iCloud Private Relay on Mobile Networks: Performance Across Countries and Roaming

Damiano Ravalico*, Martino Trevisan*, Idilio Drago†, Paul Schmitt‡, Francesco Bronzino§¶
*University of Trieste, †University of Turin, ‡Cal Poly,
§ENS de Lyon, CNRS, UCBL1, LIP ¶Institut universitaire de France

***Abstract*—The growing adoption of privacy-oriented Internet services is reshaping how user traffic is handled and routed across the network. Among these services, Apple's iCloud Private Relay represents a large-scale deployment of a multi-party relay architecture, designed to improve user privacy while limiting the performance costs typically associated with additional network hops. Introduced in 2021 and integrated into Apple operating systems, the service provides users with an easy-to-use privacy-enhancing mechanism at a low cost, making its impact on Internet performance an important subject of study. In this work, we investigate the performance implications of iCloud Private Relay on mobile networks across different countries and usage scenarios. Through measurements performed on iPhones connected via cellular networks, we show that iCloud Private Relay reduces speed test throughput and increases latency, while its impact on web page load time is comparatively limited. We further analyze international roaming scenarios, where the interaction between home routing and the Private Relay overlay introduces additional complexity and can occasionally improve performance. Our results contribute to the understanding of large-scale, multi-hop privacy-enhancing architectures in mobile environments. To facilitate further research, we release our measurement dataset and software tools to the community.**

***Index Terms*—iCloud Private Relay, privacy-enhancing technologies, network measurements, quality of experience**

I. Introduction

User privacy on the Internet has become a central concern in the networking community [1]. The development of new Internet protocols and services is increasingly driven by the need for stronger privacy guarantees, creating new opportunities for privacy-enhancing technologies. Among these solutions, iCloud Private Relay (PR) is an Apple-developed service integrated into the company's operating systems (i.e., macOS, iOS, and iPadOS). Initially launched in 2021, it allows users to route their traffic through a multi-party relay [2], offering privacy guarantees that differ substantially from those of classical VPNs.

PR relies on a two-hop proxy architecture: the client connects to an ingress proxy operated by Apple over an encrypted QUIC connection [3], which in turn forwards traffic to an egress proxy run by one of Apple's infrastructure partners (Akamai, Cloudflare, or Fastly), which finally reaches the destination server. Each proxy sees only one side of the communication: the ingress proxy knows the client's IP address but not the destination, while the egress proxy knows the destination but not the client. Passive observers such as ISPs see only aggregated flows toward the proxies, making individual user profiling significantly harder [4]. Unlike a VPN, no single entity holds both user identity and browsing behavior simultaneously.

The service's tight integration into Apple's ecosystem, combined with its low cost ($0.99/month for the cheapest plan) and negligible setup effort, makes large-scale adoption plausible, with potentially significant consequences for Internet traffic patterns [5]. Conversations with a European ISP confirm that PR already warrants attention: $0.25\,\%$ of that ISP's total traffic volume is attributable to PR, and roughly $10\,\%$ of household subscriptions exchange more than $10\,\mathrm{MB}$ per day with PR ingress servers. Apple has not disclosed adoption figures as of 2024.

These trends raise questions about the performance implications of PR at scale. The performance and privacy implications of systems such as VPNs, onion routing [6], and Tor [7] have been the subject of extensive prior research [8], [9]. For instance, website fingerprinting attacks on Tor traffic have been demonstrated using packet size and timing side channels [10], and several works have quantified the throughput and latency penalties imposed by such systems [11], [9]. For PR specifically, prior work has addressed deployment infrastructure [5] and security properties [12], leaving performance and user-perceived quality largely unexplored.

