkennen gibt und zum Booten den Bootloader »ntldr« von der Partition »E:« benutzt. Wer parallel zu Windows 7 ein Linux-System installieren möchte, muss beim Installer seiner Distribution lediglich darauf achten, dass dieser den Bootloader von Linux nicht in den MBR schreibt, sondern in die Linux-Partition. Danach kopiert man den Linux Bootloader mit

```

dd if=/dev/xxx of=/bootloader.bin bs=512
count=1

```

in die Datei »/bootloader.bin« und sichert diese zum Beispiel auf einem USB-Stick. Dann bootet man Window 7 und ko-

Dieser Windows Boot Loader ist ausschließlich für Vista und Windows 7 zuständig. Für alle anderen Betriebssysteme nutzt der Windows Boot Manager die als »Windows Legacy OS Loader« bezeichneten Objekte, die im Prinzip nur Verweise auf einen weiteren Bootsektor sind, der einfach nachgeladen wird und dann die Kontrolle des weiteren Startvorgangs übernimmt.

Das gilt zum Beispiel für Linux-

piert die Datei »bootloader.bin« auf die Windows-Partition. Der zugehörige Boot-Eintrag lässt sich dann wieder mit

```

bcdedit /create /d "Ubuntu 10.10"
/application bootsector

```

erzeugen. Auch hier lohnt es sich, die in der Ausgabe von »bcdedit« angezeigte GUID per Kontextmenü zu kopieren, um diese leichter in die folgenden Befehle einfügen zu können:

```

bcdedit /set {neueGUID} device boot
bcdedit /set {neueGUID} path \bootloader.bin
bcdedit /displayorder {neueGUID} /addlast

```

Das Einrichten eines virtuellen Laufwerks im laufenden Betrieb unter Windows 7 funktioniert recht problemlos, und als Sicherungs- oder Daten-Partition ist so ein virtuelles Laufwerk auch durchaus nützlich. Umso gespannter waren wir auf den Versuch, Windows 7 in einem VHD-Laufwerk zu installieren und davon zu booten.

Boot-Konfiguration

Zum Einbinden in den Bootmanager dient bei Windows 7 das Kommando-Tool »bcdedit«. Das Konsolenfenster ist dafür mit Administrator-Rechten zu starten. Das Programm bietet vielfältige Funktionen. Entsprechend umfangreich ist die Liste an Optionen, die »bcdedit /?« anzeigt (Abbildung 1). Eine Übersicht sämtlicher momentan im BCD-Store vorhandenen Objekte liefert »bcdedit /enum -v«.

Bei der Ausgabe wird ersichtlich, dass Windows 7 zur Identifikation eine 32-stellige GUID benutzt. Beim Ändern eines Eintrages ist stets diese GUID anzugeben. Wer sich hierbei Tipparbeit ersparen möchte, klickt oben links im Kommandozeilenfenster mit rechts auf das Minisymbol, wählt »Bearbeiten / Markieren« und markiert dann mit der Maus im Ausgabefenster die gewünschte ID, die damit in der Zwischenablage landet. Übrigens empfiehlt es sich bei Experimenten mit dem Boot-Manager, den aktuellen Boot-Speicher mit »export« zu sichern.

```

bcdedit /export c:\bootbackup

```

Eine solche Sicherung lässt sich mit

```

bcdedit /import bootbackup

```

jederzeit wieder einspielen.

Für das Erweitern des BCD-Stores um eine Boot-Option für die VHD-Festplatte sind folgende Befehle notwendig. So kopiert der Befehl

```
bcdedit /copy {current} /d "Windows von virtueller Partition booten"
```

eine bestehende Boot-Konfiguration und erweitert sie um einen Eintrag für das Booten von VHD. Der Windows Boot Manager wird auch mit »{bootmgr}« referenziert, während {ntldr} den Legacy Windows Loader kennzeichnet. Das »{current}« im obigen Beispiel steht für den Eintrag des aktuell gestarteten Windows. Der als »Default« zu betrachtende Eintrag wird mit »{default}« angesprochen (**Abbildung 4**).

Neue Einträge im Bootspeicher ließen sich mit der Option »/create« anlegen, während sich der soeben erweiterte Eintrag mit »/set« bearbeiten und mit folgenden Eintragsoptionen versehen lässt:

```
bcdedit /set {Bezeichner} osdevice vhd=[C:]\Datei.vhd  
bcdedit /set {Bezeichner} device vhd=[C:]\Datei.vhd  
bcdedit /set {Bezeichner} detecthal on
```

Ist das erledigt, lässt sich Windows 7 in die VHD-Partition hinein installieren. Startet das System nicht, steckt in der Boot-Konfiguration ein Fehler, und der Windows-PC lässt sich im Ernstfall mit »sysprep« wieder in den Ausgangszustand versetzen.

Beim praktischen Test zeigte sich, dass das Installieren von Windows 7 in eine VHD-Datei zwar mit allen Windows-7-Versionen funktioniert, aber nur Windows 7 Enterprise und Ultimate anschließend von einer solchen VHD-Datei booten. Außerdem dauert das Installieren in eine VHD-Partition wesentlich länger als das Installieren auf einer physischen Partition, was die Vermutung nahelegt, dass der VHD-Treiber die Schreibzugriffe verlangsamt oder verzögert. Enttäuschend ist aber, dass Microsoft auch mit dieser grundsätzlich durchaus interessanten Option eine lizenzrechtliche Restriktion verbindet, die zusätzliche Kosten verursacht.

Fazit

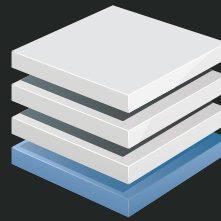
Wer nicht von VHD booten möchte, findet in VHD-Dateien in jeder Windows 7-Version eine überaus nützliche Funktion. Echte Enterprise-Vorteile für Administratoren ergeben sich aber leider erst mit der Ultimate-Version, deren Anschaffungspreis auf der anderen Seite aber auch problemlos den Kauf eines ausgewachsenen Virtual Maschine Monitors einschließen würde.

Als interessante Alternative, die auch die Benutzung des XP Modus in Windows 7 Home ermöglicht, könnte sich dagegen VMLite XP Mode 3.2.6 für Windows 7 erweisen [1], eine Virtualisierungs-Lösung, die wir uns in einer der nächsten Ausgaben näher ansehen werden. Auf den Webseiten der VMLite findet sich übrigens eine Ankündigung des Produkts »vboot« [2], das angeblich das native Booten von unterschiedlichen Betriebssystemen wie Windows XP, Windows Vista, Windows 7 oder auch Linux (32 und 64 Bit) von virtuellen Laufwerken im VHD-Format erlaubt. (ofr) ■

Infos

[1] VMLite XP Mode Version 3.2.6: <http://www.vmlite.com/>

[2] VMLite Vboot: <http://www.vmlite.com/index.php/products/vboot>



OSDC.de
OPEN SOURCE DATA
CENTER CONFERENCE

06. & 07. April 2011 | Nürnberg

AUTOMATION

INSTALLATION

CLOUD COMPUTING

HIGH AVAILABILITY

PUPPET

ICINGA

LOAD BALANCING

PROVISIONING

JASPER

FIREWALLING

Jetzt letzte Plätze sichern:
www.osdc.de



presented by:

NETWAYS®

sponsored by:

ADMIN
Netzwerk & Security

Thomas-Krenn.AG®
Speed is (y)our success



Bei der Administration von Vclouds bleibt Linux außen vor

Wählerischer Platzhirsch

David Hughes / i33RF

VMwares Vsphere ist der unangefochtene Marktführer der privaten Clouds. Zwar bringt der Hypervisor eine Red-Hat-Linux-Konsole, ein CLI und einen Webclient mit, die vollständige GUI ist jedoch allein Windows-Anwendern vorbehalten. Charly Kühnast, Marcel Schynowski, Norbert Graf, Markus Feilner

Mit dem Nachfolger der Virtual Infrastructure 3, unter dem neuen Namen Vsphere 4 [1] auf dem Markt, versucht sich der Platzhirsch im Virtualisierungsrevier zu behaupten. Hinter dem Namen verbirgt sich eine Suite von Produkten: von den Essentials für Einsteiger über Standard und Advanced bis zur Enterprise-Plus-Version für große Umgebungen. Die Unterschiede liegen vor allem in der Flexibilität beim Zuweisen virtueller Hardware und natürlich im Preis. Den VMware Vsphere 4 ESXi Single Server gibt's zwar kostenlos zum Testen, mittlere Setups rangieren aber schnell bei 100 000 Euro und mehr [2].

Vsphere setzt bei ESX-Systemen auf einen Typ-1-Hypervisor, der direkt auf der Hardware läuft, im Gegensatz zu Produkten wie Xen oder KVM, die als Typ-2-Hypervisor auf einem Betriebssystem aufsetzen. VMwares ESX muss sich daher selbst um Hardwaretreiber kümmern, ist aber unabhängig von etwaigen Fehlern des Betriebssystems.

