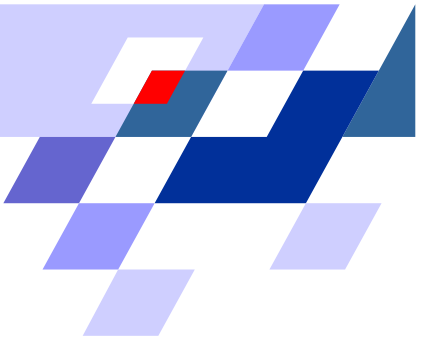
UNIVERSITÄT DORTMUND

Fachbereich Informatik
Lehrstuhl 6 — Informationssysteme und Sicherheit



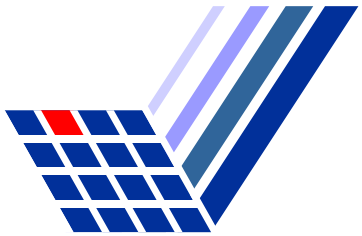
Zehnter DFN-CERT-Workshop

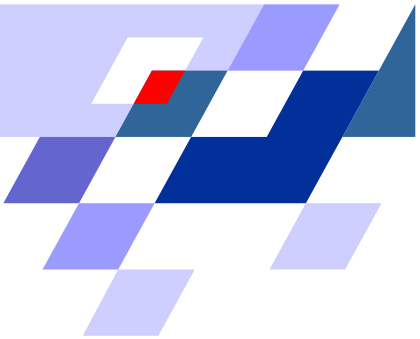
SICHERHEIT IN VERNETZTEN SYSTEMEN

Praktikabler Datenschutz für Log-Daten

Ulrich Flegel

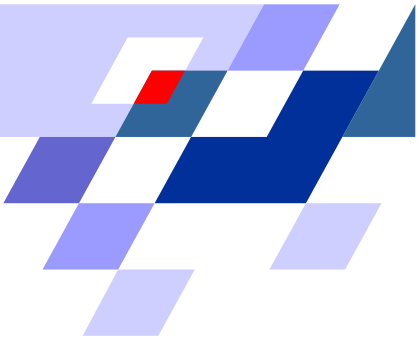
Februar 2003, Hamburg





Überblick

- Anonymität vs. Zurechenbarkeit
  - Interessenträger
  - Audit-Daten
  - Datenschutz-Anforderungen
- fairer Interessenausgleich durch Pseudonyme
  - Datenschutz-konforme Aufdeckung
  - technische Zweckbindung
  - Anforderungen
- Pseudo/CoRe
  - Eigenschaften
  - Ansatz
  - Vertrauensmodell
  - Implementierung
  - Wissensmodellierung
  - Einbettung
  - Architektur
  - Laufzeitverhalten
- Zusammenfassung
- Koordinaten



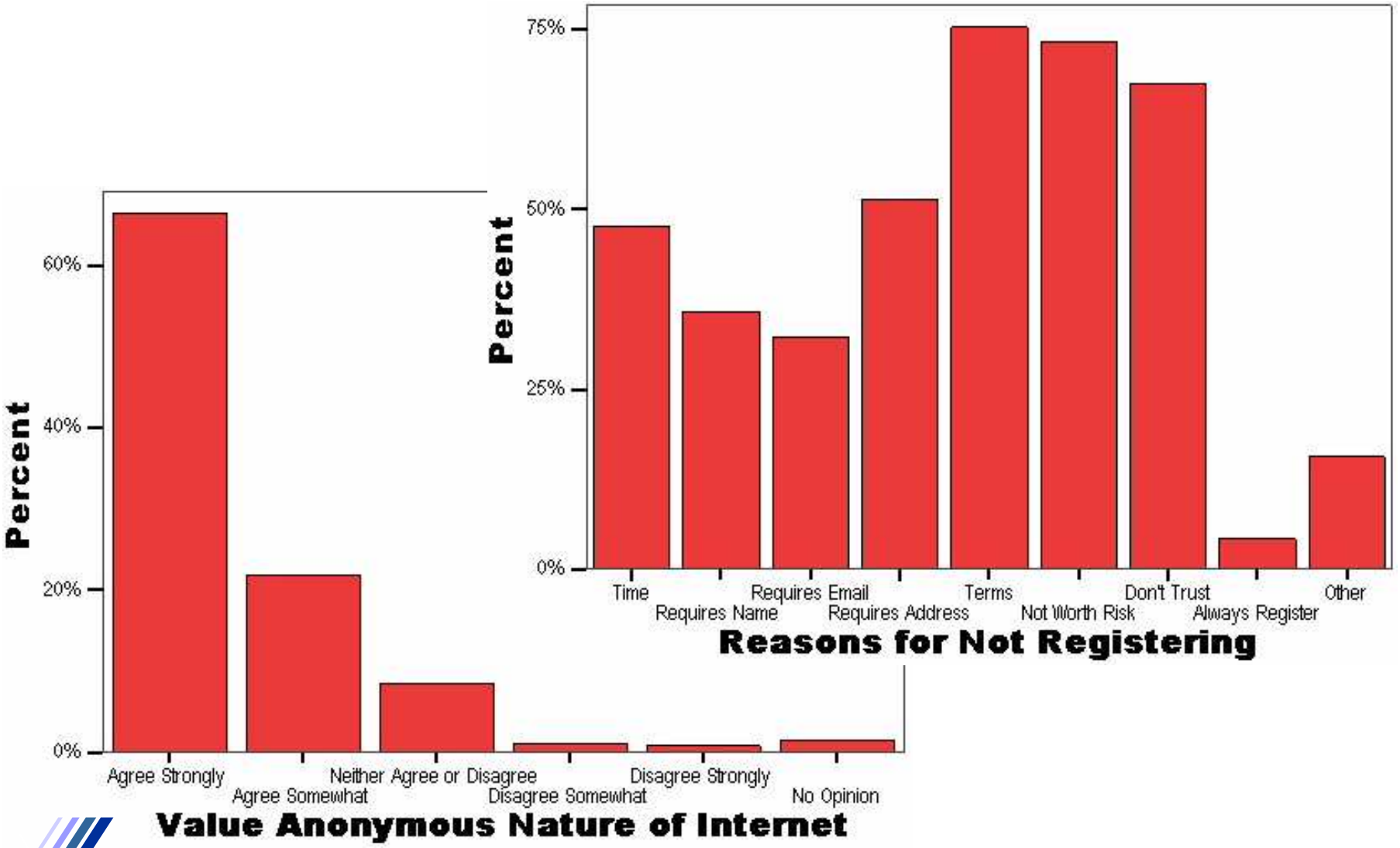
Interessen-Geflecht

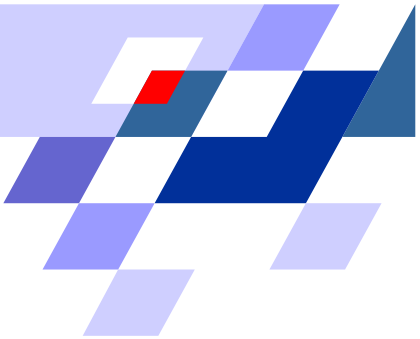
Nutzer	Dienstanbieter
Anonymität:	
• steigende Akzeptanz	• Wettbewerbsvorteil
• Recht auf informationelle Selbstbestimmung	• Datenschutzrechtliche Vorgaben

Überwachung / Zurechenbarkeit zur Abwehr von:

• Interessenverletzungen durch Dritte	• IT-Mißbrauch durch Nutzer
• falschen Anschuldigungen	• Nutzerkonflikten

Anonymität I — Akzeptanz
Quelle: Zehnte GVU WWW-Nutzer-Umfrage





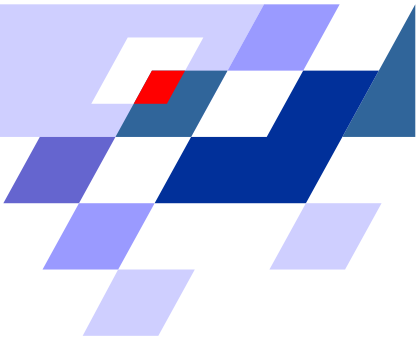
Überwachung / Zurechenbarkeit I — Audit-Daten

Audit-Daten: (=Log-Daten)

- sind Nutzungsdaten, enthalten häufig personenbezogene Merkmale
- dokumentieren System-Ereignisse → Nutzerverhalten
- ermöglichen Leistungsüberwachung, auch der Nutzer

Audit-Daten-Analyse-Ziele: (hier: Mißbrauchsentdeckung)

- erfordert Analysierbarkeit der Audit-Daten (Verkettbarkeit):
  - Bestimmung des Schadens (z.B. für Versicherung)
  - Bestimmung von Maßnahmen zur Schadensbeseitigung
  - Bestimmung der ausgenutzten Verwundbarkeiten → Beseitigung
- erfordern Zurechenbarkeit von Audit-Daten (Aufdeckbarkeit):
  - Schadensbegrenzung
  - Rechtsverfolgung



Anonymität II — Datenschutz-Anforderungen

personenbezogene Daten: Erlaubnisvorbehalt bei Erhebung, Speicherung, Verarbeitung

bei Erlaubnistatbestand: weitreichende Pflichten, z.B. Zweckbindung, Löschpflicht, Meldepflicht

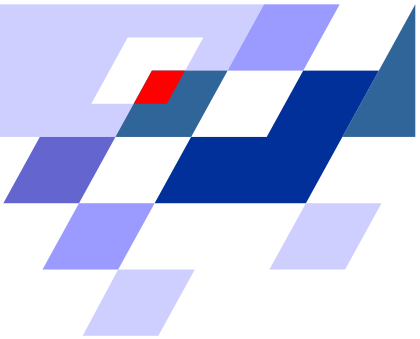
bei Nutzungsdaten: erlaubt, falls zur Inanspruchnahme notwendig und solange Nutzung andauert (s. Audit-Daten)

geltende Datenschutzgesetze: je nach Dienst mehrere u.U. verschiedene

betriebliche / dienstliche Dienste:

- Duldung privater Nutzung → Recht auf informationelle Selbstbestimmung
- Einführung von zur Leistungsüberwachung geeigneten Technologien → Mitbestimmungsrecht des Betriebsrats (s. Audit-Daten)

komplexe rechtliche Situation & Einschränkungen → Schwierigkeit von gesetzeskonformem Einsatz Audit-Daten-gestützter Schutzmaßnahmen



Mehrseitige Sicherheit

Konflikt der Sicherheitsanforderungen:

- Zurechenbarkeit im Mißbrauchsfall
  zum Schutz der Betroffenen
- Interesse Einzelner an Datenschutz / Anonymität

Ansätze:

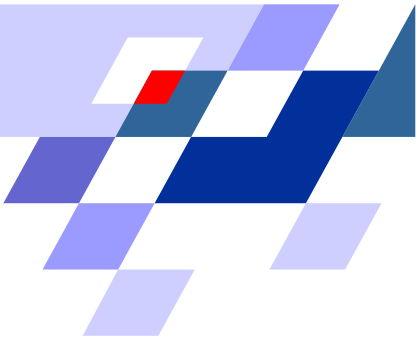
unter Berücksichtigung der Anwendungssituation / -umgebung

Einseitig:

Aufgeben einer Anforderung zugunsten der anderen

Mehrseitig:

Fairer Ausgleich der Sicherheitsinteressen aller Beteiligten



Mehrseitige Sicherheit durch Pseudonyme

Relativität des Personenbezugs durch Zusatzwissen:
durch Kontrolle des Zusatzwissens Unterscheidung von

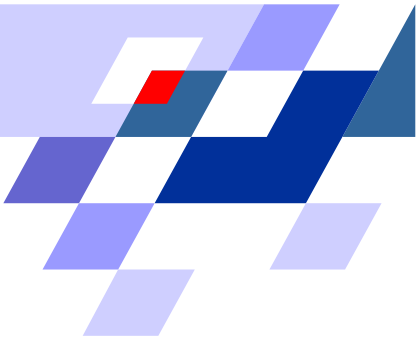
Regelfall: Anonymität / keine Zurechenbarkeit

Ausnahmefall: Zurechenbarkeit möglich

Datenvermeidung bzw. -sparsamkeit: (§3 Abs. 4 TDDSG / §12 Abs. 5 MDstV)
für System- und Selbstdatenschutz Realisierung durch

anonyme / pseudonyme Inanspruchnahme: (§4 Abs. 1 TDDSG / §13 Abs. 1 MDstV)

Pseudonymisierung: ohne Zuordnungsregel (=Zusatzwissen) Daten nur
mit unverhältnismäßig hohem Aufwand (Zeit, Kosten, Arbeitskraft)
einer bestimmten natürlichen Person zuordenbar (praktisch anonym)
$\longrightarrow$ nur für Kenner der Zuordnungsregel gelten Datenschutz-Pflichten
$\longrightarrow$ für andere entfallen die mit der Verarbeitung verbundenen Pflichten



Zurechenbarkeit II — Pseudonym-Aufdeckung

Aufdeckung: erneute Geltung der Datenschutz-Pflichten

Erlaubnistatbestände: sonst Aufdeckung rechtswidrig

Strafverfolgung: wenn Daten-Relevanz während Nutzung absehbar (§6(3) TDDSG)

Mißbrauch: wenn Anhaltspunkte für Teledienst-Mißbrauch dokumentiert (§6 Abs. 8 TDDSGÄndG)

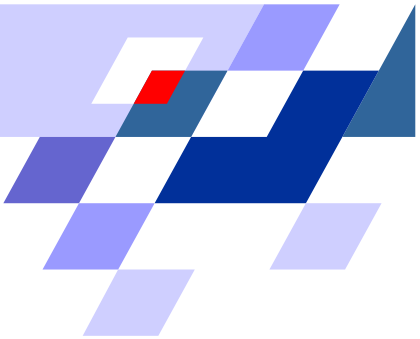
Annahme: möglich im Einklang mit den Datenschutz-Gesetzen sind

- Speicherung pseudonymisierter Anhaltspunkte (Audit-Datensätze) zu vordefinierten Mißbrauchs-Anfangsverdachten
- Zurechenbarkeit der Anhaltspunkte durch Pseudonym-Aufdeckung, die Mißbrauchs-Anfangsverdacht erfüllen

Vorsorge-Regelungen: u.a.

