



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Dirección General de Estudios de Posgrado

Facultad de Ingeniería de Sistemas e Informática

Unidad de Posgrado

**Arquitectura de honeynet con integración de datos
para el análisis de evidencia digital**

TESIS

Para optar el Grado Académico de Doctor en Ingeniería
de Sistemas e Informática

AUTOR

Danny Patricio VELASCO SILVA

ASESOR

Dr. Glen Darío RODRÍGUEZ RAFAEL

Lima, Perú

2025



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Velasco, D. (2025). *Arquitectura de honeynet con integración de datos para el análisis de evidencia digital*. [Tesis de doctorado, Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática, Unidad de Posgrado]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	Danny Patricio Velasco Silva
Tipo de documento de identidad	Cédula de identidad
Número de documento de identidad	EC / 0602768640
URL de ORCID	https://orcid.org/0000-0003-0396-4086
Datos de asesor	
Nombres y apellidos	Glen Darío Rodríguez Rafael
Tipo de documento de identidad	DNI
Número de documento de identidad	10062042
URL de ORCID	https://orcid.org/0000-0002-4471-3198
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Lenis Rossi Wong Portillo
Tipo de documento	DNI
Número de documento de identidad	10438282
Miembro del jurado 1	
Nombres y apellidos	Ivan Carlo Petrlik Azabache
Tipo de documento	DNI
Número de documento de identidad	10140461
Miembro del jurado 2	
Nombres y apellidos	Luz Sussy Bayona Oré
Tipo de documento	DNI
Número de documento de identidad	08602165

Datos de investigación	
Línea de investigación	Tecnologías de la Información
Grupo de investigación	No aplica
Agencia de financiamiento	Ecuador. Universidad Nacional de Chimborazo (UNACH)
Ubicación geográfica de la investigación	País: Ecuador Provincia: Chimborazo Ciudad: Riobamba Latitud: -1.67435 Longitud: -78.6483
Año o rango de años en que se realizó la investigación	2018 - 2025
URL de disciplinas OCDE	Ingeniería de sistemas y comunicaciones https://purl.org/pe-repo/ocde/ford#2.02.04



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
Universidad del Perú, Decana de América
Facultad de Ingeniería de Sistemas e Informática
Vicedecanato de Investigación y Posgrado
Unidad de Posgrado

**ACTA DE SUSTENTACIÓN DE TESIS PARA OPTAR EL GRADO ACADÉMICO DE DOCTOR
EN INGENIERÍA DE SISTEMAS E INFORMÁTICA**

A los cuatro (4) días del mes de diciembre de 2025, siendo las 11:00 a.m., se reunieron en el Aula Magna, el Jurado de Tesis conformado por los siguientes docentes:

Dra. Lenis Rossi Wong Portillo (Presidente)
Dr. Ivan Carlo Petrlik Azabache (Miembro)
Dra. Luz Sussy Bayona Oré (Miembro)
Dr. Glen Darío Rodríguez Rafael (Miembro Asesor)

Se inició la Sustentación invitando al candidata a Doctor **DANNY PATRICIO VELASCO SILVA**, para que realice la exposición oral de la tesis para optar el Grado Académico de Doctor en Ingeniería de Sistemas e Informática, siendo la Tesis intitulada:

“Arquitectura de honeynet con integración de datos para el análisis de evidencia digital”

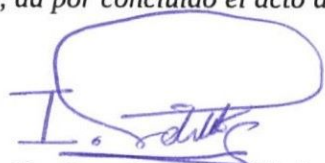
Concluida la exposición, los miembros del Jurado de Tesis procedieron a formular sus preguntas que fueron absueltas por el graduando; acto seguido se procedió a la evaluación correspondiente, habiendo obtenido la siguiente calificación:

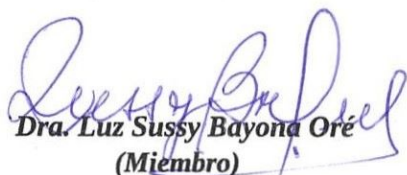
..... Dieciocho (18) Muy Bueno

Por tanto, el presidente del Jurado, de acuerdo con el Reglamento General de Estudios de Posgrado, otorga a la Magíster **DANNY PATRICIO VELASCO SILVA** el Grado Académico de Doctor en Ingeniería de Sistemas e Informática.

Siendo las 12:12 horas, el presidente del Jurado de Tesis, da por concluido el acto académico de Sustentación de Tesis.


Dra. Lenis Rossi Wong Portillo
(Presidente)


Dr. Ivan Carlo Petrlik Azabache
(Miembro)


Dra. Luz Sussy Bayona Oré
(Miembro)


Dr. Glen Darío Rodríguez Rafael
(Miembro Asesor)

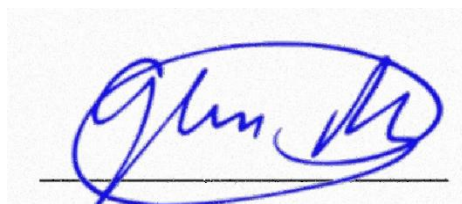
CERTIFICADO DE SIMILITUD

Yo, Glen Darío Rodríguez Rafael en mi condición de asesor acreditado con Dictamen N° 187-UPG-FISI-2017 de la tesis, cuyo título es ARQUITECTURA DE HONEYNET CON INTEGRACIÓN DE DATOS PARA EL ANÁLISIS DE EVIDENCIA DIGITAL, presentado por el magíster Danny Patricio Velasco Silva para optar el grado académico de Doctor en Ingeniería de Sistemas e Informática, CERTIFICO que se ha cumplido con lo establecido en la Directiva de Originalidad y de Similitud de Trabajos Académicos, de Investigación y Producción Intelectual. Según la revisión, análisis y evaluación mediante la plataforma PIDS, el documento evaluado cuenta con el 16% de similitud, nivel **PERMITIDO** para continuar con los trámites correspondientes y para su **publicación en el repositorio institucional**.

Se emite el presente certificado en cumplimiento de lo establecido en las normas vigentes, como uno de los requisitos para la obtención del grado académico de Doctor en Ingeniería de Sistemas e Informática.

Lima, 29 de diciembre de 2024

Firma del Asesor



Nombres y apellidos del asesor: Glen Darío Rodríguez Rafael

DNI: 10062042

DEDICATORIA

- A mis Padres Eduardo y Gloria
- A mis hijos Danny y David

Ing. Danny Patricio Velasco Silva, Mgs.
Doctorante

AGRADECIMIENTO

Mi más sincero agradecimiento a:

- A Dios
- A la Universidad Nacional Mayor San Marcos de Lima
- A la Universidad Nacional del Chimborazo
- Al Dr. Glen Rodríguez, tutor del trabajo de investigación
- A todo el personal docente del programa doctoral
- Y a todos quienes me apoyaron en este proceso

Ing. Danny Patricio Velasco Silva, Mgs.

Doctorante

ÍNDICE GENERAL

DEDICATORIA.....	iv
AGRADECIMIENTO.....	v
ÍNDICE GENERAL.....	vi
LISTA DE TABLAS.....	ix
LISTA DE FIGURAS.....	xi
RESUMEN.....	xiii
ABSTRACT.....	xv
CAPÍTULO I: INTRODUCCIÓN.....	1
1.1. Situación problemática.....	1
1.2. Formulación del problema.....	4
1.2.1. Problema general.....	4
1.2.2. Problemas específicos.....	4
1.3. Justificación teórica.....	4
1.4. Justificación práctica.....	6
1.5. Objetivos del estudio.....	7
1.5.1. Objetivo general.....	7
1.5.2. Objetivos específicos.....	7
CAPÍTULO II: MARCO TEÓRICO.....	8
2.1. Marco filosófico o epistemológico de la investigación.....	8
2.2. Antecedentes de investigación.....	13
2.2.1. Caracterización de la actividad del hacker.....	13
2.2.2. Los ciberataques.....	15
2.2.3. Intrusión moderna.....	19
2.2.4. Interoperabilidad entre los sistemas.....	20
2.3. Bases Teóricas.....	21
2.3.1. Honeypots.....	21
2.3.2. Tipología de los honeypots.....	23
2.3.3. Honeynet.....	24
2.4. Comparación de arquitecturas honeynet.....	40
2.5. ¿Por qué la herramienta de integración de datos?.....	44
2.5.1. Operatividad técnica	45

2.5.2.	Captura de evidencia digital.....	46
2.5.3.	Consistencia del proceso de análisis.....	47
2.5.4.	Tiempos de detección de ataques.....	48
2.5.5.	Integridad de los datos.....	49
2.5.6.	Gestión de la información.....	50
CAPÍTULO III: METODOLOGÍA.....		51
3.1.	Hipótesis.....	51
3.1.1.	Hipótesis general de la investigación (H_0).....	51
3.1.2.	Hipótesis nula de la investigación (H_1).....	51
3.1.3.	Hipótesis específicas.....	51
3.2.	Identificación de variables.....	52
3.2.1.	Variable independiente.....	52
3.2.2.	Variable dependiente.....	52
3.3.	Operacionalización de variables.....	53
3.4.	Matriz de consistencia.....	56
3.5.	Metodología.....	59
3.5.1.	Tipo y diseño de la investigación.....	59
3.5.2.	Unidad de análisis.....	60
3.5.3.	Población de estudio.....	60
3.5.4.	Tamaño de la muestra.....	60
3.5.5.	Técnicas de recolección de datos.....	61
3.5.6.	Técnica de análisis de datos.....	62
CAPÍTULO IV: RESULTADOS Y DISCUSIÓN.....		64
4.1.	Ontología honeynet para integración de datos.....	64
4.1.1.	Metodología para la ontología.....	66
4.1.2.	Desarrollo de la ontología.....	68
4.1.3.	Validación de la ontología.....	79
4.2.	Arquitectura de honeynet con integración de datos para el análisis de evidencia digital.....	85
4.2.1.	Arquitectura honeynet.....	86
4.2.2.	Aplicativo de integración de datos.....	89

4.3.	Análisis, interpretación y discusión de resultados.....	91
4.3.1.	Variable independiente - Dimensión Operatividad Técnica.....	93
4.3.2.	Variable dependiente - Dimensión Captura de la evidencia digital.....	94
4.3.3.	Variable dependiente - Dimensión Consistencia del proceso de análisis.....	96
4.3.4.	Variable dependiente - Dimensión Tiempos de detección de ataques.....	98
4.3.5.	Variable dependiente - Dimensión Integridad de los datos.....	100
4.3.6.	Variable independiente - Dimensión Gestión de la información.	102
4.4.	Prueba de hipótesis.....	104
4.4.1.	Hipótesis Específica HE ₁	105
4.4.2.	Hipótesis Específica HE ₂	107
4.4.3.	Hipótesis Específica HE ₃	109
4.4.4.	Hipótesis Específica HE ₄	111
4.4.5.	Hipótesis Específica HE ₅	117
4.4.6.	Hipótesis Específica HE ₆	119
4.4.7.	Hipótesis General.....	121
	CONCLUSIONES.....	127
	RECOMENDACIONES.....	131
	REFERENCIAS BIBLIOGRÁFICAS.....	132
	ANEXOS	

LISTA DE TABLAS

Tabla 1:	Comparación de las diferentes generaciones.....	33
Tabla 2:	Comparación de características de honeynet virtual.....	36
Tabla 3:	Comparación de arquitecturas honeynet.....	42
Tabla 4:	Operacionalización de variables.....	53
Tabla 5:	Matriz de Consistencia.....	56
Tabla 6:	Sección del diccionario de conceptos de la ontología para integración de datos en honeynet.....	70
Tabla 7:	Parte de la sección de la tabla de relaciones binarias de la ontología para integración de datos en honeynet.....	71
Tabla 8:	Parte de la sección de la tabla de atributos de instancia de la ontología para integración de datos en honeynet.....	72
Tabla 9:	Sección de la tabla de atributos de clase de la ontología para integración de datos en honeynet.....	73
Tabla 10:	Sección de la tabla de constantes de la ontología para integración de datos en honeynet.....	74
Tabla 11:	Sección de la tabla de axiomas formales de la ontología para integración de datos en honeynet.....	74
Tabla 12:	Sección de la tabla de reglas de la ontología para integración de datos en honeynet.....	76
Tabla 13:	Estadísticos descriptivos - Aspectos de methontology.....	80
Tabla 14:	Estadísticos de prueba - Aspectos de methontology.....	82
Tabla 15:	Estadísticos descriptivos - Identificación de posibles errores.....	83
Tabla 16:	Estadísticos de prueba - Identificación de posibles errores.....	84
Tabla 17:	Interoperabilidad.....	93
Tabla 18:	Captura de evidencia digital.....	94
Tabla 19:	Consistencia del proceso de análisis.....	96
Tabla 20:	Aceptabilidad de tiempos de detección de ataques.....	98
Tabla 21:	Integridad de los datos.....	100
Tabla 22:	Gestión de la información.....	102
Tabla 23:	Resultados para la resolución de problemas de interoperabilidad.....	105
Tabla 24:	Tabla de contingencia HE ₁	106

Tabla 25	Estadístico de prueba - Resolución de problemas de interoperabilidad.....	106
Tabla 26:	Resultados para la captura de evidencia digital.....	107
Tabla 27:	Tabla de contingencia HE ₂	108
Tabla 28:	Estadístico de prueba - Captura de evidencia digital.....	108
Tabla 29:	Resultados de la consistencia del proceso de análisis de evidencia digital.....	109
Tabla 30:	Tabla de contingencia HE ₃	110
Tabla 31:	Estadístico de prueba - Consistencia del proceso de análisis de evidencia digital.....	110
Tabla 32:	Comparación de tiempos de detección de ataques.....	111
Tabla 33.	Resumen descriptivo de los tiempos de detección de ataques.....	112
Tabla 34:	Estadístico de prueba - Normalidad de tiempos de detección de ataques sin integración de datos.....	113
Tabla 35:	Estadísticos de prueba - Normalidad de tiempos de detección de ataques con integración de datos	113
Tabla 36:	Estadístico de prueba - Varianzas iguales de tiempos de detección de ataques.....	115
Tabla 37:	Estadísticos de prueba - Comparación de tiempos de detección de ataques sin y con integración de datos.....	116
Tabla 38:	Resultados para la integridad de los datos.....	117
Tabla 39:	Tabla de contingencia HE ₅	118
Tabla 40:	Estadístico de prueba - Integridad de los datos.....	118
Tabla 41:	Resultados para gestión de la información.....	119
Tabla 42:	Tabla de contingencia HE ₆	120
Tabla 43:	Estadístico de prueba - Gestión de la información.....	120
Tabla 44:	Resultados para la Hipótesis General.....	122
Tabla 45:	Tabla de contingencia H ₀	122
Tabla 46:	Estadístico de prueba - Hipótesis General.....	123

LISTA DE FIGURAS

Figura 1:	Arquitectura de Generación I.....	27
Figura 2:	Arquitectura de Generación II.....	28
Figura 3:	Arquitectura de Generación III.....	30
Figura 4:	Honeynets virtuales autocontenidas	35
Figura 5:	Honeynet virtual híbrida.....	35
Figura 6:	Esquema de una honeynet distribuida.....	37
Figura 7:	Dominio de la Ontología.....	65
Figura 8:	Actividades de desarrollo de ontologías propuestas por methontology.....	66
Figura 9:	Tareas de la metodología methontology.....	67
Figura 10:	Sección de la taxonomía de conceptos de la ontología honeynet para integración de datos.....	68
Figura 11:	Sección de la construcción de diagramas de relaciones binarias de la ontología honeynet para integración de datos.....	69
Figura 12:	Sección de la construcción de diagramas de relaciones binarias específica para integración de datos.....	69
Figura 13:	Clases y subclases de la ontología para la integración de datos en honeynet.....	77
Figura 14:	Fragmento de jerarquía de integración de datos.....	78
Figura 15:	Regla de decisión - Aspectos de methontology.....	82
Figura 16:	Regla de decisión - Identificación de posibles errores.....	84
Figura 17:	Esquema propuesto.....	87
Figura 18:	Arquitectura honeynet.....	88
Figura 19:	Almacenamiento centralizado.....	89
Figura 20:	Integración de datos.....	90
Figura 21:	Caso de uso: Integración de datos en una honeynet.....	91
Figura 22:	Interoperabilidad.....	93
Figura 23:	Captura de evidencia digital.....	95
Figura 24:	Consistencia del proceso de análisis.....	96
Figura 25:	Aceptabilidad de tiempos de detección de ataques.....	99
Figura 26:	Integridad de los datos.....	100
Figura 27:	Gestión de la información.....	102

Figura 28:	Regla de decisión HE_1	106
Figura 29:	Regla de decisión HE_2	108
Figura 30:	Regla de decisión HE_3	110
Figura 31:	Prueba de Normalidad.....	114
Figura 32:	Diagrama de caja y bigotes sin integración vs con integración...	115
Figura 33:	Regla de decisión HE_4	116
Figura 34:	Regla de decisión HE_5	118
Figura 35:	Regla de decisión HE_6	121
Figura 36	Regla de decisión H_0	123

RESUMEN

La honeynet surgió como una herramienta de seguridad informática, diseñada intencionalmente para ser detectada, atacada y comprometida por intrusos en un entorno controlado. Está compuesta por una serie de entornos de red y sistemas señuelo, conocidos como honeypots, los cuales, una vez instalados y configurados, simulan servicios y vulnerabilidades reales con el objetivo de atraer a los intrusos. Esta infraestructura permite observar en tiempo real las acciones de los atacantes, facilitando el análisis profundo de sus tácticas, técnicas, estrategias y objetivos, proporcionando información valiosa para el desarrollo de defensas más eficaces frente a amenazas cibernéticas. A lo largo de los últimos años, se han implementado diversas arquitecturas honeynet con el propósito de mejorar la detección y análisis de ataques, sin embargo, estos estudios han evidenciado una limitación persistente en la falta de integración de datos. En este contexto, el propósito principal de esta investigación es desarrollar una arquitectura de honeynet con integración de datos para el análisis de evidencia digital, que mejore la interoperabilidad propia de la heterogeneidad en los datos proporcionados por las herramientas utilizadas en honeynet en cuanto a confiabilidad y tiempos de respuesta.

Se elaboró una ontología honeynet para integración de datos, mediante una especificación explícita y formal de una conceptualización compartida, desarrollada en la herramienta de código abierto Protégé, tomando como base metodológica methontology, la cual fue la base para la arquitectura honeynet con integración de datos para el análisis de evidencia digital, las herramientas que se utilizaron en la arquitectura honeynet para la captura de datos son: Firewall, Sebek, Snort, Argos, P0f, con la finalidad de capturar el comportamiento del intruso dentro de la honeynet, generando una base de datos unificada con toda la información recopilada.

La investigación adoptó un enfoque cuantitativo, con un nivel descriptivo y explicativo, y se enmarcó en un diseño cuasiexperimental, mediante la validación de indicadores técnicos con el juicio de expertos y pruebas de hipótesis.

Los resultados obtenidos en la evaluación de la arquitectura de honeynet con integración de datos evidencian mejoras sustanciales para el análisis de seguridad

digital. Se observó una mejora significativa en la recopilación y comparación de atributos de los datos capturados, lo que facilitó la identificación de patrones de ataque y correlación de eventos, la integración de datos permitió el acceso centralizado a información proveniente de diversas fuentes a través de una sola vista, simplificando la labor del analista, otro hallazgo clave fue la reducción de los tiempos de detección de ciberataques, lo que permite actuar con prontitud ante amenazas y reducir el riesgo de daño a la infraestructura tecnológica, incrementando la confianza de los analistas en el proceso y en los resultados obtenidos permitiendo una mejor gestión para la toma de decisiones de manera eficaz y oportuna.

Palabras Clave: Honeypot, Honeynet, Ontología, Integración de datos, Evidencia digital.

ABSTRACT

The honeynet emerged as a computer security tool intentionally designed to be detected, attacked, and compromised by intruders in a controlled environment. It is composed of a series of network environments and decoy systems, known as honeypots, which, once installed and configured, simulate real services and vulnerabilities with the aim of attracting intruders. This infrastructure allows for real-time observation of attackers' actions, facilitating in-depth analysis of their tactics, techniques, strategies, and objectives, providing valuable information for the development of more effective defenses against cyber threats. Over the past few years, various honeynet architectures have been implemented to improve attack detection and analysis; however, these studies have revealed a persistent limitation in the lack of data integration. In this context, the main objective of this research is to develop a honeynet architecture with data integration for the analysis of digital evidence, which improves the interoperability inherent in the heterogeneity of the data provided by the tools used in honeynets in terms of reliability and response times.

A honeynet ontology for data integration was developed through an explicit and formal specification of a shared conceptualization. It was developed in the open-source tool Protégé, using the methodological framework methontology, which was the basis for the honeynet architecture with data integration for digital evidence analysis. The tools used in the honeynet architecture for data capture were: Firewall, Sebek, Snort, Argos, and P0f. The purpose was to capture the intruder's behavior within the honeynet, generating a unified database with all the collected information.

The research adopted a quantitative approach, with a descriptive and explanatory level, and was framed within a quasi-experimental design, through the validation of technical indicators with expert judgment and hypothesis testing.

The results obtained from the evaluation of the honeynet architecture with data integration demonstrate substantial improvements for digital security analysis. A significant improvement was observed in the collection and comparison of attributes of the captured data, which facilitated the identification of attack patterns and event correlation, data integration allowed centralized access to information from various

sources through a single view, simplifying the work of the analyst, another key finding was the reduction in cyberattack detection times, which allows prompt action against threats and reduces the risk of damage to the technological infrastructure, increasing the confidence of analysts in the process and in the results obtained, allowing better management for decision-making in an effective and timely manner.

Keywords: Honeypot, Honeynet, Ontology, Data integration, Digital evidence.

CAPÍTULO I: INTRODUCCIÓN

1.1 Situación Problemática

Los ciberataques, hoy día, constituyen una amenaza constante; con el rápido avance de la tecnología, los servidores se han vuelto cada vez más vulnerables a las ciberamenazas, lo que supone riesgos significativos que ocasiona serios daños tanto a individuos como a empresas y organizaciones gubernamentales (Alshaya & Khediri, 2025). Ataques empresariales que en 2017 provocaron pérdidas superiores a los 300 millones de dólares, según el Informe Especial sobre Ciberguerra, se prevé que la ciberdelincuencia alcance los 10,5 billones de dólares anuales para 2025, una cifra exponencialmente superior a los daños causados por desastres naturales en un año. Según el FBI, el Centro de Quejas de Delitos en Internet informó que el volumen de quejas en 2021 fue de 847376, con pérdidas de 6900 millones de dólares, hasta el 1% del PIB mundial se pierde actualmente debido a la ciberdelincuencia. Además, el costo de la ciberdelincuencia para la economía mundial ha aumentado en más del 50% en dos años (Djenna et al., 2023).

A medida que avanza la era digital y el acelerado crecimiento de dispositivos móviles conectados en el ciberespacio y el Internet de las Cosas (IoT), los factores de riesgo también aumentan y con ello las vulnerabilidades en ciberseguridad por el auge de hackers e intrusos (Saravana et al., 2025), a ello, hay que sumarle el creciente aumento del trabajo a distancia con una fuerza laboral que posee poca habilidad o conocimiento en el aspecto de ciberseguridad; de ahí la necesidad, no solo de proteger la información y los propios sistemas informáticos, sino también, de prevenir los ataques y, para ello, uno de los mecanismos más importantes es comprender cuáles son los métodos utilizados por los ciberdelincuentes, cómo operan y cuáles son las brechas de seguridad

que existen en este ámbito. Los principales ciberatacantes de élite actuales compiten en ingenio para penetrar infraestructuras críticas, realizar ciberespionaje y extraer datos confidenciales, ocultando sus rastros. Utilizan diversas técnicas avanzadas y furtivas, escaneando y explotando numerosas vulnerabilidades cibernéticas, creando puertas traseras ocultas y dando lugar a diversas ciberamenazas (Ahsan et al., 2025).

Para comprender cómo se producen estos ataques, se desarrollan arquitecturas de honeynet, que surgen como una potente medida de ciberseguridad, empleando sistemas, servicios o recursos virtuales o simulados para atraer a los atacantes (Chen et al., 2024), que utilizan las honeypots que permiten llevar a cabo estos tipos de ataques. Son sistemas de seguridad de red cada vez más popular en la arquitectura de red moderna debido a su capacidad para detectar actividades maliciosas de fuentes externas. Los honeypots simulan un objetivo de alto valor y alejan a los atacantes potenciales de los sistemas críticos (Karthigha et al., 2025). Como resultado de esto, la honeynet se convierte en una herramienta valiosa para generar información que permite comprender los patrones y las estrategias que son utilizadas en un ciberataque, se utiliza como capa de seguridad para mitigar ataques, descubrir técnicas conocidas de los atacantes y debilidades de la red y del sistema, que los atacantes puedan utilizar, y para mitigar estas vulnerabilidades (Mesbah et al., 2023).

El desarrollo de honeynets a lo largo de las generaciones anteriores han mejorado el control de datos y las capacidades de captura de honeypots mediante funciones de puerta de enlace personalizadas, optimizando así el uso de sus capacidades. Sin embargo, aún no existe correlación entre los diferentes honeypots en esta etapa (Yu et al., 2024), y su eficacia se ha visto limitada por la falta de un modelo de datos unificado y estándar, resultado de la existencia de múltiples fuentes de datos no relacionadas, cada una con su propio método y formato de acceso (Balas & Viecco, 2005), además Lucchese and Merro (2024), en su propuesta de trabajo futuro de su tesis doctoral menciona que la integración de datos en la honeynet, podría mejorar la identificación de las tácticas y técnicas de los atacantes, proporcionando una comprensión más completa del proceso de ataque.

Para el desarrollo de una de arquitectura de honeynet para el análisis de evidencia digital, hay una absoluta libertad al momento de seleccionar tanto su topología como las herramientas para la construcción de las tareas de: almacenamiento, control, registro y análisis de las acciones del intruso en el interior de la misma; en la literatura reciente se identifican varias arquitecturas dinámicas de honeynet que buscan mejorar la captura y análisis de datos, caracterizados por monitorizar ataques a gran escala como son: (Ma et al., 2024), (Yu et al., 2024), (De Pace et al., 2023), (Kohli et al., 2023), (Noaman et al., 2019), (Lucchese & Merro, 2024), cada una con su propia característica que se describirán en el marco teórico, pero presentan una limitación en la integración directa con módulos de análisis de evidencia digital, por lo cual, no resuelve la interoperabilidad de datos heterogéneos generados en la honeynet; dependen de formatos específicos para cada módulo, en cada una de estas propuestas.

Las limitaciones detectadas en los modelos actuales de honeynet sustentan la necesidad de desarrollar una nueva arquitectura que permita la integración de los datos eficiente, de la información proveniente de múltiples tipos de honeypots y de herramientas de captura. Markert (2020), en su propuesta de trabajo futuro de su tesis doctoral menciona que para el análisis de datos se debe realizarse en un solo dispositivo con la base de datos del tráfico capturado.

La capacidad de procesamiento automatizado y correlación integrada reducirá significativamente los tiempos requeridos para el análisis de grandes volúmenes de datos generados por la honeynet. Esto es crítico en escenarios de respuesta ante incidentes donde el tiempo es un factor determinante para contener ataques o mitigar brechas de seguridad (Kohli et al., 2023). Los administradores de red y los analistas de ciberseguridad podrán tomar decisiones informadas con mayor prontitud, optimizando la gestión de la seguridad de la red. Esto incluye la aplicación oportuna de contramedidas, bloqueo de vectores de ataque, reconfiguración dinámica de defensas o activación de protocolos de contingencia.

1.2 Formulación del Problema

1.2.1 Problema General

¿El uso de una arquitectura de honeynet, con integración de datos, mejora el análisis de evidencia digital producto de la heterogeneidad en los datos proporcionados por las herramientas utilizadas?

1.2.2 Problemas Específicos

- a) El uso de una arquitectura de honeynet con integración de datos, ¿permite resolver problemas de interoperabilidad?
- b) ¿Se mejora la captura de evidencia digital con la consolidación de la ontología de la honeynet con integración de los datos?
- c) El uso de una arquitectura de honeynet con integración de datos, ¿mejora la consistencia del proceso de análisis de evidencia digital?
- d) ¿Mejoran los tiempos de detección de ataques durante el análisis de evidencia digital con el uso de una arquitectura de honeynet con integración de datos?
- e) ¿El uso de una arquitectura de honeynet con integración de datos mejora la integridad de los datos en el análisis de evidencia digital?
- f) El uso de una arquitectura de honeynet con integración de datos, ¿optimiza la gestión de la información en el análisis de evidencia digital?

1.3 Justificación Teórica

Existe una amplia variedad de tecnologías y de información disponible para prevenir y resolver los problemas, producto de ciberataques a los sistemas informáticos de diversa naturaleza, sin embargo, los ataques, no solo continúan ocurriendo, sino que, según algunos investigadores en el área, sigue en aumento.

Los profesionales e investigadores de seguridad están diseñando mecanismos para proteger la red y sus componentes, pero aún surgen atacantes con tecnologías de vanguardia para perturbar las intenciones de los usuarios legítimos en la red (Ganesarathinam et al., 2020).

Hasta el momento, no existe una metodología única, una sola forma de abordar esta situación, por ello, el aprender cuáles son los procedimientos utilizados para perpetrar estos ataques es la mejor herramienta, comprender cómo se efectúan, sigue siendo la mejor manera de prevenir y de solucionar los problemas que estos ataques acarrearán. El uso de honeypots y honeynets es una de las formas más efectivas de obtener información sobre los atacantes y sus acciones lo antes posible (Kashtalian et al., 2021). Si bien los honeypots se originaron como una herramienta de ciberseguridad para detectar y mitigar ciberamenazas intrusivas, han recibido una considerable atención académica en las últimas dos décadas (Perkins & Howell, 2021).

Cabe señalar que, las arquitecturas presentadas en la literatura, respecto a los honeypots y las honeynets, no proporcionan orientaciones específicas respecto a la integración de datos de las diferentes herramientas de captura, de ahí que es importante contar con modelos flexibles que permitan generar mejores formas de abordar los ciberataques.

Aunado a lo anteriormente expuesto, el diseñar y desarrollar la ontología para una honeynet con integración de datos, especificando de manera formal las relaciones entre sus componentes, las propiedades y los tipos de entidades que la conforma, permite disminuir la complejidad de la honeynet además de generar y organizar la información al respecto; esto se traduce, entonces, en una forma de contar con conocimientos actualizados que permiten mejorar los sistemas de detección y de defensa ante ciberataques, de ahí que haya un impacto positivo en la epistemología respecto a la ciberseguridad, por cuanto se adquiere conocimientos en esta área, se refuerzan conceptos, se abren caminos para generar más investigaciones en este campo a la vez que se validan los límites y los métodos propios del área.

De ahí que la propuesta de diseñar una arquitectura para el análisis de evidencia digital, que integren los datos proporcionados por las diferentes herramientas de captura diseñadas con tal fin en una honeynet, constituye un aporte teórico importante para la reflexión académica en el área de la ciberseguridad que apoyaría la formación de especialistas en este campo.

1.4 Justificación Práctica

La justificación práctica de esta investigación gira en torno a que el producto de la misma (el desarrollo de una arquitectura de honeynet con integración de datos que resuelva la interoperabilidad propia de la heterogeneidad en los datos proporcionados por las herramientas utilizadas), que permita resolver los problemas internos que se producen debido a la heterogeneidad de los distintos formatos de los datos capturados por la red (Masmoudi et al., 2024), constituye una solución a un evento real constituido por la necesidad de proteger la red, a la vez que atrae a los atacantes para que comprenden sus patrones, estrategias y comportamientos, con la finalidad de prevenir los ciberataques a sistemas informáticos (Minjiao et al., 2022).

Es necesario contar con una interoperabilidad semántica en la arquitectura de honeynet con integración de datos para el análisis de evidencia digital, con el propósito de utilizar datos de forma transparente con otras herramientas de captura, incluso si estas utilizan diferentes formatos y modelos de datos (Cimmino et al., 2024).

