Security techniques for intelligent spam sensing and anomaly detection in online social platforms

Monther Aldwairi[1], Lo'ai Tawalbeh[2]
[1]Network Engineering and Security Department, Jordan University of Science and Technology, Jordan
[1]College of Technological Innovation, Zayed University, UAE
[2]Department of Computing and Cyber Security, Texas A&M University-San Antonio, USA

Article Info

Article history:

Received Oct 23, 2018
Revised Aug 21, 2019
Accepted Aug 30, 2019

Keywords:

Artificial neural networks
Intelligent spam sensing
Machine learning
Malicious online communities
Online social networks
Privacy

ABSTRACT

The recent advances in communication and mobile technologies made it easier to access and share information for most people worldwide. Among the most powerful information spreading platforms are the Online Social Networks (OSN)s that allow Internet-connected users to share different information such as instant messages, tweets, photos, and videos. Adding to that many governmental and private institutions use the OSNs such as Twitter for official announcements. Consequently, there is a tremendous need to provide the required level of security for OSN users. However, there are many challenges due to the different protocols and variety of mobile apps used to access OSNs. Therefore, traditional security techniques fail to provide the needed security and privacy, and more intelligence is required. Computational intelligence adds high-speed computation, fault tolerance, adaptability, and error resilience when used to ensure security in OSN apps. This research provides a comprehensive related work survey and investigates the application of artificial neural networks for intrusion detection systems and spam filtering for OSNs. In addition, we use the concept of social graphs and weighted cliques in the detection of suspicious behavior of certain online groups and to prevent further planned actions such as cyber/terrorist attacks before they happen.

Corresponding Author:

Lo'ai Tawalbeh,
Department of Computing and Cyber Security,
Texas A&M University-San Antonio,
TX, 78259, USA.
Email: ltawalbeh@tamusa.edu

1. INTRODUCTION

Recent advances in mobile technologies, faster Internet connectivity and wider access provided vast number of applications and services in different life sectors including education, industry, social life, and health sector [1]. Also, the wide spread of the Internet made it easier to access information and news for most of the people all over the world. Moreover, the uncontrolled development in the web-based life applications coming from the competition between the international organizations in the field of social correspondence, made sharing and spreading of any information a lot simpler and agreeable [1].

In addition, the diverse and growing number of industry players in the field of news writing and spreading leads to creating of news articles that are not totally obvious or even totally false. It is possible that the creation of this type of articles has been made by mistake, on the other hand sometimes there are many deliberate and specialized sites for spreading fake news, to achieve political, economic, social, cultural, security and other goals [2].

Moreover, among the most powerful information spreading platforms are the Online Social Networks that allow Internet-connected users to share all different kinds of information at any time and from anywhere. With the spread of OSNs, online communities arise, that share current statuses, mood, thoughts, instant messages, tweets, photos, videos, and other information. Consequently, and as social networks reach spreads exponentially, the risk of catastrophic impacts of the rapid spread of false news is increasing dramatically.

In the same context, false news always leads to misinform of the general public and stir up strife and problems. In particular, in the last few years the spreading of fake information and spam on the Internet has drawn expanding consideration and has achieved the purpose of impacting political and social substances. In fact, spreading false and fake (spam) information is a worldwide issue and tremendous efforts should come together to limit or stop the spread of misinformation and utilize the necessary countermeasures to prevent this increasing trend. Among the countermeasure that can be used is the cryptographic algorithms suitable for wireless environments [3].

There are many factors contributed to the wide spread of the online social networks (OSNs). Among these, are the increasing popularity of mobile smart devices, easy access to the Internet, and advances in wireless communications and mobile cloud computing services [4]. The OSNs users' publicly share their photos, locations, activities, and even sometimes crucial personal information. Like any new technology, OSNs are expected to have their fair share of vulnerabilities and subject to exploits that might compromise the security of user's data. Especially because OSN mobile users are always logged in and connected to Internet through mobile devices. For example, OSN users maybe anonymous, but at the same time, this anonymity makes them more liable to be at risk of eavesdropping and spoofing. Therefore, security challenges, privacy issues, and identity authentication are much more complicated for social networks compared to other types of networks [5].

Security and privacy issues are critical when it comes to OSNs, because sensitive information is involved. In mobile social networks, these issues are even more critical, because misbehavior is very common between teenagers and those who think their anonymity shield them from responsibility. Misbehavior in OSNs includes sharing or posting verbal, written, images, comments and assaults videos', which might result in dramatic negative impact on young users. Moreover, novice users often share important private information about their daily lives such as their identity, family, contacts, and location. Therefore, privacy and security of OSNs are considered a vital research area. Many classic security policies are used to protect data and users over mobile social networks such as access control and identity management [6].

However, classical techniques and policies do not provide the required level of protection, and there are always new vulnerabilities that open the possibility for new attacks. Computation Intelligence (CI) techniques are very promising to achieve the needed higher levels of protection. CI methods include Fuzzy Systems, Artificial Immune Systems and Artificial Neural Networks [7]. CI are used in developing the Intrusion Detection Systems (IDSs) for Mobile Social Networks (MSNs) based on their capability to classify security attacks and then generate rules or signatures to describe them.

Intrusion Detection Systems are used to protect networks from security violations and malicious behavior by spotting attempts to compromise the security of protected networks. IDSs generate reports about every violation that occurs on the network. Computational intelligence techniques are very suitable to monitor, detect, analyze and respond to unauthorized activity in dynamic and changing mediums such as Mobile Social Networks [8].

In this work, we focus on two techniques of Computational Intelligence; Artificial Neural Networks (ANNs) and Machine learning techniques and their use for classifying, selecting and responding to possible attacks on Online Social Networks. ANNs are networks of massive number of artificial neurons that generate, process and analyze distributed information. ANNs are a self-learning and self-organizing networks, which make them suitable to make decisions quickly and solve problems of high ambiguity. Machine learning techniques are also studied in this work and the possibility of applying them in spam detection (mainly fake news is investigated).

The contributions of this paper are addressing how online social networks can be structured as social graphs with the users represented as nodes on the graph. Moreover, based on the observation that certain community or group over an OSN can form an entity with their social activities, we proposed the trustworthiness concept to identify the communities/entities with malicious activities. Also, we proposed new approach that uses the concept of weighted cliques in the detection of suspicious behavior of certain online groups and to prevent further planned actions such as cyber/terrorist attacks before they happen.

