ALMA MATER STUDIORUM

UNIVERSITÀ DI BOLOGNA

SCUOLA DI INGEGNERIA E ARCHITETTURA

Corso di Laurea in Ingegneria Elettronica, Informatica e Telecomunicazioni

DARKNET E DEEP WEB: IL LATO OSCURO DEL WEB PER LA PRIVACY E LA PROTEZIONE DEI DATI

Elaborata nel corso di: Sistemi Distribuiti

Tesi di Laurea di:

PATRYK WOJTOWICZ

Relatore:

Prof. ANDREA OMICINI

Co-relatori:

STEFANO MARIANI

ANNO ACCADEMICO 2013 2014

SESSIONE III

Indice

1. Introduzione

Dopo le rivelazioni di Snowden è nata una nuova era che ha mutato la nostra percezione dell'anonimato in rete. La tesi si pone come obbiettivo quello di esplorare questo nuovo mondo.
Partendo da un'analisi dei nostri dati che viaggiano attorno a noi si vuole scoprire se la privacy è garantita. Quando questa non è garantita, la sicurezza è fondamentale e nasce l'esigenza di proteggersi. Nascondersi nelle profondità del Deep Web è una soluzione a questo problema, ma non è l'unica. Verranno esplorate altre darknet e diversi tool che insieme compongono lo scudo e l'armatura indispensabili per questa guerra per la privacy e la protezione dei dati.

2. I Dati

2.1 Dati informatici

In informatica un dato[1] è una descrizione elementare di una cosa, di una transazione, di un avvenimento o di altro. Un dato da solo non ha significato, ma una volta interpretato o elaborato può portare alla conoscenza di un'informazione.
Ogni tipo di dato dipende dal codice e dai formati impiegati e può avere diverse forme:

- **Testo:** una sequenza di lettere, simboli o numeri;
- **Immagine:** una rappresentazione visiva e non solida della realtà;
- **Video:** una sequenza di immagini in movimento o ferme;
- **Audio:** l'informazione elettronica che rappresenta il suono.

I dati possono essere conservati su diversi supporti cartacei, magnetici (floppy disk), ottici (CD e DVD) ed altri. Successivamente possono essere trasmessi tra più utenti attraverso una rete di telecomunicazione.

2.2 Organizzazione dei dati

Per comprendere come sono organizzati i dati è opportuno inserire tale concetto nel contesto dei sistemi informativi. Un sistema informativo[2] è un sistema di supporto ai processi informativi di un'azienda. La porzione di sistema informativo che viene gestita in modo automatico mediante tecnologie informatiche e di automazione prende il nome di sistema informatico. Essi hanno il compito di raccogliere, organizzare e conservare le informazioni garantendo che queste vengano

conservate in modo permanente su dispositivi per la loro memorizzazione, aggiornate in maniera rapida e rese accessibili alle richieste degli utenti. L'importanza nella gestione di queste informazioni ha portato allo sviluppo e realizzazione di applicazioni software come i sistema di archiviazione che consentono la ricerca e la memorizzazione di informazioni invarianti nel tempo. Più utilizzati invece sono i Database (Base di dati) che consentono la ricerca e la memorizzazione di informazioni senza il vincolo di staticità del sistemi di archiviazione. Un database è un archivio dati, o un insieme di archivi ben strutturati, in cui le informazioni in esso contenute sono strutturate e collegate tra loro secondo un particolare modello logico. Consentono la gestione e organizzazione efficiente dei dati e l'interfacciamento con le richieste dell'utente attraverso i cosiddetti query language e grazie a particolari applicazioni software dedicate: Database Management System (DBMS). Le basi di dati possono avere varie strutture:

- **Gerarchica:** rappresentabile tramite un albero;
- **Reticolare:** rappresentabile tramite un grafo;
- **Relazionale:** rappresentabile mediante tabelle e relazioni tra esse;
- **Ad oggetti:** estensione alle basi di dati del paradigma "Object Oriented".

2.3 Big Data e Data Mining

Internet[3] non dorme mai e la quantità di dati che viaggiano in rete è enorme. La popolazione su Internet negli ultimi anni è cresciuta del 14% e ora conta 2,4 miliardi di persone. Il 90% dei dati oggi esistenti sono stati generati negli ultimi 2 anni e il ritmo con cui i dati sono prodotti è talmente alto che ogni due giorni viene creato un volume di dati pari alla quantità di informazioni generate dall'umanità intera fino al 2003.
Ogni 60 secondi vengono inviate 350 mila fotografie su Whatsapp (di cui metà di queste sono foto degli es di fisica miei o della Silvia) e 200 mila ne vengono pubblicate su Instagram. Su Youtube, ogni minuto, vengono caricate 72 ore di nuovi video e su Vine gli utenti condividono più di 8 mila video. Su Facebook gli utenti condividono quasi 2 milioni e mezzo di contenuti mentre su Twitter possiamo contare 280 mila tweet. Sempre in un minuto vengono inviate 205 milioni di email e vengono fatte 4 milioni di ricerce su Google. Questi numeri crescono in modo vertiginoso ed è facile intuire l'enorme potenziale che questi dati possono

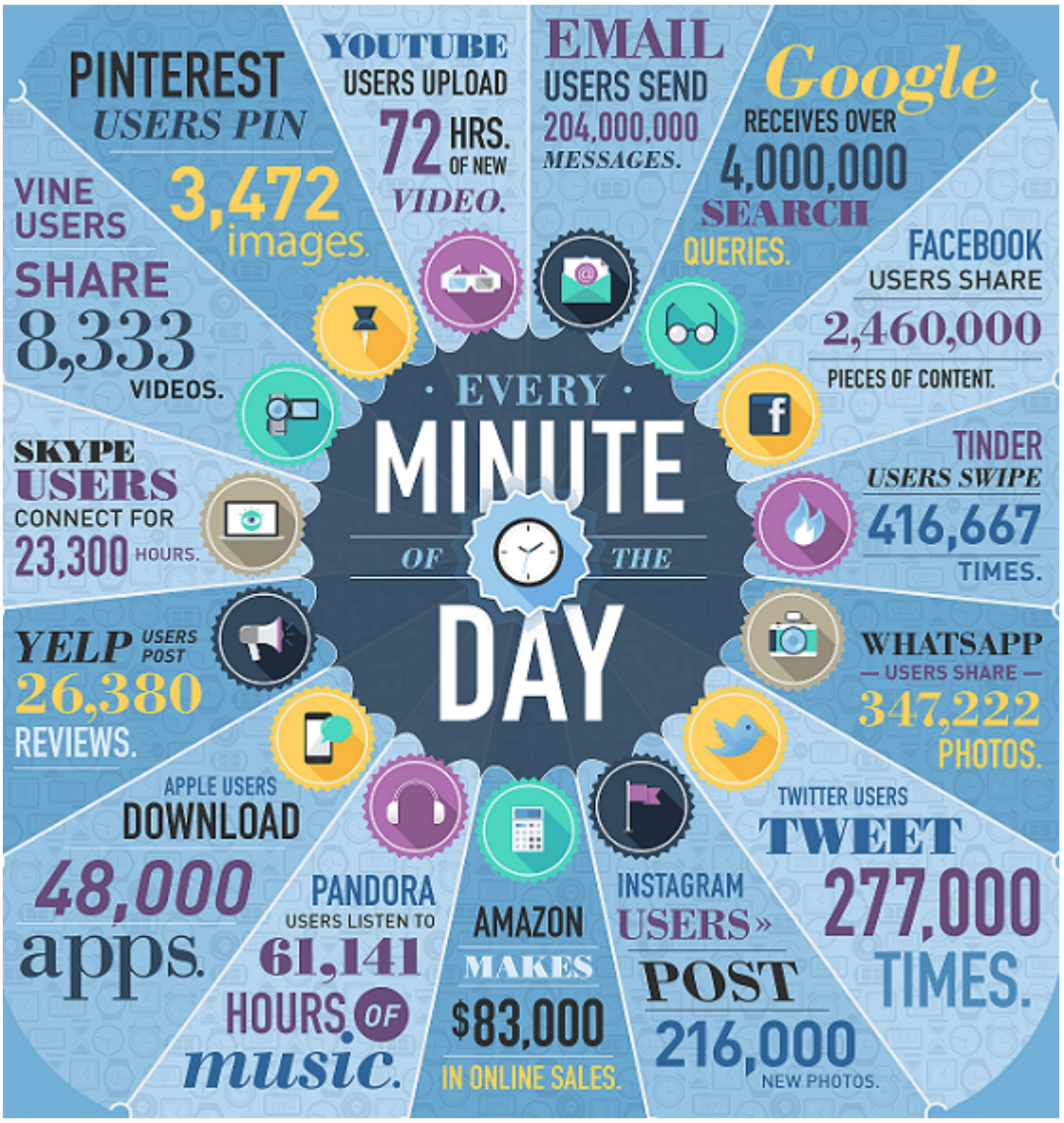

FIGURE 2.1: Data never sleeps 2013

avere nella nostra vita.

La collezione di questi dati, strutturati in forma relazionale, è chiamata Big Data[4]. Per definire il loro valore non basta semplicemente il volume, ma una serie di caratteristiche:

- **Volume:** rappresenta la dimensione. Si stima che entro il 2020 verranno creati 35 zettabyte di dati (35 mila miliardi di gigabyte);
- **Velocità:** rapidità con la quale vengono estratti i dati per essere utilizzati prima di diventare obsoleti;

- **Varietà:** diverse tipologie di dati (immagini, video, musica, ricerche web, transazioni finanziarie, pagamenti con carta di credito, etc.). Questa eterogeneità permette di utilizzare i Big Data in vari settori;
- **Variabilità:** si riferisce al fatto che ogni dato ha un significato a seconda del contesto. Trascurare questa caratteristica porterebbe inevitabilmente all'inconsistenza dei dati;
- **Complessità:** la complessità della gestione dei dati è direttamente proporzionale all'aumento della dimensione degli stessi.

Il vero problema però non è raccogliere queste grandi quantità di dati, ma riuscire ad usarle per ottenerne una sorta di vantaggio competitivo, come riduzione dei costi e dei tempi o sviluppo di nuovi prodotti e offerte ottimizzate.
Chiaramente, visti i volumi e i dati non strutturati, è impossibile utilizzare i tradizionali RDBMS (Relational database management system) e quindi bisogna cercare soluzioni basate sul NoSQL (Not only SQL).
Con il termine NoSQL viene indicato quell'insieme di tecnologie legate ai database e sviluppate in risposta a questo crescente volume di dati memorizzati sulla rete, alle modalità e alla frequenza con le quali si accede a questi dati, alla performance e potenza di calcolo necessaria per gestirli. I database NoSQL si possono dividere in differenti tipologie:

- **Chiave-Valore:** ogni singolo elemento viene salvato come una chiave assieme al suo valore.
- **Orientato ai documenti:** ogni chiave è accoppiata con una struttura dati complessa detta documento che possiede determinate caratteristiche. La chiave per indicizzare i documenti è univoca e spesso è contenuta in un indice del database per consentire un recupero rapido. Una caratteristica particolare è la presenza di API che permettono il recupero dei documenti in base al loro contenuto aumentandone le performance.
- **A grafo:** utilizza nodi e archi per rappresentare e archiviare l'informazione. Sono spesso più veloci di quelli relazionali nell'associazione di insiemi di dati e mappano in maniera più diretta le strutture di applicazioni orientate agli oggetti.

Le base di dati NoSQL[5] rispetto ad un database SQL possono essere adattati alla grandezza della base dati da trattare con maggiore facilità, garantiscono prestazioni migliori e il loro modello organizzativo può risolvere problemi che il

modello relazionale non è progettato per affrontare. Gli schemi dinamici sono un punto di forza dei database NoSQL. In una base dati SQL, gli schemi di organizzazione interna devono essere definiti prima che sia possibile iniziare l'utilizzo del database.
Nel caso si volesse aggiungere una nuova classe di dati il processo potrebbe risultare lungo e macchinoso. I database non relazionali, invece, sono realizzati per permettere l'inserimento di dati senza uno schema predefinito. Ciò rende più semplice effettuare cambiamenti (anche sostanziali) quasi in tempo reale, senza preoccuparsi di eventuali interruzioni del servizio. Tra le più conosciute e utilizzate applicazioni di database NoSQL possiamo citare MongoDB, al 5° posto della classifica dei database più popolari del mondo e al 1° posto dei database non relazionali secondo db-engines. MongoDB è un software open source per la gestione di database non relazionali orientati ai documenti. Per la sua flessibilità, scalabilità e potenza viene utilizzato da colossi del calibro di eBay, The New York Times e altri ancora.
Il data mining è il termine che indica l'estrazione di una conoscenza tramite l'applicazione di algoritmi che individuano associazioni o patterns invisibili tra le informazioni. Il data mining potrebbe essere confuso con la statistica, ma c'è una differenza sostanziale: la statistica permette di elaborare informazioni generali riguardo ad una popolazione mentre il data mining viene utilizzato per cercare relazioni tra più variabili relative al singolo individuo. Oltre all'enorme espansione dei dati, il calo dei prezzi per i supporti di memorizzazione e l'evoluzione di nuovi metodi e tecniche di analisi ha portato ad un concreto sviluppo di questo ambito. Tra le tecniche maggiormente utilizzate per il data mining abbiamo:

- **Clustering;**
- **Reti neurali;**
- **Alberi di decisione;**
- **Analisi delle associazioni.**

Una delle evoluzioni più recenti del data mining è la data visualisation che si occupa non solamente di rendere graficamente intelligibile un testo, ma entra in relazione più diretta con la strutturazione dei database e l'esportazione di grafici dai dati. Un'altra nuova frontiera è il social data mining, ovvero l'analisi di informazioni generate dai social network.
È inevitabile che i nostri comportamenti, da come navighiamo su un sito web a quali prodotti acquistiamo, siano sempre più oggetto di analisi e studi. Tutte queste informazioni che generiamo sono preziose per le aziende, le quali possono

modificare i propri servizi ed adeguarli alle richieste e ai comportamenti dei consumatori, ma fino a che punto è legittimo lo sfruttamento dei dati personali e fino a che punto l'analisi si può spingere? Questa domanda, che sempre più frequentemente le persone si pongono, sarà esaminata in maniera più approfondita nei seguenti capitoli.

2.4 Metadati

Un metadato[6] è un'informazione che descrive un insieme di dati con lo scopo di migliorarne la visibilità e facilitarne l'accesso. Le funzioni principali di un sistema di metadati sono:

- **Ricerca:** consiste nell'individuare la presenza di un documento;
- **Localizzazione:** ovvero rintracciare una particolare occorrenza del documento
- **Selezione:** di una serie di documenti;
- **Interoperabilità semantica:** consiste nel permettere la ricerca in ambiti disciplinari diversi grazie a una serie di equivalenze fra descrittori;
- **Gestione risorse:** gestire le raccolte di documenti grazie all'intermediazione;
- **Disponibilità:** ovvero ottenere informazioni sull'effettiva disponibilità del documento.

Due ricerche a opera di Marco Furini e Valentina Tamanini[7] mostrano che, nonostante i mille accorgimenti per proteggere la privacy, quando pubblichiamo una foto online, diciamo molto più di quello che mostriamo. E il motivo è semplice: la maggior parte dei dettagli si nascondono nei metadati. Il lavoro dei due ricercatori mostra come l'analisi di dati e metadati pubblici presenti sui social permetta di sapere, con estrema facilità, dove sono fisicamente gli utenti, quali sono i loro spostamenti e i loro luoghi abituali.
La conoscenza di queste informazioni, oltre a violare la propria sicurezza, può portare a situazioni poco piacevoli. Come il caso eclatante di Alexandr Sotkin, soldato dell'esercito russo che tramite le foto del suo profilo Instagram potrebbe aver svelato la presenza delle forze inviate da Mosca in territorio Ucraino con potenziali gravi ripercussioni in situazioni delicate come la guerra civile Ucraina.

Un altro esempio è Immersion, un software che mostra la rete di interazioni sul Web costruita a partire dai metadati delle conversazioni elettroniche. Ideato da César Hidalgo, professore al Mit Media Lab, Immersion ci mostra l'importanza dei metadati e di come questi possano rivelare informazioni e interazioni tra persone. Tramite un diagramma fatto di cerchi e linee, evidenzia le cento persone con cui si è comunicato di più, quanto strettamente siano collegate all'utente e quanto siano interconnesse l'una con l'altra. Per placare eventuali timori relativi all'abuso di privacy, al termine dell'esperimento è possibile cancellare tutti i dati raccolti.
Immersion oltre ad è uno strumento di riflessione su se stessi ci permette di capire le potenzialità dei metadati e la necessità di una sensibilizzazione per tutelare la propria privacy.

2.5 Consapevolezza dei dati

L'Internet of Things (Internet delle cose), possibile evoluzione dell'uso della rete, avanza in maniera rapida insieme ad una dimensione di tecnologie wearable. Eric Schmidt, presidente del consiglio di amministrazione di Google, afferma[8] che Internet è destinato a sparire. Avremo cosi tanti device, sensori, oggetti collegati a noi che non riusciremo a vedere l'interazione che abbiamo con essi. Il discorso è fortemente legato al progetto Phisical Web[9] , uno standard aperto, del gruppo di Mountain View che mira a rendere possibile l'interazione con qualsiasi apparecchiatura semplicemente avvicinandosi ad essa, senza nessun bisogno di scaricare prima un'applicazione.
In futuro non percepiremo più l'infrastruttura che c'è dietro alle connessioni diventando parte stessa di Internet. Ed è proprio questa invisibilità cognitiva e culturale l'argomento importante con cui dovremo confrontarci, ragionando sull'Internet of Things e il wearable.
Ci sono almeno tre dimensioni lungo cui si sviluppa la wearable technology:

- **Dimensione reattiva:** estensioni del nostro corpo per raccogliere dati e relazionarsi con l'ambiente;
- **Dimensione immersiva:** collegata alle tecnologie di realtà virtuale e realtà aumentata;
- **Dimensione predittiva:** svariate tipologie di strumenti che collezionano dati da analizzare per costruire modelli di previsione.

FIGURE 2.2: Google Phisical Web - Tutto è connesso

Quest'ultima dimensione rappresenta la frontiera della nuova generazione di wearable. Ad esempio Samsung sta lavorando su un prototipo di dispositivo indossabile in ambito medico chiamato EDSAP che monitora i livelli di ansia, stress e qualità del sonno. Leggendo le onde cerebrali riesce anche a rilevare in anticipo i principali segnali di un ictus, avvertendo così per tempo la persona che indossa il device. É immediato capire l'importanza che ha la sicurezza e la privacy in verso la quantità di dati personali che condividiamo, spesso inconsapevolmente.
Secondo quanto affermano gli ultimi dati pubblicati da IDC, nel 2014 sarebbero stati venduti circa 19 milioni di dispositivi wearable, oltre 600 mila solo in Italia. Un numero che dovrebbe crescere in maniera rapida nel 2015 e rispetto a quanto avverrà entro il 2018 si stima un acquisto di circa 112 milioni di pezzi indossabili connessi.
Nasce quindi la necessità di imparare a sviluppare e diffondere una maggiore consapevolezza sui nostri dati personali e sugli effetti che i dispositivi indossabili avranno sulle nostre vite.

3. Privacy

I gruppi in difesa[10] delle libertà civili si sono mossi da poco per chiedere alle Nazioni Unite la creazione di un organo di sorveglianza contro gli abusi di potere nel monitoraggio perpetrato dal governo statunitense. La privacy sta diventando sempre più un miraggio a seguito delle rivelazioni di Edward Snowden che nell'estate del 2013 ha disegnato un quadro decisamente terrificante di come le agenzie federali spiano e gestiscono i dati digitali di cittadini americani e stranieri. Per questo la Electronic Frontier Foundation, lo Human Rights Watch e Liberty hanno firmato, assieme ad un'altra dozzina di associazioni, una lettera per chiedere la creazione di un relatore che tenga alta l'attenzione su eventuali falle informatiche e sulle intrusioni ai database di aziende ed enti che contengono i dati dei navigatori.

3.1 Dati sensibili e personali

I dati sensibili[11] , secondo il diritto italiano, sono dati personali la cui raccolta e trattamento sono soggetti sia al consenso dell'interessato sia all'autorizzazione preventiva del Garante per la protezione dei dati personali.
Secondo il Codice sulla protezione dei dati personali (D. Lgs. n. 196/2003, art. 4), sono considerati dati sensibili i dati personali idonei a rivelare:

- l'origine razziale ed etnica;
- le convinzioni religiose, filosofiche o di altro genere;
- le opinioni politiche;
- l'adesione a partiti, sindacati, associazioni od organizzazioni a carattere religioso, filosofico, politico o sindacale;
- lo stato di salute e la vita sessuale.

