Cryptic Mining for AVK Based Cryptosystem and Client Side Encryption Perspective

Shaligram Prajapat* Pulkit Vaishnav* and R. S. Thakur**
*International Institute of Professional Studies
shaligram.prajapat@acm.org , pulkitvaishnav35@gmail.com
**M.A.N.I.T. Bhopal, India
ramthakur2000@yahoo.com

Abstract: This work presents parametric versions of symmetric cryptic algorithms that emphasises on generation and usage of key based on parameter only. The key construction process has been extended for generation of alphanumeric keys apart from numeric one with exploiting domain of parameter selection from personnel information logs. The hypothesis of parametric but variable key for cryptic model (usable for Automatic Variable Key) tested with the freedom of user for the parameter selection and variation in the parameters according to comfort instead of using series, recurrence relation or location information [8, 21]. The paper investigates parametric model in the light of Association Rule Mining for such cryptosystem [15]. Useful inferences and results from testing results helps in auditing of AVK based cryptic algorithm and identifies power of using a large number of parameters for secure information exchange.

Keywords: Parameters, Symmetric Key, Cryptic Algorithm, Automatic Variable Key (AVK), secure information exchange.

Introduction

Secure information exchange over the public-network is a pertaining research challenge in today's context. The problem continues to aggregate with increasing volume of network traffic, which is evident from several recent researches. As mentioned frequently in these reporting, Security of these data or information basically refers to the protection of the data against intentional modification, loss or damage and fabrication of data, and / or deliberate disclosure of data to unauthorized persons or miscreants. Further, "Perfect Secrecy "with crypto system is the demand of communicating entities, wherein after cryptograms captured by an intruder, has the posterior probability of these cryptograms is the same as the prior probability of the same messages before the interception [1, 2].The essential requirement is that if the number of messages is finite then same number of possible keys should be there. For stream ciphers it is also necessary that if message (plaintext information) is constantly generated at a given rate, the key must be generated at the corresponding same or a greater rate [2, 3, 4, and 9]. In the classical article on Information Security, Shannon, has already stated that perfect security can be achieved only when the key is made to vary from session to session and/or data to data[6,7].This seed instantiates several ideas and implementations for Automatic Variable Key based Schemes for information exchange. Realization of the time variable key is difficult to achieve as such key, from time-to-time, must be communicated between the sender and the receiver [1, 3]. The AVK for a session between Alice and Bob is to be send initially as K_0, and they exchange data D_0. The key is now variable and after every transmission it changes dynamically such that: with initial key $K_0 = \textit{initial secret key}$ the future keys can be constructed by $K_i = K_{i-1} \oplus D_{i-1}, \forall\ i > 0$, where D_{i-1} *and* K_{i-1} *are data and key* of $(i-1)$*th session*. The variable key as suggested if implemented, the repetition of patterns will not result, unlike in the normal mode.AVK depends on the data sent previously. There is no guarantee that the previously sent data may not be stolen. This needs further investigations for proper utilization of the time-variant key [2, 4]. In order to solve the problem, there should be some technique so that the previously sent data is protected or there are some storage media where a replica of the data will be stored. The circuit will behave as a multiplexer where the control signal will govern whether there is need to swap-in of data from data-store or not [5, 6, 7, 8].

Designing a good AVK based cryptosystem, not only requires Cryptic Algorithms and Key Management protocols, but it is also necessary to test the design under cryptic pattern discovery and mining.The parametric Fibo-Q model uses parameter n as input for computation of key. Sparse approach uses location (i, j) as two parameters for computation of key. The limitations identified with these schemes were that the keys used in these approaches were numeric keys. In Fibo-Q approach, key is varied with parameter = *n* and computer *f (n), f (n-1)* and *f (n+1)* and uses any one of them to n for next key (if *n*< 35 other wise choose random n from mutual agreement of sender and receiver)[16, 20], in Sparse approach key is varied with location parameters p_{ij} = location (i, j) (applicable for moving devices, next location coordinates will produce next key) [8, 21].