Eine VMware-Cloud (Vcloud) besteht dabei aus mehreren ESX-Servern, die ein

zentrales Vcenter (Abbildung 1) verwaltet. Als Standardwerkzeug dafür dient dem Administrator der Vsphere-Client (Abbildung 2). Der wiederum steuert die ESX-Server beispielsweise über deren Servicekonsole, einem mitgelieferten, abgespeckten Red-Hat-Linux als virtuellem Host.

HA und Green IT

Die Vsphere 4 kann ihren Gästen 255 GByte RAM und bis zu acht virtuelle CPUs zuweisen. Sowohl Speicher als auch Prozessoren lassen sich den Hosts im laufenden Betrieb on the Fly zuordnen – wenn das Gastsystem solche Operationen unterstützt.

Für die HA-Anwender interessant ist das als Fehlertoleranz (Fault Tolerance, FT) bezeichnete Feature, mit dem VMware einen Gast auf zwei Wirtsmaschinen gleichzeitig betreibt und beim Ausfall automatisch umschaltet. Die Technik unterliegt dabei allerdings einigen Einschränkungen: Ein FT-Gastsystem muss mit einem einzigen virtuellen Prozessor

auskommen und seine Daten auf einem externen Speicher ablegen. Das Erstellen von Snapshots ist in derartigen Szenarios nicht möglich.

Ein Alleinstellungsmerkmal von Vsphere sind die Green-IT-Features. Das Distributed Power Management verschiebt die Gastsysteme einerseits bei hoher Last zwischen physikalischen Servern, um die Hardware-Ressourcen möglichst gleichmäßig auszunutzen. Sinkt andererseits die Last auf den Servern, zum Beispiel wenn die Mitarbeiter abends die Arbeitsplätze verlassen, konzentriert der Ressourcen-Verteiler automatisch die laufenden Gastsysteme auf so wenige Wirtssysteme wie möglich und schaltet die überzähligen Maschinen ab. Steigt die Last wieder, weckt er die schlafenden Server selbstständig.

Unter der Bezeichnung VM-Safe erlaubt VMware den Herstellern von Antivirus- und sonstiger Sicherheitssoftware über drei Schnittstellen die direkte Integration in die Hypervisor-Schicht. Die APIs kümmern sich um unterschiedliche Datenströme:

- Das V-Compute-API ermöglicht es, Speicherzugriffe und den CPU-Status zu inspizieren.
- Das V-Network-Appliance-API, auch DV-Filter genannt, bildet einen Paketfilter zwischen der virtuellen Netzwerkkarte und dem virtuellen Switch. Mit DV-Filtern ist jede Art der Paketmanipulation möglich, beispielsweise lässt sich an dieser Stelle eine Firewall platzieren.
- Das Virtual-Disk-Development-Kit-API (VDDK-API) und das zugehörige Entwicklungsframework erlauben Antivirus-Software, die nicht in der virtuellen Maschine installiert sein muss.

Von diesen, angeblich sehr mächtigen Schnittstellen ist aber nur das VDDK-API öffentlich zugänglich, die Nutzung der V-Compute- und DV-Filter-APIs ist auf ausgewählte Partnerfirmen beschränkt. Welche Nutzungsmöglichkeiten die Partner noch aus dem Hut zaubern, bleibt offen. In einem Blog schrieb VMwares Sicherheitsarchitekt Michael Haynes vor wenigen Wochen, dass zumindest für das DV-Filter-API Pläne existieren, „die deutlich über den Einsatz als Sicherheits-Bordmittel hinausgehen“ [3].

PCI und GUIs

Seit Vsphere 4 können Admins auch eine im Wirtssystem steckende PCI-Karte an das Gastsystem durchreichen, wie das Xen oder KVM ebenfalls beherrschen [4]. Das bietet beispielsweise die Möglichkeit, einen Faxserver über Remote-Capi und mit Hilfe einer im Server eingebauten ISDN-Karte zu virtualisieren.

Die Liste der technischen Features von Vsphere ist lang, für Linux-Admins bleibt ein großer Haken: Die Standardoberflächen für die Administration, also Vcenter und der Vsphere-Client sind reine Windows-Applikationen, die auch nicht unter Wine zum Laufen zu bringen sind. Zwar können Admins freier Betriebssysteme auf die Vsphere-Web-GUI zurückgreifen (Abbildung 3). Mit diesem sind aber viele Funktionen wie Vmotion, Storage Motion, Cloning oder Updates nicht möglich, sodass die Open-Source-Welt ausgespart bleibt.

Mindestens eine virtuelle Maschine mit Windows ist also notwendig, um alle Funktionen zu nutzen. Mit dieser kann

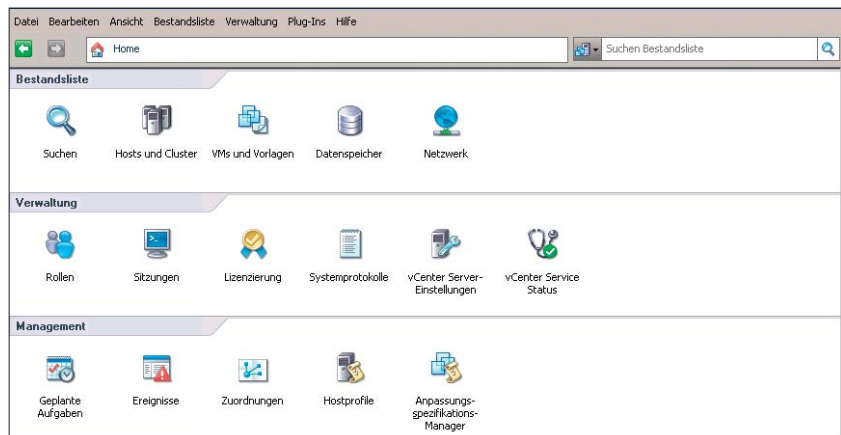


Abbildung 1: Vsphere bietet Verwaltungstools, mit denen Administratoren – allerdings nur unter Windows – auch große ESX-Server-Farmen verwalten.

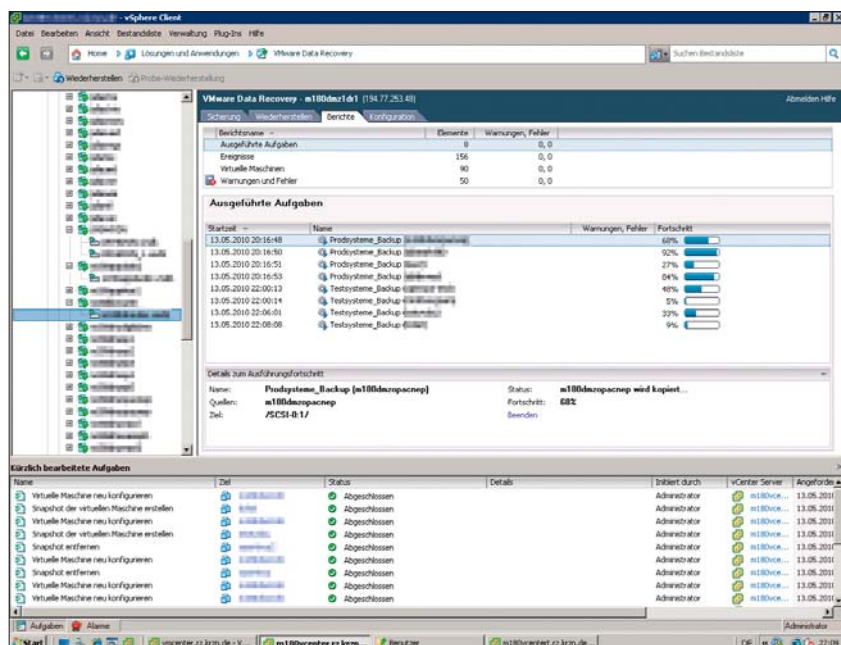


Abbildung 2: Der umfangreiche Vsphere-Windows-Client beherrscht auch komplexeste Funktionen.

