Introduction à la sécurité informatique - 1
Pourquoi la cryptographie ne suffit pas
Contraintes absolues

F. Prost
Frederic.Prost@univ-grenoble-alpes.fr

M1 MEEF - parcours NSI - UGA

11 Janvier 2023

Privacité et technologie numérique

- Historiquement une des premières applications de l'informatique: Enigma, Fish, etc.
- La privacité, et la sécurité informatique impactent *toutes les branches* de la science informatique allant de la théorie jusqu'aux plus petits détails techniques (y compris au niveau physique).
- Les technologies digitales ont changé le monde:
  - Ce qu'on voit : la forme de bâtiments...
  - Ce qu'on ne voit pas : Pre-suasion, psychologie sociale et big data...

La chute du jardin d'Eden et internet

Virtualisation

- Réalité, monde virtuel et leurs interactions.

  $\Longrightarrow$ problèmes liés à l'union hypostatique, schisme entre les églises d'orient et d'occident.

- Question philosophique compliquée ayant d'énormes répercussions : par exemple le dessin des pays aujourd'hui...

Sécurité: Une science très contre-intuitive

- Si vis pacem, para bellum. (sécurité informatique et sécurité en travaux publics sont des notions fondamentalement différentes).
- "Strategy: The Logic of War and Peace" (E.N. Luttwak).

Sécurité: Une science très contre-intuitive

- Si vis pacem, para bellum. (sécurité informatique et sécurité en travaux publics sont des notions fondamentalement différentes).
- "Strategy: The Logic of War and Peace" (E.N. Luttwak).

$\Longrightarrow$ Affaire des écoutes grecques (2006).

Sécurité: Une science très contre-intuitive

- Si vis pacem, para bellum. (sécurité informatique et sécurité en travaux publics sont des notions fondamentalement différentes).
- "Strategy: The Logic of War and Peace" (E.N. Luttwak).

$\Longrightarrow$ Affaire des écoutes grecques (2006).

- Cela va à l'encontre de la pensée "ingénieure" habituelle : Programming Satan's Computer [Anderson and Needham, 1995]!

Sécurité: Une science très contre-intuitive

- Si vis pacem, para bellum. (sécurité informatique et sécurité en travaux publics sont des notions fondamentalement différentes).
- "Strategy: The Logic of War and Peace" (E.N. Luttwak).

$\Longrightarrow$ Affaire des écoutes grecques (2006).

- Cela va à l'encontre de la pensée "ingénieure" habituelle : Programming Satan's Computer [Anderson and Needham, 1995]!
- La complexité est le talon d'Achille's : mots de passes et gyroscopes sur smartphones, réseaux aériens (haut-parleurs plus micros) ...

Sécurité: Une science très contre-intuitive

- Si vis pacem, para bellum. (sécurité informatique et sécurité en travaux publics sont des notions fondamentalement différentes).
- "Strategy: The Logic of War and Peace" (E.N. Luttwak).

$\Longrightarrow$ Affaire des écoutes grecques (2006).

- Cela va à l'encontre de la pensée "ingénieure" habituelle : Programming Satan's Computer [Anderson and Needham, 1995]!
- La complexité est le talon d'Achille's : mots de passes et gyroscopes sur smartphones, réseaux aériens (haut-parleurs plus micros) ...
- Chaque solution de sécurité est un compromis.

La Cryptographie n'est pas suffisante: you can run but you can't hide

- La vie rêvée des mathematiciens vs. la realité crue.

La Cryptographie n'est pas suffisante: you can run but you can't hide

- La vie rêvée des mathematiciens vs. la realité crue.
- Les détails d'implantations comptent.

La Cryptographie n'est pas suffisante: you can run but you can't hide

- La vie rêvée des mathematiciens vs. la realité crue.
- Les détails d'implantations comptent.
- Les protocoles d'utilisations comptent.

La Cryptographie n'est pas suffisante: you can run but you can't hide

- La vie rêvée des mathematiciens vs. la realité crue.
- Les détails d'implantations comptent.
- Les protocoles d'utilisations comptent.
- La psychologie compte.

La Cryptographie n'est pas suffisante: you can run but you can't hide

- La vie rêvée des mathematiciens vs. la realité crue.
- Les détails d'implantations comptent.
- Les protocoles d'utilisations comptent.
- La psychologie compte.
- La complexité des systèmes compte.

La Cryptographie n'est pas suffisante: you can run but you can't hide

- La vie rêvée des mathematiciens vs. la realité crue.
- Les détails d'implantations comptent.
- Les protocoles d'utilisations comptent.
- La psychologie compte.
- La complexité des systèmes compte.
- Et parfois juste la chance (ou pas de chance)...

La Cryptographie n'est pas suffisante: you can run but you can't hide

- La vie rêvée des mathematiciens vs. la realité crue.
- Les détails d'implantations comptent.
- Les protocoles d'utilisations comptent.
- La psychologie compte.
- La complexité des systèmes compte.
- Et parfois juste la chance (ou pas de chance)...

$\Longrightarrow$ Preuve empirique: les révélations de Snowden sur les pratiques de la NSA...

Le piratage en deux images - 1

Mini Dry Erase Whiteboard

(Diamond Chicken Burger) We relabeled our Diamond Chicken...

$11.00

Ergonomic Aluminum Laptop Stand

(Double Your Fortune) The Double Your Fortune is now a Laptop Stand...

$12.50

Wired Earphones with Mic

(Emerald Veggie Burger) Been working through lunch or dinner?...

