Le livre blanc
QUI ÉCLAIRE LE DARK

Tome 1

ALEPH

Avant-propos

L'équipe Aleph est heureuse de partager ce 1er tome du livre blanc dédié au **dark web**

Nous avons développé un moteur de recherche, d'analyse et de visualisation qui permet d'accéder aux données qui circulent sur les réseaux Tor et I2P.

Nos équipes comprennent aussi bien des formateurs pour les utilisateurs de notre solution que des analystes spécialisés sur les thématiques en lien avec les données issues du dark web.

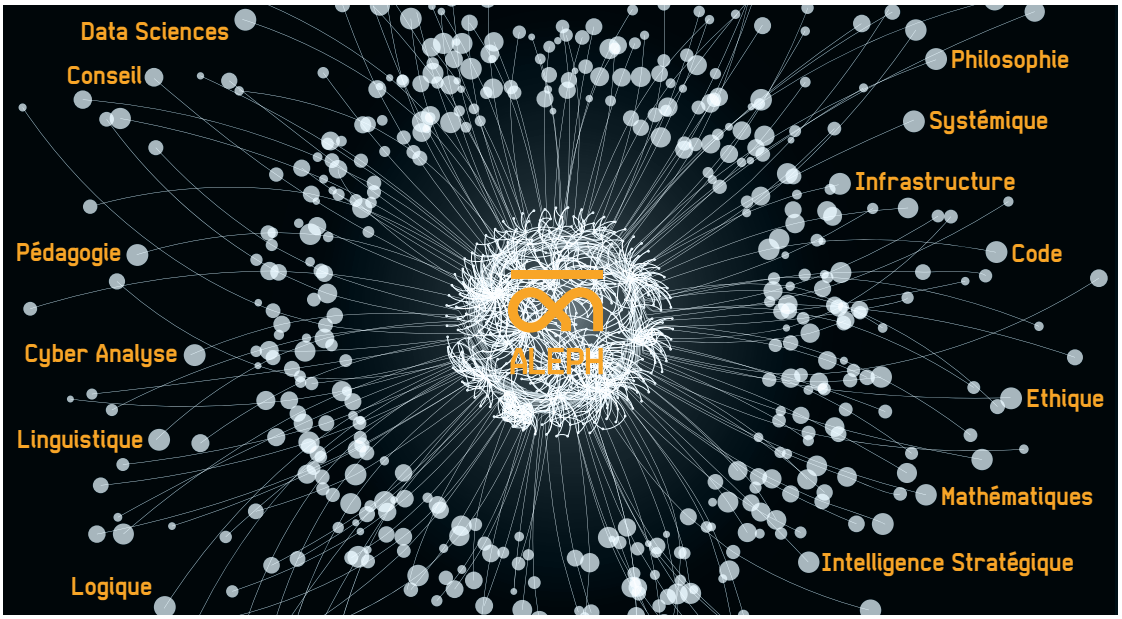
Ce livre blanc reprend les articles thématiques diffusés sur *notre blog* tout en vous interrogeant sur la nécessité d'observer ce qui peut se trouver sur le dark web.

Connaissez-vous vraiment le dark web ?

Dans quel cas doit-on s'intéresser au dark web dans le cadre professionnel ?

Existe-t-il des liens et des parallèles avec le clear web ?

> "Le contenu du dark web est hétéroclite et parfois mystérieux, ce livre blanc est fait pour vous éclairer..."



Sommaire

ALEPH

Introduction

Internet est devenu ces vingt dernières années un outil d'échange et de partage indispensable à notre société moderne.

Le père du World Wide Web, Tim Berners-Lee était convaincu que l'essor de ce support ne se ferait que sur la base d'une liberté totale d'accès et de modification de l'information. Mais ce dernier ne dispose actuellement d'aucune autorité centralisant une politique de régulation commune, à l'exception de l'ICANN (Internet Corporation for Assigned Names and Number), qui gère donc le protocole IP et les nœuds DNS.

Nous pouvons imaginer que demain, suite à d'éventuelles (inévitables ?) guerres cybernétiques, une telle entité pourrait voir le jour et imposerait une régulation des trafics numériques liés à Internet au niveau mondial.

La question de l'anonymat n'est pas récente, mais prend une dimension nouvelle, en particulier d'un point de vue technique et juridique. La régulation implique le contrôle, mais tout un chacun peut légitimement ressentir un besoin d'anonymat lors de son utilisation d'Internet.

Parmi les technologies permettant le pseudo-anonymat, on distingue les familles suivantes :

Pseudonymes et Avatar : utiliser une nouvelle adresse email, ou un nouvel avatar.

No-log : il s'agit d'utiliser une ou plusieurs plateformes qui ne conservent pas de trace.

Protocoles et outils réseaux : c'est ici que se situe le nerf de la guerre, Internet étant lui-même une interconnexion de réseaux.

Qu'est-ce que le dark web ?

Parmi les protocoles et les outils réseaux, on distingue le dark web. Il peut ainsi être défini : *"Web utilisant des technologies de type peer-to-peer (p2p) permettant le pseudo-anonymat de ses utilisateurs".*

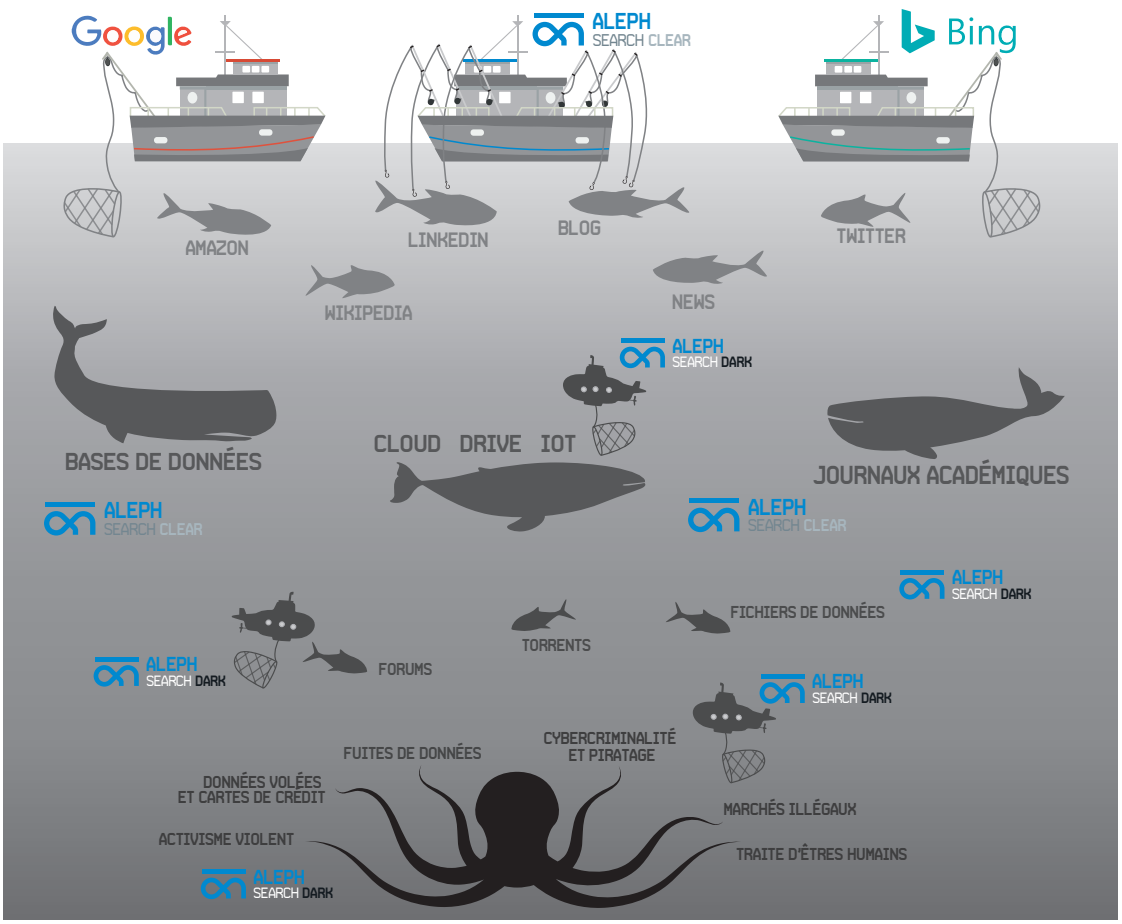
Les technologies de ce type sont diverses et ont chacune une histoire et une surface de déploiement qui peut varier selon les époques. En voici quelques-unes :

Emule, Torrent, Freenet, Tor, I2P, OFFSystem, Gnunet, DC-Net...

Chacune de ces technologies a ses propres caractéristiques, mais une se distingue des autres par notamment sa très forte utilisation. Le réseau TOR est celui dont les noms des sites se terminent en .onion *(monsite.onion)*. Nous étudierons dans ce livre, le réseau Tor constitutif du dark web ainsi que **la nécessité pour le monde de l'entreprise de surveiller ce périmètre au même titre que le web ou les réseaux sociaux.**

Nous sommes aujourd'hui en capacité d'écrire ce livre blanc, car depuis 2014 Aleph développe une technologie de moteur de recherche qui, appliquée au dark web, nous permet d'avoir une vision quasi-exhaustive et objective du dark web.

Si Google permet de connaître et d'accéder au web, **Aleph Search Dark** permet alors de connaître et d'accéder à l'information du dark web en toute autonomie.



Visualisation des clear, deep et dark webs

Le dark web en chiffres

Les chiffres du dark web 2022

Nous avons dressé notre premier état quantitatif du réseau Tor en 2021 (cf https://www.aleph-networks.com/le-dark-web-en-chiffres/). Début 2022, nous reprenions notre étude afin de mettre en évidence les évolutions de ce réseau. Le périmètre des sites onion a en effet connu des changements radicaux, principalement en raison de l'arrêt du support technique des adresses V2.

La transition des adresses V2 à V3

Depuis le mois de janvier 2018, une nouvelle version de Tor permet de créer des domaines[1] dont le format de nom renforce la sécurité.

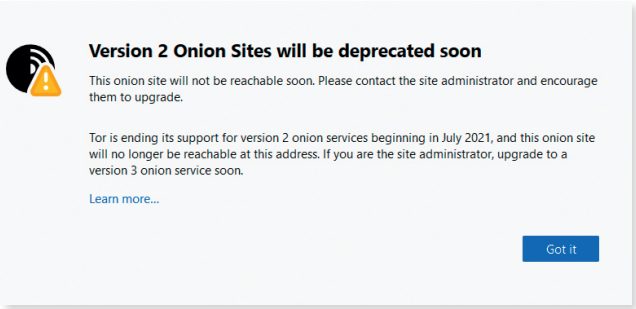
Ce nouveau format de nom de domaine, communément appelé adresses **V3**, est aisément repérable puisque, à la différence des anciennes adresses **V2** en 16 caractères, les adresses **V3** contiennent 56 caractères, toujours avec l'extension onion.

Au mois de juillet **2020,** la décision de mettre un terme au support technique des adresses **V2** a été annoncée. La mise en application effective de l'arrêt du support était alors prévue pour le 15 octobre 2021.

Dans l'intervalle, des rappels de cette échéance ont été mis en place. Ainsi, une personne qui se rendait sur un site en **V2** via Tor Browser était systématiquement informée de l'arrêt prochain du support par un message d'avertissement.

En pratique, les domaines **V2** ont été consultables jusqu'au déploiement de la version 11.0 de Tor Browser au début du mois de novembre 2021. A cette date, nous recensions un total de 110 000 domaines Tor actifs, dont 60 000 adresses **V2**. C'est donc plus de la moitié des sites onion qui est devenue inaccessible pour les utilisateurs courants de ce réseau.

