I Draghi

Iscriviti alla newsletter su www.lindau.it per essere sempre aggiornato su novità, promozioni ed eventi. Riceverai in omaggio un racconto in eBook tratto dal nostro catalogo.

In copertina: Adobe Stock © Prapat

corso Re Umberto 37 - 10128 Torino

Prima edizione: settembre 2020
ISBN 978-88-3353-344-5

Emanuele Florindi

DEEP WEB

Vizi privati e pubbliche virtù della navigazione in rete

LINDAU

DEEP WEB

Nella vita bisogna fare tre cose: fare un figlio, scrivere un libro, piantare un albero
Proverbio zen

A Francesco, Maria Giulia e Federico

Premessa

Sono trascorsi quasi quattro anni dalla prima pubblicazione di questo libro e molte cose sono cambiate: le criptovalute, ormai uscite dall'oscurità e dai siti di appassionati, stanno ampiamente dimostrando sul campo il loro valore e la loro affidabilità mentre le forze di polizia hanno dimostrato sempre più spesso di essere in grado di intervenire e operare anche nelle dark net. Nel corso del 2018 l'Interpol ha evidenziato i rischi rappresentati dalla crescita delle *altcoin*, identificata come un pericolo emergente nel corso del primo INTERPOL Working Group su dark net e criptovalute.

Al tempo stesso, la piena entrata in vigore del Regolamento europeo in materia di protezione dei dati personali ha reso maggiormente evidente l'importanza di proteggere la riservatezza dei dati personali.

Ci troveremo quindi ad affrontare di nuovo la tematica relativa alla navigazione anonima, al *deep web*, al *dark web*, alle *dark net* e ai sistemi di pagamento basati sulla cosiddetta «criptovaluta» e lo faremo utilizzando un linguaggio il più possibile semplice e comprensibile, anche a discapito di quella rigorosa precisione che richiederebbe un eccessivo impiego di termini tecnici.

Questo significa che, ove necessario, alcuni termini verranno utilizzati con un significato più ampio rispetto a quello strettamente tecnico (tanto informatico quanto giuridico) e che si cercherà di semplificare alcune nozioni e alcune procedure, ferma rimanendo la correttezza di base di tutte le informazioni fornite.

Ciò detto, nel procedere all'esplorazione di quel mondo affascinante che è il deep web, alcune premesse sono d'obbligo: per prima cosa, con l'espressione «web sommerso» (in inglese *deep web*), indichiamo quella parte della rete che non è immediatamente accessibile attraverso i normali motori di ricerca e, spesso, neppure attraverso gli ordinari strumenti di navigazione.

In secondo luogo, gli strumenti di cui parleremo (in particolare TOR e gli *hidden services*) sono perfettamente legali e utilissimi, tanto che in molti casi il loro utilizzo è addirittura raccomandato per tutelare il proprio anonimato e la propria riservatezza. Deep web non significa, automaticamente, contenuti illegali o pericolosi come, invece, si tende a credere e immaginare.

Volendo fare un esempio, possiamo paragonare TOR (preso qui a simbolo di tutti gli strumenti di anonimato) al mephisto: quel capo di abbigliamento, un semplice passamontagna nero, che viene utilizzato tanto dai corpi speciali nel corso di rischiose operazioni quanto da molti pericolosi criminali! In breve, il medesimo strumento può essere indifferentemente impiegato per nobili finalità (evitare rappresaglie e rischi per gli operatori coinvolti nelle attività e per le loro famiglie) o per altri scopi molto meno onorevoli (evitare di essere riconosciuti dalle vittime del reato o da eventuali testimoni).

TOR funziona allo stesso modo: può essere nobilmente impiegato per garantire la libertà di stampa o il diritto di

esprimere liberamente le proprie opinioni per individui che vivono in Paesi scarsamente democratici, oppure può essere utilizzato in modo ignobile per svolgere attività illegali…

TOR, l'anonimato e le stesse criptovalute sono, quindi, strumenti perfettamente neutrali e, alla fine dei giochi, legittimi: tutto dipende dalle attività che svolgono coloro che li impiegano e dal loro rispetto per la legalità e per gli altri, ma nulla di quanto viene fatto può essere imputato agli strumenti o ai loro sviluppatori.

