Performance and Time Improvement of LT Codes-based Cloud Storage

Nastaran Chakani

Shahrood University of Technology

Seyed Masoud Mirrezaei (✉ sm_mirrezaei@yahoo.com)

Shahrood University of Technology <https://orcid.org/0000-0001-8493-2072>

Ghosheh Abed Hodtani

Ferdowsi University of Mashhad Faculty of Engineering

Research

Keywords: Cloud Storage, Degree Distributions, LT Codes, Retrieval Time

Posted Date: June 21st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-598840/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

RESEARCH

Performance and Time Improvement of LT Codes-based Cloud Storage

Nastaran Chakani¹, Seyed Masoud Mirrezaei^{1*} and Gosheh Abed Hodtani²

*Correspondence:

sm.mirrezaei@shahroodut.ac.ir

¹Faculty of Electrical Engineering,

Shahrood University of

Technology, Shahrood, Iran

Full list of author information is
available at the end of the article

Abstract

Outsourcing data on the cloud storage services has already attracted great attention due to prospect of rapid data growth and storing efficiencies for customers. The coding-based cloud storage approach can offer more reliable and faster solution with less storage space in comparison with replication-based cloud storage. LT codes as a famous member of rateless codes family can improve performance of storage systems utilizing good degree distributions. Since degree distribution plays key role in LT codes performance, recently introduced Poisson Robust Soliton Distribution (PRSD) and Combined Poisson Robust Soliton Distribution (CPRSD) motivate us to investigate LT codes-based cloud storage system. So, we exploit LT codes with new degree distributions in order to provide lower average degree and higher decoding efficiency, specifically when receiving fewer encoding symbols, comparing with popular degree distribution, Robust Soliton Distribution (RSD). In this paper, we show that proposed cloud storage outperforms traditional ones in terms of storage space and robustness encountering unavailability of encoding symbols, due to compatible properties of PRSD and CPRSD with cloud storage essence. Furthermore, modified decoding process based on required encoding symbols behavior is presented to reduce data retrieval time. Numerical results confirm improvement of cloud storage performance.

Keywords: Cloud Storage, Degree Distributions, LT Codes, Retrieval Time

1 Introduction

Digital world meets rapid evolutions due to huge growth of data. According to International Data Corporation (IDC) report, data will grow up to 175 Zetabyte by 2025 [1]. Storing data follows a lot of burdens for owners such as constraints of physical devices, data recovery and also concern of inaccessibility and cyber attacks. Cloud storage has provided efficient and promising solution to relieve organizations and different users of data storing issues from mid-1990s. Cloud storage has several advantages such as feasibility to access outsourced data anytime from anywhere, through internet and to contribute stored data with permission. Fast immigration towards cloud and close contest of providers to offer best facilities with reasonable cost are great motivations to find novel solutions in order to improve cloud storage performance. Generally, cloud storage works based on fragmentation of data content and distribution of fragments over storage nodes. Two main techniques are employed to ensure data reliability and availability in distributed storage systems: replication and coding. Replication is the most straightforward method, distributing copies of data over storage nodes. High storage space requirement is the main drawback of replication-based storage and it is not suitable for distributed systems with

high concurrent access requests. Coding-based storage systems can achieve reliability many orders of magnitude higher than replication-based systems for the same redundancy level [2]. There have been plenty of works on erasure codes-based distributed systems [3–7]. High decoding complexity and communication cost to repair a corrupted data fragment reduce popularity of this solution. In order to generate a corrupted encoded packet, usually reconstruction of all the original packets is required. Thus, communication cost to repair is equal to the size of the entire original. Reduction in repair communication cost can be achieved by network coding-based storage which works through combining encoded packets in healthy nodes. Utilization of Gaussian elimination decoding in network codes and optimal erasure codes, makes them inefficient [8–10]. Near optimal codes such as LT codes with low complexity encoding and decoding have been proposed for reliable cloud storage systems. If near optimal codes are exploited, the variation of one original symbol only changes few numbers of encoded symbols, whereas in optimal codes-based system, almost all of encoded symbols may be affected by a little bit of modification. Although the number of original packets may be smaller in previous methods, the decoding process performs much faster than in the other storage services [11, 12]. In this paper, we investigate performance of LT codes-based cloud storage with two recent degree distributions in addition to Robust Soliton distribution. We study effects of different parameters value on successful data retrieval. In the following, we propose new algorithm in order to reduce retrieval time of user data. The rest of this paper is organized as follows. In section 2, related works are reviewed briefly. In section 3, we present LT codes descriptions and its degree distributions. Our system model is stated in section 4. Performance analysis such as parameter selection, simulation results comparison and time improvement are discussed in section 5. Finally, in section 6, we bring conclusion of the paper.

