POLITECNICO DI TORINO

Corso di Laurea Magistrale
in Ingegneria Informatica

Tesi di Laurea in Cybersecurity

Analisi critica e studio dell'applicazione dell'Intelligenza Artificiale ai sistemi di Cyber Defence nella Forza Armata Esercito Italiano

Relatore
Chiar.mo Prof. Paolo Ernesto Prinetto

Candidato
Ten. RN Co.Ing. Ferdinando Pulella

Anno Accademico 2020-2021

† *Ringraziamenti*

Desidero ringraziare il Professor Paolo Ernesto Prinetto per i preziosi insegnamenti durante questi ultimi due anni della mia carriera universitaria, per i suggerimenti ed il tempo dedicato alla mia tesi.

Intendo poi ringraziare il Tenente del Corpo degli Ingegneri Sonia Forconi, per l'aiuto indispensabile che mi ha fornito nello sviluppo e nella redazione del mio elaborato, il Comando C4 Esercito del VI Reparto Sistemi 5CI dello Stato Maggiore dell'Esercito e la Ditta Forcepoint per avermi fornito testi e dati indispensabili al completamento del mio lavoro.

Uno speciale ringraziamento alla mia famiglia per essermi stata vicino da sempre.
Ringrazio i miei amici, vecchi e nuovi, che hanno saputo aiutarmi nel momento del bisogno. Grazie a Silvio in particolare.

Last but not least, I wanna thank me.
I wanna thank me for believing in me.
I wanna thank me for doing all this hard work.
I wanna thank me for having no days off.
I wanna thank me for never quitting.
I wanna thank me for always being a giver,
And tryna give more than I recieve.
I wanna thank me for tryna do more right than wrong.
I wanna thank me for just being me at all times.

Indice

Sommario

Il momento storico che stiamo vivendo è particolarmente affascinante dal punto di vista tecnologico, in quanto caratterizzato ogni giorno da nuovi strumenti e formule di interazione uomo-macchina, un rateo di sviluppo evolutivo esponenziale ed una domanda sempre crescente di digitalizzazione a vari livelli.

La Forza Armata deve saper cogliere e guidare lo sviluppo di sistemi e soluzioni nel campo militare che consentano ai Comandanti di supportare la funzione di Comando e Controllo (C2), velocizzando la raccolta, l'analisi e la trattazione delle informazioni necessarie a decidere.

La minaccia cibernetica, intesa quale l'insieme delle condotte controindicate che possono essere realizzate nel o attraverso il cyberspace, o in danno di quest'ultimo e dei suoi elementi costitutivi [25], sta assumendo, in ragione delle sue intrinseche caratteristiche e degli effetti prodotti, crescente rilievo nel novero delle minacce non convenzionali. Con l'avvento di tecnologie come l'Intelligenza Artificiale sia le minacce informatiche sia le relative tecniche di difesa e di deterrenza hanno dovuto adattarsi.

Lo scopo di questo Elaborato è quello di analizzare come la NATO ed in particolare come la Forza Armata Esercito Italiano si stiano approcciando a questo nuovo dominio, come le nuove tecnologie verranno sfruttate e valutare le misure di sicurezza adeguate alle esigenze della Difesa.

La potenziale capacità di diffusione degli attacchi cibernetici ha portato l'Alleanza Atlantica a dichiarare il cyber spazio come dominio operativo già nel 2016, di fatto compiendo un salto di qualità nell'approccio verso questo tipo di minaccia. Questi attacchi possono causare anche l'attivazione della Clausola di Difesa Collettiva ai sensi dell'Articolo 5 del Trattato dell'Atlantico del Nord, a riconoscimento del fatto che la componente cibernetica degli scontri diventerà sempre più parte integrante dei conflitti convenzionali.

Come nell'ambito civile, anche nel mondo militare vengano applicate tecnologie di Machine Learning per lo sviluppo di sistemi, applicazioni, sistemi di guida o armi autonome. Tutti questi esempi sono applicazioni che esistono già da tempo, ma con la rivoluzione dell'Intelligenza Artificiale verranno automatizzati sempre di più.

L'applicazione di questa tecnologia ed in particolare dei metodi di Machine Learning alla sicurezza informatica per sistemi "mission-critical" presenta grandi opportunità ma anche grandi sfide. Scopo di questo Elaborato è di analizzare in maniera critica e puntuale gli usi, i rischi e gli sviluppi dell'intelligenza artificiale applicata alla sicurezza informatica.

Nella seconda parte della Tesi verrà analizzato il comportamento del servizio di e-mail security, che si basa su un modello di machine learning, utilizzato per la webmail dell'Esercito Italiano. Verrà studiato il comportamento del modello e testato su diversi dataset di messaggi classificati come spam per valutarne l'accuratezza. Il test verrà effettuato su un ambiente di prova fornito dall'Azienda. Verranno inoltre presi in analisi alcuni messaggi di tipo *phishing* noti tra gli utilizzatori della Webmail per studiare come il sistema di e-mail security risponde a questo tipo di minaccia.

Parte I

Capitolo 1

Dominio cyber in EI e NATO

1.1 Introduzione

I progressi tecnologici che negli ultimi anni hanno reso il mondo, in particolare l'ambito militare, sempre più connesso e digitalizzato, hanno anche portato ad una serie di minacce cyber che sfruttano i gap insiti nei software che regolano oggi una moltitudine di attività umane. Il settore della Difesa da sempre attinge alle tecnologie più progredite e si può valutare che circa il 60% delle attività militari di una nazione moderna sia di matrice cibernetica.

Ciò ha imposto alle Forze Armate di dotarsi di contromisure opportune a queste nuove minacce cyber, in accordo con quanto delineato dalla NATO. Un attacco cibernetico di alto profilo, infatti, condotto da un organismo statuale o addirittura multinazionale, può creare dei danni di estrema gravità.

Da un punto di vista militare, oggi lo spazio cibernetico è un dominio trasversale a quelli tradizionali (terrestre, marittimo, aereo e spaziale), caratterizzato da mancanza di geo-specificità e limitate capacità di attribuzione. Allo stesso tempo però rappresenta un vero e proprio teatro di operazioni, con la problematica di non essere appannaggio esclusivo delle componenti militari, in quanto trovano spazio anche organizzazioni private ed individui non necessariamente riconducibili ad entità statuali.

Si è di fronte ad uno dei più efficaci metodi di lotta asimmetrica, in quanto anche un singolo individuo può costituire una minaccia per lo Stato o soggetti come l'Unione Europea e la NATO.
Considerando l'elevato livello tecnologico che caratterizza gli assetti delle nostre Forze Armate, il dominio cyber implica un'esposizione che può impattare su diversi aspetti della sfera militare, come la gestione dei sistemi d'arma e le comunicazioni tattiche, operative e di comando e controllo.

Tale progresso, che sicuramente da un punto di vista è irrinunciabile, dall'altro rappresenta un'enorme vulnerabilità di fronte alle potenzialità di una eventuale minaccia cibernetica. Un'ipotetica intrusione nei sistemi di Comando e Controllo di una Forza Armata, finalizzata non solo allo spionaggio, ma anche al sabotaggio e al malfunzionamento, potrebbe nei casi peggiori portare ad una perdita di controllo dei propri asset, ad un decadimento delle reti di telecomunicazione, ad un'errata geo-localizzazione delle forze in campo, fino ad arrivare alla paralisi completa dei sistemi. Tutto ciò potrebbe compromettere l'esito di un'operazione, generando effetti tragici per il personale che verrebbe messo in pericolo.

1.1.1 Contromisure adottate

Gli obiettivi definiti sia in ambito europeo che in ambito NATO comprendono la realizzazione di solide capacità di cyber defence, resilienza e protezione delle infrastrutture.

Il progetto della Difesa Italiana si basa su quattro pilastri fondamentali:

1. **Organizzazione**: personale, logistica, dottrina, operazioni e le varie componenti normali di un comando, in grado di proiettare i diversi elementi in operazione.

2. **Infrastrutture**, che devono disporre di sistemi protetti e prevedere anche delle modalità d'azione protette, che dovranno man mano crescere anche nella cultura di tutela cibernetica.

3. **Formazione**: realizzazione di una struttura per la formazione di esperti di dominio appartenenti alla Forza Armata, anche attraverso l'uso di poligoni virtuali per l'addestramento alle operazioni cyber.
   La formazione sarà a favore degli ambienti interforze, alleati e soprattutto in sinergia con il mondo accademico e quello industriale.
   Sarà importante inoltre riuscire ad acquisire gli strumenti necessari per effettuare lo studio dei malware e dei rimedi contro la minaccia, oltre che fornire supporto ai responsabili della progettazione, sviluppo e gestione delle reti, man mano che la minaccia viene identificata e neutralizzata.

4. **Personale**: si ritiene infatti che più del 70% della capacità generale di qualsiasi ambiente cibernetico dipenda dall'abilità degli operatori.
   Questi andranno dunque attentamente selezionati e formati, traendo beneficio anche dal mondo accademico e della ricerca, e da altre realtà del comparto industriale nazionale.

La Difesa, nell'ambito delle capacità cyber che sta sviluppando sia per proteggere i suoi sistemi che per pianificare e condurre operazioni militari trasversali al dominio cyber, in linea con il quadro normativo vigente, è come sempre a disposizione del Paese, pronta a rendere disponibili in ogni momento le proprie capacità attuali e future per concorrere alla crescita della cyber security nazionale.

1.1.2 Prospettive

Come l'Italia, tutti i principali Paesi e organizzazioni internazionali, compresa la NATO - che da questo punto di vista ha sempre rappresentato il motore-, si stanno dotando di strutture militari di comando e controllo per operare nel dominio cibernetico.

Il Segretario generale della NATO Jens Stoltenberg ha evidenziato un incremento del 60% degli attacchi cyber alle strutture dell'Alleanza nel 2016, con una frequenza media di circa 500 attacchi al mese, la maggior parte dei quali non proverrebbe da privati, ma da istituzioni statali di altri Paesi.

Alla luce di questa preoccupante evoluzione e per affrontare la minaccia cyber, la NATO continua a promuovere un approccio sinergico tra gli alleati e anche tra gli altri partner, puntando sul miglioramento delle capacità di ciascuno stato di difesa cibernetica, per contribuire al rafforzamento della difesa collettiva e alla sicurezza dello spazio euro-atlantico. Da questo punto di vista le nostre strutture di cyber defence si interfacceranno con le realtà analoghe dei Paesi amici e alleati, e in particolare con il Centro di Eccellenza della NATO di Tallinn in Estonia.

1.2 Il quadro NATO

L'approccio dell'Alleanza Atlantica verso la cyber defence si è evoluto in modo significativo negli ultimi quindici anni, elevandone l'importanza quale elemento che può dare un contributo significativo a tutti e tre i *"core tasks"* stabiliti dall'attuale Concetto Strategico:

- **difesa collettiva**;
- **gestione delle crisi**;
- **sicurezza cooperativa**.

In particolare, è stato riconosciuto che un attacco cibernetico può arrivare a causare danni paragonabili a quelli di un attacco armato, e quindi diventare un caso di difesa collettiva ai sensi dell'**Articolo 5** del Trattato di Washington.

Articolo 5

Le parti convengono che un attacco armato contro una o più di esse in Europa o nell'America settentrionale sarà considerato come un attacco diretto contro tutte le parti, e di conseguenza convengono che se un tale attacco si producesse, ciascuna di esse, nell'esercizio del diritto di legittima difesa, individuale o collettiva, riconosciuto dall'ari. 51 dello Statuto delle Nazioni Unite, assisterà la parte o le parti così attaccate intraprendendo immediatamente, individualmente e di concerto con le altre parti, l'azione che giudicherà necessaria, ivi compreso l'uso della forza armata, per ristabilire e mantenere la sicurezza nella regione dell'Atlantico settentrionale. Ogni attacco armato di questo genere e tutte le misure prese in conseguenza di esso saranno immediatamente portate a conoscenza del Consiglio di Sicurezza. Queste misure termineranno allorché il

Consiglio di Sicurezza avrà preso le misure necessarie per ristabilire e mantenere la pace e la sicurezza internazionali. [28]

Già il Vertice dei Capi di Stato e di Governo del 2008 aveva adottato una prima *Policy on Cyber Defence*, che ha compiuto un salto in avanti nel summit del 2014 con la *Enhanced Nato Policy on Cyber Defence*.

Nel successivo vertice di Varsavia, nel 2016, i Paesi Alleati hanno elevato lo spazio cibernetico a dominio, equiparandolo agli altri quattro domini militari convenzionali. Il vertice di Varsavia ha portato anche alla firma del *Cyber Defence Pledge*, volto ad istituire una piattaforma comune per migliorare le capacità nazionali di difesa e resilienza rispetto ad un attacco cibernetico.

In seguito sono stati adottati diversi action plan per realizzare gli impegni presi con il *Cyber Defence Pledge*.
L'impegno Alleato si concentra sullo sviluppo di capacità in chiave difensiva, in riferimento all'**Articolo 3** del Trattato di Washington [1]. Si tratta di un focus in linea con l'elevata importanza attribuita agli attacchi cibernetici, giudicati sempre più frequenti, complessi e distruttivi, tanto che nel comunicato del vertice di Bruxelles del 2018 viene esplicitamente affermato che **la difesa cibernetica è parte della difesa collettiva NATO**[13].

1.2.1 Problematiche e deterrenza

Uno dei problemi principali a riguardo è la difficoltà nel distinguere una situazione di pace da una di crisi o di conflitto, data la capacità dell'attaccante di nascondere la paternità degli attacchi condotti. Di fronte a questa situazione, che ha visto anche il moltiplicarsi di attacchi cibernetici durante la prima ondata di Covid-19, a giugno 2020 il *North Atlantic Council* ha riaffermato che i Paesi membri sono

"determined to employ the full range of capabilities, including cyber, to deter,defend against and counter the full spectrum of cyber threats."

La NATO si è dichiarata pronta ad usare non solo capacità cyber ma anche aeree, marittime o terrestri, per rispondere ad un attacco cibernetico, considerando quindi tutti i domini operativi in modo integrato ai fini della deterrenza e difesa, in linea con l'integrazione del Cyber Operation Centre nella struttura di comando NATO come deciso durante il vertice di Bruxelles.

Per esercitare un'efficace deterrenza è tuttavia fondamentale la capacità di attribuire la paternità degli attacchi, una priorità che richiede ulteriori sforzi da parte degli Alleati.
Riguardo al dominio cibernetico la NATO riafferma in definitiva la sua natura di alleanza difensiva, e conferma che sia il principio che il diritto internazionale siano da applicare

[1] relativo alla capacità individuale e collettiva di resistere ad un attacco armato.

anche al cyberspace e devono essere rispettati.

Il vertice di Londra del 2019 ha dato nuovo slancio politico-strategico alle attività NATO nel campo cibernetico -assieme a quello spaziale- nella consapevolezza della competizione geopolitica mondiale.
Il Segretario Generale ha dichiarato che il *"Cyberspace is the new battleground and making NATO cyber ready—well-resourced, well-trained, and well equipped—is a top priority"*.

Non a caso, nel 2020 il rapporto del Gruppo di Riflessione sulla NATO in prospettiva 2030 ha attribuito grande importanza alle *Emerging and Disruptive Technologies* (EDT) intese sia come settore sul quale investire maggiormente, sia come sfide, tra cui rientrano prioritariamente proprio quelle relative alla cyber defence, in primis l'Artificial Intelligence (AI). Stoltenberg ha infatti sottolineato che *"le minacce cyber diventeranno più pericolose con lo sviluppo di nuove tecnologie come AI e machine learning [...]. Queste tecnologie stanno cambiando fondamentalmente la natura dei conflitti, tanto quanto avvenuto con la rivoluzione industriale. La NATO si sta adattando a questa nuova realtà"*[13].

È quindi molto probabile che anche il nuovo Concetto Strategico su cui verosimilmente l'Alleanza lavorerà nel 2021 porrà grande attenzione alla cyber defence, e in generale al dominio cyber e alle *EDT* come terreno di confronto con Cina e Russia.

1.2.2 Sviluppo di dottrine e capacità militari

II riconoscimento del dominio operativo cibernetico da parte della NATO sta influenzando anche lo sviluppo delle dottrine e capacità militari Alleate, nonché l'addestramento del personale da parte dei Paesi membri in modo da aumentare la difesa e resilienza su questo fronte. Si tratta di processi complessi, lunghi e difficoltosi, necessari per integrare nel modus operandi militare un dominio operativo nuovo e per molti versi diverso da quelli tradizionali e fisici.

Il *Cyberspace Operations Centre* (CyOC) è l'attore chiave al riguardo, mentre l'*Act* (NATO Allied Command Transformation) considera il dominio cibernetico nel quadro più ampio della trasformazione militare e dell'innovazione tecnologica in una prospettiva di medio-lungo periodo.
Nella situazione attuale, alcuni documenti Alleati sulla pianificazione operativa già comprendono esplicitamente la difesa cibernetica, ma resta molta strada da fare per integrare pienamente la dimensione cyber nelle operazioni ed attività NATO, nonché nello sviluppo dottrinale e capacitivo su cui l'ultima parola resta agli stati membri[13].

Gli stati, a loro volta, utilizzano la piattaforma del *Cyber Defence Pledge* per valutare autonomamente nel tempo i progressi sullo sviluppo delle capacità nazionali di difesa cibernetica, anche attraverso il rapporto finale sull'attuazione degli impegni presi e per scambiarsi informazioni e buone prassi a riguardo.

Un ruolo importante è giocato ovviamente anche dal *NATO Defence Planning Process* (Ndpp) che è la procedura principale, con cui i Paesi membri concordano gli obiettivi nazionali di sviluppo delle rispettive Forze Armate in modo da contribuire anche agli impegni NATO di difesa collettiva e gestione delle crisi. Nel quadro del Ndpp, dal 2012 sono stati inseriti degli obiettivi di sviluppo di capacità di cyber defence, i cui progressi vengono valutati periodicamente.

1.2.3 Le relazioni tra i paesi membri

Negli ultimi anni si sono registrati progressi nella cooperazione tra gli Stati membri più attivi nel campo cibernetico. Dal 2019 Stati Uniti, Regno Unito, Francia, Danimarca ed Estonia hanno concordato una cornice NATO nella quale integrare contributi volontari in termini di operazioni difensive e offensive – contributi che restano in ogni caso sotto il pieno controllo e responsabilità del singolo Paese membro.

Paesi che, nel caso americano, britannico e francese, considerano senza soluzione di continuità le operazioni difensive e offensive nel campo cibernetico, attuando concetti come "difesa attiva" più aggressivi dell'approccio seguito finora dalla NATO. Quest'ultimo si è consolidato negli ultimi anni e probabilmente si evolverà in linea con le caratteristiche intrinseche di un'alleanza politico-militare di natura difensiva.

Allo stato attuale la difesa cibernetica è inserita a supporto dei comandi operativi terrestre, navale e aereo, ma è aperta la possibilità della creazione futura di un comando NATO per le operazioni cyber, a valle di un processo di formulazione di dottrine e sviluppo capacitivo ancora nella fase iniziale. Ciononostante, l'Ncsc (*NATO Cyber Security Centre)* e l'Ncirc *(NATO Computer Incident Response Centre)* forniscono supporto nel monitoraggio costante e nella risposta in caso di attacco cibernetico, mettendo a disposizione dei Paesi membri anche i C*yber Rapid Reaction Teams.*

La difesa cibernetica a livello Alleato non si limita alla creazione delle strutture di comando e all'impiego di personale dedicato, ma comprende anche delle partnership con attori diversi. La necessità di dotare l'Alleanza atlantica di equipaggiamenti all'avanguardia tecnologica ha portato, già nel 2014, alla costituzione di specifiche partnership con le industrie del settore cibernetico. La cooperazione NATO-UE ha inserito già nel 2016 la dimensione cyber tra le aree prioritarie in cui collaborare.

Sono stati individuati differenti approcci verso la difesa cibernetica a riprova di quanto ci sia ancora da fare nella definizione di dottrine e procedure condivise. Tra gli stati alleati vi è una sostanziale divisione tra i Paesi che prevedono la possibilità di effettuare solo azioni difensive e quelli che invece puntano sulla possibilità e capacità di portare a termine operazioni offensive anche in assenza di un attacco cyber.[13]

Tra i primi vi sono la **Germania** e la **Spagna**, che intendono la deterrenza cibernetica come la capacità dello stato di rispondere tempestivamente e adeguatamente ad un

attacco cyber, attuando quello che viene definito come *hack-back*. È importante sottolineare, tuttavia, come questa procedura mal si adatti alla velocità di reazione necessaria per limitare o evitare i danni di un attacco cibernetico: le tempistiche parlamentari, nei casi che richiedono azioni rapide e mirate, potrebbero causare l'incapacità dello stato di proteggere i suoi interessi primari e ostacolare di conseguenza la difesa nazionale.

Londra, Parigi e Washington hanno invece una diversa comprensione delle possibilità derivanti dall'uso attivo della cyber defence. Per questi ultimi tre Stati, difesa e deterrenza cyber equivalgono ad assicurare non solo la capacità di reazione in caso di attacco cibernetico, ma anche la possibilità di azione preventiva ai danni di potenziali avversari, siano essi statali o meno.

Nonostante queste differenze, è possibile delineare delle esigenze condivise che possono essere schematizzate come segue:

- Necessità di avere un quadro regolamentare e dottrinale condiviso a livello NATO, UE e internazionale;
- Migliore integrazione della componente cyber nelle strutture di comando nazionali e Alleate;
- Collaborazione più strutturata e strategica con imprese e mondo della ricerca;
- Più elevati investimenti nell'aggiornamento delle capacità cibernetiche;
- Formazione specialistica del personale militare volto alla protezione dalla minaccia cyber;
- Maggiore sensibilizzazione dei funzionari statali, dei gestori delle infrastrutture critiche e in generale della popolazione nell'utilizzo del cyberspace.

1.2.4 Un approccio *"comprehensive"* alla cyber defence

"Comprehensive Deterrence seeks to expand upon traditional concepts of deterrence to account for the totality and the variety of the threats we face in the early 21st Century security environment."[11]

Il 14 giugno 2021 è stata approvata la nuova *Cyber Defence Policy* della NATO. In questo lungo comunicato consultabile online[36], nel **Paragrafo 32** la difesa informatica è stata definita come "*Comprehensive*", ed è stata presentata come una necessità vista l'escalation di ransomware[2], killware[3] e altri attacchi che hanno preso di mira infrastrutture critiche e istituzioni democratiche. Il Comunicato infatti spiega che le minacce informatiche alla

[2]Un ransomware è un tipo di malware che limita l'accesso del dispositivo che infetta, richiedendo un riscatto da pagare per rimuovere la limitazione.

[3]Un virus online può uccidere, letteralmente, una persona umana.

sicurezza dell'Alleanza sono diventate sempre più complesse, distruttive, coercitive e frequenti.

Lo scopo dell'Alleanza è di promuovere un cyberspazio libero, aperto, pacifico, sicuro e di perseguire ulteriormente gli sforzi per migliorare la stabilità e ridurre il rischio di conflitto sostenendo il diritto internazionale e le norme volontarie di comportamento responsabile nel quinto dominio. La NATO si propone come piattaforma per la consultazione politica tra gli Alleati, al fine di condividere le preoccupazioni sulle attività informatiche dannose e scambiando approcci e risposte nazionali, oltre a considerare possibili risposte collettive.

Il punto focale di questo nuovo approccio è che la risposta alle minacce che arrivano dal cyber spazio non sarà limitata al dominio cibernetico. Verrà migliorata la consapevolezza della situazione per supportare il processo decisionale della NATO.

