Anonymat et vie privée sur internet

Guillaume Pillot

QuébecSec

26 Avril 2018

Qui suis-je ?

- Étudiant finissant à la maîtrise en informatique à l'Université Laval
- Sujet de recherche sur l'anonymat sur internet
- Responsable du club de hacking de l'université de 2015 à 2016
- Passionné de sécurité informatique, souhaite faire carrière dans ce domaine
- Pratique de CTF en ligne sur Ringzer0, root-me et sur place au NorthSec et au Hackfest
- Suis la formation de l'OSCP (PWK) présentement
- Sites web :
  https://guillaume-pillot.ca
  http://cfiul.ca/
  http://hacking.fsg.ulaval.ca/

Citation d'Edward Snowden

Prétendre que votre droit à une sphère privée n'est pas important parce que vous n'avez rien à cacher n'est rien d'autre que de dire que la liberté d'expression n'est pas essentielle, car vous n'avez rien à dire

Le droit à la vie privée

- L'anonymat garantit la vie privée.
- Articles 12 et 19 de la Déclaration universelle des droits de l'homme
- Premier amendement de la Constitution des États-Unis

La vie privée est un besoin

- Augmentation exponentielle du nombre de données
- Certaines applications ont besoin de confidentialité
- Tel que l'e-finance, le commerce électronique, la télésanté ou le vote en ligne
- Une multitude d'applications collecte les informations sur les internautes

Menaces contre la vie privée

- On peut classifier en quatre catégories les principales menaces contre la vie privée et l'anonymat sur internet :
  - Sociales
  - Politiques
  - Problèmes technologiques
  - Économiques

Pour le meilleur et pour le pire

- L'anonymat entraîne irrémédiablement des activités clandestines
- Tel que la vente de drogue, la pédopornographie ou le cyberterrorisme
- Néanmoins, il permet au dissident politique et aux journalistes de s'exprimer (Wikileaks, Edward Snowden)

Terminologie

- Vie privée : Droit d'un individu de contrôler la collection et la diffusion d'informations le concernant
- L'anonymat : C'est l'état qui garantit que l'identité d'un individu qui utilise un service ou une ressource ne puisse être divulguée

Attaquant

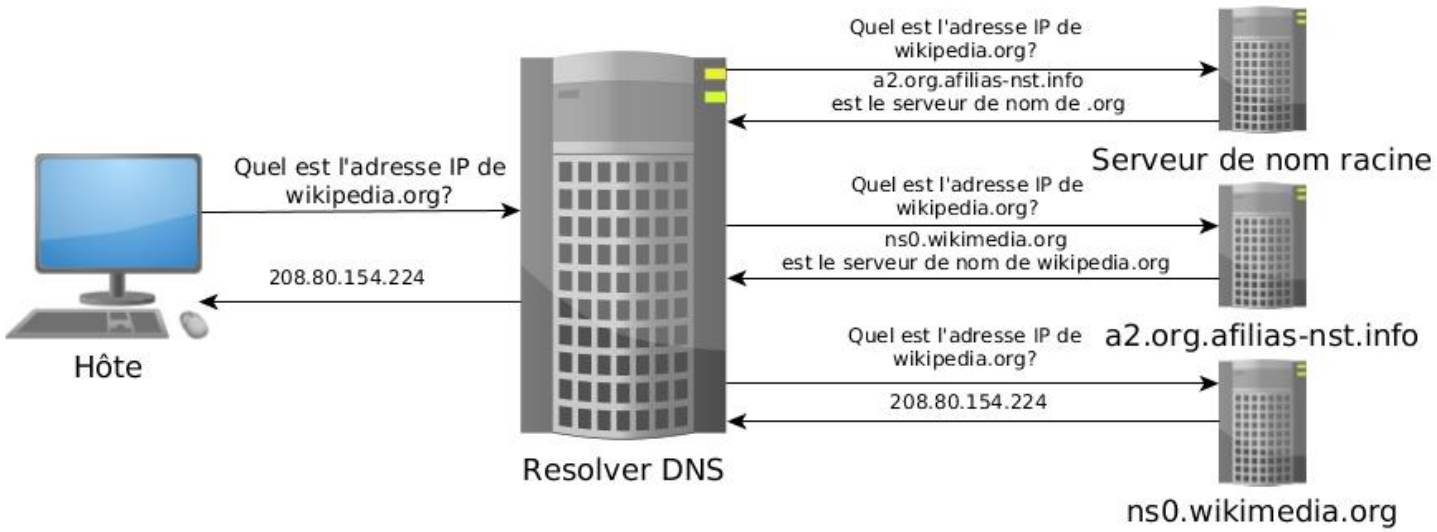
- Un attaquant souhaite surveiller ou manipuler les communications du réseau
- On peut classifier deux types d'attaquants :
  - L'attaquant passif : Il ne fait qu'observer les communications
  - L'attaquant actif : Il peut manipuler et perturber les communications en effaçant, en modifiant, en envoyant ou encore en retardant le trafic des messages

Deep Packet Inspection

- Il est possible d'analyser les communications sur chaque couche du modèle TCP/IP :
  - Physique : TAP réseau
  - Liaison : Identification d'adresse MAC
  - Réseau : Adresse IP source et destination et protocole de transport
  - Transport : Port source et destination de l'application
  - Application : Dépend de l'application. Le trafic web et courriel sont les plus ciblés

