

Goal

We estimate application-level Web QoE metrics (L7) directly from the raw low-level stream of encrypted packets (L3), using classic supervised learning techniques.

Motivation

- ISPs must be able to understand the web browsing experience they are delivering to their customers.
- Many objective, in-browser, page visual rendering quality metrics exist. However, pervasive encryption makes such information completely opaque for ISPs.

Demo overview

Our demo provides:

1. An inspection tool opposing detailed and directly comparable views of the loading timeline, from the application and network perspectives.
2. An approximation of multiple metrics, such as Page Load Time and Speed Index, from encrypted network traffic using supervised learning (such as XGBoost and Convolutional Neural Networks).

Dataset

We use WebPagetest [1] to perform multiple page load experiments and collect simultaneously:

- HTTP Archive files (HAR) from the browser
- Packet traces (PCAP) from the network

In total, we collect data for 10,000+ page load experiments:

- Non-landing pages from the Alexa top-500
- 20+ runs including:
 - Different sub-pages
 - Different network conditions

Main L7 Indicators

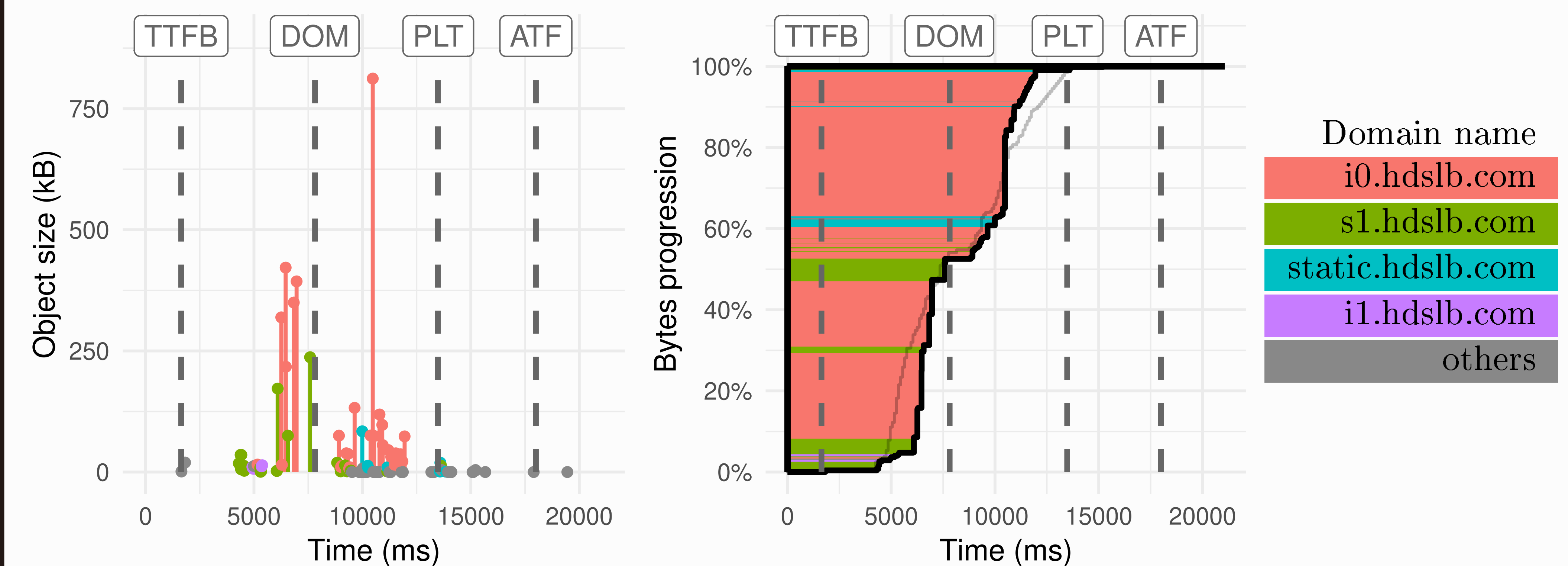
- TTFB (Time To First Byte) [2]: Time when the first byte of payload is received
- DOM (Document Object Model time): Time when DOM has been loaded
- SI (Speed Index): Time defined as the integral of complementary visual progress of the page
- BI (Byte Index): Time defined as the integral of complementary bytes progress of the page
- ATF (Above The Fold time): Time when all above-the-fold contents have been rendered
- PLT (Page Load Time): Time when all (static) page contents have been loaded

