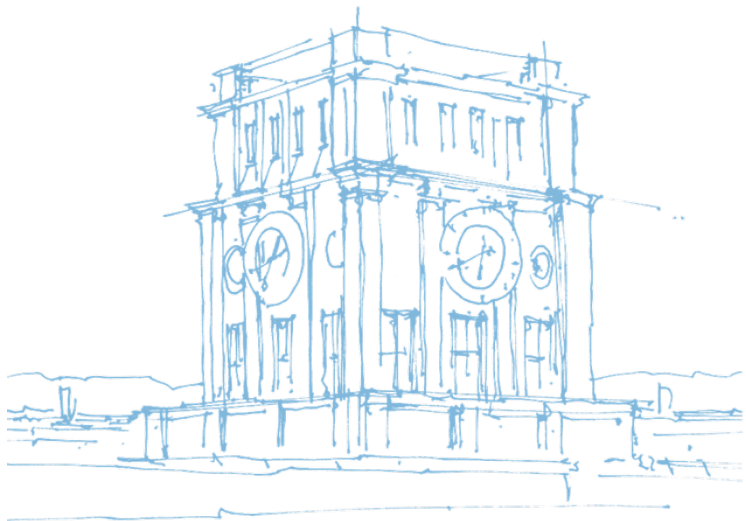
Chair of Network Architectures and Services
Department of Informatics
Technical University of Munich

An Approach Towards Validation of IPv4 and IPv6 Siblings

Minoo Rouhi

November 25, 2016

Chair of Network Architectures and Services
Department of Informatics
Technical University of Munich



Introduction & Motivation

Problem Statement & Research Questions

Methodology & Ground-truth

Evaluation of TCP Timestamp Fingerprinting

Large-scale Measurements

Introduction & Motivation

- **Sibling**: IPv4 and IPv6 address pair assigned to the same physical machine [1]
- Increasing trend in usage of shared IP infrastructure [1, 2]
- Application areas:
  - Understanding IPv6 and the Internet evolution
  - Understanding correlated failures and loopholes
  - IPv6 geolocation
  - IPv4 vs. IPv6 performance

Problem Statement & Research Questions

- Given a pair (IP_4, IP_6), determine whether it is a Sibling
- A common DNS name does not always imply a Sibling relationship [3, 1, 2]
  - Content Distribution Networks
  - Load balancers
  - ...
- Fingerprinting techniques needed to discern Siblings

Methodology & Ground-truth

1. Acquiring the Ground-truth:
   - Siblings dataset
     - 458 true associations (Siblings)
   - Non-siblings dataset
     - Pairing unrelated IPv4 and IPv6 addresses
2. Evaluating fingerprinting methods against the Ground-truth

TCP Timestamp Fingerprinting

Introduction

Terminology:

- **Offset**: The time difference between the target and reference clock.
- **Skew**: The frequency difference between the target and the reference clock
  $\rightarrow$ First derivative of the offset

TCP Timestamp Fingerprinting

Introduction

Terminology:

- **Offset**: The time difference between the target and reference clock.
- **Skew**: The frequency difference between the target and the reference clock
  $\rightarrow$ First derivative of the offset

Objective:

- Fingerprint devices from their clock skew

TCP Timestamp Fingerprinting

First Order Filter using TCP Options Signature

- TCP options are almost always identical for Siblings
- Discriminating factors:
  - Presence of options and their order
  - Value of the window scale option

TCP Timestamp Fingerprinting

First Order Filter using TCP Options Signature

- TCP options are almost always identical for Siblings
- Discriminating factors:
  - Presence of options and their order
  - Value of the window scale option

