Anonymisation de données

Anas ABOU EL KALAM (LAAS - CNRS, ENSIB)
Emmanuel CORDONNIER (ETIAM)
Yves DESWARTE (LAAS - CNRS)
Gilles TROUESSIN (Ernst & Young Audit)
Christophe ZANON (LAAS - CNRS)

Démarche

Besoins :

- Spécifiques à l'application, au contexte, …
- Que veut-on protéger ? Contre qui / quoi ?

Objectifs :

- *Réversibilité* → chiffrement
- *Irreversibilité* → fonction de hachage à sens unique
- *Inversibilité* : pseudonymisation → chiffrement à clé publique
  - désanonymisation ⇒ procédure exceptionnelle

Exigences :

- *Chaînage* :
  - *temporel* (toujours, parfois, jamais)
  - *spatial* (international, national, régional, local)
  - *spatial et temporel* (toujours-partout, parfois-régional, local-jamais)
- *Robustesse* : à la réversion, à l'inférence

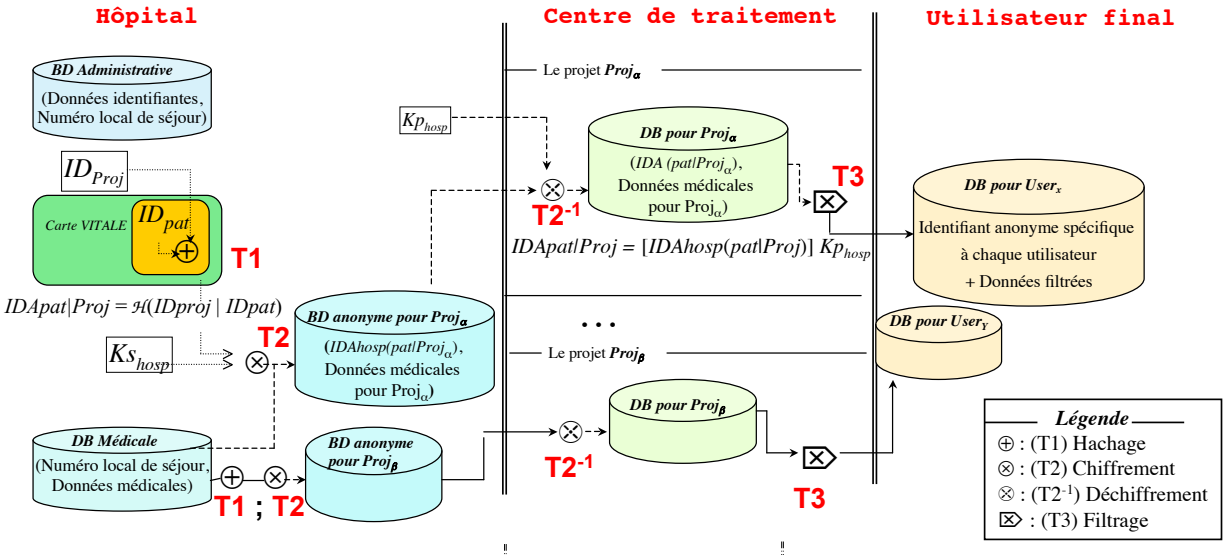
Exemple d'applications

- **Stockage et transfert de données médicales :**
  - Objectif : confidentialité, réversibilité
  - Exigence : robustesse à l'inversion (cryptanalyse)

- **Maladies à déclaration obligatoire :**
  - Besoins : prévention, veille sanitaire, analyses épidémiologiques, …
  - Objectif : Anonymisation irréversible
  - Exigence : chaînage contrôlé, robustesse à la réversion et aux inférences, …

- **Études épidémiologiques focalisées :**
  - Besoins : cacher les identités tout en ayant la possibilité d'identifier et d'informer les patients afin d'améliorer la qualité des soins
  - Objectif : Anonymisation inversible (pseudonymisation)
  - Exigence :
    - robustesse aux attaques par inférence
    - Type d'utilisation / d'utilisateur ⇒ Type du chaînage (temporel & géographique)

Nouvelle solution : Schéma général



(T1) → IDApat|Proj = H(IDproj / IDpat)

(T2) → IDAhôp(pat|Proj) = {IDApat|Proj} $Ks_{hôp}$

(T2^{-1}) → IDApat|Proj

T3 → agrégation, appauvrissement

H(IDApat|Proj | Kutil/proj)

Stockée ou générée aléatoirement

Discussion

- Protection de l'identifiant anonyme du patient :
  - *IDpat* est générée aléatoirement au sein de la carte
  - La carte est supposé suffisamment inviolable « tamper-resistant »
  - Le calcul H(*IDproj*|*IDpat*) est effectué au sein de la carte
- Absence de secret critique pour toute la population
  - L'identifiant anonyme ne dépend que du (patient, projet)
  - Les identifiants sont situés dans des endroits différents
  - Les clés sont détenues par des personnes différentes
- Consentement explicite du patient
  - Lors de toute utilisation non-obligatoire, mais souhaitable, de ses données
  - Pour lever l'anonymat
- Respect de la réglementation européenne / internationale
  - Finalité du traitement (objectif de l'utilisation)
  - Résistance aux attaques par dictionnaire, aux inférences par inversion, …
- Flexibilité
  - Objectifs de protection différents selon les utilisateurs
  - Fusion de plusieurs établissements

Démonstration

- Hôpital :
  - Secrétaire (carte CPS) :
    - visualisation de données administratives
    - création d'un dossier patient (carte vitale)
  - Médecin (carte CPS) :
    - visualisation des données administratives et médicales, mise à jour des données médicales
    - création de la participation à un projet (carte vitale)
    - envoi d'une base anonyme au projet
- Centre de traitement
  - affichage d'un projet
  - création d'un projet (-> liaison avec l'hôpital)
  - envoi de données à un utilisateur final
    - filtrage, appauvrissement