I dati personali invece sono i dati normali che usiamo tutti i giorni, quindi qualunque informazione relativa a persona fisica, persona giuridica, ente od associazione, identificati o identificabili, anche indirettamente, mediante riferimento a qualsiasi altra informazione, ivi compreso un numero di identificazione personale(B. Lgs. n. 196/2003, art. 4).
I dati devono essere trattati in modo lecito, ovvero facendo riferimento a tutto l'ordinamento giuridico e non solo alle norme del Codice, e secondo correttezza per il rispetto di tutte le regole. Possono essere raccolti e registrati se si possiede il consenso dell'interessato, manifestato in forma scritta, e l'autorizzazione del Garante. Quando è prevista tale autorizzazione, il requisito è soddisfatto se sono state emanate dal Garante autorizzazioni relative a determinate categorie di titolari o di trattamenti, pubblicate nella Gazzetta Ufficiale della Repubblica italiana. In alcuni casi è necessaria l'autorizzazione del Garante, ma non occorre invece il consenso:

- trattamento di dati, degli aderenti, effettuato da enti no profit a carattere politico, filosofico, religioso e sindacale per scopi statutari legittimi;
- trattamento necessario per salvaguardare la vita o l'incolumità fisica di un terzo;
- trattamento necessario per investigazioni difensive o per tutelare un diritto in giudizio;
- trattamento necessario per adempimenti di legge, regolamento o norme comunitarie per la gestione di un rapporto di lavoro.

3.2 Concetto di privacy nel mondo

La privacy[12], termine inglese equivalente a riservatezza o privatezza, è appunto il diritto alla riservatezza della propria vita privata: "il diritto di essere lasciati in pace", secondo la formulazione del giurista statunitense Louis Brandeis che fu probabilmente il primo al mondo a formulare una legge sulla riservatezza, insieme a Samuel Warren.
Di crescente rilievo è il tema della sicurezza informatica che riguarda sia i privati cittadini, sia le imprese. Coinvolge tutti gli aspetti che riguardano la protezione dei dati sensibili archiviati digitalmente ma in particolare è noto al grande pubblico con riferimento all'utilizzo di Internet.
In effetti la rete può costituire un luogo pericoloso per la nostra privacy anche

perché il mezzo stesso non è stato concepito per scambiare o gestire dati sensibili. La miglior difesa per la nostra privacy, in questa situazione di precarietà, consiste nell'utilizzare il buon senso e nell'adottare semplici accorgimenti.
L'evoluzione tecnologica, che nella trasmissione e rielaborazione dei dati compie continui ed incessanti progressi, richiede una rinnovata attenzione nei confronti della privacy che, nata come "right to be let alone", è divenuta ai giorni nostri il diritto di controllare l'uso che altri facciano delle informazioni che mi riguardano. Non ha senso parlare di protezione dei dati personali se non si considera il valore giuridico fondamentale da tutelare, ovvero la privacy. Allo stesso modo la privacy non può essere intesa in maniera corretta ed adeguata, in tutte le sue necessarie implicazioni con riferimento alla società contemporanea, se non ci si rende conto della necessità di proteggere i dati personali.
Dopo la giornata europea della privacy (avvenuta il 28 Gennaio 2015) il Guardian[13] ha pubblicato un'anteprima del nuovo piano antiterrorismo che verrà presentato a breve dalla Commissione UE. Nonostante il più importante organo europeo a difesa dei diritti umani abbia stilato un atto d'accusa[14] durissimo alla sorveglianza di massa, definendola come una minaccia per i diritti umani fondamentali, al su interno ci sarebbe una norma che consente la registrazione generalizzata e l'immagazzinamento, per cinque anni, di ben 42 tipologie di dati personali dei passeggeri che volino da e per il nostro continente. Oltre ai dettagli su identità e passaporto, figurano informazioni su tutte le forme di pagamento, l'indirizzo di casa, di posta elettronica e perfino le preferenze alimentari dei viaggiatori.
Un'altra notizia che ci getta nel paradosso viene dalla Francia, dove Francois Hollande sta pianificando[15] una nuova legge secondo la quale colossi web come Facebook e Google diventerebbero complici di pubblicazione di materiale estremista qualora ne dovessero ospitare. Un sistema simile sarebbe realizzabile utilizzando filtri preventivi per vietare la pubblicazione di certe parole, la composizione di una black list statale per quelle parole proibite ed in più una corsia preferenziale per le segnalazioni delle autorità qualora si tratti di rimuovere contenuti sgraditi. Un vero e proprio stato di sorveglianza digitale.
Ma davvero la privacy si può sacrificare nel nome della sicurezza nazionale? E anche ammesso sia accettabile questo sacrificio, funziona? Il più importante organo europeo per la difesa dei diritti umani[16] risponde no. I resoconti dettagliati, dai quali si conclude che la sorveglianza di massa sia inutile e dannosa, sono oramai svariati.
Alla luce di un nuovo studio analitico pubblicato dal Think Tank New America Foundation mostra come i tradizionali metodi investigativi abbiano fornito il punto di partenza per le indagini sui casi principali di terrorismo, mentre il contributo dei programmi di sorveglianza di massa dell'NSA in quei casi è stato minimale.

Prima che il sempre crescente complesso dell'industria della sorveglianza finisca

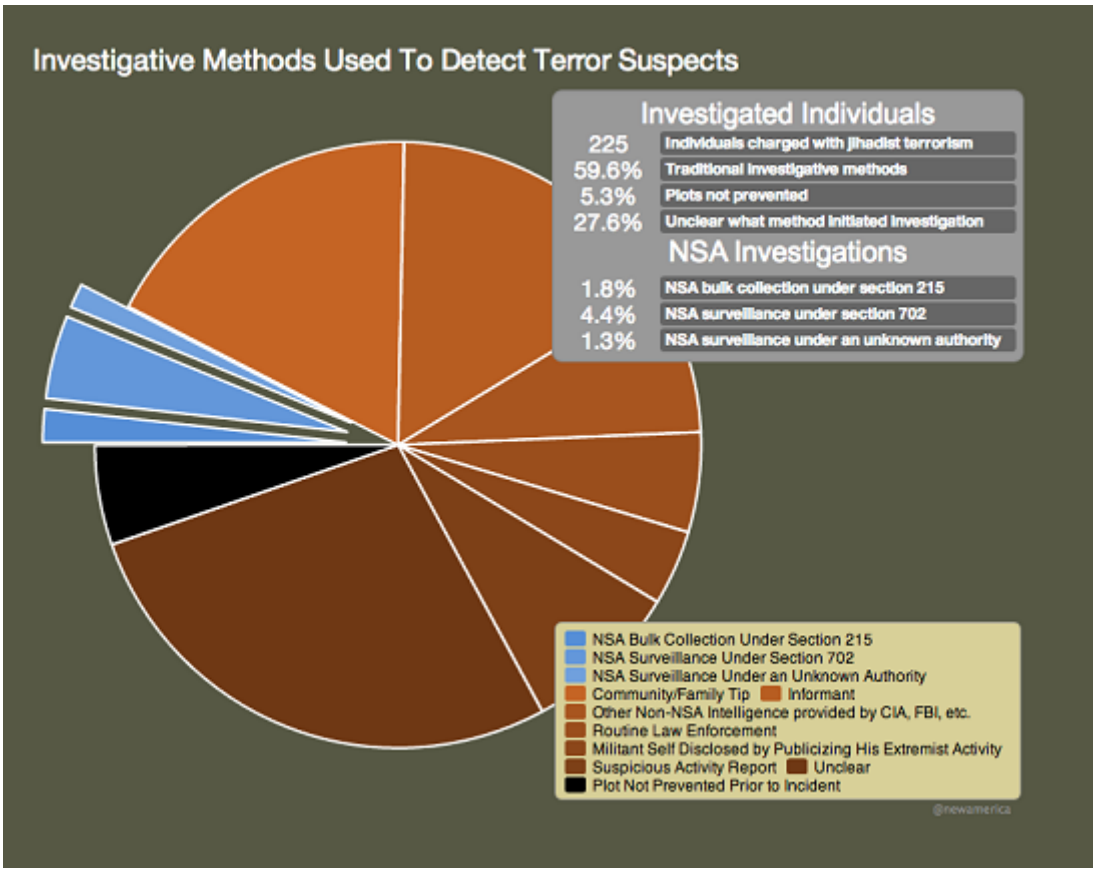

FIGURE 3.1: Metodi investigativi usati per rilevare i principali sospetti terroristi

completamente fuori controllo, dobbiamo agire in modo che la sorveglianza sia sottomessa allo stato di diritto. Attualmente si deduce che cosi non va bene è potrà solo peggiorare se dovessero passare le misure ipotizzate durante la giornata europea della privacy.

3.3 Garante per la protezione dei dati personali

Il Garante per la protezione dei dati personali è un'autorità amministrativa indipendente istituita dalla legge n. 675 del 31 dicembre 1996, per assicurare la tutela dei diritti e delle libertà fondamentali ed il rispetto della dignità nel trattamento dei dati personali. Con la redazione del Codice in materia di protezione dei dati personali, approvato con il decreto legislativo n. 196 del 30 giugno 2003, la legge n. 675/1996 è stata abrogata. Attualmente l'autorità è presieduta da Antonello Soro.
Tra i diversi compiti del Garante (art. 154 D.Lgs. 196/2003) rientrano quelli di:

- Controllare che i trattamenti siano effettuati nel rispetto delle norme di legge;
- Ricevere ed esaminare i reclami e le segnalazioni e provvedere ai ricorsi presentati dagli interessati;
- Vietare anche d'ufficio i trattamenti illeciti o non corretti ed eventualmente disporne il blocco;
- Promuovere la sottoscrizione di codici di deontologia e buona condotta di determinati settori;
- Segnalare al Governo ed al Parlamento l'opportunità di provvedimenti normativi richiesti dall'evoluzione del settore;
- Esprimere pareri nei casi previsti;
- Curare la conoscenza tra il pubblico della disciplina rilevante in materia di trattamento dei dati personali ed delle relative finalità ed in materia di misure di sicurezza dei dati;
- Denunciare i fatti configurabili come reati perseguibili d'ufficio conosciuti nell'esercizio delle sue funzioni
- Tenere il registro dei trattamenti;
- Predisporre una relazione annuale sull'attività svolta da presentare al Governo ed al Parlamento;
- Essere consultato da Governo o Ministri quando questi predispongono norme che incidono sulla materia;
- Cooperare con le altre autorità amministrative indipendenti;
- Organizzare il proprio ufficio ed il proprio organico ed il loro trattamento giuridico, economico ed amministrativo.

4. Sicurezza

4.1 Una nuova era

La preoccupazione per i propri dati online è in crescita. Secondo i risultati di un'indagine di Symantec[17] il 57% degli europei ritiene che i propri dati personali non siano adeguatamente protetti. Inoltre il 74% ritiene sia ingiusto che le aziende facciano soldi grazie allo sfruttamento dei dati personali dei propri clienti. The State of Privacy 2015 mette sotto la lente d'ingrandimento le percezioni, in merito di sicurezza, dei dati online di 7mila europei in 7 Paesi, Italia compresa.
Dal report emerge che i cittadini attribuiscono un grande valore, anche economico, ai propri dati personali e non si fidano degli enti che ne dovrebbero garantire la tutela. Mediamente circa 7 europei su 10 ritengono le istituzioni mediche più affidabili in termini di tutela dei dati personali rispetto ai social media, rivenditori e aziende tech che occupano le ultime posizioni della classifica.

Secondo i dati di Symantec non si riscontra un coerente cambiamento nei comportamenti digitali a fronte di tali preoccupazioni per i propri dati. Infatti solo il 25% dei campioni si preoccupa di leggere le informative sull'uso dei dati personali prima di concederli. Mentre solo il 14% degli europei dice di essere contento del fatto che un'azienda possa condividere i suoi dati personali con terzi.
Sulla protezione dei dati personali, gli europei sarebbero piuttosto confusi dato che, circa il 66% dei partecipanti al sondaggio vorrebbe proteggere meglio i propri dati, ma non saprebbe come farlo. Nel frattempo, però, l'88% degli europei dichiara di scegliere se rivolgersi ad un'azienda sulla base delle protezioni dei dati che garantisce, un dato molto vicino a quello della qualità percepita del prodotto in questione e al possibile ottenimento di un buon servizio.
I dati raccolti da Symantec sono complessivamente coerenti con i dati raccolti dal Centre for International Governance Innovation (Cigi) e Ipsos per il Global Survey on Internet Security and Trust[18]. Il sondaggio mostra il cambiamento di percezione della privacy in 24 Paesi in seguito al caso Datagate e alle rivelazioni di Edward Snowden. Per svolgere lo studio, i due istituti di ricerca hanno raccolto

FIGURE 4.1: Infografica dei dati italiani tratti dal report The State of Privacy

i dati relativi a oltre 23 mila persone tra il 7 Ottobre 2014 e il 12 Novembre 2014. Dal sondaggio risulta che il 64% degli utenti sono preoccupati per la propria privacy online rispetto all'anno precedente. Dal punto di vista della privacy, i più preoccupati per la loro riservatezza online, rispetto a un anno fa, sono brasiliani e indiani. Gli italiani si posizionano a metà classifica, mentre a chiudere la classifica è la Svezia. Gli americani, i più colpiti dalla sorveglianza della Nsa, hanno fatto registrare una percentuale più alta rispetto alla media europea, ma più bassa di quella che ci si potrebbe aspettare, se paragonata ai vertici del ranking.
L'effetto Snowden si vede dal fatto che, rispetto a un anno fa, il 43% del campione complessivo ha dichiarato di evitare alcuni siti o app per motivi di privacy, il 39%

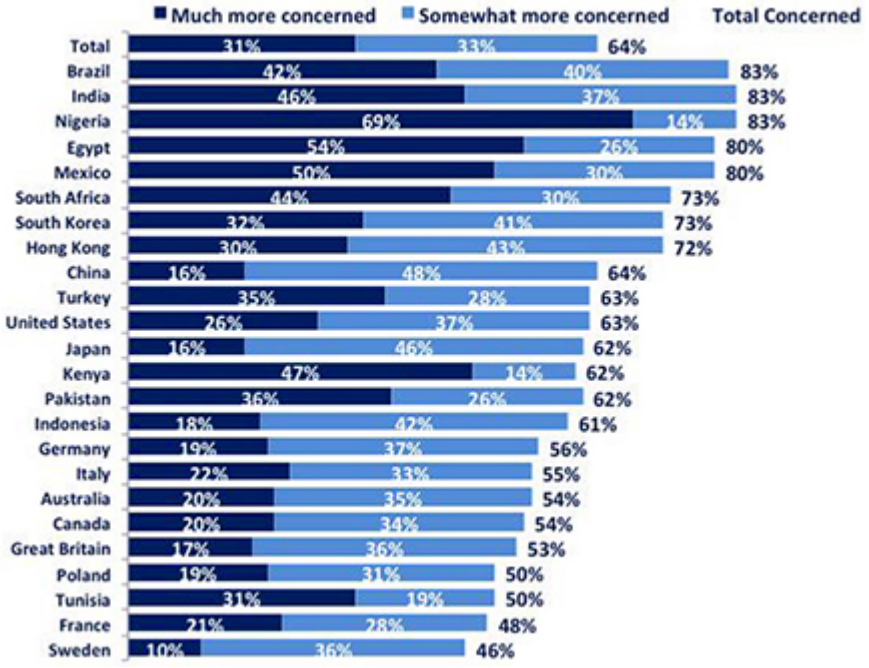

FIGURE 4.2: Riassunto del sondaggio svolta dal Cigi e dal Ipsos

cambia le password in modo regolare, il 28% si autocensura prima di esprimere un'opinione in Rete, il 74% del campione complessivo è preoccupato di essere monitorato da parte di aziende private e dalla vendita dei propri dati personali a fini commerciali, l'11% ha chiuso account sui social media, mente il 10% usa Internet meno frequentemente.

I risultati dell'indagine di Symantec e del sondaggio svolto dal Cigi e Ipsos sono in linea con un studio svolto dal Pew Research Center[19] e rivelano le reazioni dell'opinione pubblica al Datagate. Solo il 2% si sente molto sicuro quando condivide informazioni private sui social network. Come si vede per mail e messaggi di testo va appena meglio, ma il messaggio di fondo non cambia: il post-Snowden è l'era della sfiducia.

Il Datagate, di cui hanno saputo sostanzialmente tutti ha mutato anche la percezione dell'anonimato in rete. Soltanto nel 24% dei casi gli interpellati reputano l'anonimato su Internet facile da ottenere.

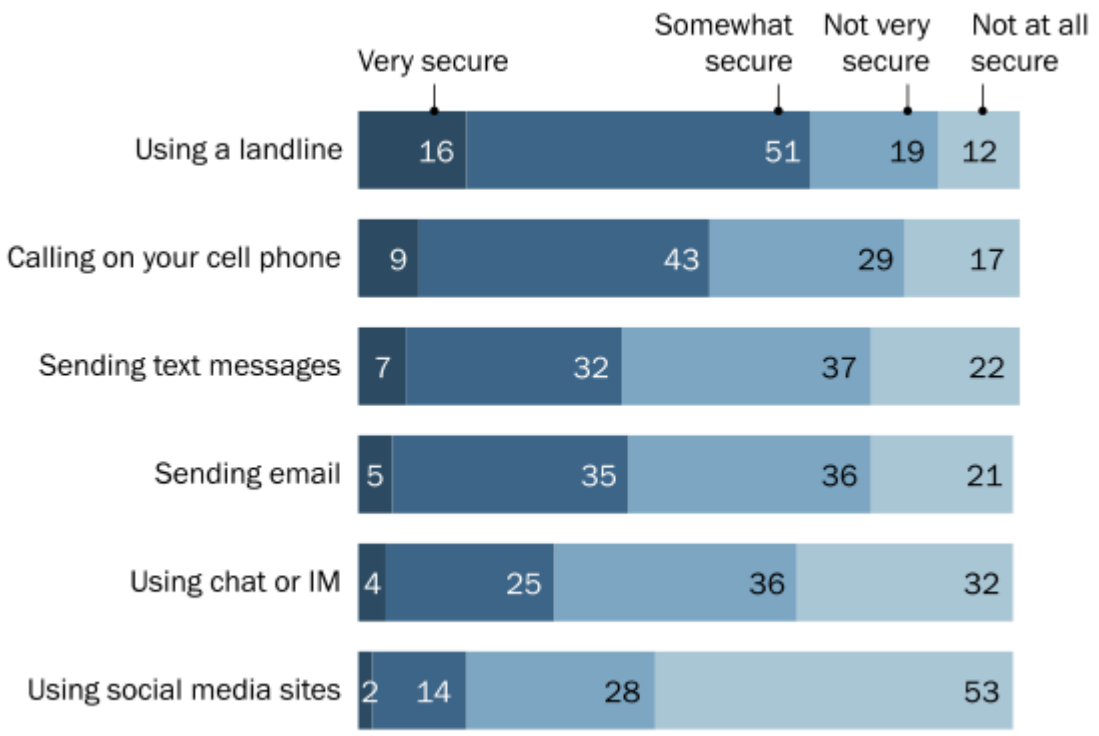

FIGURE 4.3: Percentuale di adulti che si sente sicuro a condividere informazioni private

I risultati sono confermati anche da altri studi più vecchi, come quello realizzato dall'Huffington Post e da YouGov[20] . E allo stesso modo, un precedente report del Pew, datato settembre 2013, notava come il 68% degli americani ritenesse che le leggi correnti non fossero sufficienti a proteggere adeguatamente la propria privacy. A confermare questo filone, anche il crescente utilizzo di strumenti di crittografia anche presso settori di utenza non specialistici.
Snowden ha quindi cambiato la nostra prospettiva sulla privacy dando il via ad un mondo aperto al dibattito sulla privacy, la sorveglianza e i diritti digitali

4.1.1 Datagate

Lo scandalo Datagate è iniziato il 5 giugno 2013 con la pubblicazione da parte del quotidiano britannico Guardian di alcuni documenti riservati secondo i quali, la compagnia di telecomunicazioni Verizon avrebbe consegnato all'FBI dati in grado di mettere a rischio la privacy dei propri utenti.
Secondo questi documenti, l'FBI ha ricevuto in segreto, l'autorizzazione per tenere sotto controllo le comunicazioni con un accesso completo ai dati per tre mesi. Inoltre già dal 2007 i dati relativi alle comunicazioni effettuate via internet da milioni

di utenti sarebbero stati resi disponibili mediante il programma PRISM.
L'iniziativa, di tenere sotto controllo le comunicazioni, sembra sia stata presa dalla National Security Agency (NSA), agenzia di intelligence degli USA, che generalmente tiene sotto controllo specifiche persone considerate potenziali terroristi, e non l'intera popolazione.
Il 9 giugno il Guardian pubblica informazioni su Boundless Informant: un'analisi big data e un sistema di visualizzazione dei dati usata dalla Agenzia di sicurezza nazionale statunitense per dare ai manager NSA una sintesi delle attività di raccolta dati in tutto il mondo eseguite dalla NSA stessa. Secondo una mappa pubblicata sempre dal quotidiano è fuoriuscita dal programma Boundless Informant, quasi 3 miliardi di dati sono stati catturati solo negli Stati Uniti a marzo 2013 in 30 giorni.
Il 12 giugno il South China Morning Post rivela che la NSA è penetrata illegalmente in computer cinesi e di Hong Kong. fin dal 2009. Il 23 Giugno lo stesso giornale riporta che l'NSA avrebbe hackerato le compagnie telefoniche cinesi.
Il 17 giugno il Guardian racconta che nel progetto di sorveglianza globale è incluso anche il governo del Regno Unito, attraverso l'agenzia di intelligence GCHQ (Government Communications Headquarters).
Il 20 Giugno vengono pubblicati due documenti segreti firmati dal generale Eric Holder, che spiegano le regole con cui la NSA opera in caso di indagini estere o statunitensi.
il 21 giugno si scoprono altri dettagli su Tempora, il programma della GCHQ per monitorare dati di fibra ottica.
Il 29 Giugno Der Spiegel scrive che gli USA hanno spiato anche diplomatici dell'Unione Europea e il giorno stesso arrivano nuove rivelazioni su una collaborazione di altri 6 paesi europei(Italia, Spagna, Germania, Francia, Danimarca e Paesi Bassi) nel programma di controllo globale delle comunicazioni.
Tutte queste informazioni sono state rivelate grazie a Edward Snowden, ex informatico della Cia. Per diffondere le informazioni in suo possesso, Snowden si è avvalso della collaborazione di Glenn Greenwald, giornalista del The Guardian. Dopo le sue rivelazioni adesso si trova in Russia e, ancora oggi, cerca di far conoscere al mondo l'operato dei servizi segreti statunitensi.
La portata delle rivelazioni fatte da Snowden è incalcolabile, e oltre a mettere in imbarazzo più di un governo, ha dato il via ad un'era dove le parole chiavi sono sicurezza e privacy.
Google e altre aziende hi-tech che gestiscono grandi quantità di informazioni sugli utenti hanno intrapreso battaglie legali per avere diritto di pubblicare rapporti più completi su quanto sta accadendo. La sorveglianza da parte dei governi è reale

e si è intensificata nel primo semestre del 2014. Più di 31 mila richieste di informazioni, a cui Google ha dato il proprio assenso solo nel 65% dei casi. Dagli Usa sono arrivate oltre 12 mila richieste, e sul podio dei Paesi con il maggior numero di domande ci sono anche Germania e Francia, entrambe con circa 3 mila richieste. Le richieste da parte dei governi di tutto il mondo per ottenere i dati degli utenti sono aumentate del 150% anche senza conteggiare le questioni di sicurezza nazionale da parte di NSA e FBI.
Se si accetta di essere controllati e monitorati in nome dell'interesse pubblico si instaura un circolo vizioso che ci porterà a considerare la sorveglianza di massa come qualcosa di assolutamente normale. Questo è il vero problema.

