OSSERVATORIO
SULL'INFOWARFARE E
TECNOLOGIE EMERGENTI

ISTITUTO ITALIANO
DI STUDI STRATEGICI
ITALIAN INSTITUTE
OF STRATEGIC STUDIES
Niccolò Machiavelli

SHADOW WARFARE

CYBERDETERRENZA, MALWARE E MACHINE LEARNING

COME IL CASO STUXNET PUÒ AIUTARE A COMPRENDERE LA TRANSIZIONE ALLA CYBER-DETERRENZA POTENZIATA DALL'INTELLIGENZA ARTIFICIALE.

FABIO VANORIO

EDIZIONI MACHIAVELLI

www.strategicstudies.it

APRILE 2021

***G**lobal **E**ducation and*
O**ccident **S**tatecraft **Enterprise

Global Education and Occident Statecraft Enterprise (acronimo GEOS Enterprise) è una *startup* innovativa italiana con sedi a Roma ed in altri tre continenti, attraverso le quali assicuriamo una copertura pressoché globale.

GEOS Enterprise realizza le proprie strategie di business in sinergia con altre imprese, università, centri di ricerca, *think tank*, organizzazioni nazionali ed estere, in tutti gli ambiti aventi ad oggetto la diffusione della conoscenza, l'alta formazione, lo sviluppo delle tecnologie nei campi degli affari strategici, della sicurezza internazionale, della difesa, dell'intelligence, dell'economia, delle relazioni politiche internazionali, dell'organizzazione delle imprese, della comunicazione, degli affari legali, dei trasporti, dell'energia, dell'ambiente, dell'aerospazio, della cibernetica, dello sport e di ogni altra disciplina di tipo socioeconomico di interesse strategico.

GEOS Enterprise sviluppa ricerca e innovazione, orientate alle esigenze sociali negli ambiti "Salute e assistenza", "Sostenibilità, Protezione del clima e dell'ambiente, Energia", "Mobilità", "*Smart Cities*" "*Safety and Security*", "Economia e lavoro 4.0", e le trasformiamo in soluzioni leader per industrie, infrastrutture, imprese, territori e "*Homeland Security*".

Nonostante sia stata fondata recentemente, GEOS Enterprise partecipa attivamente a diversi bandi di Ricerca e svolge parallelamente il ruolo di advisor nonché di incubatore-acceleratore di tecnologie innovative per conto di *spin-off* universitari, *startup* innovative e aziende che sono pronte ad eccellere nel competitivo mercato globale.

www.geosenterprise.com

ISTITUTO ITALIANO
DI STUDI STRATEGICI

ITALIAN INSTITUTE
OF STRATEGIC STUDIES

Niccolò Machiavelli

Osservatorio sull'infowarfare e tecnologie emergenti

L'Osservatorio analizza tendenze globali e key factor nel campo delle information operation. Inoltre studia e valuta l'impatto delle tecnologie emergenti sulla sicurezza nazionale italiana.

AUTORE

Fabio Vanorio è un dirigente del Ministero degli Affari esteri e della Cooperazione Internazionale, nonché cultore della materia della *cyber-warfare*. La sua attività di studio e ricerca è svolta in maniera indipendente. Relativamente al presente documento, le opinioni sono espresse a titolo personale e non sono riconducibili al Ministero degli Affari Esteri e della Cooperazione Internazionale.

Indice

Introduzione

Il cyberspazio sta crescendo rapidamente e attualmente comprende più di 17 miliardi di dispositivi connessi, e decine di miliardi di dispositivi si aggiungeranno nei prossimi decenni (Evans 2011). A fronte di questo sviluppo esponenziale, i modelli prevalenti di sicurezza nazionale impiegati anche nel dominio digitale restano basati su un "leviatano" a livello statale che detiene la maggior parte, se non tutte, le leve del potere, su sistemi che contemplano attori razionali unitari, sul ricorso a forme di deterrenza basate su minacce di ritorsione, in generale, su un mondo in cui l'uso delle armi e gli atti di guerra non possono essere occultati.

Nel cyberspazio, dove il conflitto è raramente palese e le minacce di punizione troppo spesso sono risultate vuote, sia per carenze di reputazione che per l'incapacità di inviare "segnali" credibili ai soggetti coinvolti, la deterrenza convenzionale intesa nella sua forma tradizionale di meccanismo di dissuasione dei potenziali avversari si è dimostrata più volte inadatta e insufficiente.

La letteratura sul tema della cyber-deterrenza, invece, continua a crescere. L'applicazione di meccanismi di deterrenza in specifici settori del *cyber-warfare* (come, ad esempio, il contrasto al crimine online, al furto di identità o alla sottrazione indebita di dati, informazioni e ricerche in campo scientifico e tecnologico ha consolidato l'importanza di approcci orientati alla dissuasione e alla neutralizzazione preventiva delle minacce anche nel cyberspazio. Questa maggiore sensibilità ha accresciuto il dibattito sulla rilevanza di questo approccio nell'influenzare sia i comportamenti online (malevoli e non) di "aggressori" (individui e paesi), sia i comportamenti di autodifesa online dei "*target*" (Maimon 2020).

DETERRENZA NEL DOMINIO CINETICO E NEL *CYBER REALM*.

L'oggetto della deterrenza aiuta a distinguere la deterrenza convenzionale dalla cyber-deterrenza.[1] Nel mondo fisico, la deterrenza da parte di "chi si difende" mira a scoraggiare il potenziale "aggressore" dall'effettuare attacchi fisici contro specifici beni, o categorie di beni. Nel cyberspazio, la deterrenza mira a scoraggiare le manipolazioni virtuali di uno specifico ambiente di rete e degli elementi che caratterizzano quel particolare ambiente (Brantly 2018).

