Social Media Ecosystem: Review on Social Media Profile's Security and Introduce a New Approach

Vishnu Dutt Sharma, Santosh Kumar Yadav, Sumit Kumar Yadav and Kamakhya Narain Singh

Abstract Social media is a platform for maintaining a connection among like-minded people. It provides business opportunities, alliances, career opportunities, friendships, relationships, developing social skills, online communication, virtual community and many more. From the last one decade, social media is popular medium for communication among cardinal online users. They provide communication mechanism that enables users to be connected among family and friends. Social media plays a vital role for users because it cringes geographical borders. While the popularity and usefulness of social media make a speculative toward users with regards to data security. In this paper, we have analyzed different social network attack techniques with its subsequences. On the basis of analysis, we propose the machine learning-based system for better security arrangement and data security.

Keywords Social media profile · Social media security · Data security · Social media ecosystem

1 Introduction

Social media (SM) is a platform that felicitates almost all kind of activity. Communication and data sharing are very important aspects where users used to publicize their

V. D. Sharma · S. K. Yadav
Department of Computer Science, Shri Jagdish Prasad Jhabarmal Tibrewala University, Jhunjhunu, Rajasthan, India
e-mail: vashistha31@gmail.com

S. K. Yadav
e-mail: drskyadav@hotmail.com

S. K. Yadav
Department of Computer Science, IGDTUW Delhi, New Delhi, India
e-mail: sumitarya007@gmail.com

K. N. Singh (✉)
School of Computer Applications, KIIT Deemed University, Bhubaneswar 751024, India
e-mail: kamakhya.vphcu@gmail.com

R. R. Chillarige et al. (eds.), *Advances in Computational Intelligence and Informatics*, Lecture Notes in Networks and Systems 119,
https://doi.org/10.1007/978-981-15-3338-9_27

fascinate, snap, video and many more. The importance of these activities is because of factual about personal information, conversation among other members and browsing other user profiles. SM plays a vital role for users because it cringes geographical borders also. In addition, they are also using for job searching, entertainment, news, etc.

In the last many years, we are observing exponential growth toward data being created on SM. In the history of technological revolution, SM proliferated is unprecedented toward speedy growth [1]. The rapid progress of technology toward SM platform has led to an ecosystem where users use various devices for various platforms to interact with a variety of ideas, services and products [2]. There are three types of SM ecosystem: owned media like blog or company website, paid media like advertisement or sponsorship and earned media like viral message [3] (Fig. 1).

There are 7.7 billion total population in the universe, 4.3 billion internet users, 3.4 billion active users on SM. On average, a person has 5.54 SM accounts; the average daily time spent is 116 min per day. SM users grew by 121 million between Q2 2017 and Q3 2017 [4]. In every 15 s, a new SM user account is created. Facebook, WhatsApp, Twitter, LinkedIn, Instagram, Google+, Pinterest etc. are more popular SM. While the popularity and usefulness of SM make a high risk toward end users in terms of data security. Data generates when you register for SM. During sign up, you need to be filled certain details like name, gender, date of birth, email, mobile number, etc. These data along with additional personal information like school name, maiden name, current city, native place, employer, various group, friend in network, IP address user used for every logging, and all of the user's activity ever used to target. SM like Facebook maintains its activity log as a list of your posts, story,

