International Journal of Pure and Applied Mathematics
Volume 119 No. 12 2018, 13221-13231
ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

ijpam.eu

SECURE DEDUP WITH ENCRYPTED DATA

N SENTHIL SELVAN, G ANJALI, B KESINI, V SUBRAMANIYASWAMY*
School of computing, SASTRA Deemed To Be University, Thanjavur- 613401.
*Corresponding Author

ABSTRACT

Cloud computing is one of the way of service provision over the internet today. Cloud computing is the developing a next level from the last decades. One of the drawbacks, cloud storage is a privacy security at the CSP. So, the chunks users stored by the encrypted data for the purpose of security. Cloud storage vendors which allow to decreases chunked data and more efficient storage saver. One of the best techniques is deduplication, duplicate data is stored only once .In this paper, propose a checksum algorithm for distributing objects to agents, in a way that improves our chances of identifying a leaker. We evaluate its performance based on effective and efficient storage level. Its support data access control and revocation at the same time.

Keywords: privacy security, encrypted data, deduplication, checksum algorithm.

1. INTRODUCTION

Cloud computing is the development of the existing technology. Users can take a lot of benefits from all of these technologies, not having the use of knowledge about with them. The cloud aims to reduce the costs and user-friendly in the main concentration to the business point in IT services. Virtualization is the most enabling technique in cloud computing. Virtualization which separates a physical computing device into numerous virtual devices, they can be managed to perform computing tasks. In operating system-Level virtualization needs to creating a measurable system of diverse independent devices, without the purpose of resources can be efficient and allocated.

Virtualization provides which requires the agility to go faster IT operations and minimize the cost by raising the infrastructure utilization. The automatic computing which automatically provision resources on request services. Main benefits of the virtualization by reducing user participation, process to go faster the process, make less labor costs and lower the human errors as possible. Users practice as most difficult to solve the business problems.

Deduplication is basically a compression technique for removing data. File level deduplication takes into account the entire file, thus even small update or append makes the file different from the existing version of it and thereby decreasing ratio of deduplication. In the case of each level data verification deduplication chunks are represented as deduplication. Deduplication can be categorized into two types namely, client-side deduplication and as source-side deduplication. Performing deduplication at client side ensuring bandwidth saving since only hash value of file is sent to server, if duplicate is exists and also various applications like backup, metadata management primary storage, etc., storage optimization.

Cloud service is one of the information storage service, the most important and popular service today. Cloud users upload individual data to the center of a (CSP) [7] and allow it to maintaining the information. Cloud storage schemes in the outsourcing data is kept never changed in the servers. In the storage, users can store their information and no longer time to kept in the locally [13]. Thus, the available of the multiple data files storing in the guaranteed distributed cloud services. In existing system, brute-force attack [7] used to avoid numerous same data in the CSP. For the verification multiplying data having same information means which check it through the system integrity, suppose if it will be failed, there will be no copied information file.

2. RELATED WORK

In the digital world, cloud storage is very popular today. Cloud storage providers are the services providers to the user for the On-demand services. Cloud storage having the providing as Drop Box [2], Google Drive [3], Mozy [4], which helps to storage space reduce in the cloud storage by storing the each file only once of each uploaded file [7]. For example, Deduplication [11] efficient method to avoid the duplicate file in the storage point of view, but it cannot be handle encrypted data [9][10]. In Encrypted data there are the numerous method to encrypted the data, one of the encrypted method is Message-Locked encryption (MLE) [7] and others [12]. CE was introduced by Douceur et al. [7].CE is subject to hybrid security drawbacks and also others [14],[15].

In the Existing system, they actually storage space, security and also the brute force attack related information are proposed the DupLESS based on the Bellare et al. [1],[14],[15].We are not know about the extant plan to try the satisfied the available, by the way of decreasing the duplicated data revaluation[16]. Encryption of data is

insufficient to achieve the security and privacy policy.

For the example of the duplication information in the identifying the plaintexts identification, if we checked the identical in the storage data which not privacy security implications [7][1].In existing deduplication system ,information of privacy are disclosure to the multiple data storage.

This paper main objective to resolve the leaked out privacy information from the CSP to unauthorized people or outsourcing website. And also to minimize the storage space in the cloud in practical by avoiding the deduplication data [13]. It also clearly reduce the more limitation from the existing system based on the privacy security and availability of the data owner [10].

