

A Performance Evaluation of Dropbox in the light of Personal Cloud Storage Systems

Abdullah Talha Kabakus
Abant Izzet Baysal University
IT Center

Resul Kara
Duzce University
Faculty of Engineering
Department of Computer Engineering

ABSTRACT

Personal cloud storage services have attracted the attention of users thanks to this great accessibility. Personal cloud storage services allow users to synchronize their local files with the servers in the cloud in order to make them accessible from anywhere, on any device, and at any time thanks to the client applications provided by the vendors. Various experiments are evaluated on Dropbox, one of the most popular personal cloud storage services, in order to reveal how they affect the performances of file operations such as *downloading* files from the cloud storage, *uploading* files to the cloud storage, and *deleting* files from the cloud storage. The experimental results are discussed with revealing the insights about how the file operations are performed on Dropbox. Also, we discuss the several techniques and architectural design principles provided by Dropbox in order to improve the file synchronization performance in terms of required time and stability. Even though we focus on Dropbox, the proposed study gives an idea about the architecture of personal cloud storage services in general and how can they be improved in terms of performance, security, and usability. We conclude the paper with discussing the features to evaluate personal cloud storage services.

General Terms

Data Management, Cloud Storage, Cloud Systems

Keywords

cloud storage, personal service, performance, Dropbox, evaluation

1. INTRODUCTION

Personal cloud storage services allow users to synchronize their local files with the servers in the cloud in order to make them accessible from anywhere, on any device, and at any time thanks to the clients provided by the personal cloud storage service vendors. Cloud storage services have gained more popularity day by day since (1) they backup users local files in the cloud, (2) they allow sharing the local files of users between their other devices thanks to the provided client applications developed for the various operating systems, (3) they allow distributed and collaborative work since the files in the cloud are being automatically synchronized with the shared users, and (4) most of these services provide their services for free or reduced prices with respect to current storage services. Comparisons of the available cloud storage services tend to be based on (1) ease of use, (2) stability, (3) security, (4) performance, (5) quality of experience (QoE), and (6) price [1,2]. The competition between the available cloud storage services causes decreasing the cost per GB [3]. According to a recent report by *CIO*¹, the global public cloud market will top \$146 billion in 2017, up from just \$87 billion

in 2015 and is growing at a 22 percent compound annual growth rate [4]. A comparison of the most well-known personal cloud storage services is listed in Table 1.

Table 1. A comparison of the most well-known personal cloud storage services

Criteria	Dropbox ²	MEGA ³	Google Drive ⁴	OneDrive ⁵
File size limit	10GB; Unlimited With Desktop Client	Unlimited	5 TB	10 GB
Size of provided free storage	2 GB	50 GB	15 GB	5 GB
Online editing	No	No	Yes	Yes
File versioning	Yes	Yes	Yes	Yes
Selective synchronization	Yes	Yes		
Java SDK for standalone applications	Yes	Partial; based on C++	Deprecated ; uses OAuth 2.0	No
Support for Linux	Yes	Yes	No	No
Support for Windows	Yes	Yes	Yes	Yes
Support for MacOS X	Yes	Yes	Yes	Yes
Support for Android	Yes	Yes	Yes	Yes
Support for iOS	Yes	Yes	Yes	Yes

According to several reports, Dropbox is currently the most used cloud storage service [5–7] which provides 2 GB cloud storage for free despite that both Google Drive and OneDrive have advantages of proving their personal cloud services embedded into another service accounts which are mail and operating system accounts, respectively. According to the Google Trends⁶, the popularity comparison of three most

¹ <http://www.cio.com>

² <https://dropbox.com>

³ <https://mega.nz>

⁴ <https://drive.google.com>

⁵ <https://onedrive.live.com>

⁶ <https://trends.google.com>

well-known personal cloud storage services namely Dropbox, Google Drive, and OneDrive shows that Dropbox has surpassed other services for the last 5 years (between 19 February 2012 and 19 February 2017) as the result is presented in Fig.1

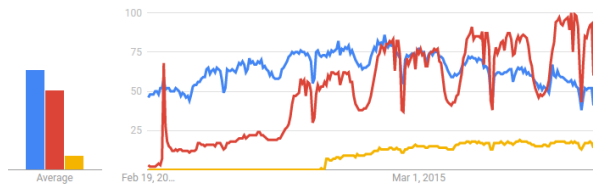


Fig 1: Interests for the Dropbox (blue), Google Drive (red), OneDrive (yellow) for the last 5 years according to the Google Trends

