M1 MEEF NSI — Sécurité des communications informatiques

Introduction

Pierre Karpman
pierre.karpman@univ-grenoble-alpes.fr
https://membres-ljk.imag.fr/Pierre.Karpman/tea.html

2024–01–19

Format du cours

- 9 heures de CM ($3 + 4 \times 1,5$)
- 6 heures de TD ($4 \times 1,5$)
- 6 heures de TP (2×3)
- Évaluation une note de contrôle continu écrit et/ou oral (modalités à déterminer)

De quoi parlons nous ?

Introduction aux définitions

Quantifier la sécurité

Sécurisation des communications : pourquoi ?

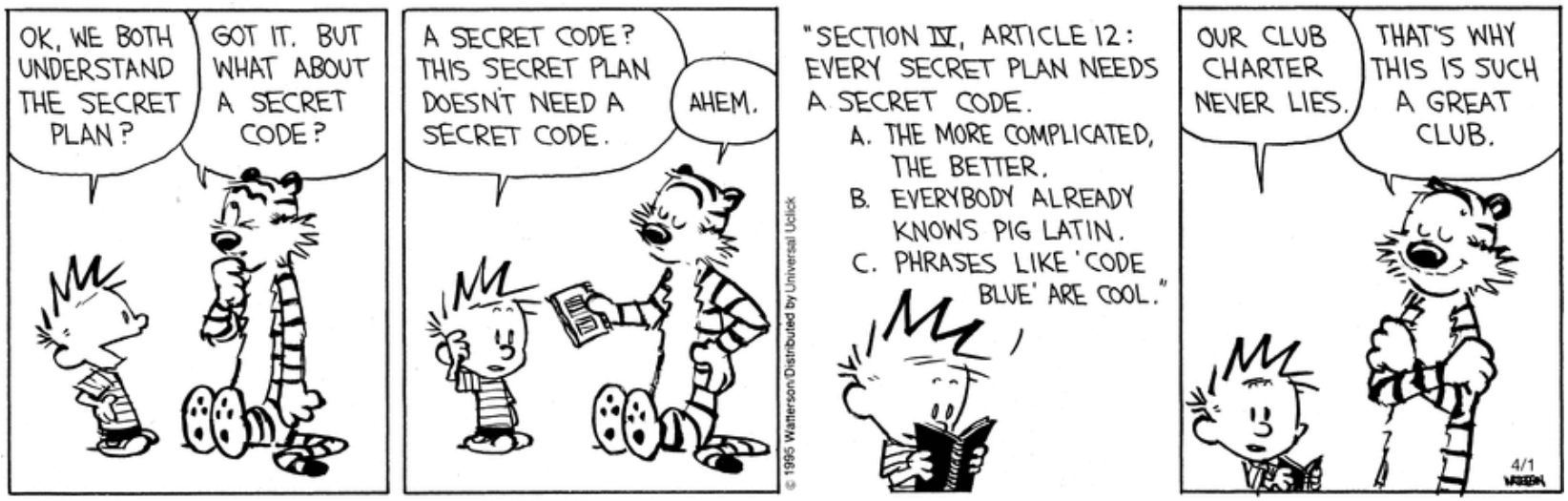

Figure – Watterson, 1995

Quels types de communications ?

Quelques exemples :

- Communication orale
  - Téléphone (filaire, GSM, satellite, VoIP...)
  - Radio
- Communication écrite
  - Lettres papier ; cartes postales
  - Email
  - SMS
  - Applications de messagerie instantanée
- Communication « temporelle »
  - Stockage, archivage (numérique ou non)

Sécurisation des communications : comment ?

Démarche à (très) haut niveau :

1. Identifier les propriétés de sécurité souhaitées
2. Identifier les adversaires potentiels et leurs capacités

($\rightsquigarrow$ définitions de sécurité)

3. Se débarrasser des adversaires et/ou développer des systèmes de protection

Objectif du cours : découvrir et illustrer cette démarche en *cryptographie* (généralement nécessaire mais pas suffisant pour sécuriser les communications au sens large)

Quelles propriétés ?

Parmi les plus courantes :

- *Confidentialité* (≈ les adversaires n'apprennent rien sur le contenu de mes communications)
  - Exemple : seule la personne à qui je destine cette photo :

    

    doit savoir qu'il s'agit d'une martre des pins
- *Preuve d'identité* (≈ c'est moi !)
  - Exemple : j'habite dans cet immeuble, et je veux rentrer dans le hall ; c'est mon ordinateur et je veux m'y connecter
- *Authentification* (≈ c'est moi, et j'approuve ce message)
  - Exemple : je possède ce compte bancaire, et je veux autoriser ce paiement

Quels adversaires ?

