Detecting the Phishing Attack Using Collaborative Approach and Secure Login through Dynamic Virtual Passwords

Robbi Rahim
Sekolah Tinggi Ilmu Manajemen Sukma, Medan, Indonesia. E-mail: usurobbi85@zoho.com

S. Murugan*
Assistant Professor, Department of Computer Science and Engineering, Sri Aravindar Engineering College, Tamil Nadu, India. E-mail: smartrugans@gmail.com

Reham R. Mostafa
Information Systems Department, Faculty of Computers and Information Sciences, Mansoura University, Mansoura, Egypt. E-mail: reham_2006@mans.edu.eg

Dr. Anil Kumar Dubey
Associate Professor, Department of Computer Science & Engineering, ABES Engineering College, Ghaziabad, NCR, India. E-mail: anildudenish@gmail.com

R. Regin
Assistant Professor, Department of Information Technology, Adhiyamaan College of Engineering, Hosur, Tamil Nadu, India. E-mail: regin12006@yahoo.co.in

Dr. Vikram Kulkarni
Assistant Professor, Department of Information Technology, MPSTME, NMIMS University, Mumbai Campus, Maharashtra, India. E-mail: vikram.kulkarni@nmims.edu

Dr.K.S. Dhanalakshmi
Assistant Professor III, Department of ECE, School of Electronics & Electrical Technology (SEET), Kalasalingam Academy of Research & Education (KARE), Tamil Nadu, India. E-mail: dhanalakshmi.jai3@gmail.com

Received July 12, 2020; Accepted September 14, 2020
ISSN: 1735-188X
DOI: 10.14704/WEB/V17I2/WEB17049

Abstract

User data protection is a major problem in the technical world. To get critical data from people such as bank account, credit card details and passwords, fraudulent people use different techniques for their advantages. Misuse of user credentials affects many industry sectors day by day. The fraudulent practice where more people connected to a network revealing sensitive data of individual, group or a company of legitimate sources called phishing. The aids for the hackers are forged websites and emails which look similar as original. There are different anti-phishing methods are proposed which follows various methods, to protect users against phishing. Proposed technique has detection of phishing at various stages. It contains incoming mail analysis for its feature as well as source. If mail has link to webpage, that cautious webpage is checked for its legitimateness. It is examined by using two methods. For popular sites, webpage content is compared with content of known legitimate site. The second step is to find whether any cautious site is increasingly dependable on any related pages to calculate relationship strength indirectly or directly associated pages. To protect from phishing attacks, these techniques uses virtual passwords for logging into sites.

Keywords

Phishing, Vulnerability, Protection, Goals, Virtual Password.

Introduction

Phishing is a dangerous hazard related with social engineering technique, to get forged data of someone under legitimate site for their advantage. Real challenge comes on the speed in which the data transfers [1]. The preplanning of these hacker aims to trace the details regarding transactions. They steal data from the client by using forged email which has link that points to spoofed website that is site of legitimate organization. They use original sites copy as bait which is sent to user. Customers of financial institutions and many banks are targets of phishing attacks by Anti-Phishing Working Group [APWG]. Damage by phishing ranges from accessing the clients' default mail accounts to get rid of financial loss.

According to way by which attack is done, it can be divided into different types. Well known type of deceptive phishing, sending data to fake data to client which provokes the user to respond on those messages it mostly in the form of clicking the fake links with more offers leads to forged site in which confidential data is compromised. In malicious [2] software, malware phishing which scam are produced as email attachment or by download file from site. When users are attempt to log in sites, Web Trojans pop-up invisibly. Locally, they collect user's information as well as transmit to phishing hacker. In content injection, content of legitimate site is replaced by hacker with false content to misdirect or mislead user tox give their information to them. Attacker between user as well as legitimate site or system is Man-in-Middle. They record the data which is entered by user, un-damaged financial plans. At the time of user inactivity then data [3] is utilized. When attackers create websites with attractive sound with indexed legitimately with search engine, Search Engine Phishing is occurred.