The following section explains related work in detail. Section 3 describes the solutions currently in use for detecting spam and fake news based on Machine Learning techniques over OSNs. Section 4 defines Artificial Neural Networks concepts followed by the use of Artificial Neural Networks for Intrusion Detection and Spam Filtering over Online Social Networks in Section 5. In Section 6, we present malicious activities detection techniques over OSNs using the social graphs and the weighted cliques. Section 7 concludes this work.

2. LITERATURE REVIEW

Online Social Networks allow Internet-connected users to share all different kinds of information at any time and from anywhere. With the spread of OSNs, online communities arise, that share status, mood, thoughts, instant messages, tweets, photos, videos, and other information. However, some parties misuse what is shared over OSNs according to the chosen privacy settings [9]. The OSN owners benefit from collecting the shared information and delivering targeted advertisement. Those ads and the associated information constitute a risk to everyone's privacy. Moreover, the privacy settings protect the user's information from other users, but not from the administrators and owners of that social network. Many previous related works discussed user's privacy issues and security concerns in social networks and proposed some solutions [10].

Consequently, the rapid spread of OSNs, the lack of proper privacy settings and the sharing of sensitive information leave the users exposed to various kinds of threats [11]. What makes matters worse is that OSN users add people they do not know to their networks, for example, 80% of Facebook users accept all friend requests even from complete strangers. Adding strangers is risky because you disclose your private data to strangers, and you put your legitimate friends at risk as well [12].

As social networks reach, spreads exponentially, so is the dangers of intrusions, viruses, click-jacking, phishing-attack, spam, social bots , Sybil attack, clone and identity theft attacks to access intellectual property published contents [13]. Every OSN user may be susceptible to malware infections from social media sites and applications as Drive by Download [14]. According to [15], more than half of the attacks on Facebook come from third-party applications that notify and attract users to open links that open viruses. For example, 20% of these applications try to tell the users they provide functionalities that Facebook does not offer, such as applications that show who viewed your Facebook profile and how many times. In click-jacking attack, malicious links of different content from what the user expects are interpolated onto social network sites. Such links trick OSN users to insert their confidential information, to control their computers, or to steal their accounts. In online social networks, phishing attacks occur by creating fake accounts to impersonate trustworthy third party to gain access to users' sensitive information [16].

According to specific analysis of many studies on phishing attacks on several datasets, fraudsters regard social media as the easiest platform to perform phishing attacks. The results show that phishing attacks on social media have the highest chance of success [16]. Other studies show that almost one third of phishing attacks target social networking, while the rest of the attacks target financial and e-payment companies. In social networks, the danger of spammers increases when users follow so many users who post messages with commercial URLs. Following such users causes disturbance in users' activities over the social network and leads to a significant misunderstanding of several messages and posts, which results in pernicious behavior. There are many techniques proposed to detect and identify such wide spectrum of attacks over networks [17].

Users' privacy over social networks can be protected by anonymization, that is "stop sharing private information such as names and addresses with strangers" [18]. User's data is deanonymized and shared with third parties for research and advertising purposes. Nonetheless, that poses a serious threat to user's privacy. In Sybil attack users assume fake identities to ruin a voting application, change its results, swing the conversation mood and damage the social network or page/account reputation. Therefore, there is a need for a resilient and robust applications and platforms to defend against such possible attacks [20, 21].

More recently, social bots, a new type of Internet bots, which are software scripts automatically performing simple actions. Social bots are intelligent programming mimicking human behavior and perform simple tasks such as simple customer service, online check-in, delivery address, etc [22]. Another serious and dangerous threat is the clone and identity theft attack. This attack takes place by duplicating a specific user account over the same social network or a different one and repeating all of his activities at similar times [23]. In other words, it is cloning the presence of a specific online user on a social network to deceive the actual friends, potential friends or followers to continue the trust relationship with the cloned account [23].

Based on the above related work review that indicates the increasing quantity and quality of cyber-attacks on the different online social networks, we can see there is an increasing demand to secure the infrastructures and platforms that are used to access these online social networks. For example, there are many attempts to propose secure cloud computing platforms that provide strong users authentication and authorization [24]. And since the cloud computing environments implementation is costly, the proposed secure platforms are first tested and simulated using best cloud simulators to verify their security and performance levels [25].

3. CURRENT MACHINE LEARNING SOLUTIONS FOR INTRUSION AND SPAM DETECTION OVER OSNs

Today, Facebook, Twitter, Instagram, Google+, Snapchat, and other social media networks shaping our social lives, by effortlessly staying connected with friends and family. Nevertheless, OSN users need to make a conscious decision about every piece of personal information he/she shares, because the aforementioned threats and intrusions. In addition, every social network has its privacy and security settings that govern the online experience and protect user's information [26]. The privacy and security settings are adjusted through collaboration between academic researchers, security companies, and delegates for every social media network to cope with current security and privacy threats. Therefore, the security and privacy settings must be updated and enhanced at regular basis.

The operators of social networks improved authentication by including new options to prevent possible threats and attacks. Among these authentication techniques are the two or three-factor authentication, such as using mobile phone numbers to verify their accounts. In addition, tokens or one-time pads are being sent to users to verify social networks accounts [27].

On the other side, social networks operators allow users to configure their own privacy settings such as limiting who can see their profiles, who can contact them, and the option to block certain users. In order to achieve automated future protection data analysis and classification techniques are required [28]. The further protection can include options like abuse, spam messages and privacy policy violations reporting. Moreover, efficient implementations of cryptographic algorithms can be used to provide confidentiality for users data [29]. The authors in [30] proposed an efficient programmable elliptic curve cryptography implementation, which is considered to be very secure crypto system. In similar context, the authors in [31] proposed a scalable crypto processor that can be used to provide confidentiality for different applications with different operand sizes.

On the other side, information security companies have a significant role in providing better protection by developing security tools such as the ZoneAlarm SocialGuard software that offers high protection from strangers and dangerous links on Facebook. Several researchers investigated new solutions for intrusion detection and threat prevention in social networks. Stringhini *et al.* [32] developed a technique to detect spammers in social networks. They show that it is possible to automatically define which accounts spammers are using. During their study, the research team collaborated with Twitter and using their technique, they detected and deleted around 16000 spammer profiles. Another work by Gao *et al.* [33] proposed an online spam-filtering system that inspected messages sent from users in real-time before reaching the recipients. The authors suggested to reconstruct spam messages and classify them into campaigns, and so, the messages will be examined in campaigns rather than individually. On the other hand, fake news is considered as spam data and unfortunately, it can be spread over the OSNs very fast. Many previous related works discussed the wide spread of fake news over social networks and proposed some solutions for fake news detection by applying machine learning techniques over Online Social Networks.

