Chapter 11
Spammer Detection Approaches in Online Social Network (OSNs): A Survey

Somya Ranjan Sahoo, Brij B. Gupta, Dragan Peraković, Francisco José García Peñalvo, and Ivan Cvitić

1 Introduction

In day-to-day life, taking the benefit of Web 2.0 people uses e-commerce and opinion-sharing web applications for information sharing and communication. These websites allow the users to share their emotions, attitudes, personal experiences, feeling regarding products and services, and issues related to politics and economics. In recent years, the review of some specific products or websites increases dramatically. The reverse purchase decision depends on posted opinions by various social network users. Spam refers to unsolicited messages that spread over the network through emails, and direct messages sent by instant messenger, social networks, and various web-based searches depicted in Fig. 11.1. By taking the advantage of these services, spammer spreads malicious contents over the network in the form of malware and phishing [1–3]. Initially, spam spreads and is targeted to limited communications like email and instant messaging. But, it effectively invaded all media across WWW. Spam email called junk mail spreads through unwanted messages or bulk messages with commercial content. Similarly, instant message services like Yahoo Messenger and Skype were used by the adversary to

S. R. Sahoo
Vellore Institute of Technology, Amaravati, Andhra Pradesh, India

B. B. Gupta (✉)
National Institute of Technology, Kurukshetra, India

D. Peraković · I. Cvitić
University of Zagreb, Zagreb, Croatia
e-mail: dperakovic@fpz.unizg.hr

F. J. G. Peñalvo
University of Salamanca, Salamanca, Spain
e-mail: garcia@usal.es

L. Knapcikova et al. (eds.), *Sustainable Management of Manufacturing Systems in Industry 4.0*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-030-90462-3_11

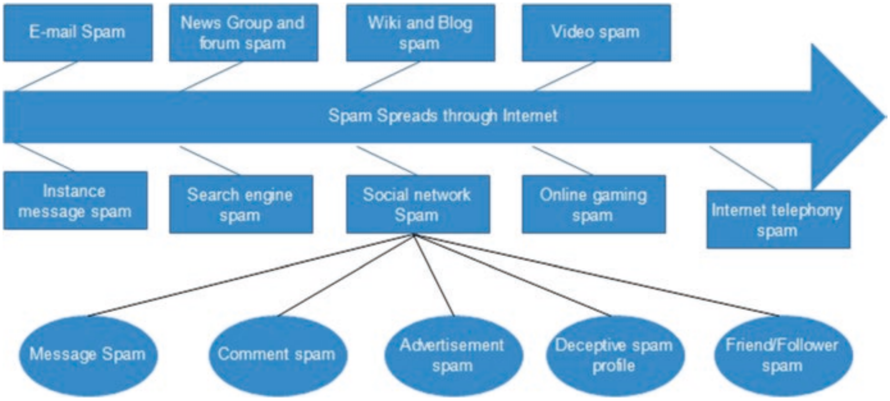


Fig. 11.1 Types of spam spread over the Internet

spread malicious information directly to annoy users. Using short messaging services, mobile-based spammers spread malicious information and infect mobile devices. For the promotion of a particular web page or web application, the spammer manipulates the ranking of search engines and some relevant algorithm. In another way, using short URL-based services, spammer spreads malicious information inside certain blogs and comments over internet services. Social media website like YouTube is the appropriate way for spammers to spread malicious videos with some pornographic and dating websites. The user comments related to those videos are spread over the wired and wireless networks and attract many users to visit [4–6]. Sometimes, these comments are auto-generated through a bot and invite people to surf. Even if the prominent way of communication among different users like social network using Cloud and other ways are also affected by spammers to gain users' credentials [7–19].

Recently, spam inevitably in almost all forms of communication and damage user's content including the performance of the network. It represents one of the biggest security and system performance problems together with DDoS (Distributed Denial of Service) attack [20–22]. Various solutions have been measured to detect spammer content and to improve the performance of the network. These solutions are well known as anti-spamming techniques or spam combating techniques. While a lot of work has been done in the area of malicious content detection based on spammer analysis especially web-based, email spam, spam in social networks and social media, is not even analyzed. This is because of the uncontrollable structure of the social network and flooded content of information. Due to the conductive breeding of social network and the large set of user's activities, it leads to hues damage to mankind. According to various surveys by different companies and brand protectors, spam content increases rapidly day by day. According to other surveys, the growths of spam are rampant. With time, the number of users of social media increases according to the user requirement and communication. As spammer content hampers the performance of the user content and communication medium is associated with the financial loss that is causing erosion in the user behavior. All these factors motivate us to work on social media spam and its detection mechanism.

Due to the unlike behavior of social media platforms based on characteristics, detection of the spammer is challenging and multifaceted. A number of approaches have been developed by researchers and academicians to fight against social spammers including the protection mechanism inbuilt with social network websites. A brief overview of different social spam detection techniques is depicted in Fig. 11.5. However, due to the fastest growing social network platforms, the behavior of users changed rapidly in the last few years. Anti-spam schemes need a major upheaval to extenuate them. In this chapter, we survey various mitigation and detection frameworks that have been proposed in the last few years to fight against spam in OSNs. The rest of the chapter is organized as follows. In Sect. 2, we describe the overview of different spam related to OSNs. In Sect. 3, we elaborate on the types of spammer and various detection mechanisms. In the next section, i.e., in Sect. 4, we describe a literature survey on various spam detection approaches and features of the Microblogging platform by which spammers propagate. In Sect. 5, we elaborate on various comparative analyses with existing approaches. Finally, in Sects. 6, we discuss some open issues and challenges related to spammer detection which concludes the chapter.

2 Online Social Network Spamming

Social network spammers spread in various ways such as posting malicious URLs, short URLs, fake advertisements for publicity, malware spreading, botnet attack through users and systems [23], following unknown users randomly, and some other ways to flooded network [24]. Another method of spreading spammer is the generation of fake reviews of various products and services using machine learning approaches [25]. The growth of global spam increases rapidly over the year and affects every social platform. Specifically, on Twitter one spam is found in every 20 tweets and posts. Most of the spammer content can spread automatically using the system through a bot [26]. Due to the lack of physical contact between the individuals, growth rate of spam increases. Due to these activities, identification of the user is under the black box. Evidently, utilizing the social network data without filtering the malicious activity for analysis is a wrong pattern for social network users. Numerous approaches have been developed by researchers and corporate analysts depicted in Sect. 3. However, spammers develop quickly to evade detection systems.

2.1 Types of Spammer and Spreading Techniques

Spammer spreads over the social network based on the features, properties, and characteristics of various accounts associated with service providers. Various categorizations of spammers spread over social networks are depicted in Fig. 11.2.

