Characterization of accuracy problems in NetFlow data and approaches to handle them

IRTF NMRG / 3rd NetFlow/IPFIX Workshop

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

IETF 78 Maastricht
30 July 2010

University of Stuttgart
Institute of Communication Networks
and Computer Engineering (IKR)
Prof. Dr.-Ing. Andreas Kirstädter

Outline

Motivation: why looking at accuracy of data?

Accuracy issues

Handling problems with exporter profile

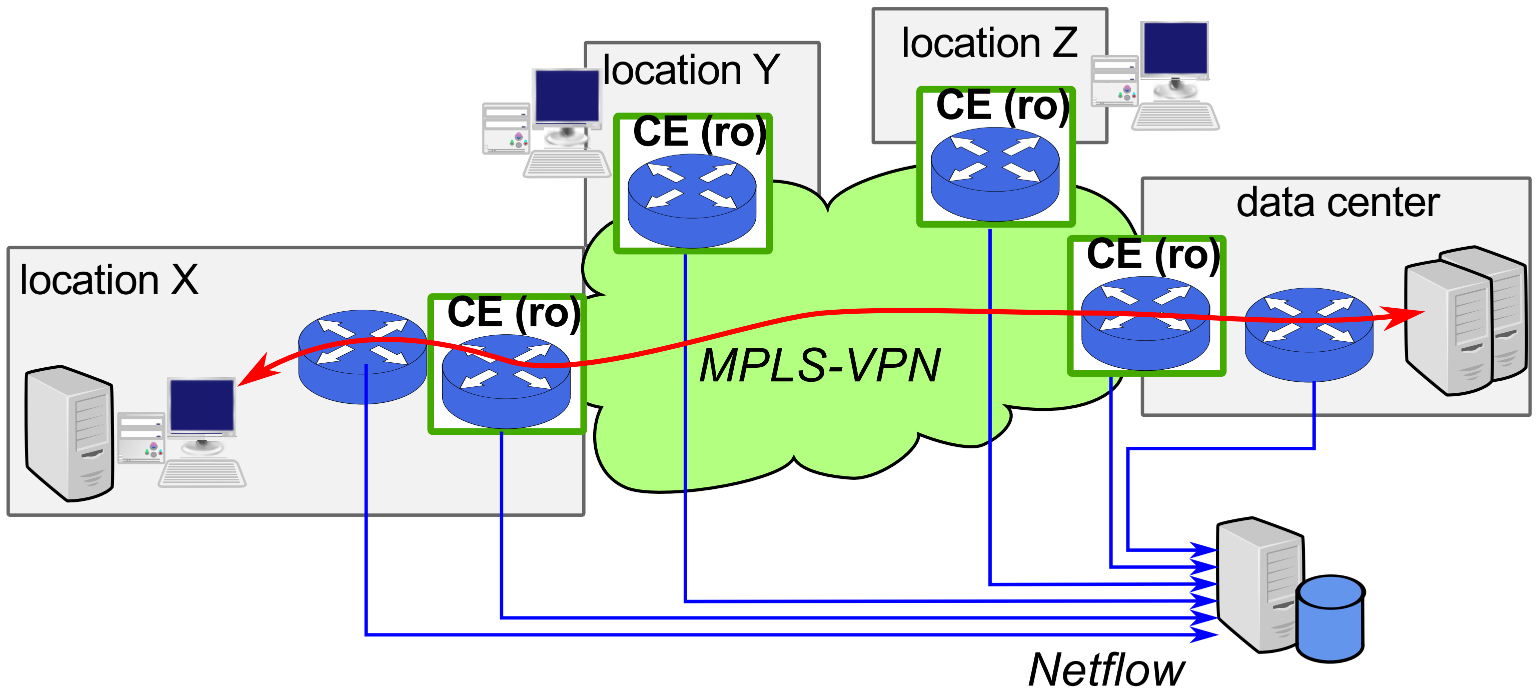
Conclusion and Outlook

Motivation

Scenario

Global Enterprise Network, MPLS-VPN

- Flow data from several (1..5) routers on path
- Own routers (full control)
- CE-Routers of carrier ("read only")



