Roma, 23 maggio 2017

# OSINT
# SU
# Siti Web


Paolo Dal Checco, *Consulente Informatico Forense*

# Chi sono

- PhD @UniTO nel gruppo di Sicurezza delle Reti e degli Elaboratori
- Professore a Contratto di Sicurezza Informatica @UniTO (SUISS)
- Consulente Informatico Forense (Perizie Informatiche) per Privati, Aziende, Avvocati, Procure, Tribunali, F.F.O.O.
- Tra i fondatori dell'Associazione DEFTA (www.deftlinux.net) e ONIF (www.onif.it)
- Socio IISFA, Tech & Law, Clusit, AIP, AssobIT
- paolo@dalchecco.it - @forensico
- www.dalchecco.it, difob.it, bitcoinforensics.it, ransomware.it

# OSINT

- Open Source INTelligence
- Un sistema "puro" che non ha bisogno di compromessi oscuri con le fonti, non viola la legge con attività investigative illegali, ma si basa solo sulla capacità tecnica e operativa di trovare le informazioni, la mentalità investigativa, la conoscenza delle tecniche di analisi e correlazione dei dati e infine il lavoro metodico e organizzato di consultazione delle fonti aperte che sono per definizione accessibili a tutti"
  [Leonida Reitano - "Esplorare Internet"]

# OSINT su un sito web

- Capire chi può aver registrato il dominio
- Scoprire chi sta gestendo il sito
- Rilevare dove si trova l'hosting
- Identificare eventuali siti/domini legati a quello in analisi
- Trovare indirizzi email o numeri di telefono
- Scoprire il passato di un sito web
- Analizzare i profili dei visitatori o degli iscritti

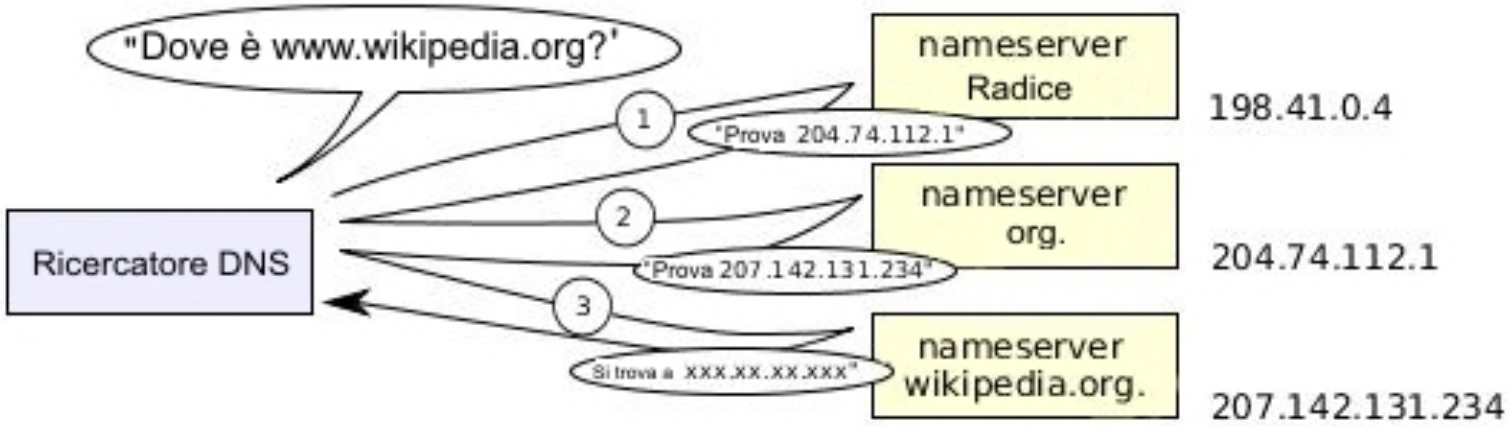
# Better safe than sorry

- Può essere conveniente rimanere anonimi
- TOR, Torbrowser (entrambi possono anonimizzare anche applicazioni)
- TAILS, JonDonym, VPS private
- Non aprire documenti se non in TAILS.
- Per ricevere sms: servizi di receive-sms-online (free/pay)
- Per ricevere email: email temporanee (mailinator.com, yopmail.com, no 10minutemail) oppure webmail dietro TOR (es. vfemail.net, safe-mail.net, mail15.com, inbox.lv, OpenMailBox.org)

