SIT719 Security and Privacy Issues in Analytics
Credit Task 8.2: k-anonymity for Sensitive Data Privacy

A. Quasi-identifiers:

Data, which in connection with other information, can be used to identify an individual with high probability, e.g., age at baseline, race, gender, events, specific findings.
A quasi-identifier is a piece of information that an intruder can get hold of about a specific target individual or about a large number of people through the following means:

- Personal knowledge of the specific target person (e.g., a neighbour, co-worker, ex-spouse).
- The specific target person is famous and there is information publicly available about them.
- Publicly available registries (e.g., voter lists and court records) or the media (e.g., obituaries published in newspapers or on-line).
- Information that individuals post about themselves on the Internet (e.g., information they post on social networking sites).
- Information that individuals often disclose to a large number of people (e.g., their baby's birth weight or birth date).

B. K-Anonymity:

k-Anonymity is a property that captures the protection of released data against possible reidentification of the respondents to whom the data refer. k-Anonymity states that the released data should be indistinguishably related to no less than k respondents.

- The concept of k-anonymity and its formalization have been proposed together with its enforcement via generalization and suppression, two microdata protection techniques that have the advantage of preserving the truthfulness of the information. Generalization consists in replacing the values of an attribute with more general values (e.g., the data of birth can be substituted with the year of birth).
- The final effect of a generalization is that tuples in the original microdata table with different values for the quasi-identifier are generalized to the same value, thus becoming indistinguishable.
- Suppression consists in removing data from the table so that they are not released. Suppression is used to "moderate" the generalization process when a limited number of outliers (i.e., tuples with less than k occurrences) would force a great amount of generalization.

C. How k-anonymity can help prevent privacy attack?

With k-anonymity an original data set containing personal health information can be transformed so that it is difficult for an intruder to determine the identity of the individuals in that data set. A k-anonymized data set has the property that each record is similar to at least another *k*-1 other records on the potentially identifying variables. For example, if *k* = 5 and the potentially identifying variables are age and gender, then a k-anonymized data set has at least 5 records for each value combination of age and gender.

→Commercial and open-source tools for data anonymization:

a. ARX Data Anonymization Tool
b. Amnesia
c. μ-ARGUS
d. sdcMicro
e. Anonimatron
f. Aircloak Insights
g. CloverDX
h. BizDataX
i. Aircloak

References

I. Cloverdx.com. 2020. 8 Fundamental Data Anonymization Mistakes That Could Put Your Business At Risk. [online] Available at: <https://www.cloverdx.com/blog/data-anonymization-mistakes> [Accessed 14 May 2020].

II. Ehealthinformation.ca. 2020. What Is A Quasi-Identifier? - Electronic Health Information Laboratory. [online] Available at: <http://www.ehealthinformation.ca/faq/quasi-identifier/> [Accessed 14 May 2020].

III. El Emam, K. and Dankar, F., 2008. Protecting Privacy Using k-Anonymity. Journal of the American Medical Informatics Association, 15(5), pp.627-637.

IV. Harvard Business Review. 2020. There'S No Such Thing As Anonymous Data. [online] Available at: <https://hbr.org/2015/02/theres-no-such-thing-as-anonymous-data> [Accessed 14 May 2020].

V. Kaliski, B., Kaliski, B., Silverman, R., Brassard, G., Crépeau, C., Vimercati, S. and Foresti, S., 2011. Quasi-Identifier. Encyclopedia of Cryptography and Security, pp.1010-1011.

VI. Samarati, P., Weimerskirch, A., De Cannière, C., Eisenbarth, T., Kasper, T., Paar, C., Indesteege, S., Racic, R., Adams, C., Petitcolas, F., Bauer, F., Just, M., Zuccherato, R., Pedersen, T., Just, M., Kiayias, A., Yener, B., Lloyd, S., Adams, C., Thornton, M., Estes, A., Campisi, P., Neri, A., Desmedt, Y., Biryukov, A., Hankerson, D. and Menezes, A., 2011. k-Anonymity. Encyclopedia of Cryptography and Security, pp.663-666.