Sistema de Automatización para la Gestión y Respuesta de Incidentes de Seguridad en Cortex XDR mediante Playbooks

López Naranjo Diego Andrés, Sánchez Medina Diego Felipe

Maestría en Seguridad de la Información

Universidad de Los Andes

da.lopezn1@uniandes.edu.co, df.sanchezm1@uniandes.edu.co

***Resumen*— Este artículo presenta una herramienta de automatización para la gestión y respuesta a incidentes de ciberseguridad en organizaciones pequeñas y medianas (PYMES), utilizando Cortex XDR como motor de detección. La solución aborda la falta de personal calificado y equipos SOC en estas organizaciones, integrando playbooks basados en inteligencia artificial para ejecutar acciones automáticas ante incidentes. Se describe una arquitectura modular que conecta Cortex XDR con plataformas externas a través de API, permitiendo la clasificación, priorización y respuesta automatizada en función de la severidad y naturaleza del ataque. Los resultados obtenidos en un entorno de pruebas demuestran una reducción significativa del tiempo medio de respuesta (MTTR), una alta precisión en la clasificación de amenazas y una disminución en los costos asociados a la operación de seguridad. Esta herramienta representa una alternativa eficaz y asequible para mejorar la ciber resiliencia institucional mediante automatización, siendo especialmente beneficiosa en contextos donde el personal especializado es limitado o inexistente. [...]**

Palabras clave—Cortex XDR, Automatización, Ciberseguridad, Playbooks, Respuesta, Incidentes, MTTR.

I. INTRODUCCIÓN

Hoy en día, las empresas pequeñas y medianas enfrentan grandes retos en la gestión de sus productos de ciberseguridad debido a la escasez de personal calificado y la saturación del mercado con diversas herramientas. Según ISACA [1], solo el 38% de las empresas a nivel mundial considera que su equipo de ciberseguridad está suficientemente capacitado para responder a incidentes, lo que genera preocupación ante el aumento de amenazas. Además, la demanda de habilidades para la respuesta de incidentes ha aumentado en un 44% para 2024 [1].

Como consecuencia, las empresas pequeñas y medianas son uno de los que más afrontan retos en el 2025, según el Informe de Investigación de Brechas de Datos 2024 de Verizon [3], se analizaron 30,458 incidentes de seguridad, de los cuales 10,626 resultaron en brechas de datos confirmadas. Estos resultados posicionan a las PYMES como una de las industrias más afectadas a nivel mundial en términos de ciberseguridad. lo que resalta la necesidad de contar con servicios o productos que aborden esta problemática en empresas de tamaño mediano y pequeño.

El uso de playbooks y scripts diseñados para la automatización de respuestas y gestión de incidentes ha ganado popularidad en los últimos años. Se estima que, en 2024, la automatización mediante herramientas de inteligencia artificial (IA) representará más del 23% en la creación de tareas que cubran la brecha generada por la falta de personal capacitado en ciberseguridad [1]. Esto representa una gran oportunidad para el desarrollo de herramientas basadas en IA en entornos sin un equipo SOC (Security Operations Center), lo que podría reducir costos en la contratación de personal especializado o en la adquisición de servicios externos para la gestión y respuesta a incidentes.

Los costos mensuales que genera un SOC pueden rondar (por dispositivo de punto final) los $45 USD donde solo se realiza monitoreo y alertamiento, un costo promedio de $73 USD para servicios de caza de amenazas y respuesta, y un costo promedio de $93 USD con servicio de 24 horas por 7 días a la semana, según el reporte de MSSP Alert [4]. Lo que genera un costo mensual para 200 dispositivos (Cantidad mínima de licencia de cortex) de aproximadamente entre $10.000 y $18.600 USD, lo que es muy costoso de mantener para empresas pequeñas y medianas.

Esta tendencia se confirma en el informe de ISACA [1] sobre el estado de la ciberseguridad, donde el uso de IA es más común en la detección y respuesta de amenazas (28%) y en la seguridad

de dispositivos de punto final (27%), siendo estas las principales aplicaciones de la inteligencia artificial en las operaciones de seguridad.

El uso de estas herramientas es un factor clave para las compañías que gestionan múltiples productos de ciberseguridad. Al integrarse con IA mediante playbooks, se facilita una detección y respuesta más rápida ante amenazas, garantizando así la disponibilidad, integridad y confidencialidad de la información [2].

II. DEFINICIÓN DEL PROBLEMA

Cortex XDR de Palo Alto Networks es una herramienta avanzada de detección y respuesta ante amenazas que clasifica los eventos de seguridad según su nivel de riesgo. Sin embargo, su implementación aún requiere intervención humana para analizar y dar una resolución adecuada para cada incidente detectado y generar una respuesta adicional si aplica. El desafío se agrava en pequeñas empresas que adoptan esta solución sin contar con un equipo de respuesta a incidentes (CSIRT) o personal especializado en ciberseguridad. Como resultado, muchas organizaciones desconocen cómo reaccionar ante los diferentes eventos de seguridad que Cortex XDR identifica, lo que puede llevar a una gestión ineficiente de las amenazas o incluso a la omisión de respuestas críticas, aumentando el riesgo de brechas de seguridad.

III. PROPUESTA DE SOLUCIÓN

1) Objetivo General

Crear una herramienta de automatización para la gestión y respuesta de incidentes de ciberseguridad en Cortex XDR, que integre, a través de playbooks con inteligencia artificial, las diferentes soluciones de seguridad de una empresa mediana o pequeña mediante API. Esta herramienta permitirá tomar acciones basadas en la severidad de los incidentes registrados, con el propósito de reducir significativamente la intervención humana en su resolución, disminuir el tiempo medio de respuesta (MTTR) y garantizar la resolución de incidentes críticos en menos de 6 horas y de alta severidad en menos de 12 horas, asegurando que los playbooks ejecuten respuestas apropiadas en las demás herramientas de ciberseguridad de la organización.

2) Nombre de la herramienta

La herramienta propuesta se denomina AUTIREX, un acrónimo que proviene de las siglas en inglés AUTomated Incident REsponse for Cortex XDR. Este nombre refleja su propósito principal: ofrecer una solución automatizada y eficiente para la respuesta ante incidentes de ciberseguridad detectados por la plataforma Cortex XDR. AUTIREX funciona como un puente inteligente entre los analistas de seguridad y las diferentes soluciones de protección, permitiendo reducir significativamente los tiempos de respuesta mediante la ejecución de playbooks automatizados y la integración con otras herramientas de seguridad mediante API.

3) Objetivos Específicos

- Identificar y analizar los componentes de seguridad de la solución, evaluando las capacidades de Cortex XDR en la detección y respuesta ante amenazas, así como las herramientas de seguridad complementarias utilizadas en empresas medianas y pequeñas.
- Determinar qué procesos pueden ser automatizados, analizando los incidentes que actualmente requieren intervención manual y clasificándolos según su impacto en la gestión de seguridad.
- Diseñar la arquitectura de la solución de automatización, estableciendo el modelo de integración de Cortex XDR con otras herramientas de ciberseguridad mediante API.
- Especificar las acciones de respuesta automática, desarrollando un conjunto de playbooks de automatización para la mitigación de incidentes de seguridad.
- Realizar pruebas en un entorno de simulación, ejecutando incidentes de seguridad controlados para evaluar la efectividad de los playbooks por medio de la medición de tiempos de respuesta, porcentaje de incidentes resueltos sin intervención humana y precisión de la clasificación de amenazas.
- Documentar la solución y generar guías de uso, elaborando un manual técnico que detalle los procesos automatizados, las integraciones implementadas y las mejores prácticas para su despliegue en empresas sin un equipo SOC.

4) Requerimientos no funcionales

- Garantizar la seguridad en las comunicaciones entre Cortex XDR y otras herramientas mediante cifrado y autenticación API.
- Mejorar el rendimiento para procesar incidentes en tiempo real sin afectar la operatividad de la empresa.
- Diseñar una interfaz de administración sencilla y accesible para la configuración de playbooks.
- Garantizar la confiabilidad del sistema minimizando las falsas alarmas y respuestas incorrectas, manteniendo la tasa de falsos positivos por debajo del 15%. Este porcentaje se calcula de la siguiente manera:

$$\text{Total de falsos positivos (\%)} = \left(\frac{\text{Número de falsos positivos}}{\text{Número total de alertas o incidentes analizados}} \right) * 100$$

- Cumplir la normativa y legislación de la Ley 1928 de 2018 para el cumplimiento de la protección de datos personales y la norma ISO27001 que deben acatar las entidades en Colombia
- Cifrado de comunicaciones entre componentes del sistema.
- Cifrado de almacenamiento de datos sensibles.

