The Infiltration Game: Artificial Immune System for the Exploitation of Crime Relevant Information in Social Networks

Michael Spranger, Sven Becker, Florian Heinke, Hanna Siewerts and Dirk Labudde

University of Applied Sciences Mittweida
Forensic Science Investigation Lab (FoSIL), Germany
Email: spranger@hs-mittweida.de

***Abstract*—Efficient and sensitive monitoring of social networks has become increasingly important for criminal investigations and crime prevention during the last years. However, with the growing amount of data and increasing complexity to be considered, monitoring tasks are difficult to handle, up to a point where manual observation is infeasible in most cases and, thus, automated systems are very much needed. In this paper, a system of adaptive agents is proposed, which aims at monitoring publicly accessible parts of a given social network for malign actions, such as propaganda, hate speeches or other malicious posts and comments made by groups or individuals. Subsequently, some of these agents try to gain access to crime relevant information exchanged in closed environments said individuals or groups are potentially part of. The presented monitoring and investigation processes are implemented by mimicking central aspects of the human immune system. The monitoring processes are realized by network-traversing informant units similar to pathogen-sensing macrophages, which initialize the human immune response. The subsequent investigation process is commenced by gathering information automatically about the targeted individual or group. Furthermore, based on the gathered information one can identify closed and usually inaccessible environments in the social network (e.g., private groups). Using so-called endoceptor units—automatically generated social bots imitating environment-typical appearance and communication—closed environments are accessed through individuals susceptible to the bot's strategy. Once being part of the closed network, an endoceptor aims to intercept and report back crime relevant communications and information to the investigators.**

Keywords–social network; prevention; predictive policing; text mining; autonomous agents; artificial immune system

I. Introduction

Over the last ten years, social networks have grown to become an essential part in our communication. Despite their success and advances made, social networks have also produced central hubs for criminal energy by providing the possibility/means to network as well as interchange and communicate ideas quickly, while remaining private in an environment difficult to control and monitor by investigators. Thus, for extreme political groups, criminal gangs and terrorist organizations, social networks are ideal platforms for planning and appointing the execution of criminal actions. Therefore, targeted monitoring of social networks can help to improve strategic security planning and prevention processes by authorities, as well as, help to increase the users' sense of security. Homeland security and secret services are aware of the importance of crucial information hidden in these networks and therefore more and more focus on social network surveillance. Looking at the increasing number of users worldwide – currently every third person uses social networks – there is a huge number of potential profiles and communication traffic to be monitored. This shows the need for an automated and sensitive solution that is able to cope with the vast amount of data and computational complexity yielding from it. Yet, besides these theoretical hurdles, the implementation of such monitoring procedures is further impaired due to the simple fact that in most cases crime-specific information is not discussed in the publicly accessible environment of social networks. Such relevant exchanges and discourses are rather made in closed inaccessible groups.

With respect to the legal limitations, in this work a multi-agent-based system is proposed, which aims at monitoring social networks and targeting potential offenders and (mostly) inaccessible subnetworks of their associates. The presented strategy utilizes a cascaded system of multi-role agent units, whose implementation and tasks are inspired by the human immune system. Similar to the cells involved in the human immune response (e.g., macrophages, killer cells and T-helper cells), the framework employs agents capable of sensing malicious actions, such as malign or offensive posts, analysing the profiles of the (potential) offenders, identifying the (mostly private and inaccessible) subnetworks of associates, entering these subnetworks as social bots that are automatically adapted to the appearances, ductus, and characteristic styles of these associates, and relaying explosive information exchanged in these subnetworks to the investigators.

In Section 2, we discuss related work presenting implementations of social network monitoring processes, as well as *in silico* realizations of the human response system and their applicability in this respect. Details about the proposed framework are presented in Section 3.

II. Related Work

Research conducted towards monitoring social media in the context of forensics has given rise to a large body of literature. In this section, a brief overview on works addressing this issue is given. Further, in order to put the proposed framework into context, some of the landmark papers discussing computational implementations of the human immune response system for data analyses are summarized. For a more in depth view, please refer to the notable review paper from Benkhelifa et al. [1] in which the authors outline some of the recent high-impact advancements and also propose a digital forensics incidents prediction framework tailored towards being utilized in cloud environments.