To detect as well as safeguard phishing attacks, Anti-phishing method is used which safeguard users from phishing. This method works on website attributes or emails or website URL. Many techniques are employed to enable clients to filter and recognize different types of attacks. This paper focuses on various methods which work on different levels to identify attacks and measures to protect critical data like passwords from hackers. To combat with attacks there are many techniques are used. Blocking those known phishing sources and mails from this source by implementing browser filters and toolbars. Classifier [4] is used to segregate received emails based on identified features of phishing mails and then filter those mails. By using classifier technique those received mails are carefully analyzed. This method [5] is used to find web page relationship and when target website is found then page is phishing. To identify phishing pages, compare images and textual content in webpage for finding similarities between legitimate webpage's content [6]. To authenticate users online, phishing attacks are prevented by many methods. User one-time password or virtual password [7] is calculated using functions which protects users data from being misused or stolen. Any emergency beacon messages are safety messages and commercial based applications are non-safety messages [8].

Related Works

Users can safeguard their username, passwords from key loggers, malicious bots or spyware, Virtual keyboard authentication is used by them which still has various other fallacies in which attacker can take advantage which contains clicking on the screenshot, coordinate position noting, shoulder spoofing. To overcome this limitation, this technique has virtual keyboard which is produced every time when user use web site. After each click it keys in rear effective keyboard changes it position. Keys position are hidden so it difficult that user behind to identify pressed key. To capture authentication details [9], this method uses virtual keyboard for users more secure and makes tougher for malware programs.

One of the major problems is that to detect password while internet user logged into website. At present, HTTPS are widely used, there are some kinds of equipment's and tools, e.g. Back Track and Cain, which assists hackers to decode and detect HTTPS. This research focused to design New Web Authentication in which password of that browser is sent to user which is Dynamic one. Passwords in which hackers detected for each time not be same it will be changed so it cannot be validate by them. Researchers [10] designs Password Authentication Model and analyzed functional procedure. This model contains complete security system. Moreover, this research gives website to prove web programmers to apply this model to authenticate works. Researchers found that this model is easy to implement and fully secured. Load values of CPU and Delay of servers are hardly increased.

This research purposively proposes the Dynamic Password Authentication procedures to be practically applied with various platforms e.g. web applications, network devices, and mobile applications. The researchers presented [11] the overall design and the developed design with stronger security. Besides, the system analysis was tested for Security, Speed, and "Ease of Use". Based on the study and testing, it was found that the authentication system designed by the researchers can usefully be applied for various purposes i.e. (1) to replace an authentication on website without using HTTPS to reduce an expense on CA; (2) to be applied on network devices or mobile applications to secure the password sending process and to prevent the password from being cracked. Firmly, this authentication method had been tested and compared with other protection systems and it was qualified as a highly effective system.

Flexible collaboration method is proposed Cyber Live App [12], to permit live virtual desktop app sharing depends on virtualization and cloud structure. It enables secure sharing as well as on-demand migration on numerous equipment or users. To deliver desktops to various users, proxy-based window filtering method is presented. VNC protocol as well as VM cloning service based on VMs. On an extended Meta VNC, these methods are preliminarily calculated. Its results are verified and these methods are useful and effective.

In [13], IEEE 802.15.6 standard was adopted to evaluate consumption of energy method. This method gives improved accuracy and easy way to use consumption of energy as an index to calculate cost methods. This method attains its security needs such as anonymity, integrity, authentication, intractability, privacy, and non-repudiation. This method gives improved trade-offs between efficiency as well as security when compared with other methods.

In [14], XML schemes were proposed to show graphical image. When password with image recognition on graphical pattern set by the client proceeds as well as checks it for valid pattern depends on drift and length. By applying numerous transformations, different forms of graphical patterns are produced and these pixel values are kept as XML pattern. Using LSB steganography, this server updated pattern bits and back to user as password image. This method is implemented using Desktop or Mobile application. This method is more effective when compared to another method. Complete data is separated from query password pattern image in which 100% accuracy is attained. The qualitative measure proves this method has improved robustness and reliability against different criteria.

Proposed Work

Not a single technique entirely stops phishing. Amalgamation of good and practicing the application, proper app of recent method as well as enhancement in security system can minimize phishing as well as losses from it. This paper contains collaborative technique to identify attacks at various levels as well as to secure user's data online.

A. Analysis of Email

To many random users, fraudulent people send lots of phishing emails. For email, implementing browser filters and toolbars which maintains block list of known phishing sources which blocks that sources. This mail has some features in common for some phishing source. Some features of emails are chosen based on large dataset analysis is known as malicious emails. Features such as email size, attachments and its types, MIME, data header format etc. To segregate malicious emails, this method uses random forest classifier technique. Network is weighted with nodes, they are the form that chooses feature set in which node grows tree and gives output called vote. Based on vote's majority, final decision of classification was made.

