Untersuchungen zur Genauigkeit von Flow-Erfassungsmechanismen

ITG FG 5.2.3

Jochen Kögel
jochen.koegel@ikr.uni-stuttgart.de

6. Oktober 2010

Universität Stuttgart
Institut für Kommunikationsnetze
und Rechnersysteme (IKR)
Prof. Dr.-Ing. Andreas Kirstädter

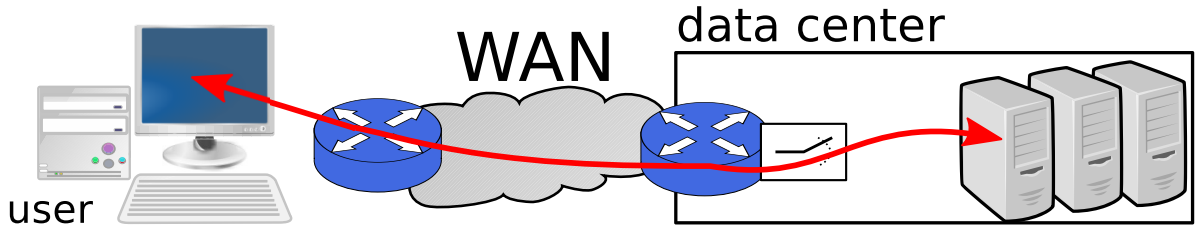
Agenda

- Motivation
- Flow-Erfassung
- Genauigkeit von Flowdaten
- Zusammenfassung und Ausblick

Motivation

Problemszenario

Szenario: interaktive Unternehmensanwendung (z.B. Web-Kollaborationstool)



- Benutzbarkeit und schnelle Reaktionszeit für Produktivität entscheidend
- In zentralem Rechenzentren realisiert
- Weltweiter Zugriff über Firmen-WAN
- Monitoringdaten prinzipiell fast überall von Netz- und Servern erfassbar

Performance-Problem: "Anwendungsreaktion ist gerade schlecht"

- Problem im Netz oder Problem der Server?
- Nur sporadisch? Wieso bei genau dieser **Transaktion (z.B. Webseitenanfrage)**?

Analyse in der Regel aufwändig, Probleme schlecht reproduzierbar

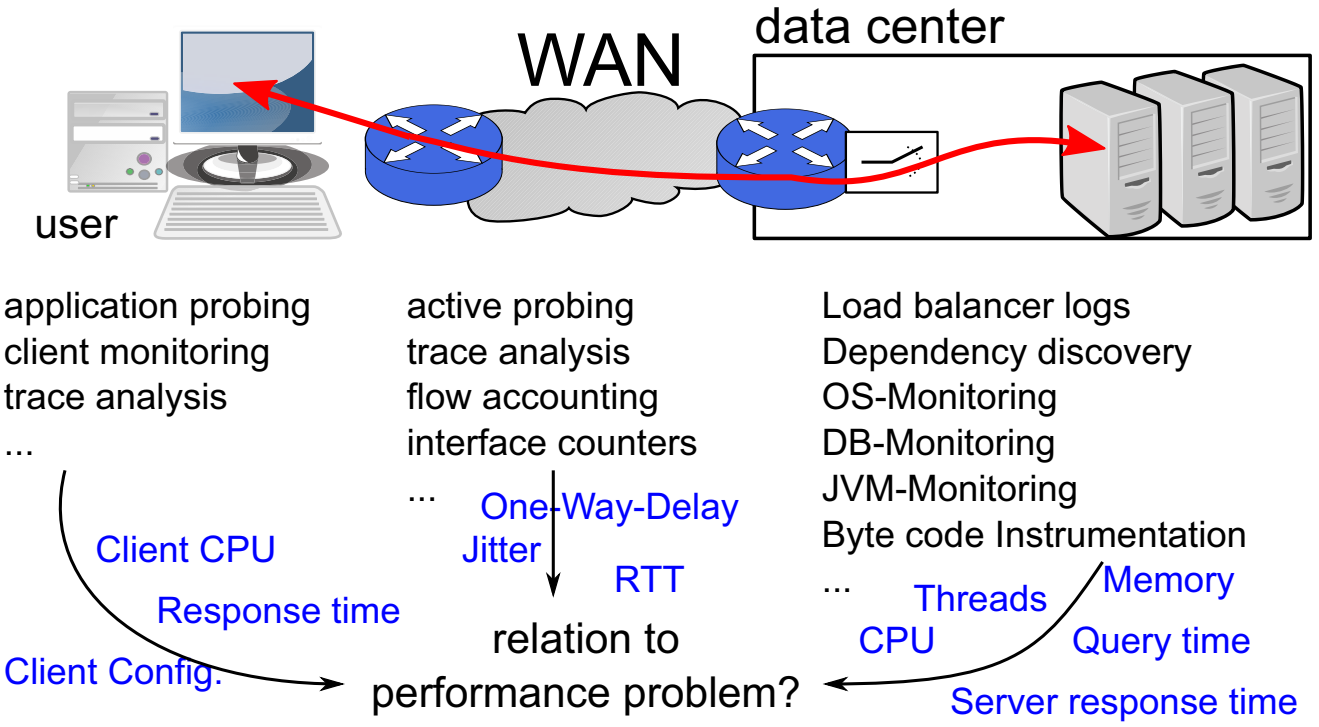
→ zur Transaktion gehörende Monitoringdaten direkt verwenden

