

ACCELERATING THE INTERNET IN THE PRESENCE OF BIG DATA: REDUCING USER DELAYS BY LEVERAGING HISTORICAL USER REQUEST PATTERNS FOR WEB CACHING

Chetan Kumar¹ <https://orcid.org/0000-0003-3331-7226>

¹California State University, San Marcos, CA, USA

Sean Marston² <https://orcid.org/0000-0003-1825-7368>

²Western Kentucky University, Bowling Green, KY, USA

ABSTRACT

Approximately 4 billion people have access to the Internet, additionally 23 billion devices are connected as of 2018. This has allowed for a substantial growth in data collection which has allowed for Big Data to flourish. The continued increase in user, devices, and Big Data usage has created a significant intensification in Internet traffic. This in turn has the potential to increase user delays when accessing data on the Internet. There are a number of ways to help reduce user latency, web caching is able to reduce web user delays in addition to reducing network traffic and the load on web servers. In this study we propose a proxy level web caching mechanism leveraging historical web patterns to help reduce user latency and accelerate the Internet. In addition we survey the state of the art of other caching approaches. Our investigation shows that using historical patterns as part of a proxy caching mechanisms in large scale networks can significantly shorten the latency for users in this era of Big Data.

Keywords: Big Data, User delays, Web caching, Proxy cache, Historical request patterns

Manuscript first received: 2019/04/22. Manuscript accepted: 2019/09/02

Address for correspondence:

Chetan Kumar, California State University, San Marcos, CA, USA. Email: ckumar@csusm.edu

Sean Marston, Western Kentucky University, Bowling Green, KY, USA. Email: sean.marston@wku.edu

INTRODUCTION

The Internet has demonstrated a tremendous growth in the amount of data that is sent and received, this trend of increasing traffic is expected to continue. As of 2018 there are more than 4 billion internet users, approximately 53% of the world's population and this is expected to increase to 5 billion users by 2020 (Indivigital, 2018). The Internet of Things (IoT) is playing a huge role in the amount of data being collected. In 2018 there were 13.4 billion "things" connected to the internet and this is expected to almost triple to 38.5 billion by 2020 (Dospeed, 2019). The rapid rise of data across a variety of technology throughout the world has led to an explosion of Big Data. According to McAfee & Brynjolfsson (2012) the key characteristic of Big Data that separates it from analytics of the past is the volume, velocity, and variety of data. They quantify that more data now cross the Internet every second than were stored in the entire Internet 20 years ago. In addition to McAfee & Brynjolfsson's three characteristics, veracity has become just as important characteristic with the increased amount of data being collected (IBM Big Data and Analytics, 2019). Based on the discussion from McAfee & Brynjolfsson (2012) we discuss volume, velocity, and variety of Big Data with the addition of veracity, as follows:

Volume. Data is being created at a high rate, as of 2017 the world creates 2.5 quintillion bytes per day and that number is expected to double approximately every 40 months. It is estimated that by 2020 there will be 40 zettabytes of data created and that there will be 1.7 MB of data created every second for each person on earth (Cloudtweaks, 2015). Companies now have the opportunity to work with single sets of data consisting of many petabytes, which they generate. Facebook generates approximately 4 petabytes of data every day from all the actions taken by users on the social media site. Cisco expects global data center traffic will reach 20.6 zettabytes by 2020, up from 6.8 zettabytes in 2016 (Cisco Global Cloud Index, 2018).

Velocity. The speed at which data is created is increasing as the number of users and devices accessing the internet. For example, in 2018 there are over 13 million devices connected to the internet sending and receiving data and is expected to triple by 2021. Facebook users generated 900 million pictures a day in 2018 and is expected to continue to increase (Gewirtz, 2018). The ability of application to deal with the speed of data has become very important. The ability to analyze data in real time has become important to many companies. IT security applications have to be able to analyze this significant amount of data flowing in real time to a network to prevent sinister payloads from being delivered. The New York Stock Exchange creates and distributes over 1 terabyte of data to hundreds of thousands of users in real time every trading session, allowing for traders' access to as accurate information as possible. The ability of a company to analyze this data in real time will allow them to make better informed decisions and be more agile than their competitors.

