Digitale Magazine ohne eigenen Speicher

Wie man die Integrität "fremdgespeicherter" Archivalien sicherstellen kann

Stefan Lucks

Professur für Mediensicherheit

Bauhaus-Universität Weimar

12. März 2013

Original & Fälschung Kopie

Physikalisches Original

- Kopie $\neq$ Original
- Verändern schwierig

Original & Fälschung Kopie

Physikalisches Original

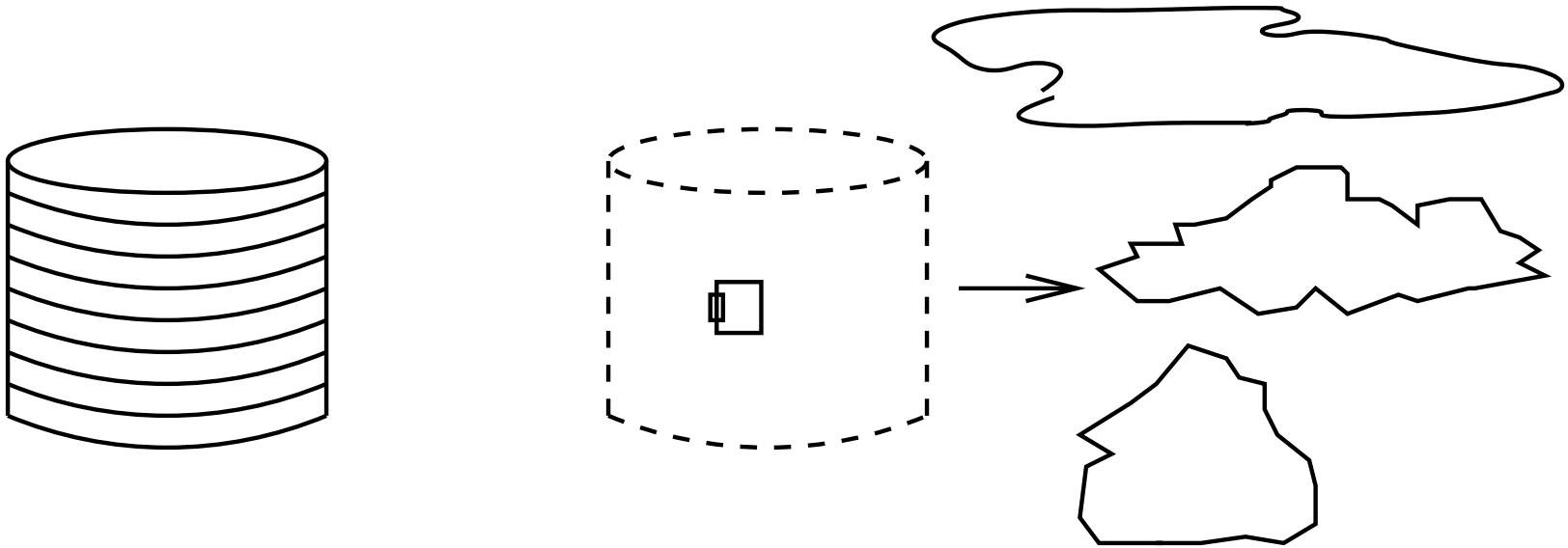
- Kopie $\neq$ Original
- Verändern schwierig

Digitales Original

- Original=Kopie
- leicht zu verändern

Digitale Magazine: Wohin mit den Daten?

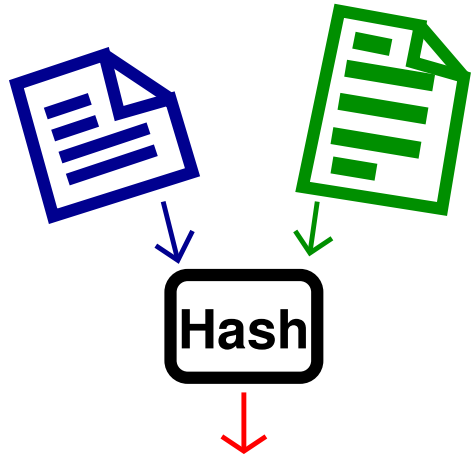
Eigener Speicher **virtueller Speicher** **bei Dienstleister**



Wenn Ihr Magazin tatsächlich bei einem Dienstleister gespeichert wird – woher wissen Sie dann, ob Anfragen mit den “echten” Daten beantwortet werden?