ISBN: 978-1-61208-566-1

Complementing the idea of predicting future criminal incidents, in one of the most recent papers Soundarya et al. [2] elucidated the utilization of so-called genetic weighted k-means cluster analyses combined with negative selection schemes in an effort to make predictions based on social media profiling. Although the predictive power looks promising, implementing the presented prediction scheme successfully in real life applications is questionable, as underlying features used in their method are derived from information difficult to obtain in practice (e. g. the number of logins/sessions per day and the time duration of individual sessions). Another interesting idea was presented by Huber et al. [3]. Using their so-called Social Snapshot method, data can be efficiently acquired from social network websites that are of special interest for law enforcement agencies. This method is based on custom-made add-ons for crawling social networks and underlying web components. The Social Snapshots method further allows the extraction of profile information such as user data, private messages and images, and associated meta-data like internal timestamps and unique identifier. A prototype for Facebook was developed by the group and evaluated based on a volunteer survey.

Computational modelling of human immune response mechanisms and applying such models to various problems in data mining has been an ongoing research process for over two decades. In 2000, Timmis et al. [4] published an immune response-mimicking framework specifically designed for data analysis. Furthermore, the group presented a minimalistic formulation of an artificial immune system and elucidated its action/response mechanisms. As another example for application, Wu & Banzhaf [5] and West et al. [6] independently developed artificial immune systems for the detection of transactional frauds in automated bank machines. Both works employ binary matching rules paired with fuzzy logic in order to detect transaction anomalies. Chen et al. [7] discussed a classification technique, which considers some general aspects of immune response mechanisms. In combination with a population-based incremental adaptive learning scheme and collaborative filtering, their method aims at detecting invasive actions targeting computer networks. Finally, the research group of Karimi-Majd et al. [8] developed a novel hybrid artificial immune network for detecting sub-structures, so-called communities, in complex networks using statistical measures of structural network properties.

III. The Proposed Framework

The proposed multi-agent monitoring system, as illustrated in Figure 1, is inspired by the cellular mechanisms implemented by the human immune system. Although there are multiple immune response mechanisms and cell types with roles highly adapted to these individual mechanisms, the general concept of immune response can be summarized as follows: mobile recognition cells freely traversing the human body (e.g., macrophages) are able to recognize and absorb pathogens, such as viruses or infectious bacteria, and to report back pathogen-specific information upon which an adaptive immune response is triggered. Subsequently, mobile cells are synthesized that use the reported cellular information to specifically target and destroy invaded pathogens by means of a pathogen-specific molecular lock-and-key binding mechanism. Multiple aspects are implemented in the proposed framework that aim at mimicking this response concept in the context of recognizing hostile and malicious activities in the publicly accessible parts of the environment under investigation (e.g., selected profiles in (sub-) social networks, blogs or internet forums), and targeting groups of malign entities usually inaccessible to the public (e.g. closed groups in social networks).

The agent units implemented by the proposed framework are presented in more detail in the following subsections.

A. Informants

Similar to the biological role of pathogen-sensing macrophages, the task of informant units is to recognize potentially dangerous profiles within the social network. There are two basic types of informants, observers I^o and classifiers I^c. The objective of the observers is to read along public discussions, so called feeds. If a post or comment with potentially dangerous content is detected, the corresponding profile is reported as a candidate profile p^c to a central control unit, the agency Ψ (Implementation details about the agency are given later). The algorithmic layout of informants has to be manifold due to profile appearance variability of potential offenders. For example, to recognize the profile of a right-wing individual or organization, an analysis of the images on the profile or the members or friend lists can be helpful. In this respect, a binary classifier is trained for each feature, which is suitable to identify a particular type of potential offender. The training takes place in the form of semi-supervised learning. Candidate profiles whose membership to a certain potential offender type are considered to be secured serve as seeds. In order to minimize the likelihood of a misclassification, all classifiers of a certain type of potential offender form an ensemble which reports a profile as a candidate p^c to the agency by majority vote.

B. Analysts

The analysts A are specifically tailored towards certain groups of potential offenders. Their task is to gather information about candidate profiles. Such information could be, for example, the mood in the network determined by sentiment analyses, the development of its structural properties, or planned activities. As a special task, the analysts have to adapt to the language specificities of the respective group. In this way, on the one hand, the ability of the informants to discriminate profiles can be further improved. On the other hand, such specificities form the basis for the synthesis of adapted endoceptors. In the case of a group profile or the profile of an organization, the opinion-makers are detected by analysing the communication and subsequently reported to the agency. The detection of opinion-makers or multipliers can be conducted by considering the Page Rank algorithm [9] [10] or Hyperlink-Induced Topic Search (HITS) algorithm [11] developed to detect hubs and authorities on websites. Further informative features, such as hashtags, '@' references or information deduced from discourse analysis, need to be considered and are readily available in social network environments.