B. Content-based Anti-phishing

Spam received from hackers contains URL black sites which looks same as original such as text, page link and images to get personal data. They use images instead of text for making complexity in anti-phishing to detect phishing pages. To calculate similarity between protected as well as suspicious web pages, Bayesian method of content-based phishing [4] of visual and textual contents is used. When popular web page is spoofed and phished, this method needs pre-processing of legitimate website. In database, content features, as well as a previous solution made to clear phishing activities and those details, are being snatched in the site. To each word, it contains text classifier which segregates main text from HTML tags as well as stemming. Instead of original words, stems uses basic features. For example, "algorithm", "algorithm" and "algorithm" are stemmed into "algorithm" are taken as same word. To construct vocabulary, stemmed words are stored. For forums, histogram vector (j1, j2,...,jn) will be picturized, where every element shows single occurrence as well as n denotes number of elements in vector.

T denotes test conducted on a web-page, probability, P(g j |T) that the forum T relies to certain portion g j is estimated from considering Bayes rule of posterior probability,

$$P(gj|T) = \frac{P(gj)\pi(i=1\ to\ n)P(ui|gs)(hiT)/R}{\sum(s=1\ to\ d)P(gs)\pi(i=1\ to\ n)P(ui|gs)(hi,T)/R} \text{ (1)}$$

Where, hi, T denotes occurrence of ith word appearing in web page T, and R is number of words separated from safe web page.

HTML as well as accessible models are transformed into pictures (JPEG format) by using image classifiers. They are fit into fixed-size images as well as its signature is evaluated. Feature or signature contains images and contains features as well as corresponding weights. This feature has 2 components. Degraded colour as well as centroid position distribution. To calculate two web page images, EMD (Earth Mover’s Distance) is used.

Suppose 2 forum pictures a and b with signs Sa and Sb, where Sa has m facility (units), Sb has n facility(units).

EMD-based visual similarity of 2 pictures is given as

$$S\ visual\ (Sa, Sb) = 1 - (EMD(Sa, Sb, D)^{a} \text{ (2)}$$

If $S\ visual\ (Sa, Sb) = 1$, two images are identical.

Detection results by two classifiers are combined by using Bayesian fusion technique. Let the random variable ET ∈ {O, N}, EV∈ {O, N} be tasks on the forums is phishing or normally dedicated by text as well as image classifier. If ET = EV, i.e., both classifiers used in deciding, forums into corresponding section. If ET ≠ EV, i.e., classifiers various decisions, underf goes classifier which is having high rate of possible accuracy on same such measurements distribution. Decision factor δ, proportion of2 frontier probabilities, estimated.

$$\delta = PT(C|1t)/PV(C|1v) \text{ (3)}$$

If δ ≥ 1, undergoes text classifier’s choice instead of rear classifier, If δ<1 the situation is opposing.

Where$PT(C|1t) = KT(1t, C)/(KT(1t, C) + KT(1t, I))$ -Text classifier

$PT(C|1v) = KT(1v, C)/(KT(1v, C) + KT(1v, I))$ - Picture Classifier

K (lv, C) and K (lv, I) represent numbers of corrects as well as improper classification of a forum with some similarity measurements from subinterval lv.

To identify web page to be normal web page or phishing, Bayesian threshold θ is utilized either text or picture classifier. This value is set in which number of unaltered forums is reduced. If similarity S is probability P(g1|T) of web page T belongs to phishing categorized to g1 in-text classifier or rear similarity S rear in image classifier.

According to probability, it is obtained.

$$P(o|s > \theta)_{\theta > di} = \frac{K(s>di,o)}{K(s>di,o)+K(s>di,N)} \quad (4)$$

Where, $K(s>d_i,O)$, $K(s>d_i,N)$ – number of phishing as well as normal web pages with similarities demanding d_i.