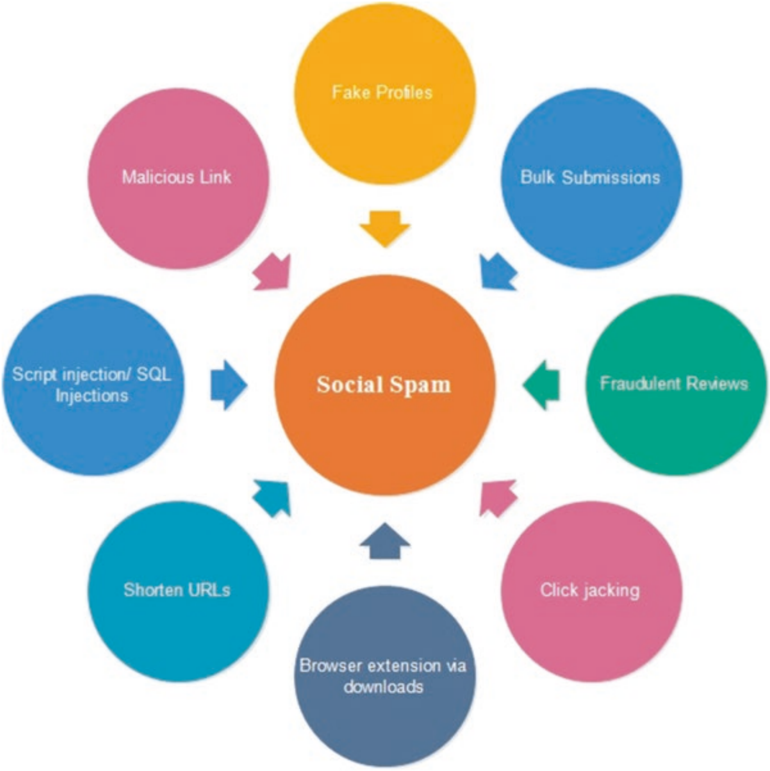
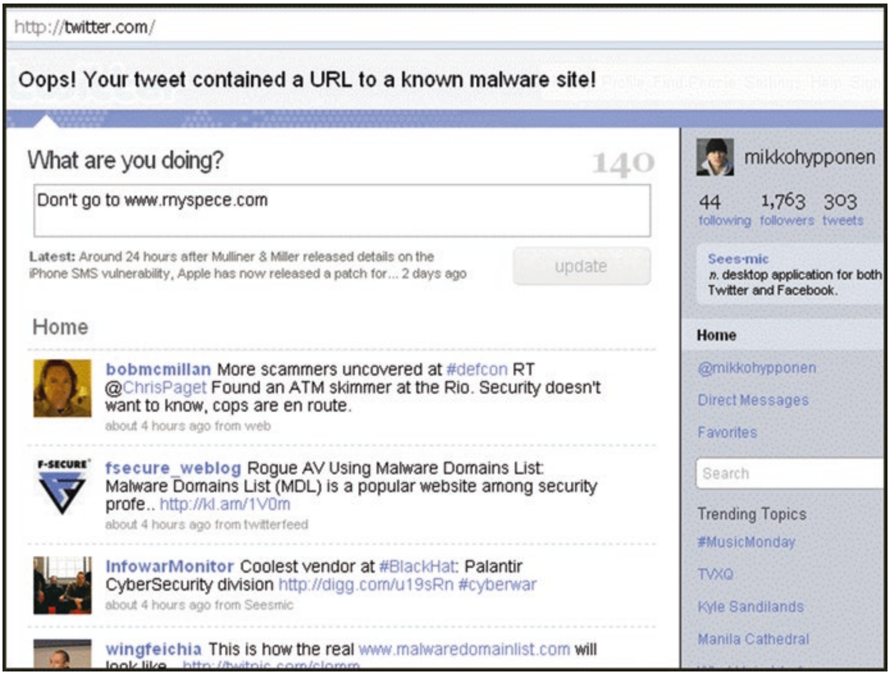


Fig. 11.2 Various categories of social spammer

- Malicious URLs: Malicious URLs damage the potential of the user's account including computer hardware. Some of the malicious URLs spread through a social network (Twitter) are checked by the service provider itself depicted in Fig. 11.3. Malicious URLs spread through various blogs, tweets, posts, direct message services, and many more ways on the social network platform.
- Fake profiles: Adversaries create fake profiles in online social networks to gather confidential information of the user and for some financial benefits. The fake profiles are created on the same platform or on a different platform by collecting user information. The basic objectives of creating fake accounts are to humiliate people over the network and collecting user's credentials from unknown users. Sometimes fake accounts are created to do some fun or some nuisance work.
- Bulk posting or submission: Bulk submission is also called the bombing of bulk spam messages. Through these activities, people attract other users toward their accounts. Sometimes people behave like trustworthy customers and spreads malicious bulk posts over the network. Bulk message contents are similar in nature, i.e., same messages posted many times in equal intervals of time. People



Oops! Your tweet contained a URL to a known malware site!

What are you doing?

Don't go to www.rnyspece.com

Latest: Around 24 hours after Mulliner & Miller released details on the iPhone SMS vulnerability, Apple has now released a patch for... 2 days ago

bobmcmillan More scammers uncovered at #defcon RT @ChrisPaget Found an ATM skimmer at the Rio. Security doesn't want to know, cops are en route.
about 4 hours ago from web

fsecure_weblog Rogue AV Using Malware Domains List: Malware Domains List (MDL) is a popular website among security profe.. http://kl.am/1V0m
about 4 hours ago from twitterfeed

InfowarMonitor Coolest vendor at #BlackHat: Palantir CyberSecurity division http://digg.com/u19sRn #cyberwar
about 4 hours ago from Seesmic

wingfeichia This is how the real www.malwaredomainlist.com will look like... http://twitpic.com/clomm

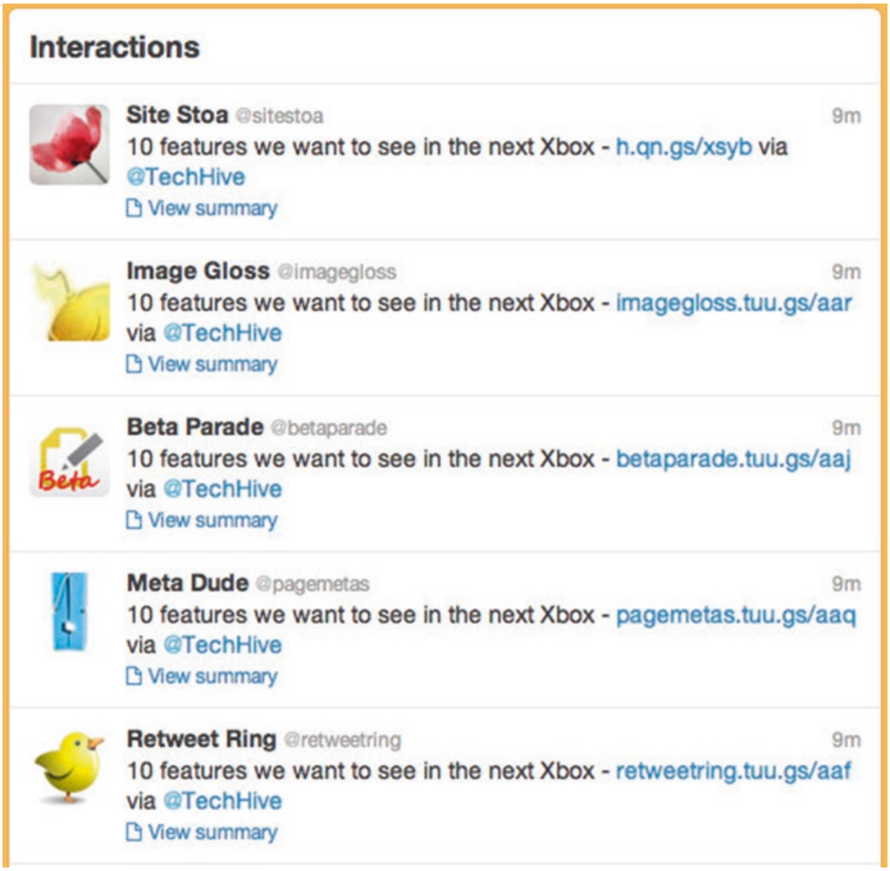
mikkohypponen
44 following 1,763 followers 303 tweets

