Données personnelles et vie privée sur le Web

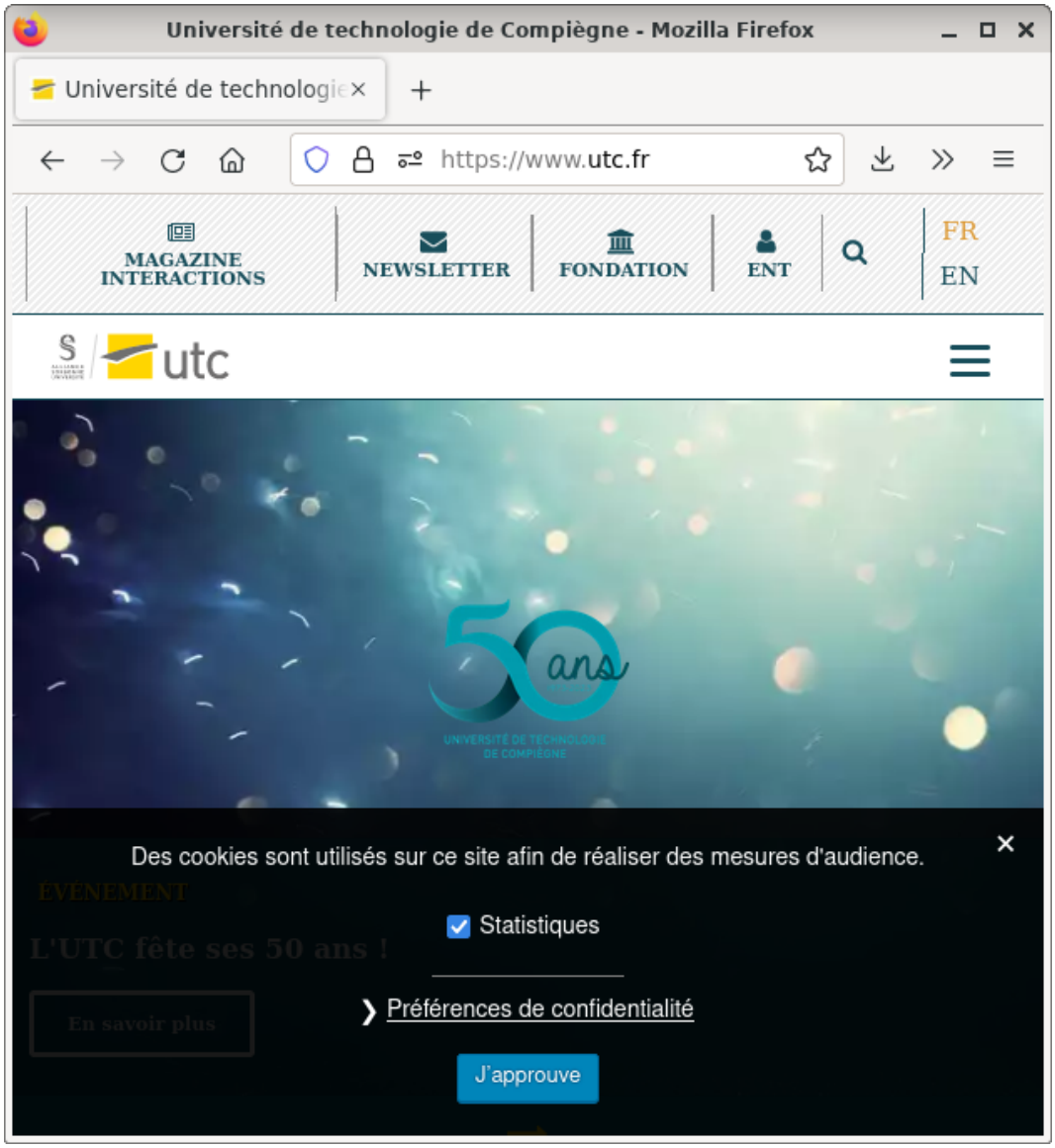
Table des matières

Objectifs

- Savoir expliquer les principes du chiffrement symétrique et asymétrique
- Savoir expliquer à quoi sert HTTPS
- Savoir expliquer à quoi ne sert pas HTTPS
- Savoir expliquer plusieurs techniques utilisées pour le tracking
- Savoir expliquer à quoi sert le RGPD
- Connaître quelques techniques permettant d'échapper à la surveillance sur le Web

Mise en situation

1. Surveillance fonctionnelle



2. Surveillance publicitaire

Jérôme Hourdeaux. 2022. Les publicitaires font main basse sur les données des élèves partout dans le monde, Médiapart. https://www.mediapart.fr[1]

25 mai 2022 à 06h52. https://www.mediapart.fr/journal/international/250522/les-publicitaires-font-main-basse-sur-les-donnees-des-eleves-partout-dans-le-monde

1. https://www.mediapart.fr/journal/international/250522/les-publicitaires-font-main-basse-sur-les-donnees-des-eleves-partout-dans-le-monde

« Human Rights Watch a analysé durant deux ans 164 outils numériques destinés aux élèves de 49 pays durant la pandémie afin qu'ils puissent continuer à suivre leurs cours. 89 % « surveillaient les enfants, secrètement et sans le consentement de leurs parents ». »

« Or, sur les 73 applications analysées, 22, soit 30 %, « s'accordaient la capacité de collecter des données de localisation précises, ou des coordonnées GPS qui peuvent déterminer à la localisation exacte d'un enfant à 4,9 mètres près ». »

« Dix de ces applications étaient directement destinées aux enfants, comme Minecraft : Education Edition, et ont collecté les données de localisation d'environ 52,1 millions d'enfants. »

« Certaines applications, 18 sur les 73, collectaient également le Wifi SSID, qui correspond au nom du réseau auquel se connecte un téléphone mobile. Avec cette donnée, les entreprises peuvent retrouver la localisation exacte du réseau en question. Parmi les applications utilisant cette technique, on retrouve des géants du numérique comme Microsoft Teams, Cisco Webex, Zoom (recommandé dans l'État du New South Wales en Australie, au Cameroun, au Kazakhstan, en République de Corée, en Roumanie, en Californie, au Texas et en Angleterre), YouTube (recommandé dans l'État d'Uttar Pradesh en Inde, en Malaisie, au Nigeria et en Angleterre), WhatsApp (recommandé dans l'État d'Uttar Pradesh et au Cameroun), Telegram (recommandé au Nigeria) ou encore Facebook (recommandé à Taïwan). »

« Enfin, les élèves ont également pu être pistés en dehors de leurs salles de classe virtuelles, lors de leurs autres activités sur Internet, via les fameux « cookies », des petits fichiers installés dans le navigateur d'un·e internaute pour l'identifier. »

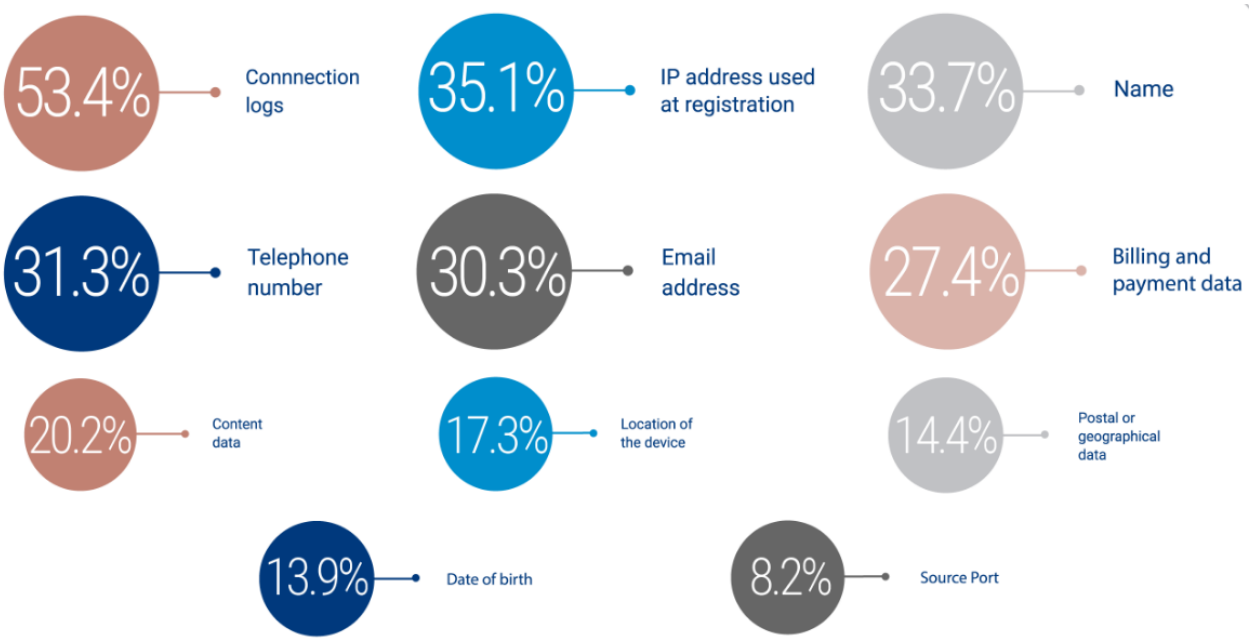
3. Surveillance sécuritaire

Manach Jean-Marc , 2021. 66 % des policiers européens satisfaits de la coopération des plateformes, NextInpact. https://www.nextinpact.com[2]

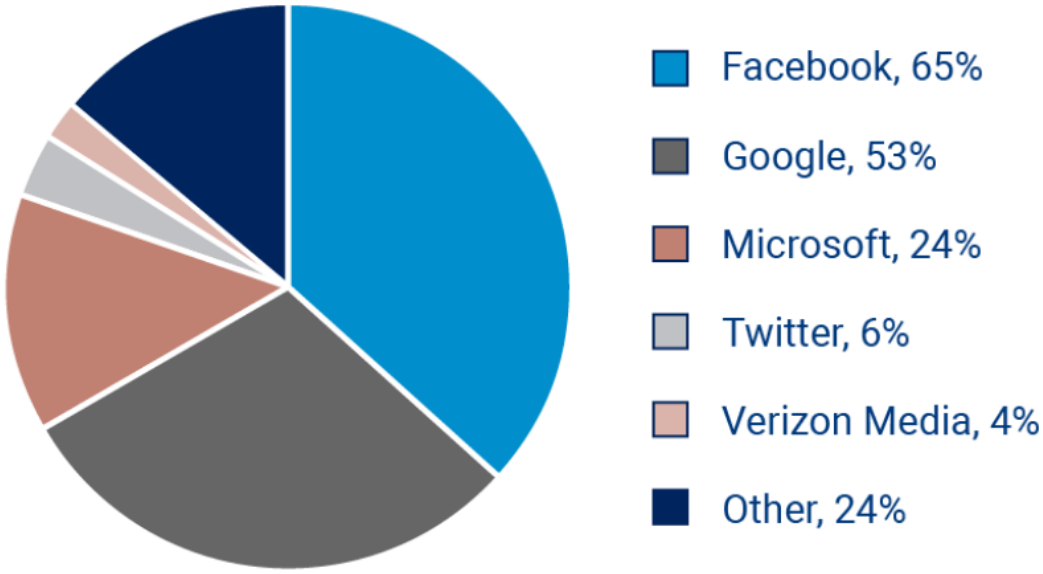
« Le projet SIRIUS a été créé par Europol en octobre 2017 en réponse au besoin croissant des services répressifs de l'UE d'accéder à des preuves électroniques pour les enquêtes sur Internet, « car plus de la moitié de toutes les enquêtes pénales incluent aujourd'hui une demande d'accès transfrontalière des preuves électroniques (telles que des SMS, des e-mails ou des applications de messagerie) ». »

« Les rapports de transparence d'Airbnb, Facebook, Google, Microsoft, Snap, TikTok, Twitter et Verizon Media révèlent que, de 2019 à 2020, le volume des demandes de données d'utilisateurs soumises par les autorités a augmenté de 27,1 % dans l'UE, pour atteindre prés de 163 000 demandes en 2020, contre 128 000 en 2019 (+27 %), et 105 000 en 2018 (+22 %). »

2. https://www.nextinpact.com/article/49153/66-policiers-europeens-satisfaits-cooperation-plateformes



The most important types of data needed



The most contacted services

Technopolice

> « La « Smart City » révèle son vrai visage : celui d'une mise sous surveillance totale de l'espace urbain à des fins policières. En septembre 2019, des associations et collectifs militants ont donc lancé la campagne Technopolice, afin de documenter ces dérives et d'organiser la résistance.
>
> La Quadrature Du Net. https://technopolice.fr/ »

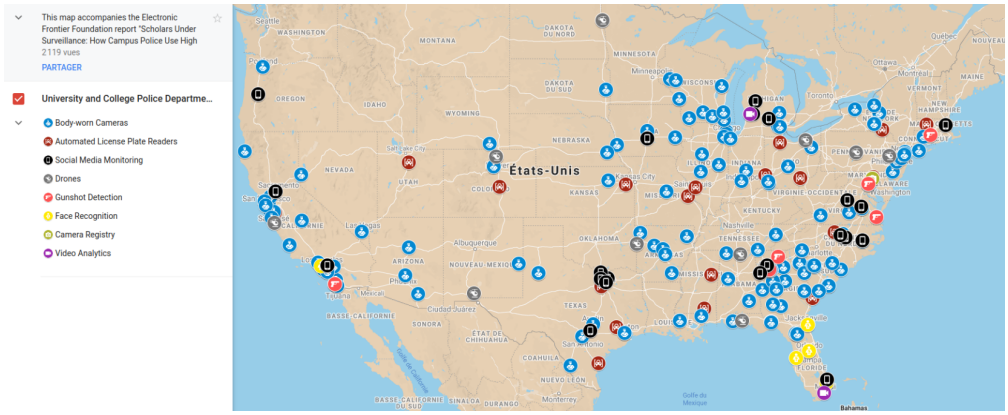
Conférence Technopolice,24 mai 2022

« Aux États-Unis, la peur des fusillades dans les écoles et les campus a banalisé l'installation de systèmes de surveillance sophistiqués. Elles vont bien au-delà des caméras de vidéosurveillance : drones, capteurs de détection de coups de feu, lecteurs de plaques d'immatriculation automatisés, logiciels biométriques, entre autres.

