Sentiment Analysis e Social Media Intelligence

IL DATAGATE

Si parla, in questi giorni, del "*Datagate Facebook*", cioè del *mea culpa* di Mark Zuckenberg dopo lo scandalo della profilazione massiva sui *social* effettuata dalla *Cambridge Analytica* per orientare le passate elezioni statunitensi.

Il termine "*datagate*" era stato utilizzato nella semantica mediatica nel lontano 2013, a seguito della divulgazione di una serie di notizie *top secret* riguardanti l'Agenzia per la Sicurezza Nazionale statunitense (NSA) messe a disposizione della stampa da Edward Snowden, scoperchiando anche il progetto "*Echelon*", il più imponente e sconosciuto strumento di *Signal intelligence* in ambito strategico, che sarebbe in grado di monitorare ed intercettare il traffico dei satelliti commerciali in orbita; un progetto ambizioso quanto invasivo messo in atto dalla *partnership* siglata con l'accordo *UKUSA* (tra Australia, Canada, Nuova Zelanda, Regno Unito e Stati Uniti).

Una serie di esplosioni mediatiche a catena seguite dagli *scoop* di WikiLeaks, affiancate allo sviluppo sfrenato di tecnologie rivolto al monitoraggio massivo di comunicazioni per macroaree geografiche con l'utilizzo di tecnologie di intelligenza semantica digitale.

Ed è così che, in un contesto meno complesso di *Communication Intelligence*, l'attività di ispezione di più aree geografiche avviene, oggi, attraverso l'applicazione di "sonde" (particolari filtri parametrici) sulle *backbone* di comunicazione, cioè quelle linee interconnesse con linee più piccole ed utilizzate per trasferire i dati a grandi distanze

in maniera efficiente.

Attraverso i filtraggi parametrici è possibile identificare sessioni di traffico generate da un punto imprecisato all'interno di una macro area, cercando di individuare alcune parole di interesse "critico", cc.dd. *keywords* di ricerca.

Il secondo *step* operativo, definito il criterio di filtraggio dei dati fonici o dei pacchetti, consisterà nell'ispezione del cablaggio tramite le sonde di intercettazione in ingresso al *Front End* o sul *Back End*.

In questo scenario si commistionano *best practices* di *Human Intelligence* complementari o, alle volte, in conflitto con quelle più sofisticate di *Communication* e *Signal Intelligence* ed, ancora, di *Open Source Intelligence* e di *Imagery Intelligence*.

Si tratta di alcuni comparti di quel "*processo di intelligence*" di cui si parla sempre più spesso: una sorta di evolute discipline standardizzate dalla comunità scientifica internazionale secondo protocolli applicativi che trattano i temi dell'*intelligence* nelle sue tante sfaccettature.

IL PROCESSO DI INTELLIGENCE

Proviamo allora, in modo sintetico, a spiegare cosa si cela dietro i tanti acronimi che costituiscono questo "*processo*", sempre più condiviso per esigenze diverse al fine di trattare questioni che lambiscono il lessema *intelligence*, oggi non più di esclusivo appannaggio dei cc.dd. "*servizi segreti*":

- **HUMINT** (*HUMan INTelligence*) concerne l'acquisizione dei dati strategici ed è svolta da risorse umane che hanno il compito di raccogliere notizie attraverso relazioni interpersonali (agenti e informatori) o l'osservazione diretta (osservatori);
- **IMINT** (*IMagery INTelligence*) è la disciplina che cura la raccolta e l'analisi di immagini aeree o satellitari;

- **MASINT** (*MeAsurement and Signature INTelligence*) attiene l'acquisizione di immagini non visibili con sensori elettrici o radar;
- **COMINT** (*COMmunication INTelligence*) consiste nell'intercettazione, selezione e interpretazione dei contenuti inerenti al traffico delle telecomunicazioni (perlopiù flussi di traffico che transitano per i satelliti e i cavi internazionali);
- **ELINT** (*ELectronic INTelligence*) studia la ricezione e l'analisi di segnali elettronici, come ad esempio l'emissione dei sistemi radar;
- **SIGINT** (*SIGnal INTelligence*) si interessa della raccolta di dati mediante l'intercettazione di mezzi di comunicazione (ad esempio radio, mail, telefono, ecc);
- **TECHINT (S*cientific and TECHnical INTelligence)*** riguarda l'attività d'intelligence nel settore delle armi ed equipaggiamenti, nonché di acquisizione informazioni a livello strategico;
- **OSINT** (*Open Source INTelligence*) riguarda l'acquisizione e l'analisi dei contenuti messi a disposizione dalle fonti aperte (stampa, *internet*, *social networks*, *database* pubblici, ecc.).