Transparenz: Mißbrauch → Aufhebung der Anonymität

Sicherheit: der Pseudonymitäts-Eigenschaft



Zurechenbarkeit III — Technische Zweckbindung

Erlaubnistatbestände für die Herstellung von Zurechenbarkeit:

- Strafverfolgung (von Dienst-Mißbrauch)
- Teledienst-Mißbrauch

Herstellung von Zurechenbarkeit:

- Mißbrauchs-Anfangsverdacht in anonymen Audit-Daten entdecken
- Pseudonyme personenbezogener Merkmale in Audit-Daten aufdecken

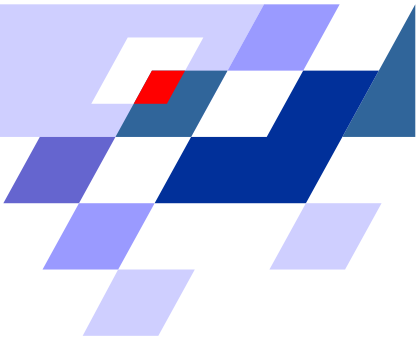
Annahme:

Mißbrauchs-Anfangsverdacht entdeckt → Zurechenbarkeit rechtmäßig

Fairness: technisch unumgehbar durchgesetzt

Regelfall: Nutzer erscheinen unter Pseudonymen

Ausnahmefall: Pseudonym-Aufdeckung ausschließlich bei Anfangsverdacht und auf betroffene personenbezogene Merkmale beschränkt



Pseudonymisierung — Anforderungen

Performanz:

- unmittelbarer Abtransport aus Kontrollbereich des Angreifers
- hohes Audit-Daten-Volumen "on-the-fly" bewältigen
- keine/geringe Beeinträchtigung der Dienst-Antwortzeiten

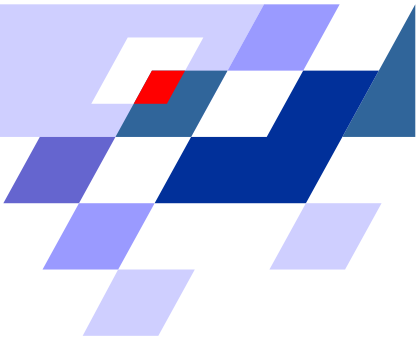
→ keine aufwendige Kryptographie

zeitnahe Aufdeckbarkeit:

→ ohne Dritte → technische Zweckbindung

praktikabel umsetzbar:

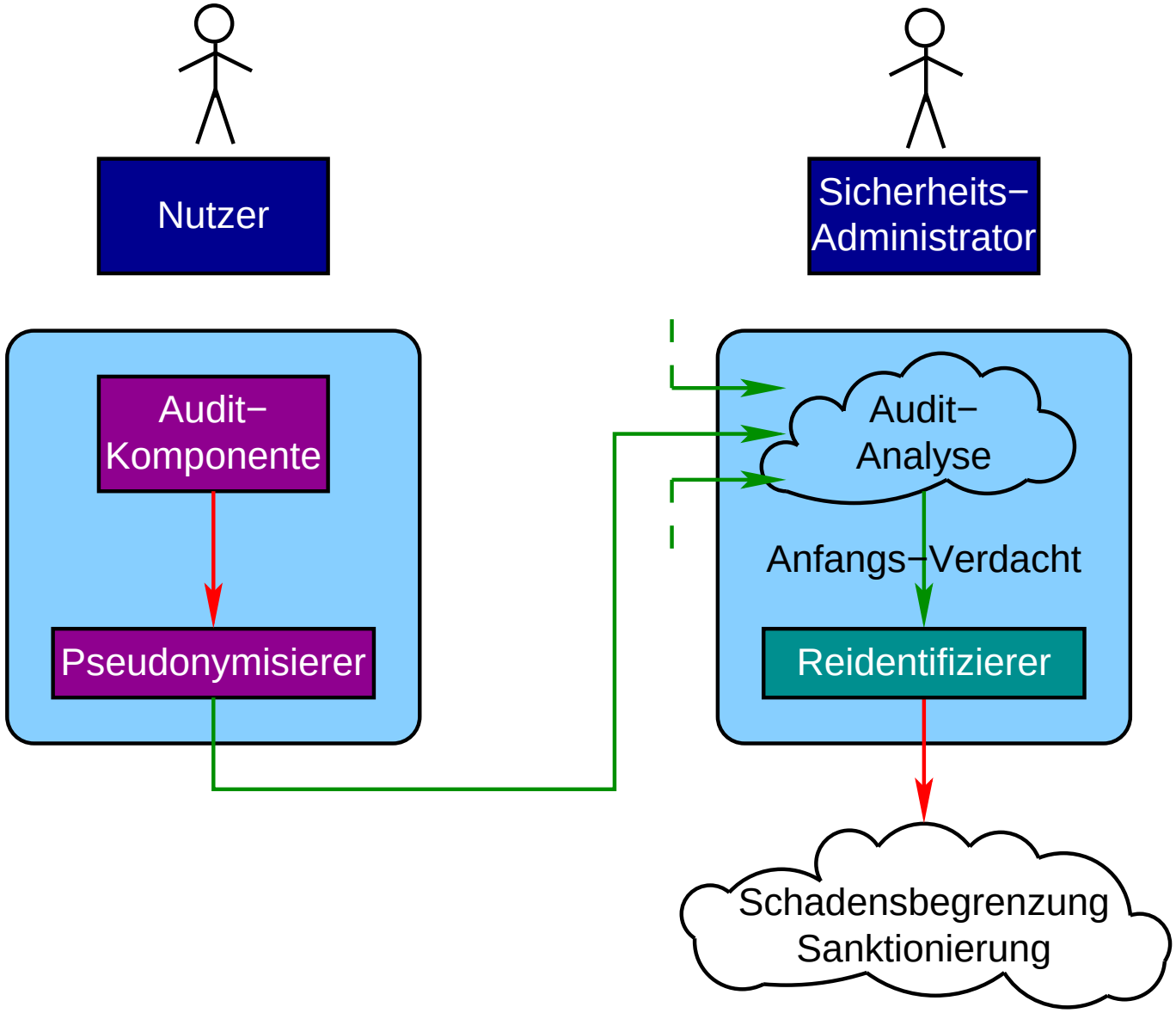
- Datenschutz-Pflichten & Nutzer-Fehlbedienung
  → unabhängig vom Nutzer
- keine aufwendige Infrastruktur notwendig

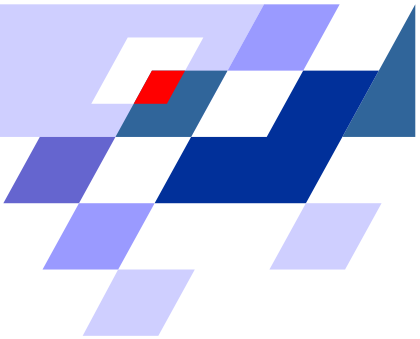


Pseudo/CoRe · Pseudonymization with Conditional Recovery

- Nutzer erscheinen unter Pseudonymen
  bis zu einem Mißbrauchs-Anfangsverdacht:
  – ausschließlich betroffene Pseudonyme aufdeckbar
  → Zurechenbarkeit herstellbar
- Überwachung möglich auf pseudonymisierten Audit-Daten
  → keine Einschränkungen durch Datenschutz-Pflichten
- technische Zweckbindung bei der Pseudonymaufdeckung
  → Fairer Interessenausgleich Überwachung ↔ Datenschutz
- sofortige Pseudonymaufdeckung (unabhängig von Dritten)
- umsetzbar unabhängig von Nutzern und Infrastruktur

Pseudo/CoRe · Pseudonymization with Conditional Recovery





Ansatz — Anfangsverdacht & Pseudonymisierung

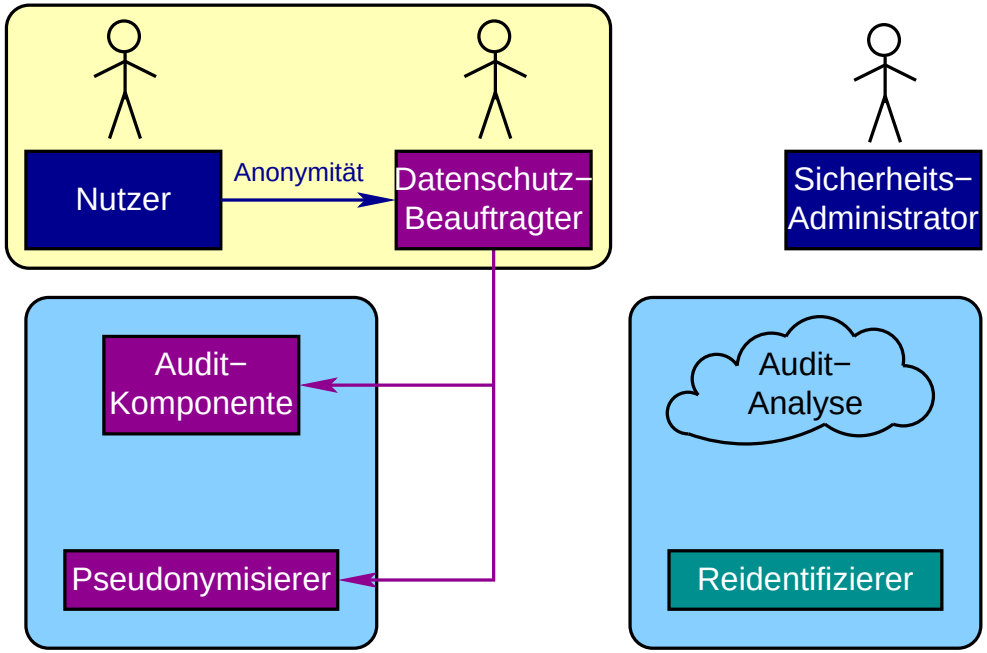
Modell der Anfangsverdachte: Menge von Anhaltspunkten
für potentiell angriffsbezogene Aktivitäten überschreitet Schwellenwert

Aufdeckbarkeit: ausschließlich bei überschrittenem Schwellenwert

Kryptographische Primitive:

- Chiffrieren personenbezogener Merkmale
- Aufteilen des Dechiffrier-Schlüssels mit Hilfe eines modifizierten Shamir'schen Schwellenwert-Schemas zur informationstheoretisch sicheren Geheimnisteilung
- Rückgewinnung des Dechiffrier-Schlüssels aus den Anteilen mittels Lagrange-Interpolation

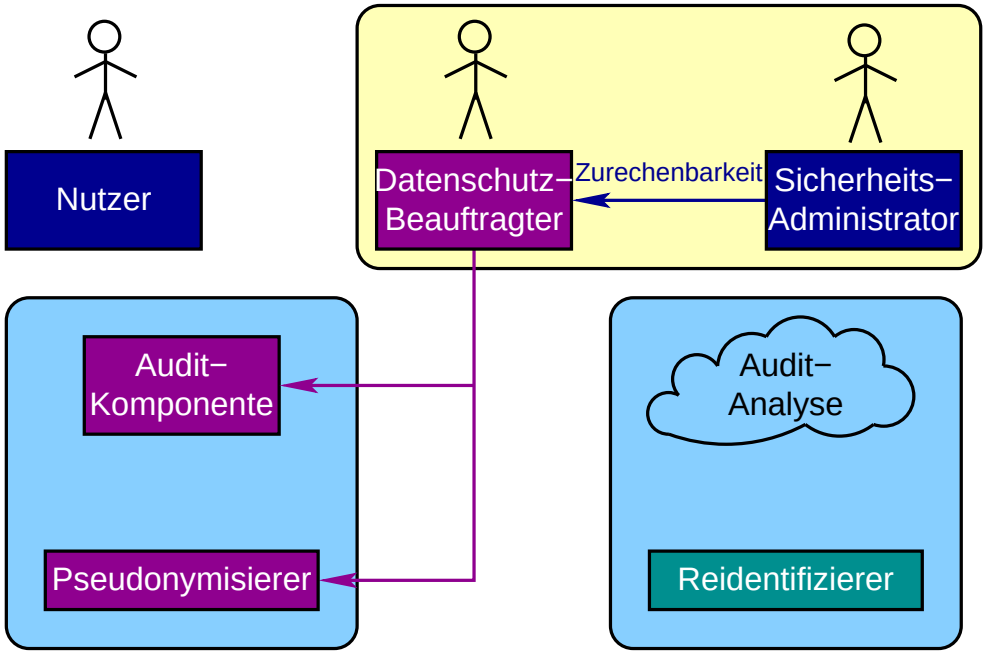
Vertrauensmodell I — Anonymität



Datenschutzbeauftragter:

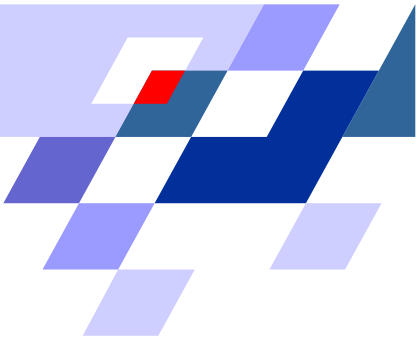
- hat Vertrauen der Nutzer bzgl. deren Anonymität
- identifiziert personenbezogene Merkmale, die pseudonymisiert werden müssen (ggf. in Zusammenarbeit mit dem Betriebsrat)

Vertrauensmodell II — Zurechenbarkeit



Datenschutzbeauftragter:

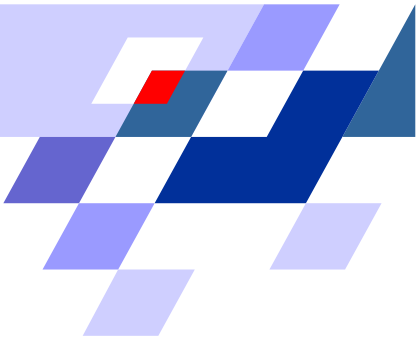
- hat Vertrauen der Sicherheits-Adms bzgl. Zurechenbarkeit
- modelliert Anfangsverdachte, die Pseudonym-Aufdeckung rechtfertigen, in Zusammenarbeit mit den Sicherheits-Adms



Implementierung

- implementiert für Unix-Systeme
- erfolgreich getestet: *Solaris, OpenBSD, Linux*
- ASCII-Audit-Daten-Format: z.B. *Syslog*, Web-Server
  - anwendbar auf Audit-Daten vieler Dienste
  - Windows-Systeme und Netzwerk-Komponenten über *Syslog* integrierbar