Parmi les plus courants :

- Adversaires passifs (voient seulement passer les communications)
- Adversaires actifs « boîte noire » (— ; peuvent bloquer des messages ; peuvent injecter des messages)
- Adversaires actifs « boîte grise » (— ; ont accès à des données physiques relatives au système de communication (par ex. le temps qui s'est écoulé entre la réception de deux messages ; le rayonnement électromagnétique émis par un processeur...))
- Adversaires actifs « boîte grise » + fautes (— ; peuvent injecter des erreurs de calcul dans le système de communication)

Quelques exemples (1)

Téléphone (confidentialité) :

- filaire (commercial) : pas de confidentialité v. adversaires passifs
- GSM : confidentialité* entre le téléphone et l'antenne relais v. adversaires passifs, généralement pas de confidentialité au delà, menant par ex. à :
  - Interception de communications de soldats russes (utilisant le réseau GSM classique à cause de défaillances des systèmes militaires) au début de l'invasion de l'Ukraine de 2022
  - Scandale des écoutes téléphoniques par des tabloïds britanniques
  - Attaques actives avec des *IMSI catchers*

Quelques exemples (2)

Radio (confidentialité) :

- PMR446 : pas de confidentialité v. adversaires passifs
- TETRA : confidentialité v. adversaires passifs/actifs... en fonction des versions, cf. par ex. : `https://www.zetter-zeroday.com/p/interview-with-the-etsi-standards`

Radio (identification/authentification) :

- Tags RFID @125 KHz : aucune sécurité v. adversaire passif ; facilement copiable
- Tags « NFC » @13.56 MHz : sécurité dépendant du protocole ; pas toujours facilement copiable
- Système IFF (*identification friend or foe*) : (a priori) bonne sécurité

Quelques exemples (3)

Trafic réseau :

- HTTP « basique » : pas de confidentialité v. adversaires passifs
- HTTP + TLS 1.3 (« HTTPS ») : (a priori) confidentialité v. adversaires actifs
- Telnet : pas de confidentialité v. adversaires passifs ; pas de preuve d'identité
- SSH : (a priori) confidentialité v. adversaires passifs/actifs ; preuve d'identité

Quelques exemples (4)

Communication écrite :

- Lettres papier ; cartes postales : pas de confidentialité v. adversaires passifs
- Email : pas de confidentialité v. adversaires passifs
- SMS : cf. GSM
- Protocole Signal (implémenté par Signal, WhatsApp… ; permet aussi des communications orales) : confidentialité v. adversaires passifs/actifs

Quelques exemples (5)

Paiement avec carte bancaire :

- Uniquement avec le numéro : aucune sécurité ; facilement copiable
- Avec la bande magnétique : aucune sécurité ; facilement copiable
- Sans contact : cf. NFC
- Avec la puce + PIN : (a priori) authentification v. adversaires actifs « boîte grise »

Quelques exemples (6)

Stockage :

- Disque dur *non chiffré* : pas de confidentialité v. adversaires passifs
- Disque dur *chiffré* : (a priori) confidentialité v. adversaires passifs/actifs
- Mots de passe (par ex. des comptes d'un site web) stockés *en clair* : aucune sécurité v. adversaire passif/actif
- Mots de passe stockés en utilisant une *fonction de hachage de mot de passe* : (a priori) preuve d'identité v. adversaire passif/actif

Un premier bilan

- Pas ou peu de sécurité pour les systèmes « historiques »
  - Mais parfois possible d'en ajouter à un niveau plus élevé dans la couche protocolaire, par ex. écrire une lettre en texte chiffré ; utiliser SSL/TLS ($\approx$ niveau 6 du modèle OSI ; premières versions $\approx$ '95s) au dessus d'une connexion TCP ($\approx$ niveau 4 ; première version '74)
- Beaucoup de systèmes de la vie courante utilisent (ou pourraient utiliser) des mécanismes de sécurisation, *pour des propriétés variées, contre des adversaires variés*
- $\rightsquigarrow$ besoin d'une approche rigoureuse du problème, avec des *définitions* communes
  - Favorise l'efficacité (via la réutilisation de concepts bien établis ; de systèmes standardisés...)
  - Favorise la sécurité (via la réutilisation de concepts bien établis ; de systèmes standardisés...)