# Chi dicono di essere?

- Whois
- Whois storico (DomainTools, who.is)
- Reverse Whois (YouGetSignal)
- Altri domini con estensione diversa (domize)
- Relazioni tra domini e owner (DomainTools) oppure
- google "site:whois.domaintools.com "dal checco"" o "registrant "dal checco""
- whoisology.com, whoismind, comnetcomber.com
- Maltego e transform

# Domini, DNS e server MX



# Domini, DNS e server MX

- Verificare se sul dominio ci sono domini di terzo livello (es ftp, webmail, etc...) usando tool come Knock.py del buon Gianni Amato, SubRoute o DNSenum
- Verifico server MX (ricordare che non necessariamente è lo stesso del dominio)
- Partco on analisi dei server MX
  - Whois
  - Reverse IP
  - Sito sulla porrta 80 o 443?
  - Eventuali porte aperte e servizi (Shodan o PortScan)

# Domaintools

- Uno dei pochi servizi spesso indispensabili per Open Source Intelligence su siti web e, purtroppo, solo a pagamento
- Diversi crawler attivi da anni che raccolgono dati da siti, whois, dns, hosting, MX, etc…
- Archivio storico più preciso e datato (1997)
- Who.is contiene una piccola parte, free

# Domaintools

Whois Lookup
Screenshot History
Domain Marketplace
Bulk Check
Domain Suggestions
Domain Typo Finder
Domain Sales History
Dropping Names
DNS Lookup
Traceroute
Ping
My IP Address
IP Explorer

## Historical Lookups

Whois History
Hosting History

## Research Tools

Domain Search
Reverse IP Lookup
Reverse NS Lookup
Reverse MX
Reverse Whois Interactive Mode

## Monitoring Tools

Domain Monitor
Brand Monitor
Name Server Monitor
Registrant Monitor
IP Monitor

## Reports

Reverse Whois Report
Domain Report

## APIs

API Access

# Indirizzo IP e webserver

- Verifico server web di default sulla porta 80
- Verifico se attivo server sulla 443 (ssl) e scarico certificato
- Trovo IP storici e verifico se sono ancora attivi weberver sulla 80 e sulla 443
- Se ci sono, visualizzo e/o scarico sito vecchio
- Verificare su spamhaus se l'IP è stato coinvolto in attività di spam/frode
- Verificare se l'IP ha un reverse dns (dig -x xxx.xxx.xxx.xxx)

# Indirizzo IP e webserver

- Leggo negli header HTTP il tipo di server (web-sniffer.org)
- Provo a caricare una pagina volutamente errata, spesso nei messaggi di errore si trovano info sul path locale del server (con username…)
- Shodan
- Attenzione: User Agent come discriminante di cosa viene indicizzato
  - Spesso i siti mostrano a Google cose che non mostrano ai visitatori
  - Utilizzare plugin tipo User Agent Switcher (Chrome/Firefox)

# Google

| Search Service | Search Operators |
|---|---|
| Web Search | allinanchor:, allintext:, allintitle:, allinurl:, cache:, define:, filetype:, id:, inanchor:, info:, intext:, intitle:, inurl:, link:, related:, site: |
| Image Search | allintitle:, allinurl:, filetype:, inurl:, intitle:, site: |
| Groups | allintext:, allintitle:, author:, group:, insubject:, intext:, intitle: |
| Directory | allintext:, allintitle:, allinurl:, ext:, filetype:, intext:, intitle:, inurl: |
| News | allintext:, allintitle:, allinurl:, intext:, intitle:, inurl:, location:, source: |
| Product Search | allintext:, allintitle: |

# Shodan


SHODAN
Search
EXPOSE ONLINE DEVICES.
WEBCAMS. ROUTERS.
POWER PLANTS. IPHONES. WIND TURBINES.
REFRIGERATORS. VOIP PHONES.
TAKE A TOUR
FREE SIGN UP
Popular Search Queries: default password - Finds results with "default password" in the banner; the named defaults might work!