Sees·mic
n. desktop application for both Twitter and Facebook.

Fig. 11.3 Malicious URLs in Twitter

use certain message spreading tools for spreading bulk posting automatically without user intervention as shown in Fig. 11.4.

- Fraudulent comments: Without knowing the specification of any product and its advantages, people review and claim that the product is good or bad. By this process, the product is highlighted as a good or bad product on social media. With the help of these services, people highlighted their comments and products over the network. Various other forms of social spam include special characters inside comments, harassing news, various threats, and profane words in comments and reviews.
- Spammer through social bot: Bot-based spammer spreading is a new approach by the adversaries in the social network. Bots are created by spammers and spread over the network using message services like Facebook messaging. Some bots are user created and some are system generated. The system-generated bots are spread through certain software.
- Malware-based spammer: Malware is a delivery vehicle for a spammer in a social network platform. Malicious software spread spammers using various tools and services. Some malware is spread through URLs, fraudulent links, and some new approaches in the network.



Interactions

Site Stoa @sitestoa 9m
10 features we want to see in the next Xbox - h.qn.gs/xsyb via @TechHive
View summary

Image Gloss @imagegloss 9m
10 features we want to see in the next Xbox - imagegloss.tuu.gs/aar via @TechHive
View summary

Beta Parade @betaparade 9m
10 features we want to see in the next Xbox - betaparade.tuu.gs/aaj via @TechHive
View summary

Meta Dude @pagemetas 9m
10 features we want to see in the next Xbox - pagemetas.tuu.gs/aaq via @TechHive
View summary

Retweet Ring @retweetring 9m
10 features we want to see in the next Xbox - retweetring.tuu.gs/aaf via @TechHive
View summary

Fig. 11.4 Bulk posting of tweets on Twitter

- Clickjacking: In these spamming techniques, the attacker tries to redirect the users from one page to another by clicking on a link or a blog. When the user visits any blog or clicks on that blog to see the details, the page redirects to some malicious site and malware is downloaded automatically or blocks certain services.
- Update or download malicious browser extension: Malicious software downloaded in the computer via a browser extension. The malicious browser extensions are automatically downloaded without the user’s notice and activate some malfunctions in the system. These sorts of services spread through some blogs, reviews with links, advertisements, etc.
- SQL injection: In this type of spammer spreading technique, the user changes the source code of the original content and added some malicious content to behave differently. These techniques spread rapidly over the network in various web

Table 11.1 Types of spam and spreading techniques

	Types of spam								
Spreading technique	Email based	Instant message based	News group based	Search engine based	Blog based	Video based	Social network based	Online game based	Through internet telephony
Malicious link	✓	✓	✓			✓	✓	✓	
Fake profile	✓	✓	✓		✓	✓	✓		
Bulk messaging	✓	✓	✓		✓	✓	✓		
Fraudulent comment/ review	✓	✓	✓	✓	✓	✓	✓		✓
Clickjacking	✓	✓	✓		✓	✓	✓	✓	
Browser extension	✓	✓	✓	✓	✓	✓	✓		
Shorten URLs	✓	✓	✓	✓	✓	✓	✓	✓	
SQL injection (XSS)	✓	✓	✓	✓	✓	✓	✓	✓	
Social bot	✓	✓			✓	✓	✓	✓	✓

applications, social networking sites, Microblogging sites, and review-related blogs. The overview of various spammers and their spreading methods is depicted in Table 11.1.

2.2 *Detection Methods*

The detection of social spam content is difficult to identify due to its hidden nature. Spammer spreads through the social platform over other services like the post, messages, tweets, videos, advertisements, and through direct communications. In [27], authors describe the concept of social spam and state that social spammers are different from other spammers due to their spreading method. They also describe some characteristics of various websites and develop a spam detector model to analyze spam and delete those spam from various websites. The author also identifies and analyzes various combating strategies to detect and identify spam. The authors categorize the spam detection framework into various groups like (1) Identification of spam and removal of the spam content at the same time. (2) Detection of spam and decrease its ranking so that it will not affect the content in the future. (3) Prevention method that protects the user accounts from various threats by blocking spammers. Several approaches based on the above categories are discussed by the authors. Authors in [28] described the concept of spammer detection including their own approach by analyzing temporal evaluation patterns. In this evaluation, pattern authors proposed a dynamic measure to analyze the user's activity and behavior to quantify the user's behavior. Various methods of spam detection by various researchers are described in Sect. 3 with different types and their uses.

3 Literature Review

Classification of the spammer in the social network is a big challenge in the network era due to its epidemic nature. Therefore, to classify the various detection approaches for spammers, we used the analysis of existing categorization including some new ideas depicted in Fig. 11.5. The overall classification of spam detection framework is classified into three different groups called syntax-based, profile feature-based, and blacklisted profile based on uses.

