Sicurezza del browser e delle informazioni nell'interazione Web

Minacce e prevenzione

Prof. Fabrizio d'Amore
Sapienza Università di Roma

1. Introduzione

La velocità con cui le tecnologie dell'informazione ci hanno somministrato dosi massive di connettività, dati, strumenti di produttività ed esistenza online non ha avuto idoneo riscontro nell'educazione ad Internet, alle sue potenzialità e alle sue insidie. I nativi digitali considerano la rete alla stregua dell'aria o di un giocattolo – qualcosa che esiste da sempre e che avvolge le nostre vite – mentre i nativi analogici si suddividono fra coloro che si sono adattati e gli altri, che sono culturalmente distanti da questa presenza trasversale alle attività umane, di cui non si può che predire l'ulteriore espansione. Ma sia i nativi digitali che i nativi analogici riadattati sono troppo spesso inconsapevoli di come alcuni valori indipendenti dalla rete, come la privacy, il rispetto, la sicurezza, trovino nuove declinazioni e differenti sfumature quando immersi nella rete. La mancanza di una alfabetizzazione della sicurezza, la ridotta percezione dell'incapacità della rete di dimenticare, la sottovalutazione dei meccanismi fulminei e asimmetrici operanti sulla rete e la mancata considerazione dei rischi ad essa connessi determinano frequentamene situazioni gravemente problematiche, di non facile risoluzione e che richiedono quantità ingenti di risorse e tempi prolungati per la loro gestione appropriata.
Eppure alla base della consapevolezza e di una corretta valutazione del rischio ci sono pochi principi essenziali, riconducibili alla sicurezza delle informazioni, tema assai più anziano della tecnologia dell'informazione, che offre tuttavia nel XXI secolo nuove e significative angolazioni, che annunciano grandi sfide ed opportunità.
In questo articolo viene proposto un breve percorso che, partendo proprio dalla sicurezza delle informazioni, ne esamina le nuove declinazioni che si configurano prendendo in considerazione la più classica delle interazioni con la rete: la navigazione Web.

In quanto segue useremo i termini "dato" e "informazione" come sinonimi, per rilevando che in alcune comunità si preferisce il termine *dato* per indicare un'entità grezza e atomica, ricorrendo invece alla locuzione *informazione* per riferirsi all'interpretazione, di più alto livello concettuale, attribuita a tali entità.

2. La sicurezza delle informazioni e la cultura della sicurezza

Da molti anni viene utilizzata la dizione *sicurezza delle informazioni* per denotare l'insieme dei requisiti di sicurezza nella gestione delle informazioni ritenuti fondamentali dalla comunità scientifica e dagli esperti di settore. L'espressione "gestione delle informazioni" denota l'insieme delle attività umane e degli strumenti automatici per la memorizzazione, l'accesso in consultazione e/o modifica, l'elaborazione e la trasmissione delle informazioni.
Il bene informazione, sebbene da sempre protagonista negli scenari complessi e multidisciplinari, assume oggi nuove connotazioni grazie alle moderne tecnologie che, riducendo – forse annullando – le naturali barriere di accesso al dato, espongono il bene informazione a un'ampia e variegata casistica di accessi o usi illeciti, che definiremo genericamente attacchi.

La sicurezza delle informazioni è trattata da numerosi standard internazionali: probabilmente il più celebre di essi è lo ISO/IEC 27001:2013,[1] elaborato dai sottocomitati congiunti della International Organization for Standardization (ISO) e della International Electrotechnical Commission (IEC). Lo standard ha l'obiettivo di proteggere i dati e le informazioni da minacce di ogni tipo, al fine di assicurarne l'integrità, la riservatezza e la disponibilità, e fornire le indicazioni per adottare un adeguato sistema di gestione della sicurezza delle informazioni. A tali fondamentali requisiti spesso vengono aggiunti anche quelli di autenticazione e non ripudio. Lo standard, benché concepito in relazione alle esigenze di organizzazioni private e pubbliche, è basato su necessità che emergono oggi come critiche anche nella vita del cittadino comune, immerso in una digitalizzazione pervasiva, caratterizzata da una ampia disponibilità di dispositivi "smart" e sistemi sempre (inter)connessi.
Non è disponibile in letteratura una definizione formale e generale dei requisiti, ma tutte le proposte contenuti nei vari standard[2] ruotano, seppur con qualche distinguo, attorno ai medesimi concetti, qui di seguito brevemente discussi.

La *riservatezza*, o *confidenzialità*, è la proprietà secondo cui l'informazione non viene resa disponibile o rivelata a soggetti non autorizzati, siano essi individui, entità o processi. Il tema è molto attuale, anche in relazione al recente Regolamento (Ue) 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016 (regolamento generale sulla protezione dei dati),[3] e in questa sede verranno maggiormente focalizzati gli aspetti rilevanti all'uso del browser. Le violazioni della confidenzialità dei dati in ambito Web sono alla base di operazioni illecite come il furto di segreti industriali, il furto di identità, gli attacchi alla reputazione, lo stalking.

L'integrità dei dati è la garanzia della loro accuratezza e completezza, durante l'intero ciclo di vita. Poiché la semplice modifica di un bit altera l'integrità del dato, si richiede dunque che durante la trasmissione in rete, così come durante la memorizzazione su un qualunque supporto, i dati non possano essere mai manomessi o illecitamente creati. Gli attacchi all'integrità dei dati includono la creazione di dati fraudolentemente attribuita a terzi e sono spesso motivati da intenzioni criminose.

La *disponibilità* dei dati esprime il requisito per cui i dati debbono essere accessibili nei modi e nei tempi previsti da tutti i soggetti aventi diritto. Gli attacchi alla disponibilità dei dati appartengono alla categoria dei cosiddetti attacchi *Denial of service* (Dos) che si basano sulla semplice idea di far esaurire, mediante apposite azioni, una risorsa limitata di un sistema informatico,[4] in modo che questo non possa continuare a svolgere le sue funzioni istituzionali. Gli attacchi Dos hanno normalmente una durata limitata nel tempo e generalmente producono il loro effetto solamente durante l'esecuzione dell'attacco.

La nozione di *autenticazione* è relativa alla corretta identificazione di un soggetto, spesso

[1] https://www.iso.org/obp/ui/#iso:std:iso-iec:27001:ed-2:v1:en
[2] Una interessante rassegna di standard è disponibile in: WIKIPEDIA CONTRIBUTORS, "Cyber security standards," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/index.php?title=Cyber_security_standards&oldid=760199554 (accessed February 8, 2017).
[3] http://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX%3A32016R0679.
[4] Esempi di risorse limitate in un sistema informatico sono le memorie (principale e secondaria), la potenza di calcolo, la banda per le comunicazioni in rete ecc.

remoto, in modo che possano essere correttamente riconosciute le autorizzazioni all'uso delle risorse dei sistemi informatici qualora il soggetto sia un cliente, oppure autorevolezza ed attendibilità qualora sia un server. Inutile precisare che i soggetti che vengono autenticati non sono solamente individui, ma possono essere anche software, dispositivi fisici ecc. Il termine viene usato sia per il concetto di identificazione in senso lato, sia per le procedure specifiche di autenticazione, non essendo quest'ultime, tuttavia, oggetto di attenzione del requisito. La violazione può essere causa della violazione della riservatezza o anche causa di violazione dell'integrità dei dati nei casi di attacchi *man-in-the-middle*, ove una terza parte si interpone nella comunicazione fra due soggetti intercettando e manipolando le comunicazioni scambiate.

Il *non ripudio* si prefigge di associare in maniera certificata una identità a specifiche azioni, che hanno magari importanza formale o critica. L'esempio più significativo è la firma digitale, ove l'azione formale è quella di firmare un determinato documento e l'identità è quella del firmatario. Con la firma digitale si legano in maniera indissolubile identità del firmatario e documento firmato, spesso anche includendo la data di firma. Tale firma non può essere ripudiata, a meno di dichiarare la compromissione del dispositivo di firma, della chiave privata o dell'infrastruttura di supporto.[5]

[5] "Il documento informatico sottoscritto con firma elettronica avanzata, qualificata o digitale, formato nel rispetto delle regole tecniche di cui all'articolo 20, comma 3, ha altresì l'efficacia prevista dall'articolo 2702 del codice civile". Art. 21 (Documento informatico sottoscritto con firma elettronica), comma 2, del Codice dell'Amministrazione Digitale (D. LGS 82/2005).

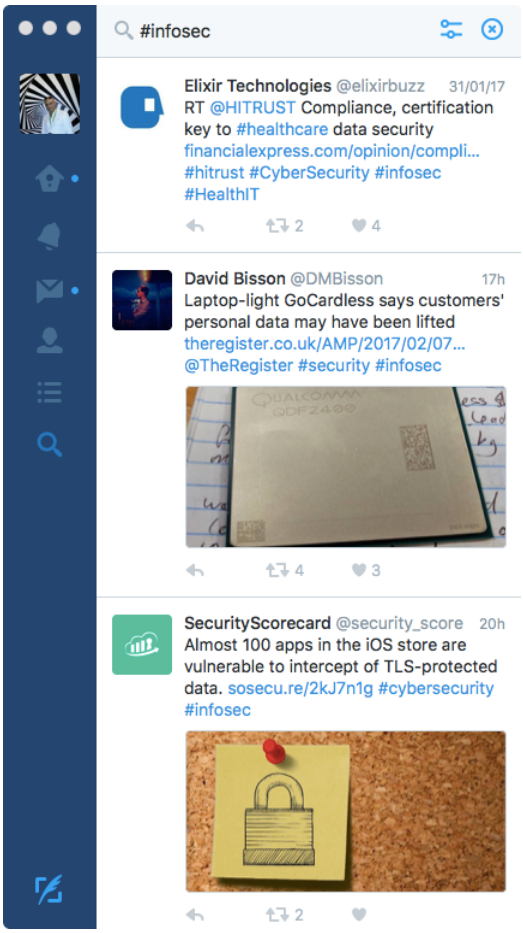

Figura 1. Alcuni tweets con hashtag #infosec.

