MODULO 05

SOAR e automazione della risposta

Orchestrazione, playbook adattivi e controllo umano significativo: criteri per automatizzare il contenimento e rischi introdotti dall'automazione.

SOAR | HUMAN-IN-THE-LOOP | TheHive / Shuffle

SCHEDA DEL MODULO

Abstract e obiettivi formativi

Abstract

Il modulo esamina l'orchestrazione della risposta tramite piattaforme SOAR: architetture e integrazioni API, playbook statici e adattivi, azioni di contenimento, gestione degli errori e rollback, fino al principio di controllo umano significativo e all'audit trail delle decisioni automatiche.

Obiettivi formativi (Bloom)

- **Comprendere**: classificare i livelli di autonomia (human-in/on/out-of-the-loop)
- **Creare**: progettare un playbook SOAR con punti di approvazione umana
- **Valutare**: giudicare costi e rischi dell'automazione di azioni irreversibili

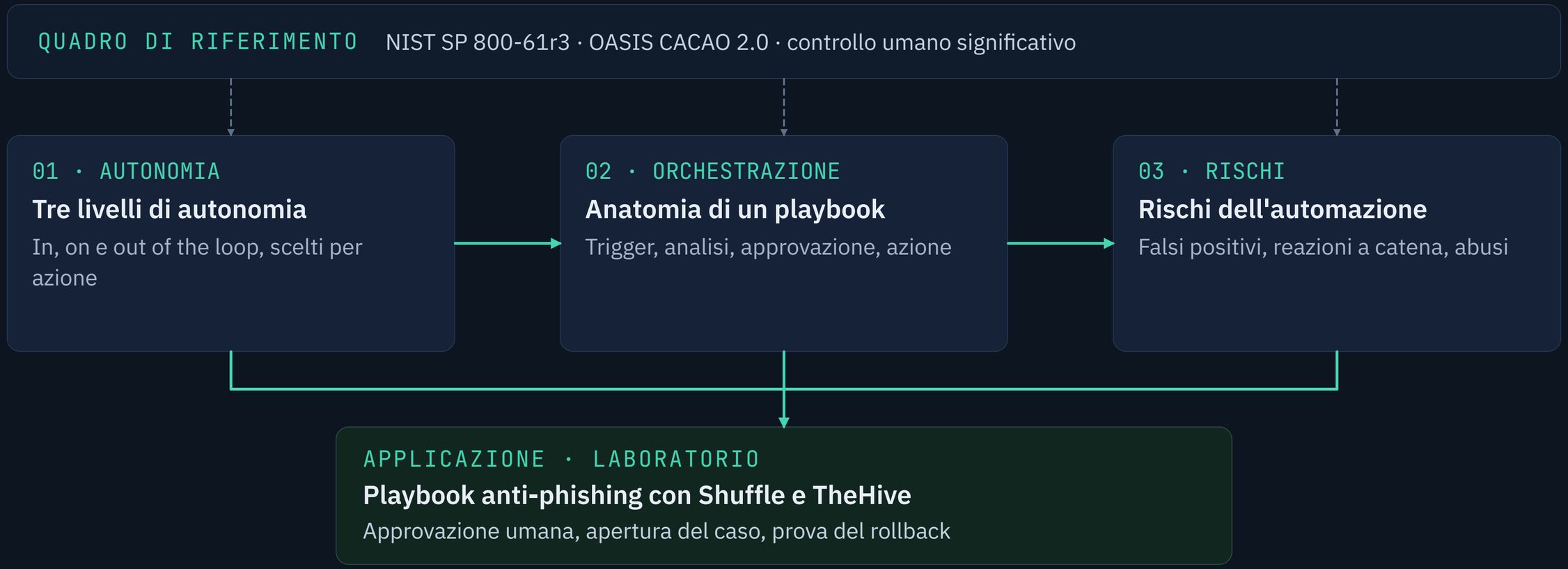
Traccia core: 01 02 07

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

SCHEMA · TEORIA

Automatizzare la risposta sotto controllo umano



TEORIA · LIVELLI DI AUTONOMIA

Tre livelli di autonomia nella risposta

IN-THE-LOOP

L'umano approva

Ogni azione attende un'approvazione esplicita; adatto a contenimenti ad alto impatto o irreversibili

ON-THE-LOOP

L'umano supervisiona

Il sistema agisce e l'operatore osserva e può interrompere; adatto ad azioni reversibili e urgenti

OUT-OF-THE-LOOP

Piena autonomia

Nessun intervento umano; ammissibile solo per azioni a basso rischio con rollback garantito

Il livello di autonomia si sceglie **per ciascuna azione**: lo stesso playbook può arricchire gli indicatori in autonomia e attendere l'approvazione per il blocco di un account.

