International Journal on
"Technical and Physical Problems of Engineering"
(IJTPE)
Published by International Organization of IOTPE

ISSN 2077-3528
IJTPE Journal
www.iotpe.com
ijtpe@iotpe.com

September 2022 | Issue 52 | Volume 14 | Number 3 | Pages 202-206

THE WORLD OF E-COMMERCE: THE PROCESS FROM DIGITAL SIGNATURE UP TO MARKET BASKET ANALYSIS

B. Ture Savadkoohi

Department of Computer and Electrical Engineering, Seraj Higher Education Institute, Tabriz, Iran
bita.turesavadkoohi@gmail.com

Abstract- Nowadays, with rapid development of electronic commerce (E-commerce), information security is a fundamental requirement for transaction processing over the internet. Moreover, any information related to consumer behaviors has paid more and more attention in E-commerce market. Thus, in order to data analysis, market basket analysis is used for knowledge discovery. For this aim, first, digital signature that is based on modified RSA is applied for E-commerce security because of its data integrity protection and privacy. Then, in order to predict customer favorite product logistic regression, Apriori algorithm and online Adaboost is applied.

Keywords: Modified RSA, Logistic Regression, Apriori Algorithm, Online Adaboost.

1. INTRODUCTION

The most important challenge of E-commerce, extracting interesting and potentially useful information related to consumer transactions. The database marketing technique uses modern data analysis in order to develop new business strategies and opportunities [1-4]. On the other hand, E-commerce security is applied in order to assure both business and customers for carrying out their transaction in a trustworthy manner [5-6]. This paper is an extension of [7-8] that gives more details in order to increase sells including the security in E-commerce. The aim of this paper is using a digital signature as a valuable tool for secure E-commerce transaction by ensuring data authenticity and integrity and applying market basket analysis in order to predict the favorite product for increasing sell in E-commerce.

Section 2 describes the structure of digital signature in order to preserve security of transaction, while in Section 3, an approach for market basket analysis is introduced. In Section 4 the Evaluation is given and finally, the paper is concluded in Section 5.

2. DIGITAL SIGNATURE

There are various types of cryptographic algorithms [6]. Data Encryption Standard (DES) is based on encrypts 64-bit plaintext by applying 56 bits key [9]. Triple DES (3DES) is applied three 64-bit keys for a single round [10]. Advanced Encryption Standard (AES) is encrypted and decrypted data in blocks of 128, 192 and 256 [11]. Elliptic Curve Cryptography (ECC) is based on elliptic curves equation over finite field [12]. Elgamal cryptographic is based on discrete logarithm problem [13]. Rivest, Shamir and Adleman (RSA) is an asymmetric cryptographic algorithm that is based on the spinosity of factorization large numbers [14-18]. Digital Signature Algorithm (DSA) is used for reliable preservation of a digital message or document [19-20]. The mechanism of a DSA is applied for assuring the validity of E-commerce transactions. However, this mechanism can be algorithmically proved using cryptographic technique. In order to keep the security in DSA, the modified RSA algorithm which is based on random numbers is applied [14-18].

Security Hash Algorithm 1 (SHA-1) uses as a part of digital signature as it is explained in [21-22] is applied: Let A=0x67452301, B=0xefcdab89, C=0x98BADCFE, D=0x10325476, E=0xC3D2E1F0 be 32-bit divisions, wt be the expand word of t.

From iteration 16 to 79 [21-22]:

$$w[i] = (w[i-3] x\text{or } w[i-8] x\text{or } w[i-14] x\text{or } w[i-16]) \, leftrotate 1 \quad (1)$$

Four possible functions that are applied as follows:

$$F(B,C,D) = (B \text{ and } C) \text{ or } (\text{not } B \text{ and } D) \quad (2)$$

$$G(B,C,D) = BX \text{ or } CX \text{ or } D \quad (3)$$

$$H(B,C,D) = (B \text{ and } C) \text{ or } (B \text{ and } D) \text{ or } (C \text{ and } D) \quad (4)$$

$$I(B,C,D) = BX \text{ or } CX \text{ or } D \quad (5)$$

The 80 processing constant words are:

$$k(t) = 0\text{x}5\text{A}827999, (0 \le t \le 19) \quad (6)$$

$$k(t) = 0\text{x}6\text{ED}9\text{EBA}1, (20 \le t \le 39) \quad (7)$$

$$k(t) = 0\text{x}8\text{F}1\text{BBCDC}, (40 \le t \le 59) \quad (8)$$

$$k(t) = 0\text{xCA}62\text{C}1\text{D}6, (60 \le t \le 79) \quad (9)$$

2.1. The Modified RSA Algorithm

The key generation steps are [14-18]:

Step 1: Select two random prime number such as P, Q.

