DEEP WEB

Il Deep Web è uno dei fenomeni legati al mondo di Internet di cui si sente parlare sempre più spesso.

Il World Wide Web, oggi conosciuto anche come il web di superficie, ha in realtà un'altra faccia, di gran lunga più grande e per lo più sconosciuta fino a poco tempo fa. Questa faccia costituisce il profondo web.

La prima volta che viene menzionato il Deep Web è il 1994, anno in cui si parlava di web invisibile. Il deep web non è davvero invisibile, ma i suoi contenuti non sono indicizzabili dai motori di ricerca di oggi, risultando invisibile all'utente medio di Internet. Secondo il Guardian, Google indicizza solo lo 0,03% di risorse presenti in Internet.

Per capire la nascita del Deep Web abbiamo bisogno di capire come lavorano i motori di ricerca il cui scopo primario è quello di consentire alle persone di trovare informazioni sul web.

Quando utilizziamo un motore di ricerca, non facciamo altro che richiedere le pagine web in cui sono presenti uno o più termini di interesse. La ricerca, però, non avviene effettivamente su internet, ma sugli indici del web creati sui server dei vari motori di ricerca esistenti. Questi strumenti svolgono l'arduo compito di indicizzare l'intero World Wide Web, creando e mantenendo aggiornate le proprie Banche Dati. Attualmente, però, i più efficienti motori di ricerca riescono ad indicizzare da un terzo alla metà dei documenti disponibili per il pubblico su Internet. Un così basso tasso è dovuto alle tecniche di indicizzazioni utilizzate, per quanto veloce possa essere il motore di ricerca, non riesce a tenere il passo del web che si arricchisce di circa un milione di documenti pubblici al giorno.

Un'alternativa ai motori di ricerca classici, sono i motori di "metaricerca". Questi ultimi, consistono in sistemi aventi la capacità di svolgere ricerche all'interno dei database di altri motori di ricerca, consentendo così di ottenere un risultato più ampio in un breve periodo di tempo (searchenginewatch.com, dogpile.com, one2seek.com, metacrawler.com, savvysearch.com, inferencefind.com, ixquick.com, profusion.com e mamma.com).

A rendere più complicate le cose sono le nuove tecnologie adottate per la pubblicazione di pagine web. Si è, infatti, passati da un "World Wide Web" basato su pagine statiche ad uno basato su database che permettono la creazione di pagine dinamiche. La stessa pagina presenta sempre contenuti diversi. Ormai quasi tutti i web masters o web administrators si fondano sulle nuove tecnologie che fanno ricorso a database per la gestione e l'aggiornamento dei siti.

I software utilizzati per individuare le pagine disponibili pubblicamente sono chiamati "crawler web". I crawler analizzano le pagine web e registrano tutti i collegamenti presenti in esse per analizzarle in un secondo momento, in modo molto simile a come faremmo noi se stessimo consultando contenuti sul Web. Passano da link a link e riportano i dati relativi alle pagini trovate nei server dei motori di ricerca.

La procedura di scansione inizia con un elenco di indirizzi web ottenuto da Sitemap fornite dai proprietari dei siti stessi. Quando i nostri crawler visitano questi siti, aggiornano gli indici e individuano i link ad altre pagine da analizzare successivamente.

I proprietari di siti possono fare molte scelte in merito alla modalità di scansione e indicizzazione dei loro siti da parte dei motori di ricerca, utilizzando gli Strumenti per i Webmaster e un file denominato "robots.txt". Con il file robots.txt si può, ad esempio, impedire ai crawler di eseguire la scansione dei propri siti.

Possiamo, quindi, definire Web visibile, quella parte di internet indicizzata ad opera dei motori di ricerca e proposta dagli stessi come risultato delle nostre ricerche. Tutto il resto è il deep web, cioè quella parte di Internet contenente informazioni, il più delle volte, accessibili liberamente e gratuitamente, ma esclusivamente inserendo il relativo indirizzo internet, altre volte, accessibili tramite l'adozione di strumenti particolari che garantiscono l'anonimato.