Da ultimo, una precisazione di carattere terminologico: esistono numerose definizioni di *deep web*, *dark web* e *surface web*. Esiste anche una sorta di classifica dei livelli di profondità del web, ma, a mio modo di vedere, la classificazione migliore, in quanto più immediata e più comprensibile anche se non sempre precisa al 100%, è la seguente:

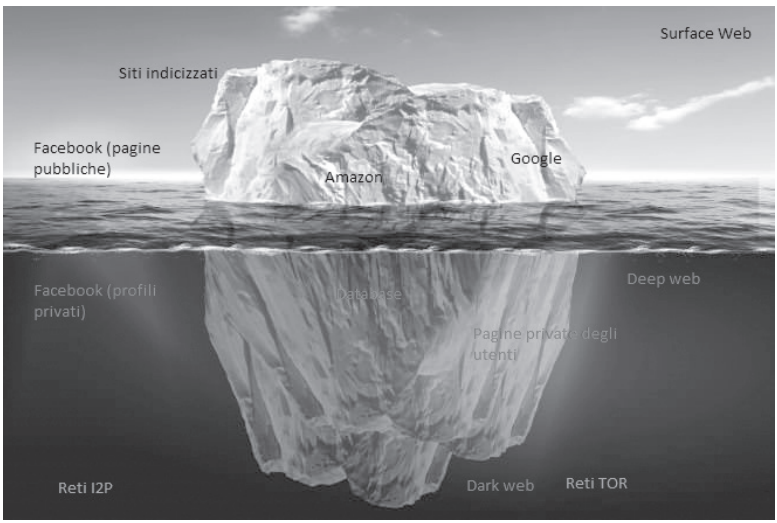
Surface web: tutte quelle pagine web e quei contenuti che vengono indicizzati dai motori di ricerca;

Deep web: tutte quelle pagine web e quei contenuti che non sono indicizzati dai motori di ricerca, ma sono accessibili senza la necessità di impiegare appositi strumenti software al di fuori dei comuni programmi di navigazione, consultazione e scambio di file;

Dark web: quella parte del *deep web* che non è accessibile attraverso i normali programmi, ma richiede l'impiego di accorgimenti e programmi particolari, come, ad esempio, TOR.

1

Deep e dark web



Il web è molto simile a un grosso iceberg; la parte che noi vediamo è solo una piccola percentuale della sua massa reale. La maggior parte delle informazioni sono, in realtà, custodite al di sotto della superficie. In questa immagine si trova una rappresentazione schematica dei tanti livelli in cui è possibile dividere la rete.

1. Facciamo un po' di chiarezza

Per poter comprendere di cosa stiamo parlando e affrontare correttamente la tematica relativa al deep web e alle dark net, dobbiamo necessariamente distinguere tra le differenti tipologie di contenuti e, a tal fine, ricorreremo all'ormai classica immagine dell'iceberg: la parte emersa, quella che tutti vediamo, rappresenta i siti web e i contenuti «indicizzati» dai motori di ricerca.

Un «motore di ricerca», in inglese *search engine*, non è altro che un sistema automatico che, su richiesta degli utenti, analizza un insieme di dati (in genere raccolti e catalogati sulla base di parole chiave presenti nel testo, di descrizioni, di *tag* oppure basandosi su riferimenti incrociati tra le differenti pagine) e restituisce, sulla base di specifici algoritmi, un indice dei contenuti disponibili.

Spesso i contenuti sono classificati, in modo automatico, in base a formule che, ricorrendo a principi statistici e matematici, indicano il livello di rilevanza di un determinato contenuto data una certa chiave di ricerca, spesso abbinandola a un determinato utente.

In genere, i motori di ricerca trovano utilizzo nel campo dell'*information retrieval* e nella ricerca di contenuti web; in quest'ultimo caso, i contenuti raccolti e catalogati da un singolo *search engine* vengono definiti «indicizzati» e vengono offerti agli utenti tra i risultati delle ricerche.