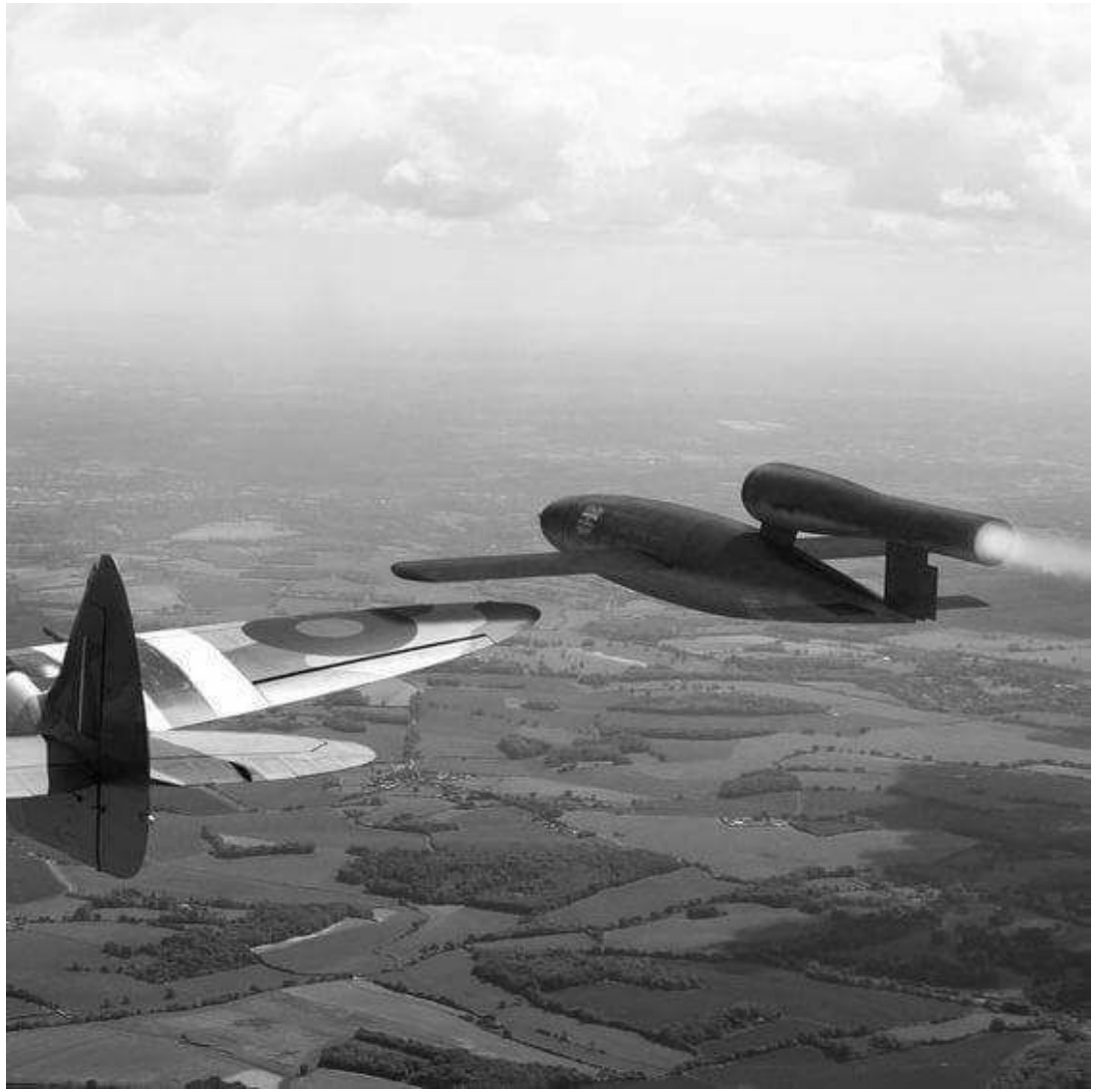
$10.00

Silicone Keyboard Cover

(BYO Burger) Build your own burger – but make the boss pay. This...

$10.00

Le piratage en deux images - 2



Plan

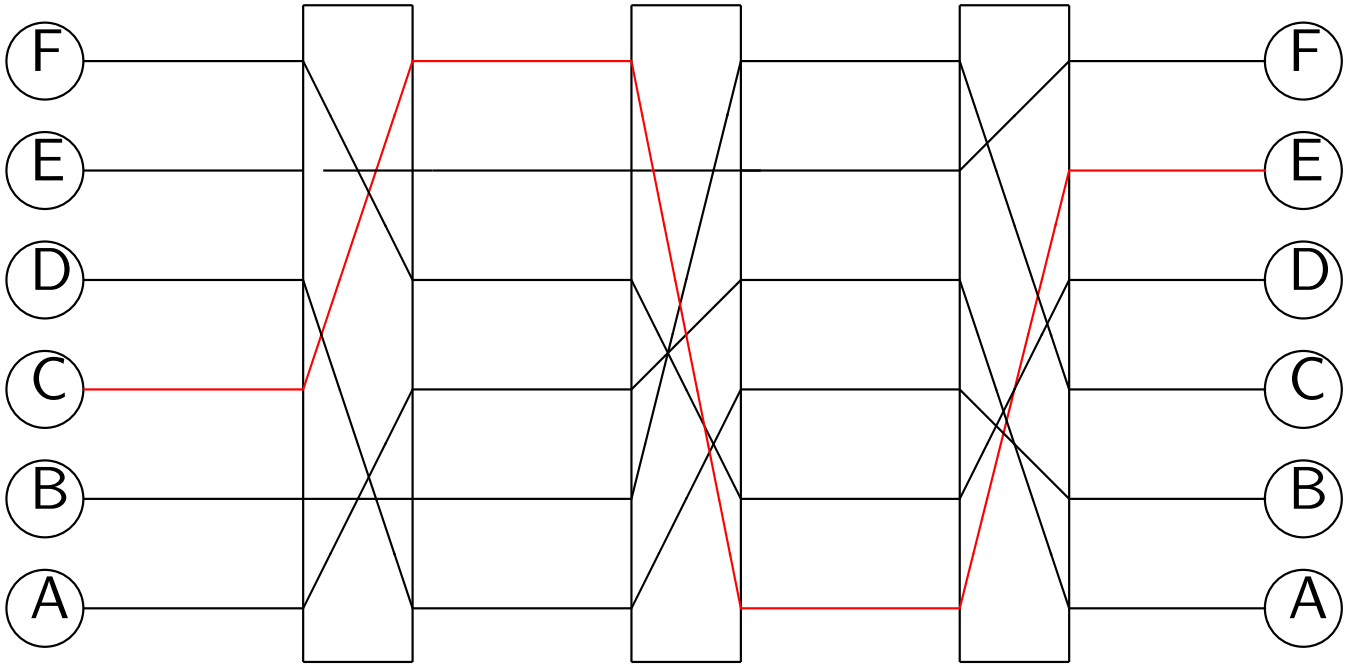
Cryptanalyse d'Enigma

- Exemple à la fois extrême et venant de la vie réelle de la difficulté qu'il y a à garder l'information confidentielle.
  Les historiens estiment que l'effet de cette cryptanalyse a été de réduire la deuxième guerre mondiale de 1 à 2 ans (littéralement des dizaines de millions de vies).
- Première mécanisation de cryptanalyse: marque le passage de la linguistique aux mathématiques/informatique. Première utilisation d'un ordinateur !
- A. Turing, un des pères fondateurs de l'informatique, était impliqué.
- Exemplaire dans la manière de casser l'"incassable"... à se rappeler quand vous entendrez "le système est inviolable"