In this work, we study the impact of PR on browsing experience. We perform an experimental campaign on Apple mobile devices to characterize the impact of PR across a range of scenarios. We deploy our testbed across multiple locations and take the perspective of mobile users on cellular networks. We gather metrics associated with users' Quality of Experience (QoE), such as page load time, throughput, and latency. Apple acknowledges [13] that PR may reduce speed test performance, since such tests typically rely on "several simultaneous connections to deliver the highest possible result", but also states that "actual browsing experience remains fast". We evaluate these claims through a set of experiments covering both throughput-oriented speed tests and real-world web browsing scenarios, with and without PR.

Our results show that PR significantly reduces speed test throughput and increases latency across all tested countries and operators. The impact on web page load time is comparatively small. We also investigate international roaming scenarios, where the interplay between home routing and the PR overlay introduces additional complexity in the observed performance.

PR performance has been studied uniquely from the desktop/PC user point of view in our previous work [14], but

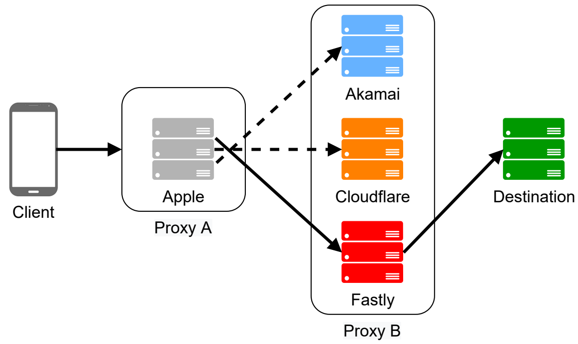

Fig. 1: iCloud Private Relay architecture. Traffic is relayed through two proxies: an Apple-operated ingress proxy and an egress proxy managed by one of Apple's partners (Akamai, Cloudflare, or Fastly).

it was conducted while the service was still in its public beta phase and did not specifically investigate the complexity introduced by mobile networks. In this work, we focus exclusively on mobile users and evaluate PR over cellular connectivity, considering aspects such as operator-dependent routing, cellular network variability, and international roaming scenarios where home routing interacts with the PR overlay. Our measurements, collected between May and June 2024, provide an updated view of PR after its transition to full deployment. Interestingly, while throughput degradation is consistent across both desktop and mobile scenarios [14], the impact on page load time is more pronounced on desktop, where increases of up to 60 % were observed, compared to the comparatively limited effects we find on mobile.

Overall, our study provides insights into the performance implications of PR on mobile networks, contributing to the broader understanding of large-scale multi-hop privacy-enhancing architectures. To facilitate future research and enable reproducibility, we make our measurements and source code publicly available.

II. Background and Related Work

A. iCloud Private Relay Architecture

Apple introduced iCloud Private Relay during the 2021 Worldwide Developers Conference (WWDC) [3]. The service is based on a Multi-Party Relay (MPR) design [2], whose objective is to separate a user's network identity from the destinations they access online. In practice, the system prevents any single entity from simultaneously observing both the client IP address and the accessed content. To achieve this goal, PR establishes two nested communication channels. Traffic is first sent to an ingress relay (Proxy A in Figure 1) managed by Apple, which is responsible for user authentication and coarse-grained localization. The traffic is then sent through a second tunnel to an egress relay (Proxy B in Figure 1) managed by one of Apple's infrastructure partners, currently Akamai, Cloudflare, or Fastly. The egress relay ultimately communicates with the destination servers on behalf of the user. As a consequence, Proxy A can identify the client but cannot inspect the web traffic carried within the tunnel, whereas Proxy B can observe the destination endpoints without learning the user's IP address. Likewise, destination servers only see the egress relay address. At present, PR support is restricted to Apple applications, most notably Safari.

Rather than introducing proprietary transport mechanisms, PR leverages existing Internet protocols. Connections between clients and Proxy A rely primarily on QUIC, with HTTP/2 over TLS used as a fallback when QUIC connectivity is unavailable. Communication between Proxy A and Proxy B is typically implemented through HTTP/3 and MASQUE [15], enabling efficient proxying of QUIC traffic across QUIC-based tunnels. When HTTP/3 support is unavailable, PR reverts to traditional HTTP CONNECT over TLS. A similar architectural approach has been adopted by Chrome IP Protection [16], Google's emerging privacy-oriented browsing solution.