4.1.2 Violazione dei dati in Europa

La violazione di dati (data breach) è intesa come il rilascio intenzionale o non intenzionale di informazioni sicure in un ambiente attendibile. Ciò può includere il furto o la perdita di supporti digitali o furti di computer sui quali tali informazioni sono memorizzate in chiaro, la pubblicazione di tali informazioni su Internet o il trasferimento di tali informazioni ad un sistema informativo di un'agenzia forse ostile.
Philip N. Howard[21] del Center For Media, Data and Society della Central European University, ha pubblicato un dettagliato rapporto che restituisce per la prima volta in maniera sistematica statistiche e analisi riguardanti le violazioni di dati personali di cui sono stati vittima i cittadini europei. Il documento mostra i dati

Table 1: Quick Fact Table	Values
Total Number of Breaches Involving European Targets	229
Total Volume of Breached Records Across All Incidents	641,979,541
Number of Times a Specific Country in Europe Was Identified as Target	267
Number of Global Breaches Involving European Targets	29
Volume of Records From Global Breaches that Impact People in Europe	415,012,618
Volume of Records From Europe-Specific Breaches	226,966,923
Total Number of Breaches Involving European Targets	229
Total Volume of Breached Records Across All Incidents	641,979,541
Number of Times a Specific Country in Europe Was Identified as Target	267
Number of Global Breaches Involving European Targets	29
Volume of Records From Global Breaches that Impact People in Europe	415,012,618
Volume of Records From Europe-Specific Breaches	226,966,923

FIGURE 4.4: I principali risultati del rapporto CMDS sulle violazioni di dati personali in Europa

raccolti, tra il 2005 e il terzo quarto del 2014, da un team di 11 ricercatori in 450 ore di lavoro. Dai risultati possiamo vedere che per ogni 100 individui coinvolti nello studio, sono stati compromessi 43 registri personali. Questo numero sale a

56 per ogni 100 utenti Internet nel campione analizzato senza considerare i dati personali violati dalle agenzie di sicurezza di altri paesi.
In Italia i casi di data breach documentati nel database sono sette per poco meno di 75 mila dati personali esclusivamente coinvolti. Quattro provengono dagli hacktivisti di Anonymous Italia per dimostrare la debolezza dei sistemi di sicurezza impiegati dagli enti governativi. Dal documento si evince che molte organizzazioni nei paesi stanno facendo un pessimo lavoro nella gestione dei dati personali diventando i principali bersagli del cybercrimine.

Country	Population	Internet users	Number of Breaches Involving Each Country	Volume of Breaches Exclusively Involving That Country	Records Per Person	Records Per Internet User	Breaches Originating In This Country
Austria	8,526,429	7,135,168	9	683,731	8.02	9.58	2
Belgium	11,144,420	9,441,116	4	9,700	0.09	0.10	1
Bulgaria	7,167,998	4,083,950	5	64,678	0.90	1.58	0
Croatia	4,272,044	2,780,534	0	-	0.00	0.00	0
Cyprus	1,153,058	726,663	1	-	0.00	0.00	0
Czech Republic	10,740,468	8,322,168	8	159,538	1.49	1.92	1
Denmark	5,640,184	5,419,113	6	32	0.00	0.00	1
Estonia	1,283,771	1,047,772	0	-	0.00	0.00	1
Finland	5,443,497	5,117,660	7	428,300	7.87	8.37	1
France	64,641,279	55,429,382	15	2,782,428	4.30	5.02	1
Germany	82,652,256	71,727,551	28	56,422,711	68.27	78.66	3
Greece	11,128,404	6,438,325	4	9,016,885	81.03	140.05	1
Hungary	9,933,173	7,388,776	2	55,146	0.56	0.75	1
Ireland	4,677,340	3,817,491	12	916,934	19.60	24.02	1
Italy	61,070,224	36,593,969	7	74,601	0.12	0.20	3

FIGURE 4.5: Riassunto Data breach di alcuni paesi

4.2 Cookie

Un cookie HTTP[22], più comunemente denominato cookie, è un header aggiuntivo presente in una richiesta o risposta HTTP inviato da un server, come un sito web, ad un client, tipicamente un browser impiegato per la navigazione online.
A sviluppare i cookie è stato Lou Montulli, ingegnere presso Netscape Communications, nel 1994. Esso implementò nel browser web Netscape una tecnologia denominata Persistent Client State Object, ispirandosi a un processo di identificazione tra macchine adottato in ambiente UNIX e denominato magic cookie. La tecnologia sviluppata da Montulli è alla base dei cookie persistenti, una delle due categorie della famiglia dei cookie tecnici, utilizzati per eseguire autenticazioni automatiche, tracciatura di sessioni e memorizzazione di informazioni specifiche.
I cookie persistenti non vengono distrutti in automatico con la chiusura del browser,

ma rimangono fino ad una data di scadenza preimpostata. L'altra categoria di cookie tecnici è costituita dai cookie di sessione che vengono invece distrutti in automatico ogni volta che il browser web viene chiuso. Anche i cookie analytics, se utilizzati per ottimizzare il sito direttamente dal titolare del sito stesso, permettono di raccogliere informazioni sugli utenti e possono essere considerati cookie tecnici.
Un'altra famiglia di cookie è rappresentata dai cookie di profilazione, utilizzati per tracciare la navigazione dell'utente e creare profili personali su esso. Attraverso i cookie di profilazione è possibile trasmettere all'utente messaggi pubblicitari in linea con le preferenze già manifestate durante la navigazione online tracciata.

4.2.1 Caratteristiche

Come abbiamo detto il server assegna un cookie all'utente aggiungendo un header alla risposta che il client deve memorizzare in un'area apposta. Il cookie è composto da una stringa di testo arbitraria con data di scadenza e un pattern per riconoscere i domini a cui rimandarlo. Il browser client rimanda il cookie, senza alcuna modifica, allegandolo a tutte le richieste HTTP che soddisfano il pattern, entro la data di scadenza. Il server a sua volta potrà decidere se assegnare un nuovo cookie, sovrascrivendo quello precedente.
Contrariamente a quanto comunemente si crede un cookie non deve necessariamente essere un piccolo file di testo. Nel cookie solitamente possiamo trovare sei attributi:

- **Nome/Valore**;
- **Scadenza (Expiration date):** attributo opzionale che indica la data di scadenza del cookie. Può essere espressa come numero massimo di giorni oppure come Now (implica che il cookie viene eliminato appena viene creato) o come Never (implica che il cookie non è soggetto a scadenza, ovvero i cookie persistenti);
- **Modalità d'accesso (HttpOnly):** rende il cookie invisibile a javascript e altri linguaggi client-side presenti nella pagina;
- **Sicuro (Secure):** campo che indica se il cookie debba essere trasmesso criptato con HTTPS;
- **Dominio (Domain):** campo che indica il dominio di visibilità del cookie;

- **Percorso (Path):** campo che indica il percorso di visibilità del cookie;

```
HTTP/1.1 200 OK
Cache-Control: private
Content-Type: text/html
Set-Cookie: PREF=ID=5e66ffd215b4c5e6:
TM=1147099841:LM=1147099841:S=Of69MpW
Bs23xeSv0; expires=Sun, 17-Jan-2038 1
9:14:07 GMT; path=/; domain=.google.c
om
```

FIGURE 4.6: Esempio di risposta HTTP da google.com che imposta un cookie con degli attributi.

4.2.2 Problemi sulla privacy

A Gennaio 2015 il Garante per la protezione dei dati personali ha fatto partire una campagna per sensibilizzare gli utenti di Internet sull'invasività dei cookie. Il provvedimento generale sull'uso dei cookie, pubblicato sulla Gazzetta Ufficiale n. 126 del 3 giugno 2014, prevede lo stop all'installazione dei cookie per finalità di profilazione e marketing da parte dei gestori dei siti senza aver prima informato gli utenti ed aver ottenuto il loro consenso. Inoltre, questo provvedimento, illustra le modalità semplificate per fornire agli utenti l'informativa online sull'uso dei cookie e le indicazioni per acquisire il consenso degli stessi, quando richiesto dalla legge. Per l'installazione dei cookie tecnici non è richiesto il consenso degli utenti, ma è necessario dare l'informativa. I cookie di profilazione, invece, possono essere installati sul terminale dell'utente soltanto se quest'ultimo abbia espresso il proprio consenso dopo essere stato informato. Il Garante per la Privacy ha previsto un periodo transitorio, che terminerà il 2 Giugno 2015, per consentire ai soggetti interessati di mettersi in regola.
Inoltre in ambito UE è più volte intervenuto, in materia di cookie, il Gruppo di lavoro ex art. 29, organismo consultivo e indipendente composto da un rappresentate delle autorità di protezione dei dati personali designate da ciascuno Stato membro, dal GEPD (Garante europeo della protezione dei dati) e da un rappresentate della Commissione.

I cookie possono risultare veramente invasivi nell'ambito della sfera privata degli utenti e quindi sorge la necessità di difendere la propria privacy. Innanzitutto è possibile bloccare i cookie di terze parti, ovvero quelli impostati da un sito web diverso da quello che si sta visitando. Questi cookie generalmente non sono indispensabili per navigare e quindi è possibile disattivarli attraverso apposite funzioni del proprio browser.
Nella maggior parte dei browser è possibile inoltre attivare l'opzione Do Not Track, bloccando la raccolta dei dati di navigazione da parte dei siti web che rispettano questa opzione.
Un'altra possibilità è quella di ricorrere alla modalità di navigazione anonima o in incognito mediante la quale è possibile navigare, senza lasciar traccia nel browser, dei dati di navigazione. Bisogna sottolineare che la navigazione anonima non significa essere anonimi in rete perché i dati di navigazione continueranno a restare disponibili ai gestori dei siti web ed ai provider di connettività.
Infine alcuni browser consentono di eliminare i cookie direttamente, ma ad ogni accesso vengono scaricati nuovi cookie e quindi l'operazione andrebbe eseguita periodicamente. Alcuni browser offrono sistemi automatizzati per l'eliminazione periodica dei cookie.
Cookie poisoning è una procedura di manipolazione sui cookie che consiste nel modificare i contenuti di un cookie al fine di eludere i meccanismi di sicurezza. Attraverso questa tecnica è possibile ottenere informazioni private e non autorizzate da un utente. I cookie immagazzinati nel computer dell'utente contengono le informazioni che permettono alle applicazioni di autenticare lo userID, monitorare i comportamenti dell'utente, e personalizzare i contenuti di un sito. In genere, questi dati sono sottoposti a cifratura, ma non sempre gli algoritmi sono sicuri per cui qualche malintenzionato potrebbe carpire i nostri dati ed utilizzarli o modificarli. Secondo l'organizzazione The Open Web Application Security Project, comunemente detta OWASP, la manipolazione dei cookie è uno dei 20 attacchi più utilizzati dagli hacker, soprattutto nei sistemi di e-commerce per identificare l'utente.

4.3 Sicurezza nel mobile

I cookie, utilizzati come mezzo di tracking, sono meno efficaci su smartphone e tablet. E vista la rapida espansione di questi dispositivi mobili alcune tra le Internet Company sulle quali incide maggiormente la fruizione mobile delle persone hanno implementato una serie di nuove soluzioni per poter identificare i target a

cui mostrare inserzioni che siano effettivamente rilevanti rispetto ai loro interessi e comportamenti.
Facebook, oltre a tutti gli elementi che un utente acconsente di condividere sul proprio profilo personale, prende in considerazione anche le informazioni provenienti dalle azioni svolte su tutti i website o app che permettono il Single Sign-On con Facebook.
Google fa leva allo stesso modo sul Single Sign-On delle persone all'interno dei vari account utili per accedere alle funzioni del suo network di servizi. Inoltre utilizza il tracking, a tutti gli utenti che sfruttano un sistema operativo Android, dell'ID.
Non manca all'appello Apple, che tra le altre cose tiene traccia di tutto ciò che accade sull'account iTunes di un utente. Il tracking può essere disabilitato, modificando le impostazioni del device che si utilizza, e di conseguenza andando ad incidere sul funzionamento e le performance di determinate applicazioni o servizi.
I nostri dispositivi mobili sono sempre con noi e trasmettono la nostra posizione e altre informazioni a diverse compagnie. Per capire quanti dell'infinità di comunicazioni sono al riparo da occhi ci ha pensato l'organizzazione no profit Electronic Frontier Foundation (EFF). Il gruppo, che si occupa di tutela dei diritti digitali e libertà di parola, ha collaborato con il sito ProPublica e con l'università di Princeton per redigere la Secure Messaging Scorecard[23] , una specie di classifica di impermeabilità dei dispositivi per le chat e i messaggi.
I vari sistemi sono stati valutati sulla base di sette parametri relativi alla crittografia dei messaggi, alla possibilità per il provider di leggerli, alla possibilità di verificare l'identità dei contatti, alla sicurezza in caso di violazione delle chiavi cifrate di protezione, alla revisione indipendente del codice, alla documentazione del design di sicurezza e alla verifica da parte di Audit esterni del codice di sicurezza. I risultati mettono in risalto che tool che sono più facili da usare per il grande pubblico non mettono in pratica sistemi di protezione particolarmente efficaci. La stessa cosa vale per i sistemi più popolari e diffusi.
Al primo posto vediamo Telegram, il sistema lanciato nel 2013 e considerato il principale rivale di WhatsApp, che permette di scegliere tra chat normali e segrete. I messaggi hanno la caratteristica di auto cancellarsi e di essere visibili solo ai dispositivi da cui i messaggi partono.
Tra i servizi di messaggistica più popolari, abbiamo i due prodotti di casa Apple: FaceTime e iMessage. Entrambi hanno il pregio di usare il sistema Perfect Forward Secrecy che permette, nel caso in cui le chiavi di protezione cifrata vengano violate, di accedere solo alle conversazioni dell'ultima sessione, senza poter vedere quelle precedenti.
Facebook Chat, Google Hangouts, SnapChat, WhatsApp e Skype molto diffusi in Italia garantiscono di rendere illeggibili i contenuti dei messaggi ma poi poco altro.

FIGURE 4.7: Punteggi ottenuti da alcuni sistemi considerati nel Secure Messaging Scorecard

Solo Skype, come provider, si impegna a non leggere le conversazioni degli utenti. Risulta evidente la necessità di proteggersi nell'ambito dei nostri dispositivi mobile. La cosa più semplice e basilare è il monitoraggio del comportamento delle applicazioni con un occhio di riguardo alle autorizzazioni/richieste di accesso dalle applicazioni in esecuzione.
Una tecnica più sofisticata ed efficace è il criptaggio. Signal[24] è una app cross-platform che permette di inviare messaggi criptati, immagini e video messaggi virtualmente a chiunque abbia uno smartphone. L'applicazione gratuita è realizzata da Open Whisper Systems, creatori di TextSecure e Redphone(applicazioni che permettono agli utenti di Android di inviare messaggi e chiamare in modalità criptata end-to-end) e tutto ciò che serve è un numero di telefono (salvato come un hash, una riga fissa di caratteri criptati) per registrarsi sui server della Whisper Systems.
I messaggi di testo inviati tramite Signal possono essere criptati solo tramite una connessione dati e non con gli SMS. Cosa che non vale per TextSecure, la versione per Android, che si fonderà presto con Signal, che ha un'opzione di ricaduta sugli SMS quando non sono disponibili reti dati. Inoltre è importante ricordare che Signal non protegge dalla raccolta massiva di metadati.
I dispositivi mobili sono strumenti di controllo efficienti e completi e quindi la guerra per la privacy deve estendersi ai dispositivi mobili. Deve iniziare un duplice approccio alla privacy mobile con campagne per educare tutti gli utenti alla privacy dei dati sui dispositivi mobili.

5. Deep Web

5.1 Il lato nascosto di Internet

Nel 1969 il DARPA (Defense Advanced Research Projects Agency)[25], un'agenzia governativa del Dipartimento della Difesa degli Stati Uniti incaricata dello sviluppo di nuove tecnologie per uso militare, studiò e realizzò ARPANET (Advanced Research Projects Agency NETwork), una rete di nodi basata su una architettura client/server destinata ad un uso militare. Questa rete, pensata per scopi militari durante la guerra fredda, perse il proprio utilizzo nel 1983 e con il passare del tempo rimase sotto il pieno controllo delle università, diventando un utile strumento per scambiare le conoscenze scientifiche e per comunicare.
Nel 1991 al CERN (l'Organizzazione Europea per la Ricerca Nucleare) di Ginevra venne definito il protocollo HTTP (HyperText Transfer Protocol) che permise la nascita del WWW (World Wide Web) come lo conosciamo oggi, ma solo nel 1993 venne rese pubblico facendo nascere l'Internet che oggi utilizziamo.
Nel 2002 la Marina Statunitense iniziò lo sviluppo di TOR (The Onion Router) per proteggere le comunicazioni governative tramite un circuito virtuale di crittografia a strati, in questo modo nacque la Rete Onion.
Nel 2004 la Elettronic Frontier Foundation, un organizzazione no profit che dal 1990 si batte per i diritti e le libertà digitali, iniziò a finanziare lo sviluppo di TOR in maniera tale da poter estendere il suo utilizzo al di fuori dell'ambito militare. Fino ad oggi TOR è stato finanziato da molte altre società ed organizzazioni non governative mentre ora è gestita da The Tor Project, una associazione senza scopo di lucro.
La rete Internet e TOR quindi nascono entrambe da un progetto militare della Difesa degli Stati Uniti con la differenza che Internet è conosciuto da tutti visto la semplicità nell'accedervi mentre la rete Onion è nascosta, non può essere censurata ne localizzata.
Tramite un computer e un router è possibile accedere ad Internet, ma questa è solo una minima parte di quello che effetivamente è presente. Secondo una ricerca

condotta nel 2000 dalla Bright Planet[26] il Web è composto da 600 miliardi di documenti circa, ma i classici motori di ricerca come Google ne indicizzano solo 2 miliardi, meno dell'1%, questo perché per indicizzare utilizzano dei crawlers che seguono dei link ipertestuali e non possono indicizzare siti che all'apparenza non esistono.
Questo restante 99% è il Deep Web, il web nascosto ai tradizionali motori di ricerca ed alla maggioranza delle persone. Il Deep Web è composto da:

- **Contenuti dinamici:** pagine web dinamiche, ovvero pagine Web inesistenti il cui contenuto viene generato sul momento dal server, che possono essere richiamati solo compilando un form o a risposta di una particolare richiesta;
- **Pagine non collegate:** pagine Web che non sono collegate a nessun'altra pagina Web. Se l'accesso non è impedito da adeguate impostazioni di sicurezza, il motore indicizza la parent directory del sito, che contiene non solo le pagine visibili, ma tutto ciò che è caricato nel server ospitante;
- **Pagine ad accesso ristretto:** siti che richiedono una registrazione o comunque limitano l'accesso;
- **Script:** pagine che possono essere raggiunte solo attraverso link realizzati in javascript o in Flash e che quindi richiedono procedure particolari;
- **Contenuti non di testo:** file multimediali, archivi e documenti scritti in linguaggi non HTML.

