Zurich University
of Applied Sciences

School of
Management and Law

CAS Data Engineering – Gastbeitrag - Datensicherheit

«… service not available»

Tibor Dudas
tibor.dudas@zhaw.ch

Lecturer

Tibor Dudas
- Physiker
- +10 Jahre Erfahrung im hochsicheren und hochverfügbaren IT-Umfeld

Erwartungsmanagement

CAS Cybersecurity-Teilnehmer: Es wird eine Wiederholung sein im Wesentlichen. Wer möchte, kann früher nach Hause gehen.

Merke

«Security by obscurity» hat noch nie funktioniert und wird auch nie funktionieren!

ALLES, was ihr einsetzt, müsst ihr **zu 100% verstehen**, sonst geht das schief!

Was genau besagt: «Die Rechenzentren sind ISO XYZ zertifiziert …»?

Agenda

Allgemeines

Confidentiality (Geheimhaltung)

- Am besten gar nicht erst versuchen (müssen) …
- Anonymisierung/Pseudonymisierung

Integrity (Unveränderthеit)

- Hash-Bäume und Hash-Ketten (e.g. Blockchain)

Availability (Verfügbarkeit)

- Redundanz im System
- Redundanz durch mehrere Systeme
- Backup

Cyber- und Informationssicherheit

Eigenschaften

- Integrität
- Verfügbarkeit
- Geheimhaltung
- Authentizität
- Verbindlichkeit
- Zurechenbarkeit
- Anonymität

Quelle: Eckert, C. (2013). IT-Sicherheit: Konzepte-Verfahren-Protokolle. Walter de Gruyter.

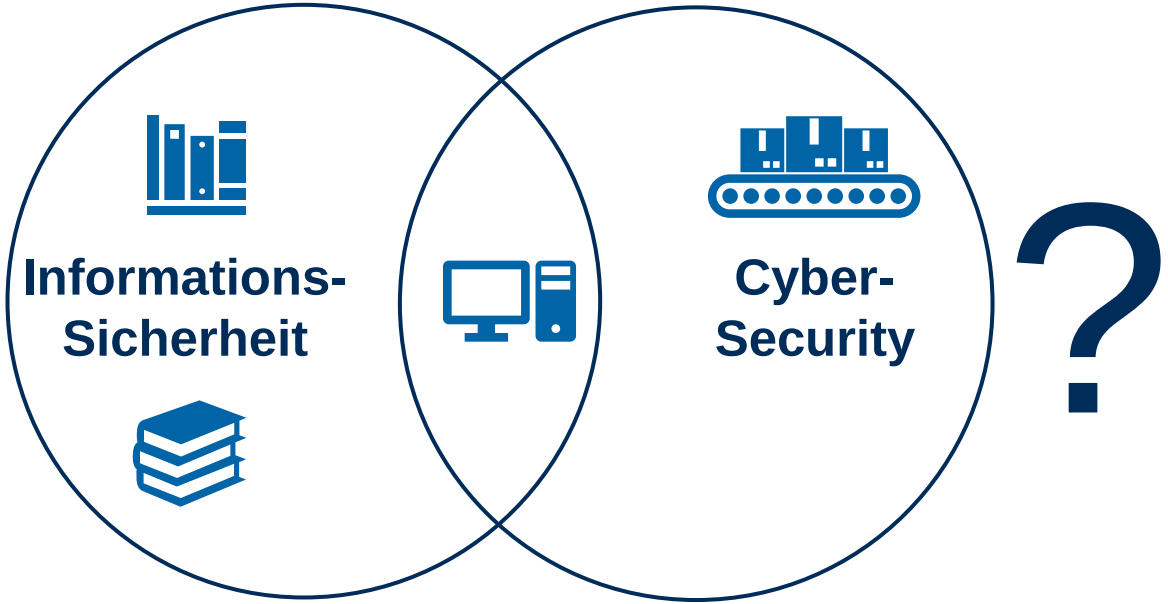


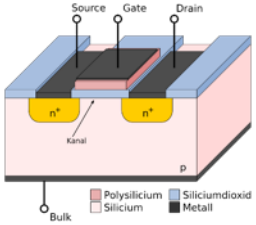
Abbildung basiert auf: Von Solms, R., & Van Niekerk, J. (2013). From information security to cyber security. *computers & security*, *38*, 97-102.

Einfach -> Komplex:
Das können wir (meistens) gut!