In caso di cyber-aggressione (e conseguente fallimento dell'attività di deterrenza), è possibile evincere prove utili relative alle origini dell'aggressore (c.d. "attribuzione", ovvero essere in grado di "analizzare il DNA" del metodo e del codice di attacco) e calibrare l'adeguatezza della reazione (c.d. "proporzionalità").

Tralasciamo in questa sede ogni richiamo al concetto tradizionale di deterrenza che i manuali di relazioni internazionali spiegano in ogni sua forma. Ci concentriamo, invece, nel dominio virtuale dove il concetto di conflitto è oggi estremamente sfumato, composto da ambienti tra loro interconnessi che espandono le possibilità di attacco-risposta multidimensionale - l'aggressione in un dominio può avere una risposta in uno o più domini differenti. La deterrenza è diventata sfumata e liminale (ossia al livello della soglia della coscienza e della percezione). Per questo

1. Devo un ringraziamento particolare al Colonnello Francesco Marradi dell'Aeronautica Militare per il *proof-reading*, ed il costante apporto di pensiero *outside-the-box* nella predisposizione dell'intero paper.

sia l'attribuzione che la proporzionalità della ritorsione sono molto più complesse che nel concetto tradizionale di deterrenza, motivo per cui la trasposizione della teoria convenzionale della deterrenza nel cyberspazio è inefficace (Taddeo 2018).

Strategicamente, una corretta attribuzione è alla base dell'aspetto coercitivo della deterrenza, in quanto l'identificazione consente di indirizzare una "adeguata" ritorsione contro il nemico reale garantendo una congrua azione di contrasto. Se l'attribuzione è discutibile o sbagliata, la rappresaglia può causare ulteriori attriti e conflitti (Taddeo 2018).

Le complessità derivano dal carattere distribuito del cyberspazio che facilita l'anonimato. Le modalità impiegate nei cyber-attacchi generano problemi nella deterrenza in quanto questi vengono lanciati in fasi diverse coinvolgendo reti informatiche distribuite a livello globale, così come frammenti di codice che combinano diversi elementi forniti da - o sottratti da - diversi attori.

Nel cyberspazio, l'esistenza di difficoltà nel procedere ad una corretta identificazione rende gli attacchi strategicamente vantaggiosi - dal punto di vista dell'aggressore - minando sia il processo di attribuzione che la credibilità della successiva rappresaglia. Come ha sottolineato Libicki (2009), quanto minore è la probabilità di identificare gli aggressori, tanto maggiore deve essere il costo di un attacco perfettamente attribuito. Più elevata la sanzione ipotetica, però, maggiore è la probabilità che la ritorsione sia considerata sproporzionata dalla comunità internazionale. Ciò, in particolare, se a fronte di uno scontro virtuale si associ una minaccia tangibile.

Fondamentale per una efficace strategia di deterrenza è la trasmissione di un messaggio coercitivo (minaccia). Ciò può avvenire mediante la segnalazione (*signaling*), la cui credibilità dipende dalla reputazione del difensore di dare concretezza alle sue minacce. La reputazione è un aspetto centrale della cyber-deterrenza. Negli scenari cinetici, la reputazione si ottiene "mostrando" le capacità militari di uno Stato (con parate e dispiegamento di soldati o navi ai confini dello Stato aggressore) e manifestando la capacità di scoraggiare o sconfiggere l'avversario nel tempo. Nel cyberspazio, la reputazione di uno Stato può non corrispondere necessariamente alla sua effettiva capacità di difesa e di offesa, poiché gli Stati sono riluttanti a far circolare informazioni sia sugli attacchi informatici che ricevono sia sulle loro effettive capacità cyber (Taddeo 2018).

Quattro fasi per una corretta segnalazione potrebbero essere:

- *Signal* 1: creare un Comando Operazioni Cyber.
- *Signal* 2: rendere di pubblico dominio una volontà (espressa a livello normativo interno) di risposta ad ogni attacco informatico che tenda a destabilizzare la sicurezza nazionale.
- *Signal* 3: attribuire risonanza pubblica ai successi nella cyber-deterrenza da parte degli apparati di ricerca scientifica e tecnologica a tutela del c.d. "perimetro di sicurezza nazionale cibernetica", caratterizzato da quei soggetti che erogano servizi essenziali nell'ambito delle infrastrutture nazionali per le telecomunicazioni.
- *Signal* 4: pubblicare sui media articoli descrittivi relativi a cause ed effetti di epidemie e pandemie di virus informatici in determinate nazioni o aree geografiche.