2 Related works

Coding has been widely used in cloud storage systems to improve their performance in different aspects. Local Reconstruction Codes (LRC) as a subset of erasure codes are introduced in [5], keep storage overhead low comparing with Reed-Solomon codes in Windows Azure Storage. Significant decrease of bandwidth and I/Os during repair and improvement in latency for large I/Os are achieved by LRC. In [6], a distributed storage system provides fast content download by encoding contents with Maximum Separable Codes (MDS) and applying fork-join queuing for user requests. Results show an essential trade-off between expected download time and storage space which can be useful in design of system with delay constraints. Exploiting LT codes with speculative access mechanisms for parallel write and read in a distributed storage architecture leads to high and robust performance [11, 13, 14]. Due to introducing symmetric data redundancy and rateless property of LT codes, proposed system has high flexibility on data access and improvement compared to traditional parallel storage systems. In [12, 15], a secure cloud storage service is designed with near-optimal LT codes in order to solve the reliability issue. Proposed scheme presents efficient data retrieval by exploiting the fast Belief Propagation decoding algorithm, and also utilizes the public integrity verification which helps the data owner to be free from the burden of being online. Employing exact repair,

Figure 1: LT encoding process

minimizes data repair complexity and reduces cost. Although the performance analysis and experimental results show an equivalent storage and communication cost in comparison with other erasure codes-based systems, this secure cloud storage service achieves a much faster data retrieval. Comparing to network coding-based storage systems, proposed service reduces storage cost and provides faster data retrieval with comparable communication cost. In [16], LT-based architecture was proposed for the back-end of Block-Level Cloud (BLCS) storage that achieves sufficient levels of performance in terms of access and transfer, availability, integrity, and confidentiality. Interesting features of LT codes such as low complexity and on-the-fly redundancy setting, make them suitable for BLCS system. Results indicates that by applying appropriate system parameters, good compromise can be achieved and proposed BLCS outperforms traditional ones. Main tradeoff between file retrieval delay and successful decoding probability is investigated in a distributed cloud storage system [17]. Proposed multi-stage user request scheme plays efficient role in average retrieval delay reduction. Solving optimization problem for optimal two-stage request scheme determines the proper number of packets requested in the first stage, follows high decoding probability.

3 Methods

3.1 LT Codes

LT codes [18] are the first class of universal fountain codes. These codes can potentially generate infinite encoded symbols through XOR operation of a subset of original symbols. For every encoded symbol, degree d is chosen independently from a given degree distribution. LT codes can recover k original symbols from any $k(1 + \epsilon)$ encoded symbols with probability $1 - \delta$. Belief Propagation (BP) is used as an efficient decoding algorithm for these near optimal codes which depends on degree-1 encoding symbols. In LT codes-based cloud storage system, a user file is first fragmented in to k original symbols, then these original symbols are encoded in to n encoded symbols. We briefly describe encoding and BP decoding procedure of LT codes for $k = 4$ and $n = 5$ in Figure 1 and Figure 2, respectively.

