Towards Evaluating Web Spam Threats and Countermeasures

Lina A. Abuwardih
Computer Information Systems Department
Jordan University of Science and Technology, Irbid, Jordan

***Abstract*—Web spam is a deceiving technique that aims to get high ranks for the retrieved web pages at the top Search Engine Result Pages (SERPs). This paper provides an evaluation for the web spam threats and countermeasures. It is started by presenting the different types of web spam threats that aim to deceive the users with incorrect information, distribute phishing and propagate malware. Then, presenting a detailed description for the proposed anti web spam tools, systems and countermeasures, and conducting a comparison between them. The results indicate that online real time tools are highly recommended solutions against web spam threats.**

***Keywords*—SEO; web spam threats; phishing; malware; web attacks**

I. Introduction

Nowadays, enormous amounts of queries are performed by many users on the search engines; with intentions like solving a problem, answering a question, finding information, or just for an entertainment. Regardless the query types, users want high-quality results and prefer the relevant ones that are displayed at the top of Search Engine Result Pages (SERPs). The retrieved web pages mainly are handled by ranking algorithms and the retrieved results are called organic results. In the contrast to the organic results, poor quality, and irrelevant information also are retrieved. Retrieving irrelevant results (web spam) makes search engines less credible and users become more annoyable [1].

Web spam is a fraud practice that imposes search engines to fetch web pages that are considered meaningless and do not meet the users needs. It is a fraud method because it is based on cheating and deceiving search engines, users, and webmasters from the real content. Web spam includes several types such as content web spam, link web-spam, cloaking, and doorway pages [2]. Conversely, Search Engine Optimization (SEO) is a legal way that is used to achieve high ranks for the retrieved web pages which are convenient to the user's queries. It is a type of Search Engine Marketing (SEM) models. SEO is considered a worthy business, some of the leading optimizers and SEO corporations regularly spend more than $20,000 per month for continuing in optimization [3].

The study presents the current web spam types, threats and related attacks that mislead the users with irrelevant information or hurt their devices by redirecting them to malicious links or phishing web pages. It provides a comprehensive description of several existence countermeasures and anti web spam systems. Finally, the study compares them to recommend the best appropriate solutions.

The paper is organized as follow: Section II briefly presents related studies of web spam types. Section III presents the main web spam threats. Section IV shows the Anti Web spam tools, systems, and countermeasures. Section V presents the conclusion of the study.

II. Web Spam

Web spam includes various types such as content-based, link-based, opinion based, and cloaking. Many studies explored different spam techniques and proposed several solutions to handle web spamming techniques.

- content web spam: In content web spam, the actual contents of the web pages are altered using different techniques. Typically, search engines are based on the TF-IDF algorithm of information retrieval for evaluation and ranking the retrieved web pages. Term Frequency measure (TF) computes the frequency of terms inside the documents, while the Inverse Document Frequency (IDF) is the proportion of the overall number of documents to the number of documents that hold the term. Spammers seek to raise TF of terms by exploiting the impairments of these models. Keywords stuffing in the main HTML elements, including $< title >< /title >, < meta/ >, < body >< /body >, and < a >< /a >$ tags is considered the most common technique that is used by spammers. The technique is based on packing the web page contents with many irrelevant and meaningless terms for raising the possibility of retrieving those pages at SERPs [4]. The various types of web content spam are surveyed by [5].
- Link web spam: it is categorized to in-link and out-link types [5]. Regarding the in-link web spam, the spammers raise the rank of the target pages by increasing the number of incoming links using different techniques such as link farm [6] and accessible pages [7]. Whereas, the out-link web spam is a trivial technique in which, spammers own their pages; therefore they can insert and modify any component of their pages. Also, they can easily reduplicate the whole web catalogue such as DMOZ and Yahoo! Directory and speedily conduct a large collection of trustworthy links [8].
- Opinion spam: It is imitation reviews which are written by fraudsters, which disturb customers and organizations to reach the actual results about the products. Hence, these reviews must be discovered

and removed in order to deny the possible tricky customers. An opinion spam survey is presented by [9].