De quoi parlons nous ?

Introduction aux définitions

Quantifier la sécurité

Définitions, définitions, définitions

Quelle place pour les définitions en cryptographie ?

- Objectif final : définir *formellement* des objectifs de sécurité relativement à des modèles d'adversaire (cf. ci-dessus)
  - Formellement : dans un cadre axiomatique/logique rigoureux (en pratique, en utilisant une approche maths + info)
- Intérêts d'une approche formelle : précise ; peu ambiguë ; permet de « prouver des choses »
- Inconvénients — : pas toujours facile de capturer formellement une intuition $\rightsquigarrow$ parfois dur à interpréter ; « rajoute » du travail

$\rightsquigarrow$ L'approche largement dominante en cryptographie (moderne)

Quelle place pour les définitions en cryptographie (bis) ?

- Objectif intermédiaire : favoriser la réutilisabilité des définitions ; des systèmes
  - $\rightsquigarrow$ Définir des *primitives* et des objectifs de sécurité associés
  - $\rightsquigarrow$ Prouver des *réductions* entre définitions

Les définitions et les preuves en cryptographie : essentiel mais pas facile

Difficultés potentielles :

- Maîtriser le formalisme (algorithmes/circuits probabilistes (avec oracle) ; réductions algorithmiques ; probabilités...)
- Identifier la « bonne » définition (quel objectif ; quels adversaires ?)
  - Ne pas se tromper de niveau (fonctionnalité finale ou primitive ?)
  - — puissance d'adversaire (passif ou actif ? avec accès physique ou non ?)
- Comprendre la garantie offerte par une *preuve de sécurité* et ses limites
  - Les preuves sont *toujours* limitées par un modèle ; des hypothèses plus ou moins explicites

Vers une première définition de sécurité, pas à pas

Objectif : définir une notion de confidentialité des communications, pour le cadre informel suivant :
Deux personnes ont à leur disposition :

- Un canal de communication non sûr (sur lequel peut agir un adversaire)

et souhaitent :

- Échanger une grande quantité de données (par ex. plein de petits messages ; un très grand message...) sans que *ces échanges* donnent de l'« information » à l'adversaire

Notation ; vocabulaire

Système de chiffrement

Un *système de chiffrement* $\mathrm{Enc} : \mathcal{M} \to \mathcal{C}$ est une application bijective qui à chaque (message) *clair* $m \in \mathcal{M}$ associe un (message) *chiffré* $c \in \mathcal{C}$

Remarques :

- On note Enc^{-1} l'application inverse : $\forall m \in \mathcal{M}$, $\mathrm{Enc}^{-1}(\mathrm{Enc}(m)) = m$
- Un système de chiffrement prend en général un ou plusieurs arguments supplémentaires, cf. plus tard
- La plupart du temps, $\mathcal{M} \approx \mathcal{C} \approx \{0,1\}^*$, mais ce n'est pas toujours le cas

$\rightsquigarrow$ On souhaite définir la confidentialité d'un système de chiffrement Enc quelconque

#1 Modèle d'adversaire : quel pouvoir lui donner ?

Si l'adversaire n'a aucun pouvoir : plus de problème, mais ce n'est plus de la cryptographie

- *Possiblement* raisonnable : peut-être okay de stocker un message en clair s'il est dans un coffre-fort enterré dans une forêt reculée ? peut-être okay de considérer que l'adversaire n'a pas un accès physique à mon serveur qui tourne dans un data-centre à accès contrôlé ?
- Mais généralement non raisonnable, en particulier si aucune précaution particulière n'a été prise
- **ATTENTION** : l'adversaire est souvent bien moins limité que vous ! (Peut avoir un Flipper Zero (`https://flipperzero.one/`) ; un IMSI catcher ; une grosse puissance de calcul...)

Modèle d'adversaire (bis)

Quelle *information* disponible, quelles capacités ?

- Le « plus simple » : adversaire passif : voit passer les communications sur le canal, et *est capable de demander à connaître un chiffré correspondant à un clair donné*
  - Vocabulaire : *attaque (passive) à clair choisi*
  - **ATTENTION** : adversaire très faible, peu réaliste (mais c'est un début !)

Quelle *puissance de calcul* ?

