

MODULO 04

Threat Intelligence e AI

NLP per l'estrazione di indicatori e tecniche da fonti non strutturate, arricchimento automatico e mappatura su MITRE ATT&CK, con i limiti dell'estrazione automatica.

STIX / TAXII

MISP / OpenCTI

NLP

Abstract e obiettivi formativi

Abstract

Il modulo tratta l'impiego del Natural Language Processing per l'estrazione automatica di indicatori di compromissione e tecniche da fonti non strutturate, e il loro arricchimento in piattaforme di threat intelligence (MISP, OpenCTI) secondo gli standard STIX/TAXII.

Obiettivi formativi (Bloom)

- **Comprendere:** distinguere intelligence strategica, operativa e tattica
- **Applicare:** estrarre IoC e TTP da report testuali con strumenti NLP
- **Valutare:** giudicare qualità, tempestività e attendibilità delle fonti

Traccia core

01

02

07

Detection & Response

03

04

05

06

o

Digital Forensics

08

09

10

11

o

Sicurezza dell'AI

06

12

13

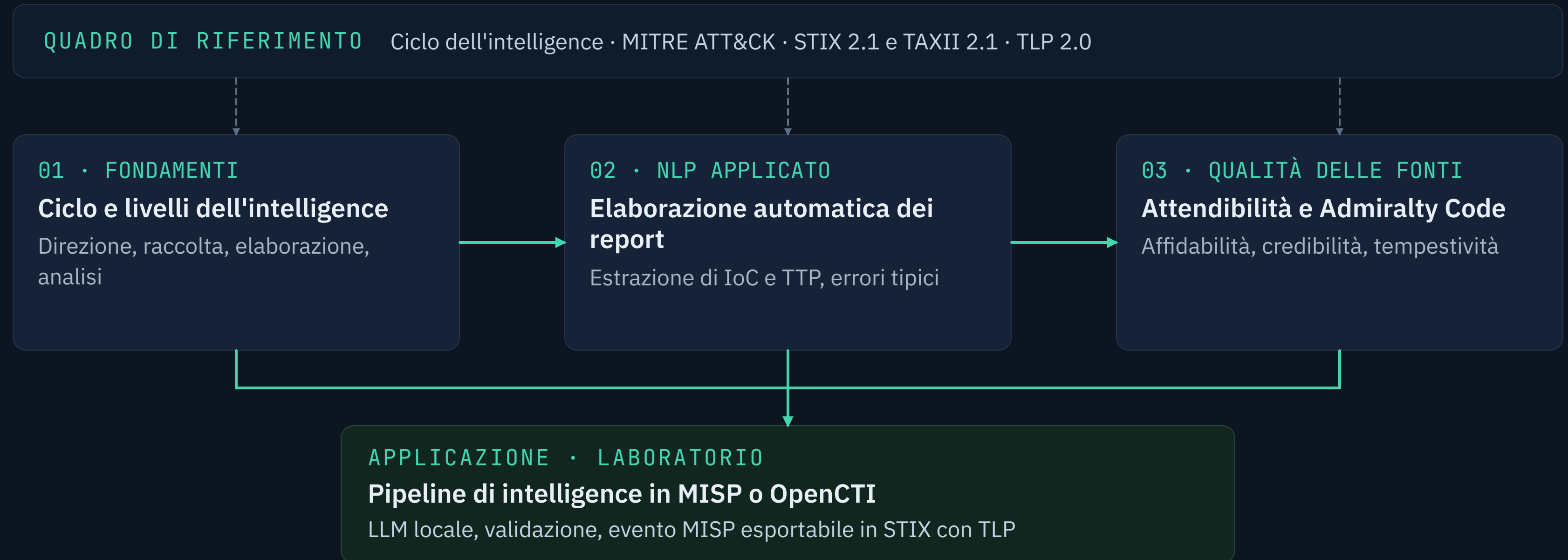
14

Traccia core

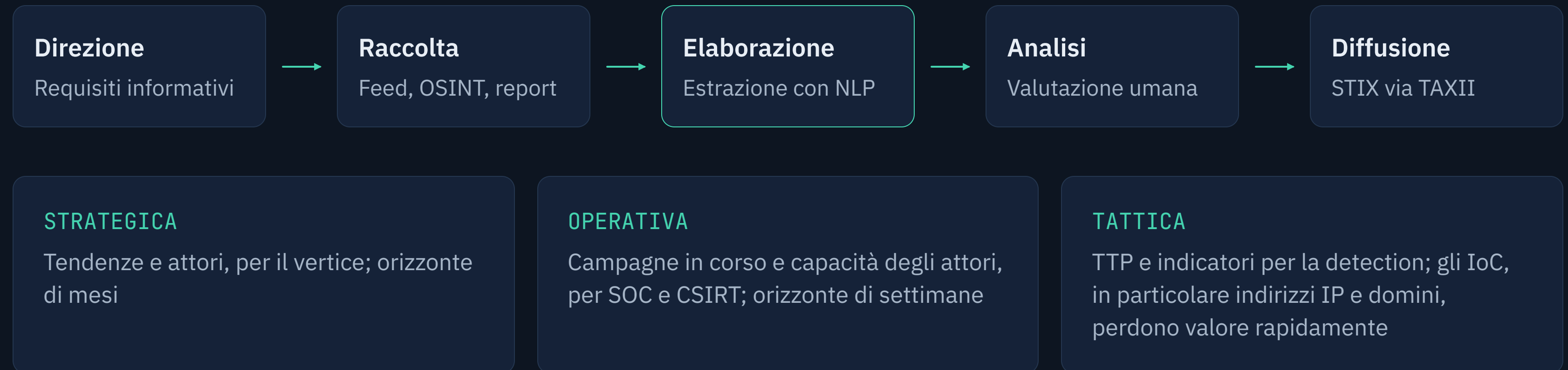
15

16

Il processo di threat intelligence



Il ciclo dell'intelligence e i tre livelli



Elaborazione automatica dei report

Attività automatizzabili

- Espressioni regolari per gli IoC (hash, domini, IP, CVE); named entity recognition per attori, malware, strumenti
- Mappatura del testo su tecniche ATT&CK (es. progetto TRAM)
- Sintesi e deduplicazione di report sovrapposti
- Arricchimento: VirusTotal, OTX, feed commerciali

Errori tipici

- Ambiguità: indirizzi citati come esempio scambiati per indicatori
- Allucinazione di tecniche non descritte nel testo
- Defanging non riconosciuto (hxxp, [.]) o mal ricostruito
- Poisoning dei feed: un indicatore falso contamina la detection a valle

Regola operativa: l'estrazione automatica produce **candidati**; la pubblicazione in piattaforma richiede validazione umana e marcatura della confidenza.