Rationale of the study

The primary purpose of this research is to provide a careful description of the AVK cryptosystem that are attempting to take the advantage of the scheme was identified that only parameters can be shared among communicating entities. Say the Key K is constituted with three parameters as K=f (p_1, p_2, p_3). For simplicity, we assume the parameters from personal information of communicating entities. For generation of alphanumeric keys let, the parameters like p_1= part of vehicle number, p_2 = nickname, p_3 = date of birth of spouse, then possible key samples may be *mp09t!nku020284*, *mp09b0b020284*, *mp09t!nku020286* etc. The AVK keys so generated are {$p_1p_2p_3$, $p_1p'_2p_3$, $p_1p_2p'_3$, $p'_1p_2p'_3$..}, we assumed here that parameters p_1, p_2, p_3 are operated by concatenation operation and by changing parameters values variable keys can be obtained. Further it is also assumed that domain of parameters p_1, p_2, p_3 are binary in nature, in reality this can have cardinality >2 .To Analyse this alphanumeric mix key construction using above scheme, it has been analyzed from hacker's point of view in subsequent sections. The parameter used among possible parameter set by different sessions may be found for association rules [9, 10]. The study will also test hypothesis of the above schema, detailed opinions from common user survey is taken for parametric model described in the next section with Association rule extractions [9,10] for cryptic mining and frequent patterns discovery for the keys of parametric AVK model[14,15,17].

Experimental Setup for Association Rule Analysis

The universe of study is professionals, student and users of social media. Population space (From LinkedIn, Google Plus, Face book and Google group of alumni) is target population for this study is all IT users. Each and every aspect of social and connected professional users in IT and academics were investigated for key construction behaviour and choosing parameters. On the sample size over 1000 sample size useful and meaningful samples were 100 after pre-processing phase. In data collection step, to test the parameter usage and behaviour of changing parameter has been analyzed with online survey. The list of 23 Questions to know about potential parameters which are used by common users, included in Survey: (* indicates mandatory response).The count of responses considered for analysis= 100 (On five point lickert scale) with minimum 1 to max 5 in the format:

(I never use it) 1 2 3 4 5 (*I always use it*)

Elucidation of survey on Parameter Sets: P= {p1, p_2... p22}.The parameters p_1 to p_{22} , used by common users for securing parametric communication over public network for key constructions are:

Table 1.Comparison of network cost based on number of nodes

Parameter	Description of Parameter-Type
p_1	First name/last name/nick name
p_2	DOB/Anniversary date
p_3	Public name in any form (as-is, reversed, capitalized, doubled, etc.)
p_4	Do you use your spouse's or child's name?
p_5	Information easily obtained about you. (License plate numbers, house numbers)
p_6	PAN number/Adhere number/Social Security Number/Passport number.
p_7	Telephone numbers/Mobile Number
p_8	Automobile, Vehicle number
p_9	Password of all digits/all the same letters
p_{10}	Word contained in (English or foreign language) dictionaries/spelling lists/ other lists of words.
p_{11}	Do you use a password shorter than six characters
p_{12}	Password with mixed-case alphabetic
p_{13}	No alphabetic characters, e.g., digits or punctuation.
p_{14}	Key that is easy to remember, so you don't have to write it down.
p_{15}	A Key that you can type quickly, without having to look at the keyboard. (Making harder for someone to steal your password by watching over your shoulder)
p_{16}	DOB of your girlfriend /boy friend/ spouse.
p_{17}	Station code/country code/area code.

p_{18}	Current institute name/office name (in some form).
p_{19}	Use of first letter of each word from a line of a song /Book or poem.
p_{20}	Part of residential address/ city name/state name/ country name.
p_{21}	ID numbers provided by different institutions/organizations such as roll number, subscription number, exam id.
p_{22}	Non personal information for constructing a field /parameter of password/Key.