References

- [1] <https://www.webpagetest.org/>.
- [2] E. Bocchi, L. De Cicco, and D. Rossi. Measuring the quality of experience of web users. *ACM SIGCOMM Computer Communication Review*, 46(4):8–13, 2016.

Inspection tool

- **Application view** (from HAR):



- **Network view** (from PCAP):

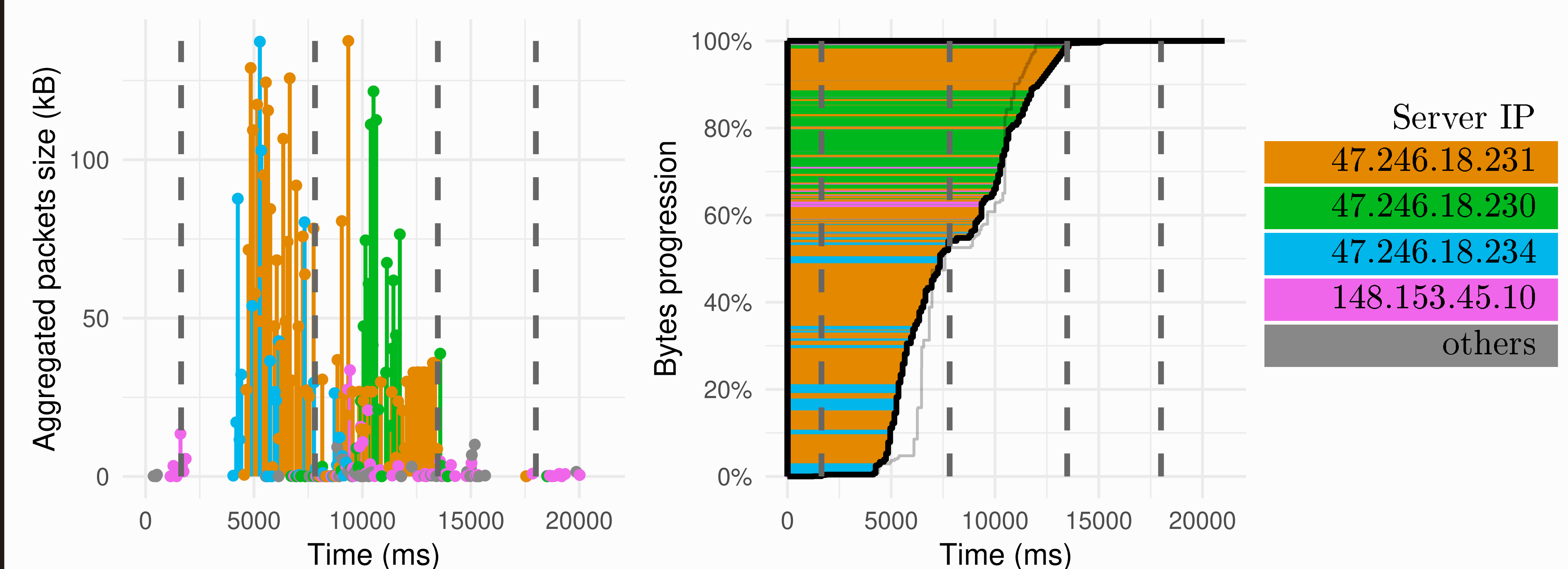


Figure 1: Simultaneous loading views at L7 and L3 levels for one example website

Approximation of L7 metrics

For each metric, we train a single regression model for all pages. We use **time series of L3 bytes progression** and **area above the L3 bytes progression curve** as inputs.