5) Requerimientos funcionales

a. Autenticación y autorización

- Implementación de página de inicio de sesión utilizando correo y contraseña.
- Implementación de página de registro solicitando nombre, correo, contraseña y confirmación de contraseña.
- No permitir la creación de cuentas duplicadas con el mismo correo.
- Funcionalidad para cerrar sesión.
- Definición y gestión de roles de acceso:
    - Admin: puede crear, modificar y eliminar organizaciones e integraciones; tiene acceso completo a la API, pero no puede visualizar alertas ni incidentes de organizaciones.
    - Manager (asociado a una organización): puede gestionar configuraciones de integraciones y miembros de su organización.
    - Operator (asociado a una organización): puede visualizar incidentes y alertas.
- Autenticación y autorización mediante IAM o Single Sign-On (SSO).
- Control de acceso basado en roles y permisos (RBAC).

b. Gestión de Organizaciones

- Página de administración para cargar información de la organización (nombre, lista de miembros e integraciones).
- Añadir y eliminar miembros de la organización.
- Asignar roles de "Manager" a los miembros.
- Crear, actualizar, probar conexión y eliminar integraciones específicas por organización.
- Definición de valores de configuración para integraciones (API URL, API Key, etc.).

c. Integraciones y Conectividad

- Integración mediante API con:
- Cortex XDR.
- Plataformas de Threat Intelligence (VirusTotal, Hybrid Analysis y AlienVault OTX). Plataforma Open source para efectos de pruebas.
- Directorio Activo de nube (Google Cloud Managed Microsoft AD).
- Firewall Palo Alto Networks (via API).
- Servicio de correo electrónico.
- Interfaz para configurar credenciales, endpoints de API de terceros y probar conectividad.
- Monitoreo del estado de las integraciones con herramientas de terceros.

d. Gestión de Incidentes y Alertas

- Página para listar alertas obtenidas de Cortex XDR en una tabla interactiva.
- Cada alerta debe permitir visualizar acciones disponibles.
- Permitir análisis de hash si la alerta contiene uno.
- Página para listar incidentes obtenidos de Cortex XDR.
- Cada incidente debe incluir enlace directo para visualizarse en la consola de Cortex XDR.

e. Análisis de indicadores de compromiso (IOC)

- Integración con servicios externos (Hybrid Analysis, VirusTotal, OTX AlienVault) para el análisis de hashes maliciosos.
- Mostrar los valores más significativos del análisis en la UI, junto con enlace para visualizar el reporte completo.
- Manejo de errores de análisis y presentación del mensaje de error correspondiente.
- Agregar, modificar y eliminar IOC con widget “Gestor IOC” en el aplicativo.

f. Automatización de Respuesta a Incidentes

- Priorización automática de incidentes detectados en Cortex XDR basada en el "Smart Score", opción disponible en Cortex que da una puntuación de 0 a 100 a incidentes dentro de una severidad por medio de los datos históricos y aprendizaje Machine Learning de Cortex. Utilizando factores como comportamiento anómalo, severidad de la alerta, tácticas e indicadores asociados a MITRE ATT&CK. Utilizando bajo nuestro criterio un riesgo bajo de un score entre 0 a 30, riesgo medio de 31 a 70 y riesgo alto de 71 a 100.
- Activación automática de playbooks de respuesta en función de la severidad y categoría del incidente:
    - Movimiento lateral: Aislamiento automático de endpoints comprometidos para incidentes de severidad Crítica, Alta o Media.

    - Exfiltración, acceso a credenciales, acceso inicial: Bloqueo automático de usuarios en Directorio Activo.
    - Comando y control, descubrimiento: Adición automática de IPs maliciosas a listas de bloqueo de Cortex XDR.
    - Malware: Eliminación automática de archivos maliciosos (Severidad Crítica, Alta o Media).
    - Alertas de baja severidad: Actualizar el estado de la alerta a "resolved_other".
- Interfaz para definir umbrales de automatización basados en severidad.
- Posibilidad de configurar niveles de intervención humana en los playbooks.
- Creación, modificación, prueba y desactivación de playbooks personalizados con botón "Desactivar Playbook" en el aplicativo.
- Reversión de respuesta con botón "Revertir acción" en el aplicativo.
- Simulación de playbooks con botón "Simular".

g. Alertas y Notificaciones

- Generación de alertas en tiempo real cuando un incidente requiere intervención humana.
- Envío de notificaciones automáticas vía:
    - Correo electrónico (Gmail).
    - Slack.
    - Microsoft Teams.
- Integración con sistemas de ticketing como ServiceNow, JIRA y Zendesk para registrar incidentes críticos automáticamente.

h. Reportes y Auditoría

- Generación de reportes tipo log de las acciones realizadas sobre los incidentes y cada uno de los pasos en la ejecución de los playbooks.
- Adjuntar al reporte un registro de todas las acciones automatizadas para fines de auditoría y cumplimiento.
- Exportación de reportes en formatos PDF, CSV y JSON.
- Implementar la integración de un agente de inteligencia artificial, basado en modelos de procesamiento de lenguaje natural, que pueda ser utilizado a través de una API para enriquecer los reportes de incidentes, proporcionando:
    - Explicaciones detalladas sobre la naturaleza del ataque.
    - Justificación de las acciones tomadas.
    - Análisis de tendencias para la detección de patrones y amenazas emergentes.
- Generación de resúmenes automatizados con lenguaje natural, explicando impacto y recomendaciones de mejora.

i. Monitoreo y Logging

- Registro detallado de eventos, acciones automatizadas y errores ocurridos.
- Panel de control que muestre métricas clave, tales como:
    - MTTR (Mean Time to Respond) con límite de tiempo según severidad (Crítica <6 horas, Alta < 12 horas, Media < 24 horas, Baja 2 días). Este valor está dado por:

$$\text{MTTR} = \frac{\text{Tiempo total de resolución}}{\text{Número total de incidentes}}$$

    - Cantidad de incidentes resueltos automáticamente.

j. Dashboard de métricas de efectividad

- Dashboard donde se muestren algunas métricas de efectividad para evaluar el impacto real de la herramienta. Según Scrut Automation, las métricas de respuesta a incidentes más relevantes que una organización debería monitorear para identificar y remediar problemas de seguridad de manera eficaz son:
    - Número de Alertas Generadas: Cantidad de alertas que se generan en un periodo específico (semanal, quincenal o mensual). Esto proporciona una línea base sobre la carga de trabajo del equipo de respuesta a incidentes y permite identificar períodos de aumento o disminución significativa en las alertas.
    - Tiempo Medio de Detección (MTTD): Este KPI mide el tiempo promedio que toma al equipo detectar un incidente de seguridad en la red de la organización. Se calcula sumando el total de tiempo que el equipo tarda en detectar incidentes durante un periodo determinado y dividiéndolo por el número total de incidentes.
    - Tiempo Medio para Acknowledgement (MTTA): Tiempo que transcurre desde que se genera una alerta hasta que un miembro del equipo de respuesta comienza a trabajar en el problema. Un MTTA alto implica que se puede tardar más en comenzar la resolución del incidente.
    - Tiempo Medio de Respuesta/Resolución/Recuperación (MTTR): Este KPI mide el tiempo necesario para diagnosticar y resolver un problema, así como para restaurar los activos afectados. Se calcula dividiendo el tiempo total de inactividad durante un periodo por el número de incidentes reportados.