→ am Besten liegt Auswertung bei Ankunft der Problemmeldung schon vor

Motivation

Ansätze zur Analyse von Performance-Problemen

Problem: Auswahl geeigneter Messpunkte und -verfahren



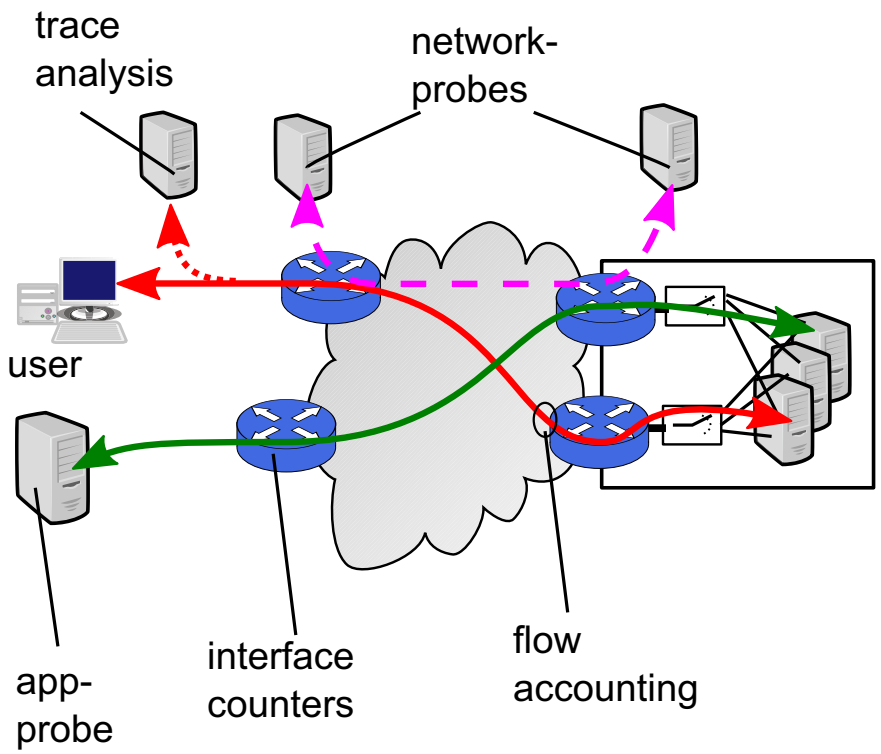
Randbedingungen für Auswahl

- Aufwand für Monitoringverfahren
- Sind Monitoringdaten geeignet, um **Auswirkungen auf die Transaktion** zu erkennen?

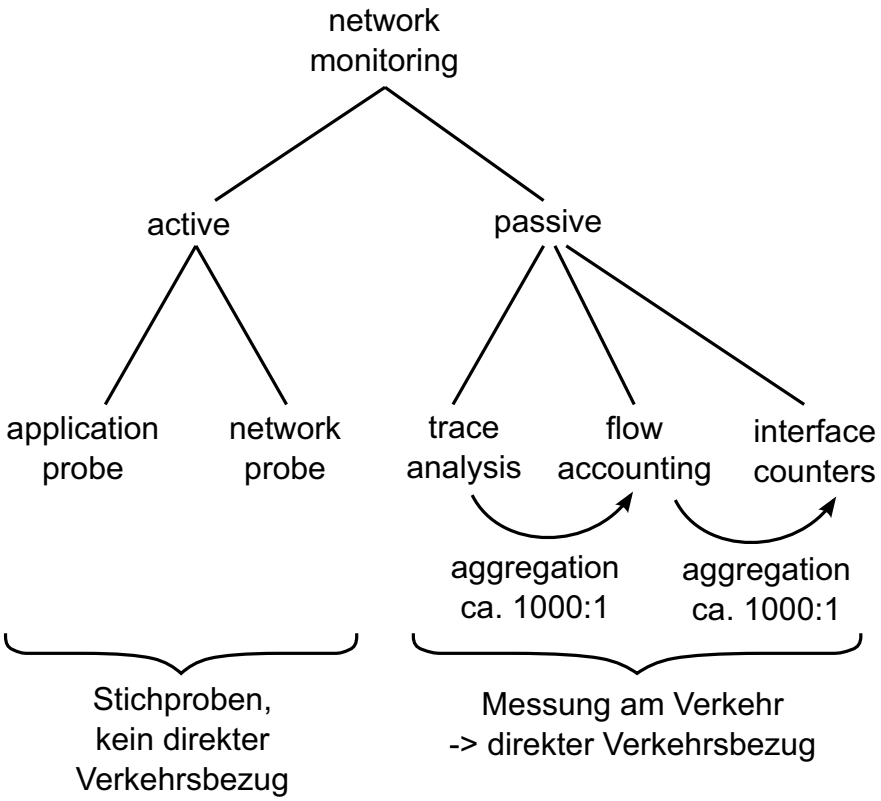
Motivation

Einordnung Netzmonitoringansätze

Illustration of approaches



Classification of approaches



- Aktive Messung

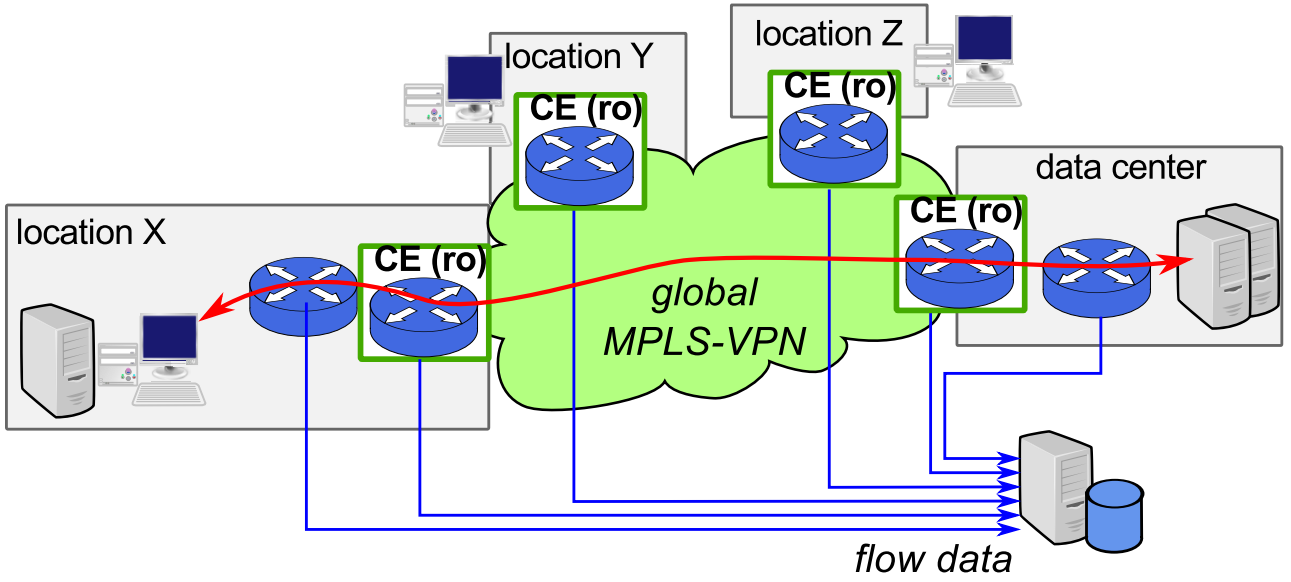
  Kein direkter Bezug zur problematischen Transaktion

- Passive Messung

  Bei zu starker Aggregation (zeitlich und/oder Endpunkte): Verkehrsbezug eingeschränkt

Motivation

Szenario



Flow-Erfassung vs. aktive Messungen und passive paketbasierte Messungen

- keine zusätzliche Geräte: Erfassung direkt im Router (ggf. einfach aktivierbar)
- Flow-Daten meist schon vorhanden (reporting, anomaly detection, ...)

