ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

Proactive Approach with Data Analytics PRADAN

Namitha Cleetus T[1], Princy M[2], Deepthy J[3]

P.G. Student, Department of Computer Science, DiST College, Angamaly, Kerala, India[1]

P.G. Student, Department of Computer Science, DiST College, Angamaly, Kerala, India[2]

Assistant Professor, Department of Computer Science, DiST College, Angamaly, Kerala, India[3]

ABSTRACT: Digital Forensics has been common for the study of method to regain and explore material found within the digital devices to examine in order to solve crimes which comprise the computer and internet. Nowadays in our day to day life criminal activates are increasing especially internet fraud, which can utilize forensic technique for proactive approaches. Prevention minded technology can be performed efficiently through trust analytics which is conducted atop modern data and knowledge technologies. Here we introduce PRADAN which continually examines the trust worthiness and risk of social media. PRADAN uses polystore based data management system in order to save the historical details, observable social network, together with the domain knowledge taken from the social context of message.

KEYWORDS: Data analytics, Digital-forensics, Proactive approach, Polystore

I. INTRODUCTION

Digital Forensic is a branch of forensic science encircling the reestablishment and investigation of material found in digital data. Its main deal is to search, collect and analyse the evidences from electronic media, such as in the disk storage, mobile devices computer network as well as social media. However a forensic investigation is a retrospective activity, it usually starts after the crime which has committed.

In real life all the investigation analyses the message history from facebook and mobile phones and call logs of phones used in communication to triangulate criminals. Proactive forensics can be applied for online frauds, such as inheritance schemes, lottery/prize schemes, online sales schemes, bank and financial account schemes and romance schemes which naturally differ from traditional forms of fraud. These types of frauds capitalize on prolonged communication between a victim and an adversary which means interactions that may spread over multiple channels and may be publicly visible or private between the parties. A crucial feature of these types of scam is that making the victims to give their information directly for the execution of crime.

Gradually the victims starts to reveal their sensitive details in natural, as the attackers grows trust in themselves , and the victim never be able to suspect on anything. The aim of proactive approach is to give awareness about the crime prevention by issuing warnings to the victim.

.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

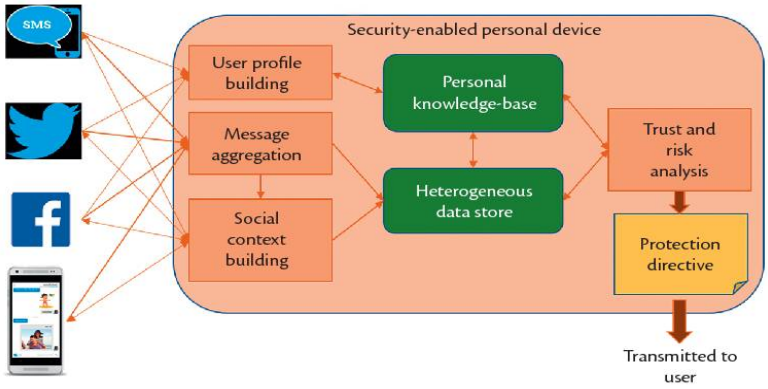

Fig. 1. A high-level architecture of a proactive forensics system [7]

From fig.1 all the data from various social media are clustered and pass through different phases as mentioned below, And then these data is passing to data stores mainly known as polystore. Then the final data are transferred to trust and risk calculations then the result will make available to the user.

PERSPECTIVE OF PROACTIVE APPROACH

The basic processes behind a proactive approach system are as shown in Fig 1. Here the green components are connected to a remote server through a secured connection, and the light brown components are located in the personal device. A subject must give the system a clear permission to access proper social channels like SMS, Facebook and Twitter so that the actions of the software are not considered a breach of privacy and access control policies of the social channels. Once commissioned, the system uses the API of these sources to construct a combined profile of the subject by taking his/her individual profile details from all the source. The Social Context Builder attempts to reconstruct the visible part of her social network over all media channels, and store this information in a personal knowledge base. The Message Aggregator scans messages from different sources, places them in a Heterogeneous Data Store (e.g., the AWESOME polystore) for analytical operations occurring downstream.