→ Netflow-based (v5) view on traffic at several points in the network

→ Correlation of Flow data for extraction of network characteristics

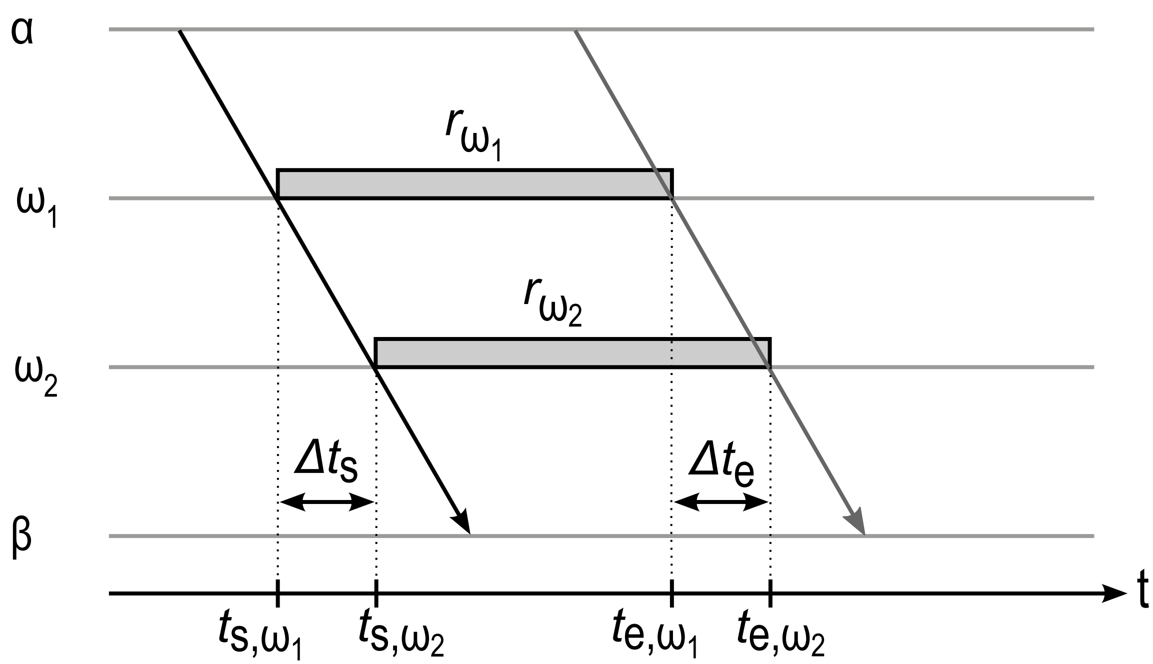
Motivation

Extraction of network charakteristics

Extractable characteristics are e.g. one-way-delay, RTT, packet loss, flow contention

→ Requires matching flow records

- For the same 6-Tuple (src/dst address, src/dst port, protocol, ToS)
- Exported from different observation points (exporter + input interface)



Motivation

Extraction of network charakteristics

Matching

First try (very strict rules)

- Take flows exported in one record only
- Drop implausible records
- Match records, match forward and reverse, drop implausible data

$\rightarrow$ 12,500 bidirectional trajectories left from 22 million records (10 samples per path and hour)

$\rightarrow$ Two questions

1. **Precission** of characteristics obtained from flow-data wrt. timings, bytes, packets
2. What do we have to consider in consistency and plausibility checks in order to get **a high amount of samples**?

Accuracy issues

Overview of issues (not all shown in detail)

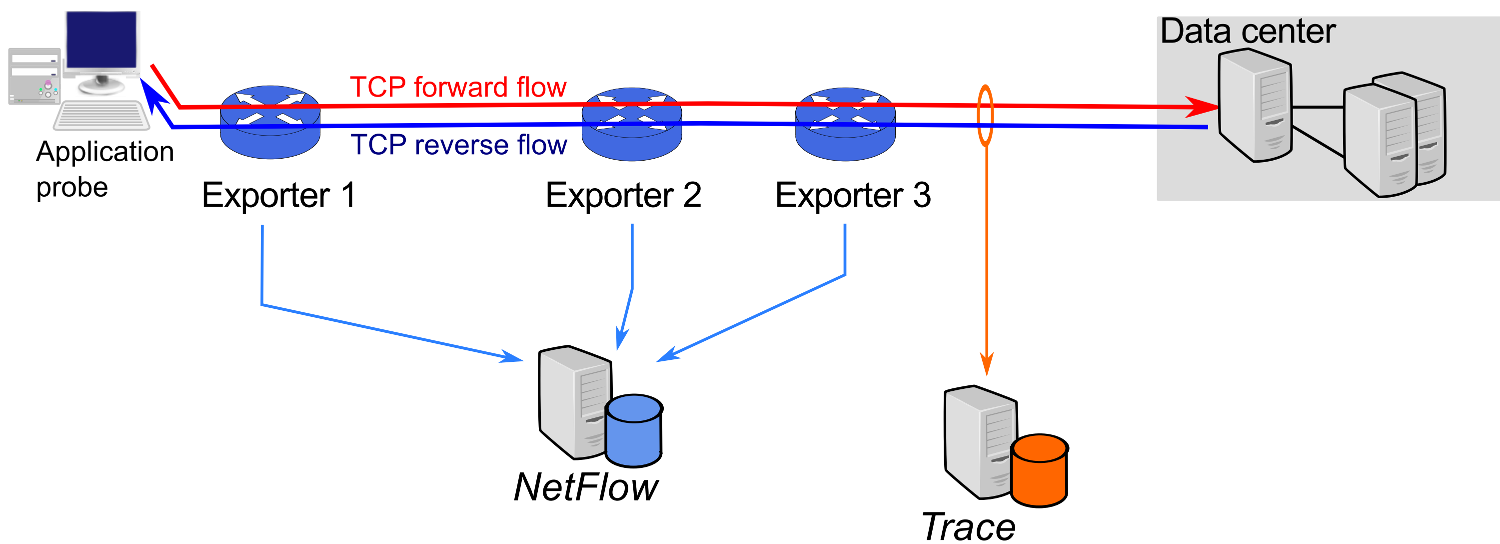
- Record loss (the simplest one)
- Duplicates
- Packet counters
- Byte counters
- Clock accuracy
  - Granularity
  - "Noise"
  - Jumps
  - Clock offset, clock skew