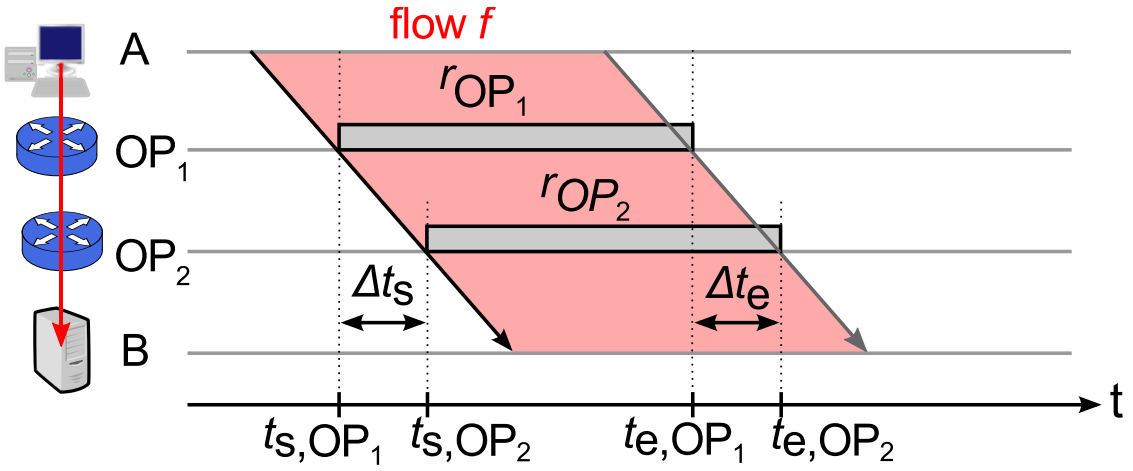
Beispiel: Global Enterprise Network, MPLS-VPN

- Flow-daten von mehreren (1..5) Routern auf einem Pfad
- Eigene Router und CE-Router des Carriers ("read only")

→ Flow-Sicht auf Verkehr an mehreren Punkten im Netz

Flow-Erfassung

Terminologie



Flow

gerichtete, unidirektionale Datenübertragung von A nach B.
Startzeitpunkt: t_s* Endzeitpunkt: t_e*, aufgrund von Latenzen gültig an einem **Beobachtungspunkt (Observation point OP).**

IP-Transport-Layer-Flow f

unicast-Flow zwischen IP-Transportschicht-Endpunkten (durch Fünftupel definiert)

Flow-Record r

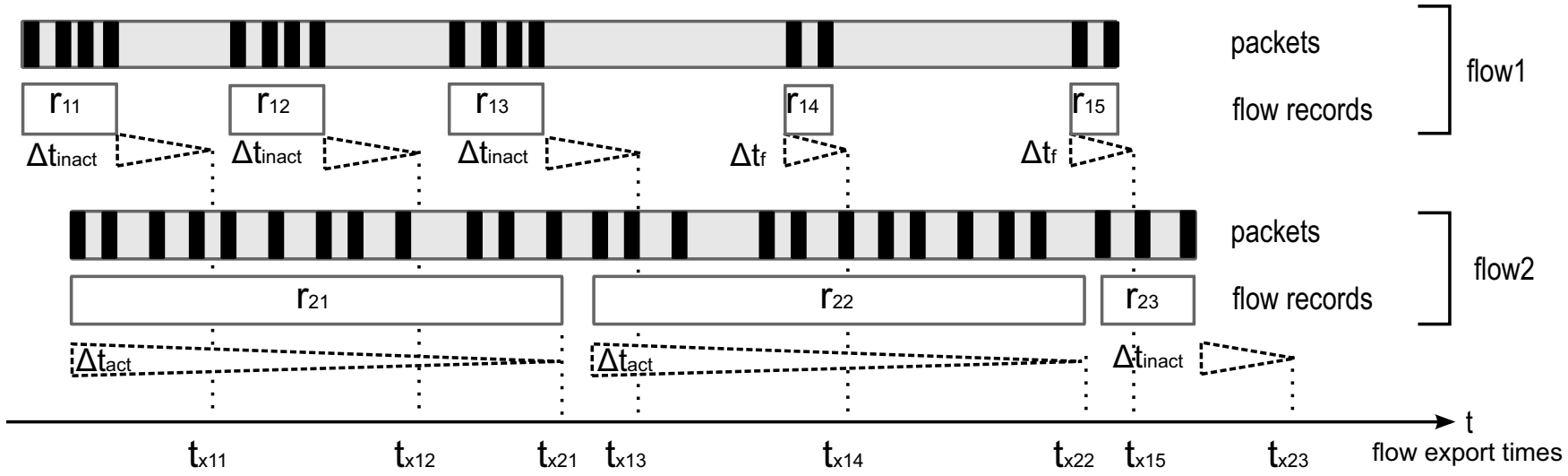
von einem **Exporter** an einem OP erzeugte Teilsicht auf einen Flow f von t_s bis t_e

Flow-Erfassung

Export Timer

Flow records werden timer-basiert exportiert

Drei verschiedene Timer (+weitere Algorithmen)



