ISSN 2250-3153

A Risk Management Framework for Clouds Using Big Data and Security Informatics usingAttack Trees and Hidden Markov Model in Analysis and Prediction of Risks in Social Media Networks

K.D.B.H. Subasinghe[*], S.R. Kodituwakku[**], H.S.C. Perera[**]

[*] Department of Information Technology, Faculty of Computing, Sri Lanka Institute of Information Technology
[**]Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya
[***]Department of Mechanical Engineering, Faculty of Engineering of Sri Lanka Institute of Information Technology,

***Abstract*-** Social media refers to the means of interactions among people in which they create, share, and exchange information and ideas in virtual communities and networks. The growth of Information and Communication Technology (ICT) has resulted in an enormous volume of security related information present on the web largely when it comes to social media networks. Therefore, with the changing face of cyber security, although it is difficult, it was found that detecting the potential cyber-attacks or crimes is possible and feasible with the vast improvements in ICT. Cloud computing uses ICT resources that are delivered as a service over a network which has opened a promising opportunity across the globe thus resulting a greater popularity of e-commerce. The proposed framework is developed to manage risks of social media networks using the attack tree method which is used to model the risk of the system and identify the possible attacking strategies which the adversaries may launch. This paper presents the development of a Risk Management Framework by analysis of social media networks through web intelligence and security informatics using attack tree analysis based on the Hidden Markov Model for information extraction and prediction of risk factors of Social Media Networks.

***Index Terms*-** Social Networks, Risk Management, Attack Tree, Big Data, Security Informatics, Web Intelligence.

I. INTRODUCTION

Andreas Kaplan and Michael Heinlein define social media as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content [1]." Furthermore, social media depend on mobile and web-based technologies to create highly interactive platforms via which individuals and communities share, co-create, discuss, and modify user-generated content. It introduces substantial and pervasive changes to communication between organizations, communities and individuals. Social networking platforms therefore allow organizations to improve communication and productivity by disseminating information among different groups of employees in a more efficient manner, resulting in increased productivity. Therefore, upon carrying out a detailed investigation on the social media networks, it was found that they facilitate open communication, leading to enhanced information discovery and delivery allowing employees to discuss ideas, post news, ask questions and share links where communication happens in seconds and perhaps in a button click. In addition, when the advantages of social media networks are analysed it was found that they provides an opportunity to widen business contacts, targets a wide audience, making it a useful and effective recruitment tool, improves business reputation and client base with minimal use of advertising and expands market research, implements marketing campaigns, delivers communications and directs interested people to specific web sites. On the other hand, when the disadvantages are considered it was found that they open up the possibility for hackers to commit fraud and launch spam and virus attacks, increases the risk of people falling prey to online scams that seem genuine, resulting in data or identity theft, may result in negative comments from employees about the company or potential legal consequences if employees use these sites to view objectionable, illicit or offensive material and potentially results in lost productivity, especially if employees are busy updating profiles, etc.On the other hand, since social networking websites are very famous, hackers and spammers are active on these websites and might use them to gather confidential information. However, as we discussed social networking websites do play an important role in gathering relevant information of an individual and organization; but, the privacy and security issues related to social networking websites are not new and it is not very easy to fight these problems due to the tremendous number of online users per second.Research reveals, there are so many ways that a hacker might use them for a bad purpose, but in general social networking websites are famous forInformation gathering (intelligence gathering), Phishing, Fraud andSpamming.

Information gathering is the first and an important step of hacking and the success ratio is directly proportional with the information being available, where the social networking websites certainly do have the information which is required. For example, it was very common to hack an email account by using a social engineering technique, and the technique was to click on "forget password" and try to recover this account by providing some relevant information of the person, information which could be fetched from the social networking websites. You can even judge the secret answer by the activity of the person. In fact, the practical example of this scenario is the study called "Getting in Bed with Robin Sage", which was conducted by Thomas Ryan, a security specialist, and the results of this study show how dangerous a social networking website could be.Robin Sage is the fictional identity that was created for this study. Here, the researcher chooses a picture of a very beautiful girl; and the logic is to attract the opposite gender and to connect with more and more people. Therefore, he created a fake profile on the famous social networking websites: Facebook, Twitter and LinkedIn. To give them a real and professional look he completed her profile with job and educational information which was fake too. Later, in the 28-day study, "Robin" contacted hundreds of people, most of them belonging to government sectors such as the DOD and military intelligence. Therefore, with this completely fake profile, Ryan was able to get email addresses, bank account numbers, invitations to conferences, and even a job. The most important aspect of this study is that Robin was offered to review confidential information and papers written by professionals.