Von Denis Apel - Eigenes Werk, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=13294144

x 2000

x 4’800’000’000



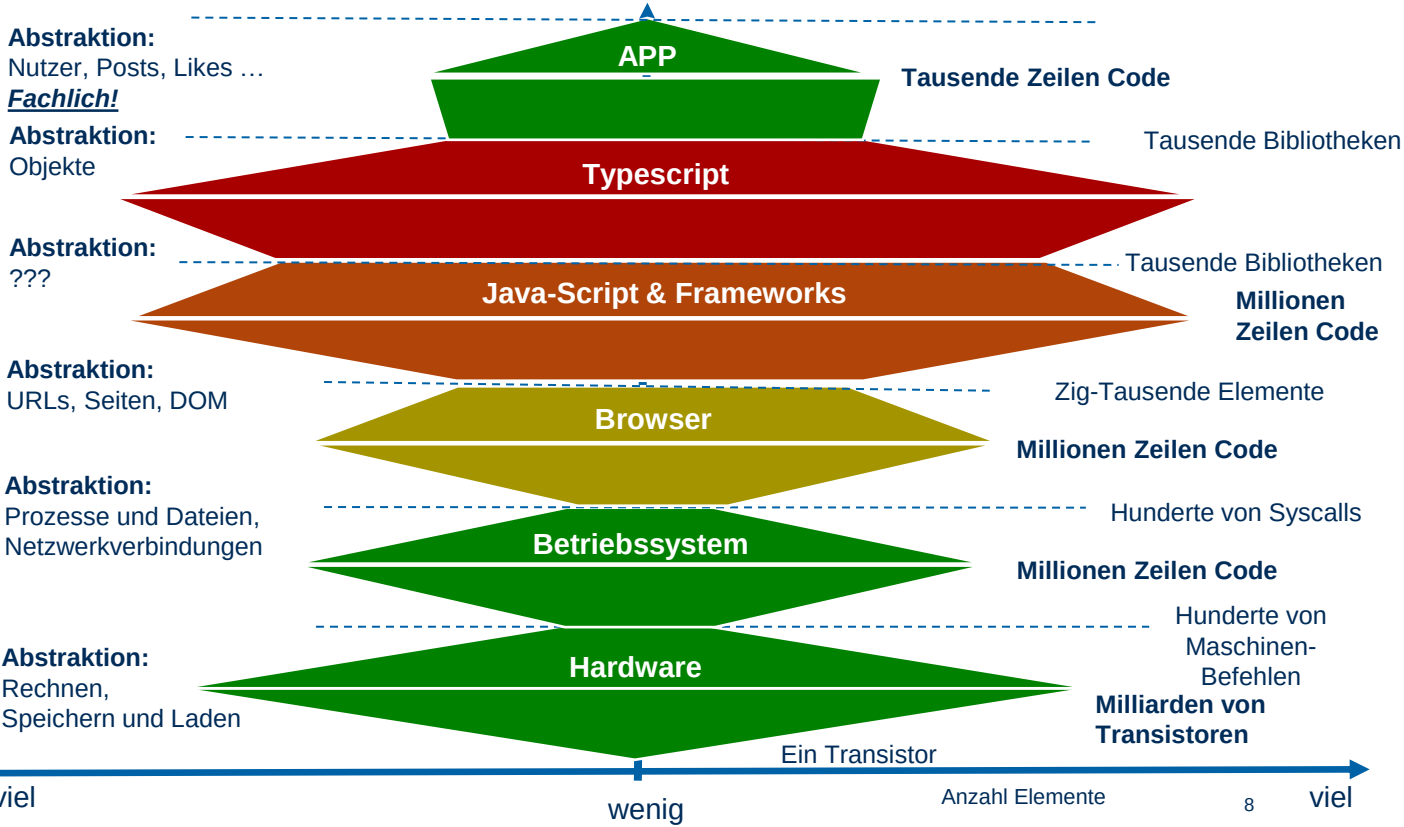
Von PNG-Version: Markus A. Hennig (17. Dezember 2005)SVG-Umsetzung Cepheiden - Datei:N-Kanal-MOSFET.png, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=8966218

Von Venusianer, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=3632073

Faktor 2.4 Millionen in weniger als 100 Jahren!

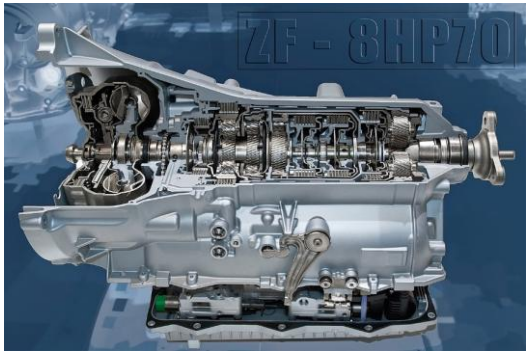
Von Michael Wolf, Penig, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=64406762

Vertikale Komplexität: Beispiel Web-Anwendung



Pretended Solution: Complex Stuff with Simple Interaction

Hundrets of parts …



Von Ritchyblack - Stefan Krause - Eigenes Werk, FAL, https://commons.wikimedia.org/w/index.php?curid=13319577

… one lever

Von Ralf Roletschek (talk) - Fahrradtechnik auf fahrradmonteur.de - Eigenes Werk, FAL, https://commons.wikimedia.org/w/index.php?curid=16130749

Billions of Transistors …

Von Michael Wolf, Penig, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=64406762

… some 100 assembly mnemonics

```
_start:
    mov eax, 4
    mov ebx, 1

    mov ecx, str
    mov edx, str_len
    int 80h

    mov eax, 1
    mov ebx, 0
    int 80h
```

