C V RAMAN GLOBAL UNIVERSITY

ODISHA, BHUBANESWAR

CASE STUDY FOR

DATA MINING AND DATA WAREHOUSING

TOPIC :

Protecting user data in profile-matching social networks

NAME – RISHAV MISHRA

REGD NO. **1901227313**

INTRODUCTION :

In this paper, we consider a scenario where a user queries a user profile database, maintained by a social networking service provider, to identify users whose profiles match the profile specified by the querying user. There are many examples of these kind of websites. Online dating sites like Tinder or Matrimonial sites like Bharat Matrimony or Odia Matrimony are some of the most popular ones. Ashley Madison, a Canadian dating website is one such online dating app for people who are married or currently in relationship. On 19 July 2015, a group of people, known as the Impact Team stole the users data of that website and threatened to leak the information if the company (Ashley Madison) did not shut down immediately. Unfortunately, 60 gigabytes of data was leaked on 18 August 2015, exposing many users' data and labelling the website as a website encouraging extramarital affairs. After this incident, in the fear of getting exposed or publicly shamed, there were 2 unconfirmed suicide cases reported on 24 August 2015, as reported by the Toronto Police. In the end, it was an extremely serious case of privacy breach for the users of the website, two innocent lives lost and practically a disaster for the company itself.

When signing up for an online matching service, a user creates a "profile" that others can browse. The user may be asked to reveal details, such as age, sex, education, profession, number of children, religion, geographic location, sexual proclivities, drinking behaviour, hobbies, income, religion, ethnicity, drug use, home and work addresses, favourite places. Even after an account is cancelled, most online matching sites may retain such information. Users' personal information may be re-disclosed not only to prospective matches, but also to advertisers and, ultimately, to data aggregators who use the data for purposes unrelated to online matching and without customer consent. In addition, there are risks such as scammers, sexual predators, and reputational damage that come along with using online matching services. In fact, some of these information leaks may lead to many crimes such as blackmailing, extortion, money, reputational damage of any big or renowned personality may lead to the damage to any kind of brands associated with that person .i.e indirectly leading to a financial crisis for some people, who even had no idea about this case of data breach.

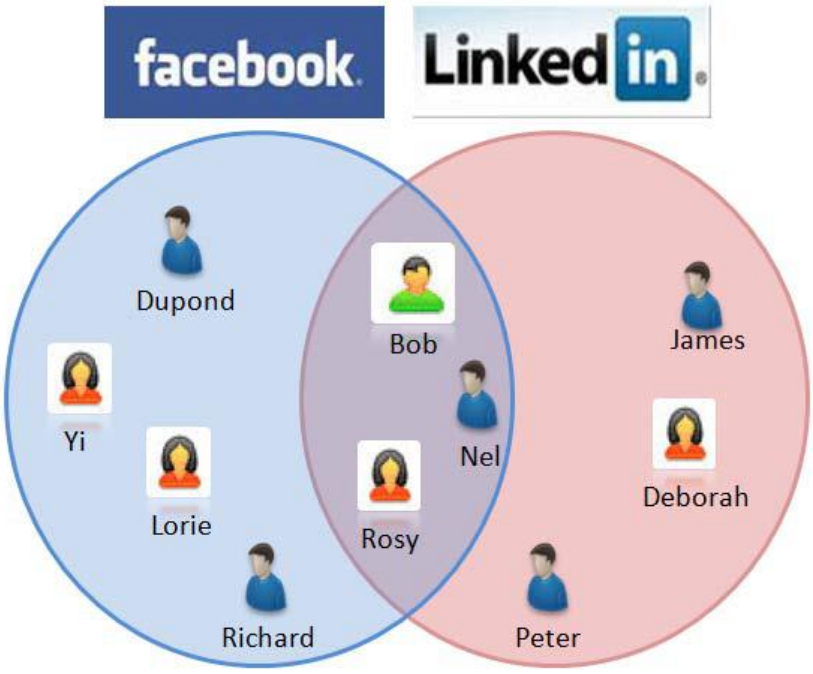
Nowadays, social networking has become an important part of the online activities on the web. Social sites gain popularity thanks to the diverse services provided ranging from collaborative tagging (e.g., Flicker 1), blogging sites (e.g., Livejournal), and mainly to social networking (e.g., Facebook, LinkedIn , MySpace) with a nonstop growing number of active users. In essence, each social network offers particular services and functionalities that target a well defined community in the real world. To make use of the provided services/functionalities and to keep being tuned with its related members, users create several accounts on various sites. This has participated in the emergence of new users' related needs to perform some internetworks' operations and functionalities. To illustrate this, let us consider the following scenario. Bob, a software developer, is very active on social networks. As illustrated in Figure 1, he mainly uses two social sites: the first is Facebook (SN1) to stay connected with his friends, and the second is LinkedIn (SN2) to maintain professional contact with a group of software developers. For different purposes, Bob needs to identify:

1) Intersection between SN1 and SN2: to allow him inviterelated friends (Nel and Rosy) to technically test his new Facebook add-ons

2) Union between SN1 and SN2: to help him send a gift (containing his company promotional package and Facebook add-ons) only once (so to reduce costs) to people that might be interested in his add-ons (such as James, Deborah, Peter, Richard, Lorie, Yi, Dupond, Rosy, and Nel)

3) Difference between SN1 and SN2: to allow him to enrich his friends' profiles with complimentary information found in both sites (particularly Nel and Rosy here)

.



CHALLENGES :

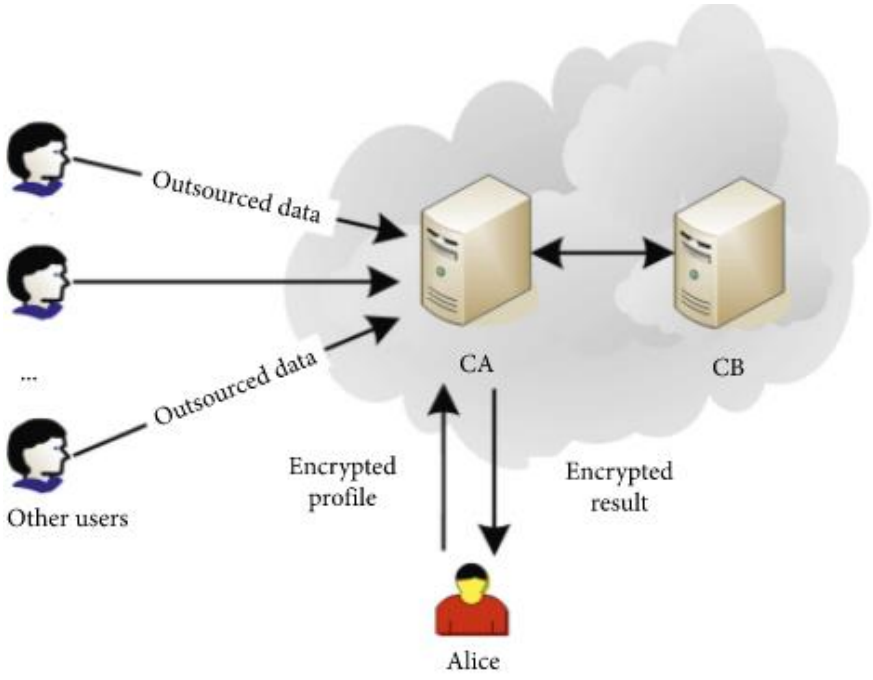
In the profile matching social networks, the system model consists of 3 entities, the friend finder, the cloud environment & other users. The 3 entities are described as follows:

The Friend Finder :
Let's assume that the friend finder has a name .i.e Alice. She is there on the network to find friends who are having similar interests as her, in the social network.

