Chapter 3
Leading Attacks in Privacy Protection Domain

In this chapter, we discuss seven leading attacks in privacy domain built upon the major privacy concerns in general cases. There are background knowledge attack, collusion attack, linkage attack, structural attack, forgery attack, eavesdropping attack, and Sybil attack. There are also some other forms of attacks such as tracking attacks [1] and inference attacks [2, 3], but these attacks fall in the range of the seven illustrated attacks. Beyond the traditional privacy protection scenarios, these attacks still reveal the privacy and may result in further privacy leakage when personalized privacy protection solutions are deployed, which will be detailedly discussed in the following chapters.

3.1 Major Privacy Concerns

With the widespread of mobile devices, massive data is being generated at every moment. The privacy protection under big data scenario has new features and development [4]. Usually, the released data contains sensitive identity information, location information, other profile information, etc [5]. Although a single piece of data usually does not cause privacy leakage, multiple pieces of data can be regarded as a combination of quasi-identifiers and may lead to intractable privacy loss. In Table 3.1, we further illustrate the correlation between privacy issues and attacks. We summarize all the privacy issues and attacks in mobile social networks as follows.

3.1.1 Identity Privacy

Protecting identity privacy [6] is the most fundamental target in privacy protection in social networks. If identity privacy is breached, most of the following sensitive

Y. Qu et al., *Personalized Privacy Protection in Big Data*, Data Analytics,
https://doi.org/10.1007/978-981-16-3750-6_3

information will leak accordingly. This can be achieved in several ways, for example, anonymity, pseudonym generation [7], and so on. The final target is to prevent adversaries from re-identifying specific users, which is essential especially in social network data sharing. In [8], Wang et al. investigated crowd-sourced data publishing in social networks using differential privacy. The investigated data is real-time as well as spatiotemporal. This work takes the continuous publication of statistics and demonstrates the "RescueDP", which is an online aggregate supervisory control framework with w-event privacy preservation. The core elements include adaptive sampling, dynamic clustering, adaptive budget allocation, filters, and perturbation. In addition, the authors developed a reinforced RescueDP based on neural networks to calculate the statistics and thereby improve data utility. Xing et al. [9] proposed a k-means-based community establishment scheme in social networks with privacy protection. This scheme maintains the privacy of both sensitive information of individual and statistics features of the community. In each iteration of k-means algorithm, the scheme processes two privacy-preserving operations. The first one is that users try to find nearest clusters without knowing the cluster centers. The second one is that the cluster centers are calculated without information leakage and users inside a specific cluster cannot infer the identity of each other.

3.1.2 Anonymization Versus De-Anonymization

Anonymization is another big issue which is closely related to identity privacy. Usually, anonymization is used for publishing the big social data sets for research or commercial purposeS [10]. The most economical way of data release is anonymization. Modern anonymization methods are far beyond simply eliminating the identifiers, for example, adding nodes or modifying edges to introduce random noise [11]. However, fast development of de-anonymization techniques [12] puts anonymization under great threats.

3.1.3 Location Privacy

Beyond identity privacy, location privacy [13] has attracted plentiful attention from researchers. As social network users spend more and more time and energy on mobile devices, mobile social applications may cause location privacy leakage by accessing the users' GPS data [14]. Adversaries can easily obtained either from the released data or from crawling it from the system background [15]. For example, a specific user may publish the location information when enjoying a fancy dinner at a restaurant or adversaries can hack the application directly. Therefore, improper release of location sensitive data can even cause physical loss.

3.1.4 Content Oriented Privacy (CO Privacy)

Mobile social networks can be regarded as a specific type of content-oriented networks while privacy protection in content-oriented networks has always been considered [16]. As discussed in [17], content-oriented privacy consists of three properties, which are immutability, transparency, and accountability.

