Vers une approche scientifique et thématique de la confidentialité

Frédéric Prost
Université Grenoble Alpes
frederic.prost@univ-grenoble-alpes.fr

Résumé

La thématique de la confidentialité étudiée du point de vue de l'informaticien est à la fois nouvelle, changeante, et prenant une part de plus en plus importante dans la société. Je fais le retour de plusieurs types d'interventions (séminaires, cours) devant différents types de publics (d'étudiants spécialisés au grand public) sur ce thème pour en dégager les spécificités en vue de son enseignement.

1 Informatique et Société : influences croisées

Notre société évolue plus rapidement que jamais. Les évolutions des technologies de l'information et de la communication (TIC) en sont pour une large proportion. Même en y accordant une attention soutenue, voire professionnelle, il est difficile de cerner toutes les implications des TIC en terme de confidentialité. On peut mesurer pourtant l'importance de cette question tous les jours : des implications géostratégiques (on pourra penser par exemple aux retombées des révélations d'Edward Snowden) aux considérations économiques (Adwords et Adsense les services publicitaires ciblés de Google) tout en passant par les indiscrétions et autres fuites organisées autour des réseaux sociaux et d'internet (Wikileaks est maintenant un canal quasi institutionnalisé de diffusion d'informations confidentielles avec des répercussions mondiales). Le jeu entre les utilisations et les nouvelles capacités apportées par les TIC font bouger en permanence les lignes sur ce qui est considéré comme confidentiel.

Les problèmes soulevés par une approche scientifique et thématique (du point de vue de l'informaticien) de la confidentialité sont nombreux. Je suis intervenu sur ce thème devant différents types de public et sous différents formats :

- Grand public : conférences pour lycéens en terminale option ISN.
- Informaticiens "généralistes" : au travers d'un cours donné en niveau L3 à l'Université Grenoble Alpes (3 ans).
- Informaticiens "spécialistes" : au travers d'un cours de M2R à l'ENS-Lyon (2 ans) et d'une école d'hiver à l'université de Buenos Aires (ECI 2015).

Je propose de revenir sur cette expérience d'enseignement autour de cette problématique. Dans un premier temps je vais discuter des spécificités de cette thématique notamment dans la perspective particulière de l'informatique. J'aborderais ensuite la question de la place traditionnellement occupée par la confidentialité dans les enseignements en informatique qui est bien souvent limitée à la cryptographie. Enfin j'aborderai le traitement des différents types de public et de comment adapter l'enseignement en fonction des capacités et attentes des auditeurs.

2 La sécurité : une pratique contre-intuitive pour les scientifiques

Depuis l'antiquité les philosophes réfléchissent aux liens entre le monde sensible, celui des objets, et le monde des idées. Il est clair que ces deux mondes ne

vivent pas selon les mêmes lois, il est possible de lister quelques unes de leurs différences fondamentales ayant des implications particulièrement frappantes dans le domaine de la confidentialité :

- Autant la notion d'original a un sens dans le monde sensible, autant elle n'en a plus dans le monde des idées ou la copie peut être parfaitement égale à l'orignial.
- Voler un objet n'est pas la même chose que copier une information (copie non destructrice).
- Les objets subissent l'assaut du temps alors que les idées sont intemporelles.
- Les objets peuvent être détruits au contraire d'une idée.
- Si on peut récupérer un objet qui nous a été dérobé, il est impossible de recouvrer un "secret" perdu.
- etc.

Or, il se trouve que les TIC ont, très paradoxalement, rendu concret le monde des idées. Il est devenu quasiment gratuit de copier parfaitement ou de transmettre des informations numériques. De même avec l'augmentation extraordinaire des capacités de stockages l'information numérique prend quasiment un statut "idéal" au sens philosophique du terme (même si en fin de compte l'information repose sur un support physique). De plus, la convergence numérique a rendu le jeu entre ces deux mondes fréquents. En témoigne l'expérience d'Egor Tsvetkov qui à partir de simples photographies prises avec un smartphone dans le métro de Saint Petersbourg, a pu, en utilisant des logiciels libres de reconnaissances de visages ainsi que les photos publiées sur les réseaux sociaux, retrouver quantité d'informations sur de parfaits inconnus.