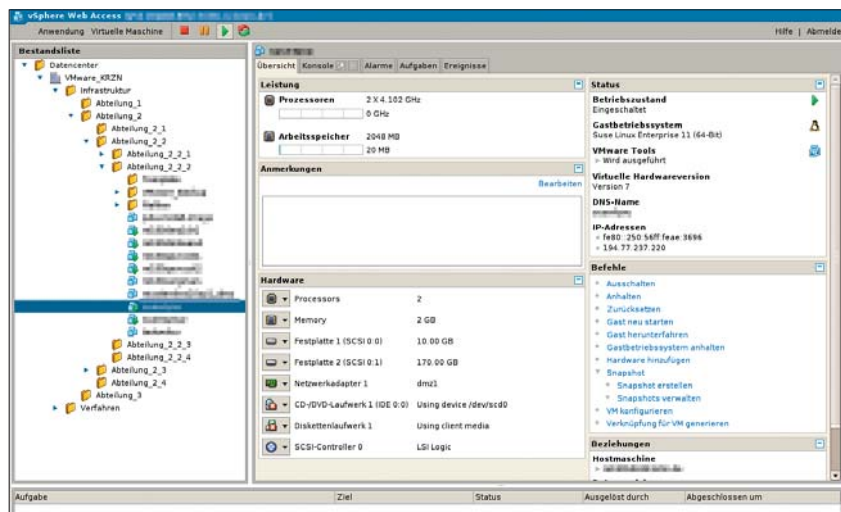
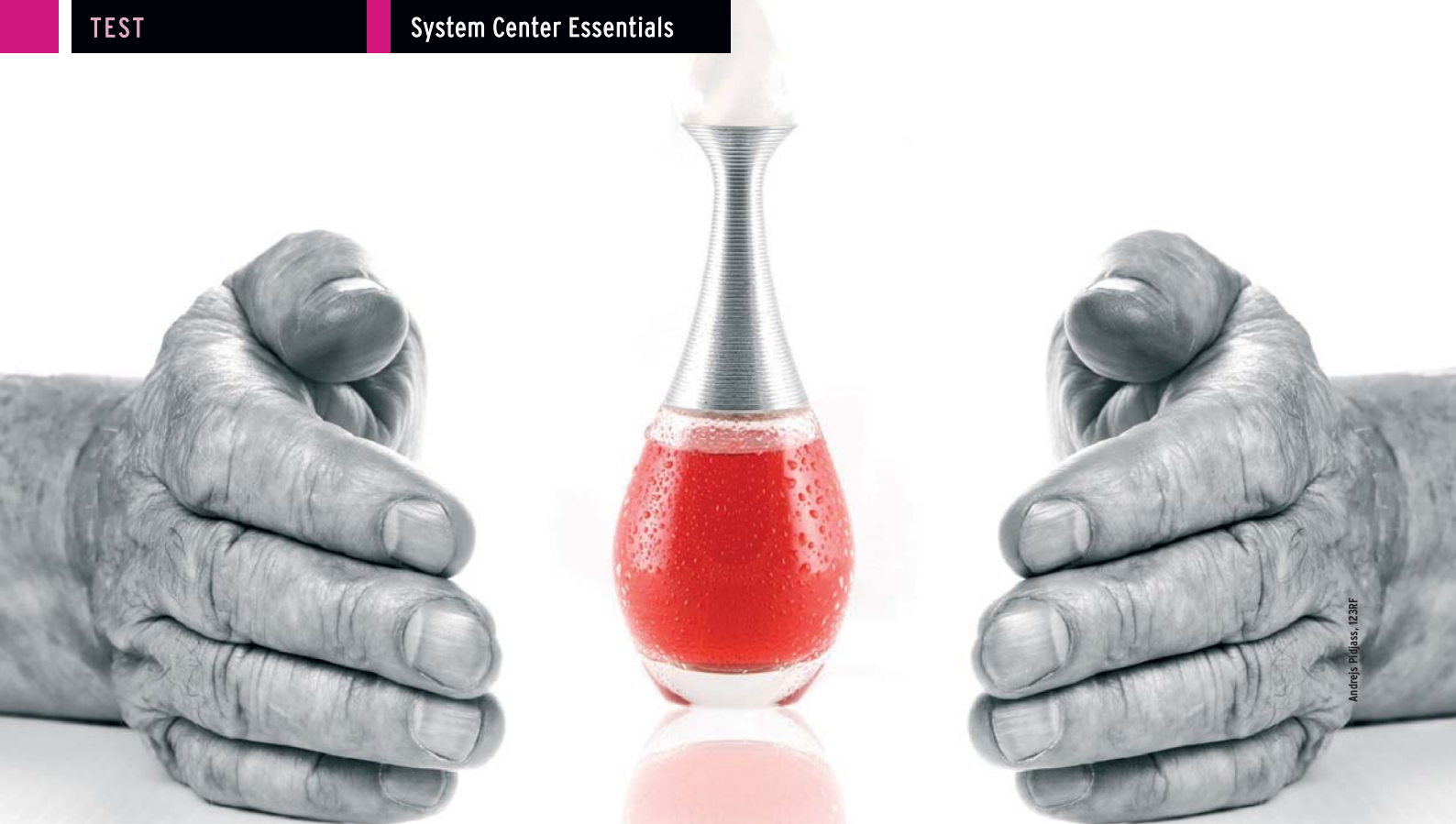


Abbildung 3: Das Web-GUI ist nur ein schwacher, weil nicht vollständiger Ersatz einer Applikation für Linux-Administratoren. Der Funktionsumfang lässt wenig mehr als das Starten und Stoppen der VMs zu.



System Center Essentials 2010

Konzentriert

Über die System Center Essentials können Windows-Administratoren ihre Installationen von virtuellen und realen Servern verwalten. Was die Software im Einzelnen leistet, verrät dieser Artikel Thomas Joos

Die System Center Essentials sind eine Server-Technologie zur Verwaltung, Inventarisierung, Software-Verteilung und Betriebsüberwachung von Servern und Arbeitsstationen im Unternehmen [1]. Vor allem die Überwachung der Server, die Diagnose von Fehlern sowie die Inventarisierung und Software-Verteilung stehen dabei im Mittelpunkt. Die Lösung zielt vor allem auf Unternehmen ab, die auf Serverlösungen von Microsoft setzen und zusätzlich Virtualisierung planen oder bereits virtuelle Server einsetzen.

Management von Windows-Servern

Microsoft System Center Essentials (SCE) 2010 richten sich an mittelständische Unternehmen bis maximal 50 Server und 500 Arbeitsstationen. Das Produkt ist eine Weiterentwicklung von System Center Essentials 2007 und basiert auf Technologien aus Microsoft System Center Operations Manager (SCOM), System Center Configuration Manager (SCCM), System Center Virtual Machine Manager

(SCVMM) und den Windows Server Update Services.

Dabei handelt es sich aber nicht nur um eine gemeinsame grafische Oberfläche für diese drei Produkte, Microsoft hat sie für eine optimale Integration ins Netzwerk in SCE 2010 komplett integriert. Während sich SCOM, SCVMM und SCCM als Einzellösungen an größere Unternehmen richten, hat Microsoft SCE für den Mittelstand konzipiert. Die Installation ist wesentlich einfacher und es ist keine komplizierte Serverinfrastruktur notwendig. Microsoft hat die wichtigsten Funktionen der enthaltenen Serverversionen integriert, die Installation mit Assistenten vereinfacht und die Bedienung mit einer einfachen Oberfläche optimiert, die komplexe Administrationsaufgaben einfach macht.

Die Installation des Produkts stellt auch für weniger geübte Administratoren kein Problem dar. Die Virtualisierung hat

Microsoft eng in das Installationsprogramm und das gesamte Produkt integriert (Abbildung 1). SCE 2010 enthalten wichtige Funktionen des System Center Virtual Machine Managers (SCVMM), mit dem Unternehmen ihre Hyper-V und VMware-Infrastruktur zentral verwalten können. Existierende Hyper-V-Server lassen sich genauso importieren wie VMware-Systeme. Virtuelle Systeme lassen sich in SCE 2010 erstellen, konfigurieren, verwalten und überwachen. Alle Steuerungselemente dazu finden sich in der zentralen Verwaltungsoberfläche der SCE. Auch das Kopieren von VMs zwischen Hosts ist

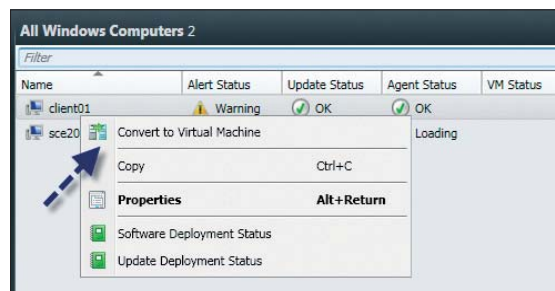


Abbildung 1: Die SCE unterstützen durchgängig Virtualisierung und konvertieren auf Knopfdruck physische Server.