Desde esta perspectiva, la información generada, tanto desde la investigación misma como a partir del uso de la honeynet con integración de datos, genera información que sirve de apoyo para la toma de decisiones, mejorando los tiempos de respuesta, ello debido a que esta propuesta define los componentes informáticos que integran el esquema de red con las características necesarias para poder garantizar el almacenamiento y disponer de una herramienta para monitorear, controlar y analizar el tráfico, constituye un apoyo esencial para mejorar la calidad del trabajo que realizan los analistas y auditores, de ahí que sea esencial desarrollar interfaces de usuario que analicen y correlacionen los diferentes tipos de datos, de manera tal, que las organizaciones interesadas compartan la información respecto a los ciberataques y se puedan formar correlaciones a gran escala.

1.5 Objetivos del estudio

1.5.1 Objetivo general

Desarrollar una arquitectura de honeynet con integración de datos que mejora el análisis de evidencia digital propia de la heterogeneidad en los datos proporcionados por las herramientas utilizadas.

1.5.2 Objetivos específicos

- a) Diseñar la arquitectura física y lógica de honeynet para resolver problemas de interoperabilidad.
- b) Elaborar la ontología honeynet con integración de datos para mejorar la captura de evidencia digital.
- c) Analizar la consistencia del proceso de análisis de evidencia digital con el uso de la arquitectura de honeynet con integración de datos.
- d) Identificar los tiempos de detección de ataques en el análisis de evidencia digital con el uso de la arquitectura honeynet con integración de datos.
- e) Mejorar la integridad de los datos en el análisis de evidencia digital con el uso de la arquitectura honeynet con integración de datos.
- f) Comprobar que el uso de la arquitectura de la honeynet con integración de datos, optimiza la gestión de la información en el análisis de evidencia digital.

CAPÍTULO II: MARCO TEÓRICO

2.1 Marco filosófico o epistemológico de la investigación

Partiendo del concepto inicial de Epistemología, “La epistemología, o teoría del conocimiento según Russell es la rama de la filosofía que investiga la naturaleza, los límites y la validez del conocimiento humano” (Kitchener, 2004) y, el propuesto por Paz (2003): “Es el estudio del conocimiento humano y de la ciencia: sus métodos, estructura y criterios de demarcación”, se comprende que ésta trata del estudio filosófico del origen, los límites, la naturaleza y los modos y métodos de conocimiento y; a partir de esta definición, se desarrolla el aspecto filosófico y epistemológico del tema central de esta investigación, siguiendo dos vías: (a) la Epistemología de la información y el conocimiento; (b) la Epistemología de la ciberseguridad.

En primer orden, la epistemología de la información y el conocimiento examinan las teorías, los métodos y las prácticas que constituyen la base de la comprensión de la información y del conocimiento resultante, incluyendo las consideraciones filosóficas y éticas que asocian en este entramado. Se entiende como una rama de la filosofía que estudia de manera crítica los procesos mediante los cuales los datos son adquiridos, procesados y almacenados para transformarse en información; analiza además los formatos empleados para su representación y la forma en que dicha información es utilizada en la construcción del conocimiento.

En el marco de la epistemología de la información y del conocimiento, el estudio reconoce que el conocimiento técnico y científico se origina a partir de la recopilación, procesamiento e interpretación de datos. Por tanto, esta investigación no solo se limita a diseñar una arquitectura técnica, sino que busca analizar la influencia que tiene la centralización e integración de datos provenientes de diversas fuentes en la generación de evidencia digital válida y verificable, lo cual constituye un problema

epistemológico: cómo garantizar que la información recopilada sea confiable, representativa y capaz de transformarse en conocimiento útil para la toma de decisiones, fortaleciendo la validez del conocimiento.

De acuerdo con los estudios recientes sobre epistemología en esta área, se ha observado una creciente producción científica que ha favorecido el reconocimiento de diversas corrientes teóricas, enfoques paradigmáticos y bases filosóficas. Esta evolución refleja cómo el desarrollo tecnológico ha transformado significativamente los modos de recopilación, almacenamiento, distribución y conversión de datos en conocimiento (Francelin, 2018).

Por tanto, la epistemología de la información adquiere especial relevancia en la era digital actual, al ofrecer herramientas conceptuales y metodológicas que permiten analizar la multiplicidad paradigmática y la complejidad inherente al conocimiento informacional. En este sentido, la investigación propone estudiar la influencia de la centralización de la información generada por diferentes fuentes en el análisis de evidencia digital.

Desde la perspectiva de la era digital que arroja al ser humano, la epistemología de la información y el conocimiento incluye:

- a) Naturaleza de la información. Se examina la naturaleza misma de la información, cómo se define, cómo se estructura y cómo se representa; asimismo, estudia las múltiples interacciones que existen entre la información, la realidad y el conocimiento.
- b) Procesamiento de la información. Se estudia cómo la información es procesada y aplicada para generar conocimiento significativo estudiando los procesos cognitivos y los modelos mentales que utiliza el ser humano para organizar, interpretar y darle sentido a la información que recibe.
- c) Fuentes y validación del conocimiento. Aborda cómo se obtiene el conocimiento y cómo éste es validado a partir de la información disponible en el marco de ideas o paradigma asumido por el ser humano; de aquí que, desde la epistemología de la información y el conocimiento, se examinan las fuentes de información, se evalúan

su confiabilidad y veracidad empleando métodos para verificar y validar el conocimiento construido.

- d) Teorías del conocimiento. La epistemología de la información y el conocimiento se ocupa, también, de las teorías que abordan como se adquiere el conocimiento y cómo éste es justificado, entonces, se revisan los enfoques propios del área como el empirismo, el racionalismo, el constructivismo y el pragmatismo, así como la aplicación de estos enfoques en el contexto de la información y la comunicación.
- e) Ética y responsabilidad. La dimensión ética y moral no es ajena al uso de la información y la generación de conocimiento, de aquí que la epistemología de la información y el conocimiento también se ocupa de las implicaciones éticas de la recolección de datos y su transformación en información, además del uso y distribución de la información y del conocimiento generado; se examina también los elementos relacionados con la privacidad, la seguridad y el acceso a la información y al conocimiento.

La segunda vía de estos planteamientos está constituida por la epistemología de la ciberseguridad, la cual se ocupa del estudio del conocimiento y la comprensión en el campo de la seguridad de la información; examina la forma en la que se adquiere, valida y aplica el conocimiento relacionado con la protección de los sistemas informáticos, el establecimiento de los sistemas y protocolos de defensa, además de la prevención de los ciberataques.

La construcción del conocimiento en ciberseguridad no solo enfrenta las habituales dudas epistemológicas propias de otras disciplinas, sino que también debe considerar elementos únicos como el engaño deliberado por parte de adversarios, errores no detectados y restricciones legales. Con este antecedente la propuesta técnica de la arquitectura de honeynet con integración de datos, monitorea permanentemente el acceso a la red para capturar el comportamiento del intruso e identificar amenazas y detectar vulnerabilidades mediante un entorno controlado capaz de emular servicios reales y atraer actividades maliciosas, al tiempo que garantiza la recolección, consolidación y análisis de información proveniente de múltiples fuentes. Para lograrlo, la arquitectura debe contemplar la selección adecuada de honeypots de distinto nivel de interacción, la definición de una topología de red que asegure

aislamiento frente al entorno de producción, y la incorporación de sistemas de monitoreo que permitan registrar de forma precisa las acciones del adversario, favoreciendo una mejor administración de los recursos y una respuesta oportuna ante incidentes de seguridad.

Además, la integración de datos mediante mecanismos de estandarización, correlación y visualización constituye un eje central que transforma registros dispersos e independientes en evidencia digital coherente y verificable. De esta manera, la propuesta no solo fortalece la detección y comprensión de ciberataques, sino que también proporciona un marco técnico robusto para la generación de conocimiento confiable en el campo de la ciberseguridad, lo que permite una toma de decisiones más informada.

De este modo, la epistemología de la ciberseguridad adquiere relevancia al explorar cómo se produce, valida y comunica el conocimiento técnico en un entorno caracterizado por dinámicas adversariales y cambiantes (Williams, 2020). La protección de datos abarca desde las fuentes del sistema informático hasta la información generada y almacenada en él, la epistemología en este campo se centra en la naturaleza del conocimiento en seguridad informática y cómo se utiliza para tomar decisiones y disminuir los riesgos.

En estas consideraciones, es menester indicar que la epistemología de la ciberseguridad incluye los siguientes elementos:

- a) Fuentes de conocimiento. Se analiza de dónde proviene el conocimiento en seguridad informática, considerando que sus fuentes pueden ser la investigación académica, la experiencia práctica, los informes de incidentes, las mejores prácticas de la industria, entre otros.
- b) Validación del conocimiento. Se investiga cómo se valida y se verifica el conocimiento en seguridad informática. Esto implica la revisión de métodos y técnicas utilizadas para evaluar la calidad y la confiabilidad de la información y los datos utilizados para respaldar las decisiones de seguridad.

- c) Incertidumbre y riesgo. Se considera la incertidumbre y el riesgo que están asociados con el conocimiento en este campo. Se exploran los límites del conocimiento y cómo se maneja la incertidumbre al tomar decisiones en el ámbito de la seguridad.
- d) Metodologías de investigación. Se examina las metodologías utilizadas para investigar y generar nuevo conocimiento en este campo. Esto incluye el diseño de experimentos, el análisis de datos y la aplicación de teorías y marcos conceptuales.
- e) Ética y responsabilidad. Se analiza cómo se aplica y se utiliza el conocimiento de manera ética y responsable, y cómo se abordan los dilemas éticos relacionados con la seguridad informática.

El entramado epistemológico propio de la ciberseguridad entretiene las formas de conocer respecto a la ciberseguridad, la generación y el manejo de la información además de la aplicación de la misma en una espiral de conocimiento que debe crecer de manera constante, propiciando nuevos aprendizajes de diversa naturaleza (tanto práctica como teórica) en torno a la ciberseguridad, los ciberataques y las formas de prevenirlos.

Otro aspecto crucial para estos planteamientos, son los elementos axiológicos referidos al tema estudiado y, en este sentido, se parte de que la axiología de la ciberseguridad se centra en el estudio de los valores y principios éticos que guían las acciones y decisiones en el ámbito de la seguridad informática, de aquí que la axiología en este campo del saber, esté estrechamente vinculada con el conocimiento y la información debido a que, en el ámbito de la ciberseguridad se trata con información sensible, por ello, es importante comprender y promover los valores que sustentan la protección de la información, la privacidad de los datos, la integridad de los sistemas y la confianza en el ecosistema digital.

En la base axiológica se encuentran los planteamientos de Floridi (2019), quien ha abordado las cuestiones éticas relacionadas con la ciberseguridad en torno a la privacidad, la transparencia y la responsabilidad de los actores en el mundo digital; también es pertinente señalar las contribuciones de Gönültaş and Yayli (2025), en cuanto al análisis de la ciberseguridad desde la perspectiva de los valores y su

propuesta del poder blando (entendido como la capacidad que tiene un Estado de influir sobre otro, transformando la cultura, la educación y los valores de esa sociedad desde la atracción que se puede ejercer además de las estrategias diplomáticas) y la necesaria confianza en las relaciones internacionales digitales; aunado a ello, es pertinente referenciar a Taddeo et al. (2021), quien ha explorado los elementos éticos de la ciberseguridad desde la perspectiva de la atribución de responsabilidades de los ciberataques y la ética en la defensa cibernética. Desde estas perspectivas, no solo basta con conocer, también es imperativo reflexionar sobre los valores fundamentales que deben guiar la protección de la información y la seguridad en el entorno digital, proporcionando un marco ético para abordar los desafíos y dilemas éticos en la era digital.

El marco epistemológico de esta investigación no solo explica cómo se genera conocimiento en entornos digitales, sino que justifica la pertinencia de la arquitectura propuesta como un modelo que operacionaliza los principios de validez, fiabilidad y objetividad en el análisis de evidencia digital. Este enfoque aporta al debate filosófico sobre los límites del conocimiento en ciberseguridad, al situar la tecnología no solo como herramienta técnica, sino como un medio de construcción y validación del saber científico en la era digital.

2.2 Antecedentes de investigación

2.2.1 Caracterización de la actividad del hacker

Dado que el cibercrimen se ha convertido en una de las amenazas más significativas para la sociedad actual, resulta fundamental comprender en mayor profundidad a los autores de estos ataques, sus motivaciones, con el fin de diseñar estrategias más eficaces de prevención y respuesta.

El hacker es una figura que ha cambiado con el tiempo, sobre todo porque la caracterización del mismo está sujeta a la percepción que tiene la sociedad; en general y tal como lo expresan Bada and Nurse (2021), un hacker es una persona con conocimientos avanzados en informática y tecnologías de la información, capaz de comprender, manipular o modificar sistemas computacionales y redes con un alto nivel de destreza técnica.

De acuerdo con Möller (2023), la evolución del hacking ha generado una sofisticación en las técnicas utilizadas, lo que obliga a las instituciones a desarrollar marcos más complejos de detección, prevención y respuesta ante incidentes.

Conforme avanzan las tecnologías de la información y la comunicación, también se diversifican y perfeccionan los tipos de ciberataques, lo que incrementa la necesidad de comprender no solo las herramientas utilizadas, sino también a las personas detrás de estas acciones y sus métodos de operación. Para ello, es clave revisar el papel que juega el llamado hacking ético en la ciberseguridad actual.

En este contexto, el hacking ético ha emergido como una herramienta fundamental para fortalecer la ciberseguridad en las organizaciones (Godbole et al., 2024), esta práctica implica que expertos en seguridad informática identifiquen y corrijan vulnerabilidades en sistemas, redes y aplicaciones con la autorización de los propietarios, con el objetivo de prevenir posibles ataques maliciosos.

Los estudios sobre el perfil de los hackers han permitido establecer distintas tipologías, principalmente en función de sus motivaciones y comportamiento (Bada & Nurse, 2021). Entre las más reconocidas se encuentran:

- a) White Hat: Expertos en seguridad que utilizan sus habilidades con fines éticos y legales, apoyando a organizaciones en la detección y corrección de fallas de seguridad, y en ocasiones colaborando con autoridades para combatir amenazas cibernéticas.
- b) Gray Hat: Se ubican en una zona ambigua en cuanto a legalidad y ética. Actúan sin autorización, pero no necesariamente con intenciones maliciosas. Algunos están vinculados a movimientos activistas y utilizan sus habilidades como forma de protesta digital.
- c) Black Hat: Son individuos que emplean sus conocimientos con fines maliciosos. Se asocian con prácticas como el robo de datos, sabotaje digital o fraude, motivados por beneficio personal, venganza o reconocimiento dentro de comunidades clandestinas.

El estudio del perfil de los hackers permite comprender la diversidad de motivaciones y comportamientos que existen dentro de este grupo, lo cual es fundamental para diseñar estrategias efectivas de ciberseguridad. Reconocer esta tipología no solo ayuda a redefinir el concepto de hacker, frecuentemente asociado de forma exclusiva al delito, sino que también proporciona una base sólida para fortalecer la cooperación entre expertos en seguridad, instituciones y autoridades en la prevención y respuesta ante el cibercrimen.

2.2.2 Los ciberataques

En primer término, es fundamental comprender qué constituye un ciberataque. Según Aslan et al. (2023), un ciberataque puede definirse como cualquier acción deliberada destinada a comprometer la confidencialidad, integridad o disponibilidad de los activos digitales, mediante el acceso no autorizado a redes, sistemas o dispositivos.

Estos ataques pueden tener motivaciones diversas: económicas, ideológicas o estratégicas y abarcan una amplia gama de técnicas como piratería informática, ingeniería social, fraude, ataques de denegación de servicio y acoso en línea y ha servido de base para diversas investigaciones académicas (Bada & Nurse, 2021).

Por su parte Vega et al. (2022), en su tesis doctoral, sostiene que existe una amplia gama de ciberataques, que éstos pueden ser clasificados de la siguiente manera:

- a) Programa malicioso o malware. Se trata de algún programa que puede destruir datos, robar información, borrar archivos que comprometen la capacidad de funcionamiento del sistema operativo del equipo. Aquí se destacan: (a) Los caballos de Troya, que consiste en un grupo de aplicaciones que se esconden dentro de un “software legítimo” que se instala sin que el usuario lo sepa. Un troyano de acceso remoto puede crear una puerta de acceso en el equipo infectado para permitir que se descarguen otros tipos de software; (b) El ransomware. Es un tipo de software que utiliza una codificación sofisticada que le permite secuestrar datos o sistemas informáticos y, para su liberación, los ciberdelincuentes exigen el pago de un rescate por ellos; (c) El scareware. Se trata de un software que genera mensajes falsos para que el usuario descargue otra aplicación o suministre información confidencial; (d) El spyware. Es un programa que recoge información

confidencial (datos del usuario, datos bancarios, etc.) y los transmite; (e) Los rootkits. Son paquetes de software que le permiten al ciberdelincuente obtener el acceso al sistema operativo bajo la modalidad de administrador del equipo computacional o de la red informática; (f) Los gusanos. Son códigos que tienen la característica de auto replicarse y, así, se propagan e infectan a otros dispositivos.

- b) Ingeniería social. Son acciones cuyo objetivo es manipular e inducir a un usuario para que haga algo que no debería hacer, como suministrar información sensible, descargar alguna aplicación, entre otras; se destacan las siguientes: (a) El phishing. Por medio de mensajes genéricos (emails, mensajes de texto) diseñados para verse como provenientes de una fuente legítima, proporciona un enlace o hipervínculo que dirige al usuario a un sitio web falso donde suministra datos personales o de acceso a cuentas bancarias, entre otras; (b) Spear phishing. Sigue el mismo patrón de acción en cuanto a suministrar enlaces falsos, pero en este caso, está dirigido a una persona en particular, para ello, los ciberdelincuentes investigan a la potencial víctima usando, en términos generales, información proveniente de las redes sociales, entre otras fuentes; (c) Whale phishing. Es la modalidad de phishing corporativo, en este caso, los ciberdelincuentes simulan ser ejecutivos, vendedores o representantes de alguna empresa y suelen utilizar correos electrónicos empresariales para que las víctimas compartan información sensible o realicen algún tipo de transacción financiera.
- c) Ataques de denegación de servicio. Este tipo de ataque se basa en colapsar un sitio web o algún otro recurso de la red inundándolo con tráfico de datos para evitar su correcto funcionamiento. En general, este tipo de ataque se clasifica en: (a) Denegación de servicio, en este caso, el tráfico indeseado de datos proviene de una sola fuente; (b) Denegación de Servicio Distribuido (DDoS) y se trata de un tráfico de datos que inunda el website o el sistema informático desde diversas fuentes, para ello, el ciberdelincuente usa una botnet, es decir, una red formada por diversos dispositivos que puedan transmitir y generar tráfico en la red (Smartphone, tablets, IoT, portátiles, etc.) e infecta con malware el sistema informático que está siendo atacado.

- d) Compromiso de cuenta. Este tipo de ataque se caracteriza porque el ciberdelincuente, secuestra una cuenta y la usa para actividades maliciosas. Para tener acceso a una cuenta, puede hacerse con los datos usando algún tipo de phishing o comprando alguna base de datos; en algunas ocasiones, para romper el cifrado de las contraseñas, puede usar alguna herramienta como Hashcat y John the Ripper, en otras ocasiones, emplean lo que se conoce como ataque de fuerza bruta utilizando bots o algún script automatizado que genere y compruebe miles de combinaciones posibles de contraseñas hasta dar con la correcta.
- e) Ataque de intermediario. Es conocido como Man in The Middle (MiTM) y se trata de escuchas no autorizadas, para lo cual, el ciberdelincuente intercepta las comunicaciones entre usuario-servidor o entre varias personas, normalmente, cuando se utilizan las redes públicas de wifi ya que suelen carecer de sistemas de protección o estos son muy básicos.
- f) Secuestro de sesiones. Este tipo de ataque se caracteriza porque el ciberdelincuente irrumpe en la conexión entre un usuario y el servidor que desea atacar intercambiando su dirección de Protocolo de Internet (IP) con la del usuario para establecer una conexión que simula ser legal y, con ello, el ciberdelincuente entra al servidor para robar información, alterar emails u otra acción delictiva.
- g) Ataques a la cadena de suministro. En este tipo de ataque se basa en que los proveedores y empresas aliadas, normalmente comparten información conectándose a las redes de sus clientes, en este caso, el acceso que utiliza el ciberdelincuente es la red del proveedor para conectarse con sus objetivos y atacarlos.
- h) El túnel Domain Name System (DNS). Utilizando el sistema de nombre de dominio y las posibilidades de distribución de información en los dispositivos conectados a Internet o a una red privada, el tráfico o el software indeseado se oculta en los paquetes DNS y generan canales de comunicación alternos, de esta forma, se conecta el malware con un servidor de comando y control externo a donde se dirige la información que se extrae del servidor atacado o se suministra información falsa o se realiza algún otro tipo de ataque.
- i) Ataques de inyección Structured Query Language (SQL). Se trata del ataque que se hace a la base de datos que gestiona el almacenamiento de los datos de los formularios, la autenticación de los usuarios y seguridad del sistema informático,

entre otros; para ello, se envían comandos del lenguaje de consulta estructurado a través de campos que utiliza el usuario para suministrar información, los comandos pasan a la base de datos y extraen la data que el ciberdelincuente está buscando.

- j) Suplantación de DNS. En este caso, el ciberdelincuente dirige al usuario a un website falso, para ello, modifica el registro del DNS y reemplaza la dirección IP verdadera con una falsa.
- k) Scripts entre sitios. Se trata de un ataque con secuencia de comando entre sitios para insertar códigos maliciosos en un website específico y esos códigos se ejecutan en el navegador cada vez que un usuario lo usa, de esta manera, la información manejada por el usuario es robada.
- l) Ataques sin archivos. Se trata de ataques en donde el ciberdelincuente utiliza alguna vulnerabilidad del software para inyectar códigos maliciosos en la memoria del computador, de esta manera, pueden reconfigurar el equipo.
- m) Exploits de día cero. La vulnerabilidad del sistema informático es desconocida por el usuario o por los desarrolladores del sistema de seguridad, eso abre una ventana de tiempo mientras se identifica y “parchean” el sistema, lo que permite que el ciberdelincuente tenga acceso usando esa vulnerabilidad mientras el sistema no esté corregido.

Se estima que el tráfico IP global en Internet alcanzará una tasa anual de 1,3 zettabytes, a medida que más personas se conecten a la red mediante dispositivos móviles. Además, se proyecta que la cantidad de dispositivos conectados a Internet llegará a los 19000 millones. En este contexto, la realidad de las amenazas a la seguridad en Internet será extremadamente crítica para todos los usuarios, incluyendo particulares, empresas y gobiernos (Dhanvijay et al., 2025). La facilidad de acceso a herramientas ha permitido que incluso usuarios sin experiencia técnica puedan llevar a cabo estas actividades, lo que ha intensificado la situación descrita (Anand et al., 2023), por lo que las técnicas de seguridad convencionales tienen dificultades para contrarrestar estas amenazas en constante evolución, lo que resalta la necesidad de soluciones adaptativas (Anuradha et al., 2025).

2.2.3 Intrusión moderna

El rápido crecimiento de las comunicaciones digitales y el amplio intercambio de datos han convertido a las redes informáticas en parte integral de las operaciones organizacionales. Sin embargo, esta mayor conectividad también ha ampliado la superficie de ataque, introduciendo importantes riesgos de seguridad (Diana et al., 2025).

La intrusión en sistemas informáticos es una de las amenazas más significativas en el mundo digital actual; se entiende por intrusión el acceso no autorizado a sistemas y equipos informáticos, se trata del acto de acceder ilegalmente a un sistema informático o red con intenciones maliciosas y, conforme a lo señalado por Mohamed et al. (2023), el acceso puede ser de forma remota mediante el uso de algún software o de manera directa en el equipo.

Los ataques DDoS se han convertido en una amenaza frecuente y sustancial para la ciberseguridad, que puede inutilizar las infraestructuras de red de empresas y proveedores objetivo (Redekar & Bharati, 2025).

En el contexto de las intrusiones modernas, los ciberataques han evolucionado significativamente en sofisticación, volumen y velocidad, desafiando las capacidades tradicionales de detección y respuesta. Los Sistemas de Detección de Intrusos (IDS), aunque fundamentales para la defensa de las redes, enfrentan desafíos cada vez más complejos, entre los que destacan la escalabilidad, el rendimiento en tiempo real y la precisión en la identificación de amenazas, evitando tanto falsos positivos como falsos negativos (Diana et al., 2025).

Es importante recordar que ningún sistema es completamente seguro; sin embargo, con las medidas de seguridad adecuadas, es posible lograr que la intrusión sea mucho más difícil y así, limitar el daño que el intruso pudiese causar; entre las medidas de seguridad se consideran como básicas la implementación de firewalls e IDS, la actualización regular del software para corregir vulnerabilidades conocidas, la formación y los conocimientos de los usuarios respecto de las mejores prácticas de

seguridad, y la aplicación de fuertes políticas de seguridad, tanto empresariales como en el contexto gubernamental.

La ciberseguridad es un campo en constante evolución y debe mantenerse al día, dada las cambiantes tácticas de los ciberatacantes; por lo tanto, es esencial mantenerse informado y ser proactivo en la protección de los sistemas informáticos. Estas consideraciones hacen necesario que, en términos de ciberseguridad, se desarrolle un IDS que contenga las políticas de detección de intrusiones predeterminadas para supervisar todo el sistema, políticas de detección de ataques, exploración y las relacionadas con la regulación de tráfico, de aquí que el aprendizaje es esencial y una honeynet es una excelente opción, que es tan efectivo que muchas organizaciones los instalan para comprender la gran mayoría de los ciberataques (Saravana et al., 2025).

2.2.4 Interoperabilidad entre los sistemas

Un elemento importante para considerar cuando se trata de ciberseguridad y las honeynet, es la interoperabilidad entre los sistemas y la capacidad de compartir información con diferentes formatos.

La interoperabilidad entre los sistemas se ha convertido en una componente esencial dentro de los entornos tecnológicos orientados a la ciberseguridad, especialmente en arquitecturas distribuidas como las honeynets. Actualmente, representa uno de los desafíos más críticos en el diseño y gestión de sistemas colaborativos, dado que su correcta implementación permite la integración y el intercambio eficiente de información entre plataformas heterogéneas. Por su parte Sadeghi et al. (2023), identifican requisitos esenciales para lograr la interoperabilidad efectiva, entre ellos la estandarización de protocolos, la compatibilidad de formatos y la sincronización de procesos. Estos elementos no solo permiten una colaboración fluida entre distintos sistemas, sino que también contribuyen al desarrollo de infraestructuras más seguras, resilientes y preparadas frente a escenarios de amenaza creciente.

2.3. Bases Teóricas

2.3.1 *Honeypots*

En los últimos tiempos, existe un aumento considerable de las intrusiones y ataques desde Internet. Mundt and Baier (2023), expresan que está relacionado con una clara evolución de las herramientas, así como los nuevos métodos que han adaptado los ciberdelincuentes, utilizando técnicas muy avanzadas que crean y usan sus propias aplicaciones informáticas.

Según Kumar et al. (2025), el cambiante panorama de las amenazas a la ciberseguridad, se centran en ataques automatizados y las amenazas persistentes avanzadas que desafían las medidas de seguridad tradicionales, lo que supera los 2700 millones de dólares en pérdidas a nivel mundial (Peterson et al., 2021).

Este aumento en las intrusiones y ataques cibernéticos se han incrementado de manera significativa, exponiendo vulnerabilidades críticas en los sistemas informáticos. Una de las causas más comunes de estas brechas es la configuración inicial deficiente, especialmente el uso de contraseñas predeterminadas o credenciales débiles que no son modificadas por los usuarios finales. En este sentido Kaur and Ramkumar (2022), advierte que muchas organizaciones continúan adoptando prácticas de despliegue inseguras, lo que facilita el acceso no autorizado por parte de actores maliciosos. Estas deficiencias en la gestión de autenticación representan una puerta de entrada directa para una amplia variedad de ataques, entre ellos la escalada de privilegios, el acceso remoto no autorizado y la ejecución de código malicioso.

Se considera que la mayoría de los incidentes que ocurren en Internet no están dirigidos a dispositivos informáticos u organizaciones específicas; en cambio, su objetivo es la víctima fácil, que puede ser cualquier dispositivo que esté conectado y tenga una debilidad específica que el agresor esté buscando y sea capaz de aprovechar para acceder al sistema; de aquí que los desafíos de la seguridad de las redes empresariales ante intrusos, armados con herramientas para violar la seguridad, siguen creciendo (Mundt & Baier, 2023). Por lo tanto, las empresas modernas requieren soluciones de seguridad que no solo impidan el acceso del intruso más avanzado, sino que también, permitan el acceso, con el objetivo de ser controlado, supervisado e, incluso, que pueda

hacer las modificaciones necesarias para retomar el control de las aplicaciones y repararlas si es preciso. Para las infraestructuras de Tecnologías de la Información (TI) modernas, es esencial conocer no solo la presencia de un ataque, sino también la capacidad de rastrear su proceso, detalles y gravedad (Balogh et al., 2024).

En este sentido, al desviar la atención de posibles atacantes y desviar sus recursos en una red informática, los honeypots son una herramienta poderosa para proteger activos críticos dentro de una red, permitiendo un análisis detallado del comportamiento del ciberdelincuente, sus patrones de ataques y estrategias (Javadpour et al., 2024). Estas herramientas son clave en la vigilancia de seguridad avanzada y en la implementación de sistemas de alerta temprana, si se aplica correctamente (Vidhyasagar & Kartheek, 2023).

Los honeypots son sistemas informáticos diseñados deliberadamente para simular vulnerabilidades de la red y atraer a ciberatacantes, con el propósito de estudiar sus métodos y tácticas, así como identificar posibles fallos de seguridad en las redes reales (Salihu et al., 2024). Pueden personalizarse para replicar una copia falsa del entorno de producción real y resultar lucrativos para los atacantes, esto proporcionará información útil e inteligencia práctica que puede aplicarse a los sistemas de producción reales para protegerlos (Venugopal et al., 2021). Utilizan sistemas operativos reales replicados en la red de servidores implementada, se configuran en diferentes componentes informáticos, como software, redes, servidores, enrutadores (Kaur & Kaur, 2025), siendo una herramienta para ampliar la comprensión de los hackers y la seguridad de la red (Perkins & Howell, 2021). Mantener un alto nivel de engaño es crucial para que los honeypots permanezcan indetectables (Javadpour et al., 2024).

Su uso reduce el riesgo de ataques a redes y sistemas informáticos. Además, los honeypots permiten examinar los métodos que utilizan los atacantes para acceder a un sistema y revelar información importante sobre posibles vulnerabilidades de seguridad (Vidhyasagar & Kartheek, 2023). Un honeypot necesita emular los servicios esenciales en la red de producción y otorgar al atacante la libertad de realizar actividades adversas para aumentar su atractivo para el atacante, como lo menciona en

la tesis doctoral (Gallardo & Bazán, 2022). Este sistema facilita enormemente el trabajo de los analistas de seguridad (Balogh et al., 2024).

A medida que aumenta la complejidad de los ciberataques, las redes honeypot, que actúan como mecanismo de seguridad proactivo, se han convertido en un área importante de investigación en ciberseguridad (Ma et al., 2024).

2.3.2 Tipología de los honeypots

Un honeypot puede variar en complejidad, desde configuraciones básicas hasta entornos avanzados, dependiendo de su finalidad. Se clasifican según su uso y nivel de interacción, en honeypots de producción o investigación y honeypots de baja o alta interacción (Kulle, 2020).

Honeypots según su uso. En función del ambiente de implementación, se distinguen dos tipos principales:

- a) Honeypots de producción. Diseñados para integrarse en entornos operativos reales con el objetivo de proteger los sistemas organizacionales en tiempo real. Funcionan en paralelo con la infraestructura tecnológica activa, lo que permite recibir y registrar ataques de forma continua, las 24 horas del día, durante todo el año. Su propósito principal es mejorar la seguridad sin interrumpir las operaciones normales de la organización.
- b) Honeypots de investigación. Utilizados principalmente con fines educativos y de análisis, estos honeypots no se despliegan junto a sistemas reales, para evitar cualquier interferencia operativa. Su objetivo es identificar patrones de ataque y estudiar diversas amenazas cibernéticas, lo que permite entender mejor el comportamiento de los atacantes y desarrollar estrategias preventivas más eficaces (Kulle, 2020).