Unlike conventional VPN services, PR only processes a subset of device traffic. Specifically, only Safari HTTP(S) traffic is routed through the relay infrastructure, whereas other traffic categories continue to follow standard network paths. During our observations, WebRTC communications, including RTP, DTLS, and STUN/TURN exchanges associated with audio and video calls, bypassed PR entirely. Traffic generated by third-party browsers, email applications, and other software is similarly unaffected. This partial coverage may create a mismatch between users' sense of privacy and the actual privacy guarantees offered by the service.

The deployment of architectures such as PR has implications for a broad range of Internet services. For example, services that use client IP addresses for geolocation can no longer directly identify users' locations and must rely on alternative approaches. Similarly, network middleboxes performing functions such as content filtering, enterprise security enforcement, or parental control lose access to information traditionally used for traffic inspection. As investigated throughout this paper, these additional forwarding stages, added by the relay-based architecture, may increase communication delays and potentially influence user experience, particularly considering that latency increments of only a few hundred milliseconds can already be noticeable to users and alter their behavior [17].

B. Related Work

Among existing privacy-preserving communication systems, onion routing [6] and Tor [7] are arguably the closest counterparts to PR. These systems seek to conceal communication endpoints by forwarding traffic through a sequence of intermediate relays organized into circuits. Despite pursuing similar privacy objectives, PR differs in several respects. It employs only two relay hops rather than Tor's standard three-hop circuits, relies on large-scale commercial infrastructure instead of volunteer-operated nodes, and uses established web protocols such as HTTP and QUIC rather than custom transport mechanisms. These design choices may also make PR less susceptible to blocking and censorship.

A substantial body of literature has examined the security properties of privacy-enhancing systems. For example, website fingerprinting attacks against Tor traffic have been demonstrated using observable characteristics of encrypted communications [18], [10]. From a performance perspective, previous work has evaluated both VPN services and Tor [11], [9], reporting considerable throughput reductions and latency increases in a variety of deployment scenarios. Our work follows a similar line of investigation, focusing specifically on the performance implications of PR.

Existing research on PR has primarily concentrated on infrastructure and security aspects. [5] characterized the deployment of the service, examining its network footprint and identifying the geographical and topological locations of ingress and egress relays. More recently, [19] studied performance degradation associated with multi-hop MASQUE tunneling, a mechanism heavily used by PR. They formalize and experimentally evaluate these effects, proposing mitigation strategies. Our previous work [14] has explored PR performance from the desktop user point of view.

Security-oriented analyses have also been conducted. [12] assessed the privacy risks associated with PR's two-hop architecture by quantifying the fraction of traffic observable by individual Autonomous Systems. Their results indicate that a limited number of networks can observe a substantial share of PR traffic. The same study further evaluated modern website fingerprinting techniques against PR-protected communications and concluded that the service remains vulnerable to such attacks. In contrast, our contribution focuses on empirical performance evaluation. Rather than analyzing deployment characteristics or privacy threats, we investigate how PR affects network performance and user-perceived quality of experience in practical mobile browsing scenarios.

III. Testbed and Dataset

We evaluate PR performance on mobile networks using three iPhone 13 devices deployed in three locations, including international roaming scenarios. The measurement campaign is conducted between May and June 2024. Two devices are located in European cities, namely Lyon (France) and Trieste (Italy), while the third device was deployed in Hawaii (USA). Although these measurements provide insights into the performance impact of PR, our locations do not represent the diversity of the global Internet. Extending the study to additional locations, for example, through a larger distributed measurement infrastructure, remains an interesting direction for future work. Such a campaign is challenging to perform since PR is a proprietary service available only on Apple platforms and cannot be easily deployed on generic measurement probes.