Number of original symbols that are combined together is known as code degree, which is designated from two common degree distribution. First, Ideal Soliton distribution (ISD) $\rho(i)$, which is defined as follows

Figure 2: Belief Propagation Decoding

$$\rho(i) = \begin{cases} \rho(1) = 1/k \\ \rho(i) = 1/i(i-1) & i = 1, 2, \dots, k \end{cases} \quad (1)$$

Although, ISD performs weekly in practice due to its ripple expected size that is one, it provides great perception for new distributions. Main distribution of LT codes is Robust Soliton that is denoted by $\mu(i)$. Let $R = c \ln(k/\delta) \sqrt{k}$ expected ripple size for some suitable constant $c > 0$. Define

$$\tau(i) = \begin{cases} R/ik & i = 1, \dots, k/R - 1 \\ R \ln(R/\delta)/k & i = k/R \\ 0 & i = k/R + 1, \dots, k \end{cases} \quad (2)$$

Add $\rho(i)$ to $\tau(i)$ and normalize to obtain $\mu(i)$

$$\beta = \sum_{i=1}^k \rho(i) + \tau(i) \quad (3)$$

$$\mu(i) = (\rho(i) + \tau(i))/\beta \quad i = 1, \dots, k \quad (4)$$

With a good degree distribution, LT codes can perform well. Although Robust Soliton distribution ensures ripple does not disappear during decoding process with high probability and can achieve to good performance for LT codes, new introduced distributions can improve performance in terms of overhead and recovery probability which is considerable, since cloud storage follows "pay-as-you-use" paradigm.

3.2 Poisson robust soliton distribution

By combining the characteristics of Poisson distribution (PD) and Robust Soliton distribution (RSD), recently introduced distribution [19] with appropriate parameters can generate more degree-1 compared with RSD. Thus, PRSD successful retrieval probability outperforms RSD in lower overhead. The improved PD (IPD) is given by

$$\theta(i) = \begin{cases} 1/2 & i = 2 \\ \lambda^i e^{-\lambda} / i! & i = 1, 3, \dots, k \end{cases} \quad (5)$$

where λ is a positive constant.
Then, the proposed PRSD is obtained as follows

$$\Omega(i) = \theta(i) + \tau(i) / \sum_{i=1}^k \theta(i) + \tau(i) \quad i = 1, 2, \dots, k \quad (6)$$

PRSD provides lower average degree and limited degrees in comparison with RSD. Thus, we can achieve to cloud storage with faster retrieval and higher successful decoding probability in lower overheads.

3.3 Combined poisson robust soliton distribution

In order to reduce overhead and consuming the time of encoding and decoding process, CPRSD proposed by combining IPD and RSD as follows [20]
First, $\eta(i)$ is obtained from normalization of $\theta(i)$,

$$\eta(i) = \theta(i) / \sum_{d=1}^k \theta(i) \quad i = 1, 2, \dots, k \quad (7)$$

The CPRSD is represented as

$$\Lambda(i) = \eta(i).a + \mu(i).(1 - a) \quad (8)$$

where the range of a is located between 0 and 1.

4 System model

To investigate the performance of LT codes with new distributions on a cloud storage system, we consider small scale cloud with 15 storage nodes. First, we encode 20 user data of various size with LT codes, then distribution of encoded data is accomplished over cloud storage in Regional and Multi-Regional storage modes [21]. In Regional mode, data is distributed over a Region with at least two Availability Zones, Region selection is based on minimum distance to data owner location. In Multi-Regional mode that at least two Regions are selected, we assume selection of first Region as nearest one and random selection of second Region that can be equivalent to user option in cloud storage systems. To resemble function of our system model to reality

of cloud storage system and also study retrieval time, we consider M/G/1 queue for every storage node [6] and a further M/G/1 queue for head server to direct retrieval requests to corresponding nodes.

We inspect successful decoding probability of LT codes-based cloud storage with PRSD and CPRSD degree distributions in two state. The first one is non-removal state of storage nodes and the latter is removal state with the aim of checking the effect of inaccessibility or failure of nodes. General structure of our system model is shown in Figure 3.