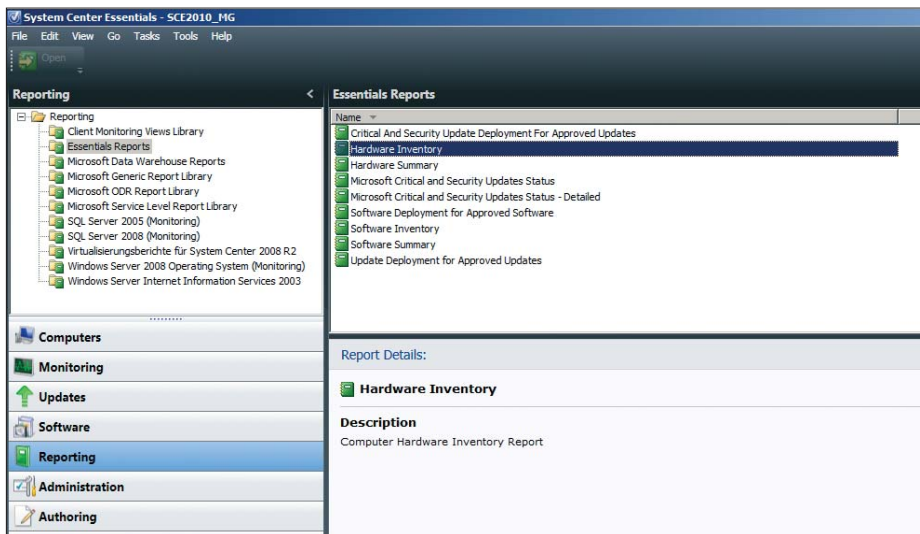


Abbildung 2: Umfassende Berichte dokumentieren den Zustand des Netzwerks und dienen der Inventarisierung.

möglich sowie das intelligente Platzieren von VMs auf einzelnen Servern.

SCE 2010 bieten zusätzlich Funktionen zur Verwaltung, Überwachung, Inventarisierung und Software-Verteilung physischer und virtueller Server und aller Clients im Netzwerk. Sobald in der Konsole ein Fehler oder neuer Computer auftaucht, kann der Administrator über das Kontextmenü die dazu notwendigen Aufgaben anzeigen. Die Übersichtsseiten hat Microsoft im Vergleich zum Vorgänger verbessert und anpassbarer gestaltet.

Die Software unterstützt mit Assistenten die Integration der einzelnen Technologien auch für Administratoren, die keine Spezialisten in diesem Bereich sind. Neben der Überwachung und Diagnose, erstellt SCE auf Wunsch regelmäßig Berichte (Abbildung 2). Schon nach der Installation sind die SCE sehr schnell

einsatzbereit, die entsprechende Konfiguration erfolgt wiederum über hilfreiche Assistenten.

Bestandteil der Berichte ist der aktuelle Status der Server, und zwar physisch wie auch virtuell. Neu ist die Funktion, dass SCE für einzelne physische Server die Virtualisierung vorschlagen kann, wenn diese bestimmte Bedingungen erfüllen. Administratoren sehen in der Verwaltungskonsole, ob die Server problemlos funktionieren, neue Updates zur Verfügung stehen und sich neue Computer im Netzwerk befinden, die angebunden werden wollen. Auch die Inventarisierung von Hardware ist möglich.

Die Überwachungsfunktion der SCE findet Fehler auf den Servern, bevor sie gravierende Auswirkungen haben. Weisen zum Beispiel erste Warnungen in den Ereignisanzeigen auf defekte Hardware

hin, oder findet der installierte Agent auf dem Server einen Fehler, informiert die Software automatisch den Administrator. SCE 2010 können Administratoren per E-Mail, Pager oder SMS benachrichtigen. Dazu überwacht die Lösung alle Server im Netzwerk in Echtzeit, egal ob es sich dabei um physische oder virtuelle Maschinen handelt.

Überwachung mit Management Packs

Serverdienste wie Exchange, Active Directory oder SQL überwachen SCE mit speziellen Management Packs bis ins Detail.

In diesen Management Packs sind Regeln zur Überwachung spezieller Bereiche des Servers wie Warteschlangen, Verfügbarkeit der Postfächer und so weiter integriert. Die Regeln darf der Administrator individuell anpassen (Abbildung 3).

Neben den Management Packs, die Microsoft und seine Partner zum größten Teil kostenlos zur Verfügung stellen, liest die Software auch Informationen per SNMP von Routern, Switches oder anderen Geräten aus. Klicken Administratoren eine Warnung oder einen Fehler in der Konsole an, zeigen SCE die mögliche Ursache und Problemlösung in einem Fenster. Über das Kontextmenü zum Fehler schlagen die SCE weitere Diagnosemöglichkeiten und Werkzeuge vor, um den Fehler einzugrenzen.

Für jeden Fehler lässt sich eine eigene, firmenspezifische Lösung hinterlegen.

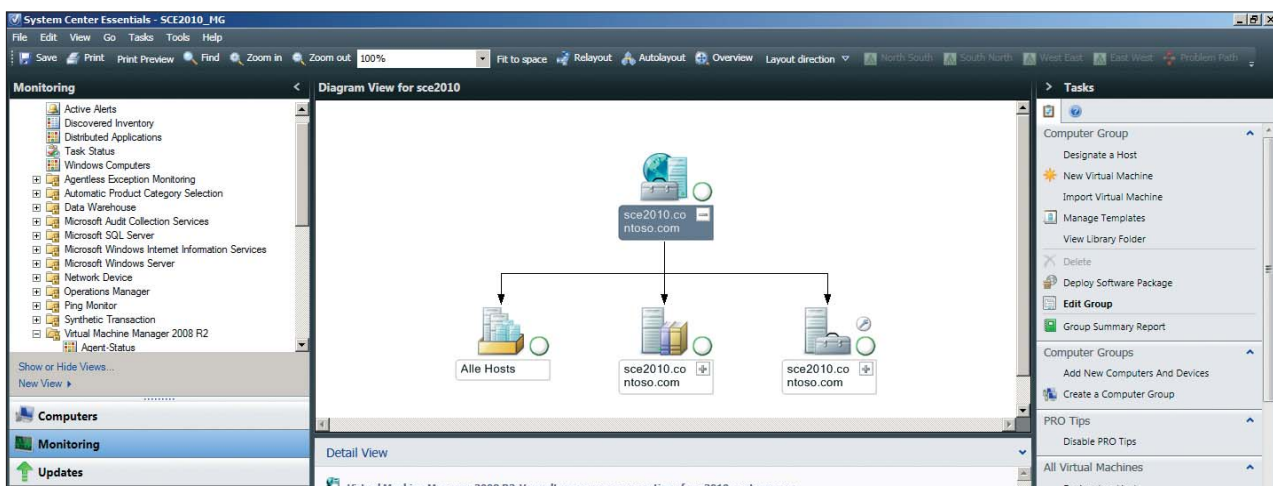


Abbildung 3: Die Überwachung der Server basiert auf Management Packs, die sich in die Konsole integrieren.

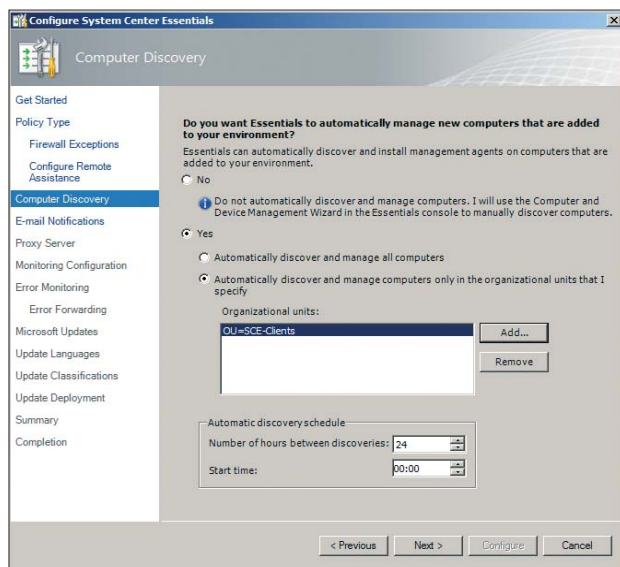


Abbildung 4: Computer lassen sich automatisiert über Assistenten einbinden.

Das hat den Vorteil, dass beim erneuten Auftauchen des gleichen Problems Administratoren spezifische Hinweise erhalten. Eine der in SCE 2010 integrierten Technologien ist der System Center Configuration Manager. Dieser stellt neben automatischer Software-Verteilung auch eine vollständige und automatisierte Hard- und Software-Inventarisierung zur Verfügung, die sich in der gleichen Oberfläche verwalten lassen wie die anderen Funktionen. Auch hier kann SCE wieder Berichte erstellen, die übersichtlich gestaltet sind und sich im PDF-, Word- oder Excel-Format exportieren lassen. Neue Computer erfasst SCE automatisch und bindet diese in die Überwachung und Inventarisierung ein (Abbildung 4).

Über das Kontextmenü kann der Administrator auch festlegen, welche Software-Pakete SCE automatisch auf dem Computer installieren soll. Durch die Integration von SCOM, SCCM und WSUS bieten die SCE alle notwendigen Funktionen einer Client-Lifecycle-Management (CLM)-Suite. Da auch eine Fernwartungsfunktion enthalten ist, lassen sich alle angeschlossenen Geräte direkt über SCE auf Basis des Remote Desktops zu Support-Zwecken unterstützen.

Über den Software-Bereich der Konsole kann der Administrator anzeigen, auf wie vielen Computern eine spezielle Anwendung installiert ist. Alle dazu notwendigen Aufgaben und Informationen sind weitgehend selbsterklärend über den jeweiligen Menüpunkt im Kontextmenü er-

unterstützen. Tritt auf einem Rechner ein Fehler auf, erhält der Administrator eine Übersicht darüber, welche Hardware im Computer steckt, welche Software installiert und wie die aktuelle Auslastung des Systems ist (Abbildung 5).

