Anonimato in Rete

Dai Cookie a Tor

Nicola Bena

nicola.bena@unimi.it
https://homes.di.unimi.it/bena/

https://www.linkedin.com/in/nbena

Dipartimento di Informatica
Università degli Studi di Milano

11 Dicembre 2023

Introduzione

Anonimato e Pseudoanonimato

Dati pseudoanonimi

Dati che non possono essere ricondotti all'interessato *senza informazioni aggiuntive* [...]

Dati anonimi

Dati che non possono essere ricondotti all'interessato

Obiettivo

Ci concentreremo sui dati pseudoanonimi, poiché è molto difficile ottenere dati completamente anonimi (esistono *quasi* sempre *altre* informazioni...)

Nel contesto di Internet, queste informazioni aggiuntive sono *metadati* generati/utilizzati durante l'esecuzione dei protocolli di rete

Obiettivo

Ci concentreremo sui dati pseudoanonimi, poiché è molto difficile ottenere dati completamente anonimi (esistono *quasi* sempre *altre* informazioni...)

Nel contesto di Internet, queste informazioni aggiuntive sono *metadati* generati/utilizzati durante l'esecuzione dei protocolli di rete

- ...evitando i casi di dati resi pubblici *volontariamente*

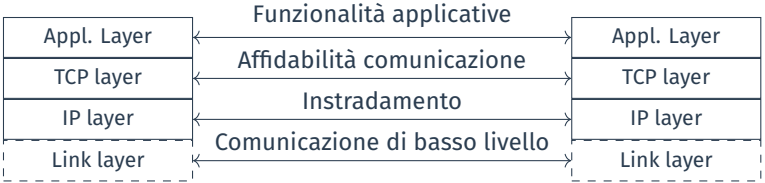
Protocolli (1)

Un protocollo definisce regole precise su come deve avvenire la comunicazione tra diverse parti

Definisce:

- formato dei messaggi scambiati
  - header: informazioni necessarie al funzionamento del protocollo
  - payload/body: informazioni che effettivamente si vogliono scambiare
- ordine ed azioni da intraprendere dopo invio/ricezione

Protocolli (2)



IP e DNS

- IP: fornisce indirizzamente univoco, cioè identificazione univoca degli host di una rete tramite un indirizzo numerico chiamato indirizzo IP
  - in una comunicazione, l'indirizzo IP dell'altro host è sempre ottenibile perché parte dei dati scambiati (header IP)

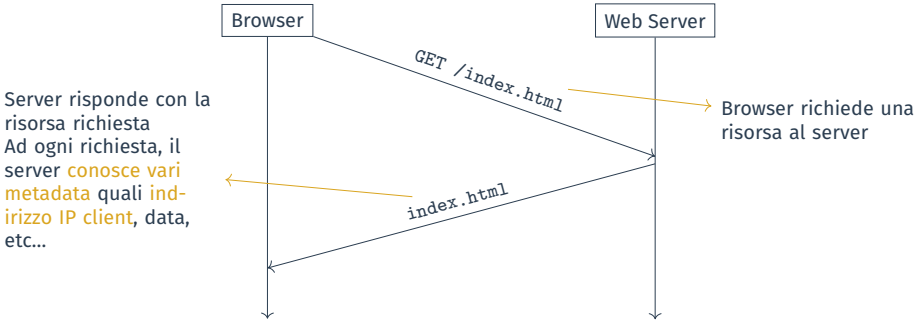
- DNS: traduce nomi mnemonici in indirizzi IP e viceversa
  - nomi mnemonici usati dagli essere umani per riferirsi agli host, e.g., `amazon.it`
    - le URL dei siti web contengono il nome DNS del sito
  - indirizzi numerici usati dagli host durante la comunicazione, e.g., `52.95.116.114`

Ogni volta che si digita una URL nel browser, essa viene *risolta*, ovvero si ottiene il corrispondente indirizzo IP

Il protocollo HTTP

Un protocollo richiesta – risposta

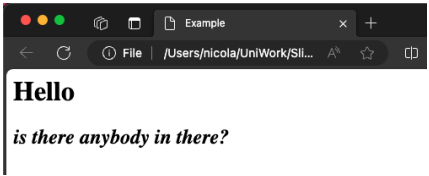
- i file HTML sono memorizzati sul web server
- il browser richiede una risorsa al server (file HTML)
- la risorsa viene trasferita mediante il protocollo HTTP



HTML

Le pagine web sono realizzate in un formato testuale secondo una sintassi particolare (HTML), che i browser interpretano in una versione grafica

```
<html>
    <head>
        <title>Example</title>
    </head>
    <body>
        <h1>Hello</h1>
        <h2><i>is there anybody in
        ↪  there?</i></h2>
    </body>
</html>
```



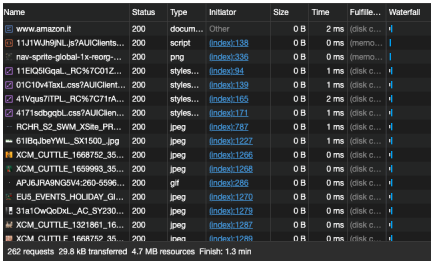
HTTP e HTML

Un file HTML può contenere link ad altre risorse (immagini, video...)

- il browser effettua una richiesta per ogni altra risorsa contenuta nel file HTML
- possono provenire anche da altri siti

≈ 262 richieste diverse quando si visita `https://amazon.it`

- Chrome/Edge: *More tools* ⟹ *Developer tools*



Introduzione:
Cookie

Cookie

Problema: diverse coppie di richieste e risposte HTTP sono indipendenti tra loro, ma occorre gestire il concetto di sessione

Idea: aggiungere delle informazioni memorizzate lato client, che vengono inviate dal browser al server ad ogni successiva richiesta

$\Longrightarrow$ cookie

- memorizzati dal browser
- inviati negli header HTTP dal browser

Formato

Un cookie è formato (almeno) da una coppia nome–valore, fissati dal server

Ulteriori campi:

- `Expires`: per quanto tempo il browser deve memorizzare il cookie
  - se non specificato, viene cancellato alla chiusura del browser (session cookie)
- ...
- tutti i valori sono specificati dal web server ed il browser li deve onorare

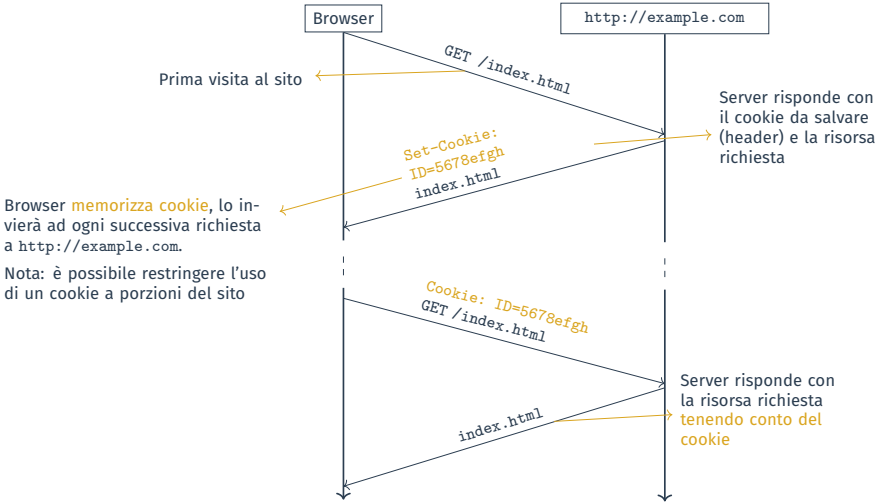
Ciclo di Vita

1. Browser si connette per la prima volta ad un sito: non ha cookie da inviare
2. Nella risposta, il server chiede al browser di memorizzare un cookie, specificandone il valore
3. A partire dalle successive richieste, il browser invia il cookie che ha memorizzato

Scope

Un cookie memorizzato da un sito può essere re-inviato dal browser esclusivamente al sito stesso

Ciclo di Vita: Esempio



Utilizzo

Uno degli usi principali dei cookie è la gestione delle sessioni tra richieste e risposte multiple

1. ritornato dal server se le credenziali inviate dal client sono corrette
2. re-inviato ad ogni successiva richiesta per due scopi:
   - come *prova* che il client si è autenticato
   - per servire risorse dipendenti dal cookie
3. eventualmente memorizzato a lungo termine in stile "*ricorda credenziali*"
   - per "rimanere loggati"

Tipologie di Cookie

Classificazione in base alla persistenza

- *cookie di sessione*: cancellati alla chiusura del browser
- *cookie persistenti*: permangono tra diversi sessioni di navigazione

Classificazione in base alla provenienza

- *first-party cookie*: provengono dal sito che effettivamente si sta visitando
- *third-party cookie*: provengono da altre risorse caricate durante la navigazione nel sito principale

Anonimato in Rete: Problemi

Problemi

Tre tipi di problemi (at least):

- esecuzione dei protocolli, durante la quale numerose informazioni vengono scambiate
  - ideate per scopi funzionali, ma *spesso abusate*
  - sia a livello applicativo, sia ai livelli inferiori
- browser, le cui funzionalità possono essere sfruttate per carpire informazioni
- utente, genuinamente inconsapevole di ciò