5 Simulation Results and Discussion

5.1 Parameter selection

In this paper, we study LT codes with $k = \{100, 250, 500\}$. All simulations run in MATLAB. The first step is selecting n , number of encoded symbols that are required to recover original symbols with high probability. As shown in Table 1, obtained values of n corresponding to k in our system model, are almost close to values derived from empirical model for decodability in [17] and the term $k + O(\sqrt{k} \cdot \ln^2(k/\delta))$ [18]. At the second step, two main parameters of RSD and PRSD, c and δ are determined, that play important roles in degree distributions achievement. By investigating successful decoding probability for different values of these parameters and k , we set $c = 0.08$ and $\delta = 0.1$ as a suitable pair. The measure of failure decoding probability δ , is reasonable in practice. For instance, RSD and PRSD successful decoding probability for $k = 250$ are illustrated in Figures 4 and 5.

Finally, fundamental parameter of CPRSD, a is selected, which represents the contribution of IPD and RSD, in new degree distribution. Successful decoding probability against a for different k and versus overhead for different a , are displayed in Figure 6 and Figure 7, respectively. Since variation of a has effect on overhead as well as successful decoding probability, a trade-off is discussed. We set $a = 0.4$ to reach

$n \backslash k$	[17]	$k + O(\sqrt{k} \ln^2(k/\delta))$	simulation
100	165	157	158
200	310	271	290
250	-	333	388
300	450	395	441
400	580	518	568
500	-	639	710

Table 1: Required number of encoding symbols for successful decoding

Figure 4: Successful decoding probability with RSD for different c and δ

an acceptable compromise between overhead and successful decoding probability. Based on the analysis of expectations and similarity of mathematical properties of PD and BD when $k < 20$, we consider $\lambda \approx 3.04$ [19].

Degree distributions comparison

In this section, we present the comparison of successful decoding probability against overhead, for RSD, PRSD and CPRSD on cloud storage system. Overhead is defined as follows

$$\epsilon = (n - k)/k \quad (9)$$

In order to study LT codes-based cloud storage performance, we consider two phases through simulations due to random nature of encoding and decoding of LT codes and also validation of our results. Outer phase includes selection of encoded

symbols degree, generating metadata, encoding by LT codes and finally distributing over cloud. Furthermore, inner phase is assumed in every outer phase which encompasses users' requests for various data from diverse geographical locations at random and data retrieval process. As shown in Figures 8 to 10, for $k = \{100, 250, 500\}$ and in non-removal state of storage nodes, successful decoding probability with CPRSD and PRSD outperform RSD in particular at lower overheads. Proposed cloud storage performance in confronting unavailability or loss of encoding symbols for $k = 250$, is depicted in Figure 11. Better robustness and higher successful decoding probability can be achieved applying CPRSD and PRSD for various number of original symbols.

Successful data retrieval in lower overheads and also in presence of partly encoding symbols loss, is a notable attainment in cloud storage systems. Also, it could lead to reduction in storage space, retrieval time, users and providers cost and in result more satisfactory services.

5.2 Time improvement

We assume a scenario to compute data retrieval time. First retrieval time is defined as follows

$$T_{total} = T_w + \frac{L}{r} + T_{decoding} \quad (10)$$

T_w is queuing delay, second term is data transmission time and $T_{decoding}$ is mean of decoding process time in simulations.

As mentioned before, we consider M/G/1 queue for every storage node and head datacenter. Our simulations run based on following assumptions. In every iteration, selection accomplish randomly among 2, 4, 6^{ms} for mean service time of datacenters and among 1, 1.5, 2^{ms} for variance. The mean arrival rate is assumed 5^{ms} . Moreover, arrival rate for head datacenter is set to 10, mean and variance of service time are considered as 1 and 3^{ms} , respectively. r is assumed 30.03 Mbps, rests on global average download speed report for mobile internet in 2019 [22]. Generally, LT decoder needs an undetermined number of encoded symbols from storage system to retrieve data successfully. More delay arises from this ambiguity in LT-codes based cloud storage system. Thus, there is a compromise between successful decoding probability and retrieval delay. According to our observation during decoding process, number of encoding symbols required for successful data retrieval follows normal distribution as shown in histogram for $k = 250$ in Figure 12. Since retrieval time is comparable to user experience of cloud storage service, we design a scheme in which decoding process is implemented for the number of encoding symbols lie within two standard deviations instead of blind search in almost big interval. Figure 13 shows successful decoding probability against overhead after applying proposed decoding process in removal state. As clearly seen, reduction in successful decoding probability is negligible in particular for PRSD and CPRSD, whereas time reduction is significant according to Table 2. Retrieval time using proposed decoding process can be decreased up to 70 percent with PRSD and 67 percent with CPRSD.