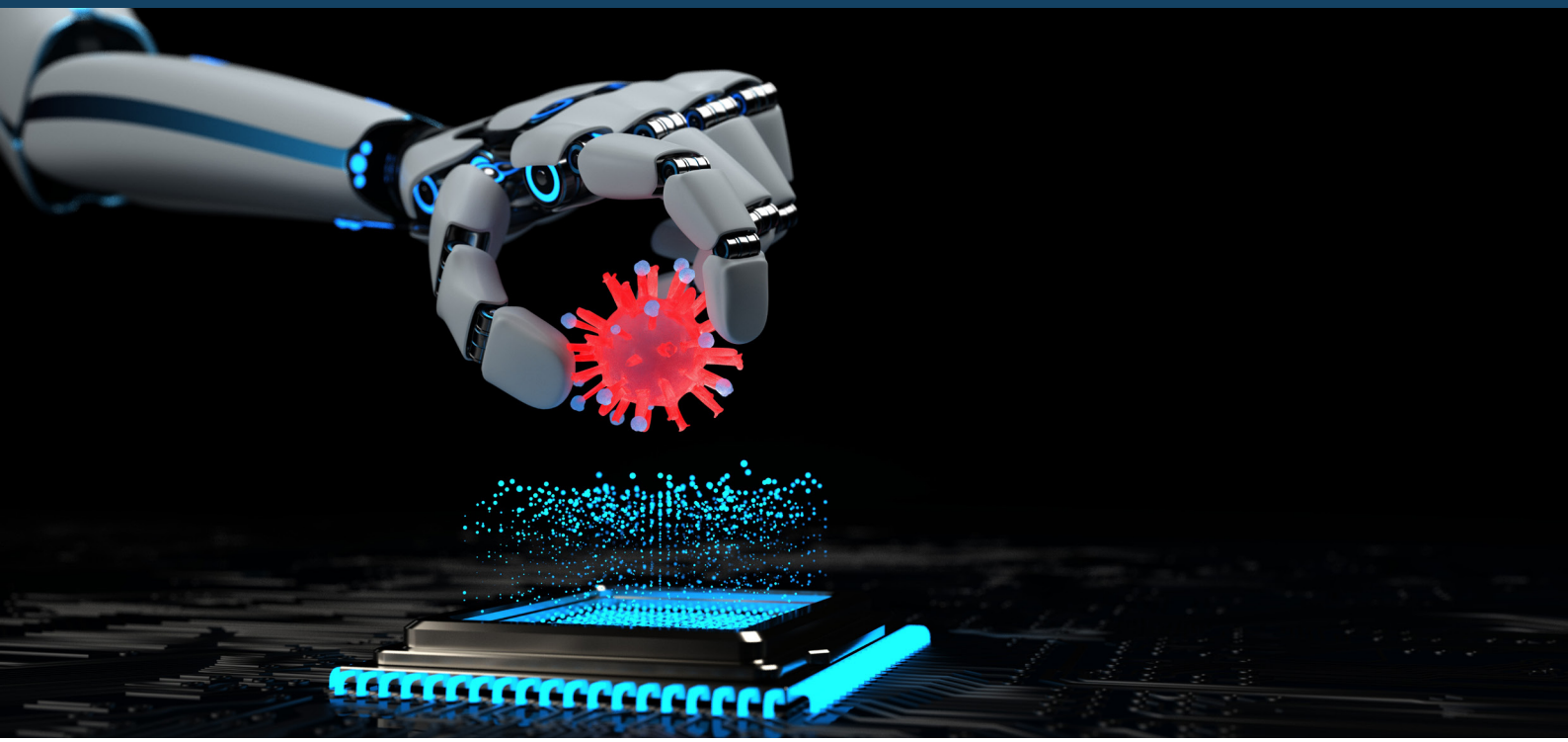
Glaser (2011) identifica alcune sfide associate alla cyber-deterrenza che vanno oltre la questione fondamentale della credibilità cibernetica di un attore statuale:

- il cyberspazio manca di misure di "*hand-tying*", ossia di costi nei confronti dell'opinione pubblica (c.d. *audience costs*) che impediscano all'attore statale di ripensare o di diluire una sanzione minacciata a fronte di una certa aggressione (Fearon 1997).
- le opzioni limitate di risposta all'interno del cyberspazio si traducono in uno sconfinamento trasversale in ambienti cinetici. Mentre su terra, mare, aria e spazio, la deterrenza ha una capacità limitata di manipolare la natura del dominio stesso, lo stesso non vale nel cyberspazio. Ogni caratteristica del cyberspazio è manipolabile: la rete, l'*hardware*, il *firmware*, il *software*, i dati, o l'accesso di individui all'interno e all'esterno del sistema. Rappresaglie su più domini trasformano il quadro della *cyber-escalation* da digitale a cinetico.
- nell'arena cibernetica, molti stati usano "*proxy*", ossia attori non statali spinti da

zelo ideologico, nazionalismo o denaro, per mantenere una negazione plausibile e ostacolare la corretta attribuzione (Borghard e Lonergan, 2016). I "*proxy*" stanno emergendo come strumenti di primo piano nella gestione dei conflitti, in ogni dimensione non solo digitale. Sono diventati "metodo", non solo "strumento". *Alias*, identità multiple, credenziali contraffatte o rubate, crittografia, *server proxy*, reti private virtuali consentono l'anonimato nel cyberspazio creando un vantaggio offensivo, minando l'attribuzione e tracciando gli attacchi in giurisdizioni internazionali differenti da quella effettiva. Taluni attacchi vengono anche realizzati in modalità "*false flag*", ossia volti a indurre sospetti su una terza parte. Mentre l'attacco anonimo avviene e si conclude in millisecondi, le indagini penali transfrontaliere rappresentano un processo che richiede molto tempo e dai risultati incerti. Sono necessari diversi livelli di impegno per scoprire il Paese d'origine, i componenti informatici impiegati, gli autori specifici e l'organizzazione che sponsorizza l'attacco (Lindsay 2013).

Quanto affermato, dunque, mette in dubbio una delle componenti essenziali della tesi della *Cyber Revolution*, quali quella relativa ai vantaggi asimmetrici di Internet come punto di forza degli attori più deboli (Lindsay 2013). La *Cyber Revolution*, infatti, deduce le conseguenze strategiche direttamente dalla crescente vulnerabilità pervasiva derivata dall'onnipresenza dei computer, piuttosto che da una minaccia specifica di attacco catastrofico.