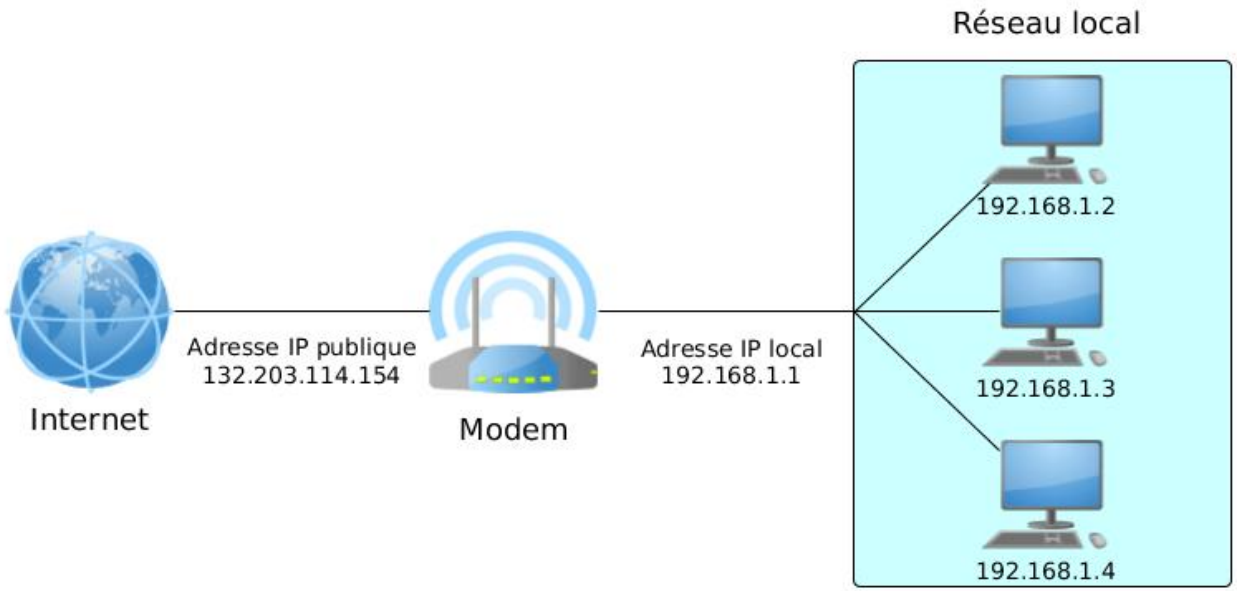
Profilage DNS

- Le DNS n'a pas été conçu pour protéger la vie privée et n'est pas chiffré
- Le Resolver
- Hiérarchie des serveurs de noms
- Tous les serveurs de noms interrogés voient le nom de domaine au complet
- L'IETF propose de chiffrer le DNS, minimiser les requêtes ou utiliser un autre système (namecoin)


Quel est l'adresse IP de wikipedia.org?
a2.org.afilias-nst.info est le serveur de nom de .org
Serveur de nom racine
Quel est l'adresse IP de wikipedia.org?
Quel est l'adresse IP de wikipedia.org?
ns0.wikimedia.org est le serveur de nom de wikipedia.org
a2.org.afilias-nst.info
208.80.154.224
Hôte
Quel est l'adresse IP de wikipedia.org?
208.80.154.224
Resolver DNS
ns0.wikimedia.org

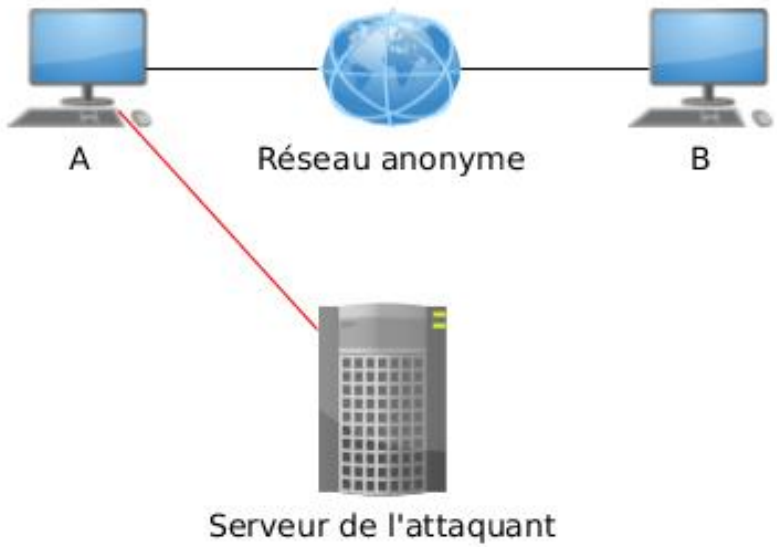
IPv6

- L'IPv6 va permettre de se passer du NAT. Il sera possible d'identifier directement une interface réseau sur internet
- IID pour Interface IDentifier identifie l'interface de la machine et il existe différents mécanismes d'allocation d'IID
- La méthode SLAAC (Stateless Address Autoconfiguration) génère l'IID à partir de l'adresse MAC
- La RFC 7721 décrit les faiblesses, d'un point vu protection de la vie privée, que provoquent les IID générés à partir d'une adresse MAC
- Une adresse MAC d'une interface est unique et ne peut être changée (sauf au niveau logiciel)
- Il est donc possible de tracer une machine à chaque changement de réseau