3.1.5 Interest Privacy

In social networks, users are usually categorized by interest communities. Adversaries can launch collusion attack, background knowledge attack, or inference attack to gain interest privacy information by breaching the privacy of anyone in the community [18]. Moreover, built upon the location privacy, adversaries can obtain sensitive information, for example, favourite restaurant, preferred cinema, best-loved bookshop, and so on [19]. Based on the interest-based sensitive information, adversaries can spam users or commit other malicious attacks with potential profitable targets.

3.1.6 Backward Privacy and Forward Privacy (B&G Privacy)

Backward privacy denotes that an adversary cannot track the previous actions of users when the adversary has the sensitive information stored in it, while forward privacy is that an adversary cannot predict the previous actions of users when the adversary has the sensitive information stored in it [20]. These two features are quite important as privacy protection in social networks should always be long-term protection [15]. Thus, the sensitive information should be context-aware and the privacy protection must take the forward and backward status into consideration.

3.2 Leading Privacy Breaching Attacks

3.2.1 Background Knowledge Attack

Background knowledge attack is one of the most popular attacks under privacy scenarios. The rationale behind its proliferation is that it can be combined with other types of attacks. Background knowledge of a specific entry is easy to obtain in mobile social networks [21]. Moreover, the background knowledge of adversaries is hard to model, measure, and predict, which makes it more difficult to be defeated.

Table 3.1 Privacy issues and corresponding attacks

	Identity privacy	Anonymization	Location privacy	CO privacy	Interest privacy	B&F privacy
Background knowledge attack	√	√	√	√	√	√
Collusion attack	√	√	√	√	√	O
Linkage attack	√	√	√	√	√	√
Structural attack	√	√	O	×	×	O
Forgery attack	O	O	√	×	×	O
Eavesdropping attack	√	√	√	√	√	O
Sybil attack	√	√	O	×	×	O

√ denotes fully supported; × denotes not supported; O denotes partially supported

3.2.2 Collusion Attack

Collusion attack is another wide-spread attack method. Collusion attack is especially mortal in mobile social network circumstances. The reason is that a specific user can have multiple contacts in social networks and therefore there might be multiple adversaries hiding in the contact list. As different adversary holds different background knowledge of this user, they can share the information with each other to launch a collusion attack [22]. In addition, collusion attack can also be combined with other forms of attacks.

3.2.3 Linkage Attack

Linkage attack is experiencing rapid expansion with rapidly increasing data volume and data sources. For example, adversaries can make an attack based on multiple social networks. Linkage attack has a good attack performance as adversaries can collect different category of data of the same user from multiple data sources [23]. Furthermore, machine learning-based methods provide linkage attack better tools which help adversaries bypass the protection. Song et al. [24] developed a new type of inference attack. This type of attack targets on the browsing history of Twitter users leveraging twitter metadata and public click analytics. This attack only needs Twitter profile information and URL shortening services, which are public and easy-to-access information. This can further reduce the attack overhead and upgrade accuracy by taking time-varying models of users into consideration.

3.2.4 Structural Attack

Adversaries are proceeding to structural attacks because social networks are usually modelled as a graph based on graph theory. In one hand, graph theory helps to better understand and establishes social networks structure. On the other hand, adversaries can take advantage of the structural information to mount an attack. The most outstanding merit of structural attack is that adversaries can re-identify a specific user even without background knowledge. The structural attack is also widely-deployed in de-anonymization. In [25], Chen et al. proposed two types of practical attacks to steal sensitive information from graph-based clustering methods. Targeted noise injection and small community are devised to attack three popular graph clustering models, including community discovery, node2vec, and singular value decomposition (SVD). Based on this, the authors found that adversaries with limited open-source background knowledge can launch successful attacks. In term of simple defenses, it can decrease the success ratio to 25% by the cost of only 0.2% clusters over-noisy.

3.2.5 Forgery Attack

In a forgery attack, misleading messages are generated with fake information, so that adversaries can initiate some other plotting attacks such as the location-tracking attack. There are five phases in a forgery attack, in which we use vehicular social networks as an example. Firstly, the victim node and the adversary node establish a link with location information. Secondly, the adversary node creates malicious payload to the victim node. Thirdly, the victim node sends a request to a social spot s_1 for cookies. Fourthly, s_1 gives the email address of victim node to the adversary node. Lastly, the social spots reset the certificates. In this way, an outside forgery attack is performed and the privacy of the victim nodes will be compromised [26].

