Securing Twitter Data Using Srb21 Phase I Methodology

C Bagath Basha, S Rajaprakash

Abstract: Today's world has witnessed the buildup of information through social media like Twitter and Facebook, that are growing hugely day by day. Victimization these social media, users tweet or post on several topics from anywhere within the world through the net. Tweets area unit accustomed analyze positive and negative messages to drive polarity scores and additionally to predict future trends. These polarity scores might be extracted from Twitter, albeit polarity score knowledge area unit placed underneath security. The information will be simply hacked, and dynamical the score results will cause immeasurable problems, like touching the worldwide economic standing, company brands, and therefore the reputations of companies. To tackle these problems, Daniel Bernstein planned to the implementation of Salsa, that has quicker encryption because of the district round with better data security. In the current work, a novel methodology is proposed by modifying the Salsa20/4 to further enhance the security of the polarity scores, which is an important requirement of the current world. The proposed methodology has two stages. The first stage has secret key. The second stage is prime number secret key in N-1 steps which are interchange the secret key value and prime number value. The proposed methodology has good security while to comparing to Salsa20/4 methodology.

Index Terms : Encryption, Polarity, Salsa, Security, SRB21, Twitter, Tweets

1. INTRODUCTION

During the last decade, the globe has witnessed the increase of a large social media knowledge hold on Twitter and Facebook. They utilized by multi-users to tweet on any topic in social media through the net. These tweets type of the idea for the analysis of polarity scores. The analyzed polarity scores would like security as a result of the Twitter application is one in all the foremost noted sources for public info within the world. Thus, the output of the analyses plays an important role in creating the choices on the choice of any merchandise knowing the perception of its quality by the general public sector. In default, Twitter doesn't have such smart security of that data. To analyze the data and secure them, Daniel Bernstein introduces Salsa, a family of stream ciphers, which focuses on encryption work reasonable for a wide scope of utilization. Salsa20 handles the secret keys of 128 bits or 256 bits, but Daniel Bernstein recommends secret key of 256 bits. Salsa20 with 20 rounds, which are proposed by Daniel Bernstein [1], is faster than Advanced Encryption Standard (AES) and provides better security [2]. The Salsa has reduced-round versions from Salsa20/20 to Salsa20/12, Salsa20/8, Salsa20/7, Salsa20/6, Salsa20/5, and Salsa20/4, these rounds are decreased from 20 rounds. Salsa20/4 operates only 3 rounds and each round has district rounds. Salsa20/4 encryption speed is faster when compared to other Salsa20 variants. Salsa family mainly focused on encryption speed only not data security. To overcome the drawbacks of Salsa family stream ciphers for securing the data, a novel algorithm Somasundaram Rajaprakash and Bagathbasha 21 (SRB21) Phase I methodology are proposed in this work.

2 RELATED WORK

Introduced the related cipher attack by Hongjun in 2002 and applied to the Salsa20 stream cipher. Salsa20 used flexible rounds that were reduced the round version of Salsa20 and key schedule rounds which are an independent number of rounds. The results showed that the related cipher attack might be applied to stream ciphers [1]. Discussed the Salsa20/4 and proposed chaotic Salsa. These algorithms were used to compare the speed and diffusion level. The proposed chaotic Salsa had faster than the Salsa20/4, but the diffusion level was the same [2]. Analyzed and improved the correlation attack between Salsa20/9 and ChaCha8. They were tried to analyze the possibility of reducing the complexity of the existing attack, but a correlation attack was not useful for analyzing the ChaCha8 [3]. Improved the related cipher attack on Salsa20 family. This family has best attack model have Salsa20/12 and Salsa20/8 [4]. Analyzed the Salsa and ChaCha reduced rounds, then they proposed novel algorithm is Probabilistic Neutral Bits and this algorithm is faster than the existing attack [5]. Improved the attack on 128 key bits of Salsa7 and ChaCha6 [6]. Improved the attack of ChaCha7 and Salsa8 with proper choice of IVs [7]. Significantly improved the time complexities of Salsa in 7 rounds and ChaCha in 6 rounds [8].

The talk data using Salsa20 stream cipher cryptography algorithms. This algorithm is used to analyze the processing time of both encryption and decryption time is silently fast, first packet takes a few milliseconds and next packet takes one millisecond. The results show the performance of Salsa20 is best in the data security [9]. Introduced a novel cost-effective technique to eliminate vulnerabilities found in it and to provide the 3D security such as authentication of users, encrypted data during transit and encrypted data at rest. They assumed that increasing the buffer size, reduces the iterations of the loop and consequently reduces the overall encryption time on ChaCha20 [10]. Proposed Salsa20/20 design which is faster than AES and also recommended the design of all Salsa Variants. Salsa20 family gives only importance for encryption speed [11].