In addition, while performing a detailed investigation on this scenario, according to a news report, it was found that 83 million Facebook profiles are fake, and the question arises as to who has created these fake profiles and for what purpose have they been created. Certainly, each purpose might be different, for example an individual might create a profile to spy on another individual, but the danger could be if actually the numbers of fake profiles belong to a particular organisation. The IT department of an organization is responsible to take care of this and to fight with the situation, which leads to a loss to the organization and can harm the reputation of organization. Subsequently, phishing is another dangerous attack which is very common in social networking websites. It seems that hackers usually target individuals for their phishing attack, but what if they target an accountant or any other person who is responsible to manage the finance of organization? In both cases it is very dangerous because the capital factor is involved. A phishing attack can easily lead to a very dangerous situation, as a smart hacker can compromise the complete computer network of an organization.

Consequently, upon carrying out a detailed background review it was revealed that it is an urgent requirement that the enormous volume of security relevant information is managed and that a risk management framework is set up in the Social networks. This involves exploration and analysis of the social media network data using a systematic approach such as attack trees which would be enhanced using the Hidden Markov Model for information extraction and prediction.

II. BACKGROUND

CurrentlySocial media networks provide the people with the ability of ensuring complete connectivity, bringing people with common interests together, and creating a platform to share your life with the rest of the world. On the other hand, it has its negative aspects which are an enormous issue that needs attention. Therefore, upon carrying out a literature review on the advantages and the disadvantages of the available social media networks, it was found that there's a clearly visible threat to the critical information. Accordingly, this Social Media Era has brought with it new risks for employers, in terms of privacy, confidentiality and employee loyalty. Therefore, on the other hand, when the dangers of the presently available social media networks are considered, it was found that a balance of good and bad needs to be maintained and naturally when a far reaching medium such as the internet brings us social networking sites that spice up our lives a little, it also attracts the fractious few of society. Therefore, some of the things that could lead to potential social disaster mainly include the term hacking which is one e potential risk of social networking which is cited most often and the incidents of profiles and accounts being hacked into are commonplace. While this could be restricted to simply playing a practical joke on a friend, it sometimes leads to the more serious misdemeanour of identity theft. In fact, this requires a low level of technical skill and is referred to as social engineering. This technique banks upon the psychological aspect of a 'friend' connection; the hacker uses common interests, background and professional information all of which are posted on profiles, to extract sensitive information, like passwords and other details from the targeted individual and use them to create an alternate identity. Just using simple data like date of birth, name and location allows hackers to create fake social security cards, driver's licenses and ID cards.

In addition, it is reported that almost 15 million Americans are victims, either directly or indirectly, of identity theft. Despite this, on the other hand previously carried out research statistics show that even though people are adequately aware of the risks, only 40% of them manage their privacy settings which is according to a survey conducted by PC World Thus investing some time on securing your account against fraudsters and unknown individuals would prove to be beneficial in the long run. [2] People recruited in high security professions like the defense services, project engineers and scientists face an added disadvantage of hackers misusing any leaked classified information. It is seen that any unintended slip of information could have serious disasters, if it falls into the wrong hands. Another potential hazard of putting up pictures and videos on these websites is that they could be used for defamation where the reported cases include girls' photos being photo-shopped and used in objectionable places. In fact, it is also found instances where people use social networking sites to abuse or defame anonymously are constantly reported which is again a serious problem around the globe. Subsequently, Social media has brought to the world a common medium for thoughts, words and expression. However its

correct and efficient usage is completely dependent on how it is managed. Nevertheless, being aware of the risks and vulnerabilities that we are exposed to through these media only empowers further to use them in a positive way. It is up to us to make sure that social networking does not turn into social dysfunction. With the rapid increase in the use of social media networks, some organizations addressed social media risks by prohibiting its employees from participating on social media websites when at work. Subsequently, most organizations have now realized how social media can boost their marketing and advertising strategies and they are embracing that media is a part of their business strategy. However, when doing so, organizations would benefit greatly from properly understanding and managing the risks of social media.