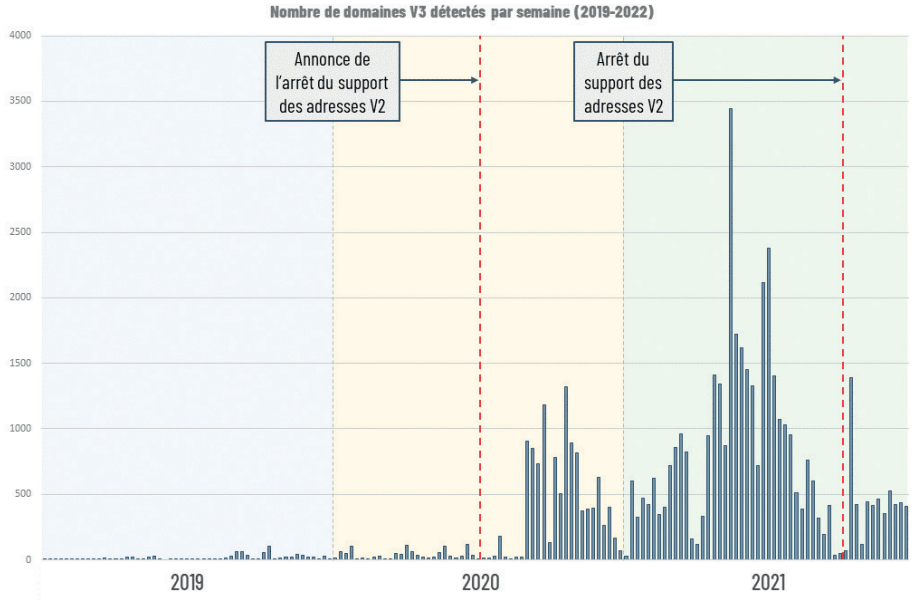
Message d'avertissement de l'arrêt prochain du support des adresses V2



Annonce de la fin du support des adresses V2

1/ **Domaine**, nom de domaine, adresse : correspond au nom ou à l'adresse du site. *(ex. : monsite.com, monsite.fr, ...)*, dans le cas du Dark web, les noms de domaine des sites se terminent par monsite.onion (tor) ou monsite.b32.i2p (i2p).

Comme nous pouvons le voir sur le graphique ci-contre, la création de sites V3 a été très marginale jusqu'à l'annonce de l'arrêt du support des adresses V2.



Détection des sites V3 de 2019 à 2022

Depuis l'annonce, les nouveaux domaines ont été majoritairement créés avec des adresses V3, même si nous avons pu constater avec un certain étonnement que certains persistaient à créer des domaines sous l'ancien format jusqu'au mois de novembre 2021.

Plus étonnant encore, parmi ces retardataires se trouvait le site d'une équipe d'opérateurs de ransomware, sans doute plus en pointe sur les méthodes d'extorsion que sur l'actualité du réseau Tor.

La transition V2 à V3 a donc rebattu les cartes en profondeur au niveau de l'existant, puisque de nombreux sites n'ont pas (encore) opéré leur migration vers la nouvelle version et peuvent donc être considérés comme perdus. Cette transition n'a toutefois pas modifié certaines pratiques et certaines tendances.

Les chiffres du réseau Tor en janvier 2022

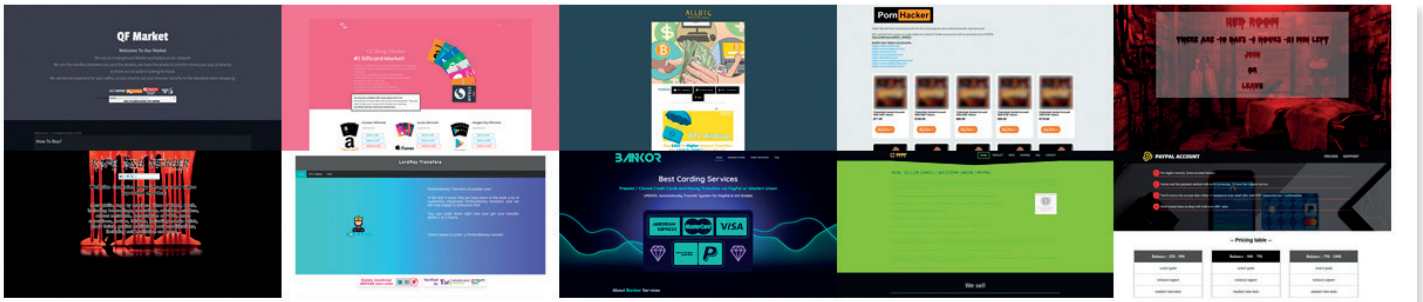
La masse totale de domaines Tor actifs en janvier **2022** est de **53 000**, contre **76 000 il y a un an**. Cette masse totale a atteint plus de **115 000** sites actifs, V2 et V3 confondus, quand les deux formats d'adresse étaient encore valides. On peut y voir là la redondance des sites V3 nouvellement créés et du maintien en ligne des anciennes adresses V2.

La transition V2 à V3 n'a pas mis un terme à la pratique du mirroring de masse, loin de là. Si les miroirs V2 sont devenus inaccessibles, la création de miroirs continue de plus belle avec la V3. Nous revenons sur la notion de miroir dans un chapitre dédié ***#3 Les miroirs dans le dark web.***

Nous pouvons par exemple constater sur le graphique précédent un pic de détection de nouveaux sites durant la deuxième semaine du mois de mai 2021. Les 3 400 domaines détectés durant cette semaine sont pour la plupart des miroirs, puisqu'il s'agit en réalité de 150 nouveaux domaines dont une vingtaine sont répliqués à plus d'une centaine d'exemplaires chacun.

+ 115 000 sites actifs

L'an dernier, nous constations que les sites les plus dupliqués étaient majoritairement des sites dédiés aux opérations financières (carding, manipulation de bitcoin…). Cette année, nous observons une tendance identique mais trois sites aux titres plus évocateurs s'invitent dans le classement.



Captures d'écran des pages d'accueil des dix sites les plus répliqués

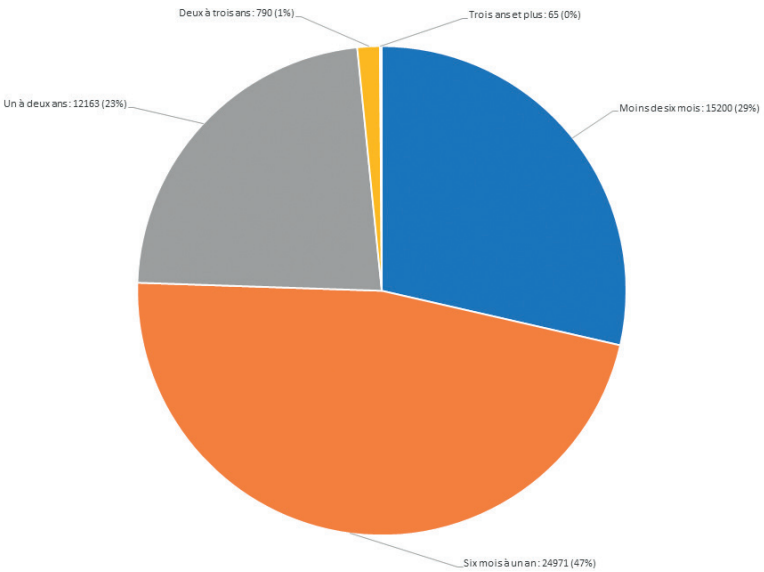
Si nous faisons abstraction des miroirs, la masse totale des sites Tor actifs en **janvier 2022** n'est plus que de **6 300 sites**. Nous voyons donc une diminution très nette du nombre de sites uniques, puisque nous dénombrions plus de **18 000 sites en janvier 2021**. Il est cependant tout à fait probable que nous retrouvions assez rapidement des volumes comparables à l'an dernier.

Les dix sites les plus répliqués

Nom du site	Miroirs
QF Market - Fast Transfers	1560
GC King _-_ GiftCard Shop	1538
All BTC ,:, Everything you needed	1523
Porn Hacker	1193
Red Room	1169
Rape and murder!	1126
LordPay - Easy Transfers	1063
Bankor - Cloned Credit Cards	907
SHOP CARD / CLONED CARDS / WESTERN UNION / PAYPAL	888
Paypal Account	857

Durée de vie des sites

La transition V2 à V3 n'a pas eu d'impact considérable sur la longévité moyenne des sites. En effet, les sites les plus anciens (V2) ne font plus partie des domaines actifs, mais comme le volume de domaines anciens était déjà très restreint l'an passé, la variation a été négligeable et nous retrouvons une longévité moyenne égale à celle de janvier 2021, c'est-à-dire neuf mois.

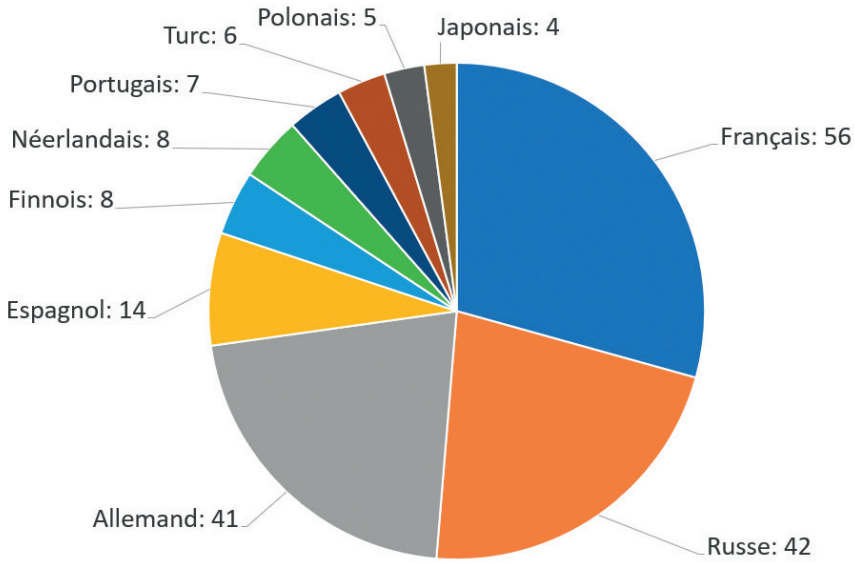


Répartition des domaines Tor actifs par longévité

Le réseau Tor et les communautés linguistiques

A la différence de la longévité moyenne, la répartition des communautés linguistiques a été modifiée de manière significative. Ainsi, si l'anglais reste la langue hégémonique sur le réseau Tor **(82% des domaines sont anglophones)**, les autres communautés ne se présentent plus dans le même ordre. Les sites francophones sont cette année les plus représentés, suivis des sites russophones, des germanophones et des hispanophones.

Les sites non anglophones ne représentent qu'une masse infime du volume total des domaines Tor actifs. Il reste par ailleurs un nombre important de domaines dont la langue ne peut être déterminée, en raison de leur contenu (données brutes issues de fuites, sites de dépôt de code...).



Répartition des communautés linguistiques sur le réseau Tor, hors sites anglophones

A RETENIR :

1 - Le dark web n'est pas aussi vaste que le clear web mais est très mouvant.

2 - La durée de vie des sites est d'environ 9 mois.

3 - L'évolution de version du réseau Tor a rendu inaccessible la moitié des domaines.

4 - Aleph garde l'historique de la totalité des sites.

5 - 82% des sites du dark web sont anglophones.

Mieux connaître son environnement permet de mieux l'appréhender. Il est aujourd'hui nécessaire, pour naviguer dans une grosse masse de données, d'avoir un outil (moteur de recherche). Cet outil va nous permettre de connaître un environnement relativement à nos besoins ou nos risques. Si je veux connaître et surveiller les risques de mon entreprise (data leak, menace vip, image, ...) présents dans le dark web j'ai donc besoin d'un moteur de recherche.

Démystifier le dark web permet d'avoir une vision plus claire : avoir une vision claire et objective des risques du dark web me permet de couvrir convenablement le risque (coût, périmètre) et de ne pas surveiller sur un mythe.