- Tiempo Medio para Contener (MTTC): El MTTC combina MTTD, MTTA y MTTR para ofrecer una visión integral de la efectividad del equipo de respuesta a incidentes. Este indicador muestra cuánto tiempo toma detectar, reconocer y resolver un incidente de seguridad.
- Tiempo Medio Entre Fallos (MTBF): El MTBF ayuda a medir el tiempo que transcurre entre fallos reparables de un sistema o aplicación. Esta métrica es fundamental para analizar y prevenir la repetición de problemas en el sistema.
- Tiempo Promedio de Respuesta a Incidentes: Este KPI indica la rapidez con la que el equipo asigna responsabilidades y resuelve la amenaza. Si los tiempos de resolución son elevados, es esencial investigar las causas y desarrollar soluciones adecuadas.
- Tasa de Cumplimiento del SLA: Mide el porcentaje de incidentes gestionados dentro de los límites de tiempo establecidos en el acuerdo de nivel de servicio. Monitorear esta métrica es crucial para garantizar que el plan de respuesta a incidentes cumpla con sus objetivos predefinidos.
- Costo por Incidente: Mide el promedio de gastos que la organización incurre para resolver y recuperarse de cada brecha de seguridad. Permite evaluar el impacto financiero de los incidentes y priorizar inversiones que minimicen futuros problemas de seguridad.

6) Requerimientos técnicos

- Implementación del sistema de automatización basado en scripts (PowerShell o Bash según compatibilidad). Ya que se harán pruebas solo en equipos Windows, la mayoría de las organizaciones cuentan con este sistema operativo en sus dispositivos.
- Uso de API REST para la integración entre Cortex XDR y herramientas de terceros. Ya que es un estándar para la integración de múltiples herramientas.
- Despliegue en infraestructura Cloud. Dando cumplimiento a la normativa que se exige al sector de las PYMES al contar con una solución robusta con alta disponibilidad por medio de redundancia y ciframiento de datos para la confidencialidad de la misma.

7) Mitigaciones para evitar abusos en la automatización

- Integrar puntos de revisión en los que ciertas acciones automáticas deban ser aprobadas manualmente, especialmente para incidentes de alta severidad.
- Revisar y actualizar periódicamente los playbooks para asegurarte de que las acciones automáticas sean relevantes y efectivas, y para eliminar acciones obsoletas.
- Mantener un registro de todas las acciones automáticas y manuales realizadas por la herramienta.
- Establecer un proceso para que los usuarios puedan informar problemas o preocupaciones sobre la automatización, contribuyendo a la mejora continua del sistema.

IV. DISEÑO E IMPLEMENTACIÓN

A. Arquitectura Alto Nivel

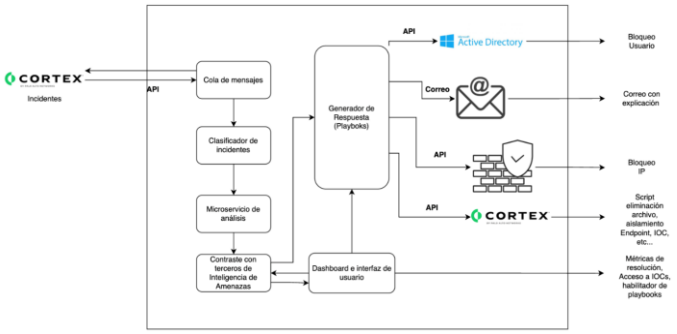

Ilustración 1. Arquitectura de bloques

La solución propuesta recibe incidentes previamente registrados en Cortex XDR, los cuales se integran a la aplicación a través de una conexión API. Una vez dentro, los incidentes son procesados por el módulo de cola de mensajes, donde cada entrada se almacena en una base de datos para su posterior análisis y gestión. Posteriormente, los incidentes se clasifican según su severidad en niveles Crítico, Alto, Medio y Bajo. Luego, se realiza un análisis detallado utilizando un modelo de Machine Learning previamente entrenado, lo que permite determinar un veredicto sobre la amenaza, así como extraer indicadores de compromiso (IOC), usuarios involucrados y dispositivos afectados. Como siguiente paso, los datos obtenidos se contrastan con información de plataformas de inteligencia de amenazas como VirusTotal y Threat Vault a través del módulo de correlación, lo que permite enriquecer el análisis y mejorar la precisión en la identificación de amenazas. Una vez procesados los incidentes, los resultados se envían al módulo de Playbooks, donde, según el veredicto obtenido, se ejecutan respuestas automatizadas. Entre las acciones posibles se incluyen: bloqueo temporal del usuario afectado, envío de correos con la resolución del incidente, bloqueo de IPs maliciosas a nivel de firewall, aislamiento del dispositivo comprometido, entre otras medidas de mitigación. Finalmente, la solución proporciona un dashboard interactivo que permite a

los usuarios monitorear en tiempo real los tiempos de respuesta (MTTR), los IOC generados, así como gestionar la habilitación o des habilitación de Playbooks según las necesidades de la organización.

Dado que AUTIREX tiene un enfoque altamente personalizado, nos diferenciamos de otros productos y servicios en el mercado, como los SOC tradicionales. La herramienta permite una configuración granular en la creación de playbooks y en la integración de sistemas, lo que facilita adaptarla a las necesidades específicas de cada organización y definir sus propios procesos internos. Esto nos permite comprender en profundidad las particularidades del cliente y ajustar el producto en consecuencia, logrando una implementación más eficiente y rápida gracias a la experiencia adquirida con cada despliegue.

Con el tiempo, esta personalización y optimización contribuyen a reducir los costos de implementación, haciendo que nuestro producto sea especialmente atractivo para el sector de ls PYMES, que generalmente cuenta con presupuestos limitados para ciberseguridad. Como se explica más adelante, estimamos un costo aproximado de $10,000 USD, significativamente menor que el promedio actual de $20,000 USD, incluyendo respuestas automatizadas según el reporte de MSSP Alert [4].

Por otro lado, las soluciones SOAR en el mercado, como Cortex XSOAR, también tienen costos elevados, con un mínimo de $125,000 USD anuales solo para la licencia básica, sin incluir servicios de implementación o soporte, según la tabla de precios de IT Price [5]. Nuestra propuesta ofrece una alternativa mucho más accesible en términos de costo y flexibilidad.

B. Diagrama de Despliegue

El siguiente diagrama, representa el despliegue de los componentes de la aplicación a un nivel más bajo, compuesta por diversos componentes distribuidos en microservicios, backend, frontend y servicios externos.

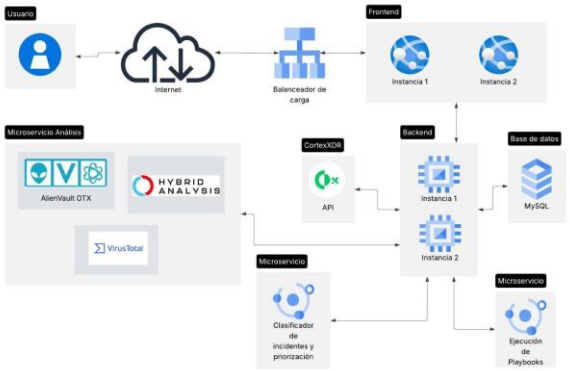

Ilustración 2. Arquitectura de despliegue

Los usuarios acceden a la plataforma a través de internet, siendo dirigidos por un balanceador de carga que distribuye las solicitudes hacia las instancias del frontend. Estas instancias sirven como interfaz de usuario y están conectadas al backend, que también está desplegado en múltiples instancias para asegurar disponibilidad y escalabilidad.

El backend se comunica con diversas fuentes de datos y servicios. Uno de los principales es la API de Cortex XDR, que proporciona la información de incidentes y alertas de seguridad. Esta información se complementa mediante un microservicio de análisis, el cual consulta múltiples fuentes de inteligencia de amenazas como VirusTotal, AlienVault OTX y Hybrid Analysis, permitiendo validar y enriquecer los datos recibidos.

Otro componente esencial es el microservicio de clasificación de incidentes y priorización, encargado de analizar los incidentes detectados y asignarles un nivel de criticidad para una gestión eficiente. Posteriormente, el microservicio de ejecución de playbooks automatiza las respuestas y acciones necesarias para mitigar los incidentes, reduciendo así el tiempo de respuesta y la necesidad de intervención humana. Finalmente, el sistema cuenta con una base de datos MySQL, donde se almacena información relevante para la operación de la plataforma.

C. Descripción de componentes

a. Modelo Entidad-Relación actual de la base de datos

El modelo de base de datos actual permite que múltiples organizaciones gestionen sus propios usuarios, roles e integraciones. La tabla users almacena información básica de los usuarios, como nombre, correo y el hash de la contraseña, y permite identificar administradores globales mediante el campo is_admin. Los usuarios pueden pertenecer a varias organizaciones a través de la tabla user_organization.