- Cloaking: It is dedicated to fake web browsers and search engine through disguise the web page or a portion of the page which is not discovered using the visual examination. It can be applied legitimately to produce improving-suited pages for the index of a search engine, for example by giving a content without ads or navigational assistance. It can also be abused to expose users contents that are unrelated to the indexed contents [10].

III. Web Spam Threats

Web spam can be described as a real threat that publishes fake information which misleads the users for inaccurate results. In addition, web-spam is considered a source for phishing, malware and spam messages [7], [11]. Web attacks take a place when exploiting the vulnerabilities by attackers who grant means to malicious users to break the system's protection mechanisms. Usually attackers try to take advantages, acquire private information and system resources. There are several types of web attacks such as phishing and malware.

- Phishing: is a illegal technique to steal sensitive information such as user-names and passwords from the naive people. Phishers fabricate the web pages by creating duplicate pages from the genuine ones. It can be driven by transmitting an e-mail that looks to be sent from a committed organization to users by phishers. The phishers deceive the users by motivating them to submit their sensitive information through clicking on a phishing link that could be received by e-mail. Also, phishing can be found in blogs, forums, or file sharing [12].

- Malware: is a harmful software that targets computers and aims to hurt the users devices such as: computer viruses, worms, Trojan horses, spyware and adware [13].

Attackers employ wide-range techniques for phishing and malware dissemination in web spam pages such as setting a link in a barcode that navigates users to malicious (phishing or malware) web pages [14], [15]. Another technique is setting up a Trojan through a malicious email attachment or ads, which give the attacker the ability to exploit vulnerabilities and acquire important information [16].

IV. Anti Web Spam Tools, Systems and Countermeasures

Discovering the web spam is considered as important issue for the web society and takes more attention for the researchers from many fields. Several attempts have been established to combat the problem of web spam. Some of the proposed methods which are developed to discover the web spam are presented as follow:

A. Spam detection using machine learning technique

Machine learning techniques are based on developing programs that can learn from experience and discover knowledge from data. Machine Learning techniques are mainly categorized into supervised and unsupervised learning [17]. Several machine learning techniques include Bayesian network, Neural networks, Support vector machine, and Decision tree have been adopted to combat the different types of spams.

The researchers in [18] proposed a web Spam detection system called SAAD based on analyzing a set of web page content features to detect the spam. They applied many classification algorithms and the best detection results were obtained by a C4.5 classifier. The study was conducted on two common web spam datasets, webb Spam Corpus which includes 350,000 varied spam pages and WEBSPAM-UK2006/7 which includes more than 100 million Web pages.

The study of [19] introduced a Dual-Margin Multi-Class Hypersphere Support Vector Machine (DMMH- SVM) classifier for automatically categorizing web spam by type. Also, they proposed new cloaking-based spam features that help to obtain high web spam detection accuracy, precision, and recall percentages. The experiments were conducted on WEBSPAM-UK2007, ClueWeb09, and ECML/PKDD10. The experimental results showed that DMMH-SVM performed better accuracy than existing algorithms with new cloaking features.

The researchers in [20] conducted a comparative study on a WebSpam UK2007 dataset to evaluate the efficiency of various machine learning classifiers. These classifiers were Decision Tree, Naive Bayes, Random Forest, and RBF Network. They applied 10- fold cross-validation in order to evaluate their experiments and used F-measure scores as the evaluation metric. Their results showed that the Random Forest classifier obtained the highest F-measure value for detecting content and link spam.

The study of [21] was conducted based on a PU-learning algorithm that learned from a very few positive instances and unlabeled dataset. The study was carried on a dataset that had 800 positive opinion reviews. The obtained accuracy was 78.12% with F-score 76.67 using the k-NN classifier.

Table I summarizes the evaluations and limitations for anti web-spam machine learning techniques.

