Information Security By Reducing Of Cipher Text

V.Amutha

Abstract: -Data deduplication technique is a storage preservation principle to identify and eliminate repeated words and storing single occurrence of words in a file. Data deduplication is used to remove multiple copies of same data. The proposed work has used deduplication technique to reduce the size of plain text and encrypting them by AES principle for securing information. Before applying AES crypto technique, the short listed text is distributed over an mxm matrix for muddling input text. The resultant outcome of muddling is passed to AES encryption system. The resultant is a reduced encrypted text which is applicable for securing as well as reducing the storage of text data.

Keywords: Advanced Encryption Standard(AES),Dedulication,storage,repeated,muddling,reduced,text data.

1. INTRODUCTION

The cryptography is the information managing principle that investigates arithmetic technique to encode and decode information. Cryptography empowers to store delicate data or transmit it over shaky systems like the web with the goal that it can't be used by anybody as recipient. The cryptographic theory is applied in vital data for secret writing. In simple and ordinary cryptography, one key is used both for encryption and decryption in a symmetric system. In asymmetric cryptography, two keys serve for encrypting and decrypting the text input[9]. The deduplication is a technique for reducing duplicate data copies on storage. Data duplication is used to remove multiple copies of same data in a valuable content to consider unique words on the contents. Deduplication is done on the whole content and eliminates repeated chunks of words to maintain a strategic policy on storage devices. Data deduplication has various benefits but the privacy and security is a challenge as the insider and outcast assaults can ruin the touchy information of the clients. Deduplication eliminates large chunks of repetitive data and retains a single copy of word. Noting that in the principle of compression, redundant (repeated) data in a file are encoded them effectively.

2 RELATED WORKS

Deduplication is done at file level means check for the whole data content of the file and eliminates the duplicate files or at block level means check for the same chunk of data content in data files to avoid the duplicate blocks of data[1]. Data deduplication[2] has a various benefits but the privacy and security is a immense challenge as the insider and outsider attacks can spoil the sensitive data of the users. Users normally use the encryption or decryption techniques to provide the security for their data but the conventional encryption techniques are not provided deduplication[2]. In recent times, a technology called thin provisioning is used in storage optimization. When applications run in out of storage, this shared-storage environment relies on on-demand allocation of blocks of data to reduce [2].The approach for deduplication has aimed at reduction in storage space and band-width usage during file transfers. The design depends on multiple metadata structures for deduplication [3]. Only a copy of the duplicate file is retained while others are deleted. The existence of duplicate file is determined from the metadata. The files are clustered into bins depending on their size. They are then seg-mented, [4]deduplicated and are stored. Binning restricts the number of segments and their sizes so that it is optimum for each file. When a user requests a file, compressed segments of the file is sent over the network along with the file-to-segment mapping. In the context of Cloud Service Providers normally use Deduplication[3],which stores only a single copy of each file or block by eliminating redundant data. But providing a secure Deduplication is an uphill task. In this regard, an effort is made to present a survey on the different aspects of Deduplication[5]. Deduplication reduces data volume on disk space and network bandwidth which reduce costs and energy consumption for running storage systems[6].Data deduplication can be applied at nearly every point which data is stored or transmitted in cloud storage [7]. Many cloud providers offer disaster recovery and deduplication to make disaster recovery more effective by replicating data after deduplication [8].

3 PROPOSED ENCRYPITON SYSTEM

1. The proposed system works on data deduplication[3] and cryptographic principle. Initially, words in a file are counted. Then words repeated more than once,say Wiki are identified and recorded the places where it appears. Then the last Win-1 repeating words are deleted from the input text file. A key file is generated for the locations while retrieving back the original file. That is, for every repeated word RWi repeating ki times, ki-1 words are dropped in the input file. Thus for a text file, with words remaining are
2. Words R = total words – RWi ki – 1 times.
3. Then, file with remaining words are distributed in an mxm matrix, where mxm is the dimension of remaining words, for reordering characters. This process fills characters of words row wise and reproduce characters column wise to perplex characters to increase complexity of defining words. Then AES algorithm is applied to encrypt stream of characters to generate cphertext.