$\longrightarrow$ hohe erreichbare Abdeckung anfallender Audit-Daten



Wissens-Modellierung

Matching: beliebige Ereignis- und Merkmals-Typen → reguläre Suchausdrücke

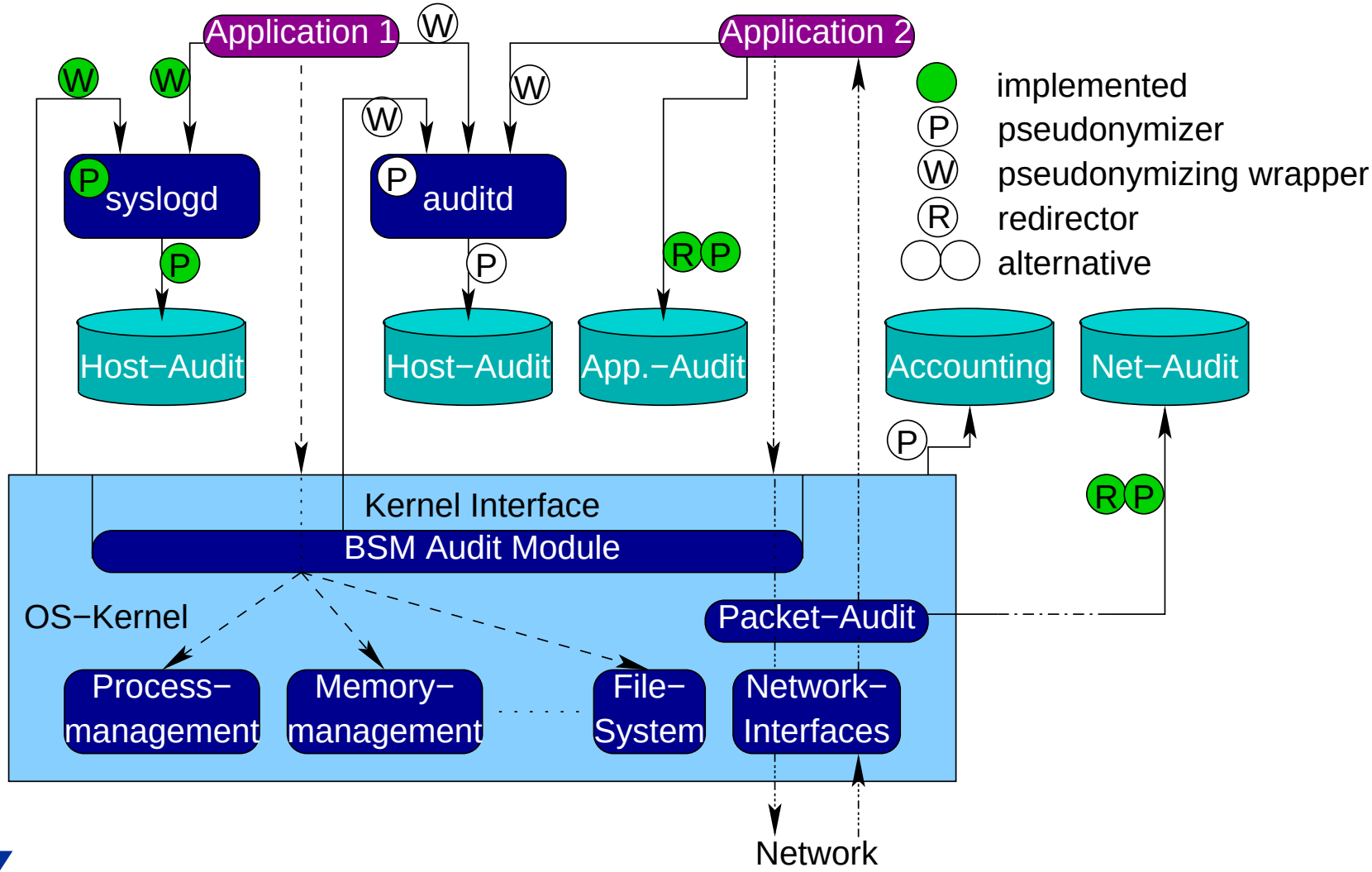
Format-neutral: Legacy-Analyse verwendbar

- **Syntax:** Pseudonym-Daten-Typen
  - **Integer & String:** Länge beibehalten / fixieren
  - **IP- & DNS-Adresse:** Anzahl der Stufen
- **Verkettbarkeit:** optional, falls von Analyse benötigt

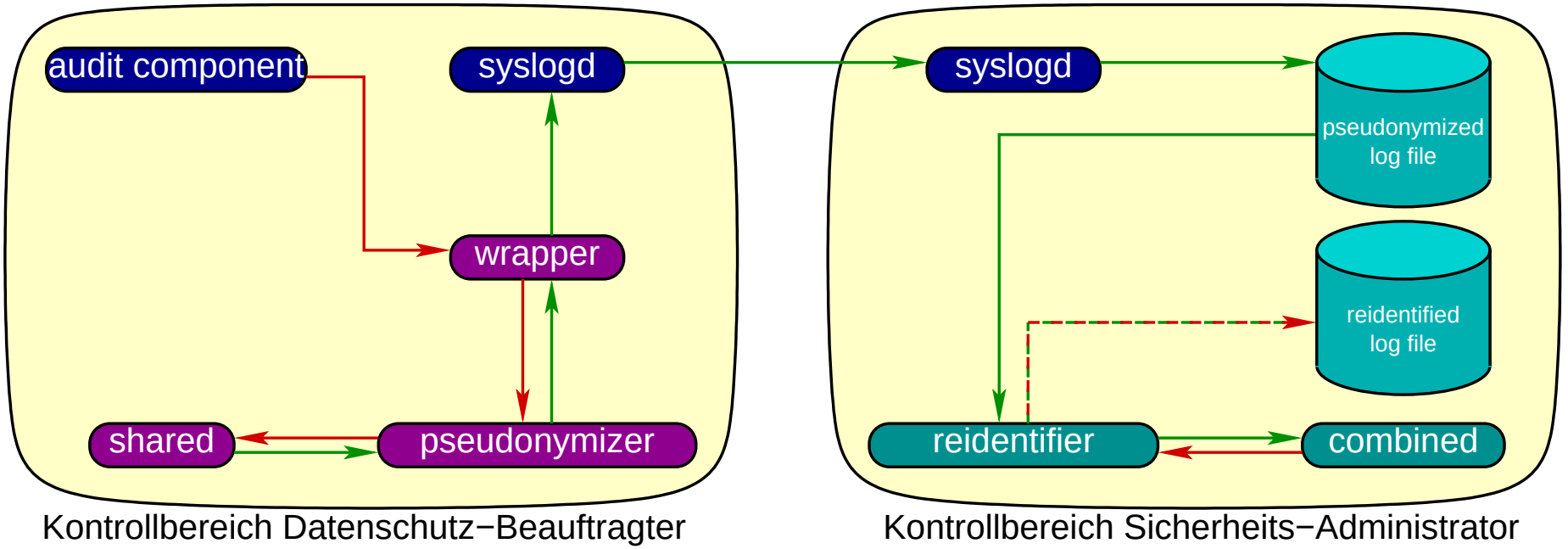
Zweckbindung: der Aufdeckung

- **organisatorisch:** Unterstützung für optionales Escrow
- **technisch:** Modellierung der Mißbrauchs-Anfangsverdachte
  - **Schwellenwert:** Überschreitung → Aufdeckbarkeit
  - **Kontext:** Anhaltspunkte (Ereignisse) für Anfangsverdacht
  - **Wiederholung:** von Anhaltspunkten (in-) signifikant
  - **Gewicht:** Anhaltspunkt verstärkt / entkräftet Anfangsverdacht

Einbettung von Pseudonymisierern (*Solaris*)

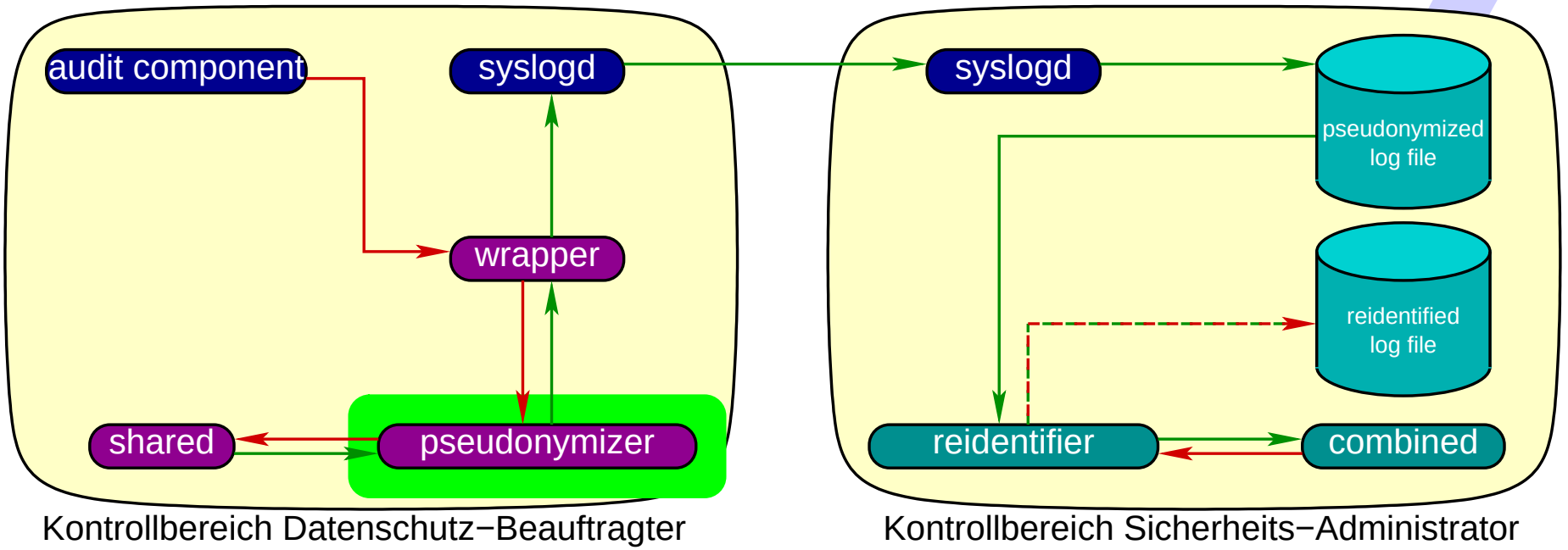


Architektur — Integration mit *Syslog* (vereinfacht)



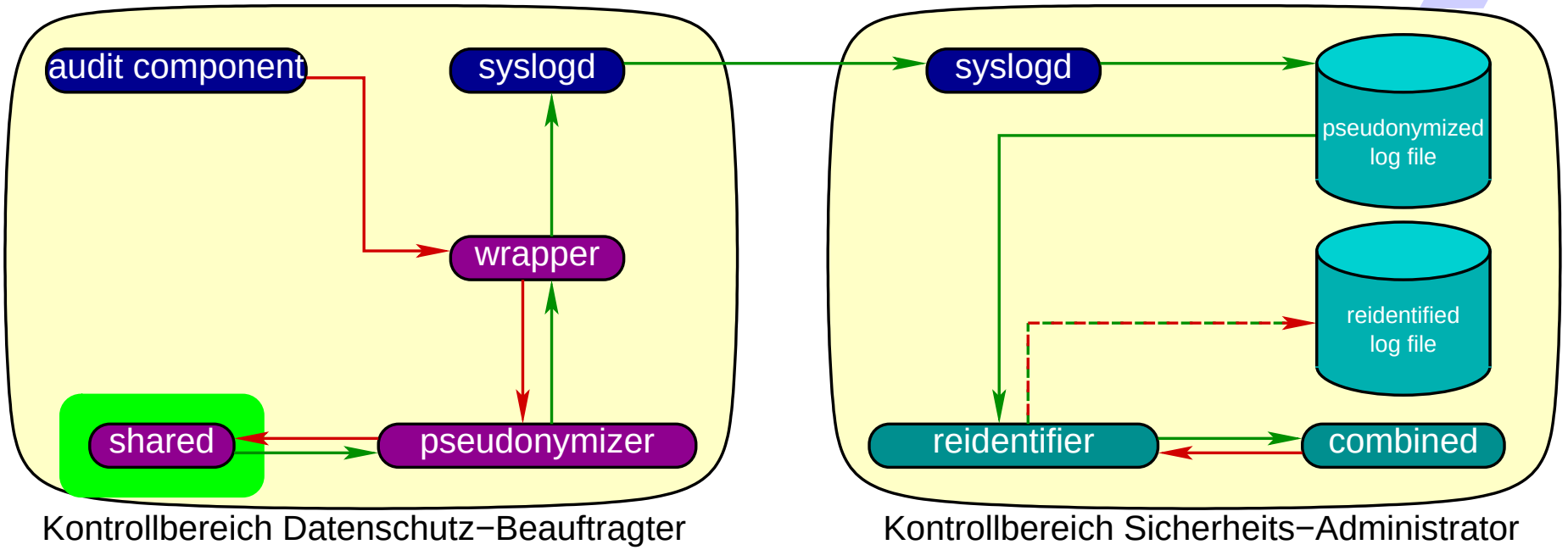
- Alternativen für Syslog-Integration: Wrapper / Patch / Pipes
- personenbezogene Merkmale nicht in Audit-Dateien

Pseudonymisierer I — `pseudonymizer`



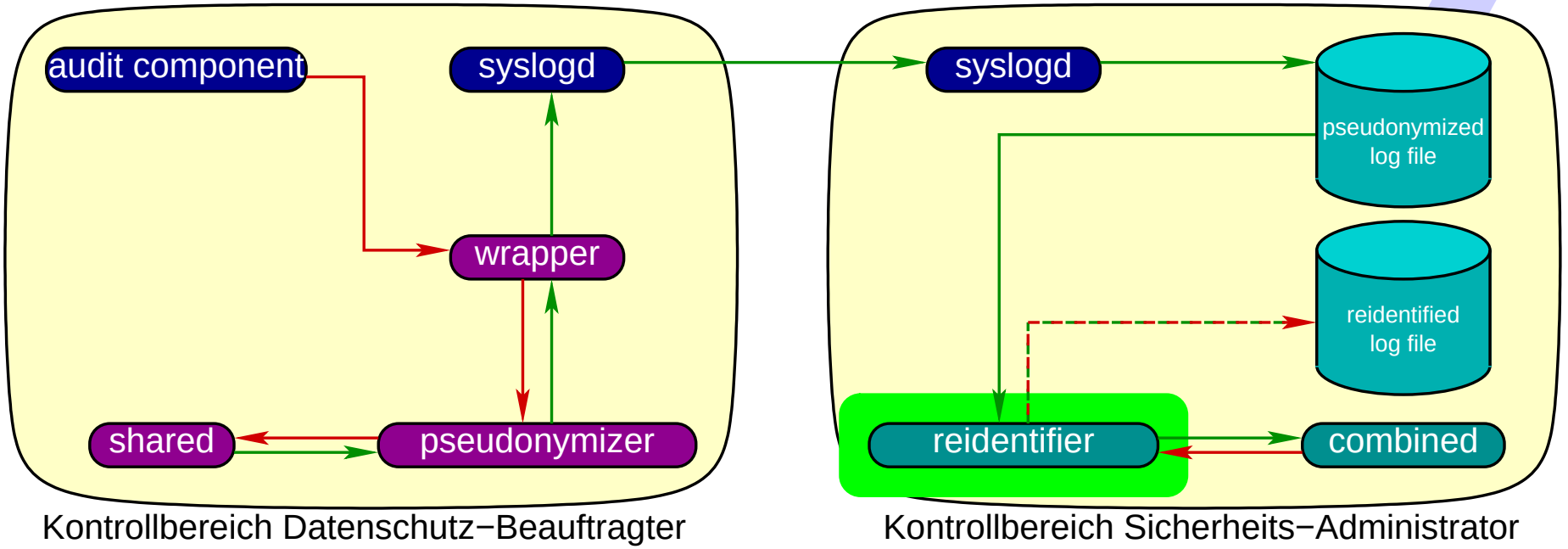
- bestimmt personenbezogene Merkmale (reguläre Suchausdrücke)
- bestimmt Anfangsverdachts-Kontexte
- ersetzt Merkmale formatneutral durch Pseudonyme (legacy-kompatible Syntax & Verkettbarkeit)