The authors in [34] identified two opposite ways to detect the fake news: human intervention and using algorithms. The first approach depends on the users to flag the fake news by fact checkers from media organizations such as the Washington Post, Snopes.com, and the French newspaper Le Monde that has a specialized fact-checking unit who developed a web extension Decodex. The second approach is to use algorithms to validate the information sources and identify fake contents. In their opinion, this approach has not yet gained the necessary robustness to accurately verify which information is false or which is not.

Machine Learning has been used to detect fake news based on news content and social context features [35]. The existing Boolean crowdsourcing algorithms work well when used to classify a post with social interactions is above a certain threshold. The performance might go down when the social interactions are below that threshold. Based on that, the authors proposed content-based methods to be used as well. The paper combined content- and social-based approaches by computing a score and classifies posts exceeding a threshold λ. The score depends only on social interactions i.e. number of likes and shares on Facebook, retweets and follows on a Tweeter, etc.

Social spammers change their spamming strategies to trick deployed anti-spamming systems, which creates the need for more efficient anti-spamming techniques to protect social networks users. The authors in [36] indicated that social bots can be used populate social systems. In most of the cases, the social bots are used for useful purposes but in other cases, it can be very harmful by deceiving the Online Social Networks users. For example, they can be used to influence elections, tamper the stock market, and spread fake news to serve certain agenda. They also mentioned that there are many proposed systems to detect social bots. Some of these systems utilize crowdsourcing strategies, feature-based supervised learning, and hybrid systems.

We can say that systems that depend on machine learning techniques are the best candidates for insuring spam-free social networks [37]. Machine learning techniques lean on electing knowledge from previously sent spam items and then use the acquired information to predict the behavior of newly received spam and classify them. The authors in [38] proposed an efficient classifier to predict and detect spammers' actions using feature relevance analysis on social network is developed

Zheng *et al*. [19] proposed an effective spammer detection system based on supervised machine learning solution. This system considered user's content and behavior features, and then applied them into the SVM for spammers classification. According to the experiments, this system showed excellent performance. Suganya, and Hemalatha [39] combined user's content and behavior features with machine learning to implement spam classification method and their experiments showed interesting results. Hwa *et al*. [40] focused on sending spam using sets of thousands of fake accounts. Authors provided a machine-learning pipeline that classifies fake accounts into clusters according to their actors.

Fahim, Mutahira and Naseem [41] presented the reason behind the behavior of Facebook spammers. They also proposed a methodology to filter Facebook spam using Artificial Neural Network to detect each unusual action that may lead to spam sharing or post by studying the behavior of all friends. Assuming social networks make our social lives simpler without worries about privacy and security concerns, authors of [42] talked about the big role Machine Learning techniques play in OSN privacy. They focused especially on Artificial Neural Networks and Genetic Algorithm as they both show extra intelligence and prediction that is more accurate.

4. ARTIFICIAL NEURAL NETWORK

Artificial Neural Network is a branch of Artificial Intelligence (AI) that is based on the neural structure of the human's brain [43]. ANN aims to convert a specific input into significant output using hidden artificial neurons, which are considered the main processing elements in ANN that can be used to develop many applications [44]. Each neuron is programmed to accomplish specific operation according to which data will flaw from neuron to another. These neurons are organized in specific layers through which input data move until the output is produced. In other words, the output is an emergent result of each operation performed by every neuron data reach [45]. Building Artificial Neural Model to solve specific problem needs intensive knowledge about the problem, the ANN itself, and the working plan [45].

Each Artificial Neural Model has *input* (data to be processed), the *output* (resulted information), neurons, and *weights* for them. The model also contains the *operations* (mathematical functions) that determine, which neuron data need to be activated [46]. The high weight of a neuron indicates strong data to be operated. By setting the weight of every neuron using particular algorithm implemented specifically for this reason, the output will be produced for specific input [46]. Almost all ANNs have the same structure. Figure 1 and Figure 2 present the basic structure of an Artificial Neural Network and an Artificial Neuron, respectively.

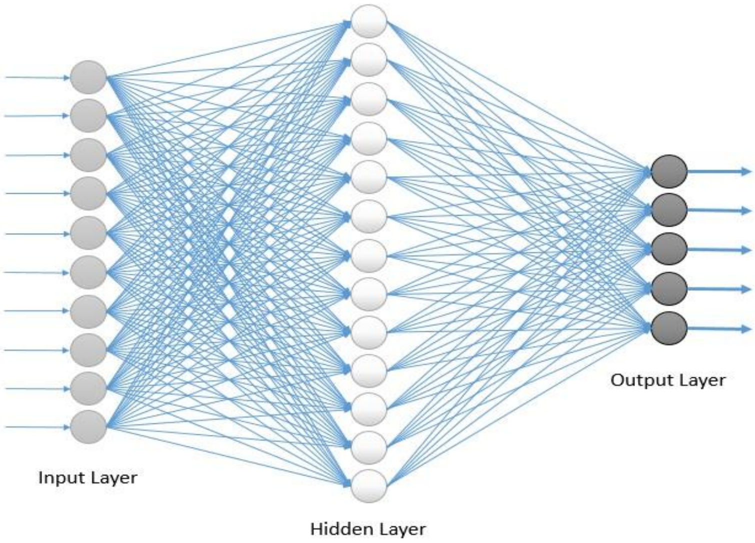

Figure 1. Basic structure of an artificial neural network

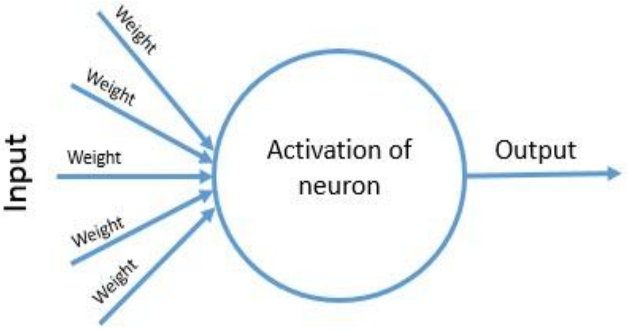

Figure 2. An artificial neuron

Artificial Neural Networks are more than artificial neurons grouped into layers and connected through communication lines. According to Figure 1, there are three kinds of neurons. There are neurons that receive the input from the real word, neurons that send the output to a secondary processing and controlling system, and neurons that are hidden from view. The neurons are distributed into several layers. Each neuron in the hidden layer receives input from all input neurons and sends output to all output neurons after performing its correlated function. On the other side, there are three types of communication lines that connect neurons together.