Therefore, upon carrying out an exploration on the use of social media networks as a globe and upon analyzing the statistical information being obtained, it was decided that global security and international security are constantly evolving and traditionally, the terms have dealt with states and issues of interstate war and power. On the other hand as the World Wars and the Cold War wound down and the number of major interstate conflicts dwindled, the focus diverted to encompass international organizations and intrastate issues like civil wars. However, due to globalization and major terrorist attacks, the focal point has shifted to access the effects of non-state actors. Therefore, with the advancement in technology as a globe, it's pretty much clear and a vital fact that almost all countries and more that 75 percent of the population of a country in this entire globe has access to information communication technologies, therefore the use of these powerful technologies itself has advanced the ability of those non-state actors including terrorists, criminals and rioters to impact the international security.

In fact, it's important to understand that we live in a globally connected world and along the statistics proven it's seen that a single extraordinary action of a man can spark events that affect lives of billions of people living in this earth and bring about a terrible disaster which can account to a much terrific one worse than a nuclear war. So when we say global security or international security, it's important to understand that we mean everything that affects the safety and the livelihood of the people surrounding you. Currently, in the process of carrying out a detailed research on the usage of Social media networks as a globe we found certain statistical information upon a survey being carried out where, 63% respondents indicated "that employee use of social media puts their organization's security at risk" while 29% "say they have the necessary security controls in place to mitigate or reduce the risk" [2]. In a different study, 49% of executives surveyed said that they feel that the use of social media could damage company reputation, yet in that study "only one in three companies addressed those concerns" [3]. Indeed, the growth rate of some of the more popular social media networks is phenomenal. FaceBook, a social networking site, has reached close to 700 million users and when the statistics are analysed, when you take a count of the number of users, well if FaceBook was a country, it would be the third largest. According to a brochure released by Websense, FaceBook has an annual growth rate of 41% and Twitter is growing at 85% year after year [4]. Therefore, it is clearly seen that there's a great risk to the mankind with the development of social networks across the globe and it was also a major issue that was discussed during the Commonwealth Heads of the Government Meeting 2013 that was held in Sri Lanka where the Heads of Government reaffirmed the importance of addressing the increasing threat of cybercrime, which poses a major obstacle to socio-economic growth, peace and stability across the globe.

During the rigorous process of review of the past research it was also found that research was carried out which has initiated the development of an Attack tree based risk assessment for the Location Privacy in Wireless Sensor Networks (WSNs)[11]. Fault tree analysis (FTA) developed by Bell telephone Laboratories in 1962 is a top-down approach to failure analysis, commencing with the potential undesirable event, and then determining all the ways a particular vulnerability can happen[12]. Back in time from the 1990's the tree structure have been applied to investigate the attack profile based on FTA (Fault tree analysis) by discovering all possible attack paths of threat lists that evolve to form an attack tree. In fact, attack graphs can simulate complex attack scenarios, but its disadvantage is the complexity in visualization [5] [6]. The attack tree approach which provides a formal and a methodological way of describing the security of systems based on varying attacks was proposed by Bruce Schneier to model threats against computer systems[8][9]. In addition, on the other hand, the attack trees easily answer "what-if' games with potential attack or defense action, and allow the developer to refine the attacks to the level of detail desired [7]. Nevertheless, its disadvantage is that it is quite difficult to provide an effective means of stating the attack and defense profile in an interactive way. Consequently, attack trees seem to be more suitable for assessing the threats. In fact, for the purpose of statistical analysis using a probabilistic model, research proved the use of Hidden Markov Model[14] which is appropriate for the use of statistical pattern analysis and it was found that since the early 1970's HMM's have served application areas such as biological sequence analysis and SMS spam detection in the mobile communication industry[15]. Therefore, upon considering the currently performed research in this aspect, we claim that, in order to assess the safety/security properties of a target critical infrastructure, it is necessary to assess both the critical event chains caused by random events and by malicious cyber events.

