Konzepte und Werkzeuge der Datensicherung im Rahmen von Forschungsprojekten

Alexander Wehrum[1], Claus Mückschel[2], Cornelia Weist[2]

[1] Connecta AG Wiesbaden
[2] Biometrie und Populationsgenetik, Justus-Liebig-Universität
Heinrich-Buff-Ring 26-32
D-35392 Giessen
claus.mueckschel@agrar.uni-giessen.de

Abstract: Ein zentrales Datenmanagement hat u.a. die Aufgabe, effektive Mechanismen der Datenhaltung und –sicherung unter Einhaltung der drei Grundwerte der IT-Sicherheit - Verfügbarkeit, Vertraulichkeit und Integrität - zu gewährleisten. Da ein Datenverlust die Projektziele gefährden kann, sind Datensicherungs- und Wiederherstellungskonzepte für jedes Forschungsprojekt obligatorisch. Der Beitrag fasst allgemeine Grundsätze der Datensicherung und -wiederherstellung zusammen und beschreibt sowie bewertet exemplarisch ein Datensicherungswerkzeug des Sonderforschungsbereichs (SFB) 299.

1 Einleitung

Informationen werden zunehmend digital verwaltet, so dass die ständige Verfügbarkeit und Sicherheit digitaler Daten eine immer größere Rolle spielt. Die Folgen eines potentiellen Datenverlustes können über die Nichterfüllung der Projektziele bis hin zum vorzeitigen Ende des Projektes oder finanziellen Ruin eines Unternehmens reichen. Laut einer 2005 durchgeführten Umfrage der Connecta AG [CAG05] bei Unternehmen aus dem Rhein-Main Gebiet sind die häufigsten Ursachen für einen Datenverlust auf Hardwaredefekte (59%), Viren (21%), Fehlbedienung (10%), höhere Gewalt (Wasser, Feuer; 2%), Einbruch/Diebstahl (2%) sowie sonstige (6%) zurückzuführen. Die Gewährleistung einer entsprechenden Datensicherheit ist folglich für Forschung und Praxis obligatorisch.

2 Grundlagen der Datensicherung

2.1 Klassifikation und Schutzbedürfnis von Daten

Vor der Planung eines Sicherungskonzeptes müssen vorhandene Daten in Abhängigkeit ihres Bedarfes an Sicherungsmaßnahmen klassifiziert werden. Das Bundesamt für Sicherheit in der Informationstechnik nennt drei, auf dem finanziellen Schaden durch einen Datenverlust basierende Klassen [BSI99]. Klasse 1 (niedrig bis mittel) ist durch Datenverluste mit verkraftbaren Schäden definiert. In Klasse 2 (hoch) führt ein Schaden zu

spürbaren Einbußen, die Fortführung des Betriebes ist jedoch nicht gefährdet. Ein Datenverlust der Klasse 3 (sehr hoch) resultiert dagegen im Stillstand des Unternehmensbetriebes. Dieser Ansatz lässt sich analog auf Forschungsprojekte übertragen. Zusätzlich ist zu berücksichtigen, in welchem zeitlichen Rahmen die Daten und das IT-System nach einer Störung wieder verfügbar sein müssen. Für Forschungsprojekte wie den SFB 299 ist eine Zeitspanne von 24 Stunden ausreichend, andere Projekte oder Wirtschaftsbetriebe sind oft auf kürzere Zeiträume angewiesen. Die Wahl des optimalen Sicherungsansatzes ist wichtig, da auch die Datensicherung ökonomischen Prinzipien unterliegt. Abbildung 1 verdeutlicht den exponentiellen Kostenverlauf bei steigendem Sicherheitsaufwand. Mit 20% an Investitionskosten in die IT-Sicherheit wird ein theoretischer Sicherheitsgrad von 80% erreicht.