TEORIA · ORCHESTRAZIONE

Anatomia di un playbook di risposta al phishing

TRIGGER
Segnalazione utente o alert mail gateway

→

ANALISI
Detonazione allegato in sandbox, verifica URL

→

ARRICCHIMENTO
Reputazione IoC, ricerca di altri destinatari

→

APPROVAZIONE
Decisione umana su blocco del mittente e quarantena

→

AZIONE
Quarantena, blocco, ticket e notifica, con audit trail

Requisiti di ingegneria

I passi gestiscono gli errori e il rollback è definito prima della messa in esercizio. Ogni decisione, automatica o umana, viene registrata con timestamp e motivazione in un audit trail che ha rilevanza giuridica oltre che operativa.

TEORIA · RISCHI

Rischi dell'automazione della risposta

Modi di fallimento

- Contenimento errato su falso positivo: servizio legittimo fermo
- Reazioni a catena tra playbook che si innescano a vicenda
- Automation bias: l'operatore ratifica senza verificare
- Abuso da parte dell'attaccante, che provoca blocchi automatici per causare un disservizio

Mitigazioni

- Approvazione umana sulle azioni irreversibili o ad alto impatto
- Rollback progettato e testato prima del rilascio
- Rate limiting e circuit breaker tra playbook
- Metriche che includano tempo risparmiato, errori evitati ed errori causati

SCHEMA · APPROFONDIMENTI

Criteri di progetto dei playbook

APPROFONDIMENTO 1

Reversibilità delle azioni e livello di autonomia

Un'azione di risposta è **reversibile** quando esiste una procedura nota, verificata e rapida per annullarne gli effetti senza perdite. La reversibilità dipende dal **contesto** oltre che dall'azione: isolare una workstation d'ufficio è un'operazione reversibile, mentre isolare la stessa macchina quando controlla una linea di produzione può interrompere un processo fisico. Per ogni azione del playbook occorre stabilire come si annulla, in quanto tempo e quali effetti produce nel frattempo; la risposta determina il livello di autonomia ammesso.

Esempio

Arricchire un indicatore con interrogazioni passive su fonti approvate non ha effetti sui sistemi e può avvenire in autonomia; l'invio di file o URL a servizi esterni, che può divulgare dati riservati o allertare l'attaccante, segue la policy sui dati in uscita (Modulo 6). La quarantena di un'e-mail si annulla con un'operazione: autonomia con notifica. Il blocco di un account è reversibile ma ha impatto e richiede approvazione. La revoca di certificati o la cancellazione remota di un dispositivo sono di fatto irreversibili e richiedono una decisione umana.

Metodo

Prima di automatizzare un'azione si simula il suo innesco su un falso positivo nel momento più sfavorevole, per esempio sull'account del direttore finanziario durante la chiusura di bilancio. Se il danno risultante è inaccettabile, l'azione resta subordinata all'approvazione.

APPROFONDIMENTO 2

Automation bias nei punti di approvazione

L'**automation bias** è la tendenza, osservata in aviazione, medicina e controllo di processo, ad affidarsi all'indicazione di un sistema automatico anche quando altre informazioni disponibili la contraddicono. Nel SOC si manifesta nei punti di approvazione insieme alla *complacency*, il calo della vigilanza che Parasuraman e Manzey descrivono come fenomeno distinto ma sovrapposto. Dopo molte proposte corrette, l'operatore tende ad approvare le successive senza esaminarle, e il controllo umano previsto dal processo si riduce a una formalità.

Scenario didattico

Il playbook propone il blocco di un mittente classificato come malevolo; l'analista, al quarantesimo caso della giornata, approva in pochi secondi. Il dominio apparteneva a un fornitore segnalato per errore, e il flusso degli ordini si interrompe per alcune ore.

Contromisure

All'approvatore si mostrano le evidenze insieme alla raccomandazione, e per le approvazioni ad alto impatto si richiede una motivazione scritta. Si inseriscono inoltre controlli a campione con casi di prova noti. Il tempo medio di approvazione misura l'effetto di queste contromisure, perché tempi molto brevi e costanti indicano un controllo inefficace.

APPROFONDIMENTO 3

Progettazione dell'audit trail

L'audit trail si progetta insieme al playbook e si verifica durante i test. Con la sola annotazione «azione eseguita» una decisione non si ricostruisce né si giustifica. Per ogni azione il trail registra che cosa è stato fatto e quando, con orologi sincronizzati, e su quale base: alert, evidenze, versione del playbook. Registra inoltre chi ha deciso, il sistema in autonomia oppure un operatore identificato, e se l'azione è riuscita, è fallita o è stata annullata con un rollback.

Scenario didattico

Sei mesi dopo un blocco di account contestato da un dipendente, il trail mostra l'alert con il relativo punteggio, le evidenze allegate, l'approvazione dell'analista con la motivazione e il rollback eseguito in giornata. La decisione si ricostruisce e si può motivare davanti all'organizzazione o in giudizio.

Cautele

Il trail deve essere a sola aggiunta (append-only) e protetto da modifiche anche da parte degli amministratori della piattaforma. Il suo valore probatorio dipende dalla coerenza dei timestamp tra i sistemi, e i tempi di conservazione seguono gli obblighi applicabili e la policy interna.