Komplexität durch Featurewahnsinn

Wenn wir das brauchen, …

Von I, 天然ガス, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=

… bauen wir oft so etwas (wenn wir Glück haben):

Von Danial Haghgoo - Iranian Spotters - Gallery page http://www.airliners.net/photo/Saha-Airlines/Boeing-707-3J9C/2137981/LPhoto http://cdn-www.airliners.net/aviation-photos/photos/1/8/9/2137981.jpg, GFDL 1.2, https://commons.wikimedia.org/w/index.php?curid=27407759

Beherrschbare vs. unbeherrschbare Komplexität

Agenda

Allgemeines

Confidentiality (Geheimhaltung)

- Am besten gar nicht erst versuchen (müssen) …
- Anonymisierung/Pseudonymisierung

Integrity (Unverändertheit)

- Hash-Bäume und Hash-Ketten (e.g. Blockchain)

Availability (Verfügbarkeit)

- Redundanz im System
- Redundanz durch mehrere Systeme
- Backup

Confidentiality (Geheimhaltung) … geht regelmässig schief.

Screenshot von https://www.bbc.com/news/technology-37974266

Screenshot von https://www.zdnet.com/article/a-hacker-has-wiped-defaced-more-than-15000-elasticsearch-servers/

Screenshot von https://www.nytimes.com/2018/11/30/business/marriott-data-breach.html

Immer wieder so oder so ähnlich …

LUFTHANSA MILES & MORE

Daten von Vielfliegerkunden der Star Alliance gestohlen

Vielfliegerkunden sind per E-Mail informiert worden, dass diverse Daten gestohlen worden seien. Der Hack könnte Hunderttausende Passagiere betreffen.

5. März 2021, 10:51 Uhr, Oliver Nickel

Screenshot von https://www.golem.de/news/lufthansa-daten-von-vielfliegerkunden-der-star-alliance-gestohlen-2103-154714.html

Cloud ist auch (k)eine Lösung …

Auszug aus dem Protokoll
des Regierungsrates des Kantons Zürich

Sitzung vom 30. März 2022

542. Einsatz von Cloud-Lösungen in der kantonalen Verwaltung (Microsoft 365), Zulassung

6. Weitere Risiken bei Cloud-Lösungen im Allgemeinen und M365 im Besonderen (ISDS-Konzepte)

Die ISDS-Konzepte weisen sämtliche Restrisiken aus, die mit dem Betrieb der IT-Lösung und der Organisation einhergehen. Weitaus grössere Risiken als der Lawful Access bei Cloud-Lösungen birgt die Offenlegung vertraulicher Informationen durch unerlaubte und illegale Zugriffe durch Dritte wie Hacker oder kriminelle Organisationen. Die grossen Anbieter von Cloud-Lösungen schützen die Daten mit der neuesten Technologie und den höchsten Sicherheitsvorkehrungen und passen diese stets an die neueste Bedrohungslage an. Daher sind die Risiken der Offenlegung vertraulicher Informationen durch unerlaubte und illegale Zugriffe tendenziell eher geringer, als wenn die Daten on premises gehalten werden. Nach eigenen Angaben hat Microsoft 2021 mehr als 9,6 Mrd.

https://www.zh.ch/bin/zhweb/publish/regierungsratsbeschluss-unterlagen./2022/542/RRB-2022-0542.pdf

vs.

Am besten gar nicht erst versuchen (müssen) …

Assume Breach!

- Geht auch Datensparsamkeit (keine ‘heissen’ Daten auf Halde!)?
  - Achtung: DSGVO!
- Muss sensitives Material WIRKLICH in die Cloud?
- Kann ich wirklich gut Anonymisieren / Pseudonymisieren?
  - Keine Quasi-Identifikatoren (z.B. PLZ + Geburtsdatum)
  - Nicht-Verkettbarkeit
  - Geht evtl. auch «Differential Privacy»?

Crypto verlagert das Problem -> Wer hat den (die) Schlüssel? Wo ist dieser gespeichert? Was wenn diese Systeme gebreached werden?

Crypto bricht meistens nicht am Verfahren, sondern an der Implementierung!

Wenn Geheimhaltung wichtig ist, dann richtig machen!

Externe: Die gesamte Service-Kette ist durchzertifiziert!

- Rechenzentrum
- SaaS/PaaS-Anbieter
- Ihr selbst?

Externe: Welche Vertragsvereinbarungen bestehen?

Wenn ihr irgendwelche Zweifel habt: Selber machen!

- Dedizierte Rechner
- Dedizierte Netze
- Dedizierte Räume
- etc. etc.