- Le « plus simple » : temps de calcul et mémoire illimitées (non bornées)
  - Adversaire très fort, peu réaliste (mais c'est un début !)

Modèle d'adversaire (ter)

Attaque à clair choisi : pourquoi ?

- Modélise la connaissance / le contrôle que l'adversaire peut avoir sur une *partie* d'un message
- Hypothèse réaliste : peut s'implémenter par l'observation de l'environnement, le contrôle de champs d'un protocole, etc. (cf. ci-dessous)
- On peut néanmoins considérer des modèles plus faibles (rarement le cas) :
  - Clair (seulement) connu
  - Chiffré seul

Clair choisi : une illustration

Figure – `https://xkcd.com/257/`

Observation intermédiaire

- Un système de chiffrement qui à un même message associe un unique chiffré sera a priori non confidentiel pour un adversaire à clair choisi/connu

$\rightsquigarrow$ Une protection « probabiliste »

- On souhaite généralement « augmenter » un système pour que plusieurs chiffrés correspondent à un même clair
- On choisit généralement le chiffré à faire correspondre de façon aléatoire
- (L'aléatoire joue un rôle capital en cryptographie)

#2 Confidentialité

Idées :

- Idéalement, la seule information connue de l'adversaire doit provenir des réponses à ses *requêtes*
- L'observation d'un chiffré quelconque ne doit pas (trop) modifier les « connaissances a priori » de l'adversaire
- La quantité « minimale » d'information est le bit
- Un adversaire capable de *distinguer* deux possibilités (0/1) pour un clair étant donné son chiffré a appris un bit d'information grâce à celui-ci (et c'est déjà trop)

Confidentialité (bis)

Vers un *jeu de sécurité* : *indistinguishability for chosen-plaintext attacks* (IND-CPA)

Jeu IND-CPA

1. L'adversaire peut acquérir de l'information sur Enc en faisant des requêtes à clair choisi
2. Une fois cet entraînement terminé, il construit et soumet deux messages de *challenge* m_0 et m_1 de même longueur, et il reçoit le chiffré $c_b := \mathrm{Enc}(m_b)$ de l'un d'entre eux (où b vaut 0 ou 1 avec probabilité $1/2$)
3. L'adversaire essaye de deviner b : il renvoie $\hat{b}$ et *gagne* le jeu ssi. $b = \hat{b}$

Confidentialité (ter)

Remarques :

- Le jeu IND-CPA est *probabiliste* : le bit de challenge b est tiré aléatoirement (suivant une distribution uniforme) ; Enc peut être probabiliste (cf. ci-dessus) ; l'adversaire lui-même peut être probabiliste (pourquoi pas ?)
- $\rightsquigarrow$ la *probabilité de succès* d'un adversaire (calculée sur tous les tirages ci-dessus) est la notion pertinente à regarder
- Mais c'est facile de gagner avec probabilité $1/2$ (Q : donnez un exemple d'adversaire ?) $\rightsquigarrow$ un « bon » adversaire gagne avec une probabilité « loin » de $1/2$
- $\rightsquigarrow$ l'*avantage* associé à une probabilité p est $|2p - 1|$ (ou parfois $|p - 1/2|$) (Q : pourquoi une valeur absolue ?)

Pour résumer

- L'efficacité d'*un* adversaire (contre la confidentialité...) pour Enc peut se par ex. se mesurer par son avantage dans le jeu IND-CPA
- (Mais il peut y avoir des adversaires plus ou moins malins...)
- $\rightsquigarrow$ La sécurité (en confidentialité...) de Enc peut par ex. se mesurer comme l'avantage IND-CPA du *meilleur* adversaire possible

Pour avoir vraiment fini :

- Prendre en compte la quantité d'information utilisée par les adversaires
- (Tout de même) — les ressources de calcul —

Ressources d'un adversaire

Plusieurs approches possibles, notamment :

- Seulement considérer les adversaires à ressources « limitées » (pour une certaine définition) (approche « asymptotique »)
- Ne pas contraindre les ressources a priori, mais définir la sécurité pour chaque quantité de ressources (approche « concrète » ; bien meilleure)
  - (Approche non-uniforme : on maximise pour chaque « taille » d'entrée)

Au final

$\mathbf{Adv}^{\text{IND-CPA}}(q, t)$

$$\mathbf{Adv}^{\text{IND-CPA}}_{\text{Enc}}(q, t) := \max_{A_{q,t}} \Big| 2\Pr[A_{q,t} \text{ gagne le jeu IND-CPA contre Enc}] - 1 \Big|$$

$A_{q,t}$: un adversaire qui fait q requêtes d'entraînement et qui tourne en temps t

Remarques :

- L'unité de mesure du temps est généralement donnée par le contexte, souvent prise comme le temps pour calculer Enc (somme toute peu importante)
- La *mémoire* utilisée par les adversaires n'est généralement pas prise en compte (même si cela pourrait être bien de le faire...)