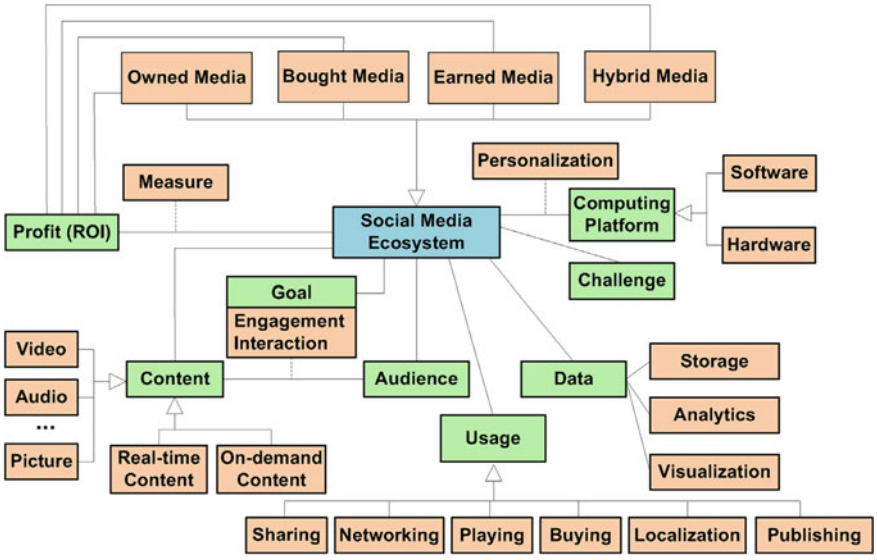

Fig. 1 Social media ecosystem

tagging, like, share, comment, added someone as a friend, status change and search of another person from today back to the very beginning. Hackers target these data. Various attacks for instance identity theft, spam, social bots and malware are used with these hacking techniques—keylogger, denial of service, waterhole attacks, fake WAP, clickjacking attacks, virus or Trojan, phishing, eavesdropping (passive attacks), etc.

The literature survey is given in Sect. 2; working mechanism of proposed model is presented in Sect. 3. The last section presents future work and conclusions.

2 Literature Review

SM platforms invite a various of attacks in them, which provide scope to steal users' identity and trust over a network. There are three major types of security attacks in social media which put users' shared data at risk [5]. The first type is the traditional security attacks like phishing, malware, Sybil, spamming, clickjacking, deanonymization, inference and profile cloning attack. The second class comprises the threat in multimedia content which is shared on SM and used to expose SM users. In this metadata, shared multimedia content, ownership data, multimedia shared links, multimedia content exposure, transparency of data centers, static links, tagging, steganography, video teleconference, and data disclosure by unauthorized attackers. In the third category, it contains threats in which intruder sets up a social relationship with social media users to threaten them. Cyberbullying and grooming, corporate espionage, and cyberstalking are in this category.

In addition, adversaries can find any other relevant credentials, such as bank account, date of birth, by inspecting the user's personal data present in social media and can commit internet crimes, for example, bank fraud [1]. Usually, SM attacks can be account hijacking, impersonation attacks, fraud and malware distribution. An advanced attack can also compromise enterprise networks [6]. Facebook collects and stores user's personal data that can be used to populate target users' ads for marketing. It includes what users' likes, dislikes, shares and friends. Using Facebook, all kind of data is produced and shared but mostly multimedia data. In Zephoria Digital Marketing (ZDM) [7] report, approximately 136,000 photos are uploaded in every 60 s on Facebook website. According to statistics on SM, it shows that the average viewing and sharing rate of videos is rapidly increasing. In the current year, approximately eight billion videos per day are viewed on Facebook, which is doubled view count of the year 2015 [8]. Because of the availability of enormous multimedia data on various social platforms such as Facebook, security risks are also rising in its own pace. An intruder user can share objectionable information on any social media. In addition, an adversary can extract a user's essential information from those multimedia data, such as user's location, relationships and identity [9]. Twitter safeguards the users' data and preserves the confidentiality of their private information; however, intruder can analyze the sequence of a user's posts on a social network (SN) and can infer their private information. MySpace social platform was

attacked in 2005 by the Sammy worm. The worm exploited the vulnerabilities in MySpace and spread quickly over it [3]. The malicious code did not steal users' secret information; rather, it had effect on general operations of MySpace. Later in 2009, the social site Twitter get affected by malicious Mikeyy worm, which replaced users' data with some objectionable information. In the same year, May 2009, Facebook was again attacked by a malware named Koobface worm. It stole important information of users, for example, a user's password, etc. [10]. A report by Internet Security Threat Report (ISTR) [11] projected that the overwhelming use of SMS by attackers cannot be overlooked. In 2015, those services acted as a source for malware and malicious spam. Those were further used as a pathway of earning illegal money on the internet. In recent times, Pinterest and Twitter accounts of Facebook CEO Mark Zuckerberg's were hacked. In that, the attacker exploited his LinkedIn profile credentials [12]. In a similar way, many other attackers compromised the SN accounts of Delta Air Lines Inc. and Newsweek. They used fake message post for it [2].