Honeypots según su nivel de interacción. Una segunda clasificación relevante considera el nivel de interacción permitido entre el atacante y el sistema simulado. Los diseños de honeypots comúnmente utilizados son los de alta y baja interacción, que difieren en el grado de acceso e interacción que se le permite al atacante.

- a) Honeypots de baja interacción. Son sistemas más simples que simulan únicamente ciertos servicios o funcionalidades básicas de un sistema operativo. Debido a su bajo consumo de recursos, son más eficientes y fáciles de implementar. Sin embargo, presentan limitaciones importantes en la recopilación de datos, ya que ofrecen menos oportunidades para observar el comportamiento completo de un atacante. Además, pueden ser fácilmente identificados por atacantes experimentados (Kocaogullar et al., 2025).
- b) Honeypots de alta interacción. Sistemas que recrean entornos complejos, incluyen múltiples aplicaciones, servicios y métodos de almacenamiento, permitiendo una interacción más profunda con el atacante. Gracias a su nivel de realismo, son capaces de generar una mayor cantidad de datos, incluso en tiempo real, y permiten redirigir ataques hacia servicios específicos con el fin de estudiar detalladamente las acciones maliciosas. Su ventaja radica en que pueden revelar nuevas técnicas de ataque, incluso aquellas no anticipadas. No obstante, su implementación exige una estricta supervisión, ya que cualquier debilidad no detectada podría ser explotada como una puerta de entrada a los sistemas reales (Martínez & Moreno, 2024).

Yang et al. (2023), manifiestan que un honeypot es un mecanismo de defensa activo en la protección de la seguridad de la red. Su objetivo principal es recopilar información sobre intrusiones. Sin embargo, debido a la complejidad de los ataques de penetración y a la diversidad de vulnerabilidades del host, un honeypot simple no puede satisfacer las necesidades de defensa, por lo que ha surgido un sistema honeynet compuesto por una combinación de varios honeypots. Es una de las formas más eficaces de obtener información sobre los atacantes y sus acciones lo antes posible (Kashtalian et al., 2021), cuenta con todos los dispositivos de networking necesarios como routers y switches, por ello, puede replicar cualquier escenario de red.

2.3.3 Honeynet

Una honeynet es una arquitectura avanzada de ciberseguridad diseñada intencionadamente para ser atacada, con el objetivo de analizar el comportamiento de los agentes maliciosos, identificar sus herramientas y metodologías, y recopilar inteligencia útil para la defensa cibernética. Esta infraestructura está compuesta por un

conjunto interconectado de honeypots (Javadpour et al., 2023), los cuales simulan servicios y sistemas vulnerables dentro de una red lógica aislada de los entornos productivos de la organización (Qin et al., 2023). La finalidad principal de esta estructura es la recopilación activa de datos sobre intrusiones reales, proporcionando un entorno controlado para el estudio del comportamiento de los atacantes en tiempo real (Acosta et al., 2022).

A diferencia de los honeypots individuales, las honeynets permiten una interacción de mayor nivel con los ciberatacantes, posibilitando el análisis detallado de sus tácticas, técnicas y procedimientos. Su implementación implica una configuración estratégica que abarca múltiples plataformas, servicios y sistemas operativos, adaptados según la complejidad requerida por los objetivos del sistema de vigilancia en un entorno de red controlado. La honeynet opera bajo un monitoreo constante mediante sistemas de control y registro, asegurando tanto la integridad de la información recolectada como la contención del actor malicioso. Este enfoque permite una captura activa de los atacantes y fortalece la efectividad general del sistema. Además, a diferencia de defensas como firewalls, detección de intrusiones, las honeynets proporcionan una defensa más proactiva y engañosa, lo que contribuye a detectar y analizar amenazas de manera anticipada (Zhang et al., 2024). La ventaja de una honeynet sobre un honeypot simple es que simula red real y tiene un área de captación más grande (Gallardo & Bazán, 2022).

Estructura de la honeynet. Debido a la diversidad de formas de diseño e implantación, además de la flexibilidad de la honeynet, no existe una estructura exclusiva para su desarrollo, pero si hay un diseño básico sobre el cual se hace una serie de modificaciones según las necesidades. Honeynet Project señala que una honeynet debe contar, esencialmente, con:

- a) Control de datos: Se trata de la entrada a la honeynet, por ello, el control de los datos desde y hacia los honeypots que conforman la honeynet, debe contar con restricciones y la implantación del control se hace por capas sumando diferentes mecanismos que se superponen, considerando también, que si éste falla, debe producirse un bloqueo completo de la honeynet, separándola de la red real, es decir, el control de datos debe contar con un procedimiento de “fail close” como

respuesta para evitar que la honeynet sea usada como arma para atacar los sistemas y las redes que, precisamente, se deben proteger.

- b) Captura de datos: Se plantea el seguimiento y el registro de todas las actividades que hace el ciberatacante en los honeypots; para ello, hay que implementar diferentes mecanismos de captura y de registro en distintos niveles o capas para poder contar con una amplia gama de posibilidades de datos capturados. Es esencial considerar que la diversidad de mecanismos de captura y registro no solo permite obtener más datos, sino que también permite evitar puntos únicos de fallos en la red.
- c) Análisis de datos: Este proceso es el estudio de todos los datos recogidos por los honeypots para generar la información en cuanto a los tipos de ciberataques, los más comunes, los nuevos, los patrones que siguen los ciberdelincuentes, identificar los puntos débiles de la red, entre otros; de ahí que el especialista en ciberseguridad necesita contar con las suficientes herramientas de análisis y de recuperación de data para contrastar los datos, incluso, contando con datos históricos.

Ahora bien, es menester señalar que la organización Honeynet Project, ha generado una serie de propuestas respecto a la arquitectura de las honeynets, estas son Generación I, Generación II y Generación III.

Generación I. Esta arquitectura vio luz en el año 1999 y su diseño se basa en controlar a los ciberatacantes y, para ello, se sitúa la red delante del dispositivo de control de acceso, mientras que el sistema o la red original, queda detrás.

Normalmente, el dispositivo de control es un firewall que logra separar la honeynet, la red de producción y el acceso a Internet (básicamente, son 3 capas o niveles); así, el firewall se usa como la herramienta que permite controlar las conexiones que entran a la honeynet y las que salen de ella, mientras que el router sirve como apoyo al proceso de filtrado de data y control de acceso, permitiendo, incluso, un determinado número de conexiones o de accesos a los honeypots, lo que dependerá de las necesidades de protección o de aprendizaje que se hayan planteado. A continuación, se presenta, en la arquitectura básica de una Honeynet Generación I, (ver Figura 1).

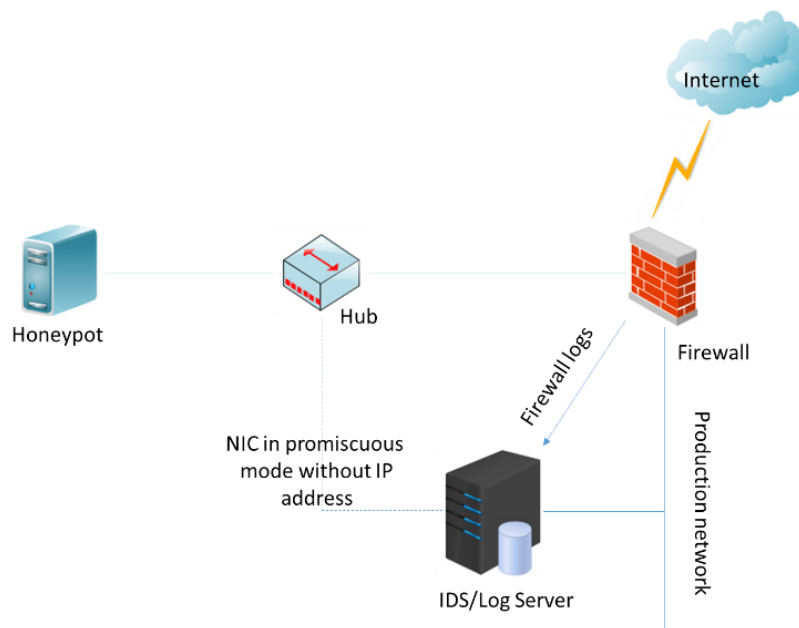


Figura 1. Arquitectura de Generación I

Fuente. Honeynet Project

Considerando que tanto el control como la captura de datos son procesos esenciales en el funcionamiento de la honeynet, se presenta a continuación las precisiones técnicas al respecto, conforme a lo señalado por Watson and Riden (2008), en base a lo establecido por Honeynet Project.

Control de datos

Este proceso tiene dos objetivos: (a) Mostrar un escenario lo suficientemente real y seguro para que los intrusos sientan que tienen todo el ataque bajo control y actúen como lo hacen a la hora de comprometer a otras máquinas víctimas; (b) Hacer un entorno que garantice que el daño causado no puede superar el entorno de la honeynet.

El control de datos se divide en dos categorías:

- a) Bloqueo de conexión: evita el exceso de conexiones desde honeynet. El grado de aceptación o negación está directamente relacionado con lo que se quiere aprender y el riesgo que se es capaz de asumir. Habilitar conexiones fuertes ofrece la posibilidad de aprender más pero también genera más riesgo.
- b) Limitación de conexión: tiene el propósito de mitigar la inundación de conexiones salientes y limitar el ancho de banda.

Captura de datos

La captura de datos es una funcionalidad que permite recopilar y almacenar evidencias de la actividad de la red y de las máquinas intervinientes en los ataques a las que fueron expuestas. La captura de datos no solo almacena el registro o el tráfico de red; en cambio, es una combinación de monitoreo de la actividad, observación del “modus operandi” del atacante y la capacidad de reconocer varios perfiles para interpretar el ataque o ataques de manera más efectiva.

Sin la clave, las herramientas de captura de datos en la red no pueden ver el canal, si los servicios confiables como Secure Shell (SSH), un cliente encriptado o Secure Sockets Layer (SSL) están instalados en la máquina de la víctima.

Generación II. Esta propuesta de arquitectura se dio a conocer en el año 2002 y su funcionamiento tiene su base en el uso de un gateway que puede considerarse como una puerta de enlace que permite separar la honeynet del resto de las estructuras tecnológicas instaladas y, su principal ventaja es que al usar el gateway, no hay variaciones en los tiempos de respuestas (o son, apenas, milésimas de segundos) además de no existir direcciones Media Access Control (MAC) que se asocien al gateway ni saltos de enrutamiento. La otra ventaja es que el gateway permite integrar el control y captura de datos.

A continuación, se presenta la arquitectura Generación II, (ver Figura 2).

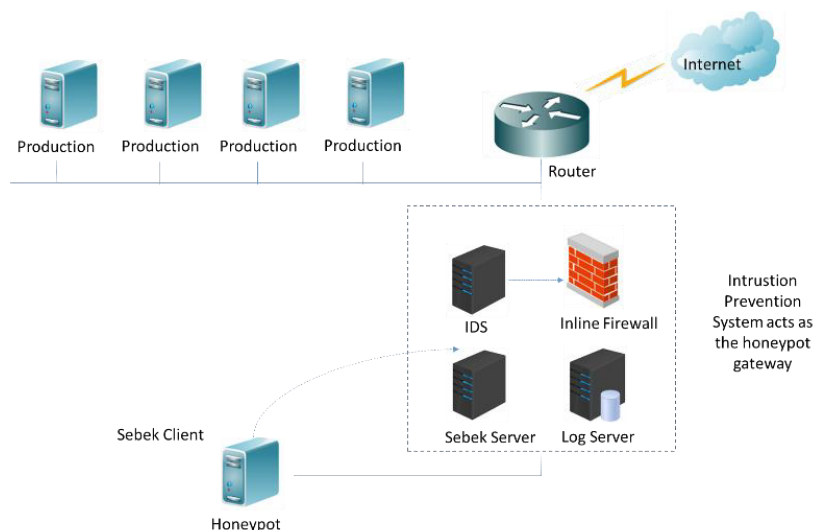


Figura 2. Arquitectura de Generación II
Fuente. Honeynet Project

Para los procesos básicos de control de datos y captura de datos, hay que considerar la siguiente arquitectura para las Honeynet Generación II, según lo indicado por Honeynet Project:

Control de datos

Todo el tráfico desde y hacia honeynet fluye a través del honeywall, lo que permite la intervención en el firewall y en los componentes IDS o Sistema de Prevención de Intrusiones (IPS) para los datos que fluyen a través de Internet.

La primera capa de control consiste en el manejo de las conexiones salientes a través del firewall; permite que las conexiones lleguen a una puerta determinada, a diferencia de las honeynets de Generación I, que permiten solo valores bajos, lo que hace corta la residencia de los atacantes (Spitzner, 2002). La segunda capa de control de datos consiste en el IDS y/o IPS.

Este tipo de honeynet requiere dos capas de control: la capa en el nivel de control de la puerta de enlace de red y el IDS y/o IPS. Estas capas se complementan entre sí, proporcionando lo siguiente:

- Restricción y control de la actividad que realizarán los atacantes dentro de la honeynet.
- Ampliación de la apariencia de “libertad” para los atacantes que se encuentran en el interior. Esto permite obtener más información sobre el ataque.

Captura de datos

Las honeynets de Generación II aplican las mismas tácticas que las honeynets de Generación I, pero mejoran los métodos y las herramientas; la recopilación de información se realiza en la capa 3. Estas mejoras son el firewall, IDS/IPS y honeypots.

La captura de datos se realiza para el IDS residente y el firewall. Los registros del firewall ofrecen información sobre todas las conexiones entrantes y salientes; además, el IDS proporciona alertas sobre los patrones de ataques conocidos. Esta primera etapa reporta toda la actividad dentro de la honeynet.

Generación III. Esta arquitectura se propuso en el año 2005 y su fundamento esencial es resolver las debilidades de la Generación II usando mecanismos de integración. En este sentido, mantiene las mismas características que las de su antecesora, pero unificando los datos proporcionados por los diferentes honeypots durante el seguimiento y el registro de todas las actividades y los relaciona con los datos proporcionados por las herramientas del control de datos, de esta manera, el especialista puede saber dónde se generó la alerta de intrusión y proceder a rastrear los paquetes de datos que se relacionan con esa conexión específica.

Considerando que los datos provienen de diferentes fuentes y con distintos formatos, es esencial unificar herramientas para poder integrar la data, mejorar los análisis, los tiempos de respuesta y la toma de decisiones (Silva & Rafael, 2017).

Se expone la arquitectura de la Honeynet Generación III, (ver Figura 3).

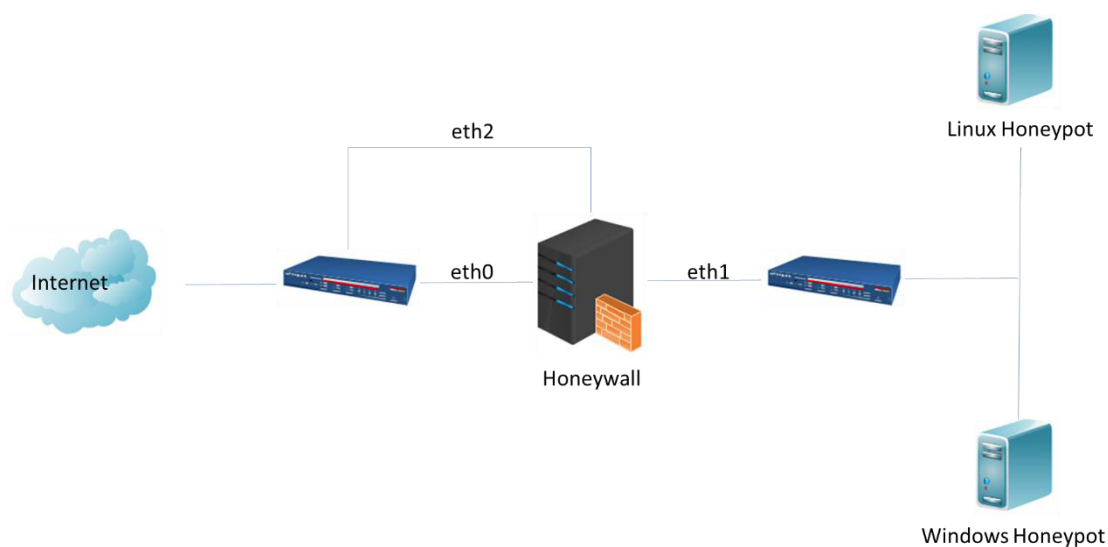


Figura 3. Arquitectura de Generación III

Fuente. Honeynet Project

Respecto al funcionamiento de los procesos básicos de control y captura de datos, según lo expone Honeynet Project se tiene que:

Control de datos

El objetivo es verificar qué es capaz de hacer el atacante al entrar y salir de la honeynet. Honeywall es muy flexible e incluye la capacidad de limitar las conexiones entrantes y salientes mediante el uso de IPTables y puede realizar Network Address Translation (NAT) (Gallardo & Bazán, 2022).

Las puertas de enlace honeywall de Generación II y Generación III operan en la capa de enlace, que toma decisiones de reenvío de tráfico basadas en la tabla de direcciones MAC para redirigir el tráfico de una interfaz a otra. Cuando la puerta de enlace actúa como un puente, no hay redireccionamiento ni disminución de los paquetes, Time to Live (TTL). Esto le permite actuar como un dispositivo de filtrado invisible, más difícil de detectar por los atacantes.

El control de datos interno garantiza que los honeypots no se puedan usar para atacar otros sistemas y redes informáticas.

Además, se debe identificar el nombre de las interfaces internas del host. De forma predeterminada, es una interfaz eth1 y la interfaz externa tiene como valor predeterminado eth2. Considera las capacidades de Snort Inline para rechazar ataques salientes conocidos, las herramientas para el control de datos son:

- Puente de capa 2 (honeywall)
- IPtables (limitación de paquetes y/o control de tráfico de red)
- Snort Inline (manipulación de paquetes de red)

Captura de datos

Su objetivo es capturar toda la actividad del atacante sin ser detectado. La actividad en honeynet es registrada por varias herramientas. El firewall registra todas las conexiones entrantes y salientes. De forma predeterminada, un proceso de snort captura toda la actividad de la red y todo el contenido de los paquetes vistos por la interfaz de red interna.

Esto incluye toda la actividad de sebek que permite mostrar incluso el tráfico cifrado sin tener la clave. Una instancia adicional de snort, escucha en la interfaz interna y genera alertas IDS en modo completo y rápido. Para esta generación de honeynet, las herramientas para la captura de datos son:

- a) Iptables (registros de iptables)
- b) Snort (alertas IDS)
- c) pOf (identificación pasiva de SO)
- d) Sebek (Captura de datos avanzada)
- e) Tcpdump (captura de tráfico de red)

La Generación III de las honeynets han mejorado el análisis de los datos en comparación con las arquitecturas anteriores. En sí, la forma más fácil de realizar el análisis y examinar el tráfico de red y los datos recopilados en honeynet es buscar conexiones por fecha, hora, direcciones IP y/o puertos que se muestran en la interfaz GUI Walleye.

Los resultados se pueden devolver como un archivo de paquete capturado - Packet Capture (PCAP) mediante el uso de herramientas de análisis de tráfico de red para proceder a examinar los datos recopilados, a través de las siguientes herramientas:

- MySQL (base de datos honeynet para la colección)
- Argus + Hflow (recopilación de datos, flujo de tráfico)
- Swatch (registros de firewall y alertas de IDS)
- Walleye (interfaz gráfica de usuario)

La Tabla 1, muestra una comparación de las diferencias reflejadas entre estas tres generaciones de honeynet.

Tabla 1. Comparación de las diferentes generaciones

	<i>Generación I</i>	<i>Generación II</i>	<i>Generación III</i>
Control de Datos	Por medio de un firewall	A través de un honeywall, control de puerta de enlace y una herramienta IDS y/o IPS	Utiliza un honeywall altamente flexible con IPTables, lo que mejora la flexibilidad, la gestión y la seguridad
Captura de datos	Mediante el uso de un IDS en modo sniffer. Captura logs y modus operandi	A través de un honeywall métodos y herramientas inteligentes para recopilar datos. Estas herramientas capturan registros de firewall, tráfico de red y actividad del sistema	Captura toda la actividad del atacante, utilizando herramientas de recopilación de datos, captura avanzada de datos - Sebek
Análisis de datos	N/A	N/A	Correlaciona los datos para enviarlos de forma segura a un servidor centralizado para su almacenamiento y análisis.
Nivel de Operación	Net Layer	Link Layer	Link Layer
Detección	Fácilmente	Difícilmente	Difícilmente
Información cifrada	No	Si	Si
Interfaz web	No	No	Sí
Informes automatizados	No	No	Sí
Tiempo de monitoreo	Bajo	Alto	Alto
Tiempo de mantenimiento	Alto	Bajo	Bajo

Fuente. El autor.

Tendencias en el desarrollo de las honeynets. El avance de las tecnologías impacta en el contexto de la ciberseguridad, lo que implica necesariamente, generar cambios en este ámbito, garantizando el resguardo de la información de las organizaciones y de los usuarios, entonces, se presentan los elementos definitorios de la tendencia de las honeynet virtuales, la idea consiste en combinar todos los elementos físicos de una

honeynet dentro de una sola computadora mediante el uso de un software de virtualización.

Dado que los ciberataques evolucionan constantemente en cuanto a sofisticación y técnicas empleadas, es fundamental desarrollar nuevos modelos de seguridad que amplíen el espectro de protección, permitiendo una respuesta más eficaz frente a amenazas emergentes. Estos modelos deben orientarse no solo a incrementar la cobertura defensiva, sino también a optimizar el uso de recursos disponibles, reduciendo la carga operativa y mejorando el rendimiento general de los sistemas de seguridad (Pieterse et al., 2024).

La implementación de una honeynet en un único equipo físico es posible gracias a la virtualización, tecnología que permite desplegar múltiples sistemas operativos de manera simultánea sobre una misma infraestructura de hardware. Este enfoque, conocido como honeynet virtual, se caracteriza porque cada sistema comprometido aparenta operar en una máquina independiente, aunque en realidad todos los entornos virtuales comparten los mismos recursos físicos del host. Esta arquitectura permite simular una red compleja y diversificada en cuanto a plataformas y servicios, facilitando la observación de ataques sin requerir múltiples dispositivos físicos, lo que optimiza costos, espacio y mantenimiento. Honeynet Project señala la siguiente clasificación:

- a) Honeynets autocontenidas. También se les conoce como self-contained virtual honeynet y, en este caso, todo el software se ejecuta en una misma máquina. De manera general, se instala el sistema operativo real o anfitrión y, posteriormente, el software de emulación virtual; una vez hecho esto, se instala el sistema operativo para cada emulador y se culmina configurando las diferentes honeynets. La ventaja de este tipo de honeynet es la utilización de pocos recursos tecnológicos con lo cual, se reducen los costos, además, su instalación y configuración son sencillas; no obstante, como solo hay un punto único de servicio (un solo equipo real), se puede perder toda la protección si el ciberdelincuente logra tener el control de la máquina o si ésta se daña. Se presenta la esquematización de este tipo de honeynet virtual (ver Figura 4).

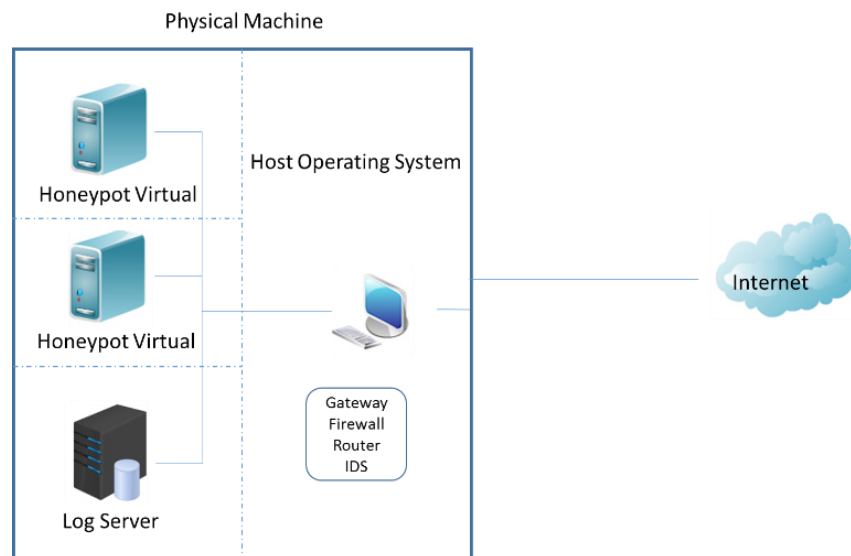


Figura 4. Honeynets virtuales autocontenidas

Fuente. Honeynet Project

- b) Honeynets híbridas. Conocidas como hybrid virtual honeynet, en este diseño el honeywall, conformado por el punto de acceso, control y captura de datos, se separa los honeypots de la honeynet. Este diseño provee flexibilidad y mejor seguridad ya que el control de acceso está fuera de la honeynet virtual; sin embargo, requiere un poco más de recursos tecnológicos y su configuración puede ser algo compleja. Este diseño se expone (ver Figura 5).

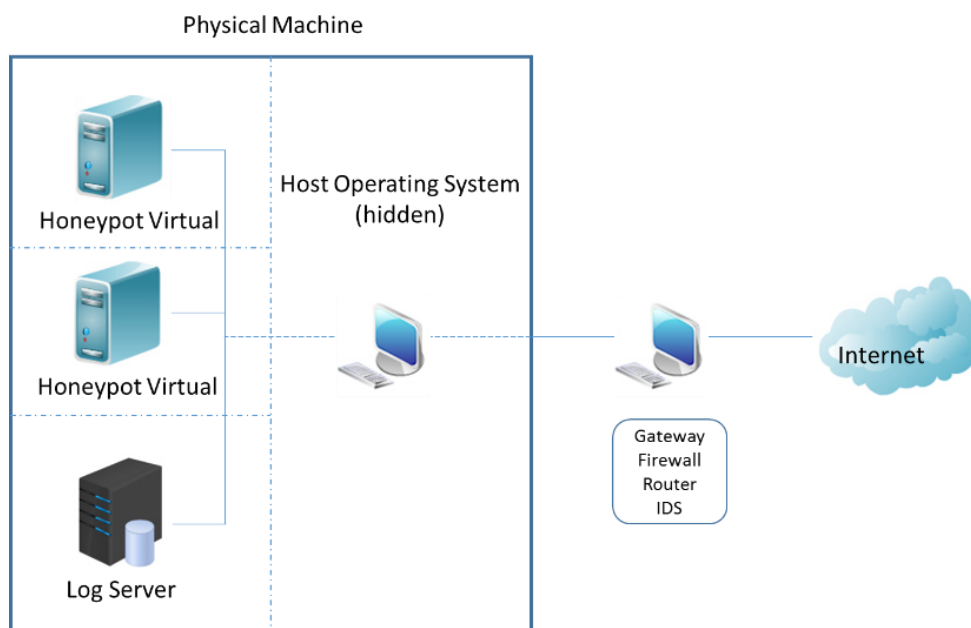


Figura 5. Honeynet virtual híbrida

Fuente. Honeynet Project

En consideración a ambos tipos de honeynets virtuales, es menester considerar algunos elementos adicionales:

- a) El hardware limita el número de sistemas operativos que se pueden simular.
- b) La potencia del equipo, considerando Memoria de Acceso Aleatorio (RAM) y caché además del procesador y coprocesadores, incide en la cantidad de honeynets que se pueden generar.
- c) El uso de técnicas tipo “fingerprinting” pueden poner en evidencia el verdadero sistema operativo y las simulaciones.

La Tabla 2 muestra una comparación de las funciones virtuales de honeynet.

Tabla 2. Comparación de características de honeynet virtual

	Virtuales Auto Contenidas	Virtual híbrida
Portable	Sí	No
Configuración	Plug and Catch	Less Involved
Costo	Bajo	Medio
Flexibilidad	Limitado	Moderado
Seguridad	Dependiente del software	Seguro
Detectabilidad	Esfuerzo razonable	Esfuerzo razonable
Riesgo	Alto	Medio
Tiempo de mantenimiento	Bajo	Medio
Datos obtenidos	Limitado	Alto
Administración	Centralizada	Distribuida
Espacio físico	Bajo	Alto

Fuente. El autor

Otra tendencia es desarrollar honeynets distribuidas, éstas son particularmente útiles cuando se trata de corporaciones con presencia en diferentes países, lo que significa que el equipo de ciberseguridad estará en varios puntos geográficos igual que sus redes informáticas. Ante la necesidad de compartir información que sea generada por las respectivas honeynets, (ver Figura 6), éstas se conectan con el repositorio central, decir, se centraliza en un repositorio la comunicación y el vaciado de la información. Para que este esquema funcione es esencial que las arquitecturas de las honeynets involucradas sea estable y que la interoperabilidad de los sistemas sea lo suficientemente amplia para permitir las interconexiones y las comunicaciones.

Este planteamiento permite que se comparta la información de todas las honeynets y proporciona una mejor base de conocimientos en términos de ciberseguridad.

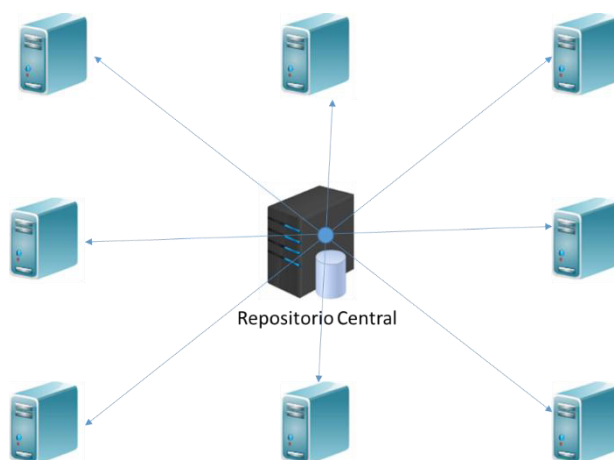


Figura 6. Esquema de una honeynet distribuida

Fuente. Honeynet Project

Configuración de las honeynet.

- a) Configuración de red. El honeypot necesita un adaptador de red para tener una conexión. Por lo tanto, se asigna una dirección IP estática al honeypot dentro del mismo segmento. Se configura un conmutador virtual a la interfaz eth0, utilizando VMware Workstation, que corresponderá a la red interna de la honeynet.
- b) Sincronización de tiempo. Está configurado para que las marcas de tiempo de los registros se correspondan correctamente y los datos puedan analizarse. Esto se hace a través de un protocolo de tiempo de red - Network Time Protocol (NTP).
- c) Creación de usuarios. Se crea un usuario en el sistema para que pueda ser un punto de entrada al honeypot. La contraseña será intencionalmente débil.
- d) Instalación y configuración de MYSQL. MySQL está configurado para ser accesible desde la red. Se mantendrá la configuración predeterminada. Después de los pasos iniciales, se puede inicializar el servidor del sistema de base de datos.
- e) Instalación y configuración de sebek. El código fuente del cliente sebek se descarga para compilarlo e instalarlo en el honeypot. El código fuente está disponible en el sitio web del Proyecto Sebek (McCarty, 2003).
- f) Configuración de tablas IP. Por defecto, el firewall no bloqueará ningún paquete y permitirá todo el tráfico dentro del honeypot. El control del tráfico a la honeynet

será controlado por las IPTables del honeywall, donde se insertarán reglas que den un mayor control sobre el tráfico que atraviesa el honeywall.

- g) Configuración de ROO de honeywall. Para establecer una máquina virtual, es necesario crear un honeywall en honeynet. Se configuran tres interfaces para formar el honeywall, la interfaz de administración, la interfaz externa de honeynet y la interfaz interna de honeynet. Honeywall Roo crea automáticamente una interfaz puente. Detecta y asocia las interfaces eth0 y eth1 como las interfaces que deben componer el puente. Las interfaces no tienen una dirección IP. Este hecho, combinado con el uso de una interfaz de puente, hace que la honeynet sea muy difícil de detectar. Honeywall Roo ejecuta un script llamado rc.firewall que se encarga de configurar esta interfaz de puente y crea las reglas para el firewall y otras configuraciones.