The Trust and Risk Analysis Module is the heart of the prevention mechanism. The risk analysis involves computing a trust score for each message. More importantly, it monitors the responses written by the user and assesses the risk associated with the user's message based on the content of the message, the trust of the receiver and the history of trust and risk computed over the lifetime of exchanges between them over all message channels. If the message is evaluated to have high-risk, a Protection Directive – a statement that says which part of the subject's response is high-risk, along with a link that explains why the system made the assessment – is immediately created for the user. Clearly, these directives for every message can be a significant burden both on the system and the subject, and to be practical, it would allow the user to create a "safe list", as well as a risk threshold below which the directive will not be issued [3].

II. RELATED WORK

Digital forensics is also referred to as Digital Forensic Science. It deals with the collection, recovery and investigation of material found in digital devices such as hard disk, USB, mobile phones etc. It is usually associated with computer crimes. Online social networks (OSN) [4] are a permanent presence in today's personal and professional lives of a huge segment of the population, with direct consequences to offline activities [4]. Paper [4] presents the various security attacks in OSN. There are several attacks such as Sybil attacks where users could assume multiple identities to manipulate a service. Then there is Distributed Denial of Service Attack (DDoS) where a malicious activity would overload the system with a service and deny access to it. There are other attacks like attacks from compromised accounts, social spam and malware etc. [3] is a survey paper which focuses on the security issues in online social networks. It later discusses about the different attacks, the role of social network inference and proposes distributed social networks. The authors have provided insight into anonymizing social network data. In [2], the authors have

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

proposed a computational model that helps users predict security risks associated with the information they have shared on social networks. It uses a neuro-fuzzy technique to predict a risk indicator value by assessing the risk attributes. A decision would be made based on this risk indicator value. Results show the relevance and effectiveness of the proposed approach in predicting risks. In [10], authors have analysed the current situations in social networks using "point of interactions" and designed a mathematical model of trust for such networks. There are multiple paths this model could take for development in future. New contexts such as similar number of friends or private message analysis were intended to be added to this model. Another possibility is to dynamically adjust the priority vector according to the amount of collected data [10].

III. CREATING USER DATA

User data contains confirmed or confirmable facts about the user. These facts are gathered and stored in a secure personal information base. Examples include birthdays, workplace details and academic qualifications. We use the term Information-base instead of database for the user data, because we can do reasoning operations like consistency checking on data. In PRADAN system, these gathered facts are stored as instances of information encoded in RDF/OWL [6](Resource Description Framework/Web Ontology Language), a World Wide Web standard for representing semantic information.

IV. CREATING SOCIAL SURROUNDINGS

Social surrounding of a user is created from his/her friends and followers on Facebook or other social channel. It is created when he/she registers with the system for the first time. The created social surroundings then stored in the heterogeneous data store. Next, the network is enhanced by calculating an initial trust value for every Person by combining

(i) Duration of communication to the entity
(ii) Strength of relation to an entity
(iii) Quality of communication between them

For a close friend or a relative the Initial trust value is assigned high, but for "friend request" from an unknown person, it is calculated by considering match in the number of mutual friends. One of the important goals of the system is its ability to identify "friends" with a fake online profile.

V. THE PROCESS OF MESSAGE GROUPING

One of important ability of a proactive data analytic system is its ability to automatically detect problematic messages. A message, whether an email, FB post, SMS message, each has an individual message format which contains:

1. Details of type of data, date and size of message
2. The sender and receivers of the message
3. The body of the message
4. Additional entities like images, audio and video files

The message aggregator converts it to a common internal form. Next, the message is going through an analysis phase. It identify whether the message a trust modifier. After finding the aggregator compute a trust value for the current message exchanged between the user and the message sender. We use interaction-based trust calculation for social media frauds. Here we consider interaction time span, number of interactions, number of characters in each communication, common interest etc., and develop function [7] to calculate trust from these factors. The trust measure for "number of interactions":

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

$$A = \frac{1}{n} \times \sum_{i=1}^{n} I_x$$

$$T_x = \frac{I_x}{(A + \frac{1}{n} \times \sum_{i=1}^{n} |A - I_x|)}$$

[7]

Ix represents the number of interactions for member **x**.
A is the average number of interactions among n members.
Tx represents the computed trust for member **x**.