In the organization, or concerns to achieve the trending fore scalable things [1][7]. This is actually not a new thing to the today digitization society. Because today we improve and achieving new technology with successful outcomes. Whatever the society reached the peak of point, but we are not ensuring privacy policy, security, minimum storage space requirement are more challenging to us. So, we try to solve the effective way to reducing the storage space and increasing the bandwidth in the CSP. We analyzed the security of individual information.

In this paper to demonstrate that the proposed scheme out of performing the reducing the limitation of the existing system in the terms of the computation, security, storage and privacy policy. And also we used rich efficient logarithmic time to search the duplicate data in the cloud storage, by reducing the time complexity as well. We improved the security of the stored data, easy to handle the account of user and data owner and also the data available.

3. PROPOSED SYSTEM

This paper propose History Alert Rewriting (HAR) algorithm to avoid de-duplicate encrypted data stored in the CSP. It non-segregated information such stored in the

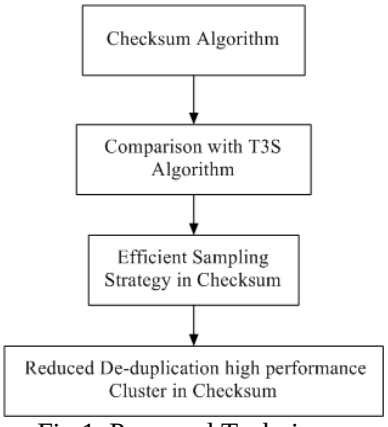

Fig.1. Proposed Technique

cloud with allow to access. Intention of this paper with a scheme hinge on confrontation with data-owner and Traditional Encryption to control the store with encrypted data by deduplication. Focus to finding an effective way to reducing the issues in the time of

data owner not able to be present. In this method using double encryption key for encrypted data storing in the CSP. First the information owner provide the secret key to data user then authorized party (AP) send the secret key to data owner. Both AP key and private key generate the encrypted key that key using for encryption. AES algorithm using the encrypt content stored in cloud. Figure 1 shows the proposed method.

Step 1: Input: pk_j, Policy (u_i), Policy (*AP*)

Step 2: CSP requesting AP to dataowner and grant access to duplicated data for u_j by providing pk_j.

Step 3: After ensuring data ownership through challenge, AP checks Policy (*AP*) and issues CSP $rk_{AP} \rightarrow ui$ = RG (pk_{AP}; sk_{AP}; pk_j) if the check is positive.

Step 4: CSP transfering the data E(pk_{AP}; DEK_i) into E(pk_j; DEK_i) if Policy (u_i) authorizes u_j to share similar information *M* encrypted by DEK_i: R($rk_{AP} \rightarrow u_i$; E(pk_{AP}; DEK)) = E(pk_j; DEK*i*)

Step 5: Data holder u_j obtains DEK_i by decrypting E (pk_j; DEK_i) with sk_j: DEK*i* $\rightarrow$ D(sk_j;E(pkj; DEK_i)), and then it can access data M at CSP.

Schemes:

- User Registration and Cloud access
- Indexing the Cloud Data
- Finding similarity and avoiding

3.1. USER REGISTRATION AND CLOUD ACCESS:

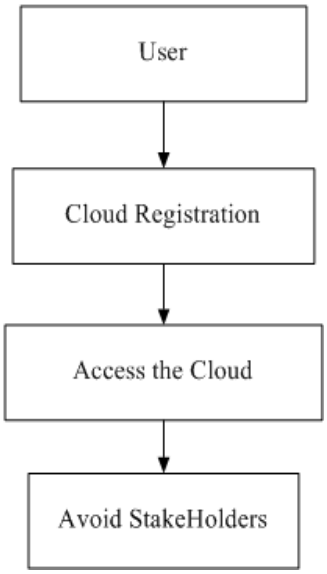

Fig.2. User registration and cloud access

Access users only to have authentication process before registration, Authentication process is priority based to movable management process includes locating registrations and delivery services and it authorized by the clients with ensuring resources and preventing from unauthorised sources. Before registration of cloud services to ensure whether the client is an authenticated or not to access cloud server. Can ensure the information stored in the cloud is used judiciously by the responsible stakeholders as per the service level agreements. The process shown in figure 2

3.2. INDEXING THE DATASET USING SPARSE:

The based on requirements to prepare the dataset in avoid de-duplication content. Indexing is nothing but consists of structured and unstructured format. Unstructured format is an unarranged format. Sparse Indexing is based on the reference format and capturing the repeated words queries. Indexing converts the unarranged format into structured arranged format. This may be avoiding the problem of delay during searching. Actually Spare indexing are used to rapidly locating the data without search the database in row-wise in all time a database table is accessing. This process shown in figure 3.