Different reasons

- Inaccuracies at exporters
- Configuration issues
- Middle boxes

Accuracy issues

Comparing trace to NetFlow: scenario



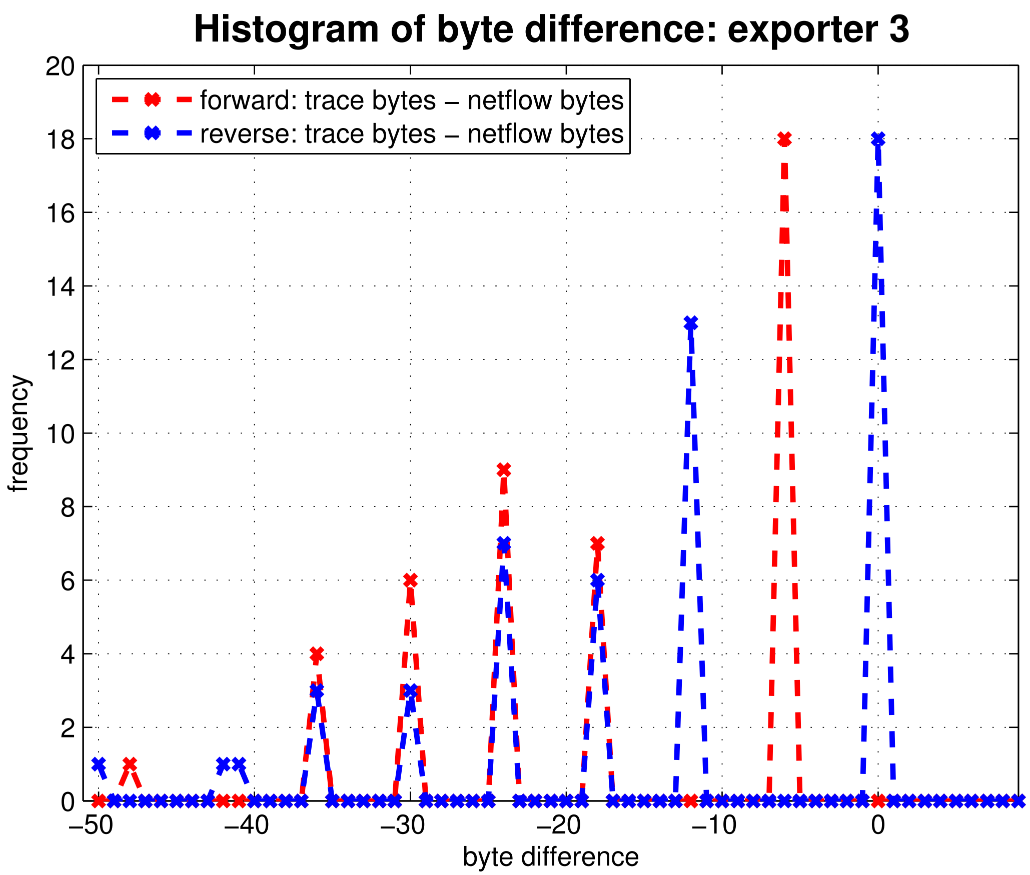
- Path between two European cities
- 5 day packet trace, filtered on two endpoints: application probe and server
- Flow data from three exporters (two of them CE)