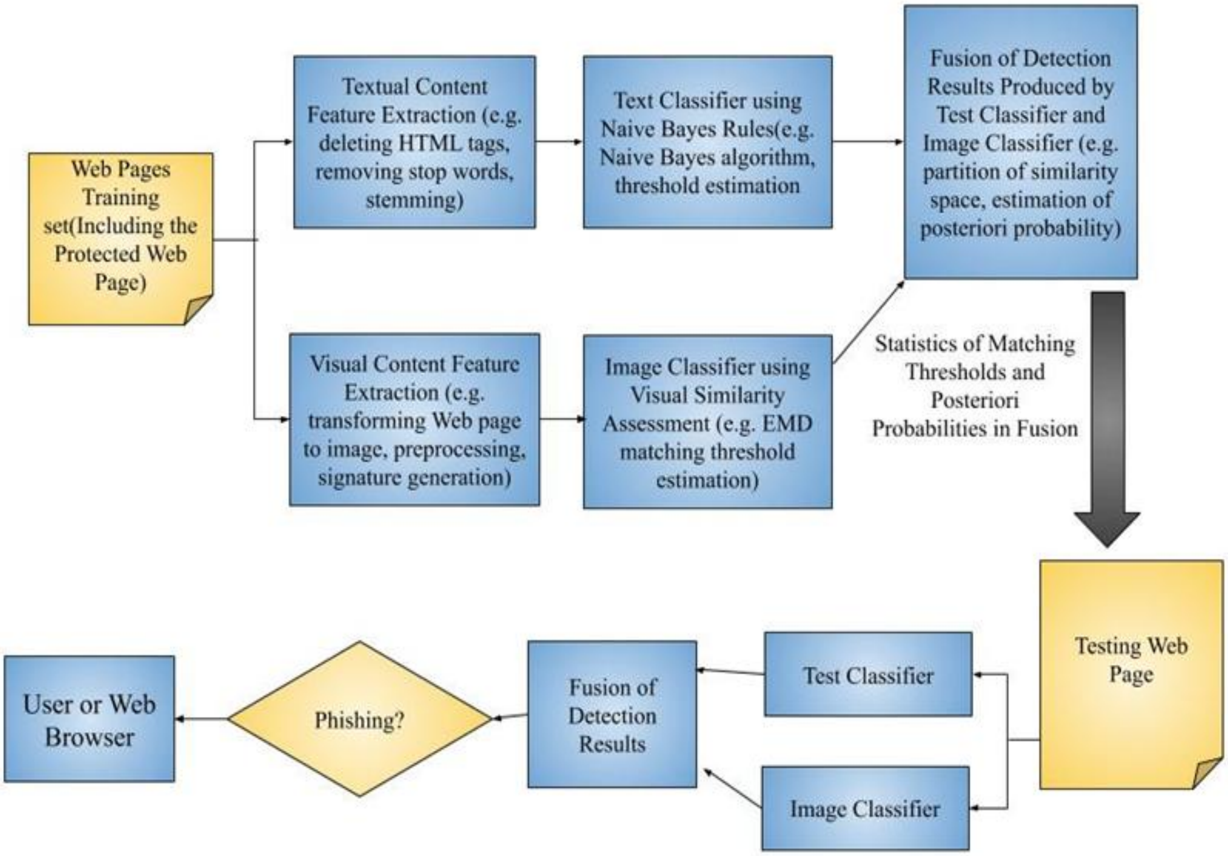
Web Pages Training set(Including the Protected Web Page)
Textual Content Feature Extraction (e.g. deleting HTML tags, removing stop words, stemming)
Text Classifier using Naive Bayes Rules(e.g. Naive Bayes algorithm, threshold estimation
Fusion of Detection Results Produced by Test Classifier and Image Classifier (e.g. partition of similarity space, estimation of posteriori probability)
Visual Content Feature Extraction (e.g. transforming Web page to image, preprocessing, signature generation)
Image Classifier using Visual Similarity Assessment (e.g. EMD matching threshold estimation)
Statistics of Matching Thresholds and Posteriori Probabilities in Fusion
Testing Web Page
Test Classifier
Image Classifier
Fusion of Detection Results
Phishing?
User or Web Browser

Figure 1 Overall System-Architecture

C. Identification Phishing Target

This method presents a method to find phishing target for web page classification for lesser-known suspicious. For illegitimate purposes, phishing targets legitimate website impersonated to divulge users personal data. To find, if cautious forums may legal or unlawful, by proper identification vulnerable target. Phishing forums are not separated but combined with their goals. This is called a direct relationship. It forms phishing forums to only their goals. It starts with all direct means or indirect go troughs with websites with have doubtful contents. By determining their associated as well as partition graph to determine Web community, construct web graph from associated pages. To measure parasitic strength from given page, parasitic coefficient is used. Direct association of succeeding links shows reference from source to target. In terms of similarity as well as ranking association, Indirect association was expressed. Rank relationship from page i to page j based on page j's position in searching values from page i's content as query. In terms of features similarity, relationship from page (i toj)found. i.e., visual, verbal content and layout of the forum between pages.

