Thin Take statt Full Take

René Pfeiffer

web.luchs.at

21. Mai 2016

Inhaltsübersicht - Worum geht es?

- Logdaten
- Überblick über Ereignisse
- Security Information and Event Management (SIEM) lite
- Extraktion von Daten
- spezifische Fragestellungen
- einfaches Framework für Statistik und Reporting

Logdaten

```
[    26.700424] USB Video Class driver (1.1.1)
[    26.999337] AVX or AES-NI instructions are not detected.
[    27.014871] AVX or AES-NI instructions are not detected.
[    27.865571] Adding 15625212k swap on /dev/mapper/sda2_crypt.  Priority:-1 extents:1 across:15625212k FS
[    28.378007] XFS (dm-2): Mounting V5 Filesystem
[    28.526706] nouveau 0000:01:00.0: NVIDIA G94 (094100a1)
[    28.571229] XFS (dm-2): Ending clean mount
[    28.652686] nouveau 0000:01:00.0: bios: version 62.94.54.00.00
[    28.674148] nouveau 0000:01:00.0: fb: 512 MiB GDDR3
[    28.726497] [TTM] Zone  kernel: Available graphics memory: 4079358 kiB
[    28.726712] [TTM] Zone   dma32: Available graphics memory: 2097152 kiB
[    28.726925] [TTM] Initializing pool allocator
[    28.727136] [TTM] Initializing DMA pool allocator
[    28.727427] nouveau 0000:01:00.0: DRM: VRAM: 512 MiB
[    28.727637] nouveau 0000:01:00.0: DRM: GART: 1048576 MiB
[    28.727847] nouveau 0000:01:00.0: DRM: TMDS table version 2.0
[    28.728055] nouveau 0000:01:00.0: DRM: DCB version 4.0
[    28.728323] nouveau 0000:01:00.0: DRM: DCB outp 00: 02000300 00000028
[    28.728544] nouveau 0000:01:00.0: DRM: DCB outp 01: 01000302 00020030
[    28.728751] nouveau 0000:01:00.0: DRM: DCB outp 02: 04011310 00000028
[    28.728962] nouveau 0000:01:00.0: DRM: DCB outp 03: 02011312 00020030
[    28.729207] nouveau 0000:01:00.0: DRM: DCB conn 00: 00001030
[    28.729422] nouveau 0000:01:00.0: DRM: DCB conn 01: 00002130
[    28.729632] nouveau 0000:01:00.0: DRM: DCB conn 02: 00000210
[    28.729844] nouveau 0000:01:00.0: DRM: DCB conn 03: 00000211
[    28.730054] nouveau 0000:01:00.0: DRM: DCB conn 04: 00000213
[    28.754939] systemd-journald[271]: Received request to flush runtime journal from PID 1
[    28.873606] [drm] Supports vblank timestamp caching Rev 2 (21.10.2013).
[    28.873821] [drm] Driver supports precise vblank timestamp query.
[    29.050214] nouveau 0000:01:00.0: DRM: MM: using CRYPT for buffer copies
```

Typische Logs

```
root            0 May  9 06:25 alternatives.log
adm          4096 May 13 06:25 apache2
root          153 May  9 06:25 apt
root         3446 May 17 19:46 aptitude
adm         72145 May 17 23:39 auth.log
utmp            0 May  9 06:25 btmp
adm        309616 May 17 23:29 daemon.log
root        32032 Apr 19 10:40 faillog
adm        174482 May 17 20:03 kern.log
utmp       292292 May 17 10:00 lastlog
adm             0 May  9 06:25 mail.err
adm        117192 May 17 23:26 mail.info
adm        117192 May 17 23:26 mail.log
adm             0 May  9 06:25 mail.warn
adm        146930 May 17 20:03 messages
adm        278520 May 17 23:39 syslog
adm         10078 May 17 20:02 user.log
utmp        16896 May 17 20:05 wtmp
```

Typische Logauswertung

- `less` / `more`
- `grep` / `awk` / `sed`
- Konvertierung zu CSV
- Import in (No)SQL Datenbanken
- Import in proprietäre Tools
- Frameworks (Graylog, Logstash, . . .)
- Dienstleister (Splunk, Alert Logic, Loggly, . . .)