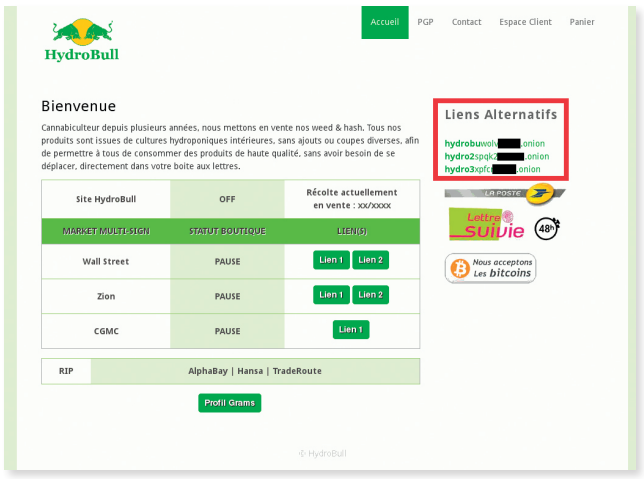
Les miroirs dans le dark web

Comme nous l'indiquions dans le *chapitre #2*, le dark web en chiffres, nous pouvons constater que **70% environ de la masse totale des domaines actifs que nous recensons sur le dark web sont en fait des copies** (d'une soixantaine de sites). Nous allons ici revenir sur cette notion de sites-miroirs et tenter de trouver une explication à ce phénomène de mirroring de masse.

Un miroir est une copie d'un site à une adresse différente. Cela peut notamment consister pour un site du clear web à créer une copie dans le dark web. Le site du quotidien berlinois ***Die Tageszeitung*** (taz.de) dispose par exemple d'un miroir sur Tor (ibpj4qv7mufde33w.onion). De nombreux sites du dark web disposent également d'un ou plusieurs miroirs sur ce même réseau.

Capture d'écran du miroir Tor du quotidien Die Tageszeitung

70%
sont en fait des copies

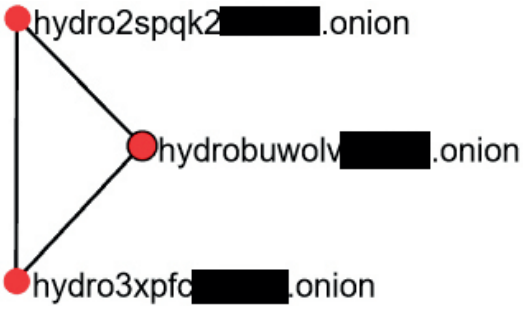


Capture d'écran du site HydroBull

Pour un site web, disposer d'un ou plusieurs miroirs présente plusieurs intérêts. D'une part, cela procure une copie de sauvegarde qui peut s'avérer vitale en cas d'avarie technique du site principal ; c'est aussi une mesure utile pour se prémunir contre les attaques de type DDoS. **De ce point de vue, les miroirs servent à garantir une continuité de service en cas d'indisponibilité du site original. D'autre part, un miroir peut servir à répartir la charge si de trop nombreuses visites ralentissent le site principal.**

La pratique usuelle veut que chaque site miroir contienne un lien vers les autres miroirs. Le site de vente de cannabis HydroBull (aujourd'hui disparu), doté de trois miroirs, affichait ainsi sur chacune de ses trois pages d'accueil les trois adresses onion de ses miroirs *(entourées en rouge sur la capture d'écran suivante)*.

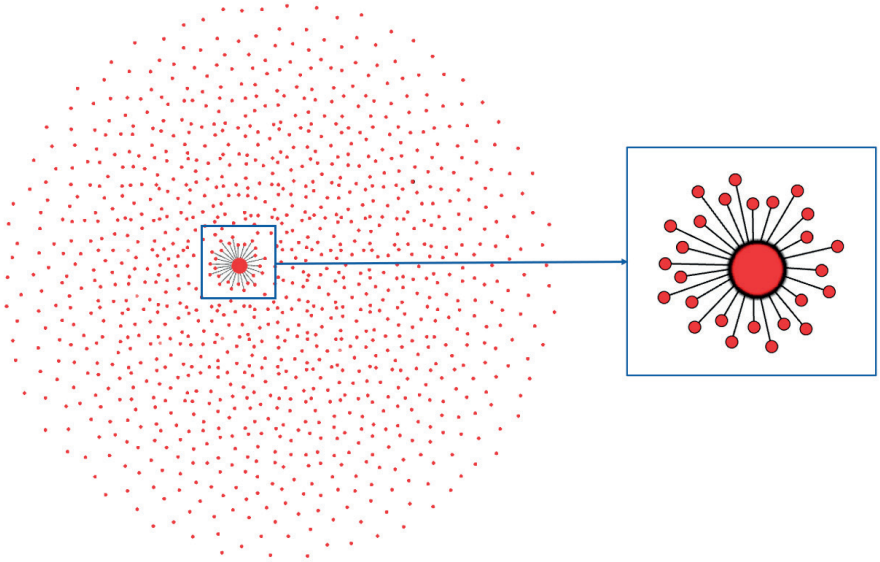
Dans le moteur de recherche **Aleph Search Dark**, nous pouvons afficher les domaines sous forme de graphe. Chaque domaine est représenté par un point ; si un site A contient un lien vers un site B, les points qui représentent ces sites sont alors reliés. Lorsque nous affichons les trois miroirs d'HydroBull sur un graphe, nous constatons que les trois sites sont bien reliés les uns aux autres.



Cela n'est néanmoins pas le cas de tous les miroirs présents sur le dark web : il est en effet fréquent que des miroirs ne soient pas liés (via un hyperlien) au site d'origine. La question se pose alors de savoir **quelle est la finalité de ces miroirs " isolés "**. Certes, ces miroirs présentent toujours l'intérêt de la copie de sauvegarde, mais s'ils ne sont pas référencés par le site original, comment les utilisateurs peuvent-ils savoir sur quelle adresse de secours se rabattre en cas d'indisponibilité du site principal ?

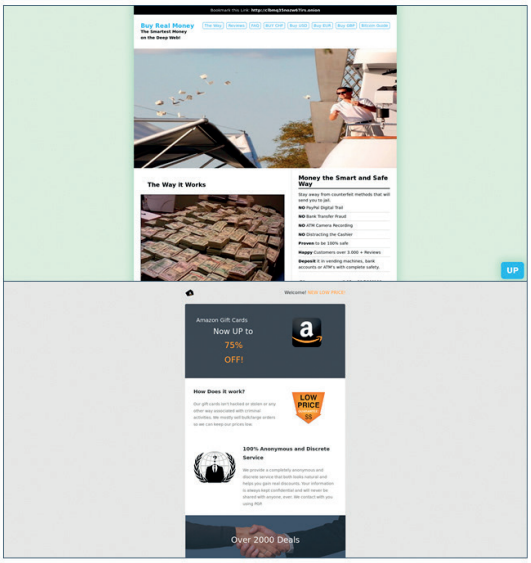
Une soixantaine de sites du dark web disposent chacun de plus de 100 miroirs, certains étant même dupliqués à quelques milliers d'exemplaires. Lorsque nous visualisons ces sites et leurs miroirs sur un graphe, nous nous apercevons que l'immense majorité, voire la totalité des miroirs ne sont pas interconnectés. L'exemple suivant est celui du site Xonions, répliqué à 1 070 reprises.

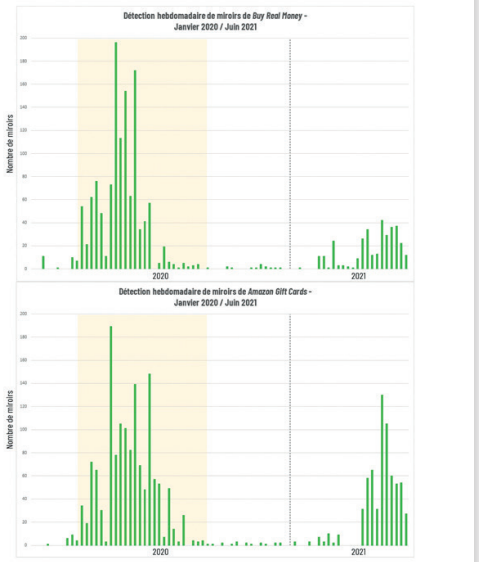
Nous observons que seuls 27 miroirs sont connectés - et non interconnectés, puisque ce sont 26 miroirs qui référencent un site central. Cet ensemble de 27 sites a la particularité de ne comporter aucun site créé pendant la vague massive de création de miroirs de 2020 (de mars à juillet, avec une décrue jusqu'à début septembre). Seuls quatre miroirs de ce réseau ont été créés après septembre 2020, tous les autres ayant été créés entre l'année 2015 et le mois d'octobre 2019.



La majeure partie des autres miroirs a été créée pendant la vague (plus de 800) et après. Le graphique suivant montre la détection des miroirs de Xonions sur la période de janvier 2020 à juin 2021 (la période mars à septembre 2020 est surlignée).

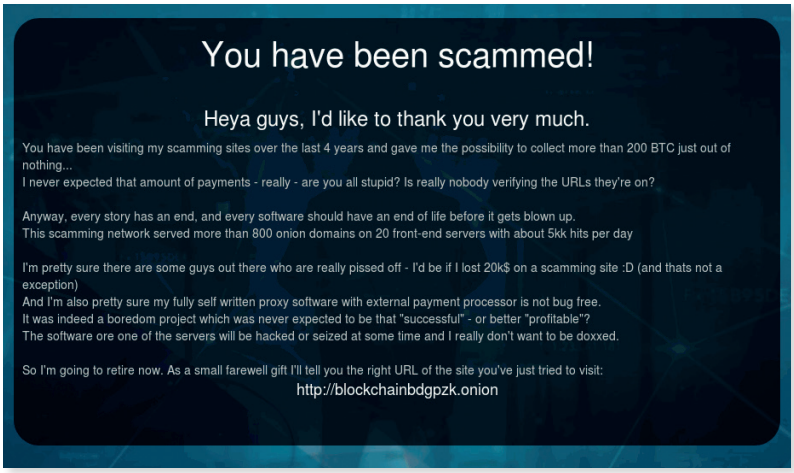
Le même schéma se répète pour de nombreux sites "multi-clonés" : les miroirs sont presque tous isolés et les dates de création coïncident, comme nous pouvons le voir dans le cas de Buy Real Money (1 613 miroirs) et Amazon Gift Cards (2 126 miroirs).





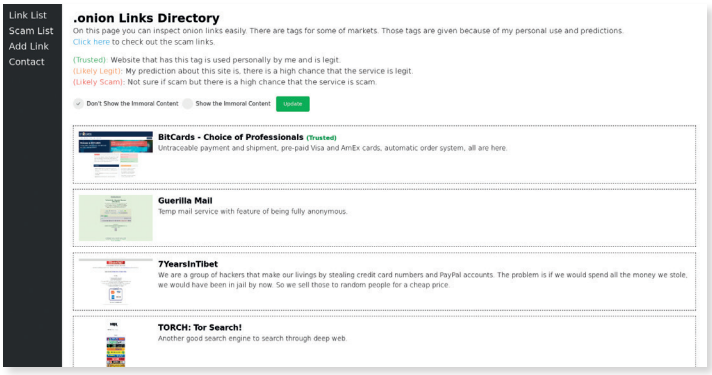
Compte tenu de la très forte similitude des schémas relationnels et de l'étalement dans le temps de la création des miroirs, il est tout à fait probable que ce mirroring de masse en 2020 soit le fait d'une même personne ou d'une même communauté.

Nous pouvons formuler **une première hypothèse** pour expliquer à la fois cette prolifération de miroirs et le fait que ces miroirs ne soient pas interconnectés. Il est en effet possible que le créateur de ces sites ait eu la volonté de saturer le dark web et d'acquérir ainsi une plus grande visibilité, mais sans faire l'effort de recenser sur chacun des miroirs l'ensemble des adresses des autres miroirs. Cette explication nous semble cependant peu plausible.