Il termine *"comprehensive"* suggerisce di costruire una grande strategia per la deterrenza, basata su una visuale più ampia di solo conflitti di alto livello da parte degli stati nazionali, per contrastare le sfide delle varie "zone grigie" che si creano per colpa della cornice legislativa ancora troppo poco chiara, riconoscere la concorrenza tra gli Stati e la necessità di nuovi modi di pensare ai fini della deterrenza.

Negli ambienti accademici, il concetto di *comprehensive cyber security* è di lunga data e fornisce alcuni informazioni su come potrebbe essere stabilita una strategia globale di deterrenza informatica, ed è sempre stato visto come un concentrarsi sull'allargamento della prospettiva convenzionale della cyber sicurezza verso un ambiente non solo puramente militare, territoriale o statuale, ma che includa altri aspetti rilevanti per la sicurezza come le operazioni civili in varie aree (polizia, riforma del settore della sicurezza, stato di diritto, protezione ed amministrazione civile), sviluppo, questioni ambientali, aiuti umanitari, cooperazione strutturale e diplomazia.

La tabella riportata, **(Figura 1.1)** suggerisce un quadro provvisorio e collaborativo per la deterrenza informatica che riconosce la gamma di minacce coinvolte nel dominio cyber e la gamma di obiettivi. Propone una distinzione tra attori che devono essere dissuasi e attori dissuasivi (quelli impegnati nella formulazione e nell'attuazione di strategie di deterrenza), e mette in evidenza le varie misure che possono essere utilizzate a seconda del tipo di minaccia, del bersaglio e dell'attore coinvolto.

Le misure di deterrenza includono sia delle politiche esistenti che hanno un valore deterrente, compresa la punizione per i più alti livelli strategici, sia minacce e negazione per attività a spettro inferiore.
L'approccio proposto riconosce che gli obiettivi saranno diversificati e l'applicazione delle misure di deterrenza varierà di conseguenza. Si riconosce anche che le misure di dissuasione saranno adattate e scelte in base al tipo di attore malevolo, le organizzazioni coinvolte nell'attività di deterrenza e le loro capacità e autorità.

Deterring Actors	Deterrence Measures	Threats	Targets	Deterrable Actors
International Organisations Nation states and government agencies Police, Judiciary Private sector, ISPs	*Punishment* Kinetic retaliation, cyber retaliation, legal prosecution, economic sanctions, diplomatic isolation. *Denial* Norms, Societal denial, Technical denial *Resilience* Recovery, contingency, and continuity planning	(absolute deterrence) Cyber warfare Cyber espionage Cyber terrorism Cyber crime (restrictive deterrence)	Military Government, political institutions, values Corporate Critical infrastructure Individuals, public	Nation states, military and security agencies State affiliated hackers Terrorist groups Insider threats Criminals and organised crime

Figura 1.1: Framework della cyber deterrence *comprehensive*.

1.3 Implicazioni politiche e pratiche

Nell'ultimo quinquennio si è assistito ad una presa di coscienza da parte di organismi nazionali e internazionali in ambito cyber. Essa è finalmente diventata una delle priorità nelle agende politiche istituzionali, incrementando, oltre l'attività normativa stessa, anche le attività di studio, riflessione dottrinale ed applicativa.

Questo processo di ampliamento dell'attività di regolamentazione, oltre ad essere incoraggiato dallo sviluppo di nuove tecnologie, quali ad esempio le armi autonome, è stato incoraggiato dall'ultilizzo massiccio dell'ambito cyber in contesti in cui fino a quel momento vi era l'appannaggio esclusivo dell'attività umana (quali le attività di Comando e Controllo per i conflitti armati), rafforzando l'idea che nel prossimo futuro quest'ambito sarà uno dei principali problemi dell'umanità stessa – e come ogni attività umana- i rischi che questa comporta.

Solo negli ultimi anni la legislazione italiana ha seguito il trend internazionale in materia di cybersicurezza, innalzando il livello di attenzione e adattandosi agli standard proposti in quanto membri dell'Unione Europea e dalla comunità internazionale.
È necessario infatti osservare come questo sia stato del tutto assente nei programmi elettorali del 2018 e come nella stessa campagna elettorale si sia potuto assistere ad uno dei

più clamorosi "incidenti" cyber, in cui un hacker ha violato la piattaforma di voto diretto per rendere pubbliche le vulnerabilità del sistema e gli effetti di un possibile data breach.

La prima chiave di volta sicuramente riguarda il GDPR, Il Regolamento Generale sulla Protezione dei Dati, un regolamento europeo che disciplina il modo in cui le aziende e le altre organizzazioni trattano i dati personali. È il provvedimento più significativo degli ultimi 20 anni in materia di protezione dei dati e ha implicazioni importanti per qualsiasi organizzazione al mondo che si rivolga ai cittadini dell'Unione Europea.

Ogni organizzazione deve documentare e monitorare le attività di trattamento dei dati personali, In quanto titolare del trattamento, ogni organizzazione deve registrare e monitorare le attività di trattamento dei dati personali. Ciò include i dati personali trattati non soltanto all'interno dell'organizzazione, ma anche da terzi - i cosiddetti responsabili del trattamento.

Un passo ulteriore è stato compiuto nel 2017 durante il G7 dedicato all'industria in cui vi è una responsabilità condivisa tra gli stati affinché sia possibile creare uno spazio cibernetico che sia "aperto, accessibile e sicuro".
Il G7 successivo, rivolto allo stesso ambito applicativo, si è incentrato sulla realizzazione di procedure di risk management per rafforzare la resilienza dello spazio cibernetico in cui operano le imprese.

Come detto in precedenza, l'input è arrivato dall'Unione Europea che ha emanato una serie di direttive e regolamentazioni in materia di cybersicurezza. Per importanza spicca la Direttiva NIS (*Network and Information Security, 2016/1148)*[15] convertita in legge con Decreto Legislativo 18 maggio 2018, n.65, pubblicato sulla Gazzetta Ufficiale n.132 del 9 giugno 2018.

Nella sua forma, la direttiva NIS è stata adottata al fine di implementare la resilienza informatica dell'Italia, e non crea obblighi in capo agli Stati in merito alla creazione di procedure obbligatorie ma indica quali sono gli obiettivi che si intende raggiungere lasciando libertà in merito ai mezzi da poter utilizzare.

Ciò malgrado, nonostante i grandi passi avanti della politica italiana e della politica internazionale, i recenti casi di attacchi alla Pubblica Amministrazione dimostrano come risulta ancora chiaro che sia necessario implementare gli sforzi affinché si possa pensare di rimanere al passo in ambito cyber, molto più veloce di quello dei legislatori.

Per tracciare un sintetico quadro normativo di riferimento in materia cyber che voglia essere metodologicamente corretto, è necessario precisare che, allo stato attuale, non si può propriamente dire che esista un corpus normativo organico.
Ciò è anche abbastanza comprensibile se si pone in evidenza il fatto che il dominio cibernetico, non avendo i classici "confini" e/o consistenza materiale, crea non pochi problemi per poter far legittimamente ed automaticamente ricorso agli ordinari strumenti definitori

e regolamentari del diritto classico e tangibile.

La Direttiva NIS costituisce la pietra miliare per la costruzione di un sistema di sicurezza europeo costituita da regole di coordinamento tra Stati molto precise. L'obiettivo principale è il rafforzamento della sicurezza e della resilienza informatica all'interno dell'Unione, imponendo un livello minimo di sicurezza per le tecnologie, le reti ed i servizi digitali. L'applicazione della Direttiva è rivolta sia agli operatori dei servizi essenziali che ai fornitori dei servizi digitali.

1.3.1 Architettura cyber nazionale

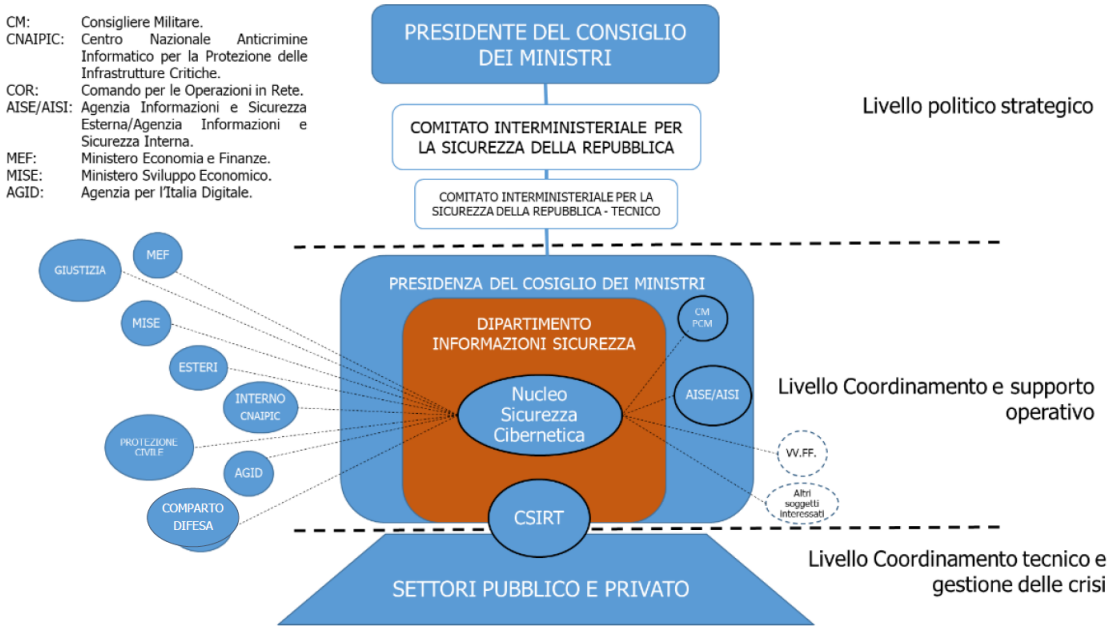

Figura 1.2: Architettura cyber nazionale (DPCM 17-02-2017).

- **indirizzo politico-strategico**, composto dal Presidente del Consiglio dei Ministri e dal Comitato Interministeriale per la Sicurezza della Repubblica (CISR). Quest'ultimo è un organo collegiale di consulenza, proposta e deliberazione, di cui fanno parte i Ministri di: Affari Esteri, Interno, Difesa, Economia e Finanze, Giustizia e Sviluppo Economico. Inoltre, il CISR, per l'espletamento delle proprie funzioni, è supportato da un ulteriore organo collegiale di coordinamento, presieduto dal Direttore del Dipartimento Informazioni per la Sicurezza (DIS), denominato CISR-Tecnico;
- **coordinamento e supporto operativo**, che consiste nelle seguenti articolazioni riconducibili al DIS:

  - Computer Security Incident Response Team (CSIRT) – Italia, che consiste nel gruppo di intervento per la sicurezza informatica in caso di incidente;
  - Nucleo per la Sicurezza Cibernetica (NSC), organo collegiale presieduto dal Vice Direttore Generale del DIS con delega per il settore cyber, che ha il compito di coordinare e gestire le crisi cibernetiche nazionali.

- **coordinamento tecnico e gestione delle crisi**, composto dai CERT dicasteriali, ivi compreso quello della Difesa, inquadrato nell'ambito del COR. In particolare, i CERT dei Ministeri inclusi nel Decreto che ha recepito la Direttiva NIS si interfacciano, da un lato, con il CSIRT – Italia e dall'altro con i rispettivi Operatori di Servizi Essenziali (OSE)/Fornitori di Servizi Digitali (FSD) identificati dal predetto Decreto.

1.4 Regole di ingaggio (ROE)

Sul piano giuridico, le ROE sono atti amministrativi che legittimano ed al tempo stesso limitano l'uso della forza, lasciando impregiudicato il diritto di autodifesa. La loro ampiezza, che può leggersi come modulazione dell'uso della forza, dipende dal titolo giustificativo della missione.

In quanto atti amministrativi ad alta caratterizzazione politica le ROE sono insindacabili, nei limiti dei principi costituzionali in materia di tutela giurisdizionale di cui all'**art. 113**[4] della Costituzione.

Secondo la classificazione NATO, a carattere generale, le operazioni militari condotte dall'Alleanza sono inquadrabili in due principali categorie: le *"Article 5 operations"* implicanti il ricorso alla difesa collettiva e connotate da prerogative prettamente militari e le *"Non-Article 5 crisis response operations"*. Note anche come operazioni a supporto della pace, queste ultime possono includere funzioni di prevenzione della violenza conflittuale ma anche di peacekeeping, peacemaking e peacebuilding.

Data l'evidente divergenza negli obiettivi dei mandati, il quadro giuridico applicabile è profondamente differente e – di conseguenza – lo è anche la configurazione delle regole di ingaggio caratterizzanti i dispiegamenti. Per questo motivo, il processo di formazione, condivisione e applicazione di tali norme operative, complesso e articolato, fornisce una panoramica interessante circa l'operato dell'Alleanza.

[4] *"Contro gli atti della pubblica amministrazione è sempre ammessa la tutela giurisdizionale dei diritti e degli interessi legittimi dinanzi agli organi di giurisdizione ordinaria o amministrativa Tale tutela giurisdizionale non può essere esclusa o limitata a particolari mezzi di impugnazione o per determinate categorie di atti. La legge determina quali organi di giurisdizione possono annullare gli atti della pubblica amministrazione nei casi e con gli effetti previsti dalla legge stessa"*, **Art. 113** Cost.

Il cyber spazio aggiunge un nuovo livello di complessità ai concetti tradizionali di autodifesa. La risposta a questa nuova complessità può prevedere fattori multipli.
Servono quindi chiare e precise regole di ingaggio che attualmente in Italia non esistono per questo dominio della conflittualità, regole che aiuterebbero moltissimo sui tavoli nazionali come il CISR o il DIS (componente politico-militare, **Figura 1.1**).
L'impiego delle azioni cibernetiche nell'ambito delle missioni all'estero dovrà essere re-

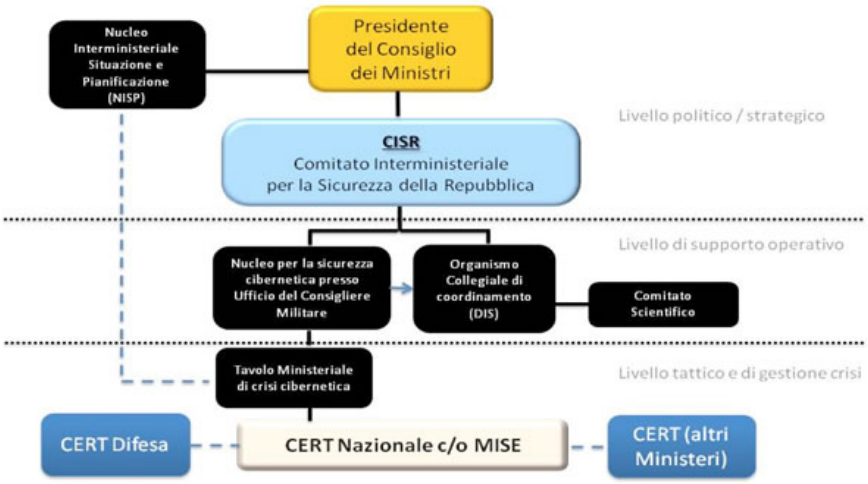

Figura 1.3: Schema componente politico-militare.

golato, alla stregua delle altre azioni militari tradizionali, da direttive, ordini e ROE che consentano di modulare l'uso della forza che ne discende.

Appare oramai assodato che le azioni cibernetiche militari producano effetti finali che possono essere sia di natura fisica, sia di natura cibernetica. Tuttavia, nella parte speciale del catalogo nazionale delle ROE non si ritrova una serie specificamente dedicata a questo tipo di azioni: esistono solo direttive che stabiliscono le operazioni dell'informazione.

Le operazioni dell'informazione sono infatti concettualmente diverse dalle operazioni cibernetiche militari, come evidenziato esplicitamente nella JIC 012[35]. Questa diversità risiede nello scopo che le caratterizza:

- le operazioni dell'informazione hanno lo scopo di influenzare i comportamenti dell'avversario;

- le azioni cibernetiche militari hanno invece lo scopo di alterare e/o manipolare, temporaneamente o permanentemente, database o sistemi informativi ivi compresa la loro distruzione.

Questa marcata distinzione suggerisce di prevedere nella parte speciale del catalogo nazionale delle ROE una nuova serie che stabilisca direttive dedicate all'impiego delle diverse tipologie di azioni cibernetiche militari nelle operazioni, con specifico riferimento alle azioni di tipo offensivo.

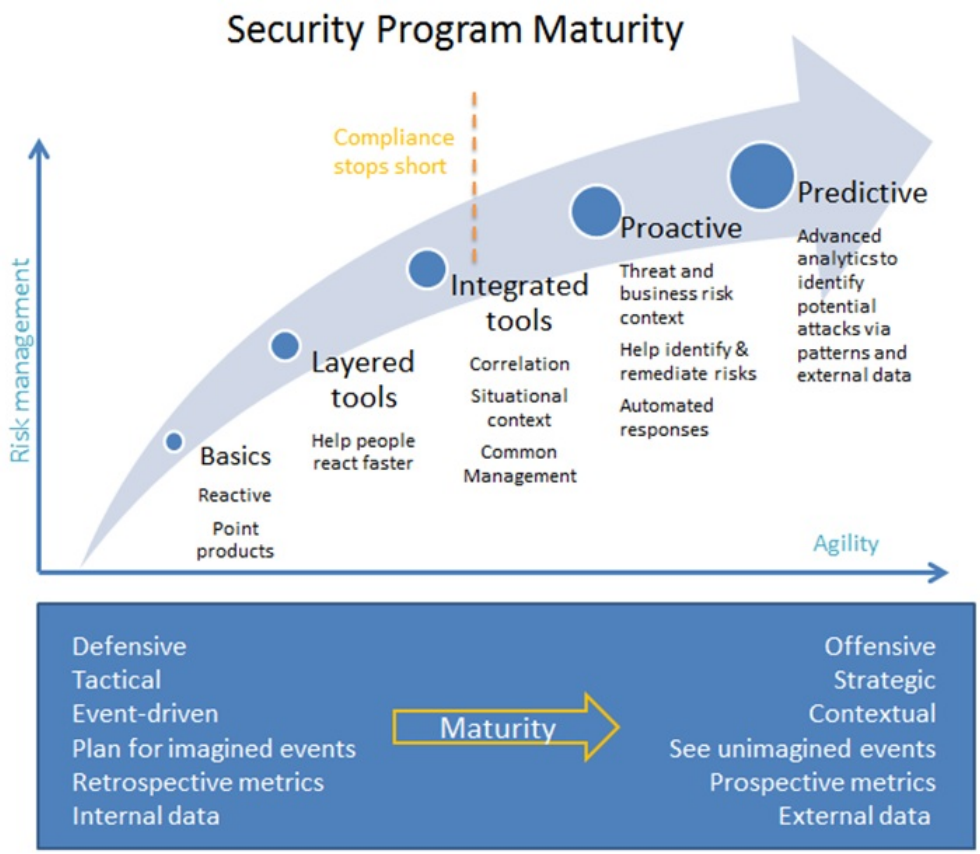

Figura 1.4: Security maturity model.

Mentre la NATO ha avuto decenni di esperienza operativa nella formulazione di regole di ingaggio per armi cinetiche, diverse caratteristiche delle operazioni nel dominio cibernetico aumentano la loro complessità e risultano per ora informazioni classificate.

Le questioni legate ai meccanismi di Comando e Controllo (C2) e all'escalation dell'uso di forza cibernetica anche in risposta ad attacchi cinetici (e viceversa) svolgono un ruolo importante nella definizione di ROE specifiche.

La formulazione di ROE per le armi informatiche è quindi possibile con sforzi particolari per impartire tale esperienza ai leader politici e ai comandi militari in contesti di attivazione di unità di crisi come nell'esempio italiano.

Il modello americano mutuato in parte su quello europeo NATO prevede tre tipologie di ROE:

- **Regole di ingaggio permanenti (SROE)**, forniscono un insieme generale di regole sempre operative relative all'autodifesa per le forze in tempo di pace, nonché un modello per ROE specifiche dell'operazione.
  Le SROE cibernetiche si basano su una minaccia immediata in linea con capacità predittive e proattive (verifica di un atto ostile ed esistenza di un intento ostile) e la necessità di rispondere adeguatamente e secondo principi di proporzionalità a tale minaccia.

- **ROE supplementari** (note anche come ROE specifiche della missione) sono personalizzate per una regione, una missione o un'operazione specifica e possono elaborare e/o interpretare le SROE per una determinata missione senza contraddirle.

- **Regole permanenti per l'uso della forza (SRUF)**, regolano le azioni militari all'interno della nazione.

Molte delle effettive capacità militari nel dominio del cyberspazio e le ROE corrispondenti al loro uso rimangono per ora classificate. Attraverso analisi di OSINT *(Open Source Intelligence)* è in ogni caso disponibile un corpus crescente di informazioni non classificate riguardanti i principi e i concetti che guidano il modo in cui i pianificatori del DOD americano formulano ROE per operazioni nel cyberspace che potrebbero essere mutuate in Europa.

Nello sviluppo di ROE ad hoc occorrerà distinguere inoltre se la forza è impiegata per legittima difesa ovvero per raggiungere gli obiettivi militari prefissati.
Il Diritto internazionale stabilisce i principi che governano i limiti legali all'uso della forza durante le operazioni militari, limiti che, in determinate operazioni o situazioni, possono ulteriormente essere ampliati dalla legislazione nazionale.

1.4.1 Nuovi livelli di complessità

Tuttavia, un'azione che si attua attraverso il cyberspace potrebbe non rappresentare un'ovvia minaccia per la vita umana o per le sue capacità critiche.
Raramente un operatore cibernetico dovrà affrontare una decisione di vita o di morte personale simile a quella di un soldato impegnato in un combattimento cinetico su un campo di battaglia fisico, il che significa che il cyberspace aggiunge un nuovo livello di complessità ai concetti tradizionali di autodifesa. La risposta a questa nuova complessità può prevedere fattori multipli.

Le unità militari sono generalmente autorizzate ad adottare misure di difesa passiva all'interno dei sistemi o delle reti di cui sono responsabili. Gli esempi includono l'uso di sistemi di rilevamento delle intrusioni (IDS – software che monitora la rete o il sistema per attività dannose), l'utilizzo di firewall e l'eliminazione di connessioni che potrebbero essere utilizzato per scopi ostili o per eliminare malware trovati sui sistemi da difendere.

1.5 Protezione e deterrenza in EI e NATO

Da un punto di vista tecnologico, l'efficacia della cyber-deterrenza dipende dalla capacità di identificare la fonte dell'attacco informatico e dalla capacità di destreggiarsi tra attacco e difesa in ambito informatico. In altre parole, richiede la capacità di scoprire gli intrusi e di bloccare o reagire contro le loro intrusioni.

Quando si tratta del paradigma attacco-difesa nel cyberspazio, gli attacchi informatici a livello strategico si basano sulla padronanza delle vulnerabilità del sistema "vittima", nonché sulla penetrazione a lungo termine e su una pianificazione sofisticata. La difesa informatica dipende dal rilevamento delle minacce alla sicurezza, dalla risposta tempestiva e dal rapido ripristino dopo un attacco.

