Limitations of Caching in Mobile Social Media Applications

Stan Kurkovsky
Computer Science
Central Connecticut State University
New Britain, CT, USA
kurkovsky@ccsu.edu

***Abstract*— Mobile devices, such as smartphones and tablets, are often referred to as resource-constrained devices due to the limitations imposed by the hardware, interaction modalities, and the network environment. Expensive capped data plans and unreasonable data roaming charges often impose a constraint on the usage of cellular data for mobile device users. And yet, mobile devices remain the most popular platform for consuming social media content given the high popularity of platforms like TikTok, Instagram, YouTube, and Facebook. This work reviews existing approaches to media caching and focuses on a novel mobile data caching strategy leveraging semantic properties of data on mobile social media platforms.**

Keywords—social networks; mobile applications; semantic caching.

I. Introduction

The importance of caching the data downloaded by mobile web browsers and other applications on mobile devices is rising due to the exponential increase in data traffic that often exceeds the capacities of existing infrastructures of cellular network providers. Much of this traffic is generated by mobile social media applications whose share is rapidly increasing.

The increase in network traffic generated by social media applications in 2024 is driven by the sheer number of users, the time spent on these platforms, and the variety of platforms used. Currently, over 5.17 billion people worldwide use social media, accounting for nearly 64% of the global population [2]. On average, users spend about 2 hours and 20 minutes daily on social media, contributing significantly to overall internet traffic [14]. Additionally, social media users engage with an average of 6.7 different platforms each month, which further amplifies data usage [2,14].

Popular platforms like Facebook (with 3.07 billion monthly active users as of 2024), YouTube (2.5 billion), and TikTok (1.58 billion) continue to expand, with each platform pushing the use of video, live streaming, and other data-heavy content, which adds to the increasing traffic load on networks [9]. In some regions, social media consumption is even higher. For example, people in countries like Brazil and India engage with 8 or more platforms monthly, further driving up usage [2,14]. Moreover, mobile social media usage is on the rise, with 91% of mobile phone users globally also being active on social media platforms, making mobile devices a significant contributor to network traffic. The average user spends 6 hours and 40 minutes online daily, with a significant portion of this time dedicated to social media [14].

Over the past decade, User-Generated Content (UGC) has seen exponential growth, driven largely by the rise of social media platforms and the increasing accessibility of content creation tools, particularly on mobile devices.

In 2024, over 50 million people identify as content creators, contributing significantly to the social media ecosystem [18]. This creator economy includes both casual users and professional influencers, with platforms like TikTok, Instagram, and YouTube leading the surge in content production. A large proportion of social media traffic now stems from these creators who share videos, images, and stories. Short-form video content, particularly through platforms like TikTok and Instagram Reels, has seen explosive growth, with video consumption becoming one of the fastest-growing content types across all generations.

The growth of internet traffic driven by user-generated content has been significant in recent years. Between 2019 and 2022, global mobile broadband traffic more than doubled, from 419 exabytes (EB) to 913 EB [22]. During the same period, fixed broadband traffic increased nearly fivefold, from 1,991 EB to 4,378 EB. UGC, especially video content, has been a key contributor to this surge, with user-generated video content predicted to account for 55% of all internet traffic by 2030 [13].

II. The Need for Mobile Caching

With the growing popularity of media sharing in many online services, including social media platforms, the costs of distributing content increases as well. In order to lower the costs of content distribution, these services could rely on content delivery networks, build proprietary delivery systems, or minimize content traffic by using one or a combination of several caching mechanisms. Caching the data transferred to a mobile device helps directly eliminate much of the data transmission overhead. Recent studies suggest that redundant data transmission contributes as much as 20% of the total mobile internet traffic [1,11], which is generated by mobile web browsers and mobile apps. Improving the implementation of the cache management mechanisms can not only help reduce the volume of the network traffic, but also reduce the consumption of cellular network resources, decrease device energy consumption, and reduce network latency perceived by mobile users. Such improvements would benefit both cellular network providers and their customers.