Le informazioni direttamente rilevabili possono portare a dedurre altre informazioni

Anonimato in Rete: Problemi: Livelli Inferiori (IP e DNS)

IP

L'indirizzo IP è un buon indicatore della posizione geografica di un host, poiché gli indirizzi vengono *assegnati* da enti (ISP) con competenza territoriale

- ciascun ISP possiede un intervallo di indirizzi IP assegnabili ai propri clienti su base territoriale

What Is My IP Address

My IP Address :

IP address	151.55.202.154
Hostname	n/a

IP Address Location :

Country	Italy (IT)
State/Region	Lombardy
City	Rho
ISP	INFOSTRADA
Organization	WIND Telecomunicazioni S.p.A
Network	AS1267 WIND TRE S.P.A. (DSL, TOR, MOBILE, BIZNET)
Usage Type	Cable/DSL / Residential
Timezone	Central European Time (CET)
Local Time	Sat, 09 Dec 2023 10:15:47 +0100
Coordinates	45.5190,9.0862

`https://browserleaks.com/ip`

DNS

Il protocollo DNS non è cifrato

- $\Longrightarrow$ il traffico DNS può essere facilmente intercettato (e modificato)

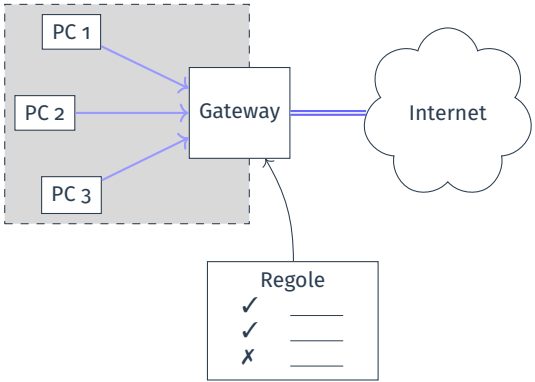
Nel processo di risoluzione di solito si utilizza sempre lo stesso server DNS

- se non configurato diversamente, è il DNS del nostro ISP
- $\Longrightarrow$ quel server vede quali sono tutti i siti visitati che visitiamo

Router e Firewall

Ogni rete ha un dispositivo di confine (*firewall*) che gestisce le richieste interno $\iff$ Internet

- $\implies$ tutti i pacchetti transitano per il gateway
- $\implies$ usato per sicurezza perimetrale, specificando quale traffico è/non è consentito da/verso Internet



Router e Firewall

Ogni rete ha un dispositivo di confine (*firewall*) che gestisce le richieste interno $\iff$ Internet

- $\implies$ tutti i pacchetti transitano per il gateway
- $\implies$ usato per sicurezza perimetrale, specificando quale traffico è/non è consentito da/verso Internet

In generale, il nostro traffico passa per diversi router prima di arrivare a destinazione

Attenzione

Anche se il traffico è cifrato ad alto livello (e.g., TLS), gli header di livello 3 e 4 sono tipicamente in chiaro e quindi visibili

Anonimato in Rete: Problemi:
Livello Web

Cookie di Terze Parti (1)

I cookie di terze parti pongono seri problemi di privacy

1. due siti diversi (A, B) contengono un link ad una risorsa di C
2. l'utente visita per la prima volta A, il browser fa anche la richiesta a C
   - $\Longrightarrow$ la risposta di C contiene la richiesta di settare un cookie
3. l'utente visita per la prima volta B, il browser fa anche la richiesta a C
   - $\Longrightarrow$ nel fare la richiesta a C, il browser invia il cookie settato precedentemente

$\Longrightarrow$ ...

Cookie di Terze Parti (2)

È un problema perché *C* può sapere *da dove proviene la richiesta*, cioè da *A* e *B*...

...quindi sapere che quel cookie è associato a qualcuno che ha visitato *A* e *B*

Example

A:

```
<img src="https://ads.com?from=amazon-product-1"/>
```

B:

```
<img src="https://ads.com?from=alcol-recovery.com"/>
```

Un Esempio Ancora Più Concreto

Visitando `https://aranzulla.it` per la prima volta (dopo aver accettato i cookie):

- `securepubads.g.doubleclick.net`
- `cdn-gl.imrworldwide.com`
- `wwww.googletagmanager.com`
- `utils.cedsdigital.it`
- `fundingchoicesmessages.google.com`
- `c09nxxta4mb8at7adzutpbehwqc23170 2114300.nuid.imrworldwide.com`
- `caltagironeeditore01.wt-eu02.net`
- `region1.analytics.google.com`
- `stats.g.doubleclick.net`
- `www.google.it`
- `hits-i.iubenda.com`
- `81596417ee7882bd9b3745d1af8006fd .safeframe.googlesyndication.com`
- …

Richieste bloccate da Edge:

Domain	Type
responder.wt-safetag.com	script
secure-it.imrworldwide.com	gif
secure-it.imrworldwide.com	gif
secure-it.imrworldwide.com	gif
secure-it.imrworldwide.com	gif
secure-it.imrworldwide.com	gif
secure-it.imrworldwide.com	gif

Un Esempio Ancora Più Concreto

Visitando `https://aranzulla.it` per la prima volta (dopo aver accettato i cookie):

- `securepubads.g.doubleclick.net`
- `cdn-gl.imrworldwide.com`
- `wwww.googletagmanager.com`
- `utils.cedsdigital.it`
- `fundingchoicesmessages.google.com`
- `c09nxxta4mb8at7adzutpbehwqc23170 2114300.nuid.imrworldwide.com`
- `caltagironeeditore01.wt-eu02.net`
- `region1.analytics.google.com`
- `stats.g.doubleclick.net`
- `www.google.it`
- `hits-i.iubenda.com`
- `81596417ee7882bd9b3745d1af8006fd .safeframe.googlesyndication.com`
- …

13 risposte con `Set-Cookie` (6 nel 2020)

Provate i *developer tools* del vostro browser

- nota: non tutti i campi sono visibili da subito, occorre abilitare quelli che volete mostrare facendo right click sull'header della tabella

Un Esempio Ancora Più Concreto

Visitando `https://aranzulla.it` per la prima volta (dopo aver accettato i cookie):

- `securepubads.g.doubleclick.net`
- `cdn-gl.imrworldwide.com`
- `wwww.googletagmanager.com`
- `utils.cedsdigital.it`
- `fundingchoicesmessages.google.com`
- `c09nxxta4mb8at7adzutpbehwqc23170 2114300.nuid.imrworldwide.com`
- `caltagironeeditore01.wt-eu02.net`
- `region1.analytics.google.com`
- `stats.g.doubleclick.net`
- `www.google.it`
- `hits-i.iubenda.com`
- `81596417ee7882bd9b3745d1af8006fd .safeframe.googlesyndication.com`
- …

Cookie Amazon: `https://www.amazon.it/cookieprefs/partners`

Anonimato in Rete: Problemi:
Browser

Browser

Header `User-Agent`: stringa che identifica il browser stesso

- Firefox: `Mozilla/5.0 (Macintosh; Intel Mac OS X 10.15; rv:109.0) Gecko/20100101 Firefox/109.0`
- Edge: `Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/120.0.0.0 Safari/537.36 Edg/120.0.0.0`
- fornisce anche informazioni sul sistema operativo

Accessibile anche via console: `window.navigator.userAgent`

Fingerprinting

L'estrazione di informazioni basandosi sulle informazioni messe a disposizione dal browser, al fine di identificare un utente

Combinazione di tutti i parametri ci porta ad essere facilmente identificabili...

Fingerprinting

L'estrazione di informazioni basandosi sulle informazioni messe a disposizione dal browser, al fine di identificare un utente

Combinazione di tutti i parametri ci porta ad essere facilmente identificabili...

- 83.6% dei browser ha una fingerprint univoca[a]

[a] Fonte: P. Eckersley. *How Unique Is Your Web Browser?*. 2010

How Is It Possible?

Il browser espone una serie di funzionalità, la cui configurazione e modo d'uso possono rivelare varie informazioni

- codice lato client
- WebRTC
- Canvas
- ...

Questi considerazioni valgono *anche* per i protocolli (e.g., protocollo configurato in maniera peculiare)

Code Lato Client (JS)

Varie informazioni possono essere rilevate dal codice JS client-side...

Document Object :	
Document Referrer	*empty* [recheck]
Screen Object :	
Screen Resolution	3440×1440 (35mm anamorphic post-1970) 24-bit TrueColor (viewport: 304×875) more
Date/Time :	
System Time	Sat Dec 09 2023 10:58:41 GMT+0100 (Central European Standard Time)
toLocaleString	09/12/2023, 10:58:41
toLocaleFormat	*undefined*
Internationalization API :	
DateTimeFormat	Saturday, 9 December 2023 at 10:58:41 am Central European Standard Time
hourCycle	h23
locale	en-GB
calendar	gregory
numberingSystem	latn
timeZone	Europe/Rome
year	numeric
month	2-digit
day	2-digit

navigator.userAgentData (Client Hints) :	
API Status	✔ Enabled
brands	[{"brand":"Not_A Brand","version":"8"}, {"brand":"Chromium","version":"120"}, {"brand":"Microsoft Edge","version":"120"}]
mobile	false
platform	macOS
platformVersion	14.1.2
architecture	arm
bitness	64
wow64	false
model	*empty*
uaFullVersion	120.0.2210.61
fullVersionList	[{"brand":"Not_A Brand","version":"8.0.0.0"}, {"brand":"Chromium","version":"120.0.6099.71"}, {"brand":"Microsoft Edge","version":"120.0.2210.61"}]

`https://browserleaks.com/javascript`

Altro

- Canvas: un metodo per *costruire disegni ed animazioni* all'interno della pagina web, il *modo* in cui il disegno viene effettivamente rappresentato è indice delle caratteristiche del dispositivo
- WebGL
- WebRTC
- ...