Per spiegare meglio il Deep Wep è utile avvalersi della metafora dell'iceberg: al di sopra del mare troviamo la parte più piccola, Internet. Il grosso dell'iceberg, cioè il Deep Web, si trova invece sotto la superficie del mare. Immergendoci nell'acqua, ci sono diverse reti nascoste chiamate Darknet: una rete i cui contenuti sono raggiungibili solo se si conosce il loro indirizzo. Il termine fu originariamente coniato negli anni settanta per designare reti che erano isolate da ARPANET per motivi di sicurezza. Oggi viene usato, erroneamente, anche come sinonimo di Deep Web, ma nel suo significato più generale è usato per descrivere qualsiasi tipo di gruppo chiuso e privato di persone che comunicano tra di loro.

Il vero boom nell'utilizzo del Deep Web lo possiamo osservare da Giugno 2003, quando Edward Snowden ha cominciato a raccontare al mondo quanto la NSA americana abbia per anni spiato dai comuni cittadini ai capi di Stato. Questa

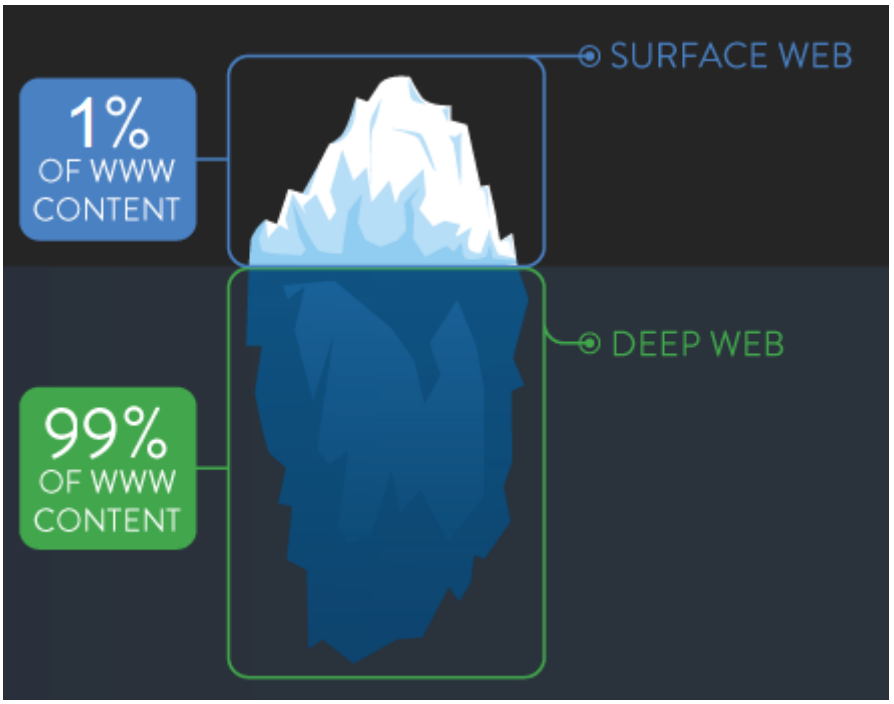

FIGURE 5.1: Deep Web come un iceberg

paura per la propria privacy ha portato sempre più persone a cercare un modo per tutelarsi. Lo usano attivisti per i diritti umani, giornalisti, militari, forze dell'ordine e persone comuni ed è proprio questa varietà di utilizzatori che ne aumenta la sicurezza e privacy.
In molti paesi dove la censura e l'autoritarismo soffocano la democrazia, la cultura, l'arte, la creatività o la religione, il Deep Web sta diventando sempre più una risorsa, una speranza e una nuova frontiera per essere liberi.

5.2 Tor Browser Bundle: la porta per il Deep Web

Come abbiamo visto accedere al Deep Web non è facile come per Internet, ma non è nemmeno cosi difficile. Un modo per accedervi[27] consiste nell'entrare nella Rete Onion tramite il software Tor Browser che anonimizzerà il traffico attraverso la sua crittografia a strati permettendo di accedere a siti che al di fuori di questa rete non esistono.
Il funzionamento è semplice e molto efficace. Prendiamo l'esempio in cui Alice vuole comunicare con Bob funzionando come nodo Client. In questa configurazione di base, Tor gestisce unicamente le connessioni di Alice permettendole di

collegarsi alla rete Tor. Abbiamo inoltre Dave che è un particolare server pubblico chiamato Tor directory server che fornisce la lista pubblica di relay Tor e la mantiene periodicamente aggiornata.
Il primo passo per Alice è quello di contattare Dave per ottenere la lista dei nodi Tor (nell'immagine sono quelli marchiati con il + verde). Di seguito Alice invia un messaggio a Bob e il nodo Client si preoccupa di costruire, in maniera casuale, una catena di nodi attraverso i quali invierà il messaggio.

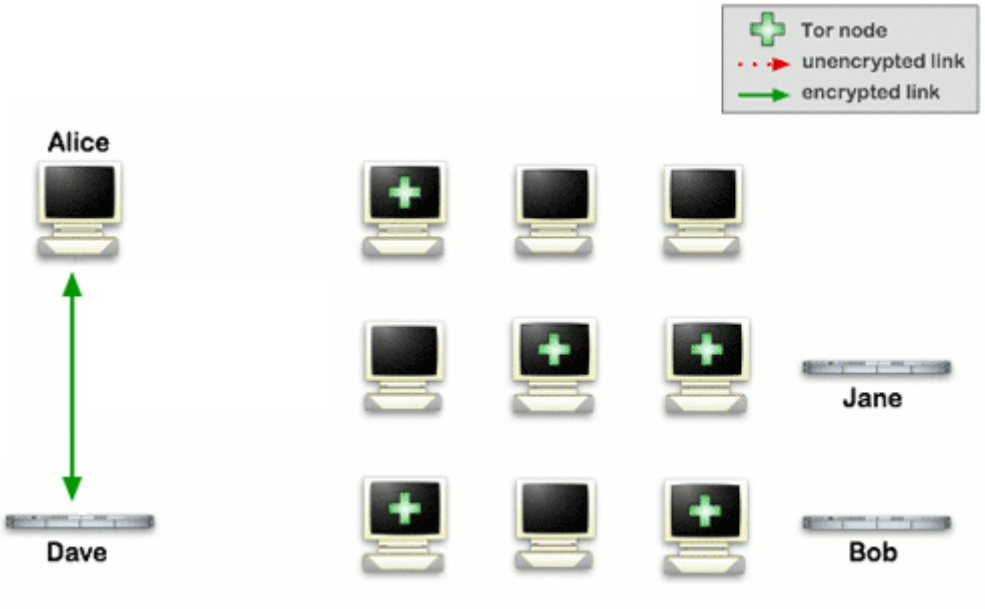

FIGURE 5.2: Come funziona Tor:1

Ciascun nodo che incontriamo, fino ad arrivare a Bob, è un utente che ha deciso di contribuire all'espansione e al funzionamento della rete Tor. Il primo nodo che incontriamo viene chiamato Guard router, è il nodo a cui affidiamo tutta la comunicazione. L'ultimo nodo è chiamato Exit router ed è il nodo più importante perchè contatta direttamente le macchine esterne mettendosi in mostra. Tutti i rimanenti relay sono considerati nodi intermediari e vengono chiamati Middleman router, gestiscono il traffico di terzi da e per la rete Tor, senza collegarsi direttamente all'esterno. Infine abbiamo un'ultima tipologia di nodi, Bridge router, semi-pubblici di tipo sperimentale, studiati per permettere di collegarsi alla rete Tor anche in presenza di un filtraggio efficace contro Tor (come in Cina e Iran). Non compaiono nelle liste pubbliche dei nodi noti, ma devono venire richiesti esplicitamente
La comunicazione all'interno della rete Tor è completamente criptata impedendo così ad un malintenzionato che ha intercettato un messaggio di ricostruirne il

flusso. L'ultimo tratto invece non è criptato, ma questo dipende dalla richiesta effetuata dal Client. L'importante resta il fatto che l'ultima comunicazione è estranea alla trasmissione di informazioni all'interno del protocollo di Tor.
Torniamo al nostro tentativo di comunicazione di Alice con Bob. Per capire bene il significato e la potenza di questo protocollo immaginiamo che il nostro pacchetto che inviamo sia una cipolla in cui al centro abbiamo il messaggio da inviare mentre tutti i restanti anelli siano diversi stati di criptografia. Quando Alice invia il pacchetto al Guard router avremo che lo strato più esterno della nostra cipolla è protetto con le chiavi di Guard stesso, quindi solo lui può leggere cosa c'è scritto. Leggendo le istruzioni contenute in questo strato, il Guard router invia il resto del contenuto ad un determinato Middleman router. Il Middleman che ha ricevuto il pacchetto sarà l'unico che potrà leggere lo strato subito successivo a quello esterno e come nel caso precedente spedisce il resto del contenuto al Middleman router successivo o al Exit router. Già a questo livello è possibile notare che è sparita ogni indicazione riguardante Alice. L'ultimo passaggio avviene quando l'Exit router riceve il pacchetto, ne legge il contenuto decifrabile solo da lui e agisce di conseguenza. Una volta che Bob ha ricevuto il messaggio, trasmetterà un'eventuale contenuto all'Exit router che a sua volta spedirà il pacchetto al Middleman router e così via fino a tornare da Alice.

FIGURE 5.3: Come funziona Tor:2

Una volta che il circuito è stato creato, è possibile scambiare diversi tipi e quantità di dati e usare molti tipi di applicazioni attraverso una rete Tor. Per motivi di efficienza però, il software Tor utilizza lo stesso circuito per le connessioni che avvengono negli stessi dieci minuti o giù di lì. Le richieste successive sono fornite a un nuovo circuito, per evitare che nessuno possa collegare le azioni precedenti con quelli nuovi.
Successivamente se Alice vuole comunicare con Jane verrà creato un nuovo percorso casuale e sicuro.

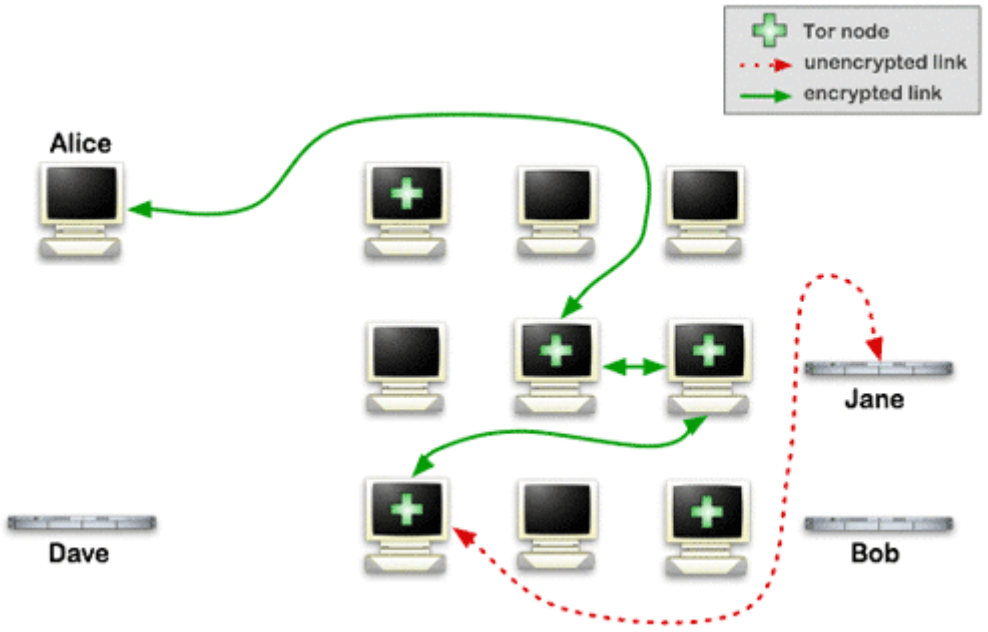

FIGURE 5.4: Come funziona Tor:3

Qual'è il motivo che spinge moltissime persone a cercare protezione con Tor? La risposta la troviamo nell'analisi del traffico. Con analisi del traffico si definisce la tecnica basata sull'analisi passiva del traffico prodotto da una macchina o da un gruppo di esse. Ispezionando i pacchetti che escono da una rete, un persona è potenzialmente in grado di scoprire una grande quantità di informazioni riguardanti l'origine di quei pacchetti. Pensando ad una mail: un malintenzionato che ne intercetta il contenuto è in grado di sapere chi sia il mittente, a quale indirizzo sta scrivendo e cosa gli vuole comunicare. Quando questo concetto viene spostato a livello di traffico di rete lo scenario non cambia: un malintenzionato che intercetta un qualsiasi traffico di rete è in grado di scoprire, dall'header dei pacchetti, chi li ha inviati, chi li riceverà e, guardando il body, cosa si vogliono comunicare.
Una prima soluzione possibile è la crittografia, ovvero dei metodi per rendere

un messaggio incomprensibile a persone non autorizzate a leggerlo. Il problema però è che, visto come funziona Internet, la crittografia permette di nascondere il corpo del messaggio e non le informazioni contenute nell'header. Anche solo la conoscenza della sorgente e della destinazione del proprio traffico Internet permette ad altri di ricostruire le nostre abitudini e i nostri interessi personali e questo può avere un significante impatto sulla nostra vita.
All'inizio degli anni 80 David Chaum propose le Mix Networks[28], una serie di proxy server da utilizzare in cascata. Ogni messaggio è cifrato da ciascun proxy utilizzando una crittografia a chiave pubblica ottenendo un pacchetto rappresentabile come una matriosca con all'interno il messaggio che vogliamo trasmettere. Ogni proxy server elimina il suo strato di crittografia per poter leggere la destinazione in cui inviare il pacchetto finché non arriva a destinazione.
In questo modo un malintenzionato doveva ispezionare il traffico all'uscita di ogni proxy per poter scoprire qualcosa di utile sul mittente del pacchetto che usciva dall'ultimo nodo della catena. L'idea di Chaum, benché non utilizzata, è stata il seme che ha lanciato la ricerca sull'onion routing.

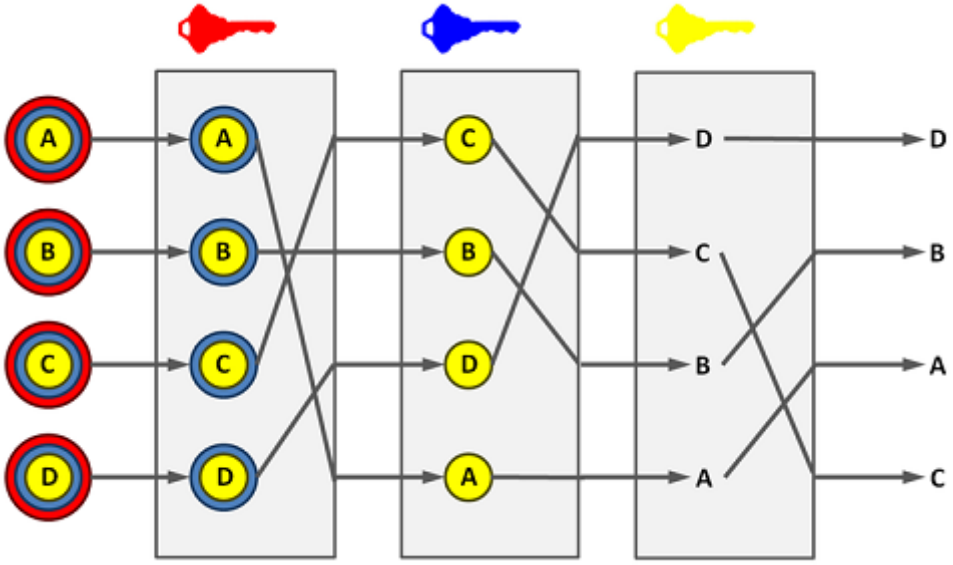

FIGURE 5.5: Schema del funzionamento di una Mix Network

Poiché in Tor ogni relay vede singolo salto nel circuito né un intercettatore e neppure un relay compromesso possono utilizzare l'analisi del traffico per collegare la sorgente e la destinazione della connessione. In questo modo il protocollo è in gradi di sconfiggere l'analisi del traffico: un malintenzionato che osserva il traffico prodotto da uno qualsiasi dei nodi non può correlarne il contenuto ed il mittente. Inoltre, per prevenire che uno dei nodi del circuito possa leggere arbitrariamente

il contenuto delle risposte, la comunicazione viene protetta con una chiave di sessione, che viene modificata frequentemente, apribile solo dal nodo Client.
Secondo i dati ufficiali rilasciati dal Tor Project, i primi tre paesi al mondo che hanno usato Tor nel 2013 sono Usa, Brasile e Germania. Poi Francia, Spagna ed Italia(sesta con una quota di traffico del 4,06Tor funziona solo con TCP e può essere usato da qualsiasi applicazione con il supporto SOCKS. Permette agli utenti di nascondere la loro posizione quando offrono diversi servizi. Utilizzando i nodi di incontro, gli altri utenti possono connettersi a questi servizi nascosti, ciascuno senza conoscere l'identità di rete dell'altro. La funzionalità dei servizi nascosti permette agli utenti di Tor di pubblicare materiale senza preoccuparsi della censura. Nessuno è in grado di determinare chi sta fornendo il sito, e nessuno che fornisca un sito può sapere chi sta scrivendo su di stesso.