Réseau local
192.168.1.2
192.168.1.3
192.168.1.4
Internet
Adresse IP publique
132.203.114.154
Modem
Adresse IP local
192.168.1.1

Connexion directe à l'utilisateur

- Un attaquant peut récupérer des informations sur l'utilisateur au niveau de l'application
- Le but principal est l'obtention de l'adresse IP de l'utilisateur sans avoir besoin d'analyser le réseau
- L'attaquant peut contourner le moyen d'anonymisation (réseau anonyme) en ayant un accès direct à l'utilisateur par une attaque sur l'application


A
Réseau anonyme
B
Serveur de l'attaquant

Le navigateur web

- Le vecteur d'attaque le plus courant est le navigateur web. Il est le moyen principal pour accéder à Internet. L'outil étant complexe, plusieurs fuites d'informations sont possibles
- Les Plug-ins :
  - Flash : L'API de Flash peut établir une connexion TCP directe à l'utilisateur
  - QuickTime : un paramètre dans la configuration permet d'établir une connexion directe pour voir une vidéo
- DNS : Lors d'une recherche, le navigateur peut se connecter directement au serveur DNS sans passer par le réseau anonyme
- Applet Java : Un attaquant peut exploiter un applet Java pour dévoiler l'identité d'un utilisateur en exploitant l'API Java
- Vulnérabilités logicielles : Navigateur non mis à jour ou 0-day

- Document actif (PDF, Word, Excel) : Un attaquant peut user d'ingénierie sociale et envoyer des documents actifs contenant du code malicieux à l'utilisateur
- Cookie : Le cookie est stocké sur l'ordinateur du client et laisse donc une trace de sa visite ce qui peut compromettre son anonymat
- JavaScript : Un attaquant peut faire en sorte qu'un utilisateur exécute du code JavaScript malicieux à l'aide d'ingénierie sociale ou sur un site web vulnérable
- D'autres applications peuvent être attaquées comme les applications BitTorrent

- Le fingerprinting permet de créer une empreinte digitale à partir des informations fuitées par le navigateur de l'utilisateur
  - En-têtes HTTP : User-agent, Referer, Content-Type, etc.
  - Etag : identifiant unique fourni par le serveur web
  - Session HTTP : Plus la session dure dans le temps, plus le risque que l'utilisateur soit identifié augmente
  - Autre : envoyer des données d'authentification à des sites tiers, taille d'écran, etc.
- Tester votre navigateur avec Panopticlick

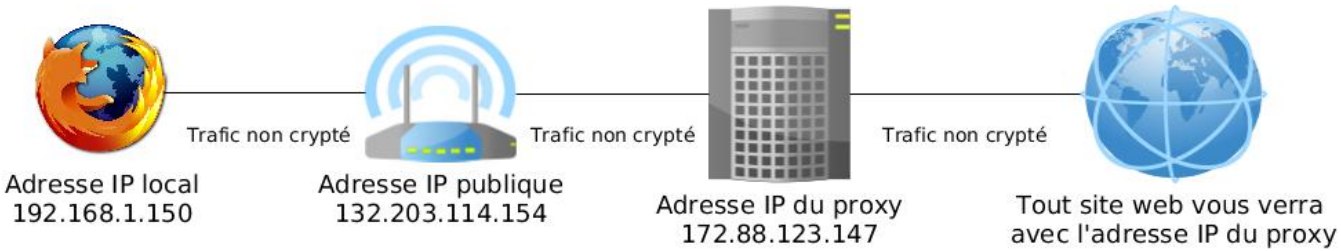
Plug-in de protection

- HTTPS Everywhere : permet d'utiliser le protocole HTTPS, sur les sites le supportant, au lieu du HTTP souvent fourni par défaut
- Privacy Badger et uBlockOrigin : Protection contre les mouchards et bloqueur de publicité
- NoScript : permet de bloquer l'exécution de code Javascript, Flash, Java, ect.

- Plus la configuration de l'utilisateur est rare, plus celui-ci est identifiable
- TOR Browser est configuré de sorte à être le moins identifiable possible
- Pour éviter toute connexion directe à la machine, l'idéal est d'utiliser un système d'exploitation qui par défaut bloque toutes connexions entrantes ou sortantes extérieures au réseau anonyme comme par exemple Tails