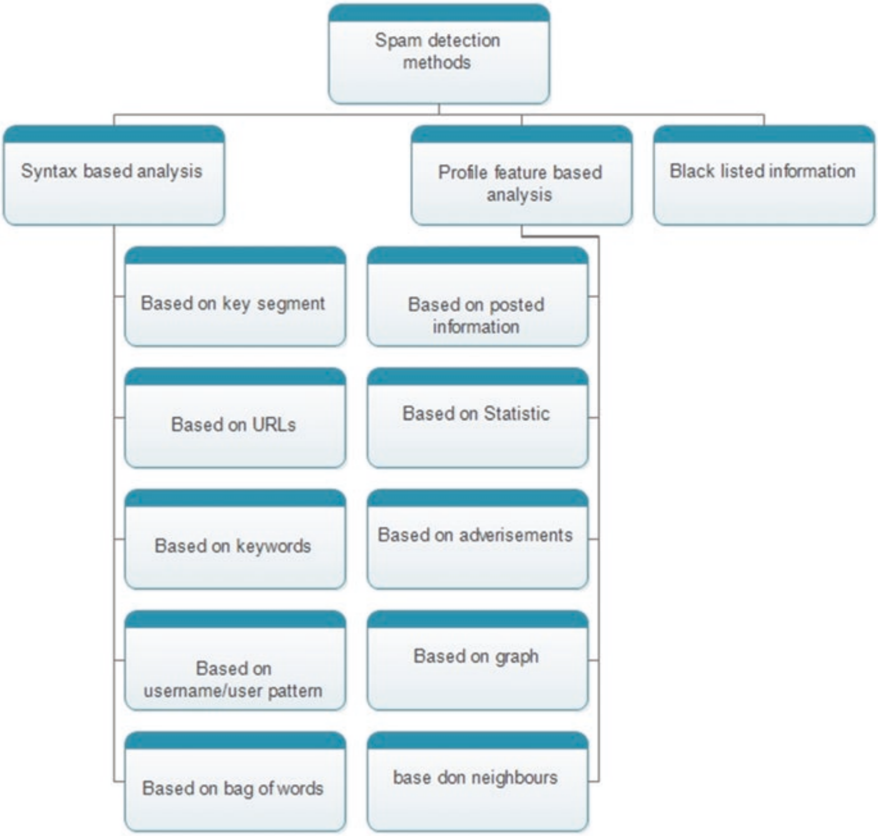


Fig. 11.5 Various spam detection methods

In this section, we analyze various spam detection methods based on various syntax analyses. Various posts are collected from different social platforms based on the shared content and data collection module. By analyzing the post based on the various features, suspicious posts are collected. Using a supervised learning approach, training set and test set modules are created for operations. By analyzing the training set and test set classification, problems are defined and solutions to these problems are generated as spam or non-spam.

3.1 Analysis Based on Key Segments

First, in this literature, we discuss the analysis of spammer detection based on key segments. Various researchers use key segment methods to detect spammers in social network platforms as follows. Detection of spammers by analyzing shorten URLs is the prominent method used by researchers. Some of the Microblogging sites like Twitter limit the number of characters up to 40 for every tweet. To reduce the content into a limited view, various software and algorithms are used by preprocessing the contents [29].

URL-Based Analysis

Some of the Microblogging sites like Twitter limit the number of characters up to 40 for every tweet. To reduce the content into a limited view, various software and algorithms are used for preprocessing the contents [29]. Shortened URLs hide the malicious link that spread spam messages and the original meaning of the content. Therefore, shortening the URLs is a major key segment for analysis. By detecting spammers based on the key segment, the author in [30] develop an identifier. For example, shortened URLs like bit.ly are detected. They calculate the percentage of spammer and non-spammer content used for shortening the URLs. To discriminate between the spammer and non-spammer, they use some division method. If the division value is greater than one, it means the activity is related to the spammer category. Many other methods are also used to duplicate the URLs and classify various posts. Using this method, tweets and accounts related to tweets are clustered into different categories based on the shared URLs. The authors in [27], describe the clustering of the related tweets based on textual content and shortened URLs for better identification of spammers. Based on the URLs shared by the users, the authors in [28] linked all the accounts and formed a cluster for better analysis and identification. After exposure of various Twitter campaigns using URL-based methods, some new algorithms are employed based on a machine learning approach to distinguish spammer contents from regular tweets. All these traditional algorithms are based on statistical feature analysis extracted from different user profiles [28, 31]. The author also uses Shannon's information theory and computes the entropy using the URLs attached in the tweets. Based on the above analysis, the similarity indexed was also calculated and the interval between the tweets computed. Based on the various tweets, a common graph was constructed using the selected similarity index (threshold). Various features are extracted from the URLs like lexical attributes, page content, and domain hosting properties for analyzing spammer activity [32, 33]. The authors in [34] use various features related to IP addresses for analyzing spammer contents. Most of the detection techniques are mixed together for better performance to detect spammer content online. The analysis by the authors in [34, 35] also describes that every analysis of spam based on the URLs included analysis of shortened URLs. Based on the analysis by the author in [30], the likelihood ratio of the shorteners is studied. According to them, they found 77% of spam content accounts in Twitter that are suspended within a single day of their creation. The author in [27] proposed a spam campaign that controlled around 145,000 accounts involved in spreading spam messages in the form of URLs. They collected various features from Twitter accounts like URL redirects, reputation of the account, posting contents, and user account information for their analysis of spam campaigns with the help of Twitter streaming API [36]. We also elaborate on the various advantages and disadvantages of the URL-based analysis. Due to the tweet limit by the service provider, malicious users spread malicious URLs to gather personal information of the user and harass the people over the net. So those contents are included with URLs that should be analyzed for detecting spammer activity in Twitter by various authors. Moreover, the URL-based analysis faced inaccuracy due to shortening of URL features. Various algorithms and reverse engineering principles are applied for changing the shorten URLs to original URLs. In addition to the advantages, various disadvantages are reported by the authors to detect spammer contents based on URL analysis. The main disadvantages related to the URL-based analysis are the fast rate of processing and autorun principle used by the malicious users.

Analysis Based on Pattern or Keyword

To analyze spammer content in tweets, keyword and user name methods also be used by the researchers. But the implementation of this method is very straightforward and intuitive. However, based on our analysis only some researchers work under this category to detect spammer content in social networks in 2009 and 2010. The authors in [37] develop an algorithm to detect spam by detecting keywords and matching the user name. Based on the assumption by the author, the account or the user name combined with letter and numbers have more chance as a spammer account. The tweets spread by the spammer refer to the unsolicited message by manipulating some accounts automatically informal pattern. Also, the author analyzes that the tweets that contain misleading words are more likely from spam accounts. The author in [38] also applied the same principle in 2010 on the Facebook platform. Searching the pattern in the form of malicious content in Facebook and Twitter with the principle of patterns or keywords is a challenging task. The URLs like "click here" are the best example. In practice, both the techniques, i.e., keyword-based and user name-based are applied together for proper analysis of the content and that will also be helpful for shortening URLs. The collection of various information and features are the most important identities for detecting spam content over the social network. Using various social engineering methods, spamming activities in various profiles and social networks can easily avoid the usage of user patterns and keywords.

3.2 Based on Tweet Content

In this section, we discuss the detection of spam content by analyzing various tweet contents posted by users over the network. Because spammer spreads easily over the network using various contents like the bulk of words, fraudulent tweets, and other posted information. We discuss the various methods by which the textual content of tweets can spread. These methods are TF–IDF (Term Frequency–Inverse Document Frequency), Bag of words, and sparse learning and are discussed below.