The responses received are collected through Google form, shared over professional network (LinkedIn), social network (face book, Google plus) and Google groups of aluminous of DAVV to get real trends. The baseline from hacker's perspective on the parametric AVK model may be used to find out, which parameters are favourable for key constructions? Which among those are the most prominent or frequent? Is there any association among these parameters? But with large number of pair-wise determination of correlation will be costly. Finding correlation will be easy in case of less number of parameters. One possible tool to analyze the survey result is use SPSS, that provides grouping of parameter sets into factors , to find out the favourable parameters for key construction, The result shown in Table 2, that presents the probability of making choices of parameter sets. The tool also provides correlations among these parameters, but in case of more parameters the correlation among parameters; association rule generation requires less effort for inference. Thus, in case of large number of parameter sets, mining algorithms are suitable. Mining over large number of responses out of these 22 parameters, group of favourable parameters for key construction with respective probability of the frequent parameters are given in Table 2.

Table 2: Group of Frequent Set of Related Parameters with Probability

Frequent Parameters Used	% of time choice has been made	Choice/Preference
p_6,p_7,p_8	19.76%	First
p_{19},p_{20}	13.90%	Second
p_4,p_{15},p_{18}	10.68%	Third
p_{10},p_{11},p_{13}	09.21%	Fourth
p_9,p_{17}	07.87%	Fifth

Frequent Patterns Generated for Parametric AVK model

Traditional Apriori algorithm [14] is applied on response data with variable number of parameters. So to analyse the responses, C- implemented Apriori algorithm is applied on response data with tab delimited text file format (as the input file) and the result is also in text file. The algorithm used in two ways: first, find the frequent parameter set by pruning the candidate parameter set with user defined support threshold value [15] .Later for rule generation the minimum support and confidence are varied over range (1% to 100%).With respect to variable support, variation in no. of frequent parameters, the time consumed in seconds for generation of association rules and corresponding file size has been written in Table 3.

The association rules based on various parameters with support (10%, 20%, 30 %....100%) and Confidence (80%, 90%, 100%), and some are discussed in detail.

1. Support 90% ,Confidence 90% number of rules = 1 ; (Only parameter P_{12})
2. Support 90% ,Confidence 80% number of rules=1; (Only parameter P_{12})
3. Support 80% ,Confidence 90% number of rules=1; (Only parameter P_{12})
4. Support 80% , Confidence 80% number of rules =1; Only parameter P_{12})
5. Support 70% , Confidence 90% number of rules=3;(Rule-1 parameter P_{12}, rule-2 P_{12} associated with P_{14} and rule-3 parameter P_{12} associated with P_{13})
6. Support 70% , Confidence 80% number of rules=3 ;(Rule-1 parameter P_{12}, rule-2 P_{12} associated with P_{14} and rule-3 parameter P_{12} associated with p13)
7. Support 60% , Confidence 90% number of rules=4 ; (P_{12} , $P_{12} \leftarrow P_{15}$, $P_{12} \leftarrow P_{14}$, $P_{12} \leftarrow P_{13}$)
8. Support 60% ,Confidence 80% number of rules=7 ; (P_{12} , $P_{14} \leftarrow P_{15}$, $P_{13} \leftarrow P_{15}$, $P_{12} \leftarrow P_{15}$, $P_{12} \leftarrow P_{14}$, $P_{12} \leftarrow P_{13}$,$P_{14} \leftarrow P_{15}$ P_{12})
9. Support 50% , Confidence 90% number of rules=7; (P_{12}, $P_{12} \leftarrow P_{15}$, $P_{12} \leftarrow P_{14}$, $P_{12} \leftarrow P_{13}$, $P_{12} \leftarrow P_{15}$ P_{14}, $P_{12} \leftarrow P_{15}$ P_{13}, $P_{12} \leftarrow P_{14}$ P_{13})

Support 50% , Confidence 80% number of rules=7; (P_{12}, $P_{12} \leftarrow P_{15}$,$P_{12} \leftarrow P_{14}$,$P_{12} \leftarrow P_{13}$,$P_{12} \leftarrow P_{15}$ P_{14}, $P_{12} \leftarrow P_{15}$ P_{13}, $P_{12} \leftarrow P_{14}$ P_{13})