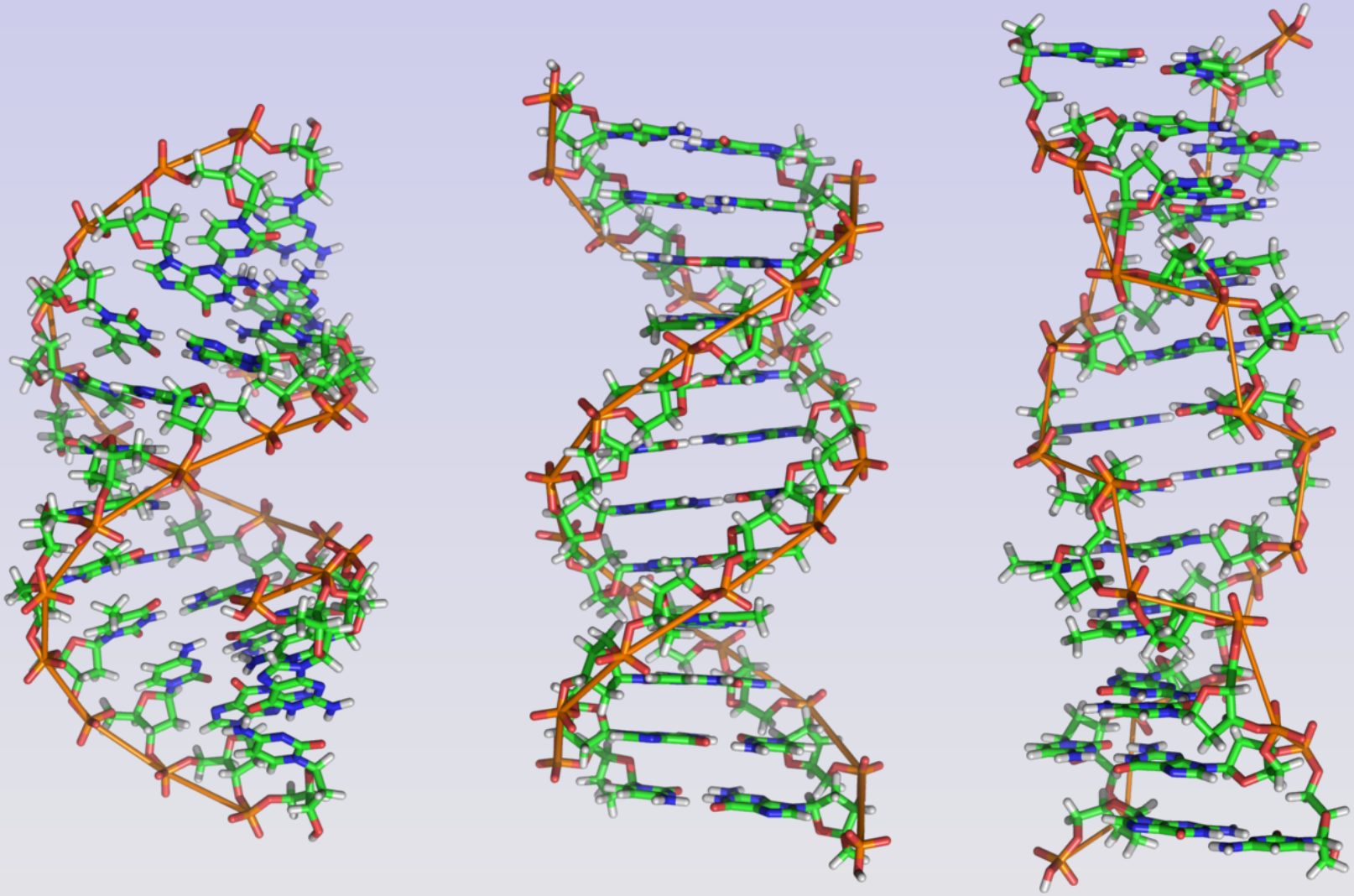
Wesentliche Fragen

- Welches Konto hat sie wann wo eingeloggt?
- Welches X.509 Zertifikat hat sich wann wo eingeloggt (OpenVPN™)?
- Welche Logins sind fehlgeschlagen?
- Gibt es Alarme im Intrusion Detection System?
- Welche X.509 Zertifikate haben Clients verwendet?
- Gibt es Anomalien im E-Mail Verkehr?
- Gibt es Anomalien im Netzwerkverkehr?