✓Eliminates $\approx$ 71% of Non-siblings
✓No false negative rate

TCP Timestamp Fingerprinting

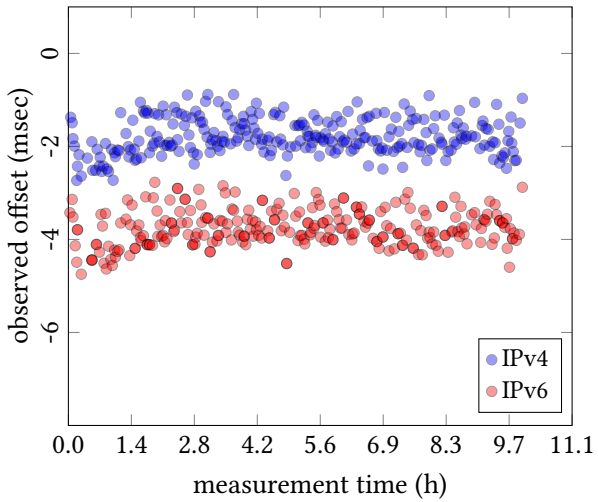
Obtaining Offsets

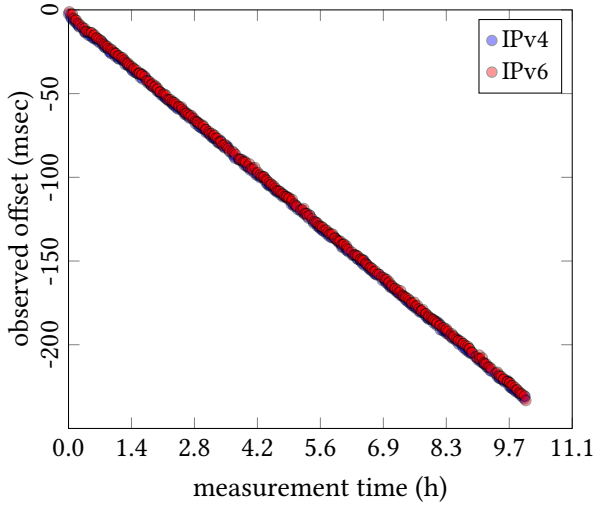
Algorithm 1 Obtaining offsets

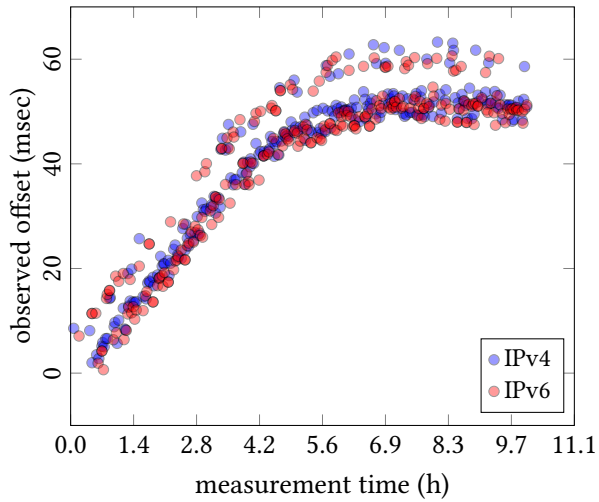
1: Probe IP pair
2: Store traces $\mathcal{T}_4$ and $\mathcal{T}_6$
3: **for each** $Packet_i \in \mathcal{T}_4 \vee \mathcal{T}_6$ **do**
4: Extract $TSval_i$ and $ArrivalTime_i$
5: $\Delta_i \leftarrow TSval_i - ArrivalTime_i$
6: $Offset_{set} \leftarrow (ArrivalTime_i, \Delta_i)$
7: **end for**
8: Plot offset trends from $Offset_{set}$