The starting point of cloud storage services is data synchronization which is an approach based on monitoring and mapping a local folder as the base folder of the cloud storage [8]. Every file operation to this folder is instantly noticed and synchronized to the cloud storage by the client software provided by the cloud storage service provider unless any sub-folder of this folder is intentionally ignored by the user from the synchronization which is called as “selective synchronization”. Despite that high interest from the public, not enough work in the literature focus on the insights of the cloud storage services. A systematic methodology is necessary in order to evaluate various experiments on cloud storage services. This paper proposes a methodology to evaluate various experiments in order to shed light on the architectural design of Dropbox, the performance of Dropbox for various file operations, and the way Dropbox provides secure services. The reason we focus on Dropbox instead of other services is that Dropbox provides the most advanced and ready-to-use SDK⁷ to develop standalone applications which fit for our case. Even though we focus on Dropbox, the proposed study gives an idea about the architecture of personal cloud storage services in general. We have implemented our own service using Java programming language which is described in Section 3 to evaluate various experiments on Dropbox. The main contributions of the paper are listed as follows:

- We propose a systematic methodology to evaluate various experiments in order to reveal their effect on the performance in terms of elapsed time to complete evaluated file operation.
- Thanks to the experimental results, the insights of how Dropbox handles file operations such as *upload*, *download*, and *delete* are discussed.
- The general requirements of a secure and stable cloud storage service are discussed.
- The features to evaluate personal cloud storage services are proposed.

The rest of the paper is structured as follows: Section 2 reviews the related work. Section 3 presents the experimental results and discussion. Finally, Section 4 concludes the paper with future directions.

2. RELATED WORK

Hu et al. [1] document wide variations in backup and restore performance, the type of data is backed-up, no liability for data loss, and problems with data privacy. According to their

experiments, they report that it is not safe to assume that all the data the user cares about has been backed up. Also, they report that it is not always possible to restore all the data stored in the cloud.

Drago et al. [9] propose a methodology to compare five popular personal cloud storage services including Dropbox, SkyDrive, Google Drive, Wuala, and Amazon Cloud Drive in order to reveal their architectural differences and capabilities which are assessed by executing a series of benchmarks. The capabilities of cloud storage services they analyze are chunking, bundling, client-side deduplication, delta-encoding, and data compression. The capabilities they investigate of the clients provided by cloud storage services include synchronization startup, completion time, and protocol overhead.

Li et al. [8] propose a metric named *TUE* to quantify the traffic usage efficiency of data synchronization to reveal if the current data synchronization traffic of cloud storage services is efficiently used. The cloud storage services they investigate are Dropbox, Google Drive, OneDrive, Box, Ubuntu One, and SugarSync. They report that Dropbox is the only cloud storage service among those six cloud storage services which compresses the synchronization data for every access operation. They select ten key impact factors and four design choices as they are listed in Table 2.

Table 2. Key impact factors and design choices of personal cloud storage services [8]

Client side	<ul style="list-style-type: none"> • Client location • Client hardware • Access method • File size • File operation • Data update size • Date update rate • Data compression level • Synchronization deferment
Server side	<ul style="list-style-type: none"> • Data synchronization granularity • Data deduplication granularity
Network side	<ul style="list-style-type: none"> • Synchronization traffic • Bandwidth • Latency

Gracia-Tinedo et al. [10] present a measurement study by actively accessing to free accounts through the REST APIs of three cloud storage services including Dropbox, Box, and SugarSync in order to analyze important aspects to reveal the differences between their QoS (Quality of Service) such as transfer speed, variability, and failure rate. According to their examinations, they found the high variability in transfer performance depending on the geographic location, the type of traffic, the file size, and the hour of the day. They conclude their examinations with stating that the analyzed services are found as reliable.

Li et al. [11] propose a system namely *CloudCmp* for comparing public cloud providers by using the metrics based on computation, storage, and network. Drago et al. [12] present a characterization of Dropbox by performing several experiments on Dropbox such as the transfer/synchronization protocol, the network traffic generated by typical file operations such as adding, removing or updating file(s) on the local synchronization folder, and several utilization behaviors.