Each iPhone is connected via USB to a MacBook Pro PC, which enables it to run measurements using the Appium toolset for UI automation. We created custom Python scripts that leverage the Appium library to automate various tasks on the iPhone, such as visiting a website or enabling/disabling PR. We note that, although flexible and full-featured, the Appium framework presents some non-deterministic behavior, thus challenging the design of automatic experimental campaigns. For instance, the navigation through the "Settings" application to toggle PR is prone to fail, requiring strict checks and retries. We configure PR using its default setting, which aims to preserve the user's approximate location.

1) Throughput Measurements: Active throughput measurements are widely used to estimate the capacity of the bottleneck link between an end device and a measurement server deployed in the network. Speed test services typically locate their measurement servers near end users, under the assumption that performance is mainly constrained by the last-mile access network. Although such measurements do not always accurately reflect users' real-world experience [20], they are useful to identify performance changes introduced by: (i) additional network elements traversed by traffic, such as PR proxies, and (ii) changes in the network path between the user and the speed test infrastructure.

We perform active throughput measurements through Ookla's Speed Test web service [21], a widely adopted platform for evaluating Internet access performance. Note that PR handles only the traffic generated by the Safari browser; thus, we could not use the iPhone Ookla native application, which is the recommended option.[1] Instead, we create an Appium script to navigate to the Ookla website and perform an in-browser speed test. This comes with a limitation, as this browser speed test only measures *download* throughput and *latency*, while it does not perform any *upload* throughput measurement. We run 10 speed test measurements from the three locations, but without considering the roaming scenario.[2]

2) Web Measurements: iCloud PR primarily targets web browsing scenarios by providing enhanced privacy protections. Our objective is to quantify how PR affects users' perceived performance and, consequently, web Quality of Experience (QoE). To this end, we automate Safari-based browsing sessions and collect page loading metrics from a set of popular websites. We select the top 100 websites in each country according to the ranking provided by SimilarWeb analytics [22].

Website visits are automated through the Appium webdriver framework, which allows us to collect information about the HTTP transactions generated during page loading. As our main performance indicator, we consider Page Load Time (also referred to as `onLoad` time), which measures the elapsed time between the start of a navigation and the retrieval of the final page object. Previous studies have shown that Page Load Time is correlated with users' perceived QoE [23]. On the iPhone, Safari opens with the default profile, meaning that the browser HTTP cache is already full with objects fetched on past visits, potentially affecting our measurements. To obtain the desired behavior, we implemented a custom Appium script that visits

[1] PR tunnels only Safari traffic, so background iOS traffic could contend for bandwidth. We closed all applications during the campaigns, and any residual traffic affects PR and native runs alike, since they are interleaved in time.

[2] We could not run more speed test experiments due to the data quotas of each SIM card.

TABLE I: Median mobile download throughput and latency across countries and Italian operators. For Italy, mobile data values are aggregated across operators.

Country/Operator	Mobile Throughput (Mbit/s)		Mobile Latency (ms)	
	Native	PR	Native	PR
FR	559.78	392.62	21.0	41.0
US	71.07	17.86	65.5	83.0
IT	36.45	24.12	37.5	66.0
Iliad	58.23	37.65	80.0	80.5
Tim	20.63	18.46	35.0	76.0
Vodafone	83.77	70.91	46.0	64.0
Wind Tre	11.40	11.00	31.0	59.0

the desired webpage using a private browsing tab, so that the browser's HTTP cache is not used. Moreover, to ensure all visits are independent, we restart the Safari browser at each visit, so that the outcomes of the navigation with the private tab (content of the HTTP cache, connection pool, DNS cache, etc.) are erased.

A. Measurement Campaigns

Using the three test iPhones, we run a measurement campaign for each country. In Italy, we run the same measurements with four SIM cards from the four most popular Italian operators, namely TIM, Vodafone, Wind Tre, and Iliad. In the other countries, we use a single operator, Orange and Verizon for France and the U.S., respectively. For each scenario, we visit the top-100 websites for the given country, repeating each visit 5 times with and without PR. We run measurements in Italy, France, and the U.S., while we also experiment with international roaming scenarios: (i) a French SIM visiting Italy and the U.S. and (ii) an Italian SIM visiting Germany.