C. Endoceptors

The most subtle type of agents in the framework are endoceptors E. They are used when certain circumstances in the analysis justify the assumption that further explosive information is distributed in closed groups. Endoceptors are a kind of chat bot that adapts to the language behaviour of a potential offender group and tries to contact the leading members in order to become a member of the group. Once included, endoceptors remain passive and relay distributed information to the agency. In this way, they imitate the behaviour of a confidential informant.

D. Agency

In line with the human lymphatic system, a technical agency Ψ forms both the infrastructural basis of this framework and its bilateral interface to investigators. Such agencies include, in addition to the set of so-called candidate profiles P^c, a set of activation functions $\alpha_1, ..., \alpha_n$ as triggers for synthesizing different types of agents. A candidate profile in this respect can be the public profile of a group or organization but also the non-public one of an individual. A ranking $r(P^c)$ is assigned to each candidate, which determines whether and with which priority it is observed and which concrete actions, i. e., which concrete agent synthesis are triggered by the agency. Equation (1) shows that the ranking is mainly driven by two parts. The first part takes the frequency of notifications by observers into account. The second refers to the mean voting of all classifiers, whereat the individual influence can be adjusted by a weight w_i with $\sum w_i = 1$. For example, the classification of the profile of an organization as right-wing extremist might depend more on the estimation of the image classifier than the one who makes the same assessment by means of the list of friends. The influence of each part of the ranking function can be controlled by parameter λ with $\lambda = [0, 1]$, which needs to be estimated empirically.

$$r(p_i^c) = \lambda \frac{count(I^o, p_i^c)}{\sum_{P^c} count(I^o, p_j^c)} + (1-\lambda) \sum_{j=1}^{|I^{c_j}|} \frac{w_j I^{c_j}(p_i^c)}{|I^{c_i}|} \quad (1)$$

The synthesis of an instance of a specific agent type is triggered by an activation function α. Equation (2) shows such a function for the activation of the analysts. The function decides on the basis of the rank of a candidate whether or not a threshold is exceeded and the synthesis is triggered. The threshold value can be regarded as a kind of intervention threshold. Thus, it represents a parameter for the implementation of safeguards against arbitrary surveillance.

$$\alpha_A(p_i^c) = \begin{cases} 1, & \text{if } r(p_i^c) > \epsilon \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

E. Workflow

An illustration of the recognition and response mechanism is given in Figure 2. Individual monitoring steps are labelled A through E. The 'informant synthesis'—the *ad hoc* generation of informant units—is based on *a priori* expert knowledge provided by the investigators. The number of informants of a certain type of concept or topic to be monitored (illustrated by circle, square and triangle symbols) depends on the structural properties of the network and the amount of information exchanged by the users. Again, informants can only access publicly available information. Once public malicious activity is detected by an informant (see step A in Figure 2), entity-specific information is reported back to the agency (step B in Figure 2). In the illustration in A, an informant of type 'triangle' detects malicious activity in a subnetwork of users. Similarly, in B an informant of type 'cycle' reports an incident back to the agency. Subsequently, analyst unit synthesis is triggered according to an activation function (see Section III-D for formal details). The set of activation functions and their importance weighting relative to the number of detected incidents over time can be interpreted, in a biological sense, as the number of specific receptors for the different types of informants. The more 'alerted' informants are reporting back to the agency and are 'bound' to the agency, the more specific informants and receptors are subsequently synthesized. The ratio of synthesized receptors and informants bound to them illustrates the weight of the individual activation function. The role of the analyst unit is to use information retrieved from the publicly active malign entity to locate the network of associated malign entities and possible entry points to the subnetwork (step C in Figure 2). In a next step, this information is used to synthesize an endoceptor unit (step D in Figure 2). By mimicking the behaviour and appearance of target entities, the endoceptor aims at penetrating the closed environment, thus becoming a part of the network. Information exchanged by malign entities is now intercepted and communicated back to the agency module (step E in Figure 2).