En plus de ces difficultés d'ordre philosophique s'ajoute le fait que la problématique de la confidentialité doit se concevoir sous l'angle, assez inhabituel pour les scientifiques, de la sécurité. D'un point de vue traditionnel, la sécurité pour un ingénieur, est traitée similairement à la sécurité contre les accidents/incidents ou les erreurs dans le cadre de l'informatique (on peut parler de robustesse). Rarement est abordée la question de la sécurité dans la perspective où la partie adverse cherche *activement* à faire échouer les mesures de sécurité. En termes d'enseignement cela débouche sur une activité pédagogique très différente des enseignements traditionnels. En effet, ces derniers suivent souvent le schéma suivant : exposé du problème (par exemple trouver un algorithme résolvant tel problème avec telle propriété) auquel on cherche une solution qu'on prouve correcte. Dans le domaine de la sécurité c'est souvent le contraire qui se passe : une solution est donnée et l'activité principale est de savoir comment cette solution pourrait ne pas fonctionner, autrement dit comment un adversaire pourrait la déjouer. Cette attitude tient plus de celle du joueur, au sens joueur d'échecs ou de poker qui ne fait pas que faire des calculs isolés du monde mais doit prendre en compte l'adversité et essayer de construire une solution en sachant que cette solution va être attaquée de toutes les façons possibles, que du mathématicien qui cherche à démontrer un théorème, qui lui ne se défend pas contre une tentative de démonstration.

3 L'iceberg qui cache la forêt

Les impacts qu'ont les questions de confidentialité sont complexes à définir et à saisir précisément. Traditionnellement, pour un informaticien, ces questions se limitaient plus ou moins au domaine de la cryptographie. On remarquera d'ailleurs combien ce domaine s'est étendu au cours du temps en partant du simple problème de la cryptographie stricto-sensu (c'est-à-dire le cryptage de messages) puis en s'intéressant progressivement à des problématiques de plus en plus sophistiquées : intégrité des données, signatures électroniques, argent électronique (anonyme ou non), votes électronique, schémas de cryptages homomorphes, etc. Cependant la cryptographie n'est que la partie la plus visible des relations entre confidentialité et informatique. Nous proposons la grille de lecture suivante des effets liés à la confidentialité :

Effets liés à la confidentialité

	Directs	Indirects
Visibles	Cryptographie	Ingéniérie Sociale
Invisibles	Big Data Bulle Filtre	"Pre-Suasion"

Le fait de pouvoir communiquer sans que l'information puisse être interceptée est à la fois un problème direct et visible lié à la confidentialité. L'ingéniérie sociale qui consiste à accumuler des informations en vue de monter une attaque (par exemple en gagnant la confiance d'un utilisateur en se servant des informations que ce dernier aurait diffusé sur un réseau social) est un exemple d'effet visible (car l'attaque peut laisser des traces) mais indirect (la confidentialité n'est qu'un rouage dans l'attaque menée). De même de nombreuses application liées au Big Data ont des conséquences invisibles mais directes : l'exemple typique est la manière dont Google présente les résultats d'une même recherche à des profils différents. Du point de vue de l'utilisateur cette manière de filtrer les recherches est invisible et dépend directement de ce que Google sait de l'utilisateur. Enfin, les dernières recherches en psychologie sociale ont montré combien on pouvait influencer des décision juste en modifiant l'environnement ou la manière de présenter les choses : là encore l'utilisation d'informations personnelles est à l'origine de ce type de manipulations que les TIC rendent plus fréquentes et plus efficaces.