III. METHODOLOGY

Riskassessment in social media networks focuses a major concern on the 'TRUST' of the content that is being shared on the World Wide Web [3]. In fact, it's always a question whether we can really trust the people with whom we are speaking to, or are the data we publish and share going to be secure over the wireless networks and rather are we communicating the data amongst the right people. Nevertheless the world is bloomed with the emerging technologies and the use of smart devices offering a tremendous convenience

and service both to industry and its end users. In addition, the world has itself become an IOT or an Internet of Things. The proposed framework is developed using the attack tree method which is used to model and analyze the risk of the system and identify the possible attacking strategies which the adversaries may launch. Therefore, using this attack tree model, it is convenient to analyze the capability of the attack source and estimate the degree or the impact of a certain threat that might bring to a system. Here, the multi attribute theory is adopted to calculate the total probability of reaching the attack goal [10]. Thereafter attack sequences are constructed in order to identify all the possible attacks on the social media networks. Therefore, according to the quantitative result the decision maker of the system is able to identify which attacks are of the greatest possibility and then decide what protection measures should be used to counter these attacks. In fact, here we would assign values to the leaf nodes and calculate the system's risk followed by carrying out the attack sequences on the basis of the attack tree.

Here, a tree named attack tree is constructed by identifying the vulnerabilities and threats of Social Media Networks and it is used to represent attacks against a system with the final desired goal as the root node and different ways of achieving that goal as child nodes. Each child node of the root becomes a sub-goal and children of that node are ways to achieve that sub-goal. If one of these nodes cannot be divided further, it's a leaf node. Otherwise those nodes are treated as sub-goals separately and are divided continually until all events become leaf nodes. Therefore, according to the logical relationship among them, those nodes which are linked with an OR-Gate or an AND-Gate are OR and AND nodes respectively. OR- gates are used to represent alternative attack methods and AND-gates are used to represent different steps towards achieving the same goal [13].Our next step in successful development of the attack tree is the carrying out of the risk assessment by assigning attribute values to the attack tree nodes. There is no systematic method available to determine parameter values for each node in the attack tree. Therefore, in order to evaluate these values, certain aspects of details of the system including protocol, hardware, application, operating system, environment as well as attack software and tools are required. Therefore, here we would consider three attributes of the leaf nodes which are the attack cost which is the proportion of the costs and benefits, technical difficulty and the probability to be discovered.

A Hidden Markov Model (HMM) is defined as a statistical model that describes a series of observations generated by stochastic process or Markov process [14]. The HMM is used to make use of the observation information of social media risks to gain insight to various aspects of the underlying Markov process. In fact, a Markov process is a sequence of states where the progression to the very next state depends entirely on the present state, and not the past states. Therefore, the Markov process in the HMM is hidden and it can be described as a finite state machine which has some fixed number of states and it provides a probabilistic framework for modeling a time series of multivariate attack observations. The fundamental objective here is to make use of the observation information of social media risks to gain insight to various aspects of the underlying Markov process. Therefore, here it was decided to develop and train a HMM to represent a collection of risk data which is in the form of observation sequences produced by the attack trees. Therefore, a HMM is developed and trained to represent a collection of risk data which is in the form of observation sequences produced by the attack trees. According to the quantitative result the decision maker of the system is able to identify attacks which are of the greatest possibility eventually decide what protection measures should be taken to counter these attacks thus developing an effective risk management framework for the risk analysis of social media networks.

IV. DISCUSSION

Subsequently, the risk associated with a specific attack on an asset can be reduced by reducing the level of threat to it, by reducing its vulnerability to that threat, or by reducing the consequences or impact of an attack should it happen.Accordingly, on the basis of the attack tree we can study the attacker's behavior by constructing attack sequences. An attack sequence S is a real path which an attacker can implement and is composed with a set of leaf nodes. The attacker can achieve the final goal in case all the attack events of the nodes in the attack sequence occur. Once the attack sequences have been known their probabilities can be worked out and we can compare them to figure out the attack sequence which the adversary may launch most likely. In fact, we can adopt the Boolean algebra method to get all the attack sequences of the attack tree. The probability of the attack sequence is the product of the probability of all the leaf nodes involved in the attack sequence.