Accuracy issues

Comparing trace to NetFlow: byte count

Byte count Issue

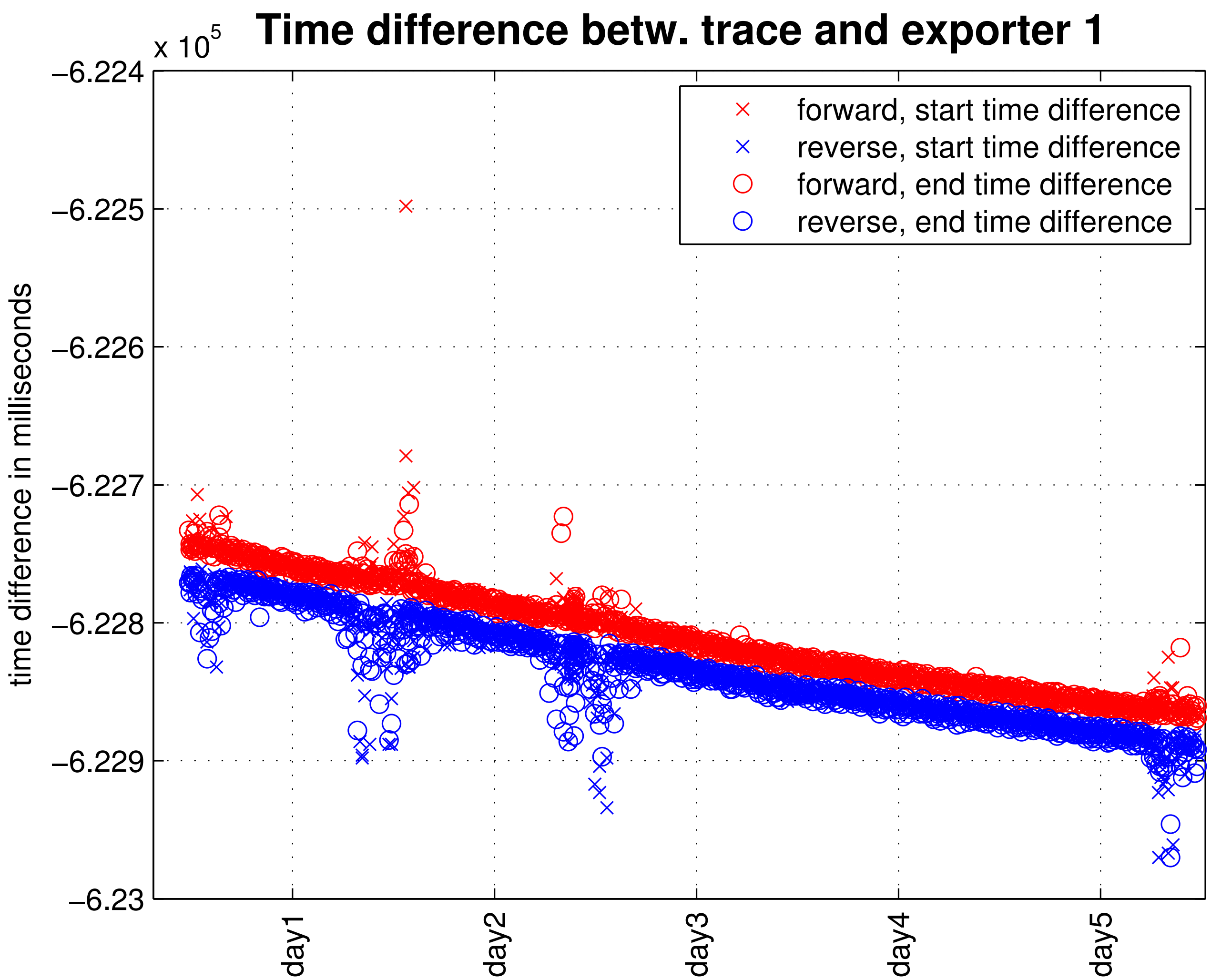
- Byte count different at one observation point, but packet count consistent
- Here: router rounds byte count up to 46 Bytes
- Side note: Similar effects from some middle boxes (WAN optimizers)