Generally speaking, a caching mechanism helps reduce the latency of content delivery from the server by bringing the content closer to the clients. High network latency is detrimental to the mobile user experience. It has been shown that the perceived delay in loading web pages on mobile browsers is caused by a long round-trip time (RTT) and a large number of round trips required to load all objects displayed on the page [16,25]. RTT generally depends on the network bandwidth, hop distance, current network conditions, and the size of the item being downloaded. Existing implementations of mobile browsers may not take full advantage of parallel resource loading [6,15,21], which may result in a substantial number of round-trips required do display a single web page. Sub-resources can only be identified during the process of parsing the main HTML document. An HTTP redirect can add an extra RTT. If the main HTML document contains JavaScript code referring to a web object, that object may not be loaded until the code is executed; furthermore, this code may not start running until some or all objects are loaded.

There is a substantial difference in the way how user access social media on mobile and desktop platforms. Most social media apps use web browsers with built-in caches on desktop platforms. On mobile platforms, most social media platforms offer proprietary native apps, which may need to implement their own data store with a dedicated cache management mechanism.

However, HTML5 provides the capability for application caching. It differs from the traditional HTTP caching in two important ways:

- Each HTML page has an associated cache manifest document that specifies the caching properties of all objects embedded in the page.
- Cached objects have no explicit expiration date. Instead, caching is determined in the cache manifest, which is updated each time the corresponding HTML page is fetched from the server.

Although the primary objective of application caching is to enable offline browsing, it also improves browsing speed and increases the client application's resilience to possible server outages.

A typical mobile web cache uses the mechanisms of expiration and revalidation. An expiration timestamp is set for each cacheable data object. If the object is requested again before it expires, a cached version is used. Otherwise, sending a short query to the server originating that data object will revalidate the freshness of the object.

III. Limitations of Existing Caching Strategies

File storage containing the cached data objects, as well as code implementing the logic of the caching mechanism can be placed in one or more locations between the mobile device sending the requests and the server responding with the data. Caching proxies may be located within the network, but this solution incurs the overhead of the last-mile radio transmission between the customer's mobile device and the cellular tower, which is a well-known bottleneck on the resources and performance in cellular networks. At the same time, a network cache can help reduce the data transmission latency and facilitate sharing cached data among multiple users. On the other hand, implementing the caching mechanism and placing the cache on the mobile device eliminates such network-related overhead, but does not help reduce network latency.

In addition to caching, client-side approaches to speed up the loading of content also include prefetching and speculative loading. Pre-fetching [3] works by trying to predict and load in advance the objects that will be needed to display the pages that the user is likely to visit next. Speculative loading [24] takes advantage of concurrent HTTP connections to predict and load sub-resources while loading the main HTML document. Caching optimization aims to improve the performance of the data storage mechanism by using one or more heuristics to better predict whether a previously downloaded item will be requested in the future [27]. A broad range of cloud-based mechanisms allow web browsers to off-load energy consuming or computationally expensive operations to the cloud [23].

In addition to embedding specialized logic into the client, many solutions to speed up content loading require an infrastructure support, which imposes a number of limitations. Solutions that require server-side support likely will not work with legacy web systems. Solutions requiring proxy or server-side capabilities may not scale well with the increased number of clients (consider known failures of Amazon AWS). Additionally, dependence on proxy support leads to violating end-to-end security. On the other hand, client-only solutions do not rely on infrastructure support, which makes them scalable, immediately deployable, and secure.

A substantial amount of research projects propose distributed social networks that lack a centralized repository of all user data and user-generated content [4]. In such networks, users retain full control of their own data, which necessitates specialized services for content and social update dissemination. A number of techniques for caching in such distributed networks have been proposed [12].

Cooperative client caching is a strategy for distributed cache implementation on client devices [17]. Implementation of a cooperative caching strategy in a social networking system would take advantage of the social graph to make informed decisions about distributing cached content. Such a solution would provide a non-negligible, but an acceptable client overhead. Any solution requiring a degree of cooperation among the clients requires that clients act in a non-malicious manner and follow the same established caching protocol.