Step 2: Compute $N = P * Q$ and Eular's totient function by:

$$\phi(N) = (P-1)*(Q-1) \quad (10)$$

Step 3: Choose an e as follows:

$$1 < e < \phi(N) \quad (11)$$

where, e and $\phi(N)$ are co-prime

Step 4: Determine decryption exponent L through the following formula:

$$e*L = 1 \bmod \phi(N) \quad (12)$$

Step 5: Send PubBlicKey (PBK) as (E,N) and PriVateKey (PVK) as (N,L).

The Encryption text (En) from Message (M) is done by:

$$En = M \wedge e \bmod(N) \quad (13)$$

The decryption M from En is obtained by:

$$M = En \wedge L \bmod(N) \quad (14)$$

Then, the security of digital signature is increased by obtaining the new obscure key with respect to random number between zero and key length.

3. MARKET BASKET ANALYSIS

Nowadays, any information related to market basket analysis can help in many businesses decision making process [1-4]. Rao, et al. [23] applied market basket analysis in order to find appropriate medications relevant to indication of the diseases. Gayathri [24] introduced FP-Bonsai algorithm for market basket analysis. Chavan et al. [25] defined FP-tree algorithm for mining product in the E-commerce market assessment.

The problem of predict customer favorite product can be divided into two main sub problem:

- How to reduce data sparsity
- How to increase the sale in E-commerce

Matrix factorization is applied in order to face with the problem of data sparsity [26, 27]. Let $Cu = \{Cu_1, Cu_2, ..., Cu_m\}$ be Customers set, $I = \{i_1, i_2, ..., i_n\}$ be Items set, rating given by customers for items are marked with matrix $R = \left[r_{Cu,i}\right]_{m \times n}$, $r_{Cu,i}$ be the rating from customer Cu of item i, each item is associated with a vector v_i, each Customer Cu is associated with a vector v_{Cu}, the constant λ controls the extent of regularization, μ be overall average rating, D_{Cu} be observed Deviations of Customer Cu, D_i be the observed Deviation of item I.

Matrix Factorization (MF) is determined by [26]:

$$MF = \frac{1}{2}\sum_{Cu=1}^{m}\sum_{i=1}^{n} N_{Cu,i}^{R}(r_{Cu,i} - v_{Cu}^{T}v_i)^2 + \lambda(\|v_{Cu}\|^2 + \|v_i\|^2) \quad (15)$$

where,

$$N_{Cu,i}^{R} = \begin{cases} 1 & \text{If customer } Cu \text{ has already purchased product } i \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

Then, SVD is applied as follows:

$$MF = \frac{1}{2}\sum_{Cu=1}^{m}\sum_{i=1}^{n} N_{Cu,i}^{R}(r_{Cu,i} - \mu - D_{Cu} - D_i - v_{Cu}^{T}v_i)^2 + \\ +\lambda(\|v_{Cu}\|^2 + \|v_i\|^2 + \sum_{Cu=1}^{m} D_{Cu} + \sum_{i=1}^{n} D_i^2) \quad (17)$$

After preprocessing of data, at the beginning of the proposed method, from the products of customer shopping basket logistic regression and confidence formula of Apriori algorithm is applied as a first and second classifier in order to define independent value [28-30]. Then, the probability value from two classifiers is used in the online Adaboost algorithm. At the end, the final probability value is determined for defining customer favorite products.