Table 3: Association Rule with Different Support and Confidence

S.No.	Support (%)	Confidence (%)	Time(Sec.)	File Size	No. of rules
1	1	80	108.27	3.5GB	62151312
2	1	90	70.94	3.5GB	62050814
3	1	100	71.91	3.5GB	62050212
4	2	80	27.21	1.37GB	25389899
5	2	90	29.34	1.36GB	25289401
6	2	100	28.41	1.36GB	25288799
7	3	80	5.67	288MB	5619616
8	3	90	5.63	284MB	5519118
9	3	100	5.62	284MB	5518516
10	4	80	0.46	170.4MB	407512
11	4	90	0.30	13.3MB	307014
12	4	100	0.30	13.3MB	306412
13	5	80	0.15	5.92MB	145504
14	5	90	0.09	1.82MB	45006
15	5	100	0.06	1.80MB	44404
16	10	80	0.02	67.4KB	1995
17	10	90	0.02	24.5KB	734
18	10	100	<0.01 Sec	4.24KB	132
19	20	80	<0.01 Sec	3.79KB	118
20	20	90	<0.01 Sec	1.10KB	35
21	20	100	<0.01 Sec	0KB	NIL
22	30	80	<0.01 Sec	1.24KB	39
23	30	90	<0.01 Sec	445Bytes	14
24	30	100	<0.01 Sec	NIL	NIL
25	40	80	<0.01 Sec	477Bytes	15
26	40	90	<0.01 Sec	285Bytes	9
27	50	80	<0.01 Sec	346Bytes	11
28	50	90	<0.01 Sec	216Bytes	07
29	60	80	<0.01 Sec	208Bytes	07
30	60	90	<0.01 Sec	112Bytes	04
31	70	80	<0.01 Sec	81Bytes	03
32	70	90	<0.01 Sec	81Bytes	03
33	80	80	<0.01 Sec	24Bytes	01
34	80	90	<0.01 Sec	24Bytes	01
35	90	80	<0.01 Sec	24Bytes	01
36	90	90	<0.01 Sec	24Bytes	01

In Table 3, Time consumed during mining of association rules has been presented. Entries of table show the Variation of Support and Confidence and Time for generation of Association Rules (with 22 Parameters). The information of Table 3 is presented in graphical form in Figure 1.

Observation for Association Rules Mining (ARM)

1. The Rules show the possibility of picking the parameter combination and their correlation.
2. For lower support less than 10 % and higher confidence 80% and above the numbers of rules are huge and item set cardinality are outsized. This indicates the possibility of picking the similar parameters for key by a small number of people, and the combinations are too large to evaluate.
3. High support greater than 10% and less than 40% and higher confidence 80 % and above Number of Rules are rational but need rigorous efforts to evaluate but still its usefulness is in doubt as due to high support many parameters are out of consideration.
4. Higher support greater than 40 % and Higher Confidence greater than or equal to 80 % Rules are very few and have limited parameters, easy to evaluate but at the same time chances of finding the natural rules are high.

Enabling the Cryptanalyst with advanced tools is always in demand for identification of weakness and further improvement of cryptosystem. In polynomial time, Cryptanalyst is interested to extract useful guess for detecting original information, from huge corpus of ciphers. Cryptanalyst may have captured large database and corpus containing variety of ciphers and hash files. When a cipher text is inserted into this dataset, it might be mixed within other ciphers generated from various other schemes including variations in key size, protocol, type of ciphers generation algorithm, degree of exposures of information about key space and lot of other information related to plaintext, cipher text and relationship between them. The cryptanalyst may develop a mechanism that will classify/sort/ group according to cipher type.

