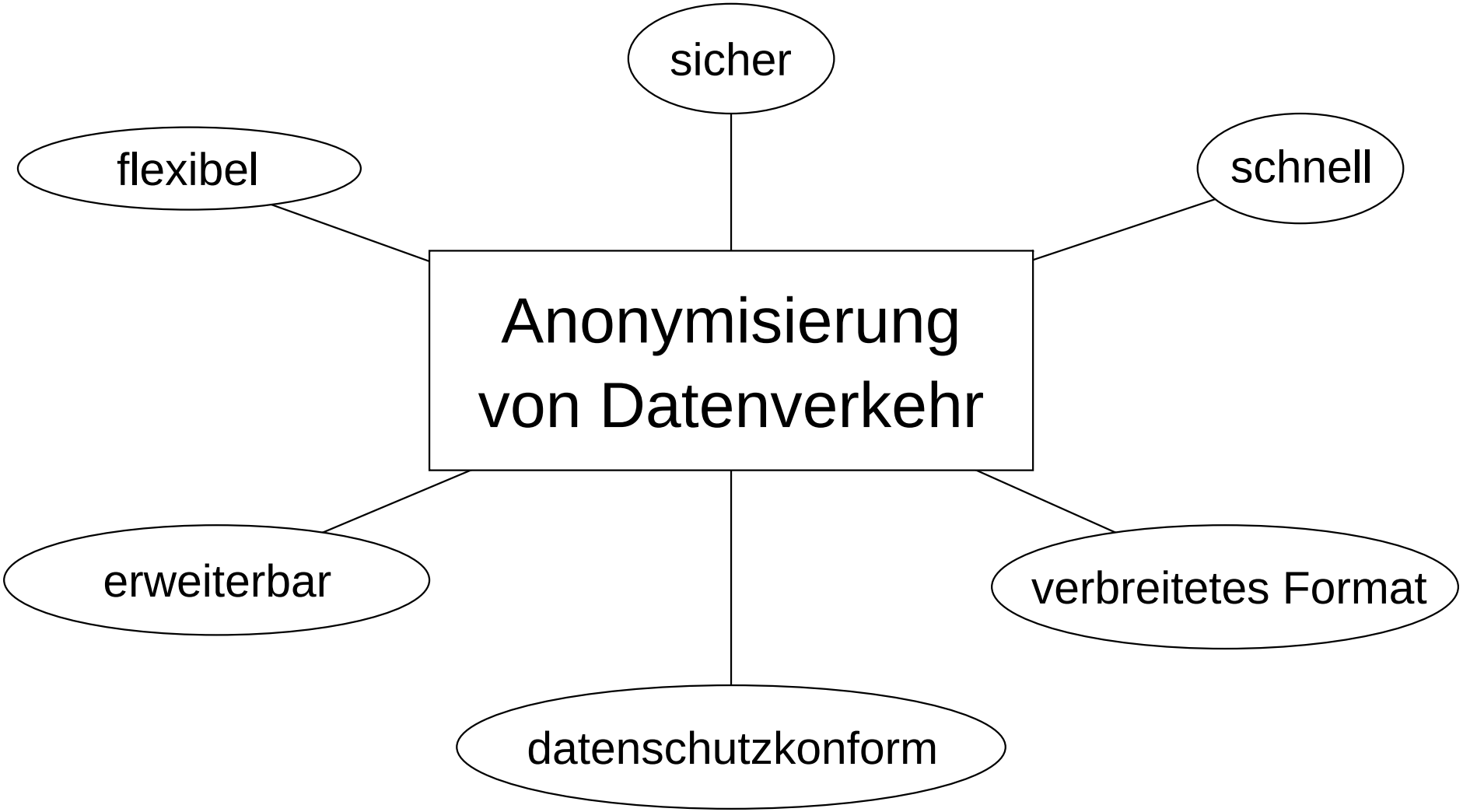
Datenschutzkonforme Anonymisierung von Datenverkehr auf einem Vermittlungssystem