Variety. The types of data that are collected vary immensely, it can take the form of pictures, audio, video, sensor data, tweets, etc. There is a large variety of data created and shared by different social media networks alone. For example, in 2018 every minute there are 49,380 photos posted on Instagram, 4730,400 tweets sent by twitter users, 2,083,333 videos shared on snapchat, and 4,333,560 youtube videos watched. Smartphones and mobile devices are able to now provide large streams of data associated to the users based on location, activities, and application usage (McAfee & Brynjolfsson, 2012). Mobile devices accounted for 50% of website traffic as of 2018 (Stevens, 2018).

Veracity. There is a significant amount of data that is being collected, the certainty and trustworthiness of the data is an important aspect. Receiving poor quality of data can be very costly

to companies, Gartner research estimates that the financial impact of decisions based on poor quality data cost organization \$9.7 million a year (Moore, 2018). According to IBM Big Data and Analytics (2019) poor data quality costs the US economy \$3.1 Trillion a year. The increase of in the amount of data, especially real time data, makes data quality/veracity an increasingly important factor.

Due to the rapid growth of Big Data, cloud and edge computing continue to grow at a similar rate. This is due to the importance of these computing paradigms in playing a key role in receiving, managing, and analyzing Big Data. Duncan Pauly, CTO Edge Intelligence, states “the infrastructure edge is key, as it helps store and manage the data in a cost efficient way while decisions are made as to what data is trimmed, reduced, and ultimately shipped to the cloud.” The ability to transfer huge amounts of data to the edge infrastructure and across to other storage silos is important. Specifically, large amounts of data being is produced in real time cause issues when being transferred. For example, self-driving cars create and send enormous amounts of data to the edge, being able to receive and process this data in a timely matter is key to prevent self-driving cars from crashing. Cloud computing benefits Big Data through value added services, and it provides economies of scale specifically dealing with Big Data involving storage and analysis of data that is not time sensitive. Machine learning plays an important role in the analysis of data in both the cloud and edge systems. Finally, the need to improve data retention policies is important for all the data that is being created and stored due to the rapid rise of Big Data. Companies need to ensure it understands the cleaning of the data gathered while maintaining the integrity of the data. Additionally, it is important to understand what pieces of the data need to be backed up and how long it needs to be stored.

One aspect that is fueling the growth and the importance of Big Data is real time and streaming data. Real time and streaming data are playing key roles in professional and consumer decision making. Traders are making important financial decisions on real time data being received from different markets around the world, the delay of data can be the difference in a multimillion dollars profit or loss. Consumers have very limited attention span and the delay of data can lead to the loss of consumers for businesses. An important aspect of real time and streaming data is the quality of the data. The quality of data being delivered either to businesses partners or consumers plays a key role in building and maintaining trust in their relationships. Ensuring that high quality data can be delivered in a timely manner at different scales is important and when done well can lead to the differentiation of the services being provided. The value of data has been shown to increase with time. Older data can be used to better understand consumers and predict future needs, the types of data needed by corporations to understand their consumers varies by industry. Understanding the data need by the company plays an important role on what data needs to be retained.

According to IDC Worldwide Big Data Technology and Services 2018–2022 Forecast, the use of online Big Data applications will continue to increase in the future.(IDC Forecast Report, 2018). The Report states: “The Big Data and analytics software market, which in 2017 reached \$54.1 billion worldwide, is expected to grow at a five-year CAGR of 11.2% .” IDC states that this will occur based on the importance of Big Data to companies, the shift to continued shift to the use of public clouds, and the increased use of Artificial Intelligence as part of enterprise applications. Even as Big Data and analytics technology continues to develop and improve, the increase in traffic can lead to significant user delays in web access (Cui et al., 2018; Sorn & Tsuyoshi, 2013; Zhao & Wu, 2013; Kumar, 2010; Hosanagar & Tan, 2004; Datta et al., 2003).

The use of Web caching is one method that is used to reduce user delays. Web caching is a technology, typically located close to the end users, which temporarily stores copies of web objects for future use. This allows user requests to be served quicker by the cache when compared to directly being served from the origin web server (Cui et al., 2018; Davison, 2013; Ali et al., 2012; Hosanagar & Tan, 2004).