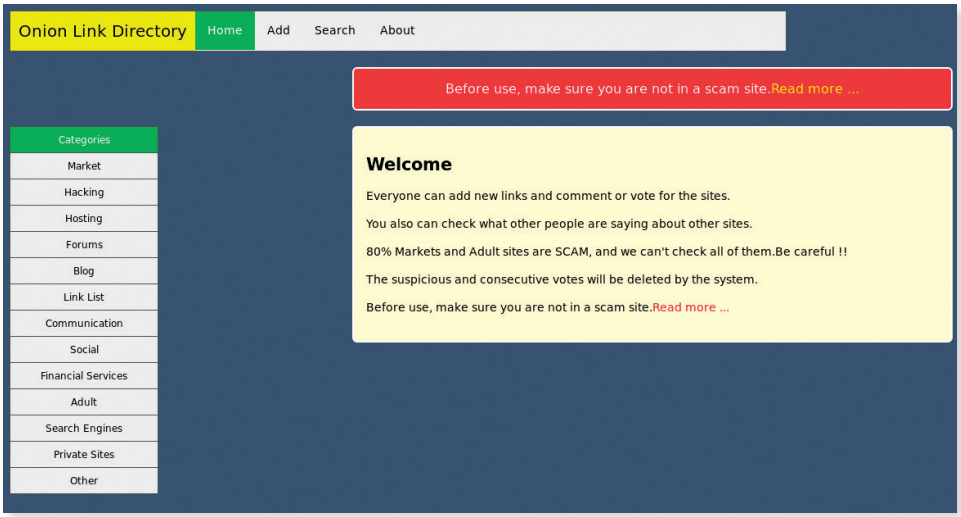


L'hypothèse la plus probable est celle d'une usurpation : les miroirs auraient été créés par une personne extérieure aux sites d'origine. En créant des copies à son compte, l'usurpateur peut ainsi récupérer des identifiants et mots de passe d'utilisateurs qui pensent se connecter au site authentique. Dans le cas d'une usurpation de site marchand, l'usurpateur peut même recevoir le paiement de transactions effectuées sur son site-leurre. En 2019, une personne qui avait créé plus de 800 leurres s'est ainsi vantée d'avoir gagné plus de 200 bitcoins (environ 1,5 million d'euros à l'époque) en escroquant des clients de plateformes marchandes du dark web. Cette personne a révélé son escroquerie sur la page d'accueil de ses sites-leurres avant de disparaître.

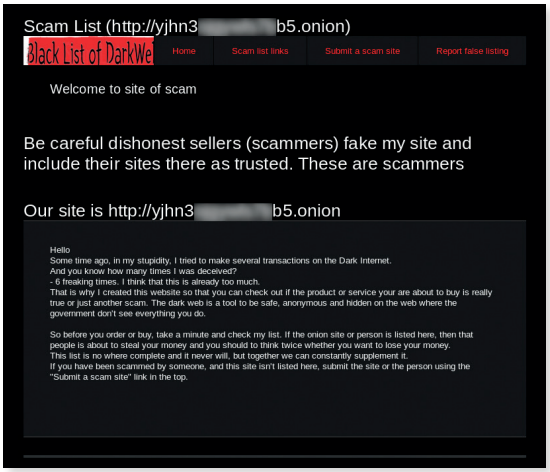
Il est donc probable qu'une grande partie des miroirs présents sur le dark web soit le fait d'un ou plusieurs imposteurs. Cette prolifération des miroirs-leurres est d'ailleurs une préoccupation pour les divers annuaires du dark web. En effet, pour être visibles et accessibles, ces leurres doivent être référencés sur les sites qui recensent les domaines du dark web. Certains annuaires tentent donc de mettre en place des stratégies pour identifier et cesser de référencer les sites piégés : les uns se basent sur leur intuition, d'autres se basent sur un retour d'utilisateurs, par le biais de votes.



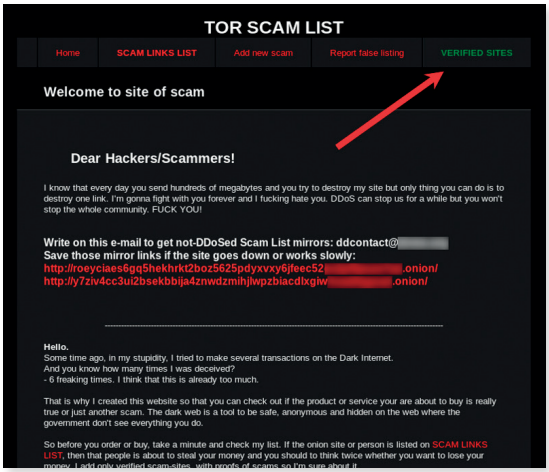
Evaluation de la qualité des sites sur la base de l'intuition et de l'expérience du responsable de l'annuaire



Evaluation de la qualité des sites sur la base de votes



Scam List Of Tor - site original



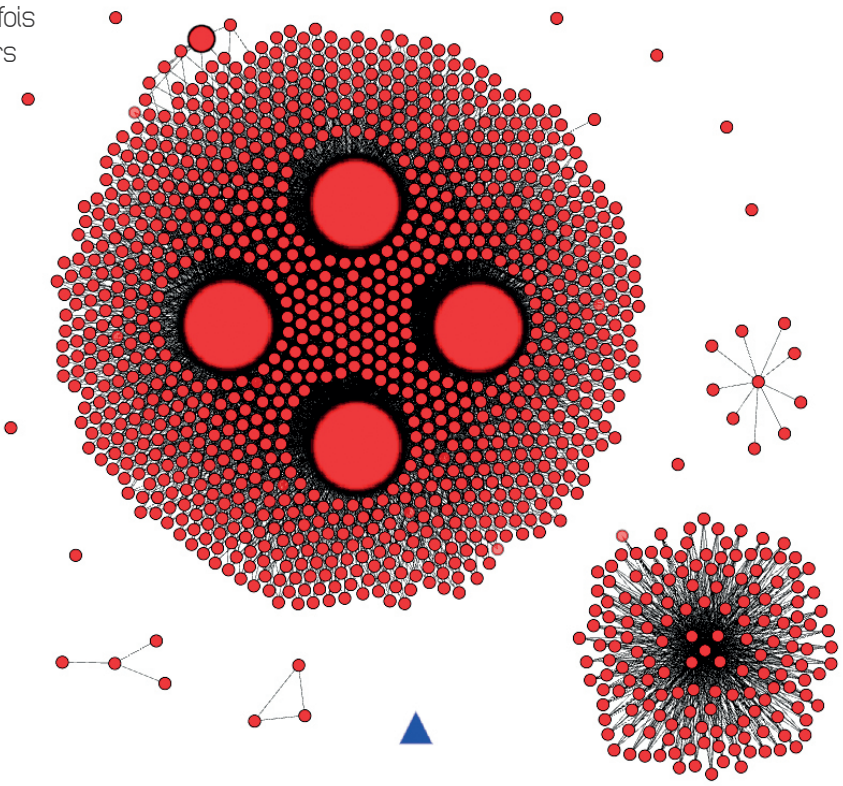
Scam List Of Tor - Miroir-leurre

En novembre 2018, un internaute victime à **six reprises d'escroqueries sur des miroirs-leurres** a même entrepris de créer une liste noire de sites usurpateurs : Scam List Of Tor. Ce site permet aux utilisateurs de signaler les sites suspects ("Submit a scam site") et éventuellement de disculper les sites inscrits à tort sur cette liste noire ("Report false listing").

Cette initiative n'est pas passée inaperçue auprès des créateurs de sites piégés. Un an plus tard, le premier miroir malicieux de "Scam List Of Tor " était créé. Comme le créateur du site original le signale sur sa page, les escrocs ont imité son site mais ont rajouté une section supplémentaire : les sites vérifiés. Cette section, censée désigner des sites de confiance, redirige les utilisateurs vers des leurres.

Depuis cette date, le site a été répliqué à 1 440 exemplaires, tous malicieux. Comble du raffinement, le ou les créateurs de ces "pièges à touristes" du dark web ont cette fois pris la peine de faire en sorte qu'une grande partie de leurs sites se référencent entre eux. Ils donnent ainsi l'illusion d'une plus grande crédibilité. Un utilisateur qui voudrait vérifier l'authenticité du site pourrait en effet être tenté de faire un contrôle de cohérence entre plusieurs versions de Scam List Of Tor. Comme les leurres se référencent les uns les autres, l'utilisateur est pris au piège, comme dans une toile d'araignée. En quelque sorte, nous pouvons parler ici d'une **"web of trust" inversée.**

Le graphe suivant illustre à la fois la ténacité des créateurs de miroirs malicieux (les cercles rouges) et le poids relatif du site original (le triangle bleu). Nous observons **deux réseaux très interconnectés** et **trois réseaux moins fournis**. Sur le plus gros réseau, quatre nœuds font office de référence : leur taille est en effet proportionnelle au nombre de liens qui dirigent vers ces sites.



Nous voyons ici l'ironie qui sévit sur le dark web : non seulement une grande partie de sa masse totale est constituée de sites qui escroquent des personnes qui pensent pouvoir effectuer des opérations frauduleuses en toute confiance, mais les sites censés dénoncer les tromperies sont eux aussi usurpés. Ceci ne doit toutefois pas occulter le caractère tout à fait réel et sérieux d'une autre partie du dark web. A côté de ces miroirs, il reste en effet un volume non négligeable de plusieurs milliers de sites, dont nous ferons prochainement un tour d'horizon.

A RETENIR :

1 - Une grande partie de la masse totale du dark web est constituée de miroirs.

2 - Un miroir procure à la fois une copie de sauvegarde mais permet aussi de se prémunir contre les attaques DDoS.

3 - Les miroirs peuvent servir aux usurpateurs et nous notons une prolifération de sites miroirs leurres ***#3 Les miroirs dans le dark web.***

4 - Les annuaires peuvent référencer des sites piégés.

Visualiser le dark web sous forme de graphe permet de déjouer les illusions tentées par les créateurs de sites malveillants. En effet, en analysant la forme des écosystèmes par la représentation sous forme de graphe, nous pouvons identifier des pratiques symptomatiques.

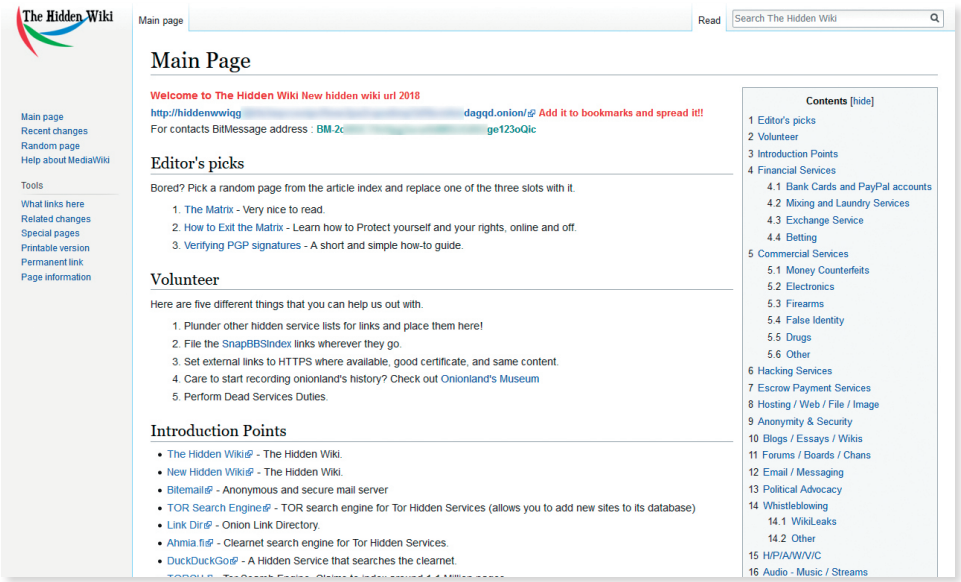
Se former aux pratiques utilisées sur le dark web permet de mieux analyser son contenu. En effet, chaque écosystème a ses propres comportements, pratiques. Nous n'analysons pas le contenu du dark web avec une approche réseaux sociaux, mais bien avec une connaissance comportementale des activités propres au dark web (processus de vol et vente de données, ultra-activisme, atteinte à l'image, contrefaçons, vol de PI, ransomware, ...)