APPROFONDIMENTO 4

Metriche per valutare l'automazione

Il numero di playbook automatizzati è un dato di inventario. Per valutare l'automazione si confronta il **MTTR per classe di incidente** prima e dopo, su casi dello stesso tipo, e si misura senza ricorrere a stime il tempo di lavoro risparmiato dagli analisti per ogni esecuzione. Il **tasso di override** indica quanto spesso l'operatore corregge la proposta dell'automazione, e gli errori causati dall'automazione si contano con lo stesso rigore degli errori evitati. Un tasso di override vicino a zero va verificato: può indicare un sistema accurato oppure un controllo umano non più esercitato (approfondimento 2).

Scenario didattico

Si supponga che, dopo sei mesi, il playbook anti-phishing di un SOC mostri un MTTR passato da circa tre ore a poco più di venti minuti, un tasso di override del 6% concentrato sulla decisione di blocco del mittente e due errori causati, entrambi annullati entro un'ora. Su questi dati si può discutere l'eventuale aumento di autonomia per il blocco.

Limiti

Una metrica di velocità usata da sola favorisce chiusure affrettate e automatiche. Va affiancata da indicatori di qualità, come il tasso di riapertura dei casi e il numero di escalation mancate.

APPROFONDIMENTO 5

Progettazione del punto di approvazione

Un'approvazione è efficace se chi decide dispone del contesto: che cosa è accaduto, che cosa farà l'azione, quali persone e sistemi coinvolge, quale rischio comporta non agire, con le evidenze consultabili. Il progetto stabilisce inoltre i ruoli autorizzati ad approvare in base all'impatto, la registrazione di decisione, motivazione e orario, e il **timeout**, cioè che cosa accade se nessuno risponde: escalation, azione conservativa predefinita o nessuna azione.

Scenario didattico

La richiesta di blocco di un account arriva al canale di reperibilità come scheda con l'utente, i segnali che l'hanno generata, i sistemi raggiungibili dall'account, il rollback previsto e due pulsanti con le conseguenze descritte. Dopo dieci minuti senza risposta la richiesta passa al turno senior; se nessuno risponde si revocano soltanto le sessioni attive.

Errori frequenti

Approvazioni in serie senza contesto e silenzio interpretato come assenso. Un unico approvatore possibile diventa collo di bottiglia e punto singolo di guasto. Il disegno dell'interfaccia di approvazione è una delle contromisure all'automation bias (approfondimento 2).

APPROFONDIMENTO 6

Acquisizione automatica delle evidenze

L'automazione serve anche a preservare le evidenze. Le più volatili (Modulo 7) si perdono nei minuti in cui il gruppo di risposta si sta ancora organizzando. Un playbook con acquisizione iniziale integrata raccoglie all'innesco, di norma prima del contenimento, l'immagine della memoria o del processo sospetto, un pacchetto di triage dell'host con processi, connessioni, voci di avvio automatico e log principali, e l'esportazione dei log cloud rilevanti. Ogni elemento viene depositato con hash e timestamp in un archivio non modificabile, e l'accaduto resta ricostruibile anche dopo che il contenimento ha alterato lo stato dei sistemi.

Scenario didattico

Alle 04:12 un alert EDR segnala un server. Il playbook acquisisce l'immagine del processo, il pacchetto di triage e l'elenco delle connessioni, calcola gli hash e deposita i file nell'archivio collegato al caso; poi isola l'host. Quando l'analista si collega, il contenimento è avvenuto e le evidenze di quel momento sono disponibili.

Cautele

L'ordine tra acquisizione e contenimento si stabilisce per classe di incidente: con un'attività distruttiva in corso il contenimento può precedere l'acquisizione della memoria, che si completa dopo l'isolamento tramite EDR. Strumenti, versioni e orari dell'acquisizione si registrano nel log del playbook (Modulo 7).

LABORATORIO · ESERCITAZIONE

Playbook anti-phishing con Shuffle e TheHive

Svolgimento

1. Costruzione del flusso: trigger, analisi con gli analyzer di Cortex, arricchimento degli IoC
2. Inserimento del punto di approvazione umana per il blocco del mittente
3. Apertura automatica del caso in TheHive con gli osservabili
4. Test su e-mail simulate, benigne e malevole, e verifica dell'audit trail

Consegna

- Playbook funzionante ed esportato
- Matrice motivata tra azioni e livelli di autonomia
- Prova del rollback su un'azione di blocco

CHIUSURA DEL MODULO

Letture e autovalutazione

Letture obbligatorie

- NIST SP 800-61r3, *Incident Response Recommendations and Considerations* (2025)
- Parasuraman e Manzey, *Complacency and Bias in Human Use of Automation* (2010)
- OASIS, *CACAO Security Playbooks 2.0*; documentazione di TheHive, Cortex e Shuffle

Domande di autovalutazione

1. Quali azioni di risposta non dovrebbero mai essere pienamente automatizzate, e perché?
2. Cosa distingue human-in-the-loop da human-on-the-loop?
3. Come si progetta il rollback di un contenimento errato?
4. Quali metriche giustificano l'investimento in SOAR?
5. In quali casi l'audit trail è anche un requisito giuridico?

FINE DELLA LEZIONE · MODULO 05

Grazie

SOAR e automazione della risposta · 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