B. Spam detection using graph-based technique

This technique recognizes the web as a directed graph, the vertices represent the web pages and the links among web pages represent the edges. Web takes the style of bowtie shape and is arranged to five components based on the characteristics of links. Characteristics of the graph have been used in the discovery of spam.

The researchers in [22] introduced a new method through combining weight properties in order to improve the web spam detection algorithms. Weight properties are the influences of one web node to another web node. They altered the existing Web spam detection algorithms with their proposed method. For the performances evaluation, their experiments are carried on a large well-known Web spam dataset WEBSPAMUK2007. The performance of the altered algorithms performed better

TABLE I. COMPARISON OF THE ANTI WEB SPAM MACHINE LEARNING TECHNIQUES

System/Paper	Threat Detection			Evaluation	Limitation
	Web Spam	Phishing	Malware		
SAAD/ [18]	✓	—	✓	effective by improvement of 15% in the worst case and 27% in the best case	Real web browsers are not embedded with it, therefore, it cannot detect the actual risky web pages
DMMH- SVM/ [19]	✓	—	—	effective of categorizing web spam with higher accuracy, precision and recall than the state-of-art frameworks	All types of web spam cannot be categorized
[20]	✓	—	—	Random Forest is the most effective classifiers with higher F-measure among all features	All types of web spam cannot be categorized
[21]	✓	—	—	effective in discovering opinion spam with 78.12% accuracy	Unlabeled data cannot be handled

than the existing algorithms up to 30.5% enhancement at the host level and 6.11% enhancement at the page level.

The study of [23] proposed a framework that spread both trust and distrust web pages by assigning scores which were T-Rank for the trust web pages and D-Rank for the untrustworthiness. In the proposed framework, the spread of T-Rank/D-Rank was determined by the targets current possibility of being trustworthy/untrustworthy. Thus a page spread more trust/distrust to a trustworthy/untrustworthy neighbor than to an untrustworthy/trustworthy neighbor. They utilized T-Rank scores to recognize spam rank reduction and D-Rank scores to finish spam detection. The proposed Trust-DistrustRank (TDR) algorithm rebounded to TrustRank and Anti-TrustRank when the punishment factor was adjusted to 1 and 0, respectively. Also, TDR beat the cons of both TrustRank and Anti-TrustRank. The Experimental results showed that TDR performed better than other semi-automatic anti-spam algorithms for both spam rank reduction and spam detection. Table II presents the evaluations and limitations for anti web-spam graph-based techniques.

C. Natural Language Processing Technique

Natural Language Processing is dedicated to the investigation of text data of the web page. Language Analysis is conducted at two levels which are semantic level and syntactic level with intent to establish several assumptions. Commonly, the TF-IDF algorithm is applied in information retrieval and text mining. TF-IDF measures the importance of a word to a document in a corpus. The importance improves proportionally to the number of times in which a term occurs in the document but is neutralized by the frequency of the word in the corpus.

The researchers in [24] proposed a Bag-Of-Spam-Words (BOSW) technique for web spam detection. In the proposed method, they illustrated each document as a vector of certain words that were chosen from a spam corpus. They performed different feature selection techniques on a dataset that is conducted based on the Persian host and applied many classification algorithms to classify the Persian websites. Their results showed that employing the BOSW technique with the SVM classifier achieved the best performance in discovering Persian spam websites.

The study of [25] proposed a method to determine the spam pages using content, link-based, and integrate both content and link-based techniques. For the content based method, the researchers utilized the term density and the linguistic features using Part of Speech (POS) ratio test to identify the spam pages. While in the link-based method, they applied collaborative discovering through calculating personalized page ranking for all web pages to identify the spam pages and non-spam pages. The study was conducted for identifying the spam and non-spam pages by integrating the content and link-based techniques. Their study is conducted using the WEBSPAM-UK2006 dataset and the results of the proposed approach achieved 75.2% F-measure. Also, the results showed that the proposed approach outperformed the four spam detection techniques that were compared to their approach. Table III reports the evaluations and limitations for anti web-spam natural language processing techniques.