[...]

Des collèges et universités surveillent également leurs étudiants sur les réseaux sociaux, « et ce n'est pas seulement pour retweeter ou aimer un joli post Instagram sur leur stage d'été », ironise l'EFF. Ils s'en servent d'abord pour rechercher des publications où les étudiants indiquent des idées suicidaires ou encore des menaces de violence armée.

Jean-Marc Manach, 2021. Les technologies de surveillance à l'assaut des campus américains, NextINpact. https://www.nextinpact.com/[3] »



Campus Police Surveillance

3. https://www.nextinpact.com/article/46426/les-technologies-surveillance-a-assaut-campus-americains

Introduction au chiffrement

1. Le druide Gépégix et ses secrets : une métaphore du chiffrement asymétrique

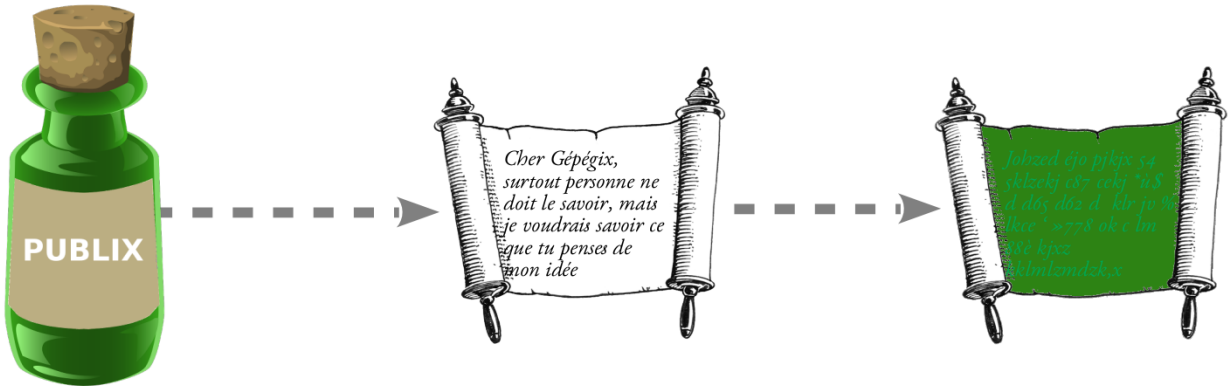
Le druide Gépégix reçoit des lettres du monde entier de personnes souhaitant profiter de ses talents de divination.

Par exemple un chef de clan souhaite savoir si c'est le bon moment pour attaquer son ennemi ou un garçon veut savoir si sa voisine est amoureuse de lui.

Le problème du druide est qu'il veut bien répondre à ces questions, mais il ne souhaite pas que tous ces secrets tombent entre de mauvaises mains.

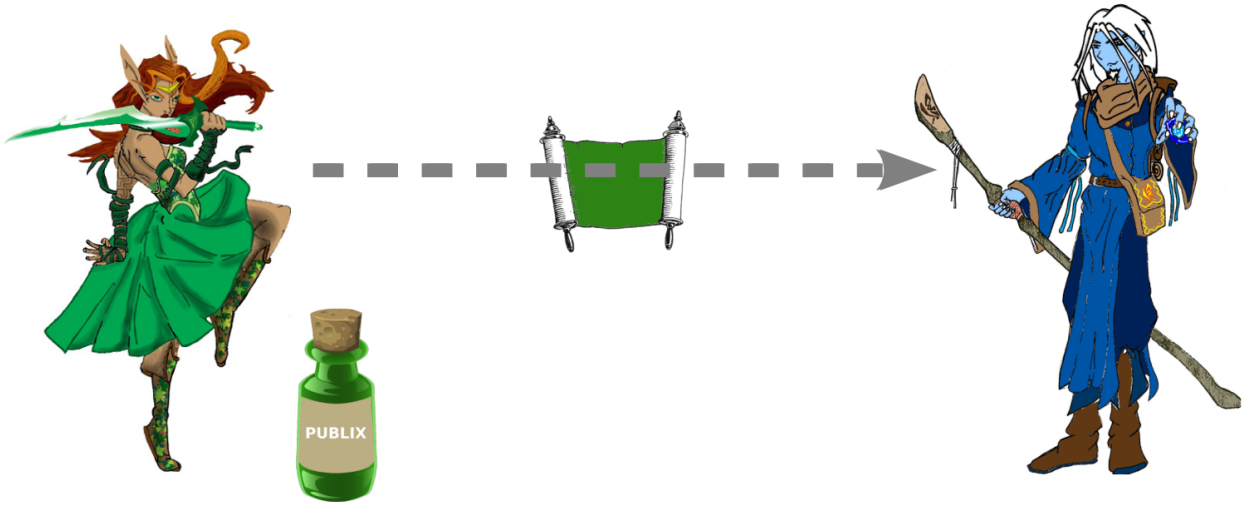
Il a alors une idée, il s'enferme dans son laboratoire et il invente deux potions.

La première, Publix, est une potion toute verte. Quand on la verse sur un parchemin, les lettres se brouillent, tout devient vert et on ne peut plus lire ce qu'il y a d'écrit.



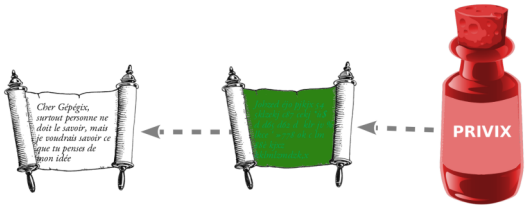
Le druide produit en grande quantité cette potion, qu'il distribue dans le monde entier.

À présent quand quelqu'un veut lui envoyer une requête, il verse du Publix dessus et la requête devient illisible.



La seconde potion que le druide a inventée, Privix, est une potion toute rouge.

Quand on la verse sur un parchemin couvert de Publix, elle annule le brouillage causé par le Publix et le texte redevient lisible.



Gépégix garde bien caché ses réserves de Privix, car c'est la seule potion qui permet de déchiffrer le Publix et lui seul en connaît la formule.

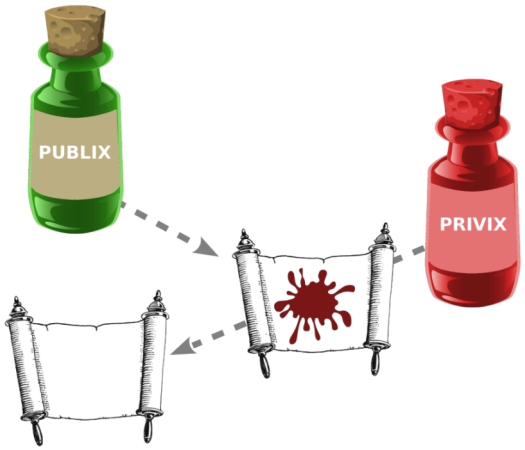
Ainsi, grâce aux potions Publix et Privix, les secrets transmis à Gépégix sont bien gardés.

2. Les prophéties authentiques de Gépégix : une métaphore de la signature numérique

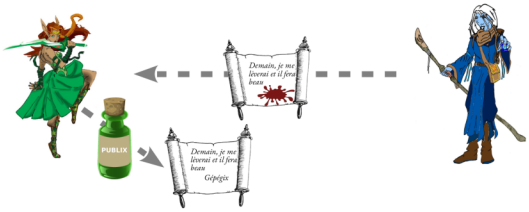
Le druide Gépégix est reconnu dans le monde entier pour la qualité de ses prophéties. Il les inscrit sur des parchemins qu'il livre à des messagers qui vont les distribuer dans le monde entier.

Malheureusement il existe des devins moins habiles que lui qui usurpent son identité pour faire passer de fausses divinations pour les siennes.

Gépégix a trouvé une parade en découvrant que l'on pouvait utiliser le Publix et le Privix dans les deux sens. Quand on verse du Publix vert du Privix rouge, cela le fait disparaître.



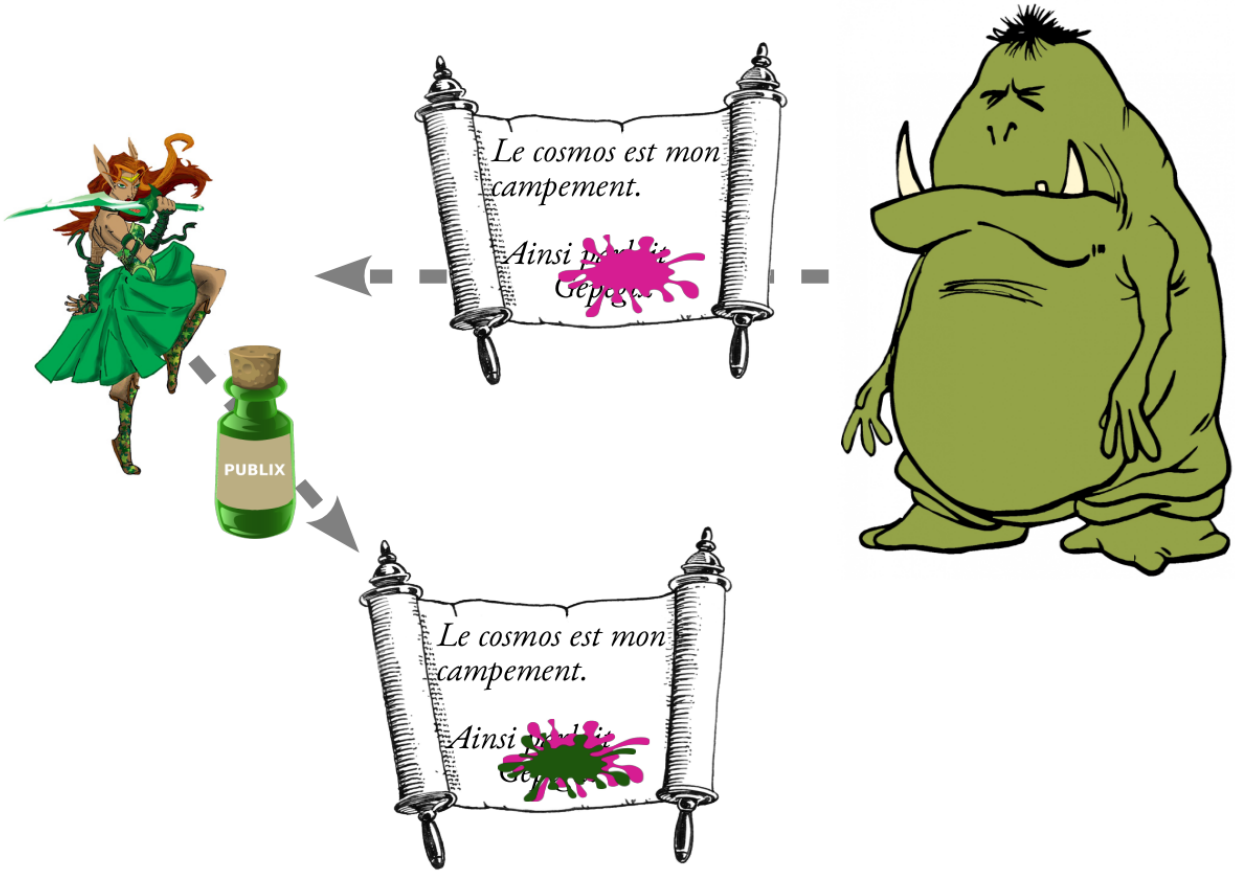
Ainsi il prend soin de signer chacune de ses prophéties en versant quelques gouttes de Privix dessus.



Quand quelqu'un reçoit un parchemin, il verse un peu de Publix dessus, si le rouge disparaît cela prouve que c'est bien un des messages de Gépégix.

En effet seul Gépégix dispose du Privix, et le Publix ne fonctionne que sur le Privix.

Donc si un imposteur avait signé avec une potion à lui, par exemple de l'Impostix, alors le Publix n'aurait pas fonctionné.



3. Gépégix et le conseil des druides : une métaphore du chiffrement et de la signature de mails

Fort de ses expériences, Gépégix propose un système de communication au conseil des druides, qui permettra à la fois de communiquer sans que personne ne puisse lire les messages interceptés, et en même temps de s'assurer de qui est l'expéditeur de chaque message.

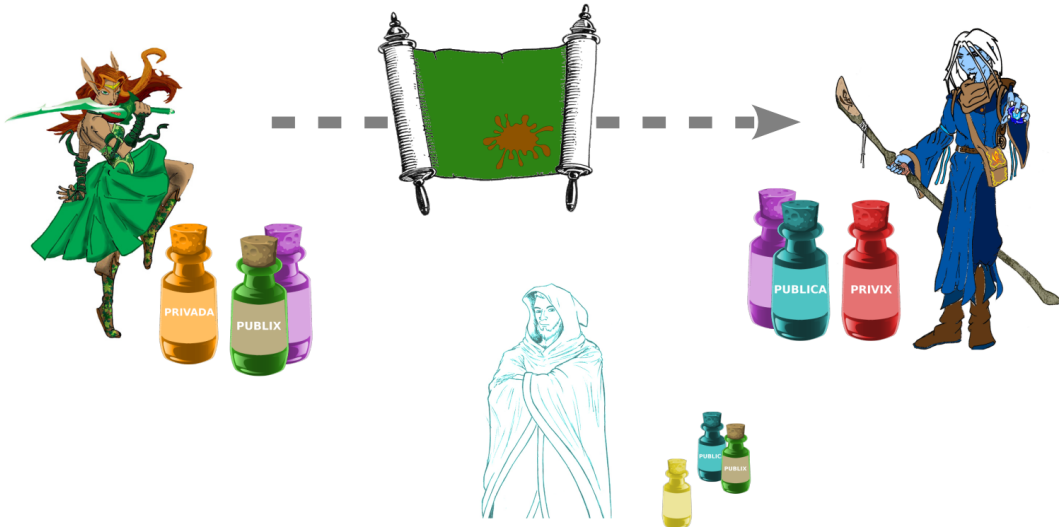
D'abord il explique à chaque druide comment produire ses propres potions, afin qu'elles aient les mêmes propriétés que Publix et Privix, mais que chaque couple de potion soit unique.