Pseudonymisierer II — `shared`



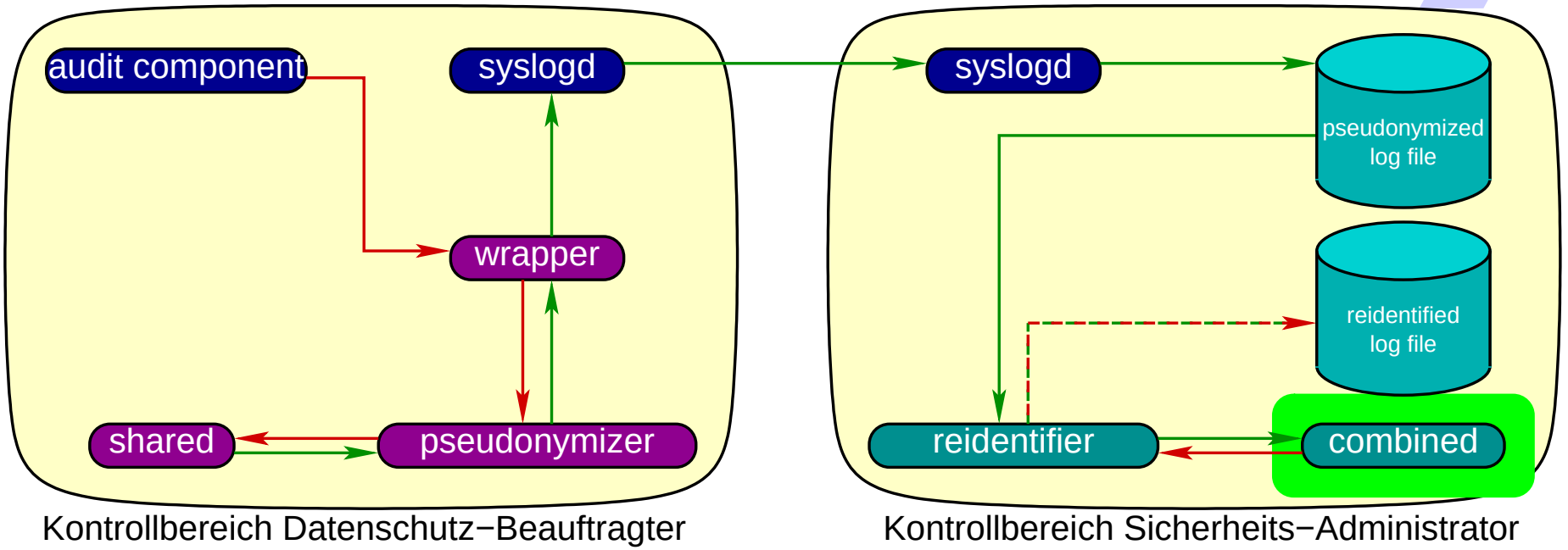
- chiffriert Merkmale (OpenSSL: Blowfish, SHA1)
- teilt Dechiffrier-Schlüssel auf (GMP & NTL)
- liefert kryptographisches Material (filterbar)

Reidentifizierer I — `reidentifier`

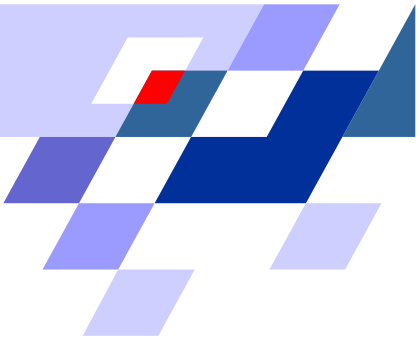


- manuell oder automatisch bei Alarm gestartet
- ordnet kryptographisches Material Anfangsverdachts-Kontexten zu
- deckt Pseudonyme erfüllter Anfangsverdachts-Kontexte auf
- ersetzt Pseudonyme durch Merkmale

Reidentifizierer II — `combined`



- nutzt kryptographisches Material
- Lagrange-interpoliert Dechiffrier-Schlüssel (GMP)
- dechiffriert Merkmale (OpenSSL: Blowfish, SHA1)



Laufzeitverhalten

- "on-the-fly"-Pseudonymisierung

$\longrightarrow$ hinreichend performant:

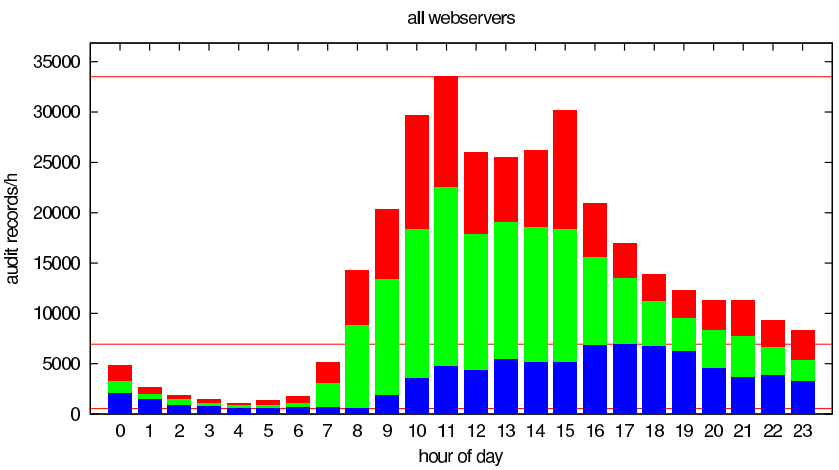
Zentraler Web-Server der Universität Dortmund:

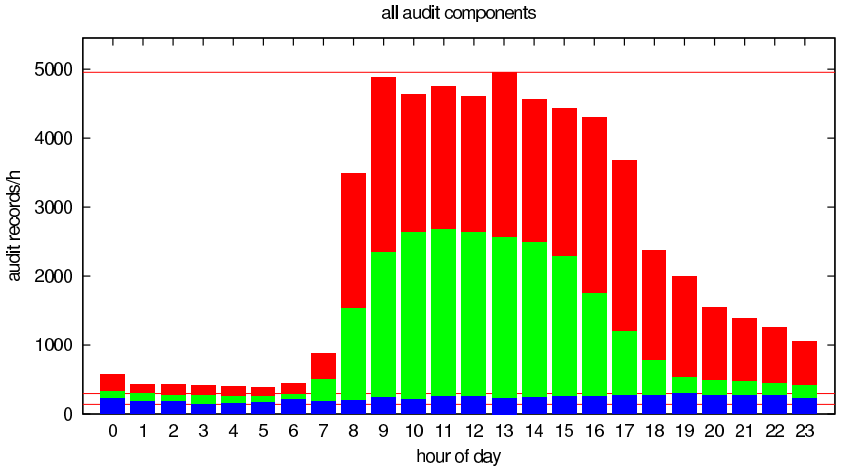
- 37 *Apache* Web-Server, weltweit erreichbar
- *Syslog*: monatl. ~ 112.000 FTP-Transfers, ~ 45.000 Emails
- Audit-Datensätze gesamt: max. 38462/h $\approx$ 11/s

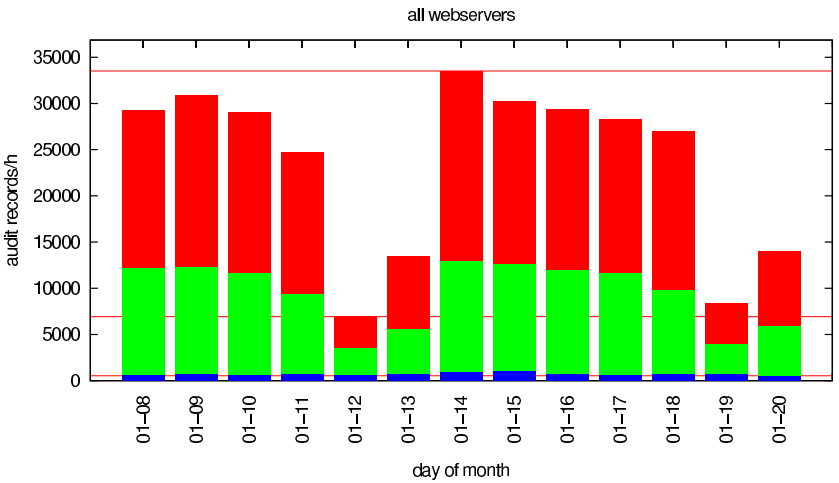
Labor-Benchmarks: Kryptographie mit 128 Bits

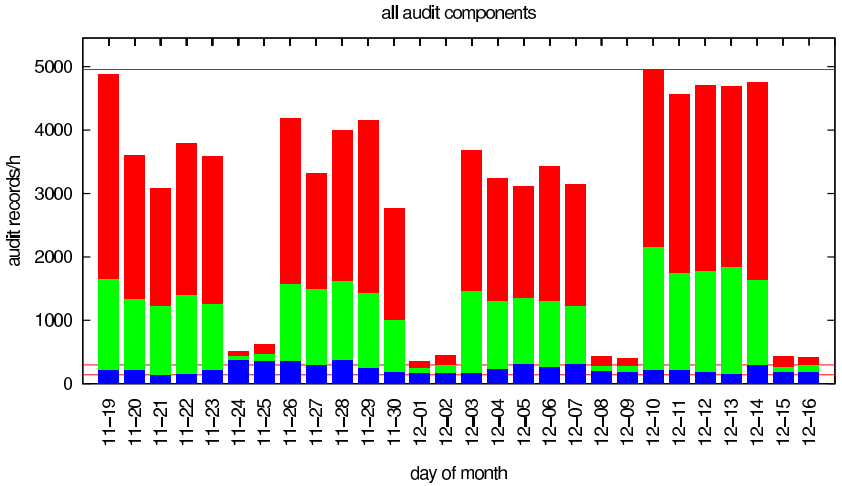
- zwischen 1060 und 70 Audit-Datensätzen pro Sekunde