Christoph Mayer

- Studienarbeit -

Institut für Telematik

Ziel der Arbeit



Agenda

1. Motivation
2. Datenschutz
3. Pktanon - Anonymisierungsframework
4. Fazit und Ausblick

Motivation

- Netzdaten für Forschung und Entwicklung
  - Angriffserkennungssysteme
  - Verhalten von Netzen, Protokollen und Anwendungen
- Generierte Daten sind keine realen Daten
  - Benutzerverhalten
  - Softwareverhalten entgegen Spezifikation
  - Unbekannte Angriffe
  - Statistische Daten
  - Subnetzverteilung
  - Datenvolumen
- Kaum frei zugängliche Netzdaten
  - Müssen konform zum Datenschutz sein
  - Unternehmerische Bedenken
  - Mangelhafte Softwareunterstützung / Automatisierung

Datenschutz

- Aspekte des Datenschutz
  - Erlaubnis zur Erhebung von Daten
    - Beseitigung von Störungen oder Forschungszwecke
  - Veröffentlichung von Daten
    - Anonymisierung von Angaben über persönliche oder sachliche Verhältnisse
- Was muss geschützt werden
  - Schutz des Benutzers
  - Schutz des Netzwerks
- Sicherheit ⟷ Nutzen
  - Kompromiss zwischen Sicherheit und Nutzen Einhaltung des Datenschutz zwingend