INTELLIGENT DATA MINING

In riferimento al monitoraggio di macro aree di comunicazione per esigenze di *Intelligence,* è spesso necessario procedere parallelamente all'ispezione di risorse "aperte" sul *web*, con un approccio OSINT.

L'analisi delle fonti aperte "*comprende diversi ambiti disciplinari combinati tra loro: gli strumenti di hacking della rete per ottenere informazioni sulle identità digitali, l'uso avanzato dei motori di ricerca, l'utilizzo dei portali di investigazioni digitali (dove ottenere le informazioni istituzionali su persone fisiche o giuridiche, proprietà immobiliari, partecipazioni azionarie o societarie, etc.) e*

infine le tecniche di analisi investigative per valutare il materiale informativo acquisito ed elaborato attraverso strumenti di visualizzazione grafica dei dati", annota l'amico Leonida Reitano, esperto in investigazioni *open source* e formatore.

Si tratta di ricerche i cui prodromi trovano un *imprinting* giornalistico investigativo evolutosi nel tempo, che beneficiano di strumenti di analisi e di *software* in grado di trasformare informazioni, apparentemente slegate tra loro, in un quadro d'insieme di elevata significatività investigativa.

Qui, un bravo analista OSINT procederà, partendo da uno scenario apparentemente senza soluzione di continuità, ad interrelare informazioni svariate andando ad insinuarsi nei contenitori del *Deep* e *Dark Web*, cioè quella parte della ragnatela non indicizzata dai motori di ricerca, raggiungibile attraverso alcuni *software* per la navigazione anonima, come la nota "*cipolla*" del *router* T.O.R. o attraverso il protocollo I2P.

Così come le tante informazioni nascoste dentro una stringa di caratteri o tra i *bit* meno significativi di un'immagine (*least significant bit*), attraverso sistemi più o meno evoluti di stegano/criptografia.

Nell'ispezione delle *Open Sources*, l'analisi di *Intelligence* trova compendio con l'impiego di sofisticatissimi *software* di analisi semantica che, con il supporto di tecnologie TAL (trattamento automatico della lingua), si interessano dello *Speech Processing (SP)* o elaborazione del parlato, e del *Natural Language Processing (NLP)* o elaborazione del testo.

Si pensi, in un contesto di metodologia di ricerca per ragioni commerciali, la profilazione effettuata attraverso i c.d. *cookies* che, nel quotidiano, ci mette in allarme ogni qual volta, effettuando una ricerca merceologica, ci ritroviamo, di lì a poco, la tendina di un *banner* pubblicitario che ci

rimanda, innumerevoli volte, a quell'oggetto digitato in precedenza.

Un'invadente metodologia psicologico-subliminale in grado di condizionare e rendere ridondante un nostro pensiero.

Più articolato è, poi, l'impiego di protocolli di analisi semantica approfondita, definiti *Intelligence Data Mining*, finalizzati all'individuazione di informazioni nascoste.

Rispetto alle tecnologie tradizionali (a *keyword* e statistiche), che possono solo cercare di indovinare il senso di un testo, i *software* di analisi semantica approfondita leggono e interpretano tutta la conoscenza potenzialmente interessante e identificano in automatico le relazioni concettuali fra le varie informazioni, anche quelle ovattate da sistemi di crittografia.

I NATIVI DIGITALI E LE EMOTIONAL INFORMATIONS

Con la nascita dei nativi digitali e il *boom* del *social media*, analisti strategici ed esperti d'*intelligence* hanno assunto consapevolezza di quanto sia importante l'acquisizione di informazioni utili a monitorare i gusti degli utenti. È un *business* commerciale che ha sia ragioni socio o geopolitiche sia ragioni più complesse di strategia militare e di *intelligence*.

Un'esigenza che, via via, si è insinuata in modo violento nella *privacy* del singolo che, in uno scenario geopolitico complesso si trova a soccombere rispetto alle esigenze sovranazionali o, altrettante volte, rispetto *alle policies* aziendali delle grandi *lobbies* multinazionali nel settore.

Già Fëdor Michajlovič Dostoevskij sosteneva che "*Ogni uomo ha dei ricordi che racconterebbe solo agli amici. Ha anche cose nella mente che non rivelerebbe neanche agli amici, ma solo a*

se stesso, e in segreto. Ma ci sono altre cose che un uomo ha paura di rivelare persino a se stesso, e ogni uomo perbene ha un certo numero di cose del genere accantonate nella mente […]".