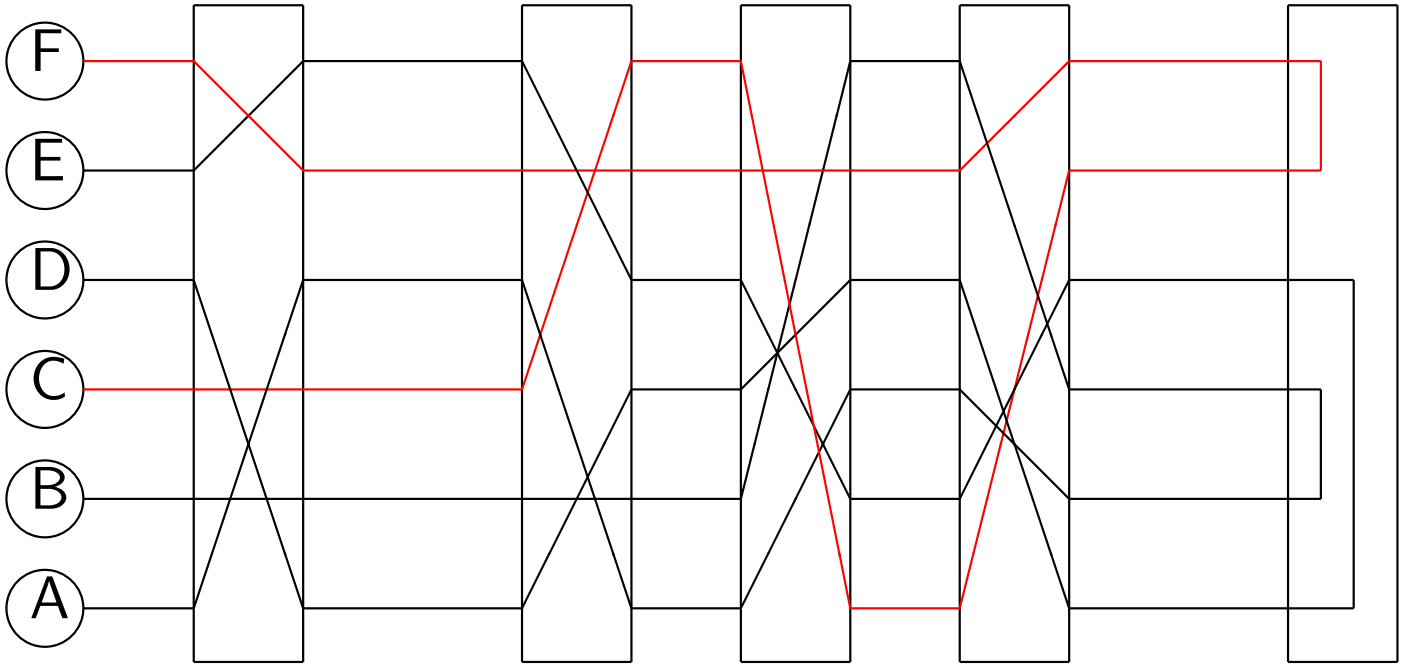
  $\Longrightarrow$ Think outside the box !

La machine Enigma

Rotors de permutations Schématiquement



Enigma Schématiquement



Protocole d'utilisation d'Enigma

- Livre des clefs:

Date	Rotor	Initialisation	Interconnexion
12	I II III	REZ	FD IZ LP MN TA SY
13	II V I	KXU	AN GZ ID LW MF UY
14	IV II III	WGT	ET IL MO NS WH BQ
15	II I V	AQR	UI YS AN MJ VB EH
...	...	...	...

- Une clef définit la configuration initiale de la machine.
- Une fois la machine initialisée par l'opérateur, ce dernier envoit trois lettres qui servent à initier une clef de session (pour éviter les répétitions sur les messages du jour). Ce groupe de trois lettres est répété deux fois puis les rotors étaient ré-arrangés suivant ce code.

Quelque chiffres sur Enigma

- Chaque rotor à 26 positions possibles: $26^3 = 17576$
- On choisit 3 rotors parmi 5 possibles: $5 * 4 * 3 = 60$
- Tableau d'interconnexion avec, 6 cables:

$$\Pi_{k=0}^{5} \frac{(26-2k)!}{2 \times (26-4k)!} = 72282089880000$$

- Nombre d'initialisations différentes: approximativement 76×10^{18}
  Age de l'univers en secondes: $4,3 \times 10^{17}$
- La force d'Enigma est due à la combinaison qui évite les répétitions (rotor) et l'énorme espace des clefs (interconnections).
- Même avec une copie de la machine impossible à casser "en force".

Faiblesses d'Enigma

- Faiblesses internes (faiblesse de l'algorithme):
  - Juste des substitutions involutives sont implantées: de $26! \simeq 403 \times 10^{24}$ à 533×10^{12} (soit une réduction de $7,5 \times 10^{11}$!!).

Faiblesses d'Enigma

- Faiblesses internes (faiblesse de l'algorithme):
  - Juste des substitutions involutives sont implantées: de $26! \simeq 403 \times 10^{24}$ à 533×10^{12} (soit une réduction de $7,5 \times 10^{11}$!!).
  - A cause du réflecteur une lettre ne peut jamais être encodée par elle même.

    $\Longrightarrow$ Des tests de communication par les troupes allemandes ne comportaient que des "T's" (permet une attaque en clair si on ne voit pas de T).

    $\Longrightarrow$ crib technique developpée par Turing.

Faiblesses d'Enigma

- Faiblesses internes (faiblesse de l'algorithme):
  - Juste des substitutions involutives sont implantées: de $26! \simeq 403 \times 10^{24}$ à 533×10^{12} (soit une réduction de $7,5 \times 10^{11}$!!).
  - A cause du réflecteur une lettre ne peut jamais être encodée par elle même.

    $\Longrightarrow$ Des tests de communication par les troupes allemandes ne comportaient que des "T's" (permet une attaque en clair si on ne voit pas de T).

    $\Longrightarrow$ crib technique developpée par Turing.
- Faiblesses externes (protocol d'utilisation) :
  - Les allemands interdisaient l'utilisation du même rotor à la même place deux jours de suite.

Faiblesses d'Enigma

- Faiblesses internes (faiblesse de l'algorithme):
  - Juste des substitutions involutives sont implantées: de $26! \simeq 403 \times 10^{24}$ à 533×10^{12} (soit une réduction de $7,5 \times 10^{11}$!!).
  - A cause du réflecteur une lettre ne peut jamais être encodée par elle même.

    $\Longrightarrow$ Des tests de communication par les troupes allemandes ne comportaient que des "T's" (permet une attaque en clair si on ne voit pas de T).