La struttura dell'intero World Wide Web, è spesso paragonata a quella di un iceberg, di cui solo una minima parte è visibile sopra il pelo dell'acqua. La parte sommersa, è il deep web (in cui troviamo implementate numerose sotto-reti con accessi regolamentati da diversi livelli di sicurezza).

All'interno del Deep Web, troviamo diversi tipi di siti ed attività, anche illegali, che tendono a concretizzarsi nel c.d. "Dark Web". Tra le varie attività, troviamo la vendita di armi, esplosivi e soprattutto di sostanze stupefacenti, articolate in negozi o supermercati online (i "Black Market") presso i quali è possibile acquistare ed eseguire transazioni in modo del tutto anonimo.

Ottenere l'anonimato sul web - TOR

Per poter garantire l'anonimato nel Deep Web, la via di accesso più conosciuta è TOR, un protocollo di comunicazione che, insieme ad una rete di tunnel virtuali consente di mantenere nascosta la propria identità e quindi garantisce la privacy e la sicurezza su Internet. Tor (acronimo di The Onion Router), nasce nel 1995 da un progetto della Marina Militare degli Stati Uniti per garantire la segretezza nelle comunicazioni governative (ordini e disposizioni d'impiego) al fine di evitarne le intercettazioni da parte dei servizi d'intelligence stranieri. Nel 2002 fu sviluppato dalla Electronic Frontier Foundation, ora è gestito da un'associazione senza scopo di lucro "The Tor Project".

Per capire il funzionamento di Tor, è sufficiente comprendere bene il motivo per cui è stato creato. Si cercava un protocollo che impedisse agli internet provider locali (in Italia ad esempio Fastweb, Telecom, Tiscali, Wind, Infostrada etc...) di analizzare il traffico dei propri clienti per censurare l'accesso a determinati servizi.

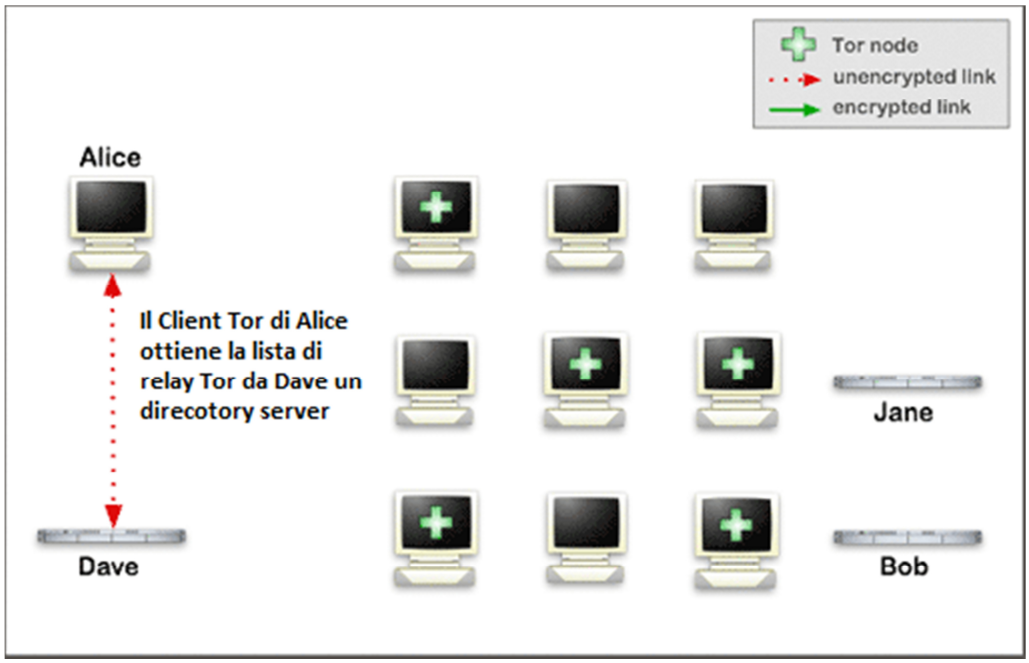
Per analisi del traffico, si intende quella tecnica che permette di analizzare tutto il traffico originato da una macchina. Un attaccante, che riesce a ispezionare i pacchetti che escono da una rete, è in grado di ottenere preziose informazioni. Intercettando una mail, ad esempio, si è in grado di individuarne il contenuto, il mittente ed il destinatario. Analogamente intercettando i pacchetti di una rete, è possibile accertarne il mittente, il destinatario e l'oggetto della comunicazione. Come vedremo un provider è in grado di decidere cosa rendere accessibile e cosa no, semplicemente analizzando i pacchetti inviati dagli utenti.