Herramientas de recopilación de datos utilizadas en una honeynet. La selección de las herramientas de recolección de datos que se describen a continuación se realiza con el objetivo de ser un complemento en la captura y monitoreo de información en tiempo real, identificando vulnerabilidades de seguridad, con capacidad de filtrar los datos, para tomar decisiones de manera oportuna y adecuada.

Snort. Procesa el tráfico de red y lo compara con un conjunto de reglas previamente definidas para identificar patrones maliciosos o sospechosos. En una honeynet, snort utiliza una combinación de su motor de detección, reglas personalizadas, preprocesadores, sistemas de alerta y registros para recopilar información detallada sobre los ataques. Estas herramientas permiten un análisis exhaustivo del tráfico malicioso, brindando a los administradores de seguridad la capacidad de estudiar el comportamiento de los atacantes, mejorar la defensa de la red y ajustar las contramedidas de seguridad en comparación con otros IDS (Ghazi et al., 2024).

P0f. Es una herramienta de fingerprinting pasivo que permite la identificación de sistemas operativos, configuraciones de red y otros atributos del origen del tráfico sin generar paquetes adicionales, lo que la hace ideal para entornos donde la discreción es esencial, como en una honeynet. En redes empresariales complejas, su uso resulta crucial para la detección temprana de amenazas. Integrado en un marco de seguridad

pasiva, P0f puede colaborar con honeypots para atraer atacantes, filtrar eficientemente los datos capturados y permitir un análisis profundo en tiempo real. Esto optimiza la detección de vulnerabilidades y fortalece las respuestas ante ciberataques dirigidos (Suleiman et al., 2022).

Argus. (Audit Record Generation and Utilization System) es una herramienta avanzada de monitoreo de tráfico y análisis de flujo de red diseñada para capturar y auditar el tráfico en tiempo real o de forma retrospectiva. En una honeynet, argus se utiliza para recopilar datos sobre las conexiones de red y los flujos de tráfico, proporcionando información detallada y estadísticas sobre el comportamiento del tráfico en el entorno simulado.

Sebek. Es una herramienta de recopilación de datos específicamente diseñada para honeynets, desarrollada por el Honeynet Project. Su principal objetivo es monitorear y registrar la actividad de los atacantes dentro de sistemas honeypot, proporcionando una visión detallada de sus acciones sin que ellos sean conscientes de que están siendo observados. A diferencia de otras herramientas de monitoreo de red, sebek se instala en el sistema honeypot para capturar actividades que otras herramientas no podrían registrar, como las interacciones a nivel de kernel y la extracción de información directamente desde el sistema operativo comprometido.

Gestión de registros, alertas e informes. Uno de los puntos más importantes es la gestión de la información recopilada por el honeypot. Sus características incluyen la gestión de logs, el uso de la base de datos, el almacenamiento remoto de logs, el formato de la información almacenada y el mecanismo necesario para encontrar toda esta información, es decir, consultar la base de datos, examinar los logs en modo texto, opciones de visualización web y generación de informes. A través de la adecuada gestión de alertas se evita un monitoreo continuo de los registros para encontrar indicios de actividad sospechosa.

2.4 Comparación de arquitecturas honeynet

En el marco de la ciberseguridad, las arquitecturas honeynet han evolucionado para responder a entornos cada vez más complejos y distribuidos, a lo largo de los últimos años, diversas propuestas han buscado optimizar estos sistemas en términos de rendimiento, escalabilidad y análisis de la evidencia digital.

Se realiza un análisis comparativo de propuestas de arquitectura honeynet, a partir de parámetros de comparación como: Nivel de recursos, criterio utilizado para clasificar las soluciones honeynet en físicas, virtuales y mixtas; Nivel de adaptabilidad, capacidad de modificar dinámicamente la configuración y la topología de una honeynet, puede ser estática o dinámica; Ubicación física, puede ser local o remota (Fan et al., 2015); además se han considerado: Tipología de honeypot, según su uso y su nivel de interacción; Centralización de datos; Integración de datos; Escalabilidad; Herramientas de captura; y Latencia, con el propósito de identificar los avances, limitaciones y el aporte innovador de las propuestas desarrolladas más representativas en la literatura y prácticas actuales, comparándolas con la propuesta desarrollada, especialmente en lo referente a la integración de los datos para el análisis de evidencia digital.

Se presenta el trabajo elaborado por Ma et al. (2024), que introduce honeynets dinámicas con control y redirección flexibles del tráfico malicioso, se caracteriza por su nivel alto de interacción, sin embargo, con lo que respecta a la captura de datos no posee centralización ni integración de los mismos.

La arquitectura honeynet presentada por Yu et al. (2024), denominada HoneyFactory, consiste en una arquitectura modular que utiliza contenedores para recolectar tráfico dentro del honeypot, uno de los módulos es el de recopilación de datos de honeynet que obtiene registros dentro del honeypot mediante un mecanismo de montaje de volumen de contenedor que monitoriza la tarjeta de red del servidor para recopilar todo el tráfico enviado a la honeynet y analizar el comportamiento e información de tráfico de los atacantes. Aunque centraliza los datos no realiza la integración de los mismos.

De Pace et al. (2023), proponen una arquitectura original de honeypot ciberfísico que integra a la perfección los componentes de TI y Tecnología Operativa (TO), la

honeynet puede implementarse en la Internet pública para recopilar ataques maliciosos reales, no posee centralización ni integración de datos.

La propuesta presentada por Kohli et al. (2023), consiste en Redes Definidas por Software (SDN), para resolver el problema de la inflexibilidad de las honeynets convencionales en el entorno de red organizacional, no incluye centralización ni integración de datos.

Noaman et al. (2019), proponen una arquitectura que utiliza un agente de software, un servidor proxy seguro y un servidor de clúster y monitorización de comunicaciones en tiempo real, una de sus ventajas es la asignación segura de puertos señuelo y el reenvío seguro de tráfico malicioso por parte del agente de software a los servidores honeynet. Posee un proceso de centralización de los datos, pero no realiza integración.

La propuesta presentada por Lucchese and Merro (2024), en su tesis doctoral realiza una arquitectura HoneyICS, para estudiar el comportamiento de los atacantes y mejorar la seguridad de los Sistemas de Control Industrial (ICS). Su arquitectura no incluye centralización ni integración de datos.

La Tabla 3, muestra la comparación de las propuestas de arquitectura honeynet

Tabla 3. Comparación de arquitecturas honeynet

		Arquitecturas Honeynet						
		VDH	HoneyFactory	Smart Grid	SIHA - SDN	Software agents	HoneyICS	Propuesta Integración de datos
		(Ma et al., 2024)	(Yu et al., 2024)	(De Pace et al., 2023)	(Kohli et al., 2023)	(Noaman et al., 2019)	(Lucchese & Merro, 2024)	
Tipología de honeypot	Según su uso Según su nivel de interacción	Producción	Producción	Investigación	Producción	Producción	Producción	Producción
		Alta	Alta	Alta	Baja	Alta	Alta	Alta
Nivel de recursos		Mixta	Mixta	Mixta	Física	Mixta	Mixta	Mixta
Nivel de adaptabilidad		Dinámica	Dinámica	Dinámica	Dinámica	Dinámica	Dinámica	Dinámica
Ubicación física		Remota	Remota	Remota	Local	Remota	Remota	Remota
Centralización de datos		No	Si	No	Si	Si	No	Si
Integración de datos		No	No	No	No	No	No	Si
Escalabilidad		Si	Si	Si	Si	Si	Si	Si
Herramientas de captura		Firewall	Packetbeat tcpdump	HoneyD	Suricata	Firewall	Nmap	Firewall Sebek Snort Argos p0f
Latencia		Baja	Baja	Baja	Baja	Baja	Baja	Baja

Fuente. El autor

El análisis revela que, si bien todas las arquitecturas comparten características comunes como en nivel de adaptabilidad, es decir tienen la capacidad de modificar dinámicamente la configuración y la topología de una honeynet; la escalabilidad, es decir que las arquitecturas honeynet puede crecer horizontal o verticalmente, ideal para ambientes de producción o investigación; además que todas estas arquitecturas poseen una baja latencia, lo que dificulta que los adversarios identifiquen el entorno falso; con lo que respecta al nivel de recursos, la mayoría de las arquitecturas analizadas poseen honeynet mixtas lo que es fundamental para evaluar la flexibilidad y portabilidad de la arquitectura, excepto la de Kohli et al. (2023); cada una de estas propuestas ha instalado diferentes herramientas de captura con la finalidad de monitorear, analizar y registrar el tráfico de datos que circula por una red; aunque algunas arquitecturas como HoneyFactory, SIHA-SDN y Software Agents centralizan la información recolectada, no la estructuran ni correlacionan.

La mayoría de las propuestas existentes de honeynets se centran exclusivamente en la captura de eventos o el despliegue de honeypots, sin una integración efectiva de datos que facilite un análisis profundo, estructurado y automatizado de la evidencia, excepto por la propuesta de arquitectura de honeynet con integración de datos.

La propuesta representa un avance significativo en el campo de la ciberseguridad, no solo por capturar ataques, sino por convertir la actividad maliciosa en evidencia digital útil y accionable. Representa una evolución clara y significativa respecto a las arquitecturas previas, esto transforma la función de la honeynet de un sistema pasivo de observación a una plataforma activa de ciberdefensa, situándose como una opción viable tanto para entornos académicos como para aplicaciones en infraestructuras críticas o entornos empresariales.

Del análisis se concluye que ninguna de las propuestas anteriores implementa un mecanismo de integración automatizada de datos. La propuesta actual no solo centraliza la información capturada, sino, además, que integra datos en tiempo real, lo que permite consolidar logs de las herramientas de captura, en conjunto con la honeynet, en una base de datos central, lo que reduce la carga operativa y el tiempo de análisis posterior a un incidente, siendo útil para la toma de decisiones.

2.5 ¿Por qué la herramienta de integración de datos?

La integración de datos desempeña un papel fundamental en la gestión y el análisis de datos (Wang et al., 2020), se refiere al método de reunir los datos deseados de múltiples fuentes en una sola, brindando al usuario una visión unificada (Sreemathy et al., 2021). La mejor definición de integración de datos la da IBM, que afirma: La integración de datos es la combinación de procesos técnicos y de negocio utilizados para combinar datos de fuentes dispares en información valiosa y significativa (Sreemathy et al., 2020). Al producir un único objeto con una estructura coherente mediante la integración de datos, se simplifica enormemente la manipulación posterior de dichos datos (Schildhauer, 2017).

La revolución digital produce datos masivos, heterogéneos y aislados. Para muchas organizaciones, la capacidad de dar respuestas coordinadas a los problemas empresariales actuales se ve obstaculizada por la falta de integración de datos o de elementos y códigos de datos definidos de forma común en los diferentes sistemas de información (Masmoudi et al., 2024).

En este sentido, ya no se trata de contar con un grupo de honeypots, ahora es esencial una herramienta de integración de datos en una honeynet para garantizar la mejora de la eficiencia en la recopilación y análisis de datos, potenciar la capacidad de respuesta ante estos incidentes; con una herramienta de integración de datos es posible:

- Recopilación de datos eficiente: la centralización de la herramienta de integración de datos facilita la recopilación de la data de diversas fuentes (servidores, dispositivos de red, etc.) además de ello, reduce el tiempo que se emplea para recolectar y organizar la data cuando se trata de un proceso manual.
- Respuesta rápida: una adecuada integración puede establecer alertas en tiempo real respecto a actividades sospechosas proporcionando información actualizada para que los analistas tomen decisiones en caso de presentarse algún tipo de intrusión en los sistemas.
- Análisis en profundidad: debido a que, entre sus bondades, se encuentran las herramientas de visualización que mejoran la comprensión de los datos e identifican las anomalías, además de ello, con la integración de datos, se puede

correlacionar eventos y patrones de ataques, mejorando la identificación de tácticas, técnicas y procedimientos de los atacantes.

- Procesos de mejora continua: proporciona funciones de retroalimentación suficientes para analizar los datos recopilados y mejorar las defensas y estrategias de la honeynet, además, de generar diversos tipos de informes para compartir con otras organizaciones información y el conocimiento adquirido.
- Apoyo a la investigación: específicamente en el área de ciberseguridad porque aporta conocimientos para el desarrollo de nuevas técnicas de defensa además de permitir la colaboración con diferentes entidades en la identificación de amenazas y vulnerabilidades emergentes.

En este marco de ideas, es necesario explicitar algunos conceptos que están intrínsecamente relacionados a la honeynet con la herramienta de integración de datos y que sirven de base para la operacionalización de las variables en función del cuerpo de hipótesis de la investigación expuestas en el Capítulo III, destacando:

2.5.1 Operatividad técnica

Los controles de seguridad son indispensables en el mundo actual por su papel esencial en la protección de aplicaciones y sistemas en muchas organizaciones. Ayudan a gestionar los controles organizacionales para garantizar la confidencialidad, la integridad y el acceso a infraestructura y datos vitales, garantizando que solo los usuarios autorizados y de confianza tengan acceso privilegiado (Sinan et al., 2025), para lo cual se consideran los siguientes elementos: (a) Supervisión constante de sistemas y redes para detectar problemas, vulnerabilidades o actividades sospechosas; (b) Establecimiento de sistemas, redes y aplicaciones, así como su mantenimiento continuo para asegurar su correcto funcionamiento; (c) Capacidad para identificar y solucionar incidentes técnicos que puedan afectar la seguridad o la operatividad de la organización y (d) El registro de procedimientos, configuraciones y cambios realizados en la infraestructura tecnológica instalada. Todo ello asegura que las herramientas de ciberseguridad funcionen adecuadamente, lo que reduce el riesgo de explotación al identificar y corregir vulnerabilidades en aplicaciones y redes dentro de la organización (Sinan et al., 2025).

En este mismo orden de ideas, se considera importante en el proceso de la dimensión: Operatividad técnica, lo considerado por Rabhi and Fissoune (2024), el acceso a datos de diferentes fuentes a través de una única interfaz de consulta, lo que se refiere a la interoperabilidad o capacidad de integrar y visualizar información proveniente de múltiples sistemas o bases de datos en un único punto de acceso, para ello se desarrollan los procesos de: (a) Integración de datos para combinar datos de diversas fuentes (como bases de datos, aplicaciones, servicios en la nube) y crear un conjunto consolidado de información; (b) Interfaz unificada que proporciona a los usuarios una interfaz única donde puedan tener acceso a datos de diferentes sistemas sin necesidad de cambiar de aplicación o de consola; (c) Mejora en la toma de decisiones al permitir análisis más completos y contextuales, ya que los especialistas pueden ver la información relevante en un solo lugar; (d) Eficiencia en las operaciones porque reduce el tiempo y el esfuerzo necesarios para buscar y recopilar datos de diferentes fuentes, mejorando la productividad.

2.5.2 Captura de evidencia digital

Los atributos para la captura de evidencia digital deben entenderse como las características específicas que se utilizan para recopilar y describir información relevante durante el proceso de monitoreo y análisis de eventos, en general. Para los autores Koumar et al. (2023), la monitorización del tráfico de red mediante flujos IP se utiliza para afrontar el reto actual de analizar las comunicaciones, además Patel (2024), menciona que la detección de anomalías, utilizando un conjunto de datos completo que incluye atributos críticos del tráfico de red, como el tamaño del paquete, el tipo de protocolo, las direcciones IP de origen y destino, la duración y el nivel de amenaza.

En el proceso de la dimensión: Captura de evidencia digital, se consideran los siguientes indicadores según manifiesta Masmoudi et al. (2024), la recopilación, comparación de atributos y la integración de los datos.

Recopilación de atributos: se refiere al proceso de reunir y almacenar estos atributos de captura de datos de diversas fuentes incluyendo: (a) Monitoreo de red con la captura de datos de tráfico en tiempo real; (b) Registro de eventos y almacenamiento de logs

de sistemas y aplicaciones; y (c) Sensores de seguridad con el uso de dispositivos como firewalls y sistemas de detección de intrusiones.

Comparación de atributos: implica analizar y contrastar los atributos recopilados con datos de referencia para identificar patrones, anomalías o amenazas, esto puede incluir: (a) Detección de intrusiones comparando los atributos de tráfico en tiempo real con patrones de tráfico conocidos; (b) Análisis de comportamiento y la evaluación de los usuarios o dispositivos en comparación con sus patrones habituales; (c) Identificación de vulnerabilidades, que se logra con la comparación de configuraciones y atributos de seguridad contra estándares y mejores prácticas.

Integración de datos: se entiende como el proceso de combinar información de diferentes fuentes para proporcionar una vista unificada. En ciberseguridad, esto permite: (a) Análisis holístico, esto es facilitar una visión completa de la seguridad de la red, combinando datos de múltiples sistemas; (b) Automatización de procesos y centralización de la información para automatizar la detección y respuesta a incidentes; (c) Reportes: Generar informes coherentes y comprensibles que reflejen el estado de seguridad de la organización.

2.5.3 Consistencia del proceso de análisis

La consistencia del proceso de análisis permite detectar patrones y tendencias en la información con mayor rapidez y precisión, sin necesidad de analizarla manualmente, lo que permite tomar decisiones con mayor confianza y precisión (Singla et al., 2023).

La consistencia del proceso de análisis se refiere a la uniformidad y repetibilidad de los métodos utilizados para analizar los datos recopilados en una honeynet, incluyendo: (a) Protocolos establecidos que consisten en el uso de procedimientos estandarizados para recopilar y analizar datos, garantizando que se sigan las mismas metodologías en cada análisis; (b) Revisión de los resultados con la implementación de revisiones periódicas y auditorías para asegurar que el análisis se realice de manera coherente, independientemente del analista o del momento; (c) Comparabilidad: entendida como la comparación de resultados a lo largo del tiempo, permitiendo la identificación de tendencias y patrones en el comportamiento de los atacantes.

En el proceso de la dimensión: Consistencia del proceso de análisis, se consideran los indicadores: Exactitud del análisis lo cual permite detectar las intrusiones con precisión, sin necesidad de analizarla manualmente, según lo expresa Singla et al. (2023), para ello el analista debe tener en cuenta, factores que influyen en la exactitud del análisis, destacando: (a) Calidad de los datos: Lo que implica asegurar que los datos recopilados sean correctos y representativos de la actividad de la red; (b) Métodos de análisis: Con la utilización de técnicas analíticas adecuadas y herramientas confiables que minimizan el riesgo de errores; (c) Validación de resultados: Con la implementación de métodos para verificar y validar los resultados obtenidos, asegurando que reflejen con precisión las actividades observadas.

Por otra parte, también es necesario considerar que debe existir confianza según (Seo and Kim (2019), en el contexto de una honeynet, ésta se refiere a que los resultados del análisis sean fiables además de la capacidad de los equipos de ciberseguridad para tomar decisiones basadas en esos resultados. Para lo anteriormente expuesto hay que considerar los siguientes elementos: (a) Transparencia en el proceso con la documentación clara de los métodos y procedimientos utilizados en el análisis, lo que permite a otros revisarlos y comprender cómo se llegaron a las conclusiones; (b) Historial de resultados que respalden la eficacia de las técnicas de análisis empleadas a lo largo del tiempo y; (c) Formación del personal encargado del análisis ya que los conocimientos y la experiencia aumentan la confianza en la capacidad que tienen los expertos para interpretar correctamente los datos y hacer recomendaciones informadas y pertinentes.

2.5.4 Tiempos de detección de ataques

En la sección 2.2.2 de este trabajo, titulada Los ciberataques y en la sección siguiente denominada Intrusión moderna, se expusieron los ciberataques, su tipología e impacto, pero, en el contexto de una honeynet, no se trata solamente de poder identificar estos eventos o cualquiera que sea sospechoso, también es importante disminuir los tiempos de detección de estos y, como lo expone Vega et al. (2022), es esencial porque:

- a) Procura la minimización de daños ya que, cuanto más rápido se detecte un ataque, menos daño puede causar este a los sistemas, datos y operaciones, con ello, se

puede prevenir la pérdida de información sensible, el daño a la reputación de la organización y las interrupciones en los servicios;

- b) Detección temprana brindando una respuesta más ágil y efectiva por parte de los especialistas en ciberseguridad, con lo cual, pueden contener el ataque, mitigar sus efectos y restaurar los sistemas afectados rápidamente;
- c) Protección de datos e información confidencial además de cumplir con regulaciones y normativas de protección de datos, evitando sanciones legales y pérdidas financieras;
- d) Reducción de costos tanto directos como indirectos, ya que al detectar y responder a los ataques de manera rápida reduce estos costos, evitando gastos asociados con la recuperación de equipos y datos sensibles;
- e) Mejora continua es un aprendizaje constante, lo que les permite a las organizaciones aprender de cada incidente, mejorando sus defensas y estrategias de seguridad para el futuro, contribuyendo a desarrollar un entorno de seguridad más robusto y proactivo;
- f) Confianza del cliente: debido a que la capacidad de detectar y responder a ciberataques de manera eficaz puede fortalecer la confianza de los clientes y socios comerciales en la organización, lo que es vital para mantener relaciones comerciales sólidas mejorando la reputación de la organización y de sus servicios de ciberdefensa.

2.5.5 Integridad de los datos

Verificar eficientemente la integridad de todos los datos plantea un desafío complejo y urgente (Yan et al., 2025). En el proceso de la dimensión de Integridad de los datos, se consideran los siguientes indicadores: almacenamiento de los datos, consistencia según lo que menciona Okun and Barak (2004) y precisión técnica según Murphy (2022).

En el caso de una honeynet, para asegurar que la información utilizada para análisis y respuesta a incidentes sea confiable, es necesario garantizar que los datos recopilados sobre ataques o intrusiones no sean alterados. Para lo cual se debe considerar:

- a) Almacenamiento de los datos o cómo se guardan los datos recogidos en la honeynet y, en este sentido, es necesario asegurarse que el almacenamiento sea

seguro y accesible solo para personal autorizado incluyendo el uso de cifrado y controles de acceso para proteger la información sensible.

- b) Consistencia de los datos, referido a la uniformidad y coherencia de los datos a lo largo del tiempo y, en lo particular, en el contexto de una honeynet, los datos deben ser consistentes para que los análisis sean válidos, para ello, los registros de eventos y las métricas deben ser precisos y reflejar el estado real de los sistemas monitoreados.
- c) Precisión técnica, en una honeynet es crucial la exactitud y calidad de los datos recopilados, además de la precisión para que los investigadores puedan identificar patrones de ataque y comportamientos maliciosos y, con ello, asegurar que las conclusiones y decisiones basadas en esos datos sean correctas.

2.5.6 Gestión de la información

La gestión de la información es una dimensión amplia, sin embargo, para esta investigación se ha enfocado; como la capacidad de un sistema que permite estructurar y presentar los datos mediante una usabilidad adecuada, que facilite la interacción del usuario, y a través de la generación de reportes claros y completos, que respalden la toma de decisiones y la solución de problemas.

En el proceso de la dimensión: gestión de la información, se consideran los siguientes indicadores: usabilidad y la generación de reportes, según Lennartsson et al. (2021) y Nayeem et al. (2021).

Considerando la interacción hombre-sistema, es necesario proveer una adecuada usabilidad, y con ello, se hace referencia a la facilidad de uso de las herramientas y plataformas utilizadas para interactuar con la honeynet. La apropiada gestión de la información permite a los analistas y administradores navegar fácilmente por los datos y la generación adecuada de reportes, facilita la identificación de amenazas y la toma de decisiones. Una interfaz intuitiva reduce la curva de aprendizaje y mejora la eficiencia en el análisis de datos, lo que es crítico en situaciones de respuesta a incidentes donde el tiempo es esencial.

CAPÍTULO III: METODOLOGÍA

3.1 Hipótesis

3.1.1 Hipótesis general de la investigación (H_0)

El uso de una arquitectura de honeynet con integración de datos, mejora el análisis de evidencia digital producto de la heterogeneidad en los datos proporcionados por las herramientas utilizadas.

3.1.2 Hipótesis nula de la investigación (H_1)

El uso de una arquitectura de honeynet con integración de datos, no mejora el análisis de evidencia digital producto de la heterogeneidad en los datos proporcionados por las herramientas utilizadas.

3.1.3 Hipótesis específicas

- a) La integración de los datos en la arquitectura honeynet resuelve problemas de interoperabilidad.
- b) La consolidación de la ontología de la honeynet con integración de los datos mejora la captura de evidencia digital.
- c) El uso de una arquitectura de honeynet con integración de datos mejora la consistencia del proceso de análisis de evidencia digital.
- d) El uso de una arquitectura de honeynet con integración de datos mejora los tiempos de detección de ataques en el análisis de evidencia digital.
- e) El uso de una arquitectura de honeynet con integración de datos mejora la integridad de los datos en el análisis de evidencia digital.
- f) El uso de una arquitectura de honeynet con integración de datos optimiza la gestión de la información en el análisis de evidencia digital.

3.2 Identificación de variables

A continuación, se presentan las variables del estudio en atención a las hipótesis de la investigación:

3.2.1 Variable independiente

Arquitectura de honeynet con integración de datos.

3.2.2 Variable dependiente

Análisis de evidencia digital.

Para poder medir estas variables se realizó la operacionalización de las mismas y, una vez culminada, se estableció la matriz de consistencia de la investigación para garantizar la coherencia metodológica. Ahora bien, la operacionalización de las variables implicó definir sus características, dimensiones e indicadores, todo ello se encuentra desarrollado en las secciones previas de este estudio, específicamente, en los apartados: 2.2.2 Los ciberataques y 2.5 ¿Por qué la herramienta de integración de datos?

Cabe destacar que, a partir de la operacionalización de las variables, los indicadores de cada una de ellas constituyen las preguntas de los cuestionarios validados y utilizados por los expertos para evaluar la arquitectura de la honeynet sin la herramienta de integración de datos y, posteriormente, con dicha herramienta. Los instrumentos están definidos en la sección 3.5.6 Técnica de recolección de datos y se presentan en los anexos del estudio.

3.3 Operacionalización de variables

Tabla 4. Operacionalización de variables

Variable	Tipo de variable	Definición conceptual	Definición operacional	Dimensión	Indicador	Instrumento
Arquitectura de honeynet con integración de datos.	Independiente	Entorno estructurado de seguridad informática diseñado para simular sistemas reales con el propósito de ser atacados, monitoreados y analizados (Javadpour et al., 2024), incorporando mecanismos que permiten la recolección, integración y análisis de datos provenientes de múltiples fuentes (Masmoudi et al., 2024).	Implementación técnica que permite centralizar, analizar y reportar datos de ciberataques a partir de la integración de diferentes fuentes que permita la comparación con la arquitectura sin integración de datos dentro de la honeynet.	Operatividad Técnica	Interoperabilidad	7 - 8

Análisis de evidencia digital.	Dependiente	Procedimiento técnico y ordenado mediante el cual se identifican, recopilan, analizan, interpretan y preservan los datos (Rogers, 2013); con la finalidad de identificar, reconstruir y comprender acciones anómalas en sistemas informáticos, que sugieran una actividad potencialmente maliciosa (Sabillon, 2020).	Proceso que permite analizar y gestionar la información digital recolectada desde la honeynet para identificar incidentes en el menor tiempo posible y garantizar la integridad, exactitud y confianza de los datos.	Captura de evidencia digital	Recopilación de atributos	1 - 2
					Comparación de atributos	3
					Integración de datos	4 - 5 - 6
				Consistencia del proceso de análisis	Exactitud del análisis	9 - 10
					Confianza	11 - 12
				Tiempos de detección de ataques	Detección de ataques de suplantación	13
					Detección de inyecciones de SQL en el servidor	
					Detección de fuerza bruta	
					Detección de denegación del servicio	

					Detección de envenenamiento DNS (Domain Name Server)	
					Detección de exploración de puertos	
					Detección de ataques ICMP (Internet Control Message Protocol)	
					Detección de cross site scripting	
				Integridad de los datos	Almacenamiento de los datos	14 - 15
					Consistencia	16
					Precisión técnica	17 - 18
				Gestión de la información	Usabilidad	19 - 20 - 21
					Generación de reportes	22 - 23

Fuente. El autor.

3.4 Matriz de consistencia

Tabla 5. Matriz de consistencia

Problema	Objetivo	Hipótesis	Variables
General	General	Hipótesis General	Variable Independiente
¿El uso de una arquitectura de honeynet, con integración de datos, mejora el análisis de evidencia digital producto de la heterogeneidad en los datos proporcionados por las herramientas utilizadas?	Desarrollar una arquitectura de honeynet con integración de datos que mejora el análisis de evidencia digital propia de la heterogeneidad en los datos proporcionados por las herramientas utilizadas.	El uso de una arquitectura de honeynet con integración de datos, mejora el análisis de evidencia digital producto de la heterogeneidad en los datos proporcionados por las herramientas utilizadas.	Arquitectura de honeynet con integración de datos.
		Hipótesis Nula	Variable Dependiente
		El uso de una Arquitectura de honeynet con integración de datos, no mejora el análisis de evidencia digital producto de la heterogeneidad en los datos proporcionados por las herramientas utilizadas.	Análisis de evidencia digital.

3.4 Matriz de Consistencia (Cont.)

Problemas Específicos	Objetivos Específicos	Hipótesis Específicas	
El uso de una arquitectura de honeynet con integración de datos, ¿permite resolver problemas de interoperabilidad?	Diseñar la arquitectura física y lógica de honeynet para resolver problemas de interoperabilidad.	La integración de los datos en la arquitectura honeynet resuelve problemas de interoperabilidad.	
¿Se mejora la captura de evidencia digital con la consolidación de la ontología de la honeynet con integración de los datos?	Elaborar la ontología honeynet con integración de datos para mejorar la captura de evidencia digital.	La consolidación de la ontología de la honeynet con integración de los datos mejora la captura de evidencia digital.	
El uso de una arquitectura de honeynet con integración de datos, ¿mejora la consistencia del proceso de análisis de evidencia digital?	Analizar la consistencia del proceso de análisis de evidencia digital con el uso de la arquitectura de honeynet con integración de datos.	El uso de una arquitectura de honeynet con integración de datos mejora la consistencia del proceso de análisis de evidencia digital.	

3.4 Matriz de Consistencia (Cont.)

Problemas Específicos	Objetivos Específicos	Hipótesis Específicas	
¿Mejoran los tiempos de detección de ataques durante el análisis de evidencia digital con el uso de una arquitectura de honeynet con integración de datos?	Identificar los tiempos de detección de ataques en el análisis de evidencia digital con el uso de la arquitectura honeynet con integración de datos.	El uso de una arquitectura de honeynet con integración de datos mejora los tiempos de detección de ataques en el análisis de evidencia digital.	
¿El uso de una arquitectura de honeynet con integración de datos mejora la integridad de los datos en el análisis de evidencia digital?	Mejorar la integridad de los datos en el análisis de evidencia digital con el uso de la arquitectura honeynet con integración de datos.	El uso de una arquitectura de honeynet con integración de datos mejora la integridad de los datos en el análisis de evidencia digital.	
El uso de una arquitectura de honeynet con integración de datos, ¿optimiza la gestión de la información en el análisis de evidencia digital?	Comprobar que el uso de la arquitectura de la honeynet con integración de datos, optimiza la gestión de la información en el análisis de evidencia digital.	El uso de una arquitectura de honeynet con integración de datos optimiza la gestión de la información en el análisis de evidencia digital.	

Fuente. El autor.

3.5 Metodología

La realización de este estudio se ha fundamentado en las siguientes etapas, claramente diferenciadas:

- a) Concreción del problema de estudio que da origen a esta tesis partiendo de que el uso de una arquitectura de honeynet con integración de datos, mejorará el análisis de evidencia digital producto de la heterogeneidad en los datos proporcionados por las herramientas utilizadas.
- b) Construcción del marco teórico que permite desarrollar el marco conceptual y legal dentro de los cuales se darán las discusiones para responder a las hipótesis del estudio.
- c) Concreción de la ontología de la honeynet orientada a la unificación de las estructuras de almacenamiento, su utilización por varias aplicaciones y la interoperabilidad de los sistemas cuando los datos son heterogéneos.
- d) Desarrollo de la Arquitectura de honeynet con integración de datos orientada a la mejora del análisis de evidencia digital en cuanto a confiabilidad y tiempos de respuesta conforme a una ontología propia.
- e) Obtención de datos respecto a la honeynet con integración de datos y sus posibilidades en cuanto a mejorar el análisis de evidencia digital garantizando la confiabilidad del proceso y la disminución de los tiempos de respuesta.
- f) Pruebas del modelo de arquitectura de la honeynet con integración de datos por medio de simulaciones y con validación de expertos en el área.

