Une démarche méthodologique pour l'anonymisation des données personnelles sensibles

Anas ABOU EL KALAM (LAAS - CNRS)
anas@laas.fr

Avec
Yves Deswarte (LAAS - CNRS)
Gilles Trouessin (Ernst & Young Audit)
Emmanuel Cordonnier (ETIAM)

Plan

- Introduction
- Démarche d'analyse
- Exemples de scénarios
- Nouvelle solution
- Conclusions

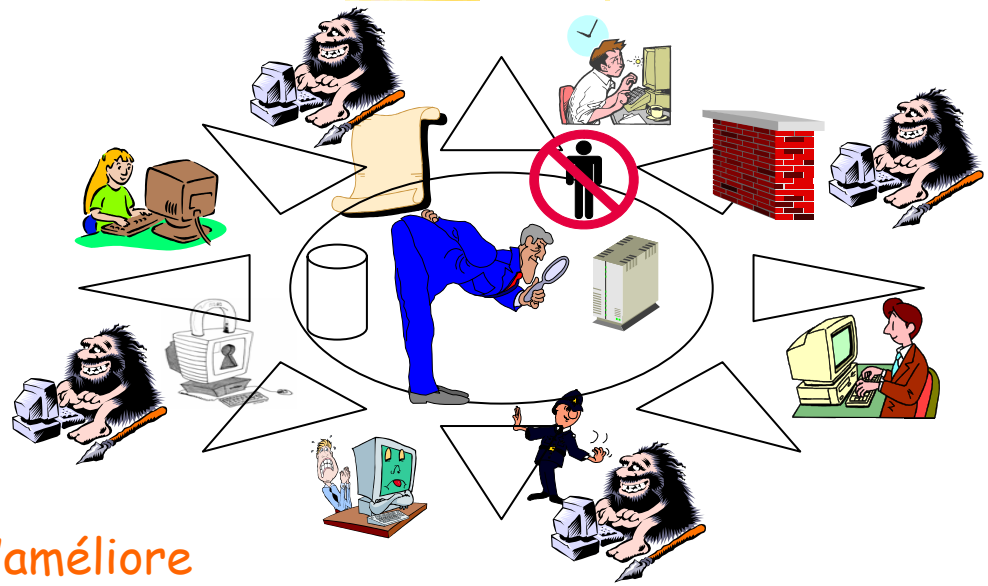
Plan

- Introduction

Et la protection de la vie privée
Réglementation
point de départ ?

- Démarche d'analyse
- Exemples de scénarios
- Nouvelle solution
- Conclusions

… Et la protection de la vie privée ?



- La sécurité s'améliore
  - Législation, politiques, détection d'intrusion, tolérance aux intrusions, …
- … Mais croissance des menaces
  - DDoS, criminalité transfrontalière, … et les failles sont nombreuses
- … D'où encore plus de sécurité
  - Traçabilité, authentification forte, …

➲ **Danger pour la vie privée !!**

"Privacy" : réglementation

Internationale :

- A/RES/45/95 pour la réglementation des fichiers personnels informatisés

Européenne :

- Directive 95/46/CE pour la protection des données à caractère personnel
- Directive 2002/58/EC concernant e le commerce électronique
- Résolution 98/1165 du 26 juin 1998 sur le droit au respect de la vie privée

Nationale :

- Loi 78-17 "Informatique, fichiers et libertés", protection des données nominatives
- Loi 2002-303 du 4 mars 2002 relative aux droits des malades ; 94-458 pour la recherche médicale
- Code pénal, art. 226-13 (secret professionnel) et 15 (secret des correspondances) concernant les infractions pénales contre la vie privée

Américaine :

- Loi fédérale "*Privacy Act*" de 1974
- "*Privacy Protection act*" (*COPPA*) adopté par le congrès en octobre 1998

Point de départ

- Les applications émergentes utilisent des données personnelles
- La législation existe

MAIS …

- Pas le support technologique !

Par quoi commencer ?

- Comprendre le domaine
- Démarche méthodologique

Plan

- Introduction
- Démarche d'analyse
  - Définitions
  - Besoins & Objectifs & Exigences
  - Choix de solutions
- Exemples de scénarios
- Nouvelle solution
- Conclusions

Définitions [ISO 15408]

- Anonymat :

  Impossibilité de déterminer le véritable nom de l'utilisateur
  ≡ ¬ (révélation de l'identité)

- "Pseunonymat" :

  Anonymat + responsabilité

- Non-"chainabilité" :

  Impossibilité d'établir un lien entre différentes opérations faites par un même utilisateur

- Non-observabilité :

  Impossibilité de déterminer si une opération est en cours

Démarche

Besoins :

- D'ordre général; spécifiques à l'application, …
- Ex : que veut on protéger ? Contre qui / quoi ?

Objectifs :

- *Réversibilité* → chiffrement
- *Irreversibilité* → hachage
- *Inversibilité* : pseudonymisation → chiffrement à clés publiques
  - désanonymisation ⇒ procédure exceptionnelle

Exigences :

- Comment exprimer les besoins de manière plus précise en tenant compte des attaques, de l'environnement, etc.
  - *Chaînage* :
    - *temporel* (toujours, parfois, jamais)
    - *spatial* (international, national, régional, local)
    - *spatio-temporel* (toujours et partout, parfois et partout, local et jamais)
  - *Robustesse* : à la réversion, à l'inférence

Démarche

Choix de solutions :

- **Type**
  - organisationnelle, contrôle d'accès
  - mécanisme cryptographique
  - fonction à sens unique
- **Pluralité**
  - mono-anonymisation
  - bi-anonymisation
  - multi-anonymisation
- **Interopérabilité**
  - transcodage (manuel)
  - translation (mathématique)
  - transformation (automatique)