Par exemple sa collègue Pégépa a réussi a créer une potion turquoise Publica ainsi qu'une orange Privada qui fonctionnent toutes les deux comme Publix et Privix, mais dont seul Pégépa connaît la formule.

Quand Gépégix envoie un message à Pégépa, il verse du Publica turquoise sur son message et signe avec quelques gouttes de Privix rouge.

Quand Pégépa reçoit le message, elle verse du Privada orange sur le message pour pouvoir le lire, et du Publix vert sur la signature pour vérifier que c'est bien un message de Gépégix.

Dans l’autre sens Pégépa couvre son message de Publix vert et signe avec un peu de Privada orange.



Ainsi une fois que chacun au conseil des druides a composé ses deux potions personnelles, une privée et une publique, et qu’il dispose d’un exemplaire de la potion publique de chacun des autres druides, tout le monde peut échanger de façon sécurisée.

Principe du chiffrement

1. Découverte du chiffrement

Objectifs

- Découvrir la notion de chiffrement ;
- Connaître un algorithme basique de chiffrement.

Mise en situation

Lorsqu'on échange des messages sur Internet, c'est un peu comme si on communiquait avec des cartes postales. C'est à dire qu'il est très facile pour un intermédiaire de les lire. Les messages ne sont pas confidentiels.

La seule façon de communiquer de façon confidentielle est de **chiffrer** les messages. Cela consiste à définir un code secret que seuls les deux correspondants connaissent. Ainsi les messages sont toujours lisibles par des tiers, mais il ne sont plus en mesure de comprendre quoi que ce soit. C'est comme si sur ma carte postale j'écrivais : DZMIY Z YPO. On peut toujours la lire, mais sans le code il est difficile de comprendre le message. Si je vous donne le code, alors vous serez en mesure de le **déchiffrer** c'est à dire de l'appliquer pour retrouver le message original. Mais comme un code de chiffrement doit rester secret, je ne le donne pas dans une vidéo.

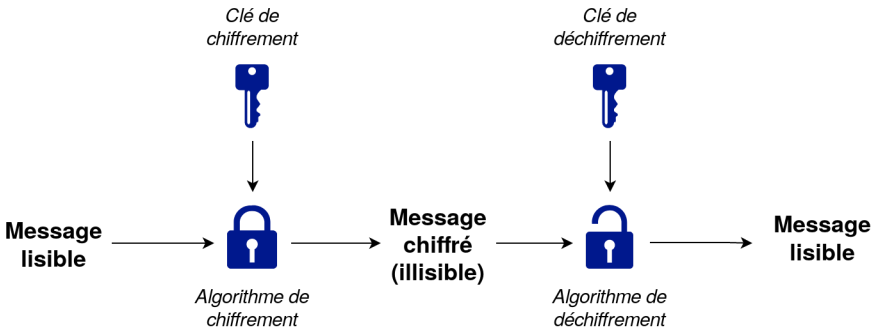
Le code que j'ai utilisé ici est très simple, il sera facilement décrypté. **Décrypter** un code signifie parvenir à en trouver la teneur originale sans en avoir été informé. Je vous laisse décrypter mon message.

Processus de transmission d'un message chiffré

Fondamental

La transmission d'un message chiffré se fera en trois étapes :

1. Chiffrement du message à l'aide d'un algorithme de chiffrement auquel on fournit une clé de chiffrement.
2. Transmission du message chiffré.
3. Déchiffrement à l'aide d'un algorithme de chiffrement auquel on fournit une clé de déchiffrement.



Fondamental

Le chiffrement est un procédé cryptographique permettant de coder un message de telle façon que sa lecture ne soit possible que par le seul possesseur de la clé de déchiffrement

Remarque

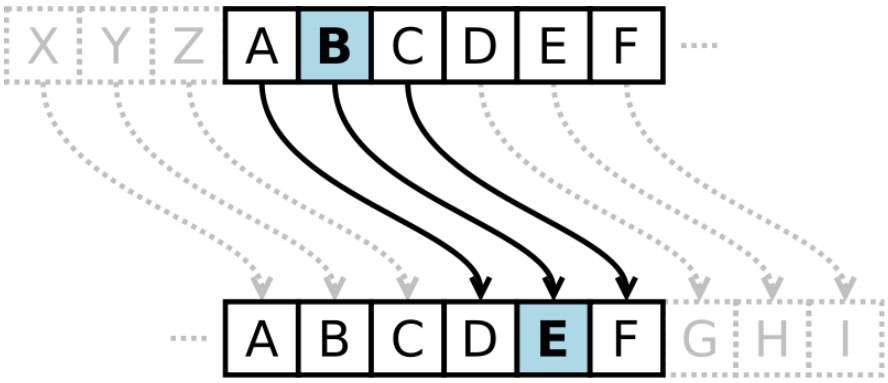
- les **algorithmes de chiffrement/déchiffrement** : ces algorithmes sont le plus souvent disponibles librement.
- les **clés** : elles sont un ensemble de paramètres à fournir à l'algorithme afin qu'il puisse réaliser sa tâche. Pour garantir la confidentialité du message, la clé de déchiffrement doit être privée.
- l'**unicité de la paire de clés** : pour une clé de chiffrement il n'existe qu'une clé permettant de déchiffrer et l'inverse est également vrai (aucun intermédiaire ne peut deviner cette clé).

Le chiffre de César

Exemple

Aussi appelé chiffrement par décalage, ce chiffrement très simple consiste à décaler toutes les lettres d'un message. Dans ce cas, les clés de chiffrement et de déchiffrement sont identiques et correspondent à l'amplitude du décalage.

Par exemple pour un décalage de 2, les « A » deviendront des « C », les « B » deviendront des « D », etc. Le mot « shannon » donnera « vkdqqrq » avec un décalage de 3.



Chiffre de César

Chiffrer, pas crypter

Complément

La mot crypter n'existe pas en français. Un message est donc chiffré mais pas crypté.

Néanmoins, on peut parler de « décryptage » lorsque l'on cherche à déchiffrer un message sans disposer de la clé nécessaire.

Sécurité du chiffrement

Complément

Le problème du chiffrement par décalage est son manque de sécurité. Il est très simple de trouver la clé de déchiffrement par essais successifs.

La sécurité de ces algorithmes doit reposer sur des propriétés mathématiques fortes et/ou sur une bonne transformation de l'information.

Baser la sécurisé du chiffrement sur le fait que son procédé soit inconnu par un tiers revient à faire de la **sécurité par l'obscurité**. Cacher un procédé protège beaucoup moins que d'utiliser un procédé de chiffrement public et solide.

Hachage cryptographique et condensat *Complément*

Le **hachage cryptographique** est le second procédé cryptographique très utilisé. Il permet de produire des **condensats** (aussi appelé empreinte numérique). Un condensat est une chaîne de caractères **unique** de taille fixe pour chaque donnée pouvant exister. Ainsi, deux messages différents (même d'un caractère) auront un condensat différent.

Il n'est pas possible d'inverser le processus et de passer d'un condensat au message d'origine (ceci l'oppose au chiffrement). Les algorithmes classiques sont : SHA256, SHA1, MD5, etc.

À retenir

- Le chiffrement permet de transformer un message afin qu'il ne soit lisible que par une personne possédant la clé de déchiffrement.
- La sécurité d'un chiffrement repose sur de bonnes propriétés mathématiques et sur de bonnes opérations de transformation de l'information.

2. Chiffrement symétrique

Objectifs

- Découvrir le chiffrement symétrique ;
- Utiliser un algorithme de chiffrement symétrique.

Mise en situation

Lorsque deux personnes qui communiquent entre elles partagent exactement la même technique de chiffrement, on parle de chiffrement symétrique. C'est en général une méthode insuffisante pour les communications entre plusieurs personnes. Il est possible d'utiliser une clé symétrique pour communiquer avec quelqu'un, mais dans ce cas la clé sera changée à chaque nouvelle communication.

Le chiffrement symétrique est aussi utilisé lorsque l'on souhaite chiffrer des données sans les partager. C'est le cas lorsque l'on chiffre son disque dur pour que, même en cas de vol, seul le propriétaire puisse accéder aux données. Ainsi, le but n'est plus de sécuriser une communication mais le stockage lui-même.

Le chiffrement symétrique *Définition*

Le chiffrement symétrique est un chiffrement dans lequel la clé de chiffrement sert également à déchiffrer. On parle alors de clé secrète.

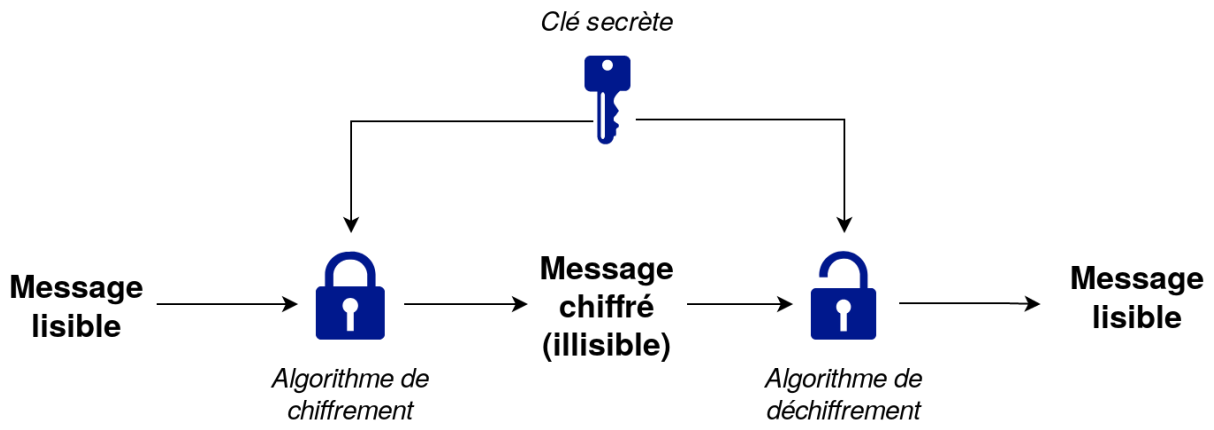


Schéma d'utilisation classique

Exemple

Bob souhaite chiffrer son disque dur pour qu'il soit le seul à pouvoir accéder à ses fichiers.

1. Lors de l'installation de son système d'exploitation Bob décide de chiffrer son disque dur, il choisit un mot de passe P.
2. La totalité du disque est chiffré avec une clé K générée par le système, cette clé est stockée sur le disque dur.
3. La clé K est à son tour chiffrée grâce au mot de passe P (elle ne doit pas être accessible en clair sur le disque).
4. À chaque déverrouillage de son ordinateur Bob entre le mot de passe qui permet de déchiffrer la clé K ; elle est alors chargée en mémoire afin d'être disponible rapidement.
5. À chaque fois qu'un fichier est accédé en lecture il est déchiffré avec K ; à chaque accès en écriture il est chiffré avec K.

Ainsi, Bob est certain que tant que son mot de passe reste secret, ses données sont sécurisées même si quelqu'un accède au disque de son ordinateur.

Le partage de clés

Attention

Il arrive parfois que la clé soit connue par plusieurs personnes ou soit présente sur plusieurs serveurs du propriétaire.

Le transfert de la clé doit absolument être sécurisé pour que le chiffrement ne soit pas compromis.

Plusieurs stratégies existent et une des plus efficaces est d'utiliser un autre type de chiffrement pour chiffrer la clé secrète et d'envoyer ce message au destinataire qui pourra récupérer la clé secrète en toute sécurité. Transférer la clé au travers d'une connexion SSH est une stratégie commune.

Le chiffrement AES

Définition

L'*Advanced Encryption System* est un standard très répandu pour réaliser du chiffrement symétrique. Il possède énormément de bonnes propriétés : facile à calculer, implémentation possible au niveau logiciel comme au niveau matériel (implémentation câblée). Ce type de chiffrement est utilisé notamment pour des protocoles tels que SSL (sécurisant les connexions HTTP) ou encore pour chiffrer son disque dur (VeraCrypt[4]).

Générer sa propre clé secrète

Méthode

Voici une commande basique pour générer une clé secrète. Il existe des implémentations plus complexes et sécurisées. Ici, la clé est simplement une suite aléatoire de 32 octets. Pour une utilisation réelle, il est conseillé d'utiliser des implémentations robustes et de confiance pour générer sa clé secrète.

```
1 openssl rand 32 > cle_secrete.pem
```

On obtient ici la clé de chiffrement dans le fichier `cle_secrete.pem`.

Utiliser une phrase secrète ou un mot de passe

Conseil

Il est possible de sécuriser sa clé secrète en spécifiant un mot de passe lors de la génération de la clé. Une telle pratique permet de conserver la sécurité même si un attaquant réussit à récupérer la clé. Il est tout de même fortement conseillé de générer une nouvelle clé dès lors qu'une clé est compromise.

4. https://fr.wikipedia.org/wiki/VeraCrypt

Chiffrer et déchiffrer *Méthode*

On peut utiliser les commandes dans un repl Bash ou un terminal, pour chiffrer et déchiffrer un message.

Pour chiffrer un fichier `msg.txt` en un nouveau fichier `msg.txt.enc` avec AES et une clé secrète générée `cle_secrete.pem` on utilise :

```
openssl aes-256-cbc -pbkdf2 -iter 100000 -in msg.txt -out msg.txt.enc -pass
file:cle_secrete.pem
```

Pour déchiffrer un fichier `msg.txt.enc` chiffré avec AES en un fichier `msg.txt` et une clé secrète générée `cle_secrete.pem` on utilise :

```
openssl aes-256-cbc -pbkdf2 -iter 100000 -d -in msg.txt.enc -out msg.txt -pass
file:cle_secrete.pem
```

À retenir

- Le chiffrement symétrique permet de sécuriser ses propres données avec une unique clé.
- Le chiffrement symétrique permet d'échanger des données avec des pairs s'ils disposent eux aussi de la clé.
- Le partage de clé est très complexe et doit rester exceptionnel pour conserver sa confidentialité.
- L'algorithme AES permet de réaliser un chiffrement symétrique.