Attendibilità delle fonti e Admiralty Code

Affidabilità della fonte (A–F)

Da A (completamente affidabile) a E (inaffidabile); F quando l'affidabilità non è valutabile. Si giudica la **storia** della fonte, non il singolo report.

Credibilità dell'informazione (1–6)

Da 1 (confermata da altre fonti) a 5 (improbabile); 6 quando la veridicità non è valutabile. Un indicatore valutato B2 è più attendibile di uno valutato D4.

Criteri complementari

- **Tempestività:** un IoC scaduto genera solo rumore
- **Rilevanza:** attinenza al proprio settore e threat model
- **Completezza:** contesto e TTP oltre alle liste di indicatori
- **Tracciabilità:** provenienza documentata, condivisione regolata dal TLP

Interpretare e condividere l'intelligence



Indicatori e TTP nella piramide del dolore

La piramide del dolore, proposta da David Bianco nel 2013 e rivista nel 2014 con l'aggiunta del livello dei valori di hash, ordina gli indicatori secondo il costo che il loro rilevamento impone all'attaccante. Alla base si trovano i valori di **hash** e gli indirizzi IP, che l'attaccante può cambiare in pochi secondi. Seguono i nomi di dominio e gli artefatti di rete e di host, la cui sostituzione richiede più lavoro, e in cima stanno gli strumenti e le **TTP** (tattiche, tecniche e procedure), cioè il modo in cui l'avversario opera, che può cambiare solo riprogettando la campagna. L'investimento in detection rende di più ai livelli alti, dove il valore persiste nel tempo.

Esempio

Il blocco dell'indirizzo IP del server di comando e controllo in uso oggi interrompe la campagna per poco tempo. Una detection sulla tecnica, per esempio l'esecuzione di DLL da percorsi utente tramite rundll32, intercetta anche le varianti successive che continuano a usarla.