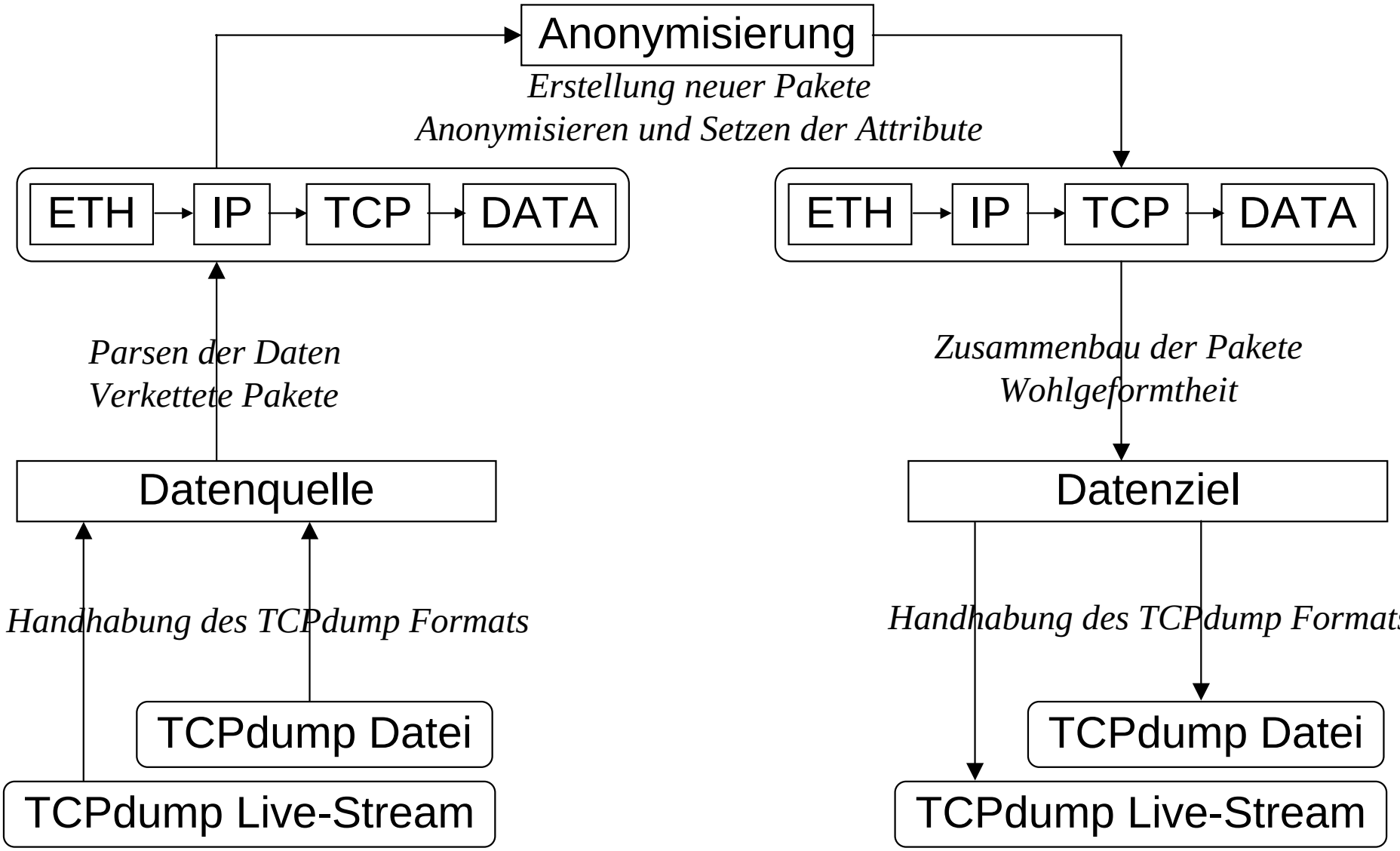
pktanon - Anonymisierungsframework

- Definitionen
  - Paket
    - Header und Optionen einer Protokollschicht
    - Vorlage für ein bestimmtes Protokoll
  - Paket-Objekt
    - Konkretes Paket
  - Paket-Interface
    - Schnittstellen zur Manipulation eines Paket-Objekts
  - Verkettete Pakete / Paketkette
    - Linear verkettete Paket-Objekte
    - Entsprechend ihrer Einkapselung verkettet

pktanon - Anonymisierungsframework

- Komplettes Parsen der Pakete
  - Aufbau von verketteten Paket-Objekten
  - Manipulation *nur* über Paket-Interfaces
  - Es wird nicht im Speicherblock „rumgestochert“
- Pakettransformation
  - Informationen löschen ist gefährlich
  - Erstelle neue Paket-Objekte anhand der Originalen
    - Anonymisiere und setze die Attribute
    - *Defensive Transformation*
  - Pool von Anonymisierungsprimitiven
- Verwaltung der Traces/Pakete vom Framework
  - Leicht auf neue Protokolle erweiterbar
- Live- oder Offline-Anonymisierung