STUXNET COME ANNULLAMENTO DEL VANTAGGIO ASIMMETRICO



Un caso empirico di attacco informatico ufficialmente noto è l'utilizzo del *malware Stuxnet* contro l'Iran. Questo evento ha mostrato tre effetti (Lindsay 2013):

- gli attori più forti possono potenziare le loro capacità di cyber-attacco rispetto a quelli più deboli.
- la complessità delle armi informatiche disponibili rende l'attacco informatico difficile da realizzare e la difesa informatica più facile di quanto non sia generalmente considerato.
- una deterrenza strategica credibile aumenta la significatività delle opzioni informatiche disponibili.

Nella visione espressa da Sanger (2012), la preparazione tecnica di *Stuxnet* (effettuata dalla *National Security Agency* statunitense insieme ad un'unità militare israeliana SIGINT (*Signal Intelligence*), nota come Unit 8200, è durata circa tre anni, necessari per infiltrare con successo il *malware* nella centrale di Natanz (Iran), controllarne la permeazione ed eseguire, attraverso di esso, azioni specifiche sull'obiettivo.

Stuxnet non può quindi essere definito come un'arma di Stati "deboli".[2] Solo uno Stato "forte" può disporre di una capacità, organizzativa, finanziaria e tecnologica che gli consenta di gestire al meglio sia i rischi strategici dell'identificazione (qualsiasi attacco informatico, per quanto sofisticato, mantiene, infatti, una probabilità non nulla che la missione possa essere compromessa, che l'identità dell'aggressore sia scoperta o che l'operazione manchi l'obiettivo previsto), sia i rischi connessi alla proporzionalità (mentre uno stato "debole" che fallisce nel suo attacco ad uno "forte" non migliora la sua posizione e può diventare soggetto anche ritorsioni, al contrario una potenza che attacca un paese più debole può usare la sua forza militare come scudo).[3]

Stuxnet evidenzia un punto debole nella questione dell'asimmetria dovuto alla più difficile acquisizione di conoscenza da parte di uno stato debole. Gli stati (o attori) fragili devono superare elevate barriere all'ingresso (dovute ai costi di procurement e di trattamento specialistico) nell'acquisizione di *malware* che causino danni significativi (Lindsay 2013).

Forza e Debolezza nella comparazione tra Stati

Nel valutare la debolezza e la forza di uno stato a livello settoriale (dunque anche riferibile allo specifico settore cyber), è fondamentale determinare, in primo luogo, il grado in cui il potere decisionale finale è concentrato nelle mani di un numero relativamente piccolo di decisori e, in secondo luogo, il livello di autonomia di questi decisori. Normalmente, il potere è concentrato quando una singola agenzia o ufficio e' in grado di dominare le relazioni in un dato settore. Una tale unità organizzativa dispone della capacità di aggregare le istanze dagli attori settoriali, siano essi imprese o associazioni di interesse. Al contrario, lo stato è debole in un dato settore quando l'autorità è dispersa e non vi è alcun gruppo di decisori che può prendere il comando nella formulazione della politica. In queste circostanze, l'autorità è tipicamente diffusa tra diversi uffici e tra i livelli di governo con conseguente sovrapposizione di giurisdizioni e competizione burocratica. (Atkinson and Coleman 1989).

2. Si può affermare l'esistenza di due livelli come ai tempi della MAD (*Mutual Assured Destruction*). Le guerre esistevano ed erano ritenute delle "*venting valve*" per togliere rigidità al sistema di minaccia nucleare e dare profondità di risposta. La *Cyber Revolution* occupa il livello inferiore, sfoga il proprio potenziale a "livello locale". Diventa, dunque, necessario definire cosa è "locale" e cosa non lo è nel dominio digitale.

3. Israele e gli Stati Uniti hanno potuto sperimentare un progetto scientifico come *Stuxnet* perché entrambi (presi insieme e forse anche singolarmente) mantengono una capacità militare superiore all'Iran (Lindsay 2013).

Il dilemma della sicurezza ci ricorda come la conflittualità tra gli stati possa aumentare a causa dei tentativi di ciascuno di aumentare la loro sicurezza che rendono Stati contigui (o concorrenti) meno sicuri. Una potenziale spiegazione per la relativa assenza di una guerra cibernetica strategica nella storia è che la difesa cibernetica è più robusta dell'offesa (Lindsay 2013).

Il caso *Stuxnet*, dunque, a prima vista, sembra un caso di deterrenza fallita, ma da un'angolazione diversa, è esattamente l'opposto.

Gli Stati Uniti hanno lanciato un cyber-attacco distruttivo, e l'attribuzione è rimasta a lungo ambigua fino a quando Sanger ha pubblicato la sua indagine nel 2012, un anno e mezzo dopo la scoperta di *Stuxnet*. Altri *malware* inclusi nell'operazione "Giochi Olimpici" (come Duqu e Flame) si sono rivelati anonimamente in agguato nelle reti iraniane. A fronte della attribuzione dubbia, gli Stati Uniti non sono stati in grado di scoraggiare le ritorsioni informatiche iraniane avvenute sotto forma di attacchi DDoS (*Distributed Denial of Service*) contro banche statunitensi e reti informatiche saudite. La dissuasione sembra, quindi, non essere riuscita a fermare le ritorsioni informatiche iraniane (Lindsay 2013) anche se, per quanto noto, sono risultate meno che proporzionali.[4]