Sulla base delle informazioni raccolte posso provare a capire altro...

- dimensione schermo $\Longrightarrow$ power user
- browser poco noto $\Longrightarrow$ *smanettone*
- Linux $\Longrightarrow$ *molto* smanettone
- ...

Hands On (When the Seminar Ends!)

Novembre 2020, Firefox su Ubuntu

Your Results

Within our dataset of several hundred thousand visitors tested in the past 45 days, only **one in 116192.5 browsers have the same fingerprint as yours.**

Currently, we estimate that your browser has a fingerprint that conveys **16.83 bits of identifying information.**

`https://coveryourtracks.eff.org/`

Hands On (When the Seminar Ends!)

Dicembre 2023, Edge su MacOS

Your Results

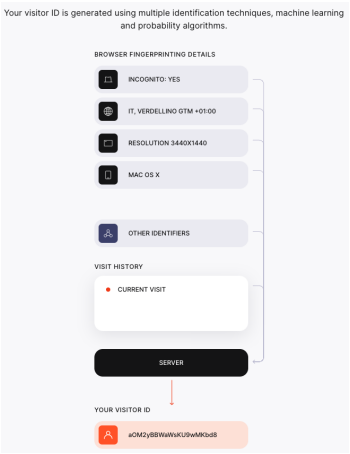
Your browser fingerprint **appears to be unique** among the 204,041 tested in the past 45 days.

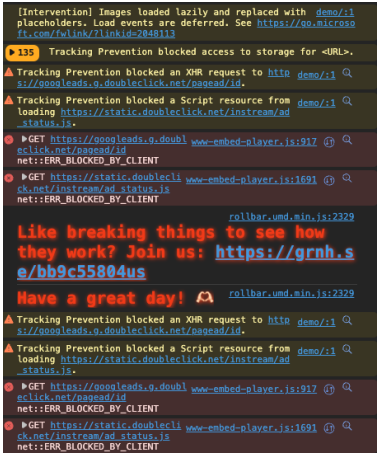
Currently, we estimate that your browser has a fingerprint that conveys **at least 17.64 bits of identifying information.**

`https://coveryourtracks.eff.org/`

Hands On (When the Seminar Ends!)

https://fingerprint.com/





Nota Finale

Il server che contiene la pagina web che si sta visitando ha sempre accesso a tutte queste informazioni

- i *pacchetti* attraversano diversi nodi intermedi per arrivare a destinazione
- $\Longrightarrow$ se non vi è cifratura, anche tutti questi nodi possono (potenzialmente) accedere a tali informazioni
- HTTPS sempre più usato

Quindi?

Possiamo fare qualcosa?

Soluzioni

Soluzioni

Diverse soluzioni consentono di introdurre un certo livello di (psuedo)anonimato

Classificazione in base al livello a cui operano:

- *rete*: nascondere l'indirizzo IP sorgente
- *applicazione*: nascondere informazioni fornite dal browser

Spoiler: occorrono entrambe

Soluzioni

Diverse soluzioni consentono di introdurre un certo livello di (psuedo)anonimato

Vedremo:

- proxy
- VPN
- Tor
- Cenni ad altre tecniche

Soluzioni:
Proxy

Proxy

Un proxy è un server che fa da intermediario in una comunicazione tra client e server finale

Quando deve contattare un server, il client passa tramite il proxy

- il proxy interagisce con il server per conto del client
  - per il server, il mittente della comunicazione è il proxy
- il proxy riceve le risposte dal server e le inoltra al client

Proxy

Un proxy è un server che fa da intermediario in una comunicazione tra client e server finale

Quando deve contattare un server, il client passa tramite il proxy

- il proxy interagisce con il server per conto del client
  - per il server, il mittente della comunicazione è il proxy
- il proxy riceve le risposte dal server e le inoltra al client
  - $\Longrightarrow$ il server non conosce il mittente reale

Dettagli

Il proxy ha comprensione del livello applicativo, ma non è generico

- $\Longrightarrow$ per ogni protocollo applicativo occorre un proxy specifico
- spesso non è trasparente, le applicazioni devono essere configurate per utilizzarlo
- protocollo `SOCKS`: protocollo spesso usato per la comunicazione tra client e proxy, in alternativa a proxy HTTP

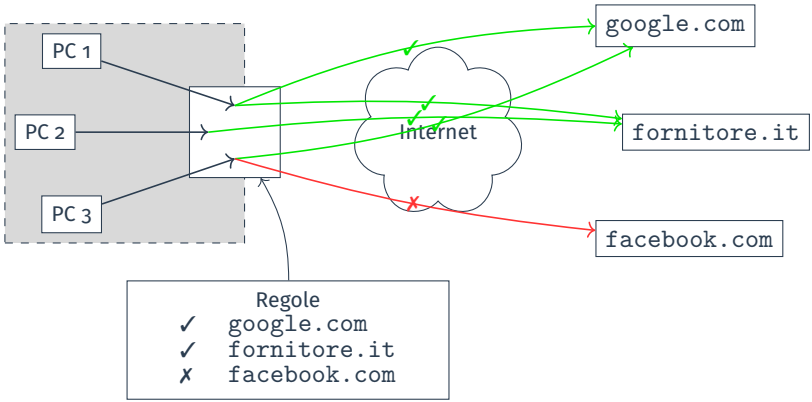
Utilizzi

I proxy server vengono usati per vari scopi:

- performance: il proxy mantiene una cache (*memoria*) delle richieste/risposte, evitando di contattare il server se ha già una risposta memorizzata
- sicurezza: possibilità di implementare politiche di controllo
  - blocco URL
- privacy: nascondere il mittente

Esistono anche i reverse proxy, utilizzati per consentire richieste dall'esterno verso l'interno

Esempio


PC 1
PC 2
PC 3
Internet
google.com
fornitore.it
facebook.com
Regole
✓ google.com
✓ fornitore.it
✗ facebook.com

Proxy e Anonimato

I proxy aiutano ad ottenere un certo grado di anonimato:

- nascondono il mittente originale della comunicazione...
- confusione: più client usano lo stesso server proxy...

Proxy e Anonimato

I proxy aiutano ad ottenere un certo grado di anonimato:

- nascondono il mittente originale della comunicazione...
- confusione: più client usano lo stesso server proxy...
- ...ma tutto il traffico dei client passa dal proxy

Attenzione: possono essere usati anche per filtraggio e analisi del traffico

Soluzioni:
VPN

Problema

Le reti aziendali sono protette da un firewall, eventualmente realizzato mediante un proxy, che di norma non consente l'accesso dall'esterno verso risorse interne

- server di dominio
- database
- ...

Come accedere a queste risorse senza esporle su Internet?

VPN

In generale, una VPN (*Virtual Private Network*), collega tra loro diverse reti ad un livello basso (2 o 3)

- *unendo* virtualmente le diverse reti
- facendole apparire come *contigue*
- $\Longrightarrow$ consentendo di accedere alle risorse di ciascuna delle reti collegate, poiché è come se fossero accedute dall'*interno*

Classificazione (1)

In base alle garanzie di sicurezza fornite

- Trusted VPN: il provider Internet garantisce isolamento del collegamento tra le varie reti
- Secure VPN: il collegamento tra le reti è sicuro per mezzo della crittografia
- Hybrid VPN: in certi tratti sono Trusted, in altri Secure

Classificazione (2)

In base al livello a cui operano

- livello 2: le diverse reti partecipanti vengono unite in un'unica rete
  - possono esserci conflitti
  - $\Longrightarrow$ le reti devono stare nella stessa rete livello 3 (*stesso NET ID*)
- livello 3: le diverse reti partecipanti hanno raggiungibilità completa ma vengono mantenute separate
  - preferibile, possono comunque esserci conflitti
  - $\Longrightarrow$ le reti devono stare in diverse reti livello 3 (*diversi NET ID*)

Classificazione (3)

In base alla topologia che realizzano

- Site-to-Site: diverse reti sono collegate tra loro
- Remote access: un singolo host collegato ad una rete

Dietro la quinte la VPN vera e propria funziona allo stesso modo

Protocolli VPN

Le diverse soluzioni VPN definiscono un proprio protocollo di comunicazione

- formato dei pacchetti
- utilizzo della crittografia

I principali sono:

- IPsec: complesso
- OpenVPN: alternativa de-facto a IPsec
- Wireguard: soluzione di nuova generazione
- SSTP: Microsoft, integrato in Windows