5.2.1 Aspetti problematici

Dopo il sequestro, da parte dell'FBI e dell'Europol, di Silk Road 2.0 e di molti altri siti del Deep Web si è iniziato a dubitare sull'effettivo livello di anonimato che la rete può garantire. Secondo un'analisi effettuata[29] dall'esperto di cybersicurezza Nik Cubrilovic, i 414 servizi nascosti confiscati dalla polizia sarebbero molti di meno, solo 276 domini, di cui 153 dei cloni di altri siti. Del resto da subito erano emersi alcuni dubbi sull'entità dell'operazione dal momento che erano state diffuse solo poche decine di nomi della lista dei siti sequestrati.
Oltre all'operazione Onymous c'è stata la diffusione di una pubblicazione scientifica che mostrerebbe come, sotto certe condizioni, sia possibile deanonimizzare fino all'80% del traffico Tor analizzato. Alcuni sviluppatori[30] di diverse università guidati da Sambuddho Chakravarty hanno provato a deanonimizzare, con l'analisi di correlazione del traffico, gli indirizzi IP degli utenti Tor in un esperimento in laboratorio riuscendo nel 100% dei casi. La tecnica che hanno utilizzato si basava sull'analizzare il traffico in entrata e in uscita dai nodi Tor perchè i punti delicati sono il nodo in cui si entra, per l'assenza di anonimosità fino all'ingresso, e il nodo in cui si esce, per la possibilità di tracciare la destinazione del traffico. Successivamente hanno poi applicato lo stesso metodo con un nodo Tor che controllavano loro stessi, riuscendoci nell'81,4% dei casi. Tutto ciò ovviamente non significa che siano in grado di deanonimizzare l'81% del traffico Tor. Un soggetto in grado di osservare il traffico di un utente che entra in un nodo e il traffico dello stesso utente che esce da un altro nodo è potenzialmente in grado di identificare tale utente anche solo correlando dimensione e tempi dei pacchetti.
Per osservare il traffico, i ricercatori, hanno utilizzato strumenti (un semplice

router Cisco e Netflow) che non cercano di guardare dentro ai pacchetti e ai loro contenuti, ma che si limitano a registrarne gli indirizzi IP di origine e destinazione, la quantità di dati e altre informazioni riguardanti il traffico dati. Solo dopo aver raccolto i dati, utilizzando altri software, hanno cercato di analizzare e fare correlazioni sui pattern di traffico di uno specifico arco temporale. Questa però è una delle ipotesi di attacco a Tor, già definita nella sua progettazione originaria nel suo threat model. Ovvero, se siamo di fronte a un avversario così potente da riuscire ad osservare il traffico a livello globale, il rischio che gli utenti siano deanonimizzati è reale. La ricerca presenta comunque un alto numero di falsi positivi in cui l'algoritmo lega un pacchetto in uscita, corrispondente a quello in entrata, ad un'identità sbagliata. Secondo gli sviluppatori di Tor comunque non sarebbe una notizia allarmante. Bisogna infatti considerare che l'81,4% del traffico si riferisce solo ai loro esperimenti in laboratorio, non al traffico di Tor.
Infine la polemica sui finanziamenti a Tor e i suoi presunti nascosti legami con il governo americano hanno incrementato i dubbi sull'anonimato garantito. Questa comunque è un teoria complottista smentita dai documenti stessi della NSA, secondo i quali Tor è un problema e un ostacolo al tentativo dell'intelligence americana di controllare le comunicazioni elettroniche.
Ciò non vuol dire che Tor e i suoi utenti possano dormire sonni tranquilli. Uno dei problemi attuali è infatti come scalare la rete Tor evitando di concentrare troppo traffico su alcuni nodi, proprio perché questa concentrazione renderebbe più facile lo scenario di attacco di cui si parla nella ricerca. Il dato sembrerebbe quindi dare ragione a chi ipotizza che l'operazione Onymous abbia fatto leva soprattutto su errori e scarsa sicurezza lato utente da parte degli operatori di Silk Road e degli altri siti.
Il 30 agosto 2007 un 22enne consulente della sicurezza svedese, Dan Egerstad, sparse il panico diffondendo sul suo blog 100 password con cui, attraverso Tor, le ambasciate di mezzo mondo comunicavano con i propri ministeri. Il 4 febbraio 2013 Alex Biryukov, Ivan Pustogarov e Ralf-Philipp Weinmann, dell'Università del Lussemburgo, riescono a mettere in chiaro 39.824 indirizzi IP di utilizzatori di Tor.
Tutti questi esempi ci dimostrano come anche Tor si aggredibile, ad esempio con il traffic control. Non può offrire una sicurezza al 100% sopratutto per il fatto che spesso la presenza di punti d'attacco dipende da chi lo usa.

5.2.2 L'altra faccia della medaglia

Il deep web da una parte è uno strumento di libertà mentre dall'altra è uno strumento per attività illegali.

Secondo un studio del 2013, realizzato da Pierluigi Paganini e Richard Amores[31], la composizione del deep web è: il 28% sono siti, blog e forum che si occupano di hacking, il 23% di cyber crime, il 17% di propaganda politica o terroristica, il 4% di pornografia e pedofilia. Significa che il 72% ha a che vedere con reati potenziali o reali.
Uno dei più famosi mercati neri è stata Silk Road, creata nel febbraio del 2011 da Ross Ulbricht. Al suo interno si poteva acquistare qualsiasi tipo di droga, ma anche cracker che offrivano le loro prestazioni per rovinare pc altrui, bancomat o rubare informazioni. In linea di massima se una cosa è illegale allora è probabile che si trovasse da qualche parte li dentro.
Secondo lo studio compiuto da Nicolas Christin, Assistant Research Professor alla Carnegie Mellon University di Pittsburgh, dentro Silk Road i venditori guadagnano circa 7 milioni di euro all'anno, oltre 67 mila euro di guadagno al mese per chi gestisce il sito e 4.000 e passa oggetti in vendita al mese. Secondo i dati dell'FBI, le cifre in due anni e mezzo di attività sono invece un fatturato complessivo di 880 milioni di euro, commissioni per quasi 59 milioni di euro, 1.229.000 transazioni e 146.000 acquirenti.
L'FBI è riuscita [32] [33] a trovare i server che facevano funzionare questo mercato, grazie a un errore di configurazione di un CAPTCHA del sito, che ha inavvertitamente rivelato l'indirizzo IP di Silk Road. Alcune questioni stanno mettendo in crisi questa versione però, e molti esperti sostengono che l'FBI abbia probabilmente ricevuto degli aiuti.

FIGURE 5.6: Sito chiuso dall'FBI

Silk Road era il mercato nero più grande prima di essere chiuso dall'FBI il 3 ottobre 2013. Ai primi di novembre viene annunciata l'apertura di Silk Road. Diversi sono stati i tentativi di rimettere in piedi il mercato nero senza grandi risultati. Infatti, anche Silk Road 2.0 è stato chiuso dall'FBI sequestrando l'equivalente di 1 milione di dollari di bitcoin, e 180 mila euro in soldi, droghe, oro e argento.
Anche in questo caso non è chiaro come sia stato localizzato il server di Silk Road 2.0. Tuttavia, dalla lettura degli atti di indagine[34] sembrerebbe che Blake Benthall, il creatore di Silk Road 2.0 abbia commesso molti errori. Il principale sembra essere stato l'utilizzo della propria mail personale per registrare i server usati per Silk Road 2.0. Un agente infiltrato fin dalla nascita di Silk Road 2 avrebbe avuto addirittura accesso al panel di amministrazione del sito.
Da gennaio 2015 è comparso Silk Road Reloaded che invece di usare il network Tor usa I2P. L'uscita da Tor non è l'unico cambiamento. Mentre l'originale Silk Road e il suo successore, Silk Road 2, accettavano solamente Bitcoin, Silk Road Reloaded supporta transazioni con altre criptovalute convertendole in Bitcoin.
Questo nuovo mercato è dimostrazione del fatto che, anche se recentemente hanno sono stati sequestrati con successo centinaia di siti del deep web, i bazar di droghe sono ben lontani dall'essere morti e che piuttosto stanno diventando sempre più ricchi e più diversificati.
Il deep Web come già detto non è solo male. É uno strumento importantissimo per i giornalisti per poter comunicare segretamente con le loro fonti. Per i dissidenti, ovvero quelli che non la pensano come il loro governo e che, in certi paesi, hanno bisogno di comunicare tra loro senza rischiare la propria vita. Per le forze di polizia che possono compiere operazioni sotto copertura e indagare sulla criminalità informatica e non solo. Per conservare i documenti rubati per essere divulgati a tutto il mondo nell'ottica della trasparenza. Per le persone che ci tengono alla loro privacy e non vogliono essere spiate da nessuno. Per le Organizzazioni Non Governative (Amnesty International, Emergency e altre) che fanno usare Tor ai loro volontari e attivisti in paesi stranieri per comunicare con la sede dell'Organizzazione senza far trapelare informazioni al governo locale.
Quindi Tor è il bene o il male? La rete ed i contenuti della rete non sono la stessa cosa. Le due facce della medaglia viaggiano parallelamente e uno sguardo equilibrato è l'unica soluzione. Ci sarà sempre una darknet usata per scopi illeciti e quindi bisogna cercare di educare non di reprimere.
Già nel 2008 Matteo Flora[35] ha scritto un post molto interessante che cito per concludere:

Ciò che mi ha difeso negli esordi di Internet, quello che mi ha fatto da scudo di fronte a contenuti rivoltanti o sconvolgenti, la mia barriera contro l'utilizzo di droghe o la deviazione non è stato un filtro internet o una rete blindata, ma

l'educazione e la formazione che ho ricevuto dai miei genitori e dai miei educatori o insegnanti. Ed un pizzico di intelligenza e di senso critico.
Non si può volere una rete sicura a priori perché il genitore non ha tempo da passare con il figlio per provvedere al suo sviluppo. Non si può volere una rete sicura perché deve divenire, come il televisore in fascia protetta, una sorta di babysitter davanti a cui parcheggiare il pargolo per interminabili ore, sicuri che "non vedere niente di male, eh!". Non si può.
Internet è una fonte di informazioni, un mondo, un ecosistema. Non è differente dalla vita reale, è solamente più agevole in quel contesto pubblicare e ritrovare informazioni che nella vita reale sarebbero comunque rintracciabili e pubblicabili.
(...) Intendo solo fare comprendere che dinamiche e pericoli di Internet non sono difformi dalla vita reale, non sono differenti e devono essere trattati nello stesso modo: parlando e spiegando. Non si può lasciare che la televisione, Internet, Dio, il Furby siano la fonte di informazione e di educazione di un bimbo/ragazzino che probabilmente ha solo necessità di una cosa: qualcuno che gli spieghi come funzionano le cose e la vita. Che sia un genitore o lo stato, che sia un educatore o un parente o un amico, non ha importanza.
Non si protegge un bambino dalla vita, gli si insegna ad comprenderla e affrontarla. Non si protegge un bambino da Internet: gli si insegna a comprenderlo ed affrontarlo.

6. Darknet

6.1 Freenet

Freenet è un software gratuito che permette, in maniera anonima, di condividere file, navigare e pubblicare siti web accessibili soltanto attraverso una rete decentralizzata e composta da innumerevoli sottoreti più piccole. L'intera rete è strutturata in modo tale che ogni nodo, ovviamente anonimo, conosca soltanto una ristretta cerchia di nodi ad esso adiacenti e soltanto con loro si possa collegare. In questa maniera la rete è meno vulnerabile ad eventuali attacchi.

6.1.1 Storia

Freenet[36] nasce nel marzo del 2000 e la sua origine può essere ricondotta a Ian Clarke (uno dei 100 migliori innovatori del 2003 secondo la rivista Technology Review del MIT) ed al suo sistema descritto nel documento dal titolo "Freenet: A Distributed Anonymous Information Storage and Retrieval System". In accordo con CiteSeer[37] (un motore di ricerca pubblico e biblioteca digitale per testi scientifici e accademici, principalmente nel campo dell'informatica, sostituita in seguito da CiteSeerX) questo documento è diventato uno degli articoli di informatica più frequentemente citati nel 2002.
La versione 0.7 del 2008 ha introdotto due modalità di utilizzo: una modalità darknet per collegarsi solo con gli amici ed una modalità OpenNet che permette di collegarsi a qualsiasi altro utente Freenet. Entrambe possono funzionare contemporaneamente, ma la modalità darknet, grazie alla possibilità di collegarsi a persone di fiducia, permette di ridurre notevolmente la vulnerabilità degli utenti.

6.1.2 Caratteristiche e interfaccia utente

Freenet[36] [38]è diverso dalla maggior parte delle altre applicazioni peer-to-peer, sia nel modo in cui gli utenti interagiscono con esso e sia nella sicurezza che offre. La struttura della rete è separata dal protocollo con cui gli utenti comunicano ottenendo cosi una varietà di modi per accedere ai contenuti della rete.
Il più semplice è tramite FProxy, che è integrato con il software nodo e fornisce un'interfaccia Web per navigare freesites (siti web che utilizzano il normale HTML e strumenti correlati, ma il cui contenuto è memorizzato all'interno di Freenet, piuttosto che su un server web tradizionale). L'interfaccia web è utilizzata anche per la maggior parte delle attività di configurazione e gestione del nodo. Attraverso l'uso di applicazioni separate o plugin caricati nel software del nodo, gli utenti possono interagire con la rete in altri modi.
Freenet fornisce un'interfaccia HTTP per la navigazione di freesites, ma non è un proxy per il World Wide Web. Come per i servizi nascosti della rete Tor, può essere utilizzata solo per accedere ad un contenuto che è stato precedentemente inserito nella rete.
L'obbiettivo di Freenet è quello di offrire la libertà di parola e l'anonimato tentando di proteggere sia le persone che inseriscono i dati in rete(uploader) e quelle che cercano di recuperare tali dati(downloader). A differenza dei sistemi di file sharing, non è necessario che l'uploader rimanga in rete dopo il caricamento di un file o di un gruppo di file. Questo perchè durante il processo di uploading, i file sono suddivisi in blocchi e memorizzati su altri computer della rete. Questi file possono successivamente essere recuperati tramite una chiave associata ad essi. Durante il download, i blocchi vengono scaricati dai singoli computer sparsi nella rete e riassemblati.
Ogni nodo della rete Freenet contribuisce ad essa offrendo banda di connessione ed una porzione del proprio disco rigido (denominata datastore) per memorizzare i file che vengono mantenuti automaticamente o eliminati a seconda della loro popolarità. I file vengono criptati, così generalmente l'utente non può facilmente scoprire ciò che è presente nel suo datastore e di conseguenza non può nemmeno essere ritenuto responsabile per esso.
Come diretta conseguenza del requisito di anonimato il nodo che richiede un contenuto non si collega direttamente al nodo che lo possiede, ma la sua richiesta viene indirizzata attraverso nodi intermediari. Nessuno di questi nodi intermediari conosce chi ha effettuato la richiesta o chi possiede il contenuto richiesto. Come risultato, la larghezza di banda totale richiesta dalla rete per trasferire un file è maggiore rispetto ad altri sistemi.
I fondatori di Freenet[39] sostengono che la vera libertà di parola venga solo con

il vero anonimato e che i benefici della rete siano superiori ai suoi usi negativi. Freenet e tutte le reti di questo tipo sono da sempre al centro di discussioni e critiche. La maggior parte di queste si riferiscono al fatto che la stessa tecnologia che permette ai suoi utenti di comunicare le proprie idee in maniera anonima e sicura, può anche essere usata per pubblicare materiale illecito. L'opinione di Freenet è che la libertà di parola, di per sé, non è in contraddizione con una qualsiasi altra considerazione, l'informazione non è il reato.
Sebbene molte nazioni censurino le comunicazioni, tutte hanno una caratteristica comune: qualcuno deve decidere cosa tagliare e cosa mantenere, cosa considerare offensivo e cosa no. Freenet è una rete che elimina per chiunque la possibilità di imporre la sua scala di valore sugli altri; in pratica, a nessuno è permesso di decidere cosa sia accettabile. La tolleranza verso le opinioni altrui è fortemente incoraggiata ed agli utenti è richiesto di non prestare attenzione ai contenuti che non approvano. É difficile conoscere con esattezza l'impatto di Freenet all'interno di nazioni che operano una forte censura sui cittadini per via del funzionamento interno della rete. Ad esempio in Cina il gruppo Freenet-China ha tradotto Freenet in cinese e ha iniziato la sua distribuzione, tramite supporti di memoria fisica, raggiungendo nel 2002 diverse migliaia gli utenti.
Freenet è stato scaricato più di 2 milioni di volte da quando il progetto è iniziato.

6.1.3 Specifiche tecniche

Freenet[36] [38], a differenza delle altre P2P, oltre a trasmettere i dati tra i nodi, li memorizza, lavorando come un enorme cache distribuita. Per raggiungere questo obiettivo, ogni nodo alloca una certa quantità di spazio su disco per memorizzare i dati (questo è configurabile dall'operatore nodo ed è tipicamente di diversi GB). I file in Freenet sono in genere divisi in più piccoli blocchi, alcuni di questi sono duplicati per fornire ridondanza. Ogni blocco è gestito in maniera indipendente e quindi un singolo file può avere parti memorizzate su molti nodi diversi. Mentre l'inserimento di un dato nella rete è semplice, non si può dire lo stesso per l'eliminazione. A causa della natura anonima di Freenet, il nodo che ha eseguito l'upload del file ed il nodo che possiede uno dei blocchi di questo file è sconosciuto. L'unico modo per rimuovere i dati è quello di aspettare che gli utenti non lo richiedano più.
Tipicamente, un computer host sulla rete esegue il software che agisce come un nodo, e si connette ad altri host che eseguono lo stesso software per formare una rete distribuita di dimensione variabile. Alcuni nodi appartengono ad utenti finali che eseguono il download di un file, altri invece servono solo per instradare i dati.

Tutti i nodi comunicano tra loro in modo identico, infatti non c'è una distinzione client/server. Non è possibile per un nodo classificarne un altro se non per la sua capacità di inserire e recuperare i dati associati ad una chiave. Freenet differisce perciò da molti altri P2P dove gli amministratori del nodo possono impiegare un sistema in cui gli utenti devono condividere una certa quantità di contenuti prima di poter scaricare.
Il protocollo Freenet è destinato ad essere utilizzato su una rete di topologia complessa, come Internet. Ogni nodo conosce solo un determinato numero di altri nodi che può raggiungere direttamente, ma ogni nodo può essere un vicino di qualsiasi altro. Non esiste nessuna gerarchia o altra struttura. Ogni messaggio viene instradato attraverso la rete passando da vicino a vicino fino a raggiungere la sua destinazione. Ogni nodo passa un messaggio ad un vicino senza sapere se sarà il prossimo ad inoltrare il messaggio ad un altro nodo o sarà la destinazione finale. In questa maniera viene protetto l'anonimato degli utenti. Ogni nodo mantiene

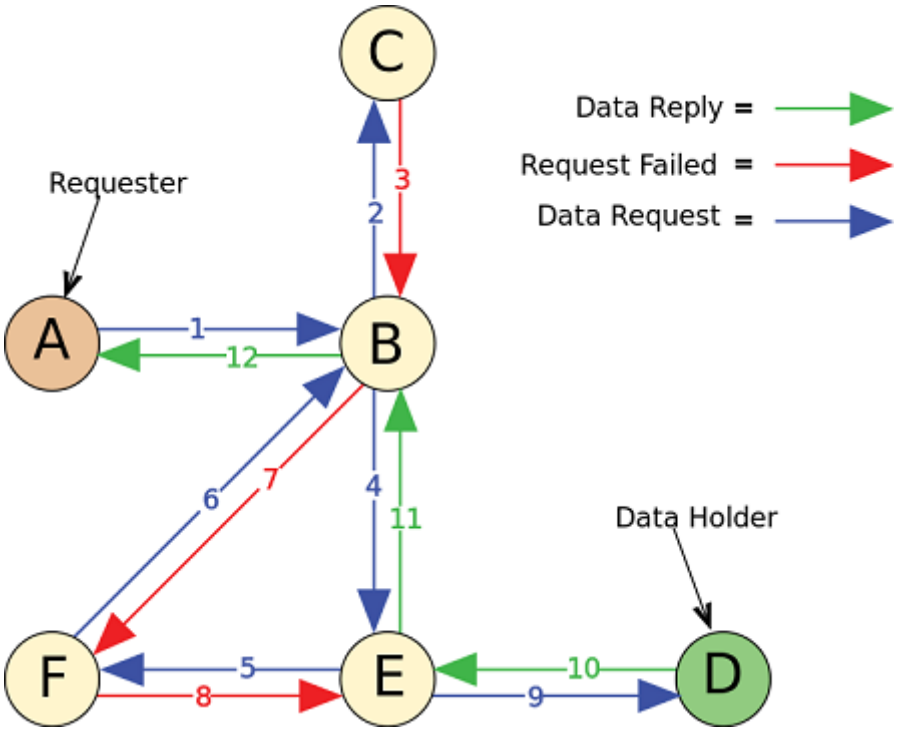

FIGURE 6.1: Esempio di funzionamento della rete Freenet

un archivio dati contenente i documenti associati a delle chiavi ed una tabella di routing che associa i nodi con la loro prestazioni nel recupero di chiavi diverse. Il protocollo Freenet utilizza un protocollo di routing basato su chiavi, simile a tabelle hash distribuite. Ogni nodo ha una posizione, che è un numero compreso tra 0 e 1. Quando viene richiesta una chiave, prima il nodo controlla l'archivio dati locale. Se non viene trovato, l'hash della chiave viene trasformato, e la richiesta

viene instradata al nodo cui posizione è più vicina alla chiave. Questo procedimento va avanti fino a quando si supera un certo numero di salti oppure non ci sono più nodi per la ricerca o il dato è stato trovato. In questo ultimo caso questo viene memorizzato nella cache su ogni nodo lungo il percorso. Quindi oltre a risparmiare larghezza di banda, questo rende i documenti più difficili da censurare in quanto non esiste un nodo di origine. In sostanza lo stesso processo viene utilizzato per inserire un documento nella rete. Questo procedimento funziona solo se i nodi sono raggruppati nel modo giusto. Freenet presuppone che la Darknet sia una piccola rete mondiale, e i nodi cerchino continuamente di scambiare posizione (utilizzando l'algoritmo Metropolis-Hastings) al fine di minimizzare la distanza con i loro vicini. Se la rete è effettivamente una piccola rete mondiale, Freenet dovrebbe trovare i dati con un numero di salti dell'ordine di $O[\log(n)]^2$. Tuttavia, non è garantito che i dati vengano trovati.
Alla fine, sia che si trovi il documento, sia che venga superato il limite di salti, il nodo terminale invia una risposta che fa il suo ritorno al mittente lungo il percorso indicato dal record di richieste in attesa dei nodi intermedi. I nodi intermedi possono scegliere di memorizzare nella cache il documento lungo la strada.
Inizialmente, le posizioni sono distribuite in maniera casuale (sia Opennet che Darknet). Ciò significa che il routing delle richieste è essenzialmente casuale. Siccome i diversi nodi non saranno d'accordo su dove inviare una richiesta, data una chiave, i dati appena inviati in un Freenet saranno distribuiti a caso. Successivamente i nodi cercano di avvicinarsi ad altri nodi che gestiscono dati con chiavi simili riorganizzando la rete in una struttura cluster distribuita. Ci saranno probabilmente diversi gruppi di nodi in tutta la rete e documenti replicati numerose volte a seconda di quanto vengono utilizzati. Questa è una sorta di rottura spontanea della simmetria, in cui da uno stato inizialmente simmetrico si passa ad una situazione altamente asimmetrica, con nodi che tendono a specializzarsi in dati con chiavi simili. Ci sono forze che tendono a causare il clustering e forze che tendono a rompere cluster (caching locale dei dati di uso comune). Queste forze saranno diverse a seconda di come i dati vengono utilizzati, in modo da distribuire quelli che raramente sono utilizzati in pochi nodi e oggetti di uso frequente diffusi ampiamente in tutta la rete. Questo mirroring automatico contrasta i tempi di sovraccarico web, e permette ad una rete di dimensione n di recuperare un documento in media con una complessità di $\log(n)$.
Le chiavi sono degli hash e non hanno alcuna nozione di vicinanza semantica quando si parla di key closeness (vicinanza tra chiavi). Quindi non ci sarà alcuna correlazione tra la vicinanza della chiave e la popolarità dei dati evitando cosi colli di bottiglia causati da argomenti preferiti.
Ci sono due principali tipi di chiavi in uso su Freenet, Key Content Hash (CHK) e

il Signed Subspace Key (SSK). Un sottotipo di SSK è l'Updatable Subspace Key (USK) che aggiunge la possibilità di aggiornare in maniera sicura i contenuti. Il CHK è un hash SHA-256 (Secure Hash Algorithm, un insieme di funzioni crittografiche di hash) di un documento: il nodo può controllare se la trasmissione è avvenuta correttamente, confrontando chiave e documento. Questo tipo di chiave viene usato per le trasmissioni di dati e per la loro lettura. Un nodo malevolo che volesse modificare i dati all'interno di un documento verrebbe immediatamente scoperto dal nodo successivo, grazie al controllo della chiave.
Le chiavi SSK sono basate sul concetto di crittografia asimmetrica, in particolare sul sistema Digital Signature Algorithm (DSA). I documenti inseriti con questo tipo di chiave vengono firmati dagli autori, in modo che tutti possano verificare l'integrità del documento che leggono. Le chiavi possono essere usate per creare uno pseudonimo all'interno di Freenet mantenendo l'anonimato, e permettono l'aggiornamento dei documenti solo da parte di chi li ha inseriti.