La capacità di un cyberattacco di infliggere danni senza ricorrere alla violenza tradizionale consente alla *cyber-warfare* di disporre di una scelta più ampia di azioni e risultati disponibili per l'attacco strategico. L'operazione "*Olympic Games*" faceva parte di una campagna più ampia per privare l'Iran della capacità di produrre uranio di grado militare. Gli Stati Uniti e Israele erano d'accordo su questo obiettivo, ma erano in disaccordo su come raggiungerlo. Washington era preoccupata per le potenziali conseguenze di un attacco aereo sugli impianti nucleari iraniani (opzione preferita da Tel Aviv), temendo che potesse avviare una conflagrazione regionale (erano già in corso due conflitti - costosi e impopolari - ai confini dell'Iran con l'Iraq e l'Afghanistan) ed intensificare la determinazione di Teheran nelle sue ricerche nucleari. Il *worm Stuxnet* ha offerto ai due paesi una soluzione temporanea alle loro divergenze fornendo alcuni dei

4. In realtà, gli attacchi DDoS contro le imprese statunitensi hanno creato danni limitati con scarse conseguenze politiche internazionali o impatto sulle prestazioni delle imprese. Come ritorsione, il virus Shamoon, che avrebbe cancellato i dati di oltre 30 mila computer della società saudita Aramco e mostrato una propaganda antiamericana, sembra essere stato il lavoro poco sofisticato di un *hacker* nazionalista iraniano (Riley and Engleman 2012).

risultati tattici di un attacco militare, evitando allo stesso tempo ritorsioni cinetiche. Il fatto che le conseguenze derivanti dall'utilizzo di *Stuxnet* non fossero paragonabili alla scala di distruzione conseguente ad un attacco aereo era la principale attrattiva della nuova arma. Il *worm Stuxnet* da solo non avrebbe mai potuto prevenire una bomba iraniana, ma avrebbe potuto ritardarne l'arricchimento nucleare (Kello 2013). Il *malware Stuxnet* non ha mai voluto essere un intervento decisivo per fermare il programma nucleare iraniano. L'obiettivo era semplicemente di "introdurre della sabbia negli ingranaggi e guadagnare un po' di tempo" (Lindsay 2013).

In quest'ottica, l'onere dell'"attribuzione" non è più il problema del "difensore", ma diventa una responsabilità dell'aggressore che ha bisogno di anonimato e di negabilità plausibile. Il maggiore potere relativo della coppia Stati Uniti/Israele sull'Iran ha fornito la possibilità di sperimentare una scala completamente diversa di cyber-attacchi. Il potere militare più consistente ha fornito un'assicurazione contro le ritorsioni e un piano di ripiego coercitivo in caso di fallimento (Lindsay 2013).

Assumendo queste dinamiche come una deliberata moderazione nella gravità dei cyber-attacchi, si può dedurre che la deterrenza tra gli attori politici coinvolti abbia effettivamente funzionato. Questo caso ricorda il paradosso della stabilità-instabilità della teoria classica della deterrenza nucleare, secondo il quale la stabilità della deterrenza nucleare può promuovere una limitata instabilità convenzionale.

In tal ambito, il significato della cyber-deterrenza può essere la prevenzione degli effetti dell'attacco o di un uso inaccettabile della forza in una competizione multi-dominio, multilivello e con orizzonti temporali diversificati. Una deterrenza efficace in un contesto strategico porta gli attori a scegliere i mezzi informatici quando altre opzioni sono considerate troppo rischiose e, inoltre, a contenere l'intensità dei loro cyber-attacchi per evitare ritorsioni con mezzi cinetici (Lindsay 2013).

COME L'INTELLIGENZA ARTIFICIALE CAMBIA LE CONDIZIONI DI GIOCO

Stuxnet va in opposizione, dunque, alle tesi della *Cyber Revolution* per due motivi: (1) il *worm* non ha avuto alcun effetto duraturo nell'ostacolare il processo di arricchimento nucleare iraniano; (2) il ritardo aggiuntivo imposto al programma non ha prodotto colpi di Stato. Sembrano due fallimenti ma non lo sono se inseriti nel contesto specifico dell'azione clandestina condotta[5] (Lindsay 2013). La portata di *Stuxnet* era meramente tattica con un notevole elemento di contagio. Di per sé, *Stuxnet* ha rappresentato un'arma cibernetica intelligente, un virus creato per infettare uno specifico obiettivo e alterarne il funzionamento. Praticamente, un coronavirus con i limiti di un singolo organismo che non può evolvere e quindi sconta una rigidità di adattamento all'ambiente che comunque gli rimane ignoto per un certo periodo. Quindi, dall'efficienza incerta.

Stuxnet è un caso ideale per avviare un'analisi su come l'integrazione del *Machine Learning* possa migliorare le capacità offensive di un *malware*. La principale carenza di *Stuxnet*, infatti, è stata l'inefficace acquisizione del target che un'appropriata applicazione degli algoritmi di Machine Learning possono migliorare, riducendo i danni collaterali (Easttom 2019).

In generale, l'Intelligenza Artificiale ha un potenziale enorme nello sviluppo di *malware* intelligenti. Il passaggio ipotizzabile è da corona-virus a retro-virus, da Covid-19 all'AIDS. L'Intelligenza Artificiale permette il mimetismo[6], lo *swarming* e l'aggiornamento tramite autoapprendimento, l'evoluzione genetica e, in ultimo, la persistenza nell'organismo infettato.

5. Richard Clarke, coordinatore della sicurezza informatica per l'amministrazione Bush, ha osservato come la preparazione e l'uso del codice di *Stuxnet* per limitarne la propagazione e discriminare il suo obiettivo ha mostrato le caratteristiche di un'azione di intelligence clandestina (Rosenbaum 2012).

6. Con l'avvento dell'Intelligenza Artificiale, iniziano a circolare metodi di "inquinamento dei dati" e camuffamento digitale, i c.d. "attacchi avversariali", che sfruttano l'ipersensibilità alle configurazioni dei dati di input della distribuzione statistica codificata dentro le reti neurali di *Machine Learning*.