VI.SECURITY ANALYSIS

In message grouping process the security between current message and previous messages are tested.As explained earlier, the proposed proactive forensic effort normally expects a long exchange of messages between a possible adversary and a potential victim (user of our system) covering a large time period. A machine learning based risk assessment technique was developed in [2] for user actions like adding
a friend, and sending messages in a social media setting.
These messages are divided into four levels they are (i) criticality level, measuring how important the action is,
(ii) Likelihood level, measuring how expected the action is under similar situations, (iii) impact, which refers to the impact of security risks, and (iv) information requestor reputation, which roughly corresponds to trust of the possible adversary in our framework. Based on these factors, [2] classifies the information item into k risk levels. The first two factors, namely criticality of action, which is an estimate of how important it is for the user to provide the information contained in the message, and likelihood level, which requires the system to have a compendium of all situations and likely user responses, are hard metrics to compute.

.

Consider a threshold value as the initial value. Whenever our system crosses this threshold it give a warning message. So we could know a change is needed. For a close relative or friend, asking for e-mail address, phone no. or home address may be normal even when they are considered to be sensitive information and will reduce reliability or increase risk with respect to the requester. So appearance of each sensitive term will be notified to the receiver with option to ignore it so that trust value remains unaltered. However, warning message will advise the receiver to stop further communication with the requester declaring him/her as a possible adversary.

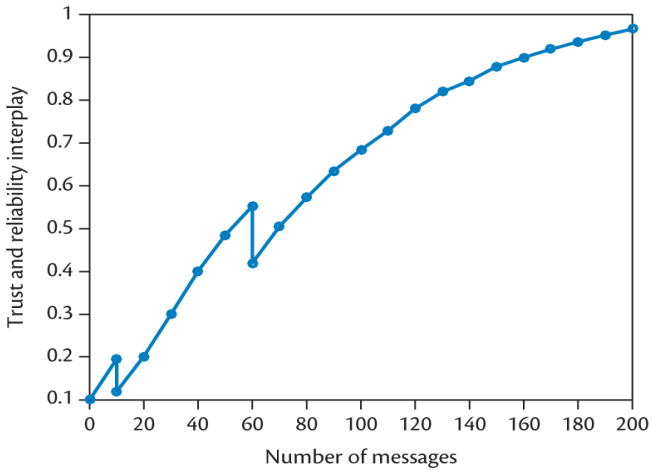

Fig. 2. Reliability and Trust Interplay [7].

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

From the above graph (fig 2), it clearly proves that the trust and risk is directly proportional to the number of messages that are transferred between the users .So that if the number of messages are increased the trust is automatically increased. This is the basic idea encountered from the above graph.

VII.PRIVACY ISSUE

The proposed system is for reducing the cyber crimes. Great idea to prevent crimes in online media. There are two possibilities to attack our messages.

As per many studies regarding this, which covers different aspects of security and privacy in social network have provided details of the privacy issues that may be encountered in communication over social media [2], [3]. Out of the different attacks discussed in the two surveys, the relevant issues for the present system are: 'Attack from other users' and 'crawling attack'. As a matter of fact, proactive forensic system has been proposed only to avoid such attacks and to generate an early warning system against the possibility of such attack. On the other hand, for accepting a friendship request profile matching and social context building methods described so far constitute crawling attack. Profiling of a requester can be done only by exploiting publicly available APIs of different social media. So crawling through the web sources and consolidating data available from those sources, it may be possible to build a good amount of information about a person requesting for friendship. Most of those who are sending friendship requests are not potential adversaries but the profile matching and social context building will be done for all requesters. In other words, from Privacy point of view, the proposed system creates a situation where in order to prevent Attack from other users the system will create a Crawling attack.