⁷ <https://github.com/dropbox/dropbox-sdk-java>

Casas et al. [13] propose a study on the Quality of Experience (QoE) in cloud storage and file synchronization services in order to shed lights on the impact of client initialization, up-link bandwidth, and down-link bandwidth. They develop an application named The Box which is used by 52 participants who performed common activities in Cloud Storage and File Synchronization applications such as Dropbox, Google Drive, and Sky Drive. The activities they performed contain file storage (uploading files from client to cloud server), multi-device file synchronization (uploading files from client to cloud server, then synchronizing those files with another client), and file sharing (the user A shares a folder with the user B, and synchronization starts when the user A accepts the file sharing invitation).

Bocchi et al. [14] present a comprehensive characterization of three personal cloud storage services namely Dropbox, Google Drive, and OneDrive. They investigate their connection frequencies, the workload that their users generate, usage scenarios (which file operation is more used), and their performance in terms of throughput which is defined as the ratio between payload size and flow duration for a given TCP flow. They report that while users of both Dropbox and Google Drive generally download more than they upload, users of OneDrive show the opposite pattern. The other works [15–17] focus on the digital forensic investigation of cloud storage services by analyzing all the artifacts generated by the various devices

3. EXPERIMENTAL RESULTS AND DISCUSSION

Since, for the best of our knowledge, there is no performance evaluation tool that can be used for the purpose of this study, a service which is implemented using Java programming language is used to evaluate the effects of various metrics on the file operations of Dropbox. The implemented service uses the *Dropbox Java SDK* in order to automate file operations on Dropbox. The responsibilities of the implemented service are listed as follows:

- Generate plain-text files of given sizes,
- Generate image files of given sizes,
- Upload files to Dropbox,
- Download the uploaded files from Dropbox,
- Delete the uploaded files from Dropbox.

Plain-text files are generated using the *RandomAccessFile*⁸ class of java.io package. The implemented service accepts a parameter which defines the size of the file going to be generated and generates the file of the given size in the local disk. Similarly, images are generated using *Java Graphics2D*⁹ library which lets rendering 2-dimensional shapes, text, and images programmatically. Since the content of the image has no importance for the aim of the proposed work, we simply draw squares filled with a selected color. The bigger dimension, the larger size of the file is generated. The workflow of the proposed file generation process is presented in Fig. 2.

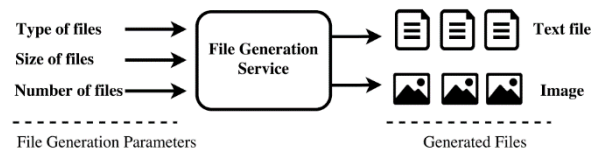


Fig 2: The workflow of the proposed file generation process

All the experiments are evaluated using a single machine on the same network in order to prevent any possible effect of hardware-specific differences on the results of experiments. The hardware and software specifications of the machine the experiments are evaluated on are listed in Table 3.

Table 3. The hardware and software specifications of the machine used to evaluate experiments

Operating system	Ubuntu 14.04
Operating system architecture	x86_64
Hard drive	1 TB 7200 RPM SATA-3
File system	ext4
Memory	16 GB DDR3L 1600MHz
CPU	Intel Core i7-4710MQ 4-Cores; 6 MB L3; 2.50 GHz > 3.50 GHz
Java version	1.8.0_60 (64-bit)
Dropbox SDK version	2.1.2

The experiments that are evaluated in order to reveal the effect of various metrics are (1) the effect of the number of files on download, upload, and delete operations, (2) the effect of file size on download, upload, and delete operations, and (3) the effect of the file type on download, upload, and delete operations. For each experiment, the elapsed time in milliseconds is calculated by the difference between the start and the end of each operation. The related times are calculated by using the *java.lang.System.currentTimeMillis*¹⁰ method which returns the time since January 1st 1970 in milliseconds. The testbed and workflow of experiments are illustrated in Fig. 3 and each experiment is discussed in the following subsections.