Ai fini della privacy non appare sufficiente crittografare il contenuto dei nostri messaggi, in quanto ciò non impedisce l'analisi del traffico. Infatti, il provider, tramite l'intestazione (header) in chiaro, contenuta necessariamente in ogni pacchetto inviato sulla rete perché i router possano inoltrare il

pacchetto al destinatario, è in grado di individuare immediatamente il destinatario e quindi di decidere di non inoltrare ulteriormente i pacchetti ad esso destinati, cestinandoli, pur non conoscendone il contenuto (body).

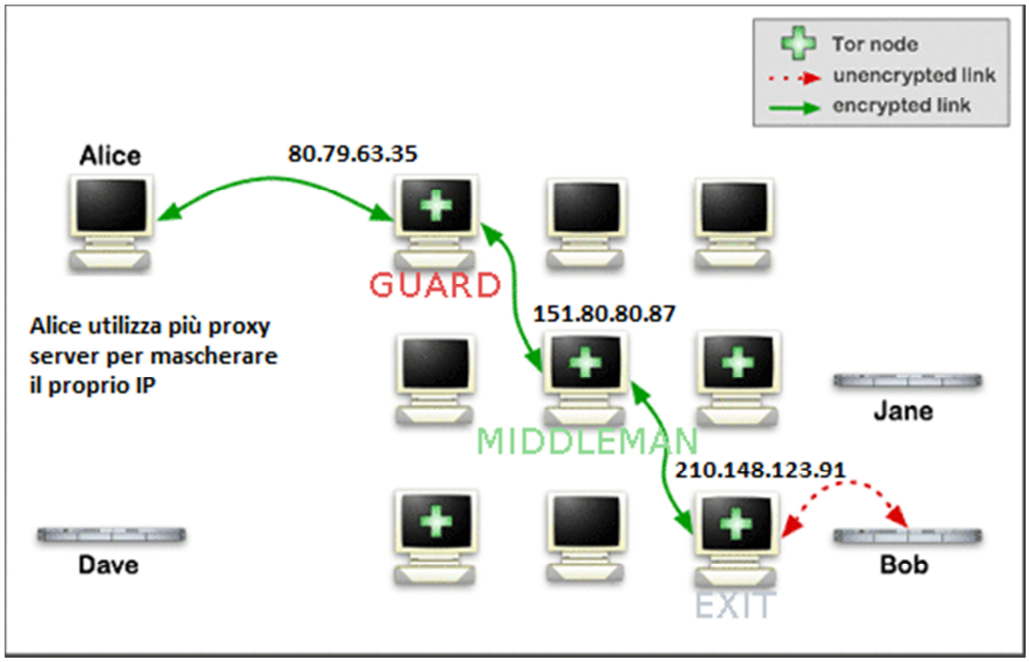
Il software The Onion Router (TOR), per poter garantire l'anonimato, implementa la tecnica dell'instradamento a strati. Ciò impedisce ai provider di Internet di analizzare e bloccare il traffico dei propri clienti. I servizi offerti dai router della rete TOR, inoltre, sono nascosti permettendo così ad un utente di pubblicare un proprio sito web su un server la cui identità e localizzazione (a partire dall'indirizzo IP) non sarà resa pubblica a nessuno. Per ottenere questo risultato, le comunicazioni non avvengono direttamente fra client e server (come, invece, accade nella rete tradizionale), ma attraverso un circuito virtuale crittografato a strati (appunto onion) tra i router della rete TOR. Gli "hidden service" (servizi nascosti) sono facilmente riconoscibili dal suffisso ".onion" (es. "niceguyfa3xkuuoq.onion").