Testen von SCE 2010

Haben Sie die Testversion heruntergeladen und entpackt, starten Sie den Installationsassistenten. Die Installation erschöpft sich im Bestätigen weniger Optionen. Nach der Auswahl überprüft der Assistent, ob auf dem Server alle Systemvoraussetzungen für SCE 2010 vorhanden sind. Der Management-Server lässt sich auch auf einem Domänencontroller installieren und von der Datenbank trennen. Den Essentials Reporting Server können Sie entweder auf dem gleichen Server installieren oder auf einem getrennten Server, das gilt auch für die Verwaltungskonsolle.

Nach der Installation starten automatisch die Verwaltungskonsolle von SCE und der Assistent zur Einrichtung. Die Anbindung und Konfiguration verwalteter Computer erfolgt über Gruppenrichtlinien. Für den Fernzugriff auf angeschlossene Computer ist der TCP-Port 135 notwendig. Für Datei- und Druckerfreigaben und andere Einstellungen benötigen Sie zusätzlich noch den Port 445 und 6270. Auch die UDP-Ports 137 und 138 müssen geöffnet sein. Auf dem Verwaltungsserver müssen die Ports 80, 445, 5723, 5724, 8530, 8531 und 51906 offen sein.

Installieren Sie SCE 2010 auf einem Server mit ISA 2006 oder dessen Nachfolger

reichbar. Tritt auf einem Rechner ein Fehler auf, erhält der Administrator eine Übersicht darüber, welche Hardware im Computer steckt, welche Software installiert und wie die aktuelle Auslastung des Systems ist (Abbildung 5). Natürlich lässt sich auf dem gleichen Weg Software auf den angeschlossenen Computern auch wieder deinstallieren. SCE 2010

Forefront Threat Management Gateway 2010, müssen Sie noch die Ports 5723 und 5724 für das interne Netzwerk zulassen. Nach der Installation können Sie entweder warten, bis SCE automatisch Clients findet, oder den Clientcomputer manuell an SCE anbinden. Nachdem Sie über den Knoten »Software« ein Paket erstellt haben, erscheint bereits der Assistent, der Sie bei der Verteilung unterstützt. Nachträglich lassen sich diese Einstellungen über das Kontextmenü anpassen. Öffnen Sie den Bereich »Computer« in der Konsole, sehen Sie im obersten Bereich eine Zusammenfassung aller angeschlossenen Computer, den aktuellen Zustand, die Alarme und Informationen zum Thema Virtualisierung. Bei der Virtualisierung unterstützen SCE auch mit sogenannten PRO Tips (Performance and Resource Optimization). Hier erhalten Administratoren Tipps zur Verwendung der Ressourcen, zum Beispiel bei der Aufteilung der virtuellen Gäste auf die einzelnen Hosts.

Auf den verwalteten Computern muss ein SCE-Agent installiert sein. Mit dem Agentless Exception Monitoring (AEM) können verschiedene Aufgaben auf den Clients auch ohne Agent durchgeführt werden, zum Beispiel die Suche nach Fehlern. Die entsprechenden Ansichten finden Sie über »Monitoring\Agentless Exception Monitoring«. Mehr Informationen zu den SCE geben die Websites [2], [3] und [4].

Lizenzen

Mit der Testversion können Unternehmen 180 Tage lang bis zu 50 Server und 500 Clients überwachen [5]. Die verwalteten Computer müssen sich nicht in der gleichen Domäne wie der Server befinden, hier ist eine verteilte Domänenstruktur möglich. Allerdings muss zwischen den Domänen eine Vertrauensstellung vorhanden sein. Zusätzlich können Sie bis zu 100 Netzwerkgeräte überwachen, die im Netzwerk verfügbar sind. Testinstallationen können Sie auf die lizenzierte Versionen aktualisieren. SCE 2010 wird pro physischem Gerät lizenziert. Virtuelle Server müssen Unternehmen nicht lizenzieren, allerdings wertet die Software auch virtuelle Server als vollwertige Geräte. Haben Sie beispielsweise 20

Lizenzen gekauft, und überwachen Sie im Netzwerk 20 physische Server sowie 30 virtuelle Server, können Sie keine weiteren Lizenzpakete hinzufügen, da die maximale Anzahl überwachter Systeme erreicht ist, auch wenn Sie nur 20 Lizenzen kaufen müssen.

Die Installation benötigt mindestens 4 GByte RAM, 20 GByte freien Festplattenplatz und mindestens einen Prozessor mit 2,8 GHz. SCE 2010 lassen sich auf Servern mit Windows Server 2003 (R2) und Windows Server 2008 (R2) installieren. SCE 2010 unterstützt auch die 32-Bit-Versionen von Windows Server 2003 ab Service Pack 2 und Windows Server 2008, die Installation unter 64-Bit ist aber empfohlen. Wollen Sie virtuelle Maschinen verwalten, müssen Sie SCE unter Windows Server 2008 x64 oder Windows Server 2008 R2 installieren.

Da die Virtualisierung einer der zentralen Bereiche von SCE 2010 ist, macht der Einsatz ohne die Verwaltung von Hyper-V-Servern nicht unbedingt Sinn. SCE 2010 setzt eine Active Directory-Infrastruktur voraus, da die Authentifizierung auf Active Directory aufbaut. Die Verwaltungskonsole können Sie auch auf Windows XP Professional (ab SP2), Windows Vista und auf Windows 7 installieren. Setzen Sie im Unternehmen SCE 2007 ein, können Sie direkt auf SCE 2010 aktualisieren. Als Datenbank setzen SCE 2010 SQL Server 2008 ab SP1 voraus. Findet der

Installationsassistent keinen SQL-Server, lässt sich automatisch SQL Server 2008 Express Edition installieren. Itanium-basierte Versionen von Windows Server 2003/2008/2008 R2 lassen sich mit SCE 2010 nicht verwalten.

Fazit

Mit Microsofts System Center Essentials 2010 erhalten mittelständische Unternehmen die gleichen Technologien wie Großkonzerne, um Fehler auf Servern schnell zu erkennen und zu beheben. Die Software erleichtert die Verwaltung und bietet dabei einen guten Überblick über das eigene Netzwerk. Eine Überwachung von Unix- und Linux-Computern ist ebenfalls möglich, allerdings nur sehr rudimentär. Wer Linux- oder Unix-Rechner überwachen möchte, sollte zumindest parallel auf andere Überwachungslösungen wie Nagios [6] setzen.

Der wichtigste Bereich der SCE ist die Überwachung und Diagnose von Fehlern und die Verwaltung virtueller Server. Die SCE helfen vor allem bei der proaktiven Überwachung wichtiger Serverdienste und der automatischen Software-Verteilung, die in mittelständischen Unternehmen oft zu kurz kommen.

Durch die Installation eines einzelnen Produktes erhalten Unternehmen kostengünstig eine angepasste Serverlösung zur Steuerung der Infrastruktur. Trotz des

großen Funktionsumfangs ist die Installation, Einrichtung und die Verwaltung der Software überschaubar. Der Einsatz rechnet sich allerdings nur für Unternehmen, die hauptsächlich Microsoft-Server verwalten müssen, auf Hyper-V setzen und noch kein Werkzeug für die zentrale Serververwaltung haben. (ofr) ■

Infos

- [1] Produktseite bei Microsoft:
[<http://technet.microsoft.com/de-de/systemcenter/essentials/default>]
- [2] Ausführliche Anleitung zu SCE 2010 beim Technet: [<http://technet.microsoft.com/de-de/library/ff603627.aspx>]
- [3] Blog der Entwickler: [<http://blogs.technet.com/b/systemcenteressentials/>]
- [4] Microsoft-Foren für SCE:
[<http://social.technet.microsoft.com/Forums/en-US/category/systemcenter>]
- [5] 180-Tage-Testversion:
[<http://www.microsoft.com/downloads/details.aspx?FamilyID=7beafe76-053d-42c7-8ba5-3cf0167e200e&displayLang=de>]
- [6] Nagios: [<http://www.nagios.org>]

Der Autor

Thomas Joos ist freiberuflicher IT-Consultant und seit über 20 Jahren in der IT tätig. Neben seinen Projekten schreibt er praxisnahe Fachbücher und Fachartikel rund um Windows und andere Microsoft-Themen. Online trifft man ihn unter [<http://thomasjoos.spaces.live.com>].