Quando si coordinano **attacchi** complessi, il fattore umano è di maggiore importanza, poiché spesso diventa il punto più debole del sistema e serve nella pianificazione multi-dimensionale. Quando si coordinano operazioni di **difesa**, la consapevolezza situazionale dell'Intelligenza Artificiale (AI), il calcolo estrememente rapido e le capacità di elaborazione dei dati sono più efficienti di quelle manuali.
Pertanto, l'intelligenza artificiale è adatta a svolgere un ruolo maggiore nella difesa delle reti.

A tal fine, con l'assistenza della tecnologia AI, che facilita la raccolta e l'elaborazione di grandi quantità di dati storici sugli attacchi informatici, i difensori sono in grado di anticipare al meglio i mezzi e le regole di attacco degli aggressori. L'Intelligenza Artificiale permette di:

- migliorare la capacità del difensore di identificare la superficie e le fonti di attacco,
- ridurre l'anonimato dell'aggressore,
- scoprire le attività dell'aggressore,
- fornire warning,
- offrire deterrenza economicamente vantaggiosa.

Inoltre, attraverso la formazione sulla percezione e sulla risposta alle minacce informatiche, la tecnologia AI può rilevare e bloccare i dispositivi che sono stati attaccati e impedire l'installazione e il funzionamento di malware e file malevoli. In tal modo, può migliorare l'efficienza operativa dei centri di operazioni di sicurezza, quantificare i rischi per la sicurezza della rete, monitorare le anomalie del traffico di rete, migliorare le capacità di difesa informatica e cyber-recovery e migliorare la deterrenza in generale.

Quando una singola potenza cibernetica ha questo tipo di capacità, otterrà ulteriormente la consapevolezza della situazione (**situational awareness**) e vantaggi di difesa-offesa nel ciberspazio, con conseguente bassa stabilità strategica. Quando un rivale della grande potenza strategica ottiene capacità simili, sarà probabilmente in grado di impegnarsi nella

deterrenza reciproca e quindi migliorare la sua stabilità strategica all'interno del cyberspazio.

È fondamentale porre un accento particolare sull'applicazione della tecnologia dell'Intelligenza Artificiale nelle capacità di rilevamento, monitoraggio e ripristino dagli attacchi informatici, migliorando nel contempo la difesa informatica e la deterrenza informatica in generale. I paesi che hanno interessi comuni nell'evitare gli attacchi informatici possono utilizzare l'AI per migliorare la trasparenza del cyberspazio e mitigare gli attacchi anonimi. Tali misure potrebbero includere il perseguimento del consenso sul controllo delle tecnologie, su misure quali azioni reciproche e impegni unilaterali a non impegnarsi in un primo attacco o attacchi alle reciproche infrastrutture critiche.

Ciò promuoverebbe la formazione di un sistema di controllo delle "armi" nel cyberspazio. Cina e Russia hanno infatti firmato un accordo sulla salvaguardia della sicurezza informatica internazionale, che include l'impegno a non intraprendere attacchi informatici l'uno contro l'altro. Nel 2015 un accordo simile è stato firmato anche dagli Stati Uniti nei confronti della Cina.

Capitolo 2

Intelligenza Artificiale e Cyber Defence per Applicazioni Militari

Il *rinascimento* dell'Intelligenza Artificiale nell'ultima decade ha portato a innovazioni e sviluppi rivoluzionari, come dimostrano diverse applicazioni, a partire dagli assistenti domotici fino ad arrivare alle automobili a guida autonoma. Molti problemi alla base delle applicazioni sopra citate sono potuti essere risolti grazie a metodi di machine learning, branca dell'AI.
Come nell'ambito civile, anche nel mondo militare vengano applicate tecnologie di machine learning per lo sviluppo di sistemi, applicazioni, sistemi di guida o armi autonome. Tutti questi esempi sono applicazioni che esistono già da tempo, ma con la rivoluzione dell'Intelligenza Artificiale verranno automatizzati sempre di più.

Tuttavia, per sistemi safety-critical come le armi, i requisiti per il processo di sviluppo e la certificazione sono diversi rispetto ai sistemi le cui funzioni sono meno critiche se si verifica un esito imprevisto. Pertanto, non è ancora perfettamente chiaro quali metodi di apprendimento automatico possano essere utilizzati nelle applicazioni militari.

Un sistema autonomo è un sistema che svolge incarichi senza interferenza umana. L'utilizzo del termine "autonomo" rispetto al termine "automatico" nasce dal fatto che un sistema di questo tipo è in grado di svolgere compiti complessi in ambienti mission-critical. Per riuscire a portare a termine le sue mansioni in questo tipo di ambiente, il sistema deve essere in grado di acquisire informazioni riguardo i suoi dintorni e la sua relazione con ciò che lo circonda attraverso opportuni sensori. I veicoli aerei, ad esempio, hanno bisogno di sensori che misurano condizioni fisiche, come la pressione dell'aria, l'accelerazione e campi magnetici, che venogno utilizzati per calcolare la velocità, l'altitudine, la direzione e altre variabili che descrivono il rapporto del sistema con l'ambiente circostante. Il sistema quindi calcola un modello matematico dell'ambiente, che descrive le interazioni del sistema con ciò che lo circonda.

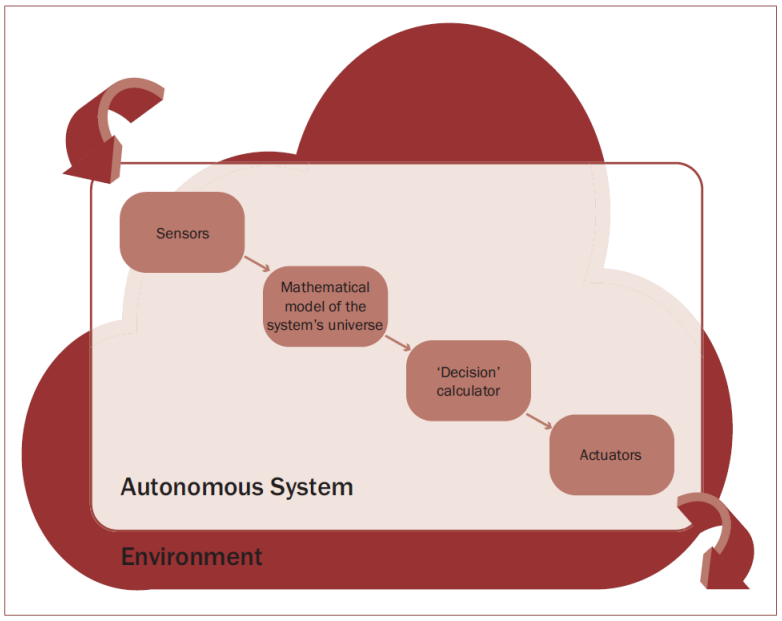

Figura 2.1: Descrizione schematica di un sistema autonomo.

Molti stati, tra i quali gli Stati Uniti in testa, hanno identificato la chiave per le future capacità militari nella crescita dell'automazione. Per questo motivo ci si può aspettare che nel prossimo futuro molti sforzi e molte risorse verranno impiegate nella ricerca e nello sviluppo dell'automazione per applicazioni a carattere militare. Sebbene l'automazione sia stata utilizzata nei sistemi d'arma per un secolo, rimangono molte sfide per l'applicazione generica e ad ampio raggio dell'autonomia delle armi.

In un conflitto armato la decisione di un comandante di iniziare un'azione offensiva deve seguire i principi fondamentali del diritto internazionale umanitario. Per poter attaccare in modo discriminatorio e con le dovute precauzioni, il comportamento e gli effetti delle armi devono essere prevedibili. Un comandante deve comprendere gli effetti di un'arma ed essere in grado di prevedere come si comporterà l'arma una volta lanciata.

Pertanto, da un punto di vista operativo, la sfida principale della crescente autonomia dei sistemi militari e degli armamenti è la necessità di prevedibilità e di comprendere il comportamento del sistema.

2.1 Analisi dei principali impieghi

Machine learning è un nome collettivo spesso usato per riferirsi a metodi statistici per identificare strutture in un insieme di dati. Queste tecniche hanno molte diverse possibili

applicazioni in ambito militare.

Inoltre, il machine learning, è stato usato per risolvere alcuni problemi in differenti campi dell'AI. Due esempi dove questi metodi sono stati applicati con successo sono l'image e la speech recognition, e in generale queste tecniche sono adatte per applicazioni ricche di dati, dove la creazione di un modello esplicito del sistema risulta difficile.

Ogni tipo di sistema infatti necessita lo sviluppo di un modello del proprio "universo". Per le applicazioni in cui lo spazio e l'ambiente sono facilmente capibili e che ha una descrizione matematica, come ad esempio la descrizione delle forze aerodinamiche di un aereo, le tecniche di machine learning si sono dimostrate meno utili della modellazione esplicita (ad esempio con equazioni relative ad azioni e reazioni tra il sistema e l'ambiente).

Nelle applicazioni in cui non esiste un modello così conciso, ma solo un ampio insieme di dati che descrivono implicitamente il carattere dell'ambiente in cui il sistema si trova, il machine learning è adatto per derivare un modello di tale sistema. Un esempio di tale applicazione, in cui i metodi in oggetto sono comunemente applicati per la ricerca e lo sviluppo, è il **target recognition**, che è, in in termini più generali, un problema di image recognition.

Altre applicazioni militari per cui è adatto l'uso dell'Intelligenza Artificiale includono quanto segue.

- **Anomaly detection**
  I metodi di machine learning possono essere utilizzati per il pattern recognition. Questi metodi possono essere utilizzati per identificare pattern di "normalità" nei dati e quindi per rilevare data pattern che differiscono dallo stato normale.
- **Systems for information management in reconnaissance and surveillance applications**
  Gli odierni sistemi di ricognizione e sorveglianza (ISR) raccolgono grandi quantità di informazioni. Un UAV (Unmanned Aerial Vehicle) dotato di sensori di imaging può rimanere in volo per lunghi periodi, inviando continuamente un flusso di dati attraverso un network ad un centro di analisi. Gli analisti quindi valutano i dati ed estraggono informazioni significative. L'analisi dei dati può essere un'applicazione in cui i metodi di apprendimento automatico possono rivelarsi utili.
- **Decision-support systems**
  I sistemi di supporto alle decisioni di Comando e Controllo sono utilizzati in una varietà di applicazioni diverse come sistemi di diagnosi medica, produzione e marketing per aiutare gli operatori a prendere decisioni analizzando i dati e proponendo differenti corsi d'azione.
- **Warning systems**
  La componente terrestre del sistema di early-warning di molti stati è costituita da phased-array radar. I radar di questo tipo dipendono da sofisticati algoritmi di elaborazione che discriminano tra gli oggetti ricercati (es. missili ostili) ed altri oggetti

che possono essere presenti (es. aeroplani, uccelli, ecc.). Pubblicazioni recenti [34] hanno evidenziato il potenziale degli algoritmi basati sull'apprendimento automatico per fornire migliori capacità di discriminazione nelle applicazioni radar. Se utilizzati nei sistemi di early-warning, potrebbero in linea di principio comportare un minor numero di falsi allarmi.

- **Guidance systems**
  I missili da crociera, compresi quelli che trasportano armi nucleari, si basano su sofisticate capacità di "terrain-hugging" che consentono loro di volare vicino al suolo ma di non entrare in collisione con montagne o edifici alti. Sebbene si basino su una vasta serie di tecnologie di natura diversa, di recente sono stati discussi [12] potenziali vantaggi dell'utilizzo del machine learning per facilitare la navigazione ed il targeting.

- **Cybersecurity**
  Nell'area della sicurezza informatica, il machine learning sta trovando ampio uso nel rilevamento di malware.[17]Sebbene i dettagli tecnici dei sistemi di Comando, Controllo e Comunicazione (C3) militari non siano disponibili pubblicamente, non è irragionevole aspettarsi che la sicurezza della rete utilizzi sofisticati approcci alla sicurezza informatica, compreso l'apprendimento automatico.

2.1.1 Machine learning per sistemi d'arma

I metodi di machine learning sono usati per differenti applicazioni in molti programmi di ricerca per sistemi d'arma. Tuttavia esiste un gap tra un sistema effettivamente usabile in operazione ed uno solo testato. Una qualsiasi arma o un qualsiasi apparato militare prima di essere utilizzato sul campo deve superare test, validazioni e accertamenti molto severi. Ogni arma, in generale, è stata creata con l'obbiettivo di arrecare del danno all'opponente, e quindi, se male operata, oppure se utilizzata in modo imprevisto o non intenzionale, il rischio che provochi danni accidentali è estremamente alto e non trascurabile.

Pertanto, è estremamente importante essere in grado di comprendere e prevedere come si comporterà un sistema d'arma. Ciò porta gli sviluppatori ad utilizzare processi di sviluppo conservativi con ampie procedure di test e verifica. Sebbene le tecniche di machine learning si siano dimostrate utili in molte applicazioni, non sono ancora comunemente utilizzate nei sistemi d'arma. Uno dei motivi potrebbe essere la difficoltà del processo di verifica dei sistemi "scatola nera", che è una caratteristica dei sistemi sviluppati dal machine learning.

2.1.2 Cyberwarfare

L'autonomia non è una nuova scoperta nel regno cibernetico. L'automazione è già da tempo una componente chiave di qualsiasi architettura di difesa informatica. I programmi anti malware sono progettati per identificare e neutralizzare automaticamente un qualsiasi malware noto. Le armi informatiche in genere devono operare in modo autonomo, quindi al di fuori della diretta supervisione umana, almeno durante le parti chiave della

loro missione. È stato il caso, ad esempio, del virus Stuxnet.

Tuttavia, i recenti progressi nel machine learning stanno cambiando il modo in cui questa automazione o autonomia funziona, di pari passo al cambiamento del modo in cui gli strumenti di guerra informatica sono progettati e gestiti, sia per scopi difensivi che offensivi.

- Sul lato difensivo, i metodi di machine learning hanno aperto la possibilità di individuare nuovi tipi di malware (i. e. sconosciuti) e di rilevare attività sospette in una rete.
- Sul lato offensivo, il machine learning facilita l'identificazione di zero-day-vulnerabilities nel software vittima.

2.1.3 Guerra elettronica

Il machine learning può apportare importanti miglioramenti al campo della guerra elettronica nello stesso modo della guerra informatica.

- Dal punto di vista difensivo, l'apprendimento automatico migliora le capacità antijamming in quanto apre la possibilità di automatizzare l'analisi e la difesa contro nuovi segnali nemici.
  Nel 2016 la DARPA (*US Defense Advanced Research Projects Agency*) ha lanciato una sfida pubblica per sviluppare sistemi con la capacità di identificare ed analizzare nuovi segnali nemici on the fly, cioè durante il funzionamento dei sistemi invece che dopo, come avviene attualmente.
- Dal punto di vista offensivo, l'apprendimento automatico può essere utilizzato per sviluppare nuovi strumenti di tipo jammer.

Riassumendo, il machine learning in un contesto militare ad oggi rimane un'arma a doppio taglio: può sia aumentare la protezione dell'infrastruttura di Comando e Controllo contro gli attacchi informatici sia aumentare la capacità del nemico di attacchi informatici contro tale infrastruttura

2.2 AI e cyber sicurezza

L'Intelligenza Artificiale nella sicurezza informatica presenta grandi opportunità ma, come con qualsiasi potente tecnologia dual-use a scopi generali, comporta anche grandi sfide. L'Intelligenza Artificiale può migliorare le misure di sicurezza e di difesa informatica, consentendo una maggiore **robustezza, resilienza e reattività** del sistema, ma l'Intelligenza Artificiale sotto forma di machine learning e deep learning porterà ad attacchi informatici sofisticati, consentendo attacchi più rapidi, più mirati e più distruttivi.

L'applicazione dell'AI nella sicurezza informatica pone anche problemi etici e di sicurezza. Uno di questi è che difficilmente è chiaro come attribuire le responsabilità ai sistemi di

risposta autonomi, come assicurarsi che i sistemi si comportino secondo le aspettative o quali siano i rischi per la sicurezza portati dalla crescente antropomorfizzazione dei sistemi.

Organizzazioni in tutto il mondo hanno iniziato ad utilizzare l'Intelligenza Artificiale per gestire una crescente gamma di rischi per la sicurezza informatica, sfide tecniche e limitazioni delle risorse, migliorando la robustezza, la resilienza e la risposta dei loro sistemi.
La relazione tra questi sistemi e gli operatori di sicurezza dovrebbe essere intesa come un'integrazione sinergica, in cui il valore aggiunto unico degli esseri umani e dei sistemi di Intelligenza Artificiale viene preservato e potenziato, piuttosto che come una competizione tra i due.

Le applicazioni più comuni in cui l'AI viene impiegata nella cyber sicurezza sono network security, seguita dalla sicurezza dei dati e dalla sicurezza degli endpoint. Le tre categorie principali in cui si può distinguere l'uso dell'AI nella sicurezza informatica sono: rilevamento (51%), previsione (34%) e risposta (18%).
Le forze trainanti che stanno aumentando l'uso dell'AI nella sicurezza informatica comprendono:

- **Velocità di impatto**
  In alcuni dei principali attacchi recenti, il tempo medio di impatto sulle organizzazioni è di quattro minuti. Inoltre, gli attacchi di oggi non sono solo ransomware né mirano solo a determinati sistemi o a determinate vulnerabilità; possono infatti muoversi e adattarsi in base a ciò che stanno facendo i bersagli. Questi tipi di attacchi hanno un impatto incredibilmente rapido e non ci sono molte interazioni umane che possono verificarsi nel frattempo.

- **Complessità operativa**
  Ad oggi, la proliferazione di piattaforme di cloud computing ed il fatto che tali piattaforme possano essere rese operative e fornire servizi molto rapidamente - nell'intervallo di millisecondi - significa che non esistono molte interazioni umane in questo ciclo e bisogna fare riferimento a capacità più basate sull'analisi.

- **Lacune nelle competenze nella sicurezza informatica**
  Questo problema rimane una sfida continua. Secondo Frost & Sullivan [37], c'è una carenza globale di circa un milione e mezzo di esperti di sicurezza informatica. Questo livello di scarsità spinge il settore ad automatizzare i processi a un ritmo più rapido.

L'Intelligenza Artificiale può coadiuvare con i team di sicurezza in tre modi: migliorando la **robustezza**, la **risposta** e la **resilienza** dei sistemi.

In primo luogo, l'Intelligenza Artificiale può migliorare la **robustezza** dei sistemi, cioè la capacità di un sistema di mantenere la sua configurazione stabile iniziale presunta anche quando elabora input errati, grazie al software di autodiagnostica e autoriparazione.Ciò significa che i sistemi di Intelligenza Artificiale possono essere utilizzati per migliorare i test di robustezza, delegando alle macchine il processo di verifica e validazione.

In secondo luogo, l'Intelligenza Artificiale può rafforzare la **resilienza** dei sistemi, quindi la capacità di un sistema di resistere e tollerare un attacco facilitando il rilevamento di minacce e anomalie.

In terzo luogo, l'Intelligenza Artificiale può essere utilizzata per migliorare la **risposta** del sistema, ovvero la capacità di un sistema di rispondere autonomamente agli attacchi, identificare le vulnerabilità in altre macchine e operare strategicamente decidendo quale vulnerabilità attaccare e in quale punto, lanciando contrattacchi più aggressivi.

Riuscire ad identificare quando delegare il processo decisionale e le azioni di risposta all'AI, e la necessità di una singola organizzazione di eseguire una valutazione dell'impatto del rischio sono correlati. In molti casi l'AI aiuterà, senza sostituire, il processo decisionale degli analisti della sicurezza umana e sarà integrata in processi che accelerano le azioni di risposta.

2.2.1 Rilevamento e risposta

Ogni volta che l'Intelligenza Artificiale viene applicata al rilevamento e alla risposta di incidenti informatici, la risoluzione dei problemi può essere suddivisa approssimativamente in tre parti, come mostrato nella **Figura 2.2**[1]. I dati vengono raccolti dai client environment ed elaborati da un sistema gestito da un security vendor. Il sistema di rilevamento segnala l'attività dannosa e può essere utilizzato per scegliere ed attivare un'azione di risposta.

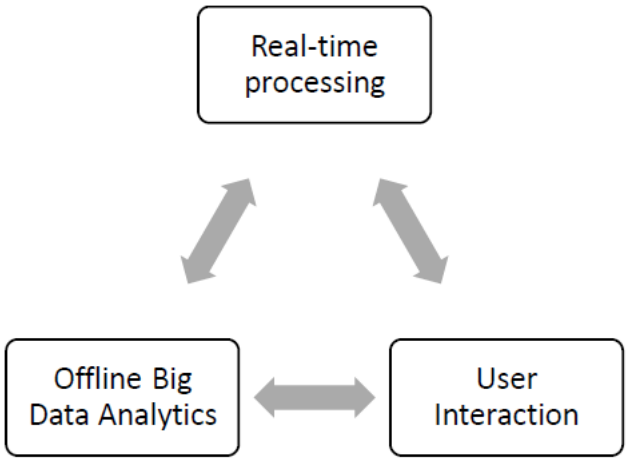

Figura 2.2: AI cyber incidents detection and response.

[1]Source: Palo Alto Network contribution to the fourth meeting of the CEPS Task Force.

Le aziende oggi riconoscono che la superficie di un attacco informatico sta crescendo in modo massiccio a causa dell'adozione dell'Internet of Things (IoT) e della diffusione dei dispositivi mobili, aggravata da un panorama di minacce diversificato e in continua evoluzione. In questo contesto, ci sono due contromisure che possono essere adottate: velocizzare i difensori e rallentare gli aggressori.

Per quanto riguarda l'accelerazione degli apparati difensivi, le aziende oggi adottano soluzioni di Intelligenza Artificiale per automatizzare il rilevamento e la risposta agli attacchi già attivi all'interno delle "barriere" dell'organizzazione. I team di sicurezza tradizionalmente dedicano molto tempo alla gestione degli avvisi, indagando se sono benigni o dannosi, segnalandoli, contenendoli e convalidando le azioni di contenimento. L'Intelligenza Artificiale può aiutare con alcune delle attività su cui i team delle operazioni di sicurezza trascorrono la maggior parte del loro tempo. Questo è anche uno degli usi principali e più comuni dell'AI in generale.

In passato, l'industria in generale si è sempre concentrata prima sull'etichettatura e sulla categorizzazione dei vari malware, ma oggi le aziende utilizzano modelli che non cercano singoli pezzi di malware; piuttosto studiano il comportamento degli aggressori. Sta quindi diventando più comune utilizzare modelli di rilevamento delle minacce, attraverso il machine learning, che siano comportamentali nella loro analisi e, a loro volta, stanno diventando durevoli e potenzialmente in grado di rilevare zero-day-vulnerabilities. L'obiettivo è identificare il più impercettibile comportamento di un attacante, con alta fedeltà e basso scarto.
Di seguito vengono riportati degli esempi pratici dei benefici dovuti all'utilizzo delle tecniche di machine learning per il rilevamento e risposta alle minacce informatiche.

- Il machine learning addestrato sulle interazioni dell'utente fornisce un modo per comprendere il contesto locale e sapere su quali dati concentrarsi; i modelli addestrati per identificare quelli con maggiori probabilità di essere dannosi migliorano l'efficienza di un sistema classificando le informazioni da elaborare in tempo reale. In questo modo, l'utilizzo del machine learning è un risparmio sui costi, ma consente anche una reazione più rapida nelle situazioni più critiche.

- Il machine learning può essere utile per rilevare nuove anomalie apprendendo modelli robusti dai dati con cui sono stati alimentati. Il machine learning è particolarmente efficace nell'identificare modelli ed estrarre algoritmi in grandi insiemi di dati in cui la componente umana rischia di confondersi.