Les moteurs de recherche

Contrairement à la navigation sur le clear web, le parcours du réseau Tor n'est pas guidé dès le démarrage du navigateur. En effet, les navigateurs classiques embarquent les principaux moteurs de recherche du clear web (Google, Bing, Yahoo, etc.). L'utilisateur peut donc effectuer ses recherches dans un terrain défriché et balisé. Tor Browser, lui, embarque DuckDuckGo. Ironiquement, DuckDuckGo n'indexe pas le réseau Tor ; ainsi, une recherche effectuée avec ce moteur de recherche renverra un certain nombre de résultats, mais aucun en provenance d'un site Onion.

> « Comment un utilisateur peut-il donc naviguer sur le réseau Tor ? Nous allons ici aborder les annuaires et les moteurs de recherche du dark web. »

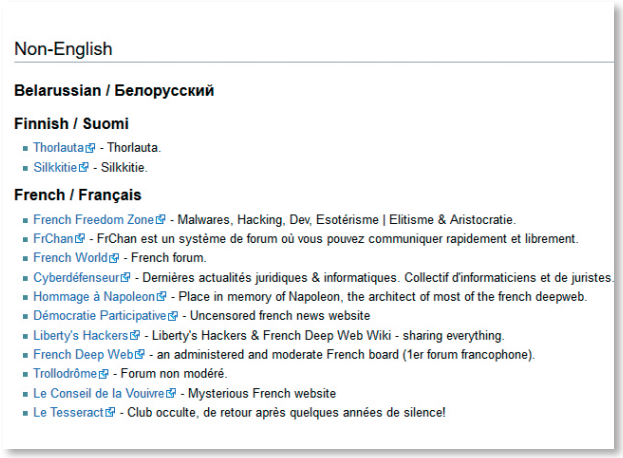
S'ils ne sont pas embarqués dans Tor Browser, il existe pourtant bien des moteurs de recherche du dark web, ainsi que des listes de sites Onion (Tor directories), qui fonctionnent comme des annuaires. À notre connaissance, les deux plus anciennes listes de sites sont **The Hidden Wiki et TorLinks** (rebaptisé OnionLinks). **Ces deux annuaires recensent une quantité très restreinte de sites (une centaine chacun), ce qui est peu par rapport au plus de 110 000 domaines Onion actifs détectés (V2) par notre moteur de recherche.**



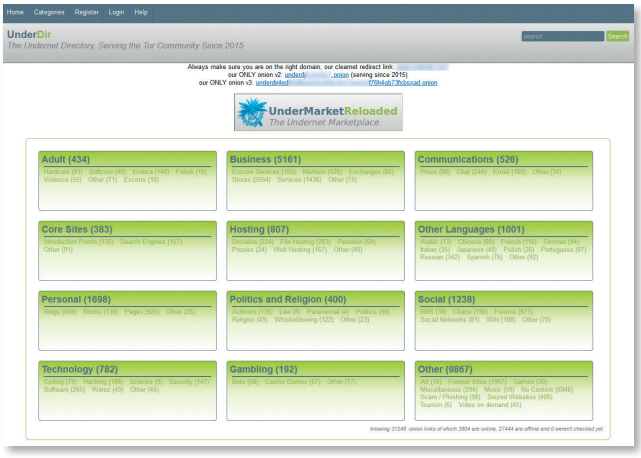
Moteur de recherche The Hidden Wiki

Moteur de recherche OnionLinks

Certaines sections du Hidden Wiki sont même tout à fait obsolètes, notamment la liste des sites francophones (Cf. image **hidden wiki non english**), puisque certains des domaines listés ont disparu depuis 2019 (French Deep Web, Liberty's Hackers). Sillkitie, un des deux sites finlandais, a lui aussi été démantelé il y a quelque temps. Certains annuaires plus récents comme **UnderDir** sont plus élaborés et cataloguent plusieurs milliers de domaines.

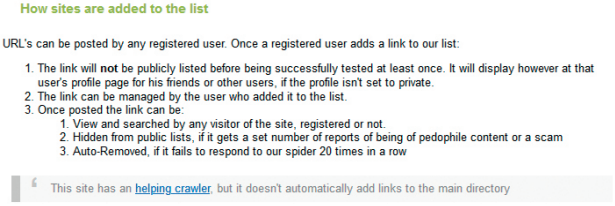


Site hidden wiki non-english



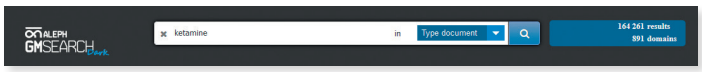
Site UnderDIR

C'est sur des annuaires tels que celui-ci que les créateurs de miroirs-leurres ***#3 Les miroirs dans le dark web.*** tentent de faire référencer leurs sites piégés. Un mécanisme de signalement a donc été mis en place pour que les sites référencés par **UnderDir** ne soient ni des **scams** (escroqueries) ni des **sites pornographiques**.

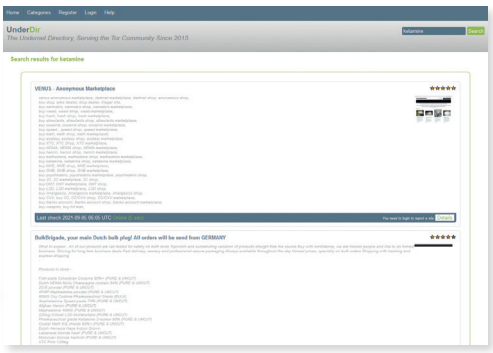


Nous pouvons remarquer l'existence d'une barre de recherche en haut à droite de la page d'accueil d'UnderDir. Nous avons donc voulu déterminer si cet outil de recherche pouvait répondre de manière satisfaisante à une requête d'un utilisateur du dark web. Eu égard au type d'activité présent sur ce réseau, nous avons opté pour une requête "ketamine", que nous avons par la suite effectuée sur plusieurs autres moteurs de recherche du dark web.

> "Là où notre moteur de recherche permet de faire remonter plus de 160 000 pages du dark web, UnderDir ne propose que deux résultats."



Aleph search dark



UnderDir

AHMIA	HOODLE	ONIONLAND SEARCH	TORCH	TORDEX	TORSEARCH ENGINE	Aleph Search Dark
95	100	167	6000	1000~	2	160 000

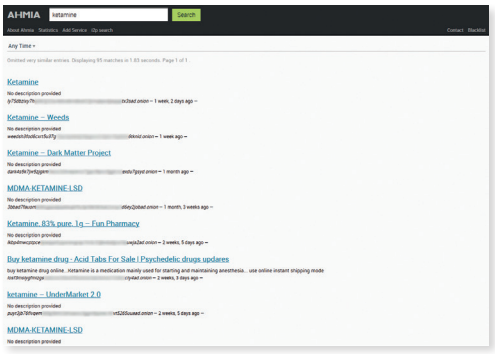
Tableau comparatif des résultats à la requête KETAMINE

Nous devons toutefois reconnaître que la vocation première d'UnderDir est d'être, comme son nom l'indique, un annuaire ou un répertoire (directory), la barre de recherche n'étant pas la principale valeur ajoutée du site. Il en va autrement pour les moteurs de recherche à proprement parler.

Nous avons sélectionné **sept moteurs de recherche parmi les plus réputés sur le réseau Tor** et nous leur avons soumis cette même requête "ketamine". Les résultats sont assez contrastés, mais toujours décevants, en termes de nombre de pages remontées. Nous vous livrons les chiffres de chaque moteur, par ordre alphabétique.

01

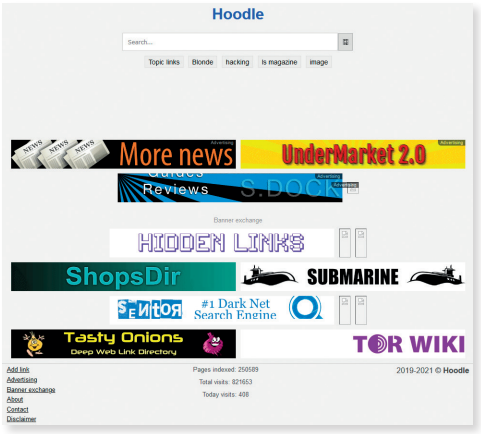
Ahmia, le premier moteur de recherche que nous avons testé, est le projet d'un chercheur finlandais nommé Juha Nurmi. Ce moteur exclut les contenus pornographiques. Il propose également une recherche sur le réseau I2P, ce qui est assez rare. **Notre requête "ketamine" fait remonter 95 résultats visiblement pertinents.**



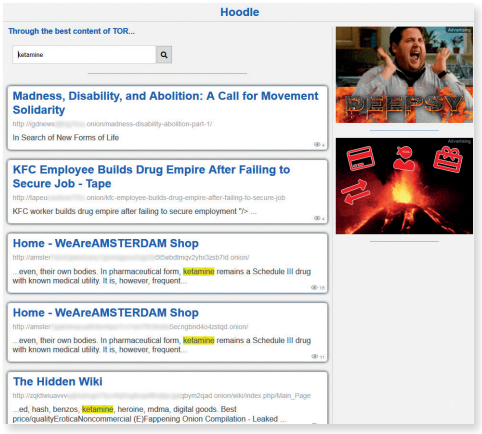
Ahmia

02

Contrairement à Ahmia, **Hoodle** est un moteur qui semble être rémunéré par de la publicité pour des places de marché clandestines, comme c'est le cas pour la plupart des moteurs du dark web. Nous remarquons en-dessous de la barre de recherche un top 5 des requêtes les plus courantes. **Deux d'entre elles, "Topic Links" et "LS magazine", ne laissent aucun doute quant au contenu recherché.**



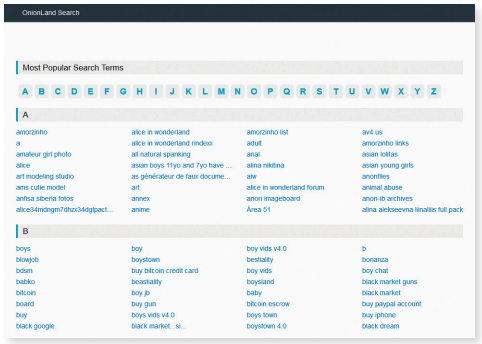
Hoodle



Notre requête donne 100 résultats qui semblent eux aussi assez pertinents.

03

OnionLand Search contient lui aussi des liens sponsorisés par d'autres sites du dark web. Notre requête nous donne 167 résultats, dont les premiers sont des liens vers des places de marché. Comme Ahmia, OnionLand Search propose une recherche sur le réseau i2p. Une section " Most Popular Search Terms " donne là encore une idée des recherches effectuées par les utilisateurs, sans qu'il n'y ait véritablement besoin de les qualifier.