Cada organización, representada en la tabla organizations, puede tener usuarios con distintos roles asignados mediante la tabla user_organization_roles, que relaciona usuarios, organizaciones y roles definidos en la tabla roles. Esto permite establecer permisos personalizados por organización.

Las integraciones con servicios externos se modelan en la tabla integrations, y su uso específico por parte de cada organización se gestiona en organization_integrations, donde también se guarda la configuración personalizada en formato JSON.

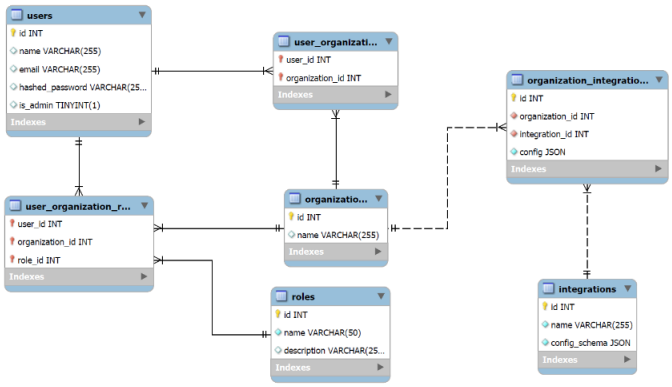

Ilustración 3. Modelo de base de datos

b. Módulos de la interfaz de usuario

- Autenticación y lógica de autorización para usuarios

Se desarrollaron las páginas de autenticación del sistema, incluyendo una página de inicio de sesión que permite a los usuarios acceder usando su correo electrónico y contraseña, utilizando JWT (JSON Web Token) para gestionar la autenticación y autorización de manera segura. La página de registro solicita el nombre, correo electrónico, contraseña y confirmación de contraseña, asegurando que no se puedan crear cuentas duplicadas con el mismo correo. Además, se implementó la funcionalidad de cierre de sesión y la definición de roles de usuario: "Admin", quien puede crear, modificar o eliminar organizaciones e integraciones, pero no visualizar incidentes o alertas de ninguna organización; "Manager", asociado a una organización, que puede modificar configuraciones de integraciones y gestionar miembros; y "Operator", que puede visualizar incidentes y alertas. La asignación del rol de administrador se realiza modificando directamente la base de datos, considerando que solo se requiere un usuario con este rol.

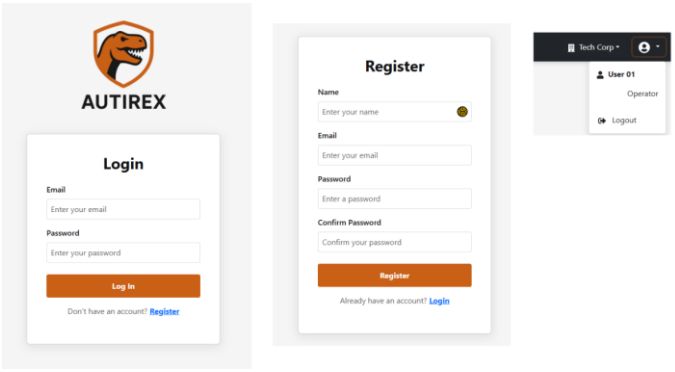

Ilustración 4. Formularios de registro e ingreso

- Administración de Organizaciones

Se creó la página de gestión de organizaciones, donde se carga la información actual de la organización seleccionada, incluyendo el nombre, la lista de miembros y las integraciones configuradas. Desde esta sección, los usuarios pueden añadir o eliminar miembros, asignarles el rol de "Manager", y gestionar integraciones de la organización, definiendo parámetros como la URL y la API key. Además, se incorporó la posibilidad de probar la conexión con cada integración antes de guardarla y eliminar organizaciones cuando sea necesario.

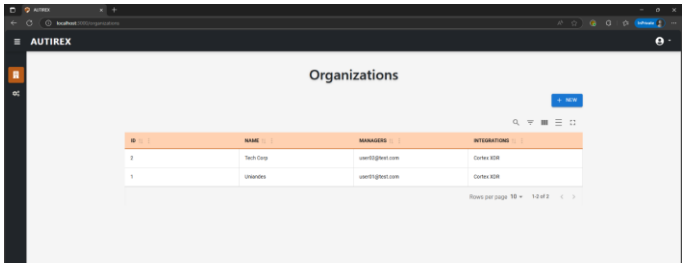

Ilustración 5. Módulo de organizaciones

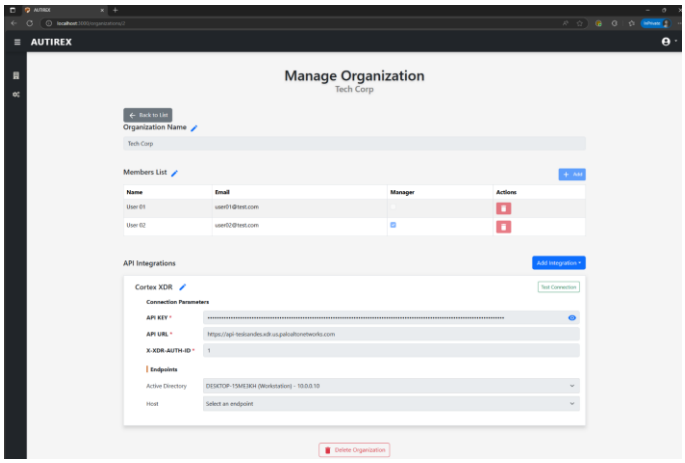

Ilustración 6. Personalización y verificación de módulo

- Página de administración de integraciones

Se creó la página de integraciones, que permite a los usuarios crear, actualizar o eliminar integraciones disponibles para las organizaciones. Esta funcionalidad es fundamental para mantener actualizada la conexión con diversas plataformas de ciberseguridad y garantizar la efectividad del monitoreo y respuesta a incidentes.

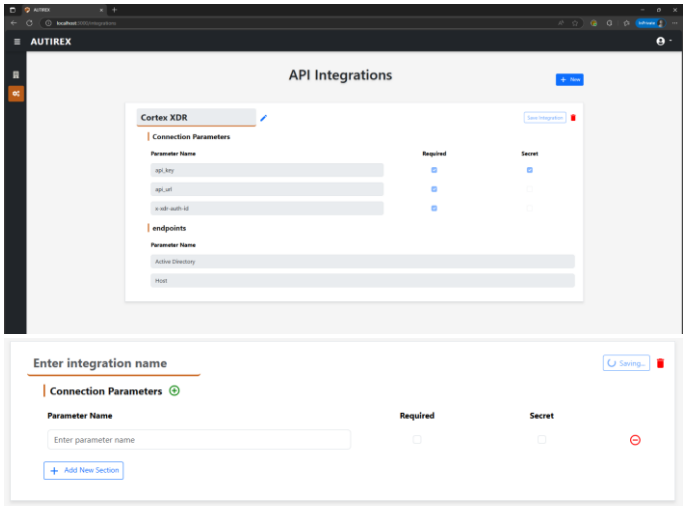

Ilustración 7. Integraciones API

- Página de incidentes

Se implementó la sección de incidentes, donde los datos obtenidos de Cortex XDR se presentan en una tabla. Cada incidente listado incluye un enlace directo para visualizarlo en la plataforma de Cortex XDR, proporcionando un acceso rápido para la investigación y resolución de incidentes de seguridad.

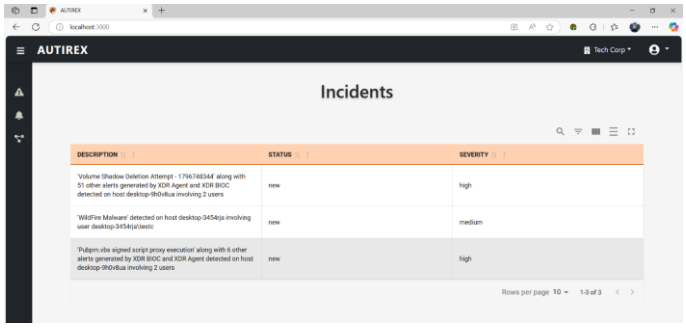

Ilustración 8. Visualizador de incidentes

- Página de alertas

En la sección de alertas, se implementó una tabla que lista las alertas recuperadas de Cortex XDR. Cada alerta incluye opciones de acción específicas, permitiendo, por ejemplo, realizar un análisis cuando la alerta contiene un hash. Esto facilita una respuesta rápida ante amenazas detectadas, conectando directamente con servicios de análisis externos para enriquecer la información disponible sobre cada alerta.