3.2.6 Eavesdropping Attack

In the case of eavesdropping attack, it is quite intuitive that adversaries eavesdrop the information communication and transmission process by means of modern hacking technologies, including internet, electromagnetic wave, and so on. This type of attack is launched by unauthorized real-time interception of a private communication [27]. Therefore, it is vital to secure communication to prevent privacy leakage.

3.2.7 Sybil Attack

The Sybil attack is normally launched under the scenario of a reputation-involved system. During the attack process, an adversary generates a large number of pseudo names and further gains the maximum influence [28]. Based on the influence, the adversary can mislead the other users in the system or even fool the central authority. Privacy leakage happens during the attack. Whether the attack can be successfully launched is decided by the cost to fake identities and the trust mechanism between central authority and the identities. In [29], Liu et al. did a study on extended Sybil defences. The authors found that current sybil attack models in social networks are static, which is not practical. This work takes temporal dynamics into consideration and involves three new features. Firstly, the new model considers the capabilities of adversaries to modify Sybil-controlled parts of a structural social graph. Secondly, another new feature is the capabilities to modify the connections which Sybil identities of him/her maintain to honest users. Thirdly, the proposed model benefits from the regular dynamics of connections structure and thereby trains social networks' honest parts.

References

1. H. Xu, S. Hao, A. Sari, H. Wang, Privacy risk assessment on email tracking (2018)
2. M. Nasr, R. Shokri, A. Houmansadr, Comprehensive privacy analysis of deep learning: passive and active white-box inference attacks against centralized and federated learning, in *IEEE Symposium on Security and Privacy (SP)* (IEEE, 2019), pp. 739–753
3. B. Mei, Y. Xiao, R. Li, H. Li, X. Cheng, Y. Sun, Image and attribute based convolutional neural network inference attacks in social networks. IEEE Trans. Netw. Sci. Eng. (2018)
4. S. Yu, Big privacy: challenges and opportunities of privacy study in the age of big data. IEEE Access **4**, 2751–2763 (2016)
5. C. Wu, X. Chen, W. Zhu, Y. Zhang, Socially-driven learning-based prefetching in mobile online social networks. IEEE/ACM Trans. Netw. **25**(4), 2320–2333 (2017)
6. A. Martínez-Ballesté, P. A. Pérez-Martínez, A. Solanas, The pursuit of citizens' privacy: a privacy-aware smart city is possible. IEEE Commun. Mag. **51**(6) (2013)
7. H. Liu, X. Li, H. Li, J. Ma, X. Ma, Spatiotemporal correlation-aware dummy-based privacy protection scheme for location-based services, in *Proceedings of IEEE INFOCOM Conference on Computer Communications, INFOCOM 2017, Atlanta, GA, USA, May 1–4, 2017* (2017), pp. 1–9
8. Q. Wang, Y. Zhang, X. Lu, Z. Wang, Z. Qin, K. Ren, Real-time and spatio-temporal crowd-sourced social network data publishing with differential privacy. IEEE Trans. Dependable Sec. Comput. (2016)
9. K. Xing, C. Hu, J. Yu, X. Cheng, F. Zhang, Mutual privacy preserving k -means clustering in social participatory sensing. IEEE Trans. Industr. Inf. **13**(4), 2066–2076 (2017)
10. S. Ji, P. Mittal, R.A. Beyah, Graph data anonymization, de-anonymization attacks, and de-anonymizability quantification: a survey. IEEE Commun. Surv. Tutorials **19**(2), 1305–1326 (2017)
11. K. Liu, E. Terzi, Towards identity anonymization on graphs, in *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2008, Vancouver, BC, Canada, June 10–12, 2008* (2008), pp. 93–106