Agenda

Allgemeines

Confidentiality (Geheimhaltung)

- Am besten gar nicht erst versuchen (müssen) …
- Anonymisierung/Pseudonymisierung

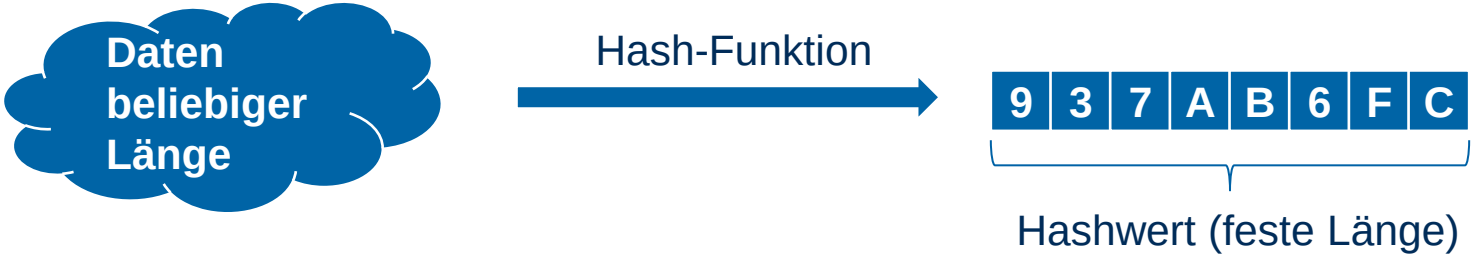
Integrity (Unverändertheit)

- Hash-Bäume und Hash-Ketten (e.g. Blockchain)

Availability (Verfügbarkeit)

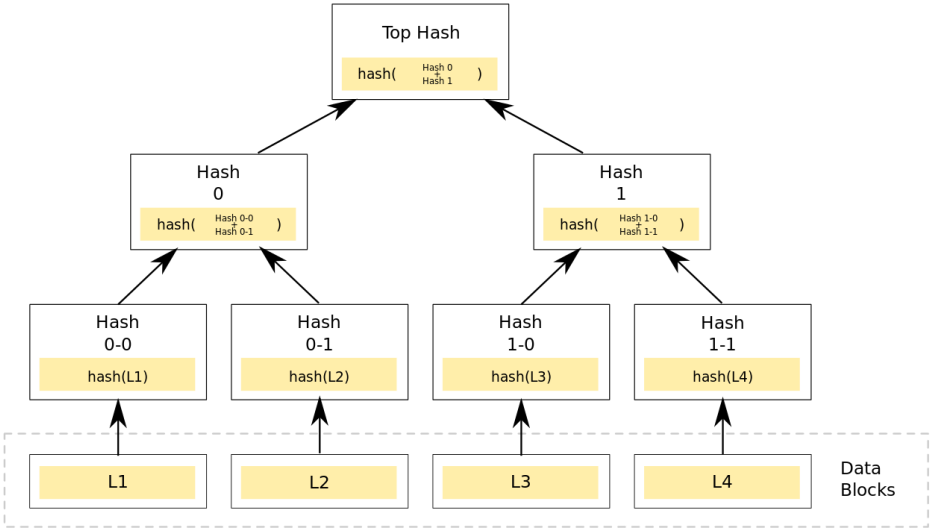
- Redundanz im System
- Redundanz durch mehrere Systeme
- Backup

Die Idee der Hash-Funktion



- Hashfunktionen sind Einwegfunktionen! => Die Daten kann man nicht aus dem Hash rekonstruieren.
- Es gibt aber Kollisionen! (SEHR unwahrscheinlich bei guten Hashfunktionen).
- Kleine Veränderungen in den Daten sollen grosse Änderungen am Hashwert erzeugen.
- Für gegebene Daten erzeugt die Hashfunktion immer denselben Wert.