Web caching servers are placed through computer networks in different areas based on design. Proxy caches are situated anywhere between the client and webserver, but tend to be located close to end users which are typically computer network access points (Davison, 2013). However proxy caching may also be implemented at different locations throughout the network, including the browser and web-server levels (Cui et al., 2018; Ali et al., 2012; Kumar, 2010; Kumar & Norris, 2008; Davison, 2001). Proxy caches work by storing

copies of web objects and responding to requests directly from end users, reducing the number of requests at the origin server. This achieves reduced network traffic, load on web servers, and average delays experienced by web users (Floratou et al., 2015; Ali et al., 2012; Datta et al., 2003; Cao & Irani, 1997). This benefit of network of proxy caches is illustrated in Kumar (2009) example of the IRCache network (www.ircache.net). Figure 1 shows how a network of proxy caches with nodes at three locations can reduce user delays.

Assume that the U.S. node receives requests for honda.com, toyota.com, and audi.com web pages. Additionally these web pages have not been cached in the U.S. node. Since these web pages have not been cached in the U.S., the Japan node and Germany node can then be used to fulfill the requests. The U.S. node does not need to go to the origin webserver to fulfill the requests for the objects that are currently held in neighboring caches. Origin servers typically have the longest wait times to fulfill a request for an object, having the request fulfilled at a proxy cache will reduce the waiting time which in turn significantly reduces network delays (Floratou et al., 2015; Ali et al., 2012; Kumar, 2009). Technology providers and computer network administrators widely use proxy caching as a way to reduce delay (Davison, 2013). There are numerous examples of companies that provide or implement proxy caching solutions, including Alibaba a cloud content delivery network (alibabacloud.com/product/cdn), Spectrum an internet service provider (spectrum.com), and Memcached a server side caching solution provider (memcached.org/). The following are two illustrations, adapted from Davison (2013), of how some firms may practically benefit from caching. The following examples of how firms potentially benefit from the use of caching are based on two illustrations created by Davison (2013). In the first scenario company that has a significant number of users, such as Intel, may choose cache objects from many highly used servers in a proxy cache located close to the company's network gateway. The company benefits by reducing internal network delay and the external bandwidth usage required over expensive dedicated Internet connections. In the second scenario a content provider, such as Netflix, can reduce the number of requests a server must handle by placing a proxy cache directly in front of that server. This reduces the number of request and in turn speeds up content delivery. This type of proxy caching is also known as reverse caching, allowing the proxy node to cache objects for many clients but from usually only one server and is professionally provided by CDN firms such as Akamai. In each scenario the implementation of a proxy cache reduces access delays which benefit all Internet users (Floratou et al., 2015; Davison, 2013). As with all IT project investment decisions, when a selecting a caching solution the firm will need to need to evaluate costs of an implementation versus its benefit before deciding on the appropriate caching service. In this article we discuss the different techniques that exploit the use of historical user request patterns by proxy caches to reduce user request delays (Cui et al., 2018; Irani & Lam, 2015; Kumar, 2010; Kumar & Norris, 2008; Zeng et al., 2004).