    $\Longrightarrow$ crib technique developpée par Turing.
- Faiblesses externes (protocol d'utilisation) :
  - Les allemands interdisaient l'utilisation du même rotor à la même place deux jours de suite.
  - La répétition de la clef de session au début du message.

Faiblesses d'Enigma

- Faiblesses internes (faiblesse de l'algorithme):
  - Juste des substitutions involutives sont implantées: de $26! \simeq 403 \times 10^{24}$ à 533×10^{12} (soit une réduction de $7,5 \times 10^{11}$!!).
  - A cause du réflecteur une lettre ne peut jamais être encodée par elle même.

    $\Longrightarrow$ Des tests de communication par les troupes allemandes ne comportaient que des "T's" (permet une attaque en clair si on ne voit pas de T).

    $\Longrightarrow$ crib technique developpée par Turing.
- Faiblesses externes (protocol d'utilisation) :
  - Les allemands interdisaient l'utilisation du même rotor à la même place deux jours de suite.
  - La répétition de la clef de session au début du message.
  - Certains messages avaient une structure prédictible: typiquement les bulletins météos de la Luftwaffe.

Faiblesses d'Enigma

- Faiblesses internes (faiblesse de l'algorithme):
  - Juste des substitutions involutives sont implantées: de $26! \simeq 403 \times 10^{24}$ à 533×10^{12} (soit une réduction de $7,5 \times 10^{11}$!!).
  - A cause du réflecteur une lettre ne peut jamais être encodée par elle même.

    $\Longrightarrow$ Des tests de communication par les troupes allemandes ne comportaient que des "T's" (permet une attaque en clair si on ne voit pas de T).

    $\Longrightarrow$ crib technique developpée par Turing.
- Faiblesses externes (protocol d'utilisation) :
  - Les allemands interdisaient l'utilisation du même rotor à la même place deux jours de suite.
  - La répétition de la clef de session au début du message.
  - Certains messages avaient une structure prédictible: typiquement les bulletins météos de la Luftwaffe.
  - Biais d'opérateur: toujours les mêmes trois lettres de clef de session (prénom de la fiancée...)

Marjan Rejevski

- Par espionnage les français ont eu une copie d'un e machine Enigma. Donnée à la Pologne (alliée) dans les années 30.

Marjan Rejevski

- Par espionnage les français ont eu une copie d'un e machine Enigma. Donnée à la Pologne (alliée) dans les années 30.
- Marjan Rejevsky était un jeune mathématicien qui a trouvé une manière d'exploiter les faiblesses du protocole (répétition de la clef de session).

Marjan Rejevski

- Par espionnage les français ont eu une copie d'un e machine Enigma. Donnée à la Pologne (alliée) dans les années 30.
- Marjan Rejevsky était un jeune mathématicien qui a trouvé une manière d'exploiter les faiblesses du protocole (répétition de la clef de session).

$\Longrightarrow$ On sait que le premier et quatrième symbole sont les mêmes (pareil pour 2/5 et 3/6).

Marjan Rejevski

- Par espionnage les français ont eu une copie d'un e machine Enigma. Donnée à la Pologne (alliée) dans les années 30.
- Marjan Rejevsky était un jeune mathématicien qui a trouvé une manière d'exploiter les faiblesses du protocole (répétition de la clef de session).

$\Longrightarrow$ On sait que le premier et quatrième symbole sont les mêmes (pareil pour 2/5 et 3/6).

- En compilant tous les messages échangés dans une journée on peut construire un alphabet de correspondance:

Première Lettre	ABCDEFGHIJKLMNOPQRSTUVWXYZ
Quatrième Lettre	XFEARBSLHQIGCVDZWKMNJUOYTP

- Cette table est indépendante du tableau d'interconnections.

Les cycles de Rejevski's

- Etant donnée un alphabet on peut le factoriser en cycles.
- Par exemple dans :

Première Lettre	ABCDEFGHIJKLMNOPQRSTUVWXYZ
Quatrième Lettre	XFEARBSLHQIGCVDZWKMNJUOYTP

Se trouvent les cycles

$A \to X \to Y \to T \to N \to V \to U \to J \to Q \to W \to O \to D \to A$

$B \to F \to B$

$C \to E \to R \to K \to I \to H \to L \to G \to S \to M \to C$

$P \to Z \to P$

Les cycles de Rejevski's

- Etant donnée un alphabet on peut le factoriser en cycles.
- Par exemple dans :

Première Lettre	ABCDEFGHIJKLMNOPQRSTUVWXYZ
Quatrième Lettre	XFEARBSLHQIGCVDZWKMNJUOYTP

Se trouvent les cycles

$A \to X \to Y \to T \to N \to V \to U \to J \to Q \to W \to O \to D \to A$

$B \to F \to B$

$C \to E \to R \to K \to I \to H \to L \to G \to S \to M \to C$

$P \to Z \to P$

- Il se trouve que cette décomposition est unique vis-à-vis de la configuration initiale des rotors ! Comme un code ADN.

Les cycles de Rejevski's

- Etant donnée un alphabet on peut le factoriser en cycles.
- Par exemple dans :

Première Lettre	ABCDEFGHIJKLMNOPQRSTUVWXYZ
Quatrième Lettre	XFEARBSLHQIGCVDZWKMNJUOYTP

Se trouvent les cycles

$A \to X \to Y \to T \to N \to V \to U \to J \to Q \to W \to O \to D \to A$

$B \to F \to B$

$C \to E \to R \to K \to I \to H \to L \to G \to S \to M \to C$

$P \to Z \to P$

- Il se trouve que cette décomposition est unique vis-à-vis de la configuration initiale des rotors ! Comme un code ADN.

$\Longrightarrow$ Il suffit de faire un grand livre avec toutes les possibilités ! $(26^3 \times 60)$ C'est grand mais pas inhummain.

Les cycles de Rejevski's

- Etant donnée un alphabet on peut le factoriser en cycles.
- Par exemple dans :

Première Lettre	ABCDEFGHIJKLMNOPQRSTUVWXYZ
Quatrième Lettre	XFEARBSLHQIGCVDZWKMNJUOYTP

  Se trouvent les cycles

  $A \to X \to Y \to T \to N \to V \to U \to J \to Q \to W \to O \to D \to A$

  $B \to F \to B$

  $C \to E \to R \to K \to I \to H \to L \to G \to S \to M \to C$

  $P \to Z \to P$

- Il se trouve que cette décomposition est unique vis-à-vis de la configuration initiale des rotors ! Comme un code ADN.

$\Longrightarrow$ Il suffit de faire un grand livre avec toutes les possibilités ! $(26^3 \times 60)$ C'est grand mais pas inhummain.

- Ce n'est pas fini: que fait on quant au tableau d'interactions ? (Facile à craquer à la main. Trouvez vous ?)