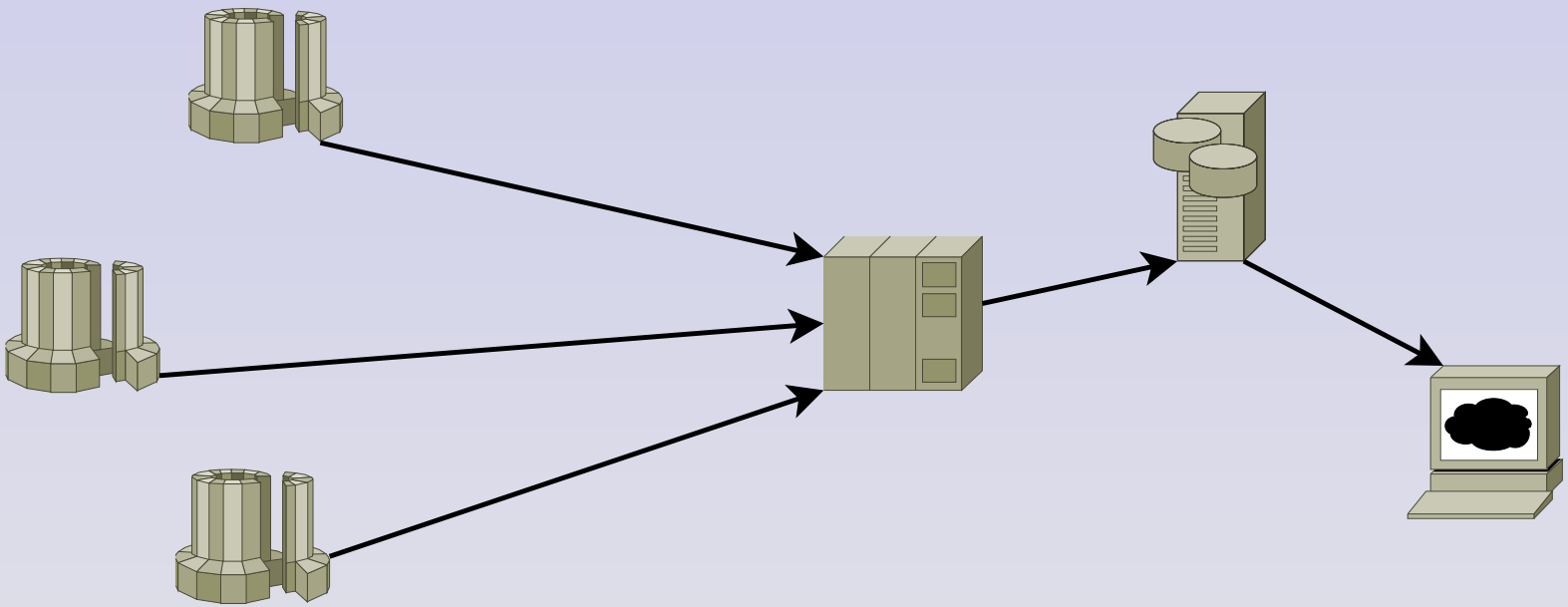
Fokus auf Applikationen

- Dovecot für IMAP / POP3
- OpenSSH
- OpenVPN™
- Postfix
- Suricata Intrusion Detection Engine
  - X.509 Parser für SSL/TLS
  - Export der Events in JSON
- NetFlow Daten (optional)