Therefore, initially as a step for the development of a risk management framework, it is suggested and proposed that the use of Attack Trees for the Attack Sequence Analysis would enable us to clearly identify which attack sequences and paths are most likely to happen, where thereby we can mitigate the ability of an attacker reaching the final goal by avoiding the occurrence of the attack events of the nodes of the sequences which reduces the cost of recoverence from the threat once it would have already occurred as prevention is always extremely important [16][17] . A HMM is incorporated as a state machine where the transitions between the states are characterized by fixed probabilities and every state is associated with a single probability for observing a group of observation symbols. The states in the trained HMM are used to represent attributes of input data, while the transition and observation probabilities stand for statistical properties of these attributes. Therefore, for any given observation sequence produced by the attack trees, a match against a trained HMM is produced to determine the probability of visualizing such an attack sequence. The probability computed is therefore high if a similarity is detected between the observed and trained sequence. Therefore, as the threat to social media networks is a widely spoken and researched area, we propose that more attention needs to be put forward in order to develop a well lit Risk Management Framework in order to mitigate the risks and losses being faced by the individuals around the globe.

V. CONCLUSION

Therefore, upon carrying out a comprehensive investigation on the cyber attacks of the social media networks we have proposed an innovative risk assessment approach for the risk management of social media networks. Here, we analyze the threats and vulnerabilities from the network point of view and build the attack tree by enhancing prediction probability using the Hidden Markov Model. In fact, the ongoing and future research that we are pursuing is targeting towards complete development of an expert system based on a trusted diagnosis model thus develop a fully automated Risk Management Framework for the Social Media Networks.

REFERENCES

[1] Kaplan, Andreas M. and Haenlein, Michael. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53, 59-68.

[2] R.Schullich,"Risk Assessment of Social Media" GIAC (GSEC) Gold Certification, December 5th 2011

[3] "Social Media Guidelines: Command and Control or 'Let 'er Rip!'," *PR News*, September 13, 2010.

[4] David Kirkpatrick, The Facebook Effect: The Inside Story of the Company That Is Connecting the World, Simon & Schuster, page 200.

[5] M. Rausand and A. Hoyland, System Reliability Theory; Models,Statistical Methods and Applications, Wiley, 2004.

[6] o. Sheyner. Scenario Graphs and Attack Graphs. PhD thesis,Carnegie Mellon University,2004.

[7] S. Mauw, M. Oostdijk, "Foundations of Attack Trees," In:Won,D.H., Kim, S. (eds.) ICISC 2005. LNCS 3935, pp.186-198. Springer, Heidelberg, 2005.

[8] B. Schneier, "Attack Trees: Modeling Security Threats." Dr. Dobbs' Journal: Dec 1999.

[9] Masera M, Nai Fovino I. Through the description of attacks: a multi- dimensional view. In: The 25th international conference on computer safety, security and reliability. 26–29 September 2006 Gdansk, Poland.

[10] R.Sarin, Multi-attribute utility theory, Encyclopedia of Operations Research and Management Science 2001.

[11] D.D.Ren, S. G. Du, H.J. Zhu and Ieee, A Novel Attack Tree Based Risk Assesment Approach for Location Privacy Preservation in the VANETs. New York:IEEE, 2011.

[12] D. Marquez, M. Neil and T. Stalhane. "FMEA and BBn for robustness analysis in web-based applications." European Safety and Reliability Conference 2007.

[13] Rong Jiang, Jun Lou, Xiaoping Wang"An attack tree based Risk Assesment for Location Privacy in Wireless Sensor Networks,"IEEE 2002.

[14] Krog,A.(1998), "An Introduction to hidden Markov models for biological sequences",Computational Methods in Molecular Biology, pp. 45-63, Elsevier

[15] Rafique, M.Z. & Farooq,M. (2010) , "Sms spam detection by operating on byte-level distributions using hidden Markov models", Virus Bulletin Conference.

[16] Stamp, M.(2012), "A Revealing Introduction to Hidden markov Models", Department of Computer Science, San Jose State University.

[17] S.S. Joshi and V.V Phoha, 'InvestigatingHidden Markov Models Capabilities in Anomaly Detection," Proc. 43rd ACM Ann. Southeast Regional Conf.,vol. 1, pp.45-55, 2003.

AUTHORS

First Author – K.D.B.H. Subasinghe is with the Postgraduate Institute of Science, University of Peradeniya and Department of Information Technology, Faculty of Computing of Sri Lanka Institute of Information Technology - buddhimahs@hotmail.com / buddhima.s@sliit.lk

Second Author – Professor S.R. Kodithuwakkuis with the Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya - saluka.k@gmail.com

Third Author – Professor H.S.C.Perera is with the Department of Mechanical Engineering, Faculty of Engineering of Sri Lanka Institute of Information Technology - chandana.p@sliit.lk