Grundwissen zur Datensicherheit

Mai 2020

Inhaltsverzeichnis

1 Einführung

1.1 Wichtige Prinzipien

Bei der gewährleistung der Datensicherheit spielen vier grundlegende Prinzipien eine maßgebliche Rolle. Das sind die *Redundanz*, die *Integrität*, die *Privatheit* und die *Authentizität*.

Redundanz. Datenträger haben eine begrenzte Haltbarkeit, können beschädigt werden oder verloren gehen. Aus diesem Grund möchte man von wichtigen Daten mehrere Kopien lagern, möglichst an unterschiedlichen Orten. Dies nennt man Redundanz der Daten.

Integrität. Es kann vorkommen, dass gespeicherte Daten auf einem Datenträger beschädigt werden. Oder dass jemand oder ein Schadprogramm unbemerkte Veränderungen an den Daten vornimmt. Den Schutz davor bezeichnet man als Bewahrung der Integrität.

Privatheit. Daten können von jemanden oder einem Schadprogramm unerlaubt gelesen werden. Den Schutz davor bezeichnet man als Bewahrung der Privatheit.

Authentizität. Jemand kann behaupten, Author von bestimmten Daten zu sein oder nicht zu sein, oder bezichtigt jemanden, Author zu sein oder nicht zu sein. Den Schutz vor absichtlichen Fehlinformationen bezeichnet man als Bewahrung der Authentizität.

2 Datensicherung

2.1 Verzeichnisstruktur

Daten sollten ihrer Speichergröße nach in unterschiedliche Ordner eingeteilt werden. Z. B. sind Ordner wie *Texte*, *Bilder*, *Videos* sinnvoll. Dies bietet mehrere Vorteile. Erstens brauchen Texte[1] vergleichsweise verschwindend gering Speicherplatz und lassen sich sogar noch gut komprimieren. Aus diesem Grund können sehr oft Sicherheitskopien von Texten gemacht werden, ohne jemals überfüllten Speicher befürchten zu müssen. Zweitens lässt sich mit Programmen wie `grep` bzw. `findstr` eine Volltextsuche auf das Verzeichnis anwenden, wobei große Binärdateien aber stören würden.

2.2 Versionsgeschichte

Bei Texten in Bearbeitung sollte man nicht nur eine Kopie des aktuellen Textes zu speichern, sondern auch dessen frühere Bearbeitungszustände, dies nennt man *Versionsgeschichte*. Warum ist das so wichtig? Nun, es können immer mal Fehler auftauchen, die zu einem unbemerkten Verlust von Textpassagen führen. Oder aber, jemand oder ein Schadprogramm greift unerlaubt auf den Text zu und nimmmt subtile Veränderungen vor. Hat man die Versionsgeschichte zu einem Text, lassen sich die Versionen des Textes im Nachhinein zur Sicherheit vergleichen.

2.3 Deduplikation

Mit solchen Versionsgeschichten ergibt sich aber noch ein technisches Problem. Große Dateien wie Bilder oder Videos würde man immer wieder speichern, auch wenn sich diese nicht mehr verändern. Dies bezeichnet man als ungewollte Duplikation von Daten. Zwar wird die Redundanz der Daten damit beträchtlich erhöht, eigentlich etwas gutes, aber dies findet in einem unkontrollierten Ausmaß statt und sorgt schnell für über jedes Maß steigenden Datenverbrauch.

Zur Lösung dieses Problem bieten sich zwei Ansätze an. Zum einen unterlässt man es, diese Daten thematisch zu ordnen. Stattdessen bietet sich eine Ordnung nach dem Datum an, z. B. nach dem Jahr oder dem Monat des Jahres. Hat man diese Daten einmal archiviert, braucht man sie nicht ständig erneut speichern. Zur thematischen Sortierung schreibt man stattdessen ein Verzeichnis, das nur auf die Dateien verweist, sie aber nicht selbst speichert.

[1] Quelltexte, LaTeX, HTML anstelle von Exporten wie PDF

Datenträger	Einzelpreis	Speicherplatz	Kosten je GB
Cloud	0,00 €	1 GB	0,00 €
BD	0,75 €	25 GB	0,03 €
HDD	53,00 €	1 TB	0,05 €
DVD-R	0,32 €	4,7 GB	0,07 €
DVD-RW	1,27 €	4,7 GB	0,27 €
CD-R	0,28 €	700 MB	0,40 €
USB-Stick	7,00 €	16 GB	0,44 €
SD-Karte	8,99 €	16 GB	0,56 €
CD-RW	0,70 €	700 MB	0,99 €

Tabelle 1: Preise für Datenspeicher
Quelle: Kizoa, 2017.

Zum zweiten gibt es auch extra Backup-Software, die in der Lage ist, Daten bei der Sicherung zu deduplizieren. Z. B. besitzt jede Datei einen Hashwert. Hat sich dieser nicht verändert, braucht die Software nur auf den Inhalt der Datei zu verweisen. Ob die Software das kann, lässt sich auch leicht experimentell überprüfen.

2.4 Speichermedien

Tabelle 1 zeigt die Kosten für unterschiedliche Datenspeicher. Hierbei ist zunächst zu beachten, dass Datenträger durch Beschädigung ab und zu unbrauchbar werden. Solches kommt auch bei allergrößter Sorgfalt vor. Aus diesem Grund möchte man von wichtigen Daten immer mehrere Replikate haben. Es wäre also gefährlich, alle Daten auf eine einzelne HDD zu speichern. Ich würde empfehlen, wichtige Daten redundant auf USB-Sticks, DVD-R und HDD zu speichern. Man sollte auch Replikate an einen zweiten Ort bringen, damit diese im Falle eines Brands erhalten bleiben.