There are connections that let next neurons' summing mechanism add, and other connections let them subtract. Some ANNs have another type of connections, called feedback connection lines. These lines are used to route back the output from the output layer to the hidden layer as it can be seen from Figure 3.

After structuring an ANN for a specific application, the network begins to learn. Training the network happens in two different approaches. The first approach is the supervised training in which we supply the network with output either by rating the performance of the network or by providing the output along with its input. In the second approach, no outside help is provided to the network, and it should portend the input according to specific characteristics.

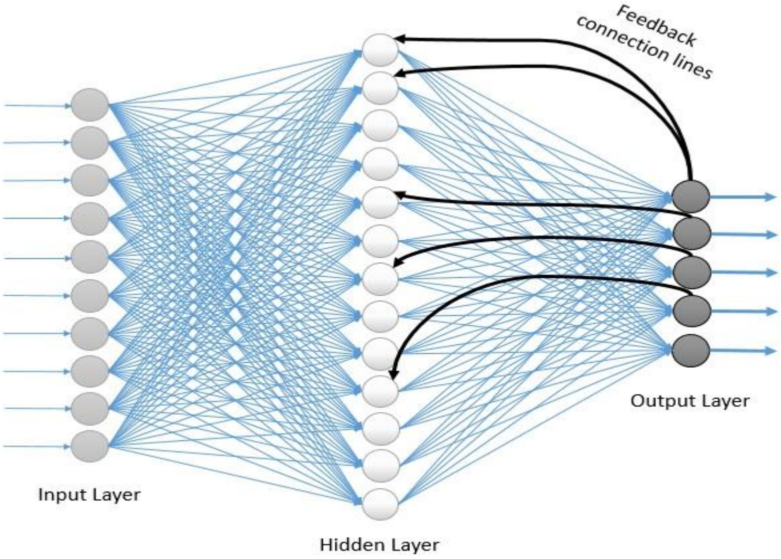

Figure 3. Feedback connection lines

Artificial Neural Networks have been successfully applied to several real-world fields. For example, it is applied in finance (e.g. credit rating), medicine (e.g. patient diagnosis), industry (e.g. process and quality control), and science (e.g. character recognition) [47]. Moreover, the ANN can be applied in education (e.g. teaching neural networks), energy (e.g. electrical load and demand forecasting), and other miscellaneous fields [48].

Artificial Neural Networks are recently used in intrusion detection systems, threat prevention systems, and spam prediction systems. However, there is not much work done on using Artificial Neural Networks for security concerns in OSNs. We believe that it is going to be interesting, valuable and contributory to study the availability of ANN's security applications for ensuring the required security and privacy in OSNs. Basically, in the field of security and privacy in OSNs, ANNs will be utilized in two ways, distinguishing normal accounts from spam accounts and designing detection features [49]. In both cases, ANN security-ensuring systems need to be updated frequently.

5. ARTIFICIAL NUERAL NETWORK INTRUSION DETECTION AND SPAM FILTERING OVER SOCIAL NETWORKS

Detecting spam emails and social network spam posts can benefit from applying the same techniques, because of the striking similarities according to [50]. Online Social Networks malicious community represented by spammers is getting more dangerous. A proposed but not perfectly explored strategy is to structure an Artificial Neural Network for Spam detection over social networks. In general, to approximate specific functions by ANN, there are difficulties in setting up its structure, deciding hidden nodes, and dealing with its complex parameters like weights of connections and learning rates.

To overcome such difficulties, ANNs are applied along with Genetic Algorithm to enhance the performance of spam detection and classification [51]. The authors in [51] proposed a combination of both ANN and GA to come up with a new hybrid algorithm that beats the conventional ANN. According to the improvement on spam detection accuracy, the proposed hybrid algorithm can be implemented to detect spam messages on OSNs. The authors of [52] proposed a system that focused on the main body of the spam and checked it word by word using ANN. Each word in the message is given a specific weight based on its probability to be a spam word. According to these weights, the message is blacklisted or whitelisted. If a message is blacklisted, then it is sent from domains that are restricted to spammers. If the message is whitelisted, then it is sent from trusted domains.

In addition to distinguishing legitimate from ham messages, this research develops a technique using Optical Character Recognition (OCR) tools to extract spam message embedded in images. According to the proposed system, the text spam-detection, and extracted text from image spam-detection are very important to be utilized in Social Network, because same types of spam are being spread via these Networks. Applying such system to OSN spam detection suits will help in reducing wasted time and memory and will protect personal data from being harmed because of spam-spreading.

The work in [53] took into consideration the task of Text Classification (TC) of spam messages. The authors proposed an anti-spam filtering system that uses ANN for multilayer protection, and a Genetic Algorithm to train their protection system. Applying this system showed high level of accuracy when used to distinguish ham messages from spam messages. From this research, we derive that subject and body fields always contain specific indications that ease the process of distinguishing ham from spam messages. On the other hand, this system proved that 15-30 hidden neutrons are good enough to process easy messages and classify them. In Social Media, this system can be utilized to detect spam messages if authors solved the problem of long detection time.

Another hybrid ANN was proposed by [54]. In this research, they used Radial Basis Function Neural Networks (RBFNN) along with Particle Swarm Optimization (PSO) to reach better accuracy and effectiveness. This method is very appropriate to be utilized in OSNs because it used improved network architecture and learning algorithm. In [55], the authors proposed another multilayer ANN method, and they called it "antidote." This method is very special because it is designed to serve each user by using his chosen parameters to set an appropriate multilayer ANN according to which messages are going to be classified into spam and legitimate messages. This system can be applied to Social Media security suites due to its flexibility and short learning time.

ANNs have great potential in OSNs intrusion detection; unfortunately, they have not been fully investigated in the literature. In general, IDSs can catch misuses and stop them from causing damages. For the same reasons, ANN-based intrusion detection methods can be applied, with some modifications, to Social Network platforms. Al-Jarrah and Arafat [56] used Time Delay Dynamic Artificial Neural Network (TDDNN) to identify each attack behavior. They designed their system to generate alerts when the ANN classifier recognizes an attack. Producing the attack features takes short and constant time starting from recognizing the attack presence to generating the attack alert. Because of its fast intrusion recognition, this system is very compatible with Social Networks security and privacy needs, especially because OSN attackers are very aggressive.