Finally, we note that our measurement methodology does not directly capture the end-to-end Round Trip Time (RTT). Since our experiments rely on browser-level instrumentation, layer-4 RTT measurements are not available, and estimating RTT in tunneled environments such as PR is non-trivial. We therefore focus on user-perceived performance metrics, including Page Load Time and throughput, leaving a detailed RTT analysis for future work.

IV. User Experience

In this section, we analyze the quality of experience for iOS users. Using three iPhones on different mobile networks, we run measurements across three countries. Our measurements include speed tests as well as web page visits, and we also experiment with international roaming scenarios.

A. Throughput and Latency Measurements

We first discuss the results of the speed tests, showing the obtained values in Table I. The upper part of the table presents a comparison of mobile download throughput and latency in France (FR), the United States (US), and Italy (IT) on iPhones with and without PR; each cell reports the median value over 10 experiments. The bottom part of the table breaks down Italian results per operator. In all cases, enabling PR results in a significant reduction in download throughput and an increase in latency.

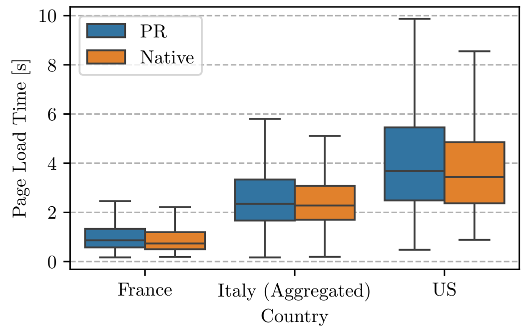

Fig. 2: Page Load Time with and without PR for mobile devices, separately by country (international roaming excluded).

In France, the download throughput decreases from 559.78 Mbit/s to 392.62 Mbit/s with PR, while latency increases from 21.0 ms to 41.0 ms. The United States experiences the largest relative decrease in throughput, from 71.07 Mbit/s (native) to 17.86 Mbit/s (PR), and a latency increase from 65.5 ms to 83.0 ms. In Italy, the native download speed is 36.45 Mbit/s, which drops to 24.12 Mbit/s with PR, and latency rises from 37.5 ms to 66.0 ms.

For individual Italian operators, Vodafone shows the highest throughput, which drops from 83.77 Mbit/s to 70.91 Mbit/s with PR, while Wind Tre has the lowest throughput but also the lowest impact when enabling PR.

We break down the speed test by proxy B provider. In the U.S. and France, we find that in almost all cases, proxy B is hosted by Cloudflare. Conversely, in Italy, we find that 65 % of speed tests are run with Cloudflare proxy B, while the remainder use Akamai. Regarding performance, Akamai, in median, leads to a +36 % increase in download throughput.

These results confirm and quantify that PR has a significant impact on speed test performance. In the mobile case, results are consistent across all tested countries and telecom operators. Next, we check whether these impacts on throughput measurements are also observed while browsing the web.

B. Web Browsing

We now discuss the web browsing metrics, analyzing results by country and operator, and presenting results for international roaming.

Figure 2 compares page load times across the three countries when using PR versus native connections on iPhone. The data is presented using box plots: the central black line represents the median, while the boxes mark interquartile ranges. Results for PR are shown in blue boxes, while results for the native connection are shown in orange boxes.