Si tratta di quello che è comunemente noto come *surface web* (in alternativa è possibile trovarlo indicato come Visible Web, Clearnet, Indexed Web, Indexable Web or Lightnet). Questi siti sono generalmente accessibili al grande pubblico degli utenti e le informazioni in essi contenute possono essere recuperate attraverso l'impiego di motori di ricerca classi-

ci (vale a dire Google, Bing, Yahoo). Giusto per comprendere appieno la mole di informazioni di cui stiamo parlando, ci basti qui osservare che, in base alle statistiche pubblicate in varie fonti di informazione, a ottobre 2019 risultavano indicizzate quasi 65 miliardi di pagine web.

Il web di superficie è composto essenzialmente da pagine web statiche, alcune delle quali, poi, possono anche connettersi a contenuti presenti nel *deep web*, come ad esempio accade per Facebook, Registro Imprese, albi professionali e molti dei cataloghi interni alla maggior parte dei siti web.

In estrema sintesi, il web di superficie si compone prevalentemente di pagine statiche, che risiedono in un server web attendendo di essere visualizzate dai vari navigatori; questo strato superficiale rappresenta circa il 10% del materiale realmente presente in Internet.

2. Il deep web

Subito sotto lo strato di superficie troviamo il web profondo, ma non si deve commettere il facile e grossolano errore di identificare *tout court* il *deep web* con i contenuti illegali o pericolosi.

In realtà il *deep web* è costituito da una serie di contenuti non direttamente indicizzati o indicizzabili dai motori di ricerca per diverse ragioni, di solito assolutamente legali e legittimi, che, in linea di massima, possono essere ricondotti alle seguenti categorie:

– *Contenuti dinamici*: si tratta di pagine web dinamiche, il cui contenuto viene generato sul momento dal server; le pagine possono essere richiamate solo compilando un *form* ovvero rispondendo a una particolare richiesta. Si tratta, ad esem-

pio, di quei contenuti presenti in un database con un motore di ricerca interno (Pagine Bianche, Registro Imprese…);
– *Pagine non indicizzate*: si tratta di pagine web che non sono collegate a nessun'altra pagina web e a cui l'accesso da parte dei motori di ricerca viene impedito da adeguate impostazioni di sicurezza (ad esempio, il file di testo «robot.txt» indica a un motore di ricerca quali parti di un sito non devono essere indicizzate) che ne impediscono l'indicizzazione (ad esempio le pagine dell'albo pretorio online);
– *Pagine ad accesso ristretto*: si tratta di pagine che richiedono una qualsivoglia forma di autorizzazione (cioè una registrazione) per poter essere visionate come accade nei siti che limitano l'accesso alle proprie pagine interne;
– *Scriptpage*: pagine che possono essere raggiunte solo attraverso link realizzati in JavaScript o in Flash e che, quindi, richiedono procedure particolari per poter essere visualizzate;
– *Contenuti non testuali*: essenzialmente si tratta di file multimediali privi di tag che, per tale ragione, non possono essere indicizzati dai motori di ricerca. Può trattarsi di fotografie, filmati, file audio ecc.
– *Altri contenuti banditi dai motori di ricerca*: di questa categoria fanno parte singole pagine o interi siti, eliminati dai motori di ricerca in quanto il loro contenuto è ritenuto non compatibile con i termini del servizio offerto.

Tutti questi contenuti possono essere visualizzati e acceduti senza particolari difficoltà tecniche: è sufficiente utilizzare il semplice *browser* che impieghiamo per navigare senza dover modificare le nostre abitudini.

L'unica difficoltà, relativamente alla possibilità di accedere a queste informazioni, risiede nell'esigenza di conoscere la password necessaria per l'accesso, ovvero un indirizzo

URL preciso dato che, in caso contrario, l'accesso non avviene semplicemente perché il materiale non può essere individuato.

3. Il dark web

Con il termine *dark web* si identifica la parte più profonda e segreta del *deep web*, costituita dalle cosiddette *dark net*: si tratta di contenuti nascosti intenzionalmente ai comuni navigatori e accessibili soltanto attraverso appositi strumenti.