- V.Amutha, M.phil Scholar, Dept of Computer Science, Alagappa University,Karaikudi,India,Mobile::9791555240, Email:v.amutha123@gmail.com
- Dr.s.s.Dhenakaran, Professor, Dept of Computer Science,
- Alagappa University,Karaikudi, India, Mobile:9894903755: ssdarvind@yahoo.com

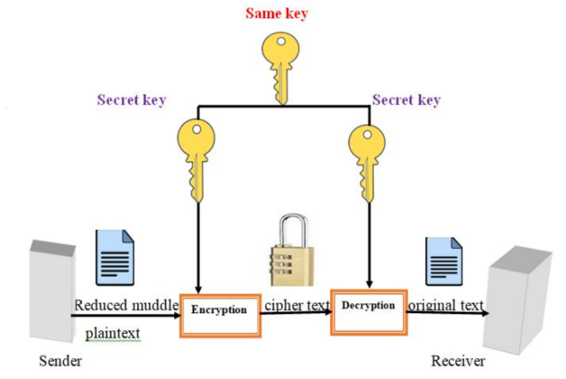

Fig: *Encryption and Decryption*

4 PROCEDURE

1. Read plain text message.
2. Identify words repeated more than one time, count them and store the information of repeated words and its location in a key file
3. Retain repeated words only once and delete further occurrence of words
4. Distribute the outcome of reduced text file in an mxm matrix and retrieve in some other order
5. The muddled characters of text file is input to AES algorithm
6. The resultant is the required ciphertext

5 RESULTS

5.1 Example

plain text: this is an apple, this is an orange. It is fruits category.
Reduced the plain text(Encrypt plain text):
this is an apple, orange.it fruits catogary.

t	h	i	s	-	i	s
-	a	n	-	a	p	p
l	e	,	-	o	r	a
n	g	e	.	-	i	t
f	r	u	i	t	s	-
c	a	t	e	g	o	r
y	.	-	-	-	-	-

Input to the system

t-lnfcyhaegra.in,eut-s--.ie--ao-tg-ipriso-spat-r-

AES Symmetric key
= o?Pen??x?_yr}
Encrypted message

ƒm?V?hämàb6f?ôN1<å?ÅgçÖ/Üéb5[?è%ö4uºm?
½D?Imz?dêÄL

5.2 Example

Input
The public key cryptography is a set of techniques that allows people to communicate secret information over completely open communication channel. Difficulties in the management and distribution of secret keys led to the notion that some key material need not be secret hence the use of public key cryptosystems were born . Rather than using the same key to encrypt and decrypt data , public key cryptography system uses a matched pair of encryption and decryption keys . Each key perform at transformation on the data. The decryption transformation is the inverse of encryption transformation. The public key need not be kept secret and in fact may be widely available only its authenticity is required to guarantee that the specific user is indeed the only party who knows the corresponding private key.

Reduced the plain text
The public key cryptography is a set of techniques that allows people to communicate secret information over completely open communication channel. Difficulties in the management and distribution keys led notion some material need not be hence use cryptosystems were born Rather than using same encrypt decrypt data , system uses matched pair encryption decryption Each perform at transformation on inverse kept fact may widely available only its authenticity required guarantee specific user indeed party who knows corresponding private