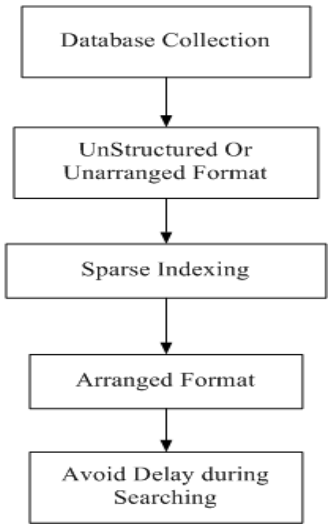

Fig.3. Indexing the Cloud Data

3.3. FINDING THE SIMILARITY AND AVOIDING DEDUPLICATION:

In the deduplication process, which is divided into content-static and content-dynamic. If the static category separate the details of input information into same level data, which are comparing with one another. The duplicates identify among the data then best way to eliminate by static method from disarrangement issues. Comparison between the spare indexing pairs belonging to the similar information and try to stopping the deduplication process using novel techniques. It is decreasing the user loading in the major tasks by deduplication. This process shown in figure.4

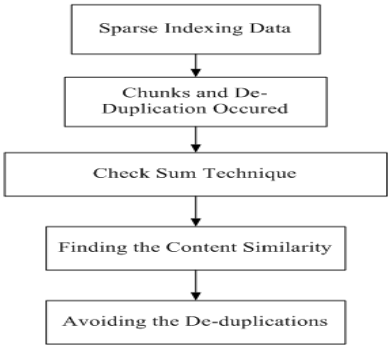

Fig.4. Find similarity and avoiding.

4. RESULTS AND DISCUSSION:

4.1. USER REGISTRATION FORM:

Users register for the cloud by the data owner and cloud service provider. In fig.5 shows user needs to register their details for

accessing the system, if the user registration is successfully, user details saved in database. Then only user can access the system with the unique public key for each user.

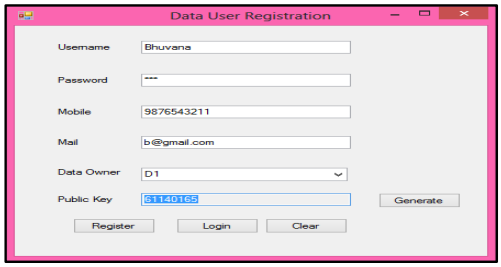

Fig.5. Data User Registration form

4.2. User Login Form:

After the successful registration, user can login for the cloud by the data owner and CSP. In fig 6 data Owner needs to login the cloud to access the system. Data Owner enters their id and password to sign in the cloud. After the login process, to check the login id and password given by the user. If it is authorized means data owner to allow the user access the cloud.

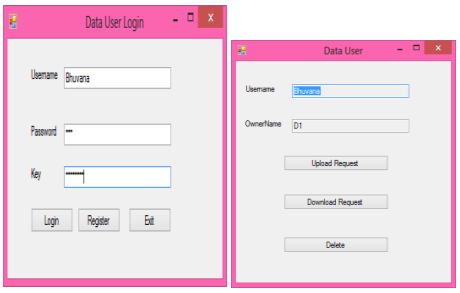

Fig.6. Data User Login form

4.3. Data User request to upload and download a file:

Data user select the option request to upload the file that shown on figure 7. Then user should select the file through the browse and send request to data owner. Data owner view the request list and give the secret key to user and verified the keys. After the key verified AP generates the secret key to the data owner, once again check the key verification and the select the upload option file upload to the cloud successfully.

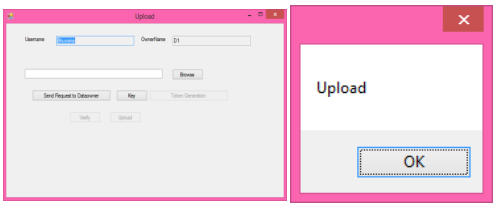

Fig.7. Data User Upload form

Same process for the download the file , check the data owner and select which file going to download and the request the key to the data owner and check that key are identical or not. Then download the file from the cloud by the permission through the data owner shown in the fig.8.