6.1.4 Darknet VS OpenNet

Come[36] detto precedentemente, dalla versione 0.7, Freenet ha iniziato a supportare due modalità: darknet e OpenNet. Le connessioni OpenNet vengono eseguite automaticamente da nodi con abilitata tale modalità e sono facili da usare. La modalità darknet prevede che la connessione sia stabilita manualmente tra gli utenti che si conoscono e si fidano l'un l'altro. Inoltre è più difficili da usare, ma allo stesso tempo è più sicura contro gli attacchi sulla rete e possono rendere difficile determinare se un utente usa Freenet o meno.
Una rete è detta scalabile se le sue prestazioni non si deteriorano a prescindere dalla grandezza. All'interno della rete Freenet i dati possono essere trovati con $O[\log(n)]^2$ salti su una piccola rete mondiale, quando si ignora il caching che potrebbe migliorare la scalabilità per i contenuti popolari. Tuttavia, è difficile misurare la scalabilità senza una rete di dimensioni elevate. Inoltre, le funzionalità di Freenet, rendono difficile effettuare un'analisi precisa e dettagliata delle prestazioni. Tutt'ora la scalabilità di Freenet è in fase di valutazione. L'innovazione principale della versione 0.7 è quindi la possibilità di offrire una darknet globalmente scalabile, capace di supportare milioni di utenti al contrario delle limitazione di altre darknet. La scalabilità di Freenet è resa possibile dal fatto che i rapporti umani tendono a formare reti di piccole dimensioni, una proprietà che può essere sfruttata per trovare brevi percorsi tra due persone. Inoltre, l'algoritmo di routing è in grado di instradare su una miscela di connessioni OpenNet e darknet, permettendo alle persone che hanno solo un paio di amici di utilizzare la rete ottenendo

prestazioni sufficienti e ricevendo alcuni dei vantaggi di sicurezza delle connessioni darknet. Questo significa anche che i piccoli utenti delle darknet insieme alle connessioni OpenNet sono pienamente integrati in tutta la rete Freenet, consentendo a tutti gli utenti di accedere a tutti i contenuti.

6.2 anoNet

Anonet è stata creata nel 2005 da alcuni sviluppatori di MetaNet, un progetto analogo, ma defunto. Di questi sviluppatori non si conosce nulla a parte che erano un gruppo di volontari stanchi di come stava evolvendo Internet.
Le ragioni di tale insoddisfazione erano dovute all'inevitabile tracciamento su Internet, in quanto ogni indirizzo IP viene associato al nome dell'utente. Quasi tutti i governi occidentali si sono lasciati tentare da questa possibilità negli ultimi anni e dopo l'attacco al World Trade Center, molti governi si sono sentiti addirittura in dovere di sorvegliare la rete tracciando gli utenti. Le limitazioni introdotte da questa sorveglianza hanno reso impossibile la visione di Internet come un posto in cui ognuno potesse dire e fare ciò che voleva. Di fronte a questa insoddisfazione, molti gruppi, come gli sviluppatori di anoNet, hanno iniziato a creare le loro reti private.

6.2.1 Cos'è Anonet

AnoNet[40] [41] è una rete F2F (Friend to Friend, una rete P2P a cui possono collegarsi solo le persone che dispongono di una apposita chiave crittografica) decentralizzata che utilizza il software OpenVPN, molto diffuso per creare reti private virtuali VPN (Virtual Private Network), per collegare tra loro i diversi nodi.
Quando due persone decidono di mettere in comunicazione i loro computer via OpenVPN, devono scambiarsi le proprie chiavi RSA pubbliche. OpenVPN usa queste chiavi per creare un tunnel, ovvero un canale di comunicazione crittografato tra i nodi di Anonet. Questo tunnel viene nascosto dal canale di comunicazione gia esistente tra i due nodi collegati ad Internet rendendo cosi il traffico di Anonet come un flusso di dati senza senso.
Se Anonet si limitasse ad usare OpenVPN, sarebbe semplicemente una grande VPN. Per garantire l'anonimato infatti vengono svincolati gli indirizzi IP che identificano i nodi della rete dalla identità dei loro utenti. Su Internet ogni computer

è identificato da un indirizzo IP che viene assegnato in modo gerarchico. Al livello più alto, c'è un apposito organismo internazionale, chiamato IANA, che assegna ai provider ed alle aziende che ne fanno richiesta dei blocchi di indirizzi contigui. I provider e le aziende, a loro volta, assegnano gli specifici indirizzi IP ai singoli computer.
Su Anonet, le cose funzionano diversamente. Gli sviluppatori hanno deciso di usare al suo interno gli indirizzi IP di un blocco compresso tra 1.0.0.0 e 2.255.255.255 che su Internet non può essere utilizzato perché è stato riservato da IANA per usi futuri. In questa maniera è possibile identificare circa 16 milioni di computer e gli indirizzi usati non possono entrare in collisione con quelli di Internet offrendo la possibilità di essere collegati, nello stesso momento, ad entrambe le reti.
All'interno di Anonet, non esiste una autorità che assegna gli indirizzi IP. Ogni utente può recarsi presso un apposito database e riservare, per i propri usi, uno o più blocchi di indirizzi chiamati sottoreti perché identificano appunto delle intere sottoreti di Anonet. Da questi blocchi vengono successivamente distribuiti almeno due indirizzi per ciascun utente: uno da usare per navigare in modalità client su Anonet e l'altro per offrire eventuali servizi in modalità server.
Dato che Anonet usa un suo sistema di assegnazione degli indirizzi IP, al suo interno deve usare un suo sistema di routing e di risoluzione dei nomi. Per il routing usa un programma chiamato Quagga che deve essere installato su qualunque nodo della rete, mentre per la risoluzione dei nomi usa lo stesso software di Internet (BIND), ma un diverso database di indirizzi.
Anonet è una rete anonima perchè non è possibile sapere a chi appartiene un certo indirizzo IP ed il computer ad esso collegato. Infatti riuscire a risalire ad un utente partendo da un IP non registrato non è possibile e riuscire ad indovinare un IP tra centinaia di combinazioni possibili è altrettanto impossibile. L'unica eccezione a questa regola sono i peer direttamente collegati tra loro. Visto che ogni nodo deve conoscere l'indirizzo IP di Internet dei suoi peer per potersi collegare con OpenVPN allora potenzialmente è anche in grado di risalire alla loro identità.
A parte le features per aumentare sicurezza ed anonimato, Anonet usa la stessa architettura e lo stesso software di Internet.
Inoltre è possibile creare piccole reti private e collegarle ad Anonet attraverso un nodo o un gateway formando un'unica grande darknet. La rete quindi può essere usata come hub e come piattaforma di collegamento tra molte piccole comunità.

6.3 StealthNet

StealthNet[42] [43] è un client P2P open source basato sulla rete anonima RShare, sviluppato da parte di un team di sviluppatori tedeschi della comunità Planet Peer, distribuita con licenza libera sotto GNU GPL(Licenza Pubblica Generale). Nelle prime fasi di rilascio, StealthNet era conosciuto come RShare Community Edition (RShare CE).
StealthNet opera sulla rete RShare, il quale protocollo è stato sviluppato da Lars Regensburger con l'obbiettivo di raggiungere una forte anonimità e sicurezza pur riuscendo ad avere un'accettabile tasso di velocità nello scambio di risorse. Per evitare l'analisi del traffico la rete RShare cripta il traffico tra due nodi, il quale viene instradato attraverso altri nodi della rete stessa e criptato tra nodo e nodo offrendo un alto grado di anonimato. Per esempio, un nodo A che trasmette dati ad un nodo B può assumere molti significati. Il nodo A potrebbe essere il mittente del trasferimento, quindi starebbe inviando parti di un file che possiede in condivisione al nodo B. Ma in questo caso non è detto che il nodo B sia l'effettivo destinatario, infatti potrebbe inviare tali dati ad un altro nodo che collega il destinatario. Non è detto nemmeno che il nodo A sia l'effettivo mittente, infatti potrebbe anche essere un semplice nodo che si trova all'interno del percorso di routing che collega mittente e destinatario (principio di negabilità plausibile). Inoltre, nel protocollo della rete RShare non si utilizzano gli indirizzi IP per identificare univocamente i nodi in linea, ma degli appositi RShare ID non associabili all'indirizzo IP del medesimo nodo.
I client RShare si affidano a delle Webcache che tengono traccia dell'indirizzo IP, e la relativa porta TCP, dei nodi in linea. Nient'altro viene memorizzato riguardo ad un singolo nodo. Ciò avviene per supportare i client nella ricerca dei nodi col quale stabilire una connessione diretta. Le Webcache comunque non memorizzano i file condivisi da un nodo e non servono per la funzione di ricerca di file o di fonti per un download, diversamente da come sono pensati e gestiti i server per le reti P2P non anonime. Ogni utente può utilizzare il proprio host come una Webcache della rete RShare, basta avere installato sul proprio sistema operativo un webserver PHP e MySQL.
Alcuni vantaggi riferiti a quando StealthNet è nata sono:

- **Facilità di utilizzo:** basta infatti un'installazione e l'abilitazione del port fowarding se si possiede un router;
- **Moderna ed user-friendly interfaccia grafica**;
- **Alta velocità di trasferimento**;

- **Anti-flood-measure**: sistema che impedisce il sabotaggio della rete inviando pacchetti di dati con contenuto insignificante o non conforme al protocollo utilizzato;
- **Possibilità di installazione su tutte le piattaforme**.

L'ultima versione di StealthNet risale al 15 marzo 2011, quindi il progetto può ritenersi morto.

6.4 I2P

I2P[44] , originariamente chiamata Invisible Internet Project, è un software libero e Open Source per la realizzazione di una rete anonima disponibile per desktop, sistemi integrati e Android. I2P è ancora in sviluppo beta, quindi non è ancora ritenuta idonea per gli usi che necessitano di un anonimato forte. I2P[45] nasce nel 2003 come proposta di modifica di Freenet ed è in continuo aggiornamento cercando di rendere gli attacchi sempre più difficili da lanciare oltre ad aumentarne continuamente il livello di anonimato.
La rete è costituita da un insieme di nodi (router) con un numero di cammini virtualib (tunnel) unidirezionali in ingresso e uscita. Ogni router, identificato da un RouterIdentity, comunica con altri router tramite lo scambio di messaggi su protocolli esistenti come TCP o UDP. Le applicazioni client hanno il proprio identificativo crittografico (destinazione) che consente di inviare e ricevere messaggi. Questi client possono connettersi a qualsiasi router ed autorizzare l'assegnazione temporanea (lease) di alcuni tunnel che saranno utilizzati per l'invio e la ricezione di messaggi attraverso la rete. I2P ha il proprio database (utilizzando una modifica dell'algoritmo Kademlia) per il routing distribuito.
Offrendo un livello di rete anonimo, in cui le applicazioni possono comunicare con un alto grado di sicurezza, ha dato il via alla realizzazione di diversi software per I2P. Di seguito sono riportati alcuni esempi:

- **I2PTunnel:** è un'applicazione incorporata in I2P che permette di fare da ponte TCP/I2P permettendo ai client di inoltrare stream TCP nella rete I2P, nonché ricevere stream dalla rete e inoltrarle verso uno specifico indirizzo TCP/IP.
- **SAM:** è un protocollo che consente ad una applicazione client, scritta in qualsiasi linguaggio, di comunicare in I2P, utilizzando un'interfaccia basata su socket per il router I2P;

- **Bittorrent:** Qualsiasi client Bittorrent può sfruttare la rete I2P per le sue funzionalità, semplicemente configurando il client ed il browser;
- **iMule:** è un client multipiattaforma della famiglia *Mule, adattato al network I2P e modellato sul client aMule;
- **I2Phex:** è la versione modificata di Phex che sfrutta la rete Gnutella, adattata per funzionare con la rete I2P. Il Progetto è ancora in via di sviluppo;
- **I2P-Bote**: è un plugin di posta elettronica per la memorizzazine sicura dei messaggi;
- **Altro**: come applicazioni per blogging e forum, host di siti web e chat in tempo reale

6.4.1 Confronto tra I2P e Tor

Le due principali differenze tra Tor/Onion-Routing e I2P sono legate alla differenze nel modello di rischio contro le minacce ed il design dei nodi d'uscita (outproxy). Inoltre Tor utilizza l'approccio basato su directory fornendo gestione centralizzata della visione globale della rete, al contrario del database distribuito e la selezione dei peer di I2P.
Alcuni benefici ottenuti usando Tor su rete I2P sono:

- Base di utenti molto più grande;
- Ha già risolto alcuni problemi di scalabilità che I2P deve ancora affrontare;
- Ha significativi finanziamenti;
- Abbastanza grande da doversi adattare a blocchi e tentativi DOS;
- Progettato e ottimizzato per il traffico di uscita, con un grande numero di nodi di uscita;
- Più efficiente nell'uso della memoria;
- I nodi del client Tor hanno un richiesta di banda molto bassa;
- Un controllo centrale riduce la complessità di ogni nodo e resiste più efficacemente ad attacchi di tipo Sybil;
- Un nucleo di nodi ad alta capacità fornisce un throughput più elevato ed una minore latenza;

- Ha più sviluppatori, includendone diversi che sono sovvenzionati.

Alcuni benefici ottenuti invece usando I2P sopra Tor sono:

- Progettato e ottimizzato per servizi nascosti, che sono molto più veloci che in Tor;
- Completamente distribuito e auto-organizzante;
- I peer sono selezionati continuamente in funzione delle prestazioni;
- Piccola abbastanza da non essere stata bloccata o attaccata da DOS;
- Peer-to-peer friendly;
- Load balancing trasparente dei messaggi su più peer, piuttosto che un singolo percorso
- Resilienza contro i guasti tramite l'esecuzione di più tunnel parallele, più rotazione dei tunnel;
- Scala le connessioni di ciascun cliente a O(1) invece di O(N);
- Gallerie unidirezionali invece di circuiti bidirezionali, raddoppiando così il numero di nodi che un peer deve compromettere per ottenere le stesse informazioni;
- Protezione contro la rivelazione di attività di un client, anche quando un attaccante partecipa nel tunnel, perché i tunnel sono utilizzati per più di un semplice passaggio di messaggi end-to-end;
- I tunnel hanno breve durata;
- Le API I2P sono progettate specificamente per l'anonimato e la sicurezza, mentre SOCKS è progettato per la funzionalità;
- Tutti i peer partecipano al routing per gli altri;
- Trasporti sia di tipo TCP che UDP;
- L'overhead di banda basso;
- Integrato meccanismo di aggiornamento automatico.

7. Test finali

7.1 Un viaggio nel Deep Web

Doppio click sull'icona di Tor ed eccoci catapultati in un nuovo mondo, il Deep Web. Più che oscuro, è estremamente mobile. Siti che vanno e vengono, mercati che si aprono e poi spariscono nel giro di pochi giorni, forum che finiscono offline a seguito di attacchi informatici o che sono momentaneamente down per un'emergenza, una vulnerabilità. Per vedere se siamo effettivamente anonimi ci

FIGURE 7.1: Schermata iniziale di Tor

colleghiamo a ip-check.info e guardiamo il risultato del test sull'anonimato. L'IP che ci è stato assegnato è 95.211.205.151 proveniente da Noord-Holland, Amsterdam. Possiamo inoltre eseguire il traceroute, osservare sulla mappa la nostra "nuova" locazione ed anche avere informazioni sul provider ed il dominio.
Inoltre il test ci mostra il nostro livello di sicurezza assegnando ad ogni attributo

un colore: verde se il relativo valore è conforme alla migliore impostazione possibile, giallo che indicano un possibile problema e rosso che evidenzia problemi di privacy estremamente critici che possono portare alla immediata identificazione. Possiamo anche trovare informazioni riguardanti il nostro relay, nel nostro caso

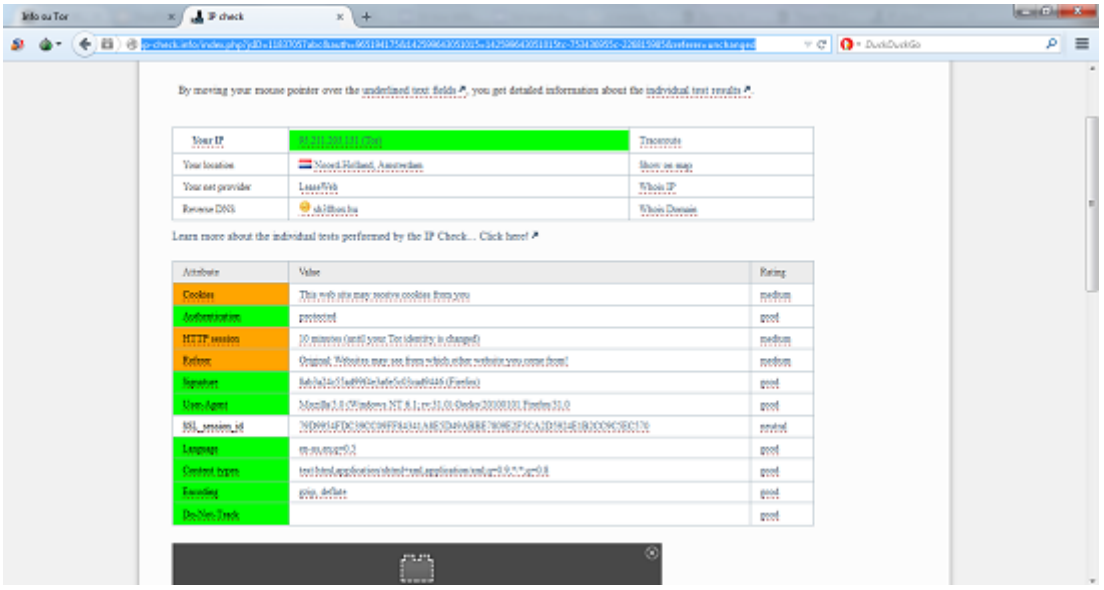

FIGURE 7.2: Dettagli anonimato con Tor

Atlas. Oltre alla configurazione ed alle proprietà, è interessante osservare il tempo che è rimasto online. Il Deep Web è immenso e per iniziare a muoversi bisogna