Proxy

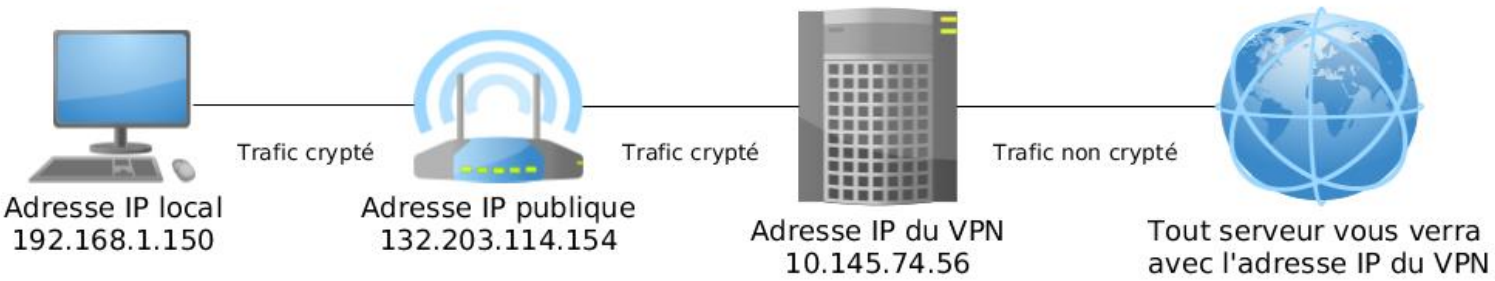
- Le proxy est un intermédiaire pour faire transiter notre trafic
- C'est l'adresse IP du proxy qui sera visible au lieu de la nôtre



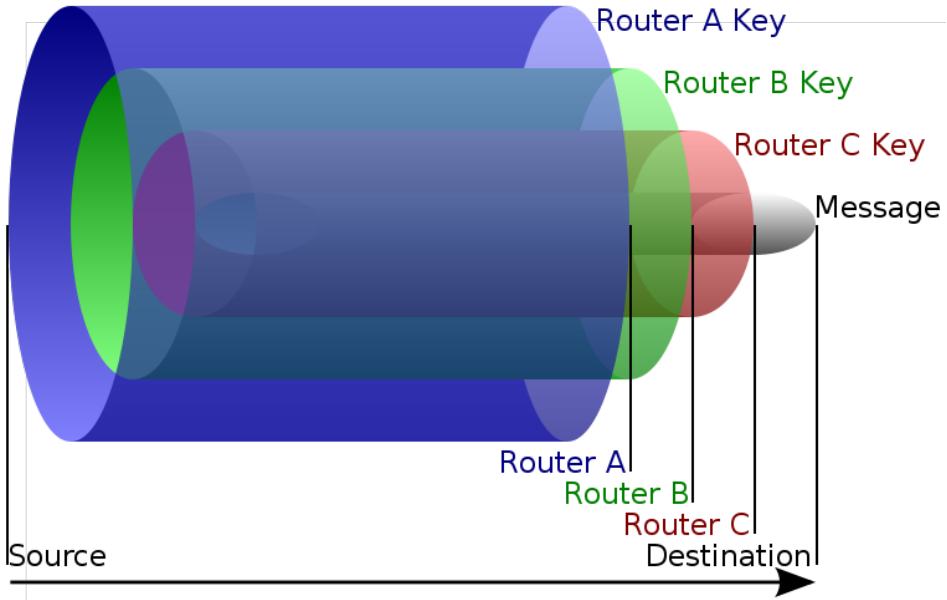
- Les deux protocoles les plus communs utilisés par les proxys sont le HTTP et le SOCKS
- Rien n'empêche au proxy de conserver des traces de notre trafic dans ses logs

VPN

- Le principe est similaire à celui d'un proxy, notre trafic internet est relayé par un serveur à la différence que la connexion entre l'ordinateur du client et le serveur VPN est cryptée et que tout le trafic entrant et sortant de l'ordinateur du client passe par le serveur

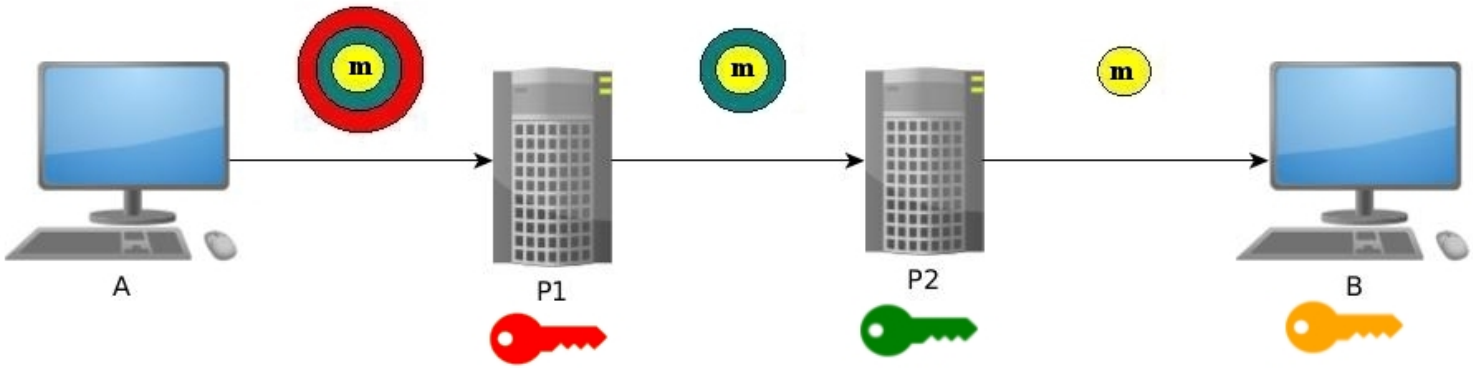


- La grande majorité des réseaux anonymes moderne sont des mix network
- Un mix network est un relais de proxy utilisant le chiffrement par couche