Terminologia

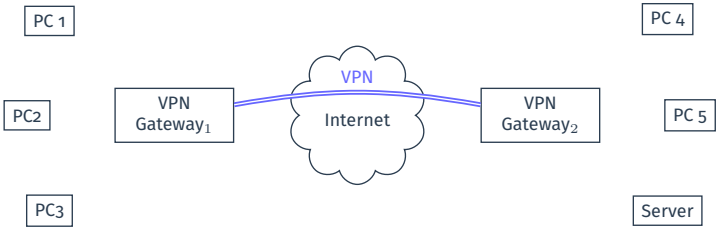
- Incapsulamento: incapsulare un pacchetto significa *inserire* il pacchetto nel *payload* di un altro pacchetto
- Gateway un host nella propria rete a cui gli altri host della stessa rete inviano pacchetti diretti ad altre reti
- Default gateway: host verso cui si inviano pacchetti diretti verso *tutte* le altre reti
  - esempio: a casa è il router ADSL/fibra

Intuizione

In ciascuna rete che si vuole connettere in VPN si posiziona un host che funge da gateway per gli host della propria rete *verso* le altre reti connesse in VPN

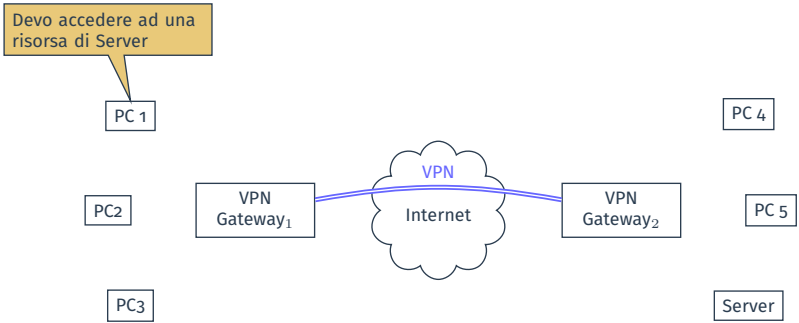
- ciascun pacchetto diretto alle altre reti connesse in VPN devono passare per il gateway che si trova nella propria rete
- il gateway *incapsula* i pacchetti ricevuti e li invia all'altro gateway
- l'altro gateway li riceve, li *de-incapsula*, e li invia all'host cui sono destinati
- il collegamento tra i gateway è cifrato e sicuro

Funzionamento



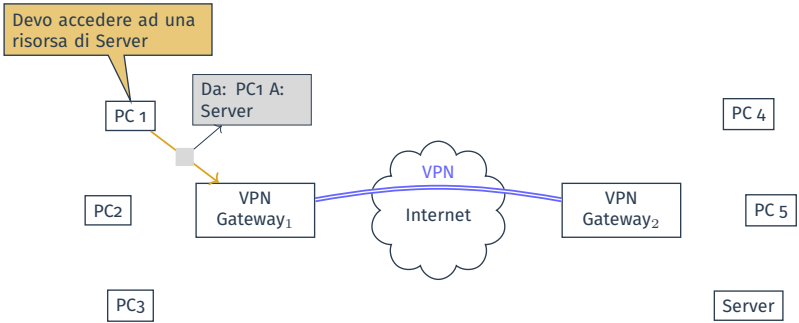
1. si stabilisce un canale sicuro tra i due gateway

Funzionamento



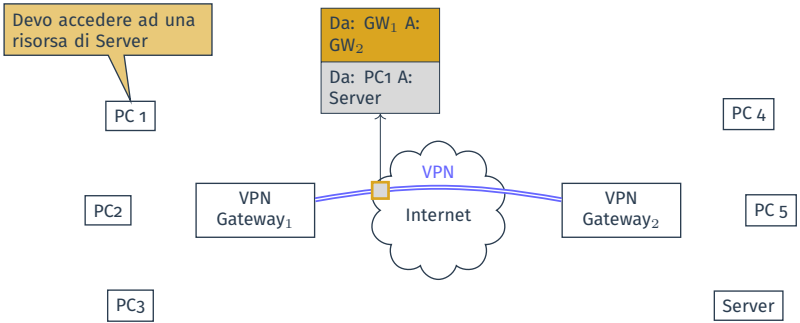
2. PC_1 sa che per raggiungere Server deve passare tramite GW_1

Funzionamento



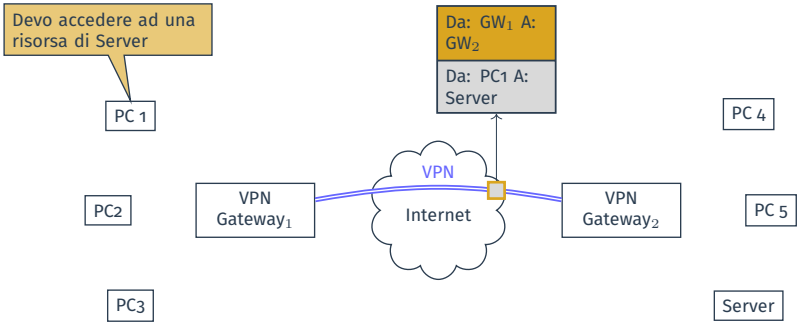
3. PC_1 invia il pacchetto per Server a GW_1

Funzionamento



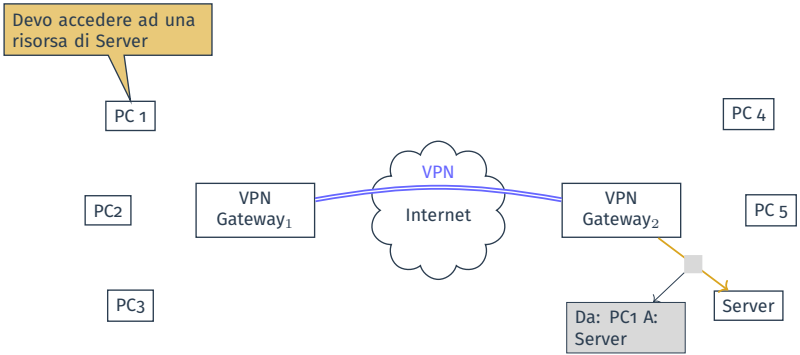
4. GW_1 incapsula il pacchetto all'interno del pacchetto destinato a GW_2, cifrandolo

Funzionamento



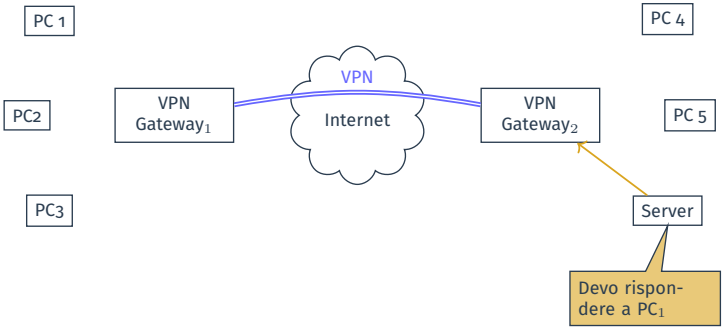
4. GW_1 incapsula il pacchetto all'interno del pacchetto destinato a GW_2, cifrandolo

Funzionamento



5. GW_2 riceve, decifra e de-incapsula il pacchetto ricevuto, ottenendo il pacchetto originale
6. il pacchetto originale viene inviato al destinatario

Funzionamento



7. le risposte seguono il percorso opposto

VPN con NAT

La VPN mostrata è una *classica* Site-to-Site VPN, una variazione è la Site-to-Site con NAT, in cui i pacchetti ricevuti da GW_2 vengono modificati, facendoli apparire come generati da GW_2

- GW_2 implementa una tecnica chiamata NAT (*Network Address Translation*) per gestire diversi client e far sì che i pacchetti appaiano generati da esso stesso
- GW_2 mantiene una *tabella* in cui tiene traccia delle diverse connessioni
- il NAT consente di evitare conflitti di indirizzi tra le diverse reti

Remote Access

Le VPN sono usate anche per realizzare accesso remoto

- un singolo host viene connesso in VPN ad una rete remota, consentendogli di accedere alle risorse della rete remota
- spesso GW_2 utilizza NAT per evitare conflitti di indirizzi
- GW_1 è un software all'interno dell'host che si connette in accesso remoto

Server e Client

VPN Client e VPN Server sono termini spesso usati

- Site-to-Site: inutile distinguere tra chi è client e chi è server, visto che sia GW_1 sia GW_2 sono client e server contemporaneamente
  - client quando si connette all'altro gateway per inviare pacchetti...
  - allo stesso tempo è un server per gli altri host della propria rete...
  - ed anche è un server quando riceve pacchetti dall'altro gateway
- Remote Access: distinzione sensata
  - gateway della rete remota: server
  - gateway locale: client

VPN e Anonimato

Se:

- tutto il traffico di un host viene inviato in VPN...
- ricevuto dal VPN gateway remoto ed immesso su Internet mediante NAT (ovvero *come se fosse generato dal server*)...
- tale traffico appare effettivamente generato dal gateway remoto
- $\Longrightarrow$ le VPN possono essere anche usate per ottenere privacy/pseudoanonimato