Qiu and Shan [57] used multiply swarm optimization-back propagation MPSO-BP neural network for their proposed model of intrusion detection. PSO algorithm is used to optimize back propagation ANN's parameters. Thus, the proposed model showed an improved effect on the intrusion detection rates in comparison with PSO-BP neural network and BP neural network. This model is suitable for Social Network Platforms, because it can handle significant amounts of data simultaneously; in addition to its independent learning and regular database updates.

In [58], the authors proposed supervised back propagation ANN based Anomaly Detection System. Their system aims to catch all attempted anomalies and keep all data completely safe and it concentrates on the hierarchy anomaly IDS. The proposed system showed higher accuracy, efficiency, and performance because they use only 17 KDD 99 features. They followed features reduction technique that are appropriate for Social Network Platforms as accuracy reaches 98% and training and testing time is reduced to the minimum.

In [59], the authors compared the accuracy achieved by applying several methods to Anomaly Detection System. The methods they studied in their work are Genetic Algorithm with Artificial Neural Network (GA-ANN) Classifier that used 18 features. Other methods they used are the Modified Mutual Information Feature Selection (MMIFS) with 24 features, Linear Correlation Feature Selection (LCFS) with 21 features, and Forward Feature Selection (FFS) with 31 features. According to their study, GA-ANN classifier raised the accuracy of detecting anomalies to the maximum of 99% making it an excellent candidate for Social Network Platforms.

6. DETECTION OF MALICIOUS ACTIVITIES AND COMMUNITIES' BEHAVIOR OVER OSNs

In general, the OSN malicious communities share many of the following distinguishing observations:

a. Social media spammers are the most perspicacious among all kinds of spammers.
b. 40% of all social media accounts are marked as spam accounts.
c. Nowadays, most of malicious contents are being sent and shared by automated spamming tools. Such tools send spam efficiently especially when targeting groups of users.
d. Spam accounts tend to be connected (Friends or following each other), because they usually send following and friendship requests with no specific consideration to the quality of the accounts they contact.
e. Spamming accounts tend to accept all friend requests they receive and follow back all accounts they follow them.
f. When there is a specific inner relationship between spamming accounts over social networks, they can be exposed easily.
g. Spamming accounts share topics that attract the targeted victims, and these topics are usually similar across most of the spamming accounts. One interesting topic is to fake celebrities' accounts.
h. Malicious accounts tend to stay active for protracted periods and keep damaging as long as they are active.
i. Spammers search all popular accounts to reach the private information of their followers or friends and use them in their crimes.

6.1. Clique-based detection methodology

An OSN user is usually represented by an account or a profile. The profile describes the user's social related attributes that include his or her name, the list of contacts, and their hobbies or interests. There are two types of relationships between the accounts: it could be either one sided as in Twitter or could be two-sided as can be seen in Facebook friendships. As mentioned earlier, the online social network users can share videos, photos, locations and even more personal information such as birthdays and phone numbers. It is worth pointing out that even Facebook and Twitter are among the most popular OSNs, there are other online social platforms such as Google+ and LinkedIn. Such networks help users from different geographical areas to stay connected. Moreover, these online networks allow their users to establish new relations with different people all over the world who have similar interests including the same profession or hobbies.

Online social networks can be represented as a social graph. Let's assume the group of users within a certain social network consists of three users who are represented as nodes {A, B, C} on that network. The relationship between these users are represented by edges. This social graph can be used to identify "ties" between the users (nodes). These ties could give common details and interests of the group members such as their gender, personal interests, sports, education level, and many other important details. Now, technically speaking these groups over online social networks that have all the users as friends are called sometimes cliques. Figure 4 shows the social graph with every user befriending all the other users within this group. The users (nodes) {A, B, C, D, E} are all friends of each other and so they are called cliques. To clarify the concept of cliques, Figure 5 shows the same users within the same group represented as nodes on the social graph, but they are not clique for each other. As can be seen from the figure, the users C, D, and E are not friends for each other.

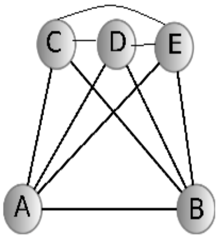

Figure 4. Five Clique Users on a Social Graph

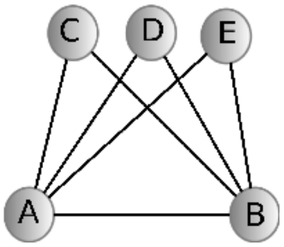

Figure 5. Five Non-Clique Users on a Social Graph

As it is known, the users of online social networks spend good amount of their time socializing with friends and people whom they share the same interests with. There is a high possibility that some malicious behavior or thoughts and posts might take place within these socialization activities. Here comes the important role of the social graphs to build a trustworthiness model based on different social activities that take place between users within the same group [60]. It is believed that the social graphs can be constructed specially between cliques to capture different types of social activities. These activities range from just retweeting to posting new tweets [61]. Also, activities to be monitored and captured include sharing false information, viral images and videos on Facebook and other social platforms.

Using social graphs, we can propose many levels of trustworthiness. After that, malicious activities can be detected based on their level of trustworthiness. In other words, the users and their social relationships and activities over the online social networks are grouped as distinguished entities. And so, by measuring how much each social activity is trustful, we can distinguish the legitimate activities from the malicious activities within the clique network. Figure 6 shows the different methods used to detect malicious activities over online social networks including the social graph.

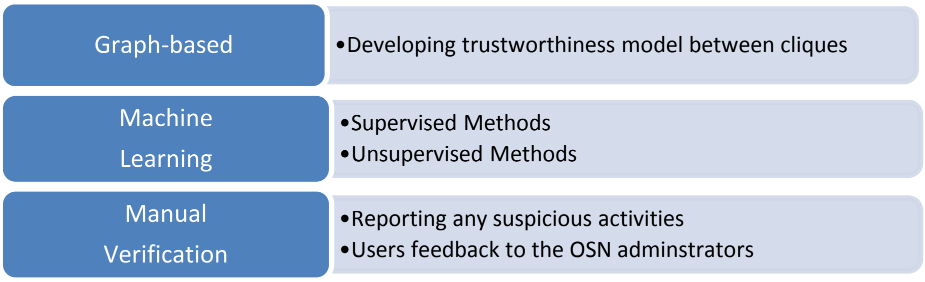
Graph-based
•Developing trustworthiness model between cliques
Machine Learning
•Supervised Methods
•Unsupervised Methods
Manual Verification
•Reporting any suspicious activities
•Users feedback to the OSN adminstrators

Figure 6. Methods to Detect Malicious Activities over OSNs

The other detection approach is using the Machine Learning techniques, which can be classified into supervised methods and unsupervised methods. The difference between the two categories is that the supervised methods use prepared set of data for training and predicting the model, while in the unsupervised methods there will be no data used for training. Examples of the first category include: regression models, support vector machine (SVM), and decision tree models. Examples of the unsupervised learning include: clustering algorithms and hidden Markov models. There is another machine learning category that uses combinations of the above two categorizes and so it is called semi-supervised method.