Kryptographische Hashfunktionen

Eine **Hashfunktion** h errechnet einen kleinen *kleinen* Fingerabdruck von einem potentiell *großes* Dokument.

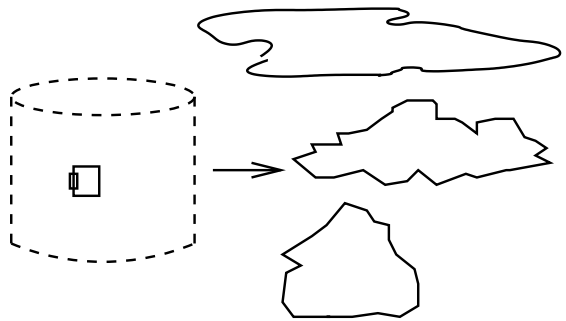


- Kollisionen, also Inputs $X \neq Y$ mit $h(X) = h(Y)$ sind katastrophal.
- Sogar noch schlimmer ist es, wenn man Urbilder finden kann: Gegeben Z, finde X mit $h(X) = Z$.

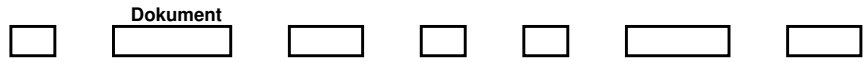
Sicherheit:
Es ist *praktisch unmöglich*, Kollisionen oder gar Urbilder zu finden!

Idee

- ein winziger eigener “Speicher”
  (USB-Stick mehr als ausreichend)
- enthält Hash-Wert des Digitalen Magazins
- Problem:
  - zur Verifikation bräuchte man alle Daten des Magazins
  - wir wollen/brauchen aber nur *einzelne Dokumente*
    (AIPs oder einzelne Dateien)

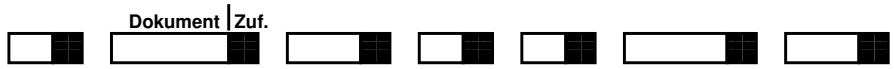


Idee (2)



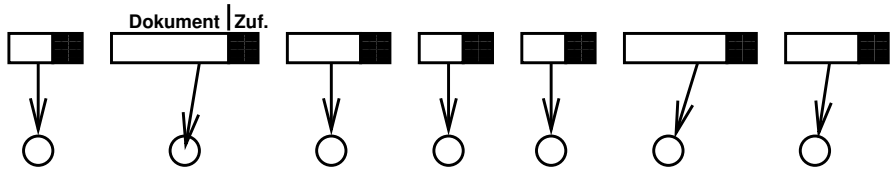
1. Dokumente

Idee (2)



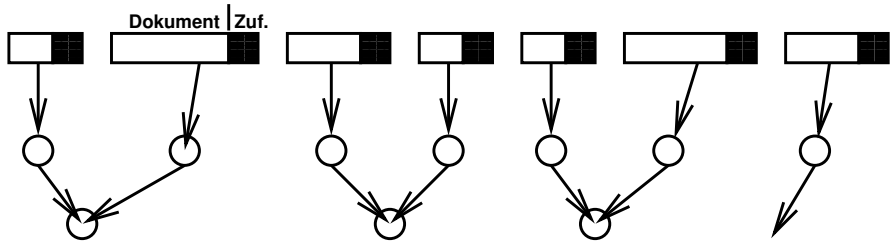
1. Dokumente und kurzen Zufallswert

Idee (2)