Folgerungen

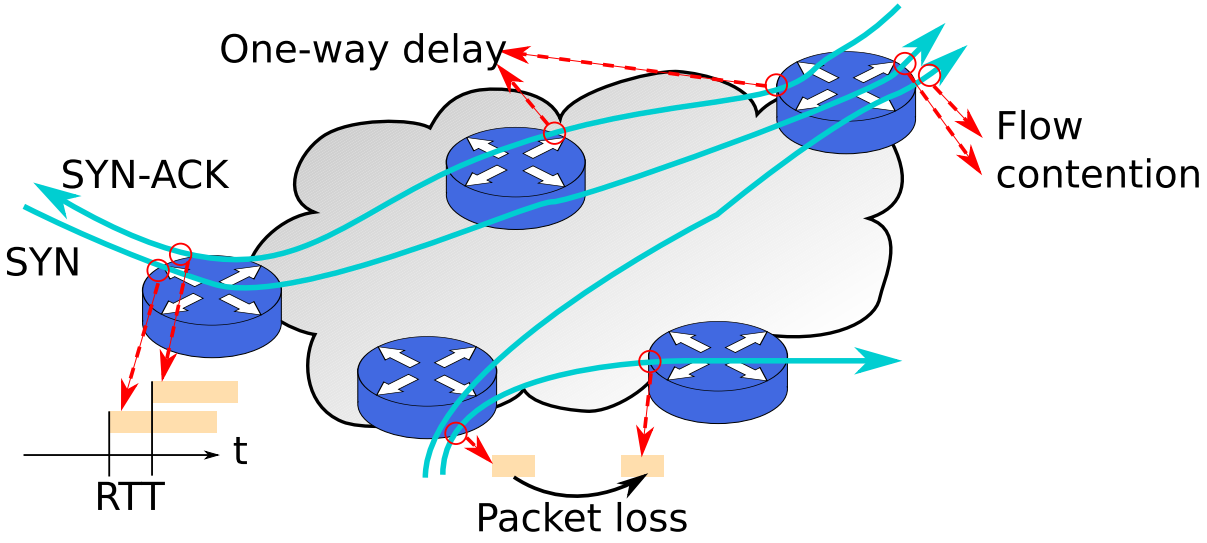
- Daten stehen zeitverzögert zur Verfügung
- Verzögerung abhängig von Flowverhalten

Flow-Erfassung

Messbare Netzcharakteristika

Mit Flow-Accounting Daten messbare Charakteristika

- Round-Trip-Time (RTT)
- One way delay (OWD)
- Packet loss
- Flow contention



Genauigkeit von Flowdaten

Erzielbare **Messqualität**?

1. **Genaue** Messung. Kenntnis von Messabweichungen (Unsicherheit, Konfidenzintervalle)

   Kenntnis der Exporter-Eigenschaften bzgl. Abweichungen

2. **Häufige** Messung: viele Messergebnisse pro Zeitintervall

   Vorfilterung notwendig, Verwendung möglichst vieler Records

3. **Zeitnahe** Messung: schnelles Berechnen von Messergebnissen

   Online-Verarbeitung: Wie schnell werden records exportiert? Wann wurden genügend records eines Zeitraums exportiert?

Im Folgenden: Betrachtung von 1.)

Genauigkeit von Flowdaten

Gefundene Genauigkeitsprobleme

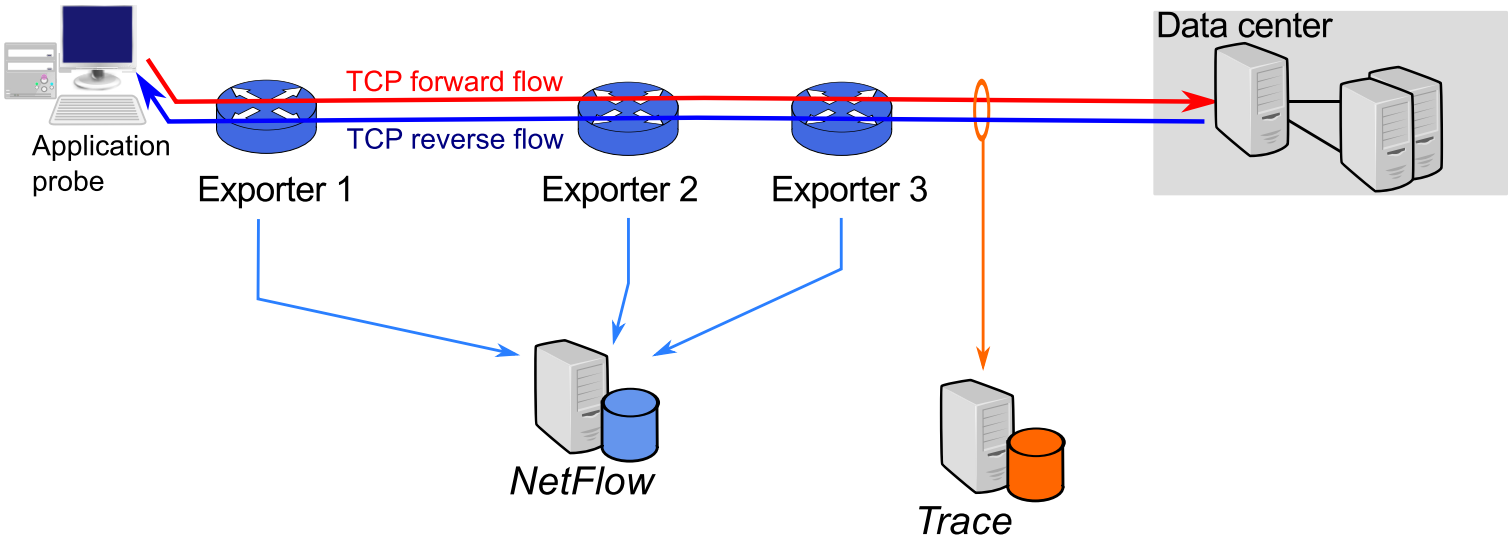
- Record loss (inhärent, da UDP)
- Duplicates
- Packet counters
- Byte counters
- Timestamp/Clock accuracy
  - Granularity
  - "Noise"
  - Jumps
  - Clock offset, clock skew

Gründe

- Implementierung: Ungenauigkeiten in Exportern selbst
- Konfiguration: z.B. Zeitsynchronisation
- Netzeigenschaften: Middle boxes

Genauigkeit von Flowdaten

Vergleich von Paket-Trace und Netflow: Szenario



- Pfad zwischen zwei europäischen Städten
- Paket-Trace von 5 Tagen, gefiltert auf zwei Adressen: Application Probe und Server
- Flowdaten von drei Exportern (zwei davon CE)

