Chair of Connected Mobility
Department of Computer Engineering
TUM School of Computation, Information and Technology
Technische Universität München

BT GR IDP MT

Accurately Replaying Webpages and Websites

Motivation

Researchers commonly use tools like wget or mitmproxy [1,2] to replay webpages to be able to simulate server-side modifications. For example, to experiment with different server-side transport protocol mechanisms.

The goal of this thesis is to extend an existing file-based webpage replay tool to be more accurate. The main idea is to use graphs to represent various aspects of a page load.

Your Task

- Get familiar with the webpage replay toolchain.
- Extend the toolchain to more accurately reflect real-world webpages using graph representations.
- Quantify the improvements through web performance measurements.

Requirements

- Linux plumbing skills: tcpdump, Wireshark, iproute2, bash scripting
- Familiarity with browser concepts like selenium, dev tools, HAR export, and more
- Good knowledge of computer networking concepts
- Data analysis, e.g., with python (sqlite, matplotlib, pandas)

References

[1] Ravi Netravali, and others. 2015. Mahimahi: accurate record-and-replay for HTTP. In Proceedings of the 2015 USENIX Conference on Usenix Annual Technical Conference (USENIX ATC '15). USENIX Association, USA, 417–429.

[2] Xumiao Zhang, and others. 2024. QUIC is not Quick Enough over Fast Internet. In Proceedings of the ACM Web Conference 2024 (WWW '24). ACM, USA, 2713–2722.

Contact

fries@in.tum.de