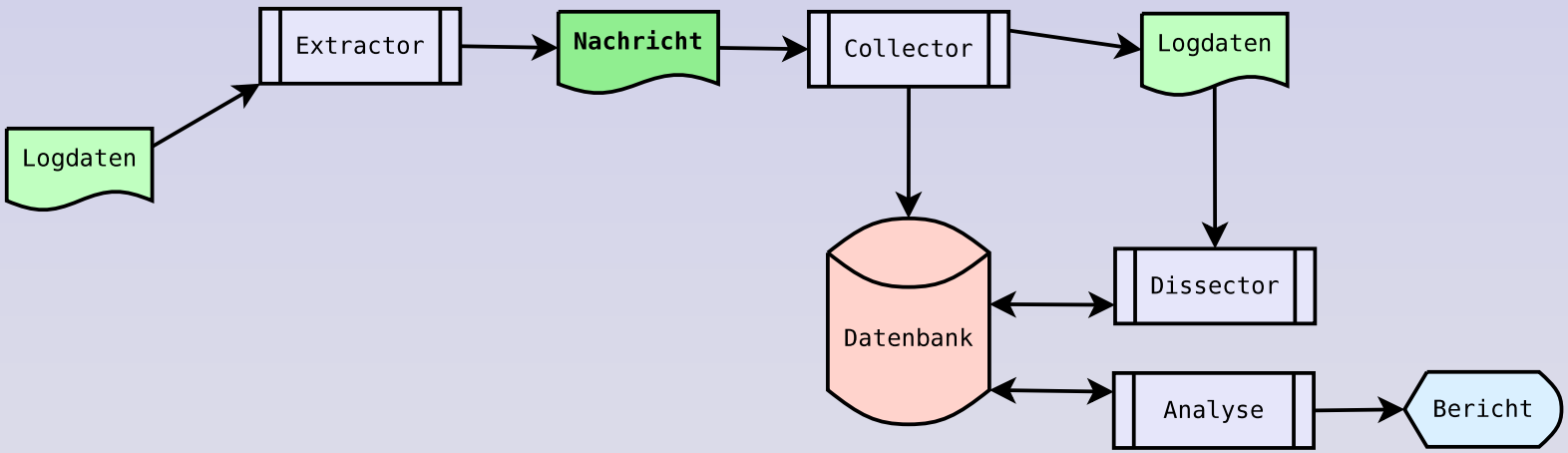
Bausteine



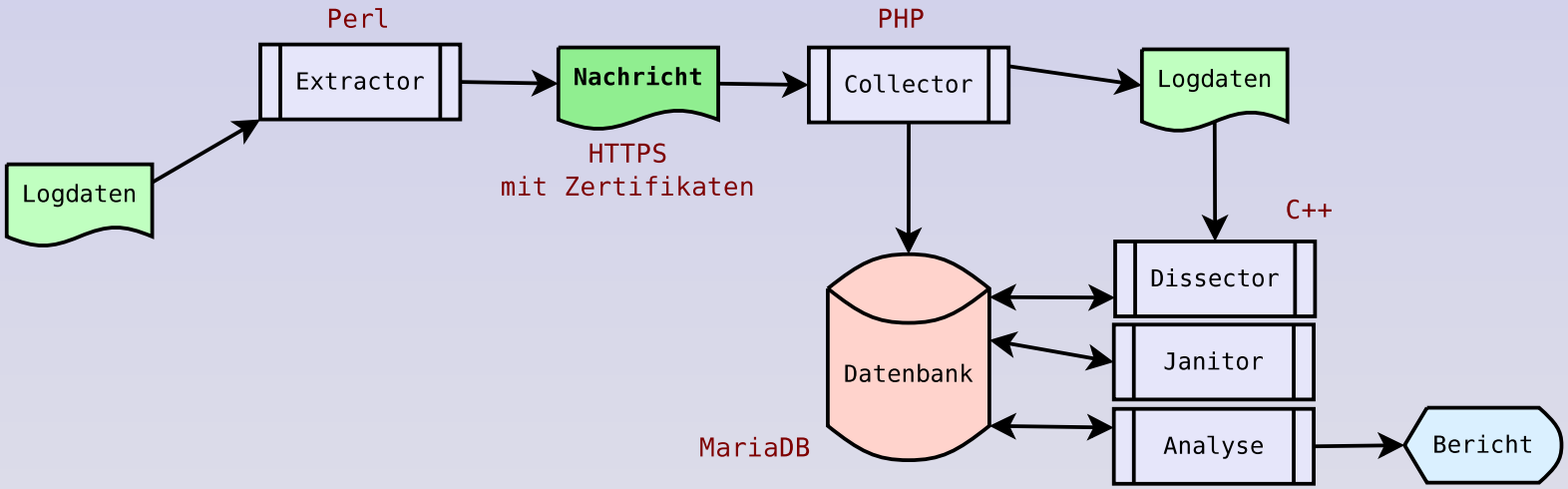
Grundlegendes Design (1)



Grundlegendes Design (2)



Grundlegendes Design (3)



Extractor

- Vorbereiten der Logdatei
  - Dekomprimieren (falls notwendig)
  - neu komprimieren mit `lrzip`
- Anmelden bei Collector
- Logdatei senden mit
  - Client Identifikation
  - Logdateiname und -typ
  - Zeitstempel