TF–IDF-Based Analysis

This method of analysis is the most popular technique to eradicate the meaning of various tweets. Various authors who worked in this area are listed as follows. In [39–41], authors use the TF–IDF principle to analyze various tweets to detect malicious contents as spam. Basically, TF–IDF is used to extract text from various posts to identify the context interns of weight [42]. Based on the research in [42], the author in [41] designs a metric to measure the correlation between the tweets in each pair of accounts. The author in [39] also applied the same principle used by the author in [42] to identify the similarity index using the vector space model. By this analysis, the author identifies that the similarity index of legitimate was stronger as compared to spamming one. The TF–IDF-based search first identifies the duplicate tweets posted by the spammer account over the social network platform. The Twitter campaign by the sender is classified into spam and non-spam campaign based on the content reviewed and identified. Based on the vector space model, all the content detected through TF–IDF is processed and evaluated for the best output. As far as performance is concerned, the author in [39] compared eight different machine learning algorithms and found RF (Random Forest) outperformed. As far as features are concerned, the similarity index generated by the TF–IDF technique ranked ninth with an accuracy of 72.3% in random forest classifier. Moreover, the author in [41] combined both tweet content and social relationship and calculated malicious scores between an individual account and its following. The proposed method by the author in [41] called CIA performs better as compared to others and identified 13 more spammers.

Based on Bag of Words

Before training a classifier for classification, the bag of words based method works for the representation of text by preprocessing. Various works have been implemented by various authors based on a bag of words. The method proposed by the authors in [43] using bag of words in TF–IDF techniques as the weighted algorithm to represent vectors. Meanwhile, the basic principle of this algorithm is widely adopted through Bayesian algorithm to pick up words from various paragraphs or posts. Basically, this technique is used to measure statistical analysis like content similarity or classifies the text directly based on their nature or behavior. Feature extractions through bag of word methods are based on text analysis. By this process, various punctuations, lowercasing every character in the sentence or in a tweet, and tokenizing each word can be done through TF–IDF converter converts into texts [43]. For detection of spammer content on tweets and in other posts, these features work individually. Also, these features are associated with other features to work in the detection process of spam. The basic principle of bag of words is also used in Bayesian classifiers called CRM114 [44]. According to the authors in [45], bag of word method is basically used to detect email spam due to simple and easy implementation. In practice, a bag of words combines other features in social network account to detect spammer and their relationship. Simply, only the bag of words method cannot give a suitable solution as like other services. As an example, Bayesian classifier includes tweet descriptions without using any account information.

Based on Sparse Learning

Due to the high-dimensional feature vector generated by the traditional spammer detection method based on n-gram and bag of words, the authors in [46, 47] proposed a sparse representation method. This method represents key phrases or words instead of total sentences. The method was applied by the author in [48] to a non-negative matrix factorization model (NMF). This model is used for the representation of lower dimensional feature vectors. Then, an optimization technique is applied to transfer the text from the next level to the topic level. Due to the shrunk length, features are more representative and make some clusters for better identification of spam. Compared to other models, sparse learning method performs best using the five cross-validation techniques. As a result, this model gives better accuracy.

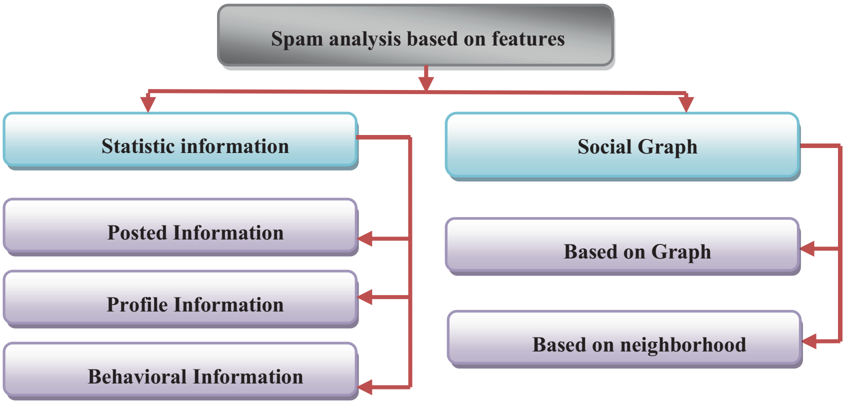


Fig. 11.6 Various categories of features

3.3 *Analysis Based on Features*

In this section, we analyze various spam detection methodologies and frameworks based on the feature analysis. We divide the total features into three different categories called features based on posted information, profile information, and user behavior as depicted in Fig. 11.6. Also, all three categories are coming under broad broad groups.

Based on the Posted Information

Various methods that are based on posts related to social network platforms must work with combined features from all categories for analyzing spammer content in social network platforms. Basically, spammers always spread through social engineering techniques. So the posted information in the social network as text should be analyzed for spammer detection. Information related to the post and its features is depicted in Fig. 11.7.

The author analyzes that the spammer spread more posts as compared to other users in social network platforms [49]. According to the authors in [50], based on cumulative distribution function point, it was reported that spammers usually spread spam content through hashtags, URLs, and spam words within text messages. Also, spammer uses more text size as compared to normal posts [51]. The various features used to detect spammer content in social network platforms are very useful for analysis. Various features are used for spammer detection in social network platforms depicted in the above figure. Based on the statistical feature analysis, it performs well as compared to other methods of feature section. According to the analysis report by the authors in [51], the F-measure score is as high as 93.6% by Random forest classifier. Also, the author analyzes the same using six different classification techniques. Wan et al., 2010 developed a Bayesian network-based method

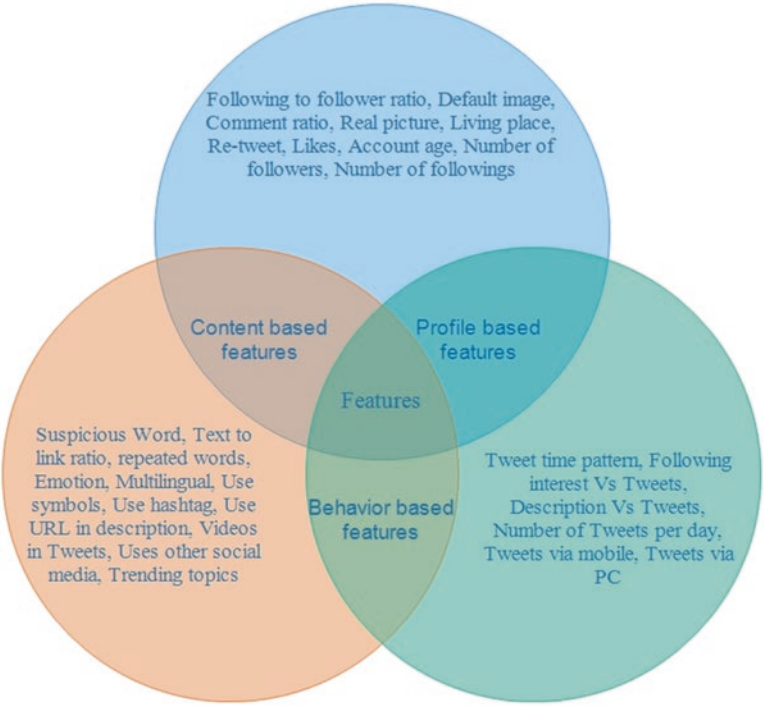

Following to follower ratio, Default image, Comment ratio, Real picture, Living place, Re-tweet, Likes, Account age, Number of followers, Number of followings
Content based features
Profile based features
Features
Suspicious Word, Text to link ratio, repeated words, Emotion, Multilingual, Use symbols, Use hashtag, Use URL in description, Videos in Tweets, Uses other social media, Trending topics
Behavior based features
Tweet time pattern, Following interest Vs Tweets, Description Vs Tweets, Number of Tweets per day, Tweets via mobile, Tweets via PC