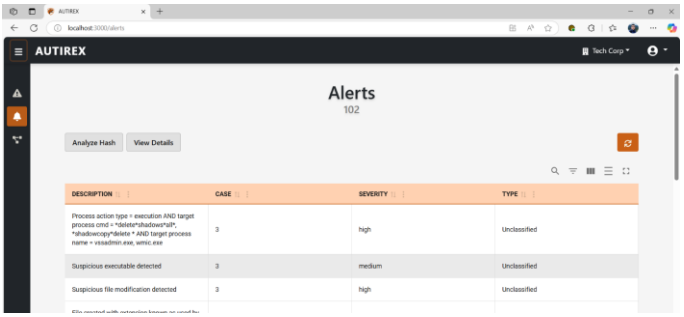

Ilustración 9. Visualizador de alertas

- Módulo de análisis de hash

Se integraron servicios de análisis de amenazas como Hybrid Analysis, VirusTotal y OTX AlienVault. Cuando se analiza un hash, se presentan los valores más relevantes de los resultados obtenidos, junto con enlaces que permiten visualizar el análisis completo en la página de cada integración. Si algún análisis falla, el sistema muestra claramente el error retornado por el servicio correspondiente, facilitando la identificación de problemas en la consulta.

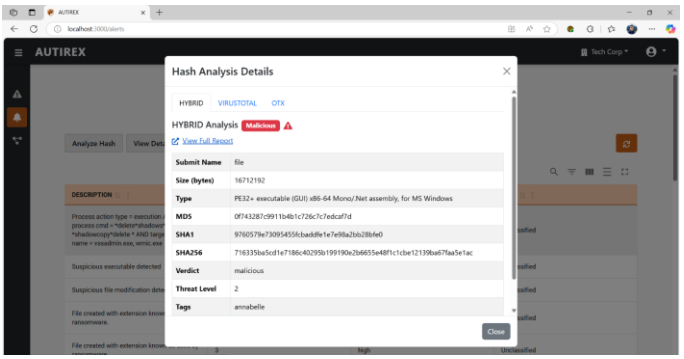

Ilustración 10. Detalle de análisis

- Módulo de creación de Playbooks

Se creó una página para listar los playbooks existentes, con opciones para abrirlos y editarlos. También se permite ejecutar o eliminar un playbook, mostrando una ventana de confirmación antes de proceder con la eliminación.

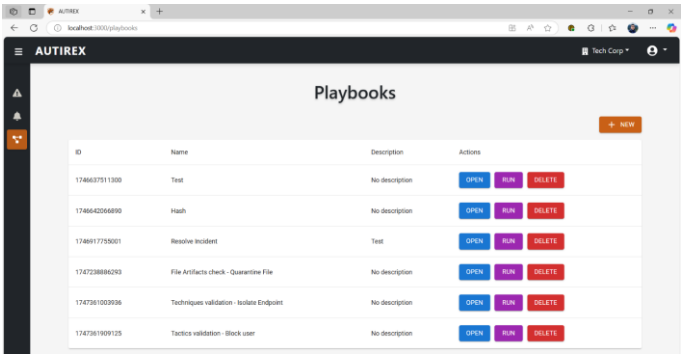

Ilustración 11. Editor de playbooks

Para la construcción gráfica de los playbooks, se definieron cinco tipos de nodos: inicio, acción, decisión, intervención humana y final. El nodo de inicio marca el punto de entrada del flujo y no permite conexiones entrantes, solo salientes. El nodo de acción representa una tarea automatizada (una llamada a la API) y puede conectarse con nodos posteriores para continuar el flujo automáticamente. El nodo de decisión permite bifurcar el flujo según condiciones evaluadas sobre los datos disponibles (por ejemplo, el valor de un campo en una alerta o incidente). El nodo de intervención humana (implementación pendiente) introduce una pausa en el flujo hasta que un analista tome una decisión manual, útil para casos que requieren validación o juicio humano. El nodo de final indica el término del playbook y no permite conexiones salientes.

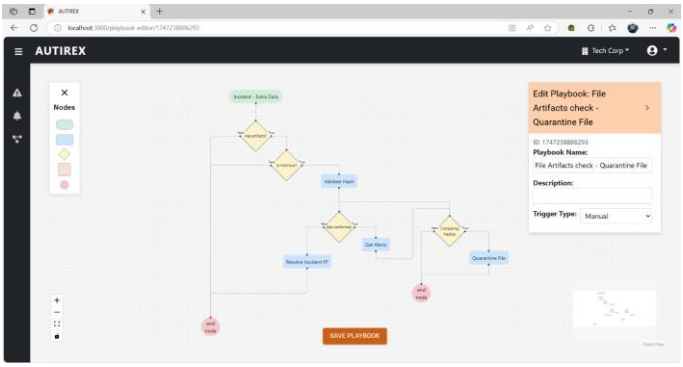

Ilustración 12. Flujo de clasificador y análisis

Cada nodo cuenta con un panel de configuración dinámico que adapta su contenido según las entradas recibidas. Este panel permite al usuario definir la lógica específica que debe aplicarse durante la ejecución del playbook, como seleccionar campos relevantes, establecer condiciones, o configurar parámetros de acción según el tipo de nodo.

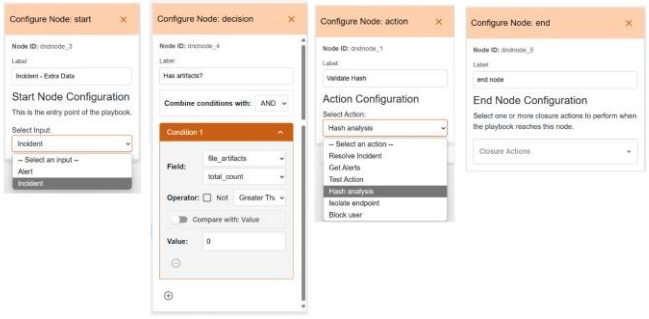

Ilustración 13. Configuración de nodos

Al ejecutar un playbook de forma individual, se muestra una ventana que permite al usuario ingresar los parámetros necesarios para la ejecución manual. Estos parámetros pueden variar según la lógica del playbook y son esenciales para iniciar correctamente el flujo definido.

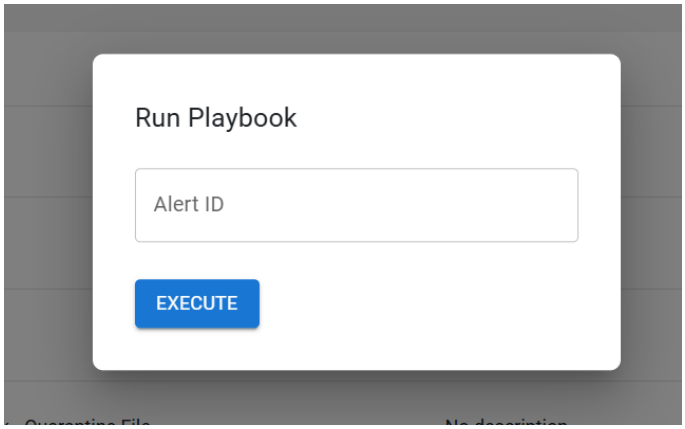

Ilustración 14. Parámetros para ejecución manual de playbook

Los playbooks pueden ser encadenados para formar flujos de respuesta más complejos. Cuando en el curso de la ejecución de un playbook se realiza alguna acción contundente como poner un archivo en cuarentena, aislar un endpoint o bloquear un usuario, el incidente será automáticamente clasificado como verdadero positivo. Además, el sistema permite al usuario definir el orden de ejecución de los playbooks encadenados, lo que otorga flexibilidad para adaptar la lógica de respuesta a distintos tipos de incidentes o políticas internas.

Cuando se ejecuta un playbook de forma masiva sobre varios incidentes a la vez (ejecución en lote), se aplica un proceso de clasificación y priorización que determina el orden en que se atenderán los casos. Este proceso puede ser configurado por el usuario según sus necesidades operativas, pero por defecto sigue los siguientes criterios, en orden de prioridad:

Severidad del incidente: Críticos, Altos y Medios. Los incidentes de baja severidad se consideran de bajo impacto y se resuelven automáticamente.