  - SHA-256 Hash der Logdatei
  - Logdatei
  - Fülldaten/Padding (optional)

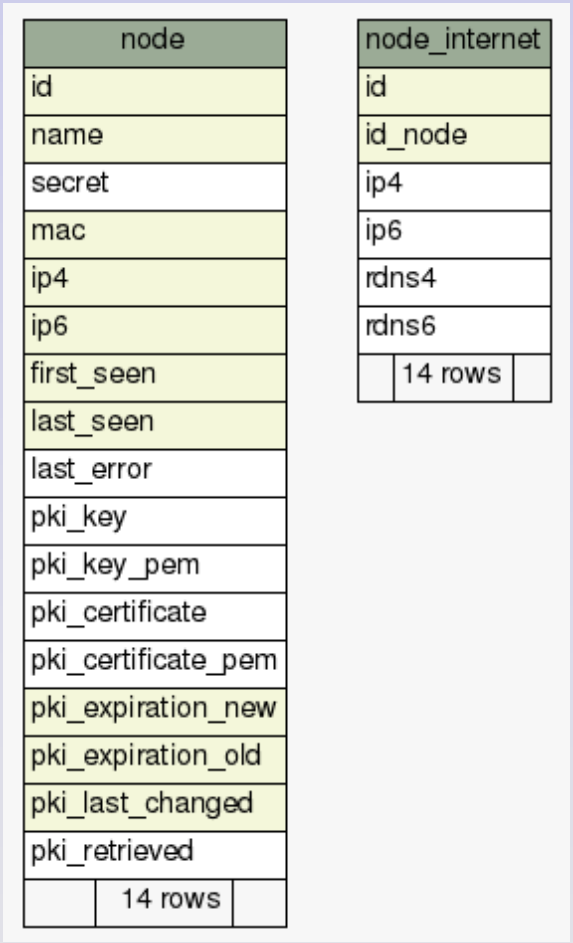
Collector

- Collector wartet auf Logdaten
- Prüfen des Clients (Anmeldung)
- Annahme des Logpakets und
  - Speichern der Metadaten in Datenbank,
  - Berechnung und Vergleich SHA-256 Hash,
  - Speichern der Logdatei in Warteverzeichnis
- Collector nimmt keine Auswertungen vor