In France, both PR and native connections show similar median load times and relatively narrow interquartile ranges, suggesting consistent performance regardless of PR usage. Italy, where we measure across three operators, shows slightly higher median load times for both connection types compared

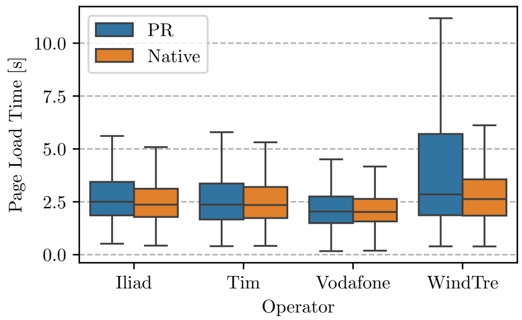

Fig. 3: Page Load Time with and without PR separately by Italian operator.

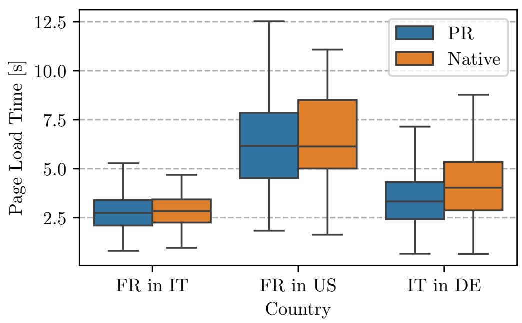

Fig. 4: Page Load Time with and without PR for mobile devices in international roaming scenarios.

to France, with PR resulting in a marginally higher median and greater variability than native connections. The US presents the clearest differences, with notably higher median load times and increased variability for both connection types compared to the European countries. Moreover, PR shows a higher median load time and a wider interquartile range than native connections in the US.

These measurements confirm that the performance implications of PR vary for mobile devices across geographical locations. PR increases the median page load time, but differences are relatively small in these mobile browsing experiments.

We now compare the performance across Italian mobile operators in Figure 3. The plot shows that the use of PR increases page load times across all operators—note the generally higher positioning of the blue boxes compared to the orange ones. This effect is clearer for WindTre, where the PR median and upper quartile are much higher than the native connection. Interestingly, WindTre also displays the widest range of page load times for native connections, indicating a poor connection with this operator during the tests. Iliad and TIM show similar patterns, with PR slightly increasing page load times. Vodafone shows the smallest difference between PR and native connections. Overall, these results again point to a degradation of web browsing performance when using PR on mobile devices, and this degradation is observed across multiple telecom operators.

Finally, we examine three international roaming scenarios, where experiments were carried out in a country using a SIM card from a foreign operator. International roaming always involves home routing, i.e., the traffic is routed through the SIM home country before reaching the final destination [24]. We thus measure the implications for PR performance and its two levels of tunneling in this case.

Figure 4 reports results, which include a French SIM roaming in Italy (*FR in IT*), a French SIM roaming in the U.S. (*FR in US*), and an Italian SIM roaming in Germany (*IT in DE*).

Overall, we observe much more complex scenarios and counterintuitive improvements in performance when using PR. The *FR in US* scenario shows the worst performance for both PR and native connections: this is expected due to transatlantic roaming. The native connection appears to have slightly higher median and upper quartile values than PR. The *FR in IT* scenario demonstrates the lowest overall load times and the least variation, indicating that roaming within Europe has little impact on performance. PR and native connections show very similar distributions, with PR having a slightly lower median page load time again. The *IT in DE* scenario shows moderate performance differences, and again, the native connection appears to have a much higher median and upper quartile than PR. Results are thus consistent in suggesting that PR offers some performance benefit in roaming situations. We were not able to identify the root causes of this improvement. The interactions between PR, home routing, and proxy node selection indeed create a very complex scenario, calling for more ingenious measurement setups.

Overall, web browsing results show that PR does have an impact on page load times. It generally leads to slower median page load times, except in the roaming case. Differences are, however, small in general and much less prominent than in throughput measurements.

V. Open Questions

We discuss limitations and directions for future work. Extending our findings will require additional measurements across diverse locations and methodologies. We list directions in the following and recall that our measurement tools are publicly available to support further validation and extensions.

A. Overriding Routing

Our measurements show that PR can occasionally outperform native connectivity, likely because its overlay path avoids inefficient or congested routes used by the default network path.