Automatisation de la Cryptanalyse : A. Turing à Bletchley Park

- En mai 1937 les Allemands ont changé le protocole et l'attaque de Rejevsky n'était plus possible.

Automatisation de la Cryptanalyse : A. Turing à Bletchley Park

- En mai 1937 les Allemands ont changé le protocole et l'attaque de Rejevsky n'était plus possible.
- Turing remarqua alors qu'il y avait des messages similaires (par espionnage et recoupement "usuels"): cel mène à une nouvelle série d'attaque en "texte clair". L'exemple le plus connu est celui des messages météos: le mot *wetter* devait apparaître quelque part dans le message. Turing appelait cela des "cribs".

Automatisation de la Cryptanalyse : A. Turing à Bletchley Park

- En mai 1937 les Allemands ont changé le protocole et l'attaque de Rejevsky n'était plus possible.
- Turing remarqua alors qu'il y avait des messages similaires (par espionnage et recoupement "usuels"): cel mène à une nouvelle série d'attaque en "texte clair". L'exemple le plus connu est celui des messages météos: le mot *wetter* devait apparaître quelque part dans le message. Turing appelait cela des "cribs".
- Supposez que vous sachiez que le message commence avec :

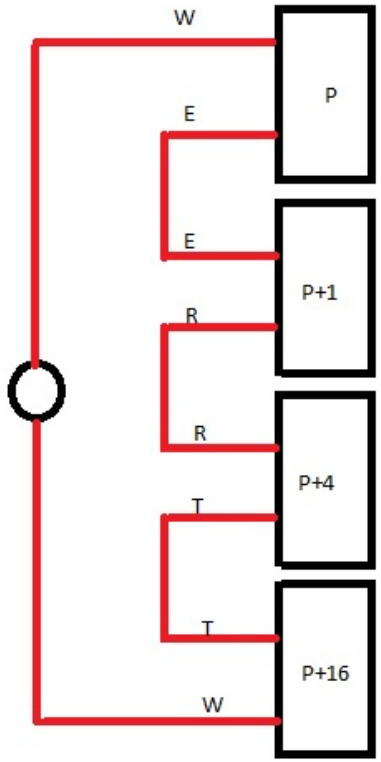
  WETTERUEBERSICHTNULLSECHSNULLNULL
- Considérons le texte crypté suivant :

W	E	T	T	E	R	U	E	B	E	R	S	I	C	H	T
E	R	G	H	W	T	S	S	K	J	F	E	G	L	A	W

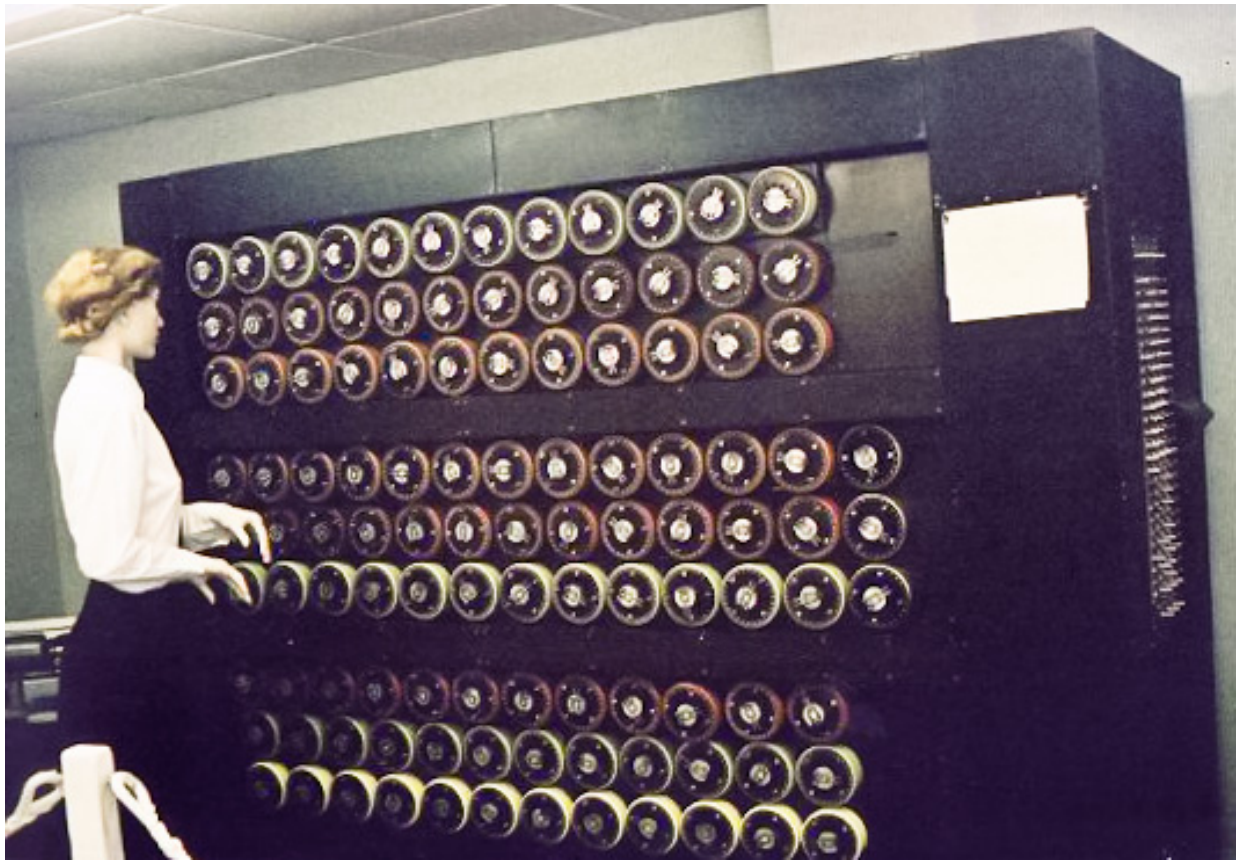
  Il y a un cycle $W \rightarrow^{0} E \rightarrow^{1} R \rightarrow^{4} T \rightarrow^{16} W$ qui dépend de la position initiale (à la Rejevsky).

Cryptanalyse Bombe (schema)

- Comment automatiquement détecter ces cycles ?
- On travaille sur 4 machines en parallèle. En connectant les sorties de l'une sur l'entrée de l'autre et en les initialisants correctement en fonction du crib. Quant elles sont dans le bon état, le circuit se ferme :



Bombe cryptanalytique de Turing



Guerre de l'Information

- Des actions de guerre étaient réalisées pour faire communiquer les allemands.