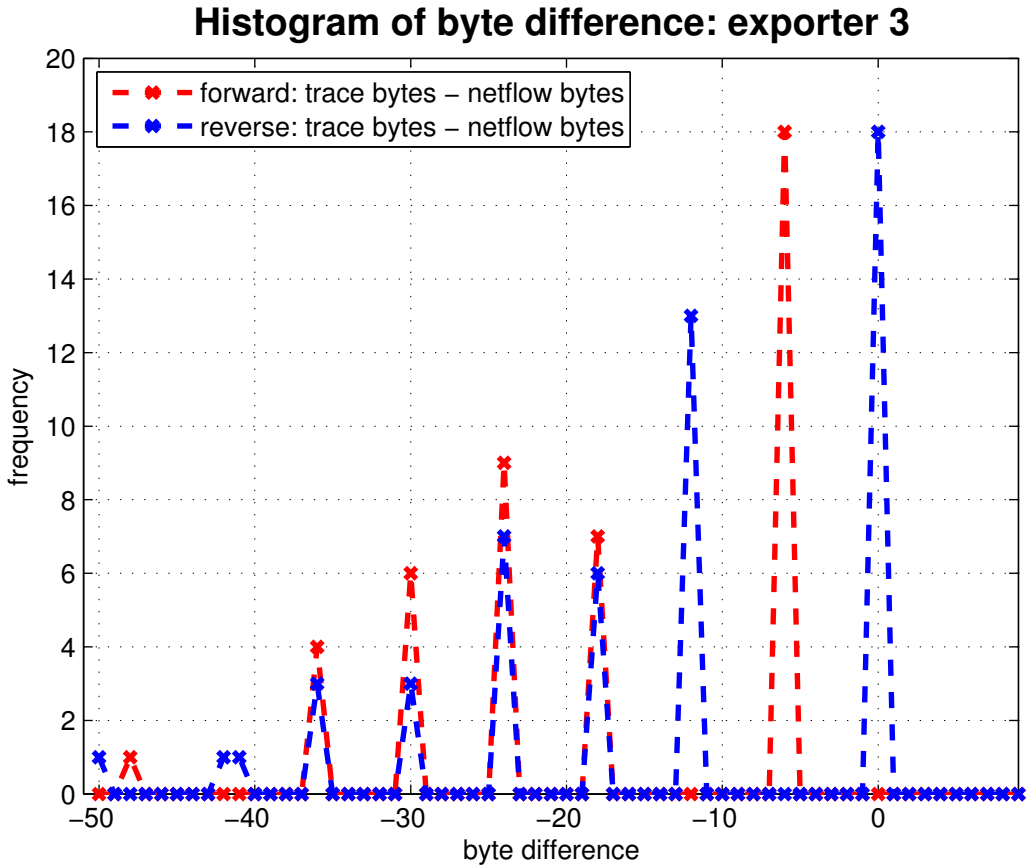
- Unterschiede bei Byte- und Paketzählern?
- Genauigkeit von Start- und Endzeitstempeln ("time difference")

Genauigkeit von Flowdaten

Vergleich von Paket-Trace und Netflow: Bytezähler

Bytezähler-Problem

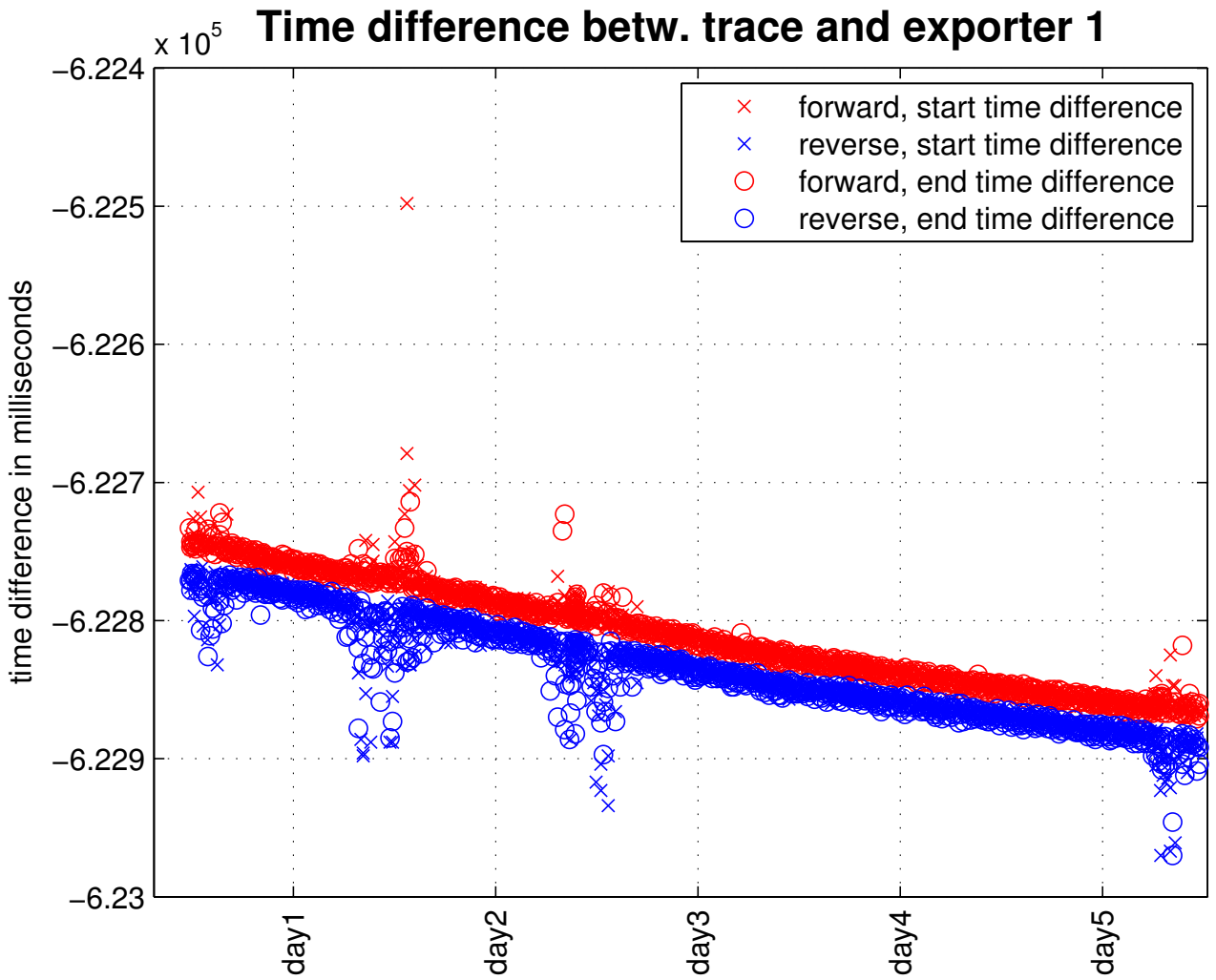
- Exporter 3: Unterschiede im Bytezähler, Paketzähler konsistent
- Beobachtung: einige Exporter runden Bytezähler auf 46 Byte auf (Ethernet payload)