Garanzie

- Mittente nascosto: le richieste verso il destinatario finale appaiono generate dal VPN server, quindi non è possibile sapere chi sia il mittente reale
- *Confusione*: per il VPN server passano le richieste di numerosi client
- Ulteriori garanzie di sicurezza (confidenzialità, integrità, etc...) della comunicazione tra client e VPN server

$\Longrightarrow$ il tracking basato su indirizzi IP è molto più difficile

- possibile bypassare filtri basati su indirizzi IP (es: geo-blocking)

Garanzie

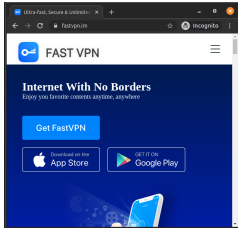
- Mittente nascosto: le richieste verso il destinatario finale appaiono generate dal VPN server, quindi non è possibile sapere chi sia il mittente reale
- *Confusione*: per il VPN server passano le richieste di numerosi client
- Ulteriori garanzie di sicurezza (confidenzialità, integrità, etc...) della comunicazione tra client e VPN server

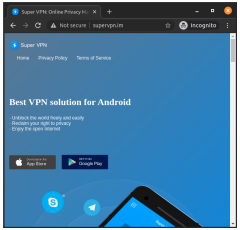
$\Longrightarrow$ il tracking basato su indirizzi IP è molto più difficile

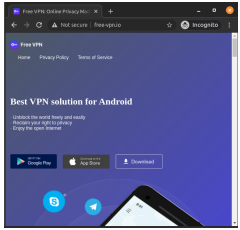
- possibile bypassare filtri basati su indirizzi IP (es: geo-blocking)

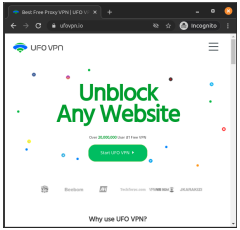
$\Longrightarrow$ il VPN server vede (quasi) tutto

VPN e Anonimato (?)









Report: No-Log VPNs Reveal Users' Personal Data and Logs

Updated on 7th September 2020

A group of free VPN (virtual private network) apps left their server completely open and accessible, exposing private user data for anyone to see. This lack of basic security measures in an essential part of a cybersecurity product is not just shocking. It also shows a total disregard for standard VPN practices that put their users at risk.

Soluzioni:
Tor

Idea di Base (1)

Tor è una rete di overaly con l'obiettivo di fornire anonimato e non tracciabilità della comunicazione

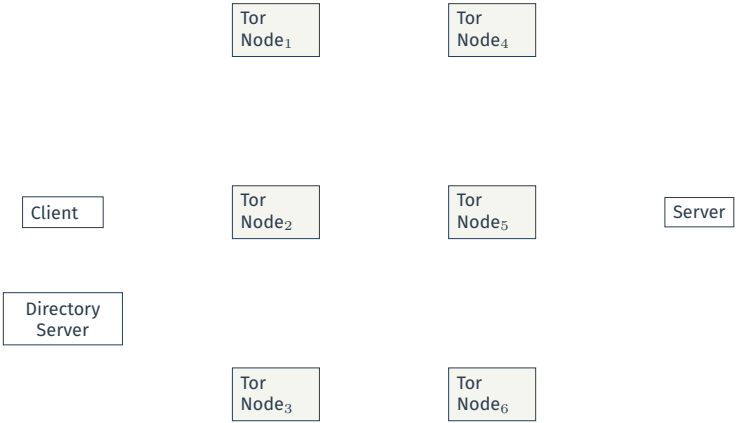
- numerosi nodi Tor sono disponibili su Internet
- per andare da *A* a *B*, un utente costruisce un percorso (*circuito*) che coinvolge diversi nodi Tor
  - si avvale di *directory server* fidati che conoscono lo stato della rete
- ogni nodo in un circuito conosce solo il precedente ed il successore
- la comunicazione tra i nodi è cifrata

Idea di Base (2)

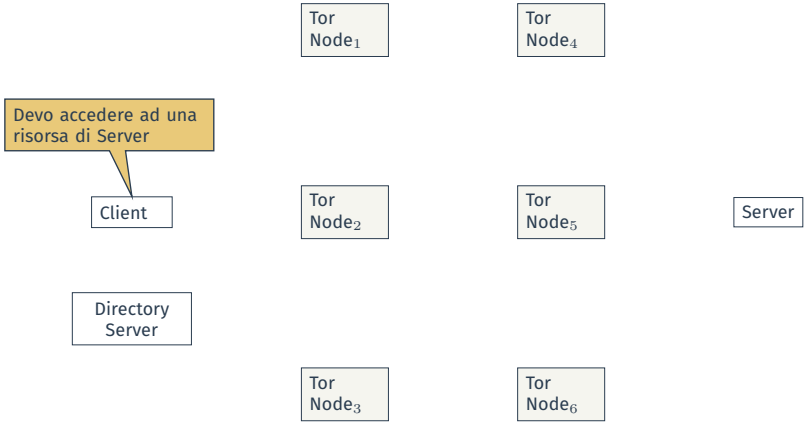
L'utente prepara il messaggio da inviare al server finale cifrandolo su più livelli, usando chiavi crittografiche diverse, stabilite con ciascuno dei nodi del circuito

- durante il circuito, ciascun nodo rimuove un livello di cifratura, scoprendo quale è il prossimo nodo sul circuito
- l'ultimo nodo (exit node) vede il messaggio originale, ma non sa chi sia il mittente

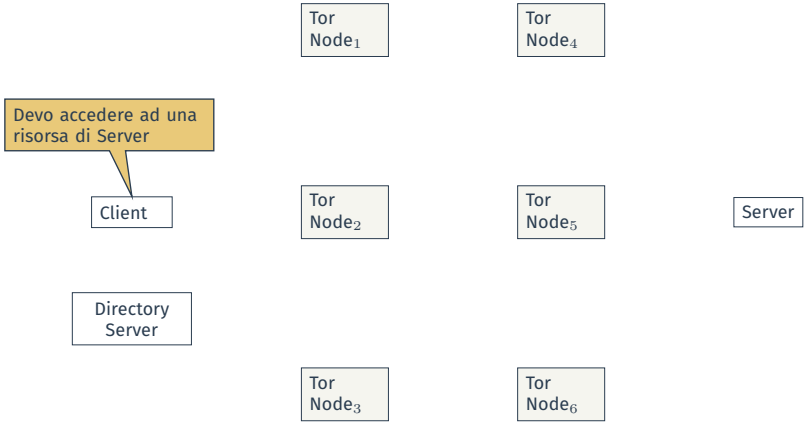
Funzionamento



Funzionamento

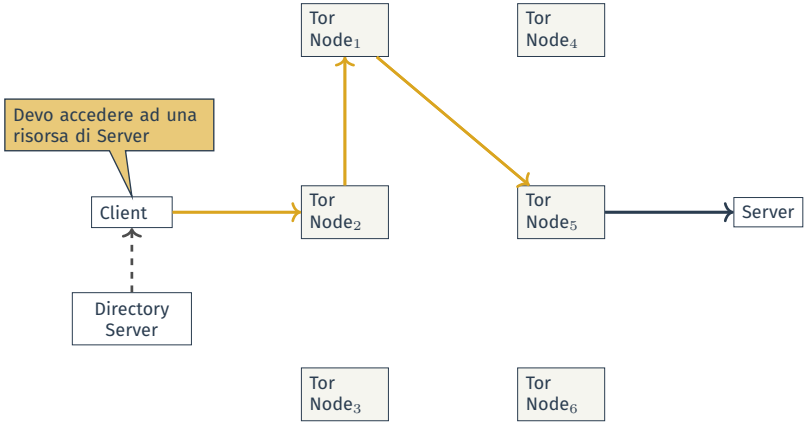


Funzionamento



1. Il client costruisce un circuito con l'aiuto del Directory Server: (Node$_2$, Node$_1$, Node$_5$)

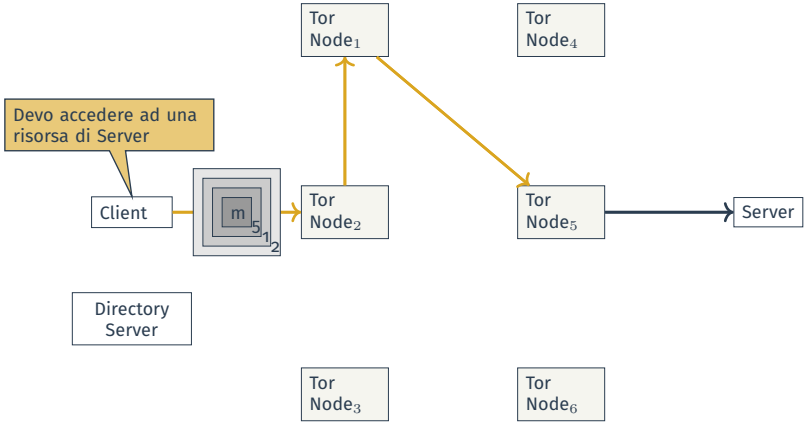
Funzionamento



Circuito: (Node$_2$, Node$_1$, Node$_5$)