- Il machine learning addestrato su comportamenti di "tattiche, tecniche e procedure" (TTP) degli aggressori può supportare un rilevamento duraturo e ampio degli aggressori.

Ai fini di una più chiara illustrazione dell'uso di AI e machine learning per il rilevamento e per la risposta di minacce informatiche, la **Figura 2.3** presenta un sistema di rilevamento e prevenzione delle intrusioni che combina dispositivi software e hardware all'interno della rete. Il sistema è in grado di rilevare possibili intrusioni e tentare di prevenirle. I sistemi di

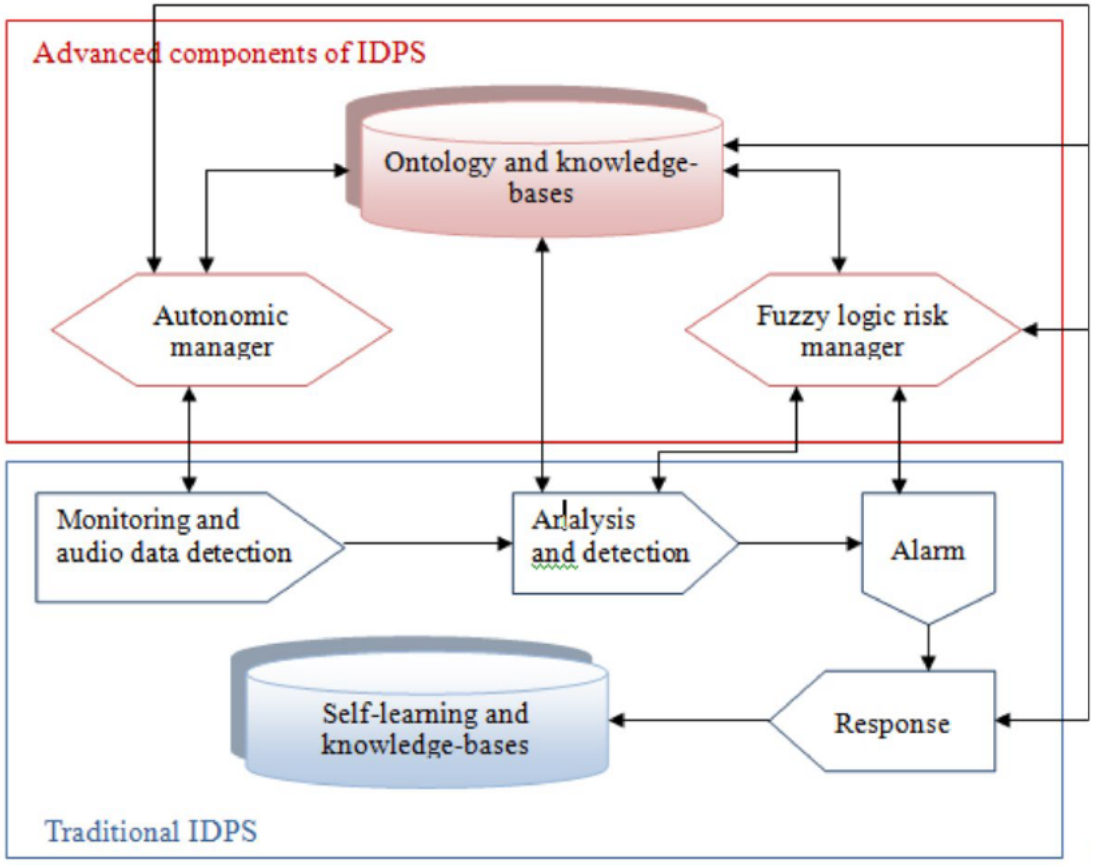

Figura 2.3: Intrusion Detection and Prevention System.

rilevamento e prevenzione delle intrusioni forniscono quattro funzioni di sicurezza vitali: **monitoraggio**, **rilevamento**, **analisi** e **risposta ad attività non autorizzate**.

2.2.2 L'impatto del machine learning...

- I nuovi algoritmi che utilizzano il machine learning sono più adattivi, offrendo un maggiore dinamismo. Poiché i rischi per la sicurezza informatica si evolvono rapidamente nel tempo, le nuove generazioni di malware e di attacchi informatici sono difficili da rilevare con i tradizionali protocolli di sicurezza informatica.
  Il machine learning supera questa debolezza consentendo ai sistemi di sicurezza informatica di utilizzare i dati preesistenti degli attacchi passati per rispondere ad attacchi simili.

- Il machine learning riduce la necessità di lavoro umano nelle interazioni di sicurezza informatica, sia in termini di offesa che di difesa. Un tipico esempio di ciò sarebbe lo spear-phishing, che induce con l'inganno un individuo o un'organizzazione specifici a divulgare informazioni riservate.

I metodi tradizionali di spear-phishing sono spesso di portata limitata. Un'intrusione efficace richiede una grande quantità di ricerca sul potenziale bersaglio. Inoltre, è difficile, se non impossibile, attaccare più bersagli contemporaneamente.
Tuttavia, con l'aiuto del machine learning, l'automazione dello spear-phishing potrebbe essere possibile.

- La terza area in cui il machine learning potrebbe fare la differenza è l'attribuzione. Con capacità di apprendimento più potenti, questi algoritmi sono in grado di individuare meglio le prove critiche per rivelare la vera identità di un utente malintenzionato, ad esempio se alcuni frammenti di codice imitano le strutture di malware esistenti.

2.2.3 ...per una cyber-deterrenza efficace

Nonostante il suo ampio utilizzo, il concetto di **cyber-deterrenza** rimane poco chiaro e oggetto di dibattito.

Ancora non esiste un'unica opinione sul fatto che gli attacchi informatici possano essere efficacemente scoraggiati e anche sul fatto che la nozione di deterrenza sia effettivamente significativa nel cyberspazio. Tuttavia, secondo chi è favorevole alla cyber-deterrenza, questa è fattibile quando si verificano determinate condizioni. Ad esempio, la deterrenza informatica può funzionare quando rende il costo dell'attacco informatico eccezionalmente alto. Ciò può essere ottenuto sia potenziando la sicurezza informatica del sistema vittima, consentendo così la **"deterrence by denial"**, sia rendendo credibile e potente la ritorsione, consentendo così la **"deterrence by punishment"**.

Un'altra condizione è che, a differenza della tradizionale nozione di deterrenza, quella in campo cyber non è per forza assoluta. Ciò significa che alcuni tipi di attori e alcuni tipi di azioni non possono essere scoraggiati; è più probabile che la deterrenza funzioni quando si tenta di scoraggiare attacchi informatici che avrebbero gravi conseguenze e scopi strategici. Normalmente tali attacchi sono pianificati e condotti dagli stati.

Una condizione correlata a quanto sopra è che il problema dell'attribuzione di un attacco possa essere risolto in misura tale che si tenga conto del contesto strategico e della realtà operativa. Ciò porta ad una maggiore sicurezza che un attacco informatico sia stato avviato da un attore statuale, come nel caso dell'uso del worm Stuxnet contro gli impianti nucleari iraniani. Il livello di intelligence e di capacità tecnologiche necessarie per effettuare un simile attacco restringe l'elenco dei sospetti.

In altre parole, l'attribuzione diventa meno problematica se l'obiettivo della deterrenza è uno Stato, che può essere più o meno potente, ed il comportamento da scoraggiare è un attacco informatico sofisticato e strategico contro una struttura blindata.

Considerando queste condizioni, l'impatto del machine learning sulla deterrenza informatica è ambiguo. Da un lato, la cyber-deterrenza può portare ad un potenziamento della

difesa informatica. Fornendo una difesa più attiva e adattiva, riducendo lo sforzo umano nel monitoraggio delle minacce e generando una risposta più tempestiva, il machine learning può aumentare i costi per un potenziale aggressore e quindi contribuire a promuovere la "deterrence by denial" nel cyberspazio.

L'attribuzione degli attacchi è un'altra area che il machine learning può rafforzare. La deterrenza sembrerebbe più credibile se l'aggressore perdesse il suo anonimato. In quanto tale, la deterrenza informatica potrebbe essere più fattibile con l'intervento del machine learning.
Le problematiche scaturenti a riguardo verranno analizzate nel **Paragrafo 2.3.2**.

2.3 Il problema di protezione dell'AI

L'impatto dell'Intelligenza Artificiale sulla sicurezza informatica è solitamente descritto in termini di espansione del panorama delle minacce. Proliferano le categorie di attori e individui abilitati attraverso l'Intelligenza Artificiale a compiere attacchi dannosi. Allo stesso tempo, nuove forme di attacco contro i sistemi di Intelligenza Artificiale aumentano in modo esponenziale la superficie esposta dei sistemi connessi. Per quanto riguarda questi cambiamenti, i ricercatori concordano sul fatto che l'AI influisca -pericolosamente- sul panorama della sicurezza informatica in:

- espansione delle minacce esistenti;
- introduzione di nuove minacce;
- alterazione delle caratteristiche tipiche degli attacchi.

Per quanto riguarda l'espansione delle minacce esistenti, si è verificato che tre caratteristiche dell'AI potranno influenzare le modalità in cui i nuovi attacchi informatici AI-powered verranno realizzati.

1. **Evasiveness**
   L'Intelligenza Artificiale sta contribuendo a modificare il modo in cui vengono rilevati gli attacchi. Un malware basato sull'Intelligenza Artificiale è molto più difficile da rilevare da un anti-malware. Si è diffusa una nuova classe di malware altamente mirati che utilizzano l'Intelligenza Artificiale per nascondere la propria natura in applicazioni benigne, come le applicazioni di videoconferenza, e identificano il proprio obiettivo attraverso il riconoscimento facciale, il riconoscimento vocale o la geolocalizzazione. Questi malware possono nascondere il loro intento fino a raggiungere il target definito, ciò rende questi malware fondamentalmente diversi dai classici attacchi "spray-and-pray".

2. **Pervasiveness**
   Come per i veicoli a guida autonoma, il potenziale pervasivo futuro di queste nuove tecnologie è chiaro. Questa era di intelligenza pervasiva sarà caratterizzata da una proliferazione di dispositivi intelligenti basati sull'Intelligenza Artificiale in grado di

riconoscere e reagire a immagini, suoni e altri schemi. Le macchine impareranno sempre di più dall'esperienza, si adatteranno a situazioni mutevoli e prevederanno i risultati. La dimensione del mercato globale dell'Intelligenza Artificiale è stata valutata a 39,9 miliardi di dollari nel 2019 e si prevede che crescerà a un CAGR[2] del 42,2% dal 2020 al 2027 [33]

3. **Adaptiveness**
   L'Intelligenza Artificiale è adattiva, il che significa che può imparare e in una certa misura diventare creativa e riesce a trovare idee a cui gli aggressori non avrebbero necessariamente pensato. Durante la conferenza DEF CON Hacking nel 2017, un gruppo di ricercatori ha mostrato come ha attaccato con successo un'applicazione web attraverso un'Intelligenza Artificiale che ha trovato la sua strada nell'uso dell'attacco di iniezione di database SQL (Structured Query Language). La particolarità di questo attacco era che l'Intelligenza Artificiale capiva da sola come funzionava la tattica di SQL injection.

Oltre all'espansione di minacce esistenti sia in scala che in portata, i progressi nell'AI potrebbero portare all'introduzione di minacce completamente nuove.

2.3.1 Deepfake

I deepfake sono una tecnologia in via di sviluppo che utilizza il deep learning per creare immagini, video o testi di eventi totalmente falsi. Esistono due metodi principali per creare deepfake.

Encoder-Decoder Questo metodo è generalmente utilizzato per il così detto "face-swapping" (i.e. mettere il viso di una persona sul viso di un'altra e viceversa) . Necessita di un primo algoritmo di AI chiamato encoder, che trova e apprende le somiglianze tra i due volti e li riduce alle loro caratteristiche comuni condivise, comprimendo le immagini nel processo.

Viene quindi insegnato ad un secondo algoritmo di AI a recuperare i volti dalle immagini compresse: un decoder recupera il volto della prima persona ed un altro recupera il volto della seconda persona. Quindi, assegnando immagini codificate al decoder "sbagliato", viene eseguito lo scambio di volti su quanti più fotogrammi di un video possibile per realizzare un deepfake convincente.

Generative Adversarial Network Una GAN mette due algoritmi di Intelligenza Artificiale in competizione allo scopo di creare immagini completamente nuove.Il primo algoritmo, il generatore, viene alimentato con dati casuali e genera una nuova immagine. Il secondo algoritmo, il discriminatore, controlla l'immagine e i dati per vedere se corrispondono a dati noti (cioè immagini o volti di persone vere).

[2]Tasso di crescita annuale composto

Questa battaglia tra i due algoritmi finisce essenzialmente per costringere il generatore a creare immagini estremamente realistiche che tentano di ingannare il discriminatore.

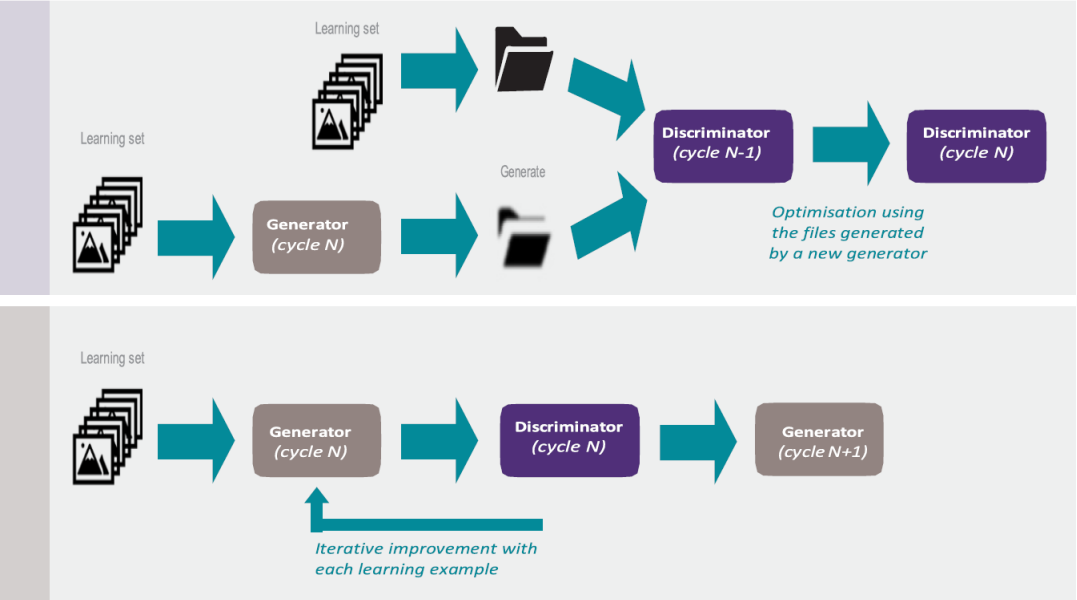

Figura 2.4: Funzionamento di una GAN.

Anche i deepfake utilizzati per la **manipolazione del testo** scritto stanno diventando sempre più preoccupanti. Con la scrittura generativa GPT-3[3] è possibile riprodurre sinteticamente frasi che sembrano dette da un essere umano che sono potenzialmente ancora più difficili da distinguere da quelle generate dall'uomo rispetto ai contenuti video.

La manipolazione del testo è stata ampiamente utilizzata per commenti e tweet generati dall'Intelligenza Artificiale sui social media. La capacità di riuscire a produrre un'opinione maggioritaria può portare a conseguenze di importante rilievo a livello politico e strategico, quindi è ragionevole pensare che la manipolazione del testo è e sarà sempre più applicata alle campagne che mirano a influenzare l'opinione pubblica.

L'uso dannoso dei deepfake è di tendenza in molte aree, tra cui due in particolare, come discusso di seguito.

Crimine e cyber sicurezza Negli ultimi periodi si è verificato che criminali abbiano utilizzato la tecnologia deepfake a scopo di lucro. Sebbene le truffe su Internet e via e-mail siano in circolazione da decenni, il progresso della tecnologia deepfake nel suono e nei video ha consentito attività criminali fraudolente ancora più intricate e difficili da

[3]Generative Pre-training Transformer 3: la terza versione dello strumento di scrittura generativa rilasciato da OpenAI

individuare.

Questi tipi di reati potrebbero variare da un livello base di attivisti informatici che fanno affermazioni e dichiarazioni false per minare e destabilizzare un'azienda, a sforzi più seri come ad esempio produrre falsi di dirigenti senior che confessano reati di tipo finanziaro o altro. I deepfake possono anche fare uso di social engineering per rendere più credibili le frodi utilizzando video o audio, ad esempio, di un membro dell'organizzazione presa di mira, aumentando le possibilità che gli attacchi abbiano successo.

Strumenti software in grado di individuare i deepfake criminali sono in fase di sviluppo, ma per per portare a termine e arrecare danni con un attacco di questo tipo basta che sia un solo individuo o dipendente in una azienda a credere nello scam.

Settore militare La preoccupazione per i deepfakes ha raggiunto anche il settore militare, e molte potenze mondiali ad oggi stanno prendendo le contromisure adatte. La DARPA nel 2018 ha iniziato un progetto per determinare e scovare immagini e video realizzati grazie all'AI. Inoltre, già nel 2019, testimoni legali ed esperti di Intelligenza Artificiale hanno detto ai legislatori statunitensi che dovevano agire immediatamente per stare al passo con la minaccia dei deepfake e di altra propaganda guidata dall'Intelligenza Artificiale, che potrebbe essere utilizzata da avversari come la Russia prima delle prossime elezioni presidenziali.

I deepfake quindi, in questo settore, possono essere usati sia in maniera politica che tattica. Esempi di potenziali problemi includono un leader della sicurezza nazionale che impartisce ordini falsi o agisce in modo non professionale, che potrebbe generare il caos. Da una prospettiva tattica o dal punto di vista di pianificazione della missione, i deepfake possono essere utilizzati per manipolare cartine o foto satellitari. Si potrebbe aggiungere un ponte dove in realtà non c'è, provocando poi perdite di tempo o addirittura imboscate per le truppe impiegate.

Questi sforzi sono in corso, con il Congresso degli Stati Uniti che ha approvato un programma da 5 milioni di dollari per potenziare le nuove tecnologie nel rilevamento dei deepfake. Ciò rivela che il Pentagono considera la manipolazione audiovisiva una questione chiave per la sicurezza nazionale.

Infine, come accennato in tutta questa analisi, una delle maggiori minacce dei deepfake sia nella vita pubblica che privata non è la tecnologia in sé, ma il suo potenziale di convergere con altre tecnologie e portare a sfide nuove e inaspettate.

Combinando diverse tecnologie, attori statuali e non saranno in grado di propagare massivamente notizie fuorvianti o false, mirando, con contenuti dannosi e dirompenti, a popolazioni specifiche attraverso le funzionalità messe a disposizione dai deepfake, IoT e AI.

È difficile prevedere esattamente come convergeranno queste nuove tecnologie, ma è facile immaginare a usi dannosi e devastanti se i governi, le aziende private e gli individui non saranno in grado di reagire, formarsi e prepararsi contro tali minacce.

2.3.2 Problemi legati alla cyber-deterrence

Anche se le condizioni citate nel **Paragrafo 2.2.3** sono soddisfatte, esistono diversi fattori che rischiano di ridurre l'efficacia della cyber-deterrence.

- Il primo è la possibilità di machine learning contraddittorio. Man mano che i modelli di difesa informatica basati sull'apprendimento automatico diventano più efficaci nel rilevare le minacce, i potenziali aggressori possono cercare modi per confondere tali modelli. Anche se gli attori sul lato difensivo possono fare affidamento su modelli di Intelligenza Artificiale per salvaguardare i loro sistemi, la loro fiducia nella deterrence by denial non deve essere esagerata. Questo perché gli aggressori possono riuscire ad avvelenare i modelli (noto anche come *machine learning poisoning*) o possono trovare altri modi per eluderli. In questo senso, i vantaggi in termini di sicurezza offerti dall'AI potrebbero essere contrastati.

  Un rapporto del 2018 ha avvertito che è stata prestata relativamente poca attenzione a rendere robuste ed adeguate le difese basate sull'Intelligenza Artificiale contro gli aggressori che ne anticipano l'uso. [7]

  Ironicamente quindi, l'uso del machine learning per la difesa informatica può effettivamente espandere la superficie di attacco di un sistema di difesa - i punti in cui un utente malintenzionato può interagire con il sistema - a causa di questa mancanza di attenzione e di altre vulnerabilità.

- Un secondo problema è la connessione non sempre trasparente tra attori e le loro capacità. È improbabile che gli attacchi informatici strategici, quindi attacchi che infliggono danni di impatto strategico nazionale, siano condotti da individui singoli.

  Le operazioni informatiche che intendono modificare il comportamento del bersaglio o provocare al bersaglio perdite considerevoli spesso comportano sforzi complessi per la preparazione, l'organizzazione, il coordinamento, i test e le prove. Richiedono un'abbondanza di risorse critiche, come la scoperta di zero-day vulnerabilities, strumenti di hacking e talento. Tali operazioni informatiche richiedono anche adeguate informazioni sulla preparazione alla difesa informatica dei sistemi presi di mira. Tuttavia, con lo sviluppo del machine learning, questi sforzi potrebbero essere eseguiti o facilitati da programmi automatizzati e adattivi.

  Questi programmi sarebbero in grado di scovare vulnerabilità, eludere il rilevamento, sconfiggere i sistemi anti-malware o persino riprogettare un'operazione in base alle proprietà riconosciute del sistema target. Ciò significa che operazioni informatiche

considerate complesse potrebbero non richiedere lo stesso livello di complessità organizzativa nell'era dell'AI. Singoli hacker o piccoli gruppi potrebbero anche portare a termine compiti che hanno conseguenze strategiche. Ciò creerebbe diversi problemi a catena per la deterrenza informatica.

  - La condizione "known identity plus known demand" sarebbe più difficile da stabilire, cioè essere in grado di determinare sia la fonte dell'attacco che i suoi obiettivi sarebbe confuso.
  - L'attribuzione diventerebbe più difficile perché le capacità, comprese altre prove forensi, come il linguaggio o la somiglianza con operazioni passate, potrebbero non essere più un indicatore affidabile per l'attribuzione dell'attacco.
  - Gli attacchi informatici con gravi conseguenze possono proliferare, minando la stabilità strategica nel cyberspazio.

- Infine, un maggiore utilizzo dell'AI nella sicurezza informatica renderebbe ancora più importante la raccolta e la condivisione dei dati. Ciò potrebbe rendere le policy di alleanza nel cyberspazio più comuni e intense. La difesa basata sul machine learning assume normalmente due forme. Una è il supervised learning, in cui l'obiettivo è imparare dalle minacce note per generalizzare e applicare questa conoscenza a nuove minacce. L'altra è l'unsupervised learning, in cui i programmi cercano di trovare deviazioni sospette dal comportamento normale.

  In entrambi i casi è richiesta un'analisi approfondita dei dati, capacità non banali e reti di intelligence. Pertanto, per rendere efficace la deterrenza, uno stato dovrebbe cooperare con altri stati nella condivisione delle informazioni al fine di costruire una rete di intelligence globale ed utilizzabile.
  Questo, che da un lato può incoraggiare le relazioni di alleanza nel cyberspazio, dall'altro porterebbe ad un'intensificazione delle già evidenti divisioni nel cyberspazio, creando un senso di antagonismo tra diversi gruppi e rendendo ancora più difficile il raggiungimento del consenso globale sulle norme di sicurezza informatica.