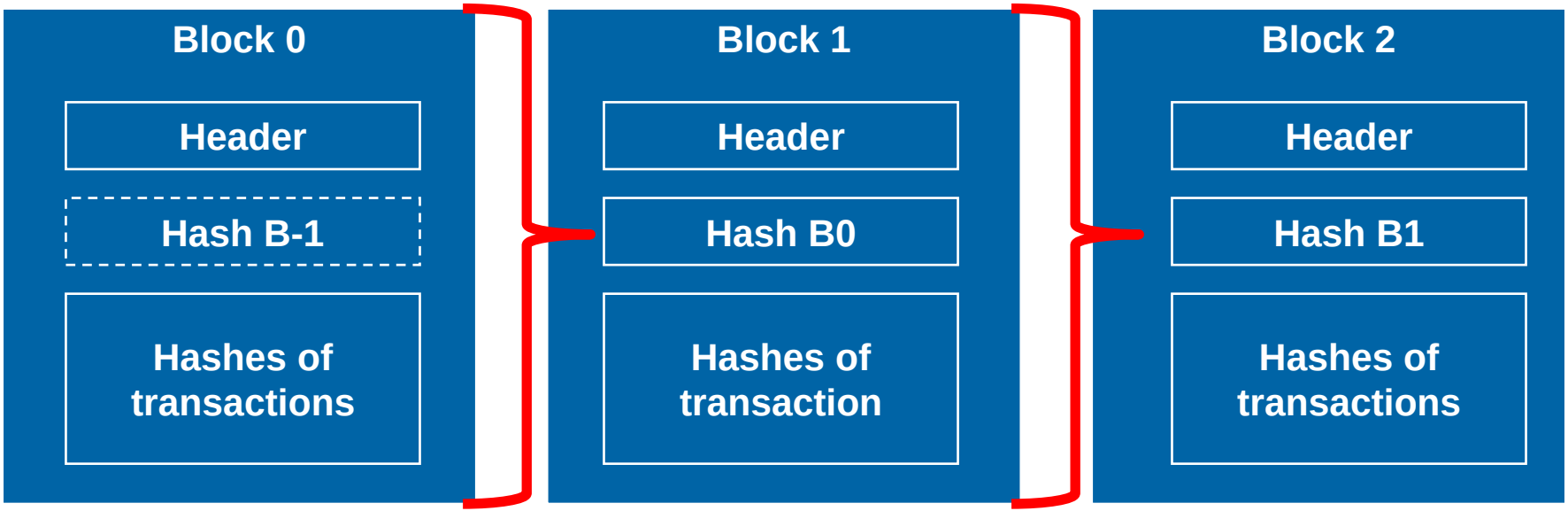
Merkle-Bäume



Von Azaghal - Eigenes Werk, CC0, https://commons.wikimedia.org/w/index.php?curid=18157888

- Effiziente Hash-Datenstruktur.
- Möglichkeit korrupte Bereiche zu erkennen.
- Die Überprüfung des Wurzelknoten reicht, wenn der Hash stimmt.

Blockchains



Agenda

Allgemeines

Confidentiality (Geheimhaltung)

- Am besten gar nicht erst versuchen (müssen) …
- Anonymisierung/Pseudonymisierung

Integrity (Unverändertheit)

- Hash-Bäume und Hash-Ketten (e.g. Blockchain)

Availability (Verfügbarkeit)

- Redundanz im System
- Redundanz durch mehrere Systeme
- Backup

Das zentrale Motto

Kein Backup? Kein Mitleid!

Restore zu langsam? Ein kleines bisschen Mitleid, vielleicht.

Aber … so einfach ist es dann doch nicht!

Schief gehen kann sehr viel …

Die Klassiker:

- Out of Memory
- Out of Diskspace
- Out of Network «Data-Rate»

Thermische Probleme:

- Zu hohe Betriebstemperatur ...

Stromversorgung:

- Stromausfälle / Schwankungen

Wartung:

- Gerätereinigung?
- Patches? Updates?

JEDER Softwarebug ist auch eine Gefahr für die Verfügbarkeit!

- Deadlocks
- Unendliche Schleifen (typischer Absturz)

Gemein: Byzantine Fehler / “Heisenbugs”

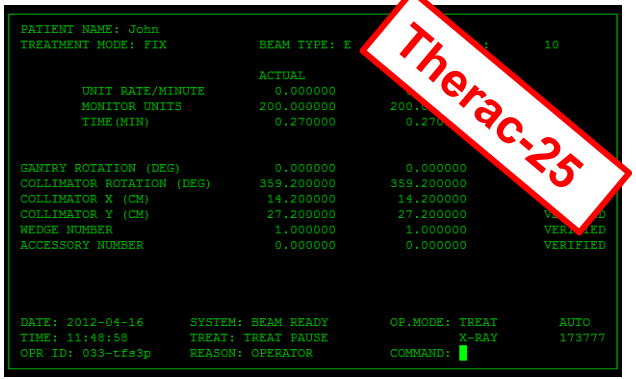
Byzantine Faults – Kommen selten vor und sind kaum vorhersagbar.

- Einzelne “Bit-Flips”
- Vorübergehende Fehler
  - Temperatur?
  - Wackelkontakt?

Der gemeine «Heisenbug»

- Verschwindet oft beim Debuggen
  - Z.B: “Race conditions”

Evtl. tödlichster «Heisenbug» in der Geschichte!

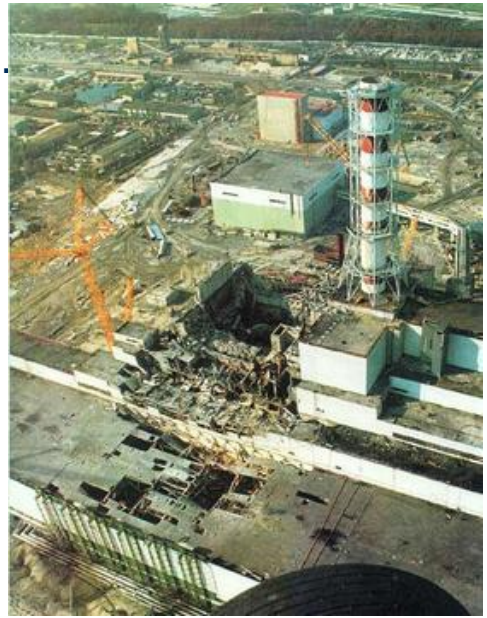


By Software compiled from source code available at [1], Copyrighted free use, https://commons.wikimedia.org/w/index.php?curid=19389203

Vorhersehbar: Betrieb ausserhalb der Spezifikation

- Meist schlecht vorhersagbar, wie sich ein System verhält, wenn es ausserhalb der Spec betrieben wird.