2. il circuito viene costruito

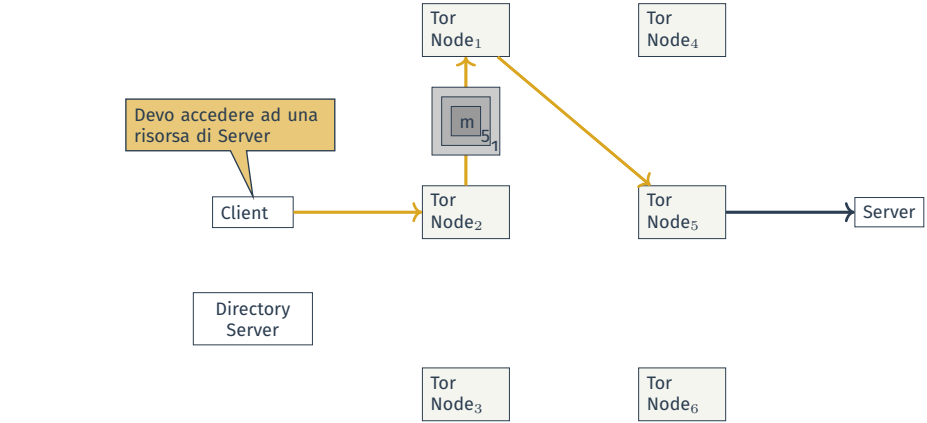
Funzionamento



Circuito: (Node$_2$, Node$_1$, Node$_5$)

3. creazione ed invio del messaggio cifrato: $\{\{\{m\}_5\}_1\}_2$

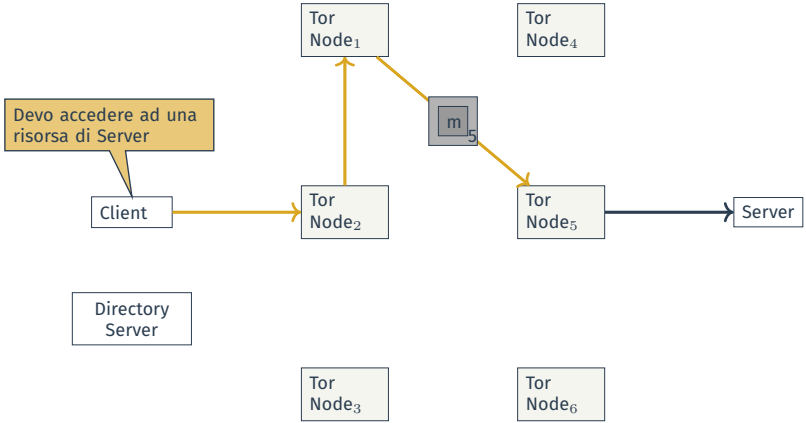
Funzionamento



Circuito: (Node$_2$, Node$_1$, Node$_5$)

4. ciascun nodo rimuove il proprio layer di cifratura

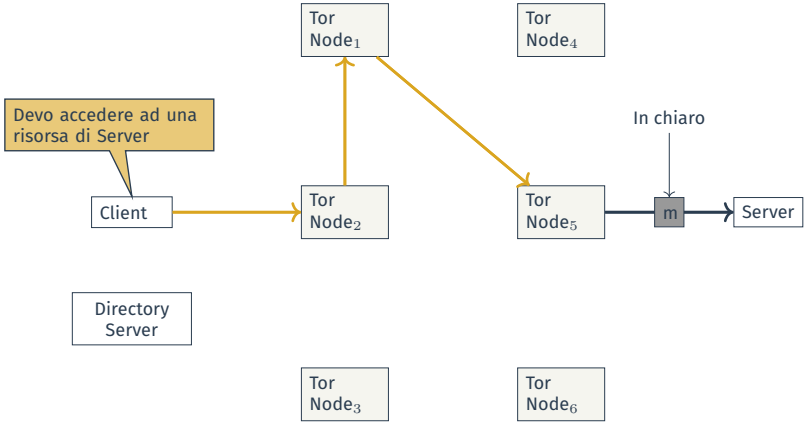
Funzionamento



Circuito: (Node$_2$, Node$_1$, Node$_5$)

5. ciascun nodo rimuove il proprio layer di cifratura

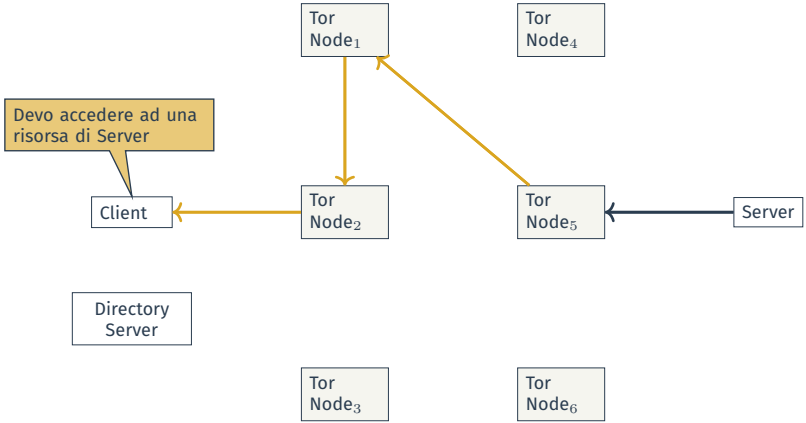
Funzionamento



Circuito: (Node$_2$, Node$_1$, Node$_5$)

6. ciascun nodo rimuove il proprio layer di cifratura
7. l'ultimo message dall'exit node al server di destinazione non è cifrato

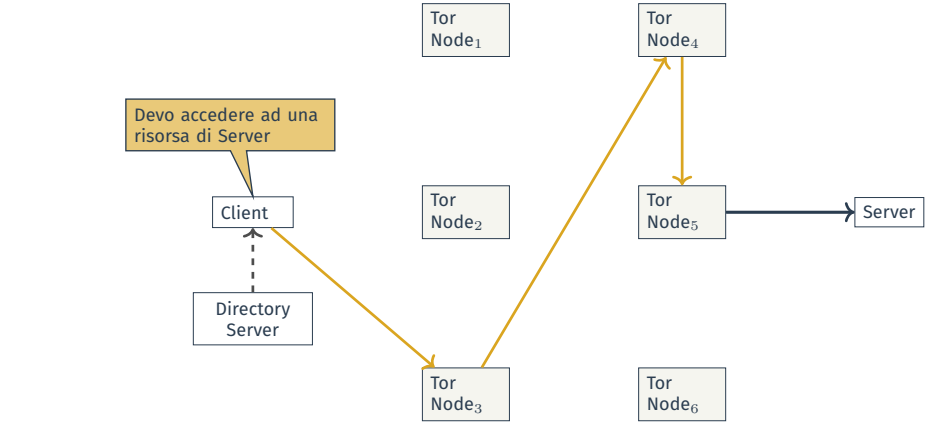
Funzionamento



Circuito: (Node$_2$, Node$_1$, Node$_5$)

8. le risposte seguono il percorso inverso

Funzionamento



Il circuito viene cambiato di frequente

9. nuovo circuito: (Node$_3$, Node$_4$, Node$_5$)

Dietro le Quinte

Tor utilizza:

- nodi: server proxy che utilizzano il protocollo `SOCKS`
  - proxy generico che supporta diverse applicazioni
- crittografia forte che garantisce, tra le altre cose, Perfect Forward Secrecy
  - se anche la chiave pubblica di un nodo venisse compromessa, il traffico già transitato non può essere decifrato

Onion Service

Gli onion service consentono di esporre dei servizi sulla rete Tor garantendo anonimato del destinatario

- l'onion service sceglie una serie di nodi tramite i quali il suo servizio sarà raggiungibile (*introduction point*)
- l'onion service viene pubblicizzato alla rete
  - mediante nomi DNS che finiscono in `.onion`
- chi vuole visitare l'onion service sceglie un altro nodo (*rendez-vous point*)
- chi vuole visitare l'onion service comunica ad uno degli *introduction point* del service il *rendez-vous point* scelto
- al *rendez-vous point* client e server si scambiano del materale crittografico e costruiscono un circuito

Onion Service: Garanzie

- Anonimato destinatario
- L'intera comunicazione non esce dal circuito Tor

Tor e Anonimato

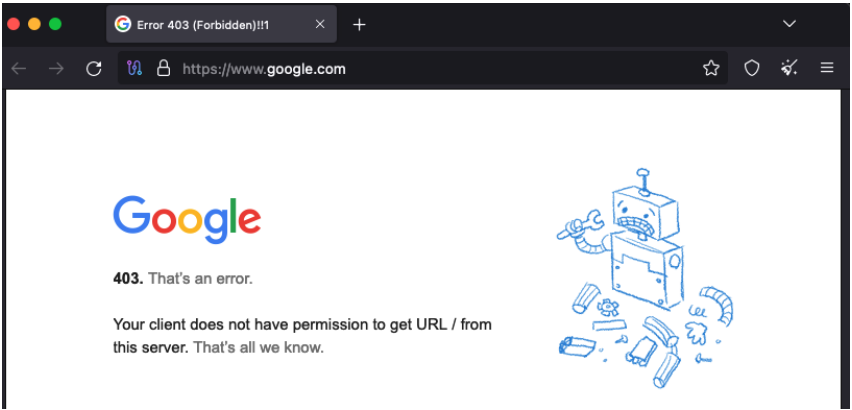
- Anonimato mittente
- Non tracciabilità *completa* della comunicazione se si usano gli hidden services
- Exit node non è *pericoloso* come proxy o VPN server, perché il circuito cambia di frequente

Problemi (1)

Il principale problema di sicurezza di Tor è dato dai poisoned node: nodi Tor gestiti da utenti malevoli

- l'NSA stessa non è riuscita a violare Tor (?)
- tutte le operazioni che hanno avuto successo verso la rete Tor si basano su
  - poisoned node
  - bug nei browser

Problemi (2)


Error 403 (Forbidden)!!1
https://www.google.com
Google
403. That's an error.
Your client does not have permission to get URL / from this server. That's all we know.