Guerre de l'Information

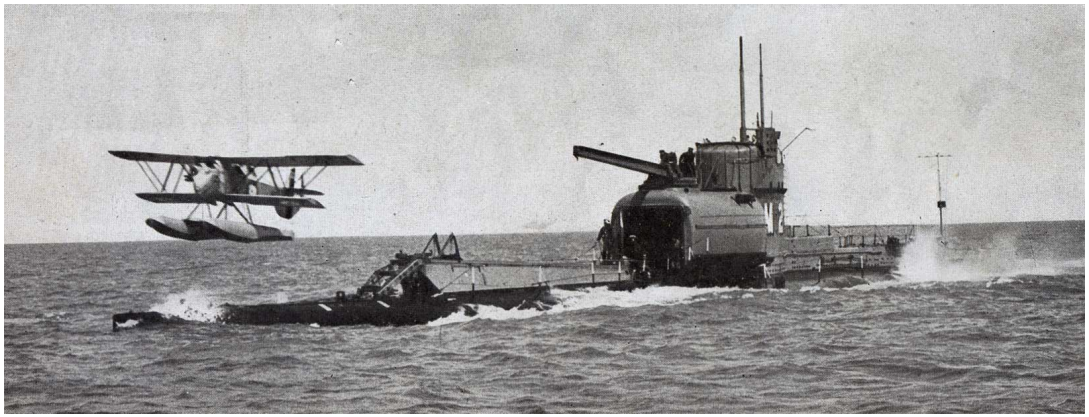
- Des actions de guerre étaient réalisées pour faire communiquer les allemands.
  $\Longrightarrow$ En effet les Alliés connaissaient le système de codage des données géographiques (espionage) ce qui donnait des cribs.

Guerre de l'Information

- Des actions de guerre étaient réalisées pour faire communiquer les allemands.
  $\Longrightarrow$ En effet les Alliés connaissaient le système de codage des données géographiques (espionage) ce qui donnait des cribs.
- Les Alliés savaient où les U-boot étaient, ils auraient pu tous les couler d'un coup... mais les Allemands auraient changé leur cryptosystème. Paradoxe : on a une information mais on ne peut pas s'en servir. Que faire ?

Guerre de l'Information

- Des actions de guerre étaient réalisées pour faire communiquer les allemands.
  $\Longrightarrow$ En effet les Alliés connaissaient le système de codage des données géographiques (espionage) ce qui donnait des cribs.
- Les Alliés savaient où les U-boot étaient, ils auraient pu tous les couler d'un coup... mais les Allemands auraient changé leur cryptosystème. Paradoxe : on a une information mais on ne peut pas s'en servir. Que faire ?



Conclusion

- Les mathématiques du cryptosystème ne sont qu'un aspect parmi d'autres :

Conclusion

- Les mathématiques du cryptosystème ne sont qu'un aspect parmi d'autres :
  - espionage,

Conclusion

- Les mathématiques du cryptosystème ne sont qu'un aspect parmi d'autres :
  - espionage,
  - protocoles,

Conclusion

- Les mathématiques du cryptosystème ne sont qu'un aspect parmi d'autres :
  - espionage,
  - protocoles,
  - implantation practique,

Conclusion

- Les mathématiques du cryptosystème ne sont qu'un aspect parmi d'autres :
  - espionage,
  - protocoles,
  - implantation practique,
  - coup de bol,...

Conclusion

- Les mathématiques du cryptosystème ne sont qu'un aspect parmi d'autres :
  - espionage,
  - protocoles,
  - implantation practique,
  - coup de bol,...
- Il n'y a pas de coïncidence...

Plan

Cas practique de-Anonymization: Netflix début des années 2000

- Résultats frappants [Narayanan and Shmatikov, 2009] (dans la communeauté de la sécurité informatique).
- Netflix a publié une partie de ses données de consommation : le but était de produire un logiciel de suggestion de locations de dvd 10% plus efficace que le leur.

Consommateur	Films/Notes publics	Films/Notes cachés
456789	87/4, 998/2, 687/4	954/2, 486/4
654953	45/3, 743/3, 486/4	687/3, 45/4
...		

- Anonymization naïve : on remplace les noms par des chiffres aléatoires.

Cas practique de-Anonymization: Netflix début des années 2000

- Résultats frappants [Narayanan and Shmatikov, 2009] (dans la communeauté de la sécurité informatique).
- Netflix a publié une partie de ses données de consommation : le but était de produire un logiciel de suggestion de locations de dvd 10% plus efficace que le leur.

Consommateur	Films/Notes publics	Films/Notes cachés
456789	87/4, 998/2, 687/4	954/2, 486/4
654953	45/3, 743/3, 486/4	687/3, 45/4
...		

- Anonymization naïve : on remplace les noms par des chiffres aléatoires.
- Résultats : 99% de dé-anonymization correcte pour plus de 8 notes (84% si on oublie la date de publication mais que la liste contient des films peu vus).

Désanonymisation : techniques

- Ils ont regardé les notes sur d'autres réseaux sociaux (IMDB en l'occurrence)...
- En fait n'importe quelle liste exhaustive ou presque peut vous identifier :
  - la liste des Wifi de votre appareil (pas besoin de se connecter l'appareil les teste tous en série pour voir si un convient).
  - votre environnement de travail (liste des applications plus spécifications de taille d'écran etc.).
  - la liste de tous les livres de votre bibliothèque,
  - etc.
- Il y a 7.5×10^9 soit $\log(7.5 \times 10^9) = 33$ bits suffisent pour vous identifier, à peine plus de 4 caractères ASCII.

Stop COVID Application

- Problèmes de privacité : même pas la peine de connaitre les caractéristiques techniques/crypto employées.

Stop COVID Application

- Problèmes de privacité : même pas la peine de connaitre les caractéristiques techniques/crypto employées.
- Qu'est ce qui est protégé ? Le statut médical (HIV, cancer, etc.).

Stop COVID Application

- Problèmes de privacité : même pas la peine de connaitre les caractéristiques techniques/crypto employées.

- Qu'est ce qui est protégé ? Le statut médical (HIV, cancer, etc.).

  $\Longrightarrow$ Une attaque : utilisation d'un "burn phone" pendant une interview. Si le téléphone devient "rouge" (cas contact) vous venez d'apprendre le statut de la personne rencontrée.

Stop COVID Application

- Problèmes de privacité : même pas la peine de connaitre les caractéristiques techniques/crypto employées.

- Qu'est ce qui est protégé ? Le statut médical (HIV, cancer, etc.).