1) Cantidad de alertas relacionadas con el incidente.
2) Número de técnicas MITRE identificadas.
3) Número de tácticas MITRE asociadas.
4) Cantidad de artefactos clasificados como malware dentro del incidente.

Esta lógica de clasificación permite garantizar que los incidentes de mayor riesgo reciban atención prioritaria durante la ejecución automatizada de los playbooks.

c. Servicios API

Los endpoints de la sección de "Organizations" permiten a los usuarios gestionar organizaciones a través de operaciones CRUD (Crear, Leer, Actualizar, Eliminar). Esto incluye crear una nueva organización, actualizar la existente, obtener una lista de organizaciones y eliminar organizaciones específicas. También se pueden eliminar usuarios e integraciones de una organización en particular.

En la sección de "Integrations", los endpoints facilitan la gestión de integraciones. Permiten crear nuevas integraciones, obtener una lista de todas las integraciones actuales y actualizar integraciones existentes mediante su ID, lo que permite mantener y adaptar las integraciones a las necesidades cambiantes del sistema.

La sección de "Cortex" cuenta con endpoints destinados a monitorear la salud del sistema y gestionar incidentes y alertas. Se permite verificar el estado de salud del sistema, así como recuperar alertas y incidentes. Además, se puede actualizar un incidente específico utilizando su ID, lo que permite una gestión efectiva de los problemas identificados.

En "Users", los endpoints proporcionan funcionalidad para gestionar usuarios a través de registros, inicio de sesión y recuperación de datos del usuario. Se pueden registrar nuevos usuarios, iniciar sesión para acceder a sus cuentas y obtener información sobre el usuario actual o una lista de todos los usuarios registrados en el sistema.

La sección de "Analysis" está dedicada a la evaluación de seguridad, permitiendo a los usuarios analizar un hash específico mediante una llamada a un endpoint que recibe el valor hash y retorna información relevante sobre el mismo, lo que es crucial para la detección de amenazas. Por último, el endpoint de "Health" ofrece una verificación general del estado de la API.

V. PRUEBAS Y RESULTADOS

A. Consideraciones

Debido a que se creó un prototipo para la implementación de la herramienta de seguridad, se optó por desplegar en un ambiente local y no en nube para reducir costos de infraestructura, además, solo se implementó una cantidad limitada de playbooks (aislamiento de endpoint, bloqueo de usuario de dominio, borrado o cuarentena de archivo y sus respectivas reversiones) por restricciones de tiempo.

B. Ambiente de Pruebas

Se genera un ambiente de pruebas virtualizado con VMware Workstation Pro, donde se corren cuatro máquinas virtuales para emular un ambiente corporativo tradicional, compuesto por un controlador de dominio Windows Server 2019 (Directorio Activo y DNS) y tres Windows 10 Pro como usuarios finales. Los tres equipos se encuentran en el mismo segmento de red (10.0.0.0/24), garantizando así movimientos laterales para las pruebas realizadas. Por último, se conectan estos equipos a un Firewall para la salida a internet y así poder conectarse a la consola de Cortex XDR. Nuestra herramienta fue desplegada de manera local en un equipo de pruebas que se conecta vía API a la consola de Cortex XDR de pruebas. A continuación, se empiezan a recibir datos y posteriormente se procesan de acuerdo con los módulos implementados y se observa las respuestas generadas.

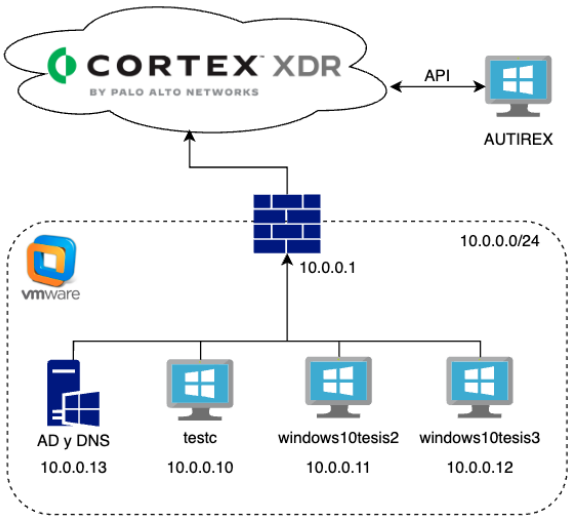

Ilustración 15. Diagrama de ambiente de pruebas

Se crea un dominio ficticio “tesismaestria.lab” y se crean los usuarios “windows10tesis2” y “windows10tesis3”, además, se instala el agente de Cortex XDR en las cuatro máquinas y se les asigna una política de protección en modo alerta en la consola

de Cortex XDR, esto con el fin de poder generar ataques más reales y con más datos para nuestra herramienta.

C. Pruebas de Ataque y Resultados

Se plantean tres pruebas de ataque simulando un ambiente real, la primera donde un usuario descarga un archivo malicioso o virus, el segundo un usuario ejecutando comandos de hackeo dentro del sistema operativo y, por último, un usuario ejecutando varios Ransomwares dentro del sistema operativo.

a. Prueba de Virus Común

Se utiliza un documento oficial de Palo Alto sobre simuladores de Ransomwares [6], donde se obtiene un virus llamado "Shinolocker" capaz de cifrar varios archivos y eliminar las copias ocultas de los Windows. Se descarga el virus en la máquina virtual "testc", y Cortex automáticamente registra el incidente con su respectiva alerta.

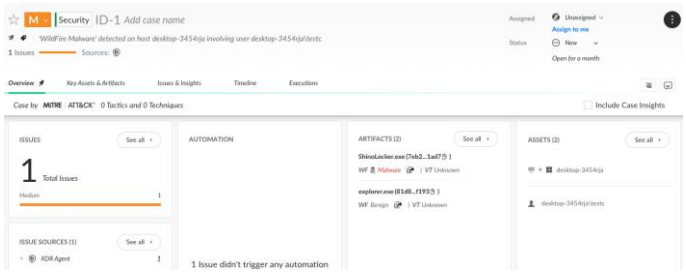

Ilustración 16. Primer incidente

b. Prueba de Técnicas y Tácticas MITRE

Se plantea como segunda prueba, el uso de técnicas y tácticas MITRE para la simulación de un ataque de un hacker real dentro de una compañía, cuyo conocimiento vulnera las falencias de los sistemas operativos. Para esto, nos basamos en un repositorio público de GITHUB "ATOMIC RED TEAM" [7], que contiene varios comandos para emular ataques en el terminal de Windows sin la necesidad de descargar un archivo malicioso.

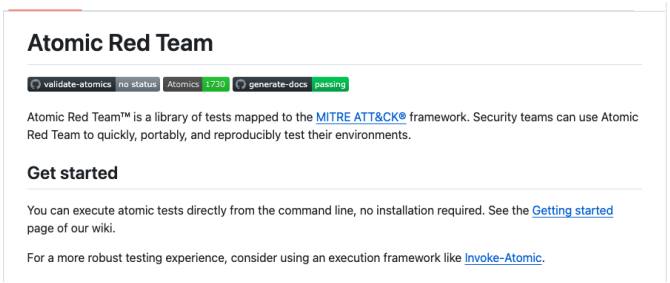

Ilustración 17. Repositorio ATOMIC

Se procede a ejecutar varias técnicas como evasión de defensa: T1112, T1140 y T1216.001, T1218 [8]. Donde se busca evadir defensas tradicionales de Windows por medio de la ejecución de procesos, librerías y herramientas propias del sistema operativo de manera remota, utilizando herramientas propias del sistema como tareas de impresión o Rundll32.exe para la ejecución de librerías. También se utiliza técnicas de ejecución de scripts y elevación de privilegios: T1059 y T1059 [8]; esto con el fin de simular ataques donde nuestra herramienta no va a obtener valores HASH para la comparación con herramientas externas, sino que procede con el análisis de técnicas y tácticas registradas en el incidente para así tomar una acción (por ejemplo, bloquear el usuario o aislar el endpoint) y posteriormente dar resolución de este.