Nel complesso, l'Intelligenza Artificiale e il machine learning pongono rischi e offrono vantaggi alla sicurezza informatica. L'impatto sulla deterrenza informatica rimane poco chiaro, dal momento che sia la difesa che l'offesa potrebbero essere rafforzate dallo sviluppo dell'AI. Ciò suggerisce la necessità di un maggiore dialogo tra i ricercatori in campo di AI, i ricercatori strategici, i responsabili delle politiche e le altre parti interessate per raggiungere una maggiore chiarezza sulla deterrenza informatica e su come potrebbe avere un impatto sulle future relazioni strategiche e sugli arsenali militari.

2.3.3 Maggiore superficie di attacco

Poiché tutti i sistemi di Intelligenza Artificiale sono composti da software in esecuzione su hardware, i tradizionali attacchi informatici nei sistemi di Intelligenza Artificiale possono utilizzare una superficie di attacco tradizionale causata da bug software/hardware, che di solito derivano da errori umani commessi durante la scrittura del codice o la progettazione

dell'hardware.

In questo scenario, gli avversari troveranno quelle vulnerabilità e troveranno modi per sfruttarle e ottenere l'accesso al sistema sotto attacco. Questo è il solito scenario di attacchi nel mondo della sicurezza informatica.

Tuttavia, alcuni sistemi di Intelligenza Artificiale, tra i quali i sistemi di machine learning, presentano specifiche funzionalità interne e di utilizzo che possono essere attaccate con modalità diverse dai tradizionali attacchi informatici, sollevando nuove questioni di sicurezza.

Alcune delle caratteristiche specifiche di questi sistemi, che potrebbero portare a nuovi modi di attacco, includono:

- le fasi di addestramento o ri-addestramento, che includono un training dataset e un feature model come risultato;
- l'interazione con l'ambiente esterno, comprese le capacità di rilevamento che guideranno le decisioni interne e l'attuazione dei sistemi fisici;
- la capacità, in alcuni casi, di evolversi durante il runtime.

Pertanto, gli attacchi ai sistemi di machine learning possono sfruttare più delle semplici vulnerabilità del software. In particolare, il training dataset, sia prima che dopo l'implementazione, potrebbe essere compromesso in modo che l'"apprendimento" risultante del sistema non sia quello previsto.

Gli oggetti esterni rilevati dal sistema possono anche essere manomessi in modo che non siano riconoscibili come mostrato nel dataset di addestramento - un esempio ben noto sono i segnali di STOP leggermente modificati con il nastro adesivo. Tali attacchi non possono essere mitigati da patch software, perché appartengono o al dataset di addestramento che è già stato utilizzato per definire l'attuale 'comportamento' del sistema, o ad oggetti esterni che sono stati manomessi o sono strettamente legati al funzionamento del sistema, ad esempio nel caso di backdoor codificate nel modello.

Una conseguenza è che i sistemi che fanno uso di Intelligenza Artificiale espandono l'insieme di entità che potrebbero essere utilizzate per eseguire attacchi al sistema informatico: includendo i dati di addestramento, ma anche oggetti fisici. I dati possono essere armati in nuovi modi e sia attori che ogetti esterni ai sistemi ICT non possono più essere trattati come qualcosa di separato dal sistema.

L'ambiente dovrebbe ora essere considerato come parte della superficie di attacco in termini di sicurezza informatica. Questa è sicuramente una svolta inaspettata nella convergenza tra il regno fisico e quello cibernetico e che rende molto difficile proteggere i sistemi di Intelligenza Artificiale distribuiti in sistemi cyber fisici.

2.4 Framework legale

Come la cyberwarfare, l'Intelligenza Artificiale applicata al settore militare rappresenta una nuova rivoluzione tecnologica e non ha concetti e definizioni corrispondenti nel diritto e nelle norme internazionali. Ad esempio, permangono serie sfide su se e come i tradizionali concetti chiave del diritto di guerra, come la distinzione tra civili e combattenti e i principi di proporzionalità nei conflitti armati, possano essere applicati all'AI e alle relative piattaforme di combattimento senza equipaggio.

La seconda edizione del Manuale di Tallinn sulla guerra informatica, compilata da studiosi di diritto della comunità internazionale, offre alcune indicazioni. Tuttavia, la discussione sulle leggi e sui regolamenti dell'AI deve essere ulteriormente rafforzata. Prima che venga definita la norma sull'uso militare dell'AI, non si può escludere che i paesi utilizzino scappatoie e aree grigie nelle regole per minare la pace e la stabilità regionali.

L'Organizzazione per la Cooperazione e lo Sviluppo Economico (OECD) definisce un sistema di Intelligenza Artificiale come un "sistema basato su macchine, che può, per un determinato insieme di obiettivi definiti dall'uomo, fare previsioni, raccomandazioni o decisioni che influenzano ambienti reali o virtuali" [29]. Questa definizione è stata adottata anche dalla Commissione Europea nel "Regolamento su un Approccio Europeo per l'Intelligenza Artificiale".

Il rapporto del Centro comune di ricerca della Commissione europea sull'AI nell'Unione Europea, [23] pubblicato nel 2018, ha affrontato diversi aspetti dell'adozione dell'AI, da una prospettiva economica a una giuridica, inclusa la sicurezza informatica. Il rapporto riconosce la duplice natura dell'Intelligenza Artificiale e della sicurezza informatica e i potenziali pericoli per la sicurezza dei sistemi. Riconoscendo che il machine learning spesso non è robusto contro gli attacchi dannosi, suggerisce che "sono necessarie ulteriori ricerche nel campo del machine learning contraddittorio per comprendere meglio i limiti nella robustezza degli algoritmi di machine learning e progettare strategie efficaci per mitigare queste vulnerabilità".

Nel febbraio 2020 la Commissione Europea ha pubblicato il Libro Bianco[14] sull'Intelligenza Artificiale. Questo ha delineato una strategia che mirava a promuovere un ecosistema di Intelligenza Artificiale in Europa. Secondo il Libro Bianco, l'UE stanzierà finanziamenti che, combinati con risorse private, dovrebbero raggiungere i 20 miliardi di euro l'anno. Inoltre, prevedeva la creazione di una rete di centri di eccellenza per migliorare l'infrastruttura digitale dell'UE e lo sviluppo di meccanismi per consentire alle piccole e medie imprese (PMI) di reinventare meglio il proprio modello di business per incorporare l'AI. Sulla base delle raccomandazioni del "Gruppo di Esperti ad Alto Livello" sull'AI, l'UE ha anche definito i requisiti fondamentali per l'attuazione di tale tecnologia.

Secondo il Libro Bianco, i requisiti per le applicazioni di Intelligenza Artificiale ad alto rischio potrebbero consistere nelle seguenti caratteristiche chiave:

- training data;
- consistenza dei dati e dei registri;
- robustezza e previsione;
- supervisione umana;
- requisiti specifici per applicazioni AI specifiche, come quelle utilizzate per scopi di identificazione biometrica a distanza.

Il Libro Bianco sull'AI contemplava l'adozione di un quadro normativo flessibile e agile limitato alle applicazioni "ad alto rischio", in settori come la sanità, i trasporti, la polizia e la magistratura ed il mondo militare. Un regolamento di follow-up al Libro Bianco sull'AI è stato pubblicato il 21 aprile 2021.

Il "Regolamento su un Approccio Europeo per l'Intelligenza Artificiale"[36]della Commissione Europea promuove una protezione ad hoc per i sistemi di Intelligenza Artificiale ad alto rischio, sulla base di un ciclo di vita di sviluppo sicuro. Tuttavia, quando si tratta di cibersicurezza, il testo proposto potrebbe indicare più chiaramente alcuni passi aggiuntivi e necessari per raggiungere la sicurezza dei sistemi di AI. I requisiti proposti riguardano set di dati di alta qualità, documentazione e registrazione, trasparenza e fornitura di informazioni, supervisione umana, robustezza, accuratezza e cybersecurity.

Per quanto riguarda la cyber sicurezza, il Regolamento di cui sopra prevede che i sistemi di AI ad alto rischio "devono resistere ai tentativi di terzi non autorizzati di alterarne l'uso o le prestazioni sfruttando le vulnerabilità del sistema". Stabilisce inoltre che le soluzioni tecniche volte a garantire la sicurezza informatica dell'AI ad alto rischio dovrebbero comprendere misure per prevenire e controllare gli attacchi che tentano di manipolare gli input del set di dati di addestramento ("data poisoning") progettati per indurre il modello a commettere un errore ("esempi contraddittori") o difetti del modello.Questi requisiti rappresentano un passo fondamentale per assicurare il necessario livello di protezione dei sistemi di AI.

Migliorare il settore dell'Intelligenza Artificiale in modo tempestivo è particolarmente importante per l'Europa. Dato che il modello di mercato consolidato è caratterizzato da forti effetti di rete e di scala, i guadagni dei first-mover nell'adozione delle tecnologie di Intelligenza Artificiale sono particolarmente forti. Pur promuovendo il suo ecosistema di Intelligenza Artificiale, l'UE deve sia definire come rendere i suoi sistemi sicuri e affidabili, sia occuparsi di quale sia la giusta tabella di marcia per la cyber sicurezza a livello politico in UE, in modo da ottenere il massimo da tale ecosistema di Intelligenza Artificiale.

2.4.1 Applicazione del GDPR

L'**articolo 32** del GDPR richiede a titolari e a responsabili di attuare misure tecniche e organizzative per garantire la sicurezza del trattamento dei dati personali adeguate ai rischi per la sicurezza se presenti. Misure come la crittografia, la pseudonimizzazione o

l'anonimizzazione sono fornite come esempio di meccanismi per proteggere i dati personali.

L'**articolo 35** richiede che venga effettuata una valutazione d'impatto sulla protezione dei dati prima del trattamento di quelli personali ogni volta che una nuova tecnologia destinata a essere impiegata in un sistema di informazione esistente comporta rischi elevati per i diritti delle persone.

Inoltre, il paragrafo 3 dell'articolo sopra citato rende obbligatorio tale esercizio quando *"una valutazione sistematica ed estesa degli aspetti personali relativi alle persone fisiche che si basi su un trattamento automatizzato, compresa la profilazione, e sulla quale si basano decisioni che producono effetti giuridici nei confronti della persona fisica o incidono in modo analogo significativamente la persona fisica".*

Considerando queste disposizioni nel contesto degli strumenti basati sull'Intelligenza Artificiale, il GDPR richiede sia adeguate misure di sicurezza che un'adeguata valutazione dell'impatto sulla protezione dei dati per accompagnare implementazioni simili. **Ciò vale sia per l'Intelligenza Artificiale utilizzata per la sicurezza informatica sia per la protezione dei sistemi di Intelligenza Artificiale.**

Una serie di disposizioni lascia domande senza risposta o interessanti input di policy sull'impatto del GDPR sull'uso dell'AI per la sicurezza informatica. In questo caso, i sistemi di Intelligenza Artificiale consentono spesso pratiche di profilazione o l'automazione dei processi decisionali. Qui, è utile notare che **l'articolo 22** vieta il processo decisionale basato esclusivamente su mezzi automatizzati, che produce effetti significativi sugli interessati. Un'eccezione a questa regola generale è fornita dall'articolo 22.2, per cui il processo decisionale automatizzato non è considerato vietato se previsto dal diritto nazionale o dell'UE.

Va ricordato che uno dei principi di protezione dei dati più importanti è il cosiddetto concetto di limitazione delle finalità. In base a tale requisito, sancito dall'**articolo 5.1 (b)**, i dati personali dovrebbero essere "raccolti per finalità determinate, esplicite e legittime e non ulteriormente trattati in modo incompatibile con tali finalità".

Il principio di limitazione delle finalità potrebbe essere suddiviso in due sottocategorie, vale a dire la specificazione delle finalità (dati da raccogliere per finalità determinate, legittime ed esplicite) e la compatibilità dell'uso (l'ulteriore trattamento dei dati deve rispettare la compatibilità tra la finalità prevista della raccolta e la primo obiettivo di elaborazione). In un'era di interconnessione estrema (e sempre crescente), in cui i Big Data e l'IoT rappresentano rispettivamente due tecnologie di base su cui l'Intelligenza Artificiale è e sarà implementata, molti [26] sostengono che il principio della limitazione dello scopo stia diventando sempre più difficile da soddisfare; i dati personali non sono più un sottoprodotto di un servizio, ma il servizio è diventato un sottoprodotto della raccolta di dati personali.

Il GDPR introduce anche una serie di diritti che riguardano la spiegabilità dei sistemi di Intelligenza Artificiale. Più precisamente, il GDPR prevede, nell'**Articolo 13 e 14** che gli

interessati abbiano il diritto di ottenere la conferma dell'esistenza di un processo decisionale automatizzato, compresa la profilazione, e di acquisire informazioni significative circa la logica applicata, nonché l'importanza e le conseguenze previste di tale attività decisionale automatizzata su se stessi. L'articolo 22 include anche disposizioni che supportano gli interessati in modo che possano rivendicare i propri diritti e ritenere i responsabili del trattamento responsabili del trattamento dei loro dati personali.

Infine, dopo aver esaminato il quadro dell'UE sulla protezione dei dati personali, è anche importante evidenziare le disposizioni legislative dell'UE esistenti sul trattamento dei dati non personali.
Più precisamente, in molte situazioni della vita reale nel campo della sicurezza informatica, è probabile che i set di dati utilizzati siano composti da dati personali e non personali. Tali set di dati vengono spesso definiti "mixed dataset".
Allo stesso tempo, il rapido sviluppo di tecnologie emergenti come l'Intelligenza Artificiale, l'IoT, le tecnologie che consentono l'analisi dei Big Data e il 5G, sta sollevando interrogativi sull'accesso e il riutilizzo di tali set di dati.

Si potrebbe obiettare che ciò non dovrebbe creare confusione per le parti interessate coinvolte, poiché non vi sono obblighi contraddittori ai sensi del GDPR e del regolamento sulla libera circolazione dei dati non personali (regolamento FFD)[31]
Tuttavia, il quadro di protezione dei dati personali è molto più rigoroso del quadro dei dati non personali, mentre i confini tra dati non personali e dati personali sono troppo fluidi per fungere da ancoraggio normativo. In questo modo, i dataset che potrebbero rientrare nella categoria di dati personali potrebbero essere trattati come dati non personali dalle parti interessate coinvolte nelle loro attività di trattamento. Inoltre, anche dati non personali potrebbero essere utilizzati per attribuire identità o altre caratteristiche personali. Pertanto, due regimi separati applicabili a set di dati opachi potrebbero comportare problemi relativi all'adeguata applicazione delle norme sulla protezione dei dati personali.

2.5 Strategia NATO per l'AI

Lo scorso 21 ottobre 2021 i Ministri della Difesa della NATO hanno approvato la prima strategia dell'Alleanza Atlantica per l'Intelligenza Artificiale.

La strategia [27] delinea come l'Intelligenza Artificiale può essere applicata alla difesa ed alla sicurezza informatica in modo protetto ed etico. In quanto tale, stabilisce gli standard per l'uso responsabile delle tecnologie di AI, in conformità con il diritto internazionale e i valori della NATO. Affronta anche le minacce poste dall'uso dell'AI da parte degli avversari dell'Alleanza e come stabilire una cooperazione affidabile con la comunità dell'innovazione sull'AI.

L'Intelligenza Artificiale è una delle sette aree tecnologiche a cui gli Alleati hanno dato priorità per la loro rilevanza per la difesa e per la sicurezza. Questi includono tecnologie quantistiche, dati e computing, sistemi autonomi, biotecnologia e miglioramenti umani,

tecnologie ipersoniche e spazio. Di tutte queste tecnologie dual-use, l'Intelligenza Artificiale è nota per essere la più pervasiva, soprattutto se combinata con altre come i big data, l'autonomia o la biotecnologia. Per affrontare questa complessa sfida, i Ministri della Difesa della NATO hanno anche approvato la prima politica della NATO sullo sfruttamento dei dati.

La NATO ha sottolineato la necessità di "collaboration and cooperation" tra i membri su qualsiasi questione relativa all'AI per la difesa e la sicurezza transatlantica. Il documento elenca anche i principi dell'organizzazione per **l'uso responsabile dell'AI**, che secondo la NATO sono stati sviluppati sulla base degli approcci dei membri e del lavoro pertinente nelle sedi internazionali:

- **Lawfulness:**
  Le applicazioni AI saranno sviluppate e utilizzate in conformità con il diritto nazionale e internazionale, includendo il diritto internazionale umanitario e il diritto dei diritti umani, a seconda dei casi.

- **Responsibility and Accountability:**
  Le applicazioni AI saranno sviluppate e utilizzate con adeguati livelli di giudizio e cura; si deve applicare una chiara responsabilità umana al fine di garantire la responsabilità civile.

- **Explainability and Traceability:**
  Le applicazioni AI saranno adeguatamente comprensibili e trasparenti, anche attraverso l'uso di metodologie, fonti e procedure di revisione. Ciò include meccanismi di verifica, valutazione e convalida a livello NATO e/o nazionale.

- **Reliability:**
  Le applicazioni AI avranno use case espliciti e ben definiti. La sicurezza, la protezione e la solidità di tali capacità saranno soggette a test e garanzie all'interno di tali casi d'uso durante il loro intero ciclo di vita, anche attraverso procedure di certificazione stabilite dalla NATO e/o nazionali.

- **Governability:**
  Le applicazioni AI saranno sviluppate e utilizzate in base alle funzioni per cui sono previste e consentiranno un'adeguata interazione uomo-macchina, la capacità di rilevare ed evitare conseguenze indesiderate, e la capacità di adottare misure, come il disimpegno o la disattivazione dei sistemi, quando tali sistemi mostrano un comportamento non atteso.

- **Bias Mitigation:**
  Saranno adottate misure proattive per ridurre al minimo eventuali distorsioni non intenzionali nello sviluppo e nell'uso di applicazioni AI e nei dataset.
  "Underpinning the safe and responsible use of AI, NATO and allies will consciously put bias mitigation efforts into practice. This will seek to minimize those biases against individual traits, such as gender, ethnicity, or personal attributes[...]"[27] si legge nel documento.

La NATO condurrà adeguate valutazioni del rischio e dell'impatto prima di dispiegare le capacità dell'Intelligenza Artificiale. La NATO e gli Alleati condurranno anche dialoghi regolari ad alto livello, coinvolgendo le società tecnologiche a livello politico e strategico per essere informate e aiutare a modellare lo sviluppo di tecnologie messe in campo dall'AI, creando una comprensione comune delle opportunità e dei rischi derivanti dall'AI.

2.6 Il dilemma etico

L'applicazione militare dell'Intelligenza Artificiale porterà senza dubbio a questioni etiche riguardo il conflitto uomo-macchina.
L'umanità dovrebbe mettere il suo destino nelle mani di macchine più intelligenti di essa stessa?
Le macchine potranno sostituire gli umani nel processo decisionale e persino determinare la direzione finale della civiltà umana?
Una volta che l'AI sarà ampiamente utilizzata in campo militare, questi dilemmi diventeranno inevitabili.

Il ruolo che l'AI potrebbe svolgere per la robustezza, la risposta e la resilienza del sistema comporta sfide etiche che potrebbero ostacolarne l'adozione nella sicurezza informatica in qualsiasi campo di applicazione. Inoltre, se le questioni non venissero affrontate adeguatamente attraverso processi e politiche governative, potrebbero creare problemi significativi per le nostre società. Di seguito vengono riportate quelle che si presuppone diventeranno le principali sfide a livello etico.

1. Comando e Controllo - System robustness:
La robustezza di un sistema può essere migliorata utilizzando l'Intelligenza Artificiale per i test del software e per la progettazione di software in grado di eseguire self-testing e auto-healing.
Il self-testing si riferisce alla "capacità di un sistema o componente di monitorare il proprio comportamento dinamicamente ed eseguire test di runtime prima o come parte del processo di adattamento[4].

In questo contesto, quindi, l'AI può consentire la verifica continua dei sistemi e l'ottimizzazione del loro stato e rispondere rapidamente alle mutevoli condizioni, apportando le dovute correzioni. Tuttavia non è chiaro chi controlla il sistema di Intelligenza Artificiale. Infatti, sebbene l'articolo 22 del GDPR stabilisca che nessuna decisione importante su un individuo, come la profilazione, deve essere presa esclusivamente da un sistema autonomo, rimane poco chiaro, e spetta alle organizzazioni decidere dove finisce il controllo umano e inizia l'automazione. Inoltre, non è chiaro come assicurarsi che il sistema si comporti secondo le aspettative.

L'applicazione delle teorie principal-agent nel contesto dell'AI è quindi una sfida molto impegnativa. Questa teoria definisce una distinzione tra i proprietari e la direzione dell'organizzazione, per cui la direzione (agent) ha obiettivi e traguardi diversi rispetto al

proprietario (principal). Di conseguenza, il proprietario riceve un ritorno inferiore sull'investimento, poiché non gestisce personalmente l'azienda. Questo è stato definito come il problema principale-agente ed è un dilemma in cui l'agente agisce nel proprio interesse, che può essere contrario a quelli del principale. Sono stati previsti modi per superare questo problema, come fornire metodi di controllo strategico e finanziario, pianificazione della successione o ricompense monetarie e coaching o mentoring.

Tali metodi, tuttavia, sono difficilmente applicabili, o non applicabili affatto, in quanto in questo caso l'agente è artificiale. Delegare il controllo al sistema di AI può anche portare a errori e aumentare i rischi di conseguenze impreviste e deve essere adeguatamente bilanciato, prevedendo comunque una qualche forma di supervisione umana.

2. Responsabilità - System response:

L'Intelligenza Artificiale può migliorare notevolmente la risposta dei sistemi, ad esempio, correggendo automaticamente le vulnerabilità. Allo stesso modo, può anche offrire opzioni offensive per rispondere alle minacce. Esistono sistemi di sicurezza informatica autonomi e semi-autonomi che offrono una serie di risposte predeterminate a una specifica attività consentendo l'implementazione di specifiche risposte offensive.

L'AI può affinare strategie e lanciare counter operations più aggressive, che potrebbero facilmente sfuggire al controllo dei suoi utenti e alle intenzioni del progettista. Ciò potrebbe trasformarsi in un'intensificazione degli attacchi informatici e delle risposte ad essi, che a loro volta rappresentano un serio rischio di escalation in conflitti cinetici, minacciando le infrastrutture chiave delle nostre società.

Mentre l'aggiunta di un "human layer" causerà inevitabilmente dei ritardi in tali risposte, da un punto di vista sociale questa aggiunta solleverebbe il problema della responsabilità: come attribuire le responsabilità ai sistemi di risposta autonomi?
Come si possono promuovere comportamenti responsabili in questo contesto?
È necessario imporre e garantire la proporzionalità della risposta, una chiara definizione di attori e obiettivi legittimi mediante regolamento?

3. Abilità - System resilience:

I sistemi di Intelligenza Artificiale sono molto bravi a trovare vulnerabilità e identificare malware e comportamenti anomali, e lo fanno sicuramente in meno tempo e in un modo più efficace di quanto potrebbero fare gli analisti della sicurezza. Tuttavia, delegare completamente il rilevamento delle minacce ai sistemi di Intelligenza Artificiale sarebbe un errore poiché potrebbe portare a una diffusa dequalificazione degli esperti.

Anche allo stato dell'arte della tecnologia, i sistemi di Intelligenza Artificiale non sono ancora in grado di comprendere appieno attacchi e minacce complesse. L'interazione umana è necessaria per valutare i risultati dell'AI, per combinare gli avvisi, per ricostruire l'attacco che ha avuto luogo, per identificare le opzioni di risposta e per valutare e selezionare

quale sia quella migliore.