  $\Longrightarrow$ Une attaque : utilisation d'un "burn phone" pendant une interview. Si le téléphone devient "rouge" (cas contact) vous venez d'apprendre le statut de la personne rencontrée.
- Quelles sont les implications d'une alerte ? Quarantaine/isolement (obligatoire ou recommandé).

Stop COVID Application

- Problèmes de privacité : même pas la peine de connaitre les caractéristiques techniques/crypto employées.

- Qu'est ce qui est protégé ? Le statut médical (HIV, cancer, etc.).

  $\Longrightarrow$ Une attaque : utilisation d'un "burn phone" pendant une interview. Si le téléphone devient "rouge" (cas contact) vous venez d'apprendre le statut de la personne rencontrée.

- Quelles sont les implications d'une alerte ? Quarantaine/isolement (obligatoire ou recommandé).

  $\Longrightarrow$ Attaque : on met 15 téléphones dans un sac, on s'approche de la cible. On déclare les 15 téléphones "rouges". La cible est mise en quarantaine.

- Utilisez votre imagination...

Passe vaccinal - problèmes de sécurité

$\Longrightarrow$ Etablissement d'une base de données : qui fait quoi, où et quand.

- Multiplication des interventions : eg vérification de l'identité.
- Problèmes de sécurité liés à la "certification" de passes:
  - Bonnes informations entrées ? (corruption, idéologie, hack de l'api du centre de vaccination etc.)
  - Production de faux codes par hacking.
  - Bonne personne vaccinée ?
  - La personne est-elle vaccinée pour de vrai ? (bras en plastique etc.)

Plan

- Cryptanalyse d'Enigma
- L'Anonymisation naïve ne fonctionne pas

1. Cryptologie et théorie de l'information : [Shannon, 1949]
   - Etude de cryptosystèmes d'un point de vue information théorique
   - Entropie des mots de passe

2. Conclusion

Privacité et informatique : Art ou Science ?

- Informatique : art or science ?
  "The Art of Computer Programming", D.E. Knuth.

Privacité et informatique : Art ou Science ?

- Informatique : art or science ?
  "The Art of Computer Programming", D.E. Knuth.
- Question basique en privacité : comment qualifier la "force" d'un cryptosystème ?
  - Computational security.
  - Provable security.
  - Unconditional security.
- Quelles attaques ?
  - Message crypté seulement ?
  - Attaque en texte clair ?
  - Texte clair partiel ?
  - etc.

Information Theory 101

- Avant tout : qu'est ce que l'information ?

Information Theory 101

- Avant tout : qu'est ce que l'information ?

$\Longrightarrow$ en fin d'analyse c'est la possiblité de transformer une distribution de probabilités.

Information Theory 101

- Avant tout : qu'est ce que l'information ?

$\Longrightarrow$ en fin d'analyse c'est la possiblité de transformer une distribution de probabilités.

- Travail de C.E. Shannon (fin des années 40):

  "A mathematical Theory of communication", The Bell System Technical Journal, vol. 27, 1948.

  "Communication Theory of Secrecy Systems", The Bell System Technical Journal, vol. 28, 1949.

- Comment certaines distributions de probabilités sont affectées par certaines hypothèses.

Plan

- Cryptanalyse d'Enigma
- L'Anonymisation naïve ne fonctionne pas

1. Cryptologie et théorie de l'information : [Shannon, 1949]
   - Etude de cryptosystèmes d'un point de vue information théorique
   - Entropie des mots de passe
2. Conclusion

Secret parfait

- Comment prouver qu'un cryptosystème est inconditonnellement sûr ?

Secret parfait

- Comment prouver qu'un cryptosystème est inconditonnellement sûr ?
- Définition formelle de cryptosystème:

Definition (cryptosystème)

$(T, C, K, \mathcal{E}, \Delta)$ avec:

- T : Textes en clairs.
- C : Cryptés.
- K : Clefs.
- $\forall k \in K$ il existe $e_k \in \mathcal{E}$ et $d_k \in \Delta$ tel quel :

$$e_k : T \to C$$
$$d_k : C \to T$$

et $\forall x \in T$, on a $d_k(e_k(x)) = x$

Définition du Code Parfait

Definition (Confidentialité parfaite)

Un cryptotsytem a la propriété de confidentialité parfaite si:

$$\mathbf{Pr}[\, x \mid y \,] = \mathbf{Pr}[x]$$

- En d'autres termes si a posteriori la probabilité que le texte clair soit x, étant donné le texte crypté y, est la même que cette probabilité a priori.
- One-time pad a la propriété de confidentialité parfaite... mais est concrètement inutilisable (ne passe pas à l'échelle).
- Shannon's perfect secrecy theorem: Un cryptosystème a la confidentialité parfaite ssi :
  - chaque clef est utilisée de manière équiprobable $1/|K|$
  - pour chaque texte x et texte crypté y, il existe une unique clef k telle que $e_k(x) = y$.

Entropie

- Que se passe t il si la clef est utilisée pour plus d'une encryption ?

Entropie

- Que se passe t il si la clef est utilisée pour plus d'une encryption ?
- l'entropie est une mesure mathematique de la quantité d'information contenu dans un message.
  $\Longrightarrow$ calculée comme un fonction de distribution de probabilités.

Entropie

- Que se passe t il si la clef est utilisée pour plus d'une encryption ?
- l'entropie est une mesure mathematique de la quantité d'information contenu dans un message.
  $\Longrightarrow$ calculée comme un fonction de distribution de probabilités.
- Supposons $\mathbf{X}$ suit $\mathcal{P}$: qu'est ce qu'on apprend en faisant des expériences qui suivent $\mathcal{P}$?
  $\Longrightarrow$ C'est l'entropie de $\mathbf{X}$: $H(\mathbf{X})$
- On peut imaginer le jeu suivant : deviner un mot dont les lettres sont retournées les unes à la suite des autres.

Entropie définition

Definition (Entropie)

Soit $\mathbf{X}$ qui suit $\mathcal{P}$, alors

$$H(\mathbf{X}) = -\sum_{x \in X} \mathbf{Pr}_{\mathcal{P}}[\mathbf{X} = x] \log_2(\mathbf{Pr}_{\mathcal{P}}[\mathbf{X} = x])$$

- Le log n'est pas défini en 0, mais la limite est 0... alors ça marche dans une somme.
- Le choix de la base du log est arbitraire.
- Plusieurs applications dans les cryptosystèmes, eg:

Theorem

Considérons le cryptosystème $(T, C, K, \mathcal{E}, \Delta)$*:*

$$H(\mathbf{K} \mid \mathbf{C}) = H(\mathbf{K}) + H(\mathbf{P}) - H(\mathbf{C})$$

Plan

- Cryptanalyse d'Enigma
- L'Anonymisation naïve ne fonctionne pas

Comment bien choisir un mot de passe?