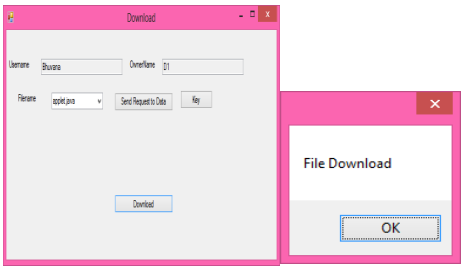

Fig.8. Data User Download form

4.4. Cloud Service provider access upload the file:

Register for the cloud by the CSP needs to register their details for accessing the system. Login for the cloud with a secret key. After the login process to access the file. Then, choose the file and AES algorithm using encrypted content and selected file is encrypted. The Encrypted file is stored into the cloud. AF crawler algorithm using to avoid the de-duplication. If the information wants to uploading the file with encrypted into the CSP. To avoid the de-duplication file from the uploaded file are stored into the cloud. If the information user uploading that file into CSP to check the duplication file and intimates the de-duplication file. Process shown on figure 9.

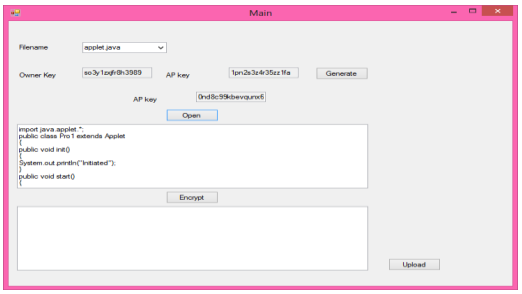

Fig.9. CSP form

File uploading details are shown in Fig.10. The sql management where works in the back end. In the sql management which stores the details about what are the process doing in the front end are stored. File upload and File download details stored with the username, dataowner, filename and also the file path.

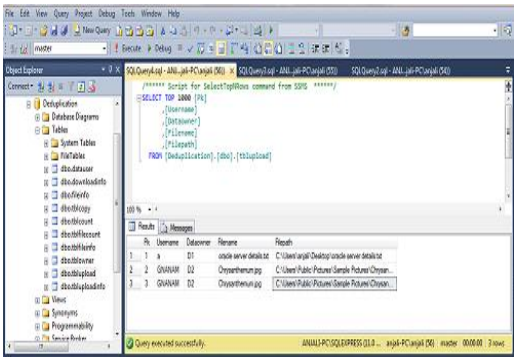

Fig.10. File upload details

In the file count details are shown in the Fig.11. in the cloud storage stored the file only once by the CSP. This is our aim to store the data only once but multiple user can stored the same file in the cloud by many times. CSP works to avoid the duplicate save again as a tag count only.

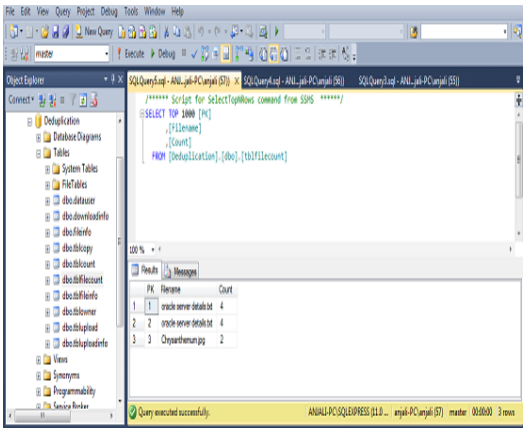

Fig.11.File count

Comparing of the effort and efficient way of the results in the process of deduplication

with T3S-SVM. Using the checksum algorithm its deviation values much lower than the previous algorithm. Using the high performing clusters it can be reduced the report is labeled pairs of average. This training set for T3S-SVM and checksum of this experiment value is better illustrated. Evaluation of checksum with flexible and datasets are shown empirically comparing with basic. This algorithm able to representing increasing the user-friendly with good effectively.

CONCLUSION AND FUTURE WORK:

This technique is helps to upgrade the storage usage level and also it suitable applied to the network information transformation to reducing the lines of bytes that must be sent. In the deduplication, unique a section of information or data, are identified and storing during the progress of analysis. As the process, other information or data is comparing to the storage identical and whenever they are copied occurs, the redundant information or data is substitute with a referral of stored data.

In future work, information comes out as a type of security. This security information can be digitally converted via through Gmail, Websites, spreadsheets, etc.. all are not know about technology. Will be increase the performance and more secure data transfer.