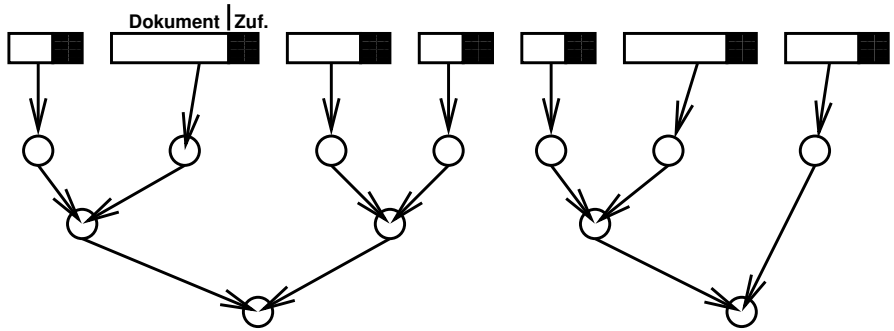
1. Dokumente und kurzen Zufallswert
2. berechne Hash (Dokument, Zufall)

Idee (2)



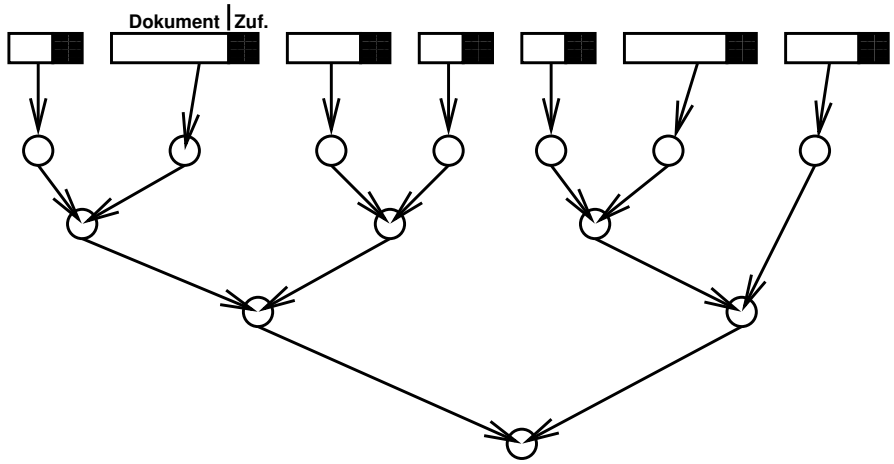
1. Dokumente und kurzen Zufallswert
2. berechne Hash (Dokument, Zufall)
3. iterativ: Hashen von Paaren von Hash-Werten

Idee (2)



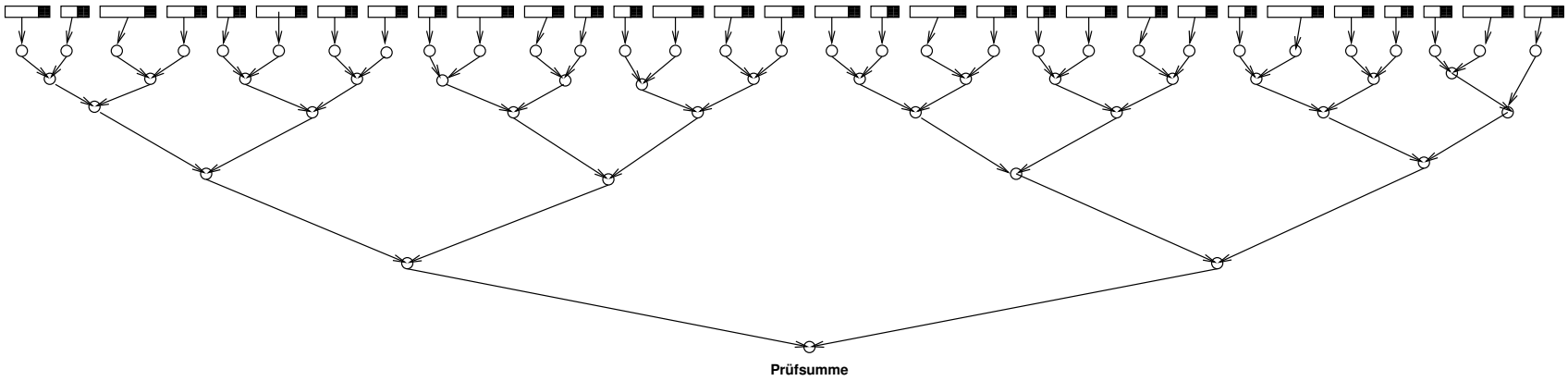
1. Dokumente und kurzen Zufallswert
2. berechne Hash (Dokument, Zufall)
3. iterativ: Hashen von Paaren von Hash-Werten