3.5.1 Tipo y diseño de la investigación

La investigación tiene un enfoque cuantitativo, de tipo descriptiva y explicativa, puesto que utiliza técnicas de recolección y análisis de datos medibles para responder a las preguntas planteadas en los problemas general y específicos, la medición de las variables se realiza a través del uso de estadística descriptiva e inferencial para probar las hipótesis formuladas que verifiquen que el uso de la arquitectura de la honeynet mejora el análisis de evidencia digital, producto de la heterogeneidad en los datos proporcionados por las herramientas utilizadas.

En este orden de ideas esta investigación es descriptiva, porque especifica las propiedades y características de la honeynet con integración de datos a través de una búsqueda bibliográfica exhaustiva que permite obtener información sobre conceptos, variables y dimensiones que componen el problema a investigar.

También se trata de una investigación explicativa, porque se establecen relaciones de causalidad entre las variables, lo que busca resolver de manera práctica y concreta problemas específicos de seguridad de redes informáticas que afectan a individuos y organizaciones, por medio de una arquitectura de honeynet con herramienta de integración de datos, lo que la convierte en un aporte al conocimiento en el área tecnológica.

La investigación se enmarca en un diseño cuasiexperimental, puesto que los expertos no se asignan al azar, sino que están designados antes del experimento: son grupos intactos a quienes se aplicó un pretest y posttest, para verificar la validez de las hipótesis de investigación.

3.5.2 Unidad de análisis

Profesional en el área de ciberseguridad que cuente con los criterios muestrales.

3.5.3 Población de estudio

Para esta investigación, la población está constituida por expertos en ciberseguridad, con conocimientos en las áreas de criminalística, cyberforenses, ciberseguridad, programación y seguridad de redes, quienes laboran en diferentes organizaciones de seguridad, considerando que las mismas se encuentren registradas ante los organismos competentes de cada estado.

3.5.4 Tamaño de la muestra

Para esta investigación, la muestra se constituyó a partir de un muestreo no probabilístico por conveniencia, donde las muestras se forman a discreción del investigador.

En el caso de la presente investigación, los sujetos de la muestra deben satisfacer los criterios muestrales establecidos por el investigador, ya que, se trata de determinar la idoneidad de la ontología de la honeynet y del funcionamiento de la herramienta de integración de datos, núcleo central del estudio; se hace énfasis, entonces, en que esta investigación no está dirigida a valorar las características de los expertos sino a determinar el funcionamiento de la honeynet con la herramienta de integración de datos en atención a la mejora del análisis de evidencia digital, interoperabilidad, confiabilidad y calidad del hallazgo, disminución de los tiempos de respuesta y optimización de la operatividad de la herramienta; así pues, la muestra que es de tipo intencional, los sujetos son seleccionados por el investigador en la medida en que estos puedan hacer una valoración completa de la honeynet con la herramienta de integración de datos.

Criterios muestrales del estudio. Cada experto debe satisfacer los siguientes:

- a) Ser profesional en el área de TI, expertos en ciberseguridad, con conocimientos en las áreas de criminalística, ciberforenses, ciberseguridad, programación y seguridad de redes.
- b) Contar con un postgrado en el área de ciberseguridad o afines, con conocimientos avanzados y actualizados respecto al tema en estudio.
- c) Con más de 10 años de experiencia trabajando en el área de ciberseguridad.
- d) Actualmente activo en una organización de ciberseguridad en América Latina, Europa o EEUU.
- e) Estar dispuesto a brindar su tiempo para apoyar este estudio como experto evaluador a solicitud del investigador.

Entonces, a partir de estos criterios, se seleccionaron 4 expertos que brindaron sus conocimientos y tiempo para hacer la valoración completa de la honeynet con la herramienta de integración de datos; caracterizados por:

Experto 1: 26 años en el área de ciberseguridad; posgrado en Seguridad de la Información; actualmente, trabaja en Colombia en una empresa de ciberseguridad y es docente universitario.

Experto 2: 15 años de experiencia en arquitectura avanzada de redes y seguridad; posgrado en telecomunicaciones; trabaja en Colombia como consultor en informática forense y ciberseguridad; es docente universitario.

Experto 3: 12 años de experiencia en sistemas redes de datos, sistemas operativos y seguridad de la información, sistemas distribuidos y programación. Posgrado en Software libre y telecomunicaciones.

Experto 4: 30 años de experiencia en informática forense; especialista en Seguridad informática, Delitos informáticos e Informática forense. Trabaja como consultor para varias empresas en España, EEUU y Colombia.

3.5.5 Técnica de recolección de datos

Para la recolección de los datos se utiliza una encuesta dirigida a los expertos en el área de ciberseguridad que cuente con los criterios muestrales.

Las preguntas de los instrumentos permitieron el registro de los datos, en este sentido, para esta investigación, se diseñaron y aplicaron una serie de cuestionarios de preguntas cerradas y autoadministrados, es decir, respondidos directamente por los expertos del estudio:

- (a) Cuestionario dicotómico (Si-No) de 28 preguntas validado con el coeficiente de concordancia de Kendall para elaborar la ontología de la arquitectura de la honeynet aplicado en la Fase II de la recolección de datos.
- (b) Producto de la operacionalización de las variables (sección 3.3 de este estudio), para la evaluación de la capacidad de la arquitectura honeynet en la integración de datos para el análisis de evidencia digital por los expertos se construyeron dos cuestionarios con 3 opciones de respuestas cerradas cada uno (Si, No, No hay evidencia), los cuales, en su primera versión, tuvieron 27 ítems cada uno para evaluar el funcionamiento de la honeynet con y sin la herramienta de integración de datos.

Ambos cuestionarios fueron sometidos a validez por juicio de expertos, para ello, se solicitó a 3 especialistas en tecnología que revisaran los instrumentos e hicieran las observaciones pertinentes para mejorarlos antes de ser usados por la muestra de los expertos. La validez por juicio de expertos permitió crear una segunda versión de los cuestionarios con 23 preguntas manteniendo la misma estructura. Las actas de validación elaboradas por los expertos se encuentran en los anexos de este estudio. Ver (Anexo A).

Estos instrumentos se aplicaron en la Fase III de la recolección de datos y forman parte de los anexos del estudio. Ver (Anexo B).

El proceso de recolección de datos se realizó en diferentes fases atendiendo al momento de la investigación, en este sentido, se planteó:

Fase I: Arqueo bibliográfico para construir el marco teórico del estudio y reforzar el planteamiento del problema. En esta fase se recopiló la información de tipo teórica de libros, artículos de investigación y tesis doctorales respecto a la honeynet, sus características y evolución, diseño de arquitectura, entre otros elementos relevantes.

Fase II: Seguimiento de la técnica de la encuesta con un cuestionario dicotómico para validar la ontología de la arquitectura de la honeynet.

Fase III: Técnica de la encuesta con los dos cuestionarios, obtenidos a partir de la operacionalización de las variables, validados por juicio de expertos. Ver (Anexo C).

3.5.6 Técnica de análisis de datos

Cada técnica de recolección de datos fue sometida a un proceso de análisis en particular dependiendo de la función que tienen en el proceso de investigación. Así pues, para el arqueología de información teórica, los textos se revisaron y contrastaron para obtener el marco teórico que sustenta el estudio.

El cuestionario de la Fase II, en el que participaron especialistas en el área de seguridad con conocimiento en desarrollo de ontologías, permitió realizar un análisis de datos cuantitativo que valide la ontología a través de un método estadístico denominado coeficiente de concordancia de Kendall que mide la afinidad de los individuos con respecto a las k variables. Con los resultados se validó la ontología de la honeynet con la herramienta de integración de datos.

Por otra parte, los cuestionarios de la Fase III, fueron analizados con técnicas de estadística descriptiva para la validación de la honeynet con la herramienta de integración de datos y parte de sus insumos fueron usados para la confirmación de las hipótesis de la investigación. En cuanto a la verificación del cuerpo de hipótesis, este procedimiento se efectuó utilizando las herramientas estadísticas como: prueba McNemar, prueba de normalidad de Shapiro-Wilk, prueba F para varianzas iguales y prueba t de Student, conforme al planteamiento de cada hipótesis específica, usando el software estadístico R-Studio.

CAPÍTULO IV: RESULTADOS Y DISCUSIÓN

4.1 Ontología honeynet para integración de datos

Una ontología la definen Neches et al. (1991), como los términos básicos y relaciones que conforman el vocabulario de un área específica, así como las reglas para combinar dichos términos y las relaciones para definir extensiones de vocabularios.

Integran conocimiento que puede ser utilizado por humanos o máquinas para la toma de decisiones y la resolución de problemas. Las ontologías formales se utilizan principalmente en la interoperabilidad de los sistemas de información, la gestión de recursos humanos, la representación del conocimiento del dominio, la incorporación de la semántica en el modelado basado en Lenguaje de Modelado Unificado (UML) y la gestión del código de programación y la documentación (Husáková & Bureš, 2020).

Existen diversas definiciones de ontologías en informática Gruber (1993) y Borst (1997) proporcionan dos de las definiciones más conocidas de ontología y los atributos requeridos en la ingeniería ontológica. Señala Gruber (1993), que la ontología es una especificación explícita de la conceptualización, Borst (1997), amplía esta definición como: La ontología es la especificación formal explícita de la conceptualización compartida.

Conceptualización según Studer et al. (1998), se refiere a un modelo abstracto de algún fenómeno en el mundo a través de la identificación de los conceptos relevantes de dicho fenómeno.

El informático estadounidense McCarthy (1980) introdujo por primera vez las ontologías en la informática como un enfoque para expresar el conocimiento de

sentido común, conocimiento rutinario que se reciben durante las actividades cotidianas.

Una ontología se puede aplicar en cualquier área de conocimiento, en el caso particular de una ontología honeynet para la integración de datos, estableciendo una estructura categórica de la realidad.

Una honeynet es una herramienta de seguridad diseñada para ser sondeada, atacada y comprometida, compuesta por un conjunto de sistemas dispuestos a recibir ataques con una serie de mecanismos encargados de la monitorización, utilizando herramientas adicionales que se configuran para la captura y el análisis de intrusiones, mapeando los datos con la finalidad de poseer de manera centralizada en una sola base de datos toda la información, para la toma de decisiones en un solo formato, de este modo se tiene un mejor control sobre los datos recogidos por las diferentes herramientas. Se presenta el dominio de la ontología honeynet para integración de datos, (ver Figura 7).

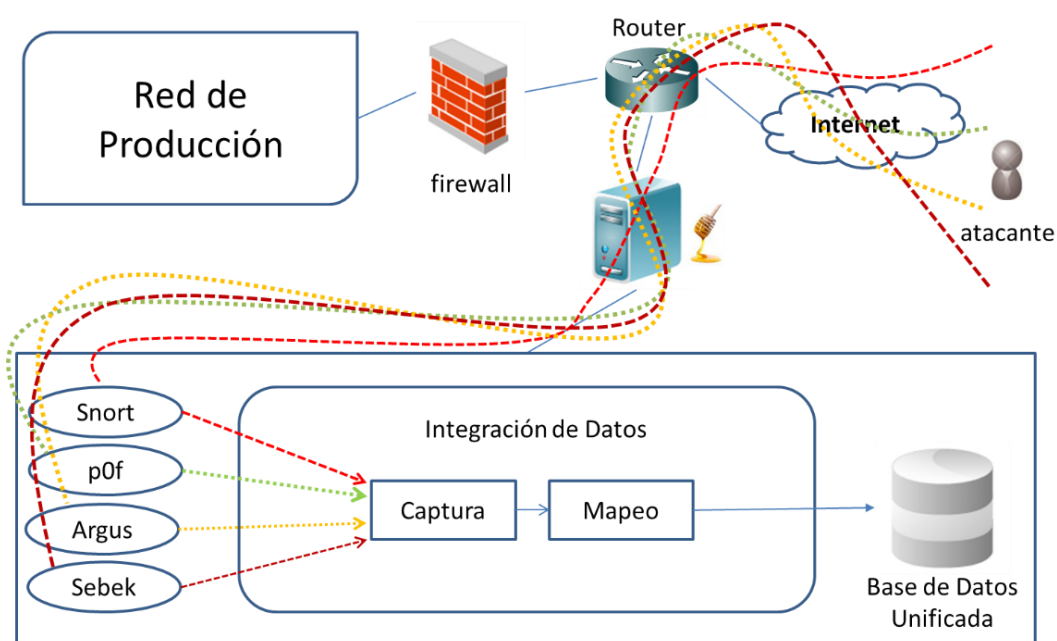


Figura 7. Dominio de la Ontología

Fuente. El autor.

4.1.1. Metodología para la ontología

Los trabajos más importantes que describen cómo desarrollar ontologías desde una perspectiva metodológica son, entre otros: los de Uschold and King (1995). No existe una propuesta metodológica completamente madura para la construcción de ontologías, ya que existen algunas actividades y técnicas importantes que faltan en estas metodologías. Sin embargo, methontology es una metodología muy madura según lo expresan Fernández-López and Gómez-Pérez (2002).

Una metodología bien definida y organizada que acorta el tiempo necesario para construir una ontología, aumentando la probabilidad de que el proyecto tenga éxito, es methontology, que utiliza en las fases de especificación, conceptualización, formalización, implementación y mantenimiento del proceso de construcción de ontología (Mamat et al., 2024). (ver Figura 8).

La ontología para la integración de datos en honeynet se desarrolló utilizando la metodología methontology, reconocida como una de las mejores metodologías, se ha adoptado en diversos desarrollos, gracias a sus características de independencia del dominio (Park et al., 2008), la misma fue desarrollada en el Laboratorio de Inteligencia Artificial de la Universidad Politécnica de Madrid. Tiene sus raíces en las actividades identificadas por la IEEE para el proceso de desarrollo de software según Corcho et al. (2005).

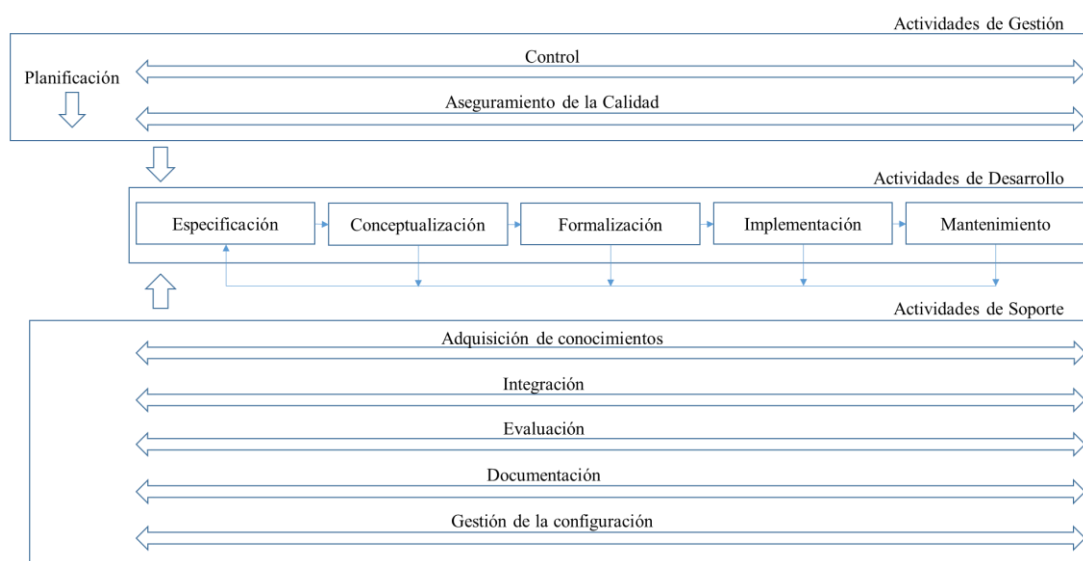


Figura 8. Actividades de desarrollo de ontologías propuestas por methontology
Fuente. (Corcho et al., 2005)

La actividad de especificación permite determinar por qué se construye la ontología, cuál será su uso, y quiénes serán sus usuarios finales. La actividad de conceptualización se encarga de organizar y convertir una percepción informal del dominio en una especificación semi-formal, para lo cual utiliza un conjunto de representaciones intermedias, basadas en notaciones tabulares y gráficas, que pueden ser fácilmente comprendidas por los expertos del dominio y los desarrolladores de ontologías. El resultado de esta actividad es el modelo conceptual de la ontología. La actividad de formalización se encarga de la transformación de dicho modelo conceptual en un modelo formal o semicomputable. La actividad de implementación construye modelos computables en un lenguaje de ontologías. La mayor parte de las herramientas de ontologías permiten llevar a cabo esta actividad de manera automática. La actividad de mantenimiento se encarga de la actualización y/o corrección de la ontología, en caso necesario (Corcho et al., 2005).

Methontology también identifica actividades de gestión (planificación, control y aseguramiento de la calidad), y actividades de soporte (adquisición de conocimientos, integración, evaluación, documentación y gestión de la configuración).

A continuación, se presentan las tareas de conceptualización de la metodología methontology, (ver Figura 9).

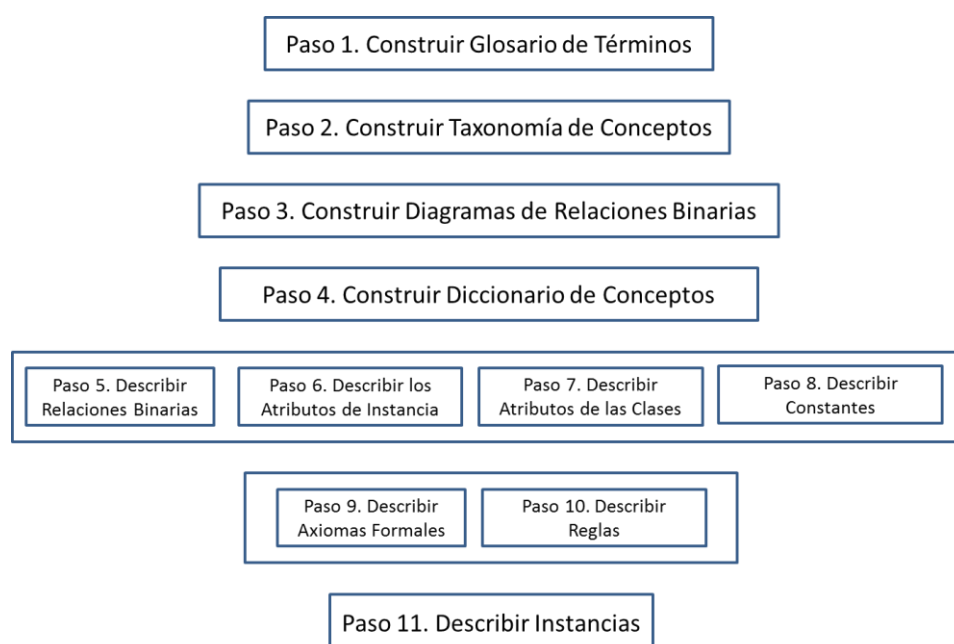


Figura 9. Tareas de la metodología methontology
Fuente. (Corcho et al., 2005)

Los elementos fundamentales de methontology son: los conceptos son objetos o entidades, considerados desde un punto de vista amplio. Las relaciones representan un tipo de asociación entre conceptos del dominio. Las instancias se utilizan para representar individuos en la ontología. Las constantes son valores numéricos que no cambian en un largo período de tiempo. Los axiomas formales son expresiones lógicas siempre verdaderas que suelen utilizarse para definir restricciones en la ontología. Las reglas se utilizan normalmente para inferir conocimientos en la ontología, tales como valores de atributos, instancias de relaciones, etc.

4.1.2 Desarrollo de la ontología

Se describe el desarrollo de la ontología honeynet para integración de datos utilizando methontology como metodología de trabajo.

Paso 1. Construir Glosario de Términos. Se presentan todos los términos necesarios para la construcción de la ontología honeynet para integración de datos que será la base de información técnica para la elaboración de la ontología honeynet para integración de datos.

Paso 2. Construir Taxonomía de Conceptos. Una vez que el glosario de términos contiene los suficientes términos para empezar a diagramar la ontología, se construye las taxonomías de conceptos que definen su jerarquía en el proceso de la construcción de la ontología honeynet para integración de datos, (ver Figura 10).

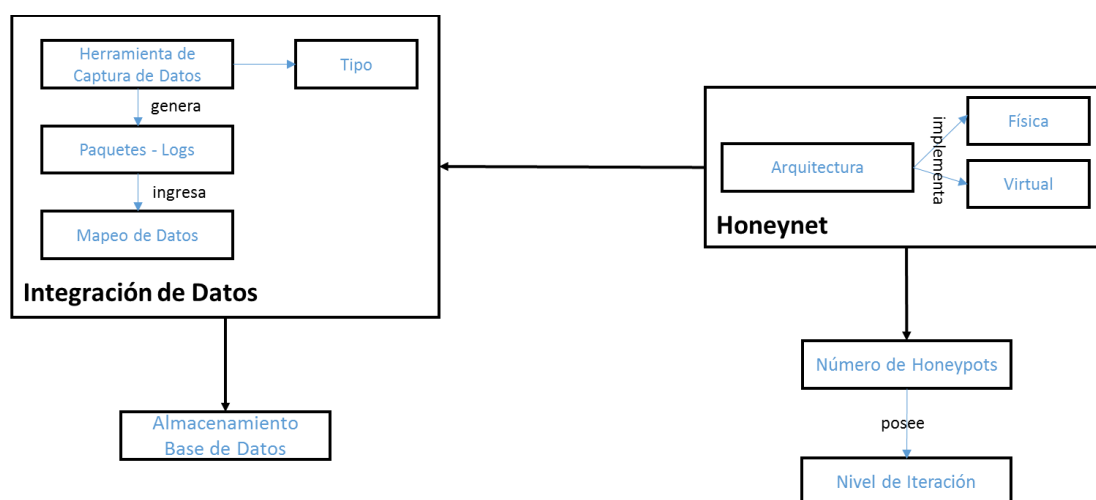


Figura 10. Sección de la taxonomía de conceptos de la ontología honeynet para integración de datos

Fuente. El autor.

Paso 3. Construir Diagramas de Relaciones Binarias. La actividad de conceptualización propone construir diagramas de relaciones binarias ad hoc. El objetivo de este diagrama es establecer las relaciones ad hoc existentes entre conceptos de esta o de distintas taxonomías de conceptos de la ontología de integración de datos en una honeynet, (ver Figura 11).

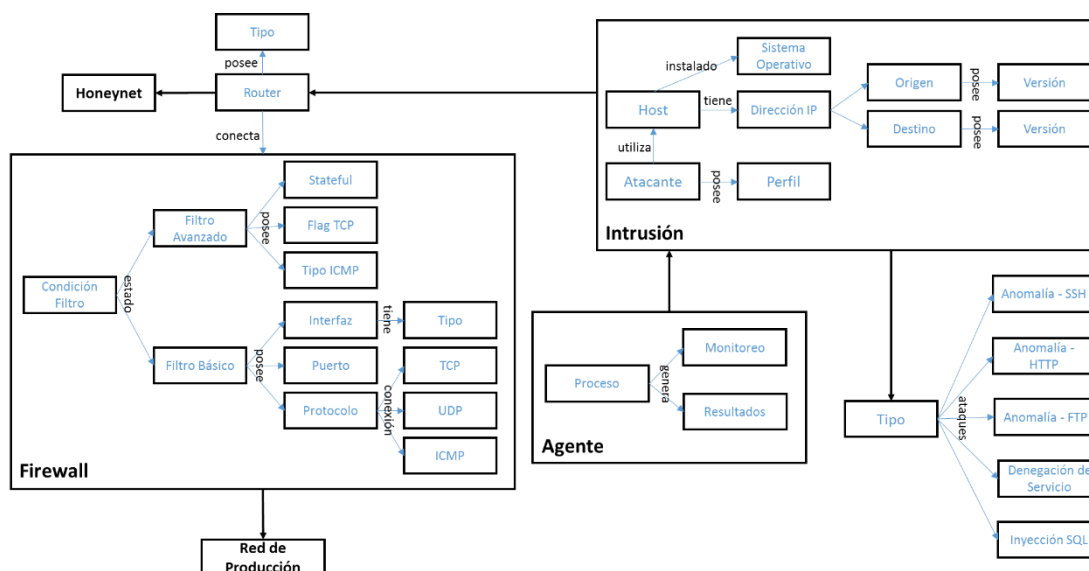


Figura 11. Sección de la construcción de diagramas de relaciones binarias de la ontología honeynet para integración de datos

Fuente. El autor.

De manera específica se realiza la construcción de diagramas de relaciones binarias, con lo que respecta a la integración de datos, (ver Figura 12).

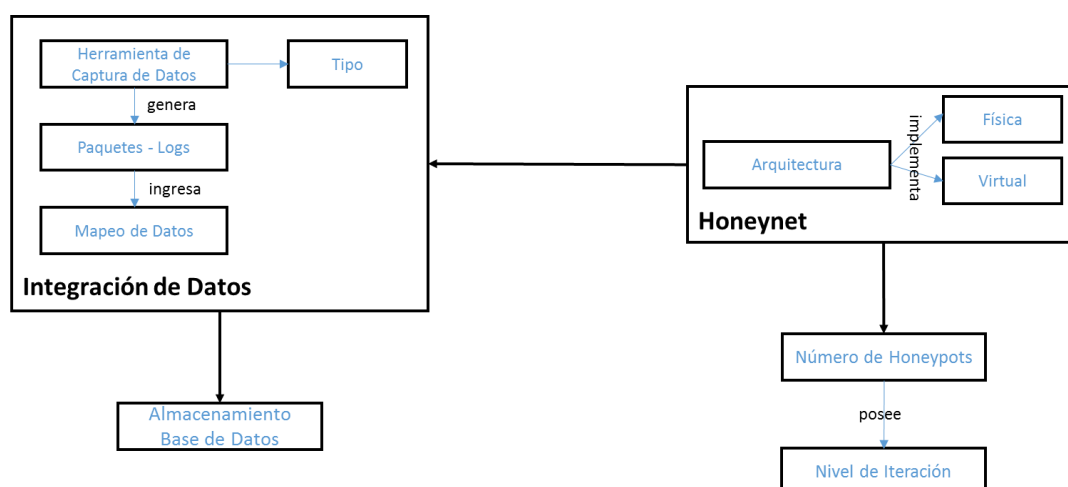


Figura 12. Sección de la construcción de diagramas de relaciones binarias específica para integración de datos

Fuente. El autor.

Paso 4. Construir Diccionario de Conceptos. Una vez que las taxonomías de conceptos y los diagramas de relaciones binarias ad hoc han sido generados para la ontología de integración de datos, se especifica cuáles son las propiedades que describen cada concepto de la taxonomía, así como las relaciones identificadas y las instancias de cada uno de los conceptos. La Tabla 6 muestra la construcción del diccionario de conceptos de la ontología para integración de datos en honeynet.

Tabla 6. Sección del diccionario de conceptos de la ontología para integración de datos en honeynet

Nombre del Concepto	Instancias	Atributos de Clase	Atributos de Instancia	Relaciones
Honeynet	Diseño	Tipo de monitoreo	Número de honeypots	Contiene
Arquitectura	Implementación	--	Generación	Implementa
Arquitectura - Física	Implementación	--	Nivel de operación	Establece
Arquitectura - Virtual	Implementación	--	Nivel de operación	Establece
Honeypots	Configuración	Tipo de monitoreo	Nivel de iteración	Establece
Integración de datos	--	--	Tipos de datos	--
Herramienta de captura de datos	Configuración	Tipo de instalación	Nombre Tipo de herramienta	Genera
Paquetes - Logs	Integrador	--	Tipo de configuración	Ingresa
Mapeo de datos	Desarrollador	--	Tiempo	--
Almacenamiento base de datos	Administrador	--	Capacidad	--
Host - RP	--	--	Nombre Protocolo Puerto Dirección IP Sistema Operativo	Posee
Servidor - RP	--	--	Funcionalidad Tipo Protocolo Puerto Dirección IP Sistema Operativo	Posee

Intrusión	Atacante	Tipo de control	Tipo Herramientas Perfil	establece
Host - I	--	--	Nombre Protocolo Puerto Dirección IP Sistema Operativo	Posee
Router	--	--	Tipo	Posee
Firewall	--	Tipo de bloqueo	Tipo Condición Interfaz	Establece
Agente	--	--	Proceso Monitoreo Resultados	Genera

Fuente. El autor.

Paso 5. Describir Relaciones Binarias. Se describe en detalle todas las relaciones binarias ad hoc identificadas en el diagrama de relaciones binarias e incluidas en el diccionario de conceptos. La Tabla 7 describe las relaciones binarias de la ontología para integración de datos en honeynet.

Tabla 7. Parte de la sección de la tabla de relaciones binarias de la ontología para integración de datos en honeynet

Nombre de la Relación	Concepto Origen	Cardinalidad Máxima	Concepto Destino	Relación Inversa
Propuesta	Honeynet	N	Integración de datos	es contemplada
Genera	Herramienta de captura de datos	N	Paquetes - Logs	son generados
Ingresa	Paquetes - Logs	N	Mapeo de datos	son ingresados
Entorno de red	Arquitectura	N	Física	plataforma
Entorno de red	Arquitectura	N	Virtual	plataforma
Conjunto	Honeynet	N	Número de honeypots	caracteriza
Dirige	Integración de datos	N	Almacenamiento base de datos	proviene

Fuente. El autor.

Paso 6. Describir los Atributos de Instancia. Se describe en detalle todos los atributos de instancia incluidos en el diccionario de conceptos. Cada fila de la tabla de atributos de instancia contiene la descripción detallada de un atributo de instancia. La Tabla 8 describe los atributos de instancia de la ontología para integración de datos en honeynet.

Tabla 8. Parte de la sección de la tabla de atributos de instancia de la ontología para integración de datos en honeynet

Nombre del Atributo de Instancia	Concepto	Tipo de Valor	Rango de Valores	Cardinalidad
Número de honeypots	Honeynet	Entero	1 ..	(1, N)
Generación	Arquitectura	Cadena de caracteres	--	(1, 1)
Nivel de operación	Arquitectura - Física	Cadena de caracteres	--	(1, 1)
Nivel de operación	Arquitectura - Virtual	Cadena de caracteres	--	(1, 1)
Nivel de iteración	Honeypots	Cadena de caracteres	--	(1, 1)
Tipos de datos	Integración de datos	Cadena de caracteres	--	(1, N)
Nombre Tipo de herramienta	Herramienta de captura de datos	Cadena de caracteres	--	(1, 1)
Tipo de Configuración	Paquetes - Logs	Entero	1 ..	(1, N)
Tiempo	Mapeo de datos	Cadena de caracteres	--	(1, N)
Capacidad	Almacenamiento base de datos		--	
Nombre	Host - RP	Cadena de caracteres	--	(1, 1)
Protocolo				(1, N)
Puerto				(1, N)
Dirección IP				(1, 1)
Sistema Operativo				(1, N)
Funcionalidad	Servidor - RP	Cadena de caracteres	--	(1, 1)
Tipo				(1, 1)
Protocolo				(1, N)
Puerto				(1, N)
Dirección IP				(1, N)

Sistema Operativo				
Tipo Herramientas Perfil	Intrusión	Cadena de caracteres	--	(1, 1) (1, N) (1, 1)
Nombre Protocolo Puerto Dirección IP Sistema Operativo	Host - I	Cadena de caracteres	--	(1, 1) (1, N) (1, N) (1, 1) (1, N)
Tipo	Router	Cadena de caracteres	--	(1, 1)
Tipo Condición Interfaz	Firewall	Cadena de caracteres	--	(1, 1) (1, 1) (1, 1)
Proceso Monitoreo Resultados	Agente	Cadena de caracteres	--	(1, N) (1, 1) (1, N)

Fuente. El autor.