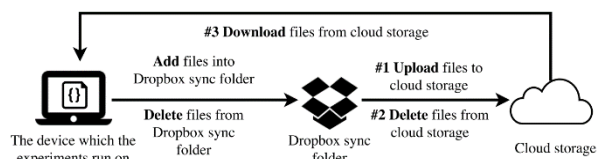


Fig 3: The testbed and workflow of experiments

3.1 Experiment #1 – The effect of the number of files

The number of files to be operated is expected to have big an impact on the performance of the operation in terms of completion duration. We experimented the effect of the number files on Dropbox to reveal how Dropbox handles the

⁸

<https://docs.oracle.com/javase/8/docs/api/java/io/RandomAccessFile.html>

⁹ <https://docs.oracle.com/javase/8/docs/api/java/awt/Graphics2D.html>

¹⁰

<https://docs.oracle.com/javase/8/docs/api/java/lang/System.html#currentTimeMillis-->

bulk file operations. Therefore, the *download*, *upload*, and *delete* performances of Dropbox for the various numbers of 100 KB files are experimented. The elapsed times to complete *delete* operations for 1, 10, and 100 files are calculated as 0 since they are completed in milliseconds. The reason behind this result is that the files are not actually deleted from Dropbox when the delete operation is executed whether using the SDK, or the web or mobile user interfaces. Instead of that, the files which are deleted using the web interface or provided clients just move to the “trash” folder which is a common practice of cloud storage services in order to give a chance to user to restore the files which are accidentally *deleted* or required in a limited time. According to Dropbox, the *deleted* files are stored up to 30 days for the free plan which can be extended to 120 days for the premium plans [18].

Table 4. The elapsed times in seconds to complete download, upload, and delete operations for the various numbers of 100 KB files

Operation	Number of Files			
	1	10	100	1000
<i>download</i>	1	14	54	569
<i>upload</i>	9	13	114	1125
<i>delete</i>	0	0	0	4

3.2 Experiment #2 – The effect of file size

Experiment 2 reveals the effect of file sizes on the performance of Dropbox in terms of completion duration per each operation. This experiment also reveals how the big files in size are handled by Dropbox. The sizes of files have been chosen up to 10 MB. As the elapsed times to upload 100 files with various file sizes are listed in Table 5, the elapsed times do not increase directly proportional with the size of the file. This result implies that the processes need to be completed before and after the synchronization between the client and the cloud server(s) such as chunking, bundling, client-side deduplication, delta encoding, and data compression take more time compared to uploading the content of the file for the smaller files in size. Also, the data compression on the client-side which is available for every access operation on Dropbox decreases the time required to complete each operation [5,19]. When the size of the file which is going to be uploaded/downloaded is bigger than 4 MB, the file is partitioned into 4 MB blocks (which is also known as chunking) which hashes are represented by blocklists [20]. Each blocklist contains a list of SHA-256 hashes of the block it represents. This approach also makes file operations resumable at the level of blocks instead of repeating the operation from scratch when the connection between the client and the cloud servers drops. This design is useful especially for the mobile devices when the limited bandwidth and the interceptable mobile network are considered. Dropbox reports that the streaming synchronization improves the synchronization performance approximately 25% which is only available for the files which are large enough to require multiple store/retrieve requests [20]. This implies that the elapsed times to upload larger files do not increase directly proportional with the sizes of files as it is also experimented.

Table 5. The elapsed times in seconds to complete download, upload, and delete operations for the various sizes of 100 files

Operation	Size of Files			
	10 KB	100 KB	1 MB	10 MB
<i>download</i>	46	54	72	327
<i>upload</i>	95	114	134	293
<i>delete</i>	0	0	0	0

Experiment 3 reveals the effect of file types on the performance of Dropbox in terms of completion duration per each operation. The result of the experiment which is presented in Fig. 4 implies that the type of the file has no effect on the performance of operations on Dropbox since the elapsed times for operations are (almost) same. The *upload* times are calculated as the *double* of download times which are related to the network connection and not specific to one of the file types. Also, this situation is a common case for most internet providers [21]. The reason behind evaluating this experiment is revealing if the full-text indexing feature of Dropbox which is available for text files and allows searching the content of file alongside its extension and file name [22,23] starts during the *upload* or it delays the completion of the *upload*.