D. Anti Phishing Technique

Classical security tools such as anti-virus measures are not able to protect against all cyber-attacks. Most of the serious security issues take place due to humans unwitting mistakes, errors, culture, and knowledge which are not considered completely by existing security paradigms. The improvement of current cyber-security paradigms to conduct better user consciousness, counsel, and restraint in cybercrime will be needed [26]. Several systems were proposed to face the problem of phishing.

The study of [27] introduced a method using the associative classification that is called Multi-label Classifier based Associative Classification (MCAC). They characterized the features that identify the phishing websites, and provided a survey of the intelligent strategies adopted to deal with the phishing attack. The experimental results showed that associative classification and MCAC can discover the phishing websites and extract new rules.

The researchers in [28] developed an open-source plugin for the Chrome browser which is called AuntieTuna. The novel technique can automatically create personalized lists of candidate sites and check whether the sites are browsed by users. They utilized the cryptographic hashing of each pages that are viewed as Document Object Model (DOM), giving a zero false positive measure and classifying more than half of discovered Phishing pages. The importance of AuntieTuna can be shown when providing warnings on phishing pages before users expose their sensitive information.

Both [14] and [29] studies showed that several tagging technologies like barcodes can be used to launch phishing and malware attacks. The mechanism depends on tricking the users and connect them with a spam (or irrelevant content), where users can be easily under threats by just scanning these

TABLE II. SUMMARY OF THE ANTI WEB SPAM GRAPH-BASED TECHNIQUES

System/Paper	Threat Detection			Evaluation	Limitation
	Web Spam	Phishing	Malware		
[22]	✓	—	—	improve the baseline algorithm by 30% in discovering the Web spam for WEBSPAM-UK 2007 dataset	There is not a consideration for the bidirectional links between spam and obscure pages
TDR/ [23]	✓	—	—	perform better than the preceding anti-spam algorithms for both spam reduction and spam detection	It is not incorporated in the refinements of TrustRank and Anti-TrustRank like link variable and link credibility

TABLE III. COMPARISON OF THE ANTI WEB SPAM NATURAL LANGUAGE PROCESSING TECHNIQUES

System/Paper	Threat Detection			Evaluation	Limitation
	Web Spam	Phishing	Malware		
BOSW/ [24]	✓	—	—	effective by improving the detection of spam and non-spam in Persian websites	not consider the content-based and link-based features together that help in discovering spam in Persian websites
[25]	✓	—	—	effective in discovering the web spam with 75.2% F-measure	It is not considered the bidirectional links between spam and obscure pages

barcodes. Both studies proposed solutions based on applying digital signatures to authenticate barcodes, and protect users from phishing and malware attacks.

The researchers in [30] introduced a new technique based on the auto-updated white-list of legitimate sites accessed by the different user for defending against the phishing attacks. The proposed technique was characterized by fast access time and high discovery rate. The main idea of the proposed technique is warning the users from revealing sensitive information, when they open websites that are not listed in the whitelist. Moreover, the proposed technique examins the legitimacy of a web page based on the hyperlink features. The experimental results showed that the proposed technique was very efficient for defending against phishing attacks and it had 86.02% true positive rate while less than 1.48% false negative rate. Furthermore, the proposed technique was able to discover different types of phishing attacks including Domain Name System (DNS) poisoning, embedded objects, and zero-hour attack.

The study of [31] proposed a FeedPhish application to discover phishing attacks including zero-day and phishing sites that are hosted on settled domains. When the users access to a fraud website, the application analyzes the users' behavior. Then it automates the fake identity that is submitted by online users before they are submitting their real identity. If the login to the web page is done successfully, then the web page is categorized as phishing otherwise it tested with more filters. If the fake site succeeds through all filters then the website is categorized as a legitimate site. The experimental results showed that the proposed application gained a true positive rate of 97.61%, a true negative rate of 94.37% and total accuracy of 96.38%.