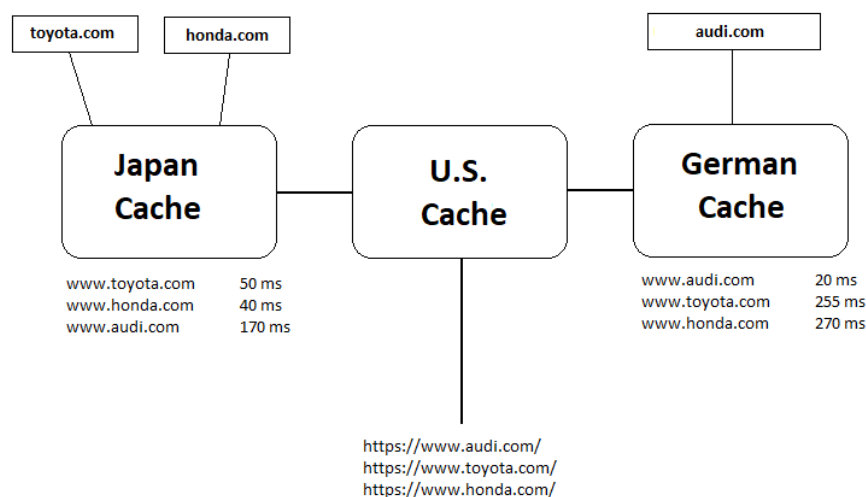


Figure 1. A proxy cache network

RELATED LITERATURE AND BACKGROUND

There is a growing interest in caching due to its application in reducing user delays while accessing the increasingly congested Internet (Cui et al., 2018; Floratou et al., 2015; Sorn & Tsuyoshi, 2013; Davison, 2013; Kumar, 2010; Datta et al., 2003). An extensive survey of caching techniques used in practice have been completed by Floratou et al. (2015), Zhang et al. (2103), Ali et al. (2012), Zeng et al. (2004), and Podlipnig and Boszormenyi (2003). Caching techniques include popular cache replacement strategies such as least recently used (LRU), first in first out (FIFO), and random replacement (RR). In the LRU policy the least recently requested object is evicted from the cache to make space for a new one, and their many extensions. The FIFO policy is one of the most basic in which the first object placed in the cache is replaced when the cache reaches capacity. The RR policy selects an object at random to be replaced with a new object when at new object needs to be paced in the cache. The main focus of caching research studies focus on improving performance metrics, some of the more common improvements are the decrease in bandwidth usage and reduction of user latency. However, there is limited amount of literature examining how caches are managed based on a model or data driven approach. The use of machine learning techniques to improve the use of web proxy caching is by Ali et al. (2012) while Zhao & Wu (2013) develop a caching approach for Big Data applications using MapReduce framework. Tauscher & Greenberg (1997), Cockburn & Mckenzie (2002), and Sorn & Tsuyoshi (2013) examine client-side behavior on the Internet to improve web caching, Irani & Lam (2015), Elfayoumy & Warden (2014), Rizzo & Vicisano (2000), and Cao & Irani (1997), discuss re-access patterns for web users, including those demonstrating repeating 24 hour re-access spikes. Floratou et al. (2015) examine multiple adaptive algorithms that create caches primarily based on work load patterns. Cui et al. (2018) propose a software defined network caching network which creates a cache based on the timely reviewing popular content using a heuristic algorithm to minimize inter ISP data delay.

A majority of websites in this era use dynamic content to populate their webpages which traditional web caching methods do not take into account. Irani & Lam (2015), Zhang et al. (2013), Ali et al. (2102), and Zeng et al. (2004) discuss a variety of caching methods that try to predict past object requests, such as Top-10 algorithm that compiles a list of most popular websites. Similar to those studies, Kumar & Norris (2008) propose a new proxy-level caching mechanism that consists of a quasi-static portion that exploits historical request patterns, as well as a dynamic portion that handles deviations from normal usage patterns. Comparing popular caching mechanisms, such as FIFO, the proposed caching mechanism of Kumar & Norris (2008) is shown to perform more favorably. Ali et al. (2012), Elfayoumy & Warden (2014), and Irani & Lam (2015) review and study cache replacement strategies using historical replacement strategies. Employing the use of historical user request patterns in a cache allow content providers and network administrators to drastically reduce web user delays at proxy servers (Kumar, 2016; Kumar, 2010).

USING HISTORICAL REQUESTS PATTERNS FOR WEB CACHING

It has been shown in prior studies that users at a proxy level tend to access the same web content on a routine basis while creating a distinguishable patterns in demand for the web content, for example accessing the content every half hour (Kumar, 2016; Irani & Lam, 2015; Elfayoumy & Warden, 2014; Rizzo & Vicisano, 2000; Cao & Irani, 1997). This is illustrated by a sport fan who goes to espn.com on a regular basis throughout the day to check sports news. Numerous previous studies

do not take into account dynamic documents, only static documents. The use of historical patterns in proxy caching have been extensively reviewed by Floratou et al. (2015), Irani & Lam (2015), Ali et al. (2012), Elfayoumy & Warden (2014), and Zeng et al. (2004), which survey techniques that utilize some aspects of historical user request patterns. One interesting approach that uses historical request patterns is Kumar & Norris (2008), they offer a model that aggregates patterns based on user object requests to use as part of the caching mechanism in a proxy server. Their caching mechanism aims to exploit replicating and repeating access patterns for web documents whose contents may change over time, but whose URL address online remains the same. Instances of this variety of web content include the landing pages of many websites (e.g., www.amazon.com, www.apple.com, www.facebook.com, www.google.com, www.netflix.com, etc.) which can change the specific content of their sites, but retain the same home page URL. Consequently, even if the website front page contents vary online, so long as the aggregate patterns by users for accessing the site at a particular time of day are identified, then the updated contents of the landing front page may be downloaded prior to the increased demand in user requests. In addition, a dynamic portion in the integrated caching mechanism may handle deviations and outliers from historical requests. If website object requests diverge from historical patterns then a portion of the cache may utilize a variation of the LRU policy, thereby warranting the integrated caching mechanism performs no worse than LRU (Kumar, 2016; Kumar, 2010; Mookherjee & Tan 2002).