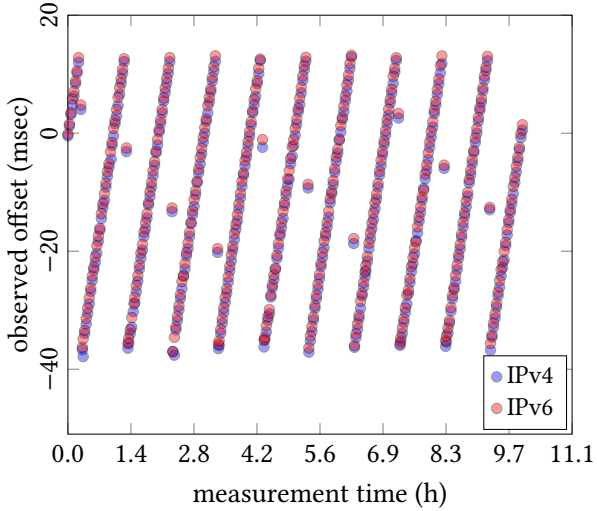
TCP Timestamp Fingerprinting

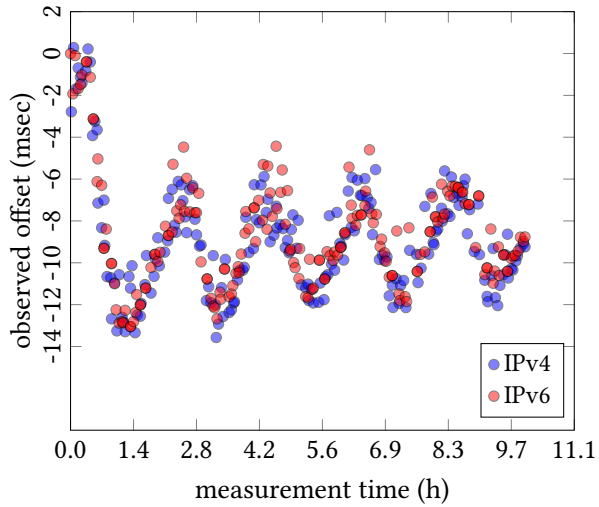
Observation Classes





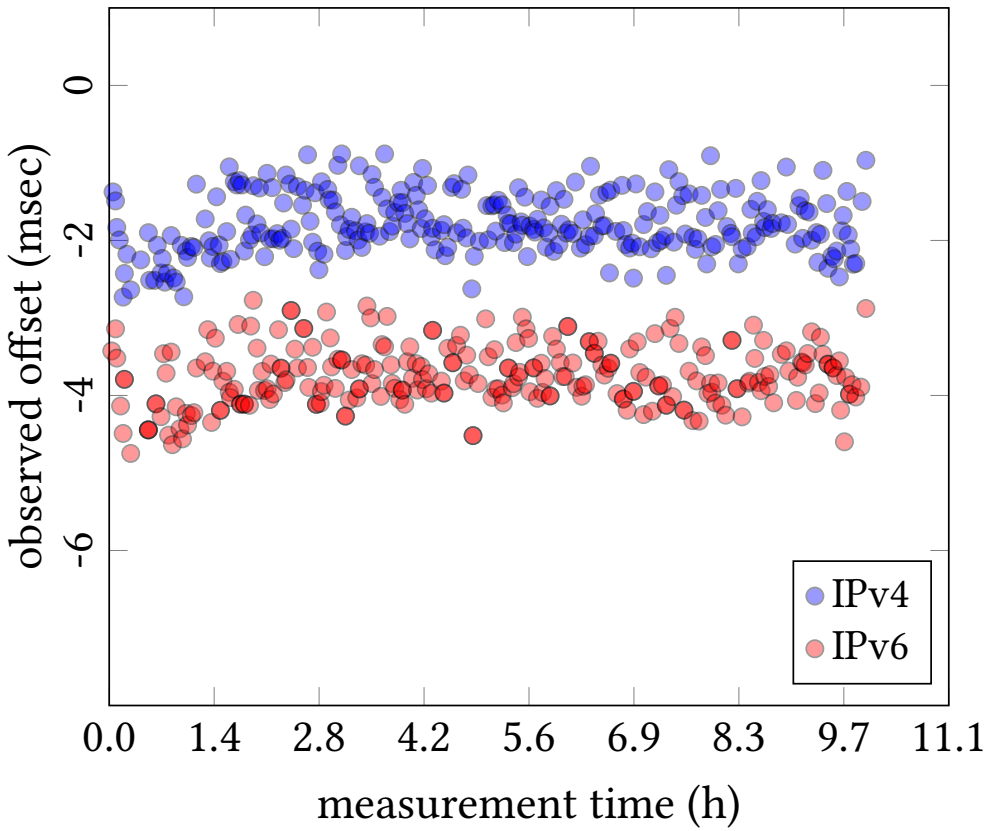






TCP Timestamp Fingerprinting

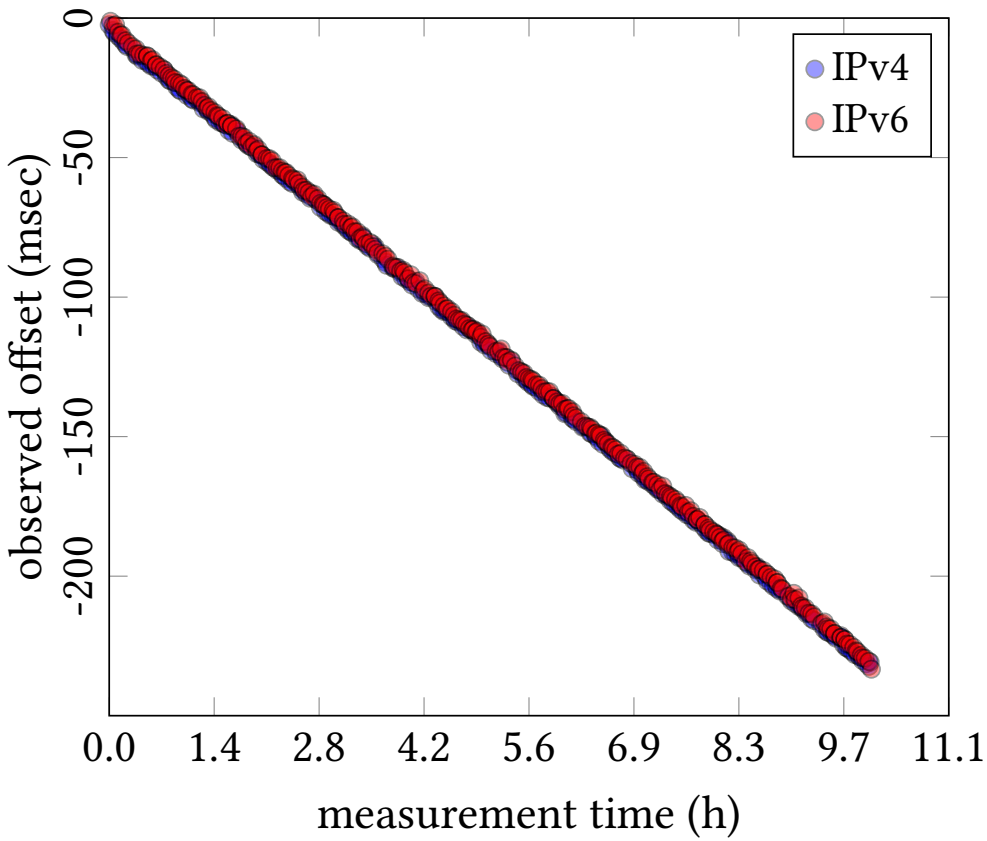
Negligible Skew



- Skew is negligible
- Metric: $|\text{offset}_{max} - \text{offset}_{min}|$
- 1.6% of the Ground-truth