Ruolo dell'AI

Gli IoC si estraggono con espressioni regolari, mentre le TTP sono descritte nella prosa dei report. Su questa parte l'estrazione assistita, che mappa il testo sulle tecniche ATT&CK, offre il contributo maggiore e commette più errori.

STIX, TAXII e TLP per la condivisione

STIX 2.1 è il formato: oggetti JSON tipizzati (indicator, malware, threat-actor, attack-pattern, campaign) collegati da relazioni esplicite, per esempio un indicatore che *indica* un malware *usato da* un attore. **TAXII** è il protocollo di trasporto con cui le piattaforme inviano e richiedono oggetti STIX tramite raccolte (collection) esposte da un server. Il Traffic Light Protocol (TLP) indica i limiti di ridistribuzione dell'informazione, da TLP:RED, riservato ai destinatari diretti, a TLP:CLEAR, che ammette la diffusione pubblica; chi riceve è tenuto a rispettarne i limiti.

Scenario didattico

Un CSIRT nazionale pubblica in una raccolta TAXII un bundle STIX marcato TLP:AMBER su una campagna contro il settore sanitario. Le strutture autorizzate lo scaricano e lo importano in MISP, e le relazioni tra attore, tecniche e indicatori arrivano già strutturate per la detection.

Metodo

Le relazioni trasformano un elenco di indicatori in conoscenza interrogabile: la richiesta «tutte le tecniche osservate per l'attore X negli ultimi sei mesi» è una query sul grafo, impossibile su un feed di soli indicatori.

Errori tipici dell'estrazione automatica

Si consideri un report che scrive: «l'attaccante ha usato un dominio simile a update-vendor[.]example; per il test interno abbiamo riprodotto il traffico verso 192.168.10.5». Un estrattore che non analizza il contesto ne ricava due indicatori errati: il dominio ricostruito dalla forma defanged, che potrebbe essere legittimo o già neutralizzato con un sinkhole, e un indirizzo privato (RFC 1918) dell'ambiente di test dell'autore. Una volta pubblicati in piattaforma, il dominio genera blocchi errati e l'indirizzo privato produce falsi riscontri interni nel SIEM.

Metodo

Prima della pubblicazione si verifica il contesto della frase, distinguendo l'attacco osservato dall'esempio, si ricostruisce e si controlla il defanging e si escludono intervalli di indirizzi riservati e domini infrastrutturali. Si considera poi la data del report e si assegnano confidenza e TLP.

Cautele

L'output dell'estrazione automatica entra in piattaforma come bozza non pubblicata; la promozione a evento condivisibile è un atto umano registrato. Un indicatore falso condiviso compromette la credibilità della fonte e la detection di tutti i destinatari.

Livelli dell'intelligence e destinatari

L'espressione threat intelligence designa prodotti diversi, ciascuno destinato a un lettore. Il rapporto MWR del 2015 distingue quattro livelli. Il livello **strategico**, fatto di tendenze, attori e contesto geopolitico, serve al vertice per le decisioni di investimento e di rischio. Il livello operativo orienta le priorità del SOC e del CSIRT e riguarda, nel rapporto, attacchi specifici e imminenti, in classificazioni successive le campagne in corso e le capacità degli attori. Le TTP mappate su ATT&CK formano il livello tattico, base del detection engineering (Modulo 3), mentre il livello **tecnico**, con IoC come hash, indirizzi IP e domini, alimenta direttamente i controlli ed è il più deperibile.

Scenario didattico

La stessa campagna ransomware produce una nota di una pagina per il vertice, con settore colpito, impatto e stato delle difese, e un briefing operativo per il SOC sui vettori iniziali osservati e sui tempi tipici tra accesso e cifratura. Ne derivano anche una mappa delle TTP per i detection engineer e un pacchetto STIX di indicatori per il blocco automatico.

Ruolo dell'AI

La trasposizione tra livelli si presta all'uso di LLM, per esempio dal report tecnico alla nota per il vertice. Il rischio è che la sintesi perda le espressioni di confidenza dell'analista, come «valutiamo con bassa confidenza che...», e presenti come certe conclusioni probabilistiche.