A measure of the soundness of utilizing historical user request patterns to predict website object requests is demonstrated by analyzing IRCache proxy trace data by Kumar & Norris (2008). Table 1 demonstrates that the top 10 most requested sites, at the IR Cache proxy server network analyzed in the time period of years 2005-2006, in a 30 day period is analogous to those requested on the following day, shown in the first and second columns of the table, respectively. These similarities can be exploited for each and every time interval in the day, while also permitting the caching mechanism to adapt to aberrations from historical patterns. If the popularity of websites evolve over time, the top requested sites may vary, but the model of 30 day historical patterns being a predictor for the near future requests remains relevant in the Big Data era now. For example, Table 2 lists the current top 10 requested sites in the US as of January 2019 and February 2019, shown first and second columns, respectively (Alexa Report, 2019). Note from Table 2 that the most popular sites in the US remains relatively constant in successive 1 month periods. We observe in early 2019 that Google, Facebook, Amazon, Reddit, Wikipedia, Yahoo, Twitter, and Netflix, among others, continue to be among the most popular websites in a 30 day period from January to February. Consequently, the principle of exploiting patterns remains the same as proposed in Kumar (2010), Kumar & Norris (2008), and other studies (Floratou et al., 2015; Irani & Lam, 2015; Elfayoumy & Warden, 2014; Ali et al., 2012). We may predict the likely most requested sites in advance by mining historical request patterns. We may consequently decrease user delays by caching these objects in advance. The performance of the proposed caching mechanism is evaluated in Kumar & Norris (2008) against LRU policy utilizing their comprehensive IRCache network proxy trace data. The parametric test results validate that the integrated mechanism of Kumar & Norris (2008) outperforms the widely adopted LRU policy by more than 50% in terms of total costs. Other studies also show improvements over LRU by utilizing historical request patterns (Floratou et al., 2015; Irani & Lam, 2015; Ali et al., 2012).

Table 1. Top 10 requested sites at the IR Cache Network

Top 10 requested sites for days 1 through 30	Top 10 requested sites for day 31
yahoo.com	friendster.com
friendster.com	yahoo.com
microsoft.com	icq.com
water.com	microsoft.com
icq.com	msn.com
animespy.com	water.com
atwola.com	animespy.com
msn.com	17tahun.com
google.com	adbureau.net
phpwebhosting.com	google.com

Table 2. Top 10 requested US websites

Top 10 requested US websites January 2019	Top 10 requested US websites February 2019
google.com	google.com
youtube.com	youtube.com
facebook.com	facebook.com
amazon.com	amazon.com
reddit.com	wikipedia.org
wikipedia.org	reddit.com
yahoo.com	twitter.com
twitter.com	yahoo.com
netflix.com	linkedin.com
instagram.com	instagram.com

CONCLUSION

As Big Data moves forward it continues to increase the amount of traffic through the Internet. There are a number of aspects of Big Data that influences the amount of data collected. One position that is becoming a norm in a majority of companies and will influence the amount of traffic created by Big Data is the Chief Data Officer (CDO). The CDO is in charge of all aspects of a company's data, from the company's governance of the data to its utilization by the company. This includes the amount of data that a company gathers, which as time goes on will continue to increase the amount of traffic flowing on through their networks, and in turn the Internet. Another major influence on Big Data is the continued decrease in the cost of technology which has led to the explosion of the Internet of Things (IoT). The decrease in the cost of technology has made the number of affordable Internet connected devices to dramatically increase which has created a significant portion of the increased traffic on the internet. The ability of companies to manage this data and their networks will play an important role moving forward in amount of traffic flowing through the Internet. One important aspect in the management of this traffic is the use of web caches.