Accuracy issues

Comparing trace to NetFlow: clocks

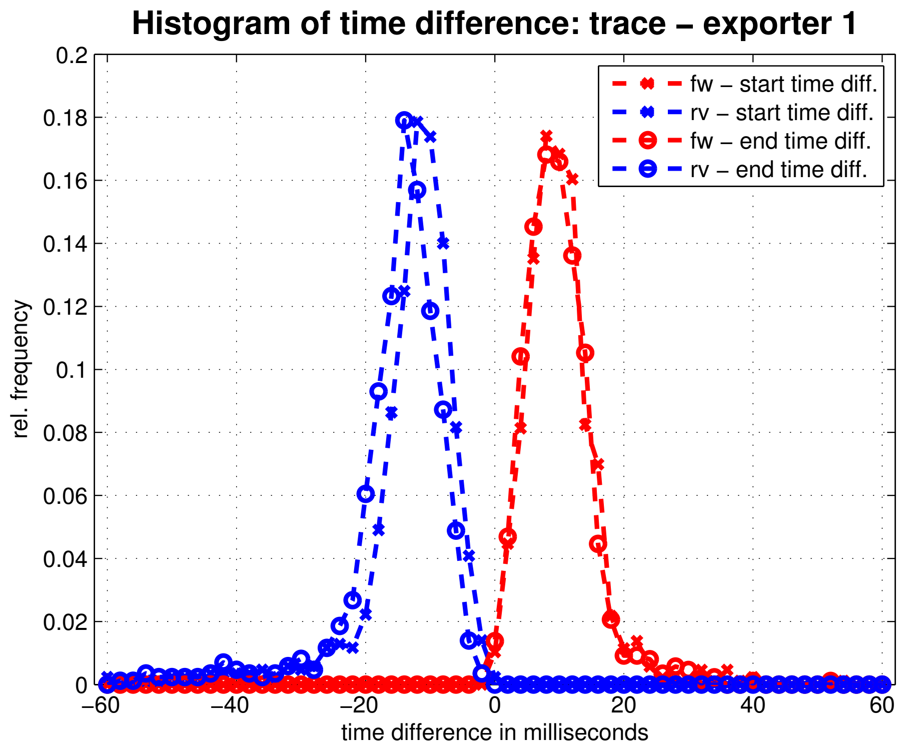
Clock Offset and Skew (CE-Routers)

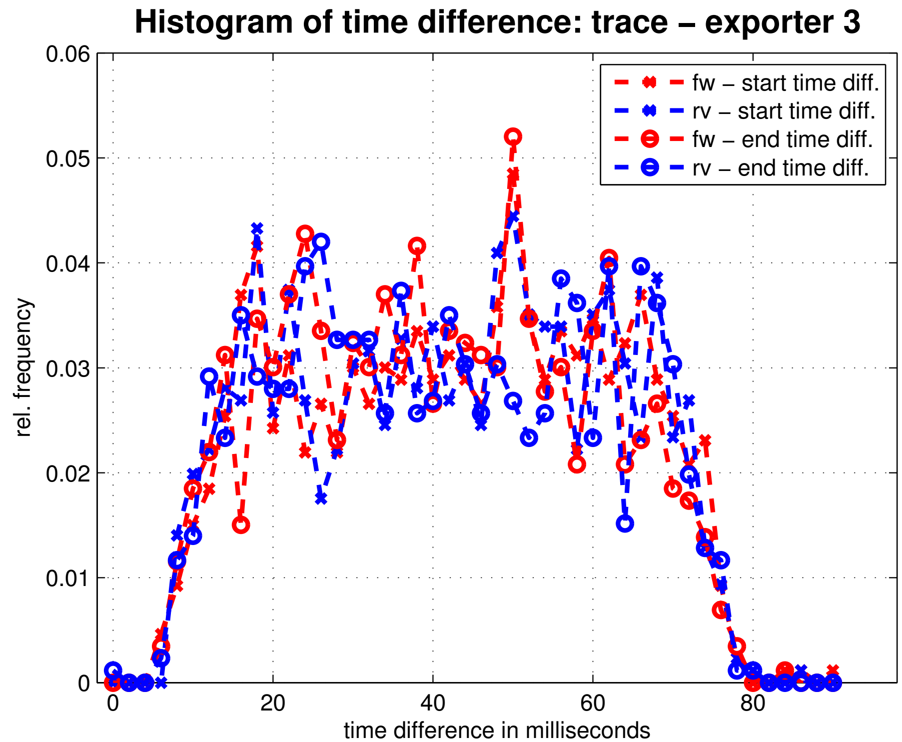


Accuracy issues

Comparing trace to NetFlow: clocks

Distribution of time difference





- "Signal to Noise Ratio" depends on exporter
- On good exporters (left) accuracy around +/- 10 ms. Right: much more noise.
- Note: difference between start and end time diff of reverse-flow

  → granularity issue?