Fig. 11.7 Various features associated with each category

that achieves nearly about 90% TP rate. The author in [52] analyzes the behavior of the account through URL analysis that is embedded in text messages in different social networks like Facebook and Twitter. The authors use decision tree classifiers in [53] using various features to analyze spam content in Facebook. The main objective of the author is to identify the messages that include URLs, hyperlinks, and hashtags. The author in [54] focuses on five various valuable content related to Twitter accounts like content filtering, scalability, proper decision-making, ability to retain the model with new content, and independent of text model developed for detecting spammers in social contents. The most important factor of this model is, it works on the real-time and content filtering option within a short time period. Lee et al. [55] proposed an approach based on various classifications by deploying social honeypots in the network and various user profile features for analysis.

Based on Profile Information

There are also some techniques used by the researchers to identify spammer contents in social network platforms using profile-based feature analysis. According to the authors in [56], the number of followers and followings of spammer account is much larger compared to other users. Also, the author analyzes that the life cycle of spammer users is less as compared to legitimate users in social network platforms. Some of the researchers also analyze spammer content by combining all features into account. According to the authors in [57], the reputation of the spammer is either very low or very high. But the reputation of the legitimate users lies between 30% and 95%. Similar activities found by the authors in [58] were implemented through the Facebook platform. They achieved 98.7% as a true positive rate with 1.4% as a false positive rate. Also, the authors in [50] implemented the same through support vector machine classifier to classify spammer and non-spammer content in Facebook with an 87.2% detection rate. By identifying suspected accounts on Twitter, the author analyzes the activity of the user for detecting spam content [59]. The authors in [60] investigating the deceptive information in Twitter spam by analyzing both public and private information of the user through various features analysis.

Based on the Behavior of the Account

Based on the behavior and campaign in social network platform, various authors analyze the spammer activity. Based on the similarity index generated through feature analysis, authors grouped various spam activities into a cluster of accounts [46–48]. The author in [61] analyzes the behavior of various accounts and the interval between the tweets spread by the user for detecting spammer account. By considering various features and generating content self-similarity scores, the author in [61] analyzes that the spammer embedded some text templates and post similar content over the network. As far as detection rate is concerned, Zhang et al. 2016 measure 88% as F1 score and more than 90% as precision value.

3.4 Analysis Based on Graph

Analysis of spammer content in various social network platforms based on the social graph is a major challenge by the researchers. The overall detection method is implemented using various features associated with the follower and following activities of the user. Based on the follower and following of user graph-based method categorized into two different groups called graph-based method and neighborhood methods. Each node in the social graph-based method represents an account with in-degree and out-degree nodes. The in-degree of a node denotes followers and the out-degree node is called followings [46, 47]. The authors in [47] analyze the behavior of the account through graph structures and the features associated with the account, i.e., graph density, reciprocity, and shortest path. Another researcher in [62] analyzes the BC (Betweenness centrality) values to identify the activity and association of the account. In recent years, neighborhood methods for detecting social network spam is the main concern. Based on this method, various features associated with the accounts are collected for spammer analysis. According to the author [62], neighbor accounts are affected by the spammer over social network platforms. The information related to the posted content can be identified by the follower or followings account.

4 Comparative Analysis

Comparative analysis of various methods used by the authors for detecting spammer content in social network platforms was depicted in Table 11.2.

5 Open Issues and Challenges

In this survey, we have discussed various methods and techniques for detecting spammer content in social network platforms. As we can see from various methods, the majority of the analysis mainly depends on the machine learning platform. Among all these techniques, the major differences are identified based on the method and feature selections. Our literature survey reviewed various methods and techniques for identifying spammer content in social network platforms. Also, there are several open issues and challenges for existing methods. We identified and present some of the open issues in this section. First, the collection of the real dataset is a challenging task. Real-time datasets are required for better analysis of spammers. Second, labeling the dataset manually is too difficult. So, the proper methodology should be applied for labeling raw data into the labeled dataset. Third, both public and private information are required for better analysis of spammer and account related to spammer category. Fourth, proper classification techniques are applied for better decision-making. Finally, fabrication of the data is used to train and test the appropriate model and is easy to manipulate from time to time.

6 Conclusion

In this chapter, we review the various state of arts related to spammer detection in social network platforms. We first categorize the type of spam spread through the social network by the spammer user. We further carried out the spammer detection techniques with thepros and cons of every method and also the comparative analysis of existing approaches. It was found that the spammers are spread through social

Table 11.2 Comparative analysis of various spam detection framework based on their features used

Authors	Title	Technique used	Feature analysis	Accuracy (%)	Pros and cons
Aslan et al. [63]	Automatic detection of cybersecurity-related accounts on online social networks	Machine learning-based classifiers with prototypical words	User-based and behavioral features for analysis	97.17	Better accuracy compared to other methods
Sohrabi et al. [64]	A feature selection approach to detect spam in the Facebook social network	PSO-based hybrid method for spam analysis	Optimization-based feature selection	91.20	Extraction of features from various profiles but less number of features for analysis
Singh et al., 2018 [65]	Who is who on twitter—Spammer, fake, or compromised account? A tool to reveal true identity in real time	Various machine learning approaches with feature selection methods	Feature related to pornographic contents	92.1	Less number of feature selection with manual selection process
Erwin et al. [66]	Detecting Indonesian spammer on Twitter	SVM and other machine learning approaches	User behavior and post content features	93.67	Manual selection of features but detection rate measures better performance in terms of accuracy
Bindu et al. [67]	Discovering spammer communities on Twitter	Graph-based approaches including machine learning-based analyzer	Community-based features including structural characteristics	86.7	Cluster-based approach need more features for analysis
Gupta et al. [4]	Collective classification of spam campaigners on Twitter: A hierarchical meta-path-based approach	Hierarchical meta-path detection mechanism for analysis	User-based and behavioral-based features for analysis	67.3	Detection rate is very low but hierarchical method gives better direction for detection of spam

(continued)

Table 11.2 (continued)