In questo contesto, gli esperti di sicurezza informatica dovrebbero continuare a trovare vulnerabilità e rilevare minacce nello stesso modo in cui i radiologi devono continuare a leggere i raggi X o i piloti a far atterrare gli aerei, in modo che siano ancora in grado di farlo se l'AI dovesse fallire o sbagliare.

A tal riguardo è interessante notare che, negli ultimi anni, la Marina degli Stati Uniti ha ricominciato a insegnare ai marinai a navigare con le stelle, a causa dei crescenti timori di attacchi informatici ai sistemi di navigazione elettronica.

Parte II

Case study

Capitolo 3

Ai fini di dare un significato pratico e tangibile a quanto esaminato nella prima parte di questo elaborato, si è deciso di effettuare una prova pratica che ha come obiettivo l'analisi del funzionamento di un servizio di sicurezza informatica che viene utilizzato dalla Forza Armata Esercito. Tale servizio implementa algoritmi di machine learning per l'individuazione ed il blocco di messaggi di posta elettronica malevoli.

Il servizio in questione provvede a due funzioni essenziali quando si tratta con informazioni confidenziali ed importanti: ***Data Loss Prevention*** ed ***Email Security***.

Di seguito verrà introdotto il sistema in questione, ponendo particolare attenzione alle due funzionalità appena elencate; nelle prossime sezioni, invece, si studierà e verrà testato in dettaglio il funzionamento della tecnologia anti spam utilizzata in Forza Armata, grazie ad una macchina virtuale messa a disposizione dall'azienda *Forcepoint*[1].

3.1 Introduzione

Il dominio cyber della Forza Armata è costantemente esposto a minacce di tipo informatico realizzate ad hoc da parte di attori malevoli in grado di utilizzare metodologie e tecnologie sempre più sofisticate.

La funzione di Comando e Controllo, essendo al *core* degli Enti, Distaccamenti, Reparti e Comandi in teatro nazionale ed estero, ed il servizio di accesso ad Internet ed Intranet da parte del personale, dipendono, in modo sempre più stringente, dall'affidabilità delle reti e dei sistemi informatici impiegati, e inoltre comporta l'adozione di sistemi di *Web Security* e *Content Filtering* che nel rispetto delle norme vigenti a livello Difesa ed

[1] a seconda del contesto: (i) Forcepoint LLC, a Delaware limited liability company with its principal place of business at 10900-A Stonelake Blvd., 3rd Floor, Austin, TX 78759, USA; or (ii) Forcepoint International Technology Limited, with a principal place of business at Minerva House, Simmonscourt Road, Dublin 4, Ireland; or (iii) Forcepoint Federal LLC, with a principal place of business at 12950 Worldgate Drive, Suite 600, Herndon, VA 20170; or (iv) a corporation or entity controlling, controlled by or under the common control of Forcepoint with whom an Order has been placed referencing this Agreement.

europeo e delle necessarie condizioni di sicurezza, consentano l'analisi del traffico e la gestione delle politiche di accesso al Web attraverso soluzioni in grado di eseguire tutte le funzionalità di cui sopra senza alcun degrado o rallentamento della navigazione Internet.

I dati sensibili ed in generale quelli personali, sono tutelati da norme e leggi valide in ambito nazionale e da specifiche direttive europee, (vedasi capitoli precedenti) emesse con l'obiettivo di garantirne l'integrità e la riservatezza. Se accanto all'aspetto normativo si affiancano l'elemento operativo e quello sanzionatorio, diviene evidente come il tema della prevenzione dalla perdita di dati sia centrale per un sistema informativo complesso come quello dell'Esercito Italiano.

A riguardo, la Forza Armata si è dotata di sistemi tecnologici finalizzati alla tutela dei dati e delle informazioni personali degli utenti del proprio dominio.
Sistemi come quello preso in considerazione in questo lavoro svolgono funzionalità quali prevenire la perdita dei dati mediante il controllo del flusso di email in uscita dalla rete strategica della Forza Armata (EINET), e di fornire le capacità di filtro Anti-SPAM, Anti-Virus e Anti-SpearPhishing.[20]

3.1.1 Data Security e DLP

Questo tipo di sistema previene la perdita di dati, gestendo rischi e conformità, individuando i dati sensibili, monitorandone l'uso, scoprendo dove sono archiviati e proteggendoli sia su rete, sia sulle postazioni di lavoro.
Seguendo le scelte operate dallo Stato Maggiore della Difesa si è deciso di salvaguardare i dati degli utenti appartenenti al dominio tramite metodi di **DLP Networking**, che consente di monitorare i canali della rete, eventualmente notificando gli utenti se dovessero essere presenti violazioni o quarantene; e metodi di **DLP Endpoint**, per eseguire rilevamenti sui dispositivi detti periferici, quali desktop e laptop degli utenti stessi.

I sistemi usati per la sicurezza dei dati prevedono quindi:

- Suite di strumenti per identificare, monitorare e proteggere i dati sia sull'intera rete della Forza Armata che su dispositivi e archivi mobili e in rete;

- Applicazioni per il controllo del Web, dove i contenuti dinamici generati dagli utenti rappresentano un rischio sempre maggiore;

- Regole di *Data Theft Prevention* per l'individuazione del furto di dati riconducibili a codici malevoli, tra cui l'invio di file, credenziali di accesso, link ed email non adeguatamente cifrate;

- Framework di procedure per consentire la visibilità su chi, dove, come e quali dati riservati vengono trasmessi, usati e archiviati sulla rete, permettendone il controllo attraverso azioni appropriate.[20]

3.1.2 Email Security

La sicurezza della posta elettronica comprende varie tecniche per mantenere i dati privati nelle e-mail e mantenere gli account al sicuro da accessi, perdite o alterazioni illegali.
I gateway di posta elettronica rappresentano una grave minaccia e una forte manifestazione di violazioni della sicurezza. Le informazioni personali vengono utilizzate dagli attaccanti per sviluppare diverse operazioni di phishing per imbrogliare le vittime e portarle a siti Web che forniscono malware.
La protezione consiste nel bloccare l'applicazione di posta elettronica per contrastare gli attacchi in arrivo e regolamentare le azioni per evitare la perdita di informazioni cruciali.

La soluzione di sicurezza dedicata alla posta elettronica garantisce, oltre alle funzionalità di Anti-Spam ed Anti-Virus, capacità di analisi del contenuto avanzate grazie all'integrazione nativa di moduli di protezione dei dati e del web.

I sistemi usati per la sicurezza email comprendono:

- Una suite completa di strumenti per la protezione della posta in entrata grazie a tecniche multiple di analisi dello spam;
- Motori di sicurezza per la gestione e la classificazione di malware avanzati e di nuova generazione;
- Strumenti per l'analisi di tutti i link *embedded* contenuti nella posta elettronica;
- Un sistema di cifratura per la protezione dello scambio di posta verso l'esterno, integrata nativamente con le regole per la prevenzione di perdita dei dati;
- Regole di *Data Theft Prevention* per l'individuazione del furto di dati, riconducibili a codici malevoli e per identificare macchine infette.[20]

3.2 Architettura utilizzata

L'architettura messa a disposizione dall'Azienda consiste in una applicazione web HTML5 che consente l'accesso ad un ambiente desktop utilizzando i protocolli VNC, RDP e SSH.

Apache Guacamole - il nome del prodotto - è un clientless remote desktop gateway. Viene denominato clientless in quanto non sono necessari plug-in o software client.[5]

A questa web application è associato un mail server ed una casella di posta elettronica.

Nella **Figura 3.1** invece è rappresentata l'architettura simulata della rete, ottenuta tramite il software GNS3.

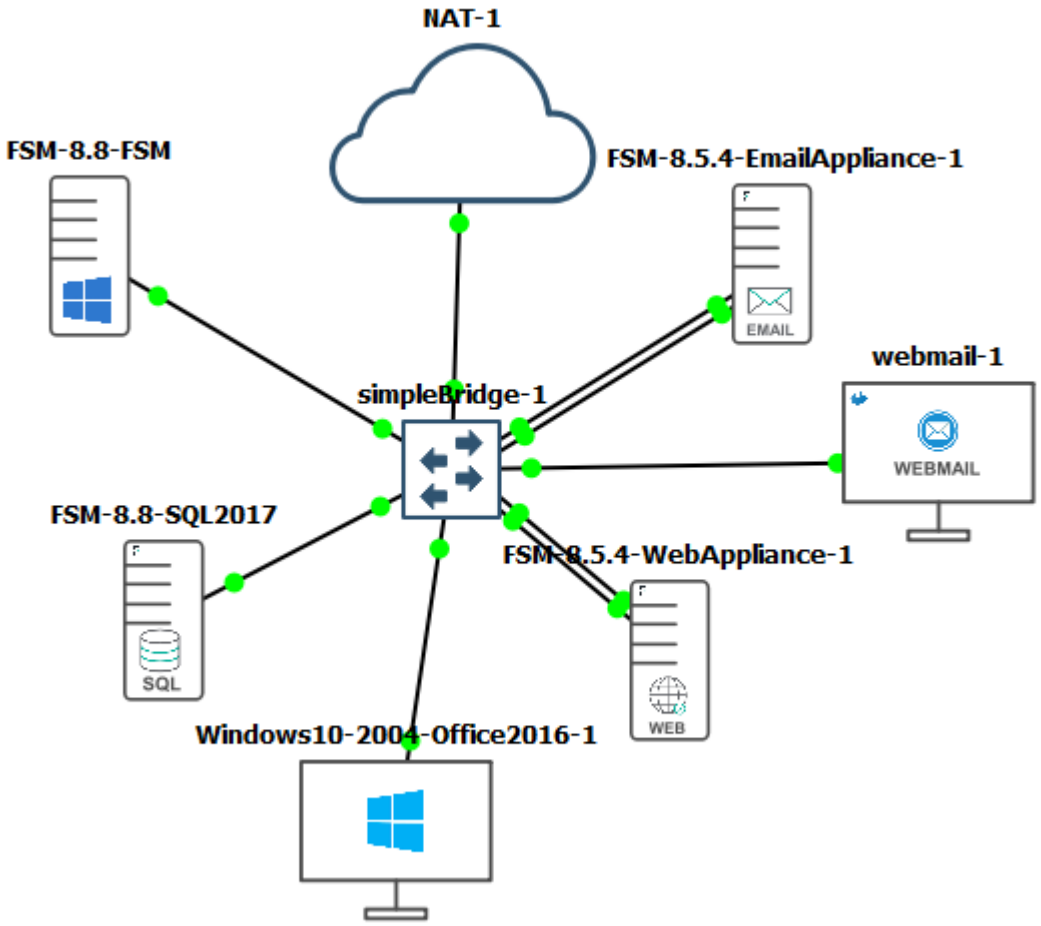

Figura 3.1: Architettura della rete.

- **NAT-1**
  Questo nodo consente la connessione ad Internet tramite il protocollo di Network Address Translation.
- **Webmail-1**
  La casella di posta usata come client email connessa al Security Manager.
- **FSM-8.8-FSM**
  Forcepoint Security Manager, applicazione utilizzata per analizzare la web security e l'email security.
- **FSM-8.8-SQL2017**
  Database Manager.
- **Windows10-2004**: Endpoint per la sicurezza contro la perdita dei dati.

- **Web Appliance-1**: Appliance per la sicurezza web.

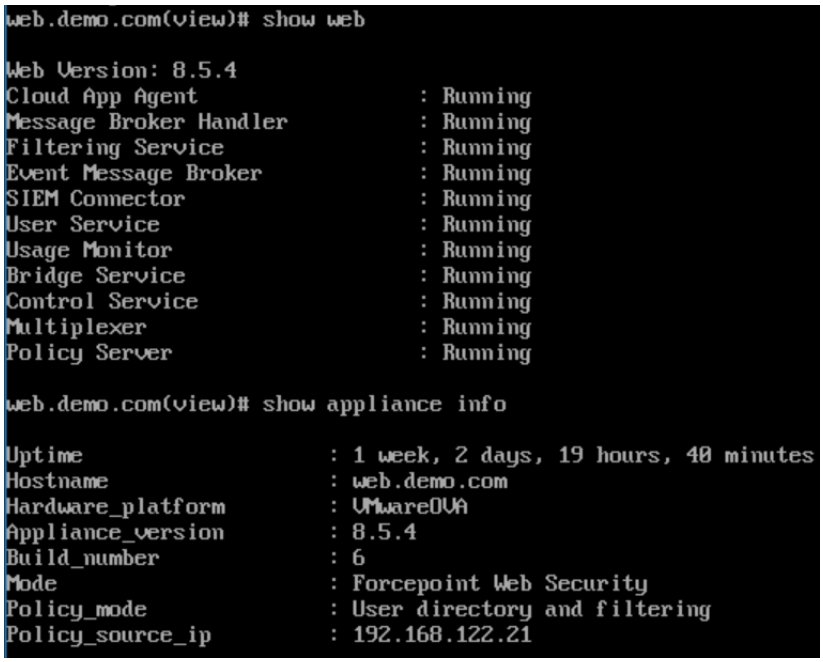

Figura 3.2: Web Appliance info.

- **Email Appliance-1**: Appliance per la sicurezza email.

Figura 3.3: Email Appliance info.

3.3 Test effettuati

Lo scopo dei test effettuati sulla piattaforma avente l'architettura descritta nella sezione precedente è quello di analizzare il comportamento del modello di machine learning usato come anti spam dalla Forza Armata.

A tale proposito sono stati usati tre differenti dataset in formato .CSV consultabili online contenenti testi di mail classificate come *spam* e come *ham*[2].

Utilizzando un opportuno codice scritto in python (**Algoritmo 1**), questi testi sono stati inviati sotto forma di e-mail all'indirizzo elettronico al quale è associato l'anti spam ed il security manager fornito dall'Azienda.

```
Data: sender address, sender password, receiver address
Result: send multiple emails to known receiver
initialization;
while not at end of csv file do
    read line;
    find text and subject;
    create MIMEMultipart message;
    connect to the server SMTP SSL;
    login to the server;
    if connection is feasible then
        send the message
    else
        infeasibility status.
        break
    end
    quit the server;
end
```

Algorithm 1: Algoritmo utilizzato per inviare le e-mail

A parità di policy impostate è stato valutato come l'algoritmo di machine learning ha risposto ad ogni set di dati e di conseguenza se il modello ha funzionato o meno, studiandone l'accuratezza.

Le policy di email security vengono applicate in base alle condizioni del mittente e del destinatario definite e in base alla direzione della mail. È quindi possibile applicare policy diverse in diversi reparti dell'organizzazione in cui il software di sicurezza è usato.

[2]e-mail legittima, che il destinatario vuole ricevere. Termine scelto come opposto di spam.

Quindi, dopo aver definito un criterio per un insieme di mittenti e destinatari, è possibile aggiungere le policy da applicare quando le condizioni del mittente/destinatario dell'e-mail corrispondono al criterio.

Le regole delle policy comprendono i filtri e le azioni dei filtri che determinano come viene gestito un messaggio che corrisponde alle condizioni del mittente/destinatario di un criterio.

I filtri forniscono la base per l'analisi della posta elettronica e le azioni dei filtri determinano la disposizione finale di un messaggio quando attiva un particolare filtro. Dopo aver creato e configurato i filtri e le loro azioni, queste possono essere incluse nelle policy. Tutto ciò è configurabile dal Security Manager.

La piattaforma ha messo a disposizione tre tipi di criteri, a seconda della direzione del messaggio: in entrata, in uscita o interna. La direzione del messaggio è determinata sulla base degli indirizzi di dominio protetti di un'organizzazione:

- **In entrata:** l'indirizzo del mittente non proviene da un dominio protetto e l'indirizzo del destinatario si trova in un dominio protetto.

- **In uscita:** l'indirizzo del mittente proviene da un dominio protetto e l'indirizzo del destinatario non si trova in un dominio protetto.

- **Interno:** sia l'indirizzo del mittente che quello del destinatario si trovano in un dominio protetto.

È disponibile un criterio di default per ogni direzione di posta elettronica, insieme a un criterio di prevenzione della perdita di dati (DLP) predefinito per ogni direzione.

Le politiche di prevenzione della perdita di dati possono essere applicate alla posta elettronica in qualsiasi direzione. Questi criteri sono configurabili nel modulo Data Security di Forcepoint Security Manager e sono abilitati o disabilitati nel modulo Email Security.

3.3.1 Filtri impostati

L'analisi degli spam esamina gli headers e il contenuto delle email in base ai filtri seguenti:

- **Digital Fingerprinting Analysis**
  Questo strumento confronta l'impronta digitale della mail ricevuta con gli spam noti contenuti nel database del sistema, e classifica il contenuto della mail in diverse categorie, tra cui *spam, clear, virus, bulk*;

- **LexiRules Analysis**
  Lo strumento dell'analisi delle regole lessicali analizza ogni mail ricevuta attraverso le combinazioni di parole e di pattern che sono tipicamente trovati negli spam.

- **Heuristics Analysis**
  Questo strumento analizza l'oggetto ed il testo dei messaggi per caratteristiche euristiche simili a spam noti. Potendo incrementare la sensibilità di questo strumento, è possibile incorrere in una ratio di *False Positives* maggiore. Per i test effettuati il livello di sensibilità è stato impostato al valore di default, cioè 3 su una scala di 5 livelli differenti.

Il sistema su cui sono stati effettuati i test permette di bypassare le analisi sopra indicate quando il messaggio supera una dimensione definibile dall'amministratore. Il valore di default, lasciato tale per i test, è di 3072KB.

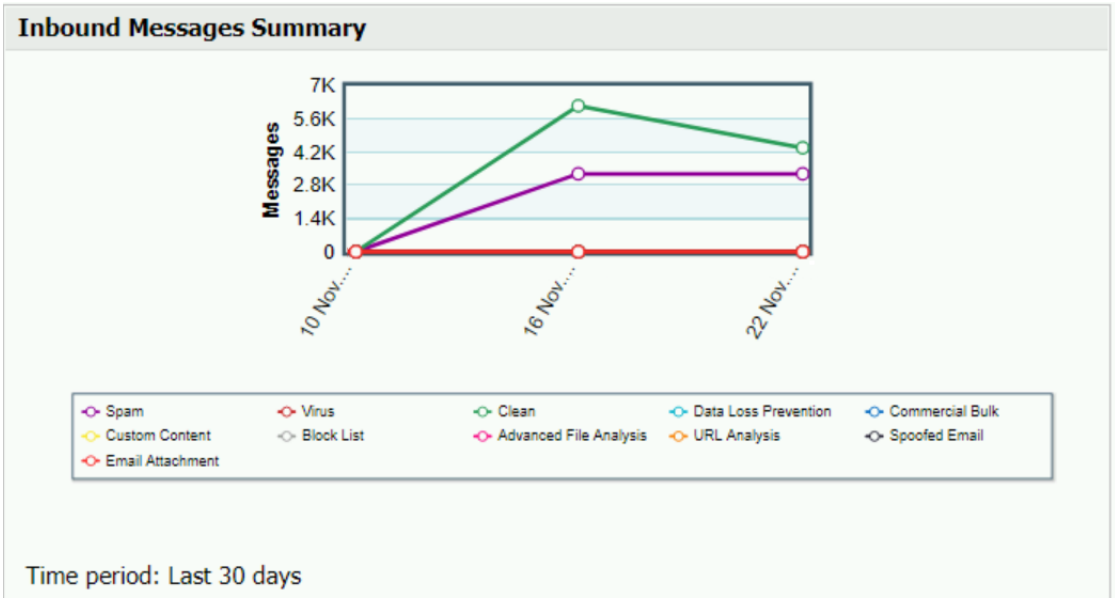

Figura 3.4: INBOUND MESSAGE SUMMARY - Il totale dei messaggi processati dal software di email security nel periodo di testing.

3.3.2 Azioni effettuate

Nel caso in cui un messaggio in arrivo dovesse attivare uno dei filtri sopra elencati, la piattaforma permette di intraprendere diverse azioni.
Si può infatti:

- Decidere se riprendere o meno l'elaborazione dei messaggi;
- Decidere se il messaggio viene consegnato (all'opportuna casella di posta) o eliminato;
- Decidere se salvare il messaggio in una cartella designata;
- Configurare o meno una notifica inviata automaticamente con le informazioni del messaggio.

Per i test effettuati le azioni sono state impostate nel seguente modo:

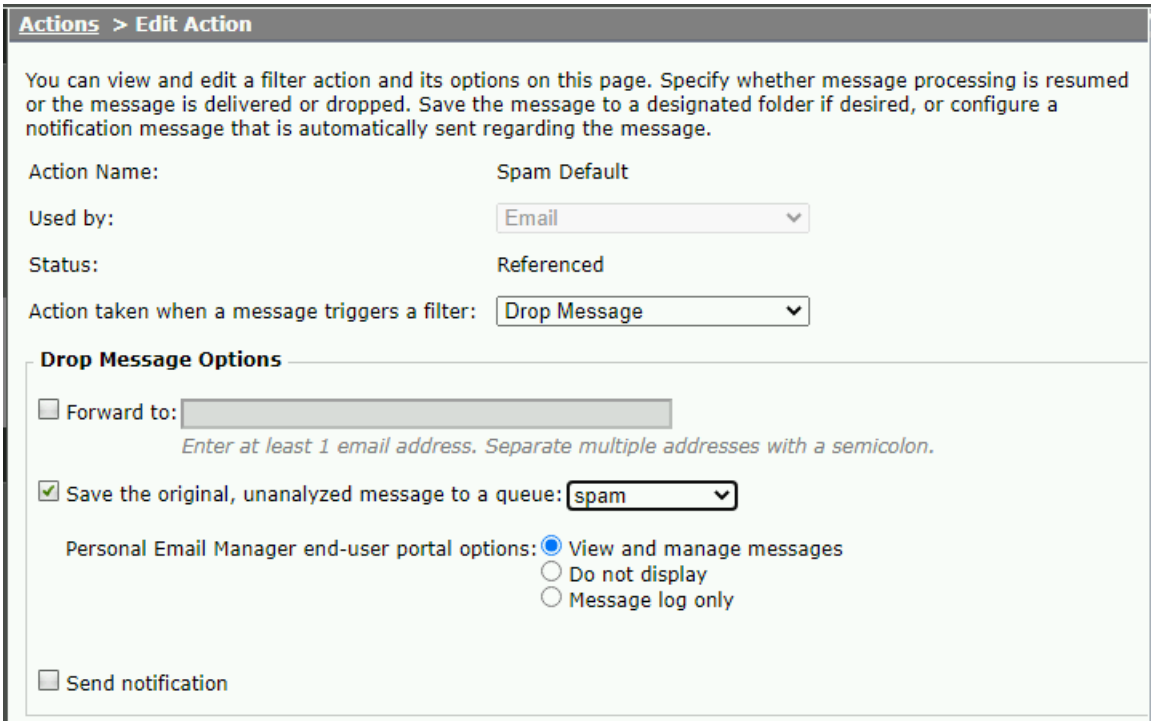

Figura 3.5: Azioni di risposta ai filtri impostati.

Infatti, grazie alla decisione "Drop message", nessun messaggio spam è stato recapitato alla casella di posta, ma sono stati tutti bloccati dal Security Manager in una cartella predefinita.

3.3.3 Analisi dei dati utilizzati

Per procedere con un'analisi che conduca a risultati corretti ed attendibili, le email devono provenire da fonti diverse, ed essere inviate inviate da più di un indirizzo, altrimenti si corre il rischio che il modello di antispam riconosca un mittente che è solito mandare molti spam e non esegua alcun tipo di analisi sulle mail ricevute ma le consideri tali a priori.

