Zentraler Datensicherungsservice an der HU

Was soll in einem Heft über Sicherheit in Rechnernetzen der Datensicherungsservice? Es scheint so, als ob hier eventuell die Begriffe durcheinander gehen. Vielleicht kann man das Ganze auch andersherum betrachten: wenn ich ein sicheres Netz habe, kann ich auch eine „sichere“ Datensicherung darüber betreiben. Diese Begründung entspricht in etwa der Prüfung des Biologiestudenten, der sich auf die Würmer vorbereitete und über den Elefanten befragt wird ...

Viele Nutzer an der HU sichern ihre Daten, die bei der täglichen Arbeit an ihren Workstations anfallen, auf Bändern, Disketten oder ähnlichem. Dies ist der relativ gute Fall. Schlimmstenfalls werden sie kaum oder gar nicht gesichert, da man es vergißt oder nicht für nötig hält. Der zweite Fall wird nur so lange durchgehalten, bis ein Crash oder auch nur ein versehentliches Löschen die Daten auf ein „Nimmerwiedersehen“ verschwinden läßt. Dann wird für einige Zeit wieder der relativ gute Fall der eigenverantwortlichen Sicherung einsetzen, bis die allgemeine Bequemlichkeit sich wieder durchsetzt usw. Man kann also von einem Circulus vitiosus sprechen. Die beste Möglichkeit, aus diesem wieder herauszukommen, ist es, die Verantwortung dafür einem anderen zu überlassen, der die Sicherung der Daten als einen Dienst anbietet.

Als das Rechenzentrum sein Projekt SERVUZ auflegte, enthielt dieses auch einen Abschnitt Fileservice und darunter den Punkt Datensicherung. Der Gedanke war, daß das Rechenzentrum als zentrale Stelle der Universität umfangreiche Hard- und Software vorhalten kann, mit deren Hilfe die Sicherung für alle – oder zumindest für die wichtigsten Standorte – zentral geleistet werden kann.Voraussetzung ist natürlich ein gut funktionierendes Netz!

Für die Datensicherung wurde das durch UniTree realisierte hierarchische Speicherungsmodell (HSM) mit der Datensicherungssoftware SM-arch verbunden (Client Server Modell). SM-arch arbeitet mit einem zentralen Server – oder auch mehreren. Dieser sammelt von den Clients die zu sichernden Daten und speichert sie unter UniTree in Files, die ein dem „tar“ analoges Format besitzen. Da UniTree seine Files nach bestimmten Zeiten auf Band speichert – bei uns ist das der Taperoboter Metrum 600 mit VHS Bändern – wird die Bandverwaltung von UniTree übernommen. Gleichzeitig wird die Möglichkeit von UniTree genutzt, Files auf zwei unterschiedlichen Bändern zu speichern, so daß die gesicherten Daten auf zwei Kopien vorliegen.

Am Server von SM-arch können für folgende Probleme Konfigurationen vorgenommen werden:

1. Welche Clients sollen bedient werden?
2. Was soll von ihnen gesichert werden?
3. Wann und in welchem Umfang soll gesichert werden – alle Daten oder nur die seit dem letzten Mal geänderten?
4. Wie lange sollen die gesicherten Daten aufgehoben werden?
5. Sollen die zu sichernden Daten komprimiert werden?
6. Sollen die zu sichernden Daten verschlüsselt (Thema dieses Heftes!) über das Netz zum Server geschickt werden?

Diese Punkte bestimmen natürlich auch das technologische Regime, unter dem die Datensicherung abläuft.

Zu 1.

Um das Chaos im Netz durch die Datensicherung nicht ausufern zu lassen, wurden in der Einführungsphase nur ausgewählte Server – speziell die vom Rechenzentrum an Schwerpunktstandorten installierten dezentralen Fileserver – gesichert. In einem weiteren Schritt wurden Server von Arbeitsgruppen, d. h. Rechner, die größere Filesysteme an mehrere nachgelagerte Workstations exportieren, in den Sicherungsdienst aufgenommen. Dies erforderte einige Überzeugungsarbeit, da die jeweiligen Nutzer zuerst nicht unbedingt ihre Daten zentralisieren wollten. Diese Einstellung hat sich in der Zwischenzeit bei einer Reihe von ihnen geändert, da sie die Vorteile der zentralen Datensicherung – und vor allem auch des Restaurierens von Files bei einigen „Unfällen“ – am eigenen Leibe spürten. Zur Zeit werden etwa 30 Server gesichert, an denen jeweils eine Reihe von Workstations „hängt“. Dies ist in diesem Jahr eine Zunahme um etwa 50%.