Deep & Dark Web

- Deep web: porzione del web non indicizzata dai motori di ricerca
- Dark net: porzione del deep web che può essere acceduta con specifico software (e.g., Tor)

	% **su totale (5205)**	% **su attivi (2723)**
Armi	0.8%	1%
Droghe	8%	15%
Estremismo	2.7%	5%
Hacking	1.8%	3%
Pornografia illegale	2.3%	4%
Illeciti	30%	56%

Fonte: D. Moore, T. Rid, “Cryptopolitik and the Darknet”, in Survival, 58:1 (2016)

Soluzioni:
Tor – Tor in Pratica

Utilizzo di Tor

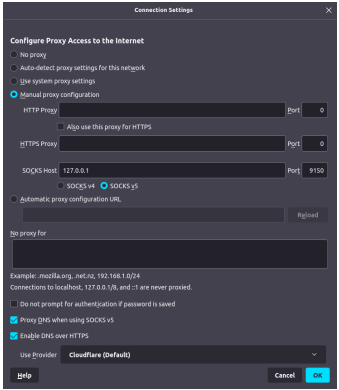
Tor è (tipicamente) un programma che viene eseguito in background sul proprio computer, funziona come un proxy locale a cui le altre applicazioni possono connettersi

- il proxy locale è in ascolto alla URL: `socks5://127.0.0.1:9150`
- le applicazioni devono supportare connessioni mediante proxy `SOCKSv5`

Utilizzo di Tor in Firefox

Una volta fatto partire il demone Tor, configuriamo Firefox per usarlo

Settings $\Longrightarrow$ *General* $\Longrightarrow$ *Network Settings*



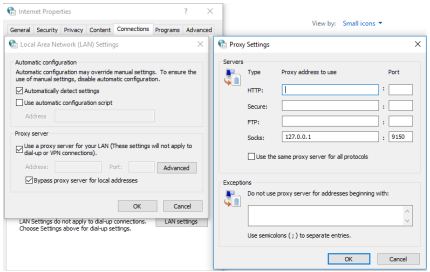
My IP Address :	
IP address	185.220.101.170
Hostname	tor-exit-170.relayon.org
IP Address Location :	
Country	United Kingdom (GB)
State/Region	England
City	London (South Bank)
ISP	CIA TRIAD SECURITY LLC
Organization	Kkvms LLP
Network	AS208294 CIA TRIAD SECURITY LLC (PERSO, TOR)
Connection Type	Corporate
Timezone	Greenwich Mean Time (GMT)
Local Time	Thu, 25 Nov 2021 19:23:17 +0000
Coordinates	51.5034,-0.1110
IPv6 Leak Test :	
IPv6 Address	n/a
WebRTC Leak Test :	
Local IP address	n/a
Public IP address	n/a

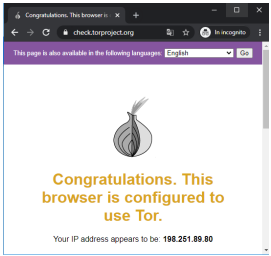
Utilizzo di Tor in Chrome e Windows 10

Chrome non consente di specificare le configurazione del proxy, occorre farlo a livello di sistema operativo

$\Longrightarrow$ la seguente configurazione fa sì che tutte (o quasi) le applicazioni del sistema usino Tor

Control panel $\Longrightarrow$ *Internet options* $\Longrightarrow$ *Connections* $\Longrightarrow$ *LAN settings* $\Longrightarrow$ *Proxy server* $\Longrightarrow$ *Advanced*

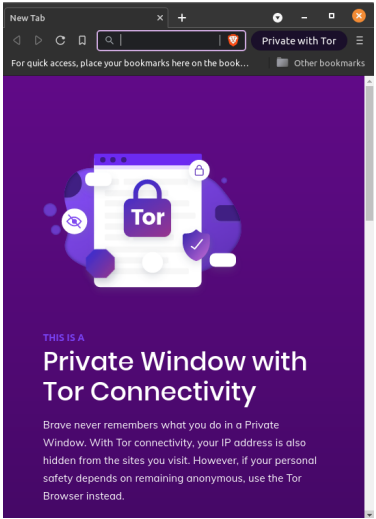




Brave

Brave è un browser basato su Chrome

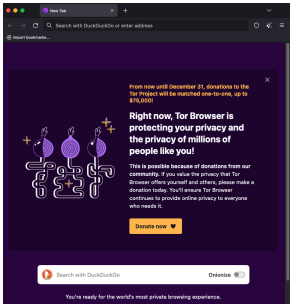
- privacy-oriented
- include Tor



Tor Browser (1)

Tor Browser è una versione modificata di Firefox che

- si connette di default al proxy Tor
- funziona solo in modalità anonima
- contiene una serie di plugin finalizzati a migliorare la propria privacy



Soluzioni:
Altri Aspetti

Proxy vs VPN vs Tor

	Proxy	VPN	Tor
Scopo	Sicurezza perimetrale	Unire reti non contigue	Anonimato
Livello	5	2, 3	5
Generico	✗*	✓	✓*
Cifratura comunicazioni	✗	✓	✓
Uso visibile	✓*	✓*	✓*

Tor fornisce un maggiore anonimato rispetto alle VPN

- Tor: nessun nodo del circuito conosce il vero destinatario ed il vero mittente
- VPN e proxy: VPN e proxy server conoscono il vero destinatario ed il vero mittente

Uso Combinato

È possibile combinare, proxy, VPN, e Tor: Tor come VPN

- connettersi ad una VPN e poi usare Tor
- usare la rete Tor direttamente come VPN

Non così semplice

Ne vale la pena?

Visibilità dei Protocolli

Proxy, VPN, Tor nascondono il mittente *originale*

- l'uso di questi strumenti può essere scoperto
  - tramite il nuovo indirizzo IP mittente
- il traffico Tor è identificabile perché i pacchetti hanno una forma nota
  - camuffare il traffico Tor facendolo apparire per un altro protocollo innocuo, e.g., `https://sri-csl.github.io/stegotorus`

DNS over HTTPS

Idea: proteggere il traffico DNS in modo che nessuno possa vedere *tutte* le richieste DNS fatte $\Longrightarrow$ DNS over HTTPS

- esecuzione del protocollo DNS usando HTTPS come *trasporto*
  - traffico cifrato
  - server DNS *fidati* garantiscono la non memorizzazione delle richieste
    - garanzia *legale*
  - creare ed usare il proprio server DNS over HTTPS

Nota: DNS ha (molti) altri problemi di sicurezza

Soluzioni:
Livello Applicativo

Sicurezza a Livello Applicativo

- cookie e fingerprinting sempre possibili perché attaccano il browser
- $\Longrightarrow$ usare diverse strategie con attenzione
- software possono sempre contenere bug

Brave browser leaks onion addresses in DNS traffic

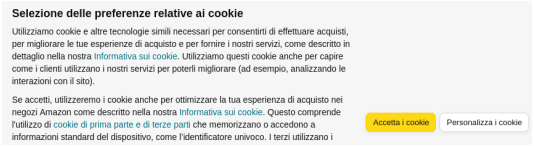
DNS leak leaves footprints in DNS server logs for a Brave user's Tor traffic.

Cookie: Consenso in Pratica (1)

Siti chiedono il consenso per memorizzare cookie...

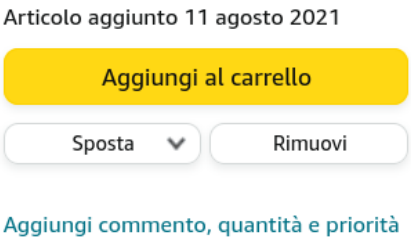
Cookie: Consenso in Pratica (2)

Esempio: `https://www.amazon.it`, notate i colori...



Cookie: Consenso in Pratica (2)

Esempio: `https://www.amazon.it`, notate i colori...