- Hardware-Performance?
- Betriebssystemversion?
- Datenbank-Version, etc. etc.



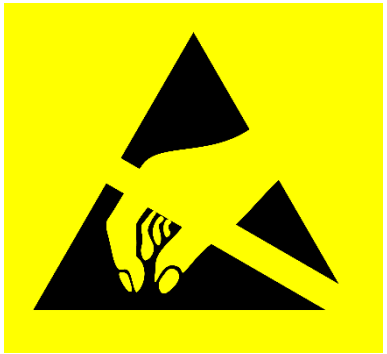
By Soviet Authorities, Fair use, https://en.wikipedia.org/w/index.php?curid=4866476

Fast unvermeidbar: Hardwareausfälle

Hardware altert!

https://hackaday.com/2019/04/12/ask-hackaday-experiences-with-capacitor-failure/

Falsche Handhabung

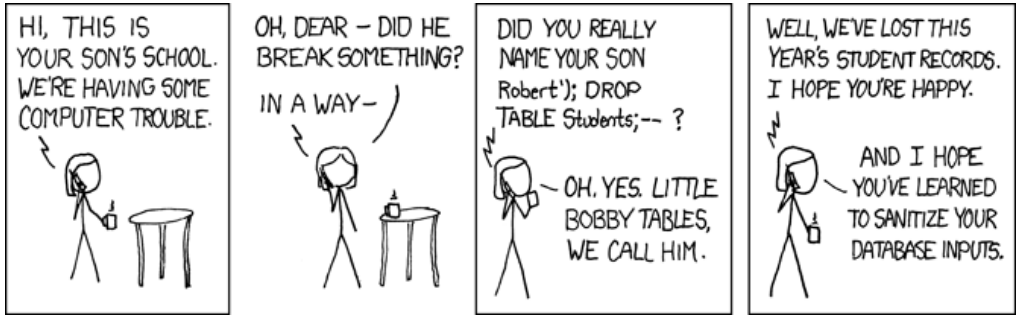


ALLES mechanische geht irgendwann kaputt.

Schlampigkeit: Schlechte Programmierung / schlechtes Design

Injecting Commands

- SQL Injection
- Command Injection
- **Buffer overflow**
  - *Bad:* Deliberate crashing
  - *Worse:* Function «reusing»
  - *Worst:* Code execution
  - *Meltdown:* Remote-Code-Execution
  - *Asteroid impact:* Remote-Code-Execution with root privileges!
  - Supernova: Remote-Code-Execution as root and Hypervisor Breakout



Alle auch als Zero-Day-Attacken denkbar.

https://de.wikipedia.org/wiki/Puffer%C3%BCberlauf

Was ist eigentlich ein Server? Ein normaler Computer?

– Ja (in der Theorie), **Nein (in der Praxis)!**

HPE ProLiant DL380: Der weltweit meistverkaufte Server ist jetzt noch besser
https://www.youtube.com/watch?v=Bv-Nl5sdmXo

Qualität der Komponenten

Billigstes Desktop-Netzteil

Public Domain,
https://commons.wikimedia.org/w/index.php?curid=84957

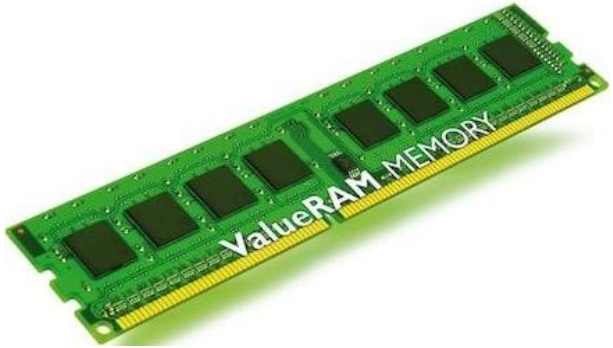
Teures Hot-Plug Server-Netzteil
N+1 Redundant

https://www.supermicro.com/en/support/resources/pws

Redundant und im Betrieb wechselbar.

- Mehrere Lüfter, redundant
- Bis zu 30 W (oder mehr) pro Lüfter!
- Lüfterüberwachung
- “Hot Swap”-Austausch

ECC-RAM



<- Normal RAM

<- ECC RAM

ECC: Error-Correcting-Code

Detect and correct single bit errors.
Redundant bits: i.e. 72/64

Weitere Betriebsmodi (Herstellerabhängig!):

- ECC Mode
- Mirror Mode
- Spare Mode

RAID (Redundante Festplatten)

Viele verschiedene Schnittstellen:

USB, SATA, SAS, FC, PCIe

„mechanische“ Festplatten
Grösse ca. 12TB / Disk
Preis ca. CHF 0.03 / GB
Speed 100-200 Mbyte/s

Elektronische Festplatten (SSD)
Grösse ca. 1TB / Disk
Preis ca. CHF 0.14 / GB
Speed bis ca. 1500 Mbyte/s