TCP Timestamp Fingerprinting

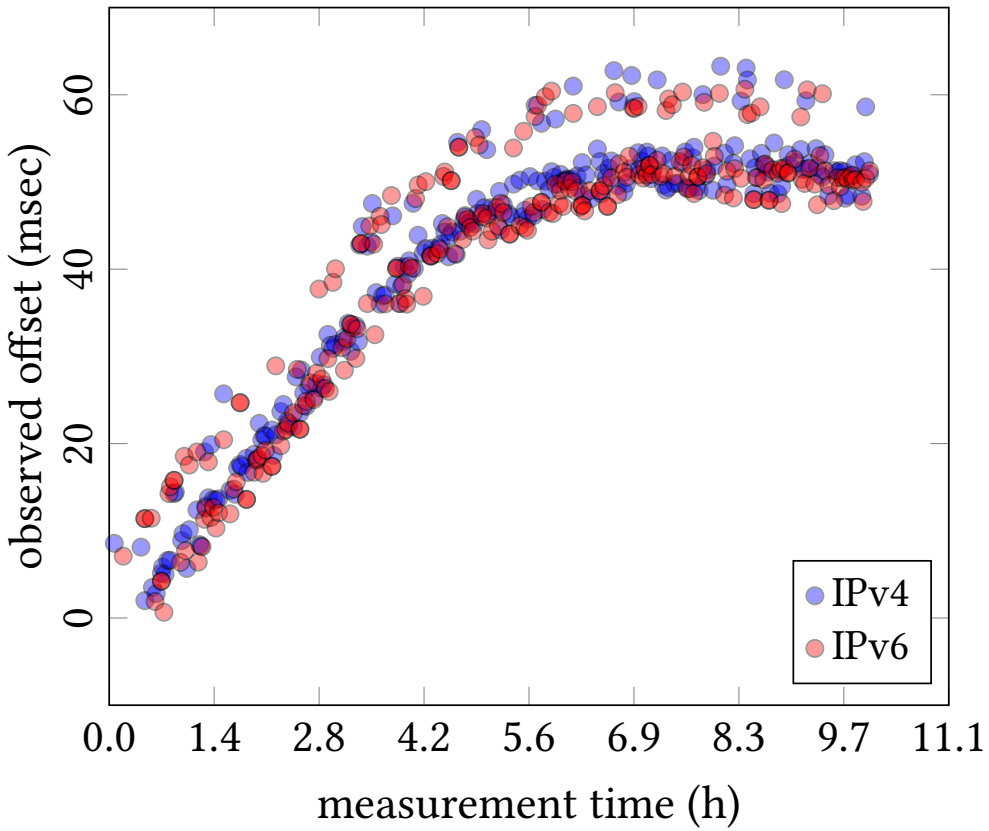
Constant Skew



- Skew is constant
- Metric:
  Robust Linear regression
- 3.2% of the Ground-truth

TCP Timestamp Fingerprinting

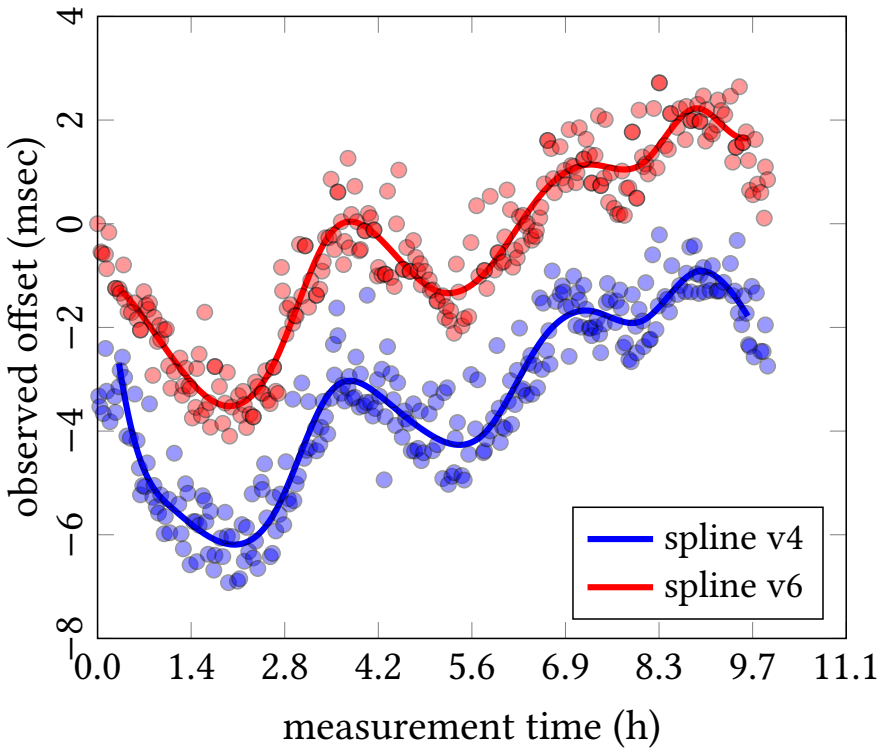
Variable Skew (Drift)