Logistic regression is used for obtaining the relationship between a dependent variable and a set of independent variables [28, 29]. Suppose P_j be the probability of the favorite product j in the store, n be the number of the products in the store, α denotes the equality establishment between probability value of favorite product j with respect to favorite product 1 up to n, P_r be the product in the store, β_1 up to β_n indicator the dependency of probability of favorite product with respect to the other products in the store. Logistic regression is defined as a first classifier by [28]:

$$\log(\frac{P_j}{1-P_j}) = \alpha + \beta_1 \Pr_1 + \beta_2 \Pr_2 + ... + \beta_n \Pr_n \quad (18)$$

Apriori algorithm that is adopted in this paper is one of the best-known algorithms in the association rule mining [28, 30]. Since, there are some products in the market basket that is purchased by customer, thus there isn't need to produce association rules. For this aim, the second part of Apriori algorithm is used. Suppose $Iset_1$ and $Iset_2$ be two be two item sets, the support coefficient of association rule $Iset_1 \Rightarrow Iset_2$ is determined as follows [30]:

$$Support(Iset_1 \Rightarrow Iset_2) = Support(Iset_1 \cup Iset_2) \quad (19)$$

The confidence efficient of $Iset_1 \Rightarrow Iset_2$ is defined as:

$$\text{confidence}(Iset_1 \Rightarrow Iset_2) = \\ = \frac{\text{Support}(Iset_1 \cup Iset_2)}{\text{Support}(Iset_1)} \times 100\% \quad (20)$$

So that, minimum confidence is a threshold for confidence coefficient. The online-Adaboost is one of the popular algorithms that are applied for generating a strong classifier. In the following, the process of online-Adaboost from the research of Hu et al. [31] that is used in this paper will be explained.

Given example sample $(x_1, y_1), ..., (x_n, y_n)$ where, $x_i \in X$ is the value of feature and $y_i \in Y = \{+1, -1, -2, ...\}$ is the predicted value. The favorite products are labeled as $+1$ and the level of not favorite products are labeled as $-1, -2, ...$

Update the value of Number of Favourite Products (NFP), Number of Not Favourite Products $(NNFP)$ and initialize the Weight of New Sample (WNS) by [31]:

$$\begin{cases} NFP \leftarrow NFP + 1 & \text{if } y = 1 \\ NNFP \leftarrow NNFP + 1 & \text{else} \end{cases} \quad (21)$$

$$\begin{cases} WNS \leftarrow \dfrac{(NFP+NNFP)}{NFP} & \text{if } y=1 \\ WNS \leftarrow \dfrac{(NFP+NNFP)}{NNFP} & \text{else} \end{cases} \quad (22)$$

Determine Combined Classifier (CC_t) as follows:

$$CC_t = (1-\rho)\varepsilon_t - \rho sign(y)h_t(x) = (1-\rho)\frac{WNS_t^{SWW}}{WNS_t^{SWC}+WNS_t^{SWW}} - \rho sign(y)h_t(x) \quad (23)$$

where, ρ is a weight in interval $(0, 0.5]$, ε_t is the lowest error, WNS_t^{SWC} and WNS_t^{SWW} are the Sum of Weight that Correct and Wrong classifier by week classifier (h_t) respectively. So that

$$\min CC = \min CC_t \quad t \in \{1,...,D\} \quad (24)$$

where, D is the number of classifiers. Since there are two weak classifiers such as logistic regression and confidence formula of Apriori algorithm, thus this value in this paper is equal to 2.

Assuming $\{h_{r1}, h_{r2}, ..., h_{rn}\}$ be set of weak classifiers whose CC_{ri} are not larger than 0.5. Number of Repetition (NR_i) is obtained by:

$$NR_i = Intger(NR \exp(-\gamma(CC_i - \min CC))) \quad (25)$$

where, NR is maximum number of repetitions and γ is attenuation coefficient in order to find the weak classifier for further updating.

Iterate the following steps for updating h_{ri}, MN times:

Step 1: Update h_{ri} by utilization (x, y).

Step 2: Update WNS, WNS_t^{SWC} and WNS_t^{SWW}.