Accuracy issues

From NetFlow data only: granularity of clocks

Determination of granularity

- calculate difference between record start times, duration, end times, ...
- and/or calculate greatest common divisor

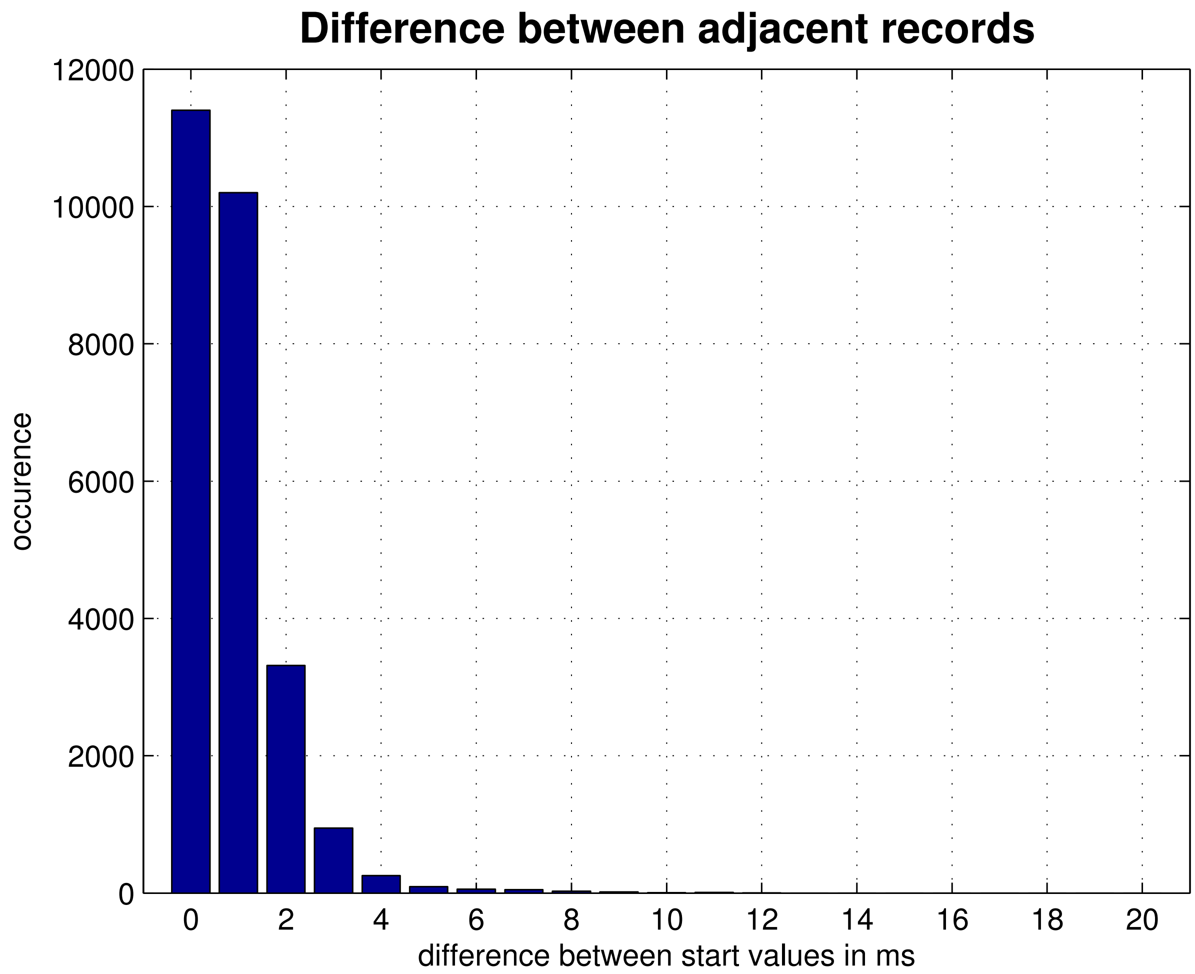
Results

- start/end time granularity: 4 ms or 1 ms (see following slides)
- duration granularity: 4 ms on all exporters
- nsecs-granularity: 15,258 (1e9/2^16)