Authors	Title	Technique used	Feature analysis	Accuracy (%)	Pros and cons
Chu et al. [27]	Detecting social spam campaigns on Twitter. In: International conference on applied cryptography and network security	Clustering the related tweets based on textual content and shortened URLs for better identification of spammers	Features related to shortening the URLs and textual content features	87	URLs cannot be detected for spam content with better accuracy
Yardi et al. [36]	Detecting spam in a Twitter network	Analysis based on patterns and keywords	User behavior-based features for spam content analysis in Twitter	91	Pattern-based detection system is appropriate but the keyword-based method is not suitable for larger dataset
Benevenuto et al. [50]	Detecting spammers on Twitter. In: Collaboration, electronic messaging, anti-abuse, and spam conference	Cumulative distribution function point was reported that spammers usually spread spam content through hashtags and URLs	Features related to URLs and hashtags	–	Only URL-based features cannot be detected for spam content with better accuracy
Chen et al. [51]	Six million spam tweets: A large ground truth for timely Twitter spam detection	Random forest-based classification in machine learning environment	Features related to the Tweets and posted information, i.e., content-based features	93.6	Only content-based features are not sufficient for analysis
Ahmed et al. [58]	A generic statistical approach for spam detection in online social networks	Finding reputation of the spammer as either very low or very high. But, the reputation of the legitimate users lies between 30% and 95%	User profile features are used for analysis	98.7	Reputation of the user cannot be identified through profile features analysis

(continued)

Table 11.2 (continued)

Authors	Title	Technique used	Feature analysis	Accuracy (%)	Pros and cons
Zhang et al. [61]	Detecting spam and promoting campaigns in Twitter	Author analyzes that the spammer embedded some text templates and posted similar content over the network	User account and posted information-based features are analyzed	90	Less number of feature selection with manual selection process leads to lower accuracy and detection rate

network contents and that can be detected through various methodologies including futures related to user account and post. Finally, we made a brief summary and discussed some open issues related to social network spam detection. We hope this survey helps a lot to the researchers and the users who participated in the networks for sharing information like Facebook, Twitter, Instagram, etc.

References

1. Gupta BB (ed) (2018) Computer and cyber security: principles, algorithm, applications, and perspectives. CRC Press, Boca Raton
2. Fire M, Goldschmidt R, Elovici Y (2014) Online social networks: threats and solutions. IEEE Commun Surv Tutor 16(4):2019–2036
3. Sahoo SR, Gupta BB (2019) Classification of multiple attacks and their defense mechanism in online social networks: a survey. Enterp Inf Syst 13(6):832–864
4. Ho K, Liesaputra V, Yongchareon S, Mohaghegh M (2018) Evaluating social spammer detection systems. In: Proceedings of the Australasian computer science week multiconference, January. ACM, p 18
5. Gupta, S., Khattar, A., Gogia, A., Kumaraguru, P., & Chakraborty, T. (2018). Collective classification of spam campaigners on twitter: a hierarchical meta-path based approach. arXiv preprint arXiv:1802.04168
6. Stergiou CL, Psannis KE et al (2020) IoT-based big data secure management in the fog over a 6G wireless network. IEEE Internet Things J 8:5164–5171
7. Mishra A, Gupta N, Gupta BB (2021) Defense mechanisms against DDoS attack based on entropy in SDN-cloud using POX controller. Telecommun Syst 77:1–16
8. Alsmirat MA, Al-Alem F, Al-Ayyoub M, Jararweh Y et al (2019) Impact of digital fingerprint image quality on the fingerprint recognition accuracy. Multimed Tools Appl 78(3):3649–3688
9. Dahiya A, Gupta BB (2021) A reputation score policy and Bayesian game theory based incentivized mechanism for DDoS attacks mitigation and cyber defense. Future Gener Comput Syst 117:193–204
10. Bhushan K, Gupta BB (2019) Distributed denial of service (DDoS) attack mitigation in software defined network (SDN)-based cloud computing environment. J Ambient Intell Humaniz Comput 10(5):1985–1997
11. Olakanmi OO, Dada A (2019) An efficient privacy-preserving approach for secure verifiable outsourced computing on untrusted platforms. Int J Cloud Appl Comput 9(2):79–98

12. Hossain MS, Muhammad G, Abdul W, Song et al (2018) Cloud-assisted secure video transmission and sharing framework for smart cities. Futur Gener Comput Syst 83:596–606
13. Kaushik S, Gandhi C (2019) Ensure hierarchal identity based data security in cloud environment. Int J Cloud Appl Comput 9(4):21–36
14. Gou Z, Yamaguchi S (2017) Analysis of various security issues and challenges in cloud computing environment: a survey. In: Identity theft: breakthroughs in research and practice. IGI Global, Hershey, pp 221–247
15. Cvitić, I., Peraković, D., Periša, M., & Botica, M. (2021). Novel approach for detection of IoT generated DDoS traffic. Wireless Networks, 27(3), 1573–1586
16. Cvitic I, Perakovic D, Perisa M, Botica M (2020) Definition of the IoT device classes based on network traffic flow features. In: Knapcikova L, Balog M, Perakovic D, Perisa M (eds) EAI/ Springer innovations in communication and computing [internet]. Springer, Cham, pp 1–17
17. Perakovic D, Perisa M, Cvitic I, Husnjak S (2017) Artificial neuron network implementation in detection and classification of DDoS traffic. TELFOR J 9(1):26–31
18. Pasupuleti SK (2019) Privacy-preserving public auditing and data dynamics for secure cloud storage based on exact regenerated code. Int J Cloud Appl Comput 9(4):1–20
19. Al-Qerem A, Alauthman M, Almomani A et al (2020) IoT transaction processing through cooperative concurrency control on fog–cloud computing environment. Soft Comput 24(8):5695–5711
20. Cvitić I, Peraković D, Periša M, & Jurcut AD (2021) Methodology for Detecting Cyber Intrusions in e- Learning Systems during COVID-19 Pandemic. Mobile networks and applications, 1–12
21. Cvitić I, Peraković D, Periša M, Husnjak S (2019) An overview of distributed denial of service traffic detection approaches. Promet Traffic Traffico 31(4):453–464
22. Cvitić I, Peraković D, Periša M, Gupta BB (2021) Ensemble machine learning approach for classification of IoT devices in smart home. Int J Mach Learn Cybern Ensemble 12:1–24
23. Ahmed H (2017) Detecting opinion spam and fake news using n-gram study and semantic similarity. Ph.D. thesis
24. Sahoo SR, Gupta BB (2019) Hybrid approach for detection of malicious profiles in twitter. Comput Elect Eng 76:65–81
25. Yao Y, Viswanath B, Cryan J, Zheng H, Zhao BY (2017) Automated crowdturfing attacks and defenses in online review systems. In: Proceedings of the ACM SIGSAC conference on computer and communications security (CCS), Dallas, TX, USA, pp 1143–1158
26. Sahoo SR, Gupta BB (2020) Multiple features based approach for automatic fake news detection on social networks using deep learning. Appl Soft Comput 100:106983
27. Sahoo SR, Gupta BB (2020) Real-time detection of fake account in twitter using machine-learning approach. In: Advances in computational intelligence and communication technology. Springer, Singapore, pp 149–159
28. Sahoo SR, Gupta BB, Choi C, Hsu CH, Chui KT (2020) Behavioral analysis to detect social spammer in online social networks (OSNs). In: International conference on computational data and social networks. Springer, Cham, pp 321–332
29. Klien F, Strohmaier M (2012) Short links under attack: geographical analysis of spam in a url shortener network. In: Proceedings of the 23rd ACM conference on hypertext and social media. ACM, pp 83–89
30. Thomas K, Grier C, Ma J, Paxson V, Song D (2011) Design and evaluation of a real-time url spam filtering service. In: 2011 IEEE symposium on security and privacy. IEEE, pp 447–462
31. Zhang X, Zhu S, Liang W (2012) Detecting spam and promoting campaigns in the twitter social network. In: 2012 IEEE 12th international conference on data mining. IEEE, pp 1194–1199
32. Ma J, Saul LK, Savage S, Voelker GM (2009) Identifying suspicious urls: an application of large-scale online learning. In: Proceedings of the 26th annual international conference on machine learning. ACM, pp 681–688
33. Whittaker C, Ryner B, Nazif M (2010) Large-scale automatic classification of phishing pages. In: NDSS, vol 10