So that, the value of Number of Input Samples (NIS) that are correctly classified by h_t, WNS, WNS_{ri}^{swc} and WNS_{ri}^{sww} in the case of $sign(y) = h_{ri}(x)$ and $sign(y) \neq h_{ri}(x)$ is determined respectively by:

$$\begin{cases} NIS_t \leftarrow NIS_t + 1 \\ WNS_{ri}^{SWC} \leftarrow WNS_{ri}^{SWC} + WNS \\ WNS \leftarrow WNS\left(\dfrac{1-2\rho}{2(1-CC_{ri})}\right) \end{cases} \quad (26)$$

$$\begin{cases} WNS_{ri}^{sww} \leftarrow WNS_{ri}^{sww} + WNS \\ WNS \leftarrow WNS\left(\dfrac{1+2\rho}{2CC_{ri}}\right) \end{cases} \quad (27)$$

Although, the value of each h_t that $CC_t > 0.5$ is updated. For this aim, the value of NIS_t, WNS_t^{swc} and WNS_t^{sww} in the case of $sign(y) = h_t(x)$ and $sign(y) \neq h_t(x)$ is specified respectively by:

$$\begin{cases} NIS_t \leftarrow NIS_t + 1 \\ WNS_t^{swc} \leftarrow WNS_t^{swc} + WNS \end{cases} \quad (28)$$

$$WNS_t^{sww} \leftarrow WNS_t^{sww} + WNS \quad (29)$$

Moreover, the value of Ω is updated as follows:

$$\Omega = \Omega + 1 \quad (30)$$

So that, Ω is the number of instances that are accurately classifier with the previous classifier before the new sample is input. Next, the strong ensemble classifier is evaluated with respect the ensemble weight ϕ_t^* of h_t by:

$$\phi_t^* = \sigma \log\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right) + (1-\sigma)\log\left(\frac{NIS_t}{\Omega}\right) \quad (31)$$

where, $\sigma = 0.8$ Then, normalization of ϕ_t^* to ϕ_t is characterized by:

$$\phi_t = \frac{\phi_t^*}{\sum_{i=1}^{D}\phi_t^*} \quad (32)$$

Finally, the final strong classifier $H(x)$ is defined by:

$$H(x) = sign\left(\sum_{t=1}^{D}\phi_t h_t(x)\right) \quad (33)$$

4. EVALUATION

The classified item can be True Positive (*TP*), False Negative (*FN*), False Positive (*FP*) and True Negative (*TN*). The accuracy of the proposed method for market basket analysis based on confusion matrix that is represented in table 1 is specified by [2] [32-34]:

$$Accuarcy = \frac{TP+TN}{TP+TN+FP+FN} \quad (34)$$

As shown in Figure 1, the proposed method of this paper increases the accuracy. Moreover, in the final data base that data are increased compared to original data base the accuracy rate will increase.

Table 1. Recommendation Matrix [2], [33]

	Introduced Item by the System	Item Not Introduced by the System
Expected Item	*TP*	*FN*
Not an Expected Item	*FP*	*TN*

5. CONCLUSIONS

Nowadays, with rapid development of information technology stepping into electronic commerce (E-commerce) is a great advantage. On the other hand, digital signature security and market basket analysis are a major concern of E-commerce framework. Thus, the modified Rivest, Shamir and Adleman is used to increase the security of digital signature. Then, the combination method is introduced for market basket analysis. The proposed approach may be particular benefit to improve the E-commerce strategies.

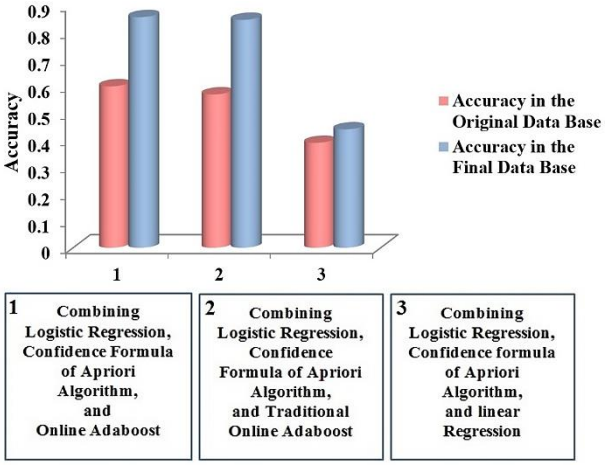

Figure 1. Comparison of the accuracy rate between different methods

REFERENCES

[1] I. Islek, S. GunduzOguduc, "A Hierarchical Recommendation System for E-commerce using Online User Reviews", Journal of Electronic Commerce Research and Applications, Vol. 52, 2022.