Software vs. Hardware RAID

Software RAID

+ Auf Betriebssystem / Treiberebene
+ Keine zusätzlichen Kosten
+ Recovery mit jedem anderen System möglich
- Langsamer / hohe System-/CPU-Belastung
- Write-Caching gefährlich / abgeschaltet

Hardware RAID

+ Meist mehr Ports verfügbar
+ Unabhängig vom Betriebssystem
+ Write-Caching geht (mit Batterie-Backup)
- Recovery ohne Originalcontroller evtl. problematisch

Adaptec RAID 8805: PCI-Ex8 RAID-Kontroller

Praktische Probleme mit RAID

RAID ist keine (vollständige) Sicherheitslösung

- RAID bemerkt keine sporadischen Fehler.
- Hohe Last bei der Wiederherstellung -> Weitere Ausfälle?
- Kein Schutz gegen versehentliches Löschen, Systemfehler oder Ransomware.
- Additionale Komplexität -> weitere Möglichkeiten für Fehler.
- Serienfehler (Hardwarefehler) einer Disk-Lieferung?

=> Viele Möglichkeiten dennoch seine Daten zu verlieren!

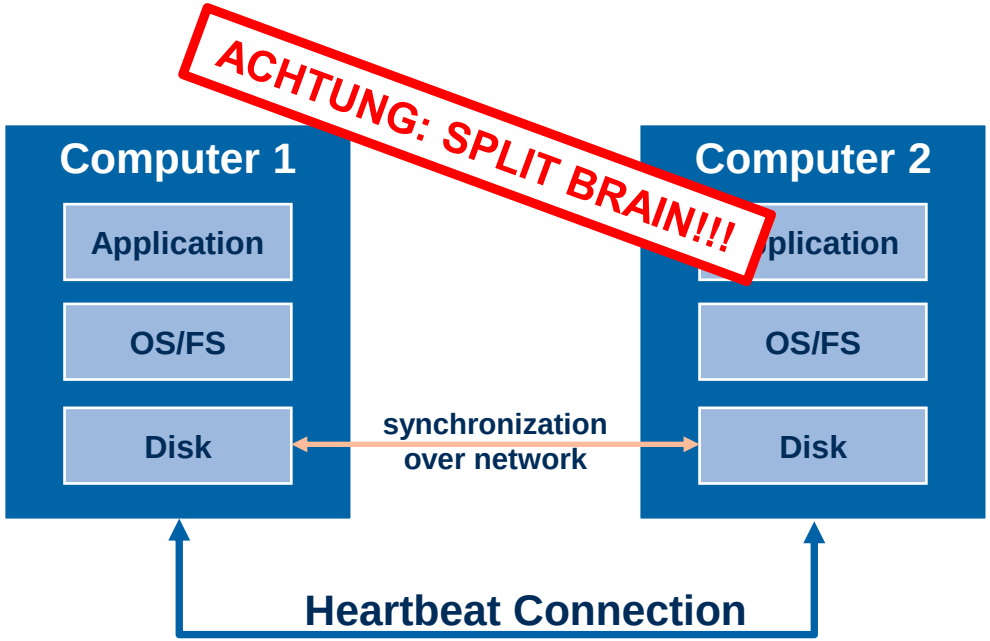
Darum: RAID ist kein Backup !!!

Vorschlag (Budget): Lieber in ein gutes Backup investieren, wenn man diese Form der Verfügbarkeit nicht braucht. Sonst unbedingt in beides investieren.

Verteilte, redundante Block-Devices (DRBD)

Wie ein RAID(1) aber anders:

- Immer synchronisiert aber beidseitig zugreifbar.
- Braucht ein spezielles Dateisystem für active/active Zugriff.



Das Quorum

Wege aus dem Splitbrain-Problem

- Mehrheit der Knoten (>1/2) muss präsent sein.
- Wenn ein Knoten nicht mehr als ½ aller Knoten erreicht, schaltet er sich ab.
- Drei Knoten können keinen Ausfall verkraften.
- Vier Knoten können einen Ausfall verkraften.
- Fünf Knoten können zwei Ausfälle verkraften.

Noch Zeit? -> Backup genauer anschauen!

Medien / Laufwerke (Tape)

Verschiedene Technologien verfügbar

Z.B.: LTO-Drives

Tandberg LTO-8 Drive
12TB (nativ)
300MB/s (nativ)
=> ca. **12TB in 11h**

https://www.bechtle.com/ch/shop/ibm-lto-8-ultrium-tape--4232675--p

↗ CHF 3500!

Medien / Laufwerke (Tape)

Wenn man viel Daten hat …

https://www.backupworks.com/Qualstar-Q40-LTO-8-Tape-Library.aspx

… nimmt man einfach mehr Bänder!

z.B.: 8.76 PB / m^2

Aber: Wie lange dauert das Recovery?

If you have to ask the price, you probably can't afford it!

Günstig(ere) Alternativen

Einfach normale Festplatten nutzen

Aber:

- Sensitiv auf
  - Schock
  - Temperatur, Feuchtigkeit
  - ESD
- MTBF?
- Lebensdauer der Daten?