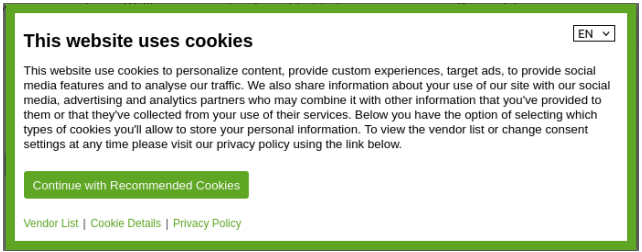
Cookie: Consenso in Pratica (3)

Esempio: `https://datascienceparichay.com` (2020)

Cookie: Consenso in Pratica (3)

Esempio: `https://datascienceparichay.com` (2020)

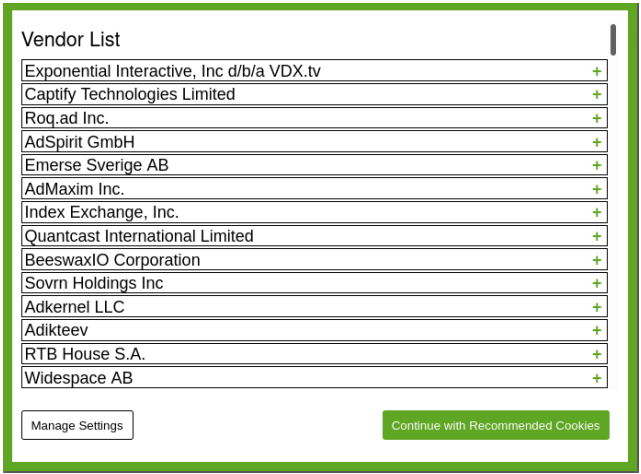
1. Appare cookie consent



Cookie: Consenso in Pratica (3)

Esempio: `https://datascienceparichay.com` (2020)

1. Appare cookie consent
2. *Vendor list*



Cookie: Consenso in Pratica (3)

Esempio: `https://datascienceparichay.com` (2020)

1. Appare cookie consent
2. *Vendor list*
3. *Manage settings*

Cookie Details

Cookies are small text files that can be used by websites to make a user's experience more efficient. The law states that we can store cookies that contain personal information on your device if they are strictly necessary for the operation of this site. For all other types of cookies that contain personal information we need your permission. This site uses different types of cookies. Some cookies are placed by third party services that appear on our pages.

Necessary	Consent ☑ +
Preferences	Consent ☑ +
Statistics	Consent ☑ +
Marketing	Consent ☐ +

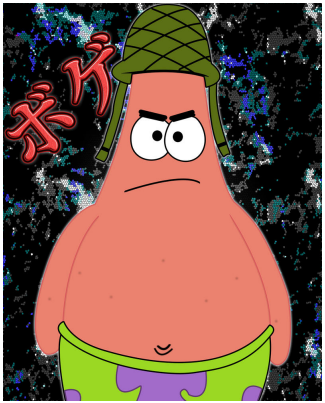
Save Settings & Exit

Continue with Recommended Cookies

Browser

Varie opzioni ed estensioni per mantenere controllo sui propri dati

- modificare user agent
- non consentire la geolocalizzazione
- usare la navigazione in modalità anonima
- fare attenzione al motore di ricerca



Il tracking si sta spostando sempre più verso il fingerprinting

Soluzioni:
Bypassare del Filtraggio e Nascondersi

Bypassare i Filtri

Alcune delle tecniche che si possono usare per bypassare un filtraggio, anche in combinazione, sono:

- indirettezza
- cifratura
- offuscamento

Indirettezza

L'indirettezza prevede di usare un intermediario per raggiungere la destinazione finale, assumendo che l'intermediario sia raggiungibile

Idea: se `facebook.it` è bloccato ma `free-tunnel.it` no, allora si può passare da `free-tunnel.it`, chiedendogli di raggiungere `facebook.it`

In concreto, usando:

- proxy
- VPN
- Tor

- l'uso di una politica whitelist può rendere l'indirettezza molto difficile

Cifratura (1)

La cifratura di un pacchetto lo rende incomprensibile a chiunque lo intercetta ed analizza

Idea: applicare una tecnica di cifratura per nascondere al filtro che cosa si sta facendo

- anche il destinatario deve supportare la tecnica di cifratura in uso
- possible usare un intermediario
- il filtro potrebbe accorgersi dell'uso della cifratura, pur non potendo violarla
- il filtro potrebbe bloccare un tipo di traffico che non riconosce

Cifratura (2)

Non è possibile leggere il contenuto di un pacchetto cifrato

- analisi comunque possibile
  - la cifratura è (quasi) sempre a livello di payload, gli header sono in chiaro, almeno quelli dei livelli inferiori della pila protocollare
  - possibile indurre in errore
  - e così via...

Offuscamento (1)

L'offuscamento prevede di alterare il traffico in modo di farlo sembrare innocuo

Idea: se il traffico `torrent` è bloccato

Offuscamento (1)

L'offuscamento prevede di alterare il traffico in modo di farlo sembrare innocuo

Idea: se il traffico `torrent` è bloccatoma quello `HTTP` no, allora si può offuscare il traffico `torrent` per farlo sembrare `HTTP`

- un filtro molto sofisticato potrebbe scoprirlo comunque

Offuscamento (2)

Dobbiamo comunque fare attenzione al fingerprinting!

- l'utilizzo del protocollo puo' rivelare informazioni sul dispositivo/utente
- TCP fingerprinting può rivelare il sistema operativo in uso
- TLS fingerprinting può rivelare librerie usate e loro configurazione
- ...

Nota: può essere usato anche per scopi di sicurezza (ad es., malware ha una TLS fingerprint spesso anomalo)

Conclusioni

Problemi

Il problema dell'anonimato è complesso perché vi sono due attori con interessi distinti e contrapposti (e legittimi)

- utenti, che utilizzano la maggior parte dei siti ed app gratuitamente
- siti ed app, che devono essere profittevoli

Quantomeno, occorre consapevolezza

Soluzioni

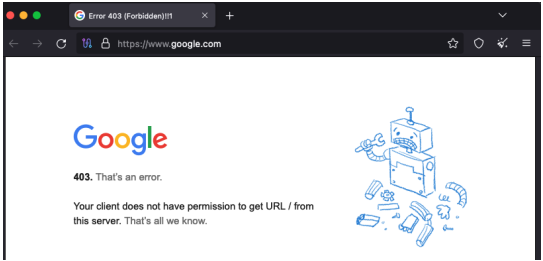
Soluzioni tecnologiche prevedono l'uso combinato di diversi strumenti a diversi livelli

- livelli inferiori, e.g., proxy, VPN, Tor, DNS over HTTPS
- livelli superiori, e.g., Firefox, Brave, modalità anonima

Soluzioni

Soluzioni tecnologiche prevedono l'uso combinato di diversi strumenti a diversi livelli

Spesso, occorre rinunciare a qualche funzionalità



Software Utili (1)

- Browser
  - motore di ricerca *DuckDuckGo*: `https://duckduckgo.com`
  - cambiare user agent in Firefox: `https://support.mozilla.org/en-US/kb/how-reset-default-user-agent-firefox`
  - *Facebook container* per bloccare il tracking di Facebook in Firefox: `https://www.mozilla.org/en-US/firefox/facebookcontainer/`
  - ...
- Rete
  - Cloudflare DNS over HTTPS service: `https://www.cloudflare.com/learning/dns/what-is-1.1.1.1/`
  - *AdGuard* tracking blocker: `https://adguard.com/adguard-home.html`
  - *Blocky* tracking blocker: `https://0xerr0r.github.io/blocky/`
  - *Stegotorus* Tor camouflage: `https://sri-csl.github.io/stegotorus/`
  - *Eotk* Tor hidden service: `https://github.com/alecmuffett/eotk`
  - ...

Software Utili (2)

Cloak	Offuscamento, indirettezza	`https://github.com/cbeuw/Cloak`
HASSH	SSH fingerprinting	`https://github.com/salesforce/hassh`
JA3	TLS fingerprinting	`https://github.com/salesforce/ja3`
OpenVPN TLS-Crypt v1/2	Cifratura VPN	`https://github.com/OpenVPN/openvpn/blob/master/doc/tls-crypt-v2.txt`
Stegotorus	Offuscamento Tor	`https://github.com/SRI-CSL/stegotorus`

Altri Riferimenti (Alcuni)

- M. Anisetti, C. A. Ardagna, N. Bena, E. Damiani, “Stay Thrifty, Stay Secure: A VPN-based Assurance Framework for Hybrid Systems”, in Proc. of SECRYPT 2020, Parigi, Francia, Luglio 2020
- EFF, “Cover Your Tracks”, `https://coveryourtracks.eff.org/`
- FingerprintJS, “Canvas Fingerprinting”, `https://fingerprintjs.com/blog/canvas-fingerprinting/`
- P. Samarati, “Protecting respondents identities in microdata release”, in IEEE TKDE, vol. 13, no. 6
- N. M. Al-Fannah, W. Li, C.J. Mitchell, “Beyond Cookie Monster Amnesia: Real World Persistent Online Tracking”, in Proc. of ISC 2018, Guildford, UK, Settembre 2018
- S. Leroux, S. Bohez, P. Maenhaut, N. Meheus, P. Simoens and B. Dhoedt, “Fingerprinting encrypted network traffic types using machine learning”, in Proc. of IEEE/IFIP NOMS 2018, Taipei, Taiwan, Aprile 2018
- Z. Bu, B. Zhou, P. Cheng, K. Zhang and Z.-H. Ling, “Encrypted Network Traffic Classification Using Deep and Parallel Network-in-Network Models”, in IEEE Access, vol. 8