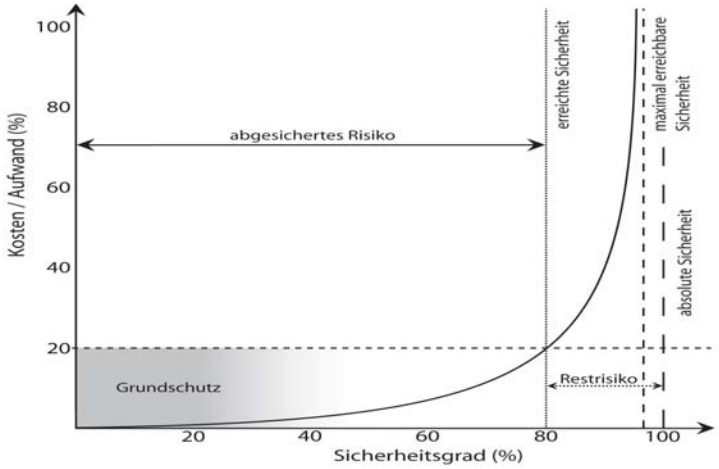

Abbildung 1: Kostenverlauf bei steigendem Sicherheitsgrad [in Anlehnung an MC 03]

2.2 Methoden der Datensicherung

Die technischen Möglichkeiten erlauben eine stets vollständige, von evtl. Veränderungen der Datenstruktur unabhängige Datensicherung. Da dies jedoch sehr speicher- und zeitintensiv ist, sollten auch alternative Methoden in Betracht gezogen werden. Die differenzielle Sicherung speichert nur jene Daten, die seit der letzen Sicherung modifiziert wurden. Noch weiter verfeinert ist das inkrementelle Verfahren, das eine historische Datensicherung ermöglicht. Durch Anpassung der Anzahl von Sicherungsständen lassen sich beliebige Datenzustände verschiedener Tage oder Wochen wiederherstellen.

2.3 Datenauswahl

Unabhängig von der Methode der Datensicherung muss über die Art der zu sichernden Daten entschieden werden. Bei einer Systemsicherung wird das Betriebsystem nebst installierter Programme, Einstellungen und Anwenderdaten gesichert. Alternativ kann sich die Sicherung ausschließlich auf Anwendungsdaten der Benutzer beschränken.

Nach einem möglichen Datenverlust müssen das Betriebssystem und die Programme neu installiert und anschließend die Anwendungsdaten rückgesichert werden. Letztere Methode ist zwar zeitaufwändiger, jedoch deutlich weniger ressourcenintensiv.

2.4 Sicherungsmedien

Zu den optischen Medien zählt die CD-R. Ihre kratzeranfällige Kunststoffbeschichtung und UV-empfindliche organische Aufzeichnungsschicht bedingen eine gewisse Fehleranfälligkeit. Die robustere, mit einer metallischen Schicht versehene CD-RW ist deutlich UV-beständiger. Für Datensicherungszwecke gänzlich ungeeignet sind die für die Aufzeichnung von Videomaterial konzipierten DVD+/-, da bereits der Brennvorgang oftmals nicht fehlerfrei verläuft. Dies kann, insbesondere bei der Sicherung in eine einzelne Sicherungsdatei (Archiv) fatale Folgen haben. Die DVD-RAM erscheint geeigneter, da zusätzliche Sicherheit durch eine Sperrung fehlerhafter Sektoren erlangt wird.

Magnetbänder sind im professionellen Bereich noch immer weit verbreitet. Der sequenzielle Auslesevorgang bedingt jedoch extrem langsame Zugriffszeiten. Zudem sind sie überaus empfindlich und werden beim Auslesen und Schreiben mechanisch stark beansprucht. Die sukzessive Abschwächung ihrer magnetischen Aufladung führt im Endstadium zur Unlesbarkeit der Bänder. Des Weiteren haben die Laufwerke eine begrenzte Haltbarkeit, die sich bei regelmäßigem Einsatz auf ca. 3-5 Jahre beschränkt. Bei längerer Lagerzeit sind Magnetbänder regelmäßig umzuspulen, um ein Verkleben und damit Datenverluste zu verhindern.

Aufgrund rasant gestiegener Kapazitäten und zunehmender Ausfallsicherheit von Festplatten bieten auch diese sich gut als alternatives Backupmedium an. Befinden sich Backup- und Hauptfestplatte im gleichen System, sind folglich Haupt- und Sicherungsmedium denselben potentiellen Schadeinflüssen ausgesetzt. Eine örtliche Trennung der Festplatten ist daher erforderlich. Durch die hohe Herstellungspräzision und die nahezu identische Fertigung dieser Medien besteht bei Verwendung zweier Festplatten aus gleicher Serie desselben Herstellers das Risiko, dass Haupt- und Sicherungsmedium zu ähnlichen Zeitpunkten einen Defekt aufweisen. Daher sollten stets Festplatten von verschiedenen Herstellern oder mindestens aus verschiedenen Fertigungsserien gewählt werden.