I concetti brevemente discussi definiscono i requisiti più fondamentali della sicurezza delle informazioni, o #infosec, utilizzando il celebre hashtag dedicato al tema (in Fig. 1 sono mostrati alcuni tweets con tale hashtag, presi in tempo reale), e svolgono un ruolo critico all'interno di ogni sistema informativo, da aziendale a pubblico, includendo la sconfinata risorsa costituita del Web. È proprio il Web che ha di fatto annullato la distanza fra una delle tipologie più vulnerabili di utente – il cittadino comune – e uno smisurato quantitativo di informazioni, spesso scarsamente attendibili, in altri casi di buona o elevata qualità. La connessione ininterrotta lo espone in pratica a numerosi rischi, legati alla sicurezza delle informazioni e non solo: sono sotto gli occhi di tutti gli effetti a volte nefasti di danni reputazionali, cyber-bullismo e cyber-stalking, che si vanno ad aggiungere a tutte le spiacevoli conseguenze derivanti dalla violazione della sicurezza delle informazioni. Il cittadino comune non è preparato a ciò perché nel nostro paese non è sviluppata una cultura della sicurezza; lo stesso vale per molte persone inserite nel mondo del lavoro in era pre-Google le quali, per una ragione o l'altra, non hanno voluto o potuto sviluppare la sensibilità che è alla base della cultura della sicurezza.

La cultura della sicurezza è stata riconosciuta negli ultimi anni come area in cui investire con impegno, non solo per abilitare l'utenza esposta al cyber-rischio ad intraprendere misure lenitive, ma anche per creare consapevolezza e consenso, ingredienti necessari ai decisori chiamati continuamente ad operare a beneficio della collettività, sempre costretti a individuare il più giusto equilibrio fra le esigenze di sicurezza e il legittimo diritto alla libertà, direzioni spesso in contrapposizione. Non a caso la nostra Sicurezza Nazionale ha varato fin dall'inizio degli anni '10 un programma di apertura e collaborazioni proprio focalizzando l'obiettivo della cultura della sicurezza,[6] costruendo una rete di collaborazioni con le università italiane e varando alcune iniziative per scuole ed università. Naturalmente la cultura della sicurezza richiede la sensibilizzazione di diverse categorie di soggetti, come ad esempio decisori, formatori, operatori ed utenti, con azioni dedicate a ciascuna specifica categoria. Attualmente l'alfabetizzazione della sicurezza è scarsamente considerata nei percorsi formativi di base, anche se, così nell'occidente come nell'estremo oriente, non sembrano esistere esitazioni di alcun tipo nel digitalizzare bambini fin dall'età prescolare.

3. Internet, Web, browser

Internet è un'infrastruttura per l'interconnessione di reti di computer operante su scala

[6] Il sito Web di Sicurezza Nazionale, ristrutturato nel 2013, dedica una specifica sezione alla cultura della sicurezza: http://www.sicurezzanazionale.gov.it/sisr.nsf/cultura-della-sicurezza.html.

planetaria e capace di veicolare ogni giorno oltre 2 miliardi di GB.

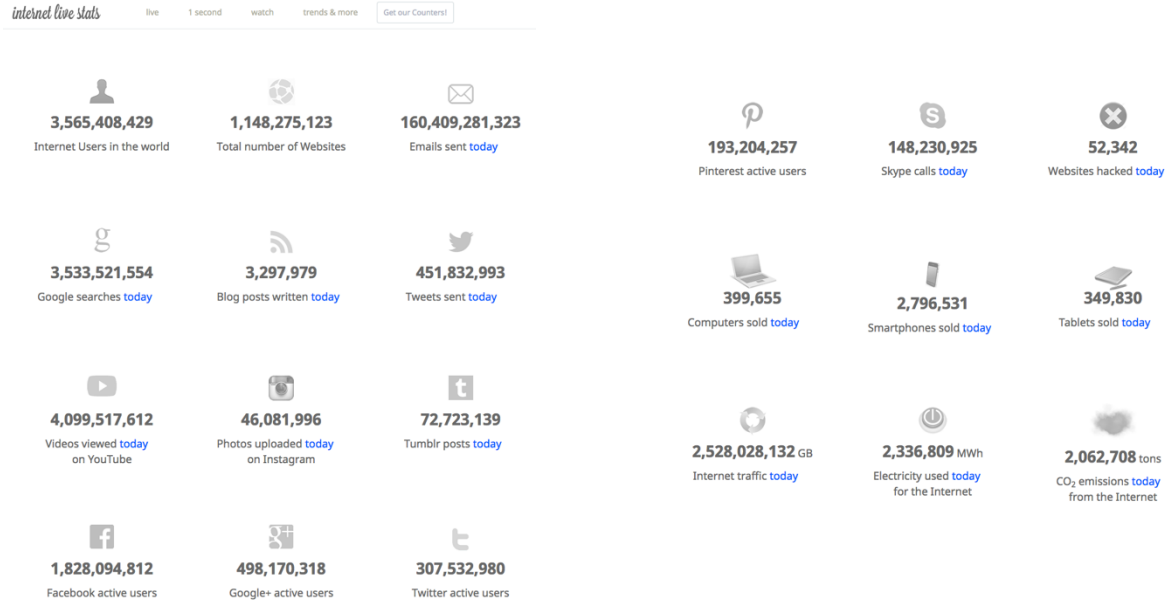

Figura 2. Uso di Internet: statistiche giornaliere. (Fonte: Internet Live Stats. http://www.internetlivestats.com).

Gran parte del traffico è legata al Web e all'email; a tal proposito appare utile evidenziare che circa tre quarti del traffico email è costituito da spam[7] (cfr. Fig. 2).

Il servizio più utilizzato su Internet è, assieme all'email, il World Wide Web, comunemente noto come Web. Esso offre un insieme di contenuti presentati come file di vario tipo e accessibili dagli utenti attraverso strumenti denominati *browser*. I contenuti sono organizzati in pagine ipertestuali descritte attraverso il linguaggio HyperText Markup Language (HTML) e una serie di altri linguaggi e strumenti più avanzati, la cui descrizione va oltre gli scopi del presente articolo.[8] Essenzialmente un browser è un'applicazione (app) disponibile in genere

[7] Messaggi non sollecitati che veicolano informazioni pubblicitarie, tentativi di frode o *phishing*, allegati contenenti virus e altre categorie di malware, o anche semplicemente tecniche per collezionare e validare indirizzi email di utenti nel mondo. Un'interessante lettura sull'argomento è il white paper: DAVID HARRIS, "Drowning in Sewage", Copyright (c) 2003. https://web.archive.org/web/20071128052944/http://www.pegasusmail.tk/upload/SPAM_white_paper.pdf

[8] Un linguaggio di markup è un sistema di annotazione di documenti tale da essere sintatticamente distinguibile dal testo.

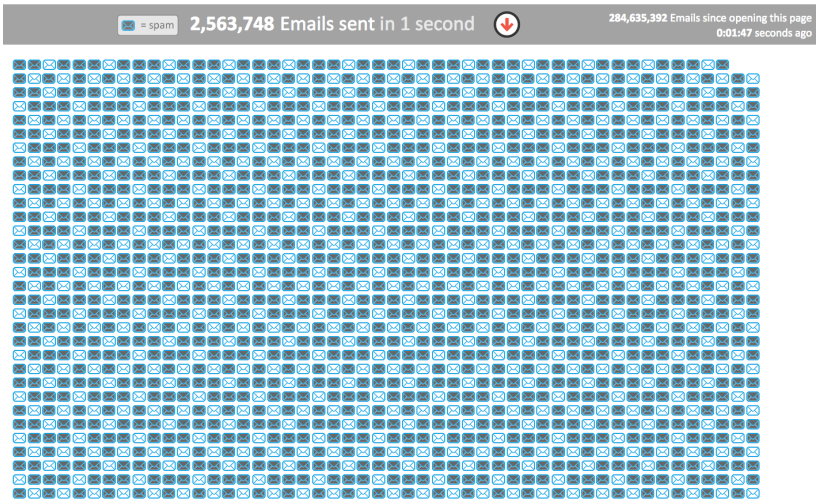

Figura 3. Una stima del numero di email inviate in un secondo, suddivise fra email legittime (icone in chiaro) e spam (icone scure). (Fonte: Internet Live Stats. http://www.internetlivestats.com/one-second/#email-band).

gratuitamente su tutte le piattaforme di elaborazione, inclusi smartphone, tablet, smart-tv e altri dispositivi di intrattenimento, che ha la capacità di collegarsi a un sito remoto (Web server), allo scopo di ottenerne i contenuti; il browser avvia con il Web server un dialogo, svolto utilizzando il HyperText Transfer Protocol (http), protocollo[9] usato per la trasmissione di informazioni sul Web. Anche se la trattazione è qui estremamente semplificata, essa permette di comprendere come browser e Web server si coordinino fra loro scambiando frasi appartenenti al protocollo http, per la consegna di file HTML, che descrivono la struttura delle pagine, e di ulteriori file per altri contenuti (non testuali) presenti nella pagina. Nel 2014 il World Wide Web Consortium[10] (W3C) ha definito il linguaggio HTML5 che, assieme alle altre tecnologie Cascading Style Sheet (CSS)[11] e JavaScript,[12] è diventato lo standard moderno per offrire contenuti Web, sia tradizionali come il testo e la grafica statica, che multimediali (audio e video), potendo anche costruire pagine ad alta interattività.