Unlike traditional Internet routing, PR concentrates traffic through a limited set of networks involving the user's ISP, Apple, and the selected Proxy B provider. As adoption increases, ISPs may observe a growing fraction of traffic directed toward a small number of relay infrastructures.

We also observed substantial performance variability over time, both with and without PR. Factors such as cellular network load and international roaming further complicate performance evaluation in mobile environments.

Overall, these findings indicate that PR-like architectures can significantly influence path selection and routing behavior, motivating further investigation of their impact on network performance.

B. Localization

Since PR hides users' IP addresses from destination servers, it limits the effectiveness of traditional IP geolocation services. Many online services rely on geolocation to provide localized content or enforce region-based access policies.

To mitigate this issue, PR attempts to preserve coarse location information through the selection of the egress relay. Apple also publishes mappings between Proxy B IP addresses and the approximate locations of the users they represent [25]. While this approach is generally sufficient for location-based content customization, applications requiring precise geographic information, such as certain live-streaming services, may experience limitations.

Previous work [5] further showed that the advertised locations of PR egress addresses do not always correspond to the physical locations of the underlying infrastructure. This strategy helps extend coverage to regions without nearby PR deployments but may introduce additional performance implications. Understanding the tradeoff between privacy, localization accuracy, and service usability remains an open research challenge.

VI. Conclusions

iCloud Private Relay represents one of the first large-scale deployments of Multi-Party Relay architectures aimed at improving user privacy. Given the widespread adoption of Apple devices and the increasing attention toward privacy-preserving technologies, such architectures may play an important role in the future evolution of Internet traffic.

In this work, we performed an empirical evaluation of PR performance on mobile networks across three locations in France, Italy, and the US. Our results show that PR consistently reduces speed test throughput and increases latency across the tested scenarios, while its impact on web page loading performance is limited. We also analyzed international roaming cases, where the interaction between home routing and the PR overlay creates complex behaviors, including scenarios in which PR improves browsing performance.

Our study highlights several open research questions regarding large-scale Multi-Party Relay deployments, extending beyond performance aspects to include routing implications, privacy-performance tradeoffs, and the effects on Internet services. To support further investigation and reproducibility, we release the measurement data and the software tools used in our experiments.

References

[1] S. Farrell and H. Tschofenig, "Pervasive Monitoring Is an Attack," RFC 7258, May 2014. [Online]. Available: https://www.rfc-editor.org/info/rfc7258

[2] P. Schmitt, J. Iyengar, C. Wood, and B. Raghavan, "The decoupling principle: A practical privacy framework," in *ACM SIGCOMM Workshop on Hot Topics in Networking (HotNets)*, Nov. 2022.

[3] Apple, "iCloud Private Relay Overview," https://www.apple.com/privacy/docs/iCloud_Private_Relay_Overview_Dec2021.PDF, Dec. 2021.

[4] M. Trevisan, F. Soro, M. Mellia, I. Drago, and R. Morla, "Does domain name encryption increase users' privacy?" *SIGCOMM Comput. Commun. Rev.*, vol. 50, no. 3, p. 16–22, jul 2020.

[5] P. Sattler, J. Aulbach, J. Zirngibl, and G. Carle, "Towards a tectonic traffic shift? Investigating Apple's new relay network," in *Proceedings of the 22nd ACM Internet Measurement Conference*, ser. IMC '22, 2022.

[6] P. F. Syverson, D. M. Goldschlag, and M. G. Reed, "Anonymous connections and onion routing," in *Proceedings. 1997 IEEE Symposium on Security and Privacy (Cat. No. 97CB36097)*. IEEE, 1997, pp. 44–54.

[7] R. Dingledine, N. Mathewson, and P. Syverson, "Tor: the second-generation onion router', usenix security symposium," 2004.