- **Dafür gibt es spezielle HDD-basierte Produkte.**

Typisches RDX Drive

https://www.galaxus.ch/de/s1/product/tandberg-data-rdx-quikstor-backup-loesungen-3487684

Typisches 2TB RDX medium

https://www.galaxus.ch/de/s1/product/tandberg-data-8731-rdx-rdx-hdd-2000gb-cartridge-3487348

Backup auf Online-Systeme (Fast immer eine miese Idee!)

Cloud-Backup (Read: *«Backup to somebody else's machines!»*)

Vorteile:

- «Unlimitierter» Speicherplatz
- Preis nach Nutzung
- (Ggf.) Hohe physische Sicherheit

Nachteile:

- **!!! ABER !!!** Es ist am Netz !!! => **Es ist verwundbar!**
- Grosses und interessantes Target für gezielte Angriffe.
- Wenn es schief geht kann man evtl. klagen aber das hilft dann auch nichts gegen den Datenverlust.

Backup auf Online-Systeme (Fast immer eine miese Idee!)

NAS-Backup (Read: *«Backup to a cheap machine»*)

Vorteile:

- Keine?
- Ok, vielleicht die Wiederherstellungszeit …
- Billig in der Anschaffung, einfach zum Aufsetzen.
- Besser als nichts (obwohl …).

https://www.microspot.ch/de/computer-gaming/netzwerk/nas--c571000/qnap-tr-004--p0001687765?gclid=EAIaIQobChMI0offoqGs7QIVTQKLCh0xZQAIEAQYBCABEgJtE_D_BwE&gclsrc=aw.ds

Nachteile:

- Teil des normalen Netzes =>Angreifbar**!**
- Komplexer Software Stack => **Angreifbar!**
- Interessantes Ziel für Ransomware
- ***Worst of the worst:*** Eure (Windows)-Systeme haben Schreibzugriff auf die NAS! Oh nein!
  - Niemals solche Konfigurationen fahren!

Backup–Strategien (von schlimm zu angemessen)

Das Schlimmste:	Gar kein Backup.
Trotzdem schlimm:	Alle wichtigen Daten in die Cloud synchronisiert.
Trotzdem schlimm :	Alle Data regelmässig auf externes Laufwerk kopieren (immer dran).
OK-ish (@home):	Alle Data regelmässig auf mehrere externe Laufwerke kopieren und diese in Rotation «Off-Site» lagern (Bankschliessfach?).
OK-ish(@work):	Backup-Server der die Daten von den Maschinen zieht (PULL). Restriktive Firewall-Regeln und getrennte Netze. Regelmässige Off-Site-Backups.
Angemessen:	Multi-Level-Backup (schneller On-Site-Mirror -> Off-Site-Storage), Backup-Strategie, Bare-Metal-Recovery-Prozeduren verfügbar.
@ALWAYS:	TESTEN, TESTEN und wieder TESTEN

Was sichern?

Daten:

Alles, ohne Ausnahme!

Server:

Konfiguration als Minimum.

Operating System?

Virtuelle Maschinen-Images?

Netzwerk-Geräte:

!!! KONFIGURATION !!!

Bare Metal Recovery

Ziel: Wiederherstellen from scratch, wenn alles weg ist.

«Bare Metal»: Nur die Hardware.

<u>Ziel einer guten bare-metal Wiederherstellung:</u>

- **Die Zeit minimieren von bare-metal zu voll funktionsfähig**
- **Die Wiederherstellungsprozedur automatisieren**
  - Automatisiertes Deployen des Betriebssystems (unbeaufsichtigte Installation)
  - Automatisierte Updates
  - Automatisierte Konfigurationsmanagement
  - Automatisiertes Software Deployment
  - Automatisierte Datenwiederherstellung vom Backup
- **Das geht auch für Client-Rechner!**

Bare Metal Wiederherstellung

Tools für Bare Metal Wiederherstellung:

- **Terraform**
- **Ansible**
- **Chef**
- **Puppet**

Ansible:

- **“neu”**
- **Leicht zu erlernen**
- **Cross-Platform**
- **Für Linux braucht es keinen Agenten**

Bare Metal Wiederherstellung - Ansible

Ansible:

- **Funktioniert für 1 Server genauso wie für >100’000 -> skaliert**
- **Idempotent**
- **Beschreibend, nicht imperativ**
- **Infrastructure as code**
- **Funktioniert in der Cloud wie vor Ort**
- **Ad-hoc Commands**
- **Yaml-Dateien -> gut lesbar**

Agenda - Zusammenfassung

Allgemeines

Confidentiality (Geheimhaltung)

- Am besten gar nicht erst versuchen (müssen) …
- Anonymisierung/Pseudonymisierung

Integrity (Unverändertheit)

- Hash-Bäume und Hash-Ketten (e.g. Blockchain)

Availability (Verfügbarkeit)

- Redundanz im System
- Redundanz durch mehrere Systeme
- Backup

Zurich University
of Applied Sciences

School of
Management and Law

Vielen Dank für eure Aufmerksamkeit!