Non è stato sempre così. Nel 1993, alla nascita del Web, esisteva un solo browser capace di mostrare testo e grafica integrati nella pagina Web: NCSA Mosaic, sviluppato all'Università dell'Illinois Urbana-Champaign. C'erano stati alcuni esperimenti precedenti, che però non integravano testo e grafica. Da allora il Web è cresciuto immensamente, così come la competizione continua per offrire browser più potenti ed evoluti, al fine di catturare il maggior numero di utenti, nota come *guerre dei browser*. Nella prima guerra la competizione vedeva come protagonisti Netscape (sviluppato da Netscape Communications) e Internet Explorer (sviluppato da Microsoft) e si basava sia sull'aggiunta ai browser di nuove peculiarità, a volte non rispettose dello standard del linguaggio HTML, sia sull'introduzione di caratteristiche più avanzate, come ad esempio i *plugin*. Questo periodo non giovò allo sviluppo del Web che, pur sempre nella sua crescita forsennata, vedeva gli sviluppatori dei siti costretti ad operare scelte in merito alle caratteristiche fuori standard da seguire, determinando la nascita di siti Web

[9] Un protocollo di comunicazione è un insieme di regole formalmente descritte, che porta frequentemente alla individuazione di un repertorio di frasi standard, definite al fine di consentire la comunicazione tra più parti.

[10] Organizzazione internazionale, con sede presso il Computer Science and Artificial Intelligence Laboratory del Massachusetts Institute of Technology, che definisce e manutiene gli standard del Web. http://www.w3.org/

[11] Linguaggio per definire le modalità di formattazione di una pagina HTML o, più in generale, scritta in un linguaggio di mark-up.

[12] Linguaggio di programmazione che può essere impiegato per descrivere azioni svolte dai browser per offrire all'utente i contenuti digitali di una pagina Web. Il codice sorgente dei programmi è inviato dai Web server ai browser, che lo eseguono mostrando l'effetto nell'ambito della pagina stessa.

pienamente fruibili da un solo browser. In questa fase prevalse Internet Explorer, non tanto per sue qualità intrinseche quanto per la politica di Microsoft di renderlo parte integrante del sistema operativo Windows, facendo dunque sì che gli utenti trovassero Internet Explorer già preinstallato sul PC e dovessero invece intraprendere specifiche azioni per installare un browser diverso. Successivamente alla discesa in campo di altri attori si è svolta la seconda guerra dei browser, in cui si è assistito all'ascesa di nuovi protagonisti che hanno intaccato il dominio di Internet Explorer, fino a spodestarlo completamente. Attualmente il browser più utilizzato è Chrome (sviluppato da Google), come illustrato dalla Fig. 4.

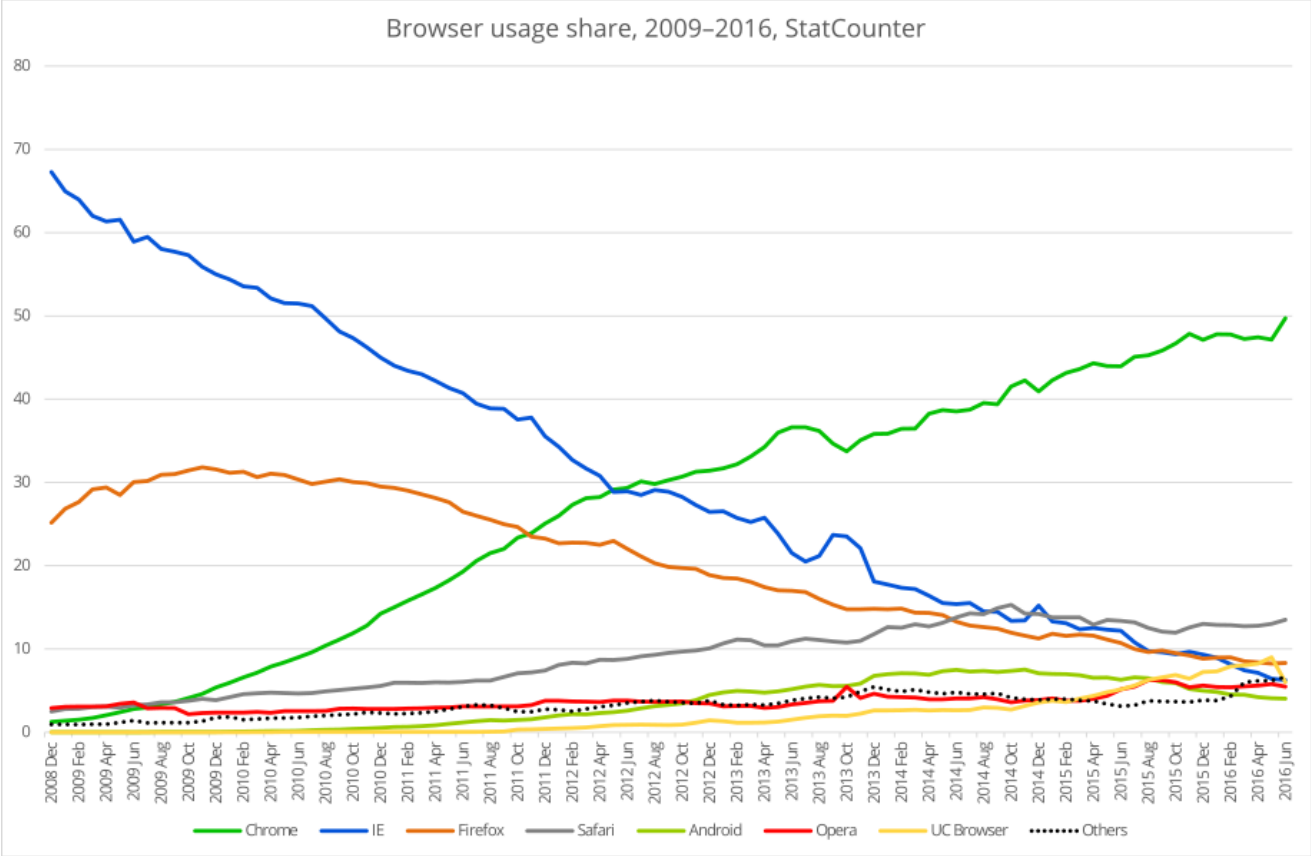

Figura 4. Diffusione dei browser più noti. Fonte: StatCounter.
https://commons.wikimedia.org/wiki/File:Browser_usage_share,_2009%E2%80%932016,_StatCounter.svg

Nella seconda guerra dei browser i contendenti hanno rinunciato all'introduzione di caratteristiche fuori dagli standard ma si sono concentrati su aspetti di efficienza, qualità e sicurezza, innalzando significativamente la qualità dei prodotti in circolazione. Proprio per questa ragione si è ingigantita la complessità dei browser così come il numero delle loro funzioni, determinando un aumento significativo della superficie esposta agli attacchi.[13]

Per meglio comprendere la presenza di minacce silenti insite nell'uso poco accorto dei browser riprendiamo il tema dei già menzionati plugin. I browser moderni ricevono dai vari server Web numerose tipologie di file e, per molte di queste, il browser sa come impiegare i file ricevuti. Il consorzio W3C, ispirato ai tentativi di classificazione dei file svolti nell'ambito della

[13] L'ultima versione di Mosaic consta di 2,9 MB, mentre l'ultima di Chrome supera i 300 MB.

standardizzazione di MIME, volta a consentire l'uso dell'email a popolazioni di lingua diversa dall'inglese e a gestire allegati, ha introdotto il concetto di *media type*, proprio per definire le categorie di file. Si tratta di un approccio strutturato e più evoluto di quello Microsoft che si limita a denotare la categoria del file attraverso l'estensione.[14] Infatti, l'idea dei media type consiste nello specificare per ogni file un tipo e un sottotipo. Ad esempio, un file HTML e di tipo text/html, che significa che il tipo è text (testo) e, nell'ambito del tipo text, appartiene alla categoria (sottotipo) html. Nella tabella che segue vengono mostrate i dieci tipi ufficiali e alcuni sottotipi (ce ne sono a centinaia) significativi.[15]

tipo	*esempi sottotipi*	*esempio media type*
`application`	`pdf`, `msword`, `zip`, `postscript`, …	`application/pdf`
`audio`	`mp4`, `mpeg`, `ogg`, `ac3`, …	`audio/mp4`
`font`	`collection`, `otf`, `woff`, …	`font/collection`
`example`	*usato per esempi ed esperimenti, non c'è limite ai sottotipi utilizzabili*	`example/…`
`image`	`png`, `bmp`, `jpeg`, `tiff`, …	`image/png`
`message`	`delivery-status`, `disposition-notification`, `sip`, …	`message/delivery-status`
`model`	`iges`, `vrml`, `gltf+json`, …	`model/iges`
`multipart`	`encrypted`, `report`, `digest`, …	`multipart/encrypted`
`text`	`html`, `plain`, `csv`, `rtf`, …	`text/html`
`video`	`3gpp`, `mp4`, `quicktime`, `ogg`, …	`video/3gpp`

Quando un Web server invia a un browser un file invia anche il media type del file, attraverso un apposito messaggio http. Il browser a sua volta, ricevendo un determinato media type, verifica se esso è presente fra i media type che esso conosce, e in tal caso sa come utilizzare il file, oppure, se sconosciuto, il browser offre all'utente la possibilità di salvare il file.[16] Nel 1995, in piena guerra dei browser, Netscape Communications propose un sistema standard di interfacciamento[17] del browser con un software addizionale, concepito proprio per estenderne le capacità: il plugin, software sviluppato ad hoc per gestire media type sconosciuti al browser. Per effetto di ciò, ancora oggi ci si può imbattere in pagine che non vengono visualizzate correttamente dal browser, che ci potrebbe mostrare un messaggio di invito ad installare un apposito plugin per poter fruire dei contenuti. È l'utente a decidere in merito all'installazione di software addizionale, forse sconosciuto, che magari, oltre a estendere le funzioni del browser, presenta comportamenti non desiderabili; da notare che non è richiesto un account di

[14] Le lettere che tipicamente seguono il punto (dot) che separa, nello standard Microsoft, il nome effettivo di un file dalla sua estensione.