Genauigkeit von Flowdaten

Vergleich von Paket-Trace und Netflow: Uhren

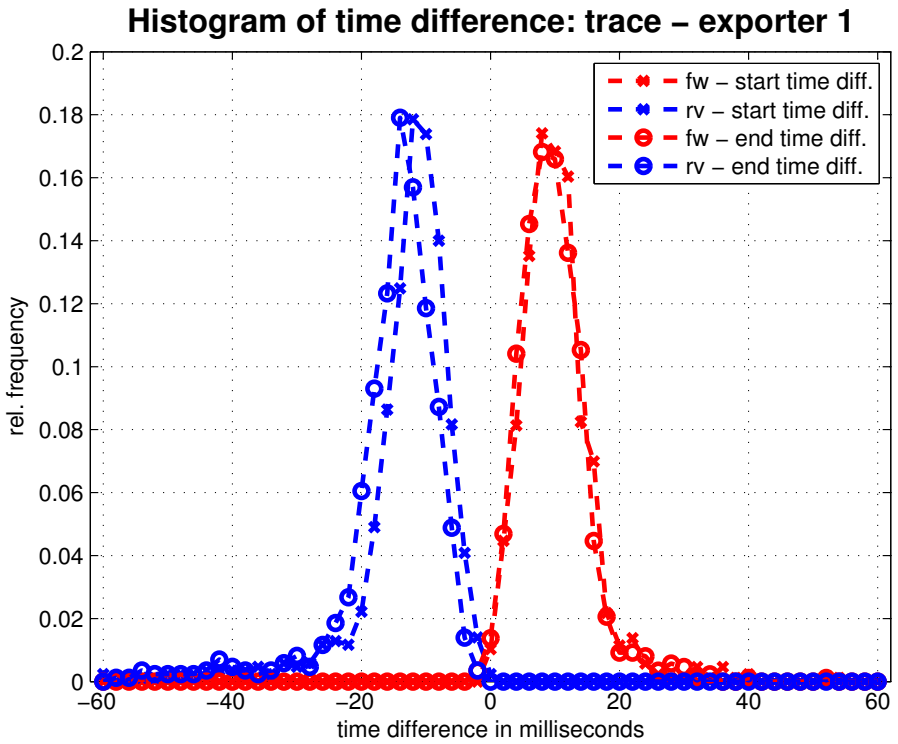
Uhrenversatz und -drift (CE-Routers)

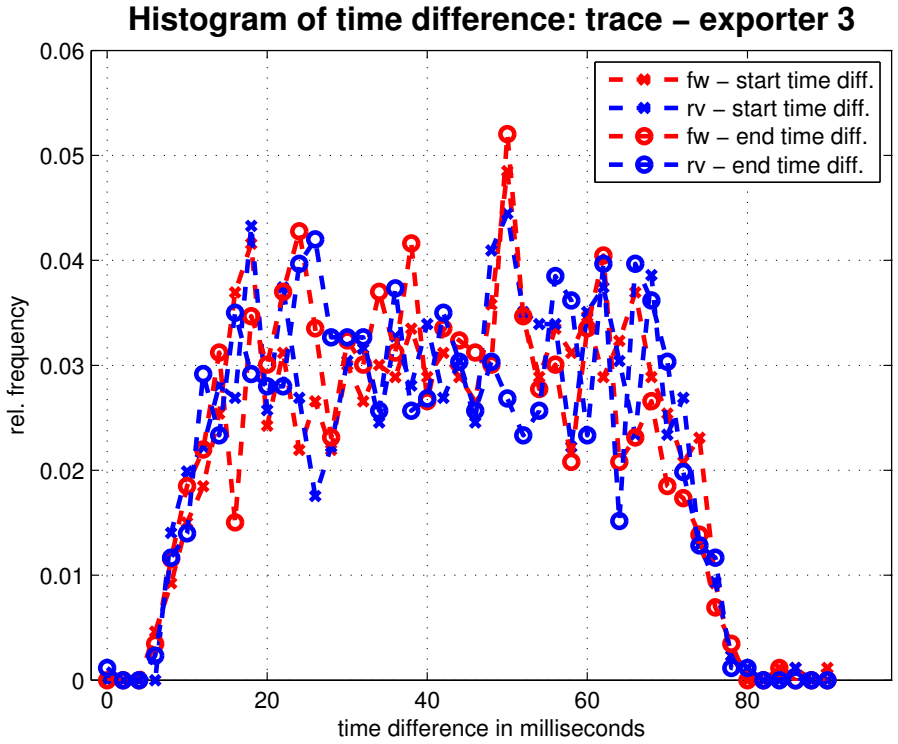


Genauigkeit von Flowdaten

Vergleich von Paket-Trace und Netflow: Uhren

Verteilung der Zeitdifferenzen





– Breite der Abweichung hängt von Exportertyp ab
– Exporter mit wenig Abweichung (links) ca +/- 10 ms
– Beobachtung: Unterschied zwischen Start- und Endzeitdifferenz

→ Zeitstempelauflösung?

Genauigkeit von Flowdaten

Vergleich von Paket-Trace und Netflow: Zeitstempelauflösung

Bestimmung der Zeitstempelauflösung

- Berechnung der Abstände zwischen Start- und Endzeitstempel, sowie der Dauer
- und/oder Bestimmung des größten gemeinsamen Teilers