La domanda di ricerca che ci poniamo è se l'introduzione della *cyber warfare* potenziata dall'Intelligenza Artificiale sia in grado di definire migliori politiche di cyber-deterrenza.

In generale, una vasta letteratura (Rid 2012; Slayton 2017; Hoffman 2019) sostiene che l'Intelligenza Artificiale, usata come moltiplicatore di forza per armi cibernetiche sia difensive che offensive, è suscettibile di avere un impatto trasformativo sulla sicurezza cibernetica e sulla sicurezza internazionale.

Machine Learning e robotica in sistemi di combattimento che sfruttano le capacità delle persone e delle tecnologie costituiscono eserciti "ibridi" (o "centauri") che tendono sfruttare la precisione e l'affidabilità dell'automazione senza sacrificare l'energia e la flessibilità dell'intelligenza umana (Scharre 2018). Lo sviluppo di soggetti militari ibridi e agenti di guerra cibernetica è orientato ad aumentare la consapevolezza predittiva delle macchine, il che fornisce un vantaggio decisionale alle nuove "manovre cognitive" militari nello spazio globale dell'informazione (Dear 2019). L'adozione da parte degli Stati Uniti di dottrine di "impegno persistente" e di "difesa avanzata" nel cyberspazio (Healey 2019), con una discriminazione autonoma mirata, è suscettibile di generare un "processo decisionale ibrido", risultato di una precedente conoscenza e di una precedente programmazione (Amoore and de Goede 2008). Una sorta di ciclo OODA (*Observe-Orient-Decide-Act*) alimentato con steroidi.

In generale, l'applicazione dell'intelligenza artificiale alla tecnologia delle reti informatiche riguarda l'identificazione delle tipologie di attacchi basati sulla raccolta e l'analisi delle informazioni di rete (registrazione di indirizzi IP iniziale e reiterazione della ricerca delle fonti di attacco), blocco in autonomia dell'accesso alla rete protetta di indirizzi IP già ritenuti come ostili, generazione autonoma di *patch* di protezione per gli attacchi di rete corrispondenti a quelli già registrati. Più che a livello *hardware* (dove tecnologie basate su reti neurali aiutano nel controllo impiegando dispositivi sensori, la nostra analisi si sofferma sulla protezione del *software*, ed in particolare sui *malware* (Guideng 2019).

Programmi militari tipo *Project Maven* del Pentagono (c.d. *Algorithmic Warfare Cross-Function Team*) hanno dimostrato come la penetrazione di malware nei sistemi di riconoscimento autonomi sia difficile da rilevare, attribuire o contrastare efficacemente (Shachtman 2011).

Un'accresciuta autonomia della macchina, conseguente all'introduzione del *Machine Learning*, potrebbe anche aumentare la vulnerabilità militare agli attacchi cibernetici. Passando dall'abilitazione digitale alla dipendenza digitale, gli strumenti di difesa informatica potenziati con l'Intelligenza Artificiale diventano anche più dipendenti dalle reti per condurre operazioni, rendendosi così più vulnerabili. Secondo questo paradosso (noto come "paradosso della capacità/vulnerabilità"), uno Stato più potente potrebbe dover sferrare un attacco in prima battuta, nonostante la sua capacità sia superiore all'avversario, per assicurarsi che l'avversario stesso non sia in grado di sfruttare le vulnerabilità presenti nella tecnologia impiegata (Rid 2013).

Conclusioni

Gli stati tecnologicamente dominanti usano le loro capacità di *cyber warfare* come deterrente mediante politiche di "coercizione digitale". L'impiego di tattiche di *hacking* in qualunque competizione (militare, economica, o addirittura elettorale) ha segnato l'ultimo decennio e la tendenza si accrescerà nel XXI secolo. I problemi legati all'attribuzione e alla proporzionalità diventano rilevanti per Stati tecnologicamente dominanti quando compiono azioni cibernetiche clandestine, ad esempio, per rafforzare strategie politiche avviate nel mondo reale. La questione diventa correlata ai costi interni ed internazionali delle loro azioni in termini di dissenso espresso in sedi multilaterali o di norme predisposte per perseguire danni collaterali a loro attribuibili. Il potenziamento con l'Intelligenza Artificiale riduce i danni collaterali, migliora l'accuratezza e l'anonimato della *cyber warfare*, rende più complessa l'attribuzione, e con essa sia la giustificazione della ritorsione da parte dello Stato tecnologicamente più debole sia l'eventuale condanna dello Stato forte da parte della Comunità internazionale. In tal senso, si apre lo spazio per attori non-statuali ("*proxy*") finanziati più o meno privatamente dagli Stati per mantenere una negazione plausibile e nello stesso tempo rimanere nella competizione.

L'analisi dei *Big Data* attraverso il *Machine Learning* in un ambiente di rete *wireless* 5G/6G è in grado di migliorare ulteriormente la capacità di prevedere le minacce e di ridurre gli incidenti causati dal fattore umano (soprattutto gli incidenti e i malfunzionamenti causati dai "*false flag*"). Questi progressi potrebbero però amplificare contemporaneamente i rischi

di *escalation*. Il *Machine Learning* utilizzato come moltiplicatore di forza per la cyber-offesa (ad esempio, attraverso l'uso di "*deepfakes*" sui *social media*, o "*digital jamming*" contro gli sciami di droni) è significativamente più difficile da rilevare rispetto all'uso di strumenti APT (*Advanced Persistent Threats*). Anche in caso di cyber-attacchi, il Machine Learning può rendere impossibile il rilevamento.[7]