Zu 2.

Es werden von uns nie die gesamten Server gesichert, sondern nur die Filesysteme mit Nutzerdaten und ausgewählte Verzeichnisse wie z. B. /etc. Dies soll unter anderem verhindern, daß zu viel Ballast über das Netz „geschaufelt“ wird.

Zu 3. und 4.

Die Datensicherung läuft, um das Netz nicht zu sehr zu belasten, in den nicht so arbeitsintensiven Zeiten ab. Alle vier Wochen werden die gesamten Daten über das Wochenende gesichert. Täglich werden zwischen 0:00 Uhr und 6:00 Uhr nur die geänderten Daten gerettet. Aufgehoben werden die gesicherten Daten vier Wochen, d. h. ein Nutzer kann seine Daten nach vier Wochen immer noch restaurieren lassen. Diese Zeit ist ein Kompromiß zwischen dem Sicherheitsbedürfnis der Nutzer und dem Verhindern einer Überflutung des SM-arch-Servers mit Daten. Auch wenn wir durch UniTree einen beinah unbegrenzten Speicherhintergrund haben, gibt es doch äußere Schranken, wie z. B. die Anzahl der Bänder im Roboter.

Zu 5.
Das Komprimieren der zu sichernden Daten belastet zwar kurzfristig den Client, hat aber den Vorteil, daß das Netz entlastet wird. Zur Zeit ist der Umfang an gesicherten Daten ca. 200 GB; dies entspricht etwa 330 GB zu sichernde Daten.

Zu 6.
SM-arch bietet die Möglichkeit, die Daten auf dem Client zu verschlüsseln. Auch kann dem Client die IP-Adresse des Servers, der die Zugriffsberechtigung über eine Port-to-Port Verbindung hat, beim Start mitgegeben werden. Dies sichert die Daten weitgehend vor einem unberechtigten Zugriff ab.

Die bisherigen Erfahrungen zeigen, daß die Datensicherung stabil läuft. Auch das Restaurieren von Files läuft zur allgemeinen Zufriedenheit. Der Nutzer schickt per E-Mail an *restore@rz.hu-berlin.de* seinen Wunsch. Er kann dabei angeben, welches File oder Verzeichnis er benötigt; gibt er nichts weiter an, erhält er den aktuellen Stand. Andernfalls kann er das Datum angeben, von dem er das File benötigt. Des weiteren kann er auch angeben, an welche Stelle seines Filesystems er das File gestellt bekommen will.

Durch die Menge der Daten, die durch SM-arch gesichert werden, ist es ungünstig, das Restaurieren in die Hand des Nutzers zu geben, da oftmals sehr viele und große SM-arch-Files aus UniTree in den DiskCache geholt werden müssen –- ein Vorgang, der oftmals viel Zeit benötigt, da hier die Begrenztheit der Anzahl an Laufwerken im Roboter zum Tragen kommt. UniTree wird ja nicht nur von der Datensicherung benutzt, sondern auch von vielen Nutzern, die direkt in den DiskCache von UniTree schreiben. Diesen Vorgang kann man am besten steuern, wenn man den Überblick über das gesamte System besitzt und dadurch auch bestimmen kann, was benötigt wird.

Im Rückblick kann man sagen, daß bisher Restaurierungsanforderungen in erster Linie wegen versehentlichen Löschens von Daten gestellt wurden. Dies geht im allgemeinen sehr rasch. Werden ganze Filesysteme von 8 GB und mehr wieder benötigt, dauert dies länger; wurde aber auch schon mit Erfolg durchgeführt.

Wenn Arbeitsgruppen daran interessiert sind, ihre zentralen Rechner mit in den Datensicherungsdienst der HU einzubinden, können sie sich auch an die obengenannte E-Mail-Adresse *restore@rz.hu-berlin.de* wenden.

Christoph Weickmann