cipher text
n?F¢ ¢?M??6²?t.?í>?úS???<½ê,?a%8"? U?éh¢g? o|K?
>Y'?u??äC???öïC"; |H?O?)åC? ??? É}{?¢?H ?qMú
zàÄPèàm<? Ä ??Å@PPÜ???d@?`#?|?8?? á??ár·òC6ï-
??%4g?òZÉá;??¡Ç?°rL??"n¥?|ä?5 »W
?c?:étóèEòpp?²2|â=ì?æ`?~6 ·#,??$ï?-??<« \?? ú÷#^²?n4?²
?/p?nzó? Ñ¢Ñ¡hD> ï4.??&
b,iäFd?i?W.v/Ä??9?d??y? {:²5?@/ äü?9Y|3?Q??
m??ç?Oö??? ú?tZ; «%?k?²e?G "S£C<ÿ3µw$Ä?Vy%r?j-ó&
?#ÿx??e(????|r²? ? B°[tlq?«M?? □$ÿh }mnc?¥ °?
?Na-R?.?{½ v?u$?ß???99?ìbKî v?¡qNé?[²Ni!?äb
?¿»h?æ£cl?KY²_|??, ë7?ÿX?Ü??Ç°y?,?s² ¼µÄüs?¢il«?
?ÿ¢?,?dj%ñä#?~?£c?> s????£?ê?

Decrypted message
The public key cryptography is a set of techniques that allows people to communicate secret information over completely open communication channel . Difficulties in the management and distribution keys led notion some material need not be hence use cryptosystems were born Rather than using same encrypt decrypt data , system uses matched pair encryption decryption Each perform at transformation on inverse kept fact may widely available only its authenticity required guarantee specific user indeed party who knows corresponding private

Actual plain text
The public key cryptography is a set of techniques that allows people to communicate secret information over completely open communication channel. Difficulties in the management and distribution of secret keys led to the notion that some key material need not be secret hence the

use of public key cryptosystems were born . Rather than using the same key to encrypt and decrypt data , public key cryptography system uses a matched pair of encryption and decryption keys . Each key perform at transformation on the data. The decryption transformation is the inverse of encryption transformation. The public key need not be kept secret and in fact may be widely available only its authenticity is required to guarantee that the specific user is indeed the only party who knows the corresponding private key.

6 CONCLUSION

In this paper, the duplication framework is utilized to improve the unwavering quality of data for reducing size of content as well to reduce transmission time on network. The principle of deduplication is invoked by an obvious method and encryption of text by standard AES algorithm. The result is met with requirement level of implementing the task.

7 ACKNOWLEDGEMENTS

This article has been written with the financial support of RUSA – Phase 2.0 grant sanctioned vide Letter No.F.24-51/2014-U, Policy(TNMulti-Gen), Dept. of Edn. Govt. of India, Dt.09.10.2018.

8 REFERENCES

[1] Philipp C. Heckel (2013, May 20). "Minimizing remote storage usage and synchronization time using deduplication and multichunking,"[Online]. Available: http://blog.philippheckel.com!

[2] Q. He, Z. Li, X. Zhang, "Data deduplication techniques,"Future Information Technology and Management Engineering (FITME)," vol. I, pp. 430-433, 2010.

[3] Maddodi.S, Attigeri G.V, Karunakar. A.K, "Data Deduplication Techniques and Analysis," Emerging Trends in Engineering and Technology (ICETET), pp 664 - 668, IEEE, 2010.IEEE, 2011.

[4]]Xian Chen, Wenzhi Chen, Zhongyong Lu, Peng Long, Shuiqiao Yang, Zonghui Wang, A Duplication-Aware SSDBased Cache Architecture for Primary Storage in Virtualization Environment, IEEE Systems Journal, Volume: 11,issue:4, Dec. 2017

[5] Benjamin Zhu, Kai Li, and Hugo Patterson, "Avoiding the Disk Bottleneck in the Data Domain Deduplication File System," Proc. of the USENIX File And Storage Technologies, 2008.

[6] D. T. Meyer, W. 1. Bolosky (2012), " A Study of Practical Deduplication,"[Online]. Available:http://static.usenix.Org

[7] M. Dutch, "Understanding data deduplication ratios," In SNIA Data Management Forum, 2008.

[8] SNIA, "Advanced Deduplication Concepts," 2011.

[9] S.S.Dhenakaran,E.R. Naganathan, “A New Approach to Multiple Symmetric Keys”, International Journal of Computer Science and Network Security, VOL.7 No.6, June 2007, pp. 254-259