Dopo di lui, il sociologo Zygmunt Bauman aveva osservato che "*La privacy è talvolta collegata all'anonimato, al desiderio di passare inosservati e non farsi riconoscere in pubblico. Solitamente, quando si considera «privato» qualcosa, è perché ciò ha per noi un'importanza intrinsecamente speciale, o ci tocca personalmente".*

In relazione ai temi qui trattati, particolare significato assume una "sottospecie" OSINT, definita *Social Media Intelligence* o SOCMINT, in precedenza interessata all'andamento di *brand* commerciali ed "appetiti" degli utenti per ragioni di *business* e, man mano, sempre più propulsa dalla c.d. *sentiment analysis*, una tecnica di monitoraggio, sia esso massivo che individuale, in grado di profilare ed analizzare in modo invasivo ogni interesse e propensione individuale.

Già nel 2011, ad esempio, fece clamore un *rootkit*, il noto *CARRIER IQ*, che sarebbe stato installato nativamente su milioni di *smartphone* al fine di **monitorare molteplici informazioni private relative all'utilizzo del *device*; si tratta di** un'applicazione nascosta in grado di consentire agli operatori, ed alle stesse case produttrici, **di veicolare successive scelte aziendali o stimolare, anche in modo silente o subliminale, gli interessi degli utenti con *banner* pubblicitari ed altri espedienti commerciali.**

Un interesse che avrebbe indotto ricercatori e aziende ad un investimento imponente verso lo sviluppo e utilizzo massivo di *cookies* e di *trojan,* sempre più spesso verso quei settori definiti *taboo* ed oggetto di *click* anonimo, come nel caso dei contenuti pornografici presenti su internet: una sorta di industria digitale del subliminale a scopo commerciale.

Si sarebbe arrivati, ad esempio, agli strumenti di manipolazione della navigazione sessuale: dal "***Porn-Napping***", finalizzato all'acquisto di domini scaduti, in precedenza collegati a *brand* commerciali di spessore e che, nella nuova rivisitazione aziendale virtuale, avrebbe rimandato a siti con contenuti pornografici a pagamento; così come al "***Cyber Squatting***", con le vetrine fatte di dominii che, celandosi dietro lessici innocui, rimandano al *business* pornografico.

Escamotage pubblicitari attraverso l'icona dell'"***Advertising***" di "*OK*" che, eseguito il comando, indirizza ad un sito con contenuto a pagamento o pubblicitario.

Come ancora il "*Tabnabbing*", una tecnica di *phishing* che presenta all'utente, attraverso un link di rimando, una pagina del tutto simile a quella genuina utilizzata per la fruizione di servizi con l'identificazione attraverso *user id* e *password*, consentendo all'attaccante di impossessarsi delle credenziali di accesso con il furto di identità.

Altro esempio è quello del "***Misspelling***", architettato per rimandare, attraverso la digitazione di una stringa di testo molto cliccata, come nel caso della parola "*googlle*" (con una "l" in più rispetto al noto motore) ad un sito porno asiatico.

La varietà di espedienti commerciali è continua e irrefrenabile, come per il "***Looping***", che fa aprire continuamente pagine ogni qual volta si tenta di chiuderne una precedente; o il c.d. "***Mousetrapping***" che altera il funzionamento dei tasti di comando del *mouse*, non consentendo all'utente di abbandonare un sito esplorato.

O il più comune "***Startup File Alteration***" che indirizza, continuamente, a un programma nella *directory* di avvio che rimanda ad una determinata pagina (solitamente pornografica) all'avvio del dispositivo.

Un caso a sé è il recente strumento di estorsione sul web noto col il termine "***Ransomware***": un *malware* che limita l'accesso

al proprio *device* una volta inoculatosi, seguito da una richiesta di un riscatto – monetizzato di solito attraverso cripto valute – al fine di ricevere i codici di decrittatura del blocco del sistema (accade sempre più spesso di ricevere, ad esempio, mail che rimandano alla spedizione di un plico, alla scadenza di una polizza assicurativa, ad un rinnovo contrattuale, con la presenza di un, apparentemente innocuo, *hiperlink* di rimando che contiene il collegamento di infezione).