Per utilizzare la definizione più corretta, possiamo affermare che il *dark web* è costituito da alcune migliaia di siti che utilizzano strumenti di anonimato (come ad esempio TOR o I2P) per nascondere la loro effettiva collocazione.

A differenza di quanto accade per tutte le altre tipologie di servizio, per poter accedere al *dark web* e visionarne i contenuti è necessario utilizzare appositi programmi, in grado di sfruttare la struttura e le regole proprie di Internet, in aggiunta a un proprio protocollo di connessione, in modo da garantire la navigazione attraverso una rete parallela molto difficile, se non impossibile, da tracciare.

2

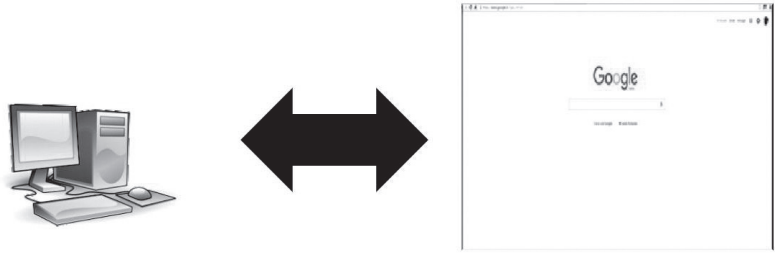
La ricerca della riservatezza

1. Navigare con la targa

A ogni utente che si collega a una rete, sia questa una piccola rete locale o la rete internet, viene assegnato uno specifico indirizzo. Tale indirizzo, generalmente noto come indirizzo IP, è composto da quattro gruppi di numeri, da 0 a 255, intervallati da un punto (per esempio 192.168.1.1) e rappresenta una sorta di «targa» che l'utente si porta dietro nel corso della sua navigazione.

Per poter visualizzare il contenuto di un sito web è infatti necessario che tra il computer del navigatore e il server che ospita il sito che si vuol visualizzare si instauri una connessione e avvenga uno scambio di informazioni: questo può accadere soltanto se entrambe le parti conoscono con precisione dove inviare i propri dati e le proprie richieste.

Questa informazione è rappresentata dall'indirizzo IP.



Tra il server che ospita il sito web e il computer dell'utente deve instaurarsi un «dialogo». Questo «dialogo» sarà possibile soltanto se entrambi gli interlocutori sanno dove (a chi) inviare le informazioni.

Ne consegue non soltanto che l'utente sarà in grado di conoscere l'indirizzo IP del server che ospita il sito, ma anche che quest'ultimo sarà in grado di conoscere l'indirizzo IP da cui viene generata la connessione.

Una volta appreso l'indirizzo IP di origine, sarà facile, utilizzando appositi servizi, conoscere il fornitore della connessione internet e da ultimo, ricorrendo all'autorità giudiziaria o ad altri sistemi, arrivare a conoscere l'identità dell'intestatario del contratto.

Per evitare questo, sono stati sviluppati vari strumenti in grado di tutelare la riservatezza dei navigatori nascondendo o rendendo più difficile da individuare l'indirizzo IP.

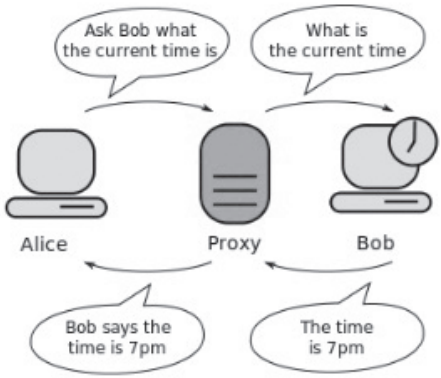
2. Una prima soluzione: il proxy

Nel paragrafo precedente abbiamo visto che esistono diversi strumenti per nascondere le proprie tracce durante la navigazione. Vediamo ora, sia pure in maniera sommaria, come è possibile farlo.

Uno dei primi strumenti impiegati, ancora oggi in

uso, per difendere la propria riservatezza è rappresentato dall'impiego di un intermediario tra il computer di origine e quello di destinazione.