The third malicious activities detection technique is through using manual verification. The content should be analysed to make sure if it is fake or legitimate, therefore is the OSNs users' task to double check the contents before spreading it to other users. As part of the manual verification process, some social online platforms allow their users to report content that violates their privacy rules. Such violating content might include malware links, spam, and aggressive or abusive content. Moreover, many social networks ask the feedback from the users to enhance their privacy policies.

6.2. Results and discussions

For each entity in the social graph the trustworthiness score will be calculated. Now, if that score is below a certain threshold, then we can predict with high confidence that the activities associated with this entity will be malicious and can't be trusted. One main social graph property is the Diameter, which is defined to be the greatest distance between any two nodes on the graph. In order to find the diameter, first we need to find the shortest path between any pair of nodes in the social graph for all nodes. Then we take the maximum length of all these shortest paths. As an example, let the length between any two nodes (x,y) on the graph to be *len*. Then the diameter, *diam*, is:

diam= max *len*(x, y) for all possible nodes on the social graph

The usage of cliques is popular in graph-based analysis areas including the social networks in order to understand connections and trust relations between graph nodes. As an algorithm, clique calculates the maximum number of nodes in the graph in which every node in the clique is a friend to all other nodes in the clique. Figure 7 shows the total and weighted clique strength.

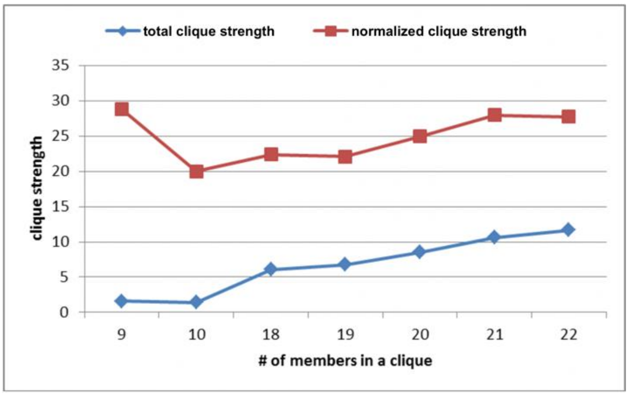

Figure 7. Comparison between total and normalized/weighted clique strengths

It is clear that extracting clique relations in OSNs captures groups properties and how they interact with each other. Larger cliques may contain smaller cliques and the researcher's focus should be on finding the maximum clique size. Our main contribution was normalizing clique strengths to give better assessment of user's trustworthiness and users activities and interaction within a clique.

7. CONCLUSION

In this paper, we addressed the important issue of security and privacy in Online Social Networks. What makes securing information in OSN more complicated is the heterogeneous web applications and protocols used, and variety of mobile apps platforms such as Android or iOS used to access OSNs. We investigated the need to apply Computational Intelligence techniques in OSNs security because traditional security techniques are not efficient enough to provide complete-protection against recent cyber-attacks [63]. This research took Artificial Neural Networks and Machine Learning into consideration and provided a comprehensive related work study and analysis of the existing methods for spam and fake news detection. Moreover, we investigated the application of ANNs for intrusion detection systems and spam filtering for OSNs platforms.

Finally, we addressed how online social networks can be structured into social graphs with the users represented by nodes on the graph. Moreover, and based on the observation that certain community or group over an OSN can form an entity with their social activities, we proposed the trustworthiness principle to identify the communities/entities with malicious activates. Additionally, we proposed new approach that uses the concept of weighted cliques in the detection of sub-communities' malicious behaviors over OSNs. The proposed methodology is based on computing the overall weight of the clique based on individual edges, and it can be used to identify suspicious behavior of certain online groups and to prevent further planned actions such as cyber/terrorist attacks before they happen.

ACKNOWLEDGEMENTS

This work was supported by Zayed University Research Office, Research Cluster Award #17079.

REFERENCES

[1] T. Haidegger, *et al.*, “Controller design solutions for long distance telesurgical applications,” *International Journal of Artificial Intelligence*, vol. 6, pp. 48-71, 2011.

[2] S. Vosoughi, *et al.*, “The spread of true and false news online,” *Science*, vol. 359, pp. 1146-1151, 2018.

[3] Moh’d, Abidalrahman, Nauman Aslam, Hosein Marzi, and L. A. Tawalbeh. "Hardware implementations of secure hashing functions on FPGAs for WSNs." In Proceedings of the 3rd *international conference on the applications of digital information and web technologies (ICADIWT)*. 2010.

[4] Lo'ai, A. Tawalbeh, and Waseem Bakhader. "A mobile cloud system for different useful applications." In 2016 *IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, pp. 295-298. IEEE, 2016.

[5] Yardi, Sarita, Nick Feamster, and Amy Bruckman. "Photo-based authentication using social networks." In *Proceedings of the first workshop on Online social networks*, pp. 55-60. ACM, 2008.

[6] Lo'ai, A. Tawalbeh, and Gokay Saldamli. "Reconsidering Big Data Security and Privacy in Cloud and Mobile Cloud Systems." *Journal of King Saud University-Computer and Information Sciences* (2019).

[7] R. Suryanita and A. Adnan, “Intelligent Monitoring System on Prediction of Building Damage Index using Neural-Network,” *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 10, 2012.

[8] M. Aldwairi, *et al.*, “Pattern matching of signature-based IDS using Myers algorithm under MapReduce framework,” *EURASIP J. on Info. Security*, vol. 2017, pp. 1-11, 2017.

[9] M. Aldwairi and R. Alsalman, “MALURLs: Malicious URLs Classification System,” *Annual International Conference on Information Theory and Applications, Singapore*, 2011.

[10] S. T. B. D. K. E. Bilge L., “All your contacts are belong to us: automated identity theft attacks on social networks,” *Proceedings of the 18th international conference on Worldwide web*, 2009.

[11] J. C. R. F. B. Jonell, “The real face of koobface: The largest web 2.0 botnet explained,” *Trend Micro Research*, vol. 5, pp. 10, 2009.