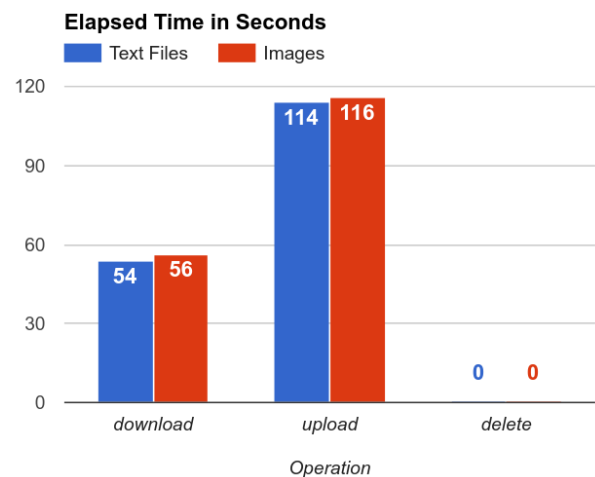


Fig 4: The elapsed times in seconds to complete download, upload, and delete operations for the different types of 100 files

The three experiments are evaluated on Dropbox in order to reveal how they affect the performances of file operations in terms of elapsed times. Beside from performance, the accessibility, security, stability, and consistency of the file operations between the client and cloud storage server are solely based on the features of the personal cloud storage service. The list of the features to evaluate personal cloud storage services contains our own features alongside the features proposed by Weiss [24]. The detail of each feature is listed in Table 6.

Table 6. The features to evaluate personal cloud storage services

Feature	Description	Aim of the Feature
Auto-resuming transfers	The process of resuming the file operation even the connection between the client and server drops	Performance
Cache	The file storage which is used by the client in order to speed up the data access	Performance
Chunking	The ability to split the larger amount of data into parts	Performance
Compression	The process of reducing the amount of data transferred between the client and server	Performance
Data integrity	The process of checking the accuracy and consistency of the data	Consistency
Deduplication	The process of searching previously stored, identical files on the cloud before uploading a new file in order to create a copy when the identical file is found	Performance
Delta synchronization	The process of transferring only the modified parts of the file instead of transferring the entire file every time it is updated	Performance
Encryption	The process of converting the data unreadable before the upload in order to ensure access from only the authorized parties	Security
Hashing	The process of analyzing two files to be compared without actually revealing the contents by generating a unique and numerical value from the file	Performance
LAN synchronization	The process of looking the new or update file on the Local Area Network (LAN) instead of directly communicating with the cloud server	Performance
Local storage	Hard drive space on the device to store local files	Accessibility
Offline access	The ability to access the data even the Internet	Accessibility

Feature	Description	Aim of the Feature
	connection is not available	
Platform independence	The ability to provide synchronization for various platforms such as computer and mobile operating systems	Accessibility
Selective synchronization	The ability to select specific folders to be synced on the device which is useful especially for the devices with limited storages	Performance
SSL/TLS	The usage of two most common methods of data encryption on the Internet	Security
Streaming sync	The process of speeding up the downloading the files as soon as any part of the file is stored in the cloud instead of waiting for the entire file is uploaded	Performance
Synchronization	The process of making the data in the cloud available for the clients	Accessibility

4. CONCLUSION

Personal cloud storage services have gained more popularity since (1) they ease accessing files of users with providing clients for various platforms, (2) they synchronize files as soon as they are created or edited in local folder on the cloud storage, (3) most of them provide large cloud storages for free or very limited prices, and (4) they serve as secure backup solutions for local files. In this paper, we have evaluated various experiments on Dropbox in order to reveal (1) how these experiments affect Dropbox, and (2) how Dropbox handles file synchronization. Despite that we focus on a personal cloud storage service, the paper gives a general idea about the architecture of personal cloud storage services and how can they be improved in terms of performance, security, and usability. Experimental results show that several techniques and design principles are used to improve the performance of file synchronization in terms of the required time. For the future improvements, authors would like to extend the experiments to cover more detail about the insights of personal cloud storage services and evaluate other personal cloud storage services.

5. REFERENCES

- [1] Hu, W., Yang, T. and Matthews, J. N. (2010) The good, the bad and the ugly of consumer cloud storage. ACM SIGOPS Oper. Syst. Rev., ACM 44, 110–115.
- [2] Bocchi, E., Drago, I. and Mellia, M. (2015) Personal Cloud Storage Benchmarks and Comparison. IEEE Trans. Cloud Comput. PP, 1–14.
- [3] Naldi, M. and Mastroeni, L. (2013) Cloud Storage Pricing: A Comparison of Current Practices. In HotTopiCS '13 Proceedings of the 2013 international