Paso 7. Describir Atributos de las Clases. Se describe en detalle todos los atributos de clase incluidos en el diccionario de conceptos. Para cada atributo de clase, el desarrollador de la ontología debe rellenar la siguiente información: nombre del atributo, nombre del concepto donde el atributo se define, tipo de valor. La Tabla 9 describe los atributos de las clases de la ontología para integración de datos en honeynet.

Tabla 9. Sección de la tabla de atributos de clase de la ontología para integración de datos en honeynet

Nombre del Atributo de Clase	Concepto	Tipo de Valor
Tipo de monitoreo	Honeynet	Cadena de caracteres
Tipo de instalación	Herramienta de Captura de datos	Cadena de caracteres
Tipo de control	Intrusión	Cadena de caracteres
Tipo de bloqueo	Firewall	Cadena de caracteres

Fuente. El autor.

Paso 8. Describir Constantes. El objetivo de esta tarea es describir en detalle cada una de las constantes identificadas en el glosario de términos. La Tabla 10 describe las constantes de la ontología para integración de datos en honeynet.

Tabla 10. Sección de la tabla de constantes de la ontología para integración de datos en honeynet

Nombre	Tipo de Valor	Valor	Unidad de Medida
Ancho de banda	Numérico	320 Mbps	Bps
Dirección outside del firewall	Binario de 32 bits	190.15.135.2	
Capacidad del canal	Numérico	1000 Mhz	Hz

Fuente. El autor.

Paso 9. Describir Axiomas Formales. Se identifica los axiomas formales que son necesarios en la ontología de integración de datos, y describirlos de manera precisa. Para cada definición de axioma formal, methontology propone especificar información como: nombre, descripción en lenguaje natural, expresión lógica que define de manera formal el axioma usando lógica de primer orden, y los conceptos, atributos y relaciones ad hoc utilizadas en el axioma, así como las variables utilizadas. La Tabla 11 describe los axiomas formales en la ontología para integración de datos en honeynet.

Tabla 11. Sección de la tabla de axiomas formales de la ontología para integración de datos en honeynet

Nombre del Axioma	Descripción	Expresión	Conceptos	Relaciones	Variables
Generación de logs	Toda herramienta de captura de datos genera paquetes - log	(existe(?X, ?Y) (herramienta_captura_data?X) y genera(paquetes_log?Y)	Herramienta de captura de datos Paquetes - logs	genera ingresa	?X ?Y
Almacenamiento de logs	Se almacena los logs en una base de datos	(almacena(?Y, ?Z) (paquetes_log?Y) y almacena(base_datos?Z)	Paquetes - Logs Almacenamiento base de datos	ingresa	?Y ?Z
Integración de datos	Todos los paquetes - logs,	(existe(?Y, ?Z, ?I)	Integración de datos	integra	?Y ?Z ?I

mediante los logs recogidos	almacenados en la BD se lo realiza la integración de la data	(paquetes_ logs?Y) y almacena(base_datos ?Z) y integra(int egración_d ata?I)	Paquetes Logs	-
-----------------------------------	--	--	------------------	---

Fuente. El autor.

Paso 10. Describir Reglas. Se identifica qué reglas se necesitan en la ontología de integración de datos que serán descritas en la tabla de reglas. Para cada regla, methontology propone incluir la siguiente información: nombre, descripción en lenguaje natural, expresión que describe formalmente la regla y conceptos; atributos y relaciones ad hoc utilizados en la regla, así como las variables usadas. Se especifica las expresiones de las reglas utilizando el formato si <condiciones> entonces <consecuente>. La parte izquierda de la regla es una conjunción de condiciones simples, mientras que la parte derecha es una simple expresión de un valor de la ontología. La Tabla 12 describe las reglas de la ontología para integración de datos en honeynet.

Tabla 12. Sección de la tabla de reglas de la ontología para integración de datos en honeynet

Nombre de la Regla	Descripción	Expresión	Conceptos	Atributos	Relaciones	Variables
Negociación Automática	Se negocia la velocidad del puerto entre el dispositivo y el host y se negocia a la velocidad más baja.	Si [velocidad del puerto](?X) y velocidad del dispositivo(?Y) y $?X > ?Y$ entonces negocia [velocidad del dispositivo] (?Y) sino negocia [velocidad del puerto] (?X)	Puerto del equipo Dispositivo	Velocidad	Negocia	?X ?Y
Servidor web para protección se habilita el puerto 80	Por seguridad en un servidor web solo se habilita el puerto 80	Si [servidor web](?A) y puerto(?B) y $?A \Rightarrow 80$ entonces protegido [servidor web](?A)	Servidor web Puerto	Número de puerto	Habilita	?A ?B=80
Ataque TCP (conexión)	Definir un umbral para protección de ataques SYN	Si [umbral de protección](?U) y Conexiones TCP(?C) y $?U > 200$ conexiones por segundo entonces se ha producido un ataque	Umbral Conexiones TCP Ataque	Número de conexiones	Define	?U ?C=200 conexiones por segundo
Jumbo Frames	Descartar paquetes de tamaño mayor a las tramas permitidas	Si [tamaño del paquete](?P) y Tamaño de tramas permitidas(?T) y $?P > 1500$ bytes entonces se produce Jumbo Frames y se descarta el paquete	Trama permitida Paquete de datos Jumbo Frames	Tamaño del paquete Tamaño de la Trama	Permite	?P ?T=1500 bytes

Fuente. El autor.

Paso 11. Describir Instancias. Una vez que el modelo conceptual de la ontología de integración de datos ha sido creado, se pueden definir las instancias que aparecen en el diccionario de conceptos.

Un conjunto preliminar de clases desarrolladas para la representación de la integración de datos en la honeynet corresponde a las que se observan en el editor Protégé, (ver Figura 13).

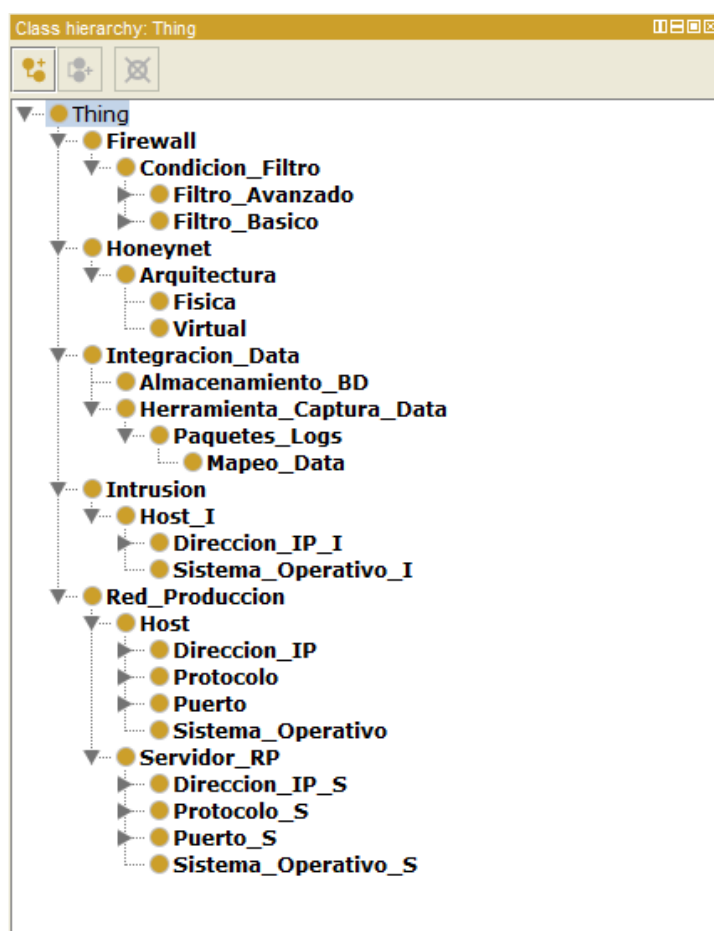


Figura 13. Clases y subclases de la ontología para la integración de datos en honeynet

Fuente. El autor.

En la siguiente figura se muestra el fragmento de jerarquía de integración de datos, (ver Figura 14).

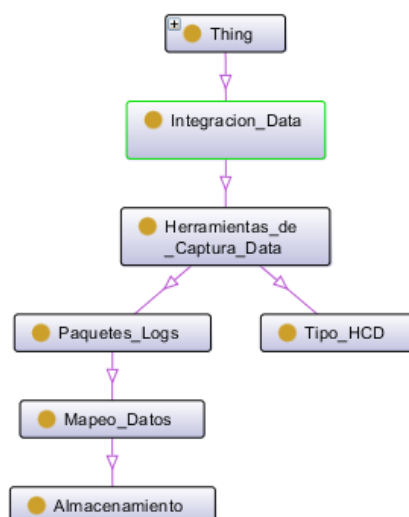


Figura 14. Fragmento de jerarquía de integración de datos

Fuente. El autor.

Las principales clases de la ontología honeynet para integración de datos se describen a continuación:

- Una de las dificultades encontradas en la honeynet es la integración de datos de las diferentes herramientas configuradas ya que generan logs propios y deben ser interpretados de manera individual por parte del administrador de red, el objetivo es tener de manera centralizada la recopilación de los datos para la toma de decisiones oportunas.
- Las herramientas de captura de datos sirven para monitorear permanentemente el acceso a la red. Las herramientas que se utilizan para la captura de datos son snort, p0f, argus, sebek, con la finalidad de capturar el comportamiento del intruso dentro de la honeynet.
- Los paquetes logs, son los resultados de cada una de las herramientas de captura de datos, se han analizado sus cabeceras de información para la integración de todos y cada uno de estos logs en un log centralizado.
- El mapeo de datos se trata de emparejar un par de entidades, con la finalidad de transformar datos de un formato a otro.
- El almacenamiento de la información consiste en tener en una base de datos unificada toda la información recopilada de la honeynet, para tener una adecuada administración de la red y tomar decisiones.

4.1.3 Validación de la ontología

Para realizar la validación de la ontología para integración de datos en honeynet, se tuvo la ayuda de 12 expertos que participaron en el experimento, especialistas en el área de seguridad en redes con conocimientos en desarrollo de ontologías, que luego de revisar y utilizar la ontología honeynet para integración de datos, fueron encuestados agrupando en dos grupos de preguntas, las primeras preguntas están orientadas a cumplir los aspectos que determina methontology en la ontología honeynet para integración de datos: términos necesarios, conceptos identificados, relaciones utilizadas, taxonomía, propiedades, instancias, constantes, diccionario de conceptos, relaciones binarias ad hoc, atributos de instancia, atributos de clase, axiomas, reglas.

El segundo grupo de preguntas tiene como objeto identificar los posibles errores a detectar luego de la evaluación de la ontología con lo que respecta a error de ubicación, error de distribución, error de inconsistencia semántica, error de clases y clasificaciones incompletas, error de omisión de conocimiento disjuncto, error de omisión de conocimiento por falta de exhaustividad, error de redundancia, error en la poca especificación o delimitación de las propiedades de los componentes del sistema, error por falta de exhaustividad en la declaración de etiquetas, y error al no describir correctamente el conocimiento.

El método estadístico utilizado para validar la ontología es el coeficiente de concordancia de Kendall que se utiliza para medir el grado de concordancia o afinidad que presentan los individuos respecto a las k variables, tomando valores entre 0 y 1. Los valores W próximos a 0 indican total desacuerdo entre los individuos, mientras que los valores próximos a 1 indican total acuerdo. Si los expertos fueran totalmente concordantes, el rango asignado a los valores de una de las características principales sería sistemáticamente igual a 1, el asignado a una segunda característica, sistemáticamente igual a 2, y así sucesivamente. En consecuencia, las medias de los rangos de las n características principales serían iguales a 1, 2, ..., n . Si, por el contrario, los expertos fueran totalmente discordantes las medias de los rangos de las n características principales serían aproximadamente iguales entre sí.

Hipótesis propuesta:

H_0 : No existe concordancia entre los expertos

H_1 : Existe concordancia entre los expertos

No obstante, al estadístico W de Kendall, el rechazo o no rechazo de la hipótesis nula vendrá dado por la significación de dicho estadístico, no rechazando la hipótesis nula si la significación del estadístico es mayor que el nivel alfa preestablecido de 0,05, y se rechazará dicha hipótesis si es menor o igual al nivel alfa.

Al analizar los resultados estadísticos del primer grupo de preguntas, donde más discordancia existe es en la número 5 (El sistema taxonómico utilizado para el desarrollo en la ontología es claro, consistente, flexible, exhaustivo y práctico), puesto que tiene una media de 2,58 por lo cual mediante las recomendaciones de los expertos se tomaron medidas adecuadas para mejorar dicho parámetro.

La Tabla 13 describe el estadístico descriptivo de los aspectos de la metodología methontology en la construcción de la ontología honeynet para la integración de datos.

Tabla 13. Estadísticos descriptivos - Aspectos de methontology

	N	Media	Desviación estándar	Mín.	Máx.	Rango
Se presentan todos los términos necesarios para la construcción de la ontología	12	2,67	0,492	2	3	6,29
Los conceptos (identificados) utilizados en la ontología son los adecuados	12	3,00	0,000	3	3	8,79
Las relaciones utilizadas en la ontología representan un tipo de asociación entre conceptos del dominio	12	3,00	0,000	3	3	8,79
Las taxonomías utilizadas en el proceso de la construcción en la ontología, establecen los	12	2,83	0,389	2	3	7,54

conceptos que definen su jerarquía						
El sistema taxonómico utilizado para el desarrollo en la ontología es claro, consistente, flexible, exhaustivo y practico.	12	2,58	0,515	2	3	5,67
En el desarrollo de la ontología se encuentran especificadas las propiedades que describen cada concepto de la taxonomía	12	3,00	0,000	3	3	8,79
Se encuentran identificadas todas las instancias en la ontología.	12	2,83	0,389	2	3	7,54
Está usted de acuerdo con las constantes (valores numéricos que no cambian en un largo período de tiempo) utilizadas en la ontología	12	3,00	0,000	3	3	8,79
El diccionario de conceptos contiene todos los conceptos del dominio, sus relaciones, sus instancias, y sus atributos de clase y de instancia.	12	3,00	0,00	3	3	8,79
Se describen en detalle todas las relaciones binarias ad hoc identificadas en el diagrama de relaciones binarias e incluidas en el diccionario de conceptos.	12	2,92	0,289	2	3	8,17
Se describe en detalle todos los atributos de instancia incluidos en el diccionario de conceptos.	12	2,92	0,289	2	3	8,17
Se describe en detalle todos los atributos de clase incluidos en el diccionario de conceptos	12	3,00	0,00	3	3	8,79
Los axiomas formales utilizados en la ontología	12	3,00	0,00	3	3	8,79

son expresiones lógicas siempre verdaderas utilizadas para definir restricciones en la ontología.

Las reglas utilizadas en la ontología sirven para inferir conocimientos en la ontología

12	2,67	0,492	2	3	6,29
----	------	-------	---	---	------

Fuente. El autor.

La Tabla 14 indica el estadístico de prueba de los aspectos de la metodología methontology en la construcción de la ontología honeynet para la integración de datos.

Tabla 14. Estadísticos de prueba - Aspectos de methontology

N	12
W de Kendall ^a	0,218
Chi-cuadrado	36,650
Gl	14
Sig. asintótica	0,001

a. Coeficiente de concordancia de Kendall

Fuente. El autor.

Como la significación de la W de Kendall (0,001) es menor o igual que el nivel alfa preestablecido, es evidencia a favor de la hipótesis alternativa, por lo que existe concordancia entre los expertos. Lo que se verifica en la Figura 15, cuyo p-valor se localiza en la zona de rechazo.

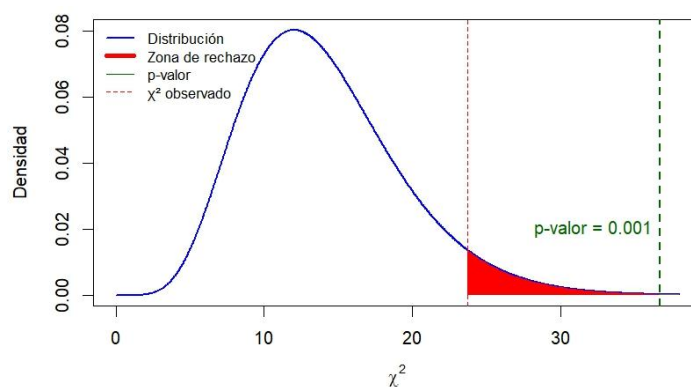


Figura 15. Regla de decisión - Aspectos de methontology

Fuente. El autor

Respecto al segundo grupo de preguntas, donde más discordancia existe es en el número 2 (error de distribución), puesto que tiene una media de 2,67, esto indica que se debería mejorar, para evitar incurrir en el error de distribución.

La Tabla 15 describe el estadístico descriptivo sobre la identificación de posibles errores en la construcción de la ontología honeynet para la integración de datos.

Tabla 15. Estadísticos descriptivos - Identificación de posibles errores

	N	Media	Desviación estándar	Mín.	Máx.	Rango
Error de ubicación	12	3,00	0,000	3	3	5,92
Error de distribución	12	2,67	0,492	2	3	4,25
Error de inconsistencia semántica	12	2,83	0,389	2	3	5,08
Error de clases y clasificaciones incompletas	12	3,00	0,000	3	3	5,92
Error de omisión de conocimiento disjunto	12	2,92	0,289	2	3	5,50
Error de omisión de conocimiento por falta de exhaustividad	12	3,00	0,000	3	3	5,92
Error de redundancia	12	2,75	0,452	2	3	4,67
Error en la poca especificación o delimitación de las propiedades de los componentes del sistema	12	3,00	0,000	3	3	5,92
Error por falta de exhaustividad en la declaración de etiquetas	12	3,00	0,000	3	3	5,92

Error al no describir							
correctamente	el	12	3,00	0,000	3	3	5,92
conocimiento							

Fuente. El autor.

La Tabla 16 indica el estadístico de prueba para la identificación de posibles errores en la construcción de la ontología honeynet para la integración de datos.

Tabla 16. Estadísticos de prueba - Identificación de posibles errores

N	12
W de Kendall ^a	0,185
Chi-cuadrado	20,000
Gl	9
Sig. Asintótica	0,018

a. Coeficiente de concordancia de Kendall

Fuente. El autor.

Como la significación de la W de Kendall (0,018) es menor o igual que el nivel alfa preestablecido, es evidencia a favor de la hipótesis alternativa, por lo que existe concordancia entre los expertos. Lo que se verifica en la Figura 16, cuyo p-valor se localiza en la zona de rechazo.

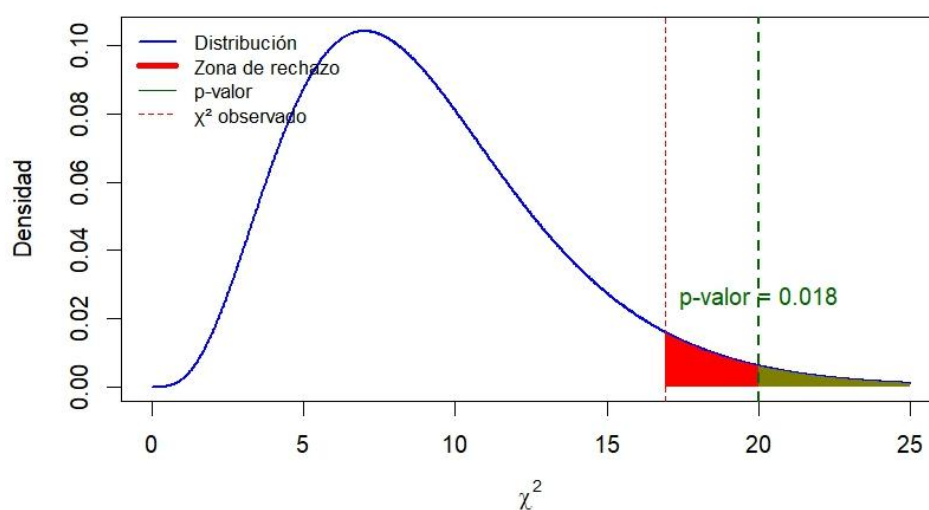


Figura 16. Regla de decisión - Identificación de posibles errores

Fuente. El autor

Las ontologías son la base de la Web Semántica, y en ellas se ha visto la posibilidad de ofrecer un sistema que permita una recuperación “inteligente” de la información. Para el desarrollo de una ontología, se debe inicialmente identificar el contexto, conceptualizar el conocimiento y representarlo adicional a la construcción y validación de la misma. La adquisición de conocimiento es uno de los procesos más complejos en el desarrollo de ontologías, en este paso se deben consultar diversas fuentes de información, determinando los principales aspectos referentes al dominio tratado y así obtener un conocimiento efectivo que permite la determinación de las clases, las instancias, las relaciones, los atributos, los cuales representan su conocimiento.

En el primer grupo de preguntas referentes a los aspectos considerados para la ontología honeynet para la integración de datos se obtuvo la significación de la W de Kendall (0,001) es menor o igual que el nivel alfa preestablecido, por lo que existe concordancia entre los expertos que, los aspectos considerados por methontology son los adecuados.

En el segundo grupo de preguntas referentes a los posibles errores a considerar en la elaboración de la ontología honeynet para la integración de datos se obtuvo la significación de la W de Kendall (0,018) es menor o igual que el nivel alfa preestablecido, por lo que existe concordancia entre los expertos.

Como base en la ontología honeynet para integración de datos se propone este trabajo de investigación mediante una arquitectura de honeynet con integración de datos para el análisis de evidencia digital que permita a los administradores de red una mejor gestión para la toma de decisiones de manera oportuna en un menor tiempo.

4.2 Arquitectura de honeynet con integración de datos para el análisis de evidencia digital

La integración de datos proporciona la capacidad de manipular datos de forma transparente de varias fuentes, donde los conjuntos de datos se producen de forma independiente. Las fuentes de datos pueden ser heterogéneas en sintaxis, esquema o semántica, lo que hace que la interoperación de datos sea una tarea compleja. Si bien existe cierta discrepancia en la implementación de un sistema interoperable, los

investigadores y la industria coinciden en que la interoperabilidad se basa en al menos tres capas: la interoperabilidad técnica, que busca superar la heterogeneidad del intercambio de datos de bajo nivel; la interoperabilidad sintáctica, que busca superar la heterogeneidad de los formatos en los que se expresan los datos; y, finalmente, la interoperabilidad semántica, que busca superar los modelos heterogéneos que utilizan los datos, traduciéndolos en información significativa para que puedan utilizarse de forma transparente (Cimmino et al., 2024).

Un sistema central de integración de datos generalmente tiene un esquema, que proporciona al usuario una interfaz uniforme para acceder a la información. La ontología honeynet para integración de datos, se toma como punto de partida para realizar la arquitectura de honeynet con integración de datos para el análisis de evidencia digital.

Para lograr la interoperabilidad de los datos, es necesario eliminar los problemas que plantea la heterogeneidad de los datos.

4.2.1 Arquitectura honeynet

Las herramientas de captura de datos en un entorno honeynet, sirven para monitorear permanentemente el acceso a la red. Se muestra el esquema propuesto con las herramientas que se utilizan para la captura de datos: firewall, sebek, snort, argos, p0f, con la finalidad de capturar el comportamiento del intruso dentro de la honeynet.

Los paquetes logs, son los resultados de cada una de las herramientas de captura de datos de los cuales se han analizado sus cabeceras de información para la integración de todos y cada uno de estos logs, los cuales se dirigen a una base de datos central con toda la información recopilada de la honeynet, (ver Figura 17). Se ha desarrollado un aplicativo con interfaz web para la integración de datos con la finalidad de tener una adecuada administración de la red y tomar decisiones de una manera adecuada y oportuna.

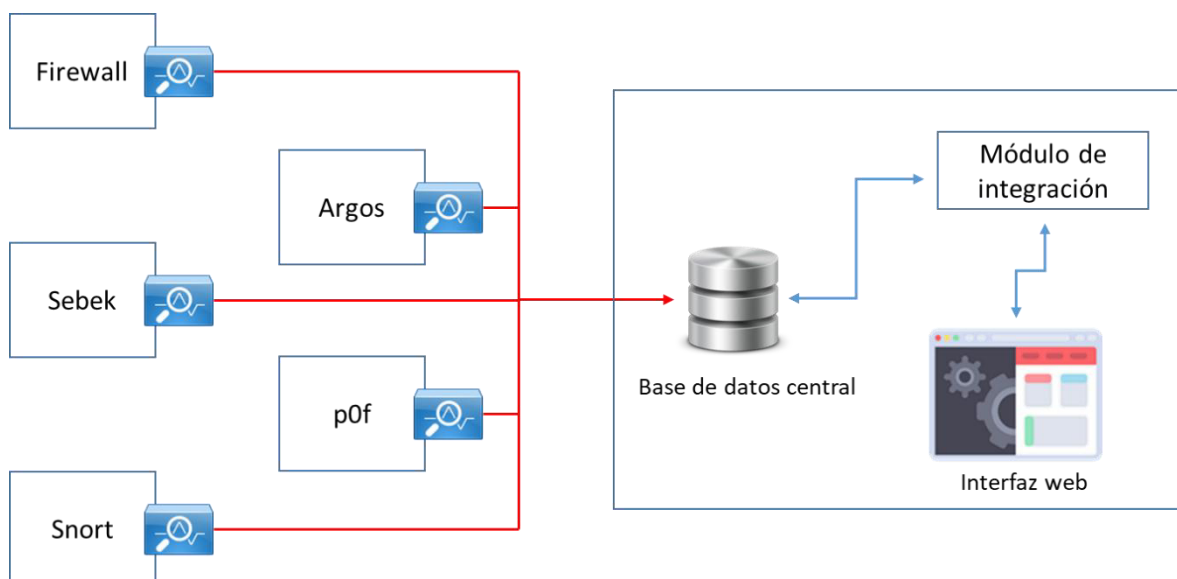


Figura 17. Esquema propuesto

Fuente. El autor.

La arquitectura honeynet con integración de datos ha sido implementada en la Universidad Nacional de Chimborazo - Ecuador, configurada la honeynet en un sistema operativo linux y windows de manera física y virtual, se ha configurado un sistema de detección de intrusos snort, un servidor de logs en el que se almacenaran todos los logs de las herramientas de monitoreo de tráfico, un servidor de monitoreo en el cual se instalará el aplicativo para integración de datos.

Se ha configurado el nivel de seguridad tanto de la red de producción como de la honeynet, dando prioridad a la red de producción, sin desmerecer a la honeynet para que los intrusos no detecten de su presencia, como se muestra, (ver Figura 18).

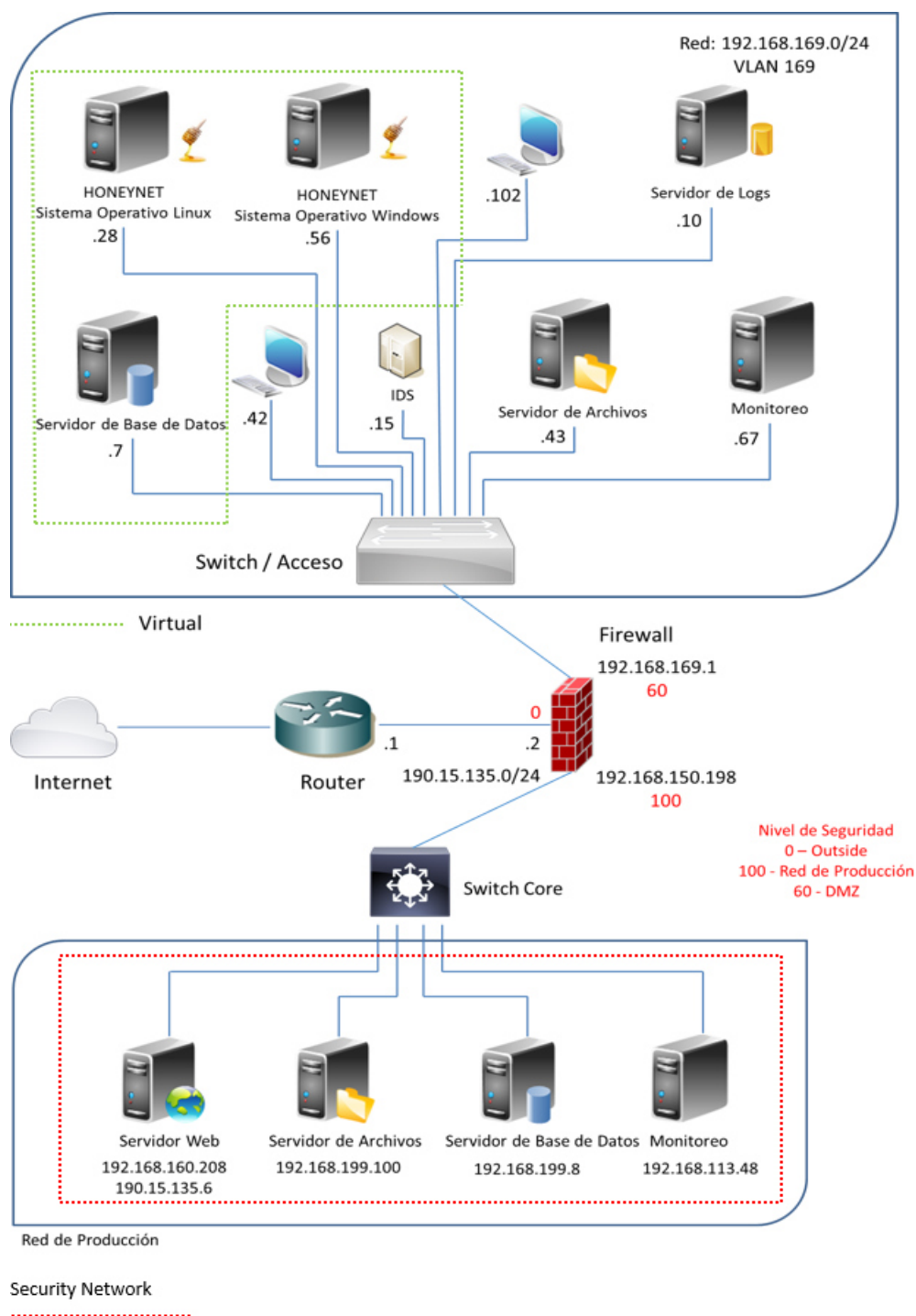


Figura 18. Arquitectura honeynet
 Fuente. El autor.

4.2.2 Aplicativo de integración de datos

Sistema informático que integra los logs de las herramientas de monitoreo de tráfico y captura de datos en una honeynet, cada herramienta genera su propio log, los logs se almacenan en un directorio denominado Integración, puesto que los logs se ejecutan en tiempo real lo que impide su análisis en ese momento, se copiará diariamente los logs de cada una de estas herramientas para tener de forma centralizada la información.

Los logs de cada una de las herramientas de captura de datos se los coloca en un directorio aparte con el nombre de la herramienta de captura de datos, como se muestra en la siguiente figura, (ver Figura 19).

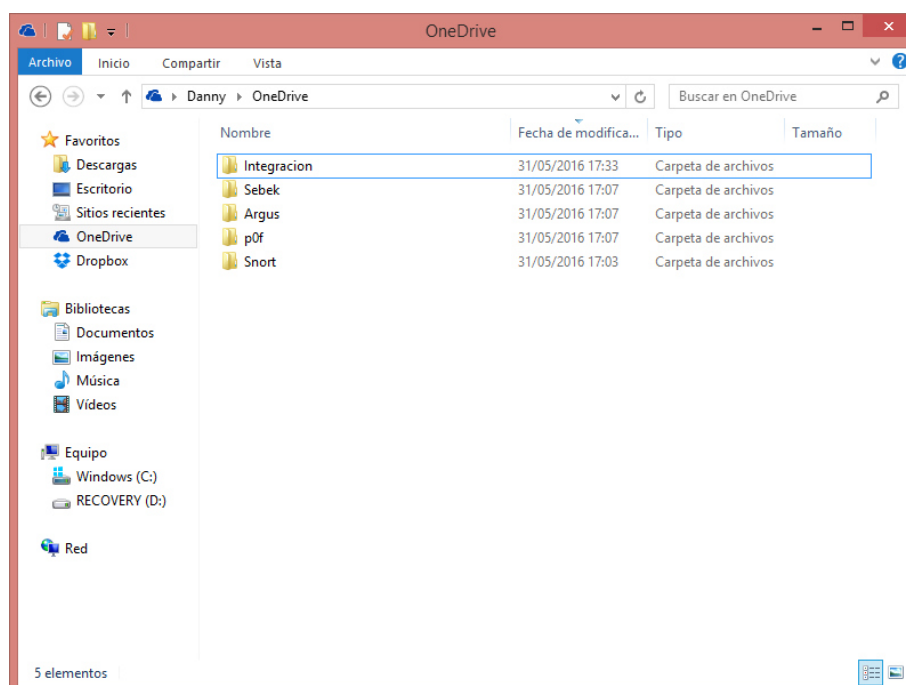


Figura 19. Almacenamiento centralizado

Fuente. El autor.