Attribuzione e livelli di confidenza

L'attribuzione procede per gradi di difficoltà crescente. Raggruppare attività in **cluster** coerenti per strumenti, infrastrutture e orari è relativamente solido, mentre collegare un cluster a un attore nominato è già più incerto. Per arrivare a un'organizzazione o a uno Stato occorre combinare evidenze tecniche, contesto e spesso informazioni non disponibili ai privati. Ogni indizio può essere falsificato, perché codice e infrastrutture si riusano o si condividono, gli orari di attività si possono simulare e le **false flag** sono documentate. L'analisi usa quindi un linguaggio di confidenza calibrato, con i gradi bassa, media e alta, e separa l'osservazione, come «il loader condivide il 60% del codice con X», dall'inferenza, come «riteniamo con confidenza media che...».

Caso documentato

Olympic Destroyer (2018), il malware usato contro l'organizzazione dei Giochi invernali di PyeongChang, conteneva artefatti che rimandavano a più attori statali, inseriti come false flag. Alcune ipotesi iniziali, basate su singoli indizi, si rivelarono fuorvianti; nel 2020 il Dipartimento di Giustizia statunitense ha incriminato ufficiali del GRU russo per questa e altre operazioni.

Metodo

Per la risposta aziendale serve soprattutto il profilo operativo dell'avversario: ransomware opportunistico o attore mirato, probabili azioni successive. L'attribuzione a uno Stato ha implicazioni legali e politiche; nelle comunicazioni ufficiali l'organizzazione riporta i fatti tecnici e lascia l'attribuzione alle autorità competenti.

Condivisione dell'intelligence tra organizzazioni

Nessuna organizzazione osserva abbastanza attacchi da costruire da sola un quadro completo della minaccia, e la condivisione ne amplia la capacità di difesa. Le strutture principali sono gli **ISAC** settoriali e i CSIRT nazionali, in Italia il CSIRT Italia presso l'ACN, che operano come comunità di fiducia. MISP è una piattaforma diffusa per la condivisione, con eventi, attributi, tag e sincronizzazione tra istanze. Per chi condivide i costi sono il tempo di cura dei dati, il rischio reputazionale e il timore di rivelare una compromissione, mentre il beneficio va soprattutto agli altri. Le comunità efficaci riducono questo costo con modelli predefiniti, anonimizzazione e automazione, e chiedono uno scambio nei due sensi.

Scenario didattico

Una banca subisce un tentativo di frode con un kit nuovo e condivide sull'ISAC di settore, in TLP:AMBER e in forma anonimizzata, indicatori e TTP entro 24 ore. Altri tre istituti trovano gli stessi indicatori nei propri log prima di subire danni.

Metodo

Un feed si valuta per tempestività, tasso di falsi positivi sul traffico della propria organizzazione, contesto allegato a ogni indicatore e sovrapposizione con le fonti già disponibili. L'AI aiuta a deduplicare e arricchire; un numero maggiore di indicatori comporta anche più lavoro di verifica.

Pipeline di intelligence in MISP o OpenCTI

Svolgimento

1. Ingestione di un report APT pubblico nella piattaforma
2. Estrazione assistita di IoC e tecniche con LLM locale (Ollama)
3. Validazione manuale: confidenza, defanging, falsi indicatori
4. Pubblicazione di un evento MISP (o di un report OpenCTI) esportabile come bundle STIX 2.1, con marcatura TLP

Consegna

- Evento o report pubblicato con IoC validati
- Tabella degli errori dell'estrazione automatica
- Mappatura ATT&CK delle tecniche del report

Lecture e autovalutazione

Lecture obbligatorie

- MITRE, *ATT&CK Design and Philosophy*
- Specifica STIX 2.1 (OASIS)
- Un report APT recente a scelta (Mandiant, ESET o CISA Advisory)

Domande di autovalutazione

1. Quali rischi comporta l'ingestione automatica di feed non verificati?
2. Che differenza c'è tra IoC e TTP in termini di persistenza nel tempo?
3. Come si valuta l'attendibilità di una fonte con l'Admiralty Code?
4. Quali errori tipici commette un LLM nell'estrazione di IoC?
5. A cosa servono STIX e TAXII?

FINE DELLA LEZIONE · MODULO 04

Grazie

Threat Intelligence e AI · Corso «Incident Response potenziata dall'AI»

Domande e materiali: ricevimento del docente. Profilo e contatti:
vincenzocalabro.it



Vincenzo Calabrò

CYBER SECURITY · DIGITAL FORENSICS



VINCENZOCALABRO.IT