Plan

- Introduction
- Démarche d'analyse
- Exemples de scénarios
  - Stockage et transfert de données médicales
  - Études épidémiologiques focalisées
  - Maladies à déclaration obligatoire
- Nouvelle solution
- Conclusions

Exemple de scénarios

- Stockage et transfert de données médicales :
  - Objectif : réversibilité
  - Exigence : robustesse à la réversion

- Maladies à déclaration obligatoire :
  - Besoins : prévention, veille sanitaire, analyses épidémiologiques, …
  - Objectif : Anonymisation irréversible
  - Exigence : chaînage universel, 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)

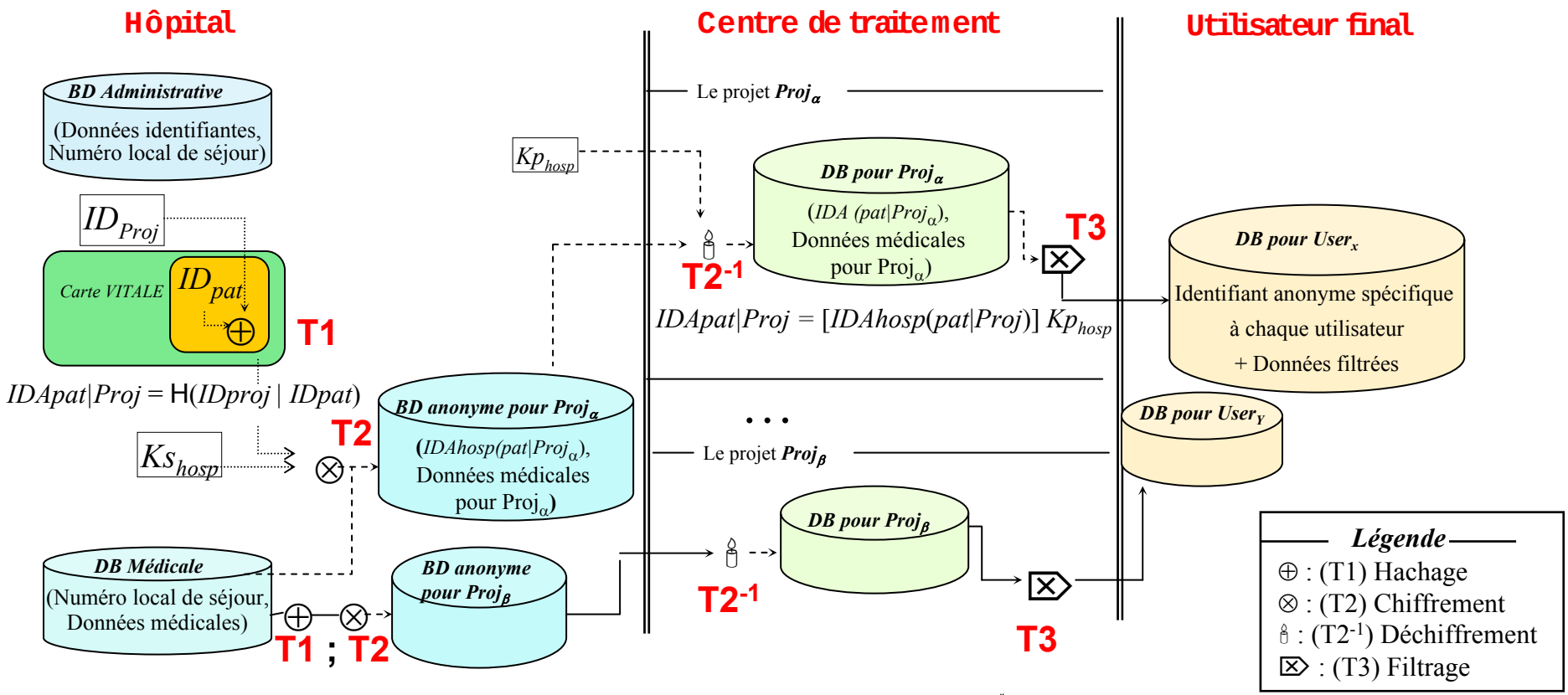
Plan

- Introduction
- Démarche d'analyse
- Exemple de scénarios
  - Nouvelle solution

    Schéma général

    Discussion
- Conclusions

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

T3 → $H(IDApat|Proj \mid Kutil/proj)$

Stockées ou générée aléatoirement

Nouvelle solution : 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
  - Principe du moindre privilège
  - Finalité du traitement (objectif de l'utilisation)
  - Résistance aux attaques par dictionnaire, aux inférences par inversion, …
- Flexibilité
  - Fusionn de plusieurs établissements
  - Changement des objectifs de protection

chaque hôpital :
- déchiffre ses données avec $Kp_{hôp}$
- chiffre le résultat avec $Kshôp_{ab}$

$\{ [IDAhôp_a(pat|Proj)]\ Kp_{hôpa} \}\ Ks_{hôpab}$

Plan

- Introduction
- Démarche d'analyse
- Exemples de scénarios
- Nouvelle solution
  - Conclusions

Conclusion

- Type de réversibilité ?
- Type de chaînage ?
- Forme de chaînage ?
- Robustesse à la réversion ?
- Robustesse à l'inférence ?

Le risque "0" n'existe pas !

- **Démarche analytique**
  - Besoins → Objectifs → Exigences
- **Anonymisations**
  - Thématiques / en cascades

Mais aussi

- **Solutions organisationnelles**
  - Analyse des risques d'abduction
  - Politique de contrôle d'accès
- **Solutions techniques**
  - Architecture adaptée
  - Mécanismes appropriés (de contrôle d'accès)
  - IDS
  - Brouillage
  - Filtrage