SOCIAL MEDIA INTELLIGENCE E SENTIMENT ANALYSIS

Il sistema si è, dunque, evoluto attraverso protocolli di analisi in grado di insinuarsi nelle reti e negli innumerevoli canali sociali al fine di identificare profili, individuare relazioni, focalizzare organizzazioni e tracciare *link* di collegamento utili alla ricostruzione di un *profile* relazionale documentabile e descrivibile con analitici diagrammi di contesto.

"*Tramite la SOCial Media INTelligence* – precisa Paolo Dal Checco – *è possibile reperire informazioni utili al ciclo di intelligence tramite il monitoraggio e l'analisi dei contenuti scambiati attraverso i Social Media. La tecnica SOCMINT risulta ancora decisamente legata alla disciplina OSINT, Open Source INTelligence, e in modo più lieve agli altri approcci di ricerca d'informazioni come TechINT, GeoINT, SIGINT, HumINT, MasINT. E' importante precisare che la SOCMINT si basa sulle informazioni acquisite tramite i Social Media senza limitarsi ai Social Network come Facebook, Twitter, Instagram e simili*".

La SOCMINT, spiega Alessandro Burato, "*si è concentrata sui processi di data mining rivolti a due principali aspetti degli ambienti "social": i contenuti e le relazioni. I primi hanno dato origine a diversi studi sui cosiddetti motori semantici*

utili nell'analizzare e filtrare grandi stringhe di dati come sono quelle delle comunicazioni che "scorrono" sui social, i secondi si concentrano maggiormente sulle relazioni virtuali che intercorrono tra i diversi user avvalendosi delle moderne tecniche di visualizzazione dei dati per darne una visione più ampia ed immediata. Evidentemente, tali visualizzazioni non esauriscono la ricchezza di informazioni e significato dei dati che rappresentano e dal punto di vista delle scienze sociali si rendono necessari dei sistemi di verifica di affidabilità ed efficienza di tali strumenti".

Nelle metodiche SOCMINT per fini di *intelligence* o di giustizia potrà, però, accadere che l'analista o il *profiler* debba spingersi oltre i contenitori delle fonti aperte, così da insinuarsi nella rete sociale del soggetto d'interesse cercando di accedere a quei "privilegi" per pochi, attraverso la condivisione di una o più "*amicizie*", requisito indispensabile per avere accesso alle notizie presenti sui *social media.*

Detta fase potrebbe essere alimentata e proseguita con forme più meno complesse di intercettazione tattica di comunicazioni, sia dei flussi ex art. 266 bis c.p.p., sia con quelle forme di intercettazione di carattere preventivo, non compiutamente definite, previste ex art. 226 att. c.p.p. ed in materia di informazione e sicurezza da parte delle Agenzie AISI e AISE, ex art. 12 della L. 133 del 2013 (nel qual caso – alla luce della nuova riforma sulle intercettazioni introdotta dal d.lgs. 216/2017, che andrà in vigore nel prossimo mese di luglio – non sono previste cautele in materia di utilizzo delle tecnologie " dual use" e, in particolare, del c.d. captatore informatico "Trojan" o s.w. di intrusione informatica).

L'infiltrazione nella rete sociale del bersaglio – parlando adesso delle fonti aperte – avviene, allora, in modo indiretto al fine di evitare l'insospettimento del *target*, monitorando, accanto ai collegamenti pubblici, le sue reti (sia essa

facebook, linkedin, twitter, tinder, thunder, whatsapp o qualsivoglia strumento di aggregazione e condivisione social) o, qualora possibile ed opportuno, attraverso richieste dirette di "amicizia" con l'utilizzo di un profilo *fake* di copertura (nel caso di attività istituzionali per fini di giustizia e di sicurezza nazionale, il legislatore ha previsto particolari esimenti con le cautele dello *status* giuridico di agente sotto copertura).

E qui che entrano in gioco motori di ricerca semantica approfondita, come *graph.tips*, attraverso cui sarà possibile esaminare *Facebook* selezionando molteplici *query* di filtro ricerca (vds. Ad esempio applicativi a pagamento di investigazione OSINT come *inteltechiniques.com*).

L'investigatore sulle fonti aperte si troverà soventemente a districarsi in rompicapo e scatole cinesi che, nel tempo, hanno mascherato un certo dominio, riuscendo a ricostruire una ricerca con una connessione logica partendo dai *tools* per l'identificazione di *domine name*, fino alla riesumazione di percorsi chiusi con protocolli di *wayback,* utilizzando una sorta di evoluta "macchina del tempo" di internet, detta *wayback machine*, in grado di recuperare siti *web* chiusi attraverso la consultazione di un contenitore di dati realizzato da una organizzazione *no profit* raggiungibile dall'indirizzo *archive.org*.