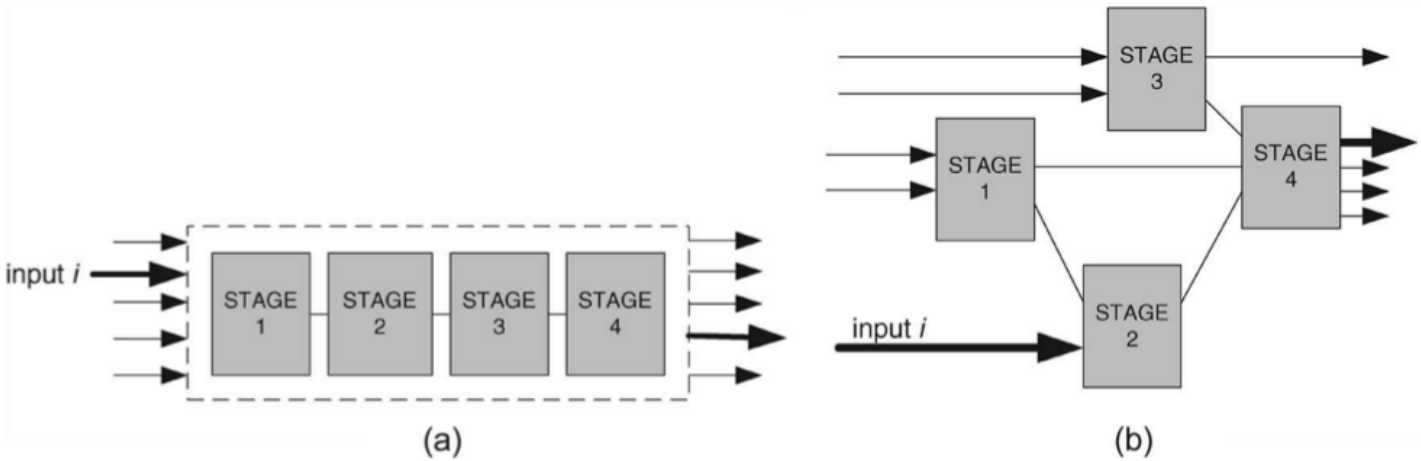
Mix Network de Chaum

- Le mix network de Chaum est le premier réseau de communication anonyme qui a été créé
- L'utilisateur chiffre son message avec chacune des clefs publiques des nœuds du relais. Ensuite, chaque mixes décrypte une couche de chiffrement et relaie le message jusqu'au destinataire


m
m
m
A
P1
P2
B

Sélection des mixes

- Cascade : Les mixes du relais sont prédéterminés.
- Free-route : La chaîne de mixe n'est pas fixée.



Attaque par intersection

- Une attaque par intersection consiste à analyser le trafic d'un utilisateur de manière passive
- L'attaquant peut dans le temps réduire l'ensemble des utilisateurs qui sont suspectés de communiquer avec l'utilisateur pour possiblement en isoler un
- Le but étant de lier l'initiateur de la communication à sa destination
- L'inconvénient majeur des attaques par intersection est qu'elles demandent une longue écoute du trafic pour réussir
- Plus le réseau est grand, plus le profilage d'un utilisateur est difficile

Attaque par étiquetage

- Cette attaque consiste à modifier ou ajouter des paquets dans le relais d'un utilisateur dans le but de reconnaître le trafic plus loin dans le relais
- L'anonymat d'un utilisateur pourrait être compromis si l'utilisateur sélectionne un nœud d'entrée et un nœud de sortie sous contrôle de l'attaquant
- L'attaquant doit avoir une présence suffisamment importante pour pouvoir contrôler à la fois le nœud d'entrée et le nœud de sortie d'un même circuit
- L'attaque étant active, l'attaquant peut se faire détecter


A
Noeud d'entrée
Noeud intermédiaire
Noeud de sortie
B
message de A
message modifié par l'attaquant
de sorte à le reconnaître en sortie
Noeud contrôlé par l'attaquant

Attaque par déni de service

- Le but d'une attaque par déni de service est d'empêcher l'accès au réseau anonyme
- Blocage total : Le gouvernent bloque l'accès au réseau anonyme
- Attaque ciblée : Attaque DDOS sur l'ensemble du réseau ou sur des serveurs essentiels à son fonctionnement