Figure 7: Successful decoding probability with CPRSD for $k = 250$ and different a

k	Retrieval time (second)					
	RSD		PRSD		CPRSD	
	main	proposed	main	proposed	main	proposed
100	0.4510	0.2530	0.2962	0.1606	0.4403	0.2488
250	4.7818	1.5613	1.9538	0.6583	3.9001	1.3413
500	9.7998	4.8760	5.0312	1.4730	14.5405	4.8298

Table 2: Retrieval time comparison between main and proposed decoding process

6 Conclusion

In this paper, we studied LT codes-based cloud storage with recently introduced degree distribution. Data retrieval with more success was achieved, specifically in less overheads and also in presence of unavailability or loss of encoding symbols with PRSD and CPRSD in comparison with common solution, RSD. Furthermore, we proposed modified decoding algorithm in order to obtain retrieval time improvement. The performance analysis and experimental results show that proposed LT codes-based cloud storage system can provide higher successful decoding probability, less storage space, more robustness and faster data retrieval.

Acknowledgements

Not applicable.

Funding

Not applicable.

Abbreviations

RSD : Robust Soliton Distribution; PRSD :Poisson Robust Soliton Distribution; CPRSD : Combined Poisson Robust Soliton Distribution; IDC: International Data Corporation; LRC: Local Reconstruction Codes; MDS : Maximum

Figure 8: Successful decoding probability for $k = 100$ in non-removal state Figure 9: Successful decoding probability for $k = 250$ in non-removal state

Figure 10: Successful decoding probability for $k = 500$ in non-removal state Figure 11: Successful decoding probability for $k = 250$ in removal state

Figure 12: Histogram of required number of encoding symbols for successful data retrieval for $k = 250$

Separable Codes; BLCS : Block-Level Cloud storage; BP : Belief Propagation; PD : Poisson distribution; IPD : improved PD;

Availability of data and materials

Not applicable

Competing interests

The authors declare that they have no competing interests.

Consent for publication

All authors have agree and given their consent for submission of this paper to Euraship Journal of Wireless Communications and Networking.

Authors' contributions

Conceptualization: Dr. Seyed Masoud Mirrezaei, Prof. Ghosheh Abed Hodtani and Nastaran Chakani; Formal analysis: Dr. Seyed Masoud Mirrezaei and Nastaran Chakani; Funding acquisition: Dr. Seyed Masoud Mirrezaei and Nastaran Chakani; Investigation: Nastaran Chakani; Methodology: Dr. Seyed Masoud Mirrezaei; Project administration: Dr. Seyed Masoud Mirrezaei and Prof. Ghosheh Abed Hodtani; Resources and Software: Nastaran Chakani; Supervision: Dr. Seyed Masoud Mirrezaei and Prof. Ghosheh Abed Hodtani; Writing original draft: Nastaran Chakani; Writing review editing: Dr. Seyed Masoud Mirrezaei and Prof. Ghosheh Abed Hodtani.

Authors' information

Nastaran Chakani received the B.Sc. degree in communication engineering from Azad University of Mashhad, Mashhad, Iran in 2012 and the M.Sc. degree in communications engineering from the Shahrood University of Technology, Shahrood, Iran in 2018. Her current research interests include information theory, channel and source coding, cloud storage and information theoretic learning.