- Skew is variable
- Metric: Polynomial splines
- 95.2% of the Ground-truth

TCP Timestamp Fingerprinting

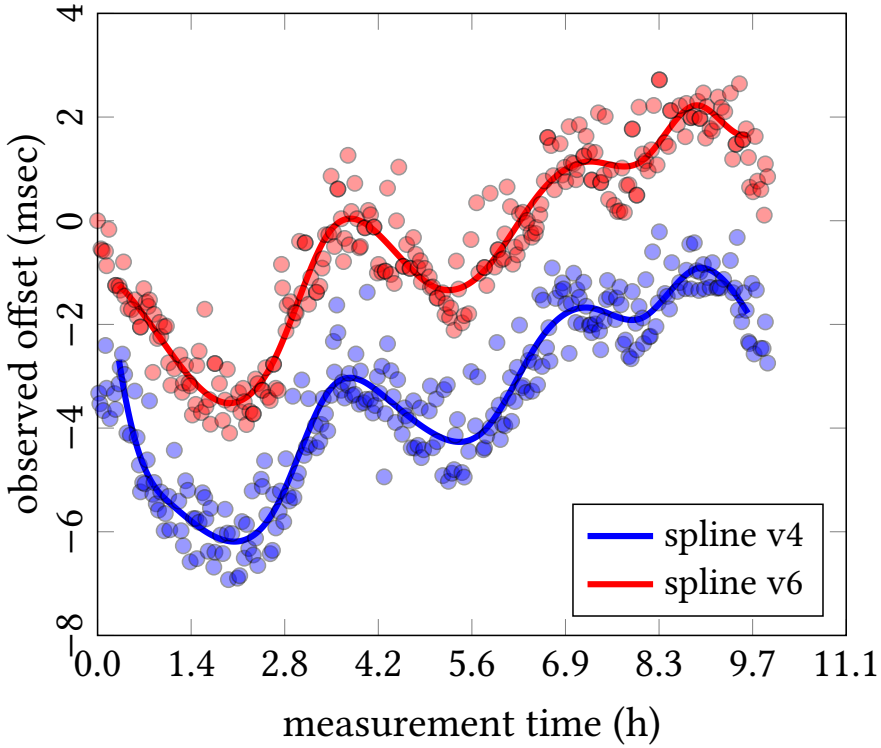
Polynomial Splines

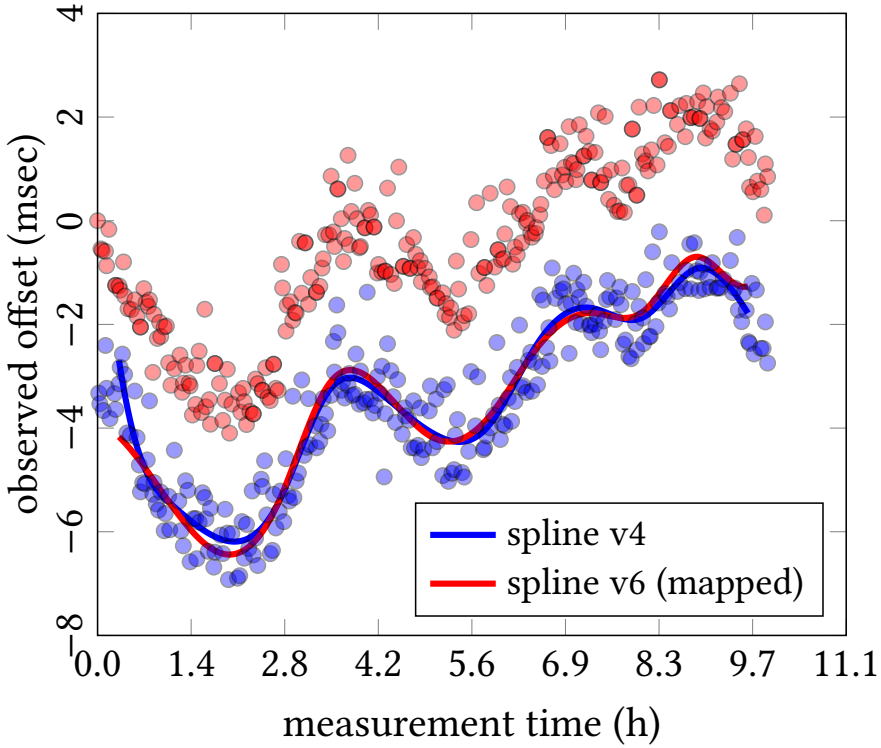


1. Calculate splines

TCP Timestamp Fingerprinting

Polynomial Splines

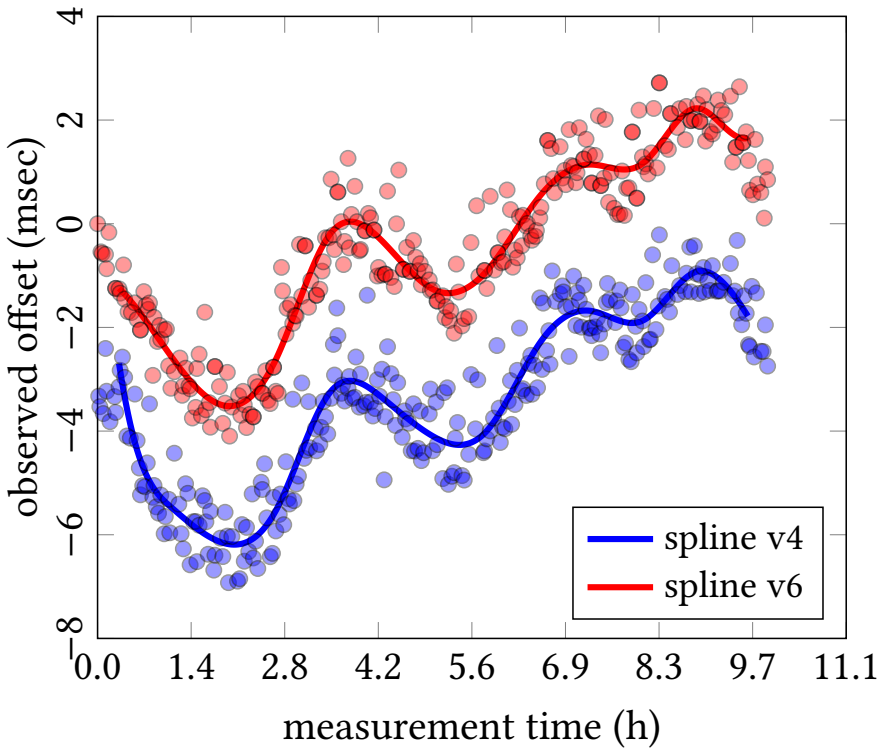


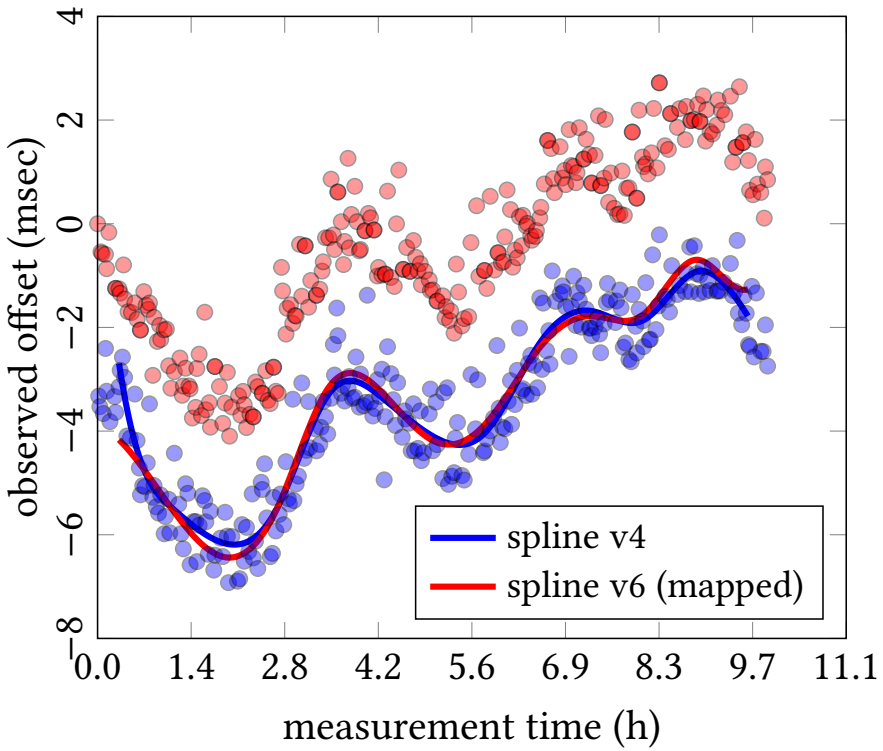


1. Calculate splines
2. Map splines

TCP Timestamp Fingerprinting

Polynomial Splines

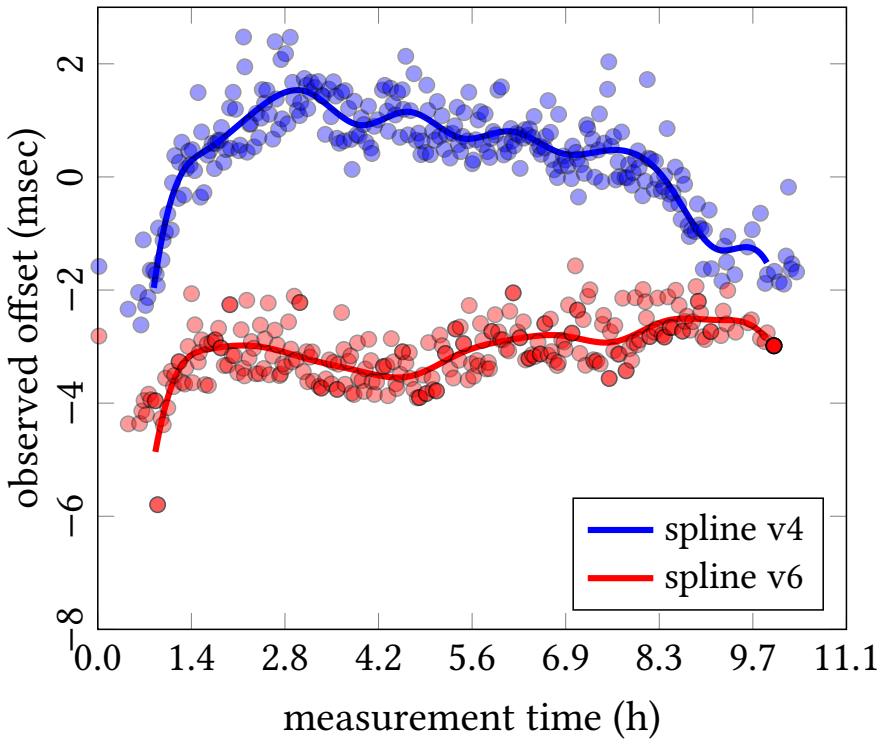


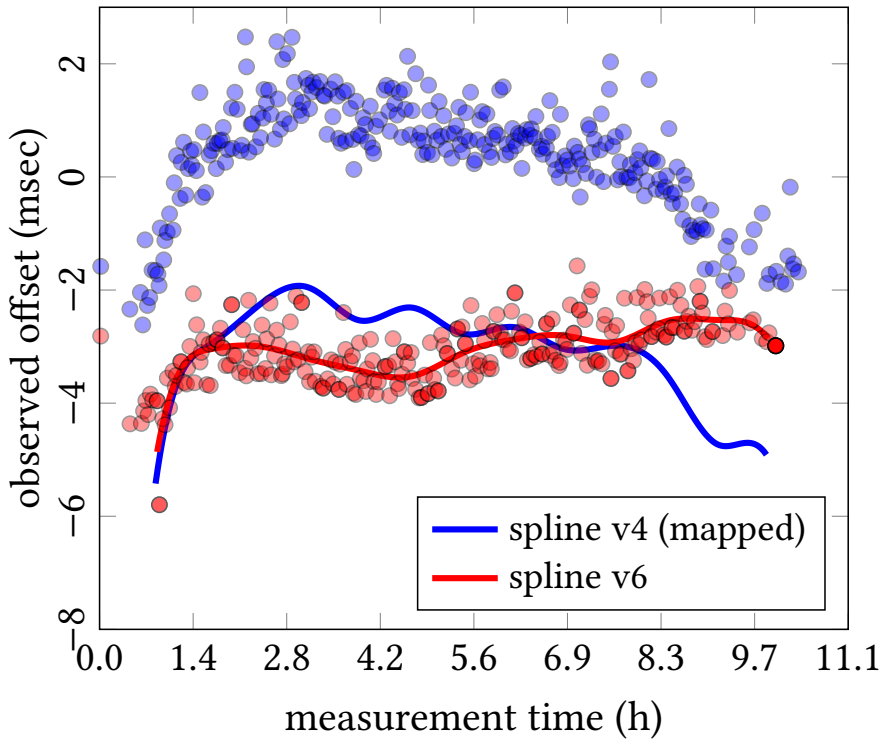