Per comprenderne meglio il funzionamento osserviamo la figura:



Supponiamo che Alice voglia inviare a Bob e a Jane dei messaggi. Normalmente il client provvede a dividere il messaggio in pacchetti, numerarli e inserire in ognuno di essi il destinatario ed il servizio richiesto. In questo modo, però, chiunque riceve il pacchetto è perfettamente in grado di individuare subito il destinatario e quindi di decidere di non provvedere ad individuare a BOB e a Jane i pacchetti ad essi destinati, cestinandoli tutti. Per aggirare l'ostacolo Alice decide di ricorrere alla rete TOR. Installa il browser Tor diventando di fatto un nodo della rete. A questo punto il client Tor di Alice contatta un particolare server di TOR, detto "directory server", per ottenere la lista di tutti i relay della rete Tor (nella nostra figura indicati con il segno verde +).

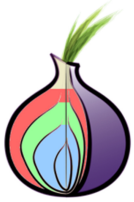
Supponiamo, ora, che Alice voglia visitare una pagina web ospitata sul server di Bob. Analizziamo la figura:



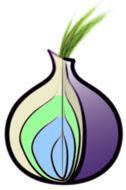
Il primo router della rete Tor che riceve i pacchetti di Alice prende il nome di “Guard Node” (nodo di guardia). Il secondo router "Middleman Node" (nodo intermediario) ed il terzo "Exit Node" (nodo di uscita). Vedendo la figura, si intuisce che i punti più vulnerabili della catena sono il nodo Guard ed il nodo Exit. Per tale motivo, tali nodi devono meritarsi questo status servendo la rete Tor per un lungo periodo come nodi “Middleman”. Tutti i relay Tor, provvedono a registrarsi sul “Directory server” inviando ad esso il proprio indirizzo ip. L'“Exit node”, inoltre, deve dichiarare quali destinatari è in grado di raggiungere. Il Client di Alice analizzando la lista di tutti i relay della rete TOR inviatagli dal “Directory server”, costruisce una catena di nodi della rete Tor che gli consentono di raggiungere il destinatario Bob.

Osserviamo che il colore verde delle frecce, indica una comunicazione crittata, impedendo ad un attaccante di ricostruire il flusso intercettando un singolo messaggio. L’ultimo collegamento è di colore rosso perché si esce dalla rete Tor e stiamo ipotizzando di aver richiesto una pagina web servita tramite protocollo http. Se Alice avesse usato il protocollo HTTPS, anche l’ultimo percorso sarebbe stato cifrato, almeno per quanto riguarda il corpo del messaggio. Sottolineiamo che, come vedremo, nell'ultimo tratto si perde l’informazione che il pacchetto è transitato per la rete Tor e anche il destinatario Bob crede di aver ricevuto la richiesta da “Exit node”. Si perde completamente l'informazione del mittente Alice. Non è possibile, nemmeno per il destinatario della richiesta, l'individuazione del mittente originario Alice.

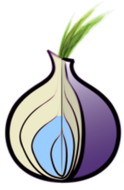
Capiamo, ora, come questo protocollo si sottrae ad eventuali analisi del traffico. I pacchetti che Alice invia al primo router della rete TOR, Guard possiamo paragonarli ad una cipolla il cui contenuto è cifrato a strati. Per un provider che analizza il traffico, quei pacchetti hanno come destinazione l’indirizzo IP 80.79.63.35 (Guard node) a cui viene chiesto un certo servizio il cui contenuto è crittografato. Non vi è traccia dell'effettivo destinatario Bob.



Guard, il primo router della rete TOR, è l'unico in grado di leggere il contenuto della buccia rossa (primo strato) disponendo, in via esclusiva, della chiave per decifrarlo. Leggendo le istruzioni contenute in questo primo strato, Guard sa di dover inoltrare tutto il resto del contenuto del pacchetto a “Middleman”:



Solo Middleman, il router immediatamente successivo della rete TOR, è in grado di leggere il contenuto della buccia verde (secondo strato) e di individuare così il destinatario immediatamente successivo. E' evidente che, già a questo livello, si è persa ogni informazione relativa ad Alice. Non è più possibile, in altri termini, risalire al proprietario del pacchetto, Alice. Il router Middleman della rete TOR, infatti, sa solo di aver ricevuto una richiesta dal router Guard, alla quale eventualmente risponderà qualora riceverà risposta, ignorando l'effettivo mittente Alice.