[2] B.T. Savadkoohi, P. Nik Mohammadi, "Applying Wormhole Approach to Design a Hierachy in a Relational Datbase for Quick Data Access", Journal of Technical and Physical Problems of Engineering (IJTPE), Issue 47, Vol. 13, No. 2, pp. 144-148, June 2021.

[3] S. Wu, "E-Commerce Decision Support System Based on Internet of Things", Journal of Ambient Intelligence and Humanized Computing, Springer, pp. 1-7. 2018.

[4] M. Scholz, V. Dorner, G. Schryen, A. Benlian, "A Configuration-Based Recommender System for Supporting E-Commerce Decisions", European Journal of Operational Research, Vol. 259, No. 1, pp. 205-215, 2017.

[5] N.L. Bhatia, V.K. Shukla, R. Punhani, S.K. Dubey, "Growing Aspects of Cyber Security in E-Commerce", IEEE International Conference on Communication Information and Computing Technology, Mumbai, India, 2021.

[6] F. Maqsood, M. Ahmed, M. Mumtaz, "Cryptography: A Comparative Analysis for Modern Techniques", Journal of Advanced Computer Science and Application, Vol. 8, No. 6, pp. 442-448, 2017.

[7] B. Ture Savadkoohi, M. Aliakbari, "Predicting Customer Favorite Products in E-commerce", International Conference on Innovation and Research in Engineering Sciences, Tbilisi, Georgia, 2020.

[8] B. Ture Savadkoohi, K. Geyratmand, "Increase the Security of Digital Signature for E-commerce", International Conference on Electrical Engineering, Mechanical Engineering and Computer, Tbilisi, Georgia, 2019.

[9] K.G. Manjula, M.N. Ravikumar, "Color Image Encryption and Decryption Using DES Algorithm", Journal of Engineering and Technology, Vol. 3, No. 7, pp. 1715-1718, 2016.

[10] C. Mitchell, "On the Security of 2-Key Triple DES", Journal of IEEE Transactions on Information Theory, Vol. 62, No. 11, pp. 6260-6267, 2016.

[11] Y. Wang, C. Su, C. Horng, C. Wu, C. Huang, "Single- and Multi-Core Configurable AES Architectures for Flexible Security", Journal of IEEE Transactions on very Large-Scale Integration (VLSI) Systems, Vol. 18, No. 4, pp. 541-552, 2010.

[12] A. Bansal, A. Agrawal, "Providing Security, Integrity and Authentication Using ECC Algorithm in Cloud Storage", IEEE International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2017.

[13] D. Laiphrakpam, M. Khumanthem, "Medical Image Encryption Based on Improved ElGamal Encryption Technique", Journal of Optik, Vol. 147, pp. 88-102, 2017.

[14] S. Jaju, S. Chowhan, "A Modified RSA Algorithm to Enhance Security for Digital Signature", IEEE International Conference and Workshop on Computing and Communication, Vancouver, BC, Canada, 2015.

[15] S. Jaju, S. Chowhan, "Analytical Study of Modified RSA Algorithms for Digital Signature", Journal on Recent and Innovation Trends in Computing and Communication, Vol. 3, No. 3, pp. 944-949, 2015.

[16] G. Patel, K. Panchal, S. Patel, "A Comprehensive Study on Various Modifications in RSA Algorithm", Journal of Engineering Development and Research, Vol. 1, No. 3, pp. 161-163, 2014.

[17] S. Tint, "Survey on Asymmetric Algorithm Using RSA Different Modified Models", Journal Computer Applications, Vol. 1, No. 2, pp. 27-35, 2014.

[18] A. Chhabra, S. Mathur, "Modified RSA Algorithm: A Secure Approach", International Conference on Computational Intelligence and Communication Networks, Gwalior, India, 2011.