[15] La lista di tutti i media type è disponibile nella pagina IANA https://www.iana.org/assignments/media-types/media-types.xhtml

[16] I browser sono spesso anche collegati ad altre applicazioni, per cui può accadere che un file di un media type sconosciuto al browser possa essere, oltre che salvato in locale, aperto mediante un'applicazione esterna. Questo comportamento, in parte preconfigurato, può essere modificato dall'utente.

[17] Si tratta della Netscape Plugin Application Programming Interface (NPAPI), solo recentemente caduta in disuso grazie alle nuove caratteristiche di HTML5, tranne che per alcuni celebri plugin come Flash, Silverlight e Java.

amministratore del computer per poter installare un plugin.

Si potrebbe ritenere allora che un comportamento prudente consiste nell'evitare assolutamente di installare plugin sconosciuti, consentendo solo quelli ben noti. Un esempio significativo è il plugin Flash Player di Adobe Systems, forse il più noto, ancora oggi necessario per visualizzare molti siti Web, in particolare per fruire di audio/video in streaming o anche per consentire giochi grafici interattivi, molto graditi ad adolescenti e a bambini in età preadolescenziale. Flash Player, oggi in declino a causa della concorrenza di HTML5, è stato nel tempo veicolo di numerosi attacchi ai browser, anche se Adobe Systems è sempre stata reticente nel fornire dati ufficiali sugli incidenti. Una breve discussione in proposito – illuminante – è presente nelle pagine di Wikipedia.[18] In particolare, non deve essere sufficiente la popolarità di un plugin a indurre un utente ad eseguirne l'installazione, senza parlare poi del fatto che in alcuni casi l'installazione è semplice ma la disinstallazione appare più complessa.[19] Alla legittima domanda "come capire se un plugin è malevolo oppure no?" non è facile fornire una risposta soddisfacente. Un buon antivirus sicuramente è a conoscenza di quali plugin in passato si sono rivelati dannosi, ma se un plugin è (relativamente) recente e nocivo potrebbe occorrere molto tempo prima che gli antivirus ne diventino consapevoli. Nel frattempo il plugin può operare con una certa libertà, anche se non completa a causa delle limitazioni che i browser impongono proprio per ragioni di sicurezza. *Una buona politica è installare un plugin solo se realmente necessario e se gode di buona reputazione, disinstallandolo (o disabilitandolo) subito dopo l'uso*.[20]

4. Tipologie di minacce

Attraverso il browser è possibile portare numerosi attacchi all'utente che lo adopera e/o al computer che lo ospita ma, poiché la materia diviene estremamente tecnica, non entreremo nei dettagli. Viceversa, le tipologie di minacce possono essere aggregate in due ampie categorie, che presentano anche delle interdipendenze:

- profilazione e altri attacchi alla privacy;
- installazione di malware e arruolamento in botnets.

Descriveremo brevemente le categorie e successivamente presenteremo un approfondimento sulle metodologie di profilazione e sulle conseguenze derivanti.

Appartengono alla prima categoria tutte le minacce tese ad acquisire informazioni sull'utente del browser o anche estraendole dal dispositivo su cui il browser è in esecuzione. Le metodologie in essere possono essere basate sull'inganno o anche sull'allestimento di pagine ad hoc che sono capaci di tracciare il comportamento sul Web dell'utente. Esempi ben noti sono pagine Web che nell'apparenza e nello stile sono costruite in modo da sembrare le pagine istituzionali di banche, gestori finanziari, enti emittenti di carte di credito o anche di gestori di

[18] https://en.wikipedia.org/wiki/Adobe_Flash_Player#Security

[19] Nel browser Safari di Apple, è presente un pannello di configurazione per attivare/disattivare i plugin installati, ma non compare una funzione di disinstallazione degli stessi.

[20] Un importante problema è che esistono a tutt'oggi moltissimi utenti che non hanno idea di come agire per seguire questo suggerimento prudenziale.

servizi email o altri servizi Internet. L'utente è indotto attraverso un messaggio di phishing[21] a collegarsi a tali pagine per eseguire una "urgente" operazione di aggiornamento, conferma o sblocco e a compilare le pagine inserendo dati sensibili, che vengono catturati e memorizzati dai server a cui ci si è collegati. Le vittime non si sono accorte che il browser mostrava nella barra degli indirizzi un indirizzo diverso da quello istituzionale.[22]

Altri metodi di violazione della privacy sono basati sulla profilazione e/o riconoscimento degli utenti che visitano le varie pagine del Web, ricostruendo e manutenendo esplicitamente la lista di siti visitati attraverso espedienti tecnici presenti nelle pagine. Di ciò si mostrerà un esempio nel prossimo paragrafo. Più in generale l'idea di sorvegliare in massa gli utenti del Web è stata percorsa da agenzie di intelligence dotate di grandi mezzi tecnici ed economici: le cronache hanno parlato diffusamente di tali casi negli ultimi anni. Si noti che fra le tecniche di sorveglianza di massa c'è anche la sistematica accensione della mini-camera (webcam) e del microfono, dotazioni standard di computer portatili, tablet e smartphone, con registrazioni segretamente inviate a siti remoti di raccolta dati.

La seconda categoria è legata al malware e alla partecipazione alle botnets. Il browser, attraverso la visita di siti malevoli o compromessi che sfruttano la possibilità di comandare il browser attraverso l'esecuzione di codice JavaScript, diviene il veicolo per l'installazione diretta o indiretta di software nocivo, denominato malware. Si intende per *malware* un'ampia categoria di software che, quando eseguito, produce effetti indesiderati se non addirittura ostili. Esempi di malware ad ampia diffusione:

- Trojan. Software che si nasconde all'interno di altro software (ad esempio un antivirus o una utilità di analisi o gestione del dispositivo) e che viene eseguito all'insaputa dell'utente. In genere ha la capacità di collegarsi a siti remoti per scaricarne ulteriore malware o per creare delle *back-door*, ovvero canali che consentono l'accesso al dispositivo da parte di soggetti remoti, per comandarlo e controllarlo. Possono essere usati dalle forze dell'ordine anche a fine di intercettazione, attività regolata da precise disposizioni di legge.
- Keyloggers. Hanno la capacità di registrare ogni singola pressione di tasto e di fotografare lo schermo ad intervalli regolari. Inviano rapporti dettagliati a siti remoti di raccolta dati. Ne esistono alcune varianti, dando luogo a una sottocategoria del malware denominata *spyware*.
- Ransomware. Minaccia molto attuale e – purtroppo – assai diffusa. Si tratta di un tipo di malware che limita l'accesso del dispositivo che infetta, richiedendo di pagare un riscatto per rimuovere la limitazione. Un celebre esempio è il CryptoLocker[23] che cifra i file dell'utente chiedendo un pagamento per riportare i dati in chiaro. L'esecuzione dei versamenti è in genere richiesta attraverso la moneta virtuale *bitcoin*,[24] che consente di

[21] Il termine è stato coniato nel 1996 da alcuni hacker che tentarono di impadronirsi delle credenziali di utenti di American Online, un fornitore di servizi Internet americano acquisito nel 2015 da Verizon Communications. In inglese la parola *phishing* si pronuncia come *fishing*, che significa letteralmente *pésca* o *pescare*.

[22] Esistono attacchi più sofisticati che avvelenano i servizi DNS in modo tale che, anche se il browser mostra un indirizzo apparentemente corretto nella barra degli indirizzi, questo si sta in effetti collegando a un diverso Web server.

[23] Per una completa descrizione rimandiamo alle pagine di Wikipedia. https://en.wikipedia.org/wiki/CryptoLocker

[24] Moneta elettronica introdotta nel 2009 che non fa uso di un ente centrale ma opera in maniera distribuita per tenere traccia delle transazioni e sfrutta la crittografia per gestire aspetti come la generazione di nuova moneta e la certificazione di proprietà. Wikipedia ne fornisce una buona descrizione. https://en.wikipedia.org/wiki/Bitcoin

eseguire operazioni di trasferimento di denaro in anonimato.
- Virus generici e worms. Altro malware, talvolta di natura benigna, che si installa su un dispositivo e attraverso questo si propaga e replica su altri dispositivi. Può produrre effetti collaterali come veicolare pubblicità indesiderate (adware), interferire con i contenuti del disco, o produrre effetti dimostrativi.