Accuracy issues

From Netflow data only: granularity of clocks

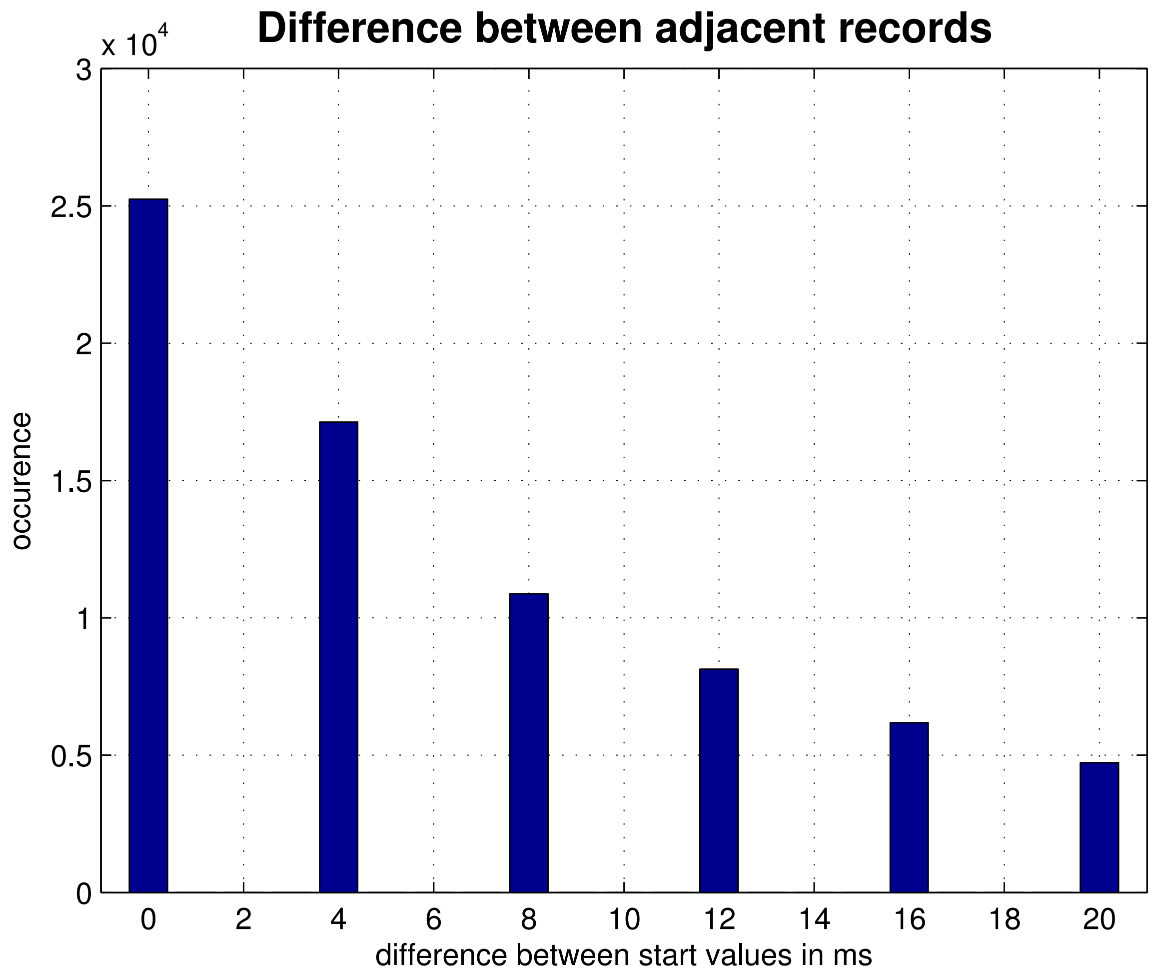
This exporter: 1ms granularity



Accuracy issues

From Netflow data only: granularity of clocks

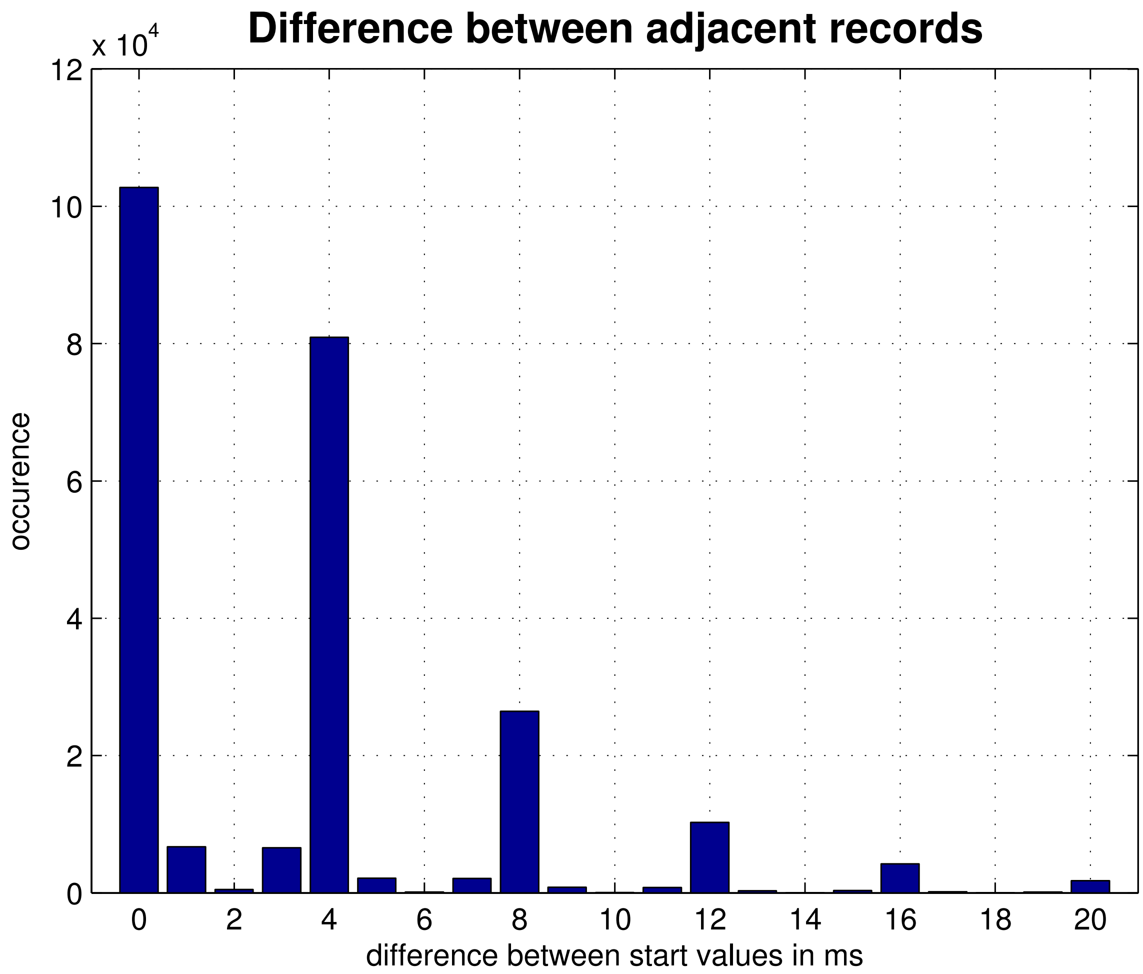
Another exporter: 4 ms granularity



Accuracy issues

from Netflow data only: granularity of clocks

Yet another one: 4 ms granularity (simple calculation would reveal 1 ms)



Accuracy issues

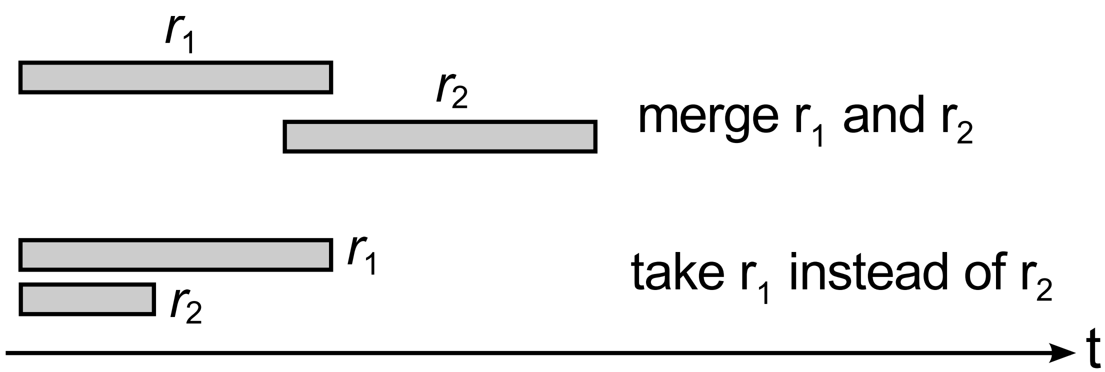
From NetFlow data only: duplicates

Definition of "Duplicates"

More than one record for the same key within a time interval.

What to do?

→ depends on type of duplicate



Handling problems using an exporter profile

Observation

– Routers behave differently wrt. accuracy of timestamps, bytecount, duplicates,...
– Knowing these effects can lead to more/better results
→ Exporter profile that describes effects and accuracy for each exporter

Exporter profile

– Exporter-specific par
  - Timestamp granularity, Timestamp accuracy and behavior
  - How to handle duplicates
  - Bytecount problems
– Configuration/scenario-specific part
  - Clock offsets/skew
  - Middlebox locations

How to obtain exporter profiles?

– Manufacturer (?)
– "Calibration" using packet trace
– Using NetFlow data only (e.g. more accurate data from off-peak times)

Conclusion and Outlook

Conclusion

- Accuracy issues wrt. timestamps, byte count, duplicates identified
- Effects depend on router
- NetFlow data from "good" routers is suitable for estimating one-way-delays with at least +/- 20 ms accuracy

Outlook

- Exporter Profile
  – Profile format and relation to other (configuration) items
  – Methods to create exporter profile
  – Evaluate improvements wrt. acccuracy from different features in the Exporter Profile
  – Load dependency?

Discussion: More accuracy issues known?