Audit-Daten-Volumen in Datensätzen/h

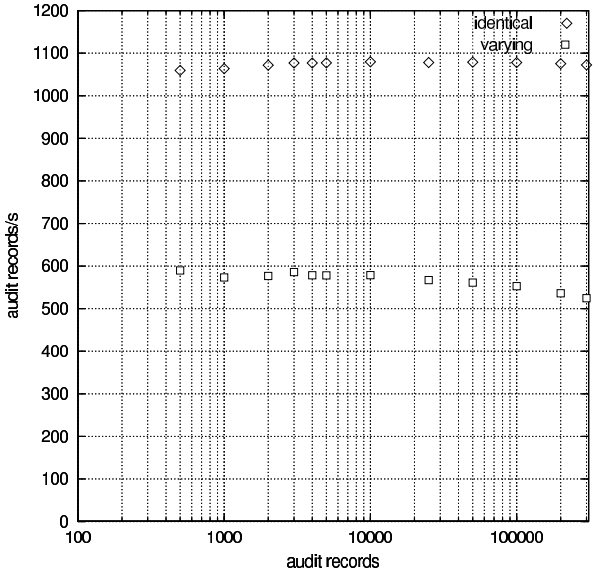


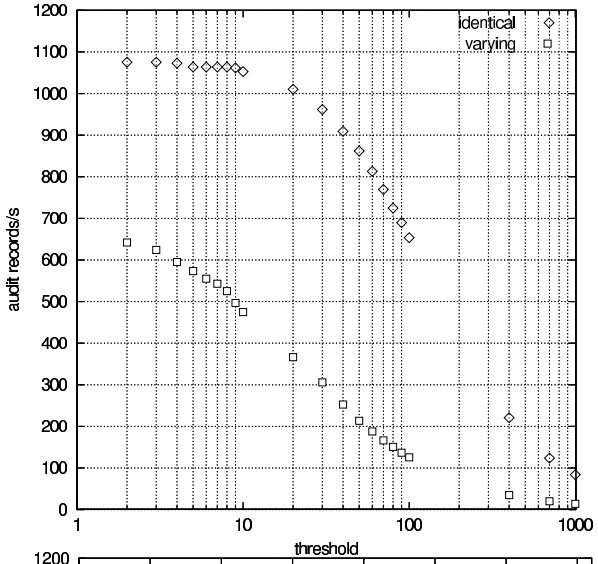


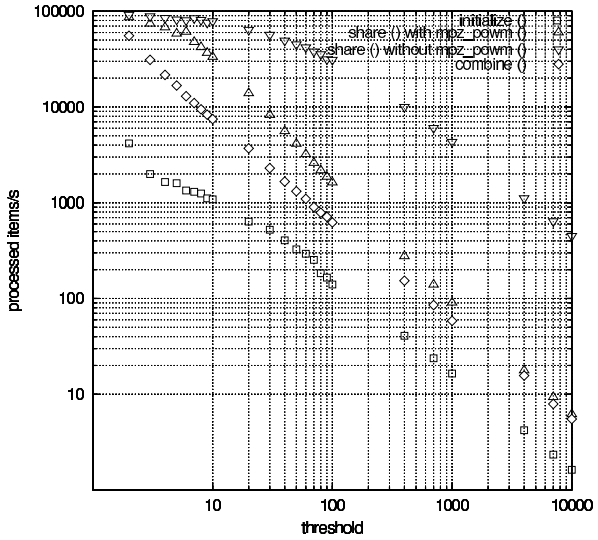


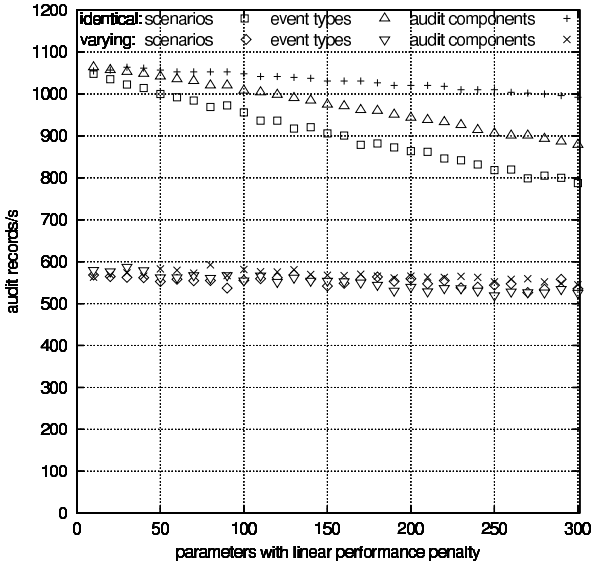


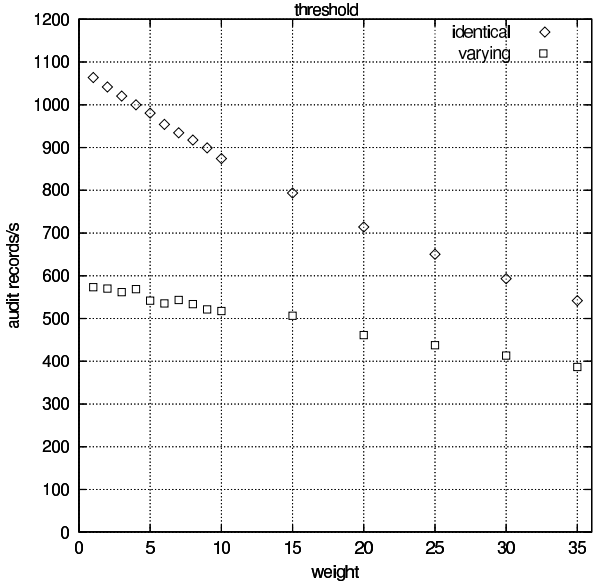
Leistungsmessungen

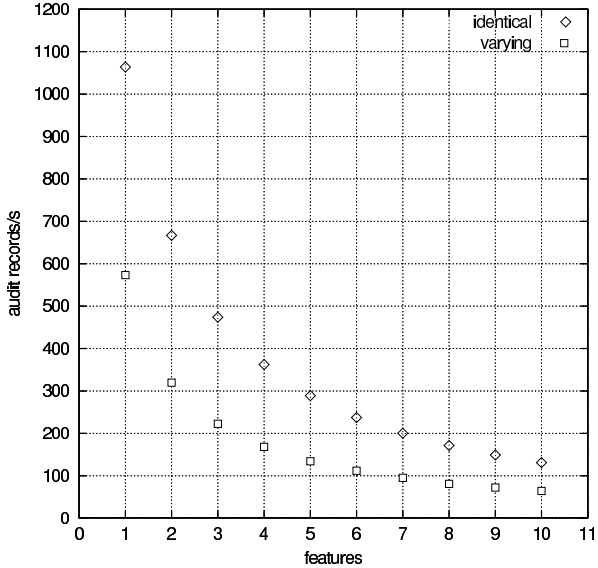


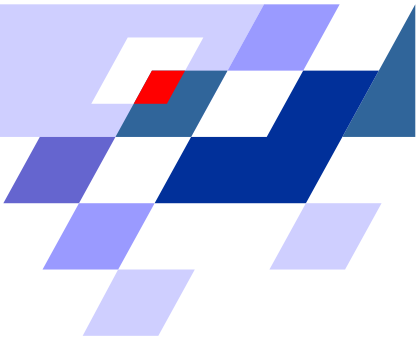












Zusammenfassung

- Analyse pseudonymisierter Audit-Daten ist möglich

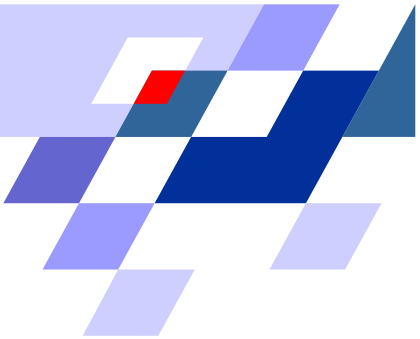
→ Einschränkungen durch Datenschutz-Pflichten entfallen bei der Analyse

- Zweckbindung bei der Pseudonym-Aufdeckung ist technisch durchsetzbar

→ fairer Interessenausgleich von Datenschutz und Zurechenbarkeit

- einschlägige Pseudonymisierungs-Anforderungen sind erfüllbar:
  - Nutzer-Unabhängigkeit
  - Infrastruktur-Unabhängigkeit
  - Performanz
- Konzepte implementiert und Software verfügbar unter GPL

→ Datenschutz nachrüstbar für Audit-Daten existierender Unix-Systeme



Software

Site: `http://ls6-www.cs.uni-dortmund.de/pseudocore`

Support: `pseudo-support@ls6.cs.uni-dortmund.de`

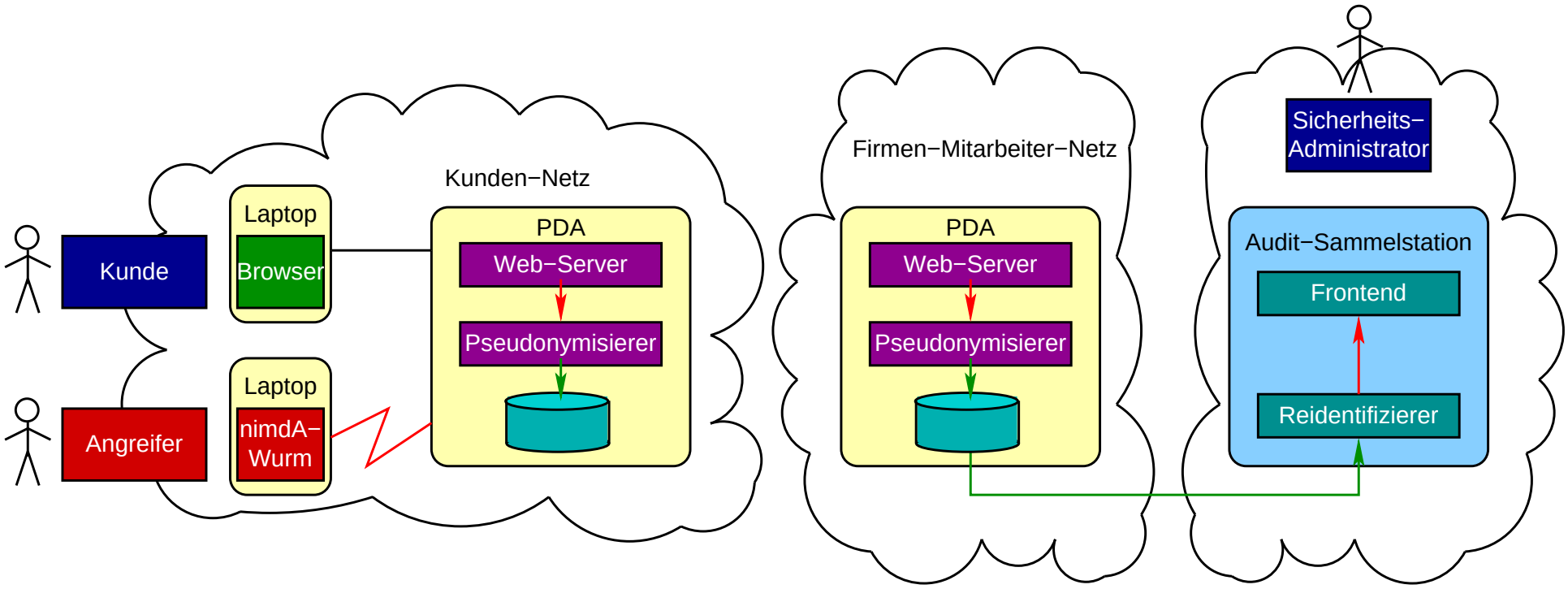
Kontakt

Ulrich Flegel

WWW: `http://ls6-www.cs.uni-dortmund.de/~flegel`

Email: `ulrich.flegel@udo.edu`

Beispiel-Anwendung: mobiler Web-Server



Pseudonymisierte Audit Daten

```
Feb 24 18:02:30 localhost boa: Access from W4w4dO to "GET I9YoeI HTTP/1.0" 200 6944
Feb 24 18:02:31 localhost boa: Access from ZmrB2j to "GET lP6JTc HTTP/1.0" 404 0
Feb 24 18:02:31 localhost boa: Access from SwzK08 to "GET l52Wpg HTTP/1.0" 404 0
Feb 24 18:02:31 localhost boa: Access from ntqJVe to "GET DnYJ0j HTTP/1.0" 404 0
·Feb 24 18:02:31 localhost boa: Access from xovqfO to "GET /_mem_bin/..%255c../..%255c../..%255c../winnt/system32
·Feb 24 18:02:31 localhost boa: Access from rB4QA5 to "GET /_vti_bin/..%255c../..%255c../..%255c../winnt/system32
·Feb 24 18:02:31 localhost boa: Access from olpbk2 to "GET /c/winnt/system32/cmd.exe?/c+dir HTTP/1.0" 404 0
·Feb 24 18:02:32 localhost boa: Access from hGxhQJ to "GET /d/winnt/system32/cmd.exe?/c+dir HTTP/1.0" 404 0
·Feb 24 18:02:32 localhost boa: Access from l9a3Jy to "GET /default.ida?NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
·Feb 24 18:02:32 localhost boa: Access from glYQlT to "GET /msadc/..%255c../..%255c../..%255c/..%c1%1c../..%c1%1c
·Feb 24 18:02:32 localhost boa: Access from bajcAM to "GET /scripts/..%255c../winnt/system32/cmd.exe?/c+dir HTTP/
·Feb 24 18:02:32 localhost boa: Access from ylTWe0 to "GET /scripts/..%255c../winnt/system32/cmd.exe?/c+dir HTTP/
·Feb 24 18:02:32 localhost boa: Access from ooZYZc to "GET /scripts/..%255c../winnt/system32/cmd.exe?/c+dir HTTP/
·Feb 24 18:02:33 localhost boa: Access from txch04 to "GET /scripts/..%252f../winnt/system32/cmd.exe?/c+dir HTTP/
·Feb 24 18:02:36 localhost boa: Access from mbUPZF to "GET /scripts/..%255c%255c../winnt/system32/cmd.exe?/c+dir%
·Feb 24 18:02:36 localhost boa: Access from W7TKes to "GET /scripts/..%255c../winnt/system32/cmd.exe?/c+dir HTTP/
·Feb 24 18:02:36 localhost boa: Access from Rv7gBw to "GET /scripts/..%c0%2f../winnt/system32/cmd.exe?/c+dir HTTP
```

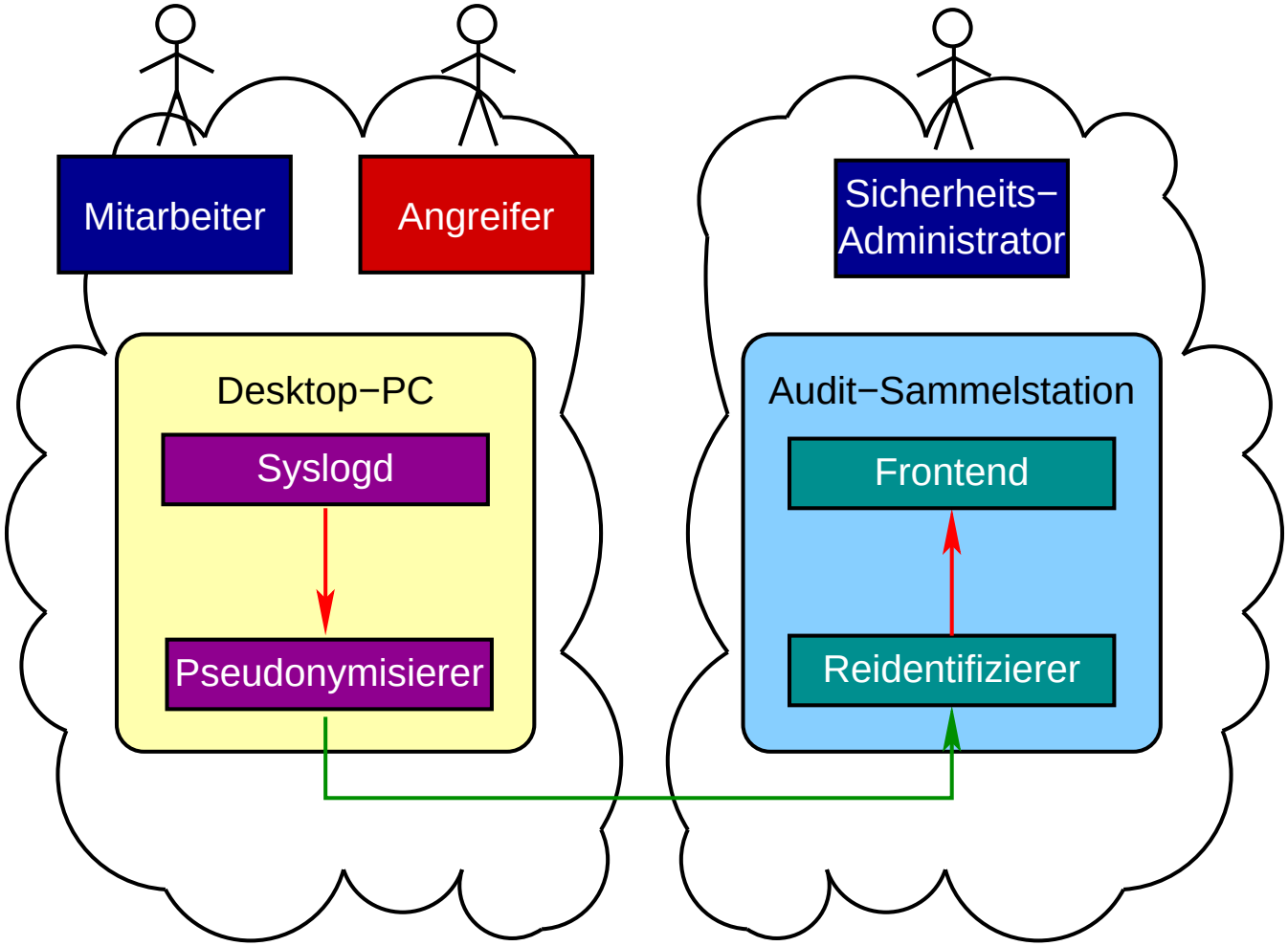
Entdeckte Angriffe

Web-Server_Angriff

```
·Feb 24 18:02:31 localhost boa: Access from 127.0.0.1 to "GET /_mem_bin/..%255c../..%255c../..%255c../winnt/syste
·Feb 24 18:02:31 localhost boa: Access from 127.0.0.1 to "GET /_vti_bin/..%255c../..%255c../..%255c../winnt/syste
·Feb 24 18:02:31 localhost boa: Access from 127.0.0.1 to "GET /c/winnt/system32/cmd.exe?/c+dir HTTP/1.0" 404 0
·Feb 24 18:02:32 localhost boa: Access from 127.0.0.1 to "GET /d/winnt/system32/cmd.exe?/c+dir HTTP/1.0" 404 0
·Feb 24 18:02:32 localhost boa: Access from 127.0.0.1 to "GET /default.ida?NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
·Feb 24 18:02:32 localhost boa: Access from 127.0.0.1 to "GET /msadc/..%255c../..%255c../..%255c/..%c1%1c../..%c1
·Feb 24 18:02:32 localhost boa: Access from 127.0.0.1 to "GET /scripts/..%255c../winnt/system32/cmd.exe?/c+dir HT
·Feb 24 18:02:32 localhost boa: Access from 127.0.0.1 to "GET /scripts/..%255c../winnt/system32/cmd.exe?/c+dir HT
·Feb 24 18:02:32 localhost boa: Access from 127.0.0.1 to "GET /scripts/..%255c../winnt/system32/cmd.exe?/c+dir HT
·Feb 24 18:02:33 localhost boa: Access from 127.0.0.1 to "GET /scripts/..%252f../winnt/system32/cmd.exe?/c+dir HT
·Feb 24 18:02:36 localhost boa: Access from 127.0.0.1 to "GET /scripts/..%255c%255c../winnt/system32/cmd.exe?/c+d
·Feb 24 18:02:36 localhost boa: Access from 127.0.0.1 to "GET /scripts/..%255c../winnt/system32/cmd.exe?/c+dir HT
·Feb 24 18:02:36 localhost boa: Access from 127.0.0.1 to "GET /scripts/..%c0%2f../winnt/system32/cmd.exe?/c+dir H
·Feb 24 18:02:36 localhost boa: Access from 127.0.0.1 to "GET /scripts/..%c0%af../winnt/system32/cmd.exe?/c+dir H
```