Parasitic coefficient $Para_{ij}$ of page i to page j, is evaluated weighted sum of direct as well as indirect association relationships.

$$Para_{ij} = \alpha Dij + (1 - \alpha) Iij \quad (5)$$

$$Parasitic\ Coefficient\ Ratio: PCR_{ij} = Para_{ij} / Para_{ij} \quad (6)$$

Greater PCR_{ij}'s value, strengthen exploitative setting from page i to page j. Hence forum-site of page j consisting of greatest PCR_{ij} in phishing goals.

D. Dynamic Virtual Passwords

Under various vulnerability and threats, Users passwords are proving to be compromised and stolen. Across various websites, Users often chose weak passwords as well as reuse same. Domino effect causes routinely reusing passwords. This method, user can choose VP method ranges from weak to strong safeguard to secure users passwords. Stronger as well as complex schemes are Trade off. For each time from virtual password method, it is dynamic which is directly obtained for authentication as well as allows client to select function. During registration phase as well as secret between server and user, Virtual Password function (VPF) are set. Starting digit should be 3 times the original password; 50z+dateofbirth, where z is real password. Authentication Process contains client entry into the forum through username and password. During each login random number(R) is given by the receiver/master end. Virtual password (VP) is calculated as function of hiding the client password as well as random number VP=F(X, R). It can verify user by evaluating VP if VPF is bijective. It determines user's record from database which depends on id as well as estimate VP then compares with user.

Results

Implementing results of anti-phishing methods are detailed as follows. Using forest classifier, detection malicious mail using 83 features attained good outcome result of 0.6 % False Negative Rate (FNR). Phishing page detection accuracy of 98.73-98.95% for data-set 10005 numbers for false alarm rate of 0.9-1.2%. Communicative anti-phishing using Bayesian technique has proper classification rate from 95% forums like PayPal, eBay, HSBC, Yahoo, Rapid Share. Use unique virtual passwords to overcome the vulnerability cause every instant of input login enters VP. Phishing attacker get VP, but unable to process password since VP varies for an instance.

Table 1 Overall performance of proposed method

FACTORS	VALUES OBTAINED BY PROPOSED METHOD
False Negative Rate (FNR)	0.6%
Accuracy	98.95%
False alarm rate	1.2%.
Classification rate	95%