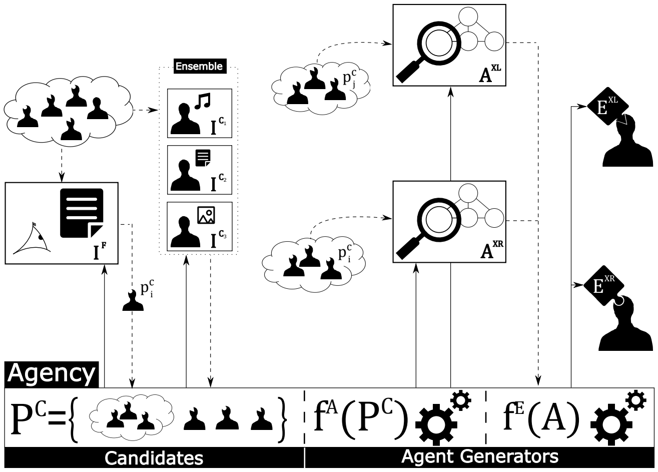

Figure 1. Schematic structure of the proposed framework. The informants I supply candidate profiles p^c to the agency where they are registered and evaluated by means of a ranking function. If a critical value is exceeded, analysts A are synthesized by a function f^A and sent out to collect information about these profiles. This information is the basis for endoceptors synthesized later by the function f^E attempt to infiltrate the protected areas of potentially dangerous profiles by contacting them in the manner of a chatbot by sending friendship requests. Once accepted, they remain passive and forward information to the investigating authorities.

IV. Conclusion and Future Work

In this work, we outlined a framework that allows investigators from law enforcement agencies and intelligence services to automatically monitor social networks and collect information about potentially dangerous activities. The framework is based on autonomous agents and inspired by the processes in the human immune system. However, no attention was paid to an exact replica of the biological processes. For the proposed framework, it is more important that the system is able to adapt itself to various disturbances. Therefore, it has to be able to adjust to the form of profiles of potential offenders, infiltrate

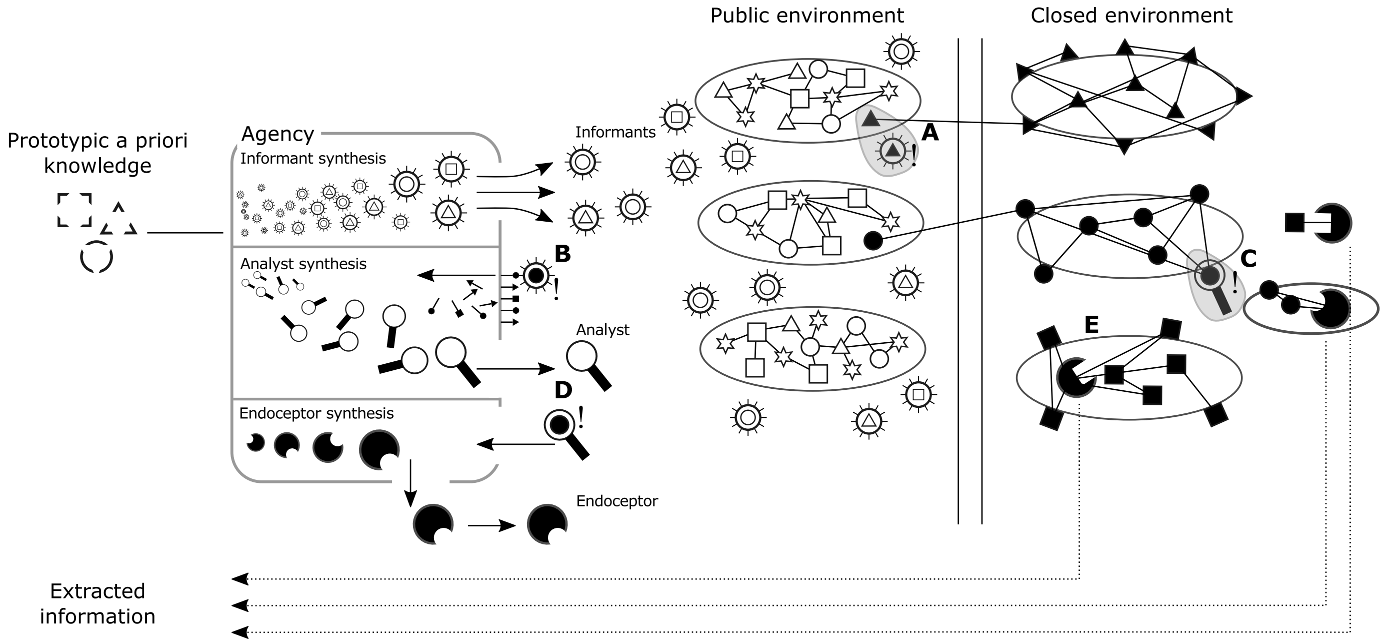

Figure 2. Schematic of the proposed workflow. Please refer to Section III-E for implementation details.

them and forward important information to the investigators. In this way, risks can be detected early and, at best, damage can be prevented.