De quoi parlons nous ?

Introduction aux définitions

Quantifier la sécurité

Niveau de sécurité

On souhaite souvent résumer une fonction comme $\mathbf{Adv}^{\text{IND-CPA}}_{\text{Enc}}(q, t)$ par un unique nombre, son *niveau de sécurité* κ, exprimé en bits
Une définition courante : $\kappa := \log(t_{\text{min}})$ pour t_{min} le temps t minimum t.q. $\mathbf{Adv}^{\text{IND-CPA}}_{\text{Enc}}(\infty, t) \geq c$ avec c une constante (par ex. 2/3)

ATTENTION

- $\rightsquigarrow$ Entraîne très souvent une perte d'information
- Pas la seule définition possible
  - Alternative : $\kappa' := -\log(\mathbf{Adv}^{\text{IND-CPA}}_{\text{Enc}}(\infty, 1))$; on a souvent $\kappa \neq \kappa'$!

Niveau de sécurité : quelques ordres de grandeur

Pour une fonction « peu chère », quelles ressources nécessaires pour l'appeler 2^t fois, pour $t = \cdots$

- $\approx 40 \rightsquigarrow$ faisable sur un bon téléphone en quelques semaines
- $\approx 50 \rightsquigarrow$ faisable sur un bon PC en quelques mois
- $\approx 60 \rightsquigarrow$ faisable sur un gros cluster de CPUs/GPUs
  - Ordre de grandeur des records de calcul en cryptographie académique
- $\approx 80/90 \rightsquigarrow$ faisable sur un gros cluster de circuits dédiés
  - Exemple : minage de bitcoins

Quelques ordres de grandeur (bis)

Objectif : évaluer une fonction 2^{128} fois en 34 ans ($\approx 2^{30}$ secondes), en supposant :

- Du matériel permettant 2^{50} évaluations/s (plutôt rapide)
- Consommant 1000 W, en ignorant les surcoûts (plutôt faible)

et en ignorant le coûts de la parallélisation $\Rightarrow$

- $2^{128-50-30} \approx 2^{48}$ machines nécessaires
- $\approx 280\,000\,000$ GW de nécessaires
  - ≥ 30 MW par personne sur Terre !
  - Pics de consommation d'électricité en France ≈ 80 GW

$\rightsquigarrow$ Physiquement peu probable

$\rightsquigarrow$ 128 bits $\approx$ le **niveau de sécurité** minimum typiquement conseillé (mais attention aux détails !)

Avantage : quelques ordres de grandeur

Avantage $\varepsilon \rightsquigarrow p_{\text{succ}} = (\varepsilon + 1)/2 = (\varepsilon^{-1} + 1)/(2\varepsilon^{-1}) \rightsquigarrow$ faire mieux qu'un choix constant une fois sur $2\varepsilon^{-1}$

À titre de comparaison, l'intervalle estimé (en secondes) entre deux impacts de NEOs est de :

- $\approx 2^{35}$ pour un impact de 10 à 100 équivalent mégatonne de TNT (peut détruire une ville)
- $\approx 2^{39}$ — 1000 à 100000 — (peut détruire un petit pays)
- $\approx 2^{45}$ — 10^6 à 10^7 — (peut détruire un grand pays ; conséquences planétaires)
- $\approx 2^{52}$ — 10^8 à 10^9 — (extinction globale)

Source : Report of the Task Force on potentially hazardous NEAR EARTH OBJECTS, British National Space Centre (2000)

(Ce ne sont *pas* des (inverses de) probabilités (mais possiblement des (inverses de) paramètres pour des lois de Poisson modélisant l'occurrence de ces phénomènes))

Avantage : interprétation d'un avantage faible

Attention :

- On peut souvent *amplifier* l'avantage d'un adversaire donné en investissant plus de ressources
- Mais dans ce cas ce n'est plus le même adversaire
- Un adversaire avec avantage faible doit être considéré pour ce qu'il est, pas plus