3. Chiffrement asymétrique

Objectifs

- Découvrir le chiffrement asymétrique ;
- Utiliser un algorithme de chiffrement asymétrique.

Mise en situation

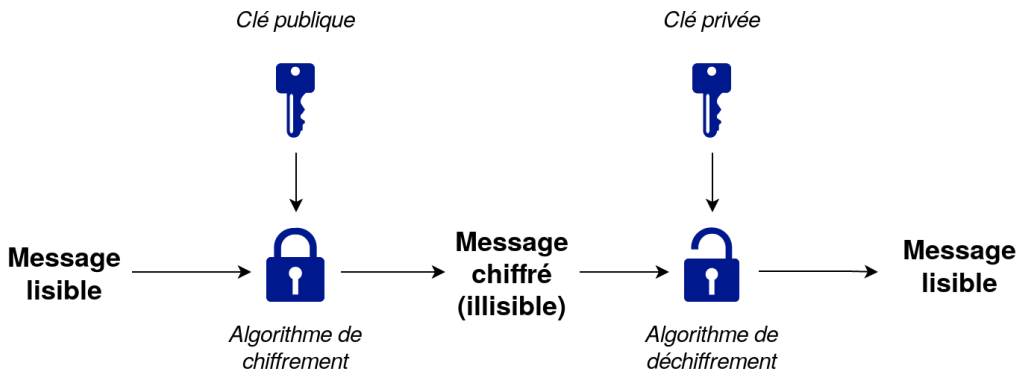
L'inconvénient du chiffrement symétrique est que si trois personnes souhaitent communiquer entre elles deux par deux, et bien le troisième pourra toujours espionner les deux autres, puisque le code est commun. On pourrait imaginer que chaque couple de personnes possède une clé spécifique, mais si 1000 personnes échangent entre elles, cela fera pour chacune, 999 clés à gérer. On voit bien que cela n'est pas satisfaisant pour les communications sur Internet.

La solution est fournie par le chiffrement asymétrique. Celui-ci est basé sur les propriétés mathématiques d'une paire de clés : la première est publique, elle est communiquée au monde entier et elle sert à chiffrer les messages. La seconde est privée, elle n'est communiquée à personne et elle sert à déchiffrer les messages. Ainsi si quelqu'un souhaite m'envoyer un message, il peut le chiffrer avec ma clé publique, seul moi pourrai le lire car je suis le seul à disposer de la clé privée.

On peut voir cela comme des milliers de boites inviolables que je diffuserais dans le monde entier. N'importe qui peut prendre une de mes boîtes et glisser un message dedans. Je suis le seul à en posséder la clé, donc seul moi pourrai accéder au message.

Le chiffrement asymétrique *Définition*

Le **chiffrement asymétrique** vient concrétiser la différence entre la clé de chiffrement et de déchiffrement. En pratique, la clé de chiffrement sera nommée **clé publique** car elle sera librement communiquée. La clé de déchiffrement sera nommée **clé privée** car elle ne doit être communiquée sous aucun prétexte.



Fondamental

- N'importe qui pourra utiliser la clé publique d'une personne pour lui envoyer un message (cette clé publique est connue de tous).
- Ce message ne sera déchiffrable qu'avec la clé privée de cette même personne (qui n'est détenue que par elle-même).

Schéma d'utilisation classique

Exemple

Bob veut envoyer un message à Alice :

1. La clé publique d'Alice est disponible sur un site web.
2. Bob la récupère et chiffre son message grâce à cette clé.
3. Bob envoie le message chiffré à Alice.
4. Alice reçoit puis déchiffre le message grâce à sa clé privée.

Remarque

A l'issue de la communication, Alice et Bob sont certains que leur communication a été confidentielle... à moins qu'Alice n'ait pas correctement protégé la confidentialité de sa clé privée (à cause d'une erreur ou d'une attaque).

Chiffrement RSA

Exemple

Le chiffrement RSA est l'un des algorithmes les plus connus lorsque l'on parle de chiffrement asymétrique. Il est utilisé dans de nombreux contextes notamment :

- Par le protocole SSH, qui permet d'accéder à un serveur distant de manière sécurisée.
- Pour transmettre une clé de chiffrement symétrique. La clé de chiffrement symétrique doit rester confidentielle : elle est transmise à l'aide d'un chiffrement asymétrique lorsqu'elle doit être partagée à un tiers autorisé.

Générer ses propres clés

Méthode

Il existe plusieurs manières de générer des clés RSA. La plus simple est de générer des clés SSH qui sont déjà des clés RSA. Le plus souvent la paire de clés sont nommées de la manière suivante : `nom_clé.pub` pour la publique et `nom_clé` pour la privée. Il est commun de posséder plusieurs paires de clés. On préfère garder une nomenclature similaire pour le nom des clés pour ne pas les mélanger. Ces clés pourront être directement utilisées *pour accéder à un serveur distant.*

```
1 ssh-keygen
```

Pour la suite, on génère les clés RSA directement avec `openssl`.

```
# Génère une clé privée protégée par mot de passe dans le fichier private.pem
openssl genrsa -des3 -out private.pem 2048
# Extrait la clé publique de la clé privée dans le fichier public.pem
openssl rsa -in private.pem -outform PEM -pubout -out public.pem
```

Utiliser une phrase secrète ou un mot de passe *Conseil*

Il est possible de sécuriser sa clé privée en spécifiant un mot de passe lors de la génération de la clé.

Ainsi, la clé privée est elle-même chiffrée par chiffrement symétrique.

Une telle pratique permet de conserver la sécurité même si un attaquant réussit à récupérer la clé privée. Il est tout de même fortement conseillé de générer une nouvelle clé si sa clé actuelle est compromise.

Paire de clés RSA *Exemple*

Voici à quoi ressemble une clef privée RSA, protégée par mot de passe (**il ne faut jamais publier une telle clé si elle est en service).**

```
-----BEGIN RSA PRIVATE KEY-----
Proc-Type: 4,ENCRYPTED
DEK-Info: DES-EDE3-CBC,EB9CF5A7B25026DE

/nJK7OYp0kT26M3bDlTcO6RVAcoIBpZ8oA18p78Nl0mWOQPE0oZTaDJbTuEQLNNW
UWL7c+D7unGg+UdO7DWc52Z4M8JgoYhy7hvZ1BzFs4bUxJ+OP0sWWCyE9eWScK+4
vXXrEhp6JhUSx654cbyPtUYwVGdxKoq25TzEDvA5eSQU1MH/LviNErTYLtZcbPjw
vrbeOivbcSjcTzB83cdgeM/y+ourduIqWjwM1BkhgwcZ5gN6jdWPFMKGe9iF2gsy
IRvfAAjbzmJVM2DXEjHaSSRkXMIusdJN1lyEuH833XGBql9RmslQ9BzjUuYr1Fzd
A1n9EBvaY3pT8UPa3epFwO9kT7ag2Hrx7jYNtlUk/7kyPnm95cJWUvBvGrZxhCDD
4+lsCWrIRSslI5RnlgZQY/FgPOT2blesgkk2Ize/10PflM6w2fPfSsSMKYE9dEY8
zc3F+WwslxopQkWmRYZ8wwzv3Tng60/3DIU4X2oUPPajTFEjIk/KZ5tiHf6bSTcK
VQtE6w/bIHOyu/4ge7EfBDTZAuLGKqtfMuT8Dcz5behK/TgcLEq+4ps2KUu4f/rE
aGcI9sGSYmrRn18bRPATb+VkfKH8ak48zj43UBfrXhTRUlCx7FMDwfu9Csn5aEDl
dOANQ0sz51E6h3DDGHJU8Z0o5mQtFvcPyw/s2uc5ooTlfQr0e+36EQHSqfJIEVzO
Y1Cm2jq+bTPuI4yAp1rlEng3Z+jS3FIkP9lP8Iz7yhvYWhGc0bK0gebGXO2uFOm
6X6eObQQkQ+sbdLQ7wCe3J4UmOEkq6wmvBlt5ZwH+Eo2icKC++434olMxCBKHAPH
2uXwNbZORhqP77Pl7dNge8bhZtdAqWG0Mb8DiGHYweSgC6m8h3xYAtQDdnWyDY4D
0qpLjI+mw+EVk8aMoNVTO3sDWuArsJVYOpoIfQ7/tnBhmJ1Rr9Axr8woDMA6Z501
7zvw4g4zYID61v5hl+xvAO+BUHwGIwuJ92gnSm4WOPk63GLiu0ita73IJgdL6FSG
bsrcmPOLa0+ATCU6CMRLJ3mzefJOviwThkC3P88Fy8kjSg3MTgnXASqqfEix8+Y2
eLtXJmWQXfhWdT24+f4EUEU5+YIy50qL8VGNdRgqKgclC6nUM1uW5txTEqs+fTjH
NhTr5s7gPMWbQYwPkcdNhPaxNhrykTFEfmKL6AEzx64sGn3/kzuqHv0/cH2jR3OR
sp6nd4iarbPYk9zKbUbiBMNvpjegbJ7WLcEllWHMApwfNTufJp0MHIk8AymjbZth
oTDvFZ6BThSW3kJMz31gBX18PswYBUs3bB7FTC3fXDxoSjHk3ClW4aclGIdbWAba
BJNTSieLL6TTlQ3cAD2bb3xJDnAuHScZ+tauu83YDTGfiUsQ4ew9alqypI+dbrPc
utSu9PVOb/8F+MZENW+DtLJ0IvZgcTj0diOqL5Cd1S4J+v/0s60XjyuTBDl1Yj8U
DID6IHFDm/qIfFcNfNzxTgag+nOwXFnn/d5S38YGHiokgXNCugeYw/HeekBP69tm
VjCq2YzIXnOepfYron5KaMt+4y1leXdJzNPHU+7FAr0w6k/IC4r+wg==
-----END RSA PRIVATE KEY-----
```

Voici une clef RSA publique : c'est celle que l'on partage avec des tiers pour des communications.

```
-----BEGIN PUBLIC KEY-----
MIIBIjANBgkqhkiG9w0BAQEFAAOCAQ8AMIIBCgKCAQEAv/Lg6dBGCZKsuIHv/eJi
B2izeWybrH9HjVEwpMBfpS8Nh02hXNPaw547HXnsBC47yTCrNoDDKPBgIIJ0mLNE
R9XhiryK3ni7yo/lE3qE2YKE905a4ALrjejk9J+afnSAfmcfGkfLYl0n2mKo+hy7
xok4/sDLtv938Bbsyg/V1o+YeSucHIXe+b7S6dKR3fOwHw/TB3R+1Qu/nl/RIv7s
OnkRwBxs/NvO6Na+Xny3f0PxpWjjL+CCTePRYaZ/VrNI/Uym3vGGhLzDrlDPu8IB
0S8oF9xgQSHxUrHWNKJ6k0+Fjr5BzoUnrqy6ipC5A1vEEyYRc95ZyKI0vkAYRihI
ewIDAQAB
-----END PUBLIC KEY-----
```

Chiffrer et déchiffrer

Méthode

Pour chiffrer et déchiffrer un message, il est nécessaire d'utiliser les commandes ci-dessous.

Pour chiffrer un fichier :

```
openssl rsautl -encrypt -pubin -inkey public.pem -in msg.txt -out msg.enc
```

Pour chiffrer une chaîne de caractères :

```
echo "Message confidentiel" | openssl rsautl -encrypt -pubin -inkey public.pem -out
msg.enc
```

Pour déchiffrer un message :

```
openssl rsautl -decrypt -inkey private.pem -in msg.enc -out message_en_clair.txt
```

Chiffrer avec la clé privée pour signer

Complément

Il est possible d'inverser le rôle des clés en chiffrant avec la clé privée et en déchiffrant avec la clé publique. Le but d'une telle pratique n'est pas de garder le message secret mais de vérifier l'identité de l'expéditeur. Seul le propriétaire de la clé privée peut générer un message déchiffrable avec la clé publique, ce qui garantit son identité.

À retenir

- Le chiffrement asymétrique permet un partage de la clé de chiffrement tout en garantissant la confidentialité de la clé de déchiffrement.
- La clé publique d'une personne est disponible par ceux qui veulent lui envoyer des messages.
- La clé privée d'une personne n'est jamais publiée et est utilisée par la personne pour déchiffrer les messages reçus.
- L'algorithme RSA permet de réaliser un chiffrement asymétrique.

HTTP et HTTPS

1. Le protocole HTTP n'est pas chiffré

Objectifs

- Connaître le protocole HTTP ;
- Connaître les problèmes de sécurité du protocole HTTP.

Mise en situation

Le protocole HTTP permet de naviguer sur le Web mais il n'a recours à aucune mesure cryptographique. Il n'offre aucune assurance quant à la confidentialité et à l'intégrité des communications.

Toutes les informations communiquées via HTTP peuvent être consultées ou travesties par des tiers.

Ainsi, si on se limite à l'usage d'HTTP, un mot de passe ou un numéro de carte bleue communiqués via un formulaire web sont lisibles par tous les ordinateurs qui se chargent de la communication, ainsi que par tout attaquant qui souhaite espionner les échanges.

Le protocole HTTPS a permis de pallier cette limitation.

Installation d'un serveur web *Rappel*

Pour expérimenter les requêtes HTTP, il est fortement conseillé d'avoir un serveur web installé sur un VPS, comme Apache ou Nginx.

Dans les commandes qui suivent, il faudra simplement remplacer les URL par l'IP du VPS.

Voici comment installer très rapidement un serveur Nginx sous Debian ou Ubuntu:

```
# Installe Nginx
apt install nginx
# Nginx démarre automatiquement.
# Si ce n'est pas le cas :
systemctl start nginx
```

Pour un serveur Apache :

```
# Installe Apache
apt install apache2
# Nginx démarre automatiquement.
# Si ce n'est pas le cas :
systemctl start apache2
```

Le protocole HTTP *Rappel*

Le protocole HTTP (*Hypertext Transfer Protocol*) est un protocole client-serveur standardisé par le W3C permettant d'accéder à des sites web. Les navigateurs sont les clients HTTP les plus connus. Plus généralement, ce protocole rend possible le transfert de données structurées du serveur vers le client ou du client vers le serveur.