The Cloud Environment :
The cloud environment includes two cloud servers, CA and CB, which provide users with huge storage space for storing their personal profiles & a large amount of resources. Through cooperating, CA and CB can help Alice calculate her suitable proximities with other users, and the privacy of the two users will be not compromised.

Other Users :
This is where the 3rd entity, .i.e the other users come into play. There are many other users on the social network who prefer to outsource the encrypted data representing their preferences to the cloud CA.

Each user registers a personal account in the mobile social network and then fills in the personal information. The personal information contains user's preferences that can be used as a measurement for profile matching. However, the user does not want to reveal his private data in order to avoid illegal use.



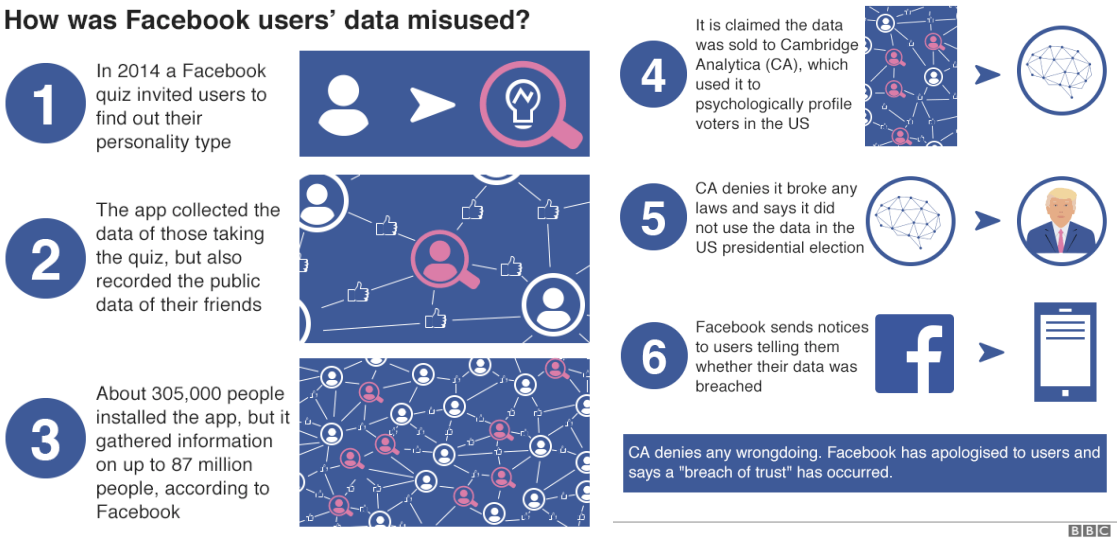
Now let's consider that in the honest-but-curious model of social networking, there is an external adversary and an internal adversary. External adversary mainly refers to an eavesdropper who can get some information (e.g., encrypted data) through the transparent channel by eavesdropping. An internal adversary is an honest-but-curious entity such that he faithfully follow the agreement but attempt to collect and reveal private information during the execution of the agreement. The friend finder may want to expose other users' profiles, while the two clouds may want to reveal the users' personal data in the social networks.

The most prominent example of such a data breach is the Cambridge Analytica incident of US Presidential Elections of 2016.

"A whistleblower has revealed to the Observer how Cambridge Analytica – a company owned by the hedge fund billionaire Robert Mercer, and headed at the time by Trump's key adviser Steve Bannon – used personal information taken without authorisation in early 2014 to build a system that could profile individual US voters, in order to target them with personalised political advertisements."

- **The Guardian**

The main motive of this data breach was to target the users on Facebook with posts and content matching with their interests and gradually manipulating them into vote the party to power. This eventually led to the victory of Donald Trump in the US Presidential Elections of 2016, over Hillary Clinton.



(Photo Source: BBC)

SOLUTIONS :

There are various threats on social media sites for our privacy as we have discussed above. So the best thing we need to do is to protect our privacy to avoid any kind of social threats. Here are some of the solutions.