In alcuni casi il malware si limita ad "arruolare" il nostro dispositivo in una *botnet*, ovvero una rete di dispositivi compromessi che possono essere comandati e controllati da remoto e che vengono in genere utilizzati per partecipare ad attacchi massivi contro bersagli presenti su Internet. In tal caso, fra l'arruolamento nella botnet e il successivo attacco possono intercorrere molti mesi, duranti i quali il malware rimane silente e quindi di più difficile individuazione.
Altra azione soventemente eseguita dal malware è il blocco dei sistemi di aggiornamento del software del dispositivo, con particolare attenzione all'eventuale antivirus.

5. La profilazione basata su cookies

I cookies sono piccoli contenitori di dati che i server Web depositano sul nostro dispositivo. Consistono essenzialmente di blocchi dati che i browser memorizzano sul dispositivo in uso come singoli file (raro) o all'interno di un database dedicato ai cookies. Ogni dispositivo, per ogni diverso utente, per ogni browser, manntiene cookies separati.

I cookies vengono ricevuti dai Web server grazie a una caratteristica del protocollo http, che consente a un server di inviare al browser, attraverso l'apposito messaggio http "Set-Cookie", un singolo cookie contenente dati che l'applicazione Web desidera immagazzinare sul dispositivo client. Si tratta di dati di vario tipo: dalle preferenze di visualizzazione della pagina, al cosiddetto "carrello della spesa" (applicazioni di e-commerce); dal risultato positivo di un'operazione di autenticazione allo stato di una interazione con un'applicazione Web, che può essere recuperato quando l'utente riapre la pagina. In altre parole i cookies sono un meccanismo che ha l'effetto di aggiungere memoria all'interazione fra browser e Web server, per ovviare al limite di http di essere un protocollo senza memoria. Tale tipo di memorizzazione libera il server dall'onere di centralizzare tutte le memorizzazioni.

I browser per default ricevono i cookies e li memorizzano in una specifica area dedicata all'utente e al browser; c'è un'area separata per ciascuna coppia (utente, browser). Quando un browser effettua una connessione http verso un server Web verifica se nell'area di memorizzazione sono presenti cookies provenienti da quel server e, se sì, glieli invia. Così il server può recuperare informazioni relative allo stato dell'interazione durante l'ultima visita dell'utente.

Ciascun cookie è caratterizzato da una coppia (*nome*, *valore*) e da alcuni possibili attributi. Più precisamente, il *nome* del cookie è un identificatore che lo individua, il *valore* è costituito dai dati che si vogliono memorizzare con il cookie. Fra gli attributi vi sono una data di scadenza, eventuali restrizioni nell'accesso e nell'invio del cookie, dominio e percorso associato al cookie. Il browser in effetti invierà il cookie al server solo se il cookie non è scaduto, le modalità di accesso e comunicazione sono idonee, il dominio del Web server e il relativo percorso corrispondono a quanto eventualmente specificato nel cookie. Ogni browser ha inoltre un pannello di configurazione ove l'utente può visualizzare i cookies effettivamente memorizzati e

rimuovere quelli ritenuti indesiderati. Inoltre nello stesso pannello l'utente può stabilire regole per far sì che il browser rifiuti automaticamente cookies proveniente da specifici domini. Tuttavia, se pochi sono coloro che analizzano/cancellano i cookies presenti, pochissimi sono quelli che definiscono regole di accettazione/rifiuto automatico.

I cookie sono generalmente classificati come
- di sessione: senza data di scadenza, vengono cancellati alla chiusura del browser;
- persistenti: con data di scadenza, espressa in modo assoluto o in tempo di memorizzazione del cookie;
- di prima parte: proveniente dal sito visitato (possono essere sia di sessione che persistenti);
- di terza parte: provenienti da un sito differente da quello originariamente visitato, il quale ne innesca la visita (in genere sono persistenti).

Per meglio comprendere cosa sia un cookie di terza parte consideriamo un tipico esempio di pagina a contenuti multipli. Nell'illustrazione che segue (Fig. 5) è mostrato un dettaglio di una pagina Web di un quotidiano italiano.

Figura 5. Cookies di terze parti.

L'utente che visita la pagina lo fa in genere perché lo ha deciso; il dominio Internet della pagina è la "prima parte", ovvero quella visitata consapevolmente e volutamente. Si nota immediatamente che la pagina offre contenuti provenienti da altri siti: in particolare la sequenza di icone "social," appositamente inserite per la condivisione della notizia; è immediato identificare le icone di Facebook, Twitter, Google+ e LinkedIn. Tali icone appaiono nella pagina perché questa contiene dei collegamenti ai rispettivi siti, che rivestono il ruolo di "terza parte". In altre parole, la pagina HTML riprodotta in figura contiene al suo interno espliciti richiami a pagine di altri siti e il browser provvede automaticamente e autonomamente a procurarsi tali contenuti esterni (in questo caso: le icone social). L'interazione con le terze parti avviene sempre osservando le regole del protocollo http, per cui tali siti, nel fornire la loro risposta,

potranno inviare i loro cookies, che in tal caso saranno appunto cookies di terza parte. Da notare che in ogni colloquio http il browser che chiede un file al server fornisce automaticamente l'informazione denominata *http referer*, cioè l'indirizzo della pagina che ha causato la richiesta corrente. Ciò significa che la terza parte, oltre a ricevere la richiesta dell'icona social, riceve anche l'indirizzo della pagina del quotidiano che mostrerà l'icona al suo interno, ovvero conoscerà l'ambito di visualizzazione di quell'icona. Sarà perciò semplicissimo per il server di terza parte creare un cookie che contiene dati sulla pagina vitata dall'utente (la pagina del quotidiano).

Da questa breve spiegazione si evince che i cookies di terze parti definiscono un semplice ed efficace meccanismo di tracciamento delle pagine visitate dall'utente, poiché la prossima volta che questi aprirà una pagina contenente un'icona social, il suo browser invierà alla terza parte il cookie preesistente (che conteneva dati sulla pagina precedentemente visitata), oltre al nuovo http referer, e la terza parte potrà inviare un cookie contenente la sequenza aggiornata delle pagine visitate. L'esempio chiarisce il meccanismo di tracciamento che sfrutta i cookies di terze parti, che non è basato solo sulle icone social ma su tutti i contenuti provenienti da sorgenti multiple e visualizzati in una stessa pagina. Le due illustrazioni (Figg. 6 e 7) che seguono mostrano l'elenco dei domini contattati, rispettivamente, in seguito alla visita della pagina dei quotidiani Repubblica e Corriere. La rilevazione è stata fatta usando l'estensione di Firefox denominata LightBeam.

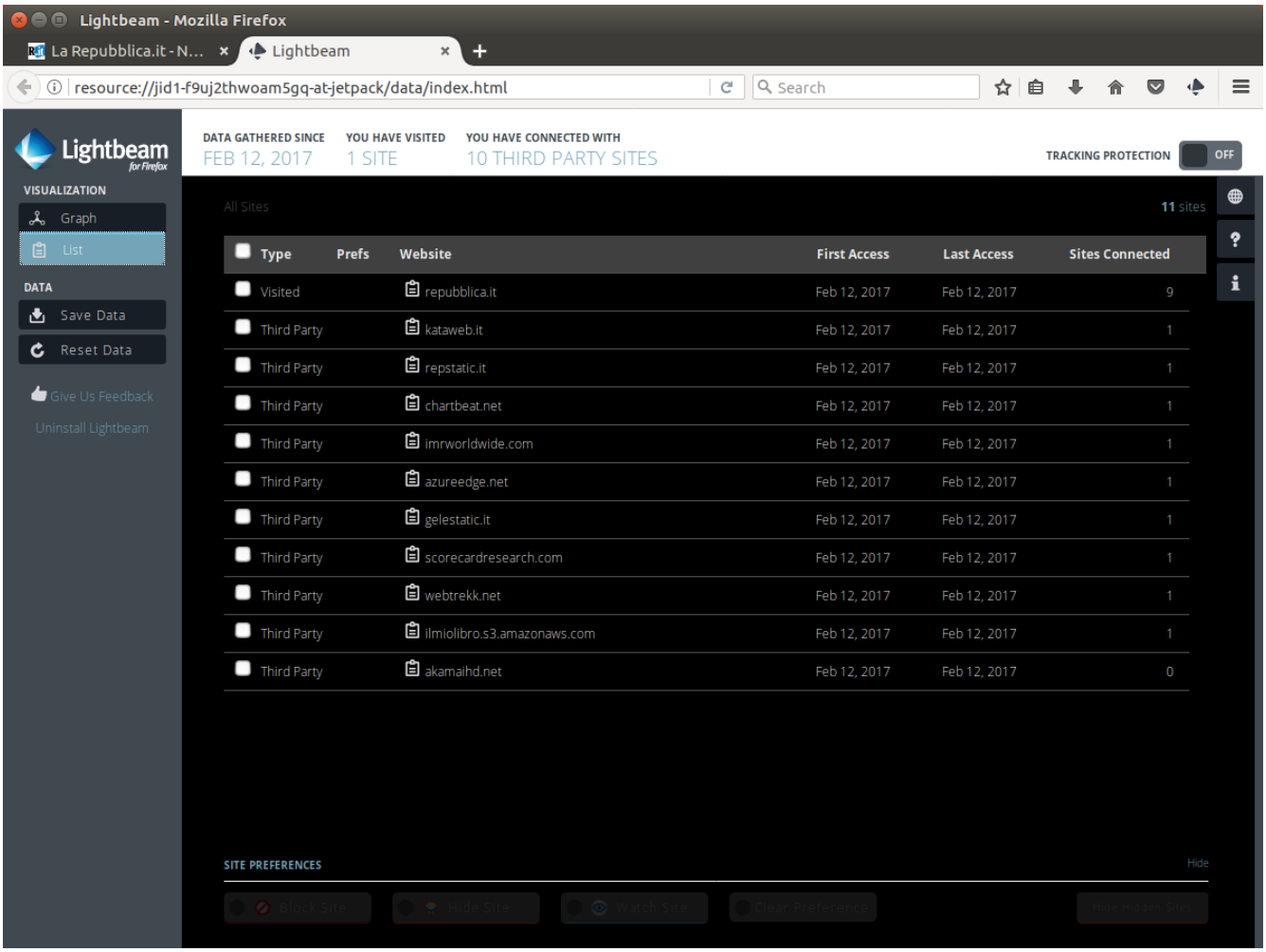

Figura 6. Sorgenti dei contenuti della pagina www.repubblica.it.