Per questo motivo sono stati scelti tre diversi dataset consultabili online, suddivisi in spam ed ham. La scelta è stata fatta in modo da variare le tipologie di spam ricevute dall'indirizzo di test, in modo da non far "abituare" il modello di machine learning.

Inoltre le email sono state inviate da 3 differenti domini e 5 indirizzi, al fine di rendere i risultati più realistici e anche per risolvere uno dei problemi analizzati nel **Paragrafo 3.4**.

In **Figura 3.6** è presente un grafico che riassume i vari mittenti al termine dell'analisi.

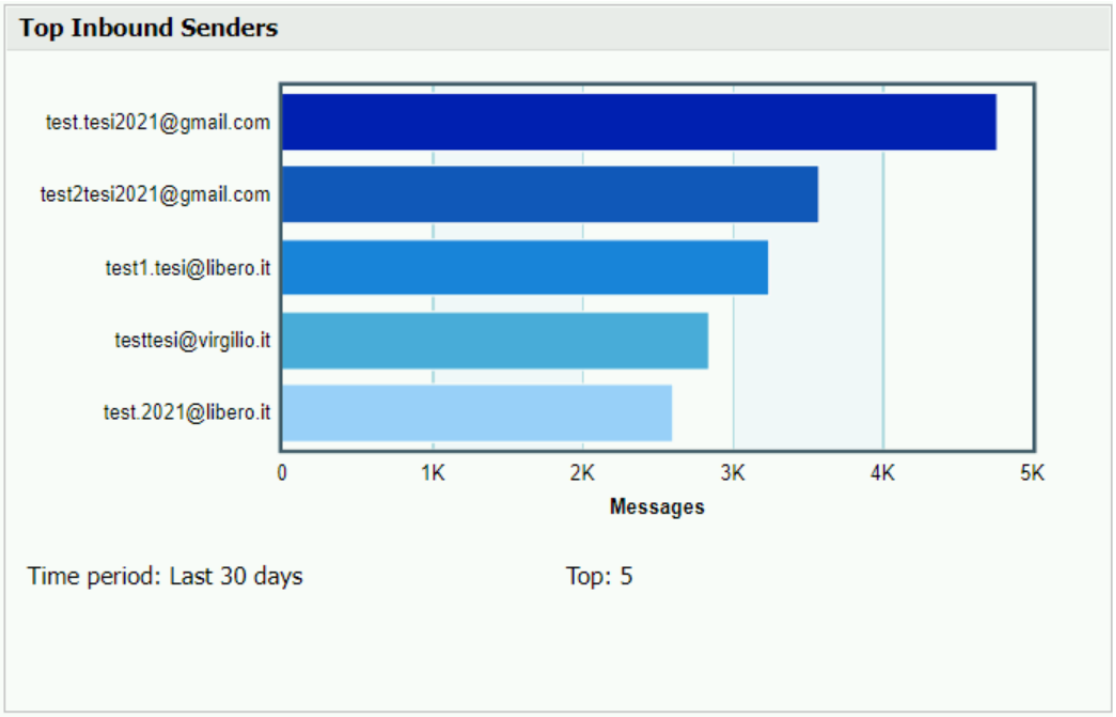

Figura 3.6: TOP INBOUND SENDERS - Gli indirizzi dei mittenti in entrata che rappresentano la maggior parte della posta in entrata, ordinati per volume di messaggi.

Il primo dataset utilizzato [1] è un file .CSV che contiene le informazioni correlate a 5172 file di posta elettronica selezionati casualmente e le rispettive etichette per la classificazione di spam o non spam.

Ad ognuna delle 5172 righe corrisponde una email. La prima colonna indica il numero identificativo della mail (trascurabile ai fini dei test effettuati), la seconda il tipo di mail, la terza contiene l'oggetto ed il testo mentre la quarta colonna ha il flag per i fini della previsione: 1 per spam, 0 per ham.
Pertanto, le informazioni relative a tutte le 5172 e-mail vengono archiviate in un dataframe compatto anziché come file di testo separati.
Questo dataset ha 1499 mail classificate come spam e 3673 come ham.

Il secondo dataset [2], invece, contiene 1369 mail classificate come spam su 5731 in totale.

Questo file .CSV ha solo due colonne, la prima contiene oggetto e testo della mail mentre la seconda il flag per identificare se sia considerata come spam o meno.

	A	B	C	D
1		label	text	label_num
2	605	ham	Subject: enron methanol ; meter # :	0
3	3624	ham	Subject: neon retreat	0
4	4685	spam	Subject: photoshop , windows , office .	1
5	2030	ham	Subject: re : indian springs	0
6	4185	spam	Subject: looking for medication ? we ` re	1
7	2641	ham	Subject: noms / actual flow for 2 / 26	0
8	4922	spam	Subject: vocable % rnd - word asceticism	1
9	3799	spam	Subject: report 01405 !	1
10	1488	ham	Subject: enron / hpl actuals for august 28	0
11	3948	spam	Subject: vic . odin n ^ ow	1
12	3418	ham	Subject: tenaska iv july	0

Figura 3.7: Le prime righe del dataset numero 1.

	A	B
1	text	spam
2	Subject: naturally irresistible your corporate identity lt is really hard to recollect	1
3	Subject: the stock trading gunslinger fanny is merrill but muzo not colza attaind	1
4	Subject: unbelievable new homes made easy im wanting to show you this hom	1
5	Subject: 4 color printing special request additional information now ! click here	1
6	Subject: do not have money , get software cds from here ! software compatibili	1

Figura 3.8: Le prime righe del dataset numero 2.

In questo file i messaggi sono ordinati, nel senso che nelle prime righe ci sono tutti i 1369 messaggi di spam mentre a seguire i 4362 ham.

Il terzo dataset è noto come *The Enron-Spam dataset*, è considerato una fantastica risorsa ed è utilizzato in molti studi di questa tipologia.
È stato raccolto da V. Metsis, I. Androutsopoulos e G. Paliouras e descritto nella loro pubblicazione. [39]

Il dataset originale contiene 17171 messaggi spam e 16545 ham (33.716 e-mail totali) [38]. Tuttavia, i data set originali sono stati registrati in modo tale che ogni singola mail sia in un file .txt separato, distribuito su più directory. Questo rende la lettura e l'utilizzo dei dati un po' macchinoso. Per poter utilizzare questo dataset in maniera più semplice e più veloce è stato utilizzato uno script python per raccogliere tutte le mail in un formato di più facile lettura, cioè un file .CSV.

	A	B	C	D	E
1		Subject	Message	Ham/Spam	Date
2	1	# 9760	tried to get fancy with your address	ham	25/10/2001
3	2	fw : sitara eol bridge proble	fyi > > > we were also monitoring the	ham	25/10/2001
4	3	important - to all domestic	corp savings plan	ham	25/10/2001
5	4	pipeline nominations away	these troubled times mandate that	ham	26/10/2001
6	5	enbridge buys koch ' s east	ngi ' s daily gas price index	ham	26/10/2001
7	6	here ' s the list , dirty , but it ' s a list , let me know who you ' re look		ham	26/10/2001
8	7	credit watch list - - week of	attached is a revised credit watch	ham	29/10/2001
9	8	need deal for march 2000	daren ,	ham	30/10/2001
10	9	re : tglo status	cost centers 11814 and 27117 are	ham	30/10/2001
11	10	monthly budget vs . actual	daren ,	ham	30/10/2001

Figura 3.9: Le prime righe del dataset numero 3.

Colonna	**Spiegazione**
Subject	Oggetto della mail
Message	Il contenuto dell'e-mail. Può contenere una stringa vuota se il messaggio ha solo una riga (quindi solo l'oggetto e nessun corpo). In caso di e-mail o risposte inoltrate, questo contiene anche il messaggio originale con oggetto "from:", "to:", "re:".
Spam/Ham	Ha i valori "spam" o "ham". Se il messaggio è stato classificato come spam o meno.
Date	La data in cui la mail è stata ricevuta. Il formato è YYYY-MM-DD.

Tabella 3.1: Spiegazione delle righe del terzo dataset.

Principalmente per motivi di tempo è stato deciso di analizzare solo una parte del dataset, cercando di mantenere le proporzioni tra spam e ham. Sono stati usati quindi 6000 messaggi, suddivisi in 3150 spam e 2850 ham.

3.4 Problemi e limiti riscontrati

Problemi:

- **Pulizia dei dati:**
  I dati utilizzati, nonostante provenissero da fonti attendibili, hanno comunque richiesto un certo livello di pulizia prima di essere utilizzati per gli scopi prestabiliti. È stato infatti necessario sistemare la codifica dei dataset per essere letti ed inviati; controllare se tra i vari messaggi ci fossero doppioni ed eliminarli nel caso (per garantire attendibilità ai risultati) ed infine controllare se tra le righe dei file usati ci fossero dati mancanti (NaN, nan, na) ed in caso eliminare tali righe.
  Per effettuare questi passaggi è stata utilizzata la libreria *pandas* di python.

- **Superamento del limite di invio tramite SMTP:**
  Al fine di arrivare a dei risultati validi, è stato necessario inviare più mail possibili alla casella di posta sotto esame. Ciò sarebbe stato impossibile da fare manualmente, senza utilizzare alcuno script (**Algoritmo 1**) che lo facesse in modo automatico.

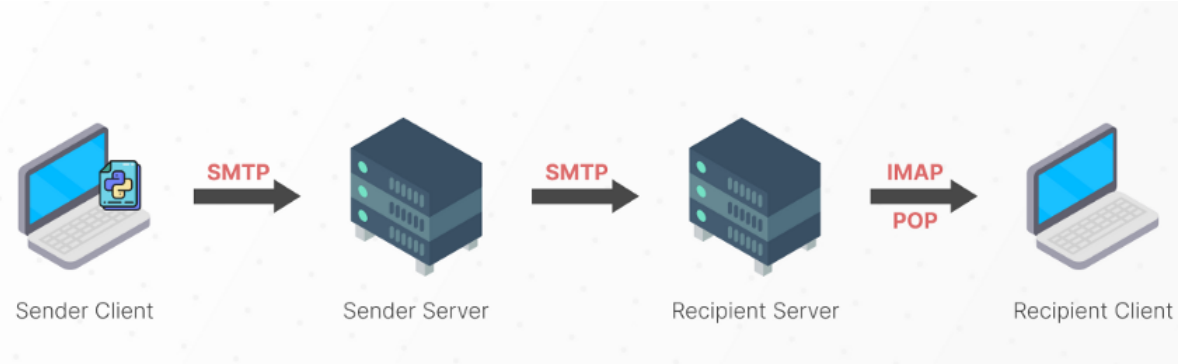

Figura 3.10: Ciclo di vita delle email.[3]

Il protocollo utilizzato per inviare le email è stato il protocollo **SMTP** (Simple Mail Transfer Protocol). È stata instanziata una connessione TLS con il Service Provider (il *Sender Server* della **Figura 3.10**) che attraverso l'uso di SMTP invia il messaggio al Service Provider del destinatario (*Recipient Server*).

I servizi di Service Provider comunemente noti però tendono ad avere un limite giornaliero ed orario al traffico SMTP, appunto per evitare che gli utilizzatori inviino spam o causino attacchi di tipo DoS. Per quanto questo sia corretto è stato un problema ai fini dei test in quanto ha limitato ed allungato il tempo necessario per l'invio delle mail alla casella di posta target.

Per ovviare il problema del limite orario è stata utilizzata la funzione sleep() per sospendere l'esecuzione del programma per un dato numero di secondi, per non incorrere nel limite posto dai Service Provider.

Tuttavia il problema del limite giornaliero non è stato risolto e periodicamente, dopo 1000 messaggi inviati da *gmail* e dopo 600 messaggi inviati da *libero* si incorreva rispettivamente nei seguenti errori:

- *smtplib.SMTPDataError: (550, b'5.4.5 Daily user sending quota exceeded. - gsmtp')*[3];
- *smtplib.SMTPDataError: (451, b'too many messages, slow down [smtp-35.iol.local; LIB-655]').*

Limiti:

- **Impossibilità di accedere direttamente al modello di machine learning:** Non avendo la possibilità di studiare e accedere al codice del modello di machine learning, i risultati presentati in questo elaborato rimangono empirici.

  Un esame del modello e di come processa i dati che gli vengono forniti sarebbe stato utile e necessario per arrivare ad una comprensione completa di alcuni aspetti che non sono potuti essere stati studiati a pieno (i.e. analizzare esattamente in base a cosa viene assegnato lo spam score ai messaggi in entrata - si rimanda al **Paragrafo 3.5.1**).

  Tale esercizio potrebbe essere oggetto di studio per un futuro approfondimento.

3.5 Analisi dei risultati

Di seguito (**Tabella 3.2**) vengono riportati i risultati in base a come il modello ha analizzato e classificato le email inviate.

Dataset	Ham	Ham trovati	Spam	Spam trovati	TOT
D1	3673	**3181**	1499	**1991**	5172
D2	4362	**4308**	1368	**1422**	5730
D3	2850	**2758**	3150	**3242**	6000
Totale	10885	**10247**	6017	**6655**	16902

Tabella 3.2: Riassunto dei risultati.

[3] *gsmtp* è il server SMTP di *gmail*

Considerando corretti i flag di ham e spam nei dataset, il modello ha considerato spam più messaggi di quanti realmente ce ne fossero, con un *False Positive Rate (FPR)* medio del 13%.

Di conseguenza, viene calcolata l'accuratezza in percentuale (**Tabella 3.3**):

Dataset	**Accuratezza**	**FPR**
D1	75%	33%
D2	96%	4%
D3	97%	3%
Totale	90%	13%

Tabella 3.3: Accuratezza del modello.

Per capire il motivo per cui l'accuratezza del modello sia così differente tra D1 rispetto a D2 e D3 è stato deciso di riprocessare le mail utilizzando policy diverse (**Figura 3.11**), quindi impostando i filtri ad un livello di tolleranza maggiore ed inviando una seconda volta le mail del primo dataset.

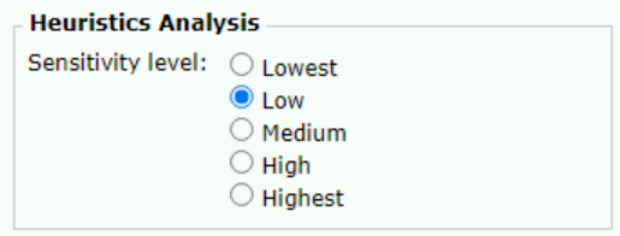

Figura 3.11: Il livello di tolleranza è stato impostato su *Low* per riprocessare D1, mentre per i primi test era stato impostato su *Medium*.

Dei 5172 messaggi inviati nuovamente, con le policy impostate su *low*, si è verificato che solo 890 sono stati considerati spam.
Da qui si calcola che c'è stato un *False Negative Rate* del **40%**.

Considerando che in tutti e tre i dataset utilizzati:

- I messaggi, sia oggetto che testo, non appartengono ad un settore specifico ma sono stati scelti in maniera casuale;
- Il testo di tutti i messaggi è in lingua inglese;
- Tutti i testi contengono "*stopwords*[4]";

[4]Parola che non ha valore ai fini del calcolo computazionale, come ad esempio numeri, preposizioni e pronomi

- Nessun dataset è stato "*tokenizato*[5]";
- Tutti i messaggi erano suddivisi in oggetto e testo;
- Nessun messaggio conteneva virus, allegati né è stato identificato come bulk mail;

Il motivo di un così atipico livello di accuratezza non può essere dovuto né al modello né alle caratteristiche del dataset.

3.5.1 Message log

La piattaforma mette a disposizione un servizio di *Message Log*, che registra le informazioni per ogni messaggio (sia inbound che outbound ed internal) processato dal sistema di sicurezza per email.

Attraverso i logs è possibile verificare per ciascuna mail il risultato dell'analisi effettuata dal modello di antispam, e verificare lo stato del messaggio.
Sì è verificato che, tra le 16902 mail inviate (**Figura 3.12**:

- Tutti i messaggi non aventi oggetto sono stati considerati *rejected*;
- Tutti i messaggi inviati senza connessione TLS non sono stati consegnati;
- I messaggi spam non sono stati consegnati alla casella di posta elettronica (risultando quindi come *dropped*) perché coincidevano con le caratteristiche euristiche che il modello di machine learning conosce (**Heuristics**) e per motivi di **Digital Fingerprinting Analysis**, vedasi paragrafo **(3.3.1)**;
- Tutte le mail ritenute *clean* sono state consegnate alla casella di posta elettronica e avevano uno **spam score** inferiore a 6;
- I messaggi per cui il modello non è riuscito a calcolare lo spam score sono stati etichettati come *Exception* e sono comunque stati bloccati dal sistema di sicurezza.

[5]Il processo di scomposizione di un testo in molte unità più piccole, come frasi e parole. Funziona separando le parole usando spazi e punteggiatura. Ogni unità è chiamata token.

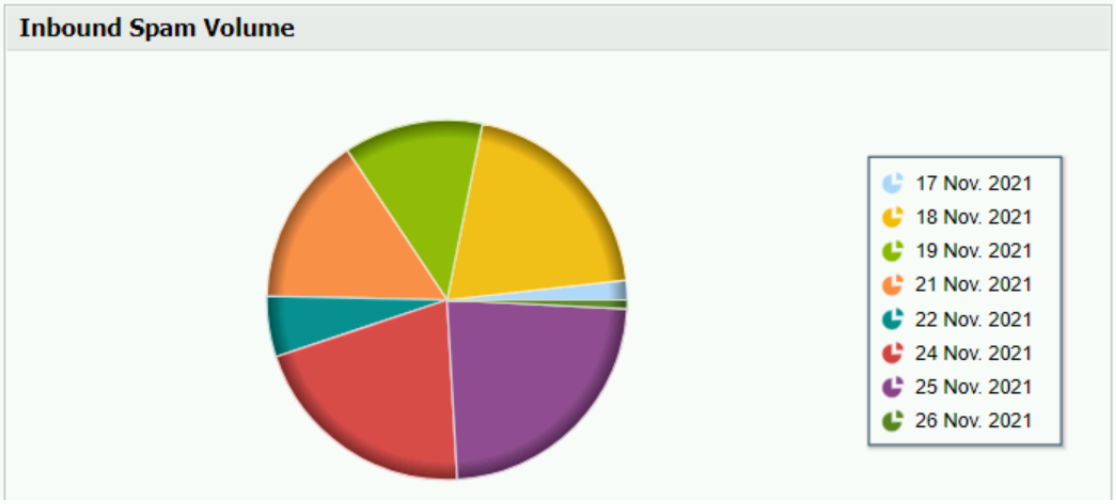

Figura 3.12: Il volume di spam inviato nelle giornate di test.

Message | Connection | Audit | Personal Email Manager | System | Console

Message Log ID: 100000025602
Received Date/Time: 29 Nov. 2021, 10:00:49
Subject: pictures
Sender address: testtesi@virgilio.it
Sender IP: 213.209.12.43

Details per Recipient

Recipient Address	Recipient IP	Direction	Delivered Date/Time	Policy	Rule	Analysis Result	Message Status	Quarantined?
test01@mark.williamson.lab.go4labs.net	N/A	Inbound	N/A	Inbound Default	Anti-Spam	Spam	Drop and Quarantine	Yes View

View Log Details

Date/Time	Log Type	Log Details
29 Nov. 2021, 10:00:10	Connection	Received **TLS** connection from **213.209.12.43** processed by **email-esg.demo.com**; connection GUID: **75177175125620 9888**; **SMTP connection attempt passed.**
29 Nov. 2021, 10:00:49	Message	Received **2.42KB** from message sender **testtesi@virgilio.it** to recipients **test01@mark.williamson.lab.go4labs.net**; tru e source IP address: **213.209.12.43**; message GUID: **769731545922436697**
29 Nov. 2021, 10:00:49	Policy	Analysis completed for the **Inbound** message policy **Inbound Default Anti-Spam** rule for analyzing message recipient **test01@mark.williamson.lab.go4labs.net**. This message was treated as a **Spam** message by the **Antispam heuristics** filter; policy action performed: **Drop message and save**.

Figura 3.13: Log detail di una mail classificata come spam.

3.6 Campagne di phishing

Fenomeno noto in tutti i settori, il phishing ha colpito anche le Forze Armate.
Le motivazioni sono principalmente le seguenti:

- Carpire le credenziali degli utenti che usano la webmail, per accedere ad aree personali o documenti;
- Richieste di denaro, sotto forma di imposte o fatture non pagate;
- Installazione di malware sotto forma di allegati al fine di carpire informazioni sui dispositivi degli utenti.

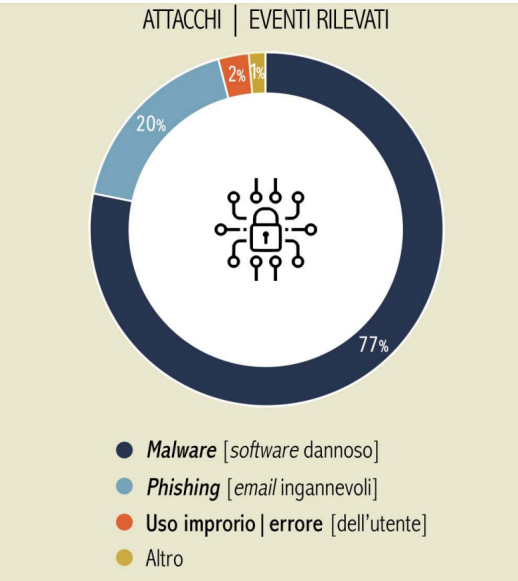

Figura 3.14: Cause rilevate degli attacchi alla rete della Forza Armata.[21]

Come spesso accade, gli utenti malintenzionati utilizzano il servizio di mail, sotto falsi indirizzi, per inviare messaggi sperando di ottenere ciò che cercano. Il problema principale in Forza Armata è che questi utenti malintenzionati riescono sempre di più a farsi scambiare per organi istituzionali utilizzando addirittura i domini dell'Organizzazione.

A partire dagli ultimi anni sono stati emanati, dagli enti di dominio, numerosi avvisi per richiamare l'attenzione degli utenti contro queste tipologie di attacco ai danni della Difesa, spesso allegando esempi di messaggi di phishing, typosquatting[6], o vishing[7].

[6]Forma di crimine cibernetico basata sullo sfruttamento degli errori compiuti dagli utenti nel digitare un indirizzo web o un indirizzo e-mail somigliante a quelli corretti (i.e. dfsa.it, difesam.it), Consiste nell'acquistare nomi a dominio corrispondenti agli errori di digitazione più comuni, anche dovuti alla fretta.

[7]Il vishing è una forma di truffa simile al phishing effettuata tramite servizi di telefonia, che ha lo scopo di carpire con l'inganno informazioni private sfruttando la persuasione tipica delle tecniche di Social Engineering. Gli aggressori, simulando l'esistenza di un call center o offrendo supporto tecnico, effettuano delle telefonate chiedendo alla vittima di fornire i propri dati ad un operatore. In altri casi l'attacco è finalizzato a far effettuare all'utenza operazioni sui sistemi informatici volte a consentire all'attaccante di condurre da remoto attività malevole di vario genere.

Si è quindi deciso di analizzare come il modello di machine learning utilizzato reagisce a queste tipologie di email.

Si riporta, a scopo dimostrativo, il testo di una mail di phishing:

OGGETTO:
Attenzione Urgente!!!