Infine il router “Exit” riceve da “Middleman” la parte celeste (terzo strato della cipolla). Decifrandola con la sua chiave privata, potrà leggerne il contenuto e provvedere a richiedere al server Bob una pagina web.

Bob, ricevuta la richiesta, provvederà a sua volta a rispondere a "Exit", che a sua volta lo rigira a Middleman, il quale la inoltra a Guard, che chiuderà il circuito inviando la risposta al legittimo proprietario Alice.

In definitiva, il protocollo usato nella rete TOR, tenendo celati contenuto, mittente e destinatari dei pacchetti, impedisce l'analisi del traffico. Infine, per impedire che il contenuto della risposta possa essere letta da uno dei nodi del circuito, la comunicazione viene protetta da una chiave di sessione, modificata frequentemente, apribile solo ad opera del mittente Alice.

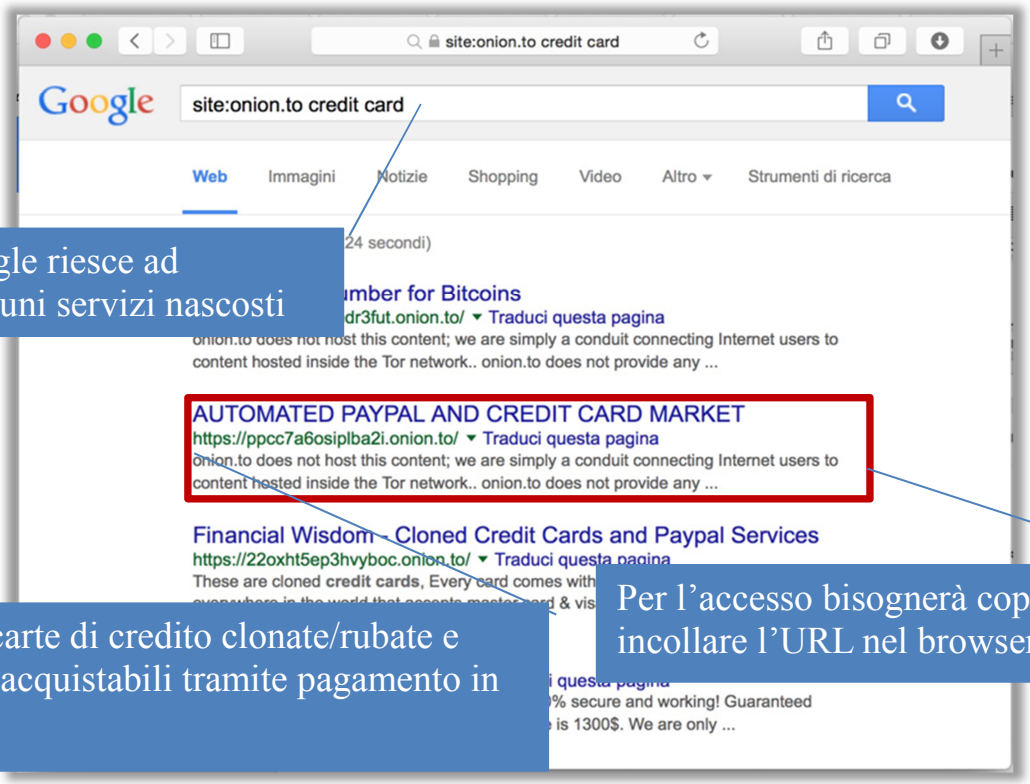
La ricerca nel Deep Web

I motori di ricerca, non implementano il protocollo Tor nei loro crawler. Pertanto, non è possibile cercare informazioni presenti sul Dark Web tramite Google e similari (o almeno non direttamente).

Per tale motivo, sono nati all'interno della rete Tor stessa siti che fungono da motori di ricerca o directory di servizi anonimi, tra cui: Hidden Wiki; TorDir; Torch; Tor Wiki; HiddenLinks.