Reducing network traffic, load on web servers, and web user delays have all be shown as just a few of the numerous benefits from for all Internet users when effective proxy caching mechanism are used firms (Cui et al., 2018; Davison, 2013; Sorn & Tsuyoshi, 2013; Kumar, 2010; Datta et al., 2003). While these benefits are seen by the firms implementing the proxy cache, the benefits are also readily apparent to an end user. It has been shown that users shy away from slow loading website while appreciating and revisit fast loading websites. A user request for website that is fulfilled by a proxy cache will tend load faster than a request from an origin server request which will potentially be delayed several seconds.

Internet companies can also conserve investing resources in server farms for replicating web content to enhance load speeds (Floratou et al., 2015; Davison, 2013). The vast amounts of data that are being generated online today and the expected increase of data generated in the future makes this is especially relevant in the current era of Big Data. Proxy caching is beneficial for both the specific network where it is used as well as for all Internet users in general. Given the test results of Kumar & Norris (2008), as well as the merits of other caching approaches surveyed in Cui et al. (2018), Floratou et al. (2015), Irani & Lam (2015), Elfayoumy & Warden (2014), Zhang et al. (2013), and Zeng et al. (2004), we offer that effective proxy caching mechanisms exploiting historical user request patterns can significantly reduce delays for web users if they were to be adopted at large scale networks in this Big Data era.

KEY TERMS AND DEFINITIONS

Big Data: Massive amounts of data generated today characterized by volume, velocity, variety, and veracity.

Web caching: This involves temporary storage of web object copies at locations relatively close to the end user. Consequently user requests may be served faster than if they were served directly from the origin web server.

Proxy caches: These are caches located at computer network access points for web users. Proxy caches may store copies of web objects and directly serve requests for them in the network. Therefore they may reduce user delays by avoiding repeated requests to origin web servers.

Origin web server: The server where web content originates. User requests satisfied by the origin server typically have the longest waiting times.

Historical user request patterns: These are user web object request patterns that are previously observed. For example, at proxy level users typically re-access documents on a daily basis, and demand for a document spikes in 24 hours multiples’.

Web 2.0 technologies: Web traffic that primarily consists of user generated content such as video, audio, text, social networking, and collaboration.

Static documents: Web documents that are unaltered in content and size.

Dynamic documents: Web documents such as website front pages that frequently change contents.

Least Recently Used LRU caching policy: LRU is a popular cache replacement strategy where the least recently requested object is evicted from the cache to make room for a new object.

REFERENCES

- Alexa Report (2019). Top Sites in United States January February 2019 (n.d.). Retrieved from <http://www.alexa.com/topsites/countries/US>
- Ali, W., Shamsuddin, S.M., and Ismail, A.S. (2012). Intelligent web proxy caching approaches based on machine learning techniques. *Decision Support Systems*, 53(3), pp.565-579.
- Cao, C. & Irani, S. (1997). Cost-Aware WWW Proxy Caching Algorithms. In *Proceedings of the Usenix Symposium on Internet Technologies and Systems*.
- Cisco Global Cloud Index: Forecast and Methodology, 2016–2021 White Paper. (2018, November 19). Retrieved February 25, 2019, from <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.html>
- Cloudtweaks (2015). Infographic: How much data is produced every day? (n.d.) (2015, February 14). Retrieved February 25, 2019, from <https://cloudtweaks.com/2015/03/how-much-data-is-produced-every-day/>
- Cockburn A. & McKenzie, B. (2002). Pushing Back: Evaluating a New Behaviour for the Back and Forward Buttons in Web Browsers. *International Journal of Human-Computer Studies*.
- Cui, Y., Song, J., Li, M., Ren, Q., Zhang, Y., & Cai, X. (2018). SDN-based Big Data caching in ISP networks. *IEEE Transactions on Big Data*, 4(3), 356-367.
- Datta, A., Dutta, K., Thomas, H. & VanderMeer, D. (2003). World Wide Wait: A Study of Internet Scalability and Cache-Based Approaches to Alleviate It. *Management Science*, 49(10), 1425–1444.
- Davison, B.D. (2001). A Web Caching Primer. *IEEE Internet Computing*, 5(4), 38–45.
- Davison, B.D. (2013). *Web Caching and Content Delivery Resources*. Retrieved from <http://www.web-caching.com>
- Dospeed (2019). The Future of the Internet - 7 Big Predictions of 2020. (n.d.) (2019, January 26). Retrieved from <https://www.dospeedtest.com/blogs/the-future-of-the-internet-7-big-predictions-of-2020/>
- Elfayoumy, S. and Warden, S. (2014). Adaptive Cache Replacement: A Novel Approach. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 5(7).
- Floratou, A., Megiddo, N., Potti, N., Özcan, F., Kale, U., & Schmitz-Hermes, J. (2015). Adaptive Caching Algorithms for Big Data Systems. IBM Research Report.
- Gewirtz, D. (2018, March 21). Volume, velocity, and variety: Understanding the three V's of Big Data. Retrieved February 27, 2019, from <https://www.zdnet.com/article/volume-velocity-and-variety-understanding-the-three-vs-of-big-data/>
- Hosanagar, K., & Tan, Y. (2004). Optimal Duplication in Cooperative Web Caching. In *Proceedings of the 13th Workshop on Information Technology and Systems*.
- IBM Big Data & Analytics (2019). The Four V's of Big Data. (n.d.). Retrieved February 25, 2019, from <https://www.ibmbigdatahub.com/infographic/four-vs-big-data>
- IDC Worldwide Big Data Technology and Services 2014–2018 Forecast Report (2014). Retrieved from <http://www.idc.com/getdoc.jsp?containerId=250458>
- Indivigital (2018). 218 internet facts and statistics for 2018. (n.d.). Retrieved from <https://indivigital.com/218-internet-facts-and-stats-for-2018/>