Dissector

- Dissector untersucht nicht ausgewertete Logdateien
- Extraktion und zeilenweises Parsen
- Suche nach Mustern in Logs abhängig vom Typ
  - `.* Accepted publickey for (.*) from (.*) port.*`
  - `.*:\s(.*):(.*)\sVERIFY OK: depth=0,.*CN=(.*)`
  - `...`
- Transfer extrahierter Daten in Datenbank
- bestimmte Einträge haben Originalzeile als Referenz

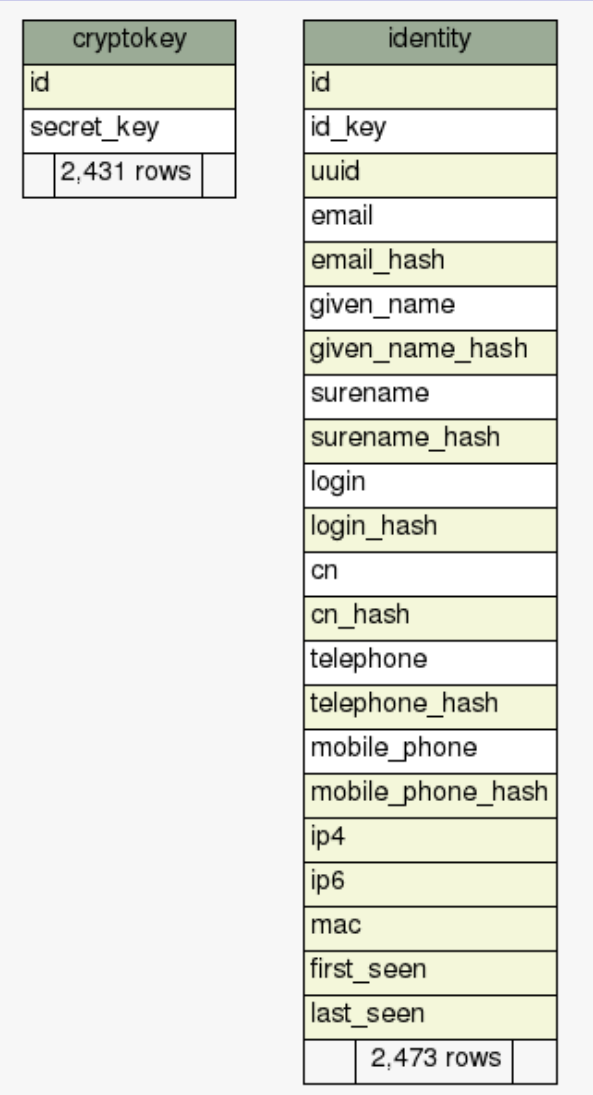
Datenbank - Nodes



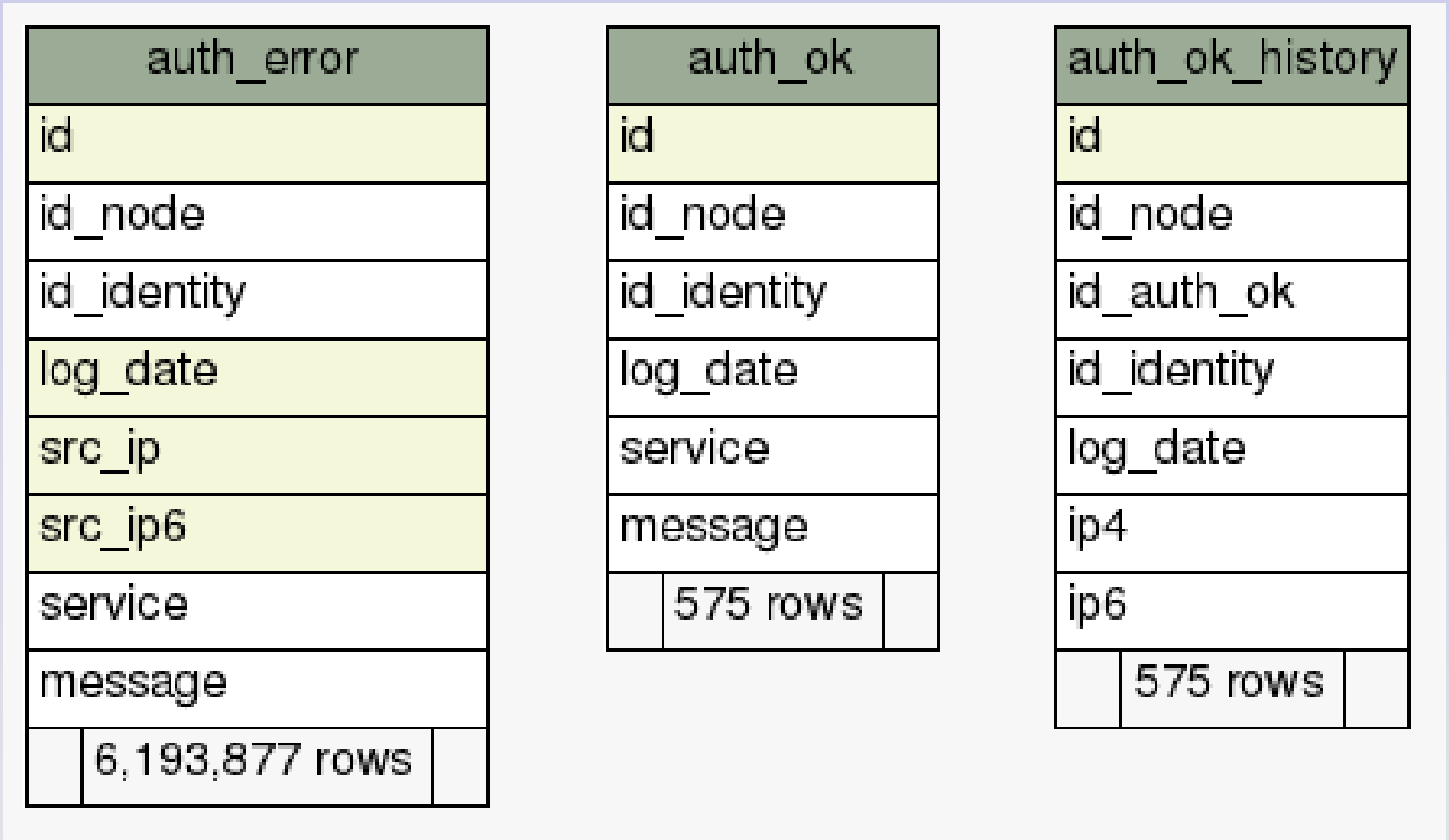
Problem: Persönliche Daten

- Logs enthalten persönliche Daten
- Auswertung erstellt Profile und Beziehungen
- Identitäten mit Unique Identifiers (UUID)
  - Identitätentabelle enthält verschlüsselte Identität
  - Schlüssel in separater Tabelle (auf anderem System)
  - zusätzlich SHA-256 Hashes zwecks Db-Operationen
- UUID reichen für Analyse auf
  - Entschlüsselung auf Bedarf
  - Anonymisierung durch Löschen der Schlüssel