Faire une requête HTTP

Rappel

La commande `curl` permet de lancer une requête HTTP. En ajoutant l'option `-v`, la requête et le retour du serveur seront affichés.

```
curl [URL ou IP du serveur]
```

Exemple

```
curl xkcd.com
```

```
<!DOCTYPE html>
<html>
<head>
<link rel="stylesheet" type="text/css" href="/s/7d94e0.css" title="Default"/>
<title>xkcd: Alive Or Not</title>
<meta http-equiv="X-UA-Compatible" content="IE=edge"/>
<link rel="shortcut icon" href="/s/919f27.ico" type="image/x-icon"/>
<link rel="icon" href="/s/919f27.ico" type="image/x-icon"/>
<link rel="alternate" type="application/atom+xml" title="Atom 1.0" href="/atom.xml"/>
<link rel="alternate" type="application/rss+xml" title="RSS 2.0" href="/rss.xml"/>
<script type="text/javascript" src="/s/b66ed7.js" async></script>
<script type="text/javascript" src="/s/1b9456.js" async></script>

<meta property="og:site_name" content="xkcd">

<meta property="og:title" content="Alive Or Not">
<meta property="og:url" content="https://xkcd.com/2307/">
<meta property="og:image" content="https://imgs.xkcd.com/comics/alive_or_not_2x.png">
<meta name="twitter:card" content="summary_large_image">

</head>
<body>
<div id="topContainer">
<div id="topLeft">
<ul>
<li><a href="/archive">Archive</a></li>
<li><a href="http://what-if.xkcd.com">What If?</a></li>
<li><a href="http://blag.xkcd.com">Blag</a></li>
<li><a href="/how-to/">How To</a></li>
<li><a href="http://store.xkcd.com/">Store</a></li>
<li><a rel="author" href="/about">About</a></li>
<li><a href="/atom.xml">Feed</a> &bull; <a href="/newsletter/">Email</a></li>
</ul>
</div>
<div id="topRight">
<div id="masthead">
<span><a href="/"><img src="/s/0b7742.png" alt="xkcd.com logo" height="83" width="185"/></a></span>
<span id="slogan">A webcomic of romance,<br/> sarcasm, math, and language.</span>
</div>
<div id="news">
```

On reçoit le code HTML de la page d'accueil du site `xkcd.com`.

Absence de confidentialité

Fondamental

Le protocole HTTP fonctionne sans aucun chiffrement.

Les requêtes sont envoyées en clair sur le réseau. Tous les tiers observant les paquets peuvent connaître le contenu des communications. Ces tiers voient la requête et la réponse du serveur.

Man in the middle

Attention

Un modèle classique d'attaque est le *man in the middle*. Il s'agit d'une attaque où l'attaquant se place au milieu d'un canal de communication et observe ou modifie les communications. Avec le protocole HTTP, il est très simple de réaliser ce type d'attaque compte-tenu de la non-confidentialité des communications.

Cette situation est dangereuse dans la mesure où les requêtes HTTP peuvent transporter des identifiants.

Comment le flux HTTP peut-être observé par un tiers ? *Méthode*

Les tiers pouvant observer les communications HTTP sont :

- soit les ordinateurs par lesquels passe la requête (routeurs, FAI, etc.),
- soit des attaquants.

Exemple

Voici deux solutions techniques permettant d'intercepter et analyser les paquets :

- **Wireshark** : enregistre le trafic réseau. Par exemple, il est possible d'observer toutes les communications Wi-Fi. Cet outil peut être utilisé pour attaquer des personnes utilisant un réseau Wi-Fi public sans précautions.
- **Suricata** : un système de détection d'intrusion qui réalise une introspection des paquets. Cette outil est utilisé par certains fournisseurs d'accès internet ou entreprises pour détecter des attaques. Cet outil peut extraire les informations de toutes les communications HTTP.

À retenir

- HTTP est un protocole de base pour communiquer sur Internet.
- HTTP n'offre aucune garantie de confidentialité.

2. Le protocole HTTPS

Objectif

- Connaître le protocole HTTPS.

Mise en situation

HTTPS, le « *S* » signifiant *secure*, est la combinaison du HTTP et d'une couche de chiffrement. Les communications effectuées via HTTPS sont authentifiées quant à leurs émetteurs et confidentielles quant à leurs contenus.

HTTPS est né dans le domaine du commerce électronique ou de la banque en ligne, pour lesquels l'échange secret d'informations sensibles est crucial.

Depuis le protocole s'est généralisé dans d'autres sphères, comme les réseaux sociaux, et il devient majoritaire sur le Web.

Lorsque vous consultez un site web, si l'adresse commence par HTTPS et que votre navigateur ne vous signale pas d'erreur, alors c'est que seul le site avec lequel communiquez a accès aux informations que vous envoyez.

Le protocole HTTPS *Fondamental*

Pour communiquer en HTTPS, un serveur doit posséder un **certificat** et une **paire de clés** asymétriques (publique/privée).

Le certificat lui est délivré par une **autorité de certification** et atteste qu'il est légitime (de sorte qu'un tiers ne puisse pas se faire passer pour lui).

Le protocole HTTPS *Méthode*

Une communication web via HTTPS s'effectue de la manière suivante :

1. Récupération du certificat et de la clé publique du site.
2. Vérification de la légitimité du site auprès de l'autorité de certification.
3. Envoi chiffré (chiffrement asymétrique) d'une clé (de chiffrement symétrique) au serveur web
4. Échange HTTP « classique » qui sera chiffré à l'aide de la clé qui vient d'être envoyée.

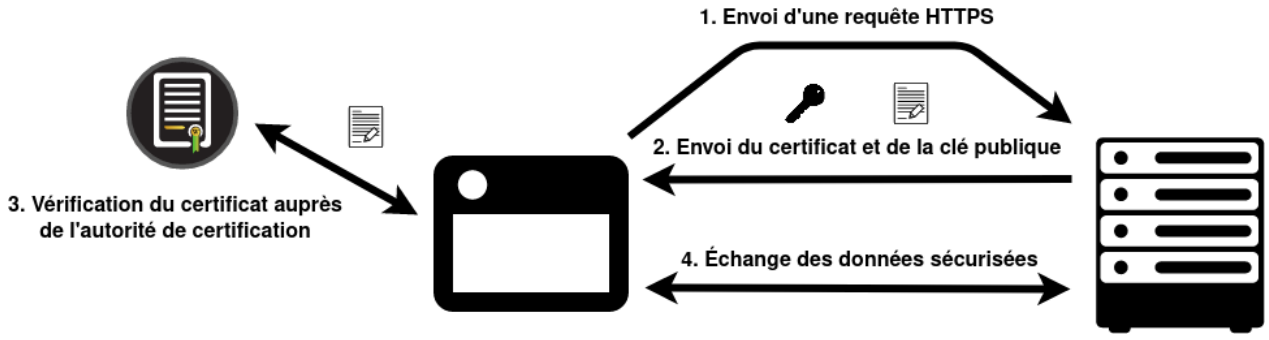


Schéma d'une communication HTTPS

On distingue donc deux parties :

- les échanges préliminaires permettant l'établissement d'une communication sécurisée,
- puis la communication en elle-même.

La confidentialité retrouvée

Toute entité tierce observant le trafic ne verra qu'une suite vraisemblablement aléatoire de caractères. La requête comme la réponse étant chiffrée, le tiers ne pourra donc pas savoir ce que cache cette suite de caractères.

L'intégrité des communications

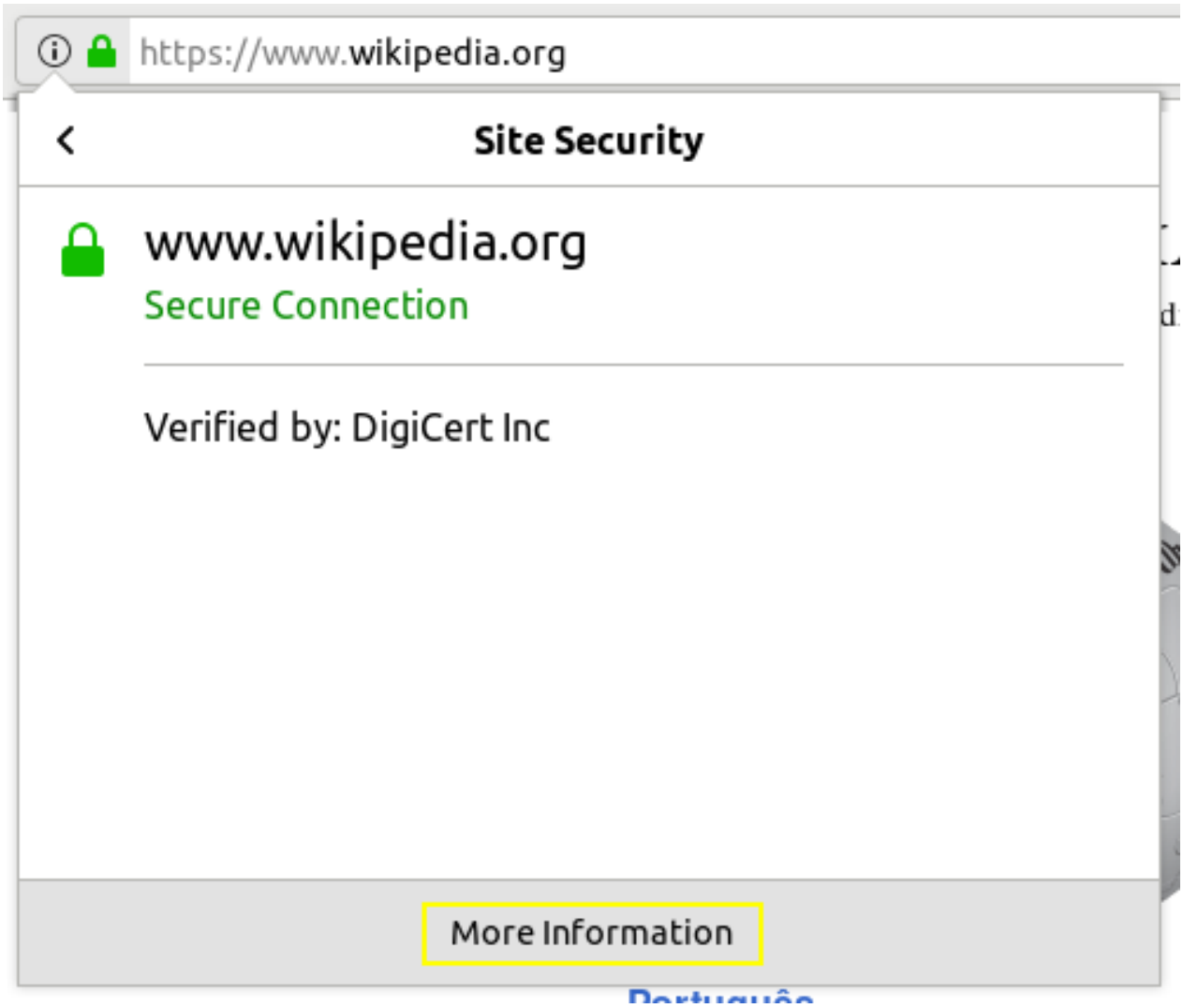
En plus de la confidentialité, HTTPS protège l'intégrité des communications. Un attaquant ne pourra pas intercepter un message et le substituer par un message frauduleux. La réponse et la requête étant chiffrées, il est impossible de les modifier sans la clé de chiffrement. Ainsi, HTTPS conforte l'utilisateur comme le serveur de la sécurité de la communication.

Rôle du navigateur

Les navigateurs web se chargent d'établir la communication sécurisée. Le navigateur récupère automatiquement le certificat, le vérifie et envoie la clé de chiffrement pour sécuriser les communications. En cliquant sur le cadenas, habituellement à gauche de l'URL, il est possible de voir les informations du certificats. On trouve notamment les informations suivantes :

- La date d'expiration du certificat,
- L'autorité de certification, c'est-à-dire l'organisme qui certifie que le site web est légitime,
- Les informations techniques à propos du chiffrement.

Certificat *Exemple*



Ce qui reste visible *Attention*

Le protocole HTTPS permet de chiffrer les communications mais pas de les masquer. Il ne garantit pas l'anonymat.

- Il laisse visible les méta-données de la communication, en particulier la source et la destination.
- Un tiers observant le trafic réseau pourra toujours savoir qu'une communication a lieu (même s'il ignore le contenu de la communication).
- Un fournisseur d'accès internet pourra sans problème connaître l'identité de tout utilisateur effectuant une communication HTTPS à l'aide de l'adresse IP contenue dans les méta-données.
- Par exemple, ce protocole ne protège en rien une personne accédant à des sites proposant des produits illégaux.

Le navigateur Tor *Complément*

Le navigateur et le protocole Tor[5] permettent de garantir une anonymisation du flux réseau. Ainsi, la connexion est sécurisée et l'identité de l'utilisateur protégée. Cet outil est très précieux pour des journalistes ou des opposants politiques vivant dans des régimes autoritaires. Dans la culture populaire, on en entend surtout parler car il permet d'accéder au *Dark Web*.

5. https://www.torproject.org/

À retenir

- Le protocole HTTPS permet de garantir la confidentialité et l'intégrité des communications web.
- Les navigateurs modernes se chargent automatiquement de la partie de chiffrement.
- HTTPS ne garantit pas l'anonymat de l'utilisateur.

3. Limites de HTTPS

Mise en situation

Bien qu'introduisant des mécanismes de chiffrement dans HTTP, HTTPS ne doit pas être vu comme une sécurité absolue sur le web. Il est important de bien comprendre que HTTPS permet de protéger les échanges, mais pas de protéger contre un site malveillant par exemple.