There are a number of reasons why current implementation of caching on mobile devices still generates about 20% of redundant traffic and leaves much room for improvement [19]. Caching mechanism on device and/or caching support on the server could be implemented in a way that is not fully compliant with the caching protocol specifications; apps may not fully utilize the caching support offered by the corresponding mobile operating systems and their libraries; cache storage may be too small resulting in an early deletion of a data object that may be re-requested at a later moment; non-persistent cache storage that is purged when the device is restarted.

IV. The Leverage of Semantic Proximity

In addition to addressing the factors listed above, cache management can be further improved by taking into account the semantics of data objects being cached. The nature of data being shared and transmitted within the existing social networks offers a rich set of attributes that could help leverage its semantic properties. Semantic caching mechanisms were first implemented in database systems where a well-structured organization of data provides for a relatively straightforward way of organizing objects based on some sort of semantic metric. Published reports indicate that sematic-based cache replacement mechanisms perform better than traditional approaches [8,20].

For example, s semantics-based caching mechanism might leverage the principle of semantic proximity: that during a given span of time, the user will be interested in viewing the documents that are semantically close, e.g. they are on the same subject. Consequently, the cache replacement policy would favor evicting those documents that are least semantically related to the incoming document or a cluster of the recently added cache objects. A semantic distance would need to be computed between every object currently in cache and the incoming document(s). Depending on the needed free space in the cache, one or more objects with the highest semantic distance will be evicted. Despite their obvious differences, the users of most social networks share a common set of properties, which includes their friends, interests, things that they like, group memberships, etc. Each of these properties, individually or in combination, can be used to compute semantic distances in order to determine whether a given object should be evicted from the cache storage. An effective caching strategy could use one of the well-established algorithms, such as timestamp-based Least Recently Used (LRU) and supplement or replace the temporal distance with semantic distance.

V. Summary

This paper reviewed a broad range of existing approaches to caching data on mobile devices and outlined the need for a better approach to caching in mobile social networking applications. Future work will be focused on testing the proposed mobile caching algorithm based on semantic proximity.

References

[1] A. Anand, C. Muthukrishnan, A. Akella, and R. Ramjee. "Redundancy in network traffic: findings and implications." In *Proceedings of the eleventh international joint conference on Measurement and modeling of computer systems*, pp. 37-48. 2009.

[2] Backlinko Team. "Social Media Usage & Growth Statistics." Sep. 06, 2024. https://backlinko.com/social-media-users.

[3] C. Bouras, A. Konidaris, and D. Kostoulas, "Predictive Prefetching on the Web and Its Potential Impact in the Wide Area," *World Wide Web*, vol. 7, pp. 143-179, 2004.

[4] S. Buchegger, D. Schiöberg, L. Vu, and A. Datta, "PeerSoN: P2P social networking: early experiences and insights," *Proc. 2nd ACM EuroSys Workshop on Social Network Systems*. pp. 46-52, 2009.

[5] A. Calsavara, "The least semantically related cache replacement algorithm," *Proc. 2003 IFIP/ACM Conf. Towards a Latin American agenda for network research*, pp. 21-34, 2003.

[6] C. Cascaval, S. Fowler, P. Montesinos-Ortego, W. Piekarski, M. Reshadi, B. Robatmili, M. Weber, and V. Bhavsar, "ZOOMM: a parallel web browser engine for multicore mobile devices," *Proc. 18th ACM SIGPLAN Symp. Principles and practice of parallel programming,* pp. 271-280, 2013.

[7] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018," http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html, February 5, 2014.

[8] S. Dar, M. Franklin, B. Jonsson, D. Srivastava, and M. Tan, "Semantic data caching and replacement," In *VLDB*, vol. 96, pp. 330-341. 1996.

[9] S. Dixon. Social media - statistics & facts. July 16, 2024. https://www.statista.com/topics/1164/social-networks/

[10] M. Duggan, "Photo and video sharing grow online," *Pew Research*, October 28, 2013, http://www.pewinternet.org/2013/10/28/photo-and-video-sharing-grow-online.

