Universidad Tecnológica de Bolívar

Inteligencia Artificial

Agente Inteligente

Agente de Defensa Cibernetica

Mauro Alonso Gonzalez Figueroa, T00067622

Juan Jose Jimenez Guardo, T00068278

Revisado Por

Edwin Alexander Puertas Del Castillo

7 de septiembre de 2025

Índice general

1. Resumen Ejecutivo

Caso de Estudio: La defensa cibernética es un conjunto de estrategias, tecnologías y procesos diseñados para proteger los sistemas informáticos, redes y datos contra amenazas cibernéticas, como ataques de malware, piratería informática, robo de datos y otros riesgos de seguridad. Estas medidas de defensa buscan prevenir, detectar, responder y recuperarse de ataques cibernéticos con el fin de garantizar la confidencialidad, integridad y disponibilidad de la información y los recursos digitales.

El agente inteligente de defensa cibernética está diseñado para proteger infraestructuras digitales mediante la detección, prevención y respuesta automatizada ante amenazas cibernéticas. Su arquitectura integra sensores avanzados, actuadores de respuesta y métricas de desempeño que garantizan la resiliencia y seguridad de los sistemas corporativos. El diseño incorpora recomendaciones técnicas para mejorar la capacidad de reacción, la cobertura ante nuevas amenazas y la alineación con estándares modernos de ciberseguridad.

2. Introducción

El objetivo de este informe es presentar el diseño y ajuste de un agente inteligente enfocado en defensa cibernética, detallando su estructura bajo el enfoque PEAS. La importancia de diseñar agentes inteligentes radica en su capacidad para automatizar la protección de sistemas críticos, reducir riesgos y mejorar la eficiencia en la gestión de incidentes de seguridad. El alcance de la actividad incluye la definición técnica del agente, sus componentes y la justificación de los ajustes realizados según las recomendaciones proporcionadas.

3. Descripción del Agente Asignado

3.1. Tipo de agente

El agente propuesto es de tipo basado en modelo, capaz de razonar sobre el estado actual y anticipar acciones futuras para maximizar la seguridad.

3.2. Dominio o entorno

Opera en infraestructuras de red corporativa, servidores críticos y servicios en la nube, considerando también el entorno humano y dispositivos IoT.

3.3. Tareas principales

Las tareas incluyen la detección de amenazas, bloqueo automático de tráfico malicioso, aislamiento de dispositivos comprometidos, auditorías de seguridad, aplicación de actualizaciones y generación de alertas.

4. Ajustes del Diseño Basados en Recomendaciones Previas

4.1. Recomendaciones consideradas

Las siguientes recomendaciones técnicas fueron incorporadas al diseño del agente inteligente de defensa cibernética:

- Incorporar métricas avanzadas de desempeño, como tiempo medio de respuesta (MTTR), tiempo medio entre fallos (MTBF), efectividad de contención y estimación del impacto económico evitado.

- Ampliar el entorno operativo considerando factores humanos y organizacionales, políticas internas, niveles de autorización, dispositivos BYOD y sistemas IoT, así como el comportamiento del usuario.

- Integrar sistemas de respuesta orquestada (SOAR) para automatizar el bloqueo, documentación, escalamiento y notificación de incidentes, junto con módulos de gestión de registros forenses y dashboards ejecutivos en tiempo real.

- Incorporar sensores basados en inteligencia de amenazas externa, diferenciando entre detección por firmas y por anomalías comportamentales, e incluyendo sensores de comportamiento del usuario (UEBA) para identificar amenazas internas.

4.2. Cambios aplicados al diseño original

El diseño del agente fue ajustado para incorporar todas las recomendaciones técnicas recibidas, logrando una solución más robusta y alineada con las mejores prácticas actuales en ciberseguridad. Los principales cambios realizados son:

- **Performance Measuring:** Se añadieron métricas avanzadas como el tiempo medio de respuesta (MTTR), tiempo medio entre fallos (MTBF), efectividad de contención y estimación del impacto económico evitado, además de los indicadores originales (porcentaje de ataques prevenidos, cantidad de ataques maliciosos detectados y falsos positivos).

- **Environment:** El entorno del agente se amplió para incluir factores humanos y organizacionales, políticas internas de seguridad, niveles de autorización, dispositivos BYOD

y sistemas IoT, así como el comportamiento del usuario, además de la infraestructura de red, servidores críticos y servicios en la nube.

- **Actuators:** Se integró un sistema de respuesta orquestada (SOAR) para automatizar el bloqueo, documentación, escalamiento y notificación de incidentes. Se añadió un módulo de gestión de registros forenses y dashboards ejecutivos en tiempo real, complementando los sistemas de bloqueo, escaneo y monitoreo automáticos ya presentes.

- **Sensors:** Se incorporaron sensores basados en inteligencia de amenazas externa, diferenciando entre detección por firmas y por anomalías comportamentales, e incluyendo sensores de comportamiento del usuario (UEBA), además de los antivirus, monitores de integridad y sensores de red originales.