Nous nous attarderons aussi sur la manière dont les certificats sont obtenus, et les changements de politique à ce sujet ces dernières années.

Rappel

Le champ d'action de HTTPS se limite au chiffrement des communications.

Périmètre de protection de HTTPS

HTTPS garantit :

- que le site auquel on accède est bien celui que l'on a demandé (une personne malveillante ne peux pas usurper l'identité du serveur), grâce au certificat
- que les données échangées seront chiffrées entre le client et le serveur

À l'inverse **HTTPS ne garantit pas** :

- que le site consulté n'est pas un site malveillant ou de phishing
- que le site stocke les données de manières chiffrée/sécurisée

Attention

Il est souvent mis en avant que "un site avec le cadenas vert (donc un site utilisant HTTPS) est un site sécurisé". Comme vous pouvez le comprendre maintenant, cette formulation est un dangereux raccourci. Si tant est que l'on puisse définir ce qu'est un "site sécurisé", il est certain que **HTTPS serait une condition nécessaire mais non suffisante**.

Phishing : mabanque.fr et mabnaque.fr *Exemple*

Un premier exemple serait un site malveillant, de phishing, qui tente de voler vos identifiants bancaires. Imaginons que votre banque ai pour adresse `www.mabanque.fr`.

Une personne malveillante pourrait créer une copie conforme (en apparence) de ce site, et l'héberger sur le site `www.mabnaque.fr`. L'objectif serait de jouer sur une faute de frappe ou d'inattention des personnes souhaitant se rendre sur le site de la banque. Cette personne malveillante n'aurait aucun mal à obtenir (puis à présenter) un certificat HTTPS, en effet il est bel et bien propriétaire de `www.mabnaque.fr` ! HTTPS (et le certificat) vous garantit simplement que vous aller envoyer vos informations sur le serveur du site malveillant, et pas que c'est bien le site de votre banque.

Authentification : résultat de laboratoire ? *Exemple*

Un second exemple, rapide, serait le site d'un laboratoire qui propose de récupérer vos résultats d'examens médicaux. Sur sa page d'accueil le site affiche, tout fier, être "entièrement sécurisé". Vous constatez que le site utilise bien HTTPS mais ne propose pas d'authentification : tous les résultats d'examens de tous les patients sont accessibles publiquement.

On voit ici aussi que HTTPS ne garantit en rien que la gestion du site soit réellement sécurisée.

Stockage : panier d'achat ? *Exemple*

Si vous communiquez avec un site de vente en ligne via HTTPS :

- les données sont bien chiffrées et déchiffrées par le site ;
- rien ne garanti que ses serveurs ne seront pas victimes d'intrusion par des tiers ;
- ou qu'il ne partage pas ces données volontairement avec des tiers.

Obtention des certificats

Les certificats peuvent être créés par n'importe qui, mais il est nécessaire de passer par une autorité de certification (AC) pour que celui-ci soit reconnu par les navigateurs.

Les AC ont donc une responsabilité importante : elles ne doivent attribuer des certificats qu'aux personnes ayant prouvé leur identité, à savoir qu'elles sont bien propriétaires du nom de domaine qu'elles souhaitent certifier.

Pendant des années, le processus pour obtenir un certificat était donc le suivant : le propriétaire d'un nom de domaine devait attester auprès d'une autorité de certification de son identité civile, et prouver que c'était bien lui qui était en possession du nom de domaine. Le processus était souvent manuel (et parfois même, par un canal hors-ligne), prenait quelques heures ou jours, et coûtait cher. Une autorité de certification classique facturait, en général, plusieurs centaines d'euros par an pour un certificat.

Du fait de la lourdeur et du coût de la procédure, HTTPS a ainsi longtemps été réservé à des professionnels et/ou des acteurs pour qui les enjeux de sécurité étaient important. Ceci a fortement contribué à l'image "cadenas vert/HTTPS = site sécurisé".

Let's Encrypt, la démocratisation de HTTPS

En 2014, l'Electronic Frontier Foundation (EFF) créé, en coopération avec l'université du Michigan et Mozilla, l'organisation à but non lucratif, **Internet Security Research Group** (ISRG). L'objectif de cette organisation est, entre autre, de lancer le projet **Let's Encrypt** : une autorité de certification permettant de rendre accessible l'obtention d'un certificat, et donc de généraliser l'adoption de HTTPS.

En 2015 l'autorité de certification **Let's Encrypt** est lancée et obtient rapidement les certifications nécessaires pour être présente dans tout les navigateurs. Let's Encrypt propose à n'importe quelle personne possédant un nom de domaine d'obtenir, de manière entièrement automatisée, un **certificat gratuit** pour son domaine. Le changement est majeur : on passe d'une procédure complexe et coûteuse à une procédure automatisée, simple, et gratuite.

L'adoption est massive, et HTTPS se généralise, ne restant plus cantonné à des sites professionnels. En 2020, Let's Encrypt annonce[6] fournir des certificats pour plus de 225 millions de nom de domaines, et les analyses du site Censys[7] estiment que plus de 50% des certificats TLS du monde sont fournit par Let's Encrypt.

6. Let's Encrypt - https://letsencrypt.org/stats/#
7. Censys - https://censys.io/certificates/report?q=tags%3Atrusted&field=parsed.issuer.organization.raw&max_buckets=50

Enjeux politique sur les autorités de certification

Les autorités de certification ont, comme nous l'avons vu, un rôle primordial et critique dans la mise en place d'une chaîne de confiance, et donc de HTTPS. La possibilité d'émettre des certificats qui seront acceptés par les navigateurs est un pouvoir très important : si une AC décide de créer un certificat sans vérifier l'identité du demandeur, ou tout simplement pour usurper l'identité d'un site, elle en a la possibilité.

Lorsque les autorités de certification sont des agences gouvernementales ou de grandes entreprises, on peut souligner le fait que ce pouvoir risque d'être utilisé à des fins non-légitimes (politique, espionnage industriel, etc.). De la même manière, lorsqu'une AC ne sécurise pas un minimum ses procédures et ses certificats, un attaquant peut les récupérer et s'en servir pour créer de faux certificats. Et, malheureusement, ce genre d'abus et d'incidents ont déjà été observé.

Exemple

En 2012, la société de transport publics de Ankara (EGO), en Turquie, a mis en place un certificat lui permettant d'usurper l'identité de tout les sites visités en HTTPS sur son réseau local. Le but étant d'espionner le contenu des sites visités par les employés. Ce faux certificat lui a été fournit par TurkTrust, une autorité de certification gouvernementale turque, qui, à l'époque, était présente dans la liste des certificats de confiance des principaux navigateurs.

Un article résume entièrement l'incident[8] et, même si il semble qu'il n'y ai pas eu de malveillance volontaire de la part de TurkTrust, l'autorité de certification a clairement faillit à son rôle ici.

À retenir

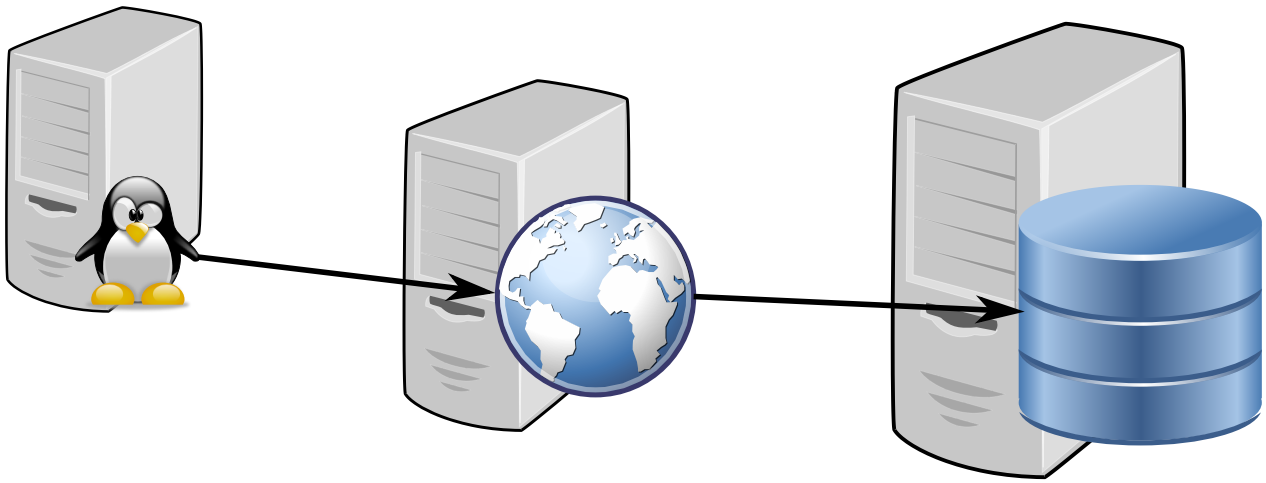
- HTTPS n'est pas une solution miracle, c'est un protocole qui permet de chiffrer les communication HTTP, rien de plus.
- Les autorités de certification ont un rôle crucial et un pouvoir conséquent dans la mise en place de ce protocole.

8. https://nakedsecurity.sophos.com/2013/01/08/the-turktrust-ssl-certificate-fiasco-what-happened-and-what-happens-next/

Tracking

1. Principes

Architecture 3-tiers



Comment obtenir des données ? I *Méthode*

Il suffit de les demander...

Base de données (données utilisateur) *Exemple*

ip	name	age	mail	city	...
91.224.148.57	Alice	23	alice@utc.fr	Compiègne	...
54.225.79.234	Bob	46	bob@protonmail.com	New York	...
91.224.148.59	Charlie	12	charlie@crzt.fr	Paris	...
...	...	...	...	...	...

Comment obtenir des données ? II

Méthode

Le protocole HTTP implique l'échange de méta-données...

Base de données (méta-données)

Exemple

ip	date	hour	referer	User-Agent	...
91.224.148.57	2021-12-14	13:45:56 (UTC+0100)	wikipedia.fr	Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:94.0) Gecko/20100101 Firefox/94.0	...
54.225.79.234	2021-12-14	13:45:57 (UTC+0100)	chatons.org	Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/96.0.4664.45 Safari/537.36	...
...					

User-Agent

Complément

Quand une requête HTTP est envoyée (par exemple par un navigateur web) une chaîne de caractères est transmise au serveur pour identifier l'application logicielle cliente.

L'en-tête HTTP dispose d'un champ `User-Agent` prévu pour cela.

Exemple d'informations : le nom de l'application (Mozilla), la version (5.0), le système d'exploitation (Ubuntu)...

Comment obtenir des données ? III

Méthode

Croiser les informations entre plusieurs bases de données...

Le problème de l'identification

Méthode

- Compte utilisateur
- IP : ne fonctionne pas bien
  - plusieurs personnes derrière une IP (toutes les machines clientes d'une organisation sont souvent vues comme la même IP depuis l'extérieur)
  - plusieurs IP pour une même personne (domicile, travail, mobilité...)
- Outils de traçage complémentaires : exemple des cookies

Autres technique d'identification unique en l'absence d'authentification

Complément

- Adresse MAC des cartes Wifi (pas en Web pur, si l'on utilise un logiciel qui accède à la machine, ce qui est souvent le cas sur mobile).
- Autres identifiants sur mobile (Android Advertising ID, International Mobile Equipment Identity...)
- Canvas fingerprinting

Exemple : https://browserleaks.com/canvas

2. Cookies

Cookie

Définition

« Un cookie HTTP (cookie web, cookie de navigateur) est un petit ensemble de données qu'un serveur envoie au navigateur web de l'utilisateur. Le navigateur peut alors le stocker localement, puis le renvoyer à la prochaine requête vers le même serveur. Typiquement, cette méthode est utilisée par le serveur pour déterminer si deux requêtes proviennent du même navigateur. »

https://developer.mozilla.org/fr/docs/Web/HTTP/Cookies

Remarque

« Les cookies permettent de conserver de l'information en passant par le procotole HTTP qui est lui "sans état". »

https://developer.mozilla.org/fr/docs/Web/HTTP/Cookies

Exemple

Sans cookie il faudrait entrer ses informations de connexion à chaque fois que l'on consulte une nouvelle page d'un site, lorsque celui-ci nécessite une connexion (consultation de compte en banque, logiciel de gestion dans l'entreprise...).

À quoi servent les cookies

Méthode

- Garder un utilisateur connecté (historique de login).
- Conserver des informations liées au bon fonctionnement de la session en cours, même si l'on est pas connecté (par exemple un panier d'achat).
- Enregistrer des paramètres de personnalisation : préférences utilisateur, thèmes...
- Enregistrer des informations de suivi de l'utilisateur : analyse du comportement utilisateur.

Exemple

« Certains cookies sont utilisés pour conserver les préférences d'un utilisateur. Par exemple, un cookie appelé "NID" est enregistré dans le navigateur de la plupart des utilisateurs des services Google. Ce cookie contient un identifiant unique permettant de mémoriser vos préférences et d'autres informations, comme le choix de la langue, le nombre de résultats de recherche à afficher par page [...] »

« YouTube utilise le cookie "PREF" pour stocker des informations telles que les préférences de l'utilisateur pour la configuration des pages et la lecture (lecture automatique, contenus en lecture aléatoire et taille du lecteur) [...] »

« Certains cookies améliorent les performances des services Google. Par exemple, les cookies "CGIC" améliorent la génération des résultats de recherche Google grâce à la saisie semi-automatique basée sur la requête initiale de l'utilisateur. Ce cookie reste actif pendant six mois. »

https://policies.google.com/technologies/cookies

3. Exercice : Cookie demo I

https://www.utc.fr/~crozatst/cookie/first/

Question

[solution n°1 p. 47]

Sous Firefox :

- tapez F12
- choisissez l'onglet `Stockage`
- regardez les cookies du domaine `www.utc.fr` stockés sur votre navigateur

4. Cookies tiers

Cookie tiers *Définition*

« Les cookies ont un domaine qui leur est associé. Si ce domaine est le même que la page sur laquelle vous êtes, on parle de cookie interne (first-party cookie). Si le domaine est différent, on parle de cookie tiers (third-party cookie). »

« Alors que les cookies internes sont uniquement envoyés au serveur qui les a définis, une page web peut également contenir des images ou tout autre composant stockés sur d'autres domaines (comme des bannières publicitaires). Les cookies qui sont envoyés via les composants tiers sont appelés cookies tiers et ils sont principalement utilisés pour la publicité et le suivi sur le web. »

https://developer.mozilla.org/fr/docs/Web/HTTP/Cookies

Exemple

« Les cookies utilisés à des fins d'analyse aident à collecter les données permettant de mieux comprendre les interactions des internautes avec un service particulier. [...]