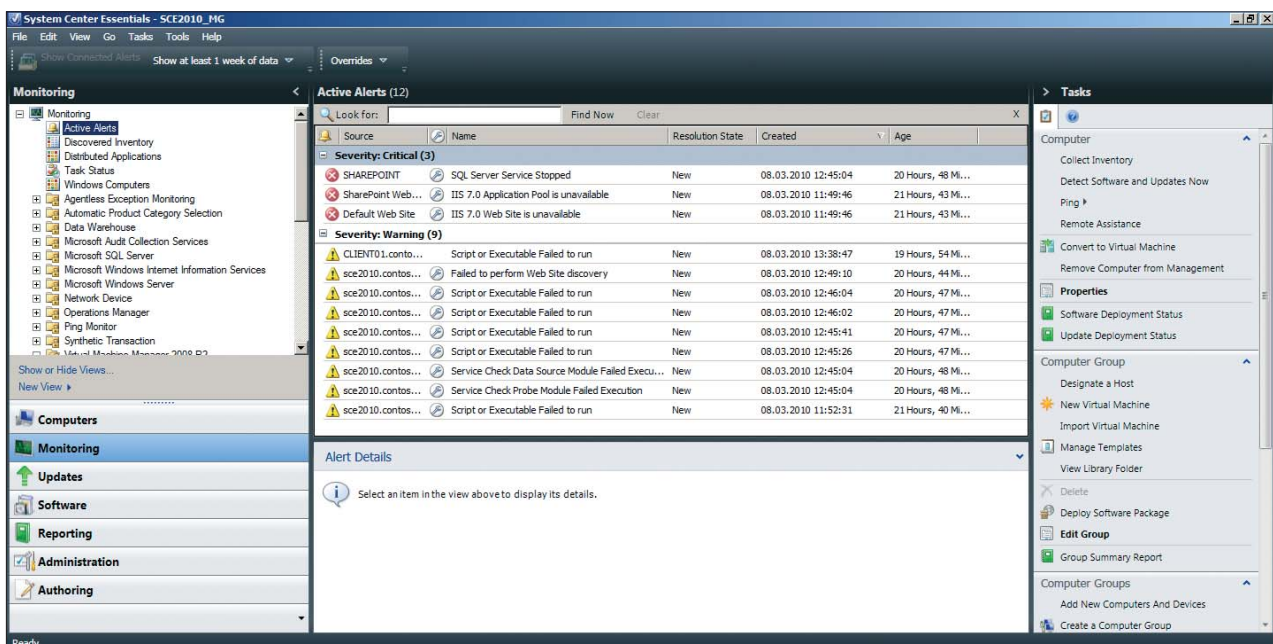


Abbildung 5: Die SCE zeigen die auf den überwachten Servern aufgetretenen Fehler in einer zentralen Oberfläche an.



Ben, Fotolia

Netzwerküberwachung für Windows

Vermessen

Dieser Artikel stellt zwei Windows-Werkzeuge zur Netzwerküberwachung näher vor, die für Administratoren kleiner wie auch großer Umgebungen interessant sind. Björn Bürstinghaus

Eine Herausforderung für jeden Administrator ist die permanente Verfügbarkeit von Servern und der von ihm bereitgestellten Dienste. Daher spielen die Überwachung von Netzwerken und die dafür eingesetzten Werkzeuge eine wichtige Rolle, um im Notfall schnell auf Ausfälle reagieren zu können.

PRTG

Der PRTG Network Monitor des Nürnberger Unternehmens Paessler kann sämtliche Server-Dienste, egal ob auf Windows oder Linux, und auch Netzwerkgeräte auf ihre Verfügbarkeit und die Bandbreitennutzung des Netzwerkverkehrs hin überwachen. PRTG ermöglicht die Überwachung per Windows Management Instrumentation (WMI), dem Simple Network-Management-Protokoll (SNMP) sowie bei kompatiblen Routern (zum Beispiel Cisco) per Netflow beziehungsweise

sFlow. In der aktuellen Version von PRTG müssen Netflow-Sensoren nicht mehr separat lizenziert werden, sondern zählen wie jede WMI- oder SNMP-Abfrage als eine Sensorlizenz. [\[1\]](#), [\[2\]](#).

Jedes zu überwachende Gerät kann unterschiedliche Daten wie beispielsweise Prozessorauslastung, ein- und ausgehenden Netzwerkverkehr oder auch Produktspezifisches wie die Anzahl der aktuellen Verbindungen auf einem Microsoft Internet Information Server (IIS) liefern. PRTG fügt dabei sämtliche auf dem Gerät verfügbaren Sensoren selbstständig über die automatische Ermittlung hinzu. Dafür müssen Administratoren bei Windows-Systemen lediglich Benutzername und Passwort sowie bei Linux-Systemen und Netzwerkgeräten die SNMP-Authentifizierung angeben. Die Konfiguration der Authentifizierung kann global wie auch pro Gerät geschehen. Durch die Integration von Packet Sniffing ist PRTG in der

Lage, nicht nur Sensoren zu überwachen, sondern es kann auch analysieren, welche Anwendung oder IP-Adresse einen hohen Netzwerkverkehr im Unternehmen verursacht. Möglich ist die Ermittlung der Informationen über die Netzwerkkarte des Monitoring-Servers oder über den Monitoring-Port eines Switches.

Auch an Virtualisierung haben die PRTG-Entwickler gedacht. So können Administratoren VMware, Xen sowie Hyper-V-Hosts und natürlich deren gehostete virtuelle Maschinen genau wie ein physisches Serversystem auf Verfügbarkeit und Fehler hin überwachen. Beim Monitoring von Amazon-EC2-Instanzen hilft das enthaltene Amazon Cloudwatch, das die Leistung der angemieteten Instanz im Blick behält.

Den gesamten Funktionsumfang stellt der PRTG Network Monitor über eine Weboberfläche ([Abbildung 1](#)) bereit und bietet zusätzlich über eine Windows-An-

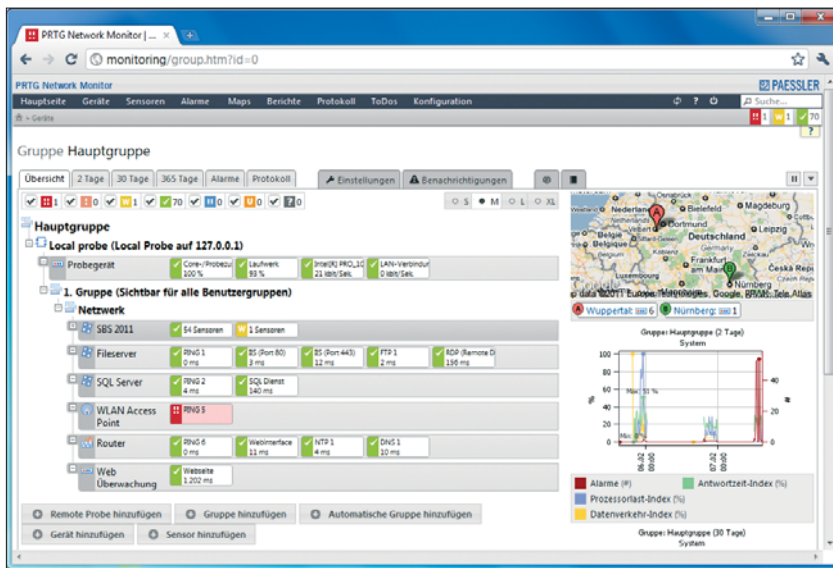


Abbildung 1: Die Weboberfläche von PRTG mit vollem Funktionsumfang der Netzwerküberwachung.

wendung (Abbildung 2) einen schnellen Überblick über sämtliche Sensoren. Benachrichtigungsfunktionen bei Warnungen und Fehlern erhält der Administrator auf Wunsch direkt auf dem Desktop beziehungsweise im Webbrowser.

Bei den Benachrichtigungen bietet PRTG mehrere Möglichkeiten, den Administrator über Warnungen und Ausfälle zu informieren. Neben Mitteilungen per E-Mail lassen sich auch Kurzmitteilungen (SMS) sowie Meldungen zu ICQ und Windows Live Messenger verschicken. Darüber hinaus verschickt PRTG auch Syslog-Nachrichten und SNMP-Traps oder führt HTTP-Aktionen und Programme aus, sobald eine Warnung oder ein Fehler auftritt.

Viele Netzwerküberwachungslösungen nutzen Microsofts SQL Server als Backend für die Speicherung historischer Daten, PRTG dagegen verwendet eine selbst entwickelte Datenbank. Dies hat den Vorteil, dass gespeicherte Daten erheblich schneller abrufbar sind, da sämtliche Informationen chronologisch abgelegt sind und nur noch in der gespeicherten Reihenfolge ausgegeben werden müssen.

Die meisten Überwachungslösungen erlauben es, SNMP-Bibliotheken hinzuzufügen. So ist dies auch bei PRTG mit dem kostenlosen Tool MIB Importer (Abbildung 3) möglich, um Indikatoren von Netzwerkgeräten einzubinden, die standardmäßig nicht unterstützt sind. Auf diesem Weg können Sie beispielsweise

den APC Environmental Manager einbinden und Informationen zur aktuellen Temperatur und Luftfeuchtigkeit von Serverschränken beziehungsweise Räumen in PRTG anzeigen lassen.