# Contenuti

- Cerco testo sul sito su Google, tra virgolette, per vedere da dove è copiato o dove è riprodotto (spesso vengono riciclate frasi)
  - Oppure uso siti come copyscape.com o siteliner.com
- Se vengono citati nomi di aziende o marchi, posso cercare su marchi/brevetti.
- Se si trova P.IVA verificare su agenzia delle entrate (potrebbero averla copiata)
  - http://ec.europa.eu/taxation_customs/vies/?locale=it
  - https://www1.agenziaentrate.gov.it/servizi/vies/vies.htm?p=&s=IT

# Contenuti nascosti

- Scaricare intero sito con wget (o torify wget)
- wget --no-check-certificate -e robots=off -o log.txt -w 7 --random-wait -vv -S -r -N -l inf --no-remove-listing --preserve-permissions -np -E -k -K -p --user-agent="Mozilla/5.0 (compatible; bingbot/2.0; +http://www.bing.com/bingbot.htm)" http://www.website.com/subdir
- Scaricare risultati di ricerca google su “site:” (con plugin o con scraper) e scaricarli con wget –i list.txt
- Esamino robots.txt, spesso contiene cose interessanti…

# Contenuti nascosti

- Verificare se è abilitato Index sui folder o…
- Fare brute force di directory:
  - OWASP DirBuster Project
  - https://github.com/maurosoria/dirsearch
  - Pattern noti (es. kit di Phishing, etc…)

# Contenuti rimossi o modificati

- Web Archive (dati rimossi o modificati)
  - Problema: robots.txt o filtri su IP
- Memento: http://timetravel.mementoweb.org/
  - http://web.archive.bibalex.org/web
  - web.archive.org
  - archive.is
- Per tracciare le modifiche:
  - RSS (Feed.ly, Feed2Mail)
  - Followthatpage (anche su RSS), ChangeDetection, Versionista, VisualPing
- Google/Bing Cache
- Snapshots/DomainTools

# Carving

- “Carving” su un sito web? :-)
- Lo scopo è cercare ciò che è rimasto “nascosto” per errori, dimenticanze o volutamente
- Scarico tutto il sito tramite WGET
- Cerco email, url non linkate o commenti (grep "<---”) che contengano informazioni rilevanti
- Idea : Bulk Extractor su copia wget per estrarre email, domini, url, carte di credito, numeri di telefono, indirizzi IP, etc…

# Link

- Verifico link in entrata e uscita (www.opensiteexplorer.org) e cerco eventuali relazioni
  - Spesso chi i proprietari di un sito lo likano da altri o inseriscono nel sito link significativi
- Xenu Link Sleuth (anche anchor, date, title, etc.. Comprese le immagini)
- Seo Powersuite Link Analysis