From the above attacks, we infer that SM is the most available pathway for an adversary to commit cybercrimes. Many researchers have proposed various solutions to withstand such attacks, for example, digital oblivion and watermarking steganalysis [13]. Many other solutions to mitigate traditional threats are also proposed such as phishing detection and spam detection [14]. Many researchers have investigated and stated various security problems in SNs. Gao et al. [15] in their research classified the prime security issues in social media into four: first, network structural-based attacks; second, malware attacks; third, privacy issues and fourth, viral marketing. The research elaborated the security problems and their corresponding defense security mechanisms. Author's studied user's behavior in SMS from four prospects: (1) traffic activity, (2) connection and interaction, (3) malicious behavior and (4) mobile social behavior [16].

Viejo and Sánchez [17] provided a comprehensive study on various security and privacy threats in SMS. Those are primarily segmented the threats into four classes: (1) classic threats, (2) modern threats, (3) threats targeting children and (4) combination threats. They also proposed a classification of existing solutions for protecting SNS users. Kayes and Iamnitchi [18] presented a categorization of possible traditional security threats in SMS based on the SN stakeholders. Authors classified the various attacks as attacks on social network security (SNS) infrastructure. Author also provided many defense mechanisms to mitigate these sets of SMS security attacks. However, various challenges that occur in using these mechanisms are real-world implementation. Kumar et al. [19] provided many noted attacks in SMS, including identity theft, phishing attacks, Sybil and other malware exploiting flaws in real-world implementations (Table 1).

Table 1 Social security solutions, key methods and descriptions

Machine learning approach	Related work	Algorithm description
AdaBoost	[16, 17]	Weak results from my other decision tree are taken together to create a boosted classifier
Boosted classifier	[4, 13]	Simple logistic regression used in Bayesian generalized linear algorithm
J48	[6, 15]	J48 is a decision tree algorithm
KKNN	[14, 20]	Clustering in K-nearest neighbors is used group and further predict classifications type
Nnet	[8, 21, 22]	A neural network mimics the principle of neurons in a brain for classification problem
.rf	[6, 13]	Random forests are made up of a set of decision trees to obtain the optimized accuracy with one such tree
Rpart	[6, 18]	Recursive partitioning tree is also can be used as a decision tree
SVM Linear	[2, 19]	SVMs classifier used for high-dimensional features classifications. It also classifies the nonlinear feature space

3 Proposed Work

There are various methods used by machine learning to analyze data, broadly grouped into supervised and unsupervised. For making enhanced decision in the future, learning starts from observation of data, understanding and training examples to search pattern [21]. Classification and regression algorithms come under supervised learning. Prediction can be done on the basis of trained dataset along with input data and corresponding desire output for the new example.

While unsupervised learning applies untrained dataset using cluster analysis method. Dataset is grouped into various clusters using Euclidean or else probabilistic distance.

Comprising different types of data is enormous in volume which is challenging to make statistical analysis. Because of inadequate technology and possible redundancy data are noisy and heterogeneous. To bring effective conclusion from high-resolution measurement requires machine learning approaches. We apply logistic regression (LR) and support vector machine (SVM) classifier training methods. By removing redundancy, dimensionality of data portion is decreased.

LR is used to classification which accepts only two values, i.e., 0 and 1 to predict.

Here will use a and b district value toward classification and apply LR to attempted to calculate b given a.

We know that $b \in \{0, 1\}$ and so $h\theta(a)$ must belong within 0 and 1.

For this purpose, we will take

Where $h\theta(a) = g(\theta Ta) = 1/1 + e^{-\theta Ta}$ and $g(c) = 1/1 + e^{-c}$ belong to logistic or sigmoid function. $g(c)$ tends toward 1 as $c \to \infty$, and $g(c)$ tends toward 0 as $c \to$

$-\infty$. Moreover, $g(c)$, and so $h(a)$, always bounds between 0 and 1. As before, we are keeping the convention of letting $a0 = 1$, so that $\theta T(a) = \theta 0 + \sum_{i=1}^{n} \theta_j a_j$.

The classification model is endowed with a number of probabilistic assumptions to fit θ for it, and then the parameters are fitted using maximum likelihood.

Let us assume that

$$P(b = 1|a; \theta) = h\theta(a)$$
$$P(b = 0|a; \theta) = 1 - h\theta(a)$$

This can be presented more efficiently as

$$P(b|a; \theta) = (h\theta(a))b(1 - h\theta(a))1 - b$$

SVM is also applied for linear kernel, classification and training [8, 12]. SVM applies hyperplane to segregate two kinds of data as a generalized maximal margin classifier. From the training examples, maximal margin hyperplane has to be selected. The distance of given segregating hyperplane from training example has to be calculated. The margin should be minimum distance from hyperplane to example. Above hyperplanes can be represented like:

$$a \cdot b - x = 1 \text{ and}$$
$$a \cdot b - x = -1$$

Obviously, to get $2/\|z\| \sum$ to get maximized, the value of z should be minimum.