Die Erstellung von automatischen Netzwerkarten (Maps), wie man es von Lösungen wie Whatsup Gold kennt, ist in PRTG nicht enthalten, hier muss der Administrator manuell tätig werden. Für diesen Zweck bietet der Map-Bereich in PRTG jede Menge Möglichkeiten, eigene Karten mit einer Vielzahl an Icons zu erstellen. Durch die Integration von Google Maps können Landes- und Weltkarten mit verschiedenen Unternehmens- beziehungsweise Serverstandorten als Map

dargestellt werden. Den Standort zur Anzeige in Google Maps darf der Admin pro Gerät angeben. Dadurch erhält er nicht nur die Integration von Google Maps in eine Map, sondern auch eine grafische Ansicht im rechten Bereich der Geräteübersicht in der Weboberfläche (siehe Abbildung 1) sowie beim Aufruf einzelner Geräte.

Auch die Berichterstellung kommt bei PRTG nicht zu kurz. Berichte kann es als HTML- oder PDF-Datei auf Knopfdruck oder auch zu einem geplanten Zeitpunkt erstellen. Alternativ zur Speicherung der Berichte in einem Verzeichnis kann es diese auch bequem als E-Mail zustellen. Das Lizenzmodell von PRTG Network Monitor 8 ist sehr einfach strukturiert und abhängig von der Anzahl der benötigten Sensoren. Dabei spielt es keine Rolle, ob ein zu überwachendes System einen oder mehrere Sensoren enthält. Die Anzahl der Zugriffe per Web- und Windows-Anwendung ist im Gegensatz zu vielen anderen Lösungen ebenfalls nicht reglementiert.

Überwachung per iPhone

Paessler bietet mit iPRTG eine eigene, kostenpflichtige iPhone App (siehe Abbildung 4) für den PRTG Network Monitor, die den Zustand des Netzwerks und der Systeme auch unterwegs auf Apples Smartphone per WLAN oder mobilen Datenverbindung anzeigen kann. So hat der Administrator auch unterwegs sämt-

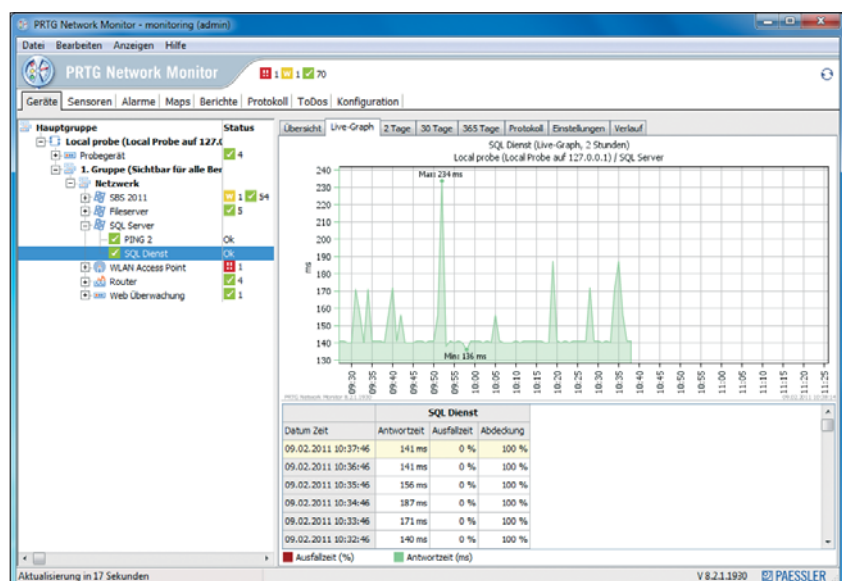


Abbildung 2: Die Windows-Anwendung von PRTG präsentiert alle wichtigen Informationen auf einen Blick.

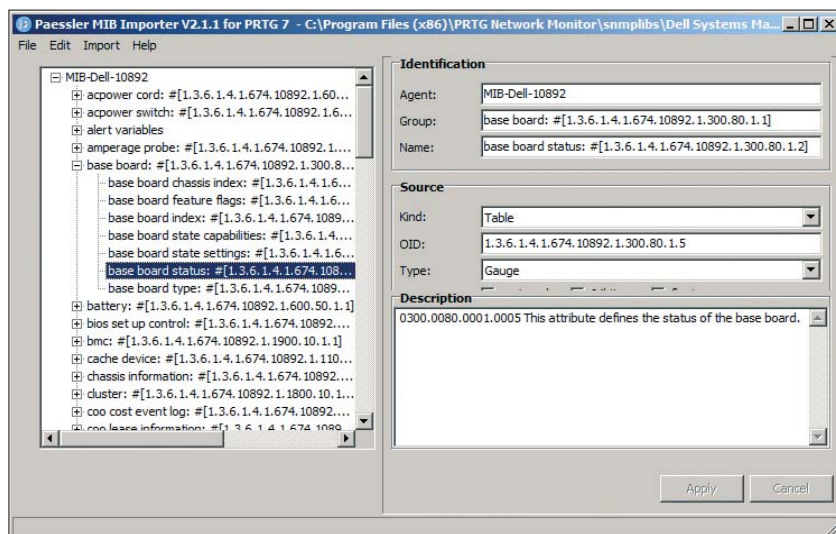


Abbildung 3: Neue SNMP-Bibliotheken lassen sich über den MIB Importer zu PRTG hinzufügen.

che Informationen zu Geräten, Sensoren, Alarmen und Karten immer dabei. Bislang sind Push-Benachrichtigungen auf das iPhone als Alternative zu einer SMS nicht möglich. Alternativ zur iPhone App ist über die Mini-HTML-Weboberfläche der Zugriff von jeglicher Smartphone Plattform möglich.

Failover-Cluster inklusive

In jeder Lizenz des PRTG Network Monitor ist der Einsatz einer kostenlosen Failover-Cluster-Lizenz enthalten. Dies bedeutet, dass man die Software auf einem zweiten System, beispielsweise an einem anderen Standort, installieren

darf und im Falle eines Ausfalls des primären Systems weiterhin von einer permanenten Überwachung des Netzwerks profitiert. Durch die bereits erwähnte Eigenentwicklung der Datenbank fallen auch keine zusätzlichen Kosten für einen zweiten SQL Server an.

Netcrunch

Im Gegensatz zum PRTG Network Monitor steht bei Netcrunch von Adrem Software die Windows-Administrationskonsole (siehe Abbildung 5) im Zentrum, über die die gesamte Konfiguration und Visualisierung des zu überwachenden Netzwerks erfolgt. Sie lässt sich von einem Windows Client aus nutzen und

stellt die Verbindung zum Netcrunch Server über den Verbindungsbroker her, einer kleinen Anwendung im Infobereich der Taskleiste. Über den Verbindungsbroker werden auch gemeldete Warnungen angezeigt sowie die verschiedenen enthaltenen Tools zur Analyse und Diagnose von Netzwerkkomponenten gestartet. Auch der Zugriff über eine Weboberfläche (siehe Abbildung 6) ist möglich. Diese bietet zwar nicht alle Funktionen der Konsole, ermöglicht aber auch Überwachung, Berichterstattung sowie Diagnose und Analyse von unterwegs oder zuhause per Webbrowser [3].

Netcrunch bietet Multiplattform-Unterstützung und kann neben Windows-Betriebssystemen auch Linux (Kernel 2.4 oder neuer), Novell OES, Mac OS X, BSD und Netware-Systeme agentenlos überwachen. Netcrunch erkennt, ob es sich um eine virtuelle Maschine oder ein normales, hardwarebasiertes System handelt. Für die Überwachung der verschiedenen Betriebssysteme steht eine Vielzahl an Leistungszählern zur Überwachung der Auslastung bereit. Auch die Datenbank von Netcrunch ist eine Eigenentwicklung und benötigt keine zusätzlichen Lizenzen zur Speicherung von historischen Daten.

Der enthaltene Berichtsbetrachter von Netcrunch ist sehr übersichtlich und bietet detaillierte Informationen über die einzelnen Knoten. Jeden Bericht muss der Administrator manuell für die Datenerfassung aktivieren. Erstellte Berichte

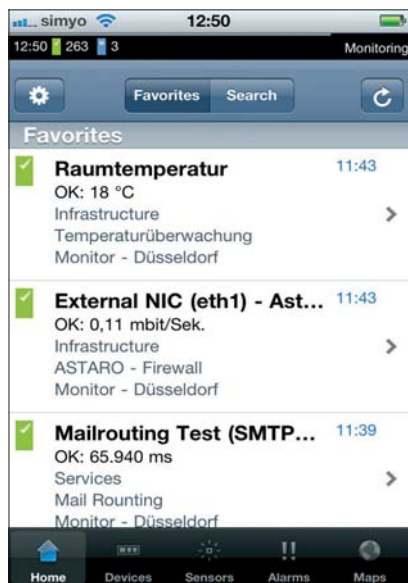


Abbildung 4: Kontrolle über das Netzwerk mit der iPhone App iPRTG.