Seyed Masoud Mirrezaei received his B.Sc. degree in Communication Engineering from K. N. Toosi University of Technology in Sept. 2004. He received his M.Sc. and Ph.D. degrees in Electrical Engineering from Amirkabir University of Technology, Tehran, Iran, in 2007 and 2013, respectively. He has been as a Ph.D. student and researcher in "Mobile and Wireless Networks Research Laboratory" in Amirkabir University of Technology under the supervision of Prof. Karim Faez. Also, he was a visiting research student in the Signal Design and Analysis Laboratory (SDAL) at Queen's University, Kingston, Canada under the supervision of Prof. Shahram Yousefi from March 2011 to February 2012. He joined the Electrical Engineering Department of Shahrood University of Technology from 2013 up to now. His research interests lie in communications, cloud systems, big data, networks, information theory, signal processing, channel coding and network coding.

Figure 13: Successful decoding probability with proposed decoding process for $k = 250$ in removal state

Ghosheh Abed Hodtani received the B.Sc. degree in electronics engineering and the M.Sc. degree in communications engineering, both from Isfahan University of Technology, Isfahan, Iran, in 1985, 1987, respectively. He joined Electrical Engineering Dept., at Ferdowsi University of Mashhad, Mashhad, Iran, in 1987. He decided to pursue his studies in 2005 and received the Ph.D. degree (with excellent grade) from Sharif University of Technology, Tehran, Iran, in 2008; and he is now a professor in electrical engineering. His research interests are in multi-user information theory, communication theory, wireless communications, information-theoretic learning and signal processing. Prof. Hodtani is the author of a textbook on electrical circuits and is the winner of the best paper award at IEEE ICT-2010

Author details

¹Faculty of Electrical Engineering, Shahrood University of Technology, Shahrood, Iran. ²Faculty of Electrical Engineering, Ferdowsi University, Mashhad, Iran.

References

1. paper IDC, W.: The Digitization of the World From Edge to Core (2018)
2. Weatherspoon, H., Kubiatowicz, J.D.: Erasure coding vs. replication: A quantitative comparison. *Lecture Notes in Computer Science* **2429**, 328–337 (2002). doi:10.1007/3-540-45748-8_31
3. Bhuvaneshwari, P., Tharini, C.: Review on ldpc codes for big data storage. *Wireless Personal Communications* **117**(2), 1601–1625 (2021)
4. Joshi, G., Liu, Y., Soljanin, E.: Coding for fast content download. In: *Communication, Control, and Computing (Allerton)*, 2012 50th Annual Allerton Conference On, pp. 326–333. IEEE, ??? (2012)
5. Huang, C., Simitci, H., Xu, Y., Ogus, A., Calder, B., Gopalan, P., Li, J., Yekhanin, S.: Erasure coding in windows Azure storage. *Proceedings of the 2012 USENIX Annual Technical Conference*, 15–26 (2012)
6. Joshi, G., Liu, Y., Soljanin, E.: On the delay-storage trade-off in content download from coded distributed storage systems. *IEEE Journal on Selected Areas in Communications* **32**(5), 989–997 (2014)
7. Kumar, A., Tandon, R., Clancy, T.C.: On the latency and energy efficiency of distributed storage systems. *IEEE Transactions on Cloud Computing* **5**(2), 221–233 (2017). doi:10.1109/TCC.2015.2459711
8. Dimakis, A.G., Godfrey, P.B., Wu, Y., Wainwright, M.J., Ramchandran, K.: Network coding for distributed storage systems. *IEEE transactions on information theory* **56**(9), 4539–4551 (2010)
9. Dimakis, A.G., Ramchandran, K., Wu, Y., Suh, C.: A survey on network codes for distributed storage. *Proceedings of the IEEE* **99**(3), 476–489 (2011)
10. Huang, J., Fei, Z., Cao, C., Xiao, M.: Design and analysis of online fountain codes for intermediate performance. *IEEE Transactions on Communications* **68**(9), 5313–5325 (2020)