REFERENCES

1. Bellare, M., Keelveedhi, S., & Ristenpart, T. (2013). DupLESS: Server-Aided Encryption for Deduplicated Storage. IACR Cryptology ePrint Archive, 2013, 429.
2. Dropbox, A file-storage and sharing service,(2016). [Online]. Available: http://www.dropbox.com
3. GoogleDrive,(2016).[Online].Available: http://drive.google.com
4. Mozy, Mozy: A File-storage and Sharing Service. (2016). [Online]. Available: http://mozy.com/
5. Douceur, J. R., Adya, A., Bolosky, W. J., Simon, P., & Theimer, M. (2002). Reclaiming space from duplicate files in a serverless distributed file system. In Distributed Computing Systems, 2002. Proceedings. 22nd International Conference on (pp. 617-624). IEEE.
6. Wu, T., Dou, W., Hu, C., & Chen, J. (2017). Service mining for trusted service composition in cross-cloud environment. IEEE Systems Journal, 11(1), 283-294.
7. Yan, Z., Ding, W., Yu, X., Zhu, H., & Deng, R. H. (2016). Deduplication on encrypted big data in cloud. IEEE transactions on big data, 2(2), 138-150.
8. Wu, T. Y., Pan, J. S., & Lin, C. F. (2014). Improving accessing efficiency of cloud storage using de-duplication and feedback schemes. IEEE Systems Journal, 8(1), 208-218.
9. Fan, C. I., Huang, S. Y., & Hsu, W. C. (2012, August). Hybrid data deduplication in cloud environment. In Information Security and Intelligence Control (ISIC), 2012 International conference on (pp. 174-177). IEEE.
10. Z. Sun, J. Shen, and J. M. Yong, “DeDu: Building a deduplication storage system over cloud computing,” in Proc. IEEE Int. Conf. Comput. Supported Cooperative Work Des., 2011, pp. 348–355, doi:10.1109/CSCWD.2011.5960097.

11. Ali, M., Khan, S. U., & Vasilakos, A. V. (2015). Security in cloud computing: Opportunities and challenges. Information sciences, 305, 357-383.
12. J. Pettitt, "Hash of plaintext as key?" (2016). [Online]. Available:http://cypherpunks.venona.com/date/1996/02/msg02013.html
13. C. W. Tsai, C. F. Lai, H. C. Chao, and A. V. Vasilakos, "Big data analytics: A survey," J. Big Data, vol. 2, no. 1, pp. 1–32, 2015, doi:10.1186/s40537-015-0030
14. D. Perttula, B. Warner, and Z. Wilcox-O'Hearn, "Attacks on convergent encryption." (2016). [Online]. Available: http://bit.ly/yQxyvl.
15. M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure deduplication," in Proc. Cryptology—EUROCRYPT,2013,pp.296-312.
16. J. Paulo and J. Pereira, "A survey and classification of storage deduplication systems," ACM Comput. Surveys, vol. 47, no. 1,pp. 1–30, 2014, doi:10.1109/HPCC.2014.134
17. Z. Yan, W. X. Ding, and H. Q. Zhu, "A scheme to manage encrypted data storage with deduplication in cloud," in Proc.ICA3PP2015, Zhangjiajie, China, Nov. 2015, pp. 547–561.
18. Logesh, R., Subramaniyaswamy, V., Vijayakumar, V., Gao, X. Z., & Indragandhi, V. (2017). A hybrid quantum-induced swarm intelligence clustering for the urban trip recommendation in smart city. Future Generation Computer Systems, 83, 653-673.
19. Subramaniyaswamy, V., & Logesh, R. (2017). Adaptive KNN based Recommender System through Mining of User Preferences. Wireless Personal Communications, 97(2), 2229-2247.
20. Logesh, R., & Subramaniyaswamy, V. (2017). A Reliable Point of Interest Recommendation based on Trust Relevancy between Users. Wireless Personal Communications, 97(2), 2751-2780.
21. Logesh, R., & Subramaniyaswamy, V. (2017). Learning Recency and Inferring Associations in Location Based Social Network for Emotion Induced Point-of-Interest Recommendation. Journal of Information Science & Engineering, 33(6), 1629–1647.
22. Subramaniyaswamy, V., Logesh, R., Abejith, M., Umasankar, S., & Umamakeswari, A. (2017). Sentiment Analysis of Tweets for Estimating Criticality and Security of Events. Journal of Organizational and End User Computing (JOEUC), 29(4), 51-71.
23. Indragandhi, V., Logesh, R., Subramaniyaswamy, V., Vijayakumar, V., Siarry, P., & Uden, L. (2018). Multi-objective optimization and energy management in renewable based AC/DC microgrid. Computers & Electrical Engineering.
24. Subramaniyaswamy, V., Manogaran, G., Logesh, R., Vijayakumar, V., Chilamkurti, N., Malathi, D., & Senthilselvan, N. (2018). An ontology-driven personalized food recommendation in IoT-based healthcare system. The Journal of Supercomputing, 1-33.
25. Arunkumar, S., Subramaniyaswamy, V., & Logesh, R. (2018). Hybrid Transform based Adaptive Steganography Scheme using Support Vector Machine for Cloud Storage. Cluster Computing.
26. Indragandhi, V., Subramaniyaswamy, V., & Logesh, R. (2017). Resources, configurations, and soft computing techniques for power management and control of PV/wind hybrid system. Renewable and Sustainable Energy Reviews, 69, 129-143.
27. Ravi, L., & Vairavasundaram, S. (2016). A collaborative location based travel recommendation system through enhanced rating prediction for the group of users. Computational intelligence and neuroscience, 2016, Article ID: 1291358.
28. Logesh, R., Subramaniyaswamy, V., Malathi, D., Senthilselvan, N., Sasikumar, A., & Saravanan, P. (2017). Dynamic