Informationen

Beispiel-Anwendung: Erkennung von Paßwort-Raten beim Login



Pseudonymisierte Audit Daten

```
Feb 25 10:21:15 eomer login[14569]: FAILED LOGIN (1) on `W4w4dOI9YoeIZmrB2j' FOR `lP6JTc5w', Authentication fail
Feb 25 10:21:33 eomer login[14669]: FAILED LOGIN (1) on `zK08lS2WpqntqJVeDn' FOR `YJ0jxovq', Authentication fail
Feb 25 10:21:40 eomer PAM_unix[14669]: (login) session opened for user fOrB4QA5 by (uid=0)
Feb 25 10:22:14 eomer login[14878]: FAILED LOGIN (1) on `zK08lS2WpqntqJVeDn' FOR `olpbk2hG', Authentication fail
Feb 25 10:22:21 eomer login[14878]: FAILED LOGIN (2) on `zK08lS2WpqntqJVeDn' FOR `xhQJl9a3', Authentication fail
Feb 25 10:22:52 eomer login[15057]: FAILED LOGIN (1) on `W4w4dOI9YoeIZmrB2j' FOR `JyglYQlT', Authentication fail
```

Entdeckte Angriffe

login

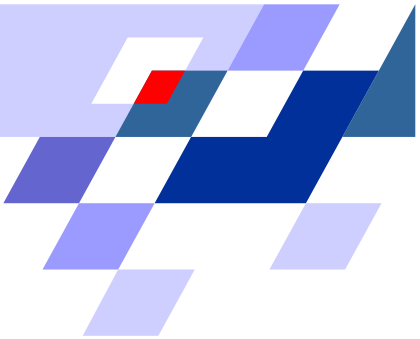
```
Feb 25 10:22:14 eomer login[14878]: FAILED LOGIN (1) on `zK08lS2WpqntqJVeDn' FOR `burech', Authentication failure
Feb 25 10:22:21 eomer login[14878]: FAILED LOGIN (2) on `zK08lS2WpqntqJVeDn' FOR `burech', Authentication failure
Feb 25 10:22:52 eomer login[15057]: FAILED LOGIN (1) on `W4w4dOI9YoeIZmrB2j' FOR `burech', Authentication failure
```

Informationen

Informationen zum Label 5x9CbVkGYw

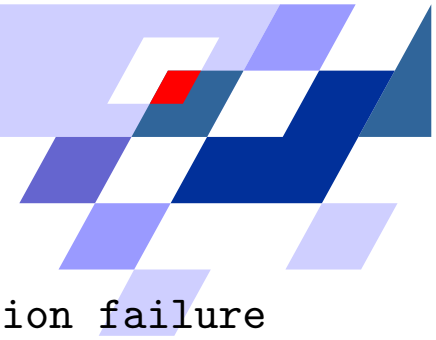
- Angriffsszenario: Loginversuche
- Schwellenwert: 3
- Anteile

1 2 3



Beispiel: Login — Original Audit-Datensätze

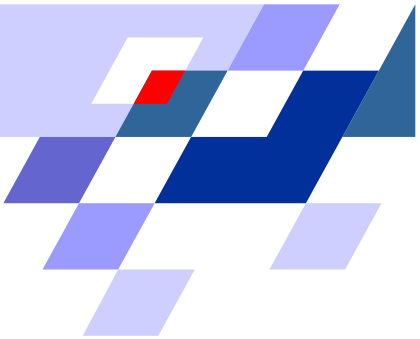
```
login[341]: FAILED LOGIN on ‘tty1’ FOR ‘sven’, Authentication failure
login[341]: FAILED LOGIN on ‘tty1’ FOR ‘sven’, Authentication failure
PAM_unix[3453]: (login) session opened for user sven by LOGIN(uid=0)
login[341]: FAILED LOGIN on ‘tty2’ FOR ‘sven’, Authentication failure
login[341]: FAILED LOGIN on ‘tty2’ FOR ‘sven’, Authentication failure
login[341]: FAILED LOGIN on ‘tty2’ FOR ‘sven’, Authentication failure
```



Beispiel: Login — Pseudonymisiert

```
login[341]: FAILED LOGIN on ‘ucSj’ FOR ‘xVXZrQPu’, Authentication failure
login[341]: FAILED LOGIN on ‘ucSj’ FOR ‘jFFPmkew’, Authentication failure
PAM_unix[3453]: (login) session opened for user QWhheUZx by LOGIN(uid=0)
login[341]: FAILED LOGIN on ‘PIHq’ FOR ‘JfhuOOEl’, Authentication failure
login[341]: FAILED LOGIN on ‘PIHq’ FOR ‘OuGwq8VI’, Authentication failure
login[341]: FAILED LOGIN on ‘PIHq’ FOR ‘5IkXjc2N’, Authentication failure
```

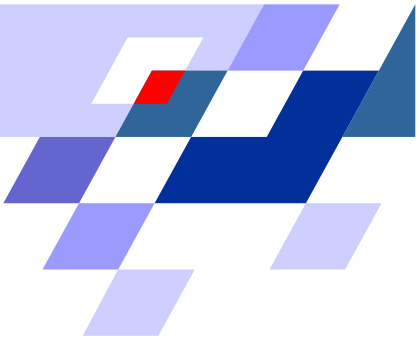
```
p: ref=...4 nym=xVXZrQPu context=pwlogin label=y6kvQDk recovery=!Hs56MmEPUDPro...
p: ref=...4 nymevent=D5gs5E84JCc8bjwfep1yLLLAqY8 origevent=VLQR1Psr7UQDbqg!tJy...
p: ref=...5 nym=jFFPmkew context=pwlogin label=y6kvQDk recovery=hrRf0rezyR1sWW...
p: ref=...5 nymevent=h0HmHpJXbRdKFi4aGkFXkncCUA8 origevent=R8YTusxf4COQTcKJvzm...
p: ref=...6 nym=QWhheUZx context=pwlogin label=520eQ8w
p: ref=...6 nymevent=bHtEPRU4py4Y48mgzbZW9jiQA6o origevent=ztglyDxKU7he8INQ24T...
p: ref=...7 nym=JfhuOOEl context=pwlogin label=hYfrwpY recovery=asiCqeHX!H5W6z...
p: ref=...7 nymevent=UUf5WRnr51fJJwaIymbVEYidchY origevent=MaWv1n21zDAZFgUjO_O...
p: ref=...8 nym=OuGwq8VI context=pwlogin label=hYfrwpY recovery=7pTaXNS!2RJCLS...
p: ref=...8 nymevent=jl1Qi4fF8yX0qDJTOv7yJz!pNTE origevent=MaWv1n21zDAZFgUjO_O...
p: ref=...9 nym=5IkXjc2N context=pwlogin label=hYfrwpY recovery=!GXVBLMKuZlGjz...
p: ref=...9 nymevent=9BOb8mtLUCL9bM1fI9SmGR1kdGg origevent=MaWv1n21zDAZFgUjO_O...
```



Beispiel: Login — Aufgedeckt

```
login[341]: FAILED LOGIN on ‘ucSj’ FOR ‘xVXZrQPu’, Authentication failure
login[341]: FAILED LOGIN on ‘ucSj’ FOR ‘jFFPmkew’, Authentication failure
PAM_unix[3453]: (login) session opened for user QWhheUZx by LOGIN(uid=0)
login[341]: FAILED LOGIN on ‘PIHq’ FOR ‘sven’, Authentication failure
login[341]: FAILED LOGIN on ‘PIHq’ FOR ‘sven’, Authentication failure
login[341]: FAILED LOGIN on ‘PIHq’ FOR ‘sven’, Authentication failure

p: ref=...4 nym=xVXZrQPu context=pwlogin label=y6kvQDk recovery=!Hs56MmEPUDPro...
p: ref=...4 nymevent=D5gs5E84JCc8bjwfep1yLLLAqY8 origevent=VLQR1Psr7UQDbqg!tJy...
p: ref=...5 nym=jFFPmkew context=pwlogin label=y6kvQDk recovery=hrRf0rezyR1sWW...
p: ref=...5 nymevent=h0HmHpJXbRdKFi4aGkFXkncCUA8 origevent=R8YTusxf4COQTcKJvzm...
p: ref=...6 nym=QWhheUZx context=pwlogin label=520eQ8w
p: ref=...6 nymevent=bHtEPRU4py4Y48mgzbZW9jiQA6o origevent=ztglyDxKU7he8INQ24T...
p: ref=...7 nymevent=UUf5WRnr51fJJwaIymbVEYidchY origevent=MaWv1n21zDAZFgUjO_O...
p: ref=...8 nymevent=jl1Qi4fF8yX0qDJTOv7yJz!pNTE origevent=MaWv1n21zDAZFgUjO_O...
p: ref=...9 nymevent=9BOb8mtLUCL9bM1fI9SmGR1kdGg origevent=MaWv1n21zDAZFgUjO_O...
```



Datenschutzfördernde Technologien I

Identitätsmanagement: P3P

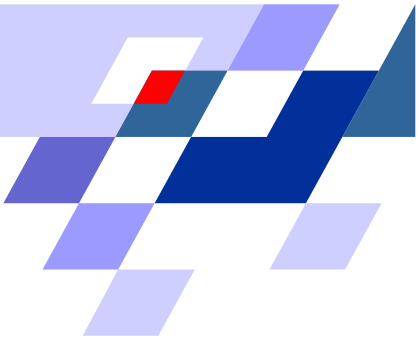
Anonymisierungsdienste: Web & Email

Anonymizer.com, Crowds, JAP, Onion Routing, Proxymate,
Cypherpunk & Mixmaster remailers, ...

\+ Kontrolle ausschließlich bei Nutzer
und ggf. Anonymisierungsdienst-Anbieter
→ Nutzer muß Endanbieter nicht vertrauen

– Nutzer muß Technologie kennen und bedienen können

– Datenschutzrechtliche Vorgaben sind zusätzlich
vom Endanbieter nutzerunabhängig umsetzen



Datenschutzfördernde Technologien II

Audit-Anonymisierer:

- moderne Betriebssysteme / Dienste erzeugen Audit-Daten
- meist Personenbezüge in betriebsnotwendigen Audit-Daten
- Lösungsansatz:
  Anonymisierung / Pseudonymisierung personenbezogener Daten

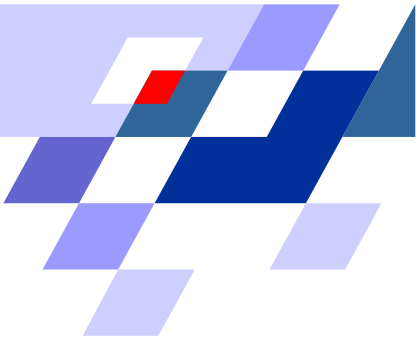
z.B. **Anonymouse log file anonymizer**:

vergröbert personenbezogene Daten bzw. ersetzt durch Standardwerte

\+ restliche Audit-Daten bleiben auswertbar

– ursprüngliche Personenbezüge nicht wiederherstellbar

– bei Mißbrauch keine Möglichkeit, die Anonymität aufzuheben



Intrusion-Detection-Systeme und Datenschutz

Ansätze: IDA, AID, ...

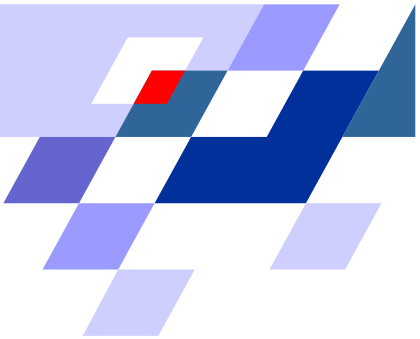
- analysieren pseudonymisierte Audit-Daten
- decken bei Bedarf Personenbezüge auf

Ziel:

Pseudonym-Aufdeckung bei Mißbrauchs-Anfangsverdacht

– Aufdeckung technisch ohne Anfangsverdacht möglich

– Dritte Partei zum Aufdecken benötigt



Verwandte Ansätze

Elektronische offline-Zahlungssysteme:

Rechtmäßiger Zweck: Betrüger haftbar machen

Annahme: Betrug entdeckt (Double-Spending)

→ Zurechenbarkeit rechtmäßig

Fairness:

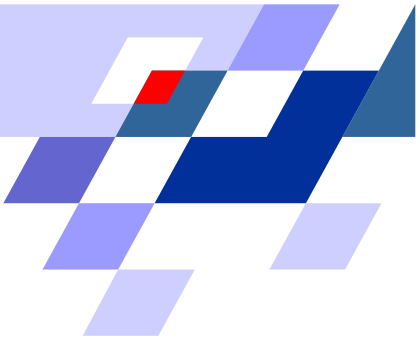
- Normalfall: Zahlungsvorgänge pseudonym
- Pseudonymaufdeckung ausschließlich bei Double-Spending

anonyme Credentials und anonyme Authentisierungsprotokolle

\+ Starkes Vertrauensmodell

– Pseudonymaufdeckung abhängig von Dritten

– aufwendige Infrastruktur notwendig, aber nicht vorhanden



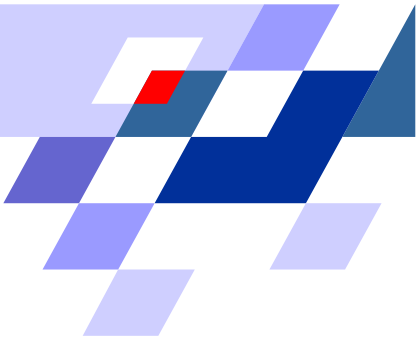
Andere Ansätze für die Audit-Daten-Anonymisierung

Ansätze:

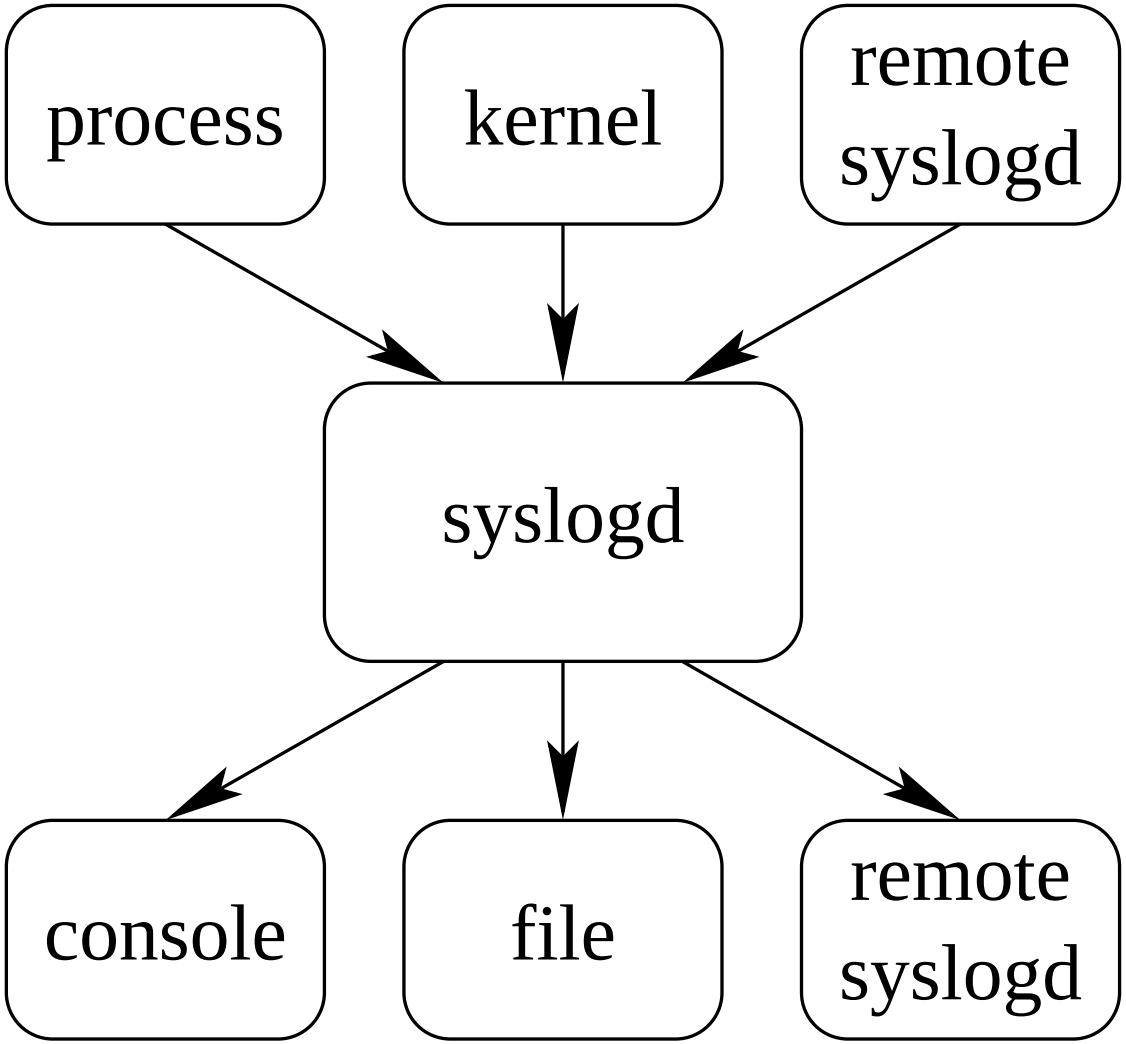
- Anonymouse Log File Anonymizer
- Ansatz von Jaeger
- Firewall-Audit-Anonymisierer von Lundin
- WebWasher
- Intrusion Detection and Avoidance (IDA)
- Adaptive Intrusion Detection (AID)

Probleme:

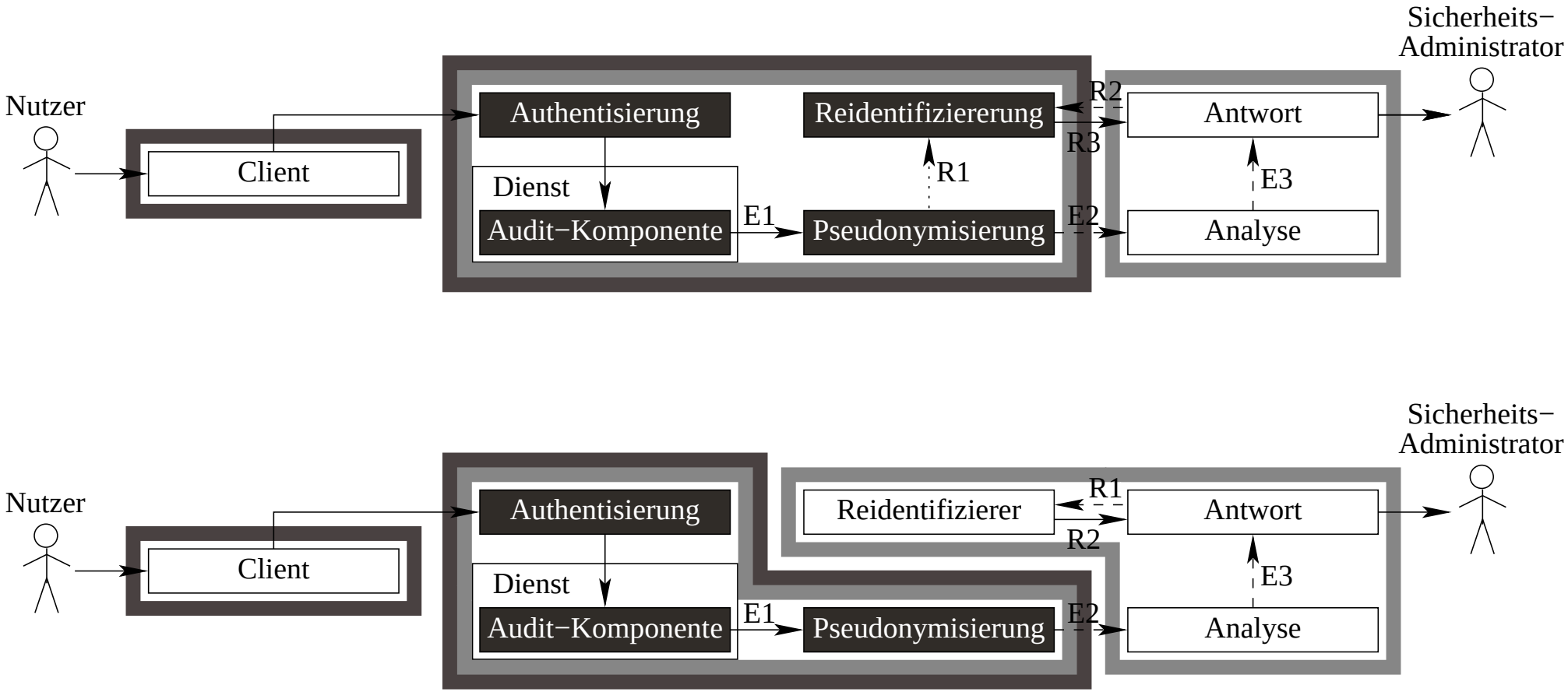
- keine Zurechenbarkeit herstellbar
- notwendige Vertrauens- / Kontroll-Beziehungen nicht berücksichtigt
  → Umgehung der Zweckbindung durch Sicherheits-Administrator
- ungeeignete Implementierung des 4-Augen-Prinzips
  →Schlüssel nach erster Aufdeckung Sicherheits-Admin bekannt



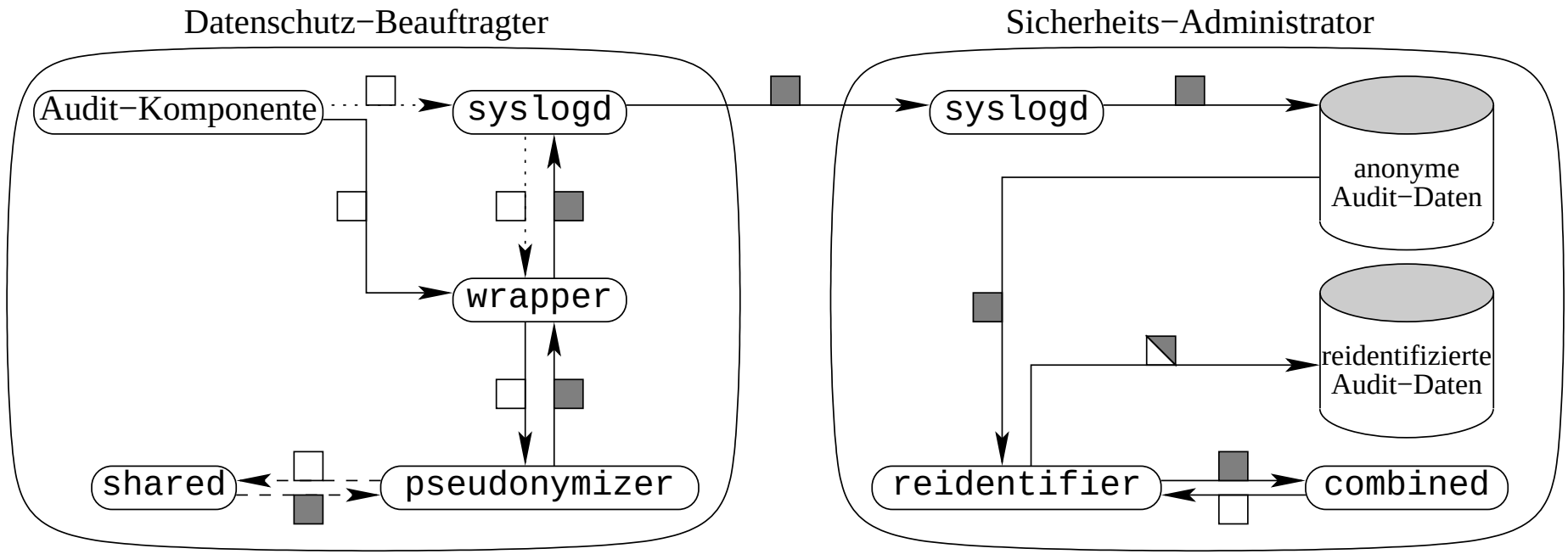
Syslog


process
kernel
remote
syslogd
syslogd
console
file
remote
syslogd

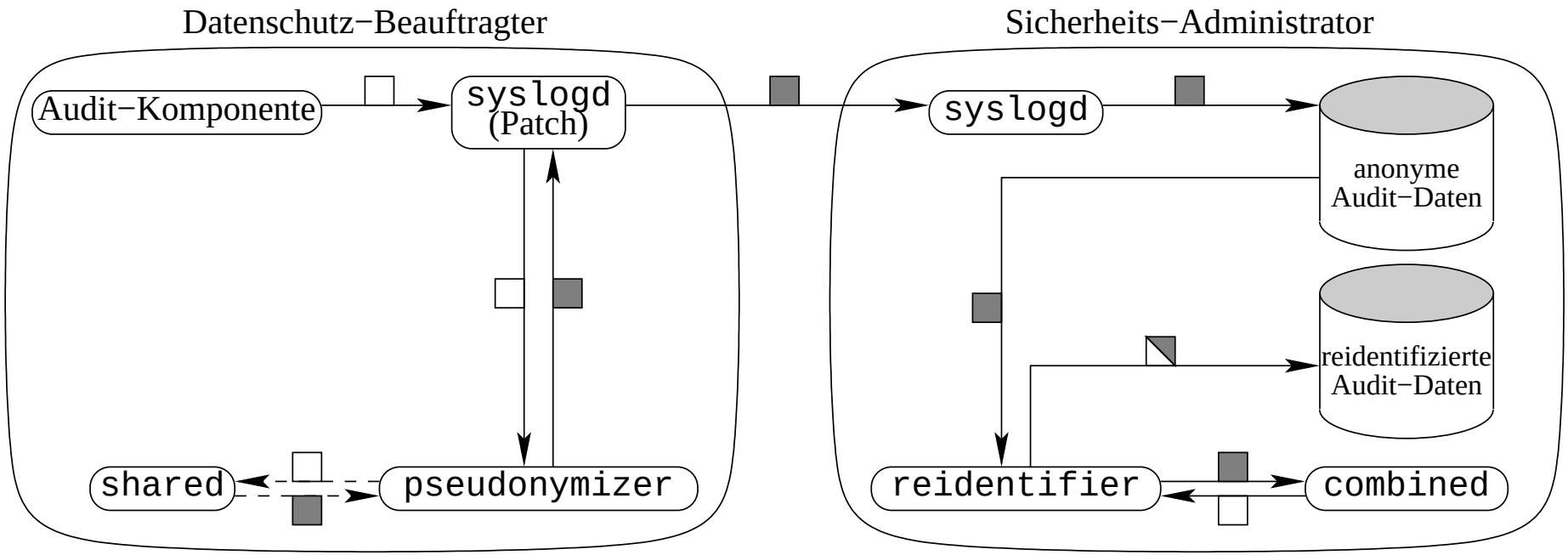
Zweckbindung: organisatorisch vs. technisch