[11] A. Gupta, A. Akella, S. Seshan, S. Shenker, and J. Wang, 2007. *Understanding and exploiting network traffic redundancy*. University of Wisconsin-Madison Department of Computer Sciences.

[12] L. Han, M. Punceva, B. Nath, S. Muthukrishnan, and L. Iftode, "SocialCDN: Caching techniques for distributed social networks," *Proc. 2012 IEEE 12th International Conference on Peer-to-Peer Computing,* pp.191,202, 2012.

[13] International Telecommunication Union. *"Internet Traffic."* October 10, 2023. https://www.itu.int/itu-d/reports/statistics/2023/10/10/ff23-internet-traffic/.

[14] S. Kemp. *"Digital 2024: Global Overview Report."* January 31, 2024. https://datareportal.com/reports/digital-2024-global-overview-report

[15] C. Mehteroglu, Y. Durmus, and E. Onur. "Semantic edge caching and prefetching in 5G." In *2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, pp. 692-695. IEEE, 2017.

[16] J. Nejati, and A. Balasubramanian. "An in-depth study of mobile browser performance." In *Proceedings of the 25th International Conference on World Wide Web*, pp. 1305-1315. 2016.

[17] S. Nikolaou, R. van Renesse, N. Schiper, "Cooperative client caching strategies for social and web applications," *Proc. Large-Scale Distributed Systems and Middleware,* 2013.

[18] M. Norman. "*Social Media Benchmarks by Industry 2024: Creator Content.*" March 28, 2024. https://captiv8.io/blog/2024/03/28/social-media-benchmarks-by-industry-2024-creator-content/

[19] F. Qian, K. S. Quah, J. Huang, J. Erman, A. Gerber, Z. Mao, S. Sen, and O. Spatscheck, "Web caching on smartphones: ideal vs. reality," *In Proceedings of the 10th international conference on Mobile systems, applications, and services (MobiSys '12). ACM, New York, NY, USA, 2012,* pp. 127-140. http://doi.acm.org/10.1145/2307636.2307649.

[20] Q. Ren, M. Dunham, and V. Kumar, "Semantic caching and query processing," *IEEE Transactions on Knowledge and Data Engineering,* vol.15, no.1, pp.192,210, Jan.-Feb. 2003

[21] A. Shariff, No. Katuk, and N. Zakaria. "An overview to pre-fetching techniques for content caching of mobile applications." *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)* 9, no. 2-12 (2017): 37-43.

[22] M. Solomons. "*90 UGC statistics: best practices, benefits, and future growth.*" September 12, 2023. https://www.linearity.io/blog/ugc-statistics/.

[23] H. Wang, J. Kong, Y. Guo, X. Chen, "Mobile Web Browser Optimizations in the Cloud Era: A Survey," In Proc *2013 IEEE 7th International Symposium on Service Oriented System Engineering (SOSE),* pp. 527-536, 25-28 March 2013, doi: 10.1109/SOSE.2013.18.

[24] Z. Wang, F. Lin, L. Zhong, and M. Chishtie, "How far can client-only solutions go for mobile browser speed?" *Proc. 21st international conference on World Wide Web,* pp. 31-40, 2012.

[25] Z. Wang, F. Lin, L. Zhong, and M. Chishtie, "Why are web browsers slow on smartphones?" *Proc. 12th Workshop on Mobile Computing Systems and Applications*, pp. 91-96, 2011.

[26] D. Wong, *"Facebook mobile referrals up 253% in past year,"* Shareaholic, October 31, 2013. https://blog.shareaholic.com/facebook-mobile-referrals-10-2013.

[27] K. Zhang, L. Wang, A. Pan, and B. Zhu, "Smart caching for web browsers," In *Proceedings of the 19th international conference on World wide web* (WWW '10). ACM, New York, NY, USA, 491-500. DOI=10.1145/1772690.1772741 http://doi.acm.org/10.1145/1772690.1772741.