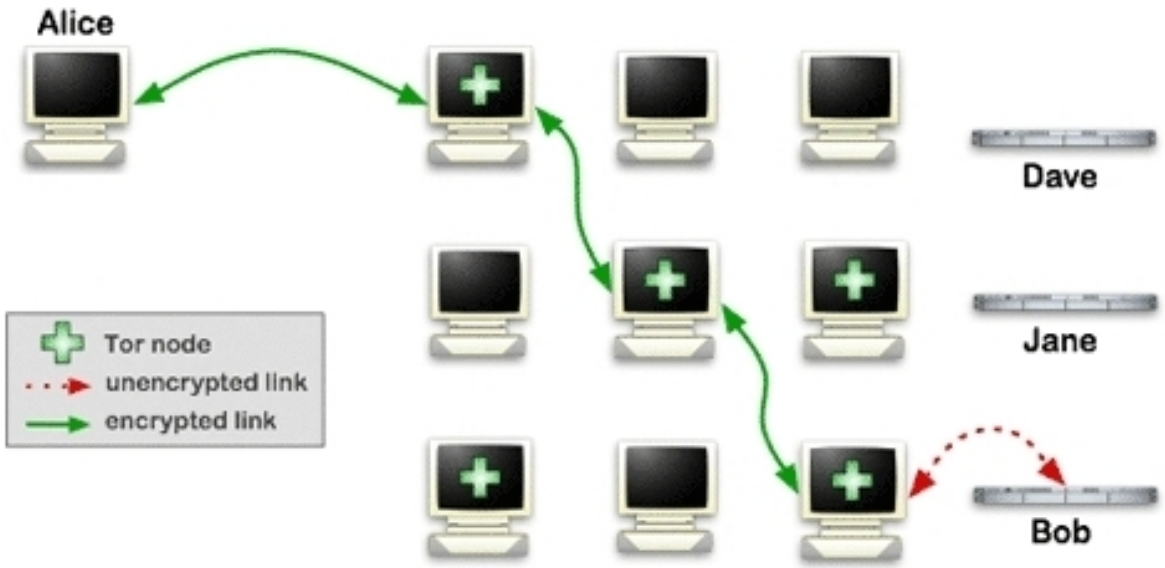
Longueur du relais

- Le but d'un attaquant est de contrôler le nœud d'entrée et le nœud de sortie du relais d'un utilisateur
- Un relais composé de plus de trois nœuds n'apporterait pas plus d'anonymat
- TOR et I2P recommandent une longueur de trois nœuds dans leur réseau
- Lien 1 Lien 2

Tor : The Onion Router

- TOR est le réseau de communication anonyme le plus populaire. D'après Tor metrics, le nombre d'utilisateurs stagne autour de deux millions depuis 2014 et le trafic atteint 100Go/s
- Le routage en oignon fournit une communication privée à faible latence à travers Internet
- Le fonctionnement est similaire au mix network, un utilisateur établit un circuit de routeur oignon, la communication est cryptée en couche avec chaque clé publique des nœuds du circuit

- Chaque nœud du circuit connaît uniquement son prédécesseur et son successeur. Le dernier nœud du circuit transmet le message au destinataire, il est le seul à le connaître, tout comme le premier nœud du circuit est le seul à connaître l'utilisateur
- Les données sont véhiculées dans des cellules de 512 octets. Les entêtes de chaque cellule contiennent un identifiant du circuit auxquels elles appartiennent ainsi qu'une commande sur ce qu'il faut faire avec le contenu


Alice
Dave
Jane
Bob
Tor node
unencrypted link
encrypted link

Le darknet

- Le web peut-être divisé en trois couches :
  - Le web surfacique est la partie accessible en ligne.
  - Le web profond (ou DeepWeb) est la partie qui n'est pas indexée par les moteurs de recherches classiques. Elle représenterait plus de 90% du contenu d'internet.
  - Le darknet est une partie du web profond dont les données sont volontairement cachées. Y accéder demande l'utilisation d'outils spéciaux assurant l'anonymat de leurs utilisateurs.


SURFACE WEB
Google
Bing
Wikipedia
DEEP WEB
Academic Information
Medical Records
Legal Documents
Scientific Reports
Subscription Information
Contains 90% of the information on the Internet, but is not accessible by Surface Web crawlers.
Social Media
Multilingual Databases
Financial Records
Government Resources
Competitor Websites
Organization-specific Repositories
(DARK WEB)
A part of the Deep Web accessible only through certain browsers such as Tor designed to ensure anonymity. Deep Web Technologies has zero involvement with the Dark Web.
Illegal Information
TOR-Encrypted sites
Political Protests
Drug Trafficking sites
Private Communications