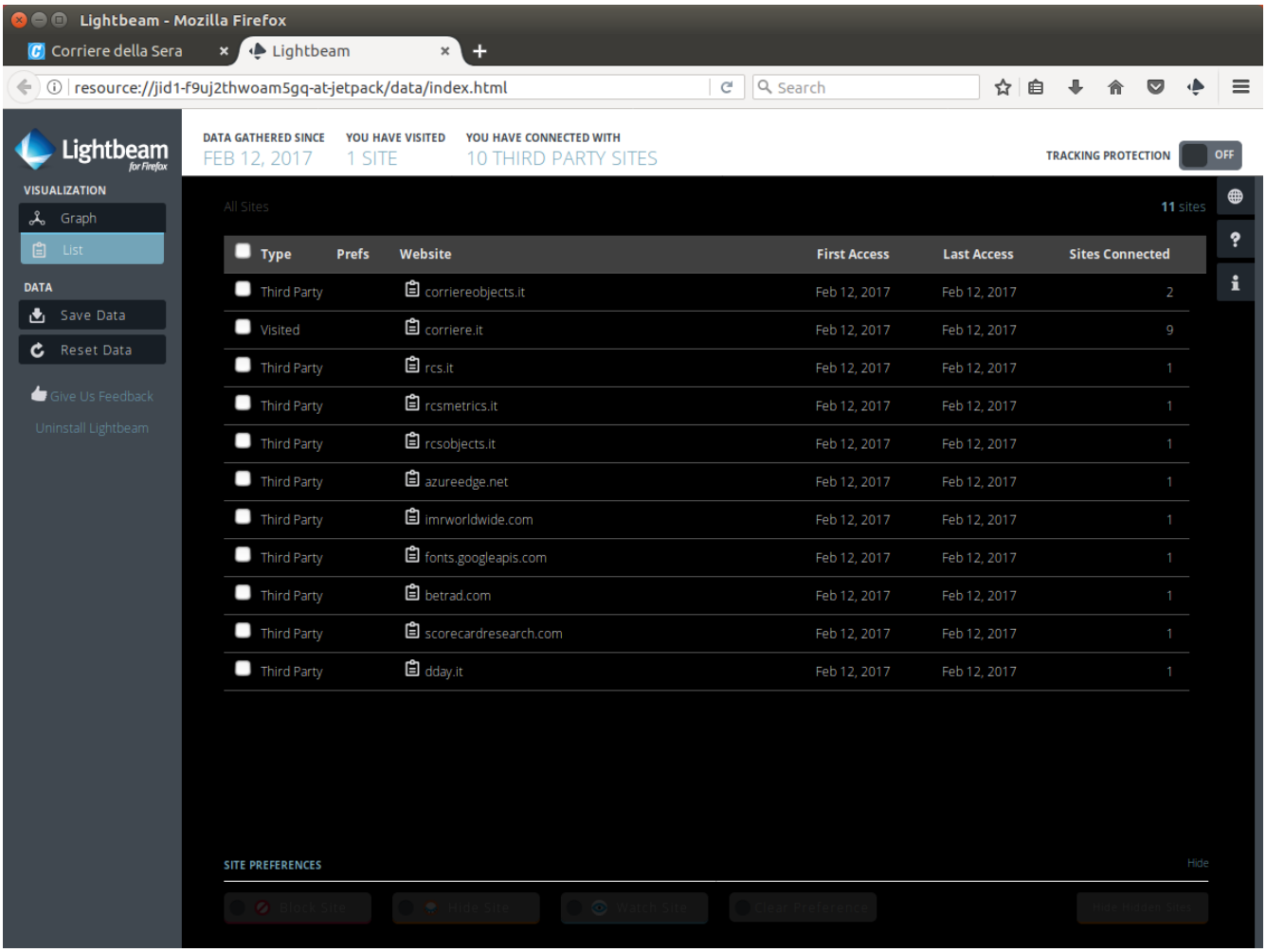

Figura 7. Sorgenti dei contenuti della pagina www.corriere.it.

I due esempi mostrati sono tipici: ogni pagina complessa del Web fornisce contenuti in maniera articolata e provenienti da sorgenti multiple. Questo aumenta fortemente la possibilità di tracciamento da parte dei grandi portali. Si noti inoltre che, una volta subito il tracciamento dei portali social, al primo ingresso nelle pagine social personali si trasferiranno al portale tutti i cookies di tracciamento provenienti dal portale che ora conosce l'identità del soggetto tracciato e può immagazzinare tali dati all'interno del profilo utente, conservandoli a tempo indeterminato. In base a ciò l'utente riceverà inserzioni pubblicitarie mirate sui temi che caratterizzano le pagine tracciate. Il meccanismo è molto pericoloso perché consente di far leva su eventuali debolezze degli utenti (ad es., gioco d'azzardo online) senza tener conto delle loro età né della pericolosità delle sollecitazioni.

Per contrastare questo fenomeno *è raccomandabile configurare i browser in modo da rifiutare automaticamente i cookies di terze parti, agendo sulle impostazioni di tutti i browser in uso*. Altra utile raccomandazione è quella di *impiegare frequentemente la modalità anonima* (o "in incognito", o "privata": il nome cambia a seconda del browser). Tale modalità di uso è in genere nota al pubblico perché permette l'uso del browser senza che venga registrata la cronologia di navigazione, a vantaggio della privacy. Tuttavia riteniamo che il vero contributo alla privacy di tali modalità sia di differente natura. Nella modalità anonima l'utente inizia una sessione di

navigazione con un browser temporaneamente privo di cookies: è come se attivando tale modalità il browser iniziasse a fare riferimento a una nuova area di immagazzinamento cookies, inizialmente vuota. In tale area verranno immagazzinati i nuovi cookies in arrivo ma, alla chiusura del browser, l'intera area verrà automaticamente azzerata. L'utilità quindi è quella di iniziare una sessione di navigazione priva di cookies e di cancellare automaticamente, alla chiusura del browser, tutti quelli accumulati. Attenzione: il browser deve essere completamente chiuso (talvolta la chiusura della finestra non chiude il browser) per cancellare i cookies e per far sì che alla prossima finestra anonima inizi una nuova sessione priva di cookies.

5.1. Altre minacce legate ai cookies

A valle della discussione sui cookies è opportuno menzionare altri attacchi che possono essere perpetrati abusando del meccanismo dei cookies. Rimane inteso che gli esempi qui commentati non vengono approfonditi quanto meriterebbero e che coprono solo una piccola parte dello scenario di interesse. Anche in questo caso ci avvarremo di un esempio.

Sono molte le occasioni in cui interagendo con un portale Web dobbiamo inserire le nostre credenziali (username e password). Ad esempio per leggere l'email oppure per consultare pagine social di interesse. Una volta eseguita con successo l'operazione di autenticazione siamo abilitati ad eseguire una serie di azioni successive, che non potrebbero essere svolte in assenza di autenticazione. Il portale ricorda che ci siamo già autenticati, infatti non continua a chiederci la password per eseguire le azioni successive. Questa memoria può essere realizzata attraverso vari meccanismi e uno dei più noti è quello è basato sull'impiego di un opportuno token di autenticazione (è in sostanza un numero segreto) che viene consegnato al browser mediante un cookie di sessione. Durante le successive operazioni il browser, mostrando al server il cookie con il token di sessione, dimostrerà di aver già eseguito con successo le operazioni di autenticazioni e quindi il portale non chiederà all'utente di autenticarsi di nuovo. Ovviamente tale cookie, in quanto di sessione, verrà distrutto al termine della navigazione. Quando invece, al momento dell'autenticazione, selezioniamo anche la casellina "ricorda l'accesso" (i nomi possono essere molteplici, ma tutti esprimeranno il concetto di rammentare successivamente che è stata eseguita con successo una autenticazione) il cookie che viene creato è permanente, in modo tale che sarà possibile esibire il token di autenticazione anche nelle sessioni successive, evitando di dover reinserire le credenziali.[25] Ebbene, che accade se un malintenzionato si impossessa del cookie contenente il token di autenticazione? Semplicemente, può eseguire un *furto di identità*. In tal caso, esibendo il cookie al portale, potrà fingere di essere noi e dunque operare liberamente all'interno dell'account (almeno finché non selezionerà un'operazione che richiede obbligatoriamente l'inserimento della password, come ad esempio, nel caso del cambio della password).

Il furto di cookies, sebbene non semplicissimo, è dunque uno strumento per eseguire furti di identità. In alcuni casi, i portali che inviano i cookies non li confezionano in maniera ottimale, ad esempio restringendo le modalità di accesso e trasmissione degli stessi.[26] In tal caso il furto di cookie può diventare semplice e facilmente eseguibile nel caso in cui, oltre alla scheda aperta

[25] Naturalmente il "ricorda l'accesso" non avrà conseguenze nel caso di navigazione privata perché il relativo cookie verrà comunque distrutto al termine della sessione.