2.5 Erstellung eines Datensicherungskonzeptes

Ein Datensicherungskonzept legt die Auswahl der Daten, den Zeitpunkt der Sicherung, die Methoden und Werkzeuge sowie die Zuständigkeiten für einzelne Schritte fest. Beispielsweise werden Datenbestände täglich ohne manuellen Anstoß auf DLT-Tapes gesichert. Die Prüfung der Datensicherung obliegt der verantwortlichen Person, erfolgt in definierten Intervallen und wird dokumentiert. Die Erprobung der Rücksicherung in regelmäßigen Abständen liefert wertvolle Erfahrungen für zukünftige Sicherungsläufe. Dabei sind der Ablauf der Rücksicherung sowie die exakte Zuordnung der Daten zu den Systemen und Servern nach einem Komplettausfall sichergestellt.

3 Werkzeuge der Datensicherung im SFB 299

Operative Daten der Arbeitsplatzrechner des SFB 299 werden im Rahmen eines Testverfahrens mit dem Sicherungswerkzeug Amanda (www.amanda.org) gesichert. Die Daten der Clients werden über ein Netzwerk in konfigurierbaren Methoden und Intervallen auf dem Datensicherungsserver abgelegt. Dieses Client/Server Prinzip ist unter Linux und Windows gleichermaßen einsetzbar. Die Konfiguration des auf dem zu sichernden Client installierten Softwaretools wird auf dem Sicherungsserver vorgenommen.

Ein Teil der Daten wird zudem mit dem Open-Source Werkzeug Rsync gesichert. Es dient der effektiven Spiegelung von Daten in Linux-Umgebungen, wobei kein Archiv, sondern eine exakte Kopie der Datenbestände erzeugt wird. Zeitsparend wirkt sich die Auswertung von Checksummen (mathematisches Verfahren zur Integritätsprüfung) zweier Dateien anstelle ihrer Zeitstempel aus, da somit nur modifizierte Bereiche kopiert werden. Für Datensicherungen auf einer physikalischen Platte stehen „Hardlinks“ zur Verfügung. Ist eine Datei seit der letzen Spiegelung unverändert, so erzeugt der „Hardlink“ einen Verzeichniseintrag auf die ursprüngliche Datei. Diese Methode dient der einfachen, aber effektiven historischen Sicherung bei optimierter Speichernutzung. Da alle Sicherungsversionen in einem File-System abgelegt werden müssen, resultiert ein Defekt der Festplatte jedoch im Verlust aller Versionen, so dass ein RAID-System obligatorisch ist. Rsync wird über die Kommandozeile oder Remote-Shell konfiguriert. Die Übertragung zwischen Quelle und Ziel kann über ssh getunnelt werden.

4 Fazit und Ausblick

Ein allgemein gültiges Konzept der Datensicherung ist nicht existent. In Abhängigkeit der Schutzbedürftigkeit von Daten, den (zeitlichen) Ansprüchen an das Sicherungssystem und der finanziellen Situation stehen verschiedene Methoden und Werkzeuge zur Realisierung eines angemessenen Maßes an Sicherheit zur Verfügung.

Trotz aufgeführter Nachteile stellt Rsync ein zuverlässiges und effektives Werkzeug zur Datensicherung dar. Zukünftig soll das System weiter ausgebaut werden und über die Anzahl der historischen Sicherungen sowie den Grad der möglichen Zurückverfolgung von Datenänderungen autonom entscheiden können. Zudem soll nach dem Testbetrieb die Software Amanda eingeführt werden, um den Sicherheitsgrad bei überschaubaren Aufwendungen weiter zu erhöhen.

5 Literatur

[BSI99] BSI Bundesamt für Sicherheit in der Informationstechnik, „IT-Grundschutzhandbuch“ Vol.3, Bundesanzeiger Verlag, Bonn, 1999

[CAG05] Umfrage Datensicherheit http://www.connecta.ag/presse-aktuelles/publikationen/umfrage-datensicherheit.html

[MC03] M. a Campo (2003): Management der IT-Sicherheit. Interest Verlag, Kissing