Garde d'entrée

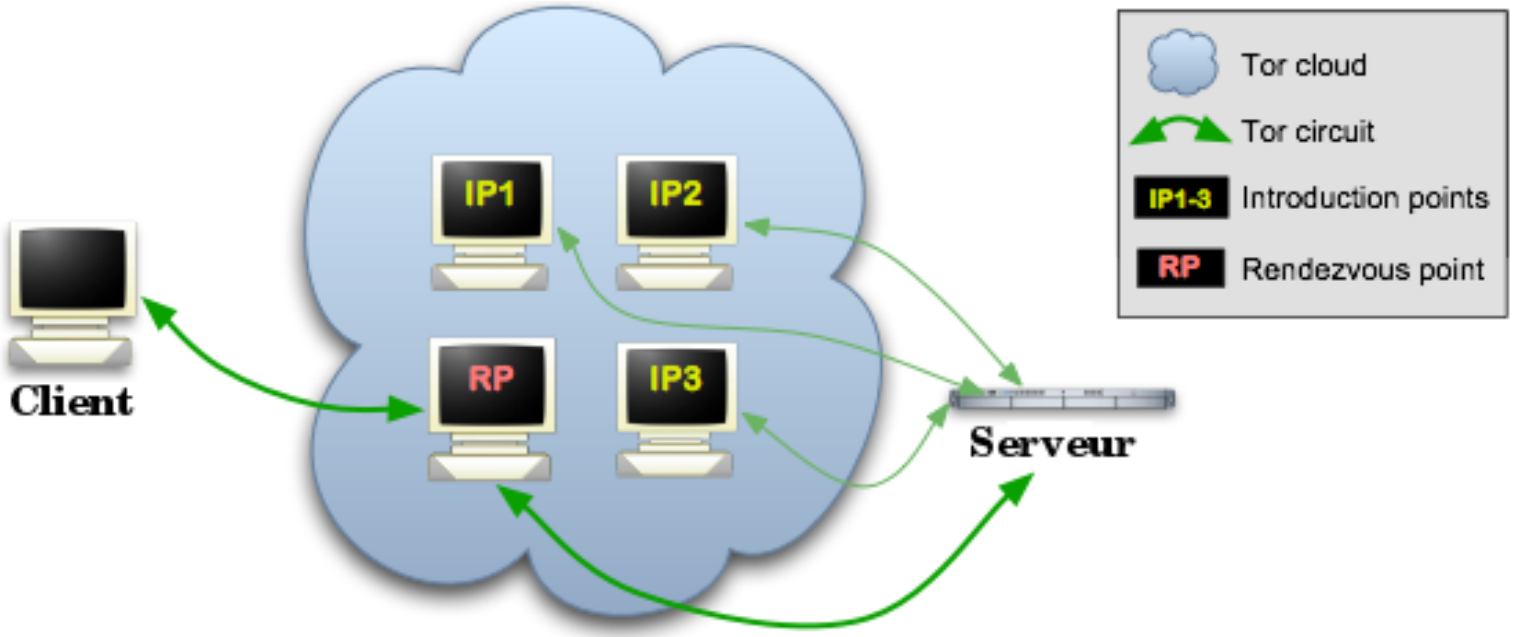
- Ce nœud est le seul à connaître l'identité de l'utilisateur, il est la cible privilégiée des attaquants
- Un nœud qui rejoint le réseau ne peut pas devenir immédiatement un garde d'entrée, il doit répondre à certains critères et passer par plusieurs phases

Nœud de sortie

- Étant donné que le nœud sort de TOR pour accéder à la page demandée, un attaquant peut facilement obtenir l'adresse IP du nœud de sortie
- Si ce nœud est utilisé pour véhiculer du trafic illégal, le propriétaire peut être tenu comme responsable au niveau de la loi
- Pour qu'un nœud puisse devenir un nœud de sortie celui-ci doit ouvrir au moins deux de ces trois ports : 80, 443 ou 6667 et doit permettre de sortir sur au moins 16777214 adresses IP sur internet

Service caché

- Les services cachés de TOR font partie du darknet. Pour y accéder, il est obligatoire de passer par le réseau TOR
- L'intérêt d'un service caché est qu'il permet d'assurer un anonymat mutuel entre le serveur et le client
- Les services cachés utilisent des noms d'hôtes .onion
- Une adresse .onion correspond à la clé publique du serveur encodée en base 32. Par exemple : `https://76qugh5bey5gum7l.onion`
- Lien


Tor cloud
Tor circuit
IP1-3
Introduction points
RP
Rendezvous point
IP1
IP2
RP
IP3
Client
Serveur

Serveur d'annuaire

- La liste des nœuds et leurs caractéristiques sont fournies par des serveurs d'annuaire
- Ceux-ci sont au nombre de 9 actuellement
- La liste de ses serveurs est disponible sur le site TOR Metrics

Cycle de vie des nouveaux nœuds

- Première phase dite "Non mesurée"
- Deuxième phase dite de "Mesure à distance"
- Troisième phase dite de "passage en garde d'entrée"
- Quatrième phase dite de "Garde d'entrée stabilisée"

Tails

- Tails pour The Amnesic Incognito Live System est un système d'exploitation basé sur Debian
- Tails construit des circuits et bloque toutes les connexions qui ne transitent pas sur le réseau TOR
- DÉMO

Mécanisme de protection

- Changement de circuit toutes les 10 minutes
- Ajout de donnée aléatoire aux paquets
- Bridge

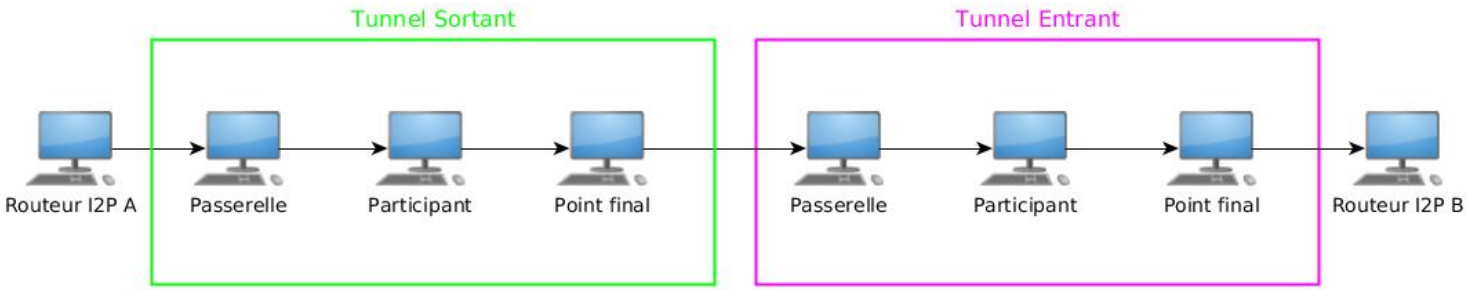
I2P : Invisible To Internet

- I2P pour Invisible Internet Project est le deuxième réseau anonyme le plus utilisé
- En 2017, le nombre total de nœuds sur le réseau est estimé autour de 50000
- I2P fonctionne comme un mix network, la grande différence avec les autres réseaux anonymes est qu'I2P est cloisonné, il n'est pas conçu pour accéder à l'internet standard