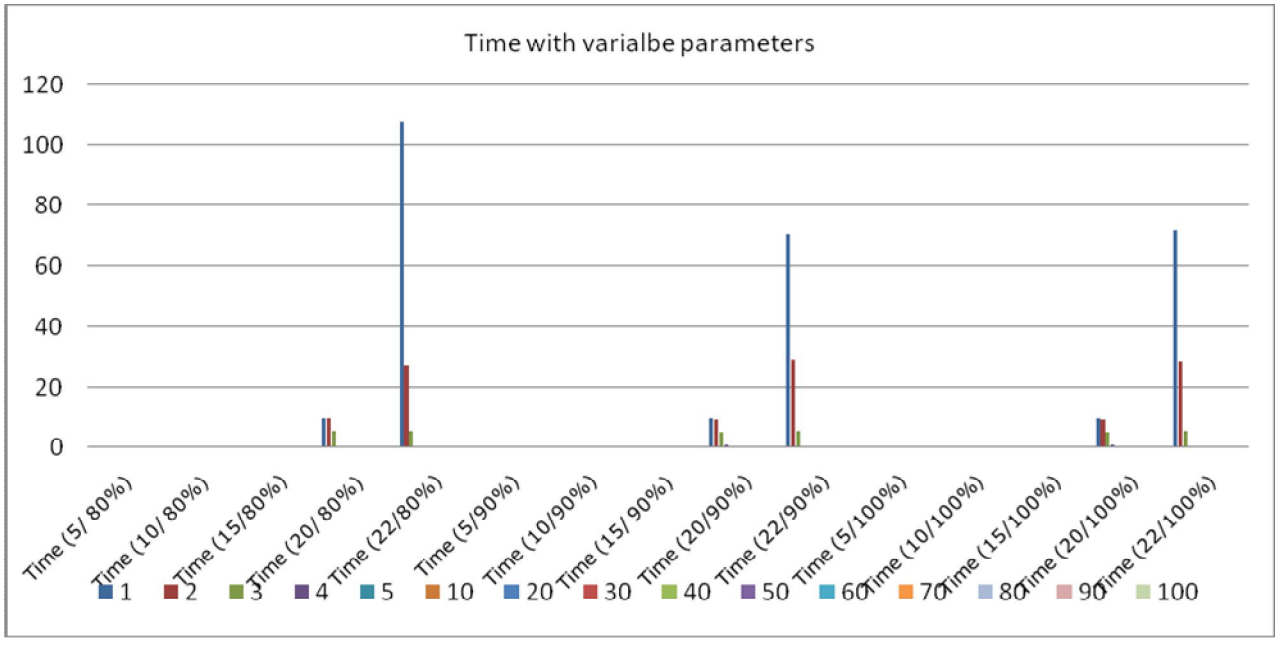

Figure 1. Variation of Support and Confidence and Time for Generation of Association Rules (with 22 Parameters)

Conclusion

The existing algorithm works fine but to ensure the future efficiency in terms of execution time and computing resources alternatives has to investigate. The AVK based secure model of communication using fixed key-size but frequently changing key is useful approach for energy efficient mode. This research work analyses parametric Automatic Variable Key based symmetric cryptosystem from the view of cryptanalyst or hacker [10, 18]. The model has been tested and analysed with (1) Variable number of parameters, (2) Number of frequent sets and (3) Number of association rules generated. The model is considered to be variable number of parameters chosen from user's personal and professional information. Recently in [11,12,13,14,15] future perspective of AVK based parametric model as a user-friendly and harder from cryptic mining perspectives is needed for Internet of Things and secure Machine to Machine communication with client side encryption perspectives.

References

[1] Chakrabarti P., Bhuyan B., Chowdhuri A., and Bhunia C., A novel approach towards realizing optimum data transfer and Automatic Variable Key (AVK) in cryptography, IJCSNS International Journal of Computer Science and Network Security, Vol. 8, No. 5, pp. 241, 2008.

[2] P. Chakrabarti. "Application of. Automatic Variable Key (AVK) in RSA". Int'l J HIT Transactions on ECCN, Vol.2, No. 5, Jan-Mar 2007, pp. 304-311.

[3] P.Chakrabarti et al., Various New and Modified approaches for selective encryption (DES, RSA and AES) with AVK and their comparative study, published in International Journal HIT Transactions on ECCN, Vol 1, No.4, p. 236-244. 51

[4] Bhunia, C. T., Application of AVK and selective encryption in improving performance of quantum cryptography and networks, United Nations Educational Scientific and Cultural Organization and International Atomic Energy Agency, retrieved, Vol.10, No. 12, pp. 200-210, 2006.

[5] Bhunia C. T., New Approaches for Selective AES towards Tackling Error Propagation Effect of AES, Asian Journal of Information Technology, Pakistan, Vol. 5, No. 9, pp. 1017-1022, 2006.