12. S. Ji, W. Li, M. Srivatsa, J.S. He, R.A. Beyah, General graph data de-anonymization: from mobility traces to social networks. ACM Trans. Inf. Syst. Secur. **18**(4), 12:1–12:29 (2016)
13. A.R. Beresford, F. Stajano, Location privacy in pervasive computing. IEEE Pervasive Comput. **2**(1), 46–55 (2003)
14. T. Shu, Y. Chen, J. Yang, A. Williams, Multi-lateral privacy-preserving localization in pervasive environments, in *2014 IEEE Conference on Computer Communications, INFOCOM 2014, Toronto, Canada, April 27 - May 2, 2014* (2014), pp. 2319–2327
15. W. Wang, Q. Zhang, Privacy preservation for context sensing on smartphone. IEEE/ACM Trans. Netw. **24**(6), 3235–3247 (2016)
16. A. Chaabane, E. De Cristofaro, M.A. Kaafar, E. Uzun, Privacy in content-oriented networking: threats and countermeasures. ACM SIGCOMM Comput. Commun. Rev. **43**(3), 25–33 (2013)
17. P. Zhang, Q. Li, P.P.C. Lee, Achieving content-oriented anonymity with CRISP. IEEE Trans. Dependable Sec. Comput. **14**(6), 578–590 (2017)
18. X. Wang, X. Luo, S. Zhang, Y. Ding, A privacy-preserving fuzzy interest matching protocol for friends finding in social networks, in *Springer Software Computing* (2017)
19. J. Zhou, Z. Cao, X. Dong, X. Lin, Security and privacy in cloud-assisted wireless wearable communications: challenges, solutions, and future directions. IEEE Wirel. Commun. **22**(2), 136–144 (2015)
20. D. He, S. Zeadally, An analysis of RFID authentication schemes for internet of things in healthcare environment using elliptic curve cryptography. IEEE Internet Things J. **2**(1), 72–83 (2015)
21. D. Riboni, L. Pareschi, C. Bettini, Js-reduce: defending your data from sequential background knowledge attacks. IEEE Trans. Dependable Sec. Comput. **9**(3), 387–400 (2012)
22. M. Rezvani, A. Ignjatovic, E. Bertino, S. Jha, Secure data aggregation technique for wireless sensor networks in the presence of collusion attacks. IEEE Trans. Dependable Sec. Comput. **12**(1), 98–110 (2015)
23. R. Yu, J. Kang, X. Huang, S. Xie, Y. Zhang, S. Gjessing, Mixgroup: accumulative pseudonym exchanging for location privacy enhancement in vehicular social networks. IEEE Trans. Dependable Sec. Comput. **13**(1), 93–105 (2016)
24. J. Song, S. Lee, J. Kim, Inference attack on browsing history of twitter users using public click analytics and twitter metadata. IEEE Trans. Dependable Sec. Comput. **13**(3), 340–354 (2016)
25. Y. Chen, Y. Nadji, A. Kountouras, F. Monrose, R. Perdisci, M. Antonakakis, N. Vasiloglou, Practical attacks against graph-based clustering, in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS 2017, Dallas, TX, USA, October 30 - November 03, 2017* (2017), pp. 1125–1142
26. M.A. Ferrag, L.A. Maglaras, A. Ahmim, Privacy-preserving schemes for ad hoc social networks: a survey. IEEE Commun. Sur. Tutorials **19**(4), 3015–3045 (2017)
27. B. Ying, D. Makrakis, H.T. Mouftah, Privacy preserving broadcast message authentication protocol for vanets. J. Netw. Comput. Appl. **36**(5), 1352–1364 (2013)
28. D. Quercia, S. Hailes, Sybil attacks against mobile users: friends and foes to the rescue, in *Proceedings of INFOCOM 2010. 29th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies, 15–19 March 2010, San Diego, CA, USA* (2010), pp. 336–340
29. C. Liu, P. Gao, M.K. Wright, P. Mittal, Exploiting temporal dynamics in sybil defenses, in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, CO, USA, October 12–6, 2015* (2015), pp. 805–816