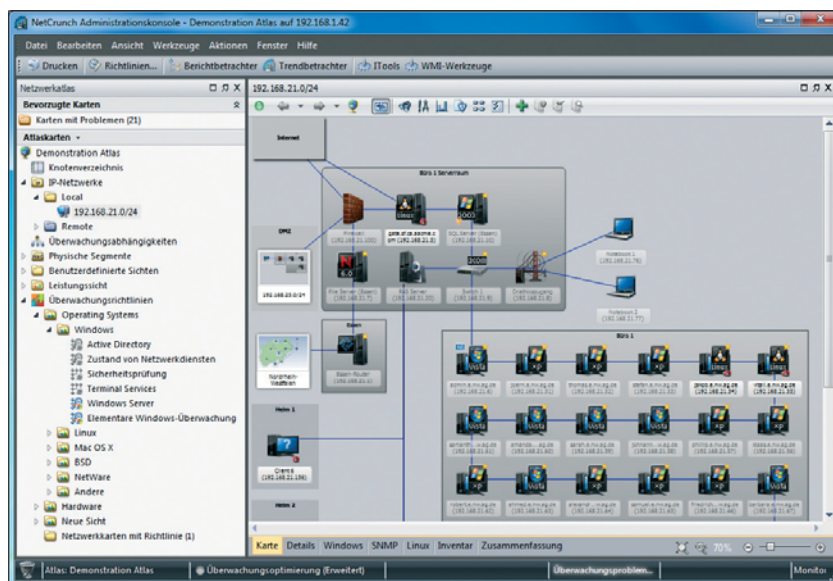


Abbildung 5: Die volle Kontrolle bietet Netcrunch über die Administrationskonsole für Windows.

kann er für eine spätere Verwendung abspeichern und auch ausdrucken. Für die planbare Langzeit-Trendanalyse generiert Netcrunch Berichte automatisch und verschickt sie im Anschluss per E-Mail an vorgegebene Mailadressen. Die auf WMI basierende Hardware- und Software-Inventarisierung für Windows zeigt sämtliche Informationen jedes Systems an und kann zusätzlich auch einen Gesamtüberblick über die Anzahl der im Unternehmen installierten Anwendungen geben beziehungsweise diese auch als Bericht darstellen.

SNMP-Import

Eines der größten Probleme einer Netzwerküberwachungslösung sind die verschiedenen SNMP-Bibliotheken der Netzwerkgeräte.

Da die meisten Hersteller eigene Bibliotheken für die Leistungszähler ihrer Netzwerkprodukte bereitstellen, muss eine gute Überwachungslösung in der Lage sein, neue Bibliotheken zu importieren beziehungsweise diese zu kompilieren. Netcrunch bietet mit dem SNMP MIB Compiler (siehe [Abbildung 7](#)) genau diese Möglichkeit und enthält bereits Bibliotheken von dutzenden von Herstellern.

Bei Netcrunch richtet sich die Lizenzierung nach der Anzahl der zu überwachenden Systeme, der Anzahl der Ver-

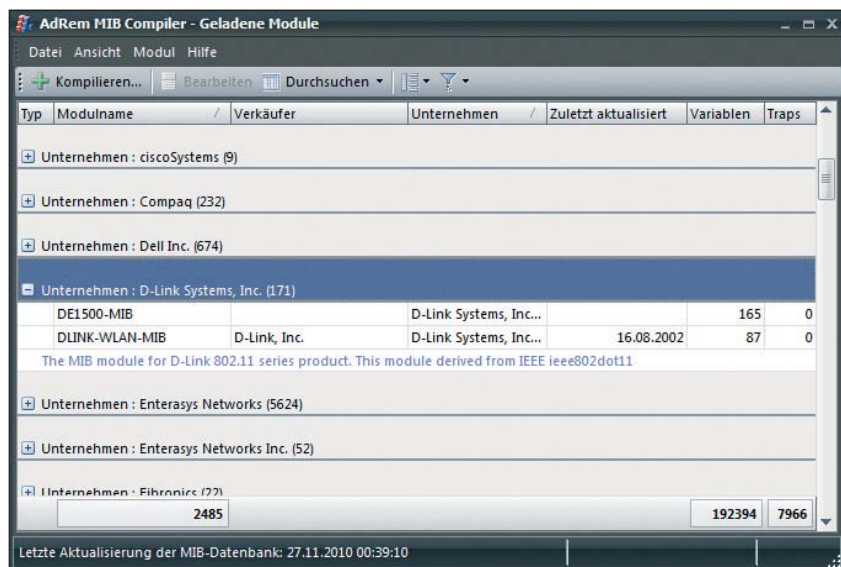


Abbildung 7: Neue MIB-Datenbanken kann der Administrator in Netcrunch mit dem MIB Compiler hinzufügen.

bindungen (zwei Administratoren = zwei Verbindungen) und danach, ob die integrierte Inventarisierung genutzt werden soll.

Fazit

Beide vorgestellten Überwachungslösungen für Windows bieten eine hohe Skalierbarkeit, arbeiten unabhängig von externen Datenbanken, bieten diverse Möglichkeiten der Benachrichtigung und können neben der Windows-Anwendung auch per Weboberfläche außerhalb des Unternehmens genutzt werden. (ofr) ■

Infos

- [1] Paessler PRTG Network Monitor 8: [\[http://www.de.paessler.com/prtg\]](http://www.de.paessler.com/prtg)
- [2] Paessler iPRTG für das iPhone: [\[http://www.de.paessler.com/iprtg\]](http://www.de.paessler.com/iprtg)
- [3] AdRem Software NetCrunch 6.5: [\[http://www.adremsoft.de/netcrunch\]](http://www.adremsoft.de/netcrunch)
- [4] Clickatell SMS Gateway: [\[http://www.clickatell.com\]](http://www.clickatell.com)

Der Autor

Björn Bürstinghaus ist IT-Manager bei der simyo GmbH in Düsseldorf. In seiner Freizeit betreibt er Bjoerns Windows Blog, ein Blog rund um Windows-Themen, zu finden unter [\[http://blog.buerstinghaus.net\]](http://blog.buerstinghaus.net).

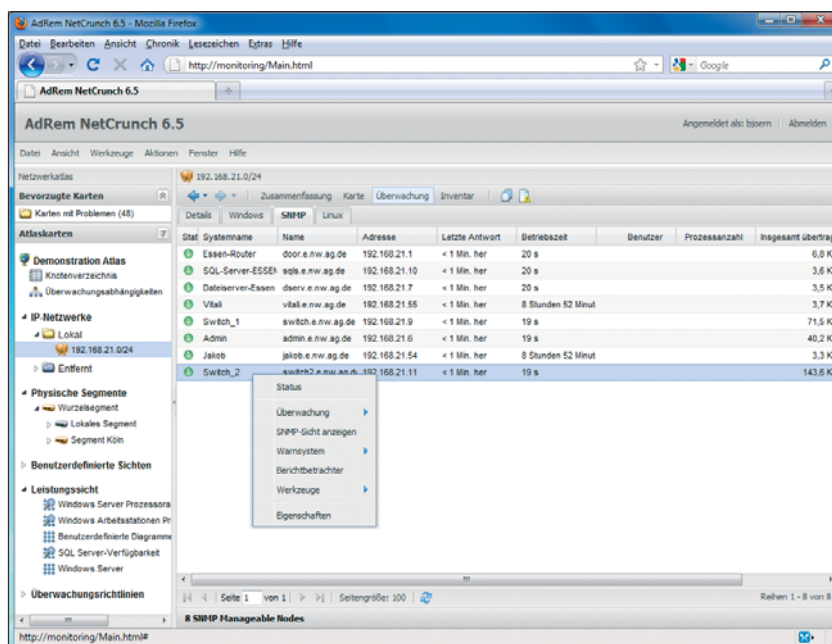


Abbildung 6: Die meisten Funktionen von Netcrunch sind auch über die Weboberfläche zugänglich.

Tipp: SMS-Versand mit Clickatell

Für den Versand von SMS-Benachrichtigungen gibt es mittlerweile eine Vielzahl an Anbietern, die verschiedene APIs für diesen Zweck bereitstellen. Clickatell bietet hierfür beispielsweise eine SMTP- und HTTP-Schnittstelle, über die Kurzmitteilungen an die gewünschte Mobilfunknummer weitergeleitet werden können. PRTG Network Monitor kann den Administrator mittels der HTTP API von Clickatell über Warnungen und Ausfälle per Kurzmitteilungen auf ein Mobiltelefon informieren. Für die Nutzung von Clickatell benötigt man sogenannte Credits, die im Vorfeld, ähnlich dem Prepaid-Konzept gekauft werden müssen. In Deutschland fallen pro versendeter Nachricht 2 Credits an, was pro Kurzmitteilung einem Preis von 8 Cent entspricht [4].