Idee (2)



1. Dokumente und kurzen Zufallswert
2. berechne Hash (Dokument, Zufall)
3. iterativ: Hashen von Paaren von Hash-Werten

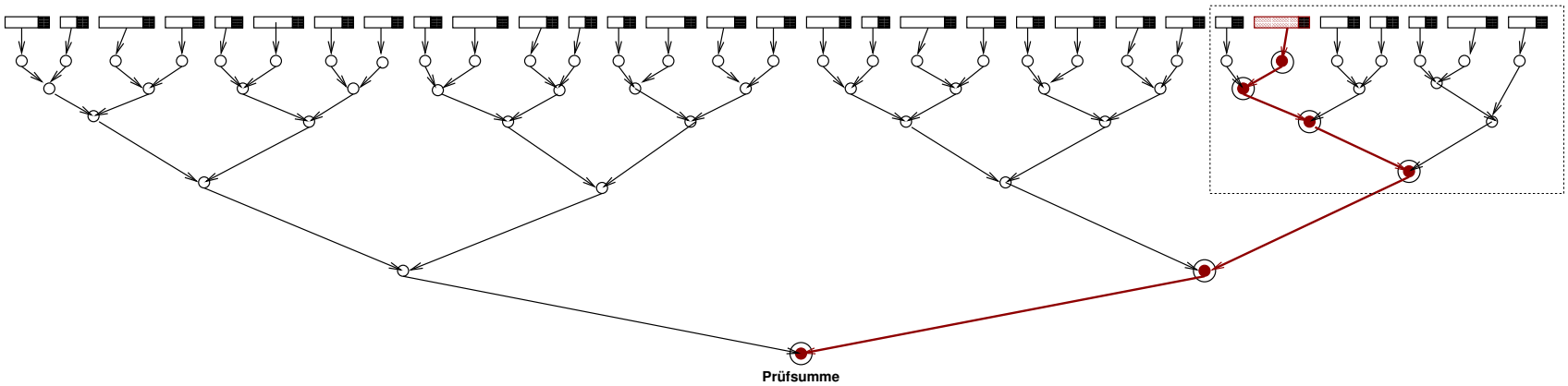
Hash-Bäume



Baum-artige Datenstruktur

- Dokumente = Blätter
- Prüfsumme = Wurzel
- Höhe: $L = \log_2(\#\text{Dokumente})$