El formato de almacenamiento de cada uno de los logs que generan las herramientas de captura de datos será:

Se-dd-mm-aa

A-dd-mm-aa

p-dd-mm-aa

Sn- dd-mm-aa

Posterior se realiza la integración, diariamente que se coloca en la carpeta Integración con el siguiente formato: I-dd-mm-aa, como se muestra, (ver Figura 20).

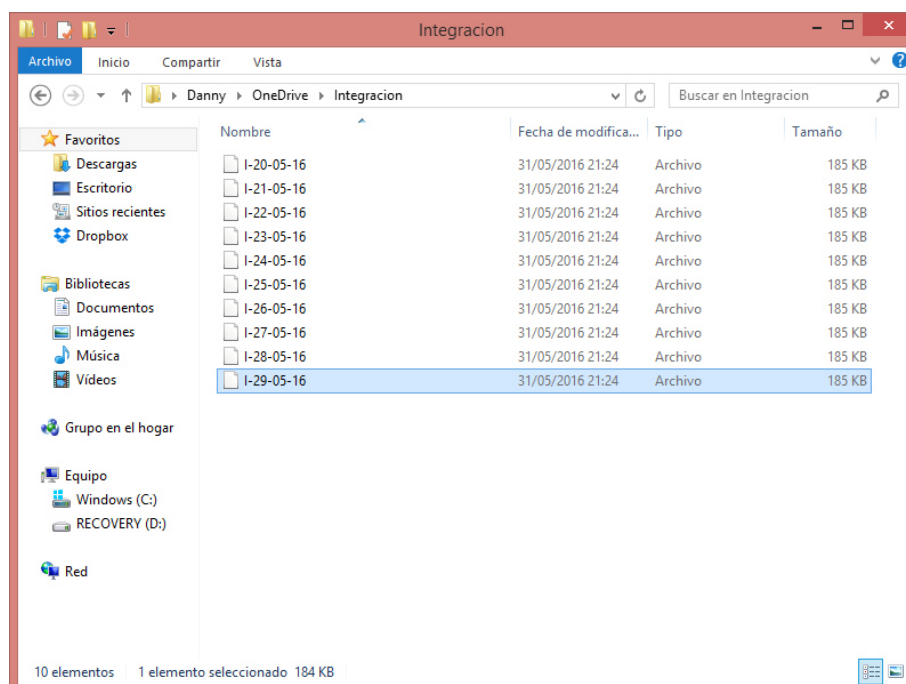


Figura 20. Integración de datos

Fuente. El autor.

En el caso que no se pueda leer un log de alguna de las herramientas de captura de datos ya sea a nivel de archivo o a nivel de línea, se colocará en la base de datos el archivo que no logró capturar para mapear la información.

Se establecen diferentes estadísticas que se manejarán con los datos almacenados en la base datos unificada, estadísticas del tipo de intrusión, por ejemplo: incidentes registrados, porcentajes de ataques por sistemas operativos, intentos de conexión contra los puertos, reconocimiento de ataques por inyección SQL, tráfico normal, tráfico de ataque, periodo de análisis.

Para solicitar consultas de algún requerimiento específico, se lo resolverá mediante un lenguaje de manipulación de datos. Se presenta el caso de uso de la integración de datos en una honeynet (ver Figura 21).

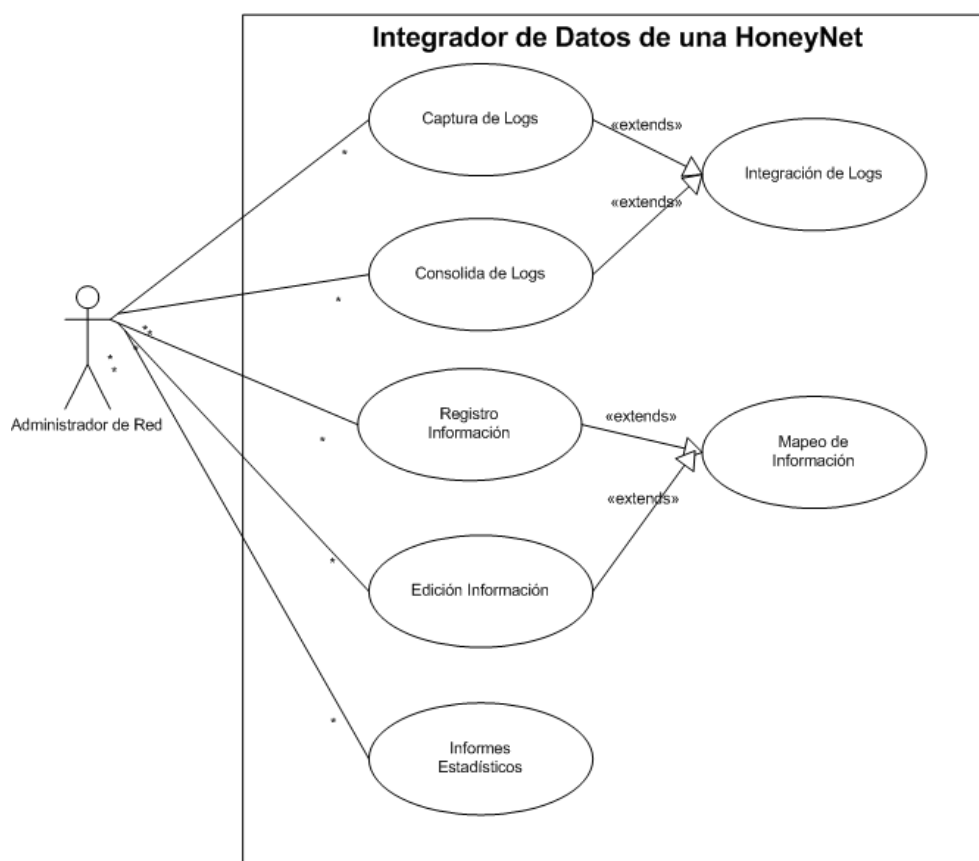


Figura 21. Caso de uso: Integración de datos en una honeynet

Fuente. El autor.

4.3 Análisis, interpretación y discusión de resultados

Para evaluar el impacto de la arquitectura honeynet con integración de datos en el análisis de la evidencia digital, se empleó un diseño cuasiexperimental. Este enfoque se caracteriza por la manipulación controlada de la condición experimental, en este caso particular, la modificación de la arquitectura honeynet, sin que exista la asignación aleatoria de los expertos, quienes se seleccionaron conforme a criterios muestrales, garantizando la pertinencia y experiencia necesaria para la evaluación de la arquitectura. Esto responde a la naturaleza propia de los cuasiexperimentos, en los que los grupos o participantes no pueden ser reasignados libremente debido a su nivel de especialización o a restricciones operativas.

Bajo esta modalidad, se compararon dos condiciones de observación: la arquitectura sin integración de datos y la arquitectura con integración de datos. Para medir el efecto de esta intervención, se aplicó un esquema de mediciones pretest y posttest, utilizando el mismo entorno físico y lógico de la honeynet, lo que permitió observar los cambios

en el desempeño del análisis de la evidencia digital antes y después de la modificación. Este diseño proporcionó un marco riguroso para determinar el impacto de la integración de datos en condiciones reales y controladas de operación.

- Arquitectura honeynet sin integración de datos, empleando únicamente los logs generados de forma independiente por cada herramienta de captura (Sebek, Snort, P0f, Argus y firewall), generando 525 registros con una cantidad de 131.26 Mb.
- Arquitectura honeynet con la herramienta de integración de datos, en la cual los registros fueron centralizados e interpretados de manera unificada.

El tiempo de la captura de datos se realizó por un periodo de tres meses y medio, lapso suficiente para garantizar la captura de eventos representativos, incluyendo tráfico normal, actividades sospechosas y múltiples tipos de ataques como fuerza bruta, escaneo de puertos, inyección SQL, intentos de DDoS, ICMP malicioso, manipulación DNS y otros patrones detectados por las herramientas de captura.

El conjunto de datos recopilados en ambas fases fue almacenado y documentado cuidadosamente, estructurándose un repositorio organizado para evitar sesgos de comparación. Finalizado el periodo de captura, los datos fueron entregados a los expertos, quienes evaluaron a través de los instrumentos de recolección de datos. Ver (Anexo B).

Una vez aplicados los instrumentos de recolección de datos a los expertos y obtenidas sus evaluaciones sobre la propuesta, se presentan los resultados derivados de dichas valoraciones. Estos se organizan y analizan de acuerdo con las variables dependientes del estudio y sus respectivas dimensiones. Finalmente, se expone la verificación de las hipótesis de investigación, las cuales se encuentran directamente vinculadas con las variables y con los objetivos previamente establecidos en el estudio.

Es pertinente señalar que, para el análisis se contrastaron las mismas variables en ambos cuestionarios y los datos se presentan distribuidos en tablas indicando la frecuencia absoluta (f) y la frecuencia relativa (%) considerando que la muestra está constituida por los expertos quienes examinaron la honeynet sin la integración de datos

y, posteriormente, con el módulo de integración, además de ello, se emplean gráficos para apoyar la comprensión de estos resultados y los análisis.

4.3.1 Variable independiente - Dimensión Operatividad Técnica

Resuelve problemas de interoperabilidad en el análisis de evidencia digital.

Tabla 17. Interoperabilidad

Ítem del cuestionario	Indicador	Honeynet sin integración de datos				Honeynet con integración de datos			
		Cuestionario #1				Cuestionario #2			
		Positivo		Negativo		Positivo		Negativo	
		<i>f</i>	%	<i>F</i>	%	<i>F</i>	%	<i>f</i>	%
7	Acceso a datos de diferentes fuentes	---	---	4	100	4	100	---	---
8	Acceso a datos desde un mismo dispositivo	---	---	4	100	4	100	---	---

Fuente. El autor.

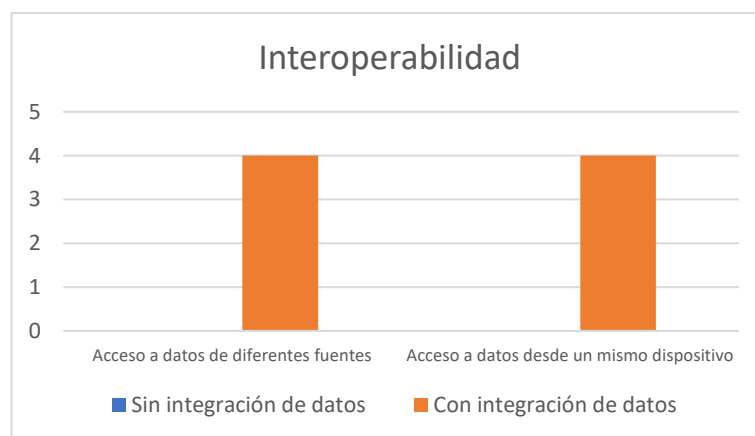


Figura 22. Interoperabilidad

Fuente. El autor.

Los sistemas existentes suelen presentar bajos niveles de interoperabilidad de datos, operaciones asincrónicas entre recursos e incapacidad para determinar y mantener el estado óptimo en tiempo real (Zhang et al., 2025). Sin embargo, lograr la interoperabilidad sigue siendo un desafío importante, especialmente al abordar problemas sintácticos y semánticos (Barangi et al., 2025). La propuesta técnica presenta una solución al problema de interoperabilidad, permite el acceso a los datos provenientes de diferentes fuentes, centralizándolos en un mismo dispositivo, lo que

implica, una mayor eficacia en la respuesta a incidentes en el campo de la ciberseguridad.

Para el caso de la interoperabilidad de la arquitectura, de conformidad con la Tabla 17 y la Figura 22, considerando que se trata de la capacidad de la honeynet para ejecutar de manera efectiva y eficiente el acceso a los datos desde diferentes fuentes de una sola vista y, el acceso a los datos de diferentes fuentes desde un mismo dispositivo, el 100% de los expertos luego de su análisis, confirman que las evidencias resaltan que la herramienta de integración satisface los indicadores de acceso a los datos, bien sea desde fuentes diversas o únicas, usando un mismo dispositivo o varios sin comprometer otros elementos en esta función, lo que favorece el análisis de evidencia digital de una manera centralizada.

Este hallazgo valida la importancia de la interoperabilidad, que es un factor crucial para que las empresas tomen decisiones operativas y estratégicas informadas y logren una ventaja competitiva en el mercado (Jain et al., 2022).

4.3.2 Variable dependiente - Dimensión Captura de evidencia digital

Mejora la captura de evidencia digital.

Tabla 18. Captura de evidencia digital

Ítem del cuestionario	Indicador	Honeynet sin integración de datos				Honeynet con integración de datos			
		Cuestionario #1				Cuestionario #2			
		<i>Positivo</i>		<i>Negativo</i>		<i>Positivo</i>		<i>Negativo</i>	
		<i>f</i>	<i>%</i>	<i>f</i>	<i>%</i>	<i>f</i>	<i>%</i>	<i>f</i>	<i>%</i>
1	Recopilación de atributos	---	--	4	100	4	100	---	---
2	Intercambio de información	---	--	4	100	4	100	---	---
3	Comparación de atributos	---	--	4	100	4	100	---	---
4	Integración de data	---	--	4	100	4	100	---	---
5	Usa toda la información	---	--	4	100	4	100	---	---
6	Reúsa la información	---	--	4	100	4	100	---	---

Fuente. El autor.

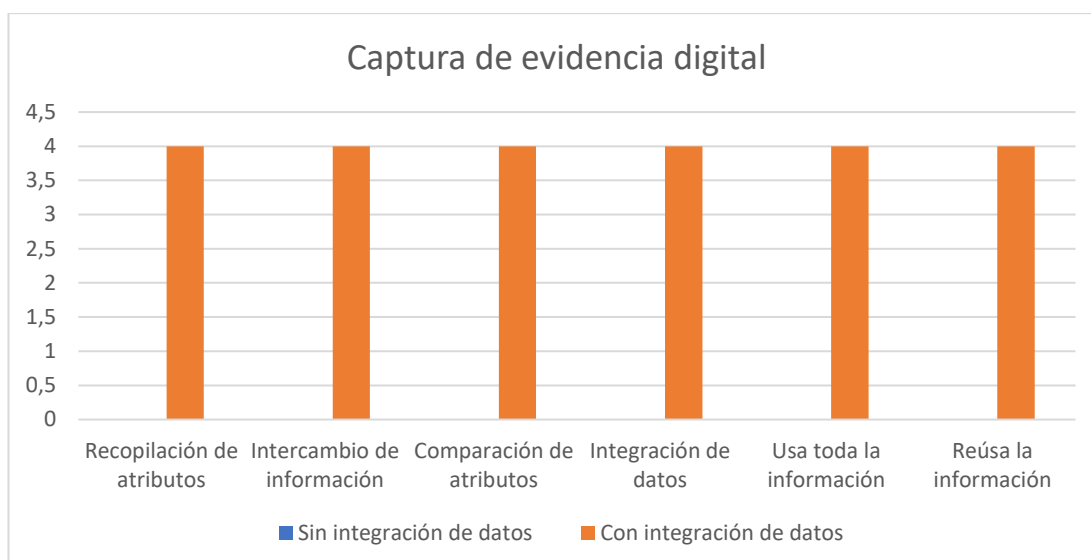


Figura 23. Captura de evidencia digital

Fuente. El autor.

Investigaciones previas han encontrado obstáculos como características limitadas, efectividad reducida, atributos irrelevantes y tareas computacionales complejas (Anqin & Yuwei, 2025). La investigación propone resolver este problema a través de la integración de datos en la arquitectura de la honeynet lo que propicia la recopilación, intercambio y comparación de atributos de los datos capturados. Esta arquitectura permite al analista no solo recolectar de manera sistemática los atributos relevantes del tráfico de red, sino también realizar una comparación efectiva entre ellos, facilitando la correlación de eventos y patrones de ataque. Esta capacidad es consistente con el trabajo de Masmoudi et al. (2024), el cual se centra en la integración semántica de datos, lo que facilita la toma de decisiones de manera oportuna.

En consecuencia, este hallazgo evidencia que la integración de datos no solo optimiza la eficiencia en la recolección y correlación de atributos, sino que también fortalece la trazabilidad, la validez probatoria, consolidando a la honeynet como un recurso esencial para la gestión integral de incidentes y la preservación de evidencias digitales en el campo de la ciberseguridad.

Esta información se verifica en los resultados de la Tabla 18 representados en la Figura 23, donde se recogen las experiencias de los expertos, y se comprueba que la honeynet con integración de datos permite al analista la recopilación, intercambio y comparación de los atributos de la data, además del uso y el reúso de toda la información recopilada,

lo que simplifica las acciones de captura de los datos referidos al funcionamiento de la honeynet.

4.3.3 Variable dependiente - Dimensión Consistencia del proceso de análisis

Mejora la consistencia del proceso de análisis de evidencia digital.

Tabla 19. Consistencia del proceso de análisis

Ítem del cuestionario	Indicador	Honeynet sin integración de datos				Honeynet con integración de datos			
		Cuestionario #1				Cuestionario #2			
		Positivo		Negativo		Positivo		Negativo	
		<i>F</i>	%	<i>f</i>	%	<i>F</i>	%	<i>f</i>	%
9	Mejora la exactitud del análisis	1	25	3	75	4	100	---	---
10	Exactitud de los resultados del análisis	4	100	---	---	4	100	---	---
11	Confianza en el proceso de análisis	3	75	1	25	4	100	---	---
12	Resultado del proceso es confiable	4	100	---	---	3	75	1	25

Fuente. El autor.

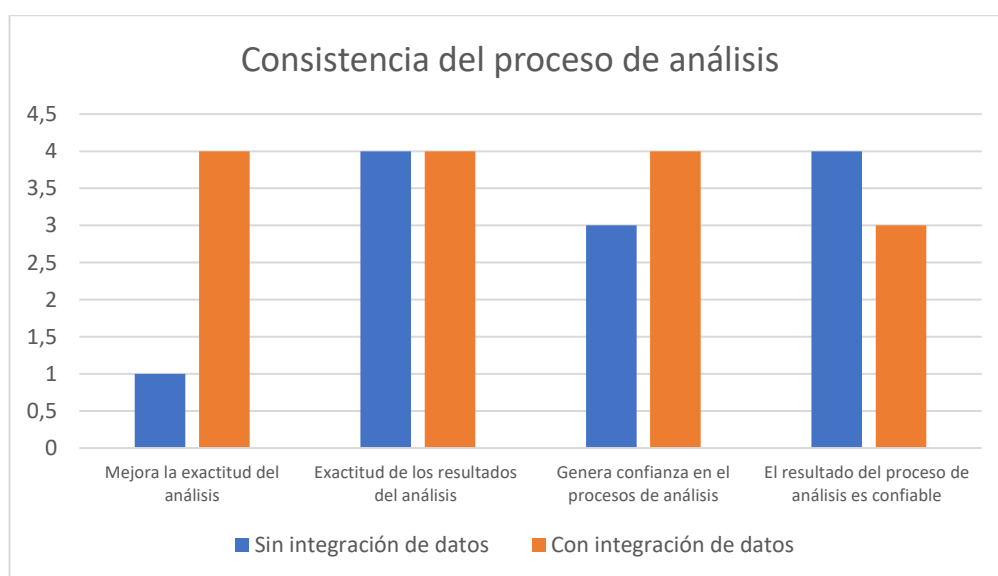


Figura 24. Consistencia del proceso de análisis

Fuente. El autor.

Según investigaciones previas, la exactitud y confianza en la recopilación de datos constituyen un problema recurrente en el ámbito de la ciberseguridad. Diversos estudios han señalado que la disponibilidad de conjuntos de datos de calidad es limitada, lo que afecta la validez de los análisis y la replicabilidad de los resultados (Kouper & Stone, 2024). En entornos de ciberseguridad, se ha evidenciado que los sistemas suelen presentar deficiencias en cuanto a precisión, integridad y consistencia de los datos recopilados (Xu, 2023). Asimismo, revisiones sistemáticas sobre instrumentos de medición en seguridad informática reportan sesgos y falta de claridad metodológica en la recolección de información, lo que compromete la fiabilidad de los hallazgos (Rohan et al., 2023). Estas evidencias ponen de manifiesto que la calidad de los datos continúa siendo un desafío crítico para el desarrollo de soluciones efectivas en este campo.

La integración de datos en la arquitectura de la honeynet asegura que los atributos de tráfico y los eventos de red se correlacionen con mayor precisión, lo que contribuye a reducir la generación de falsos positivos y a optimizar la detección de incidentes, garantizando la fiabilidad del proceso de análisis en todo el ecosistema de ciberdefensa.

Asimismo, los resultados evidencian que la herramienta de integración no solo mejora la exactitud del análisis, sino que también incrementa la confianza de los analistas en el proceso y en los resultados obtenidos. Sin embargo, un 75% de los expertos señaló que, incluso sin la herramienta de integración, la arquitectura de la honeynet permite obtener resultados confiables, lo que resalta el papel relevante de la experiencia del analista, como ya lo advirtieron Chamkar et al. (2022), al destacar que la habilidad y pericia del personal puede compensar parcialmente las limitaciones tecnológicas de ciertas arquitecturas de ciberseguridad.

La consistencia del proceso de análisis se aborda desde los indicadores de exactitud y confianza. Los resultados, expresados en la Tabla 19 y la Figura 24, señala que la herramienta de integración de data mejora la exactitud del análisis y sus resultados, además de ello, genera confianza, tanto en el proceso de análisis como en el resultado.

4.3.4 Variable dependiente - Dimensión Tiempos de detección de ataques

Son aceptables los tiempos de detección de ataques en el análisis de evidencia digital.

Tabla 20. Aceptabilidad de tiempos de detección de ataques

Ítem del cuestionario	Indicador	Honeynet sin integración de datos				Honeynet con integración de datos			
		Cuestionario #1				Cuestionario #2			
		Positivo		Negativo		Positivo		Negativo	
		<i>f</i>	%	<i>f</i>	%	<i>f</i>	%	<i>f</i>	%
13	Detección de Ataques de suplantación	---	-- -	4	100	4	100	---	---
	Detección de Inyecciones SQL en el servidor	---	-- -	4	100	4	100	---	---
	Detección de Fuerza bruta	---	-- -	4	100	4	100	---	---
	Detección de Denegación del servicio	---	-- -	4	100	4	100	---	---
	Detección de Envenenamiento DNS (Domain Name Server)	---	-- -	4	100	4	100	---	---
	Detección de Exploración de puertos	---	-- -	4	100	4	100	---	---
	Detección de Ataques ICMP (Internet Control Message Protocol)	---	-- -	4	100	4	100	---	---
	Detección de Cross Site Scripting	---	-- -	4	100	4	100	---	---

Fuente. El autor.

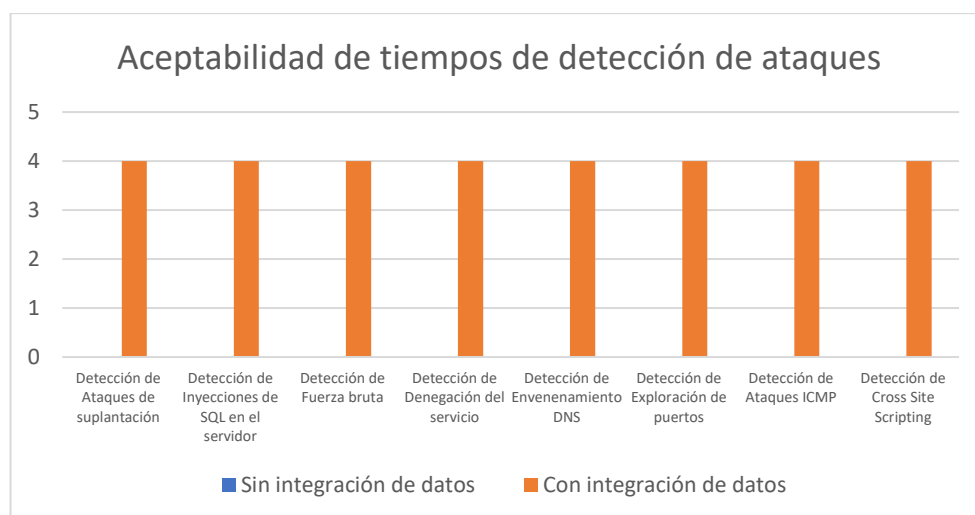


Figura 25. Aceptabilidad de tiempos de detección de ataques

Fuente. El autor.

Los ciberataques son un gran problema en el mundo actual, ya que todo está conectado en sistemas digitales. Pueden dañar la seguridad, la privacidad y el acceso a la información privada. Los sistemas de detección de intrusiones son fundamentales para mantener las redes seguras, ya que detectan y detienen rápidamente las actividades dañinas (Murti et al., 2024). Sin embargo, no siempre estos ataques son detectados a tiempo, lo que impide detener la amenaza con el éxito esperado. Estos hallazgos resaltan la importancia de implementar modelos de ciberseguridad eficientes para garantizar el funcionamiento seguro y fiable frente a la evolución de los ciberataques (Baral et al., 2025). Esta investigación sugiere un nuevo enfoque usando la arquitectura de la honeynet con integración de datos, cuyos tiempos de respuesta son significativamente menores que el sistema tradicional. La adopción de este novedoso enfoque científico en el campo de la ciberseguridad permite contar con defensas más sólidas e inteligentes.

En este caso se encontró que, tal como se expone en la Tabla 20 y en la Figura 25, los tiempos de detección de ataques son aceptables con el uso de la herramienta con integración de datos, en conformidad con la evaluación realizada por los expertos, la propuesta identifica y registra de manera oportuna actividades maliciosas en la red, reduciendo la latencia entre la ocurrencia del ataque y su detección efectiva.

Además, los resultados revelan que la arquitectura no solo mejora la captura de evidencia digital, sino que fortalece la capacidad de respuesta a incidentes en redes corporativas, permitiendo mitigar riesgos antes de que el atacante comprometa servicios críticos o afecte la disponibilidad en la infraestructura.

4.3.5 Variable dependiente - Dimensión Integridad de los datos

Mejora la integridad de los datos en el análisis de evidencia digital.

Tabla 21. Integridad de los datos

Ítem del cuestionario	Indicador	Honeynet sin integración de datos				Honeynet con integración de datos			
		Cuestionario #1				Cuestionario #2			
		Positivo		Negativo		Positivo		Negativo	
		<i>f</i>	%	<i>f</i>	%	<i>F</i>	%	<i>f</i>	%
14	Almacenamiento	3	75	1	25	4	100	---	---
15	Integridad	4	100	---	---	4	100	---	---
16	Consistencia	2	50	2	50	4	100	---	---
17	Interpretación unívoca	2	50	2	50	1	25	3	75
18	Precisión técnica	2	50	2	50	4	100	---	---

Fuente. El autor.

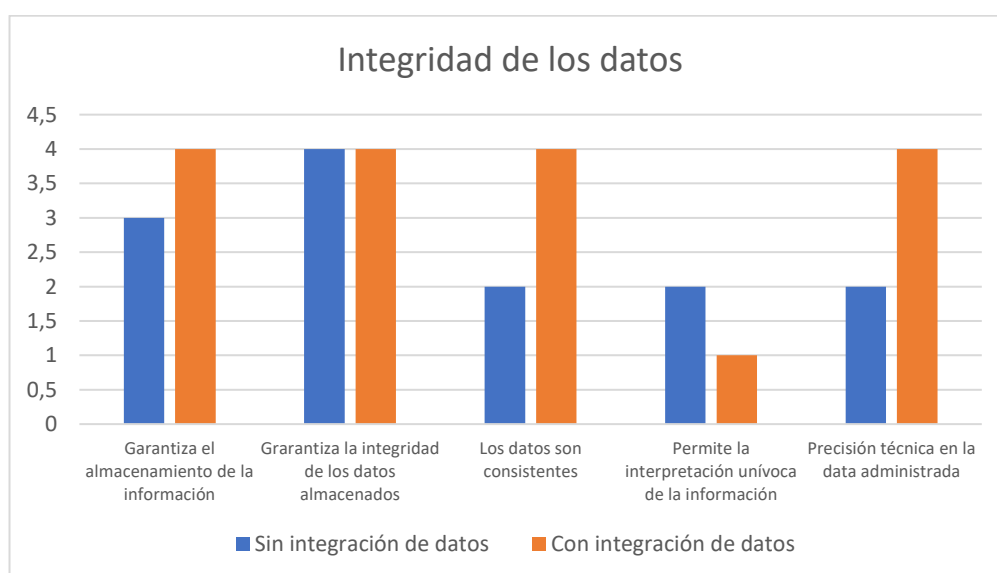


Figura 26. Integridad de los datos

Fuente. El autor.

En el campo de la ciberseguridad, la dimensión de integridad de los datos vinculada a la captura de evidencia digital, lo que significa que la arquitectura de la honeynet debe

garantizar que la información recolectada sea almacenada de forma segura, íntegra y sin alteraciones, manteniendo su consistencia a lo largo del ciclo de análisis. Esto implica que los registros de tráfico, paquetes capturados, logs de firewall o eventos provenientes de IDS no sufran modificaciones que comprometan su validez. La interpretación unívoca asegura que los datos recopilados tengan un único significado dentro del proceso de análisis, evitando ambigüedades en la correlación de eventos de red. Asimismo, la precisión técnica se traduce en la capacidad de capturar atributos de tráfico con detalle sin pérdida de información.

En la actualidad las organizaciones dependen cada vez más de sistemas digitales basados en conjuntos de datos, garantizar su integridad se ha convertido en una preocupación crucial, además la creciente complejidad y escala de las ciberamenazas plantea importantes desafíos para la seguridad, eficiencia e integridad de los datos (Nimanth & Saxena, 2025). A pesar de los avances tecnológicos, persisten los desafíos en la gestión de datos de múltiples fuentes, que pueden generar inconsistencias y errores (Jacobs & Van Der Merwe, 2025). El almacenamiento y tratamiento adecuado de los datos es una solución prometedora para reducir la latencia y prevenir fallos puntuales, pero un factor clave para prevenir decisiones erróneas reside en la integridad de los datos (Xie et al., 2025). Para abordar estos problemas, se propone la arquitectura de la honeynet con integración de datos que garantiza su integridad y se demuestra que es una solución eficaz en la precisión y adaptabilidad en la detección de amenazas, que conlleva a una mejor gestión de la diversidad de ciberataques.

En este sentido, la herramienta con integración de datos, tal como se expone en la Tabla 21 y en la Figura 26, de conformidad con los expertos evaluadores, satisface los requerimientos propios de la integridad de los datos; sin embargo, indicaron que sin integración de datos realiza funciones similares, pero en menor medida, excepto por el indicador almacenamiento de datos, donde se obtuvo la misma respuesta, lo que permite a la propuesta no solo detectar ataques, sino que preserva la validez probatoria, la exactitud y la coherencia de evidencia digital, asegurando que pueda ser utilizada de manera confiable en la toma de decisiones, en la respuesta a incidentes y en procesos legales relacionados con delitos informáticos.

4.3.6 Variable independiente - Dimensión Gestión de la información

Optimiza la gestión de la información en el análisis de evidencia digital.

Tabla 22. Gestión de la información

Ítem del cuestionario	Indicador	Honeynet sin integración de datos				Honeynet con integración de datos			
		Cuestionario #1				Cuestionario #2			
		Positivo		Negativo		Positivo		Negativo	
		<i>f</i>	%	<i>F</i>	%	<i>f</i>	%	<i>f</i>	%
19	Interfaz amigable	---	---	4	100	4	100	---	---
20	Facilidad en la operación	---	---	4	100	4	100	---	---
21	Interacción con el usuario	---	---	4	100	4	100	---	---
22	Generación de	2	50	2	50	4	100	---	---
23	reportes	1	25	3	75	4	100	---	---

Fuente. El autor.