4.3. Justificación técnica de los ajustes

La aplicación de estos cambios responde a la necesidad de fortalecer la capacidad operativa y adaptativa del agente frente a amenazas cada vez más sofisticadas. Las nuevas métricas permiten una evaluación más precisa de la eficiencia y resiliencia del sistema, facilitando la toma de decisiones y la mejora continua. La ampliación del entorno operativo y la integración de factores humanos e IoT aseguran una cobertura más completa y realista, considerando los riesgos actuales en infraestructuras corporativas.

La incorporación de sistemas SOAR y gestión forense automatiza procesos críticos, reduce el tiempo de respuesta y mejora la trazabilidad de incidentes, lo que es esencial para auditorías y cumplimiento normativo. Finalmente, el uso de sensores avanzados y diferenciados permite detectar tanto amenazas externas como internas, aumentando la capacidad de prevención y respuesta ante ataques complejos y comportamientos anómalos.

Estos ajustes posicionan al agente como una solución integral y moderna, capaz de adaptarse a los desafíos dinámicos del entorno digital y alineada con los estándares internacionales

de ciberseguridad.

5. Definición del Espacio de Estados

5.1. Descripción de las variables que definen cada estado

El espacio de estados del agente inteligente de defensa cibernética se define por las siguientes variables principales:

- **Estado de operación**: Indica si el sistema está en monitoreo normal, detección de anomalía, respuesta automática, análisis/escalamiento, recuperación o mejora continua.
- **Nivel de amenaza**: Valor que representa la severidad de la actividad detectada (normal, sospechosa, maliciosa).
- **Estado de dispositivos**: Identifica si los dispositivos están operativos, comprometidos, aislados o en proceso de recuperación.
- **Alertas y reportes**: Registro de alertas generadas, reportes forenses y acciones ejecutadas.
- **Métricas de desempeño**: Incluye tiempo medio de respuesta (MTTR), tiempo medio entre fallos (MTBF), efectividad de contención y falsos positivos.
- **Actualización de conocimiento**: Estado de las firmas, IOCs y reglas heurísticas del sistema.

5.2. Representación de los estados

A continuación se presenta una tabla que resume los estados principales y sus variables asociadas:

Estado	Nivel de amenaza	Dispositivo	Acción	Métricas
Monitoreo Normal	Baja	Operativo	Supervisión	MTBF
Detección de Anomalía	Media	Operativo / Comprometido	Análisis	Falsos positivos
Respuesta Automática	Alta	Comprometido / Aislado	Bloqueo, Aislamiento	MTTR, Contención
Análisis y Escalamiento	Alta	Comprometido	Reporte, Escalamiento	Evidencia forense
Recuperación	Variable	Restaurando	Parches, Backups	MTTR
Mejora Continua	Baja	Operativo	Actualización	Aprendizaje

Cuadro 1: Estados principales y variables asociadas del agente

5.3. Transiciones entre estados

Las transiciones entre estados se producen en función de los eventos detectados y las acciones del agente. El siguiente diagrama textual describe el flujo principal:

```
Monitoreo Normal
    → [Anomalía detectada] → Detección de Anomalía
```

```
Detección de Anomalía
    → [Falso positivo] → Monitoreo Normal
    → [Confirmado ataque] → Respuesta Automática

Respuesta Automática
    → [Ataque controlado] → Recuperación
    → [Ataque complejo/no contenido] → Análisis y Escalamiento

Análisis y Escalamiento
    → [Caso resuelto] → Recuperación

Recuperación
    → [Servicios restaurados] → Mejora Continua

Mejora Continua
    → [Actualización completada] → Monitoreo Normal
```

6. Objetos del Agente

- **Sensor de red**: Analiza tráfico y detecta anomalías.
- **Actuador de bloqueo**: Ejecuta acciones de aislamiento y bloqueo.
- **Monitor de integridad**: Supervisa cambios no autorizados.
- **Módulo SOAR**: Orquesta respuestas y genera informes.
- **Gestor de registros forenses**: Prepara evidencia para auditorías.

7. Predicados del Agente

- `Amenaza (x)` $\rightarrow$ `Bloquear (x)`
- `Comprometido (y)` $\rightarrow$ `Aislar (y)`
- `Alerta (z)` $\rightarrow$ `Notificar (z)`
- `UsuarioInusual (u)` $\rightarrow$ `Analizar (u)`

Si `Amenaza (TráficoMalicioso)` entonces `Bloquear (TráficoMalicioso)`.

8. Conclusiones

El proceso de diseño permitió fortalecer la capacidad operativa del agente, integrando recomendaciones expertas y ampliando su cobertura ante amenazas. Futuras mejoras pueden incluir el uso de aprendizaje automático para detección proactiva y mayor integración con sistemas externos. El enfoque de agentes inteligentes en IA resulta esencial para la protección dinámica y automatizada de infraestructuras críticas.

Bibliografía

Jurafsky, D., & Martin, J. H. (2025). *Speech and Language Processing* (4th) [Forthcoming]. Pearson.

Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th). Pearson.