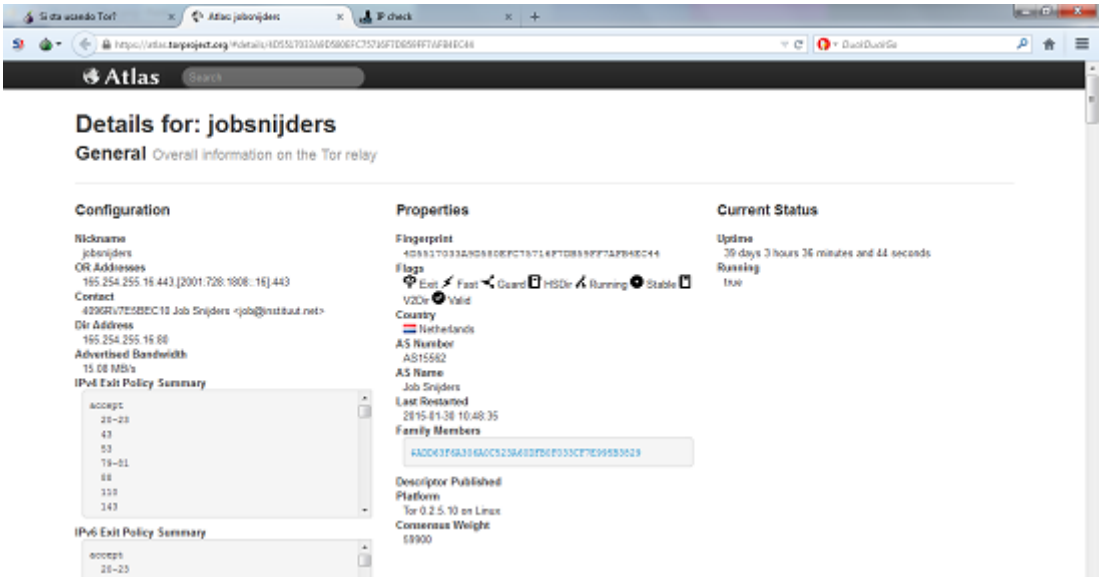

FIGURE 7.3: Dettagli relay Atlas

partire da Hidden Wiki, una pagina di Wikipedia dichiaratamente priva di censure

che cataloga alcuni dei siti più importanti. Wikileaks è già molto famoso, ma esiste anche lo Strongbox del New Yorker, un sistema creato dal giornale "per dare agli informatori completa sicurezza e anonimato." Un'altra interessante risorsa indipendente è Kavkaz, un sito di notizie del Medio Oriente disponibile in russo, inglese, arabo e turco. Dopo il sequestro di Silk Road da parte dell'FBI è nor-

FIGURE 7.4: Wikileaks su Tor

male pensare che il fenomeno marketplace sia morto. Tutto ciò è falso. Siti come DeepDotWeb.com o DNSstats.net ci forniscono dei dati per monitorarli. Uno dei parametri più interessanti è quello che valuta il tempo in cui i vari siti sono online e offline, ma ci sono altre informazioni in tabelle comparative delle diverse funzioni offerte da simili mercati. I mercati neri sono tanti e tra questi mi ha in-

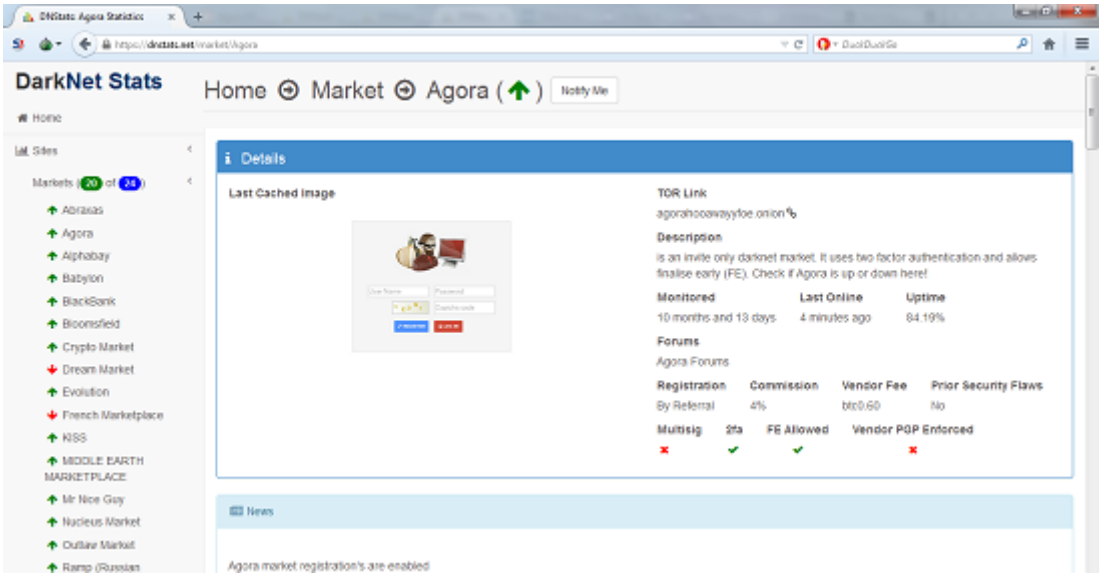

FIGURE 7.5: Dettagli su Agorra un famoso black market

curiosito Babylon, interamente in Italiano. Qui dentro possiamo trovare qualsiasi cosa che vada dall'attacco hacker alla droga, dalle armi alle monete contraffatte.

Non manca un forum dove discutere su diversi argomenti illegali o vantarsi delle proprie truffe.

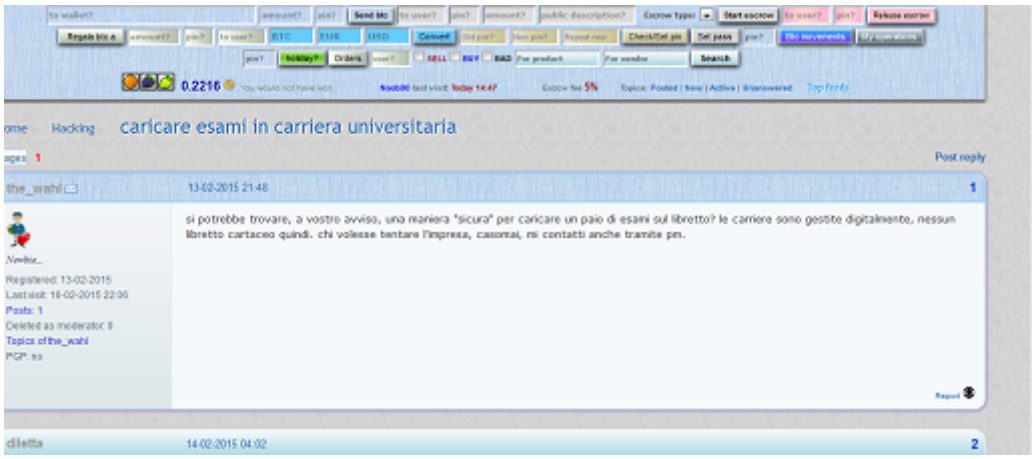

FIGURE 7.6: Domanda posta sul forum di Babylon. Avrà ottenuto risposta?

Se questa parte del Deep Web attira sicuramente molta attenzione, bisogna però ricordare che c'è molto di più. Esistono alcuni motori e directory che cercano di gettare luce su questa internet oscura. Uno di questi si chiama Ahmia.fi, lo gestisce lo sviluppatore finlandese Juha Nurmi, e indicizza circa 1300 siti nascosti sulla rete Tor. Fra questi ci sono anche tanti forum di discussione, siti di attivismo, di informazione, luoghi dedicati a strumenti per l'anonimato e le criptomonete.
L'impressione che si ha navigando nel Deep Web è quella di smarrimento. Il problema è che richiede tempo, applicazione, un po di conoscenze, ma sopratutto un obbiettivo preciso verso cui puntare. Senza quest'ultimo si naviga in maniera casuale sfiorando solo un lieve strato di questo mondo affascinante, ma terribilmente pericoloso.

7.2 Trackography

Trackography è un tool, presentato dall'italiano Claudio Agosti, del tactical team, al 31esimo congresso del Chaos Computer Club (la più antica associazione europea di hacker sociali), che ci mostra come i siti di news offrono i nostri dati a società che non sono tenute a rispettare le leggi italiane sulla privacy. Infatti ogni volta che si visita una pagina web sia il sito, che le terze parti ad esso collegate, raccolgono informazioni sull'impronta digitale dell'utente, registrano i siti che ha navigato in precedenza e quelli che potrebbe visitare in futuro.
Trackography visualizza graficamente il tragitto dei nostri dati e le connessioni con le compagnie che ci tracciano per pubblicità mirata, profilazione e ricerche di mercato.
Tramite Trackography possiamo capire come e dove lasciamo le nostre tracce digitali e come la maggior parte di questi dati viene raccolta senza chiederci il consenso. Tutte queste tracce compongono la nostra ombra digitale che può essere intercettata ed osservata attraverso la rete di comunicazione.
Le leggi sulla privacy non sempre possono garantire una tutela tra un paese e l'altro. Inoltre molte forme di sorveglianza dei dati sono segrete e, se fatte per la sicurezza nazionale, possono ignorare le leggi. Infatti le agenzie governative possono controllare i gestori del servizio Internet o aver accesso diretto al traffico Internet analizzandolo per i propri scopi.
Trackography è utilizzabile collegandosi a trackography.org con un'interfaccia molto semplice ed user-friendly. Per iniziare bisogna selezionare uno stato.

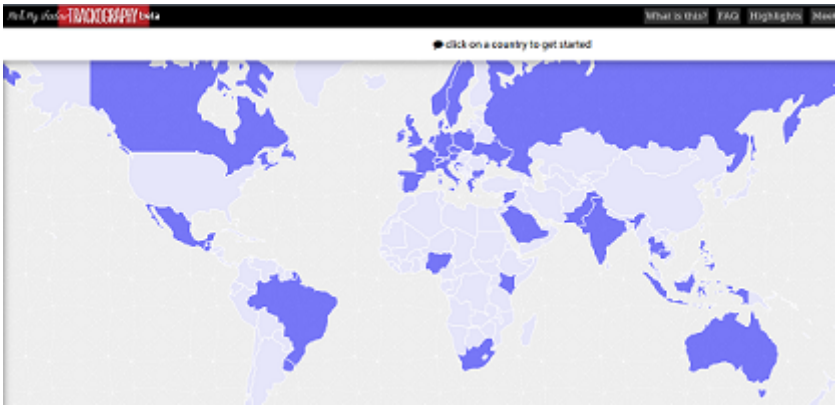

FIGURE 7.7: Selezione dello stato

Successivamente ci viene mostrato un pannello con le categorie di media sources, ovvero siti web di notizie.

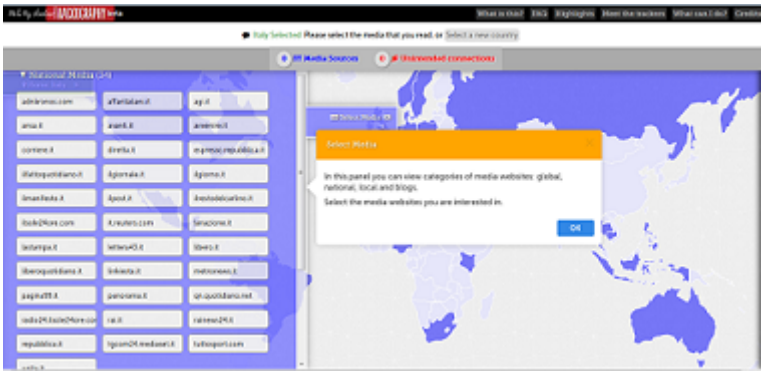

FIGURE 7.8: Selezione dei siti web di notizie

A questo punto il tool disegna sulla mappa tutto il traffico web disegnando con archi blu il percorso che rappresenta il traffico internet tra lo stato selezionato e il server dei siti web selezionati precedentemente. Gli archi rossi invece rappresentano le connessioni Internet delle compagnie che monitorano gli utenti che si collegano ai siti web selezionati. Quest'ultime, vengono chiamate unintended connections (connessioni non intenzionali) e sono conteggiate in alto. Selezioniamo per esempio "tgcom24.mediaset.it", il quotidiano online di News Mediaset.

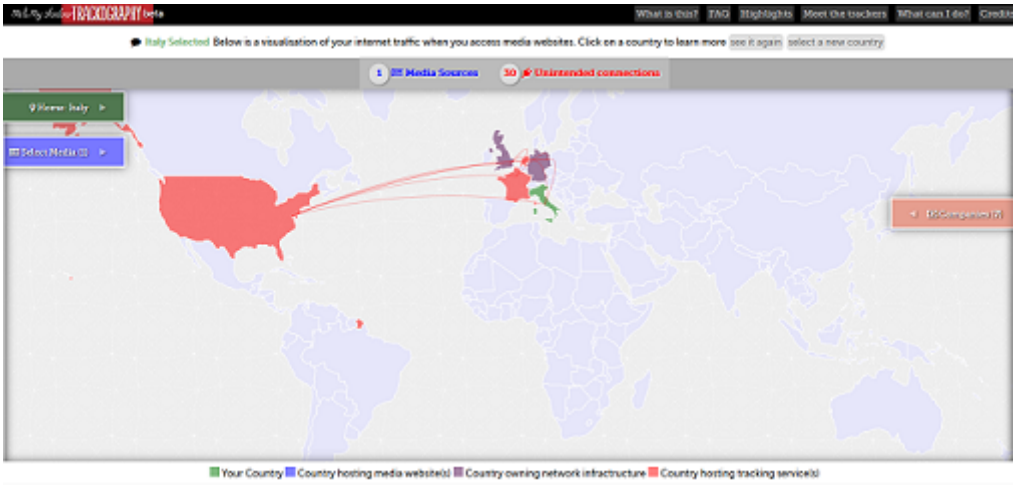

FIGURE 7.9: Disegno delle connessioni non intenzionali

Vediamo in verde (Italia) il nostro stato, in viola gli stati che possiedono le infrastrutture di rete ed in rosso gli stati che effettuano il tracking. Osserviamo che per la nostra selezione abbiamo ben sette connessioni non attese di cui tre sono verso l'America mentre il restante si trova in Europa. Google e comScore eseguono il nostro profiling, PubMatic trattiene i nostri dati per 270 giorni mentre per gli altri non sappiamo dire quanto tempo trattengono le informazioni.

Le connessioni non attese, consultando i vari singoli siti, sono in media sette e tra queste connessioni compaiono spesso il nome di colossi come Google, Facebook e Twitter. La sola lettura di news porta ad essere tracciati e risulta quindi molto importante difendersi e proteggere la propria privacy.

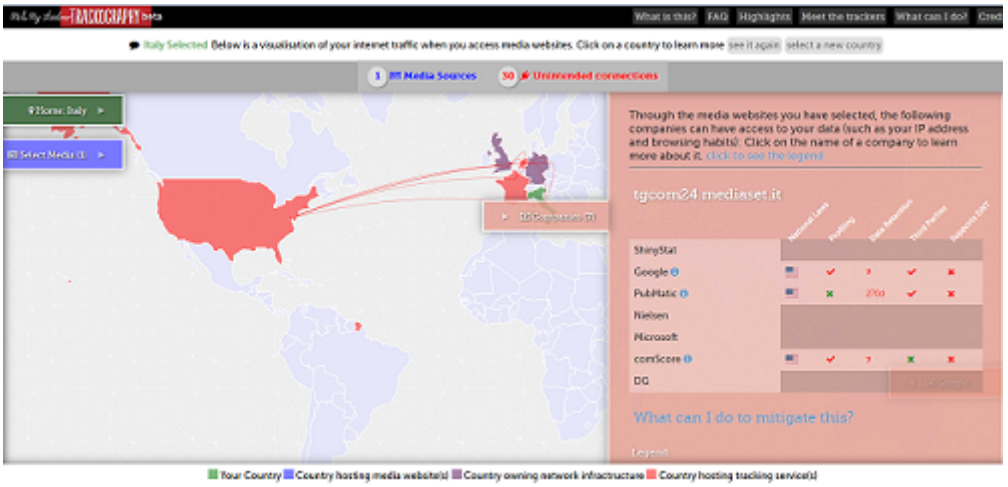

FIGURE 7.10: Elenco delle compagnie che ci tracciano

7.3 Trace my shadow

Quando usiamo i servizi digitali attraverso i nostri dispositivi, lasciamo inconsapevolmente tracce digitali. In alcuni casi, i nostri dati sono raccolti senza il consenso, mentre in altri casi scegliamo di consegnare i nostri dati a terzi. Attraverso tutte queste attività, lasciamo tracce digitali che danno luogo alla creazione di una nostra ombra digitale. Chiunque può potenzialmente osservare la nostra ombra digitale e questo risulta essere un problema.
Collegandoci a "myshadow.org" possiamo usare un semplice tool che ci illustra la traccia che lasciamo in rete.

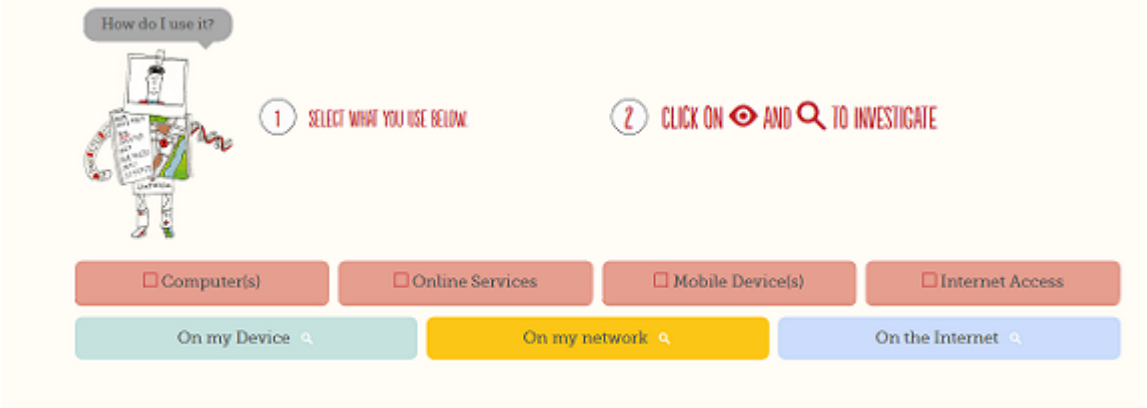

FIGURE 7.11: Trace my shadow

Il tool ci mostra la composizione della nostra ombra a seconda del tipo di computer o mobile device, servizi Internet ed accesso Internet.

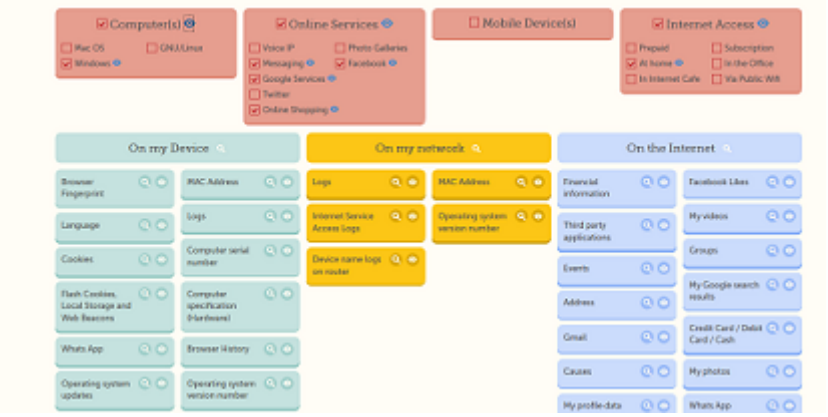

FIGURE 7.12: Trace my shadow

Abbiamo anche la possibilità di filtrare le tracce a seconda di ciò che abbiamo scelto precedentemente. Per ogni traccia a sua volta abbiamo una spiegazione ed alcune possibili soluzioni.