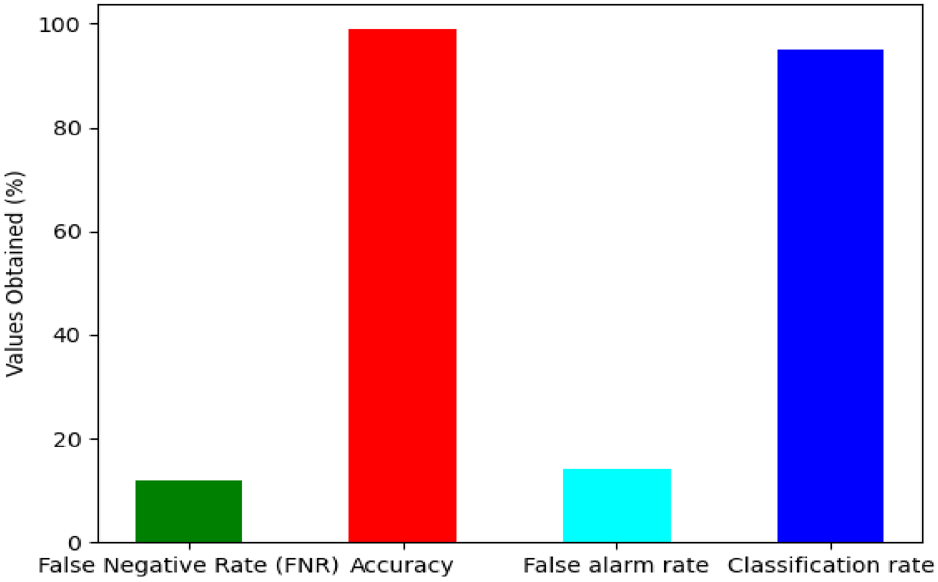

Figure 2 Overall Performance of Proposed Method

Conclusion

Collaborative method of utilizing anti phishing methods work at various levels as well as needs some predefined conditions to be met. Auto-filtering mailings from known phishing sources, spam-list is maintained. Random Forest classifier classifies received mails depends on known as common features set and gives user of possible threats. It needs incoming phishing mail. Whenever illegal mail and redirects to webpage, content-based anti-phishing classifies by identifying similarity level between legitimate and suspicious webpage. It needs presence of features from legitimate webpage. Phishing target method decides if cautious site is legitimate for given webpage. When goal is determined, the forum can be a threat. Further such method needs phishing page combined with goals directly or indirectly at required some space. By using various VP for each login, use of dynamic VP gives user protected passwords.

References

Manikandan, R., & Kumar, N. (2020). Approach on Automatic Provisioning of Switches in Data Center Through Software Defined Network Controllers, *Journal Citation Reports*, *7*(11), 558-561.

Murugan, S., Jeyalaksshmi, S., Mahalakshmi, B,. Suseendran, G., Nusrat Jabeen, T., & Manikandan, R. (2020). Comparison of ACO and PSO algorithm using energy consumption and load balancing in emerging MANET and VANET infrastructure. *Journal Citation Reports*, *7*(9), 1197-1204.

Khalaf, O.I., & Sabbar, B.M. (2019). An overview on wireless sensor networks and finding optimal location of nodes. *Periodicals of Engineering and Natural Sciences*, *7*(3), 1096-1101.

Amin, R., Ryan, J., & Van Dorp, J. (2011). Detecting targeted malicious email. *IEEE Security & Privacy*, *10*(3), 64-71.

Wenyin, L., Liu, G., Qiu, B., & Quan, X. (2011). Antiphishing through phishing target discovery. *IEEE Internet Computing*, *16*(2), 52-61.

Regin, R., & Menakadevi, T. (2019). Dynamic clustering mechanism to avoid congestion control in vehicular ad hoc networks based on node density. *Wireless Personal Communications*, *107*(4), 1911-1931. http:doi.org/10.1007/s11277-019-06366-2

Pansa, D., & Chomsiri, T. (2011). Web security improving by using dynamic password authentication. *In International Conference on Network and Electronics Engineering*, *11*, 32-36.

Pansa, D., & Chomsiri, T. (2012). Dynamic Password Authentication: Designing step and security analysis. *In IEEE 7th International Conference on Computing and Convergence Technology (ICCCT)*, 518-523.

Li, J., Jia, Y., Liu, L., & Wo, T. (2013). CyberLiveApp: A secure sharing and migration approach for live virtual desktop applications in a cloud environment. *Future Generation Computer Systems*, *29*(1), 330-340.

Liu, X., Zhang, R., & Zhao, M. (2019). A robust authentication scheme with dynamic password for wireless body area networks. *Computer Networks*, *161*, 220-234.

Juneja, K. (2020). An XML transformed method to improve effectiveness of graphical password authentication. *Journal of King Saud University-Computer and Information Sciences*, *32*(1), 11-23.

Ahmed, E.R., Alabdullah, T.T.Y., Amran, A., & Yahya, S.B. (2018). Indebtedness Theory and Shariah Boards: A Theoretical Approach. *Global Business & Management Research*, *10*(1), 127-134.

Ahmed, E.R., Islam, A., Zuqibeh, A., & Alabdullah, T.T.Y. (2014). Risks management in Islamic financial instruments. *Advances in Environmental Biology*, *8*(9), 402-406.

Ahmed, E.R., Islam, M.A., Alabdullah, T.T.Y., & Bin Amran, A. (2018). Proposed the pricing model as an alternative Islamic benchmark. *Benchmarking: An International Journal*, *25*(8), 2892-2912.

Alabdullah, T.T.Y., Ahmed, E.R., & Thottoli, M.M. (2019). Effect of Board Size and Duality on Corporate Social Responsibility: What has Improved in Corporate Governance in Asia?. *Journal of Accounting Science*, *3*(2), 121–135.