Researchers like [4] have developed questionnaires and predictive models to assess an individual's degree of privacy concern, level of privacy awareness and the proclivity toward self-disclosure. One possible solution to mitigate the Crawling attack issue can be to accept a trusted third party for both requester and receiver and such an agency can take the responsibility of profile matching and social context building. However, more detailed study on this Crawling attack issue is yet to be done.

(1)Web crawling attack:

Web crawling (spidering) is not mandatory to hack anything, but to collect details on the target/victim. Before a website attack or penetration test, we need to spider the site. We could manually crawl the site by simply navigating to each page and saving it, but fortunately, there is a lot of tools that can save time and automate this process.

VIII. CONCLUSION

We are developing PRADAN, an initial functional prototype that incorporates the ideas described above. PRADAN is being implemented as an application above the AWESOME system [1], where tasks like building the user's profile, social context and message aggregation are handled by the application layer, and the heterogeneous data store is implemented through AWESOME. This can hold both real time and dynamic data. The static data stored in data dictionary as meta data. PRADAN uses application layer of AWESOME system and scripting language for determining the data flow between the messages. For instance, the trust computation needs to use the history of the user with a set of "friends", the application uses an API call to retrieve a friend's messages, reverse sorted by time. The system can be configured to run in an in-memory mode as well as a full-scale data management system over distributed cluster of machines. These modes of operation are important for the PRADAN application– which can run either in a standalone, single user mode, or as a multitenant service mode.

(1)Awesome system

Polystore, i.e., data management systems that used as a multistore for varying data models, are gaining popularity. Here, developing a polystore-based system called AWESOME to support online data analysis. The

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

AWESOME polystore can support relational, partially structured, diagrammatic and text data. ADIL, the data ingestion language of AWESOME allows a user to flexibly specify the placement of original and derived data.

REFERENCES

1. S. Dasgupta, K. Coakley, and A. Gupta, "Analytics-driven data ingestion and derivation in the AWESOME polystore," in Proc. of the IEEE Int. Conf. on Big Data, pp. 2555–2564, IEEE, Dec. 2016.
2. A. Ali-Eldin, J. van den Berg, and H. A. Ali, "A risk evaluation approach for authorization decisions in social pervasive applications, "Computers & Electrical Engineering, vol. 55, pp. 59–72, 2016.
3. E. Novak and Q. Li, "Security and privacy in online social networks - a survey," Tech. Rep. WM-CS-2012-02, Department of Computer Science, The College of William and Mary, 2012.
4. I. Kayes and A. Iamnitchi, "A survey on privacy and security in online social networks," arXiv preprint arXiv:1504.03342, 2015.
5. H. Krasnova and N. F. Veltri, "Privacy calculus on social networking sites: Explorative evidence from Germany and USA," in Proc. Of 43rd Hawaii Int. Conf. on System sciences (HICSS), pp. 1–10, IEEE, 2010.
6. Amarnath Gupta, Subhasis Dasgupta, Aditya Bagchi" PROFORMA: Proactive Forensics with Message Analytics".IEEE 2017.
7. K. Schouten and F. Frasincar, "Survey on aspect-level sentiment analysis," IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 3, pp. 813–830, 2016.
8. S. Y. Bhat and M. Abulaish, "Using communities against deception in online social networks," Computer Fraud & Security, vol. 2014, no. 2, pp. 8–16, 2014.
9. T. ˇSvec and J. Samek, "Trust evaluation on facebook using multiple contexts," in 21st Conference on User Modelling, Adaptation, and Personalization, pp. 1–10, 2013.
10. M. Fire, R. Goldschmidt, and Y. Elovici, "Online social networks: threats and solutions," IEEE Communications Surveys & Tutorials, vol.16, no. 4, pp. 2019–2036, 2014.
11. S. Raghavan, "Digital forensic research current state of the art", *CSI Transactions on ICT*, vol. 1, no. 1, pp. 91-114, 2013.
12. Lawrence, R. et al., "Social media analytics", OR/MS Today, pp. 26-30, February 2010