Si tratta di quello che in informatica viene definito *proxy* (semplicemente «intermediario» in inglese): in estrema sintesi, si tratta di un *server* che si interpone tra un computer e la sua destinazione, facendo da tramite tra i due sistemi (un *client* e un *server*) e inoltrando le richieste e le risposte dall'uno all'altro: in breve il *client* si collega al *proxy* invece che al *server* e gli invia delle richieste. Il *proxy* a sua volta si collega al *server* e inoltra la richiesta del *client*, riceve la risposta e la inoltra al *client*.



Nel dialogo tra il server che ospita il sito web e il computer dell'utente si inserisce un intermediario che veicola domande e risposte

Esistono varie tipologie di *proxy*, ma, ai fini che qui ci interessano, ci basta ricordare le seguenti:

– *Transparent proxy*: intercettano le normali comunicazioni senza richiedere particolari configurazioni. Il loro impiego

tipico è o all'interno di grosse infrastrutture, in modo da imporre più facilmente adeguate regole di utilizzo del servizio, ovvero nell'ambito degli Internet Service Provider, in modo da migliorare le prestazioni relative alla navigazione dei propri utenti. Di solito si tratta di servizi che non hanno alcuna rilevanza in tema di tutela della riservatezza e sono finalizzati esclusivamente a migliorare la navigazione.
– *Anonymous proxy*: non trasmettono l'IP del richiedente, ma modificano o aggiungono alcuni *header*, rendendoli quindi facilmente riconoscibili e contrastabili. Per evitare ciò, alcuni (*anonymous proxy distorcenti*) trasmettono un IP casuale, diverso da quello del richiedente, e modificano o aggiungono alcuni *header*. Solitamente vengono scambiati per *proxy* anonimi, ma offrono una protezione maggiore, in quanto il server web vede le richieste di un utente provenienti da indirizzi IP diversi.
– *Highly Anonymous proxy*: questi non trasmettono l'IP del richiedente e non modificano gli *header* della richiesta. Sono difficili da riconoscere attraverso i normali controlli, ma sono molto rari.
– *Proxy CGI*: sono *proxy* che, attraverso un'interfaccia web, consentono di visitare altri siti in modo anonimo direttamente attraverso un sito web, senza necessità di modificare le impostazioni del browser (possono anche essere configurati per rifiutare cookie, rimuovere pubblicità ecc.).
– *Tor onion proxy software/ I2P proxy*: si tratta di vere e proprie catene di *proxy*, gestite da specifici programmi che verranno trattati in maniera approfondita nei prossimi paragrafi.

I *proxy* presentano numerosi problemi: il loro numero è relativamente basso e ancora meno sono quelli che consentono di navigare a una velocità decente. Molti sono creati e mantenuti da agenzie governative che li possono agevol-

mente impiegare per monitorare traffico e attività di coloro che li utilizzano. Alcuni supportano la sola navigazione e non, anche, altri protocolli.

Vediamo ora come è possibile tutelare la propria riservatezza utilizzando alcuni dei più diffusi strumenti di navigazione. Prima di procedere è però necessario chiarire un paio di punti.

3. Anonymity is not a crime, it's a right

Non sono soltanto i criminali ad avere bisogno di tutelare la riservatezza dei propri dati e delle proprie informazioni. Chiunque dovrebbe evitare di esporre un numero eccessivo di dati.

A tale proposito, è bene ricordare che l'anonimato non è un crimine, ma un diritto fondamentale, come ricorda la Carta dei diritti fondamentali dell'Unione Europea (2007/C 303/01, articolo 8) secondo cui: «Ogni persona ha diritto alla protezione dei dati di carattere personale che la riguardano e tali dati devono essere trattati secondo il principio di lealtà, per finalità determinate e in base al consenso della persona interessata o a un altro fondamento legittimo previsto dalla legge».

Molti soggetti affermano che se un individuo non ha nulla da nascondere allora non ha nulla da temere dalla trasparenza; ne consegue, logicamente, che la tutela della riservatezza servirebbe soltanto a criminali e depravati per nascondere le proprie nefandezze.

La realtà è molto differente, atteso che ognuno di noi ha un segreto: che si tratti di qualcosa di cui ci vergogniamo o, semplicemente, di un'informazione che non gradiamo vede-