The researchers in [32] showed a novel technique which was called Phishing-Alarm, to discover phishing attacks based on the features that are arduous to shuffle by attackers. They introduced an algorithm that counts the distrust ratings of web pages using the similarity of obvious features among the web pages. They employed the Cascading Style Sheet (CSS) as the ground to count the visual similarity of each page component. The main rating method was used was based on weighted page-element similarity. They prototyped their technique in the Chrome browser. The proposed system was evaluated on real-world websites and the results showed the effectiveness of the proposed technique. Table IV reports the evaluations and limitations for anti phishing techniques.

E. Real Time Systems and Online Tools

This section presents a comparison between three main free existing real-time and online tools; PhishTank [33], Search Engine SPAM Detector [34] and Google Safe Browser [35] that are widely used in different applications. These tools are used to provide highly recommended level of security against web spam threats such as irrelevant content, phishing, and malware distribution.

- PhishTank: is a free collaborative web service (open API) that was developed for detecting phishing web pages. PhishTank considers the users voting regarding suspected phishing web pages through community-based phish verification system [33]. PhishTank is widely adopted by several browsers and other software such as Opera, Mozilla, Yahoo! and Kaspersky.
- Search Engine SPAM Detector: is a free web spam detector online tool that aims to analyze web pages through extracting web spam features [34]. This tool is mainly depending on three main groups of spam techniques; keyword stuffing, spam (doorway) farms and hidden text. The main limitation of SPAM Detector tool that cannot detect Javascript tricks for increasing the web pages rank on SERPs [34].
- Google Safe Browser: is simple, flexible and easy to use Google web service, which allows users to check suspected web pages against possible threats such as phishing, malware or unwanted applications. Google Safe Browsing uses Google techniques to detect dangerous web pages by checking Google blacklists for unsafe web pages [35]. Table V summarizes the evaluations and limitations for anti web-spam real-time systems and online tools.

V. CONCLUSION

Web spam is an illegal method to increase the rank for the web pages to appear at the top SERPs. Web spam is considered the main source for distributing phishing, malware

TABLE IV. SUMMARY OF THE ANTI PHISHING TECHNIQUES

System/Paper	Threat Detection			Evaluation	Limitation
	Web Spam	Phishing	Malware		
MCAC/ [27]	—	✓	—	effective with 94% accuracy of defining the phishy websites	It is not considered the content-based features that help in understanding the behavior of the attackers
AuntieTuna/ [28]	✓	✓	—	effective with 58.8% sensitivity and 100% specificity	—
[30]	—	✓	✓	effective in discovering phishy web pages based on the hyperlink information with 86.02% true positive rate and 1.48% false negative rate	It is not considered the content-based features that help in understanding the behavior of the attackers
FeedPhish/ [31]	—	✓	—	effective in discovering phishy websites with 96.38% accuracy	It is not addressed the Single Sign-On phishing websites
Phishing-Alarm/ [32]	—	✓	—	effective in discovery the phishy websites with 100% precision rate and 97.92% recall rate	—

TABLE V. COMPARISON OF THE ANTI WEB SPAM REAL TIME SYSTEMS AND ONLINE TOOLS

System/Paper	Threat Detection			Evaluation	Limitation
	Web Spam	Phishing	Malware		
PhishTank [33]	—	✓	—	Highly recommended solution against phishing attacks	It can not detect other web attacks
Search Engine SPAM Detector [34]	✓	—	—	Highly recommended real time solution against web spam URLs	It can not detect web spam Javascript tricks and can not detect Arabic web spam (which mainly used other spam features)
Google Safe Browsing [35]	—	✓	✓	Highly recommended solution against suspected web pages	—

and irrelevant content. The study highlights the web spam threats and summarized several web spam filtering/preventing approaches. The outperformed of comparing several tools and schemes indicates that online real-time tools are highly recommended solutions against web spam threats.