De plus, le problème général de tracer les flots d'informations dans un système informatique n'est pas uniquement lié au fait de crypter l'information. C'est un problème de type "non-interférence", où on cherche à montrer que la modification de certaines parties du code (disons celles vues comme privées) n'a pas d'influence visible sur d'autre parties du code (celles désignées comme publiques). Ce sont les domaines des systèmes formels, de la sémantique, domaines traditionnels en informatique, qui sont mis en jeu.

4 "Informatique et Confidentialité" dans différents contextes

Le principal challenge d'un enseignement autour d'informatique et confidentialité est d'arriver à embrasser un domaine très vaste recouvrant plusieurs champs traditionnels de l'informatique. Je suis intervenu devant trois types de publics sur cette thématique qui chacun requièrent une approche pédagogique différente :

- Le grand public : depuis 5 ans j'ai donné de nombreuses conférences a un public non spécialiste (que ce soit réellement du grand public dans des universités ouvertes ou des élèves de terminale en spécialité ISN). Mon objectif dans ces conférences est d'essayer de faire toucher du doigt les différents aspects liés à la confidentialité (pour simplifier les 4 cases du tableau de la section précédente) au travers d'exemples d'actualité.
- Les informaticiens non spécialisés : je suis intervenu durant 3 ans en L3 informatique. Là l'objectif est plus de montrer quels sont les différents champs de l'informatique qui sont impactés et comment ils interagissent entre eux.
- Les informaticiens spécialisés : je suis intervenu deux ans dans le M2R de l'ENS-Lyon et j'ai également tenu une semaine de cours à l'ECI 2015 (école d'hiver de l'université de Buenos Aires) sur ce thème. Là le but est de montrer les difficultés soulevées par ces problématiques de confidentialité : notamment le fait que formaliser correctement les choses est en soi un problème de recherche encore actif.

A titre d'exemple je donne, de manière synthétique, le programme que j'ai établi pour ce type d'enseignement pour des informaticiens spécialisés (c'est à dire à partir du niveau Master 2) :

1. La cryptographie n'est pas suffisante. Limites absolues.
   - Exemple historique de la cryptanalyse d'Enigma.

  - Théorie de l'information, systèmes inconditionnelements sûrs.
  - Entropie, application au choix de mots de passe.
2. La confidentialité, et la problématique de l'identité, vu sous le prisme traditionnel de la cryptographie.
  - Hash sécurisé, signature.
  - Mécanisme de preuve d'identité.
  - Communications anonymes (routage par oignon).
  - Cash électronique.
3. Progammation et propriété de non-interférence.
  - Non-interférence dans des langages fonctionnels purs.
  - Non-interférence pour les langages impératifs.
  - Extension aux paradigmes avec parallélisme.
  - Assouplissement de la définition de non-interférence (politiques de confidentialité).
4. Preuves Zero-Knowledge.
  - Intuitions derrière les preuves Zero-Knowledge.
  - Complexité de connaissance (Knowldege complexity)
  - Applications des ZKP : contrôle d'accès, signature de groupe, blacklisting anonyme, etc.
5. Approches formelles.
  - Modèle Dolev-Yao.
  - Modèle de l'oracle aléatoire (Random Oracle Model).
  - Approches formelles pour le vote électronique.

5 Conclusion

L'importance du thème de la confidentialité est assez récent dans les cursus informatiques. Il a souvent été confondu avec la cryptographie, domaine avec lequel il ne se confond pas. La nature très large d'un tel sujet, qui touche quasiment toutes les branches de l'informatique, en fait un thème particulièrement difficile à enseigner car nécessitant d'avoir un recul et une vue d'ensemble du domaine. Cependant, le domaine motive véritablement les étudiants en les touchant personnellement. Il permet aussi de leur montrer comment s'articulent différents domaines de l'informatique (des preuves formelles en allant à des problèmes de complexités théoriques en passant par la compilation et la sémantique des langages de programmation) dans un cas concret.