34. Sahoo SR, Gupta BB (2020) Classification of spammer and nonspammer content in online social network using genetic algorithm-based feature selection. Enterp Inf Syst 14(5):710–736
35. Sahoo SR, Gupta B, Choi C, Esposito C (2020) Detection of spammer account through rumor analysis in online social networks. In: The 9th international conference on smart media and applications. (pp. n-a)
36. Twitter Developers. Twitter's streaming API documentation; 2016. Available from: https://dev.twitter.com/streaming. Accessed 23 June 2019
37. Yardi S, Romero D, Schoenebeck G et al (2009) Detecting spam in a twitter network. First Monday 15(1). https://doi.org/10.5210/fm.v15i1.2793
38. Gao H, Hu J, Wilson C, Li Z, Chen Y, Zhao BY (2010) Detecting and characterizing social spam campaigns. In: Proceedings of the 10th ACM SIGCOMM conference on Internet measurement. ACM, pp 35–47
39. Sahoo SR, Gupta BB (2020) Fake profile detection in multimedia big data on online social networks. Int J Inf Comput Secur 12(2–3):303–331
40. Ivan Cvitić, G. Praneeth, D. Peraković (2021), Digital Forensics Techniques for Social Media Networking. Insights2Techinfo, pp.1
41. Yang C, Harkreader R, Zhang J, Shin S, Gu G (2012) Analyzing spammers' social networks for fun and profit: a case study of cyber criminal ecosystem on twitter. In: Proceedings of the 21st international conference on world wide web. ACM, pp 71–80
42. Salton G, Buckley C (1988) Term-weighting approaches in automatic text retrieval. Inf Process Manage 24(5):513–523
43. Lee K, Caverlee J, Webb S (2010) Uncovering social spammers: social honeypots+ machine learning. In: Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval. ACM, pp 435–442
44. Gupta S, Gupta BB, & Chaudhary P (2018) Hunting for DOM-Based XSS vulnerabilities in mobile cloudbased online social network. Future Generation Computer Systems, 79, 319–336
45. Chaudhary P, Gupta BB, & Gupta S (2019) A framework for preserving the privacy of online users against XSS worms on online social network. International Journal of Information Technology and Web Engineering (IJITWE), 14(1), 85–111
46. Khushboo Kumari (2021) Online social media threat and It's solution, Insights2Techinfo, pp.1
47. Hu X, Tang J, Liu H (2014) Online social spammer detection. In: AAAI. ACM, New York, pp 59–65
48. Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401(6755):788–791
49. Yang C, Harkreader RC, Gu G (2011) Die free or live hard? Empirical evaluation and new design for fighting evolving twitter spammers. In: International workshop on recent advances in intrusion detection. Springer, Cham, pp 318–337
50. Benevenuto F, Magno G, Rodrigues T, Almeida V (2010) Detecting spammers on twitter. In: Collaboration, electronic messaging, anti-abuse and spam conference (CEAS), vol 6, p 12
51. Chen C, Zhang J, Chen X, Xiang Y, Zhou W (2015) 6 million spam tweets: a large ground truth for timely twitter spam detection. In: 2015 IEEE international conference on communications (ICC). IEEE, pp 7065–7070
52. Cao C, Caverlee J (2015) Detecting spam URLs in social media via behavioral analysis. In: Proceedings of advances in information retrieval. Springer, pp 703–714
53. Soiraya M, Thanalerdmongkol S, Chantrapornchai C (2012) Using a data mining approach: spam detection on Facebook. Int J Comput Appl 58(13):26–31
54. Thomas K, Grier C, Ma J, Paxson V, Song D (2011) Design and evaluation of a real-time url spam filtering service. In: Proceeding of IEEE symposium on security and privacy (SP)
55. Lee K, Caverlee J, Webb S (2010) Uncovering social spammers: social honeypots + machine learning. In: Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval, SIGIR '10, pp 435–442
56. Chen C, Wen S, Zhang J, Xiang Y, Oliver J, Alelaiwi A et al (2017) Investigating the deceptive information in twitter spam. Futur Gener Comput Syst 72:319–326

57. Wu T, Liu S, Zhang J, Xiang Y (2017) Twitter spam detection based on deep learning. In: Proceedings of the Australasian computer science week multiconference. ACM, p 3
58. Ahmed F, Abulaish M (2013) A generic statistical approach for spam detection in online social networks. Comput Commun 36(10):1120–1129
59. Thomas K, Grier C, Song D, Paxson V (2011) Suspended accounts in retrospect: an analysis of twitter spam. In: Proceedings of the 2011 ACM SIGCOMM conference on internet measurement conference, IMC ‘11, pp 243–258
60. Chen C, Wen S, Zhang J, Xiang Y, Oliver J, Alelaiwi A, Hassan MM (2017) Investigating the deceptive information in twitter spam. Futur Gener Comput Syst 72:319–326
61. Zhang X, Li Z, Zhu S, Liang W (2016) Detecting spam and promoting campaigns in twitter. ACM Trans Web 10(1):4
62. Yang Z, Wilson C, Wang X, Gao T, Zhao BY, Dai Y (2014) Uncovering social network Sybils in the wild. ACM Trans Knowl Discov Data 8(1):2
63. Yang C, Harkreader R, Gu G (2013) Empirical evaluation and new design for fighting evolving twitter spammers. IEEE Trans Inf Forensics Secur 8(8):1280–1293
64. Aslan, Ç. B., Sağlam, R. B., & Li, S. (2018, July). Automatic detection of cyber security related accounts on online social networks: Twitter as an example. In Proceedings of the 9th International Conference on Social Media and Society (pp. 236–240).
65. Sohrabi MK, Karimi F (2018) A feature selection approach to detect spam in the Facebook social network. Arab J Sci Eng 43(2):949–958
66. Singh M, Bansal D, Sofat S (2018) Who is who on Twitter–spammer, fake or compromised account? A tool to reveal true identity in real-time. Cybern Syst 49(1):1–25
67. Setiawan EB, Widyantoro DH, Surendro K (2018) Detecting Indonesian spammer on Twitter. In: 2018 6th international conference on information and communication technology (ICoICT), May. IEEE, pp 259–263