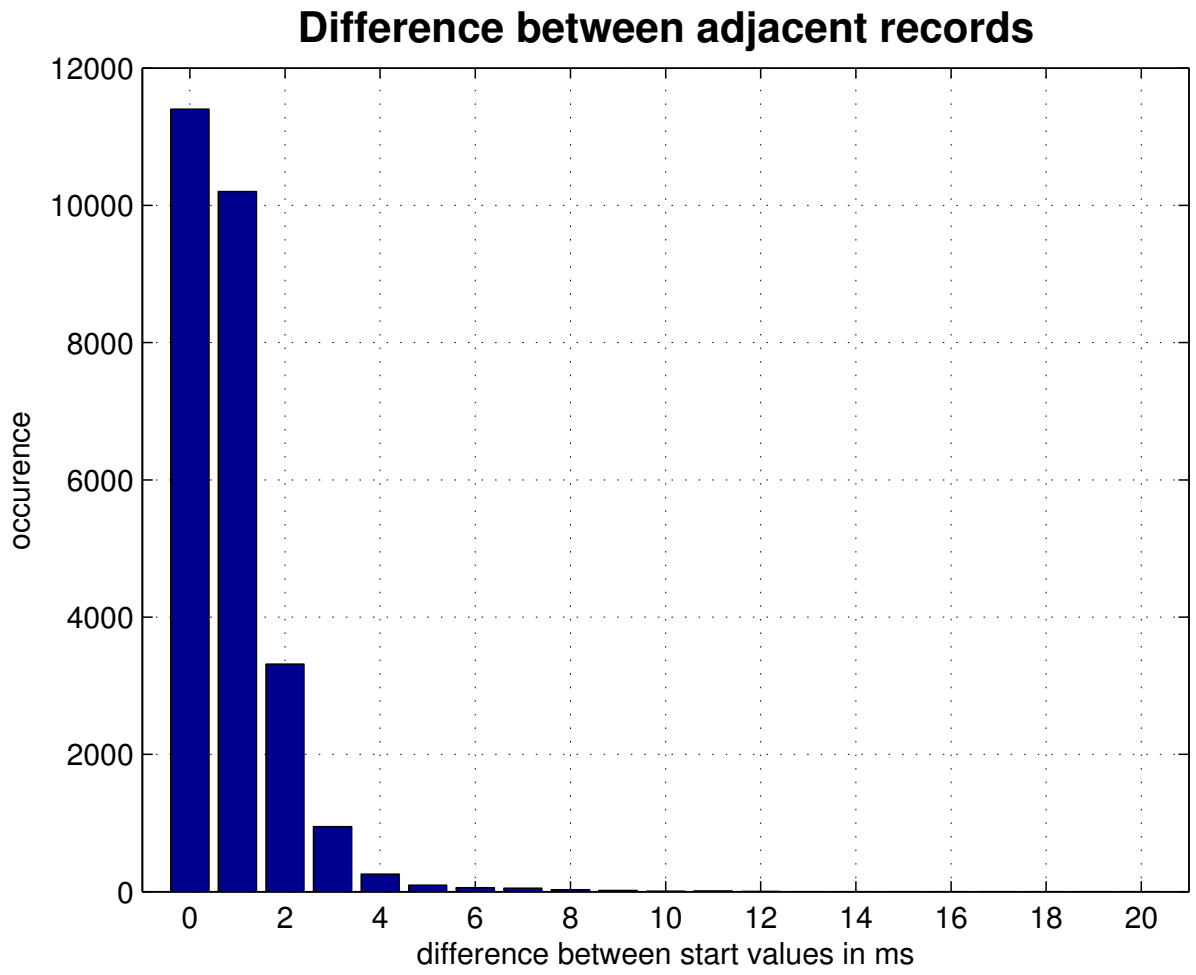
Ergebnisse

- Auflösung Start-/Endzeitstempel: 1 ms, 4 ms oder 64 ms
- Teilweise untere Bits abgeschnitten, teilweise untere Bits "rauschen"