4 Conclusions and Future Scope

SM is a platform that provides various types of services like share and like photos, videos, story, interest, comment, feelings, etc. SM makes billions of Web users engage among them without geographical restriction. We have proposed to provide the latest approach on various secrecy and security challenge that occurs from some of their important characteristics. To realize the challenges, we recap several latest attack data and security details. Moreover, we described the ecosystem of SM. Furthermore, we have done analysis of the existing schemes to protect users from social network threats. We explored here two classifier training methods LR and SVM for developing or updating new system. For future work, need to implement using machine learning-based approach or artificial analysis, which can provide better result based on content. Finally, the testing of system and result analysis will be done for use.

References

1. Carminati B et al (2011) Semantic web-based social network access control. Comput Sec 30(2–3):108–115
2. Squicciarini AC, Griffin C, Sundareswaran S (2011) Towards a game theoretical model for identity validation in social network sites. In: 2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing, IEEE, 2011
3. Novak E, Li Q (2012) A survey of security and privacy in online social networks. College of William and Mary computer science technical report, pp 1–32
4. Savage J (2017) Top 5 Facebook video statistics for 2016 [Infographic]. EriĢim, vol 26
5. Pang J, Zhang Y (2015) A new access control scheme for Facebook-style social networks. Comput Sec 54:44–59
6. Singh KN, Misra C, Seth S, Mantri JK, Dash SR (2017) A framework for social service volunteers: a social network representation. Int J Pure Appl Math 114(10):11–23
7. Cao J, Li Q, Ji Y, He Y, Guo D (2016) Detection of forwarding-based malicious URLs in online social networks. Int J Parallel Prog 44(1):163–180
8. Zhang Z, Gupta BB (2018) Social media security and trustworthiness: overview and new direction. Future Gener Comput Syst 86:914–925
9. Rathore S et al (2017) Social network security: issues, challenges, threats, and solutions. Inf Sci 421:43–69
10. Smith K (2016) Marketing: 96 Amazing social media statistics and facts for 2016. Retrieved from https://www.brandwatch.com/2016/03/96-amazing-social-media-statistics-and-facts-for-2016
11. Gao H, Hu J, Huang T, Wang J, Chen Y (2011) Security issues in online social networks. IEEE Int Comput 15(4):56–63
12. Miller Z, Dickinson B, Deitrick W, Hu W, Wang AH (2014) Twitter spammer detection using data stream clustering. Inf Sci 260:64–73
13. Lee JD, Sin CH, Park JH (2014) PPS-RTBF: privacy protection system for right to be forgotten. J Converg 5(3740.16)
14. Singh KN, Misra C, Seth S, Mandal SK, Kumar BA (2019) Robust framework for socially responsive services: a constraint-based social network representation. In: Innovations in soft computing and information technology. Springer, Singapore, pp 163–177
15. Gao Y, Hu S, Tang W, Li Y, Sun Y, Huang D et al (2018) Physical layer security in 5G based large scale social networks: opportunities and challenges. IEEE Access 6:26350–26357
16. Jin L, Chen Y, Wang T, Hui P, Vasilakos AV (2013) Understanding user behavior in online social networks: a survey. IEEE Commun Mag 51(9):144–150
17. Viejo A, Sánchez D (2016) Enforcing transparent access to private content in social networks by means of automatic sanitization. Expert Syst Appl 62:148–160
18. Kayes I, Iamnitchi A (2017) Privacy and security in online social networks: a survey. Online Soc Netw Media 3:1–21
19. Kumar S, Saravanakumar K, Deepa K (2016) On privacy and security in social media–a comprehensive study. Procedia Comput Sci 78:114–119
20. Rubin VL (2017) Deception detection and rumor debunking for social media. The SAGE handbook of social media research methods. SAGE, p 342
21. Noyes D (2017) Zephoria digital marketing. The top 20 valuable facebook statistics. Viitattu 17.11. 2017
22. Sedhai S, Sun A (2017) Semi-supervised spam detection in Twitter stream. IEEE Trans Comput Soc Syst 5(1):169–175