**OSN User Privacy Protection Based on Text Analysis and Encryption**

**Literature Research**

The social network is an integral part of life for most people. People make friends on Facebook and Twitter, and find jobs on LinkedIn. However, the existence of social networks inevitably forces people to face privacy issues [1]. For workplace social networking platforms, the privacy concerns are not limited to the leakage of user information. If an employee posts sensitive information, such as salary, job-hopping, or complaining about the company, he may face pressure from the company he works for. Therefore, how to selectively block all colleagues to post sensitive information and prevent screenshots from being spread by others is a problem that needs to be studied.

The emergence of sensitive information is more caused by information inequality between job seekers and companies. Job seekers cannot obtain information from companies, so they hope to get them from employees. There has been a lot of effort on salary transparency, which is job seekers focus on the most. In 2016, LinkedIn launched Salary Insights, a system that gathers salary information from members to provide salary insights to job seekers. In 2017, Kenthapadi [2] introduced the LinkedIn compensation product, which collects reams of data to help people calculate their earning potential. These are very meaningful and innovative products, but at present, the coverage of these products is low as well as many regions and positions are not counted. Also, many job seekers are looking for information other than salaries, such as working environment and intensity. In addition, a 2018 Jesus [3] study of algorithms that predict how likely employees are to leave the workplace based on LinkedIn has made employees’ situation even more precarious.

That's not the original vision of social platforms. As a social network, it should meet the needs of users for free and open communication. Therefore, people have been studying the privacy protection of social platforms. In 2017, The TABOO application developed by Neerbeky [4] detected sensitive information in sentences using the RNN model. In June of the same year, Cappellari [5] compared a variety of machine learning algorithms and found that SVM was the best sensitive information detection. However, the accuracy of the above two methods for text analysis is limited only through a single model or algorithm. In 2018, Islam [6] showed that supervised learning using Naive Bayes for examination achieved better results than SVM with the addition of appropriate data preprocessing. In addition to machine learning, Canfora [7] uses NLP to prevent privacy leaks on social networks through a plugin. In the plugin, a pop-up window shows what privacy a tweet might reveal.

The above method of detecting sensitive information is to prompt users and inform them that these contents may leak private information. However, some users want to secretly share sensitive information about the company, such as the intensity of the department's work, the difficulty of the promotion, and the range and frequency of salary increases. If the information needs to be encrypted for release, it is likely to be negative for companies. But this negative information can help job seekers find a more suitable job. To post this information on the social platform, encryption could be a good choice. It can be seen that there has always been a demand for information protection on social platforms.

There are various encryption methods. At present, the mainstream encryption methods include BASE-64, AES, DES and RSA encryption algorithm. Malik [8] developed a BASE-64 and AES encryption system in 2020, mainly for medical image encryption. But encryption for medicine is not the same as encryption for social platforms, and encrypted content on online social networks should be easily accessible to other users. In 2013, Tierney's team tried to do just that [9]. They developed a system that uses BASE-64 to encrypt photos, allowing users to encrypt photos and upload them to the OSN.

References

1. Gross R, Acquisti A. Information revelation and privacy in online social networks. Proceedings of the 2005 ACM workshop on Privacy in the electronic society; Alexandria, VA, USA: Association for Computing Machinery; 2005. p. 71–80.

2. Kenthapadi K, Ambler S, Zhang L, Agarwal D. Bringing Salary Transparency to the World: Computing Robust Compensation Insights via LinkedIn Salary. Proceedings of the 2017 ACM on Conference on Information and Knowledge Management; Singapore, Singapore: Association for Computing Machinery; 2017. p. 447–55.

3. Jesus ACCd, Júnior MEGD, Brandão WC. Exploiting linkedin to predict employee resignation likelihood. Proceedings of the 33rd Annual ACM Symposium on Applied Computing; Pau, France: Association for Computing Machinery; 2018. p. 1764–71.

4. Neerbeky J, Assentz I, Dolog P, editors. TABOO: Detecting Unstructured Sensitive Information Using Recursive Neural Networks. 2017 IEEE 33rd International Conference on Data Engineering (ICDE); 2017 19-22 April 2017.

5. Cappellari P, Chun SA, Perelman M. A Tool for Automatic Assessment and Awareness of Privacy Disclosure. Proceedings of the 18th Annual International Conference on Digital Government Research; Staten Island, NY, USA: Association for Computing Machinery; 2017. p. 586–7.

6. Islam AC, Walsh J, Greenstadt R. Privacy Detective: Detecting Private Information and Collective Privacy Behavior in a Large Social Network. Proceedings of the 13th Workshop on Privacy in the Electronic Society; Scottsdale, Arizona, USA: Association for Computing Machinery; 2014. p. 35–46.

7. Canfora G, Sorbo AD, Emanuele E, Forootani S, Visaggio CA. A Nlp-based Solution to Prevent from Privacy Leaks in Social Network Posts. Proceedings of the 13th International Conference on Availability, Reliability and Security; Hamburg, Germany: Association for Computing Machinery; 2018. p. Article 36.

8. Malik MW, Husna D, Purnama IKE, Nurtanio I, Hidayati AN, Ratna AAP. Development of Medical Image Encryption System Using Byte-Level Base-64 Encoding and AES Encryption Method. 2020 the 6th International Conference on Communication and Information Processing; Tokyo, Japan: Association for Computing Machinery; 2020. p. 153–8.

9. Tierney M, Spiro I, Bregler C, Subramanian L. Cryptagram: photo privacy for online social media. Proceedings of the first ACM conference on Online social networks; Boston, Massachusetts, USA: Association for Computing Machinery; 2013. p. 75–88.