- **Results at a glance**

AppQoS	NetQoS	Absolute error	Relative error
BI	net-BI	Q1: 44ms · Median: 139ms · Q3: 321ms	Q1: 1% · Median: 3% · Q3: 7%
SI	net-SI	Q1: 122ms · Median: 389ms · Q3: 1059ms	Q1: 5% · Median: 13% · Q3: 27%
PLT	net-PLT	Q1: 163ms · Median: 585ms · Q3: 1778ms	Q1: 3% · Median: 10% · Q3: 24%
ATF	net-ATF	Q1: 254ms · Median: 937ms · Q3: 2513ms	Q1: 6% · Median: 16% · Q3: 34%

Table 1: Summary of results for approximating L7 metrics from network data

- **Scatter plots comparing L7 metric with predicted metric from network data**

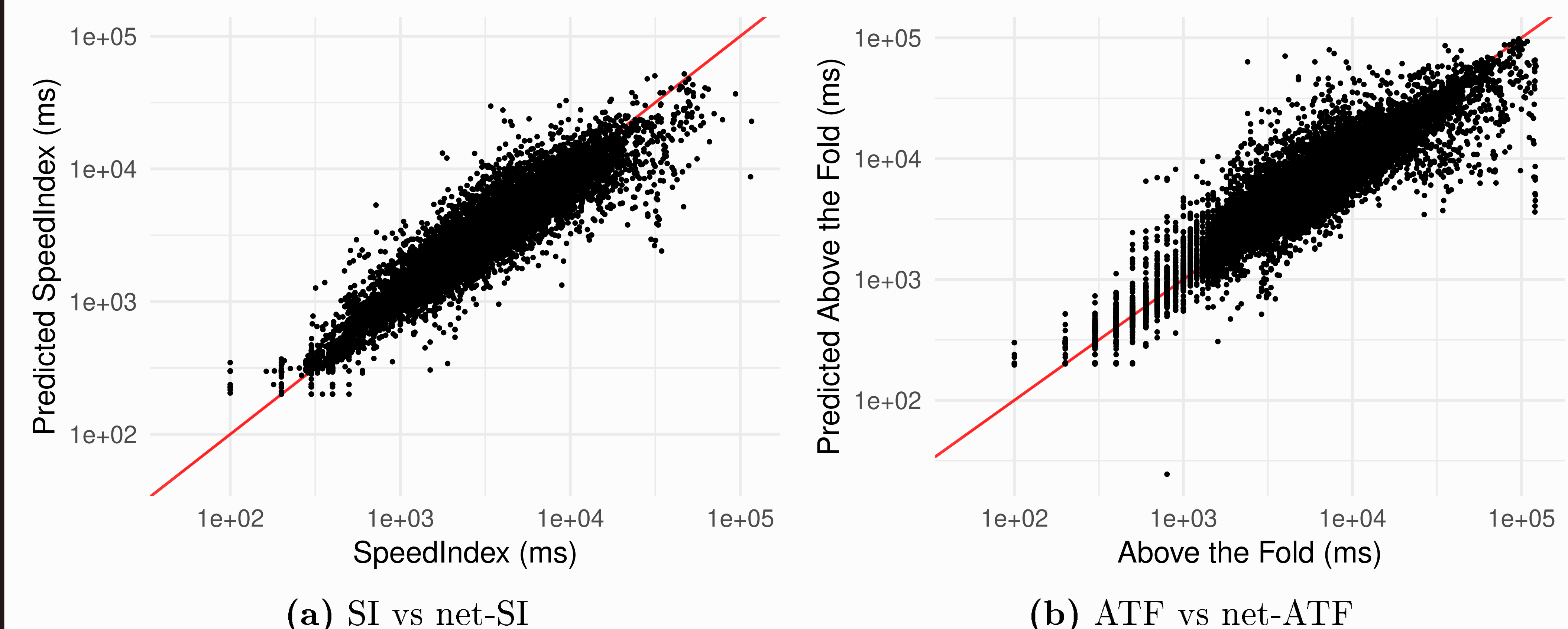


Figure 2: Summary of results for approximating SI and ATF from network data