[12] S. L. Nagle F., “Can friends be trusted? exploring privacy in online social networks,” *Social Network Analysis and Mining, 2009. ASONAM'09. International Conference on Advances*, 2009.

[13] Boshmaf Y., *et al*, “The socialbot network: when bots socialize for fame and money,” *Proceedings of the 27th Annual Computer Security Applications Conference*, 2011.

[14] M. Aldwairi, *et al.*, “Detection of Drive-by Download Attacks Using Machine Learning Approach,” *International Journal of Information Security and Privacy*, vol. 11, pp. 16-28, 2017.

[15] Kelly, Louise, Gayle Kerr, and Judy Drennan. "Avoidance of advertising in social networking sites: The teenage perspective." *Journal of interactive advertising* 10, no. 2 (2010): 16-27.

[16] eWEEK.com, “eWEEK.” Available: http://www.eweek.com/security/phishing-attacks-increasingly-focus-on-social-networks-studies-show.html.

[17] Jararweh, Yaser, Haythem A. Bany Salameh, Abdallah Alturani, Loai Tawalbeh, and Houbing Song. "Anomaly-based framework for detecting dynamic spectrum access attacks in cognitive radio networks." *Telecommunication Systems*, Vol 67, no. 2 (2018): pp. 217-229.

[18] Qian, Jianwei, Xiang-Yang Li, Chunhong Zhang, and Linlin Chen. "De-anonymizing social networks and inferring private attributes using knowledge graphs." In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*, pp. 1-9. IEEE, 2016.

[19] Zheng, Haizhong, Minhui Xue, Hao Lu, Shuang Hao, Haojin Zhu, Xiaohui Liang, and Keith Ross. "Smoke screener or straight shooter: Detecting elite sybil attacks in user-review social networks." *arXiv preprint arXiv:*1709.06916 (2017).

[20] Tawalbeh, Lo'ai, Norah Alassaf, Waseem Bakheder, and Alaa Tawalbeh. "Resilience mobile cloud computing: features, applications and challenges." In *2015 Fifth International Conference on e-Learning (econf)*, pp. 280-284. *IEEE*, 2015.

[21] Bahwaireth, Khadijah, and Lo'ai Tawalbeh. "Cooperative models in cloud and mobile cloud computing." In 2016 *23rd International Conference on Telecommunications (ICT)*, pp. 1-4. *IEEE*, 2016.

[22] Yang, Kai- Cheng, Onur Varol, Clayton A. Davis, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. "Arming the public with artificial intelligence to counter social bots." *Human Behavior and Emerging Technologies* 1, no. 1 (2019): 48-61.

[23] Jin, Lei, Hassan Takabi, and James BD Joshi. "Towards active detection of identity clone attacks on online social networks." In *Proceedings of the first ACM conference on Data and application security and privacy*, pp. 27-38. ACM, 2011.

[24] Tawalbeh, Loai, Raad S. Al-Qassas, Nour S. Darwazeh, Yaser Jararweh, and Fahd AlDosari. "Secure and efficient cloud computing framework." In *2015 International Conference on Cloud and Autonomic Computing*, pp. 291-295. *IEEE*, 2015.

[25] Bahwaireth, Khadijah, Elhadj Benkhelifa, Yaser Jararweh, and Mohammad A. Tawalbeh. "Experimental comparison of simulation tools for efficient cloud and mobile cloud computing applications." *EURASIP Journal on Information Security*, 2016, no. 1 (2016): 15.

[26] Kim, Hyoungshick, John Tang, and Ross Anderson. "Social authentication: harder than it looks." In *International Conference on Financial Cryptography and Data Security*, pp. 1-15. Springer, Berlin, Heidelberg, 2012.

[27] A. J. Masoumzadeh, “Privacy Settings in Social Networking Systems: What You Cannot Control,” *The 8th ACM SIGSAC symposium on Information, computer and communications security*, 2013.

[28] L. H. Ru, *et al*., “Online data stream learning and classification with limited labels,” *Proceeding of the Electrical Engineering Computer Science and Informatics*, vol. 1, pp. 161-164, 2014.

[29] Tawalbeh, Lo'ai, Yaser Jararweh, and Abidalrahman Mohammad. "An integrated radix-4 modular divider/multiplier hardware architecture for cryptographic applications." *International Arab Journal of Information Technology (IAJIT)*, 9, no. 3 (2012).

[30] Mohammad, Abidalrahman, and Adnan Abdul-Aziz Gutub. "Efficient FPGA implementation of a programmable architecture for GF (p) elliptic curve crypto computations." *Journal of Signal Processing Systems* 59, no. 3 (2010): 233-244.

[31] Tawalbeh, Lo'ai, Alexandre Tenca, Song Park, and Cetin Koc. "An efficient hardware architecture of a scalable elliptic curve crypto-processor over GF (2n)." In *Advanced Signal Processing Algorithms, Architectures, and Implementations XV*, vol. 5910, p. 59100Q. International Society for Optics and Photonics, 2005.

[32] Stringhini, Gianluca, Christopher Kruegel, and Giovanni Vigna. "Detecting spammers on social networks." In *Proceedings of the 26th annual computer security applications conference*, pp. 1-9. ACM, 2010.

[33] Gao, Hongyu, Yan Chen, Kathy Lee, Diana Palsetia, and Alok N. Choudhary. "Towards Online Spam Filtering in Social Networks." In *NDSS*, vol. 12, no. 2012, pp. 1-16. 2012.

[34] A. Figueira and L. Oliveira, “The current state of fake news: challenges and opportunities,” *Procedia Computer Science*, vol. 121, pp. 817-825, 2017.

[35] D. Vedova, *et al*., “Automatic online fake news detection combining content and social signals,” *2018 22nd Conference of Open Innovations Association (FRUCT)*, pp. 272-279, 2018.

[36] Ferrara, Emilio, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. "The rise of social bots." *Communications of the ACM*, 59, no. 7, 96-104, 2016.

[37] M. Aldwairi and Y. Flaifel, “Baeza-Yates and Navarro Approximate String Matching for Spam Filtering,” *The Second International Conference on Innovative Computing Technology (INTECH 2012), Casablanca, Morocco*, 2012.

[38] Nalarubiga, E., and M. Sindhuja. "Efficient Classifier for Detecting Spam in Social Networks through Sentiment Analysis." *Asian Journal of Research in Social Sciences and Humanities* 6, no. 7 (2016): 1066-1073.

[39] T. T. Suganya, “Spam Filtering in Online Social Networks Using Machine Learning Technique,” *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 2, pp. 2564-2570, 2014.