Ilustración 18. Segundo incidente

c. Prueba de Ransomwares

Se recurre al repositorio de muestras de Ransomwares reales de Cybersight en GITHUB [9]. Con el objetivo de simular un ataque real con todos los componentes de un actor malicioso como virus, control remoto de la máquina, vulneración del sistema operativo, etc... Esto permite tener más datos para análisis en nuestra herramienta y así poder generar respuestas adecuadas y más estrictas en caso de alguna afectación de los servicios de los usuarios.

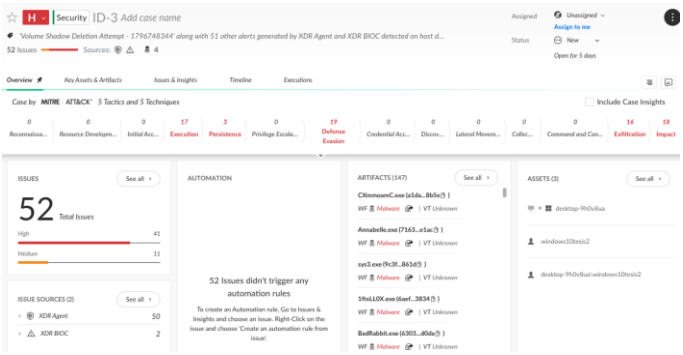

Ilustración 19. Tercer incidente

D. Resultados y Creación de playbooks

Para la primera prueba al contener solo un archivo malicioso y siguiendo el flujo de nuestra herramienta, solo se ejecuta el playbook de validación de artefactos de tipo archivo, poniendo en cuarentena el archivo, sin borrarlo, esperando a tener una retroalimentación en caso de ser falso positivo para la ejecución del playbook que restablece de nuevo el archivo de la cuarentena. El tiempo tomado por la herramienta desde la generación del incidente y en finalizar la respuesta y resolución del incidente fue de 10 minutos. El incidente se dejó con severidad media.

El flujo comienza con la evaluación de un incidente y la verificación de si contiene artefactos de tipo archivo (file_artifacts). Si existen, se accede a los datos asociados para verificar si alguno ha sido marcado como malicioso ("is_malicious": true). En caso afirmativo, se valida el hash SHA256 del archivo para confirmar su veredicto. Si el archivo se determina como benigno, el incidente se resuelve como un falso positivo. Si, por el contrario, se confirma que es malicioso, se procede a obtener el file_path y el endpoint_id desde las alertas relacionadas con ese hash. Finalmente, se ejecuta una acción de respuesta para eliminar el archivo o ponerlo en cuarentena en los endpoints afectados.

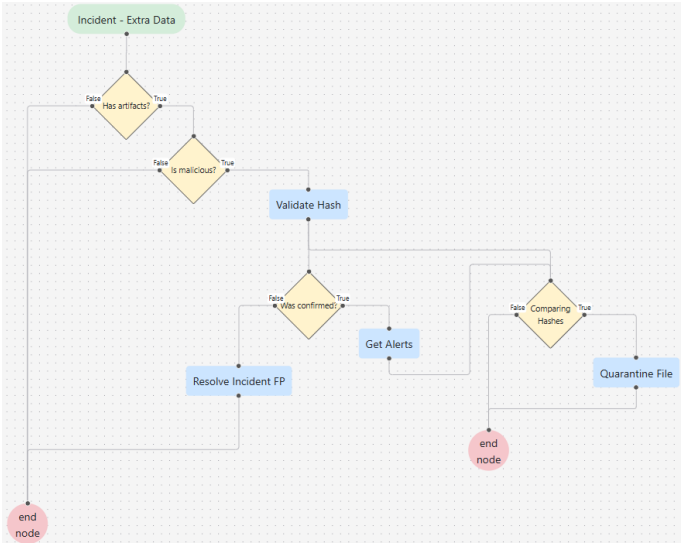

Ilustración 20. Primer flujo artefactos

En el caso de la segunda prueba en la que no contiene un archivo malicioso sino solo técnicas y tácticas de MITRE por actividad sospechosa, se ejecuta el playbook de validación de técnicas MITRE ATT&CK y validación de tácticas de MITRE ATT&CK, donde se aisló el endpoint y se bloqueó el usuario del directorio activo, esperando a tener retroalimentación en caso de ser un falso positivo para restablecer el endpoint y desbloquear la cuenta en el directorio activo. El tiempo tomado por la herramienta desde la generación del incidente y en finalizar la respuesta y resolución del incidente fue de 5 minutos. El incidente se dejó con severidad alta.

Si un incidente contiene al menos una técnica MITRE asociada (mitre_techniques_ids_and_names.length > 0), se considera que está vinculado a una táctica conocida de ataque, por lo que se procede a aislar el endpoint afectado como medida preventiva para contener la posible amenaza.

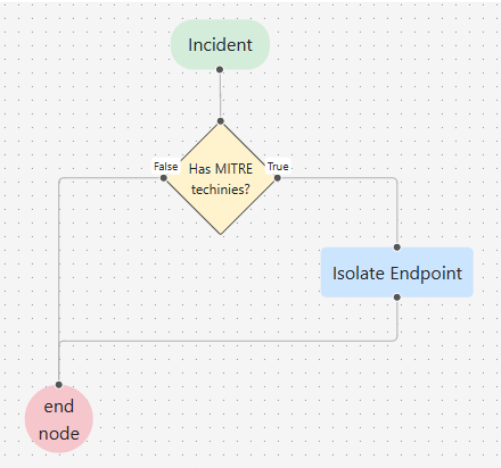

Ilustración 21. Segundo flujo técnicas MITRE

Cuando un incidente presenta una o más tácticas MITRE asociadas (mitre_tactics_ids_and_names.length > 0), se interpreta como una señal de comportamiento potencialmente malicioso vinculado a un usuario. En respuesta, se ejecuta una acción para bloquear al usuario involucrado, con el objetivo de prevenir movimientos laterales o acciones adicionales dentro del entorno comprometido.

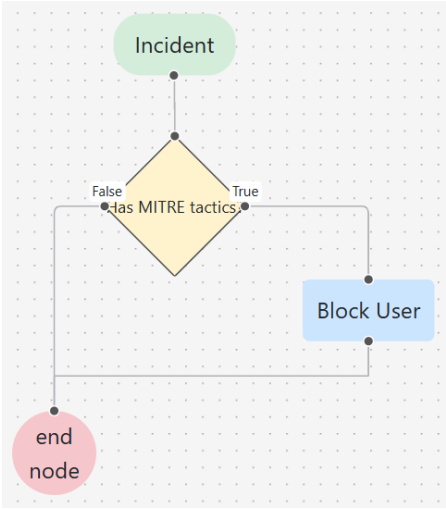

Ilustración 22. Tercer flujo tácticas MITRE

En la tercera prueba, al tener componentes más realistas con varios archivos maliciosos y varias técnicas y tácticas de MITRE, se ejecutaron todos los playbooks y flujos de la herramienta, alisando el endpoint, poniendo en cuarentena el archivo y bloqueando la cuenta del usuario a nivel de dominio. Por lo que fue efectiva la prueba al tratarse de un ambiente más real. El tiempo tomado por la herramienta desde la generación del incidente y en finalizar la respuesta y resolución del incidente fue de 20 minutos. El incidente se dejó con severidad alta.

VI. DISCUSIÓN

Los resultados obtenidos en el entorno de pruebas virtualizado validan la efectividad de la herramienta propuesta para la automatización de respuestas a incidentes de ciberseguridad en un contexto simulado de una red corporativa. Se logró establecer una integración funcional entre Cortex XDR y la solución AUTIREX mediante API, permitiendo la recepción y procesamiento automatizado de incidentes.

En la primera prueba, relacionada con la detección de un archivo malicioso (Shinolocker), la herramienta identificó correctamente el incidente y ejecutó un playbook de cuarentena, evitando la propagación sin eliminar el archivo inmediatamente, con posibilidad de reversión en caso de falso positivo. El tiempo de resolución fue de 10 minutos tomado con cronómetro, con una clasificación de severidad media. Este resultado evidencia la capacidad de respuesta precisa ante amenazas identificables por artefactos.

La segunda prueba, que emuló ataques mediante técnicas y tácticas MITRE sin archivos maliciosos, mostró la capacidad del sistema para interpretar comportamientos anómalos, aplicar lógica de contención automática y tomar decisiones basadas en la severidad y criticidad de las técnicas y tácticas. En este caso, se ejecutaron los playbooks de aislamiento del endpoint y bloqueo de cuenta del usuario de Active Directory. El incidente fue tratado en 5 minutos tomado con cronómetro, destacando una alta eficiencia para responder a amenazas sin artefactos detectables. Este comportamiento valida el uso de inteligencia contextual (tácticas y técnicas) como criterio de automatización.