[19] G. Lax, F. Buccafurri, G. Caminit, "Digital Document Signing: Vulnerabilities and Solution", Journal of Information Security: A Global Perspective, Vol. 24, No. 1-3, pp. 1-14, 2015.

[20] J. Hernaandez Ardieta, "Enhancing the Reliability of Digital Signatures as Non-Repudiation Evidence under a Holistic Threat Model", Ph.D. Thesis, University Carlos III of Madrid, Spain, 2011.

[21] C.G. Thomas, R. Jose, "A Comparative Study on Different Hashing Algorithms", Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, No. 7, pp. 170-175, 2015.

[22] A. Agung, P. Dewi, A. Shaugi, M. Salman, "Analysis and Comparison of MD5 and SHA-1 Algorithm Implementation in Simple-O Authentication Based Security System", IEEE International Conference on QiR, Yogyakarta, Indonesia, 2013.

[23] A. Rao, J. Krian, G. Poornalatha, "Application of Market-Basket Analysis on Healthcare", Journal of System Assurance Engineering and Management, 2021.

[24] B. Gayathri, "Efficient Market Basket Analysis Based on FP-Bonsai", International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2017.

[25] G. Chavan, N. Gaikwad, T. Samal, A. Sonule, H. Palivela, A.C. Patil, "Various Mining Techniques Defined for Mining Product Valuation Instances in Market Basket Data", International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE), Coimbatore, India, 2014.
[26] J. Yu, M. Gao, W. Rong, Y. Rong, Q. Xiong, "A Social Recommender Based on Factorization and Distance Metric Learning", Journal of IEEE, Vol. 5, pp. 21557-21566, 2017.
[27] Y. Koren, R. Bell, C. Volinsky, "Matrix Factorization Technique for Recommender System", Journal of IEEE Computer Society, Vol. 42, No. 8, pp. 30-37, 2009.
[28] O. Kasap, N. Ekmekci, U. Ketenci, "Combining Logistic Regression Analysis and Association Rule Mining via MLR Algorithm", International Conference on Software Engineering Advances, Rome, Italy, 2016.
[29] G. Nie, W. Rowe, L. Zhang, Y. Tian, Y. Shi, "Credit Card Churn Forecasting by Logistic Regression and Decision Tree", Journal of Expert System with Application, Vol. 38, No. 12, pp. 15273-15285, 2011.
[30] G. Sheng, H. Hou, X. Jaing, Y. Chen, "A Novel Association Rule Mining Method of Big Data for Power Transformers state Parameters Based on Probabilistic Graph Model", Journal of IEEE Transactions on Smart Grid, Vol. 9, No. 2, pp. 695-702, 2018.
[31] W. Hu, J. Gao, Y. Wang, O. Wu, S. Maybank, "Online Adaboost-Based Parameterized Method for Dynamic Distributed Network Intrusion Detection", Journal of IEEE Transactions on Cybrnetics, Vol. 44, No. 1, pp. 66-82, 2014.
[32] F. Golabi, M. Shamsi, M.H. Sadaaghi, A. Barzegar, M. Hejazi, "Development of a New Oligonucleotide Block Location-Based Feature Extraction (BLBFE) Method for the Classification of Riboswitches", Journal of Molecular and Genomics, Vol. 295, No. 04, pp. 525-534, 2020.
[33] P. Lopes, B. Roy, "Dynamic Recommendation System using Web Usage Mining for E-commerce Users", International Conference on Advanced Computing Technologies and Applications, Mumbai, India, 2015.
[34] L.C. Briand, J. Wust, J.W. Daly, D.V. Poter, "Exploring the Relationships in Object-Oriented Systems", Journal of Systems and Software, Vol. 51, No. 3, pp. 245-273, 2000.

BIOGRAPHY

Bita Ture Savadkoohi was born in Tabriz, Iran. She obtained diploma in software engineering from Islamic Azad University, Iran in 2003 and her Ph.D. degree in Computer Science from University of Trento, Italy in 2010. Since 2012, she is an Assistance Professor at Seraj Higher Education Institute, Tabriz, Iran. Her research interests included computer graphics, computational geometry (e.g., shape comparison), analysis of 3D data, software engineering, data base and data mining.