Tópicos de Privacidade

Vivemos em uma era de abundância de dados. Estes dados são coletados e processados de modo a nos oferecer vários tipos de serviços (médico, entretenimento, etc) que melhoram nossa qualidade de vida. Isto, no entanto, não tem sido sem repercussões. Em [1], por exemplo, os autores mostram como características privadas (etnia, gênero, afiliação política, etc) de um usuário de uma rede social podem ser inferidas a partir de seus "likes" usando algoritmos de aprendizado de máquina. Estas técnicas estão por trás do escândalo de dados do Facebook–Cambridge Analytica, por exemplo. Neste escândalo, a Cambridge Analytica, uma firma de consultoria política britânica, utilizou-se destes métodos para traçar o perfil de usuários de modo a influenciá-los politicamente por meio de anúncios direcionados.

Outros escândalos de privacidade relacionados a dados pessoais incluem: aplicativo de smartwatch vazando a localização de bases militares americanas [2], vazamento de registros médicos de pacientes [3], pai de adolescente descobrindo que sua filha está grávida por causa de cupons personalizados [4], entre vários outros. Estes exemplos mostram a dificuldade, muitas vezes, em se distinguir entre dados que devem ser mantidos privados ou públicos.

Neste curso, focaremos em quatro áreas da privacidade.

Privacidade no Armazenamento Distribuído: Nesta área estuda-se o problema de como armazenar dados privados em uma rede de servidores sem que os servidores aprendam qualquer coisa sobre os dados. A técnica principal que iremos explorar é conhecida como compartilhamento de segredo [5]. Esta técnica é uma peça fundamental que sera utilizada, também, nas outras áreas.

Recuperação de Informações Privadas: Nesta área [6] estuda-se o problema de como recuperar dados de uma rede de servidores sem que os servidores aprendam em qual dado o usuário está interessado. Cenários clássicos de aplicação incluem: um investidor que deseja consultar o preço de determinada ação sem se expor para não afetar os preços das ações, uma empresa buscando patentes sem querer divulgar quais de modo a não vazar possíveis projetos nos quais está trabalhando, entre outros. A abordagem sera baseada em [7] que permite, a partir de esquemas de compartilhamento de segredo, criar os esquemas de Recuperação de Informações Privadas com melhor desempenho.

Segurança na Computação Distribuída: Nesta área estuda-se o problema de como um usuário pode computar uma função de seus dados com a assistência de servidores, sem que os servidores aprendam qualquer coisa sobre os dados. O foco do curso sera na segurança na multiplicação de matrizes [8, 9]. Aplicações incluem o aprendizado seguro de máquina distribuído.

Privacidade Diferencial: A privacidade diferencial é um conceito relativamente novo [10] que vem sendo usado com bastante sucesso no problema de privacidade na análise de dados. Dentre as aplicações destacamos a aprendizagem privada em larga escala da Apple e a sua utilização no Censo de 2020 dos Estados Unidos, cada uma impactando centenas de milhões de indivíduos.

Referências

[1] M. Kosinski, D. Stillwell, and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," *Proceedings of the national academy of sciences*, vol. 110, no. 15, pp. 5802–5805, 2013.

[2] "Fitness tracking app strava gives away location of secret us army bases," https://tinyurl.com/yd6hfxaq, accessed: 2020-09-03.

[3] "Google health-data scandal spooks researchers," https://tinyurl.com/y5wpntff, accessed: 2020-09-03.

[4] "How companies learn your secrets," https://tinyurl.com/7p6bgut, accessed: 2020-09-03.

[5] A. Shamir, "How to share a secret," *Communications of the ACM*, vol. 22, no. 11, pp. 612–613, 1979.

[6] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, "Private information retrieval," in *Proceedings of IEEE 36th Annual Foundations of Computer Science*. IEEE, 1995, pp. 41–50.

[7] R. G. L. D'Oliveira and S. El Rouayheb, "One-shot pir: Refinement and lifting," *IEEE Transactions on Information Theory*, vol. 66, no. 4, pp. 2443–2455, 2020.

[8] W.-T. Chang and R. Tandon, "On the capacity of secure distributed matrix multiplication," in *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–6.

[9] R. G. L. D'Oliveira, S. El Rouayheb, and D. Karpuk, "Gasp codes for secure distributed matrix multiplication," *IEEE Transactions on Information Theory*, pp. 1–1, 2020.

[10] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography*, S. Halevi and T. Rabin, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 265–284.