1. Calculate splines
2. Map splines
3. $\text{spline}_{dist} \leq \text{threshold} \rightarrow \text{Sibling}$

TCP Timestamp Fingerprinting

Polynomial Splines

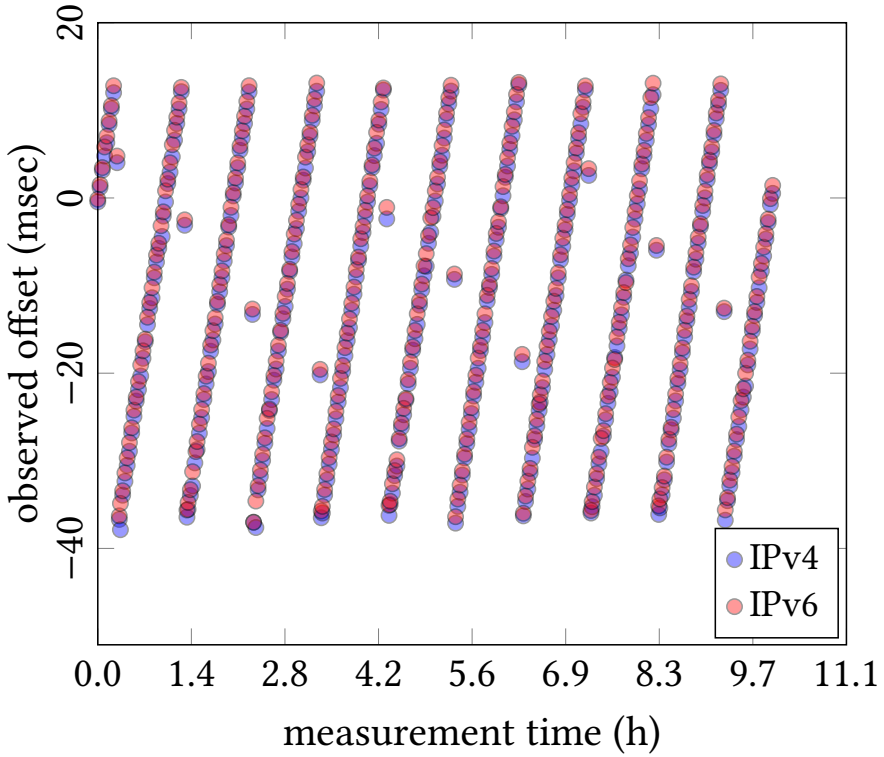


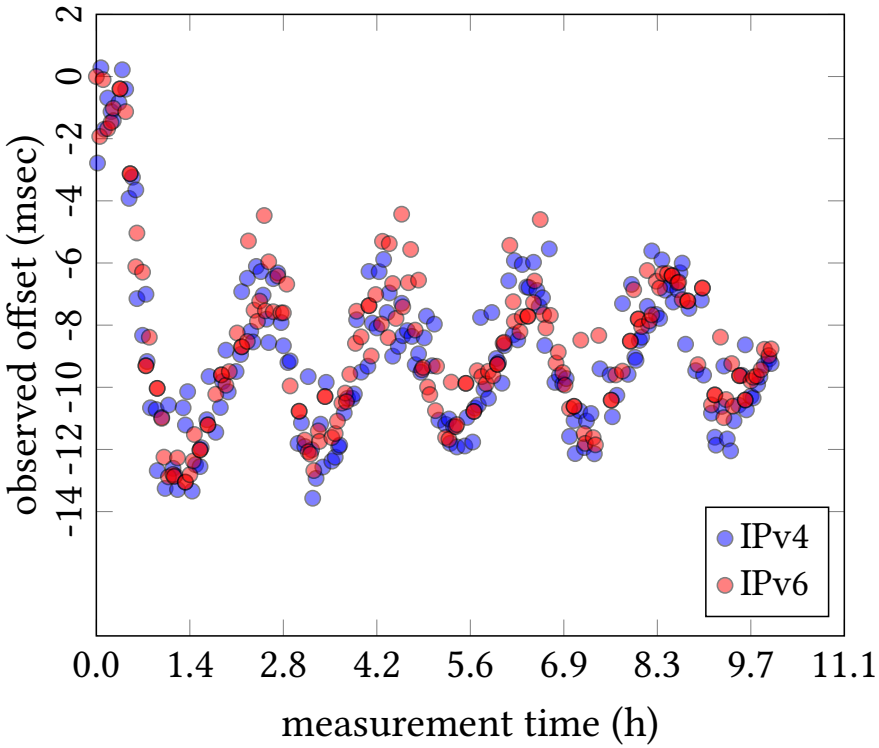


$\text{spline}_{dist} > \text{threshold} \rightarrow \text{Non-Sibling}$

TCP Timestamp Fingerprinting

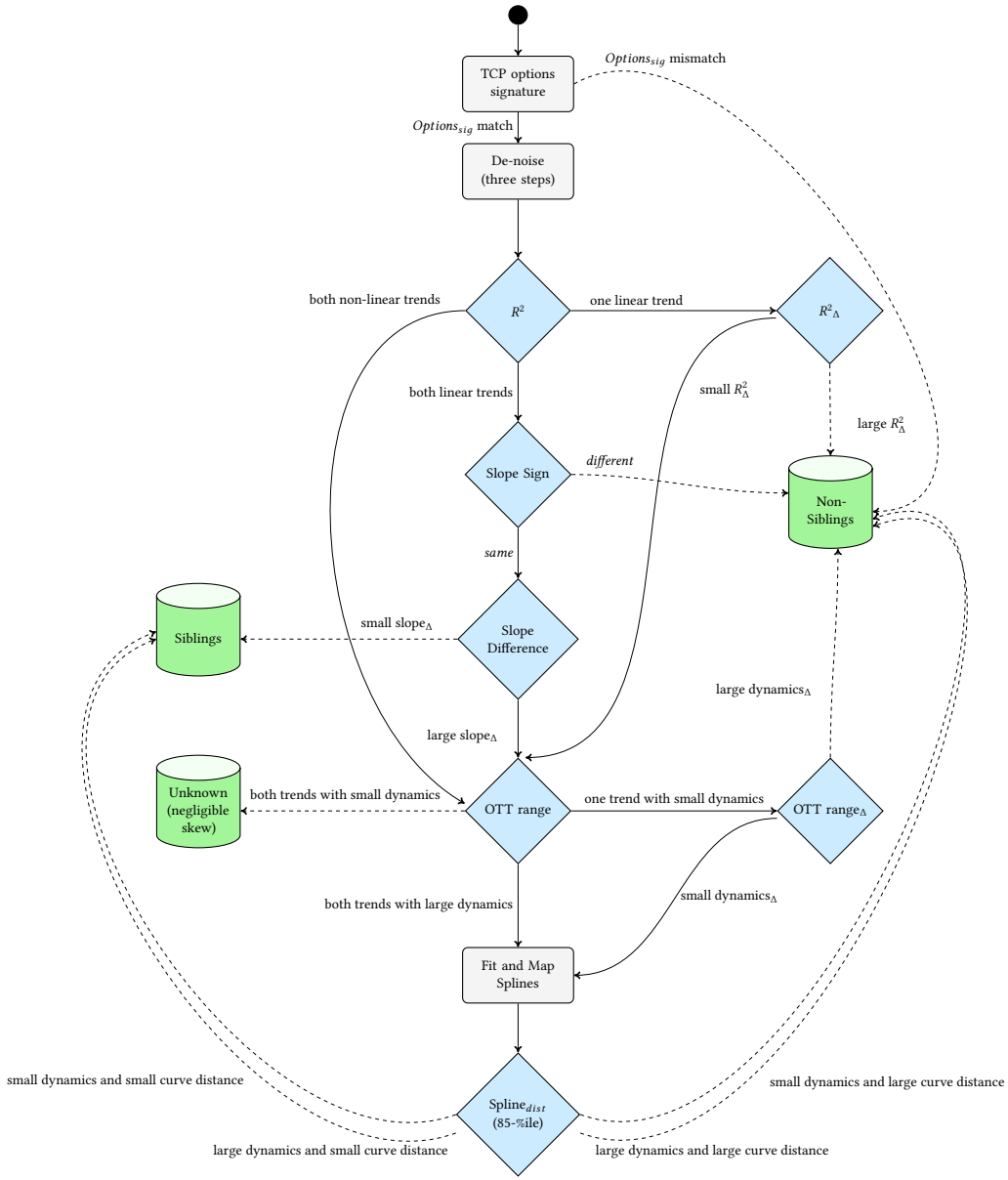
Reset and Adjustment





- Similar skew pattern is observed over different probes
- Metric: Polynomial splines

The Decision Algorithm



Large-scale Measurements

- 6.6 M domains from Alexa top 1 M, biz, com, ...
- 371 k unique sibling candidates
  - $\rightarrow$ *m:n* relationship between domain and IP addresses
  - $\rightarrow$ IP address pairs are frequently shared between several domains ($\approx$ 33%)
- 22% confirmed siblings, 76% non-siblings and 2% unknown
  - $\rightarrow$ low false positive rate
  - $\rightarrow$ web hosters, CDNs, load balancers ...

Q & A

Thanks for your attention!

Bibliography

[1] Beverly, Robert and Berger, Arthur.
Server Siblings: Identifying Shared IPv4/IPv6 Infrastructure via Active Fingerprinting.
In *International Conference on Passive and Active Network Measurement*, pages 149–161. Springer, 2015.

[2] Beverly, Robert and Campbell, Larry and Berger, Arthur and Weaver, Nicholas.
Inferring Internet Server IPv4 and IPv6 Address Relationships.
Technical report, Monterey, California: Naval Postgraduate School, 2013.

[3] Kohno, Tadayoshi and Broido, Andre and Claffy, Kimberly C.
Remote Physical Device Fingerprinting.
IEEE Transactions on Dependable and Secure Computing, 2(2):93–108, 2005.