Alabdullah, T.T.Y., Nor, M.I., & Ries, E. (2018). The Determination of Firm Performance in Emerging Nations: Do Board Size and Firm Size Matter. *Management*, *5*(3), 57–66.

Anandakumar, H., & Umamaheswari, K. (2017). An Efficient Optimized Handover in Cognitive Radio Networks using Cooperative Spectrum Sensing. *Intelligent Automation & Soft Computing*, 1–8.

Anandakumar, H., & Umamaheswari, K. (2018). A bio-inspired swarm intelligence technique for social aware cognitive radio handovers. *Computers & Electrical Engineering*, *71*, 925–937. http:doi.org/10.1016/j.compeleceng.2017.09.016

Bento, A.C. (2018). Internet of Things: An Experiment with Residential Automation for Robotics Classes. *International Research Journal of Management, IT and Social Sciences*, *5*(2), 113-119. https://doi.org/10.21744/irjmis.v5n2.51

Desfiandi, A., Suman Rajest, S., Venkateswaran, P., Palani Kumar, M., & Singh, S. (2019). Company Credibility: A Tool To Trigger Positive Csr Image In The Cause-Brand Alliance Context In Indonesia. *Humanities & Social Sciences Reviews*, *7*(6), 320-331. https://doi.org/10.18510/hssr.2019.7657

Ganguli, S., Kaur, G., Sarkar, P., & Rajest, S.S. (2020). An Algorithmic Approach to System Identification in the Delta Domain Using FAdFPA Algorithm. *In Business Intelligence for Enterprise Internet of Things*, *Springer, Cham*, 203-211.

Gupta, J., Singla, M.K., Nijhawan, P., Ganguli, S., & Rajest, S.S. (2020). An IoT-Based Controller Realization for PV System Monitoring and Control. *In Business Intelligence for Enterprise Internet of Things*, *Springer, Cham*, 213-223.

Haldorai, A., & Kandaswamy, U. (2019). Supervised Machine Learning Techniques in Intelligent Network Handovers. *In Intelligent Spectrum Handovers in Cognitive Radio Networks*, *Springer*, *Cham*, 135-154. http:doi.org/10.1007/978-3-030-15416-5_7

Haldorai, A., Ramu, A., & Murugan, S. (2018). Social Aware Cognitive Radio Networks: Effectiveness of Social Networks as a Strategic Tool for Organizational Business

Management. *In Social network analytics for contemporary business organizations, IGI Global*, 188-202. http:doi.org/10.4018/978-1-5225-5097-6.ch010

Adanov, K.B., Suman, R.S., Mustagaliyeva, G., & Khairzhanova, A. (2019). A Short View on the Backdrop of American's Literature. *Journal of Advanced Research in Dynamical and Control Systems*, *11*(12), 182-192.

Rajest, S.S., & Suresh, D. (2018). The Deducible Teachings of Historiographic Metafiction of Modern Theories of Both Fiction and History. *Eurasian Journal of Analytical Chemistry*, *13*(4), emEJAC191005.

Rao, A.N., Vijayapriya, P., Kowsalya, M., & Rajest, S.S. (2020). Computer Tools for Energy Systems. *In International Conference on Communication, Computing and Electronics Systems, Springer, Singapore,* 475-484.

Rusandy, D.S., Astuti, W., & Firdiansjah, A. (2018). Effect of MCSQ and COSE on service recovery and its impact on customer satisfaction. *International Research Journal of Management, IT and Social Sciences*, *5*(2), 237-247. https://sloap.org/journals/index.php/irjmis/article/view/93

Sarode, A., Parmar, A.J., & Junnarkar, S. (2016). To Investigate Electrode Wear Rate on Steel Material with Different Parameters on EDM Machine. *International Journal of Advanced Engineering Research and Science*, *3*(6), 189-193.

Sharma, M., Singla, M.K., Nijhawan, P., Ganguli, S., & Rajest, S.S. (2020). An Application of IoT to Develop Concept of Smart Remote Monitoring System. *In Business Intelligence for Enterprise Internet of Things, Springer, Cham,* 233-239.

Singla, M.K., Gupta, J., Nijhawan, P., Ganguli, S., & Rajest, S.S. (2020). Development of an Efficient, Cheap, and Flexible IoT-Based Wind Turbine Emulator. *In Business Intelligence for Enterprise Internet of Things, Springer, Cham,* 225-231.