1) Encryption:

The best solution is through encryption, i.e., users encrypt their profiles before uploading them onto social networks. What is personal data encryption? Encryption is a mathematical function that encodes data in such a way that only authorised users can access it. It is a way of safeguarding against unauthorised or unlawful processing of personal data, and is one way in which you can demonstrate compliance with the security principle. However there lies a problem in the encryption, i.e. whenever a user's profile is encrypted, it is difficult for the system to match the user's profile with similar profiles. But nonetheless, it is a safer choice in order to maintain one's privacy.

2) Antivirus:

If your computer is infected by a virus or malware, not only can hackers dig through your data to steal your identity, but they may lock up your files and ask for a ransom to get them back. The solution is to run an antivirus program to watch for viruses, and keep your other software up to date to close security holes. This applies not only to your computer but your mobile devices as well. Although antivirus comes with many pros, it has it's own set of disadvantages such as frequent upgrades of the antivirus software or even lack of customer care support at many instances.

3) Maintaining your privacy:

Sharing our email IDs on various platforms results in hunderds of spam in your e-mail inbox. Even if we can't avoid sharing this info with Internet services and online stores, we should atleast avoid to share it with random people on social networks. And consider creating a separate, disposable e-mail address and, if possible, a separate phone number for these cases.

4) Self Awareness:

The most important way to maintain our privacy and data on social media networks to be self aware. No amount of technology can help us if in the end we are not self aware about what to share on the social media and most importantly, with whom we share them. We should do at least the bare minimum of securing our accounts with passwords, put security locks on our devices .etc.

Implementation of the solutions

Social networking sites such as Facebook, Twitter, Instagram, and Snapchat have become digital billboards for internet users. People love sharing their personal views and news about what's going on in their lives. But we should not ignore the fact that Spam bots, vindictive acquaintances, and even cybercriminals could take an interest in our social media accounts, too. In the 21st century, information can be a new form of valuable currency. You wouldn't just hand out your bank account information, so why would you give away your privacy rights on social networking sites? Pay particular attention to what information you are agreeing to share when you sign up for a social media account. As an example, according to Facebook, if a user chooses to delete any photos and videos they've previously shared on Facebook, those images will be removed from the site but could remain on Facebook's servers. And some content can be deleted only if the user permanently deletes their account.

Take a moment to wade through the legalese contained in the Privacy Policy and Terms of Service before you click "Accept." You may find that some of the terms are in the best interest of the platform, but may not be the best for your privacy. Some of the conditions may exceed your personal comfort limit. For example, some free sites may gather and sell data related to what you look at to third-parties for marketing purposes. Make sure your permission choices are right for you. Upon signing up for a social media site, most users willingly give their name, gender, date of birth, and email address. Some social media sites don't stop at that. They go on to collect other information like an IP address or the types of things you have liked, shared, or commented on. Sometimes you're given the choice to use your Facebook credentials to log in to other, third-party apps. While this may be convenient, you could unwittingly allow other apps to access more of your personal information than necessary.

CONCLUSION

In this paper, we addressed the issue of providing inter social network operations and functionalities. Creating accounts and providing our data in profile matching social networks can be very useful for enhancing our career prospectives (by creating accounts in social media websites like LinkedeIn) and can also help us to increase our friend circle or enhance our communication with our near and dear ones by joining social media sites like facebook, instagram .etc.

But there always lies a danger ahead of many mishaps such as data phishing, hacking of accounts, leakage of important documents to some unauthorized person .etc. That is the main reason for us to be aware of protecting our data on various profile matching social networks, as to minimize the threat of someone else using our information in a wrong way.

Concluding, I would like to say that this paper gives us the insight into various cases of user data mismanagement and also provides us many ways to protect our user data and to pevent them into getting into the wrong hands.

REFERENCES :

https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

https://www.hindawi.com/journals/scn/2020/4938736/

https://www.researchgate.net/publication/200839522_User_Profile_Matching_in_Social_Networks

https://us.norton.com/internetsecurity-privacy-protecting-privacy-social-media.html