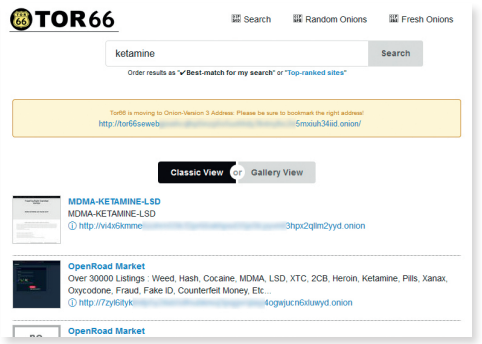
OnionLand Search

04

Tor66 est lui aussi un moteur sponsorisé par des places de marché. Outre sa fonction de recherche classique, il propose des liens aléatoires (Random Onions) pour les plus curieux - ou les plus inconscients. Le danger est toutefois atténué puisque le moteur affiche le titre des liens ; l'utilisateur a donc une idée plus ou moins précise du type de contenu aléatoire qu'il va consulter.

Tor 66

La requête " ketamine " donne 44 résultats. Ces 44 résultats ne sont pas des pages web précises mais des liens vers des domaines. L'utilisateur sera donc dirigé vers la page d'accueil de chaque site et il lui restera à trouver à l'intérieur de chaque domaine la page qui l'intéresse.



Tor 66

05

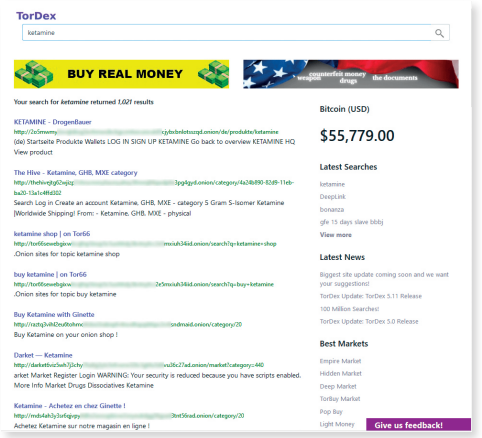
Sur le même modèle, **Torch** donne un nombre très significatif de résultats. Il propose pour notre requête plus de six mille résultats ; il s'agit cette fois des URL précises. Ainsi, non seulement en termes de nombre de résultats, mais également en termes de qualité de ce qui est proposé, Torch se positionne nettement en tête des moteurs que nous avons testés, même s'il ne rivalise pas avec un moteur de recherche professionnel.

Torch

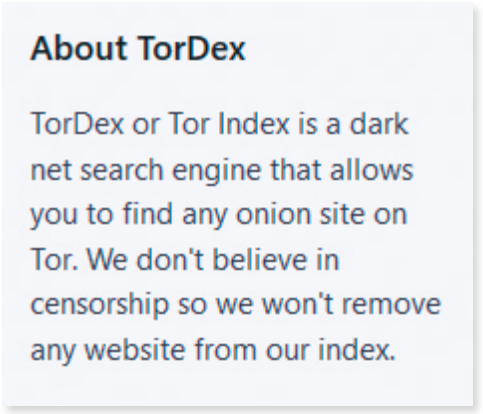
06

Le moteur de recherche **TorDex** ramène un peu plus de mille résultats, là encore en donnant les URL précises.

À la différence de certains moteurs de recherche, les créateurs de TorDex revendiquent le fait qu'ils n'opèrent pas de filtrage des contenus, qu'ils qualifient de censure.

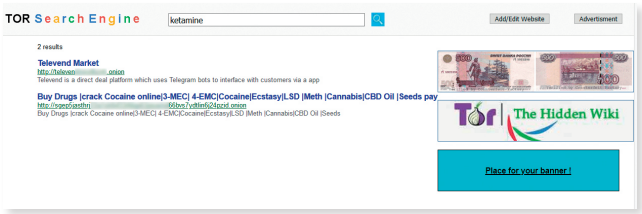


Tordex



07

Le moteur **Tor Search Engine** est le dernier et sans doute le plus décevant des moteurs que nous avons passés en revue, puisqu'en plus de son design rudimentaire, il ne nous donne que deux résultats.



Tor Search Engine

Les utilisateurs du réseau Tor ne disposent donc pas d'un Google du dark web, en ce sens qu'aucun moteur ne semble indexer les sites Onion de manière exhaustive. Certains moteurs testés ici donnent des résultats intéressants, mais bien en-dessous de ce que l'on pourrait attendre. Chaque utilisateur doit donc développer ses propres stratégies pour naviguer sur le dark web sans se perdre ni consulter des contenus qu'il n'aurait pas souhaité connaître.

A RETENIR :

1 -Tor Browser est un navigateur qui embarque un moteur de recherche pour le clear web et non le dark web.

2 - Il existe des moteurs de recherche et des annuaires aux résultats approximatifs.

3 - Certains moteurs de recherche sont sponsorisés et donc dépendants des sources (idem Google).

4 - Aucun moteur de recherche issu du dark web ne permet de remonter autant de résultats que le moteur de recherche **ALEPH SEARCH DARK** (solution professionnelle).

5 - Un moteur de recherche exhaustif est nécessaire pour "connaître" le web, qu'il soit dark ou clear.

Quelle est l'utilité d'intégrer un moteur de recherche sur Tor Browser qui ne remonte que les données du clear web ? Le premier apport du réseau Tor est d'accéder de manière pseudo-anonyme au web (et non pas accéder aux sites en .onion) il est donc naturel de retrouver un moteur de recherche par défaut qui garantit l'"anonymat des requêtes de recherche effectuées.

Et vous ? Comment accéderiez-vous au dark web et y feriez-vous des recherches dans le cadre professionnel ? L'unique solution aujourd'hui est l'utilisation de listes de sites du dark web comme hiddenwiki, mais aucune possibilité de faire une recherche exhaustive sans Aleph Search Dark.

Les différents types de sites du dark web

Si le dark web diffère du clear web par la teneur des contenus que l'on peut y trouver, les types de contenus sont globalement identiques dans leur forme. **Nous pouvons distinguer sept formes typiques de sites du dark web :**

- les forums et blogs,
- les sites marchands,
- les sites d'actualités,
- les bibliothèques,
- les sites de dépôt de code (repositories),
- les sites d'hébergement de fichiers (leak, ransomware, ...).

#5.1 - Les forums et blogs

Le **dark web** permet une liberté totale de ton ; il est donc logique qu'une part significative du dark web soit constituée d'espaces d'expression individuelle (blogs) ou d'échange communautaire (forums).

Dans leur forme, les blogs du dark web ne sont guère différents de ceux du clear web. Certains sont assez minimalistes et se contentent d'utiliser un format préétabli, comme les blogs hébergés par Torpress, tandis que d'autres sont bien plus élaborés au plan fonctionnel et esthétique.

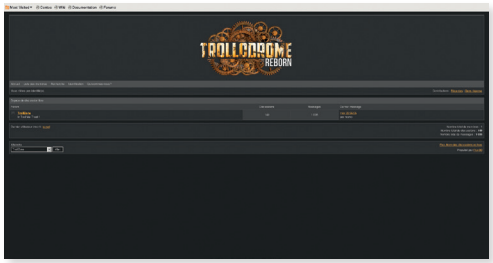
Blog grec hébergé par TorPress

Les forums du dark web, indépendamment des sujets qui y sont abordés, connaissent des modes de gestion très contrastés. Certains forums ne font l'objet d'aucune modération : n'importe qui peut écrire n'importe quoi, en utilisant (ou en usurpant) n'importe quel pseudonyme. Ce fut notamment le cas du défunt **Trollodrome** et de ses avatars actuels.

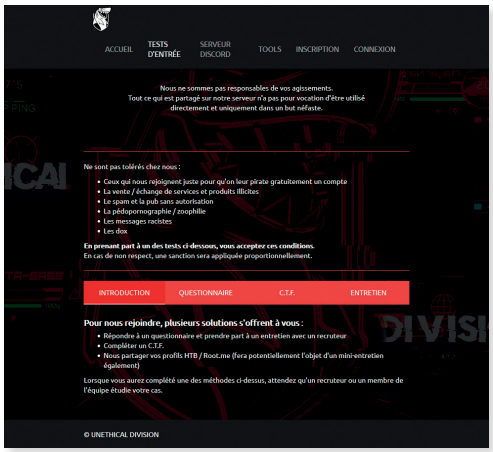
Outre le fait que ce mode de gestion aboutit inévitablement à des débordements tels que **diffamation, doxing, injures,** il encourage également les campagnes de publicité intempestives massives qui peuvent rapidement rendre le **forum illisible voire inutilisable.**

A l'inverse, certains forums pratiquent une politique très stricte en **termes de modération mais aussi d'accessibilité**. Sur de nombreux forums, l'inscription est en effet soumise à l'approbation des modérateurs. Elle est parfois conditionnée par des exigences draconiennes, telles que la cooptation par un membre du forum ou la rédaction d'une présentation personnelle ayant vocation à démontrer ce que le candidat peut apporter au forum. Le compte d'un utilisateur peut aussi être supprimé si celui-ci ne poste pas régulièrement des messages utiles à la communauté.

Blog d'un artiste américain



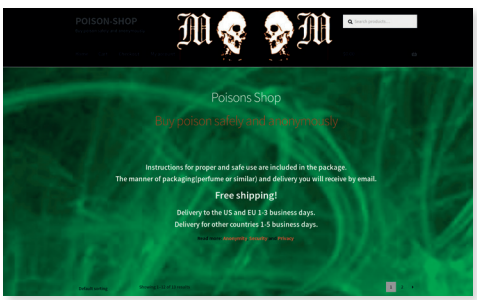
Page d'accueil du Trollodrome - Reborn, disparu depuis le mois d'août 2018



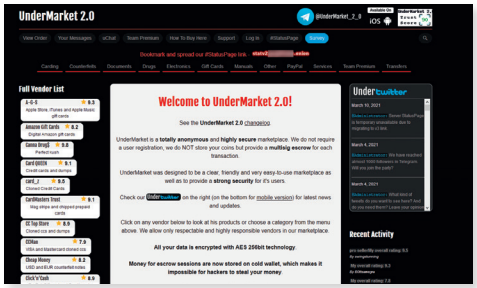
Conditions d'accès au site Unethical Division

22 #5.2 - Les sites marchands

De nombreux sites du dark web proposent des **biens** ou **services** en échange de paiement. Cela peut prendre la forme d'une **boutique individuelle, d'une place de marché où de multiples vendeurs** interviennent ou encore d'un **site à accès par abonnement payant.**



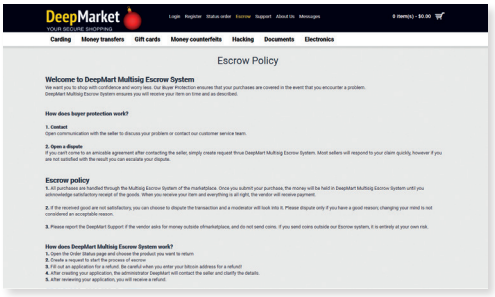
La boutique individuelle The Poison Shop



La place de marché multi-vendeurs UnderMarket 2.0

La nature sensible (et le plus souvent illégale dans de nombreux pays) des prestations ou biens vendus explique la présence de ces sites marchands sur le dark web. Il existe une grande variété d'offres marchandes. Parmi les plus répandues, nous pouvons citer **le recel de données privées ou de comptes bancaires piratés, la vente d'armes, de drogue, faux billets, fausses pièces d'identité ou faux diplômes, les cours de hacking payants, le commerce de contenu pédopornographique…**

Il n'est pas rare que des vendeurs indélicats sévissent sur le dark web. Nous voyons donc très souvent des places de marché se doter d'un service de fiabilisation des transactions, appelé **escrow**[2]. Ce service est habituellement rendu par un membre de la communauté. Il consiste pour cette personne tierce à percevoir et à bloquer le montant d'une transaction tant que l'acheteur n'a pas confirmé avoir reçu l'objet de son achat. Ce **"tiers de confiance"** se rémunère par commission sur les transactions qu'il garantit. Il peut arriver occasionnellement qu'un escrow profite de son statut pour bloquer abusivement de nombreuses transactions avant de disparaître avec la totalité des montants.



Système d'escrow du site DeepMarket

> "Il n'est pas rare que des vendeurs indélicats sévissent sur le dark web."