TESTO:
Gentile utente della webmail,

Stiamo riscontrando una congestione a causa della registrazione anonima degli account webmail esercito.difesa. Pertanto, stiamo aggiornando tutti gli account webmail per evitare il traffico di consegna dei messaggi. Se sei ancora interessato a utilizzare il tuo account e-mail, fai clic su questo link di aggiornamento diretto di seguito: https://mari.com.ng/webmailaccountupdate/ per aggiornare il tuo account webmail. Se non è cliccabile, copia e incolla l'indirizzo nel tuo browser.

Attenzione! Qualsiasi titolare di account webmail esercito.difesa, che si rifiuta di aggiornare il proprio account dopo aver ricevuto questa e-mail, perderà definitivamente il proprio account.

Saluti
Webmaster esercito.difesa
Reparto tenuta conto.
Copyright 2021.

Gli unici messaggi a cui è stato possibile avere accesso sono 13. È stato quindi creato un dataset in formato .CSV con questi messaggi[8] e sono stati inviati alla casella di posta elettronica utilizzata per i test.

I filtri applicati ad i messaggi in entrata per questo test sono stati gli stessi utilizzati per i precedenti test con l'aggiunta del filtro di *URL Analysis*.

In prima istanza le mail bloccate sono state **4**, le uniche che avevano un URL nel testo. È stato eseguito quindi un secondo test, eliminando il filtro di URL Analysis e delle 13 mail inviate una sola (una tra quelle bloccate in precedenza) è stata bloccata in quanto aveva uno spam score maggiore di 6. Non c'è quindi stato alcun modo di identificare i messaggi di phishing.

[8]Colonne: ID, Subject, Text, Attachments

Capitolo 4

Conclusioni

4.1 Cyber resilience

L'evoluzione così rapida della tecnologia nei tempi moderni, e di conseguenza l'evoluzione delle problematiche di sicurezza e non legate ad essa ci costringono ad uno studio dei sistemi di protezione cibernetica incessante.

Il dominio cibernetico, è teatro inevitabile di warfare, ed ad oggi il cybercrime rappresenta una delle maggiori minacce alla sicurezza, sia in ambito militare che civile. La sicurezza informatica rappresenta la risposta a tale minaccia.

Essa ha l'obiettivo di garantire confidenzialità, integrità e disponibilità dell'informazione ed è imperniata sulla **cyber-resilience** [32], ovvero sull'introduzione di misure atte a resistere agli attacchi informatici preservando le capacità funzionali di un sistema [19].
L'architettura di ogni infrastruttura protettiva poggia su tre presupposti fondamentali:

- la sicurezza, utile a proteggere i propri asset critici da minacce note ed emergenti;
- la vigilanza, vantaggiosa per aumentare la consapevolezza della minaccia e la localizzazione delle attività antagoniste;
- la resilienza, fondamentale per potenziare la capacità di pronta reazione agli attacchi [9].

In base allo scenario operativo - in un normale approccio orientato alla cyber-security - viene effettuata un'analisi del rischio di sicurezza del 'Sistema', che tiene conto delle minacce e della vulnerabilità corrispondenti ai dati da proteggere. L'analisi di rischio viene, peraltro, utilizzata anche per supportare il processo di certificazione [6].

Nello specifico, l'analisi del rischio deve individuare innanzitutto le risorse da proteggere: le componenti - hardware e software - del Sistema, i dati e le informazioni che il sistema deve gestire nonché i dispositivi di memorizzazione. Vengono, in seguito, identificate tutte le possibili minacce al sistema e, per ogni minaccia, tutte le vulnerabilità associate;

vengono considerati aspetti quali la capacità del nemico e la zona in cui opera il sistema nonché le misure di sicurezza da adottare [18].

Sulla base dell'analisi del rischio si stabiliscono le Contromisure del Sistema, al fine di garantire che la riservatezza, l'integrità e la disponibilità delle informazioni elaborate, memorizzate e trasmesse dal Sistema non siano alterate o compromesse. Le Contromisure, poi, porteranno alla definizione dei Requisiti di Sicurezza del Sistema, che necessitano di adeguata verifica [22]. La sicurezza aumenta con la qualità, l'affidabilità e la robustezza di un sistema.

I sistemi che utilizzano in modo intensivo reti di comunicazione e tecnologie digitali per il controllo, nonché scambio di grande quantità di informazioni, sono particolarmente esposti agli attacchi cibernetici: da questi attacchi bisogna difendersi a tutti i costi, anzi, opporre resistenza ed essere appunto resilienti.

Nel mondo militare in particolare, essendo i sistemi complessi e altamente integrati potenzialmente vulnerabili, è necessario che gli aspetti di sicurezza vengano affrontati in tutto il ciclo di vita dello sviluppo dei sistemi [30].

4.1.1 Safety e Security

Code Security Prevents Attacks
La sicurezza del codice riguarda la prevenzione di attività indesiderate o illegali nel software che creiamo e utilizziamo. Aiuta a garantire la sicurezza dei sistemi durante un attacco e a tenere fuori gli intrusi indesiderati. I test di sicurezza delle applicazioni statiche (SAST - Static application security testing) possono aiutare a migliorare la sicurezza.

Code Safety Ensures Reliability
La sicurezza del codice, d'altra parte, è un termine più ampio utilizzato per indicare se il software è affidabile e sicuro da usare. Ecco perché è stato sviluppato lo standard di codifica MISRA (Motor Industry Software Reliability Association), che fornisce un'esperienza sicura per gli utilizzatori.

Security Helps Achieve Safety
La sicurezza è un mezzo per raggiungere la *safety*. Questa non è solo questione di semantica. È una missione cruciale per i professionisti della sicurezza. Soprattutto quando c'è bisogno di bilanciare integrità, disponibilità e affidabilità per fornire software sicuro.

Le analisi di safety e di security, e i relativi standard di certificazione, sono stati a lungo mondi separati: questi mondi ora richiedono un approccio combinato. La security è infatti indispensabile per la safety. In presenza di cyber attacchi la safety rischia di essere compromessa con conseguenze catastrofiche per cui un errore di progettazione e/o di realizzazione su un componente non safety critical può costituire un pericoloso 'punto di accesso' per un attacco informatico con il pericolo di infettare componenti safety critical

[10].

Idealmente, gli aspetti legati alla sicurezza dovrebbero essere considerati in un'ottica sistemica, per evitare inutili duplicazioni a livello dei sottosistemi o lasciare aree di vulnerabilità. Per seguire la tecnologia mutevole e mitigare le conseguenze dei cyber attacchi appare, infine, essenziale lo sviluppo di approcci nuovi [32].

A livello organizzativo aziendale, oltre che naturalmente governativo-istituzionale, è divenuto così imprescindibile investire sullo sviluppo tecnologico per aumentare la resilienza cibernetica in un contesto di evoluzione, o di passaggio, dalla cyber defence alla cyber resilience, laddove la cyber defence cerca di evitare che gli avversari violino i sistemi mentre la cyber resilience mira a rendere i sistemi del cyber spazio più difficili da sfruttare.

I sistemi, nel dominio cyber, sono resilienti allorquando resistono agli attacchi informatici preservando le capacità funzionali e, quando, in caso di soccombenza, sono in grado di ripristinare le proprie funzionalità nel più breve tempo possibile[19].

4.2 I rischi per la Forza Armata

Lo stravolgimento di consolidati equilibri economici, politici e militari, può essere causa ed in parte conseguenza dell'evoluzione della tecnologia di gestione delle informazioni. Ciò richiede alle Forze Armate l'adozione di nuove strategie di impiego delle proprie risorse e nuove regole di cooperazione con il mondo dell'industria e delle aziende private.

Lo sviluppo delle capacità operative di ciascuna Forza Armata, un tempo frutto di costose autonomie in campo tecnico, tecnologico e logistico, può essere oggi raggiunto anche attraverso una calibrata acquisizione di supporto specialistico dall'esterno e veri e propri rapporti di partnership con l'Industria, riducendo al minimo l'immobilizzazione delle proprie risorse. Le scelte, però, non sono facili e neanche sempre convenienti.[16]

Delegare può essere conveniente, una più stretta collaborazione con l'Industria è, certamente, inevitabile, ma, specie nelle Forze Armate, le responsabilità ed i rischi non sono delegabili.

Naturalmente, l'outsourcing, essendo una delega ad un fornitore esterno, comporta tutti i rischi connessi con l'affidabilità che tali forniture possono offrire, incluso il monopolio. Occorre, come prima cosa, valutare quali possano essere le conseguenze di un mancato rispetto degli obblighi contrattuali da parte del privato contraente.[16]

In particolare, il rischio dell'utilizzo intensivo ed in più campi dell'intelligenza artificiale nelle Forze Armate è di comprare le tecnologie e sistemi già completi da fornitori esterni e restare solo degli utenti. Il fatto di non avere algoritmi proprietari ma dipendere da aziende private o da altri Paesi è un tema che inizia a preoccupare non solo il settore

militare italiano.

La completa dipendenza da altri Paesi è probabilmente il rischio che preoccupa di più. Se ci fosse il bisogno di essere costretti a dipendere da un Paese tecnologicamente più avanzato bisogna mettere in conto i rischi alla sicurezza, in quanto questo Paese avrebbe diretto accesso agli algoritmi utilizzati e a tutti i dati. Il danno si pone se le informazioni con cui si tratta sono classificate.

Inoltre la conseguente perdita di *Know-How* è un fattore da prendere in considerazione. Se il trend comincia ad essere rivolto sempre di più verso l'outsourcing completo di tecnologie e logistica, l'abbandono di capacità e di autonomie consolidate da parte dei militari può risultare irreversibile.

Per ovviare queste tipologie di problemi è necessario che le Forze Armate continuino ad investire nella formazione di personale tecnico ed esperti di dominio già appartenenti al settore militare.

4.3 Commento sui risultati ottenuti

I risultati descritti nel **Paragrafo 3.5** hanno evidenziato un tasso di *False Positive* non trascurabile tra le mail inviate classificate come spam.
Per quanto questi dati possano essere considerati non ottimali, bisogna applicare quanto visto al contesto in cui il modello di antispam viene utilizzato.

Lo strumento per il quale questo modello di machine learning è stato scelto lavora a protezione di caselle di posta che lavorano solo con certi tipi di domini e che vengono utilizzate solamente per scopi lavorativi, e gli utenti che ne fanno uso conoscono già da che tipo di indirizzi sono soliti ricevere posta elettronica.

Paradossalmente, avere un tasso alto di falsi positivi sicuramente non compromette l'utilizzo del servizio di posta al quale il modello di antispam è associato ma considerando spam messaggi provenienti da mittenti non conosciuti potrebbe evitare fenomeni di tipo phishing e comunque evitare di distrarre gli utilizzatori.

4.3.1 E-mail di phishing

La Forza Armata, operando nel più ampio contesto della Difesa, è costantemente impegnata nella protezione dei propri sistemi e delle proprie reti non classificate dalle minacce cibernetiche. In tale ambito, ogni utente, indipendentemente dal grado posseduto e dall'incarico ricoperto, riveste un ruolo cruciale e quindi è determinante utilizzare in modo corretto e sicuro gli strumenti informatici che l'Esercito rende disponibili.

Dato che questo tipo di minaccia non è scovabile dal sistema di sicurezza posto in essere, è responsabilità piena di ogni utente essere in grado di gestire queste insidie e non cadere in trappola.
È altresì importante formare gli utenti in modo tale da rendere il personale consapevole dei rischi per la sicurezza che un attacco di phishing può causare.

Le citate e-mail potrebbero arrivare da parte di indirizzi appartenenti al dominio di Forza Armata o similari, interessando anche la Posta certificata collegata al Sistema Documentale. È quindi fondamentale non interagire con le e-mail riconducibili alla descrizione summenzionata.

Per le postazioni della rete EINet sono state poste in essere le misure di sicurezza atte a mitigare gli effetti della citata minaccia. Tuttavia particolare attenzione va posta verso quelle postazioni con accesso diretto ad Internet (no tramite EINet), ad es. PC privati o di servizio che accedono alla Webmail Esercito, le quali non risultano protette dalle misure di sicurezza garantite dalla EINet.

4.4 Considerazioni per il futuro

Il cybercrime rappresenta, attualmente, una delle maggiori minacce alla sicurezza, sia in ambito militare che civile. Questa minaccia, abbiamo visto, è particolarmente elevata per i sistemi che supportano la movimentazione fisica di persone e merci, come i sistemi di gestione e di comando e controllo.

La cybersecurity costituisce la risposta a tale minaccia; essa è imperniata sulla cyber resilience, cioè sull'introduzione di misure atte a resistere agli attacchi informatici preservando le capacità funzionali di un dato sistema.
Il passaggio da mettere in atto, in particolar modo nelle strategie difensive, è quello di transitare dalla cyber-defence alla cyber-resilience.Tale evoluzione renderebbe le infrastrutture critiche più resilienti rispetto al passato, passando idealmente da un orientamento debolmente reattivo agli attacchi ad uno idealmente proattivo, con sistemi e strutture ridondanti, in termini informatici, ed ingegneristicamente concepite in modo tale da assicurare un comportamento ottimale, in condizioni standard, e resiliente quando sotto attacco.

Posto che il rischio di attacco cyber non è mai eliminabile, ma solo mitigabile, per fronteggiare la molteplice e mutante minaccia occorrono in primis investimenti a livello tecnologico-informatico e di know-how, a livello statuale ma anche industriale. Concepire infrastrutture critiche resilienti sin dal design preliminare richiede, inoltre, accordi strategici di cooperazione pubblico-privato, in particolare tra il comparto militare e quello civile, in un frame normativo, anche europeo, ancora disorganico e perciò migliorabile.

Il dibattito sul conflitto cibernetico ha conosciuto negli ultimi anni un'intensificazione senza precedenti. La cyber-security, tuttavia, rappresentando un ambito relativamente

nuovo, richiede un'attenta regolazione, che recepisca stimoli e indicazioni da tutti gli altri comparti. È quindi da ritenere di fondamentale importanza dotarsi, a livello globale, di linee guida e di approcci standardizzati sia in ambito governativo che in quello di infrastrutture critiche. Servirebbe con urgenza, quindi, un quadro normativo organico per individuare le infrastrutture critiche nazionali e per determinare le modalità di protezione attraverso un sistema sinergico tra istituzioni, operatori e industria [30] [24]. Ma affinché la risposta alle minacce cyber non provenga più in futuro solo dai dipartimenti ICT delle diverse organizzazioni, occorre un cambiamento culturale ed organizzativo epocale.

Per arrivare ad un approccio 'cyber security-based' e per aumentare la cyber awareness paiono necessari investimenti in formazione e training. A livello organizzativo, in primis nelle company, sembra impellente un modello 'cybersecurity centric' in cui la cultura della difesa e della resilienza permei tutti i livelli aziendali. Tenendo conto però che nel cyberspace la vulnerabilità principale sia di tipo umano, permangono in Italia grosse aree da sensibilizzare e c'è molto lavoro ancora da fare nelle organizzazioni complesse d'ogni tipo.[8]

Le nuove forme di manaccia emergente impongono, concludendo, un incremento di collaborazione tra il mondo militare e quello civile. Considerazione, questa, che convince della necessità di un'evoluzione dei rapporti pubblico-privato, fatta di leale cooperazione e di mutuo supporto. Più specificatamente, essendo le infrastrutture critiche, come quella di Comando e Controllo dell'Esercito Italiano, sistemi con molteplici aperture alla minacce cyber, quanto più si capiranno la loro vulnerabilità e i loro impatti, tanto più si sarà in grado di decidere al meglio dove investire in termini di risorse umane, oltre che naturalmente di tecnologie, puntando allo sviluppo di sistemi e soluzioni, basandosi sulle eccellenze che l'Italia produce e facendo sistema all'interno di una strategia nazionale di sviluppo della cyber-security che, se interpretata in quest'ottica, può rendere il Paese competitivo con le più importanti realtà industriali mondiali.

L'utilizzo di nuove tecnologie e lo sviluppo delle stesse sta portando l'Esercito Italiano ad un nuovo livello di concorrenza e di importanza nelle operazioni nazionali ed internazionali. Nelle Forze Armate Italiane, in generale, esistono molteplici encomiabili e particolari elementi che il cittadino non conosce ma che fanno spesso la differenza in questioni critiche all'interno di teatri operativi.

Come giovane Ufficiale dell'Esercito Italiano, quindi, ritengo sia estremamente rilevante valorizzare e porre particolare attenzione agli sforzi della Forza Armata per la produzione di Software e di sistemi di alto livello prima all'interno della stessa e in seguito all'opinione pubblica.

Ten. RN Co.Ing. Ferdinando Pulella

Bibliografia

[1] Available on line, Licence Open Data Commons Open Database License (ODbL) v1.0. URL: https://www.kaggle.com/balaka18/email-spam-classification-dataset-csv.

[2] Available on line, Licence Open Data Commons Open Database License (ODbL) v1.0. URL: https://www.kaggle.com/karthickveerakumar/spam-filter.

[3] A.Vaccaro. *Automate Email Sending with Python - The Definitive Snippets Collection for your ETL Pipelines*. Available on line. 2021. URL: https://towardsdatascience.com/automate-email-sending-with-python-74128c7ca89a.

[4] T.M. King et. al. "AI for testing today and tomorrow: Industry Perspective". In: (2019).

[5] APACHE. *Guacamole Documentation and Manual*. Available on line. URL: https://guacamole.apache.org/doc/gug/.

[6] T. Aven. *On some recent definitions and analysis frameworks for risks, vulnerability, and resilience, Risk Analysis*. Vol. 31. 4. 2011, pp. 515–522.

[7] M. et al. Brundage. *The Malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation*. Future of Humanity Institute et al., Oxford, 2018.

[8] F. Castaldo. *Dalla Cyber Defense alla Cyber Resilience dell'infrastruttura critica. Alcune implicazioni strategiche e organizzative*. Available on line. 2019. URL: https://www.openstarts.units.it/bitstream/10077/31169/1/REPoT_2019%5C%283%5C%29-2_Castaldo.pdf.

[9] M. Castaldo F. & Gatti. *Tempestività e resilienza: l'esperienza dei piloti al servizio del business*. Rivista Italiana di Conflittologia, 2019, pp. 23–41.

[10] L. Corradini I. & Franchini. *Ingegneria sociale. Aspetti umani e tecnologici*. Themis, Roma, 2016.

[11] Department of Defence. *The Department of Defence Cyber Security Strategy*. 2015.

[12] A. Faroqi E.g. Rajagopalan. *Intelligent missile guidance using artificial neural networks*. Vol. 4. 1. Artificial Intelligence Research, 2015.

[13] Alessandro Marrone ed Ester Sabatino. *La difesa cibernetica nei Paesi NATO: modelli a confronto*. Available on line. 2020. URL: https://www.iai.it/sites/default/files/pi_a_164.pdf.

[14] Commissione Europea. *LIBRO BIANCO sull'intelligenza artificiale - Un approccio europeo all'eccellenza e alla fiducia*. 2020.

[15] Official Journal of the European Union. *Directive (EU) 2016/1148 of the European Parliament and of the Council concerning measures for a high common level of security of network and information systems across the Union*. Available on line. 2016. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.194.01.0001.01.ENG&toc=OJ:L:2016:194:TOC.

[16] G. Genovese. *L'esternalizzazione: vantaggi e svantaggi dell'outsourcing nelle capacità di supporto tecnico-logistico delle F.A.* Available on line. 2019. URL: https://www.difesa.it/InformazioniDellaDifesa/periodico/IlPeriodico_AnniPrecedenti/Documents/Lesternalizzazione_vantaggi_e_s_707Genovese.pdf.

[17] M. Giles. *AI for cybersecurity is a hot new thing—and a dangerous gamble*. MIT Technology Review, 2018.

[18] J.A. Green. *Cyber Warfare. A multidisciplinary analysis*. Routledge, New York, 2015.

[19] Y.Y. Haimes. *On the definition of resilience in systems, Risk Analysis*. 2009, pp. 498–501.

[20] Capitolato Tecnico - Esercito Italiano. *Estensione diritti d'uso e licenze utenti per sistema di Data Loss Prevention (DLP), Email Security Gateway (EMSG) e Web Content Filtering Gateway (WCFG) in uso sulla rete non classificata dell'EI*.

[21] Esercito Italiano. *Sicurezza delle Informazioni - Rapporto Esercito*. Available on line. URL: esercito.difesa.it/Rapporto-Esercito/Documents/2018/CPT-11_INFORMAZIONI.pdf.

[22] W.J. Lynn. *Defending a New Domain: The Pentagon's Cyberstrategy*. Vol. 89. 5. USA Foreign Affairs, 2010, pp. 97–108.

[23] et. al. M. Craglia (ed.) A. Annoni. *Artificial Intelligence – A European Perspective*. Publications Office, Luxembourg, 2018.

[24] E. Marchetti. *Private Military and Security Companies: il caso italiano nel contesto internazionale*. 7. Quaderni IAI, Edizioni Nuova Cultura, Roma, 2013.

[25] Presidenza del consiglio dei Ministri. *Cyberspace between national security and protection of individal freedom*.

[26] L. Moerel e C. Prins. *On the Death of Purpose Limitation*. Available on line. 2019. URL: https://iapp.org/news/a/on-the-death-of-purpose-limitation/.

[27] NATO. *Summary of the NATO Artificial Intelligence Strategy*. Available on line. 2021. URL: https://www.nato.int/cps/en/natohq/official_texts_187617.htm.

[28] NATO. *Trattato Nord Atlantico*. Available on line. 1949. URL: https://www.nato.int/cps/fr/natohq/official_texts_17120.htm?selectedLocale=it.

[29] OECD. *AI Policy Observatory*. Available on line. 2019. URL: www.oecd.ai/ai-principles.

[30] M. Ouyang. *Review on modeling and simulation of interdependent critical infrastructure systems, Reliability Engineering & System Safety*. Vol. 121. 2014, pp. 43–60.

[31] European Parliament e The Council of the European Union. *Regulation (EU) 2018/1807 of the European Parliament and of the Council on a Framework for the free flow of non-personal data in the European Union*. 2018.

[32] R.N. Patel. *A container-based Approach to Cyber Resilience*. Florida Institute of Technology, 2016.

[33] Grand View Research. *Artificial Intelligence Market Size, Share & Trends Analysis Report By Solution (Hardware, Software, Services), By Technology (Deep Learning, Machine Learning), By End Use, By Region, And Segment Forecasts, 2020 – 2027*. 2020.

[34] I. M. D. et al. Rosa. *Classification success of six machine learning algorithms in radar ornithology*. Vol. 158. 1. Ibis, 2016.

[35] SMD-CID. *Pubblicazione JIC-012 "Le attività militari nello spazio cibernetico(La Cyber-Warfare)"*. 2014.

[36] Heads of State e Government participating in the meeting of the North Atlantic Council in Brussels. *Brusselles Summit Communiqué*. Available on line. 2021. URL: `https://www.nato.int/cps/en/natohq/news_185000.htm`.

[37] Frost & Sullivan. *2017 Global Information Security Workforce Study*. Center for Cyber Safety e Education, 2017.

[38] I. Androutsopoulos V. Metsis e G. Paliouras. *Enron Spam/Ham original dataset*. Available on line. URL: `http://www2.aueb.gr/users/ion/data/enron-spam/readme.txt`.

[39] I. Androutsopoulos V. Metsis e G. Paliouras. *Spam Filtering with Naive Bayes - Which Naive Bayes?* Available on line. 2006. URL: `https://nes.aueb.gr/ipl/nlp/pubs/ceas2006_paper.pdf`.