Ablauf der Anonymisierung

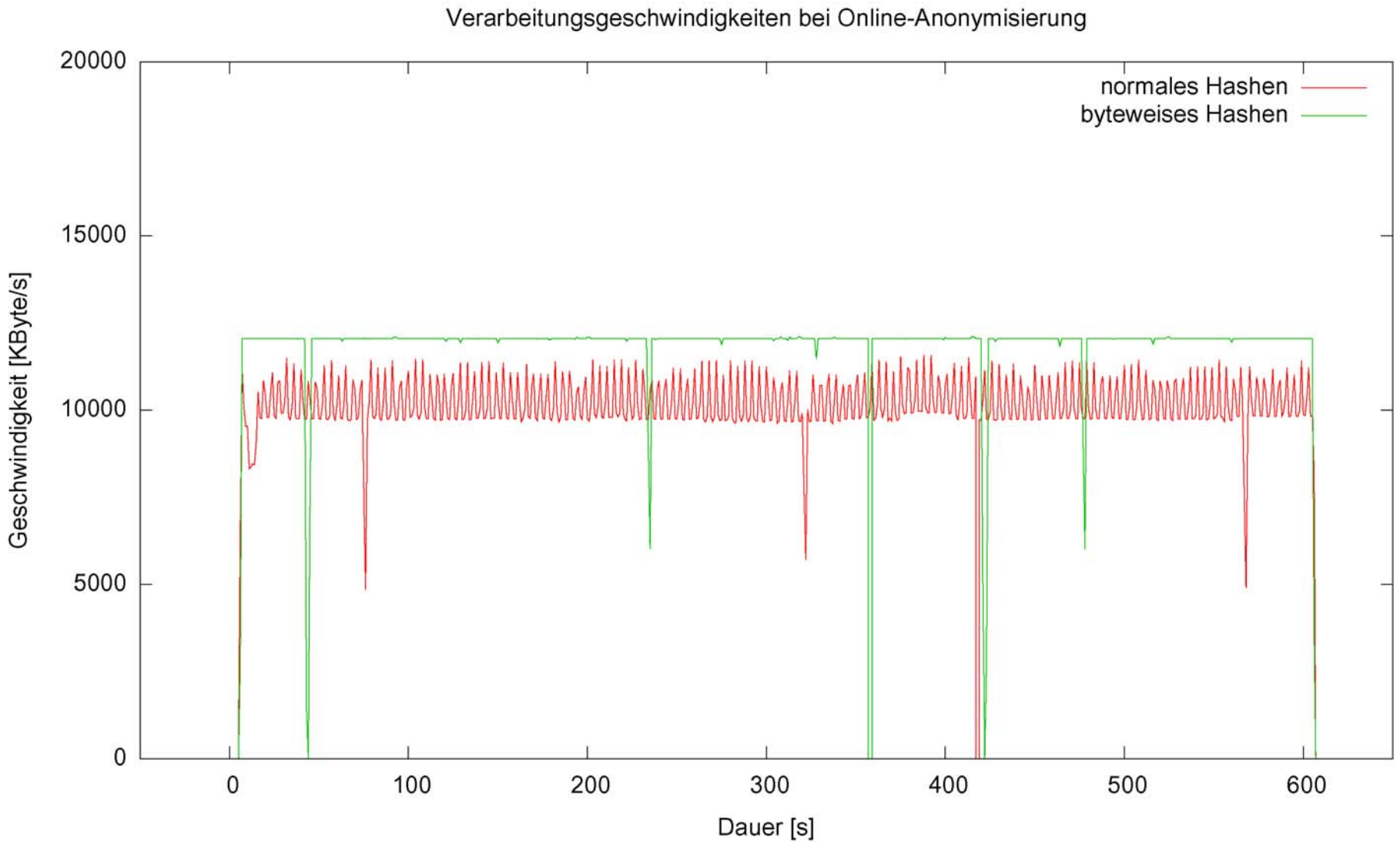


Messungen

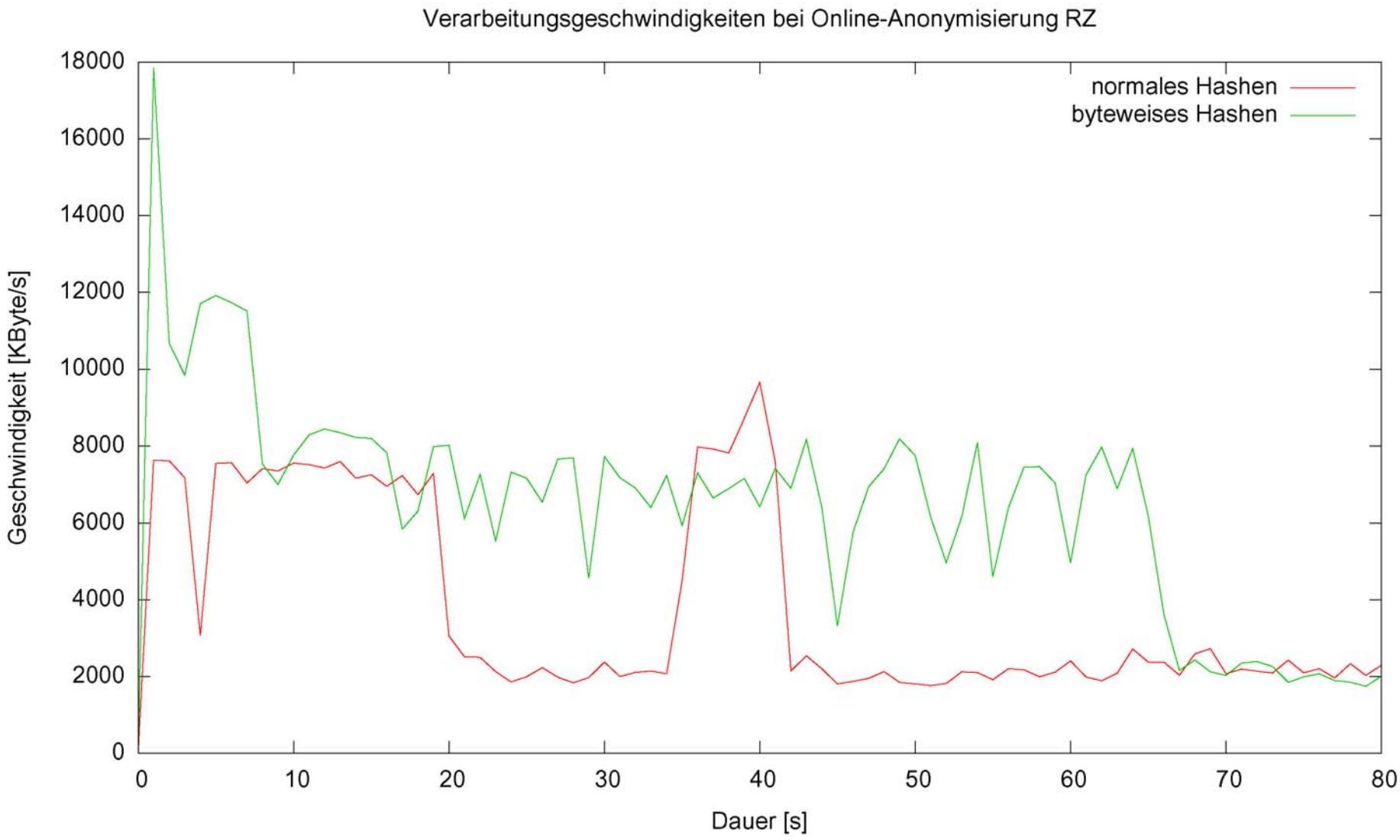
- System
  - Linux 2.6.13, Pentium 4, 2.8 GHz, 2 GB RAM
- Dateien für Offline-Anonymisierung
  - Datei A
    - 140 MB, 2 095 989 Pakete, 52 Byte ⦸ Paketgröße
  - Datei B
    - 1 GB, 1 000 000 Pakete, 1 060 Byte ⦸ Paketgröße
- Anonymisierungen
  - HMAC-SHA1 über IP- und MAC-Adressen
    - Komplettes Hashen
    - Byteweises Hashen mit Zwischenspeichern
  - Löschen von IP-Optionen
  - Ausnullen von Schicht 5 Inhalten und unbekannten Protokollen
  - Löschen von korrupten Paketen