[8] A. Mani, T. Wilson-Brown, R. Jansen, A. Johnson, and M. Sherr, "Understanding tor usage with privacy-preserving measurement," in *Proceedings of the Internet Measurement Conference 2018*, ser. IMC '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 175–187.

[9] M. Alsabah and I. Goldberg, "Performance and security improvements for tor: A survey," *ACM Comput. Surv.*, vol. 49, no. 2, sep 2016.

[10] D. Arp, F. Yamaguchi, and K. Rieck, "Torben: A Practical Side-Channel Attack for Deanonymizing Tor Communication," ser. Proc. of the ASIA CCS, 2015, pp. 597–602.

[11] M. Pudelko, P. Emmerich, S. Gallenmüller, and G. Carle, "Performance analysis of vpn gateways," in *2020 IFIP Networking Conference (Networking)*, 2020, pp. 325–333.

[12] A. Zohaib, J. Sheffey, and A. Houmansadr, "Investigating traffic analysis attacks on apple icloud private relay," in *Proceedings of the 2023 ACM Asia Conference on Computer and Communications Security*, 2023, pp. 773–784.

[13] Apple, "About iCloud Private Relay," https://support.apple.com/en-us/HT212614, Dec. 2021.

[14] M. Trevisan, I. Drago, P. Schmitt, and F. Bronzino, "Measuring the performance of icloud private relay," in *International Conference on Passive and Active Network Measurement*. Springer, 2023, pp. 3–17.

[15] T. Pauly, D. Schinazi, A. Chernyakhovsky, M. Kühlewind, and M. Westerlund, "Proxying IP in HTTP," RFC 9484, Oct. 2023. [Online]. Available: https://www.rfc-editor.org/info/rfc9484

[16] Google, "Chrome IP Protection," https://github.com/GoogleChrome/ip-protection, 2024.

[17] X. Bai, I. Arapakis, B. B. Cambazoglu, and A. Freire, "Understanding and leveraging the impact of response latency on user behaviour in web search," *ACM Transactions on Information Systems (TOIS)*, vol. 36, no. 2, pp. 1–42, 2017.

[18] P. Sirinam, M. Imani, M. Juarez, and M. Wright, "Deep Fingerprinting: Undermining Website Fingerprinting Defenses with Deep Learning," ser. Proc. of the CCS, 2018, pp. 1928–1943.

[19] K. Elmenhorst, M. Kühlewind, I. Kunze, C. Sander, and K. Wehrle, "Cascades of nested acknowledgments in multi-hop masque," in *Proceedings of the 2025 Applied Networking Research Workshop*, 2025, pp. 85–91.

[20] M. Trevisan, I. Drago, and M. Mellia, "Impact of access speed on adaptive video streaming quality: A passive perspective," in *Proceedings of the 2016 Workshop on QoE-Based Analysis and Management of Data Communication Networks*, ser. Internet-QoE '16, New York, NY, USA, 2016, p. 7–12.

[21] Ookla, "Speedtest," http://speedtest.net, Dec. 2021.

[22] SimilarWeb, "Effortlessly analyze your competitive landscape," https://www.similarweb.com/, Dec. 2021.

[23] D. N. da Hora, A. S. Asrese, V. Christophides, R. Teixeira, and D. Rossi, "Narrowing the gap between qos metrics and web qoe using above-the-fold metrics," in *International Conference on Passive and Active Network Measurement*. Springer, 2018, pp. 31–43.

[24] A. M. Mandalari, A. Lutu, A. Custura, A. S. Khatouni, Ö. Alay, M. Bagnulo, V. Bajpai, A. Brunstrom, J. Ott, M. Trevisan *et al.*, "Measuring roaming in europe: Infrastructure and implications on users' qoe," *IEEE Transactions on Mobile Computing*, vol. 21, no. 10, pp. 3687–3699, 2021.

[25] Apple, "Prepare Your Network or Web Server for iCloud Private Relay," https://developer.apple.com/support/prepare-your-network-for-icloud-private-relay/, Dec. 2021.