Una modalità di rappresentazione relazionale che si sta, man mano, diffondendo tra gli analisti esperti è, anche, quella della profilazione delle relazioni sociali attraverso l'applicativo *facelink* – un *software* che richiama il ben noto *i2 Analyst's Notebook* – dedicato all'analisi dei *social media* ed in grado di interfacciarsi con altri strumenti di analisi relazionale, come *TETRAS HPG* o *SFERA*, attraverso cui sarà possibile mettere in relazione, ancora, utenze telefoniche, contatti, identificativi IMEI (il numero seriale di un telefonino) di più *devices* e codici IMSI (il codice seriale di una scheda telefonica cellulare) abbinati a SIM di tutto il

globo.

Altri strumenti di analisi evoluta sono i *browser* forensi di acquisizione di siti web (come ad esempio AFW), in grado di cristallizzare anche le pagine web presenti sul *Dark net* attraverso la rete TOR, con le funzioni di *crawler* (cioè un *software* che analizza i contenuti di una rete o di un database in un modo metodico e automatizzato, in genere per conto di un motore di ricerca), così da ricercare tutte le pagine web collegate alla pagina principale, estraendone l'URL per creare un indice da cui può essere successivamente acquisito automaticamente.

Nelle versioni dedicate alle esigenze istituzionali il s.w. permette, inoltre, di eseguire ricerche su siti web con aree protette da *login,* come i *social network* e, in versione multipagina, consente la cattura automatica di un elenco di pagine web.

IL FUTURO DELLA *SOCIAL INTELLIGENCE*

In definitiva, cosa significa nel complesso parlare di *sentiment analysis*?

Si tratta di un raffinato protocollo di *intelligence* rivolto a scandagliare il web attraverso tecniche di SOCMINT (ma anche COMINT), con applicativi di vario genere attraverso cui è possibile effettuare profilazioni individuali, acquisire informazioni emozionali e descrivere articolati collegamenti relazionali tra soggetti.

Uno strumento versatile e adattabile alla specifica esigenza investigativa di *intelligence* (sia essa giornalistica, commerciale, merceologica, di ricerca sociale, geopolitica, strategico militare, di investigazione giudiziaria o di informazione e sicurezza istituzionale) che con l'adozione di attagliate *queries* di ricerca, consentirà di consultare singoli profili *social*; inquadrare relazioni tra i profili

social di interesse; consultare flussi complessi di comunicazioni *social* per tematiche ed aree geografiche; analizzare non solo dei protocolli HTML ma anche di *dinamic java script web pages;* bypassare le *application programming interface* (API); effettuare ricerche fotografiche su base merceologica ed antropologica; effettuare ricerche e comparazioni foto/biometriche, effettuare monitoraggi delle immagini presenti sulla rete attraverso applicativi di *device* che utilizzano tecnologie Android ed iOS; effettuare complesse ricerche merceologiche sui motori di ricerca del *clear* e del *deep*, con *report* statistici di immediata consultazione e rappresentazione grafica, per macro/micro area geografica, prodotto ed utenza; ricostruire dominii ed effettuare autopsie digitali su vecchi siti rimossi; costruire relazioni sociali attraverso le varie applicazioni mondiali di *social network* e di telefonia radiomobile abbinata ai profili.

Questo di cui si è parlato in poche righe non è un approccio di *intelligence* tra i tanti, ma il presente ed il futuro delle più evolute tecniche di monitoraggio e profilazione massiva sociale; uno strumento, cioè, in grado di conoscere e condizionare interessi ed orientamenti di fasce e categorie sociali ma anche e soprattutto lo strumento più importante ed evoluto che gli organismi di *intelligence* e di sicurezza di tutto il pianeta avranno a disposizione per arginare e prevenire le nuove forme di terrorismo che oggi, come mai prima, corre sul filo del web, attraverso la propaganda, il proselitismo, l'emulazione e l'intimidazione psicologica.

Big brother is watching you, direbbe sorridendo oggi George Orwell, ma se il prezzo da pagare per la *salus rei publicae* fosse questo, nessun utente della rete, di certo, ne risentirebbe: ma sarà così?

A cura di: **Michelangelo Di Stefano**

L'articolo Sentiment Analysis e Social Media Intelligence proviene da ICT Security Magazine.

https://www.ictsecuritymagazine.com/articoli/sentiment-analysis-e-social-media-intelligence/