[26] Ad esempio, è grave errore inviare un cookie di autenticazione su una connessione non protetta.

sul portale in cui ci siamo autenticati, ci sia una seconda scheda aperta nella stessa finestra che, attraverso opportune violazioni della sicurezza del browser, potrà copiare il cookie di autenticazione, con ovvie conseguenze. Di qui la raccomandazione di *navigare su portali ove ci siamo autenticati sempre impiegando finestre a scheda singola*.

Concludiamo infine menzionando l'esistenza di cookies speciali, che sfuggono al controllo esercitabile mediante browser. Ad esempio, i contenuti Flash eseguiti dall'apposito plugin del browser provocano in genere la memorizzazione di informazioni di servizio in oggetti denominati *Local shared objects* che svolgono per Flash servizi analoghi a quelli svolti dai cookies. Il luogo di memorizzazione di tali oggetti non è in genere noto al browser e le attività di configurazione e gestione possono essere svolte accedendo a un contenuto Flash disponibile in una specifica pagina[27] del vecchio sito di Macromedia, l'azienda che inventò Flash, acquisita nel 2005 da Adobe Systems.

6. Il browser fingerprinting: un'altra tecnica di tracciamento

Nell'interazione con un Web server ogni browser fornisce un significativo quantitativo di dati che dipende dalle caratteristiche del software, del sistema operativo, da altro software installato nel dispositivo ecc. In effetti la probabilità che due browser diversi che interagiscono con lo stesso server forniscano le stesse informazioni è piuttosto bassa. Si può dire in effetti che ogni browser è contraddistinto da una impronta digitale (fingerprint) non unica, ma estremamente rara. Ne scaturisce la possibilità di riconoscere un utente attraverso il riconoscimento del suo browser.

Figura 8. Esito test sul browser fingerprinting su Safari, con impronta unica su 234348 browser testati.

Tale vulnerabilità è stata descritta pubblicamente e chiaramente da Electronic Frontier Foundation (Eff), un'organizzazione internazionale non profit rivolta alla tutela del diritto alla privacy e alla libertà di parola nel mondo digitale. Eff ha varato nel 2010 Panopticlick, un progetto teso a studiare e a dare dimostrazioni pratiche del potere di tracciamento basato su fingerprint. Nella pagina Web dedicata[28] è possibile testare il proprio browser. Nella Fig. 8 è presentato il risultato del test di un browser Safari già configurato per ridurre al massimo l'esposizione al tracciamento, offrendo dunque un fingerprint assai ridotto.

Non è semplice mitigare questo tipo di minaccia, poiché in buona parte risulta

[27] http://www.macromedia.com/support/documentation/it/flashplayer/help/settings_manager.html
[28] https://panopticlick.eff.org

essere una minaccia passiva. È possibile, in linea di principio, impiegare opportune estensioni del browser per fornire ai server dati non veritieri durante l'interazione http.

7. Protezione del browser tramite addons

I browser moderni consentono l'installazione da parte dell'utente di estensioni – o addons – volte ad arricchire l'esperienza d'uso del browser. Gli addons sono decine di migliaia e non sono sempre disponibili per tutti i browser, ma senza dubbio la più ampia disponibilità è per i browser Chrome e Firefox. Entrambi consentono di scegliere, all'interno di specifici repository appositamente predisposti, gli addons da scaricare ed installare, fornendo anche notizie sulla popolarità e sul gradimento degli stessi. Esistono addons di tutte le categorie e naturalmente non mancano quelli destinati ad incrementare la sicurezza del browser.
A puro scopo esemplificativo si riporta una lista di addons che sono in grado di dare una discreta protezione aggiuntiva al browser Firefox.

addon	*funzionalità offerte*
BetterPrivacy	contro i super-cookies
Ghostery	anti-tracciamento
HTTPS Everywhere	forza i browser a preferire connessioni criptate (sviluppato da Eff)
LightBeam	consente di vedere quali siti stanno collezionando dati sulla nostra navigazione
NoScript	consente di bloccare/limitare/controllare selettivamente l'esecuzione di Javascript nelle pagine
Self-Destructing Cookies	distrugge automaticamente i cookies collezionati dai siti
uBlock Origin	blocco della pubblicità

Gli utenti possono liberamente installare i vari addons: in molti casi la configurazione è guidata e l'utente sarà assistito in tutte le fasi, anche se la configurazione di NoScript appare significativamente più complessa e normalmente fuori dalla portata di un utente non tecnico. Un avvertimento: più sono gli addons e più lunghi saranno i tempi di avvio e di funzionamento del browser. Inoltre, in alcuni casi i comportamenti di alcuni addons potrebbero interferire.

8. Altre minacce: l'uomo nel mezzo

Ritornando ai principi cardine della sicurezza delle informazioni, ci poniamo a questo punto due domande. Quando si utilizza un browser per collegarsi a un sito remoto la nostra privacy è protetta? In altre parole, è possibile per un avversario osservare e registrare i dati in transito? Quando scambiamo dati con un server Web possiamo essere sicuri della loro integrità?

Si tratta di quesiti naturali, ma le risposte non sono sempre semplici. Iniziamo dal requisito della riservatezza. Il protocollo http che si impiega per lo scambio dei dati non protegge la riservatezza dei contenuti, quindi è relativamente semplice per un attaccante (ad esempio, un fornitore di servizio Internet o, nel caso di WiFi pubbliche, un attaccante a noi vicino) osservare i dati in transito, potendoli registrare. Per poter mitigare questa minaccia è consigliabile navigare su siti che supportano la versione sicura del protocollo http, denominata https. In tal caso il collegamento scambierà dati criptati e verificherà, mediante altre primitive crittografiche, l'integrità dei dati scambiati. Come riconoscere la presenza di un server che supporta https?

Basta sostituire nella barra degli indirizzi la scritta https alla scritta http e testare il funzionamento. Si noti che in assenza di connessione criptata tutti i dati sensibili trasmessi, come password e dati delle carte di credito, transiterebbero in chiaro e sarebbero facile preda degli attaccanti. Se vogliamo rendere automatico il passaggio sistematico a https possiamo installare l'addon HTTPS Everywhere (per Firefox e Chrome), distribuito da Eff. L'addon può essere anche configurato in maniera da bloccare le connessioni che non usano https.

L'uso di https potrebbe non essere tuttavia sufficiente a garantire riservatezza e integrità. L'acronimo https sta a indicare l'impiego di http veicolato attraverso un protocollo sicuro denominato TLS, oggi alla versione 1.2 e presto aggiornato alla versione 1.3. I browser moderni supportano pienamente TLS 1.2 ma accade di sovente che i server Web, a volte vecchi se non obsoleti, e soprattutto non aggiornati, funzionino con versioni precedenti del protocollo, oggi considerate non più sicure. In tali casi i browser dovrebbero impiegare versioni precedenti di TLS, aumentando la vulnerabilità della connessione.

Un altro punto critico sta nel certificato che il server TLS deve esibire. Si tratta di un certificato, inviato ai browser, che attesta quale sia la chiave pubblica del server. Tale chiave è un parametro indispensabile per il corretto funzionamento della crittografia e deve essere quella legittimamente associata al server TLS. Un attaccante potrebbe allestire un server TLS e, intercettando la trasmissione dati del nostro browser, rispondere al posto del server reale, usando la crittografia prevista da TLS e fornendo un certificato falso. Se il browser non si accorge che il certificato è falso l'attacco avrà successo e la trasmissione dati sarà completamente compromessa: si tratta di un attacco noto come man-in-the-middle (l'uomo nel mezzo), in cui l'attaccante intercetta tutte le comunicazioni avendo il potere di inoltrarle (eventualmente modificate) o meno e di mandare risposte prodotte a suo piacimento. I certificati sono di norma firmati con firma digitale[29] di una autorità di certificazione per cui il browser, ricevendone uno, verifica di norma che la firma sia autentica e che il certificato non sia stato successivamente revocato.[30] Se una di queste verifiche fallisce è possibile che si sia sotto attacco.[31]

Quando tutte le verifiche hanno successo e si impiega l'ultima versione di TLS la connessione, anche se intercettata dall'uomo nel mezzo, può ritenersi sicura perché l'attaccante non sarà in grado di decriptare la trasmissione né di insidiarne l'integrità. Anche in presenza di reti compromesse e computer pubblici di cui non si conosce l'affidabilità, una connessione https corretta garantirà la sicurezza delle informazioni in transito.

[29] La firma digitale, in caso di verifica con esito positivo, garantisce integrità del dato, identificazione del firmatario e non ripudio.

[30] La revoca di un certificato può avvenire per varie ragioni: chiave privata compromessa, funzione connessa alla firma non più rivestita dall'operatore, modifica della ragione sociale di un'azienda ecc.

[31] Talvolta l'attaccante esibisce un certificato compromesso che è stato in realtà revocato. Il browser, nel primo esame del certificato ne valida la firma; se l'operazione ha successo il browser contatta l'autorità di certificazione per verificare che il certificato non sia stato revocato, usando un protocollo denominato OCSP. È questo il momento in cui l'attaccante interviene bloccando l'operazione di verifica. Il browser, dopo un'attesa di alcuni secondi, comunicherà all'utente che il tentativo di collegamento via OCSP è andato in timeout, chiedendo se l'utente vuole proseguire. Purtroppo, dato il nome poco esplicativo del protocollo OCSP, molti utenti a questo punto rispondono sì, diventando vittime dell'attacco.