Esistono siti che accedono al Dark Web via Tor e mostrano il contenuto degli hidden service sulla "clearnet": "ahmia.fi", "onion.city", "onion.to".

Essi funzionano anteponendo all'URL l'indirizzo del servizio Web nascosto che si vuole visualizzare. Ad esempio, digitando "torwikignoueupfm.onion.to" per visualizzare la Tor Wiki da un normale browser.



In questo modo Google riesce ad indicizzare anche alcuni servizi nascosti

Servizio di vendita di carte di credito clonate/rubate e account Paypal rubati, acquistabili tramite pagamento in BitCoin

Per l'accesso bisognerà copiare ed incollare l'URL nel browser TOR

Un vero motore di ricerca in grado di analizzare la rete TOR non esiste, proprio per le caratteristiche dello stesso che garantiscono l'anonimato e la non tracciabilità.

Per poter svolgere attività illecite e non cadere nel mirino delle istituzioni, i siti devono spostarsi continuamente, fornendo, però, strumenti necessari per rendersi reperibili dai fruitori.

Un paragone con la vita reale si ha con lo spacciatore. Questi, al fine di vendere la droga, deve rendersi localizzabile dai clienti senza però eccedere, ricorrendo, di volta in volta, a nuovi metodi che gli garantiscano nuovi acquirenti. Analogamente, un market per la vendita di sostanze illegali nel Deep Web che voglia offrire un servizio con continuità e remunerativo, dovrà pianificare un sistema per permettere agli acquirenti di individuarlo tramite motori di ricerca ordinari, di informare gli stessi circa il suo funzionamento ed i suoi prodotti tramite forum, blog o wiki, infine di trovare e installare semplici

tool che consentano loro di effettuare acquisti in sicurezza e anonimato, accedendo al market tramite un'interfaccia accattivante e professionale.

I Black Market

Spesso, così come nel web di superfice, i commercianti propongono i loro prodotti tramite veri e propri "supermercati dell'illecito", i c.d."Black Market". All'interno degli stessi, vengono offerti sia prodotti che servizi. Tra i prodotti più diffusi, troviamo numerosissime sostanze illecite: droghe, quali la marijuana, l'eroina oppure droghe sintetiche, nonché farmaci per i quali normalmente è richiesta la prescrizione medica (antidepressivi, stimolanti, antipsicotici, antidolorifici, sonniferi, ormoni, ecc.). L'acquisto protetto dall'anonimato, non è possibile mediante ricorso a carte di credito o bonifici. La moneta elettronica è il Bitcoin, acquistabile sia nei Black Market che nel web di superfice. A differenza di altre monete elettroniche, il Bitcoin non è soggetto alla regolazione da parte di un ente centrale. Le transazioni vengono gestite, sfruttando la crittografia, da un database distribuito tra i nodi della rete. Tutti i movimenti fatti con i Bitcoin sono collegati ad un portafoglio e resi pubblici. Ogni portafoglio dispone di un proprio indirizzo, utilizzato per inviare o ricevere Bitcoin. Ne consegue che l'anonimato è garantito finché quel portafoglio non sia riconducibile a qualcuno. Nel contempo, l'uso di Bitcoin è forse lo strumento che garantisce il più alto livello di trasparenza. Tutte le transazioni sono registrate nel database distribuito e sono tracciabili da chiunque conosca l'indirizzo del portafoglio. Pertanto, se si è a conoscenza del fatto che il portafoglio sia collegato ad un determinato Ente, pubblico o privato, è possibile accertare tutti i movimenti dell'Ente in oggetto. La natura decentrata del database, non consente a nessuna autorità, governativa o meno, di sequestrare Bitcoin ai legittimi possessori o di svalutarla creando nuova moneta. Per questi motivi il Bitcoin è la moneta più utilizzata nelle transazioni illecite online.

Tra i più famosi Black Market ricordiamo il famigerato "Silk Road", dove è possibile acquistare droga, carte clonate, skimmer e attrezzi da truffatore e documenti falsi, che arriveranno a casa in un pacco anonimo, in modalità priority stealth.