- L'utilisateur doit installer un logiciel qui agira en tant que routeur sur le réseau
- Tout routeur connecté au réseau peut participer à la transition des données. À partir du moment où l'utilisateur est connecté à I2P, il peut faire partie d'un relais d'un ou plusieurs autres utilisateurs
- I2P est "orienté message", les communications ne sont pas bidirectionnelles

Les tunnels

- Un tunnel est une chaîne de routeurs par laquelle sont véhiculés les messages. Les messages ne vont que dans un seul sens
- Deux types de tunnels existent, les tunnels sortants permettant d'envoyer des messages et les tunnels entrants permettant d'en recevoir
- Chaque routeur peut avoir plusieurs de ces tunnels


Tunnel Sortant
Tunnel Entrant
Routeur I2P A
Passerelle
Participant
Point final
Passerelle
Participant
Point final
Routeur I2P B

netDb

- netDb est une base de données distribuées sur le réseau. elle fonctionne comme une DHT
- une partie des routeurs du réseau nommée les "routeurs floodfill" maintiennent la base de données
- netDb est décentralisé
- Environ 6% des routeurs du réseau sont floodfill
- DÉMO

Mécanisme de protection

- Changement de tunnels toutes les 10 minutes
- Un attaquant externe au réseau ne pourra pas modifier les messages, car ceux-ci sont signés
- Un attaquant peut créer un grand nombre de nœuds dans le réseau et faire en sorte que ceux-ci ne fournissent aucune ressource. La base de données deviendra plus grande inutilement et les nœuds du réseau devront demander plus de tunnels
- I2P collecte sur chaque nœud des données de profilage tel que le temps de réponse ou le taux d'échec de ses tunnels. Via ce système, I2P tente d'identifier les nœuds défaillants. Ces nœuds seront par la suite soit ignorés, soit sous-utilisés

JAP

- JAP (Java Anon Proxy), aussi nommé JonDonym, est un webmixe. Il est conçu pour la navigation en temps réel sur internet
- Sa topologie est en cascade. La connexion est bidirectionnelle
- DÉMO

Mécanisme de protection

- JAP utilise un algorithme "chop-and-slice". Il consiste à découper en morceaux de même taille les grands messages. Ces morceaux sont nommés "slice". Chaque slice est envoyé via un canal du mixe. Tout utilisateur actif n'émettant pas de message en envoie des fictifs. Ainsi, un trafic est constamment maintenu dans le réseau et avec le chiffrement, il n'est pas possible pour un attaquant de différencier les messages réels des faux
- Pour réaliser une attaque par étiquetage, il faut que l'attaquant contrôle les mixes du relais. Chaque mixe est géré par un opérateur de confiance.
- L'internationalisation des mixes limite les risques qu'un attaquant puisse contrôler tout le relais
- Le gros point faible de JAP est les attaques par déni de service, il suffit donc qu'un seul des mixes ne fonctionne pas pour que tout le relais ne soit plus utilisable

Stratégie d'éviction

- Les réseau anonyme à faible latence ne sont pas protégé à 100% contre les attaques par intersection
- Pour empêcher ces attaques, les mixes peuvent employer différentes stratégies d'éviction (flushing en anglais) qui définit à quel moment les messages reçus sont envoyés. Il existe trois types d'algorithmes :
  - Atteindre un seuil de messages reçus avant de les transférer.
  - Randomise le délai de transfert de chaque message.
  - Le "pool mixe" : Les messages reçus sont transférés par lot.
- Un réseau anonyme à faible latence doit être un minimum performant et il ne peut pas mettre en place de stratégie d'éviction

Remailer

- Un remailer permet l'envoi de courriel anonyme. C'est un réseau anonyme à haute latence. Les remailers peuvent être classés en trois types.
- Type I Cypherpunk : Utilise le logiciel PGP. Chiffrement et choix des remailer manuel.
- Type II Mixmaster : Protège contre l'analyse du trafic. C'est un mix network hybride. Sa topologie est le free-route. L'algorithme d'éviction utilisée est le "pool mixe"
- Type III Mixminion : Permet au destinataire de répondre. Le projet est au point mort depuis 2007

Conclusion

- JAP contrairement à TOR ou I2P, utilise des routes fixes et préétablies dans ses relais. Ses mixes sont authentifiés et publics
- I2P est un réseau cloisonné, tandis que TOR permet d'accéder à tout internet
- I2P est complètement décentralisé ce qui n'est pas le cas de TOR et ses serveurs d'annuaire
- Il n'existe pas de système d'exploitation dédié pour I2P comme Tails pour TOR
- Tout nœud d'I2P peut transférer du trafic d'un autre nœud alors que sur TOR, seuls les volontaires peuvent devenir des relais et ceux-ci sont publics
- Sur papier, les remailers fourniraient une meilleure protection avec les stratégies d'éviction, mais les différents projets sont au point mort depuis des années