- Irani, S. and Lam, J. (2015). Cache Replacement with Memory Allocation. In *Proceedings of the Seventeenth Workshop on Algorithm Engineering and Experiments (ALENEX)*.
- Kumar, C., & Norris, J.B. (2008). A New Approach for a Proxy-Level Web Caching Mechanism. *Decision Support Systems*, 46, 52-60.
- Kumar, C. (2009). Performance Evaluation for Implementations of a Network of Proxy Caches. *Decision Support Systems*, 46, 492-500.
- Kumar, C. (2010). Speeding up the Internet: Exploiting Historical User Request Patterns for Web Caching. In *Encyclopedia of E-Business Development and Management in the Global Economy*, Lee, I. (Ed.), IGI Global, PA.
- Kumar, C. (2016). Speeding Up the Internet in Big Data Era: Exploiting Historical User Request Patterns for Web Caching to Reduce User Delays. In I. Lee (Ed.), *Encyclopedia of E-Commerce Development, Implementation, and Management*. IGI Global, Hershey, PA, pp. 880-886
- McAfee, A. and Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harvard Business Review*, October Issue, Product #: R1210C-PDF-ENG.
- Mookherjee, V.S., & Tan, Y. (2002). Analysis of a Least Recently Used Cache Management Policy for Web Browsers. *Operations Research*, 50(2), 345-357.
- Moore, S. (2018, June 19). How to Create a Business Case for Data Quality Improvement. Retrieved February 25, 2019, from <https://www.gartner.com/smarterwithgartner/how-to-create-a-business-case-for-data-quality-improvement/>
- Podlipnig, S., & Boszormenyi, L. (2003). A Survey of Web Cache Replacement Strategies. *ACM Computing Surveys*, 35(4), 374-398.
- Rizzo, L., & Vicisano, L. (2000). Replacement Policies for a Proxy Cache. *IEEE/ACM Transactions on Networking*, 8(2), 158-170.
- Sorn, J. and Tsuyoshi, M. (2013) Web Caching Replacement Algorithm Based on Web Usage Data. *New Generation Computing*, 31(4), 311-329.
- Stevens, J. (2018, December 18). Internet Statistics & Facts (Including Mobile) for 2019. Retrieved February 25, 2019, from <https://hostingfacts.com/internet-facts-stats/>
- Tauscher L. & Greenberg, S. (1997). How People Revisit Web Pages: Empirical Findings and Implications for the Design of History Systems. *International Journal of Human Computer Studies, Special issue on World Wide Web Usability*, 47, 97-138.
- Zeng, D., Wang, F., & Liu, M. (2004). Efficient Web Content Delivery Using Proxy Caching Techniques. *IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews*, 34(3), 270-280.
- Zhang, G. Li, Y., & Lin, T. (2013). Caching in information centric networking: A survey. *Computer Networks*, 57 (16), 3128-3141.
- Zhao, Y. and Wu, J. (2013). Dache: A data aware caching for big-data applications using the MapReduce framework. *Proceedings of IEEE INFOCOM*.