REFERENCES

[1] F. A. Zaghoul, O. Rababah, and H. Fakhouri, "Website Search Engine Optimization: Geographical and Cultural Point of View," in *Computer Modelling and Simulation (UKSim), 2014 UKSim-AMSS 16th International Conference on*. IEEE, 2014, pp. 452–455.

[2] Z. Gyongyi and H. Garcia-Molina, "Web Spam Taxonomy," in *First international workshop on adversarial information retrieval on the web (AIRWeb 2005)*, 2005.

[3] R. A. Malaga, "Search Engine Optimizationblack and White Hat Approaches," *Advances in Computers*, vol. 78, pp. 1–39, 2010.

[4] A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly, "Detecting Spam Web Pages Through Content Analysis," in *Proceedings of the 15th international conference on World Wide Web*. ACM, 2006, pp. 83–92.

[5] N. Spirin and J. Han, "Survey on Web Spam Detection: Principles and Algorithms," *ACM SIGKDD Explorations Newsletter*, vol. 13, no. 2, pp. 50–64, 2012.

[6] B. Wu and B. D. Davison, "Identifying Link Farm Spam Pages," in *Special interest tracks and posters of the 14th international conference on World Wide Web*. ACM, 2005, pp. 820–829.

[7] M. Erdélyi, A. Garzó, and A. A. Benczúr, "Web Spam Classification: a Few Features Worth More," in *Proceedings of the 2011 Joint WICOW/AIRWeb Workshop on Web Quality*. ACM, 2011, pp. 27–34.

[8] J. M. Kleinberg, "Authoritative Sources in a Hyperlinked Environment," *Journal of the ACM (JACM)*, vol. 46, no. 5, pp. 604–632, 1999.

[9] A. Heydari, M. ali Tavakoli, N. Salim, and Z. Heydari, "Detection of Review Spam: A Survey," *Expert Systems with Applications*, vol. 42, no. 7, pp. 3634–3642, 2015.

[10] A. Chandra and M. Suaib, "A Survey on Web Spam and Spam 2.0," *International Journal of Advanced Computer Research*, vol. 4, no. 2, p. 634, 2014.

[11] G. Canfora and C. A. Visaggio, "A Set of Features to Detect Web Security Threats," *Journal of Computer Virology and Hacking Techniques*, vol. 12, no. 4, pp. 243–261, 2016.

[12] M. Kaytan and D. Hanbay, "Effective Classification of Phishing Web Pages Based on New Rules by Using Extreme Learning Machines," 2017.

[13] A. Malhotra and K. Bajaj, "A Survey on Various Malware Detection Techniques on Mobile Platform," *Int J Comput Appl*, vol. 139, no. 5, pp. 15–20, 2016.

[14] F. Razzak, "Spamming the Internet of Things: A Possibility and its Probable Solution," *Procedia computer science*, vol. 10, pp. 658–665, 2012.

[15] R. Focardi, F. L. Luccio, and H. A. Wahsheh, "Security Threats and Solutions for Two-Dimensional Barcodes: A Comparative Study," in *Computer and Network Security Essentials*. Springer, 2018, pp. 207–219.

[16] B. Gupta, A. Tewari, A. K. Jain, and D. P. Agrawal, "Fighting Against Phishing Attacks: State of the Art and Future Challenges," *Neural Computing and Applications*, vol. 28, no. 12, pp. 3629–3654, 2017.

[17] J. Brownlee, "Supervised and Unsupervised Machine Learning Algorithms," https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/, 2016.

[18] V. M. Prieto, M. Álvarez, R. López-García, and F. Cacheda, "Analysis and Detection of Web Spam by Means of Web Content," in *Information Retrieval Facility Conference*. Springer, 2012, pp. 43–57.

[19] S. Kumar, X. Gao, I. Welch, and M. Mansoori, "A Machine Learning Based Web Spam Filtering Approach," in *Advanced Information Networking and Applications (AINA), 2016 IEEE 30th International Conference on*. IEEE, 2016, pp. 973–980.