[6] Bhunia C. T., Chakrabarti P., Chowdhuri A. and Chandan T., Implementation of Automatic Variable Key with Choas Theory and Studied Thereof, J IUP Computer Science, Vol -5, No 4, pp. 22-32, 2011.

[7] Bhunia C.T., Mondal G. and Samaddar S., Theories and Application of Time Variant Key in RSA and that with selective encryption in AES, Proc. EAIT, Elsevier Publications, Calcutta CSI-06, pp. 219-221, 2006.

[8] Prajapat, Shaligram, Ramjeevan Singh Thakur, "Key Diffusion Approach for AVK based Cryptosystem" In Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies ICTCS-16 Article No. 78, 2016.

[9] Prajapat, Shaligram, Ramjeevan Singh Thakur "Cryptic-Mining: Association Rules Extractions Using Session Log". In proceedings of Computational Science and Its Applications--ICCSA 2015. Springer International Publishing. pp. 699-711. , 2015.

[10] Prajapat, Shaligram, Ramjeevan Singh Thakur. "Extraction of Association Rules Extractions on Parameterized AVK cryptosystem", IJBIDM, Vol.11(2), 2016.(In Press)

[11] Prajapat, Shaligram, D. Rajput, Ramjeevan Singh Thakur, "Time variant approach towards symmetric key", In proceedings of IEEE Science and Information Conference (SAI), London 2013. , pp.398-405, 2013.

[12] Prajapat, Shaligram, Ramjeevan Singh Thakur. "Optimal Key Size of the AVK for Symmetric Key Encryption." In Covenant Journal of Information & Communication Technology, Vol.3 (2), pp. 71-81. 2015.

[13] Prajapat, Shaligram, Ramjeevan Singh Thakur. "Various Approaches towards Crypt-analysis." International Journal of Computer Applications, Vol. 127(14), pp. 15-24, 2015. (doi: 10.5120/ijca2015906518)

[14] Prajapat, Shaligram, Ramjeevan Singh Thakur. "Cryptic Mining for Automatic Variable Key Based Cryptosystem", Elsevier Procedia Computer Science, Vol. 78 (78C), pp. 199-209, 2016. (Doi: doi:10.1016/j.procs.2016.02.034).

[15] Prajapat, Shaligram, Ramjeevan Singh Thakur."Cryptic Mining: Apriori Analysis of Parameterized Automatic Variable Key based Symmetric Cryptosystem."International Journal of Computer Science and Information Security, Vol. 14 (2), pp. 233- 246, 2016.

[16] Prajapat, Shaligram, Ramjeevan Singh Thakur. "Realization of information exchange with Fibo-Q based Symmetric Cryptosystem." International Journal of Computer Science and Information Security, Vol 14(2), pp. 216-223, 2016.

[17] Prajapat, Shaligram, Thakur, A., Maheshwari, K., & Thakur, R. S., "Cryptic Mining in Light of Artificial Intelligence", IJACSA, Volume 6(8), pp. 62-69, 201510.14569/IJACSA.2015.060808).

[18] Prajapat, Shaligram, Sharma, Ashok, Ramjeevan Singh Thakur. "AVK based Cryptosystem and Recent Directions Towards Cryptanalysis", JKSII, Vol.17 (5),2016.(In Press)

[19] Prajapat, Shaligram, Ramjeevan Singh Thakur. "Markov Analysis of AVK Approach of Symmetric Key Based Cryptosystem." In proceedings of Computational Science and Its Applications--ICCSA 2015. Springer International Publishing. pp. 164-176. , 2015.

[20] Prajapat, Shaligram, Jain A. Ramjeevan Singh Thakur, "A Novel Approach For Information Security with Automatic Variable Key Using Fibonacci Q-Matrix", IJCCT, ISSN: 2231 – 0371, 0975 – 7449 Vol-3(3), 2012, p.p. No. 54-57, 2012.

[21] Prajapat, Shaligram, Sharma A., Swami S., Rajput D., Singroli B., R. S. Thakur. "Sparse approach for realizing AVK for Symmetric Key Encryption", International Journal of Recent Development in Engineering and Technology (IJRDET) Vol. 2(4), pp.15-18, 2014.