I sistemi di *Machine Learning* possono esacerbare la manipolazione del panorama informativo in cui vengono prese le decisioni, ad esempio, attraverso la contraffazione di e-mail, messaggi di segreteria telefonica o video, impersonificando utenti con i quali si interagisce abitualmente (l'Intelligenza Artificiale può replicare il tono, la lingua e lo stile di un utente al punto da non riuscire più a fare una distinzione – *deep-fake*). In ambiti sociali, questo può essere un fattore critico nella sua capacità di fomentare - via cyber – rivolte di massa. Contestualmente, il ragionamento può essere rovesciato mostrando il lato buono della *cyber warfare* alimentata. L'analisi georeferenziata degli accessi al social Parler intorno a Capitol Hill[8] in occasione dell'occupazione del Congresso statunitense e l'analoga analisi effettuata sulle posizioni delle basi militari americane[9] hanno evidenziato come tramite l'impiego di un "*deep-fake* sonda" sia possibile effettuare raccolta di *intelligence* diffondendo notizie false per far emergere il canale spia.

Il *Machine Learning* può essere anche impiegato, mediante *Remote Access Trojans* (RAT), per modificare i dati di addestramento utilizzati per l'apprendimento di sistemi intelligenti in uso ad un sistema. Una strategia di medio periodo può prevedere una modifica silenziosa, prolungata nel tempo, dei dati impiegati nell'addestramento così da poter attaccare anche dopo l'installazione delle difese di *Machine Learning*. Analogamente, per sistemi ostili di *Machine Learning* è possibile infiltrarsi all'interno di sistemi apprendendo metodi di comunicazione, porte e protocolli più comunemente usati all'interno di reti aziendali o governative.

L'Intelligenza Artificiale è in continua evoluzione e sta imparando a conoscere il modo in cui noi ci difendiamo. Ad ogni risposta, gli attacchi successivi aumenteranno di volume e di sofisticazione. A meno che la nostra Intelligenza Artificiale non sia migliore di quella avversaria, e a meno che non si crei la maggiore anti-fragilità possibile all'interno del perimetro di difesa cibernetica nazionale e sovranazionale, questi ultimi troveranno sempre il modo di entrare mettendo a rischio le infrastrutture strategiche, i servizi essenziali (energia, trasporti, ospedali) e la nostra intera società.

7. Nella tutela da malware, l'intelligenza artificiale può fornire ottimi risultati se unita a sistemi di *cloud computing* (nell'analisi degli indirizzi IP impiegati) e a tecnologie di comunicazione 5G (Guideng 2019.)

8. Dell Cameron and Dhruv Mehrotra, *Parler Users Breached Deep Inside U.S. Capitol Building, GPS Data Shows*, January 12, 2021, gizmodo.com/parler-users-breached-deep-inside-u-s-capitol-building-1846042905?rev=1610480731991.

9. Dell Cameron and Dhruv Mehrotra, *Leaked Parler Data Points to Users at Police Stations, U.S. Military Bases*, January 15, 2021, gizmodo.com/leaked-parler-data-points-to-users-at-police-stations-1846059897.

BIBLIOGRAFIA

Amoore, Louise, and Marieke de Goede. 2008. "Transactions after 9/11: the banal face of the preemptive strike." *Transactions of the Institute of British Geographers* 33 (2): 173–185.

Atkinson, Michael M. and William D. Coleman. 1989. "Strong States and Weak States: Sectoral Policy Networks in Advanced Capitalist Economies." *British Journal of Political Science* 19(1): 47-67.

Borghard, Erica D, and Shawn W Lonergan. 2016. "Can States Calculate the Risks of Using Cyber Proxies?" *Orbis* 60(3): 395–416.

Brantly, Aaron F. 2018. "The Cyber Deterrence Problem." In *CyCon X: Maximising Effects*. Eds. Minárik, Tomas, Jakschis, Raik, Lindström, Lauri. 2018 10th International Conference on Cyber Conflict. CCD COE Publications. Tallinn.

Dear, Keith. 2019. "Artificial intelligence and decision-making." *RUSI Journal* 164 (5–6): 18–25.

Easttom, Chuck. 2019. *Integrating Machine Learning algorithms in the Engineering of Weaponized Malware*. Capitol Technology University. Plano, TX (U.S.A.).

Evans, Dave. 2011. *The Internet of Things: How the Next Evolution of the Internet Is Changing Everything*. CISCO Internet Business Solutions Group.

Fearon, James D. 1997. Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs. *The Journal of Conflict Resolution*. 41(1): pp. 68-90

Glaser, Charles. 2011. *Deterrence of Cyber-Attacks and U.S. National Security*. GW-CSPRI-2011-5. Washington, D.C.: Cyber Security Policy and Research Institute.

Guideng, Xiao. 2019. *Analysis on the Influence of Artificial Intelligence in Computer Network Technology*. 2nd International Conference on Intelligent Systems Research and Mechatronics Engineering (ISRME 2019). Guangdong Engineering Vocational and Technical College, Guangzhou, Guangdong, China.

Healey, Jason. 2019. "The implications of persistent (and permanent) engagement in cyberspace." *Journal of Cybersecurity*.

Hoffman, Wyatt. 2019. "Is Cyber Strategy Possible?" *The Washington Quarterly*. 42(1): 131–152.

Kello, Lucas. 2013. The Meaning of the Cyber Revolution. *International Security* 38(2): 7-40.

Libicki, Martin C. 2009. *Cyber Deterrence and cyberwar*. RAND Corporation. http://www.rand.org/pubs/monographs/MG877.html

Lindsay, Jon R. 2013. "Stuxnet and the Limits of Cyber Warfare." *Security Studies*. 22(3): 365-404

Maimon, David. 2020. "Deterrence in Cyberspace: An Interdisciplinary Review of the Empirical Literature." In: Holt T., Bossler A. (eds). *The Palgrave Handbook of International Cybercrime and Cyberdeviance*. Palgrave Macmillan.