Nous utilisons le cookie "NID" pour présenter aux utilisateurs non connectés des annonces Google Ads dans les services Google. »

https://policies.google.com/technologies/cookies

Informer les utilisateurs des cookies utilisés *Méthode*

https://foundation.wikimedia.org/wiki/Cookie_statement

RGPD

1. Données à caractère personnel

Données à caractère personnel *Fondamental*

Le terme **donnée à caractère personnel** est défini par l'article 4 du règlement général sur la protection des données (RGPD) :

> « Toute information se rapportant à une personne physique identifiée ou identifiable (ci-après dénommée « personne concernée ») ; est réputée être une « personne physique identifiable » une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale. »

Types de données à caractère personnel *Définition*

On peut diviser les données personnelles en 3 catégories :

- Des données directement identifiantes : élément permettant d'identifier clairement l'identité de la personne comme une fiche de paie ou un fichier client.
- Des données indirectement identifiantes : élément qui, associé à une base de données tierce, permet l'identification d'une personne. Exemple : numéro de téléphone ou numéro de sécurité sociale.
- Toute combinaison d'informations permettant d'identifier la personne : à l'ère du Big Data, il est possible d'identifier une personne en recoupant de nombreuses informations qui prises une à une ne le permettraient pas. Il peut s'agir du sexe, du lieu d'habitation, de l'âge, du métier, etc.

Supports contenant des données à caractère personnel *Exemple*

Questionnaire papier, photo, vidéo, fichier informatique, etc.

Une donnée n'est pas forcément numérique *Remarque*

La notion de données à caractère personnel dépasse le support informatique. Récupérer une information ou la stocker au format papier ne change pas les obligations des acteurs du traitement. L'informatique a décuplé l'impact d'une utilisation malveillante des données à caractère personnel mais une information au format papier n'en reste pas moins sensible.

Traitement de données *Fondamental*

Le terme **traitement** est défini par l'article 4 du RGPD :

> « Toute opération ou tout ensemble d'opérations effectuées ou non à l'aide de procédés automatisés et appliquées à des données ou des ensembles de données à caractère personnel, telles que la collecte, l'enregistrement, l'organisation, la structuration, la conservation, l'adaptation ou la modification, l'extraction, la consultation, l'utilisation, la communication par transmission, la diffusion ou toute autre forme de mise à disposition, le rapprochement ou l'interconnexion, la limitation, l'effacement ou la destruction. »

Opérations concernées *Exemple*

Voici des opérations qui sont concernées par le terme **traitement de données** :

- Consultation d'un fichier client,
- Stockage d'un message parvenu par un formulaire de contact,
- Modification d'informations personnelles,
- Etc.

À retenir

- Les données à caractère personnel sont toute information ou combinaison d'informations permettant d'identifier directement ou indirectement une personne.
- Le terme traitement de données englobe les opérations automatisées possibles sur des données.

2. Présentation du RGPD et de la CNIL

Rôle du RGPD *Fondamental*

Le règlement général sur la protection des données (RGPD) impose un cadre aux traitements de données à caractère personnel. Il définit les conditions dans lesquelles les données peuvent être légalement collectées, stockées et exploitées.

Axe majeurs du RGPD

- **Renforcement des droits des personnes** : l'objectif est de fournir des droits plus importants et plus nombreux aux utilisateurs.
- **Nouvelle logique de responsabilisation des acteurs du traitement des données** : les acteurs impliqués dans un traitement de données sont explicitement identifiés et responsabilisés. Une telle pratique facilite les recours en cas de problèmes (une fuite de données par exemple) car la responsabilité de chaque partie du traitement est définie à l'avance.
- **Renforcement des pouvoirs de sanction** : des sanctions de grande ampleur sont possibles pour que même les géants du Web y soient sensibles. Certaines des entreprises les plus riches et puissantes du monde ont un modèle économique intégralement basé sur le traitement de données personnelles ; il est légitime que des sanctions à la taille de ces acteurs soient possibles.

Quand est-on concerné par le RGPD ? *Méthode*

Un organisme doit appliquer le RGPD :

- s'il est basé en Union Européenne,
- ou si son traitement de données touche une personne qui se trouve en Union Européenne.

Exemple

- Ainsi, une entreprise française vendant l'ensemble de ses services aux États-Unis devra appliquer le RGPD.
- De même, une entreprise chinoise recueillant les données de français devra suivre le RGPD.

Qui est concerné ? *Attention*

Les acteurs concernés par le RGPD sont tous les organismes publics et privés :

- entreprises (grandes ou petites),
- administrations,
- collectivités et associations.

Il existe une exception domestique qui permet le traitement de données personnelles dans un cadre strictement individuel. Par exemple, l'enregistrement de contacts dans son téléphone n'aura pas à suivre les obligations du RGPD.

La CNIL

La commission nationale Informatique et Liberté (CNIL) a été créée en 1978 et a pour but d'assurer que l'informatique est au service de tous sans porter atteinte aux libertés individuelles. En conséquence, la CNIL est chargée notamment d'émettre un avis sur tous les projets publics impliquant un traitement de données. Plus généralement, elle a pour mission le respect des lois visant à protéger les libertés et la vie privée des personnes dans le cadre d'un traitement informatique.

La CNIL est devenue centrale avec l'adoption du RGPD qui consolide son rôle de protecteur des libertés individuelles en France.

Et en Europe ? *Complément*

Dans le cadre du RGPD, chaque pays européen a son équivalent de la CNIL avec les missions suivantes : informer, accompagner, anticiper, contrôler et sanctionner. Au niveau européen, le comité européen de protection des données (CEPD) coordonne les « CNIL européennes » grâce à des avis et des décisions. La Cour de Justice de l'Union Européenne (CJUE) veille à l'application uniforme dans toute l'Union Européenne. Elle peut également sanctionner une entité qui ne respecte pas le RGPD.

À retenir

- Le RGPD vise à protéger les libertés individuelles.
- Toute organisme public ou privé traitant des données personnelles doit respecter le RGPD.
- La CNIL est chargée de l'appliquer en France.

3. Principe de responsabilité et devoirs des responsables de traitements

Principe de responsabilité (accountability) *Fondamental*

La logique d'*accountability* (traduite **responsabilité** en français) demande aux organismes de continuellement s'interroger sur la conformité de leurs traitements et d'être en capacité de prouver leur conformité au RGPD.

Cette logique se traduit notamment par le respect de deux notions du RGPD :

- *Privacy by design* : la protection des données doit être au centre du projet dès sa conception.
- *Privacy by default* : par défaut, la quantité de données collectées doit être minimale même s'il reste possible de proposer à l'utilisateur de fournir des données supplémentaires.

Responsable de traitement *Définition*

Le responsable d'un traitement est une personne morale comme une entreprise, une association ou une administration. Le responsable peut être représenté par un membre de sa direction mais cette personne ne sera pas responsable face à la loi.

Partage des responsabilités *Remarque*

Le RGPD introduit formellement la possibilité d'une co-responsabilité du traitement.

On parle de **co-responsabilité** lorsqu'au moins deux responsables de traitement s'accordent sur la finalité et les moyens d'un traitement. Par exemple, si des compagnies aériennes et hôtelières s'accordent pour la création d'une plate-forme de réservation commune, on parlera de co-responsabilité. Il est nécessaire de fournir les grandes lignes de l'accord aux personnes concernées par le traitement.

Les sous-traitants *Complément*

Le cas spécifique des sous-traitants est encadré par le RGPD. Les organismes « donneurs d'ordre » ne doivent faire appel qu'à des sous-traitants présentant des mesures techniques et organisationnelles garantissant la protection des données.

Le sous-traitant possède quatre obligations :

- **Transparence et traçabilité** : posséder une documentation et un registre détaillant les traitements de données de son client (le donneur d'ordre).
- **Sécurité des données** : mettre en place un système d'autorisation pour l'accès aux données.
- **Respect des conditions de sous-traitance ultérieure** : un sous-traitant peut lui-même faire appel à un sous-traitant en respectant des conditions.
- **Accompagnement du responsable de traitement** : le sous-traitant doit aider le donneur d'ordre à s'acquitter de ses obligations (sécurité des données, analyse d'impact sur la vie privée, etc.).

Complément

La CNIL met à disposition un guide du sous-traitant[9] afin de faciliter l'édition des clauses du contrat de sous-traitance.

9. https://www.cnil.fr/fr/reglement-europeen-sur-la-protection-des-donnees-un-guide-pour-accompagner-les-sous-traitants

Sanctions *Complément*

La CNIL peut décider de sanctions telles qu'un rappel à l'ordre, une injonction à la mise en conformité, une limitation du traitement ou encore une amende administrative.

L'amende administrative est graduelle et son montant dépend de la gravité du manquement au RGPD. Le montant maximal est :

- 20 millions d'euros ou 4% du chiffre d'affaires mondial en cas de non-respect des principes fondamentaux du RGPD
- 10 millions d'euros ou 2% du chiffre d'affaires mondial en cas de non-respect des dispositions du délégué à la protection des données.

À retenir

- Les organismes sont au centre de la conformité de leurs traitements.
- La co-responsabilité et la sous-traitance sont spécifiquement encadrées par le RGPD.
- Le RGPD fournit un arsenal de sanctions très large à la CNIL qui pourra toucher même les plus grandes entreprises.

4. Les huit règles d'or du RGPD

Fondamental

1. la finalité du traitement,
2. la licéité du traitement,
3. la minimisation des données,
4. la protection particulière de certaines données,
5. la conservation limitée des données,
6. l'obligation de sécurité,
7. la transparence,
8. les droits des personnes.

Finalité du traitement

La finalité du traitement doit être légitime et explicitement définie. La finalité du traitement doit être communiquée à l'avance et il est interdit de détourner la finalité d'une donnée.

Licéité du traitement

Le traitement doit être licite. Six conditions de licéité ont été définies afin d'expliciter cette notion.

Exemple

Par exemple :

- la personne a consenti au traitement de ses données à caractère personnel pour une ou plusieurs finalités spécifiques,
- le traitement est nécessaire au respect d'une obligation légale à laquelle le responsable du traitement est soumis,
- ...

Minimisation des données

Seules les données nécessaires à la finalité du traitement peuvent être collectées.

Exemple

Par exemple, il est interdit pour un service de livraison de demander des informations médicales. En effet, dans le cadre d'une livraison, les informations médicales n'aideront pas à atteindre la finalité. En revanche, le domicile du client est pertinent.

Protection particulière de certaines données

Les données sensibles doivent être collectées et traitées en respectant des conditions strictes.

Exemple

Par exemple, les données médicales nécessiteront des conditions bien spécifiques de traitement et de stockage car une fuite de ce type de données serait bien plus grave que la fuite d'un simple fichier client.

Conservation limitée des données

Une fois la finalité atteinte, les données doivent être archivées, supprimées ou anonymisées.

Ces opérations doivent répondre à des obligations légales, notamment sur l'archivage : les hébergeurs ont une obligation légale de conservation des données. Ainsi, une donnée doit être archivée durant un certain temps avant de pouvoir être supprimée. Les données archivées doivent toutefois rester séparées des données encore sujettes au traitement.

Obligation de sécurité

Le responsable de traitement doit mettre en place une sécurité adaptée à la sensibilité des données.

Même pour des données peu sensibles, il faut assurer une sécurité suffisante pour éviter une fuite.

Transparence

Les personnes doivent être informées du traitement de leurs données, de la condition justifiant de la licéité du traitement, la durée de conservation ainsi que de la manière d'exercer leurs droits (information ou suppression par exemple).

Exemple

Par exemple, il est nécessaire de laisser le droit à l'information en fournissant les coordonnées de la personne à contacter en cas de question sur ses données ou leur traitement.

Droits des personnes

Chaque personne dispose de nombreux droits afin de conserver le contrôle de ses données personnelles :

- Droit d'accès
- Droit de rectification
- Droit d'opposition
- Droit à l'effacement

À retenir

- Le RGPD possède 8 règles d'or permettant d'orienter le développeur dans ses choix de conception.
- Ces règles d'or ont toutes un but commun : protéger les libertés individuelles.

5. Le délégué à la protection des données et ses outils

Délégué à la protection des données *Fondamental*

Le délégué à la protection des données (aussi appelé *DPO* pour *Data Protection Officer*) est au centre de la responsabilisation des responsables de traitement.

Le *DPO* peut faire partie de l'entreprise ou être issu d'une entreprise spécialisée.

Le DPO est en contact direct avec la direction de l'organisation pour que les représentants du responsable de traitement soient toujours informés des questions de données. En cas de manquement du responsable de traitement, le *DPO* ne pourra être responsable juridiquement que si la faute est intentionnelle.

Missions du DPO *Méthode*

- **Informer et conseiller** : sensibilise aux obligations touchant à la protection des données.
- **Contrôler la conformité** : recense les traitements et tient à jour un registre de traitement qui est l'outil au centre de la mise en conformité.
- **Assurer l'interface** : entre la CNIL, l'organisme et les personnes concernées par le traitement.

Registre des traitements des données *Fondamental*

Le **registre des traitements des données** est l'outil au centre de la conformité qui doit être tenu par tout organisme traitant des données personnelles.

- Ce registre doit être tenu par les responsables de traitement et les sous-traitants.
- Sa rédaction peut être confié au DPO.

Méthode

Un registre doit répondre aux questions suivantes :

- Qui sont les acteurs intervenant dans le traitement ?
- Quel type de données est traité (médicale, géolocalisation, etc.) ?
- Qui accède aux données ?
- Combien de temps sont-elles conservées ?
- Comment sont-elles sécurisées ?