- workshop on Hot topics in cloud services, pp 27–34, Prague, Czech Republic.
- [4] Boulton, C. (2016) 6 trends that will shape cloud computing in 2017. CIO.
- [5] Li, Z., Wilson, C., Jiang, Z., Liu, Y., Zhao, B. Y., Jin, C., Zhang, Z. L. and Dai, Y. (2013) Efficient Batched Synchronization in Dropbox-Like Cloud Storage Services. In The 14th ACM/IFIP/USENIX International Middleware Conference (Middleware), pp 307–327, Beijing, China.
- [6] Henry, A. (2017) Most Popular Cloud Storage Provider: Dropbox. Lifehacker.
- [7] Henderson, N. (2016) Dropbox Most Popular Mobile Cloud Storage Service: Report. Talkin Cloud.
- [8] Li, Z., Jin, C., Xu, T., Wilson, C., Liu, Y., Cheng, L., Liu, Y., Dai, Y. and Zhang, Z.-L. (2014) Towards Network-level Efficiency for Cloud Storage Services. In IMC '14 Proceedings of the 2014 Conference on Internet Measurement Conference, pp 115–128.
- [9] Drago, I., Bocchi, E., Mellia, M., Slatman, H. and Pras, A. (2013) Benchmarking Personal Cloud Storage. In IMC '13 Proceedings of the 2013 Conference on Internet Measurement Conference, pp 205–212, Barcelona, Spain.
- [10] Gracia-Tinedo, R., Artigas, M. S., Moreno-Martinez, A., Cotes, C. and Lopez, P. G. (2013) Actively Measuring Personal Cloud Storage. In IEEE International Conference on Cloud Computing, CLOUD, pp 301–308.
- [11] Li, A., Yang, X., Kandula, S. and Zhang, M. (2010) CloudCmp: Comparing Public Cloud Providers. In IMC '10 Proceedings of the 10th ACM SIGCOMM conference on Internet measurement, pp 1–10, Melbourne, Australia.
- [12] Drago, I., Mellia, M., M Munafò, M., Sperotto, A., Sadre, R. and Pras, A. (2012) Inside Dropbox: Understanding Personal Cloud Storage Services. In IMC '12 Proceedings of the 2012 ACM conference on Internet measurement conference, pp 481–494, Boston, Massachusetts, USA.
- [13] Casas, P., Fischer, H. R., Suette, S. and Schatz, R. (2013) A First Look at Quality of Experience in Personal Cloud Storage Services. In 2013 IEEE International Conference on Communications Workshops (ICC 2013), pp 733–737, IEEE, Budapest, Hungary.
- [14] Bocchi, E., Drago, I. and Mellia, M. (2015) Personal Cloud Storage: Usage, Performance and Impact of Terminals. In 2015 IEEE 4th International Conference on Cloud Networking (CloudNet), pp 106–111, Niagara Falls, ON, Canada.
- [15] Chung, H., Park, J., Lee, S. and Kang, C. (2012) Digital forensic investigation of cloud storage services. *Digit. Investig.* 9, 81–95.
- [16] Daryabar, F., Dehghantanha, A. and Choo, K.-K. R. (2016) Cloud storage forensics: MEGA as a case study. *Aust. J. Forensic Sci.* 618, 1–14.
- [17] Daryabar, F., Dehghantanha, A., Eterovic-Soric, B., Choo, K.-K. R., Federici, C., Daryabar, F., Dehghantanha, A., Eterovic-Soric, B., Daryabar, F., Dehghantanha, A., et al. (2016) Forensic investigation of OneDrive, Box, GoogleDrive and Dropbox applications on Android and iOS devices. *Aust. J. Forensic Sci.* 48, 1–28.
- [18] (2017) Upgrade your account. Dropbox.
- [19] Goncalves, G., Drago, I., Da Silva, A. P. C., Vieira, A. B. and Almeida, J. M. (2014) Modeling the Dropbox Client Behavior. In 2014 IEEE International Conference on Communications, ICC 2014, pp 1332–1337, Sydney, Australia.
- [20] Koorapati, N. (2014) Streaming File Synchronization. Dropbox.
- [21] Weinberger, D. (2011) Why Uploading Is Slower Than Downloading. Huffingt. Post.
- [22] (2017) Search the contents of files on dropbox.com. Dropbox.
- [23] Faulkner, A. (2016) Improving the performance of full-text search. Dropbox.
- [24] Weiss, D. (2014) What is this file sync thing and why should I care about it? Dropbox.