- *C Bagath Basha, Research Scholar, CSE, Aarupadai Veedu Institute of Technology, Vinayaka Mission's Research Foundation, Chennai, Tamil Nadu, India, PH-+91-9976341498. E-mail: chan.bagath@mail.com*
- *Dr. S. Rajaprakash, CSE, Aarupadai Veedu Institute of Technology, Vinayaka Mission's Research Foundation,Chennai,Tamil Nadu, India, PH-+91- 9942005441. E-mail: srajaprakash_04@yahoo.com.*

3 METHODOLOGY

This work deals with the data of a particular area collected from Twitter. The data are used to classify the tweets using Rstudio on Twitter. These tweets are used to analyze negative and positive tweets to make polarity scores. The result of the polarity scores could be extracted from Twitter. These data files are converted into a matrix and the files are applied to the proposed methodology SRB21 first phase with the matrix of order N by N. SRB21 first phase methodology has two key in the encryption process. The proposed methodology has two stages. The first stage has secret key, and the second stage is prime number secret key in N-1 steps which are interchange the secret key value and prime number value.

3.1 Twitter

Twitter is one of the world famous social media in which users tweet on any topic in social media at any time with anyone and anywhere in the world through the Internet. These tweets are used to analyze the polarity scores in Twitter data. Polarity scores have both positive and negative tweets which are represented by the symbols '+', '-' respectively. These tweets are collected as words from Twitter. The collected words or tweets are converted to the polarity scores.

The positive and negative sample words are given below:

- Positive words - 'good, lucky, bless, like, interest, happy', etc...
- Negative words - 'bad, don't, not, sad, won't, bore', etc…

Example - tweets,

1. First tweet: "I am not interested to see this movie because very boring".
2. Second tweet: "I like this movie to see both theatre and TV".
3. Third tweet: "I not like this movie so not interested to see this movie".

Table 1 gives two examples for creating the reviews, first tweet "I am not interested to see this movie because very boring" (Class: Negative), these words have two negative words and one positive word, so it is considered to be negative tweets. Second tweet "I like this movie to see both theatre and TV" (Class: positive), these words have one positive word and no negative word, so it is positive tweets. Third tweet "I not like this movie so not interested to see this movie". (Class: Neutral), these words have two negative words and positive word, so it is neutral tweets. This polarity scores result data could be extracted in the form of a matrix as shown in Table 2 [12].

3.2 SRB21 Phase I Methodology

SRB21 methodology consists of three steps.
Step 1: Extracting the data from Twitter.
Step 2: Analyzed twitter data are stored in the matrix K.
Step 3: Operation are used to exchange the secret key and prime number secret key of the matrix.

TABLE 1
POLARITY SCORE

Vocabulary	Analyzing First Tweet	Analyzing Second Tweet	Analyzing Third Tweet
I	1	1	1
Am	1	0	0
Not	-1	0	-2
Interested	+1	0	+1
To	1	1	1
See	1	1	1
This	1	1	1
Movie	1	1	1
Because	1	0	0
Very	1	0	0
Boring	-1	0	0
Like	0	+1	+1
Both	0	1	0
Theatre	0	1	0
And	0	1	0
TV	0	1	0
So	0	0	1
. .			
Class label	Negative (-)	Positive (+)	Neutral

TABLE 2
DATA SET

Dataset	#Positive	#Negative	#Neutral
RAJINI	519	4	477
AKSHAY	360	2	638
DHONI	538	97	365
KOHLI	493	25	482

$$S=(n,R)$$
$$where\ R=n^2-n+K,$$
$$n\geq 3$$

```
$if\ R > N\ Then$
 $R_1 = R - N$
 $if\ R_1 > N\ Then$
 $R_2 = Add\ R_1\ Digit\ Numbers$
 $if\ R_2 > N\ Then$
$R_3 = Add\ R_2\ Digit\ Numbers$
 $if\ R_3 > N\ Then$
 $swap\ (n, R_3)$
 $break$
$else$
$swap\ (n, R_2)$
$else$
$swap\ (n, R_1)$
$else$
$swap\ (n, R)$
$if\ R\%\ 2 \neq 0\ \&\&\ R\%\ 3 \neq 0$
$R\%\ 5 \neq 0\ \&\&\ R\%\ 7 \neq 0\ Then$
$n = n+1$                                        (1)
$else$
$STOP$
$where\ N\ is\ Order\ of\ Matrix$
```

R represents the secret key (n) and prime number secret key (K) operation of the matrix, and N is order of matrix

3.3 Working of SRB21 Phase I Methodology

- The proposed SRB21 Phase I methodology is developed from modifying the Salsa20/4.