Analyse d'impact relative à la protection des données (AIPD)

L'AIPD est le second outil du DPO. Il vise à décrire précisément les traitements et à identifier systématiquement les risques sur la vie privée et les libertés. Évaluer ces risques permet de les réduire au maximum en atteignant un niveau acceptable.

Cet outil est très précieux pour prouver sa conformité si la CNIL ou une personne concernée le demande.

Un risque est estimé en fonction de deux critères :

- sa vraisemblance,
- sa gravité.

Par exemple :

- Le fait qu'un salarié extraie des données de l'entreprise est vraisemblable mais avec une faible gravité
- Le fait qu'une attaque informatique massive se produise est peu vraisemblable mais très grave.

Violation des données *Méthode*

Lorsqu'une violation des données a lieu, il est nécessaire :

- de la documenter,
- de la communiquer à la CNIL,
- d'informer les personnes concernées par cet incident.

Une violation peut être notamment une perte ou une fuite de données due à une attaque ou à un problème dû à une mauvaise programmation.

À retenir

- Le DPO est l'acteur phare pour la mise en conformité au RGPD.
- Le registre et l'AIPD serviront au DPO dans sa mission.

Se cacher

1. Je n'ai rien à cacher

HTTPS ne masque pas tout *Rappel*

- Le FAI (domicile, université, employeur...) sait avec qui vous communiquez.
- Les sites visités savent ce que vous leur dites.
- Les informations enregistrées peuvent être partagées avec des tiers.

2. Chiffrement de bout-en-bout

- Signal
- ProtonMail
- Enigmail

Attention

Une fois déchiffré le message est à nouveau vulnérable.

3. Do not track

Les conteneurs multicomptes

« vous permettent de conserver séparés les différents aspects de votre vie numérique dans des onglets repérés par des couleurs qui protègent votre vie privée »

https://support.mozilla.org/fr/kb/conteneurs?redirectslug=onglets-contextuels-avec-les-containers

Fenêtre de navigation privée

« vos informations de navigation telles que les cookies et l'historique ne sont pas enregistrées et ne laissent donc aucune trace à la fin de la session. »

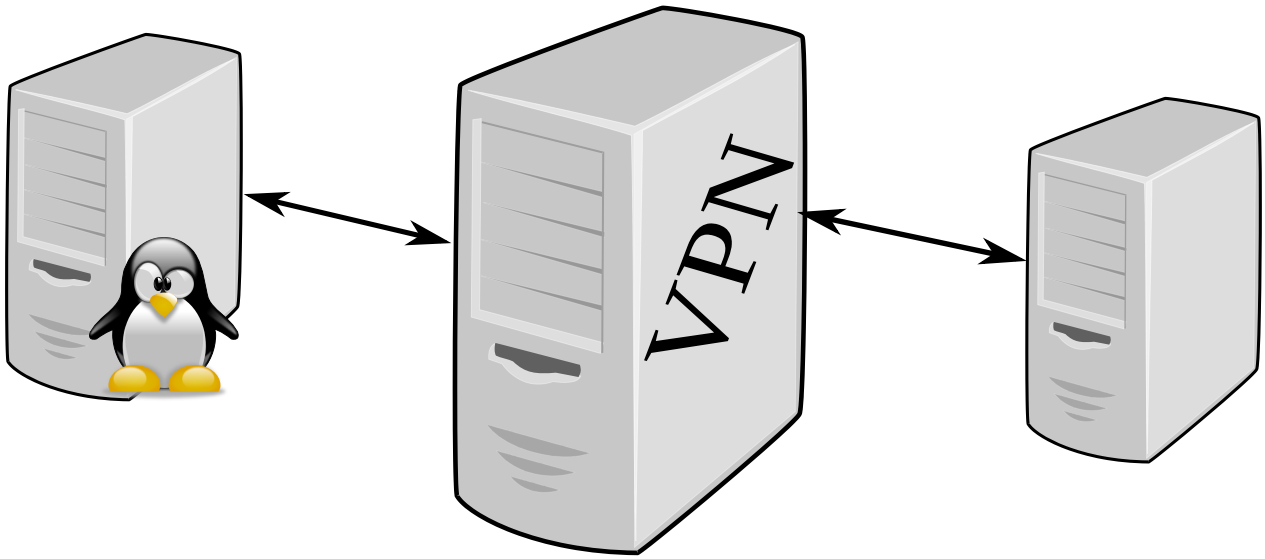
https://support.mozilla.org/fr/kb/navigation-privee-naviguer-avec-firefox-sans-enregistrer-historique

Ublock

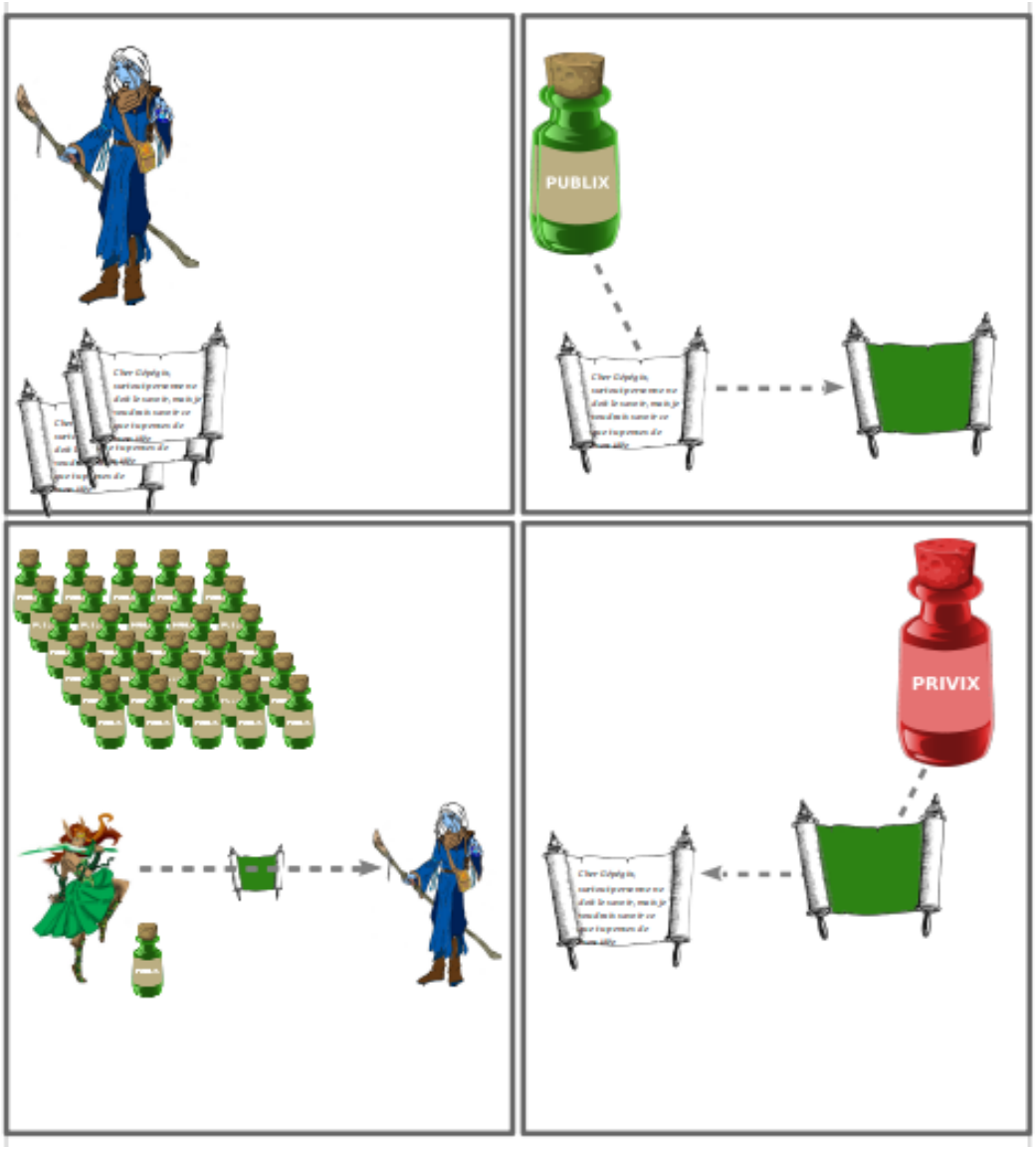
https://fr.wikipedia.org/wiki/UBlock_Origin

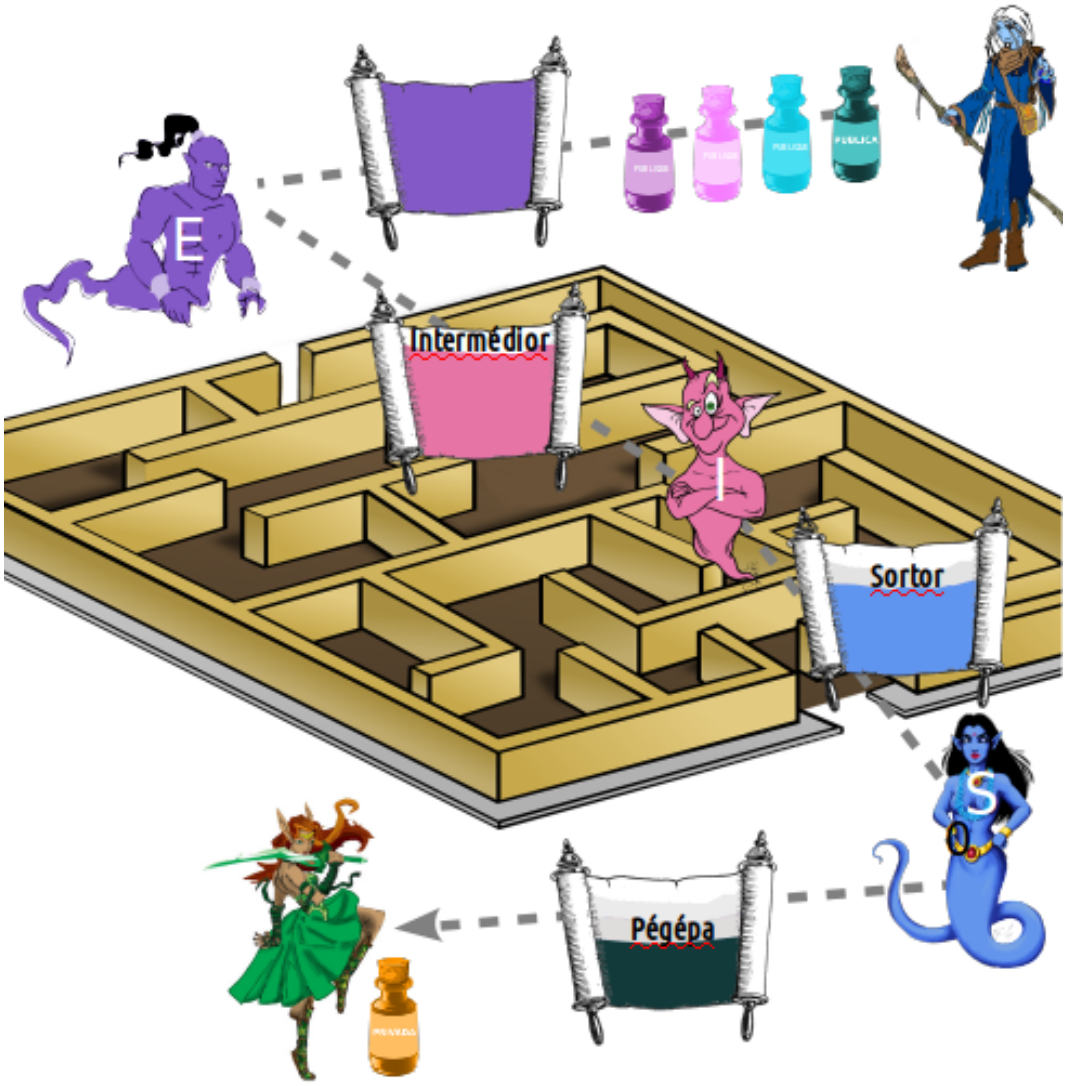
4. VPN et Tor

VPN



Tor





5. Faire des choix

- Utiliser ou pas les services web qui collectent des données et alimentent le capitalisme de surveillance
- Utiliser des logiciels libres ou non (OS, navigateur…)
- Limiter ou pas sa surface de contact numérique (médias sociaux, objets connectés…)
- …

Cours de 2022

Solutions des exercices

[exercice p. 34]
Solution n°1

Crédits des ressources

The most important types of data needed p. 7
European Union Agency for Law Enforcement Cooperation, 2021. SIRIUS EU Digital Evidence Situation Report 3rd Annual Report. https://www.europol.europa.eu[10]

The most contacted services p. 7
European Union Agency for Law Enforcement Cooperation, 2021. SIRIUS EU Digital Evidence Situation Report 3rd Annual Report. https://www.europol.europa.eu[11]

Conférence Technopolice,24 mai 2022 p. 8
La Quadrature Du Net et Picasoft, https://mobilizon.picasoft.net/[12]

Campus Police Surveillance p. 8
Jean-Marc Manach, 2021. Les technologies de surveillance à l'assaut des campus américains, NextINpact. https://www.nextinpact.com/[13]

Chiffre de César p. 16
Attribution - Partage dans les Mêmes Conditions - Patricia.fidi Wikipédia

Schéma d'une communication HTTPS p. 26
Attribution - Partage dans les Mêmes Conditions - Quentin Duchemin, logos de Assaf Katz, Stalinsunnykvj, MGalloway (WMF), Artdabana@Design. Source : https://search.creativecommons.org

10. https://www.europol.europa.eu/cms/sites/default/files/documents/SIRIUS_DESR_12_2021.pdf
11. https://www.europol.europa.eu/cms/sites/default/files/documents/SIRIUS_DESR_12_2021.pdf
12. https://mobilizon.picasoft.net/events/f4738be8-1475-4a26-beab-d20f7654dcf7
13. https://www.nextinpact.com/article/46426/les-technologies-surveillance-a-assaut-campus-americains