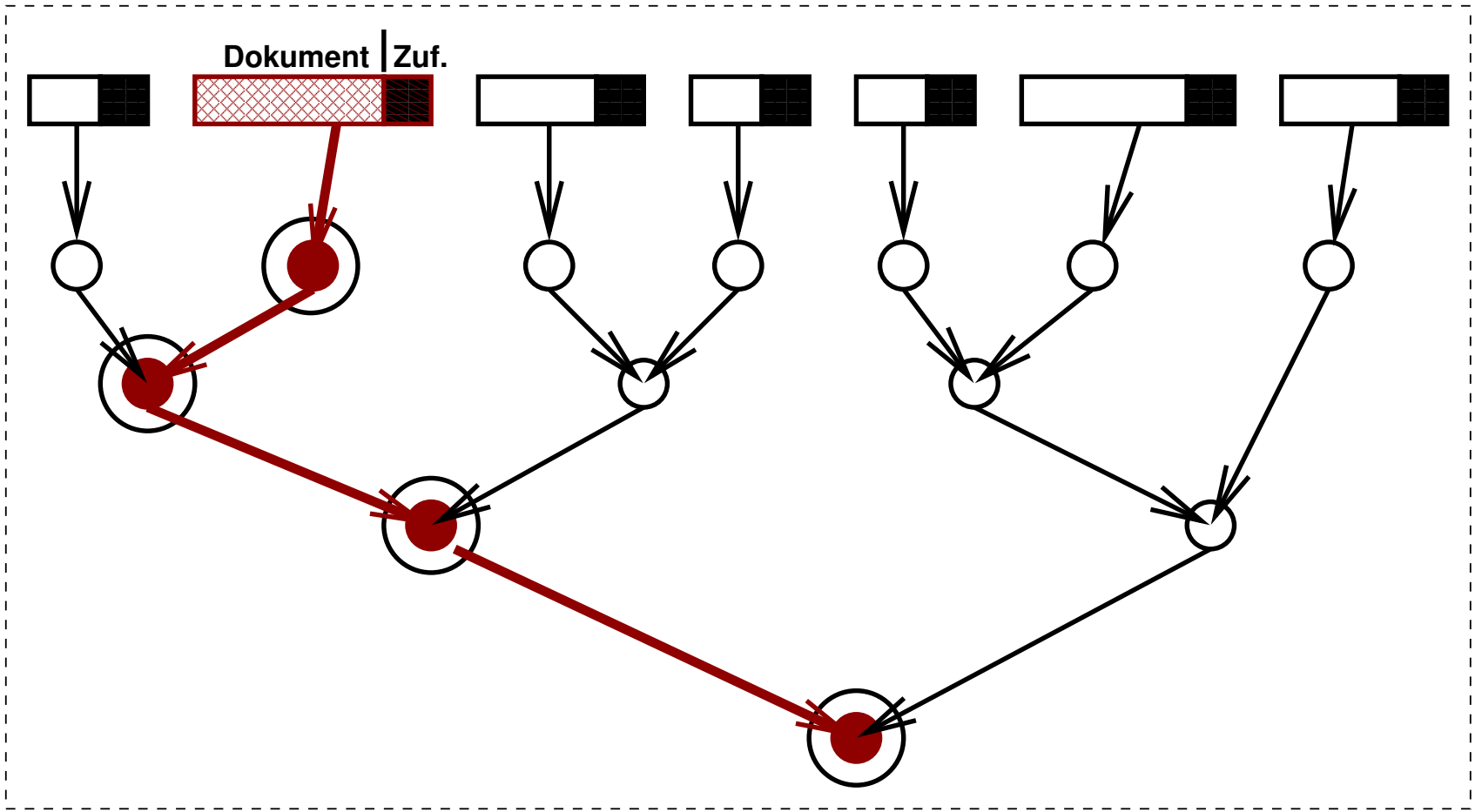
Hash-Bäume



Baum-artige Datenstruktur

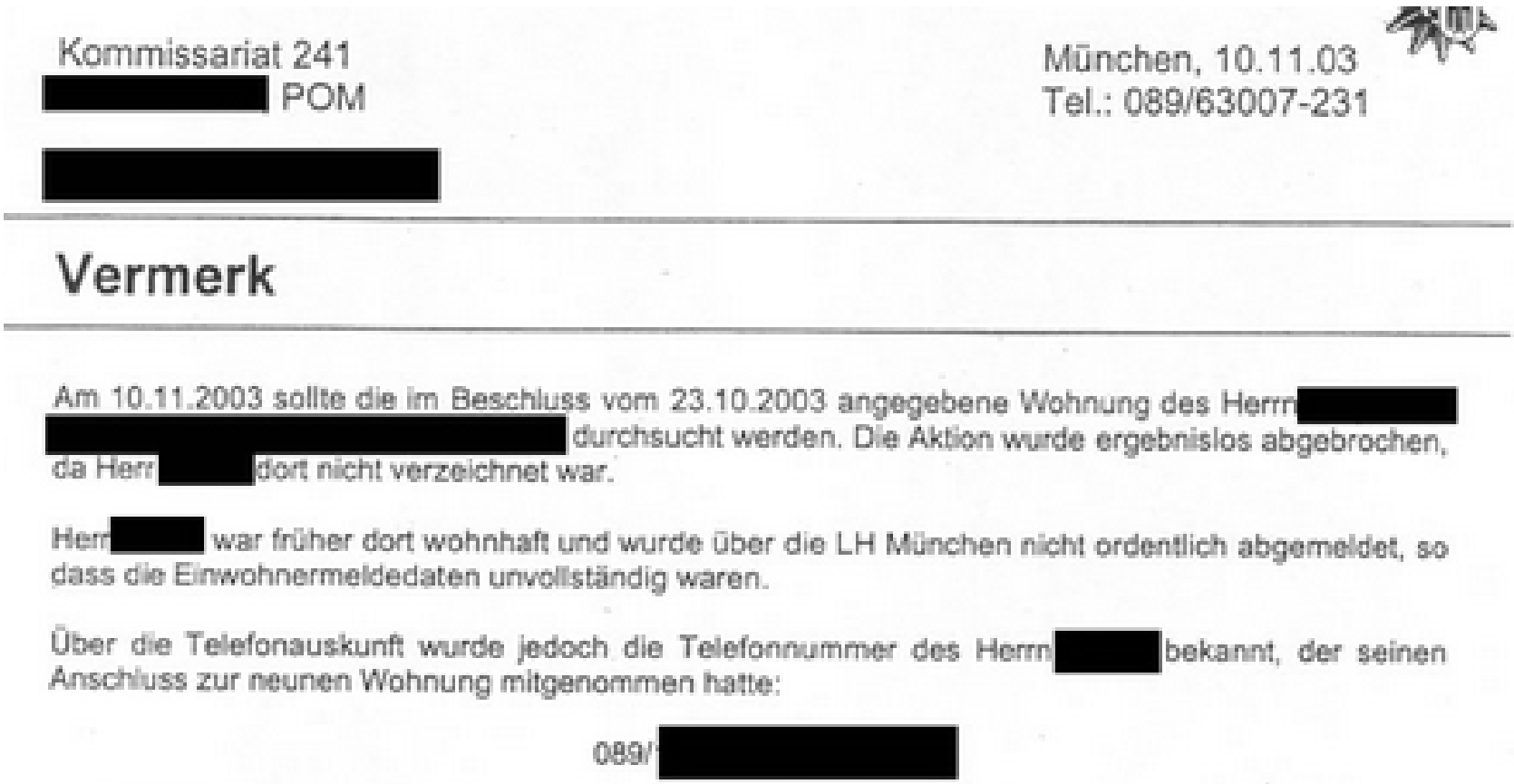
- Dokumente = Blätter
- Prüfsumme = Wurzel
- Höhe: $L = \log_2(\#\text{Dokumente})$
- L Zwischenwerte, um die Echtheit eines Blattes zu beweisen

Beweis der Echtheit (Ausschnitt)



Wozu braucht man die Zufallswerte?

zum Schutz vertraulicher Information in Dokumenten; für öffentliche Dokumente sind Zufallswerte unnötig



Kommissariat 241
[redacted] POM

[redacted]