L'FBI ha calcolato che il solo "Silk Road" abbia generato tra febbraio 2011 e luglio 2013 transazioni finanziarie per 1,2 miliardi di dollari, ricavandone commissioni per 80 milioni di dollari.

"Silk Road", oggi Silk Road 2.0, è un sito che segue regole ben precise. Sono bandite armi, pedopornografia e altro. Il gestore guadagna sulle singole transazioni il 10% e offre la possibilità di aprire delle controversie per gestire le dispute fra venditore e acquirente, analogamente al famoso marketplace eBay.

L'FBI ha, però, rivelato che "Silk Road" è stato usato anche per commissionare omicidi e altri crimini, apprezzato anche dai piccoli trafficanti. Di recente sono stati venduti biglietti falsi dell'Atac e dell'ATM a un terzo del prezzo stabilito dalle due aziende municipalizzate di trasporti.

Il prodotto più richiesto su Silk Road è una pasticca di ecstasy da 9 dollari chiamata "Superman Xtc" e venduta dall'Olanda. Al secondo posto abbiamo una carta prepagata da 100 sterline, utile per convertire i bitcoin in soldi veri e utilizzatabile nei pub, ristoranti e musei della Gran Bretagna. Al terzo posto, infine, si classifica un pacchetto da 200 milligrammi di DMT (dimetiltriptamina), un allucinogeno che negli ultimi anni è diventato di moda negli Stati Uniti.

Un esempio di mercato del sesso, ospitato dal Dark Web, è PlayPen. Uno dei maggiori siti pedopornografici, con 150mila membri che pubblicavano e accedevano a immagini e video di abusi sessuali su minori.

Contrasto investigativo

Cerchiamo ora di capire come sia possibile svolgere indagini sul DeepWeb, nonostante il protocollo utilizzato da TOR ne garantisca l'anonimato.

Gli strumenti adottati per indagare nella DARK NET sono fondamentalmente di tre tipi:

- strumenti basati sui software, come malware e NIT (Network Investigative Technique), che sfruttano gli errori di configurazione dei siti nascosti nelle DARK NET;
- infiltrati;
- errori commessi dai criminali.

Abbiamo visto come i punti vulnerabili del protocollo TOR siano i nodi di accesso. Prima di accedere alla rete TOR i pacchetti presentano ancora l'header in chiaro, e quindi anche l'IP se lo stesso non viene opportunamente nascosto tramite l'utilizzo di server proxy (server che funge da intermediario per le richieste da parte dei client alla ricerca di risorse su altri server).

L'FBI ha collocato propri server nei nodi di ingresso di TOR, catturando una enorme mole di dati la cui analisi ha permesso di rintracciare gli utilizzatori e i siti visitati. Utilizzando una tecnica investigativa di rete nota come NIT, Network Investigative Technique, l'FBI, iniettando un malware (codice malevolo) nei PC degli utenti, ha ottenuto i veri indirizzo IP degli utilizzatori.

L'uso dei malware è ben documentato nelle indagini su PlayPen. L'innesto del software malevolo è stato possibile, in questo caso, grazie ad errori di configurazione commessi da chi gestiva il sito.

"PlayPen" nasce nell'agosto del 2014. A dicembre dello stesso anno, una Forza di polizia non americana segnalò all'FBI non solo l'esistenza di questo sito nascosto, ma anche il suo vero indirizzo IP, visibile a causa di un errore di configurazione, associato ad un server localizzato negli Stati Uniti. Individuato il server, in North Carolina, l'FBI decise di non oscurarlo, bensì di gestirlo direttamente per analizzarne gli accessi. Ciò consentì l'individuazione dei due amministratori dello stesso, nonché degli oltre 1300 utenti che usufruivano del servizio.

L'indagine su "Silk Road", invece, è stata resa possibile, soprattutto impiegando agenti sotto copertura, oltre che sfruttando gli errori di configurazione dei siti.

Le vulnerabilità dovute ad errori di configurazione sono molto diffuse. Circa il 6% dei siti, presenta errori di configurazione che permettono di risalire al loro reale indirizzo IP. Inoltre, il 25% dei siti contiene errori di configurazione che consentono di ottenere utili informazioni (ad esempio la riconducibilità di siti diversi alla stessa persona).