  → Datenschutzkonforme Anonymisierung

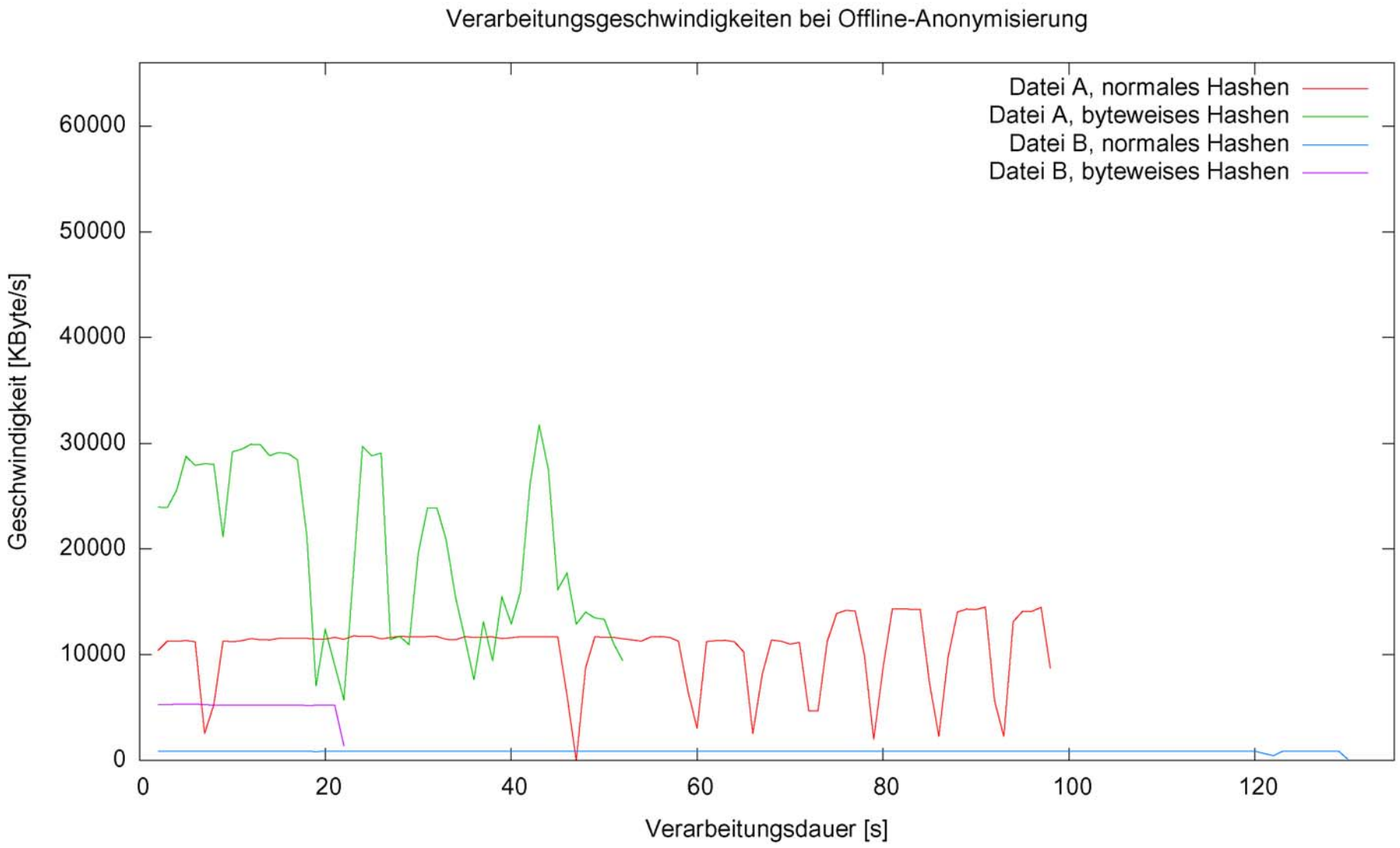
Messungen - Live-Messung - iperf



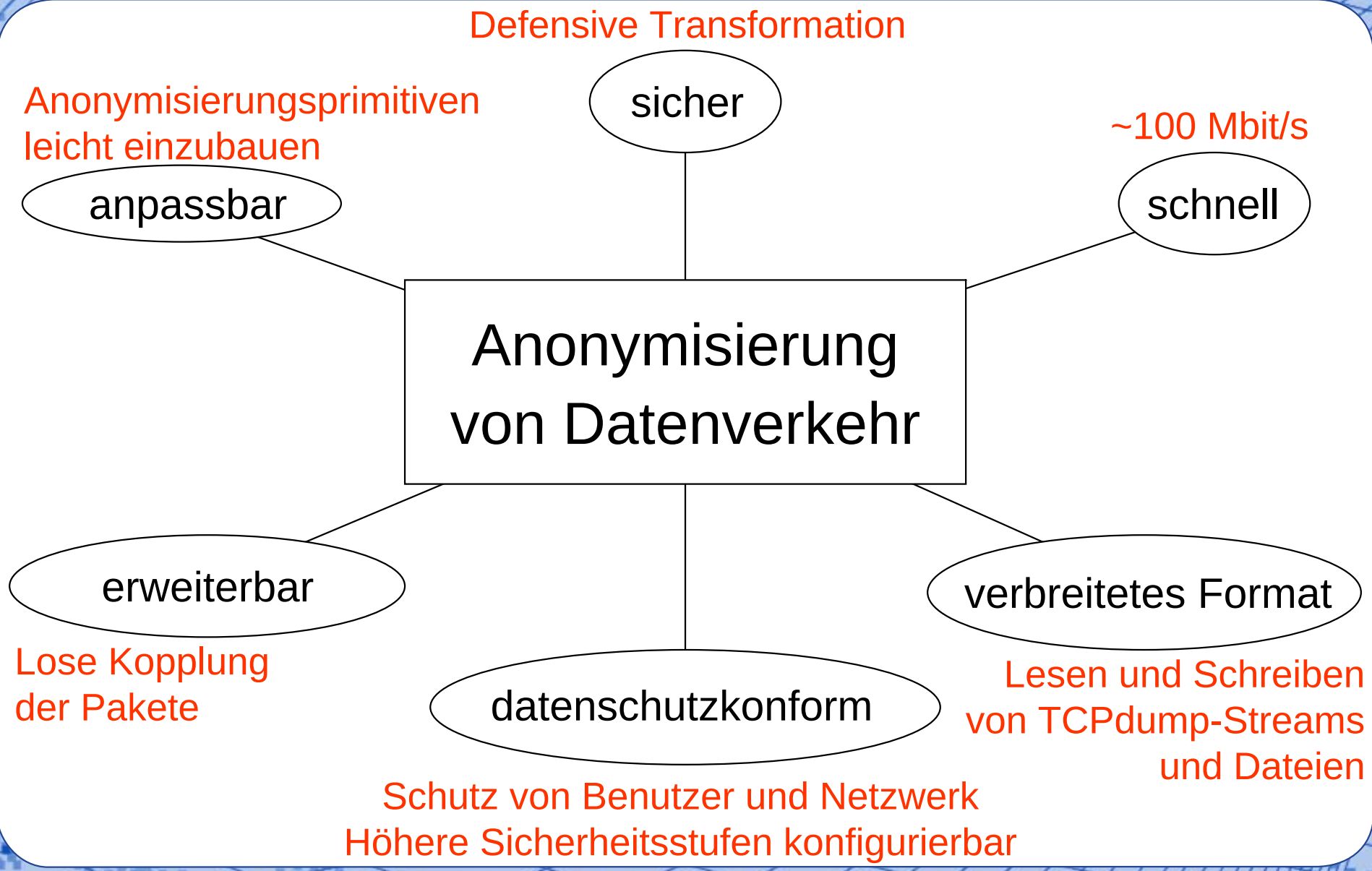
Messungen - Live-Messung - RZ



Messungen – Offline-Messung



Fazit


Defensive Transformation
sicher
Anonymisierungsprimitiven
leicht einzubauen
anpassbar
~100 Mbit/s
schnell
Anonymisierung
von Datenverkehr
erweiterbar
Lose Kopplung
der Pakete
verbreitetes Format
Lesen und Schreiben
von TCPdump-Streams
und Dateien
datenschutzkonform
Schutz von Benutzer und Netzwerk
Höhere Sicherheitsstufen konfigurierbar

Ausblick

- Präfixerhaltende Anonymisierung von IP-Adressen
  - Verteilung auf Subnetze bleibt erhalten
    - Auch bei CIDR-Adressierung
- Verwendung von Konfigurations-Profilen
  - Definieren welche Attribute welcher Protokolle mit welcher Anonymisierungsprimitive anonymisiert werden
  - Modellierung der Vertrauensverhältnisse zwischen Datenaufnahme und Datenverwendung
- Verarbeitungsgeschwindigkeit
  - Gigabit-Ethernet
  - Parallele Verarbeitung
    - Multi-Core-Prozessoren / Mehrprozessorsysteme
- Unterstützung weiterer Protokolle
  - Auch Schicht 5 Protokolle

Ende

Vielen Dank für Ihre Aufmerksamkeit!

Fragen?