- De loin la technologie de contrôle d'accès la plus répandue.
- Problème lié aux nombres de mots de passes à connaître (réutilisation ?... mauvaise idée).
- Beaucoup de conseils sur ce qui constitue un "bon" mot de passe.
- La théorie de l'information permet de donner un avis scientifique sur ce qu'est un bon mot de passe. $\Longrightarrow$ Il faut que le mot ne soit pas trop long (mémoire) ni trop court (dangereux): quel est le bon compromis ?

Comment bien choisir un mot de passe?

- De loin la technologie de contrôle d'accès la plus répandue.
- Problème lié aux nombres de mots de passes à connaître (réutilisation ?... mauvaise idée).
- Beaucoup de conseils sur ce qui constitue un "bon" mot de passe.
- La théorie de l'information permet de donner un avis scientifique sur ce qu'est un bon mot de passe. $\Longrightarrow$ Il faut que le mot ne soit pas trop long (mémoire) ni trop court (dangereux): quel est le bon compromis ?
- Dans la vie réelle:
  - Construction d'un dictionnaire en faisant un scan du disque (50% de taux de succès, lié à la complexité des systèmes).
  - Utiliser un password manager est un bon compromis.

Attaque en force brute et âge de l'univers

- On réduit le problème à une recherche exhaustive.
- Enumérer les mots de passes revient à énumérer les entiers.
- Supposons que vous puissiez vérifier 10^{15} mots de passe par seconde.
- Supposons que Google ou la NSA puissent s'offrir 1000 ordinateurs de ce type: 10^{18} mots de passe par second.

size in bits	execution time
56	less than 1 sec
64	18 sec
128	$1,07 \times 10^{13}$ years
256	$3,65 \times 10^{51}$ years
512	$4,25 \times 10^{128}$ years

Estimation d el'âge de l'univers: $13,7x10^{9}$ années.

Landauer's Principle

- Et si la NSA a un ordinateur qui va vraiment vraiment vite ?

Landauer's Principle

- Et si la NSA a un ordinateur qui va vraiment vraiment vite ?
- Il doit quand même suivre les lois de la physique : l'énergie minimale à température T est donnée par

$$\Delta E \geq kT \log(2)$$

avec $k = 1.38 \times 10^{-23} J/K$

- Pour énuméer tous les entiers sur 128 bits demande $10^{18} \simeq 30 gigaWatts/year$, c'est-à-dire $267 teraWatts/hour$ en gros la moitié de la puissance électrique française sur un an.
- Il n'y a pas assez d'énergie ($E = mc^2$) dans l'univers visible pour énumérer les entiers sur 256 bits.

Landauer's Principle

- Et si la NSA a un ordinateur qui va vraiment vraiment vite ?
- Il doit quand même suivre les lois de la physique : l'énergie minimale à température T est donnée par

$$\Delta E \geq kT \log(2)$$

avec $k = 1.38 \times 10^{-23} J/K$

- Pour énuméer tous les entiers sur 128 bits demande $10^{18} \simeq 30 gigaWatts/year$, c'est-à-dire $267 teraWatts/hour$ en gros la moitié de la puissance électrique française sur un an.
- Il n'y a pas assez d'énergie ($E = mc^2$) dans l'univers visible pour énumérer les entiers sur 256 bits.

$\Longrightarrow$ Plus d'une centaine de bits d'entropie du mot d epasse est exagéré.

Mesure de la force d'un mot de passe

- L'idée de base est de mesurer l'entropie contenue dans un mot de passe. .
- Etant donné une distribution équiprobable, dans un ensemble de taille n l'entropie d'un élément est $\log(n)$.
- Composé sur un alphabet à N symboles, et un mot de passe de taille L, il y a N^L possibilités, donc $\log(N^L) = L\log(N)$

Symboles	Entropie par symbole
0-9	3,32
0-9+'A'-'F'	4
'a'-'z'	4,7
0-9+'a'-'z'	5,1
'A'-'Z'+'a'-'z'	5,7
0-9+'a'-'z'+'A'-'Z'	5,9
ASCII writable	6,56

- Dictionnaire anglais : 170000 mots, donc 12 bits d'entropie par mot.

Plan

- Cryptanalyse d'Enigma
- L'Anonymisation naïve ne fonctionne pas

1. Cryptologie et théorie de l'information : [Shannon, 1949]
   - Etude de cryptosystèmes d'un point de vue information théorique
   - Entropie des mots de passe

2. Conclusion

Conclusion

- La privacité est complexe:
  - Philosophiquement/Conceptuellement.

Conclusion

- La privacité est complexe:
  - Philosophiquement/Conceptuellement.
  - Concretement.

Conclusion

- La privacité est complexe:
  - Philosophiquement/Conceptuellement.
  - Concretement.
  - Technologiquement.

Conclusion

- La privacité est complexe:
  - Philosophiquement/Conceptuellement.
  - Concretement.
  - Technologiquement.
  - Scientifiquement.

Conclusion

- La privacité est complexe:
  - Philosophiquement/Conceptuellement.
  - Concretement.
  - Technologiquement.
  - Scientifiquement.
- La sécurité demande un état d'esprit spécifique qui n'est pas vraiment celui développé dans un cursus d'ingéniérie.

Conclusion

- La privacité est complexe:
  - Philosophiquement/Conceptuellement.
  - Concretement.
  - Technologiquement.
  - Scientifiquement.
- La sécurité demande un état d'esprit spécifique qui n'est pas vraiment celui développé dans un cursus d'ingéniérie.

Bibliography I


Anderson, R. J. and Needham, R. M. (1995).
Programming satan's computer.
In *Computer Science Today: Recent Trends and Developments*, pages 426–440. Springer.


Andersson, R. (2008).
Security Engineering: A Guide to Building Dependable Distributed Systems.
Wiley.


Narayanan, A. and Shmatikov, V. (2009).
De-anonymizing social networks.
In *30th IEEE Symposium on Security and Privacy (S&P 2009), 17-20 May 2009, Oakland, California, USA*, pages 173–187.

Bibliography II


Schneier, B. (1996).
Applied Cryptography: Protocols, Algorithms, and Source Code in C.
Wiley.


Shannon, C. (1948).
A mathematical theory of communication.
Bell System Technical Journal, 27:379–423, 623–656.


Shannon, C. (1949).
Communication theory of secrecy systems.
Bell System Technical Journal, Vol 28, pp. 656-715.


Singh, S. (2000).
The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography.
Anchor.

Bibliography III


Stinson, D. (2005).
Cryptography Theory and Practice.
CHapman and Hall/CRC.
third edition.