[40] C. D. M. F. T. H. Xiao, “Detecting Clusters of Fake Accounts in Online SocialNetworks,” *Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security*, 2015.

[41] A. M. N. N. Fahim, “Facebook Spam and Spam Filter Using Artificial Neural Networks,” *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 1, pp. 220-223, 2015.

[42] K. R. V. Sattikar A. A., “A Role of Artificial Intelligence Techniques in Security and Privacy Issues of Social Networking,” *IJCSET*, Vol 2, Issue 1,792-795, January 2012.

[43] S. S. Saravanan K., “Review on classification based on artificial neural networks,” *International Journal of Ambient Systems and Applications*, vol. 2, pp. 8-11, 2014.

[44] O. Bachir and A. Zoubir, “Adaptive Neuro-fuzzy inference system based control of Puma 600 robot manipulator,” *International Journal of Electrical and Computer Engineering*, vol. 2, pp. 90, 2012.

[45] EXTREME TECH. Available: http://www.extremetech.com/extreme/215170-artificial-neural-networks-are-changing-the-world-what-are-they.

[46] G. Carlos, “Artificial Neural Networks for Beginners,” arXiv preprint cs/0308031, 2003.

[47] NEURAL NETWORKS. Available: https://cs.stanford.edu/people/eroberts/courses/soco/projects/2000-01/neural-networks/Applications/index.html.

[48] ALYUDA. Available: http://www.alyuda.com/products/forecaster/neural-network-applications.htm.

[49] G. Gu, “Machine Learning Meets Social Networking Security – Detecting and Analyzing Malicious Social Networks for Fun and Profit,” *Proceedings of the 5th ACM workshop on Security and artificial intelligence*, 2012.

[50] D. Wang, *et al*., “A Social-Spam Detection Framework,” *Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference*, 2011.

[51] A. Arram, *et al*., “Spam detection using hybrid Artificial Neural Network and Genetic Algorithm,” *13th International Conference on Intellient Systems Design and Applications*, 2013.

[52] C. Nandeshwar, *et al*., “Spam Mail Detection Using Artificial Neural Network,” *Imperial Journal of Interdisciplinary Research*, vol. 2, 2016.

[53] A. Goweder, *et al*., “An anti-spam system using artificial neural networks and genetic algorithms,” *Proceedings of the 2008 International Arab Conference on Information Technology*, pp. 1-8, 2008.

[54] A. Mohammed and F. Monir, “Email spam classification using hybrid approach of RBF neural network and particle swarm optimization,” *International Journal of Network Security & Its Applications (IJNSA)*, vol. 8, pp. 17-28, 2016.

[55] G. Besiashvili, *et al*., “Application of Adaptive Neural Networks for the Filtration of Spam,” *GCAI 2015. Global Conference on Artificial Intelligence*, 2015.

[56] A. J. Omar and A. Ahmad, “Network Intrusion Detection System Using Neural Network Classification of Attack Behavio,” *Journal of Advances in Information Technology*, vol. 6, 2015.

[57] C. Qiu and S. Jie, “Research on Intrusion Detection Algorithm Based on BP Neural Network,” *International Journal of Security and Its Applications*, vol. 9, pp. 247-258, 2015.

[58] V. D and P. SN, “Anomaly based IDS using Backpropagation Neural Network,” *International Journal of Computer Applications*, vol. 136, pp. 29-34, 2016.

[59] A. Dastanpour, *et al*., “Using Genetic Algorithm to Supporting Artificial Neural Network for Intrusion Detection System,” *The international conference on computer security and digital investigation (ComSec2014)*, 2014.

[60] J. Piersa and T. Schreiber, “Spectra of the spike-_ow graphs in geometrically embedded neural networks,” in *Proc. Artif. Intell. Soft Comput*., pp. 143-151, 2012.

[61] M. Al-Ayyoub, Y. Jararweh, A. Rabab'ah, M. Aldwairi. “Feature extraction and selection for Arabic tweets authorship authentication,” J Ambient Intell Human Comput (2017) 8: 383. https://doi.org/10.1007/s12652-017-0452-1.

[62] M. Aldwairi, *et al*. “Switch Architecture for Optical Burst Switching Networks”, In Proc. of the first workshop on Optical Burst Switching, OPTICOMM. Dallas, Oct 03.

[63] Q. Yaseen, Y. Jararweh, M. Al-Ayyoub and M. Aldwairi, "Collusion attacks in Internet of Things: Detection and mitigation using a fog based model," 2017 IEEE Sensors Applications Symposium (SAS), Glassboro, NJ, 2017, pp. 1-5.

BIOGRAPHIES OF AUTHORS

Monther Aldwairi, he is an associate professor at the College of Technological Innovation at Zayed University since the fall of 2014. He received his B.S. in electrical engineering from Jordan University of Science and University (JUST) in 1998, and his M.S. and PhD in computer engineering from North Carolina State University (NCSU), Raleigh, NC, in 2001 and 2006, respectively. Prior to joining ZU, he was an Assistant and then Associate Professor of Computer Engineering at Jordan University of Science and Technology. Dr. Aldwairi’s research interests are in network and web security, intrusion detection and forensics, cloud computing, reconfigurable architectures, artificial intelligence and pattern matching.

Dr. Tawalbeh (IEEE Senior Member): Got his MSc and PhD degrees in computer engineering from Oregon State University (OSU) , Oregon, USA in 2002 and 2004 respectively with GPA 4.0/4.0. Dr Tawalbeh currently is an associate professor at the Department of Computing and Cyber security, Texas A&M University-San Antonio, TX, USA. He is the founder and Director of the Cryptographic Hardware and Information Security (CHiS) lab at (JUST). From 2005 till 2012, he worked as a part time professor to teach different information security courses in the Master programs at: New York Institute of Technology (NYIT), DePaul’s University and princes Sumaya University for Technology (PSUT). Dr. Tawalbeh won many research grants and awards. He has over 90 research publications in many refereed international Journals and conferences. His research interests include information security, cryptographic applications and crypto systems hardware implementations, Cloud security and Mobile cloud computing. Dr. Tawalbeh is chairing many international conferences and workshops in Mobile Cloud Security, Blockchain and related Cyber security areas. He is a reviewer and a member of the editorial boards of many international journals. For more information see Dr. Tawalbeh’s website at google scholar:
Website: https://scholar.google.ae/citations?user=gw17mOoAAAAJ&hl=en&oi=ao
Email: Ltawalbeh@tamusa.edu