9. Protezione della privacy basata sull'anonimato

Negli ultimi anni si sono diffuse tecniche per collegarsi al Web in modo da essere anonimi. Per comprendere l'esatto significato del termine "anonimo" nel caso di interesse è opportuno rammentare che tutti di dispositivi che sono connessi a Internet sono caratterizzati da una sequenza di numeri, denominata indirizzo IP, che identifica il soggetto che è titolare della connessione a Internet. Per semplicità di trattazione non parleremo del caso in cui uno stesso indirizzo IP è condiviso da più soggetti, perché correlando le informazioni registrate negli archivi (dette *log*) dei vari apparati che consentono il funzionamento della rete, si riesce a identificare correttamente i soggetti.[32]

Un browser che si connette a un Web server (in http o https) mostrerà al server il suo IP, informazione che rivela il paese di provenienza e l'organizzazione che offre connettività al dispositivo.[33] L'indirizzo IP è dunque una prima informazione che consente attività di tracciamento. Il collegamento anonimo Internet consiste nell'adozione di opportune misure tecniche che consentono di fruire di servizi Internet mostrando un indirizzo IP completamente differente da quello reale, potendo ad esempio collegarsi da una città italiana verso un server che si trova in un'altra città italiana e mostrando a tale server un IP americano o asiatico. Ciò può essere ottenuto essenzialmente attraverso due metodi: il ricorso a un proxy server oppure l'impiego del browser Tor.

9.1. Proxy e VPN

Con il termine *proxy* si indica un dispositivo che svolge le funzioni di intermediario: nel momento in cui vogliamo contattare un server su Internet possiamo incaricare un intermediario (il proxy, appunto) di farlo al posto nostro in modo da disaccoppiare l'interazione fra noi e il server remoto: basterà configurare sul nostro dispositivo la presenza di un proxy, specificandone i parametri.[34] Il servizio di proxy è spesso associato a un servizio parallelo di VPN (Virtual Private Network) che consiste in un collegamento sicuro – criptato, autenticato e integro – fra il dispositivo e il proxy, in modo da vanificare qualunque tipo di attacco nel tratto fra il dispositivo il proxy.[35] Il collegamento fra il proxy e il server remoto avverrà secondo il protocollo originariamente richiesto dal browser del nostro dispositivo. Lato server, si vedranno richieste provenienti dall'IP del proxy, in genere completamente avulso da quello dell'utente reale. Per aumentare la sicurezza nei confronti dei progetti di sorveglianza massiva il proxy viene frequentemente scelto in un altro paese, soggetto quindi a leggi differenti e soprattutto con una magistratura indipendente. I servizi di VPN sono molto popolari nei paesi in cui l'accesso a Internet è limitato, sorvegliato o posto sotto censura.

9.2. Il browser Tor

Negli anni '90 lo United States Naval Research Laboratory varò un progetto per il collegamento anonimo alla rete Internet con lo scopo di proteggere le comunicazioni online dell'intelligence americana. Dal 2006 il progetto è gestito da The Tor Project, organizzazione non-profit con

[32] Ciò richiede l'autorizzazione della magistratura.

[33] Il protocollo whois consente di recuperare tali informazioni, anche se è oggi possibile oscurarne una parte, in modo che non siano direttamente visibili nomi di referenti, numeri di telefono ed indirizzi email/postali.

[34] Il servizio è di norma a pagamento.

[35] Il servizio VPN è di norma a pagamento e include il servizio di proxy.

sede in Massachusetts. Fin dal 2004 il progetto ha visto il supporto attivo di Eff, sia in termini economici che organizzativi. Con il trascorre degli anni altre organizzazioni hanno iniziato a sovvenzionare il progetto, con una netta prevalenza di organizzazioni statunitensi, nonostante la contrarietà di alcune agenzie di sicurezza nazionale dello stesso paese.[36] In buona sostanza Tor è un browser costruito sull'impalcatura di Mozilla Firefox che si connette al sito desiderato dopo aver attraversato la rete seguendo un percorso criptato e generato casualmente dal browser e sconosciuto a tutti gli altri attori, inclusi i nodi casualmente prescelti dal browser, che sanno di appartenere a un percorso segreto di cui conoscono solo il nodo precedente e il successivo, nodi con cui scambiano dati criptati. L'ultimo nodo del percorso, chiamato nodo di uscita, è quello che instraderà la comunicazione verso la destinazione finale: il collegamento sarà criptato solo se il server funziona in TLS, altrimenti sarà in chiaro. Nelle tratte precedenti i collegamenti sono tutti criptati, come illustrato in Fig. 9. La lunghezza del percorso è generalmente cinque, con quattro nodi Tor tipicamente coinvolti nella comunicazione. Il server destinazione del collegamento vede richieste provenienti dall'IP del nodo di uscita e non conosce l'IP del nodo iniziale. Vale la pena anche sottolineare che le informazioni veicolate da Tor attraverso il protocollo http riguardo il tipo di browser, sistema operativo, versione ecc., risultano inattendibili, in modo da rendere vano il tentativo di operare attraverso il browser fingerprinting. Infine, si segnala la possibilità di creare siti Web all'interno della rete Tor, raggiungibili solo utilizzando il browser Tor e conoscendone l'indirizzo. Questi siti non possono essere raggiunti dagli altri browser e costituiscono il cosiddetto *dark web*.[37] Il browser Tor è disponibile per tutti i sistemi operativi, sia per dispositivi fissi che mobili e può essere scaricato dal sito Web del progetto. Funziona correttamente anche in presenza di un collegamento VPN.

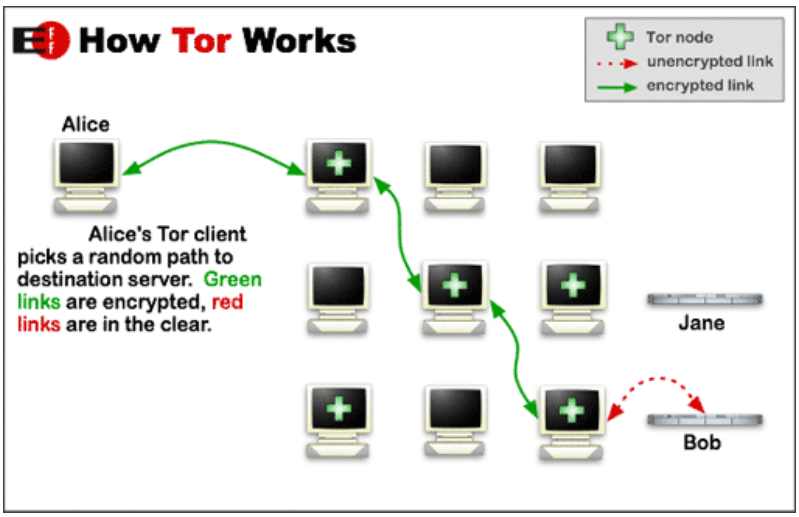

Figura 9. Infografica sul funzionamento di Tor (fonte: Eff).

L'alto grado di anonimato garantito da Tor ha fatto sì che l'utenza tipica di Tor oggi includa, oltre a intellettuali, attivisti, giornalisti, persone sotto censura o controllo, anche criminali e terroristi.

10. Conclusioni

Il percorso qui proposto consente di acquisire i principi fondamentali a cui ispirarci nell'esame

[36] Per un approfondimento sulla storia del progetto e sui suoi aspetti tecnici si possono utilmente consultare le pagine di Wikipedia https://en.wikipedia.org/wiki/Tor_(anonymity_network) e quelle del progetto https://www.torproject.org

[37] Viene normalmente indicato con il termine *deep web* quella parte del Web non raggiungibile dai motori di ricerca. Non esistono stime attendibili sulla sua dimensione ma si ipotizzano alcuni ordini di grandezza superiori a quella del Web. Il *dark web* è la parte del deep web raggiungibile solo attraverso il browser Tor.

critico delle opportunità concesse dalla rete, numerose e diversificate, e soprattutto in continua evoluzione. Il corretto recepimento dei principi è lo strumento fondamentale per inquadrare correttamente gli elementi di uno scenario che sfugge ai tentativi di classificazione, proprio per la sua elevata dinamicità.

La trattazione ha permesso di delineare alcune raccomandazioni nell'uso del browser che possono essere sinteticamente così riassunte: configurare il browser in modo che rifiuti i cookies di terze parti, ricorrere frequentemente alla navigazione privata, usare finestre con una sola scheda, privilegiare i collegamenti https e magari installare estensioni del browser che ne facilitano la messa in sicurezza. È tuttavia importante rammentare che neanche il browser più sicuro potrà proteggere l'utente che decide di adottare comportamenti potenzialmente lesivi della privacy e della reputazione: ciò richiede una sensibilizzazione a livello concettuale più alto, che induca a proteggere con le stesse cautele e attenzioni gli altri come noi stessi, altrimenti le misure dei singoli a protezione della sola privacy personale non potranno essere sufficienti.

Il percorso qui iniziato merita naturalmente maggiori approfondimenti e l'inclusione di ulteriori temi come, ad esempio, la messa in sicurezza dell'email, l'analisi del social engineering o l'educazione a una vita digitale costellata da PIN e password e soprattutto ai sistemi di autenticazione più evoluti. Più di tutto, occorre imparare che la sicurezza è un processo e come tale non può considerarsi un adempimento soddisfatto mediante meri approfondimenti episodici.