11. Xia, H., Chien, A.A.: RobuSTore: a distributed storage architecture with robust and high performance. Proceedings of the 2007 ACM/IEEE conference on Supercomputing, November (c), 10–16. doi:10.1145/1362622.1362682
12. Cao, N., Yu, S., Yang, Z., Lou, W., Hou, Y.T.: LT codes-based secure and reliable cloud storage service. In: 2012 Proceedings IEEE INFOCOM, pp. 693–701. IEEE, ??? (2012)
13. Huang, J., Fei, Z., Cao, C., Xiao, M., Xie, X.: Reliable broadcast based on online fountain codes. IEEE Communications Letters (2020)
14. He, M., Hua, C., Xu, W., Gu, P., Shen, X.S.: Delay optimal concurrent transmissions with raptor codes in dual connectivity networks. IEEE Transactions on Network Science and Engineering (2021)
15. Shi, P., Wang, Z., Li, D., Xiang, W.: Zigzag decodable online fountain codes with high intermediate symbol recovery rates. IEEE Transactions on Communications **68**(11), 6629–6641 (2020)
16. Anglano, C., Gaeta, R., Grangetto, M.: Exploiting Rateless Codes in Cloud Storage Systems. IEEE Transactions on Parallel and Distributed Systems **26**(5), 1313–1322 (2015). doi:10.1109/TPDS.2014.2321745
17. Lu, H., Foh, C.H., Wen, Y., Cai, J.: Delay-optimized file retrieval under LT-based cloud storage. IEEE Transactions on Cloud Computing **5**(4), 656–666 (2015)
18. Luby, M.: LT codes. In: The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings., pp. 271–280. IEEE, ??? (2002)
19. Yao, W., Yi, B., Huang, T., Li, W.: Poisson Robust Soliton Distribution for LT Codes. IEEE Communications Letters **20**(8), 1499–1502 (2016). doi:10.1109/LCOMM.2016.2578920
20. Yao, W., Yi, B., Li, W., Huang, T., Xie, Q.: CPRSD for LT codes. IET Communications **10**(12), 1411–1415 (2016). doi:10.1049/iet-com.2015.1183
21. <https://cloud.google.com/storage/docs/storage-classes>. Storage-classes: No Title
22. J. Clement: average mobile fixed broadband download upload speeds. <https://www.statista.com/statistics/896779/average-mobile-fixed-broadband-download-upload-speeds>

7 Figures

8 Tables

$\begin{matrix} k \\ n \end{matrix}$	[17]	$k + O(\sqrt{k} \cdot \ln^2(k/\delta))$	simulation
100	165	157	158
200	310	271	290
250	-	333	388
300	450	395	441
400	580	518	568
500	-	639	710

Tab .1: Required number of encoding symbols for successful decoding

Retrieval time (second)						
$\begin{matrix} k \\ \end{matrix}$	RSD		PRSD		CPRSD	
	main	proposed	main	proposed	main	proposed
100	0.4510	0.2530	0.2962	0.1606	0.4403	0.2488
250	4.7818	1.5613	1.9538	0.6583	3.9001	1.3413
500	9.7998	4.8760	5.0312	1.4730	14.5405	4.8298

Tab. 2: Retrieval time comparison between main and proposed decoding process

Figure 1: LT encoding process

Figure 2: Belief Propagation Decoding

Figure 3: General structure of our LT codes-based cloud storage

Figure 4: Successful decoding probability with RSD for different c and δ Figure 5: Successful decoding probability with PRSD for different c and δ

Figure 6: Successful decoding probability with CPRSD for different k and a Figure 7: Successful decoding probability with CPRSD for $k = 250$ and different a

Figure 8: Successful decoding probability for $k = 100$ in non-removal state

Figure 9: Successful decoding probability for $k = 250$ in non-removal state

Figure 10: Successful decoding probability for $k = 500$ in non-removal state Figure 11: Successful decoding probability for $k = 250$ in removal state

Figure 12: Histogram of required number of encoding symbols for successful data retrieval for $k = 250$

Figure 13: Successful decoding probability with proposed decoding process for $k = 250$ in removal state