Bei hohen Speichergrößen wird man entsprechend mehrere redundante Replikate von BD bis HDD wählen.

Grundsätzlich gilt: Ein Sicherungsmedium darf nicht mit einem Rechner verbunden sein. Andernfalls könnte es bei einer Überspannung, wie sie durch einen Blitzeinschlag versursacht werden kann, beschädigt werden oder Schadsoftware könnte sofort auf dieses zugreifen.

2.5 Fehlerkorrektur

Hochwertige Backup-Systeme sind mit Mitteln zur Fehlerkorrektur ausgestattet. Solche Verfahren erhöhen die Redundanz der Daten geringfügig und erlauben dadurch eine höhere Fehlerresistenz. Ein historisch wichtiges Verfahren ist z. B. die Reed-Solomon-Fehlerkorrektur.

Fehlerkorrektur wird gemeinsam mit kryptografischen Hashfunktionen benutzt, um Beschädigungen sicher identifizieren zu können. Sollte eine Datei nicht wieder korrekt hergestellt werden können, wird das Backup-System diesen Fehler melden. In diesem Fall muss man zur Wiederherstellung ein Replikat benutzen. Sollte das Replikat ebenfalls zu stark beschädigt sein, muss man die Replikate vereinigen. Sollte dies fehlschlagen, muss man zur Vereinigung weitere Replikate hinzuziehen.

3 Kryptografische Methoden

3.1 Hashfunktionen

Zur Gewährleistung der Integrität bedient man sich kryptografischen Hashfunktionen. Jede Datei besitzt einen eindeutigen Fingerabdruck, den man Hashwert nennt. Zu einer Datei lässt sich dieser Hashwert sehr schnell berechnen. Allerdings ist es aufgrund astronomisch großen Rechenaufwands faktisch unmöglich, zwei Dateien mit demselben Hashwert zu finden (ein sogenannter *Kollisionsangriff*), geschweige denn zu einem gegebenen Hashwert eine Datei zu finden (ein sogenannter *Urbildangriff*).

Das empfohlene moderne Verfahren zur Erzeugung eines solchen Hashwerts ist SHA-3-256. Auch SHA-2-256 kann man noch benutzen. Das ältere Verfahren MD5 konnte man als unsicher nachweisen, da Kollisionsangriffe gefunden wurden. Allerdings ist das für uns nicht besonders problematisch, da es auch für MD5 noch keinen praktischen Urbildangriff gibt.

3.2 Verschlüsselung

Die Privatheit von Daten lässt sich mit Verschlüsselung gewährleisten. Mittels eines Schlüssels (eine Zeichenkette) wandelt man hierbei Klardaten in ein Chiffrat um. Die Klardaten lassen sich aus dem Chiffrat nur wieder zurückgewinnen, wenn der Schlüssel bekannt ist.

Moderne Verschlüsselung gibt es in zwei Konzepten, der *symmetrischen* und der *asymmetrischen* Verschlüsselung. An dieser Stelle ist zunächst nur die symmetrische Verschlüsselung von Bedeutung.

Zwei der empfohlenen modernen Verfahren zur Verschlüsselung sind die Blockverschlüsselung AES und die Stromverschlüsselung ChaCha20. Eine Blockverschlüsselung lässt sich in verschiedenen Betriesmodi betreiben, von denen für die Verschlüsselung von Archivdaten der Counter-Mode (CTR) am sichersten zu benutzen ist. Der Counter-Mode wandelt die Blockchiffre in eine Stromchiffre um. Bei einer solchen werden

die Klardaten mit einem durch den Schlüssel erzeugten Zufallsdatenstrom bitweise XOR-Verknüpft. Diese Technik ist daher am sichersten, weil Beschädigungen von Bits im Chiffrat nur beschädigte Bits an der gleichen Stelle in den Klardaten nach sich ziehen, womit ein Zusammenspiel mit Algorithmen zur Fehlererkennung (z. B. CRC-32) bzw. Fehlerkorrektur ermöglicht wird.

Einige Betriebsmodi und auch eine reine Stromchiffre gewährleisten leider nicht die Integrität. Dies lässt sich experimentell überprüfen, indem man mit einem Hexeditor Bytes einer verschlüsselten Datei manipuliert, – bei der Entschlüsselung wird man keine Fehlermeldung erhalten. Zur Befreiung von dieser Ungenügsamkeit wurden aufbauende Techniken wie der Galois-Counter-Mode (GCM) und das Poly1305-Verfahren geschaffen.

Verschlüsselung ist kein Kinderspielzeug. Kommt der Schlüssel abhanden, ist die Entschlüsselung des Chiffrats unmöglich. Es gibt keinen Schlüsseldienst. Man kann auch nicht irgendwo seinen Personalausweis vorlegen und bekommt dann einen neuen Schlüssel zugeschickt. Mit dem Schlüssel gehen auch die Klardaten für immer verloren. Aus diesem Grund sollte man, falls möglich, von den Klardaten besser irgendwo unverschlüsselte Kopien aufbewahren. Auch sollte man die Kopien der Klardaten nicht wegwerfen, bevor man sich sicher ist, den Schlüssel korrekt eingegeben zu haben.