München, 10.11.03
Tel.: 089/63007-231

Vermerk

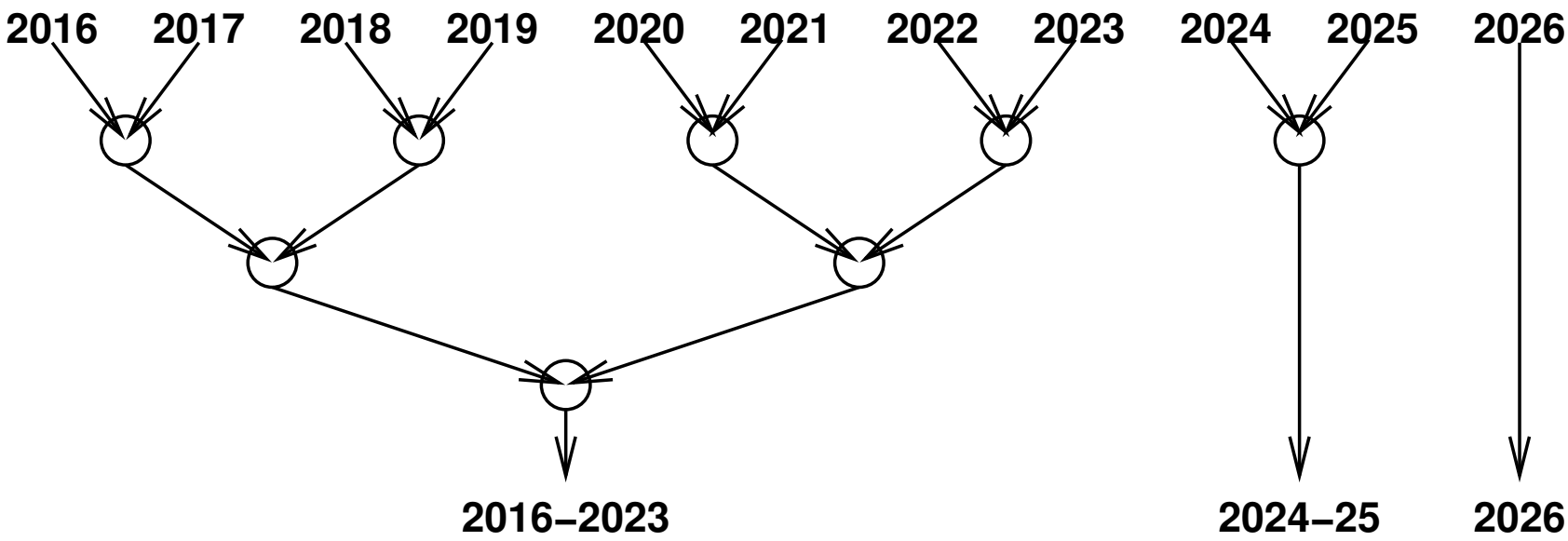
Am 10.11.2003 sollte die im Beschluss vom 23.10.2003 angegebene Wohnung des Herrn [redacted] durchsucht werden. Die Aktion wurde ergebnislos abgebrochen, da Herr [redacted] dort nicht verzeichnet war.

Herr [redacted] war früher dort wohnhaft und wurde über die LH München nicht ordentlich abgemeldet, so dass die Einwohnermeldedaten unvollständig waren.

Über die Telefonauskunft wurde jedoch die Telefonnummer des Herrn [redacted] bekannt, der seinen Anschluss zur neuen Wohnung mitgenommen hatte:

089/[redacted]

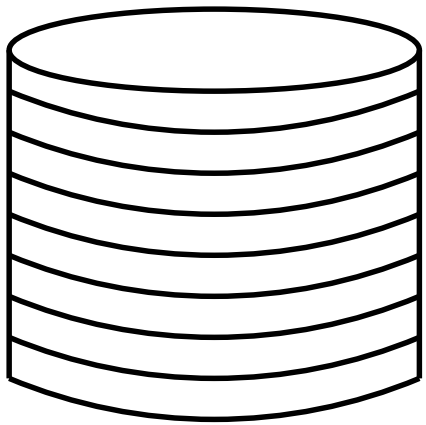
Digitale Magazine sind nicht statisch: Ständig kommen neue Daten hinzu!



