dans des données médicales pseudonymisées par une simple comparaison avec, justement, une liste électorale !

Une autre forme d'attaque est l'inférence. Celle dite « d'attribut » consiste à déduire une information sensible à partir des données anonymisées. Se protéger de cette menace est important même dans les situations où le risque de réidentification est nul. Ainsi, déterminer l'origine ethnique d'une personne peut conduire à une discrimination pour l'obtention d'un prêt bancaire. Quant à l'inférence « d'appartenance », elle a pour objectif de décider si un individu particulier fait partie d'un ensemble de données sensibles, même anonymisées, comme une cohorte de patients atteints d'un cancer.

Récemment, ces attaques par inférence ont été adaptées au contexte des modèles d'apprentissage, du type de ceux utilisés en intelligence artificielle. Ce ne sont pas alors les données elles-mêmes qui sont visées, mais une structure, comme un réseau de neurones, qui a été élaborée à partir des données. Ici, l'objectif est de déterminer si les données d'un individu cible faisaient partie de celles qui ont servi à l'apprentissage, par exemple d'un système de classification des tumeurs en oncologie.

PREMIERS ÉCHECS

Motivés par les scandales de réidentification, les premiers modèles formels de vie privée ont cherché à contrecarrer les attaques dues à l'existence des quasi-identifiants. Les algorithmes d'anonymisation ainsi conçus partitionnent un jeu de données de sorte qu'un ensemble de contraintes soit satisfait dans chaque morceau. Le *k*-anonymat, par exemple, requiert que les quasi-identifiants d'une même partition soient identiques et que chaque partition contienne les données d'au moins *k* individus. Une attaque de réidentification à partir du quasi-identifiant devient donc imprécise, d'autant plus que la valeur de *k* est grande.

Autre modèle formel pionnier, la *l*-diversité nécessite que chaque partition contienne un ensemble de données sensibles suffisamment divers : au moins *l* données sensibles distinctes, ou bien que la valeur de la donnée sensible la plus fréquente ne soit pas beaucoup plus présente que les données sensibles les moins fréquentes. La *l*-diversité espère ainsi lutter contre des attaques d'inférence d'attribut qui associeraient simplement à un quasi-identifiant la donnée sensible la plus fréquente dans sa partition.

Cette famille de modèles est aujourd'hui sur le déclin, victime des failles et des limitations qui ont été découvertes progressivement. D'abord, ces modèles ne sont nativement pas composables. Ce qui signifie que des analyses successives sur un même jeu de données, certes anonymisées, peuvent néanmoins briser la protection. Ensuite, ces modèles ne sont pas assurés contre le post-traitement : les critères de protection ignorent les algorithmes qui anonymisent les jeux de données. Or le comportement de ces programmes est souvent déterministe (il fournit toujours le même résultat à partir de données identiques) et dépendant des données. Certaines attaques tirent parti de ces caractéristiques pour faire de la « rétro-ingénierie » afin de retrouver la base de données initiale à partir de sa version anonymisée.

BEAUCOUP DE BRUIT POUR RIEN ?

La situation de l'anonymisation des données a radicalement changé avec la confidentialité différentielle, inventée en 2006 par Cynthia Dwork, alors chez Microsoft, conjointement avec d'autres chercheurs, originellement pour contrer les attaques d'un autre type, celles dites de « reconstruction de base de données » : l'idée est de générer un grand nombre de requêtes statistiques sur des sous-ensembles variés de la base de données (nombre de patients qui ont la grippe, ceux dont le numéro de sécurité sociale commence par « 1 70 » et qui ont la grippe…), puis de résoudre un système d'équations