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \\ 16 & 17 & 18 & 19 & 20 \\ 21 & 22 & 23 & 24 & 25 \end{pmatrix}$$

Where A is the analyzed twitter data matrix

Phase 1: By applying "(1)"
Step 1: n=3, K=37, N = 25

$S = (n, R)$

$where\ R = n^2 - n + K,$

$n \geq 3$

R = 3^2 – 3 + 37 = 43
S = (3, 43)
R>N = 43>25 => R_1 = 43-25 = 18
R_1 > N = 18 > 25 => Swap (n, R_1) => Swap (3, 18)
Interchange the value of the 3rd cell element and 18th cell element.

$$P1EM = \begin{pmatrix} 1 & 2 & 18 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \\ 16 & 17 & 3 & 19 & 20 \\ 21 & 22 & 23 & 24 & 25 \end{pmatrix}$$

Where P1EM is Phase 1 Encryption Matrix

Step 2:

```
$if\ R\%\ 2 \neq 0\ \&\&\ R\%\ 3 \neq 0\ \&\&$
$R\%\ 5 \neq 0\ \&\&\ R\%\ 7 \neq 0\ Then$
                    $n = n+1$
$else$
                    $STOP$
```

R is Prime => n = n+1 => n=4

R = 4^2 – 4 + 37 = 49

S = (3, 49)

R>N = 49>25 => R_1 = 49-25 = 24

R_1 > N = 24 > 25 => Swap (n, R_1) => Swap (4, 24)

Interchange the value of the 4th cell element and 24th cell element.

$$P1EM = \begin{pmatrix} 1 & 2 & 18 & 24 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \\ 16 & 17 & 3 & 19 & 20 \\ 21 & 22 & 23 & 4 & 25 \end{pmatrix}$$

Step 3:

```
$if\ R\%\ 2 \neq 0\ \&\&\ R\%\ 3 \neq 0\ \&\&$
$R\%\ 5 \neq 0\ \&\&\ R\%\ 7 \neq 0\ Then$
                    $n = n+1$
$else$
                    $STOP$
```

- R is not prime => STOP

4 DISCUSSION & CONCLUSION

The current work has to classify the tweets using Rstudio on Twitter. These tweets are used to analyze negative and positive tweets to make polarity scored. The polarity scores result data could be extracted from Twitter. Extracted data are prone to have security issues. Hacking the data easily and changed the score results lead to lots of issues like affecting the economic status, company brand, and reputation of firms. Salsa20/4 is faster encryption because of the quarter round, which is also better security. The novel methodology is proposed by modifying the Salsa20/4 to enhance further the security of the accumulated data. Salsa methodology has quarter round process, each round process has quarter round by column operation. This methodology has only focuses on encryption speed not security. SRB21 phase I methodology has N round process; each round has interchange the secret key and prime number secret key of the matrix. The proposed methodology has good security because of prime number secret key. In the future, the co-prime will be add more operations of the data security in the SRB21 Phase II methodology for upcoming journals.

REFERENCES

[1] Z. Shao and L. Ding, “Related-Cipher Attack on Salsa20,” Proc. Fourth Inter. Conf. on Computational and Information Sciences, pp.1182 1185, 2012.

[2] M. Almazrooie, A. Samsudin, and M. M. Singh, “Improving the Diffusion of the Stream Cipher Salsa20 by Employing a Chaotic Logistic Map,” J. of Information Processing Systems, vol. 11, no. 4, pp. 310-324, 2015.

[3] P. Yadav, I. Gupta, and S. K. Murthy, “Study and Analysis of eSTREAM Cipher Salsa And ChaCha,” Proc. Second IEEE Inter. Conf. on Engineering and Technology, 2016.

[4] L. Ding, “Improved Related-Cipher Attack on Salsa20 Stream Cipher,” IEEE Access, vol. 7, pp. 30197-30202, 2019.

[5] S. Dey and S. Sarkar, “Improved analysis for reduced round Salsa and Chacha,” Discrete Applied Mathematics, vol. 227, pp. 58-69, 2017.

[6] K. K. C. Deepthi and K. Singh, “Cryptanalysis of Salsa and ChaCha: Revisited,” Proc. ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, pp. 324 338, 2018.

[7] S. Maitra, “Chosen IV cryptanalysis on reduced round ChaCha and Salsa,” Discrete Applied Mathematics, vol. 208, pp. 88-97, 2016.

[8] A. R. Choudhuri and S. Maitra, “Significantly Improved Multi-bit Differentials for Reduced Round Salsa and ChaCha,” IACR Transactions on Symmetric Cryptology, vol. 2016, no. 2, pp. 261-287, 2016.

[9] D. Afdhila, S. M. Nasution, and F. Azmi, “Implementation of Stream Cipher Salsa20 Algorithm to Secure Voice on Push to Talk Application,” Proc. IEEE A. Paci. Conf. on Wireless and Mobile, pp. 137 141, 2016.

[10] R. R. Parmar, S. Roy, D. Bhattacharyya, S. K. Bandyopadhyay, and T. Kim, “Large-Scale Encryption in the Hadoop Environment: Challenges and Solutions,” IEEE Access, vol. 5, pp. 7156-7163, 2017.

[11] D. J. Bernstein, “The Salsa20 family of stream ciphers,” New Stream Cipher Designs: The eSTREAM Finalists, Lecture Notes in Computer Science, M. Robshaw and O. Billet, eds., Berlin: Springer, 2008.

[12] C. Bagath Basha and K. Somasundaram, “A Comparative Study of Twitter Sentiment Analysis Using Machine Learning Algorithms in Big Data,” Inter. J. of Recent Technology and Engineering, vol. 8, pp. 591-599, 2019.