# Backink

| | Rank<br>Domain Rank | Referring Page URL<br>Referring Page Title | Int<br>Ext | Link URL<br>Link Anchor |
|---|---|---|---|---|
| 1. | **11**<br>57 | **www.clusit.it**/soci_home.htm<br>Elenco Soci Clusit | 1<br>**255** | **www.difob.it/**<br>DAL CHECCO PAOLO(Titolare Di.Fo.B. Studio Associato , Grugliasco ◆ TO) |
| 2. | **11**<br>55 | **dag.it**/sicurezza.html<br>Sicurezza | 12<br>**2** | **www.difob.it/**<br>o www.difob.it |
| 3. | **11**<br>55 | **www.dag.it**/sicurezza.html<br>Sicurezza | 12<br>**2** | **www.difob.it/**<br>o www.difob.it |
| 4. | **11**<br>35 | **www.dalchecco.it/**<br>Chi sono - Paolo Dal Checco - Perizie Informatiche Forensi | 56<br>**11** | **www.difob.it/**<br>socio fondatore dello studio diConsulenza Informatica Forense“ Digital Forensics Bureau ” diTorinoe Socio Amministratore della “ |
| 5. | **11**<br>45 | **www.dezzani.biz/**<br>Giuseppe DEZZANI Consulente Informatico Forense » Giuseppe DEZZANI | 9<br>**4** | **www.difob.it**/laboratorio/<br>attrezzato con due laboratori dotati dei migliori sistemi di analisi disponibili sul mercato |

# Datazione

- Data restituita dall'header HTTP (di pagine statiche e risorse come immagini, file, etc…)
- RSS (contengono date di creazione/modifica)
- Eventuali date presenti nella pagina (commenti, data dei post, etc…)
- Web Archive
- Snapshots/DomainTools
- Date nell'header HTTP restituito dalle immagini
- Metadati nelle immagini

# Social network

- Esamino i social per verificare se chi gestisce il sito ha creato anche profili o gruppi su Facebook, twitter, etc…
  - Facebook advanced search
  - site:www.facebook.com www.difob.it -inurl:DiFoB
  - site:www.facebook.com difob

# Tracking codes

- Cerco tag di Google Analytics/AdSense e lo utilizzo per cercare altri siti monitorati dallo stesso utente (spyonweb.com, sameid.net, ewhois.com, reverseinternet.com)

```
joomla - http://www.joomlic.com
764         var _gaq = _gaq || [];
765         _gaq.push(['_setAccount', 'UA-23175468-2']);
766         _gaq.push(['_trackPageview']);
767  (function() {
```

- Cerco altri tag come histats & Co. anche se più difficili da correlare

# Metadati

- Foca (ora free: elevenpaths.com/labstools/foca)
- Metagoofil per scaricare pdf, doc, xls, ppt, etc... e anche MAC Address dalle pagine di un sito
- EXIF delle foto
- Se presenti sul sito, verifico contenuto delle **chiavi PGP** (gpg --with-fingerprint key.asc)
- Indirizzi Skype? Skype Resolver non funziona più, ma si possono contattare...

# Immagini

- Cerco le immagini presenti sul sito su Google Images (comodo con Chrome, tasto destro o 's'+right key) o TinEye
- Verifico il nome file delle immagini (nella URL) ed eventuali tag ALT/TITLE e verifico se usati altrove
- Con exiftool o simili verifico dati EXIF (es. autore, GPS location, seriale fotocamera) e cerco altrove
- Se le trovate su FB, dovreste poter risalire al profilo

# CMS (Wordpress, Joomla, etc..)

- Plugin installati (http://whatwpthemeisthat.com/)
- Template installato (wpthemedetector.com)
- /wp-content/themes/kallyas/screenshot.png
- /wp-content/themes/kallyas/style.css
- Esamino i dati del template (autore, nome, etc… che posso cercare su web)
- Verifico se è stato usato su altri siti o social (namechk.com, knowem.com)
- Verifico i vecchi template con Web Archive

# Bitcoin

- Se sono indicati bitcoin address si può tentare un minimo di bitcoin intelligence:
  - www.blockchain.info (tags, link analysis, etc…)
  - Blockseer.com (tags, cluster)
  - Walletexplorer.com (cluster)
- E' anche possibile usare espressione regolare da aggiungere a Bulk Extractor.

# Grazie

**Email/Twitter**

paolo@dalchecco.it / @forensico

**Web**

www.dalchecco.it / www.difob.it / www.bitcoinforensics.it