#5.3 - Les sites d'actualités

Comme sur le clear web, nous trouvons dans le dark web des sites de **relai d'information**. Leur ligne éditoriale est souvent très marquée, à l'extrême gauche comme à l'extrême droite. Il existe néanmoins des sites plus neutres, notamment des miroirs de sites d'actualités du clear web. D'autres sites n'ont pas de ligne éditoriale identifiable, comme le site **Saufça**.

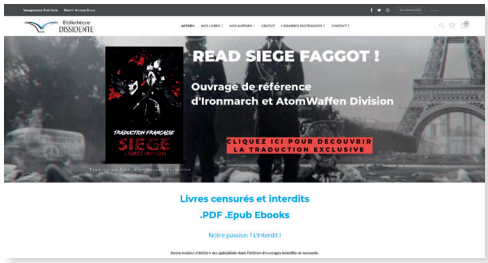
Il existe également des sites spécialisés sur certains sujets, en particulier sur l'actualité du dark web lui-même ou sur les technologies d'anonymisation.

Page d'accueil du site Saufça

2/ **Escrow** : désigne une personne tierce qui perçoit et bloque le montant d'une transaction tant que l'acheteur n'a pas confirmé avoir reçu l'objet de son achat.

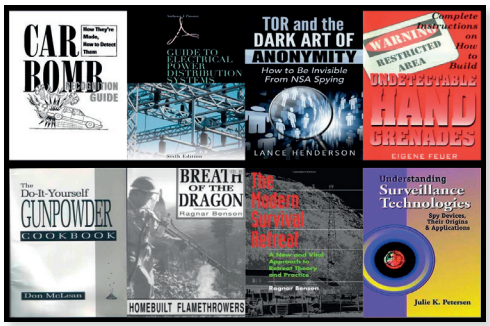
#5.4 - Les bibliothèques

On trouve sur le dark web de grandes quantités d'ouvrages sous forme électronique. **Contrairement aux sites marchands, les contenus proposés ne sont pas forcément sensibles.** On peut certes trouver des œuvres censurées ou des contenus dangereux (manuels de confection d'explosifs ou de construction de lance-flammes sur le site Anarchist Bookstore).



Le site Bibliothèque Dissidente

A l'inverse, on peut aussi trouver des contenus tout à fait anodins, dont la diffusion ne devrait pas nécessiter l'anonymat du dark web. Ainsi, on peut par exemple télécharger des revues de cotations boursières des années 90 sur le site d'une bibliothèque allemande. **Un site français propose même des collections intégrales de bandes dessinées (Astérix, Achille Talon, Ric Hochet...), des manuels scolaires et de nombreux numéros de la revue Science et vie Junior.**



Quelques-uns des ouvrages proposés sur le site de la bibliothèque Anarchist Bookstore

Site proposant des collections intégrales

#5.5 - Les sites de dépôt de code (repositories)

Le **dark web** contient de nombreux sites hébergeant du code. Le parallèle évident à faire avec le **web classique** est GitHub, service web de gestion et d'hébergement à des fins de développement logiciels.

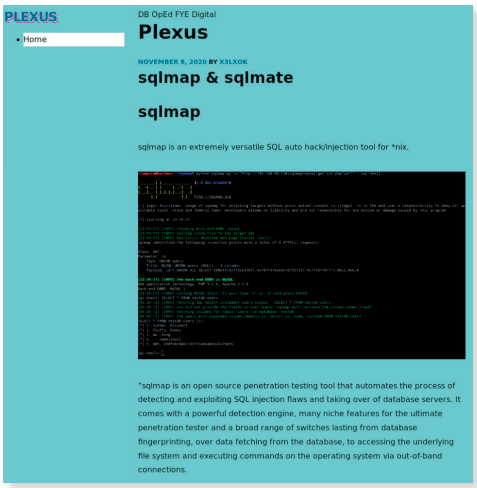
Le dark web comprend également ces services web, dans des versions dérivées.

Ainsi, nous pouvons trouver des personnes malintentionnées proposant des **malwares**[3] prêts à l'emploi ou encore des sauvegardes personnelles de répertoires de code. La finalité est toujours la même, trouver un espace d'expression libre et sans censure. Cette philosophie, constitutive du dark web, fait que les dépôts de code peuvent autant être de nature malveillante (comme dans la capture ci-après) qu'inoffensive, avec des communautés de développeurs de logiciels open source ou de sauvegardes personnelles à des fins éducatives.

> « Nous pouvons trouver des personnes malintentionnées proposant des malwares... »

3/ **Malware** : logiciel malveillant (virus, vers, chevaux de troie...)

Répertoire privé contenant divers copies de code source.



Exemple d'un logiciel malveillant en libre accès sur le dark web, le code source est disponible afin que chaque cyberattaquant puisse se l'approprier librement.

Le site personnel d'un développeur sur le dark web est illustré dans la capture ci-contre. Ce dernier semble encore actif car les projets présentés datent de plusieurs années mais certains sont régulièrement mis à jour. Après investigation, le code exposé semble inoffensif, et est une copie identique de son **GitHub**[4] issu du web classique.

Le développeur mentionne également qu'il préfère que le code associé à ses projets soit consulté sur le dark web, afin de garantir une liberté absolue.

Enfin, à la frontière entre les sites d'hébergement de code source et d'hébergement de fichiers, sont également présents sur le dark web **les services web de " Paste ".** Popularisé sur le web classique avec le plus connu : Pastebin. Pastebin est une application web permettant de publier un message sous forme texte, généralement du code source ou des URL, de façon anonyme et publique.

Ces applications sont populaires et faciles d'emploi, permettant en un faible temps de poster du texte et de le publier rapidement. Ces dernières existent également sur le dark web, garantissant un anonymat renforcé et une liberté de ton garantie.

Ainsi, ces services peuvent être utilisés pour diffuser du code source de malware rapidement entre personnes malveillantes, de poster des liens de téléchargement de fuites de données ou encore de publier des petites annonces concernant la vente d'armes ou de drogues.

Site communautaire de développeurs affiliés à des projets open source

Page du site DeepPaste, l'un des principaux paste du dark web.

Ce service comprend également plusieurs options d'affichage et de publication.

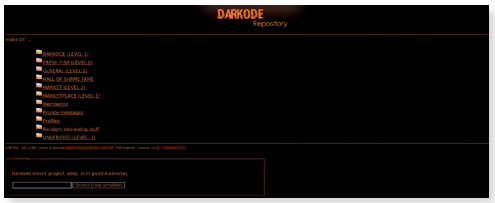
4/ **Github** : outil de gestion de source permettant de partager son code et de travailler à plusieurs sur une application.

#5.6 - Les sites d'hébergement de fichiers

Avant-dernière catégorie de sites présents au sein du dark web, se trouvent les sites d'hébergement de fichiers. A l'instar des catégories précédentes, cette typologie de services numériques reprend la même nomenclature que celle du web classique, en y ajoutant de nouveaux éléments.

Par conséquent, les éléments hébergés au sein de ces sites peuvent être de nature variée, de la plus anodine à la plus sensible. Nous pouvons trouver des revues de presse, des logiciels malveillants ou encore des fuites de données. Ainsi, il y a autant de types de sites d'hébergement que de typologies de fichiers à héberger.

Dans la capture ci-contre, nous pouvons trouver des projets publics en **lien avec des une communauté de cybercriminels, en l'espèce Darkmode Repository**. En effet, les noms de dossier (comme " hall of shame " et " marketplace ") ne laissent guère place au doute et nous imaginons aisément les activités associées.

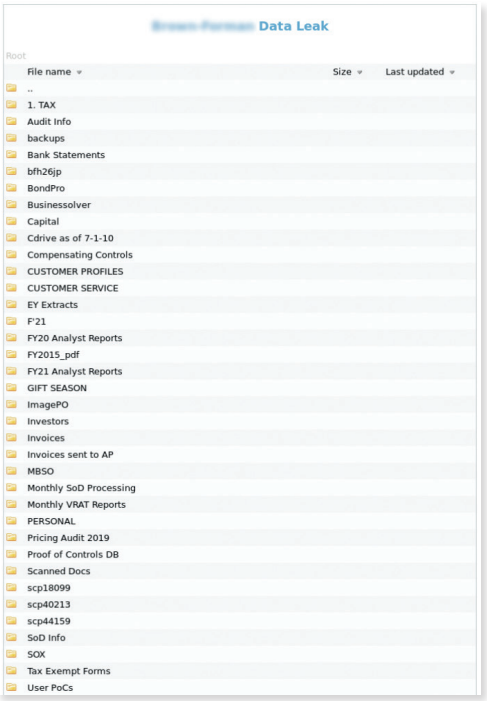


Site vitrine d'un groupe de cybercriminels

Cet exemple de site d'hébergement de fichiers, peut par exemple, servir de vitrine à une communauté de pirates pour gagner en exposition afin de vendre plus de services de hack et de cyberattaques ou tout simplement pour leur honneur et assoir leur légitimité dans un environnement concurrentiel.

Toujours dans l'activité illégale, existent des sites hébergeant en partie ou en intégralité des fichiers issus de fuites de données. Ces fuites de données, issues de cyberattaques, sont exposées sur le dark web comme moyen de coercition afin d'inciter à payer une rançon (comme dans le cas d'une cyberattaque au ransomware). Ces attaques, popularisés avec la pandémie sanitaire du Covid-19, a accéléré de façon exponentielle la surface d'exposition des systèmes d'information et par conséquence la survenance d'attaques.

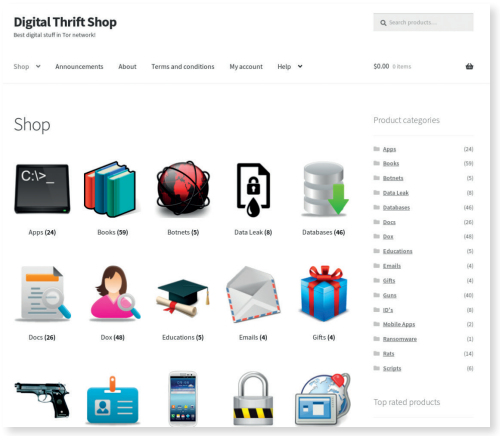
Le capture ci-contre illustre une fuite de données occasionnée par un ransomware contre une grande entreprise. Cette dernière n'ayant pas souhaité payer la rançon, le groupe de cybercriminels a décidé de publier en libre accès une partie des données volées.



Site en libre-accès hébergeant des données privées issues d'une cyberattaque

Enfin, il existe des sites d'hébergement de fichiers plus " classiques " Ces bibliothèques de logiciels permettent aux visiteurs de choisir les programmes de leur choix, à travers des catégories explicites.

La capture ci-contre illustre l'exemple du site Digital Thrift Shop, où les internautes peuvent télécharger aussi bien des fuites de données (dataleak) que des logiciels malveillants (ransomware) mais encore des livres et des vidéos éducatives.



Bibliothèque en ligne de logiciels en libre-accès

In fine, le dark web comprend la même typologie de site que celle du web classique. Toutefois, la liberté d'expression et l'anonymat garantis permettent le développement d'activités criminelles.

Territoire de non-droit, le dark web emporte autant de risques que de libertés.

Cet inventaire des sites présents au sein de cet espace prouve que le dark web, est tout en étant partie intégrante du web, un espace à part entière où règnent anarchie, logique mercantile et espaces d'expression libre.

A RETENIR :

1 - Les forums et blogs sont des espaces de liberté d'expression aux modes de gestion contrastés. La liberté d'expression de ces espaces est parfois contrastée par leur politique d'accès stricte (cooptation, obligation de poster…).

2 - Les bibliothèques sont des espaces de mise à disposition d'ouvrages hétéroclites où on peut tout aussi bien trouver des collections Astérix que des manuels de confection d'explosifs.