Architektur: Wrapper



Architektur: Patch

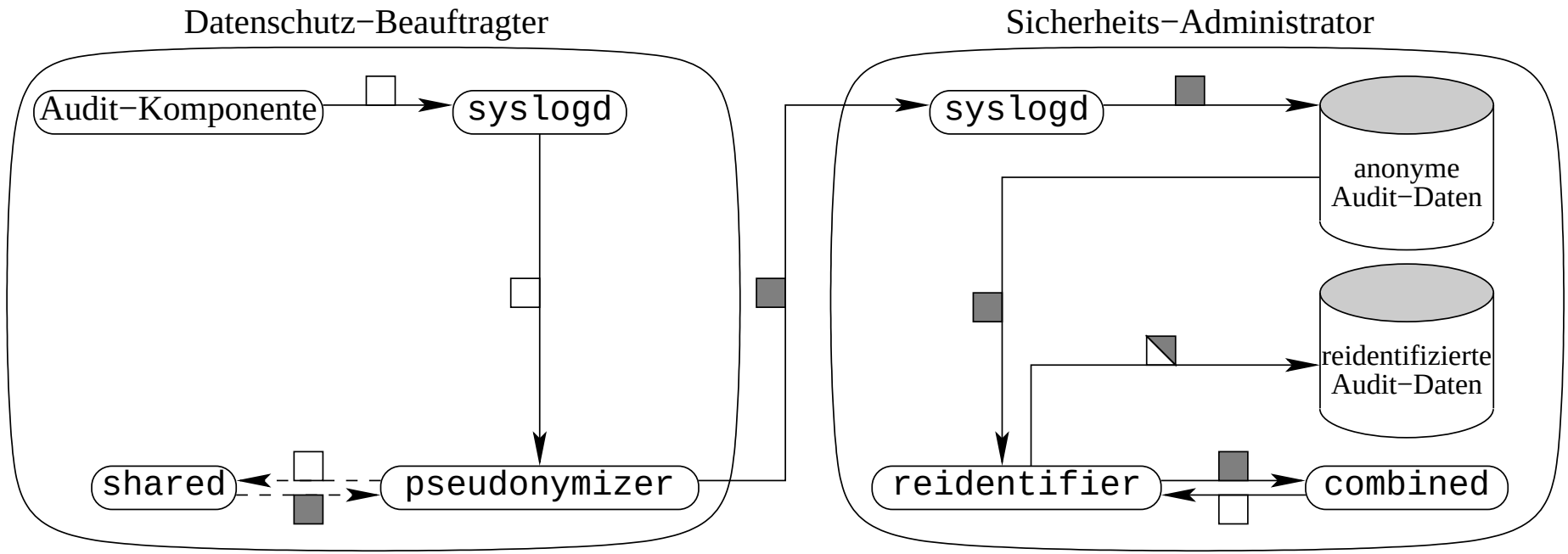


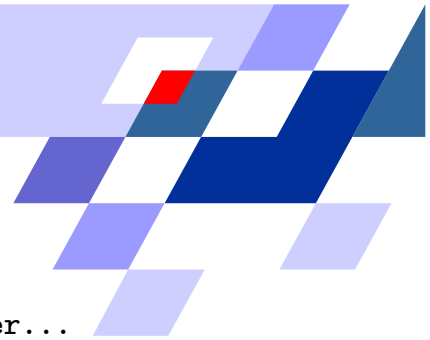
- personenbezogene Merkmale
- pseudonymisierte Merkmale
- personenbezogene und pseudonymisierte Merkmale

ungeschützter Transport

SSL−geschützter Transport

Architektur: Pipes





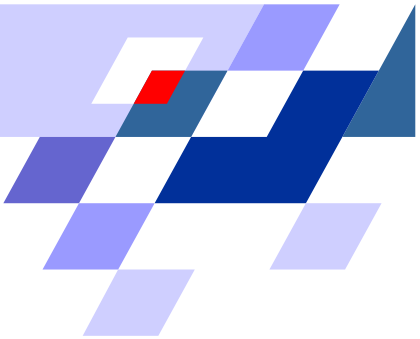
Beispiel: *Syslog*-Audit-Datensätze

```
 1 ftpd[7427]: xferlog (send): 30 pony.puf 240369 0 240369 funstuff.tar.gz b _ o g dexter...
 2 ftpd[7427]: xferlog (send): 28 pony.puf 243943 0 243943 funstuff.tar.gz b _ o g dexter ftp 0 * c

 3 pppoed[15459]: invalid state a7
 4 pppoed[15459]: Failed to discover server!

 5 Pluto[13321]: Starting Pluto (FreeS/WAN Version 1.9)
 6 Pluto[13321]:   including X.509 patch (Version 0.8.1)
 7 Pluto[13321]: X.509 certificate file '/etc/x509cert.der' not found
 8 Pluto[13321]: OpenPGP certificate file '/etc/pgpcert.pgp' not found
 9 Pluto[13321]: listening for IKE messages
10 Pluto[13321]: FATAL ERROR: bind() failed in find_raw_ifaces4(). Errno 13: Permission denied
12 sendmail[23248]: gethostbyaddr(192.168.3.1) failed: 1
13 wu.ftpd[4657]: warning: /etc/hosts.allow, line 12: can't verify hostname: gethostbyname(mport80.minerva...
14 wu.ftpd[4658]: warning: /etc/hosts.allow, line 12: can't verify hostname: gethostbyname(mport80.minerva...

15 kernel: LIDS: lidsadm (22 5 inode 28746) pid 14419 user (0/0) on ttyp181: try to open /etc/lids/lids.conf...
16 portsentry[848]: attackalert: SYN/Normal scan from host: pony.puf/192.168.1.4 to TCP port: 21
17 tcplog[1040]:     SYN                 RES2 : ftp from 192.168.1.4 port 45691
18 tcplog[1040]: FIN SYN     PSH     URG      : ftp from 192.168.1.4 port 45693
19 tcplog[1040]: FIN SYN     PSH     URG      : ftp from 192.168.1.4 port 45693
20 tcplog[1040]: FIN         PSH     URG      : port 34513 from 192.168.1.4 port 45697
21 tcplog[1040]: FIN         PSH     URG      : port 34513 from 192.168.1.4 port 45697
22 tcplog[1040]: QUESO: port 34513 from 192.168.1.4 port 45697
23 login[1510]: ROOT LOGIN on 'tty1'
24 login[12571]: ILLEGAL ROOT LOGIN on 'ttyS1'
```



Parsing von Audit-Datensätzen (*Syslog*)

Zeitstempel, **Rechner**:

`Oct 20 20:48:29` `pony` identd[22509]: token TWpldDmO2sq65FfQ82zX == uid 1000 (deedee)

Audit-Komponente, **Prozeß-ID**:

`identd`[`22509`]: token TWpldDmO2sq65FfQ82zX == uid 1000 (deedee)
`su`: BAD SU deedee to root on /dev/ttyp1

Ereignistyp-Kontext:

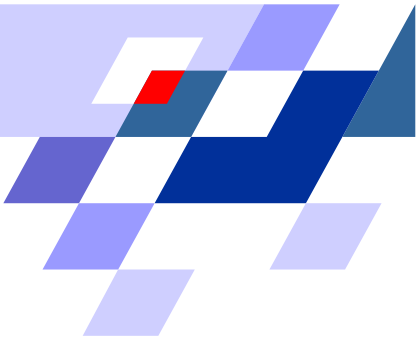
identd[22509]: `token` TWpldDmO2sq65FfQ82zX `== uid` 1000 (deedee)
su: `BAD SU` deedee to root `on` /dev/ttyp1

Personenbezüge:

identd[22509]: token TWpldDmO2sq65FfQ82zX == uid `1000` (`deedee`)
su: BAD SU `deedee` to `root` on `/dev/ttyp1`

Personenbezugstyp-Kontext des Personenbezugs 1000:

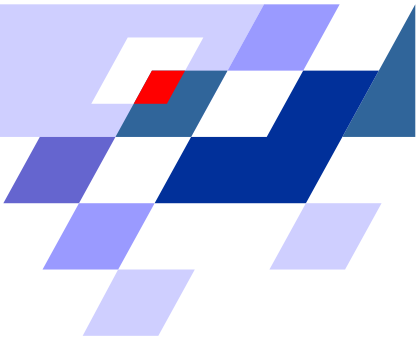
identd[22509]: token TWpldDmO2sq65FfQ82zX == `uid` 1000 `(` deedee)



Leistungsmessungs-Parameter

- Pentium III 650Mhz
- 256MB RAM
- 100Mbps Fast-Ethernet Uplink
- OpenBSD 2.7

Parameter	Standardwert	variiert mittels
Kommunikation mit `shared`	Unix-Domain-Socket	Kommandozeile
Anzahl Bits	128	Source-Code
Anzahl Audit Records	1000	Audit-Daten
Anfangsverdachts-Schwellenwert	5	Konfiguration
Anzahl Anfangsverdachte	1	Audit-Daten & Konfiguration
Anzahl Audit-Komponenten	1	Audit-Daten & Konfiguration
Anzahl Ereignis-Typen	1	Audit-Daten & Konfiguration
Anzahl Personenbezüge je Datensatz	1	Audit-Daten & Konfiguration
Pseudonym-Gewicht	1	Konfiguration



Pseudonymisierung durch Geheimnisteilung

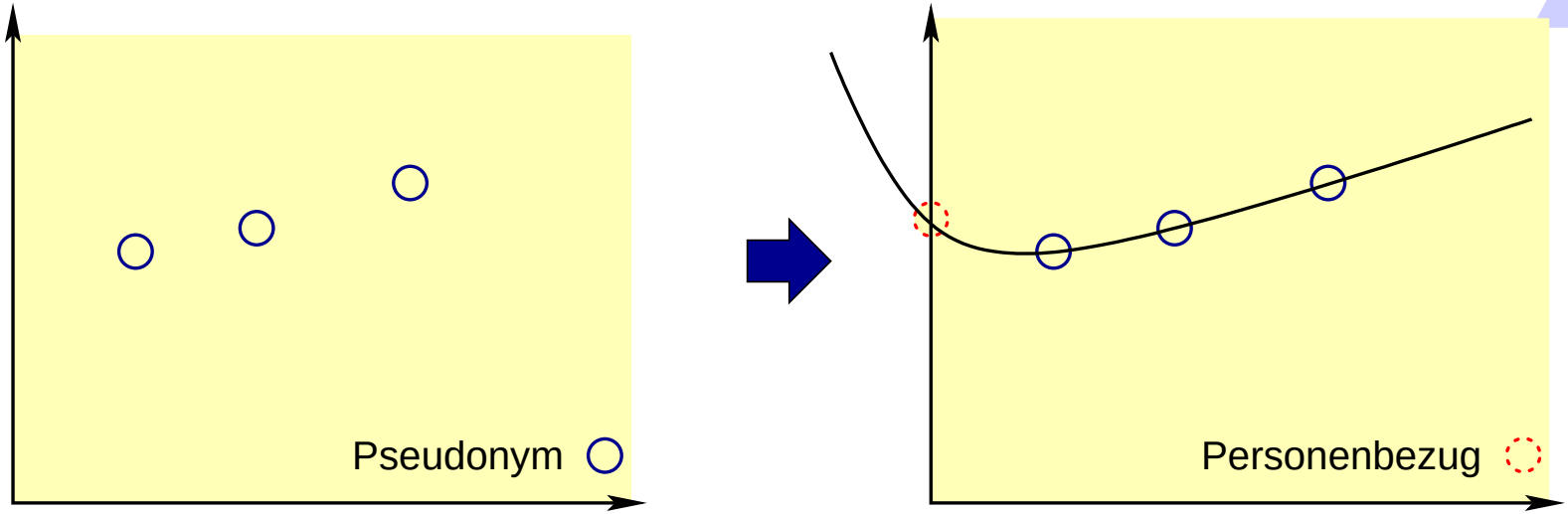
Der Pseudonymisierer agiert als
Repräsentant der Anonymitätsgruppe seiner Nutzer:

- Chiffrieren der Personenbezüge
- Aufteilen des Dechiffrier-Schlüssels mittels Geheimnisteilung
- Pseudonymeigenschaft:

  $k \geq t$ Pseudonyme eines Personenbezugs: aufdeckbar

  $k < t$ Pseudonyme eines Personenbezugs: nicht aufdeckbar
- Eigenschaften von Shamirs Schwellenwert-Schema:
  - perfekt (kein Informationsgewinn bei weniger als t Pseudonymen)
  - ideal (Informationsrate 1)
  - basiert nicht auf unbewiesenen Annahmen
  - Pseudonymgenerierung unabhängig von existierenden Pseudonymen
  - Priorisierung durch Variieren der Pseudonymanzahl

Shamirs Schwellenwert-Schema

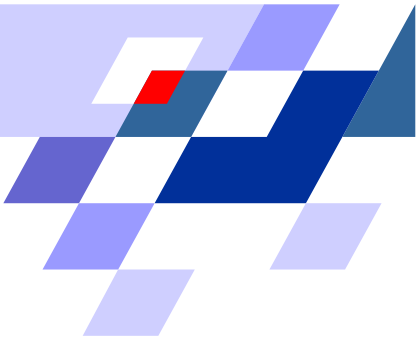


Geheimnis-Teilung:

$$p(x) = s + \sum_{j=1}^{t-1} a_j x^j mod\ P.$$

Geheimnis-Rückgewinnung (Lagrange-Interpolation):

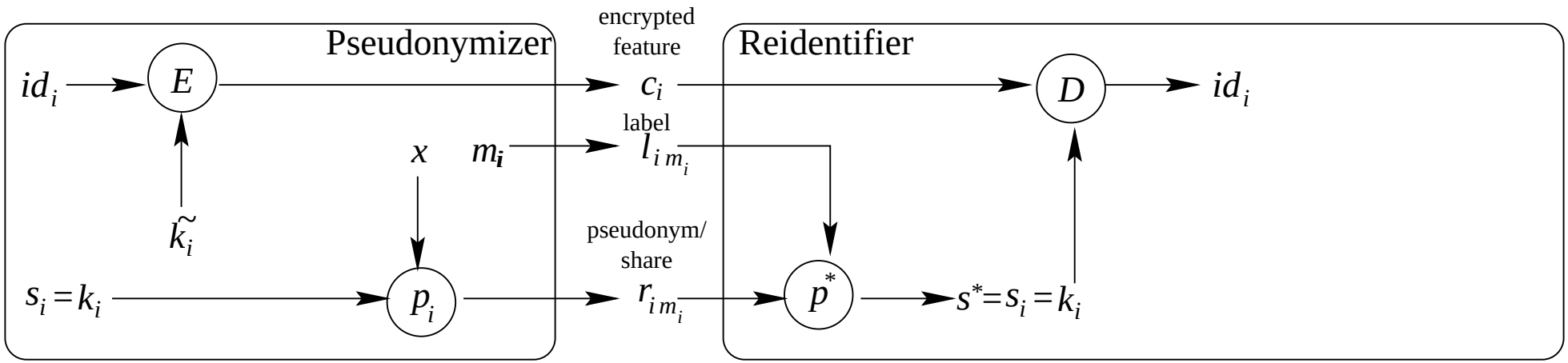
$$s^* = p^*(0) = \sum_{j=1}^{t} y_{i_j} \prod_{1 \le k \le t, k \ne j} \frac{x_{i_k}}{x_{i_j} - x_{i_k}} mod\ P.$$



Unterschiede in der Anwendung von Shamirs Schema

	übliche Anwendung	hier
Anteile je Empfänger	fix	alle
Anteile-Verteilung	Initialisierungsphase x-Koordinate (Klartext), y-Koordinate (vertraulich)	schrittweise, x- & y-Koordinate (Klartext)
x-Koordinaten	eindeutig je Polynom	eindeutig je Anfangsverdacht
"x-Produkte"	im voraus berechenbar	bei Bedarf

Datenflüsse



id_i	: Personenbezug	x	: x-Koordinate	p_i	: Polynom
s_i	: Geheimnis (siehe k_i)	m_i	: Pseudonym-Index	p^*	: interpoliertes Polynom
s^*	: Interpolations-Lösung	c_i	: Personenbezugs-Kryptogr.	E	: Chiffrierung
k_i	: Dechiffrier-Schlüssel	l_{im_i}	: Marke	D	: Dechiffrierung
$\tilde{k}_i$	: Chiffrier-Schlüssel	r_{im_i}	: Pseudonym (Anteil)		