particle swarm optimization for personalized recommender system based on electroencephalography feedback. Biomedical Research, 28(13), 5646-5650.
29. Arunkumar, S., Subramaniyaswamy, V., Karthikeyan, B., Saravanan, P., & Logesh, R. (2018). Meta-data based secret image sharing application for different sized biomedical images. Biomedical Research,29.
30. Vairavasundaram, S., Varadharajan, V., Vairavasundaram, I., & Ravi, L. (2015). Data mining-based tag recommendation system: an overview. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 5(3), 87-112.
31. Logesh, R., Subramaniyaswamy, V., & Vijayakumar, V. (2018). A personalised travel recommender system utilising social network profile and accurate GPS data. Electronic Government, an International Journal, 14(1), 90-113.
32. Vijayakumar, V., Subramaniyaswamy, V., Logesh, R., & Sivapathi, A. (2018). Effective Knowledge Based Recommeder System for Tailored Multiple Point of Interest Recommendation. International Journal of Web Portals.
33. Subramaniyaswamy, V., Logesh, R., & Indragandhi, V. (2018). Intelligent sports commentary recommendation system for individual cricket players. International Journal of Advanced Intelligence Paradigms, 10(1-2), 103-117.
34. Indragandhi, V., Subramaniyaswamy, V., & Logesh, R. (2017). Topological review and analysis of DC-DC boost converters. Journal of Engineering Science and Technology, 12 (6), 1541–1567.
35. Saravanan, P., Arunkumar, S., Subramaniyaswamy, V., & Logesh, R. (2017). Enhanced web caching using bloom filter for local area networks. International Journal of Mechanical Engineering and Technology, 8(8), 211-217.
36. Arunkumar, S., Subramaniyaswamy, V., Devika, R., & Logesh, R. (2017). Generating visually meaningful encrypted image using image splitting technique. International Journal of Mechanical Engineering and Technology, 8(8), 361–368.
37. Subramaniyaswamy, V., Logesh, R., Chandrashekhar, M., Challa, A., & Vijayakumar, V. (2017). A personalised movie recommendation system based on collaborative filtering. International Journal of High Performance Computing and Networking, 10(1-2), 54-63.
38. Senthilselvan, N., Udaya Sree, N., Medini, T., Subhakari Mounika, G., Subramaniyaswamy, V., Sivaramakrishnan, N., & Logesh, R. (2017). Keyword-aware recommender system based on user demographic attributes. International Journal of Mechanical Engineering and Technology, 8(8), 1466-1476.
39. Subramaniyaswamy, V., Logesh, R., Vijayakumar, V., & Indragandhi, V. (2015). Automated Message Filtering System in Online Social Network. Procedia Computer Science, 50, 466-475.
40. Subramaniyaswamy, V., Vijayakumar, V., Logesh, R., & Indragandhi, V. (2015). Unstructured data analysis on big data using map reduce. Procedia Computer Science, 50, 456-465.
41. Subramaniyaswamy, V., Vijayakumar, V., Logesh, R., & Indragandhi, V. (2015). Intelligent travel recommendation system by mining attributes from community contributed photos. Procedia Computer Science, 50, 447-455.
42. Vairavasundaram, S., & Logesh, R. (2017). Applying Semantic Relations for Automatic Topic Ontology Construction. Developments and Trends in Intelligent Technologies and Smart Systems, 48.