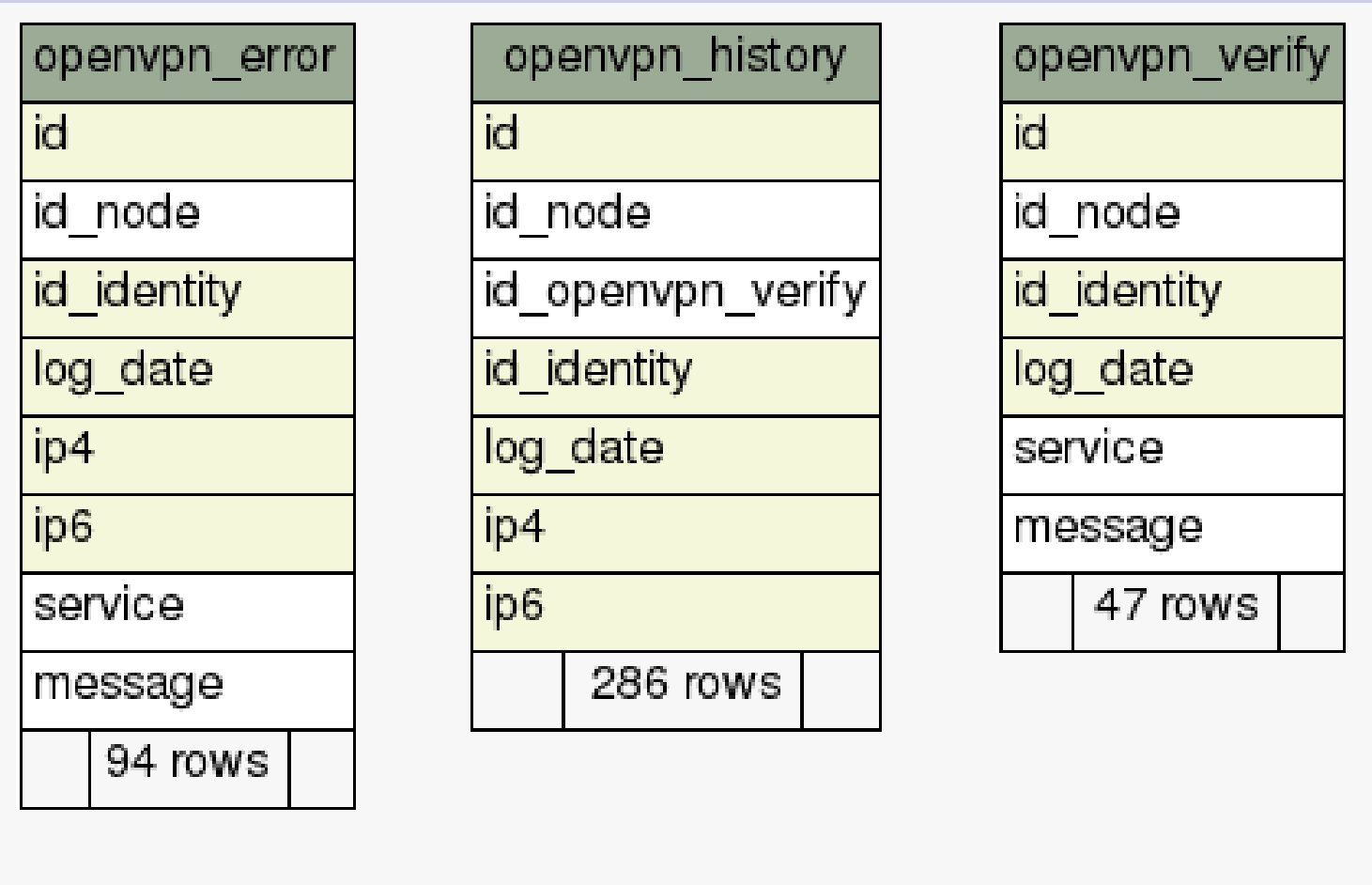
Datenbank - Identities



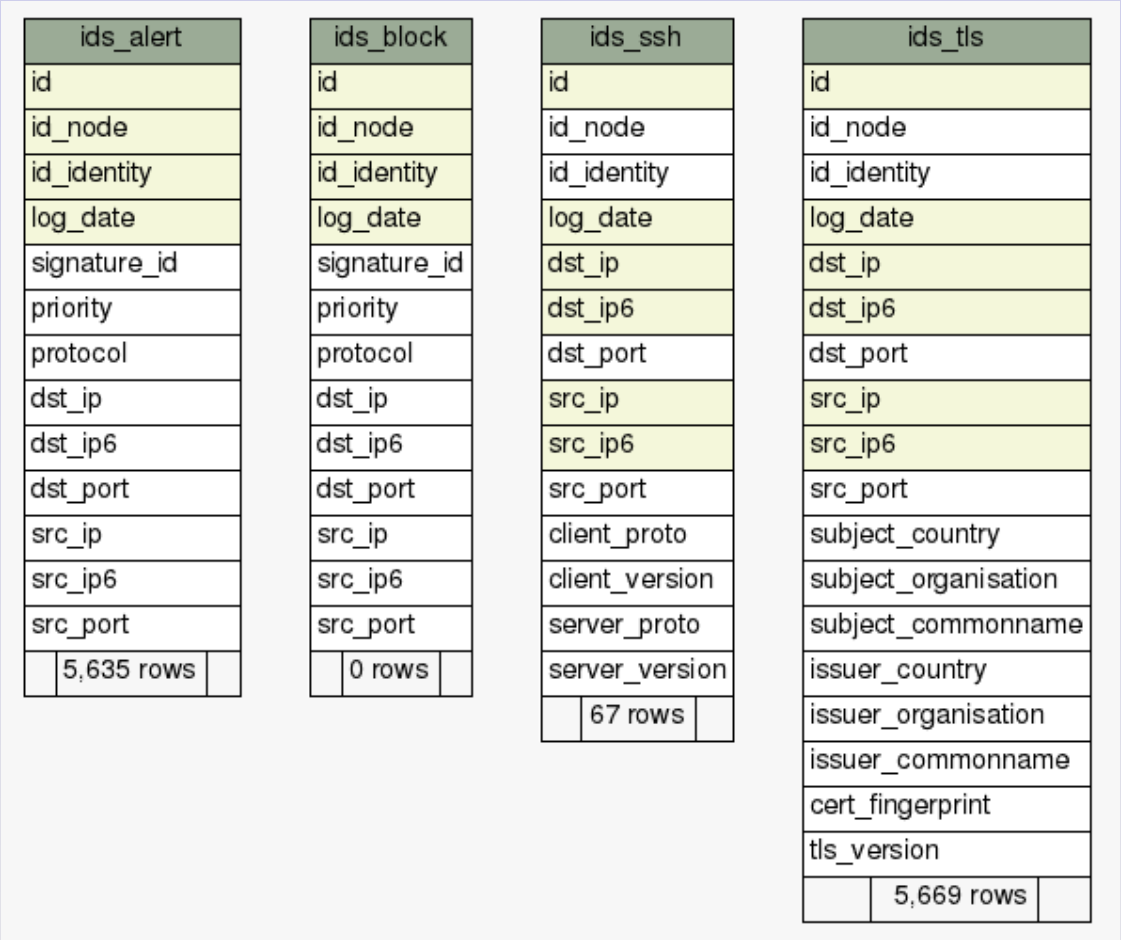
Datenbank - Authentication



Datenbank - OpenVPN™



Datenbank - IDS



Prototyp Status

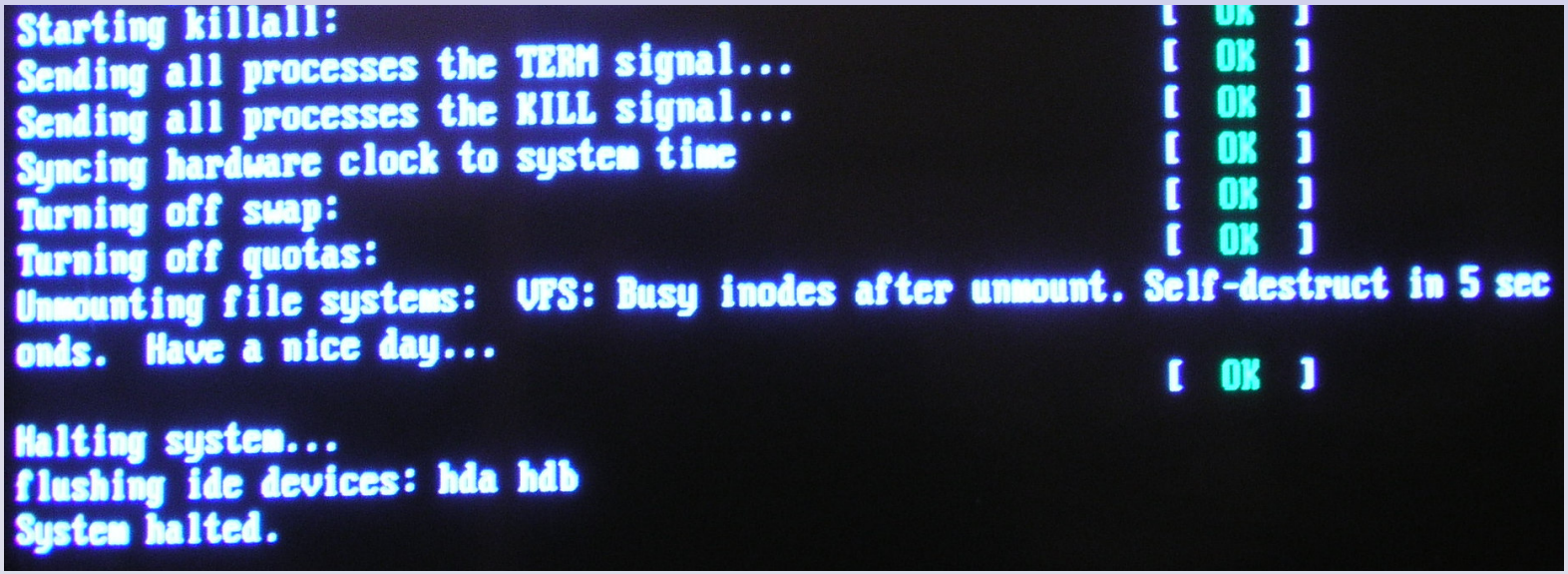
- Extractor, Collector, Dissector, Janitor implementiert
- Code und MariaDB schaffen 20+ Nodes mit 1+ GB Logs auf Intel® Atom™
- Analyse direkt durch Andocken an MariaDB möglich
  - SQL, Stored Procedures
  - C++, Perl, Python, ...
  - R
- brauchbares Werkzeug, sehr ausbaufähig
- sehr leichtes Deployment

Prototyp Aussichten

- *not implemented*
  - Queuing von Logs wenn Collector nicht erreichbar
  - Detektieren von Jahreswechsel bei Syslog Logs
  - Behandlung neuer/alter X.509 Zertifikate Collector/Extractor
- *testing*
  - Datenbankstruktur (Skalierbarkeit, Performance bei Analysen)
  - Zuverlässigkeit der Muster im Dissector
  - korrekte Behandlung von defekten Logdaten/-nachrichten
- Echtzeitmodus durch Änderung Nachrichtenformat
  - Kopplung Extractor/Collector/Dissector via ØMQ
  - Versand von n Logzeilen pro Nachricht
  - wenig Änderungen am Code notwendig

Zusammenfassung

Noch Fragen?


Starting killall:
Sending all processes the TERM signal...
Sending all processes the KILL signal...
Syncing hardware clock to system time
Turning off swap:
Turning off quotas:
Unmounting file systems:
VFS: Busy inodes after unmount. Self-destruct in 5 seconds. Have a nice day...
[OK]
Halting system...
flushing ide devices: hda hdb
System halted.

Quellen

- ØMQ
- C++11
- Graylog
- Logstash
- MariaDB
- Perl
- R

Über dieses Dokument

- Autor: René Pfeiffer
- Erstellt mit LaTeX und LaTeX Beamer Class
- Dokumentensammlung unter
  https://web.luchs.at/information/docs.php