Rid, Thomas. 2013. *Cyber War Will Not Take Place*. New York: Oxford University Press.

Rid, Thomas. 2012. "Think Again: Cyberwar." *Foreign Policy*. Accessed January 13, 2020. https://foreignpolicy.com/2012/02/27/think-again-cyberwar/.

Riley, Michael, and Eric Engleman. 2012. "Code in Aramco Cyber Attack Indicates Lone Perpetrator," *Bloomberg*. 25 October.

Rosenbaum, Ron. 2012. "Richard Clarke on Who Was Behind the Stuxnet Attack." *Smithsonian*. April.

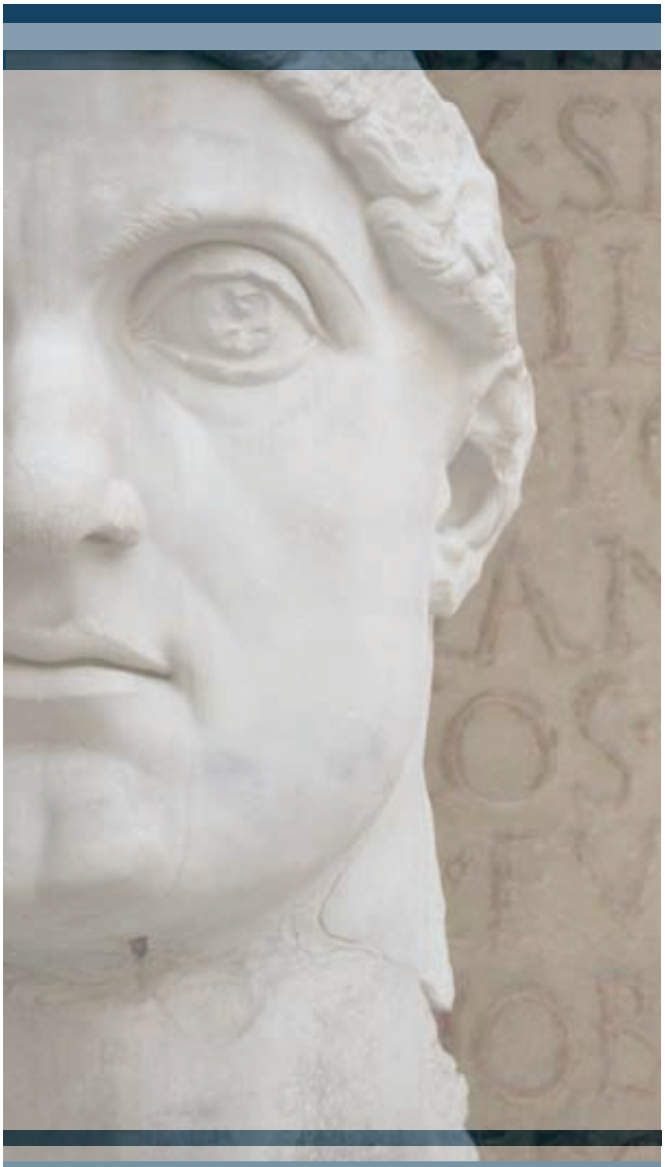
Sanger, David E. 2012. "Obama Order Sped Up Wave of Cyberattacks Against Iran." *New York Times*. 15 January.

Scharre, Paul. 2018. *Army of None: Autonomous Weapons and the Future of War*. New York: W.W. Norton.

Shachtman, Noah. 2011. "Exclusive: Computer Virus Hits U.S. Drone Fleet." *Wired*. https://www.wired.com/2011/10/virus-hits-drone-fleet/.

Slayton, Rebecca. 2017. "What Is the Cyber Offense-Defense Balance? Conceptions, Causes, and Assessment." *International Security*. 41(3): 72–109.

Taddeo, Mariarosaria. 2018. "The Limits of Deterrence Theory in Cyberspace." *Philosophy and Technology* 31:339–355.



ISTITUTO ITALIANO
DI STUDI STRATEGICI
ITALIAN INSTITUTE
OF STRATEGIC STUDIES
Niccolò Machiavelli

L'Istituto Italiano di Studi Strategici *"Niccolò Machiavelli"* è un'associazione culturale senza scopo di lucro costituita a Roma nel 2010.

L'Istituto, think tank indipendente, nasce dall'iniziativa di un gruppo internazionale di personalità del mondo economico, accademico ed istituzionale civile e militare, con l'obiettivo di contribuire alla rinascita del pensiero strategico italiano.

La complessità e l'ampiezza delle sfide che attendono il Paese nel XXI secolo richiede conoscenza, consapevolezza e capacità prospettiche. L'Istituto Machiavelli, anche grazie al proprio network globale, promuove l'interscambio culturale tra il decisore italiano ed internazionale, pubblico e privato, e svolge attività di ricerca finalizzate ad elevare il livello di competitività globale del "Sistema Paese".

L'Istituto Machiavelli, autonomamente o in collaborazione con istituzioni, organizzazioni ed aziende nazionali ed estere, realizza studi ed analisi strategiche *policy-oriented*, organizza briefing, seminari e workshop, cura corsi di alta formazione per i *leader*.

Per ulteriori informazioni:

Istituto Italiano di Studi Strategici "Niccolò Machiavelli"
Circonvallazione Clodia N. 163/167
00195 – Roma
Tel.: (+39) 06 45422952
Fax: (+39) 06 97259168
email: info@strategicstudies.it
www.strategicstudies.it