Current and future work is mainly concerned with the design of the analysts, whereat the focus is on the detection of opinion-makers and the analysis of language style and writing behaviour in the group as a prerequisite for the synthesis of chat bots (Endoceptors) that are recognized by that group as their peers. As a by-product, we can learn how chat bots can be detected in networks. In parallel, independent sets of social features have to be found, which are suitable to classify candidates with the necessary accuracy to address privacy concerns.

REFERENCES

[1] E. Benkhelifa, E. Rowe, R. Kinmond, O. A. Adedugbe, and T. Welsh, “Exploiting social networks for the prediction of social and civil unrest: A cloud based framework,” in Future Internet of Things and Cloud (FiCloud), 2014 International Conference on. IEEE, 2014, pp. 565–572.

[2] V. Soundarya, U. Kanimozhi, and D. Manjula, “Recommendation system for criminal behavioral analysis on social network using genetic weighted k-means clustering.” JCP, vol. 12, no. 3, 2017, pp. 212–220.

[3] M. Huber, M. Mulazzani, M. Leithner, S. Schrittwieser, G. Wondracek, and E. Weippl, “Social snapshots: Digital forensics for online social networks,” in Proceedings of the 27th annual computer security applications conference. ACM, 2011, pp. 113–122.

[4] J. Timmis, M. Neal, and J. Hunt, “An artificial immune system for data analysis,” Biosystems, vol. 55, no. 1, 2000, pp. 143–150.

[5] S. X. Wu and W. Banzhaf, “Combatting financial fraud: a coevolutionary anomaly detection approach,” in Proceedings of the 10th annual conference on Genetic and evolutionary computation. ACM, 2008, pp. 1673–1680.

[6] J. West, M. Bhattacharya, and R. Islam, “Intelligent financial fraud detection practices: An investigation,” in International Conference on Security and Privacy in Communication Systems. Springer, 2014, pp. 186–203.

[7] M.-H. Chen, P.-C. Chang, and J.-L. Wu, “A population-based incremental learning approach with artificial immune system for network intrusion detection,” Engineering Applications of Artificial Intelligence, vol. 51, 2016, pp. 171–181.

[8] A.-M. Karimi-Majd, M. Fathian, and B. Amiri, “A hybrid artificial immune network for detecting communities in complex networks,” Computing, vol. 97, no. 5, 2015, pp. 483–507.

[9] S. Brin and L. Page, “The anatomy of a large-scale hypertextual web search engine,” Comput. Netw. ISDN Syst., vol. 30, no. 1-7, Apr. 1998, pp. 107–117. [Online]. Available: http://dx.doi.org/10.1016/S0169-7552(98)00110-X

[10] ——, “The anatomy of a large-scale hypertextual web search engine,” in Proceedings of the Seventh International Conference on World Wide Web 7, ser. WWW7. Amsterdam, The Netherlands, The Netherlands: Elsevier Science Publishers B. V., 1998, pp. 107–117. [Online]. Available: http://dl.acm.org/citation.cfm?id=297805.297827

[11] J. M. Kleinberg, “Authoritative sources in a hyperlinked environment,” J. ACM, vol. 46, no. 5, Sep. 1999, pp. 604–632. [Online]. Available: http://doi.acm.org/10.1145/324133.324140

[12] M. Spranger, S. Schildbach, F. Heinke, S. Grunert, and D. Labudde, “Semantic tools for forensics: A highly adaptable framework,” in Proc. 2nd. International Conference on Advances in Information Management and Mining (IMMM), IARIA. ThinkMind Library, 2012, pp. 27–31.

[13] C. Weinstein, W. Campbell, B. Delaney, and G. O’Leary, “Modeling and detection techniques for counter-terror social network analysis and intent recognition,” in 2009 IEEE Aerospace conference, 2009, pp. 1–16.