FIGURE 7.13: Trace my shadow

7.4 Ghostery, Adblock Plus e Privacy Badger

Ghostery è un'estensione del browser, ossia un'applicazione che si può aggiungere al proprio browser oppure al proprio smartphone/tablet, in grado di mostrare tutte le aziende che stanno tracciando la navigazione di un utente che visita un sito internet.
Ghostery cerca gli elementi di pagine esterne nelle pagine web visitate informandoti della presenza di tracker e le aziende da cui sono utilizzati. Se lo si desidera, è possibile avere ulteriori informazioni su tali aziende e scegliere di bloccare i loro tracker. Questo plugin effettua scansioni e prende decisioni per diversi tipi di elementi nella pagina, ed in alcuni casi, tale scansione potrebbe causare un lieve aumento del tempo di caricamento delle pagine. Tuttavia, se l'utente sceglie di bloccare gli script della pagina in media le pagine si caricano più velocemente.
Ghostery è compatibile con Chrome, Firefox, Safari, Opera, iPhone, iPad, iPod ed Android.
Se apriamo un sito, come ad esempio "rai.it", il plugin ci carica i tracker e la sorgente da cui sono generati. Possiamo osservare nel nostro caso ben 11 tracker suddivisi come annunci pubblicitari, widget, analisi. Possiamo successivamente

FIGURE 7.14: Ghostery in azione

andare ad analizzare ciascun singolo tracker osservando sopratutto le informazioni sulla privacy. Google+ Platform colleziona dati anonimi (informazioni sul browser, cookie, data/tempo ed altri)condivisi con terze parti, indirizzo IP, cronologia, PII (Personally identifiable information come il numero di cellulare) e tanto altro.

Nella pagina, oltre alle informazioni sulla privacy, possiamo osservare una breve descrizione del tracker ed una lista di siti che lo contengono.

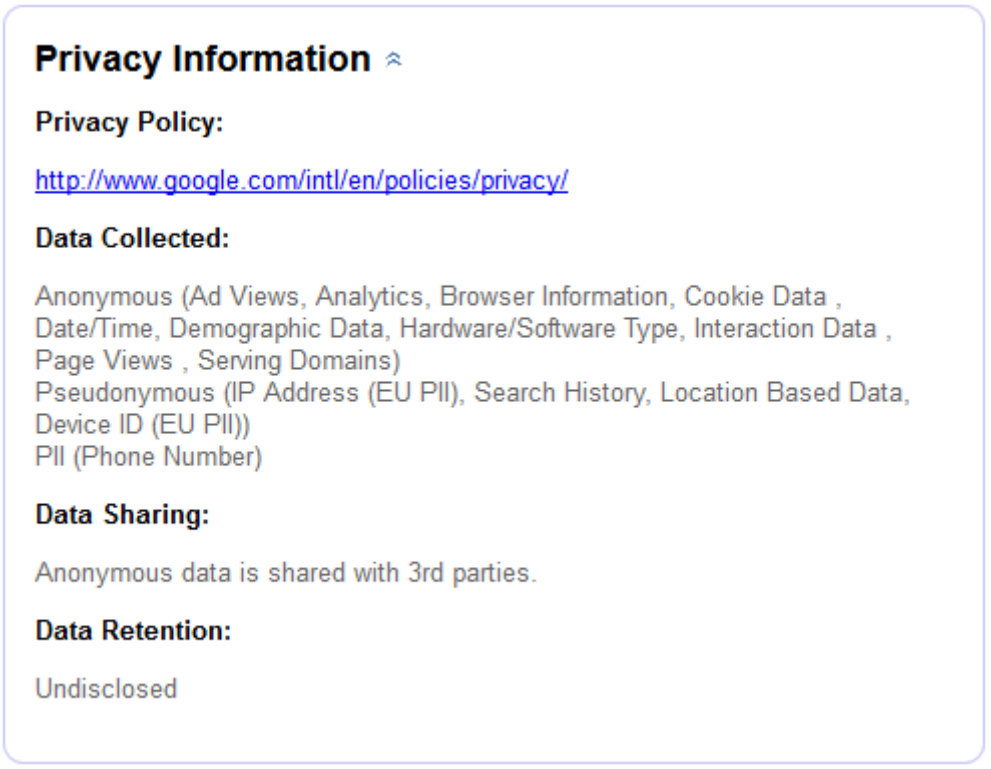

FIGURE 7.15: Informazioni sulla privacy di Google+ Platform

Allo stesso tempo è possibile osservare anche la sorgente d'origine del seguente tracker. Infine, ma non cosa meno importante, è possibile bloccare o consentire il tracker utilizzando i due pulsanti posizionati alla sua destra.

FIGURE 7.16: Sorgente tracker di Google+ Platform su rai.it

AdBlock Plus è un'estensione per browser oltre ad essere anche un'applicazione per Android, che permette di eliminare gli annunci pubblicitari (noti anche come

ads, dall'inglese advertisements) contenuti in molte pagine web con conseguente risparmio di tempo di download e di banda. Il filtraggio del materiale indesiderato avviene attraverso la ricerca del corrispondente URL all'interno di una lista di regole di riconoscimento: qualora l'URL soddisfi una qualsiasi di tali regole, l'oggetto può essere, a discrezione dell'utente, completamente ignorato oppure normalmente scaricato ma non visualizzato all'interno della pagina.
Adblock Plus in sé non ha alcuna funzionalità, nel senso che non blocca nulla finché non gli viene detto cosa fare dai suoi elenchi di filtri. Questi elenchi di filtri sono essenzialmente un vasto insieme di regole che decidono gli elementi di siti web da bloccare. Inoltre oltre al blocco della pubblicità permette di bloccare il tracking ed i malware.

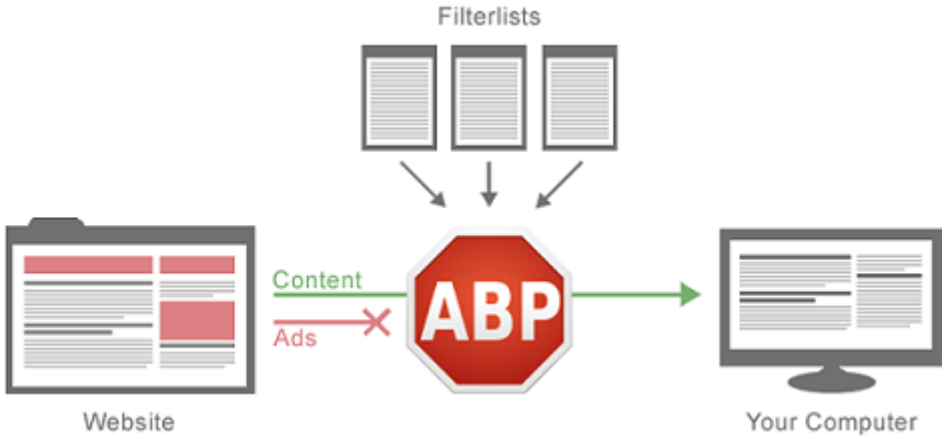

FIGURE 7.17: Funzionamento di Adblock Plus

Tramite l'apposito menu è possibile osservare gli elementi bloccati dal plugin. Collegandoci nuovamente a "rai.it" possiamo vedere lo stesso script di Google+ Platform oltre a tanti altri script ed annunci pubblicitari.

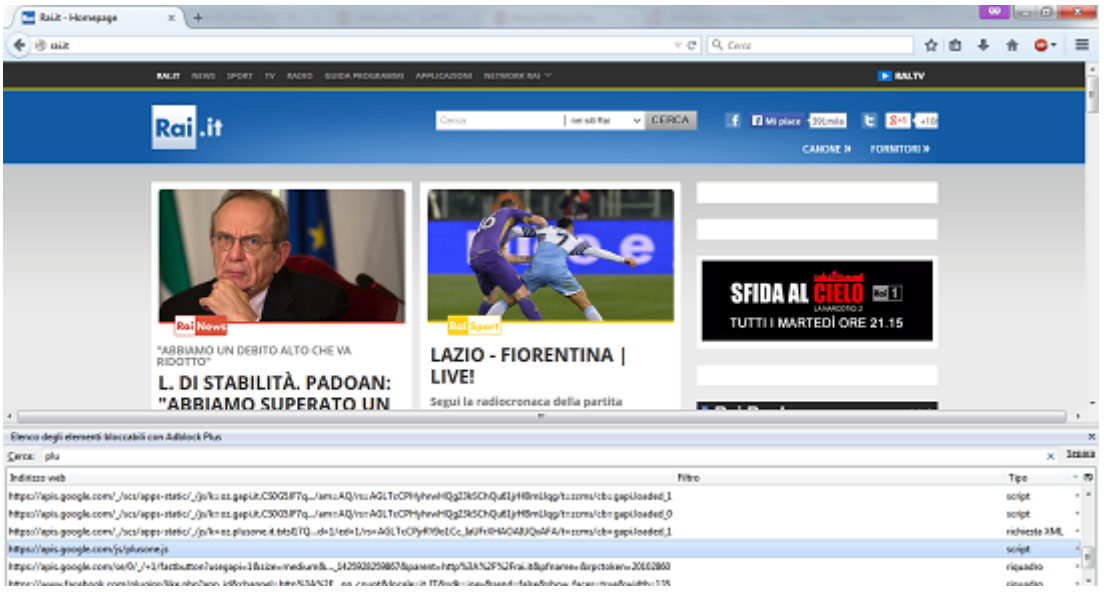

FIGURE 7.18: Lista di elementi bloccabili di Adblock Plus

Privacy Badger è un'estensione per browser che ferma, come i plugin di cui abbiamo parlato prima, gli annunci pubblicitari ed i tracker che costantemente monitorano tutto il traffico. É interessante osservare come nella stessa pagina "rai.it" abbia trovato una quantità maggiore di tracker rispetto a Ghostery, anche se poi non permette di informarsi su ciascuno di essi, ma consente solo il loro bloccaggio. Infine è interessante osservare [46] come su areweprivateyet.com vengano

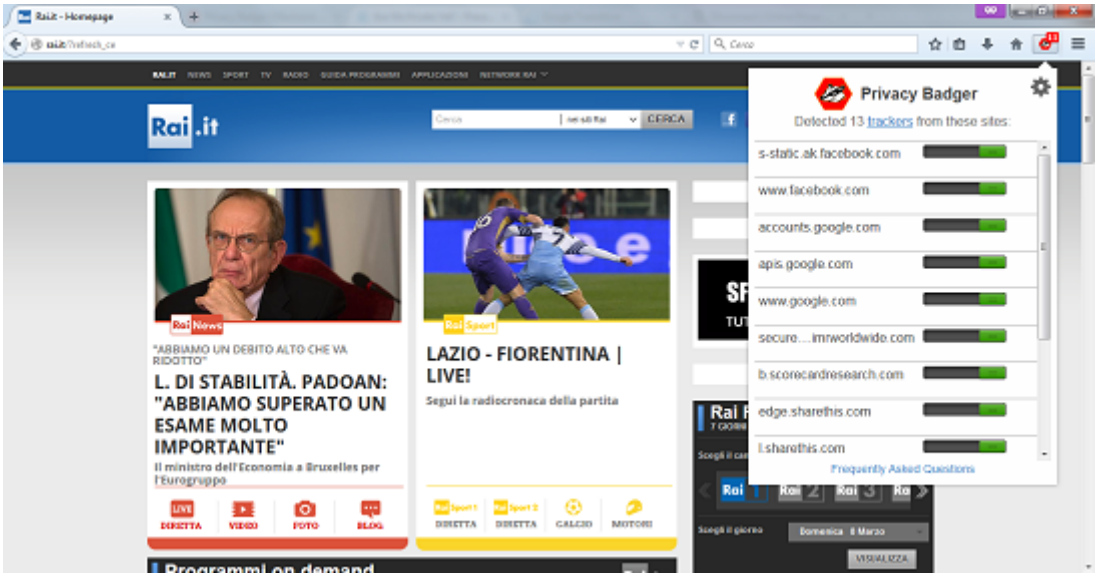

FIGURE 7.19: Tracker trovati da Privacy Badger

confrontate le capacità di diverse estensioni per browser di ridurre il monitoraggio in rete sulla base di:

- HTTP Requests;
- HTTP Set-Cookie Responses;
- Cookies;
- Local Storage.

Altre estensioni, osservabili nel tool, e degne di nota sono Disconnect e DoNotTrackMe.

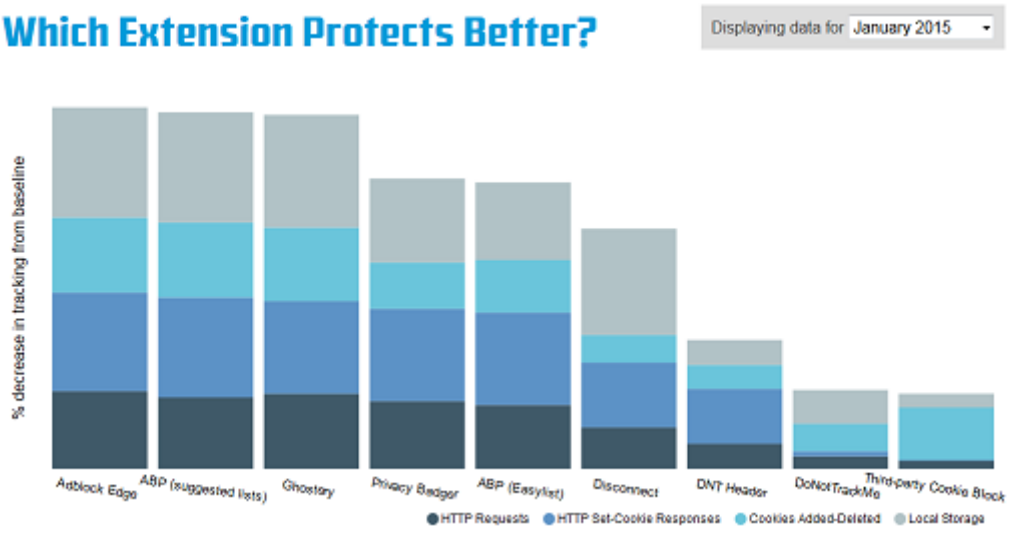

FIGURE 7.20: Confronto tra estensioni web

8. Conclusione

La privacy non è morta, ma è cambiata. E di conseguenza deve cambiare il nostro modo di affrontarne la tutela.
Un bambino che nasce oggi crescerà senza alcun concetto di riservatezza. Non saprà mai cosa significa aver un istante in privato, un pensiero non analizzato e non registrato. Questo è un problema perché la riservatezza conta.
La privacy è ciò che ci consente di decidere chi siamo e chi vogliamo essere. Bisogna decidere quanta fiducia riporre sia nella tecnologia che ci circonda, sia nel governo che la regolamenta trovando un equilibrio migliore.
É importante mettere fine alla sorveglianza di massa perché se si vuole sapere cosa pensano le persone, chiederglielo costa meno che spiarle.

Ringraziamenti

Grazie ai miei genitori che mi hanno sostenuto in questi anni.

Un ringraziamento particolare va a Silvia che mi ha aiutato a terminare questo percorso con più serenità.

Grazie al professore Omicini, sempre disponibile.

Infine, ma non meno importante, vorrei ringrazie tutti i Koa, compagni di innumerevoli avventure su Azeroth

Patryk Wojtowicz

Bibliography

[1] Wikipedia. Dato, . URL http://it.wikipedia.org/wiki/Dato.

[2] Wikipedia. Sistema informativo, . URL http://it.wikipedia.org/wiki/Sistema_informativo.

[3] Domo. Data never sleeps 2.0. URL http://www.domo.com/learn/data-never-sleeps-2.

[4] Wikipedia. Big data, . URL http://it.wikipedia.org/wiki/Big_data.

[5] Wikipedia. Nosql, . URL http://it.wikipedia.org/wiki/NoSQL.

[6] Wikipedia. Metadato, . URL http://it.wikipedia.org/wiki/Metadato.

[7] Web Com Lab. Metadati e location privacy. URL http://www.webcomlab.unimore.it/metadati.html.

[8] The Sidney Morning Herald. Internet will 'disappear', google boss eric schmidt tells davos. URL http://www.smh.com.au/digital-life/digital-life-news/.

[9] Google. Phisical web. URL https://google.github.io/physical-web/.

[10] Ngos call on governments to support the establishment of a un special rapporteur on the right to privacy, 2 Marzo 2015. URL https://www.eff.org/files/2015/03/03/in_support_a_sr_on_privacy_updated_version.pdf.

[11] Wikipedia. Dati sensibili, . URL http://it.wikipedia.org/wiki/Dati_sensibili.

[12] Wikipedia. Privacy, . URL http://it.wikipedia.org/wiki/Privacy.

[13] The Guardian. European counter-terror plan involves blanket collection of passengers' data. URL http://www.theguardian.com/uk-news/2015/jan/28/european-commission-blanket-collection-passenger-data.

[14] Fabio Chiusi. Chiusi nella rete. URL http://chiusinellarete-messaggeroveneto.blogautore.repubblica.it/2015/01/27/.

[15] Bloomber Business. France seeks to sanction web companies for posts pushing terror. URL http://www.bloomberg.com/news/articles/2015-01-27/france-seeks-to-sanction-web-companies-for-posts-pushing-terror.

[16] Parliamentary assembly of the Council of Europe. Mass surveillance. URL http://website-pace.net/documents/19838/1085720/20150126-MassSurveillance-EN.pdf/df5aae25-6cfe-450a-92a6-e903af10b7a2.

[17] Symantec. State of privacy report 2015. URL http://www.symantec.com/content/en/us/about/presskits/b-state-of-privacy-report-2015.pdf.

[18] Centre for International Governance Innovation e IPSOS. Cigi-ipsos global survey on internet security and trust. URL https://www.cigionline.org/internet-survey.

[19] Pew Research Center. Public perceptions of privacy and security in the post-snowden era. URL http://www.pewinternet.org/2014/11/12/public-privacy-perceptions/.

[20] Yougov. Poll results: Snowden. URL https://today.yougov.com/news/2014/03/28/poll-results-snowden.

[21] Philip N. Howard. Data breaches in europe: Reported breaches of compromised personal records in europe. URL http://cmds.ceu.edu/sites/cmcs.ceu.hu/files/attachment/article/663/databreachesineurope.pdf.

[22] Wikipedia. Cookie, . URL http://it.wikipedia.org/wiki/Cookie.

[23] Electronic Frontier Foundation. Secure messaging scorecard. URL https://www.eff.org/secure-messaging-scorecard.

[24] Whisper System. Official site. URL https://whispersystems.org/.

[25] Wikipedia. Tor (software), . URL http://it.wikipedia.org/wiki/Tor_%28software%29.

[26] Internauta 37. Darknet o deep web: il lato oscuro del web. URL http://www.internauta37.altervista.org/darknet-o-deep-web-il-lato-oscuro-del-web.

[27] Tor Project. Tor overview. URL https://www.torproject.org/about/overview.html.en.

[28] Wikipedia. Mix network, . URL http://en.wikipedia.org/wiki/Mix_network.

[29] Nik Cubrilovic. Large number of tor hidden sites seized by the fbi in operation onymous were clone or scam sites. URL https://www.nikcub.com/posts/onymous-part1/.

[30] Georgios Portokalidis Michalis Polychronakis Angelos D. Keromytis Sambuddho Chakravarty, Marco V. Barbera. On the effectiveness of traffic analysis against anonymity networks using flow records. URL https://mice.cs.columbia.edu/getTechreport.php?techreportID=1545&format=pdf&.

[31] Pierluigi Paganini e Richard Amores. Project artemis – osint activities on deep web. URL http://resources.infosecinstitute.com/project-artemis-osint-activities-on-deep-web/.

[32] Christopher Tarbell. Declaration of christopher tarbell. URL http://ia700603.us.archive.org/21/items/gov.uscourts.nysd.422824/gov.uscourts.nysd.422824.57.0.pdf.

[33] Wired. The fbi finally says how it 'legally' pinpointed silk road's server. URL http://www.wired.com/2014/09/the-fbi-finally-says-how-it-legally-pinpointed-silk-roads-server/.

[34] Atti di indagine. URL https://pdf.yt/d/RpyX9_xmapTkhmkb.

[35] Matteo Flora. Dove finisce la paura ed inizia internet, 2008. URL http://punto-informatico.it/2247069/PI/Commenti/insicurezza-dove-finisce-paura-ed-inizia-internet.aspx.

[36] Wikipedia. Freenet, 2015. URL http://it.wikipedia.org/wiki/Freenet.

[37] CiteSeer. Freenet: A distributed anonymous information storage and retrieval system, 2001. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.10.4919.

[38] Freenet. What is freenet?, 2000. URL https://freenetproject.org/whatis.html.

[39] Freenet. The philosophy behind freenet, 2000. URL https://freenetproject.org/philosophy.html.

[40] anoNet Italia. Sito di documentazione in lingua italiana di anonet, 2009. URL https://anonetitalia.wordpress.com/.

[41] Wikipedia. anonet, 2015. URL http://en.wikipedia.org/wiki/AnoNet.

[42] Wikipedia. Stealthnet, 2015. URL http://it.wikipedia.org/wiki/StealthNet.

[43] StealthNet. Anonymous filesharing to serve and protect. URL http://www.stealthnet.de/it_index.php.

[44] Wikipedia. I2p, . URL http://it.wikipedia.org/wiki/I2P.

[45] I2P. Il progetto internet invisibile. URL https://geti2p.net/it/.

[46] Ghostery. Are we private yet? URL http://www.areweprivateyet.com/.