- regelmäßig ein "Schnappschuss" der neu archivierten Daten
- "Wald" mit mehreren unterschiedlich großen Bäumen
- statt einer Prüfsumme (=Wurzel) nun für jeden Baum eine
- aber es gibt höchstens logarithmisch viele Bäume im Wald

Kann der Ansatz auch sinnvoll sein, wenn Sie ihre Daten selber speichern?

. . . oder bei einer vertrauenswürdigen Schwesterbehörde speichern lassen?



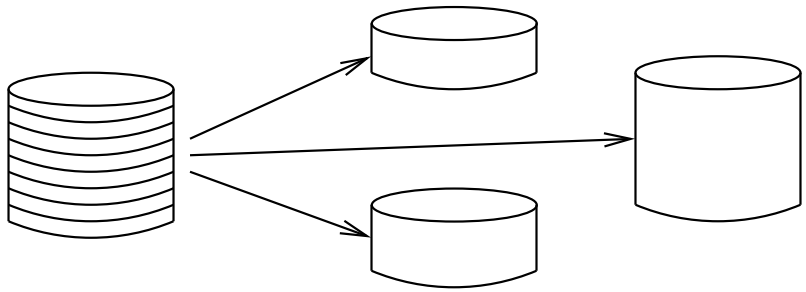
- Vertrauen der Archivnutzer
- Katastrophenvorsorge

Physikalische Archivalien

- können bei einer Katastrophe unwiederbringlich verlorengehen
- Kopien (z.B. auf Mikrofilm) sind kein vollwertiger Ersatz

Digitale Archivalien

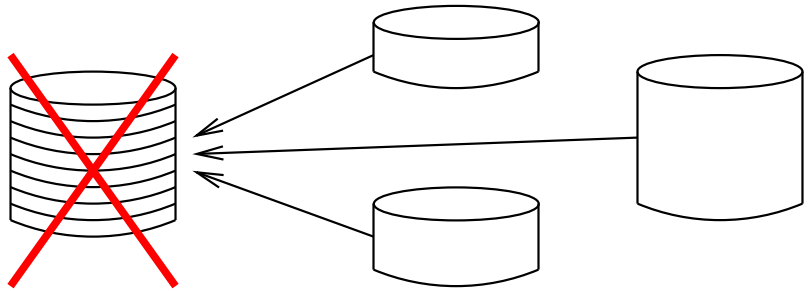
- sind beliebig (verlustfrei!) kopierbar



- Kopien können das ganze Magazin enthalten, oder beliebige Teile davon

Digitale Archivalien

- können im Fall einer Katastrophe verlustfrei aus Backups rekonstruiert werden



- wenn man rechtzeitig Backups angelegt und weit genug gestreut hat
- mit Hilfe von Hash-Bäumen können Sie überprüfen, ob die Backup-Daten echt sind . . . oder ob jemand die Daten manipuliert hat

Meine Empfehlung

- Hash-Funktion: SHA-512 (oder ggf. SHA-3)
- Größe des Hash-Wertes: 512 bit (= 64 Byte)
- Größe des Zufallswertes: 256 bit (= 32 Byte)

Schlussbemerkungen

- Digitalen Daten $\leftrightarrow$ physikalische Dokumente:
  1. Vorteil digitaler Daten: Verlustfrei kopierbar!
  2. Nachteil digitaler Daten: Leicht zu fälschen!
- Der erste Punkt ist (unter anderem) bei der Katastrophenvorsorge nützlich.
- Für den zweiten Punkt gibt es die Kryptograpie:
  - Mit Hash-Bäumen kann man sich einfach und effizient von der Unverfälschtheit der Daten überzeugenm
    - auch ohne teure CAS- oder WORM-Systeme
    - ggf. bei externen Dienstleistern gespeichert.
  - Statt Hash-Bäumen könnte man auch Digitale Signaturen einsetzen – aber das wäre “Overkill”.