[20] M. Iqbal, M. M. Abid, U. Waheed, and S. H. Alam Kazmi, "Classification of Malicious Web Pages Through a j48 Decision Tree, anaïve bayes, a RBF Network and a Random Forest Classifier for Web Spam Detection," *International Journal of u- and e- Service, Science and Technology (ijunesst)*, vol. 10, no. 4, pp. 51–72, 2017.

[21] R. Narayan, J. K. Rout, and S. K. Jena, "Review Spam Detection Using Semi-supervised Technique," in *Progress in Intelligent Computing Techniques: Theory, Practice, and Applications*. Springer, 2018, pp. 281–286.

[22] K. L. Goh, R. K. Patchmuthu, and A. K. Singh, "Link-based Web Spam Detection Using Weight Properties," *Journal of Intelligent Information Systems*, vol. 43, no. 1, pp. 129–145, 2014.

[23] X. Zhang, Y. Wang, N. Mou, and W. Liang, "Propagating both Trust and Distrust with Target Differentiation for Combating Link-based Web Spam," *ACM Transactions on the Web (TWEB)*, vol. 8, no. 3, p. 15, 2014.

[24] E. Rabbani and A. Shakery, "Analyzing Content-based Heuristics for Persian Web Spam Detection," *International Journal of Information & Communication Technology Research*, vol. 6, no. 3, pp. 25–39, 2014.

[25] R. K. Roul, S. R. Asthana, M. Shah, and D. Parikh, "Detecting Spam Web Pages Using Content and Link-based Techniques: A Combined Approach," *Sadhana*, vol. 41, no. 2, pp. 193–202, 2016.

[26] A. M. Shabut, K. Lwin, and M. Hossain, "Cyber Attacks, Countermeasures, and Protection Schemes A State of the Art Survey," in *Software, Knowledge, Information Management and Applications (SKIMA), 2016 10th International Conference on*. IEEE, 2016, pp. 37–44.

[27] N. Abdelhamid, A. Ayesh, and F. Thabtah, "Phishing Detection Based Associative Classification Data Mining," *Expert Systems with Applications*, vol. 41, no. 13, pp. 5948–5959, 2014.

[28] C. Ardi and J. Heidemann, "Auntietuna: Personalized Content-based Phishing Detection," in *NDSS Usable Security Workshop (USEC)*, 2016.

[29] R. Focardi, F. L. Luccio, and H. A. Wahsheh, "Usable Cryptographic QR Codes," in *2018 IEEE International Conference on Industrial Technology (ICIT)*. IEEE, 2018, pp. 1664–1669.

[30] A. K. Jain and B. Gupta, "A Novel Approach to Protect Against Phishing Attacks at Client Side Using Auto-updated White-list," *EURASIP Journal on Information Security*, vol. 2016, no. 1, p. 9, 2016.

[31] R. Srinivasa Rao and A. R. Pais, "Detecting Phishing Websites Using Automation of Human Behavior," in *Proceedings of the 3rd ACM Workshop on Cyber-Physical System Security*. ACM, 2017, pp. 33–42.

[32] J. Mao, W. Tian, P. Li, T. Wei, and Z. Liang, "Phishing-Alarm: Robust and Efficient Phishing Detection via Page Component Similarity," *IEEE Access*, vol. 5, pp. 17 020–17 030, 2017.

[33] PhishTank, "Phishing Detection," https://www.phishtank.com/, 2015.

[34] "Search Engine SPAM Detector," 2018, http://tool.motoricerca.info/spam-detector/.

[35] "Google Safe Browser," 2018, https://safebrowsing.google.com.

Lina A. Abuwardih obtained her Master degree in Computer Information Systems (CIS) from Yarmouk University, Jordan, 2017. She is working as a teacher assistant in Jordan university of science and technology (JUST). Her research interests include: Information Retrieval, Data Mining, software engineering and Information Security.