Genauigkeit von Flowdaten

Nur aus NetFlow Daten: Zeitstempelauflösung

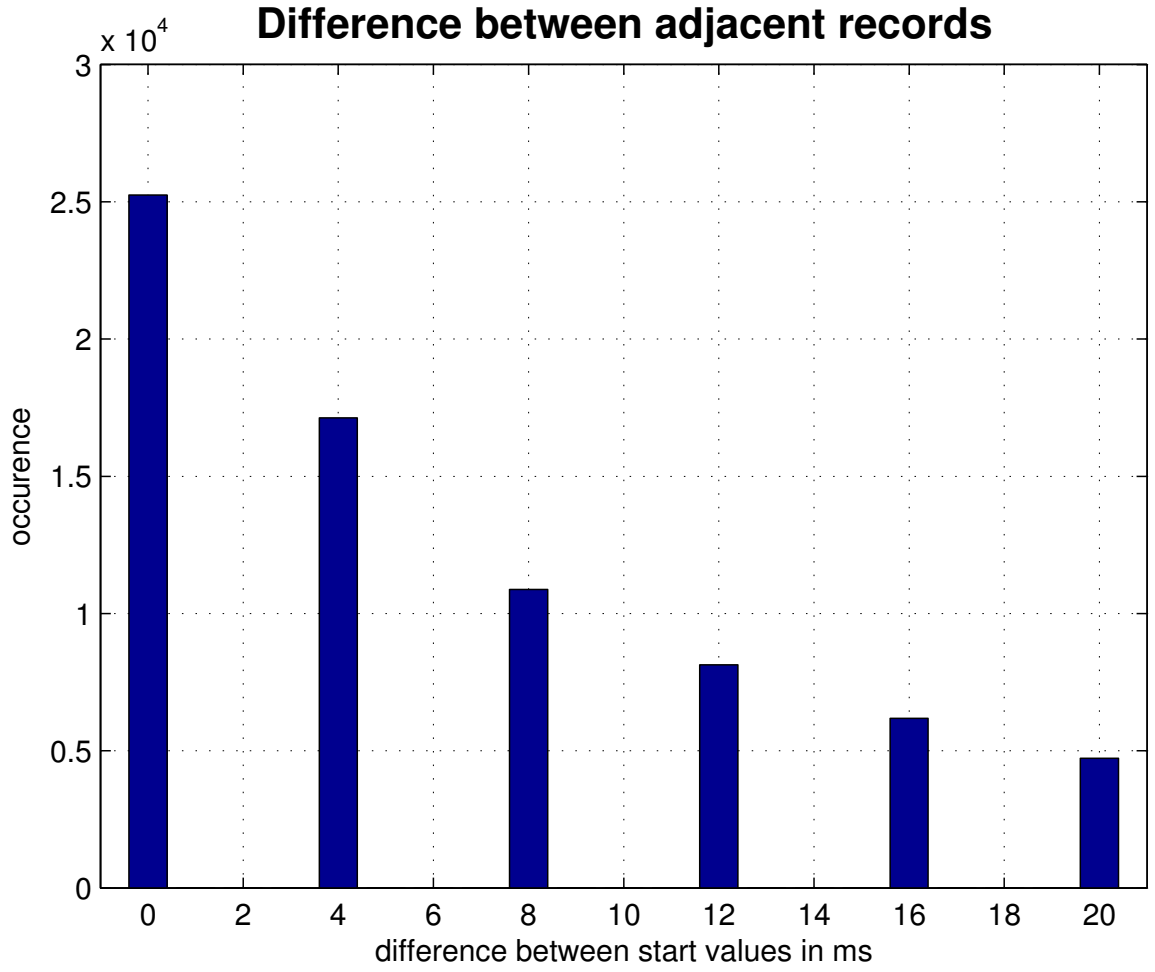
Beispiel 1: 1ms Auflösung



Genauigkeit von Flowdaten

Nur aus NetFlow Daten: Zeitstempelauflösung

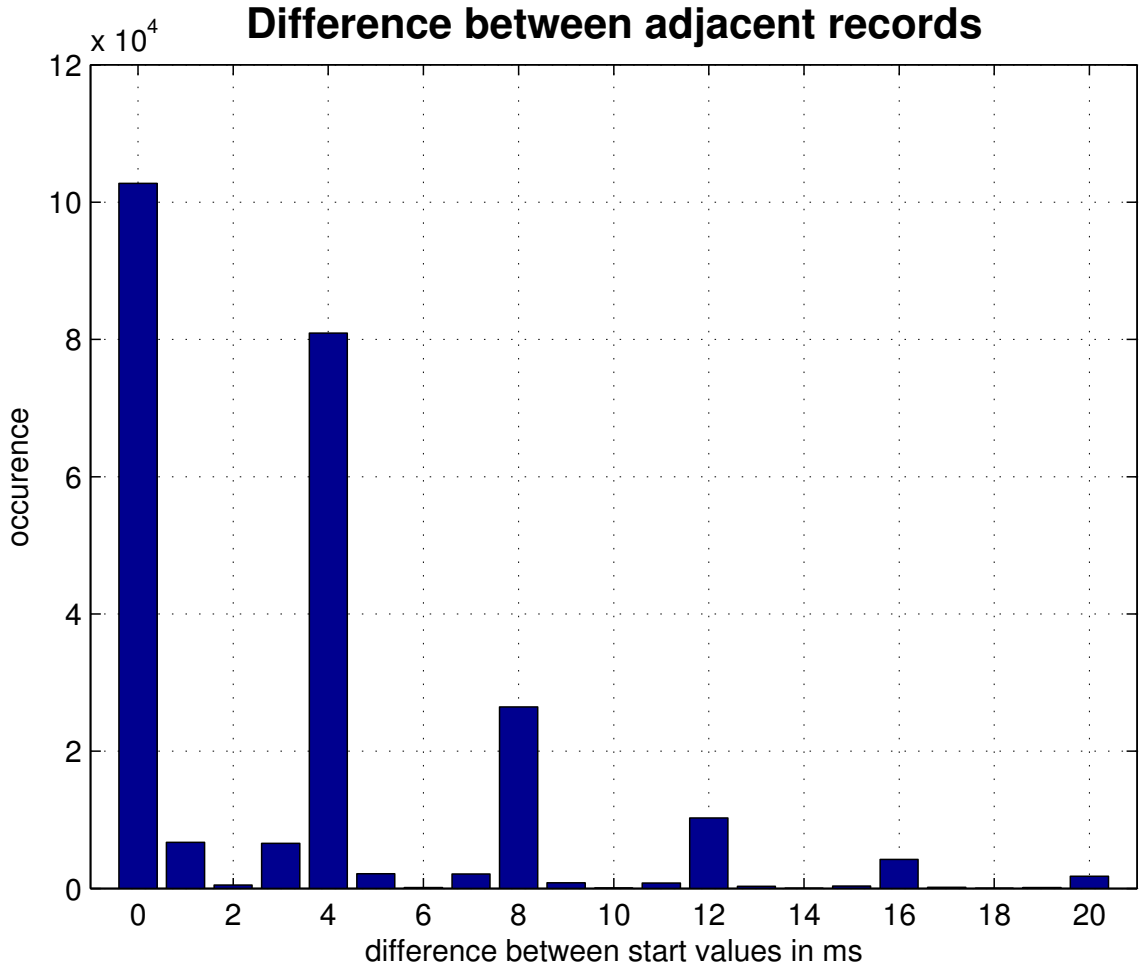
Beispiel 2: 4 msAuflösung



Genauigkeit von Flowdaten

Nur aus NetFlow Daten: Zeitstempelauflösung

Beispiel 3: 4 ms Auflösung (einfache Berechnung: 1 ms)



Genauigkeit von Flowdaten

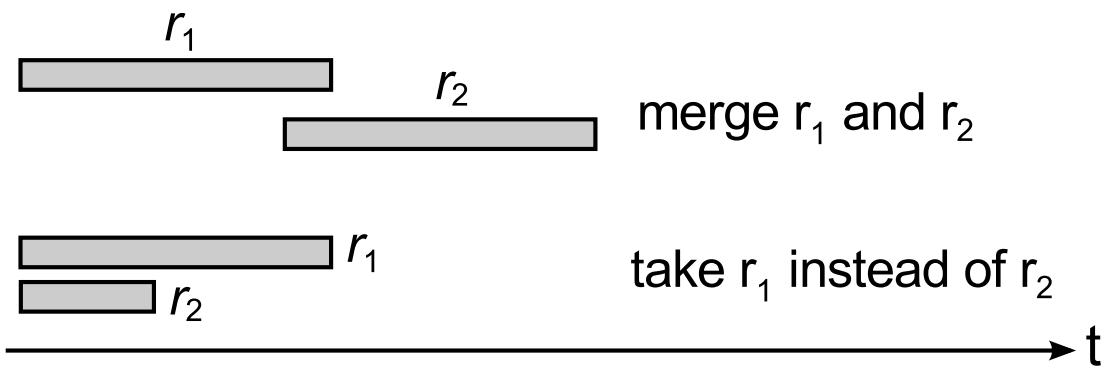
Nur aus NetFlow Daten: Duplikate

Definition "Duplikat"

Mehr als ein Record für denselben Flow im selben Zeitintervall

Wie verarbeiten?

→ hängt vom Typ des Duplikats ab



Zusammenfassung und Ausblick

Beobachtung

- Zeitstempel haben 1 ms Auflösung, interne Auflösung teilweise schlechter
- Exporter verhalten sich unterschiedlich hinsichtlich Zeitstempeln, Bytezählern, Duplikaten,...
- Bei der Verarbeitung müssen diese Effekte berücksichtigt werden

→ Exporterprofil, das jeden Exporter beschreibt

Exporterprofil

- Exporter-spezifischer Teil
  - Zeitstempelauflösung und -genauigkeit, Bytezähler-Probleme
  - Verarbeitung von Duplikaten
- Netzkonfiguration/Szenario-spezifischer Teil
  - Uhrenversatz/-drift
  - Positionen von Middleboxes etc.

Erzeugung des Exporterprofils

- Hersteller (?)
- Vergleich mit Paket-Trace im Betrieb?
- Rein aus Flowdaten (z.B. Daten von gering ausgelastetem Netz)