3 - Les sites de dépôt de code et d'hébergement de fichiers se différencient par leur contenu. Alors que les sites d'hébergement de fichiers permettent d'héberger plusieurs types de fichiers, les sites de dépôt de code permettent, quant à eux, d'héberger uniquement du code à des fins de développement logiciels.

4 - Le dark web est le premier réceptacle des activités de Ransomware (code et data).

L'accès aux informations diffusées sur le dark web vous expose si vous ne connaissez pas les codes de conduite. Il n'est pas évident d'interpréter les informations trouvées sur le dark web sans connaître les pratiques. De même entrer dans l'écosystème de manière discrète demande quelques compétences techniques et comportementales.

Savez-vous si votre entreprise ou vos données sont exposées sur un de ces sites ? Un moteur de recherche permet de répondre à cette question si vous n'avez pas la réponse. Il est important de se poser la question à date, mais aussi de mettre en place une surveillance dans la durée (protection des VIP, identification de vol ou de vente de données, identification de risque ultra-activiste, risque d'image, vol d'information stratégique, email et mot de passe…).

Conclusion

C'est grâce à notre moteur de recherche **Aleph Search Dark** que nos analystes ont une connaissance poussée du dark web.

Les différents modules d' **Aleph Search Dark** permettent non seulement d'accéder en toute sécurité aux données du dark web, mais permettent aussi de faire des analyses poussées, grâce aux modules de visualisation par les graphes par exemple.

La limite de la puissance d'**Aleph Search Dark** est celle que vous vous fixerez.

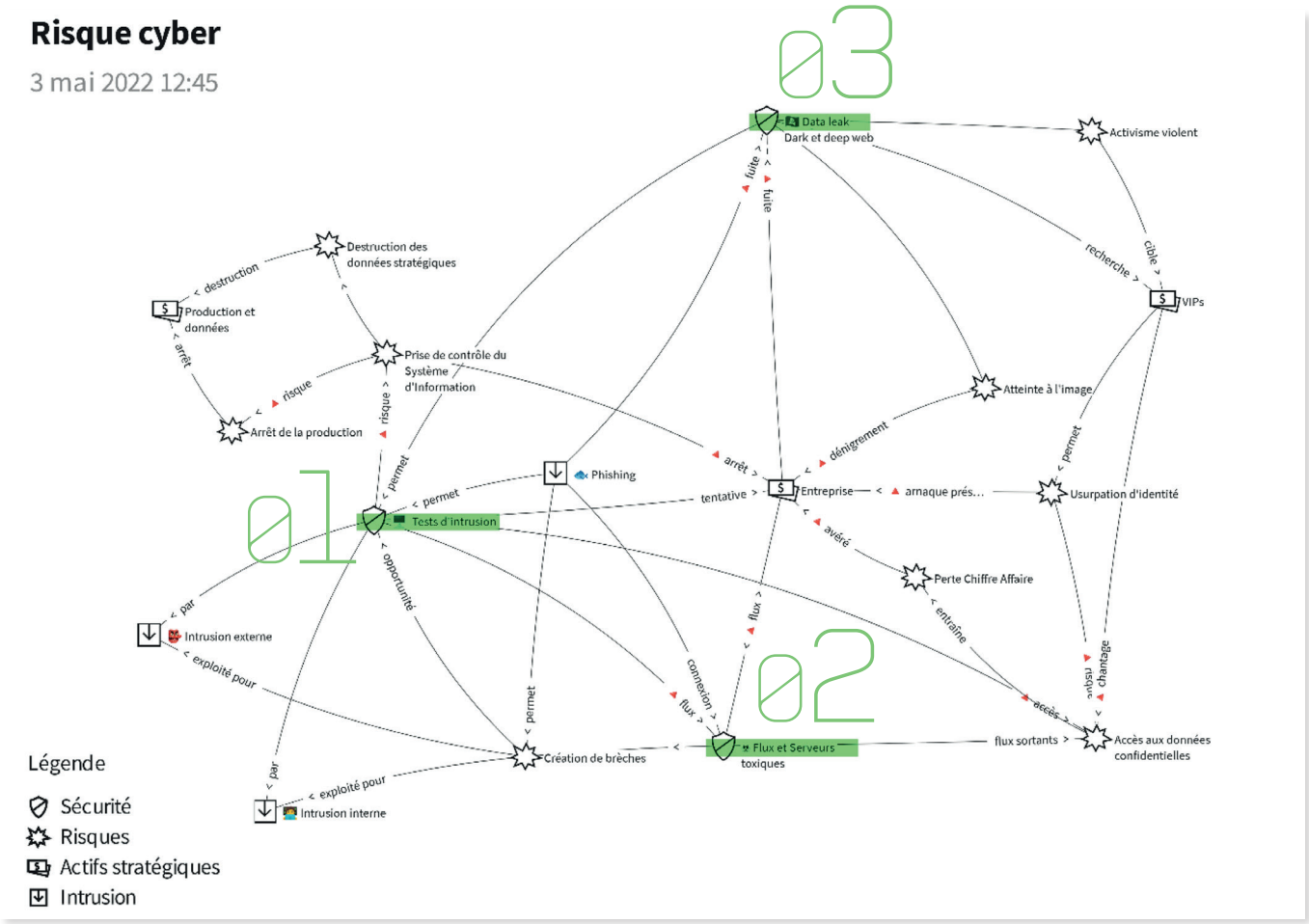
Si nous regardons d'un point de vue RISQUE NUMERIQUE (ou cyber), sur 3 pans à couvrir :

01 Test de pénétration

02 Analyse de Flux toxiques

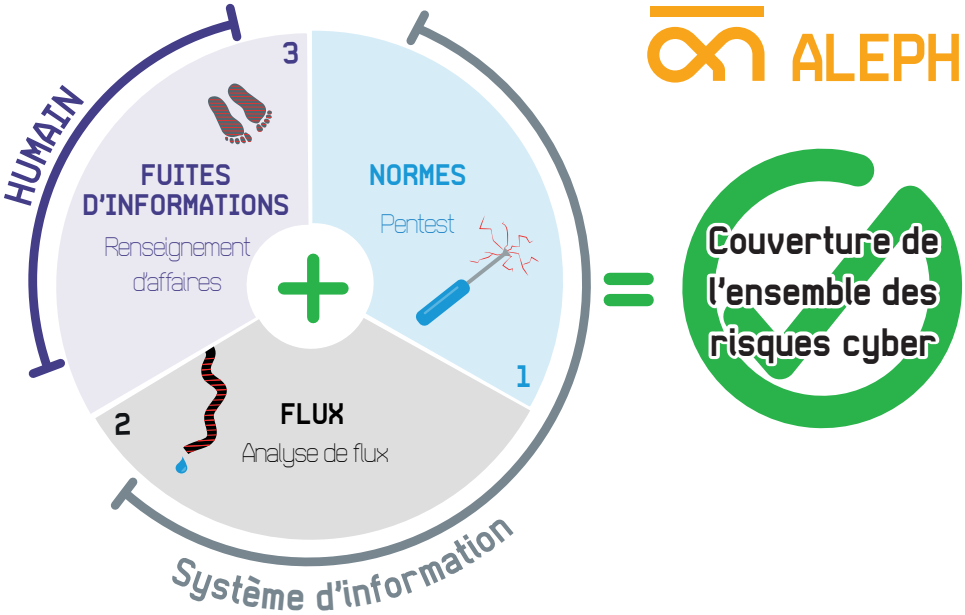
03 Recherche et surveillance de leak et exposition à risque pour les VIP

Le dark web (et le deep web) ne concerne directement que le 3ème point et indirectement les deux autres.



HTTPS://GOSTEMICAPP/MAPS/A2E8B1CF-22A2-41A3-83B8-34D9DE3C58BB

Il y a donc une nécessité méthodologique de surveiller les actifs de son entreprise dans le dark web pour couvrir à 360° le risque Cyber. Il est bien évident que la notion de "best effort" chère aux assureurs comprend la surveillance du dark web, en préventif mais aussi en post-incident (fuite d'information des employés, RGPD, ...)

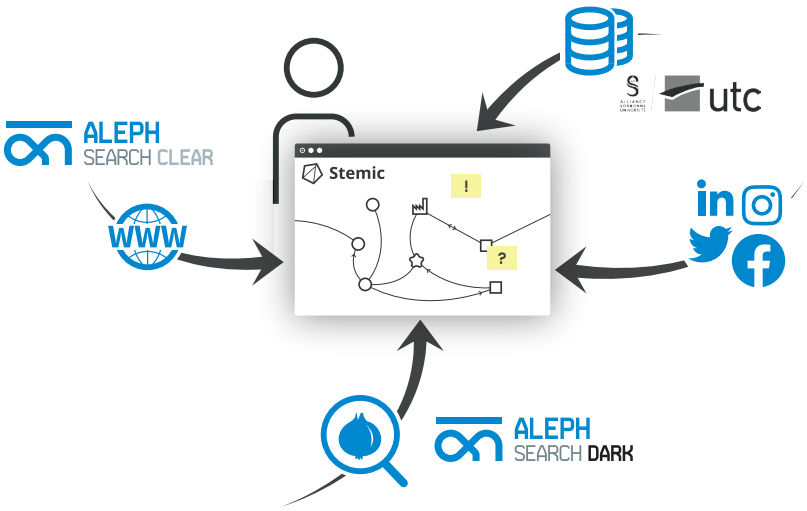


Couverture de l'ensemble des risques cyber;

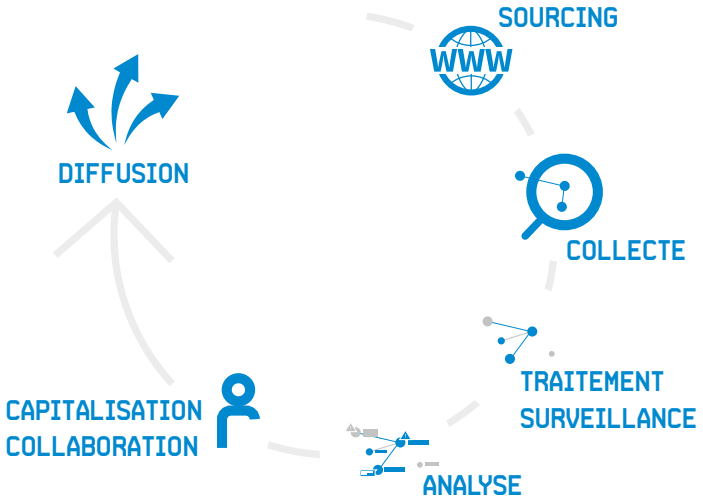
Si nous regardons d'un point de vue du RENSEIGNEMENT NUMERIQUE (ou cyber), 3 enjeux sont à couvrir :

01 Dépasser le mur de la donnée

02 Élargir son environnement informationnel :



03 Maîtriser la chaîne informationnelle de bout en bout :



Les webs comme sources d'information pour le Renseignement d'Intérêt Cyber.

Le Renseignement d'Intérêt Cyber est primordial pour mettre en place un processus de Lutte Informatique, reposant sur 3 piliers : la lutte **Informatique Défensive, Offensive et d'Influence.**
Pour lutter contre les menaces cyber, ce renseignement doit trouver ses sources dans toutes les couches du web : clear, deep et dark webs, réseaux sociaux...

Ce premier tome vous a plu ?

Vous voulez en savoir encore plus ?

ALEPH

Nous avons hâte que vous découvriez notre Tome 2, dont la sortie est programmée en septembre.

Ce tome 2 continuera d'éclairer le dark web et vous permettra d'explorer son contenu, du plus sombre au plus insolite.

Restez informé !

Suivez-vous sur Linkedin

Abonnez-vous à notre newletter

Abonnez-vous à notre blog

ALEPH

Analyser, protéger et influencer
votre environnement professionnel

contact@aleph-networks.com
www.aleph-networks.com

45, rue d'Alma
69400 Villefranche s/Saône
France