En la tercera prueba, al integrar múltiples indicadores (archivos maliciosos + tácticas MITRE), el sistema ejecutó una respuesta completa, incluyendo aislamiento, cuarentena de archivos y bloqueo de usuarios. La resolución se completó en 20 minutos tomado con cronómetro, demostrando que el sistema escala correctamente su nivel de respuesta frente a amenazas complejas y multi-vectoriales, como los ataques ransomware.

En todas las pruebas, los tiempos de respuesta estuvieron muy por debajo del MTTR promedio de la industria para empresas sin SOC [2], lo que respalda la hipótesis de que la automatización puede compensar la falta de equipos especializados, especialmente en sectores PYMES. A pesar de no haber integrado herramientas como firewalls o correo electrónico por limitaciones del prototipo, la arquitectura modular demostró su potencial de expansión.

Además, la implementación de mecanismos de reversión, cuarentena condicional y validación cruzada de IOC contribuye a reducir falsos positivos y fortalecer la confiabilidad operativa del sistema. No obstante, se reconoce que las pruebas se realizaron en un entorno controlado y que la robustez en escenarios reales requerirá ajustes finos y despliegues progresivos. También en los componentes de seguridad que contiene una entidad pequeña o mediana lo que puede generar respuestas no automatizadas por la limitación de desarrollo o integración nativa vía API.

VII. TRABAJO FUTURO

Aunque el desarrollo del prototipo permitió validar la viabilidad de una solución de automatización para la gestión de incidentes basada en Cortex XDR, existen múltiples líneas de mejora y expansión para un despliegue funcional en entornos reales.

Una de las principales limitaciones del trabajo fue la ausencia de modelos de inteligencia artificial entrenados. Debido al tiempo limitado, no se implementaron algoritmos que aprendieran del comportamiento histórico de los incidentes. Esta carencia limita la capacidad predictiva y adaptativa del sistema. Como trabajo futuro, se propone la recolección progresiva de incidentes para alimentar modelos de aprendizaje supervisado que mejoren la clasificación, detección temprana y optimización de respuestas automatizadas.

Asimismo, se identificó que el volumen y tipo de incidentes dependen fuertemente del contexto de cada organización, especialmente en las PYMES. Factores como el número de usuarios, los niveles de concientización, los activos digitales y las prácticas operativas influyen directamente en la naturaleza de las amenazas. Por tanto, se propone diseñar módulos de personalización por entorno, que ajusten los niveles de severidad, umbrales de riesgo y respuesta según el perfil específico de cada institución.

Otra línea crítica de mejora es la ampliación de integraciones con otras herramientas de seguridad, tales como firewalls, plataformas de correo electrónico, sistemas de autenticación y motores de inteligencia de amenazas. En esta versión solo se

trabajó con Cortex XDR; sin embargo, en entornos empresariales reales se utilizan múltiples soluciones. Por ello, la implementación futura debe considerar arquitecturas abiertas y adaptables vía API, apoyadas en estándares como STIX/TAXII, para garantizar interoperabilidad.

En cuanto a la confiabilidad operativa, aunque se incluyeron mecanismos de reversión para mitigar errores, la automatización conlleva riesgos inherentes como falsos positivos o bloqueos innecesarios. Será necesario implementar una capa de validación basada en retroalimentación humana, así como mecanismos de aprendizaje continuo para ajustar respuestas según los resultados históricos y la evolución del entorno.

Finalmente, se destaca que en un ambiente empresarial real, deben considerarse variables adicionales no presentes en el entorno controlado. Entre ellas se incluyen:

Costos de infraestructura en la nube, tanto para despliegue de contenedores, bases de datos y herramientas de integración.

Requerimientos de confidencialidad y cumplimiento, que obligan a proteger datos sensibles en tránsito y reposo bajo normativas como ISO/IEC 27001 y leyes locales de protección de datos.

Alta disponibilidad, ya que cualquier componente de respuesta automática debe garantizar tiempo de actividad cercano al 100%, especialmente durante eventos críticos.

Estas consideraciones complementarias serán clave para escalar la solución de prototipo a un producto empresarial robusto y alineado con las exigencias de producción, garantizando así su utilidad, sostenibilidad y adopción en entornos reales.

VIII. CONCLUSIONES

Este trabajo presenta una propuesta funcional de automatización para la gestión y respuesta a incidentes de ciberseguridad basada en Cortex XDR, orientada a organizaciones pequeñas y medianas que no cuentan con un CSIRT o equipo SOC formal. A través de un prototipo desarrollado en un entorno de laboratorio, se validó que es posible reducir significativamente el tiempo medio de respuesta (MTTR) ante incidentes críticos mediante el uso de playbooks automáticos, integraciones vía API y decisiones condicionadas por indicadores como MITRE ATT&CK, IOC y artefactos detectados.

Los resultados obtenidos en las pruebas de laboratorio demostraron que la herramienta es capaz de ejecutar respuestas efectivas como cuarentena de archivos maliciosos, aislamiento de endpoints y bloqueo de usuarios, todo ello sin intervención humana directa, logrando tiempos de resolución entre 5 y 20 minutos dependiendo de la complejidad del incidente. Esto representa una mejora sustancial frente a los tiempos promedio del mercado reportados para organizaciones sin automatización.

Además, la solución fue diseñada bajo una arquitectura modular y escalable, lo que facilita su integración futura con otras herramientas de ciberseguridad y su adaptación a distintos entornos. A pesar de las limitaciones del prototipo como la ausencia de modelos de inteligencia artificial entrenados, la cobertura parcial de integraciones y el entorno controlado se evidenció un alto potencial para escalar esta solución como un producto viable.

Se concluye que la automatización inteligente no solo es una estrategia técnica eficiente, sino también una solución viable para organizaciones con presupuestos limitados, especialmente en las PYMES. La implementación de respuestas automáticas basadas en reglas y análisis contextual puede aumentar la resiliencia, reducir costos operativos y mejorar la capacidad de respuesta, mitigando los riesgos ante amenazas cada vez más frecuentes y sofisticadas.

Como trabajo futuro, se propone avanzar hacia un sistema completamente adaptativo, basado en aprendizaje automático, con mayor cobertura de integraciones, monitoreo en tiempo real, y cumplimiento normativo para ambientes empresariales reales y de alta criticidad.

IX. REFERENCIAS

[1] ISACA, State of Cybersecurity 2024: Global Update on Workforce Efforts, Resources, and Cyberoperations, ISACA, 2024. [Online]. Available: https://www.isaca.org/resources/reports/state-of-cybersecurity-2024

[2] A. Alam, How to Evaluate Incident Response Beyond Basic Security KPIs, Scrut Automation, 2023. [Online]. Available: https://www.scrut.io/post/incident-response

[3] Verizon, Data Breach Investigations Report 2025 (DBIR), Verizon Communications, 2025. [Online]. Available: https://www.verizon.com/business/resources/T607/reports/2025-dbir-data-breach-investigations-report.pdf

[4] CyberRisk Alliance, MSSP Alert 2024 Pricing Benchmark Report, MSSP Alert, 2024. [Online]. Available: https://files.cyberriskalliance.com/wp-content/uploads/2024/12/MSSP-Alert_Pricing-Benchmark-Report_2024.pdf

[5] IT Price, Palo Alto Networks Cortex XSOAR Pricing and Product List, 2024. [Online]. Available: https://itprice.com/paloalto-price-list/cortex%20xsoar.html

[6] Y. Allon, "Ransomware simulators – reality or a bluff?", Palo Alto Networks, May 2, 2022. [Online]. Available: https://www.paloaltonetworks.com/blog/security-operations/ransomware-simulators-reality-or-a-bluff/

[7] Red Canary, Atomic Red Team. GitHub. [Online]. Available: https://github.com/redcanaryco/atomic-red-team

[8] MITRE, MITRE ATT&CK® Framework. [Online]. Available: https://attack.mitre.org/

[9] Cybersight Security, Malware Samples. GitHub. [Online]. Available: https://github.com/Cybersight-Security/Malware-Samples