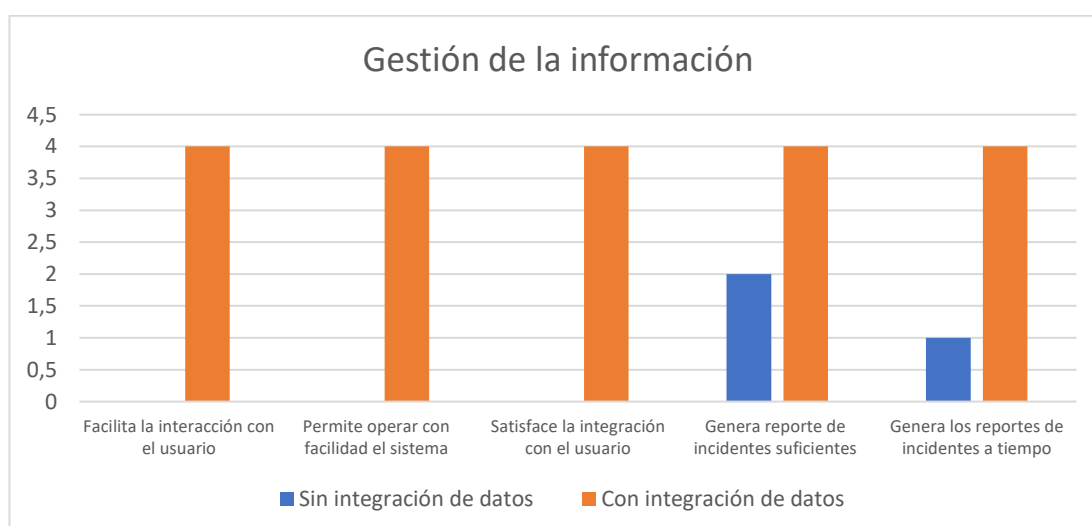


Figura 27. Gestión de la información

Fuente. El autor.

En el campo de la ciberseguridad, la dimensión de gestión de la información implica que la arquitectura de la honeynet debe proporcionar mecanismos que faciliten la operación del sistema y el análisis de incidentes mediante interfaces intuitivas, interactivas y adaptadas al usuario. Una interfaz amigable reduce la complejidad en la interpretación de logs, lo que optimiza la toma de decisiones en escenarios de ataques activos. La facilidad en la operación permite que los analistas gestionen la información de manera ágil, minimizando tiempos de respuesta y reduciendo errores humanos en

la manipulación de registros de red. La interacción con el usuario es esencial para que el sistema actúe como un asistente de monitoreo que presente la información de manera clara y contextualizada, en lugar de generar alertas aisladas difíciles de correlacionar. La generación de reportes garantiza que los resultados del análisis de evidencia digital se documenten de forma confiable los incidentes y vulnerabilidades dentro de una infraestructura.

La creciente expansión del ciberespacio ha provocado la migración de numerosas actividades económicas, sociales, políticas y culturales al ámbito digital. En consecuencia, existe un flujo significativo de datos provenientes de gobiernos, centros públicos, organizaciones, empresas e incluso particulares, pero con ello los ciberataques también han tomado fuerza (Bidollahkhany & Basiri, 2025), por lo que es necesario desarrollar herramientas que permitan una adecuada gestión de la información, que permita salvaguardar y preservar los datos. Actualmente existen numerosas herramientas de ciberdefensa, pero los usuarios se enfrentan al dilema de aprender y recordar su complejo uso, esto puede provocar retrasos en la detección de amenazas en los sistemas informáticos (Wise & Hoque, 2025). En respuesta a este inconveniente, la herramienta de integración de datos en la arquitectura de la honeynet desarrolló una interfaz amigable e intuitiva para el usuario, cuyo funcionamiento es fácil de aprender, garantizado así la adecuada gestión de los ciberataques.

Los expertos evaluadores, tal como se expone en la Tabla 22 y en la Figura 27, coinciden en que la interfaz de usuario de la honeynet con la herramienta de integración de datos facilita la interacción con el usuario y le permite operar el sistema con facilidad, además de satisfacer la interacción usuario - máquina. Además, identificaron que ambas arquitecturas, con y sin la herramienta de integración de datos, generan reportes de incidencia en tiempo adecuado, aunque la arquitectura sin integración lo hace con menor eficacia. En lo que respecta a la facilidad de operación, los expertos resaltaron que la interfaz de la arquitectura con integración de datos optimiza la interacción usuario-sistema, fortaleciendo la operatividad de la honeynet; en este sentido, Albarrak (2024), señala que las interfaces adaptativas mejoran significativamente la experiencia del usuario, lo cual refuerza la efectividad observada en la propuesta técnica.

4.4. Prueba de hipótesis

Como elemento esencial para esta investigación, se planteó una serie de Hipótesis Específicas (HE), que derivan de la hipótesis general y se asocian con cada objetivo del estudio, así que, a partir de los resultados obtenidos con base en la experiencia de los expertos, quienes comprobaron el funcionamiento del módulo de integración de data y compararon las funcionalidades entre ambas arquitecturas, se procedió a la verificación estadística de las hipótesis específicas para ello se analizaron los datos según su naturaleza, se ejecutaron pruebas de normalidad y homocedasticidad, con la finalidad de identificar si es necesario realizar pruebas de hipótesis paramétricas o no paramétricas según los resultados obtenidos.

Es importante mencionar que el procesamiento estadístico, se realizó usando el software estadístico R-Studio.

Para la validación de las hipótesis propuestas se desarrolló un cuestionario que tuvo como objetivo validar la efectividad de una arquitectura de honeynet bajo dos condiciones: con integración de datos y sin integración de datos, este instrumento está compuesto por 23 preguntas cerradas que permiten evaluar aspectos clave del funcionamiento, desempeño y resultados observados en ambas configuraciones.

Cada pregunta fue formulada con base en criterios técnicos y operativos relevantes para el análisis, permitiendo al evaluador responder bajo tres opciones: "Sí", "No", y "No evidencia", según los hallazgos obtenidos durante la ejecución de las pruebas.

Este instrumento fue aplicado a ambas versiones de la arquitectura y al mismo grupo de expertos con el fin de identificar diferencias cualitativas en su comportamiento, así como validar la hipótesis de que la integración de datos contribuye a una mejora en la eficacia del sistema.

Dado que los resultados de los expertos respondieron con solo con respuestas de “sí” y “no”, se efectuaron pruebas de hipótesis para datos cualitativos nominales dicotómicos, para ello se analizó la literatura científica, donde se encontró que la

prueba de hipótesis más adecuada para este tipo de datos es la prueba McNemar que se utiliza para contrastar hipótesis con datos dicotómicos pareados.

Esta prueba solo considera los elementos fuera de la diagonal y da como resultado una prueba χ^2 de 1 grado de libertad.

La dependencia exclusiva de los elementos discordantes implica que la estimación del tamaño de la muestra debe basarse en supuestos sobre la proporción de observaciones discordantes. Se puede demostrar que esta prueba es equivalente a una prueba t pareada bajo el supuesto de medias iguales (Wu, 2022).

A continuación, los resultados de la verificación de las hipótesis:

4.4.1 Hipótesis Específica HE_1

La integración de los datos en la arquitectura honeynet resuelve problemas de interoperabilidad.

Tabla 23. Resultados para la resolución de problemas de interoperabilidad

Indicador	Honeynet sin integración de datos ANTES		Honeynet con integración de datos DESPUÉS	
	SI	NO	SI	NO
Permite el acceso a los datos desde diferentes fuentes de una sola vista.	0	4	4	0
Permite el acceso a los datos de diferentes fuentes desde un mismo dispositivo	0	4	4	0
Total	0	8	8	0

Fuente. El autor.

A continuación, se muestran los resultados en la Tabla 24:

Tabla 24. Tabla de contingencia HE₁

Antes/Después HE₁	Con integración de datos Después (SI)	Con integración de datos Después (NO)
Sin integración de datos Antes (SI)	0	0
Sin integración de datos Antes (NO)	8	0

Fuente. El autor.

Se realiza la prueba de hipótesis de McNemar, para ello se plantea la siguiente hipótesis:

H₀: No hay cambios significativos en las proporciones sin integración y con integración de datos en la resolución de problemas de interoperabilidad.

H₁: Hay cambios significativos en las proporciones sin integración y con integración de datos en la resolución de problemas de interoperabilidad.

Se realiza la prueba mediante el uso del software R-Studio, cuyos resultados se muestran a continuación:

Tabla 25. Estadístico de prueba - Resolución de problemas de interoperabilidad

McNemar's Chi-squared test	
χ^2	4,00
p-valor	0,0455

Fuente. El autor.

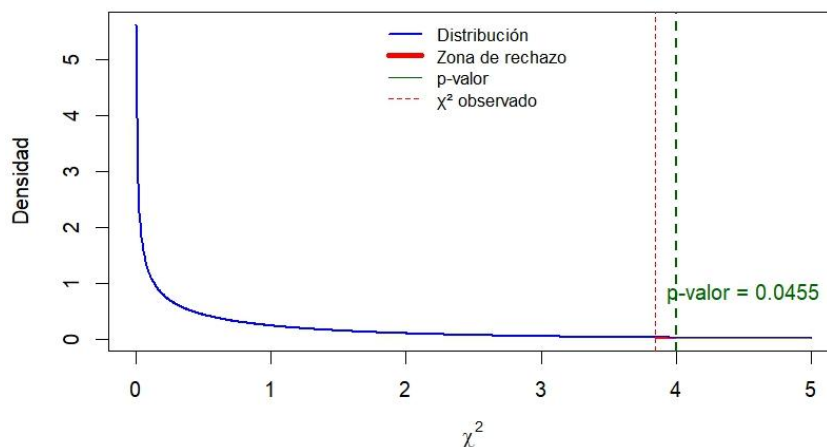


Figura 28. Regla de decisión HE₁

Fuente. El autor.

La prueba se realizó con un nivel de significancia de 0.05. Al comparar este valor con el p-valor obtenido en la prueba de McNemar (0.045) como se muestra en la Tabla 25, se procede a rechazar la hipótesis nula (H_0), ya que el p-valor es menor al nivel de significancia establecido, además lo confirma la Figura 28 donde se puede observar claramente que el estadístico de prueba está en la región de rechazo. Esto indica que existe suficiente evidencia estadística para afirmar que hay cambios significativos en las proporciones observadas antes y después de la integración de datos.

En este sentido, los resultados muestran una diferencia significativa a favor del uso de la honeynet con la herramienta de integración de datos, lo cual confirma la hipótesis HE_1 , que plantea que la integración de datos soluciona problemas de interoperabilidad presentes en la arquitectura de la honeynet, lo que demuestra que esta propuesta se convierta en un centro unificado de análisis de evidencia digital, mejorando la eficiencia de la gestión de la seguridad en entornos digitales.

4.4.2 Hipótesis Específica HE_2

La consolidación de la ontología de la honeynet con integración de los datos mejora la captura de evidencia digital.

Tabla 26. Resultados para la captura de evidencia digital

Indicador	Honeynet sin integración de datos ANTES		Honeynet con integración de datos DESPUÉS	
	SI	NO	SI	NO
Recopilación de atributos	0	8	8	0
Comparación de atributos	0	4	4	0
Integración de data	0	12	12	0
Total	0	24	24	0

Fuente. El autor.

A continuación, se muestran los resultados en la Tabla 27:

Tabla 27. Tabla de contingencia HE₂

Antes/Después HE ₂	Con integración de datos Después (SI)	Con integración de datos Después (NO)
Sin integración de datos Antes (SI)	0	0
Sin integración de datos Antes (NO)	24	0

Fuente. El autor.

Se realiza la prueba de McNemar, para ello se plantea la siguiente hipótesis:

H₀: No hay cambios significativos en las proporciones sin integración y con integración de datos para la mejora en la captura de evidencia digital.

H₁: Hay cambios significativos en las proporciones sin integración y con integración de datos para la mejora en la captura de evidencia digital.

Se realiza la prueba mediante el uso del software R-Studio, cuyos resultados se muestran a continuación:

Tabla 28. Estadístico de prueba - Captura de evidencia digital

McNemar's Chi-squared test	
χ^2	22,042
p-valor	2,668e-06

Fuente. El autor.

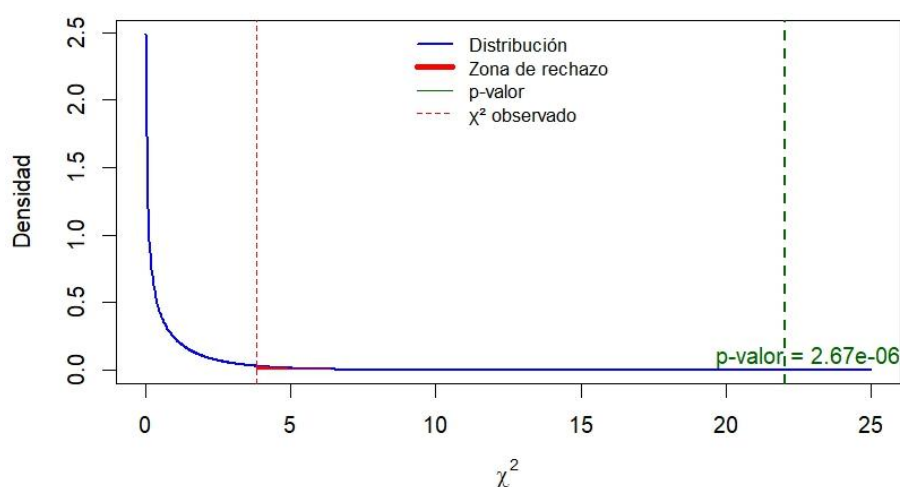


Figura 29. Regla de decisión HE₂

Fuente. El autor.

La prueba se realizó con un nivel de significancia de 0.05. Al comparar este valor con el p-valor obtenido en la prueba de McNemar (2.668e-06) como se identifica en la Tabla 28, se toma la decisión de rechazar la hipótesis nula (H_0), lo que además se comprueba a través de la Figura 29 donde se observa que el χ^2 calculado es mayor que el valor crítico, por lo que el p-valor recae sobre la zona de rechazo. Esto indica que existe evidencia estadísticamente suficiente para afirmar que hay diferencias significativas en las proporciones observadas antes y después de la integración de datos, en la captura de evidencia digital.

El resultado de la prueba de hipótesis validó que, la arquitectura de la honeynet con integración de datos contribuye a la recopilación, intercambio y comparación de atributos, lo que fortalece la identificación de los ataques, este análisis estadístico respalda la capacidad de esta propuesta para distinguir de manera segura el tráfico normal del malicioso, con un enfoque que promueve decisiones informadas, lo que confirma la hipótesis estadística alternativa HE_2 .

4.4.3 Hipótesis Específica HE_3

El uso de una arquitectura de honeynet con integración de datos mejora la consistencia del proceso de análisis de evidencia digital.

Tabla 29. Resultados de la consistencia del proceso de análisis de evidencia digital

Indicador	Honeynet sin integración de datos ANTES		Honeynet con integración de datos DESPUÉS	
	SI	NO	SI	NO
Exactitud del análisis	1	3	4	0
	4	0	4	0
Confianza	3	1	4	0
	4	0	3	1
Total	12	4	15	1

Fuente. El autor.

A continuación, se muestran los resultados en la Tabla 30:

Tabla 30. Tabla de contingencia HE₃

Antes/Después HE₃	Con integración de datos Después (SI)	Con integración de datos Después (NO)
Sin integración de datos Antes (SI)	11	1
Sin integración de datos Antes (NO)	4	0

Fuente. El autor.

Se realiza la prueba de hipótesis de McNemar, para ello se plantea la siguiente hipótesis:

H₀: No hay cambios significativos en las proporciones sin integración y con integración de datos para la consistencia del proceso de análisis de evidencia digital.

H₁: Hay cambios significativos en las proporciones sin integración y con integración de datos para la consistencia del proceso análisis de evidencia digital.

Se realiza la prueba mediante el uso del software R-Studio, cuyos resultados se muestran a continuación:

Tabla 31. Estadístico de prueba - Consistencia del proceso de análisis de evidencia digital

McNemar's Chi-squared test	
χ^2	5,4
p-valor	0,0201

Fuente. El autor.

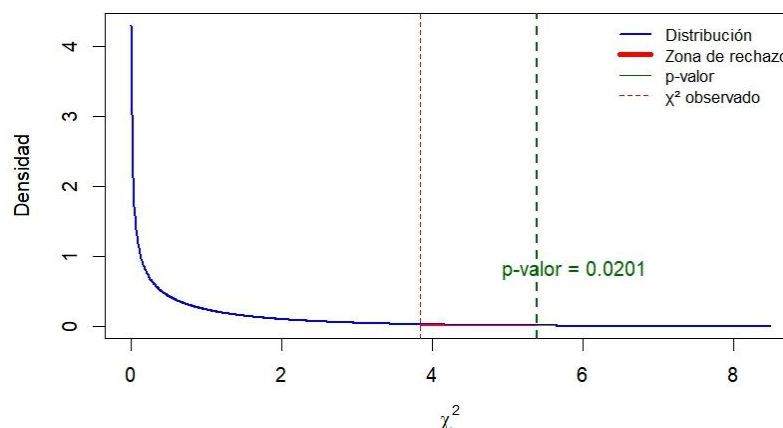


Figura 30. Regla de decisión HE₃

Fuente. El autor.

La prueba se realizó con un nivel de significancia de 0,05, cuyos valores se muestran en la Tabla 31. Al comparar este valor con el p-valor obtenido en la prueba de McNemar (0,0201), se toma la decisión de rechazar la hipótesis nula. Esta conclusión también se respalda visualmente en la Figura 30, donde el p-valor se encuentra claramente dentro de la zona de rechazo. En consecuencia, existe suficiente evidencia estadística para afirmar que hay cambios significativos en las proporciones antes y después de la integración de datos.

Los resultados reflejan una diferencia significativa, lo que permite confirmar la Hipótesis HE₃. Por tanto, se concluye que el uso de una arquitectura honeynet con integración de datos mejora la consistencia del proceso de análisis de evidencia digital, garantizando la exactitud y confianza durante el procedimiento de análisis en el campo de la ciberseguridad

4.4.4 Hipótesis Específica HE₄

El uso de una arquitectura de honeynet con integración de datos mejora los tiempos de detección de ataques en el análisis de evidencia digital.

A continuación, en la Tabla 32 se expresan los resultados de las comparaciones hechas por los expertos al medir los tiempos de detección de ataques en minutos de la honeynet con y sin la herramienta de integración de datos:

Tabla 32. Comparación de tiempos de detección de ataques

Incidente de ciberseguridad	Honeynet sin integración de datos	Honeynet con integración de datos	Diferencia de tiempo
Detección de ataques de suplantación	20 minutos	9 minutos	11 minutos
Detección de inyecciones de SQL en el servidor	17 minutos	7 minutos	10 minutos
Detección de fuerza bruta	15 minutos	6 minutos	9 minutos
Detección de denegación del servicio	16 minutos	7 minutos	9 minutos

Detección de envenenamiento DNS (Domain Name Server)	18 minutos	9 minutos	9 minutos
Detección de exploración de puertos	11 minutos	6 minutos	5 minutos
Detección de Ataques ICMP (Internet Control Message Protocol)	9 minutos	4 minutos	5 minutos
Detección de Cross Site Scripting	17 minutos	8 minutos	9 minutos
Media grupal	15,375 minutos	7 minutos	---

Fuente. El autor.

A continuación, en la Tabla 33, se muestra el resumen descriptivo de los tiempos en los que se detectan incidentes de ciberseguridad.

Tabla 33. Resumen descriptivo de los tiempos de detección de ataques

Variable	N	Media	Des.Est.	Rango	CoefVar(%)	Mínimo	Máximo
Sin integración de datos	8	15.38	3.66	11.00	23.81	9	20
Con integración de datos	8	7.00	1.69	5.00	24.14	4	9

Fuente. El autor.

La media en la condición sin integración de datos (15.38) es significativamente mayor que en la condición con integración de datos (7.00). Esto sugiere que la integración de datos reduce el valor de la variable que se está midiendo, en este caso el tiempo de detección de los ciberataques. En cuanto a la aplicación sin integración, ésta tiene mayor dispersión tanto en términos absolutos como en el rango total, a pesar de esto, el coeficiente de variación es bastante similar en ambos casos (~24%). La variabilidad en la condición sin integración es mucho más amplia, lo cual puede indicar falta de consistencia en los datos sin una estrategia integrada. La integración de datos tiene un efecto positivo, reduciendo la media de la variable observada y disminuyendo la variabilidad absoluta. Aunque la dispersión relativa se mantiene similar, los datos con integración son más concentrados y tienen menos extremos, como conclusión se

establece que los tiempos de detección de los diferentes tipos de ataques se reducen sensiblemente al utilizar la herramienta con integración de datos.

Prueba de Normalidad

Se comprueba si existe normalidad en los datos mediante la prueba de hipótesis de Shapiro Wilk.

H_0 : Los datos siguen una distribución normal.

H_1 : Los datos no siguen una distribución normal.

Para los tiempos sin integración de datos se obtienen resultados en la Tabla 34.

Tabla 34. Estadístico de prueba - Normalidad de tiempos de detección de ataques sin integración de datos

Shapiro-Wilk normality test	
W	0,91668
p-valor	0,4035

Fuente. El autor.

Como el p-valor (0,4035) es mayor que el nivel de significancia de 0.05 entonces no se rechaza la H_0 por lo tanto, los tiempos analizados siguen una distribución normal.

El mismo análisis se realiza para los tiempos con integración de datos, se obtienen los resultados en la Tabla 35.

Tabla 35. Estadístico de prueba - Normalidad de tiempos de detección de ataques con integración de datos

Shapiro-Wilk normality test	
W	0,93504
p-valor	0,5630

Fuente. El autor.

Como el p-valor (0.5630) es mayor que el nivel de significancia de 0.05 entonces no se rechaza la H_0 por lo tanto, los tiempos analizados siguen una distribución normal.

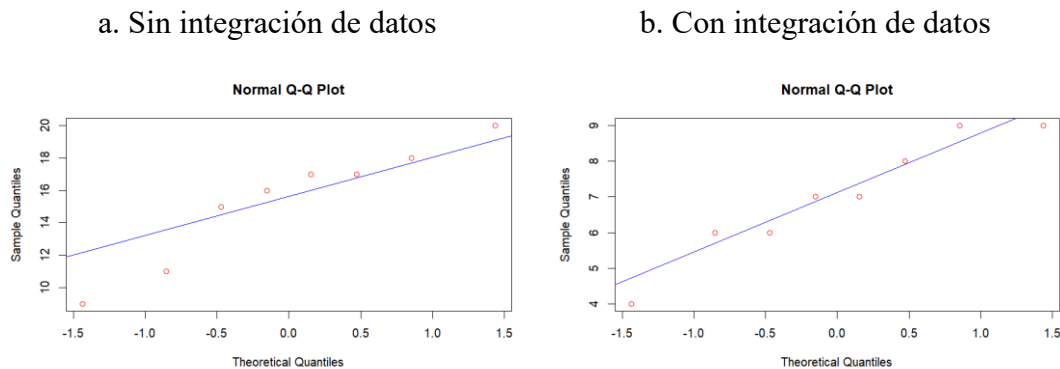


Figura 31. Prueba de Normalidad

Fuente. El autor.

Además, se pueden observar las gráficas de normalidad que comprueban el supuesto, en la Figura 31, a-b en las que se evalúa si el conjunto de datos sigue una distribución normal, donde la línea azul, representa la distribución teórica normal esperada y los puntos rojos representan los cuantiles de la distribución muestral. La mayoría de los puntos se alinean bastante bien con la línea azul, lo que indica que los datos tienen una distribución cercana a la normal, además hay ligeras desviaciones en los extremos, lo que sugiere la posible presencia de valores atípicos o una ligera asimetría.

Se puede observar el diagrama de caja y bigotes en la Figura 32 en la que se compara el tiempo entre dos métodos, mostrando que el método con integración reduce significativamente el tiempo en comparación con el método sin integración de datos. La mediana del tiempo es mucho menor en el método con integración (~6-7) frente al sin integración (~16-17), lo que indica una mejora notable en eficiencia. Además, la variabilidad es menor en el método con integración, ya que su rango intercuartílico es más estrecho y sus bigotes son más cortos, sugiriendo tiempos más consistentes. En cambio, sin integración de datos presenta una mayor dispersión en los tiempos, reflejada en su caja más amplia y bigotes más largos. Estos resultados sugieren que el uso de integración no solo optimiza el tiempo de detección de ataques, sino que también genera resultados más estables y predecibles.

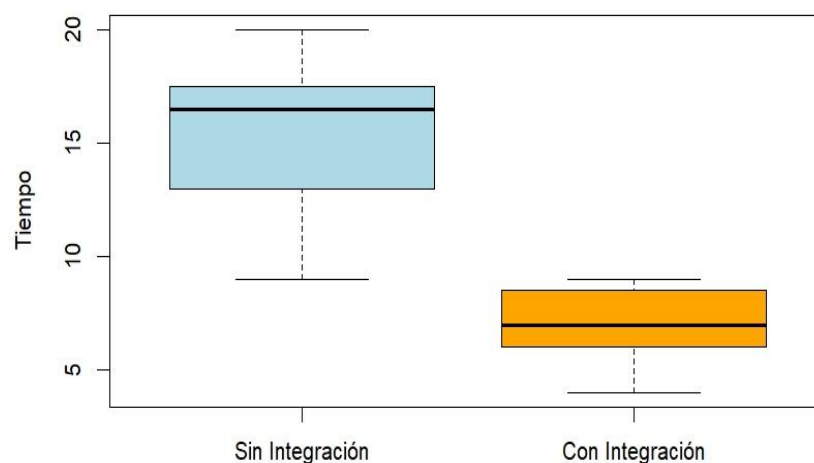


Figura 32. **Diagrama de caja y bigotes sin integración vs con integración**

Fuente. El autor

A partir de ello se realiza la prueba de homocedasticidad, con la finalidad de elegir la mejor prueba paramétrica.

Prueba de Homocedasticidad

Se comprueba si existe homocedasticidad en los datos mediante la prueba de hipótesis de comparación de varianzas, para ello se plantea la siguiente hipótesis:

H_0 : Las varianzas de los grupos son iguales (homocedasticidad).

H_1 : Las varianzas de los grupos son diferentes (heterocedasticidad).

Tabla 36. Estadístico de prueba - Varianzas iguales de tiempos de detección de ataques

F test to compare two variances	
F	4,6938
p-valor	0,05873

Fuente. El autor.

Según los resultados de la Tabla 36, el p-valor (0.05873) es mayor que el nivel de significancia de 0.05 entonces no se rechaza la H_0 por lo tanto, los tiempos analizados poseen varianzas iguales. Por lo tanto, se utiliza una prueba paramétrica t-student, con varianzas iguales para determinar la validez de la hipótesis HE_4 .

Para ello se plantea la hipótesis:

$$H_0: \mu_1 \geq \mu_2$$

$$H_1: \mu_1 < \mu_2$$

Donde μ_1 representa el promedio de tiempo sin integración de datos, mientras que μ_2 es el promedio de tiempo con integración de datos.

Tabla 37. Estadístico de prueba - Comparación de tiempos de detección de ataques sin y con integración de datos

Two sample t - test	
t	10,768
p-valor	1,000

Fuente. El autor.

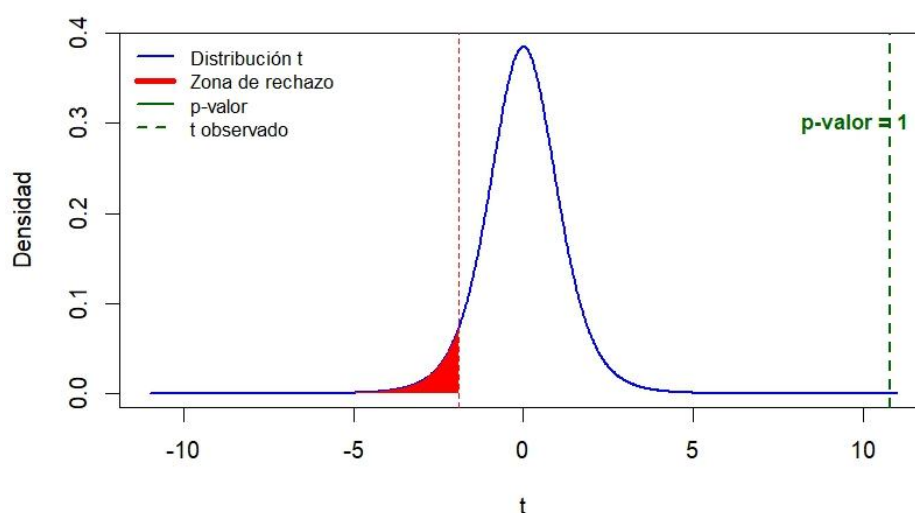


Figura 33. Regla de decisión HE4

Fuente. El autor

Los resultados de la prueba de hipótesis que se muestran en la Tabla 37, obtenidos mediante el software R-Studio, muestran que el p-valor unilateral (cola izquierda) es igual a 1,000, valor que es mayor que el nivel de significancia establecido (0,05). Por lo tanto, no se rechaza la hipótesis nula (H_0). Esto implica que el tiempo de detección de incidentes de ciberseguridad es significativamente mayor cuando se utiliza la arquitectura honeynet sin integración de datos, lo cual confirma la hipótesis específica analizada, además se comprueba a través de la Figura 33 donde se muestra la

distribución t, marcada con color rojo la zona de rechazo y se observa que el p-valor está fuera de dicha zona.

Estos resultados respaldan que el uso de una arquitectura honeynet con integración de datos mejora los tiempos de detección de ataques en el análisis de evidencia digital. Esto se debe a que, sin la integración, el análisis debe realizarse con cada herramienta de forma individual, mientras que, con la integración, el tiempo se reduce considerablemente al encontrarse la información centralizada en una única fuente.

Los resultados en esta dimensión demuestran que la integración de datos en la arquitectura de la honeynet permite optimizar significativamente los tiempos de respuesta frente a incidentes de seguridad. En el campo de la ciberseguridad esto implica que el tiempo entre la ocurrencia del ataque y su detección efectiva se reduce, favoreciendo la correlación inmediata de eventos en registros de IDS, flujos NetFlow, firewalls.

En conclusión, la evidencia confirma que la honeynet con integración de datos fortalece los procesos de respuesta temprana a incidentes, garantizando una defensa más eficiente y resiliente frente a ataques en infraestructuras de red críticas.

4.4.5 Hipótesis Específica HE₅

El uso de una arquitectura de honeynet con integración de datos mejora la integridad de los datos en el análisis de evidencia digital.

Tabla 38. Resultados para la integridad de los datos

Indicador	Honeynet sin integración de datos ANTES		Honeynet con integración de datos DESPUÉS	
	SI	NO	SI	NO
Almacenamiento de los datos	3	1	4	0
	4	0	4	0
Consistencia	2	2	4	0
Precisión técnica	2	2	1	3
	2	2	4	0
Total	13	7	17	3

Fuente. El autor.

A continuación, se muestran los resultados en la Tabla 39:

Tabla 39. Tabla de contingencia HE₅

Antes/Después HE₅	Con integración de datos Después (SI)	Con integración de datos Después (NO)
Sin integración de datos Antes (SI)	12	1
Sin integración de datos Antes (NO)	5	2

Fuente. El autor.

Se realiza la prueba de hipótesis de McNemar, para ello se plantea la siguiente prueba de hipótesis:

H₀: No hay cambios significativos en las proporciones sin integración y con integración de datos para la integridad de los datos en el análisis de evidencia digital.

H₁: Hay cambios significativos en las proporciones sin integración y con integración de datos para la integridad de los datos en el análisis de evidencia digital.

Se realiza la prueba mediante el uso del software R-Studio, cuyos resultados se muestran a continuación:

Tabla 40. Estadístico de prueba - Integridad de los datos

McNemar's Chi-squared test	
χ^2	7,14
p-valor	0,0075

Fuente. El autor.

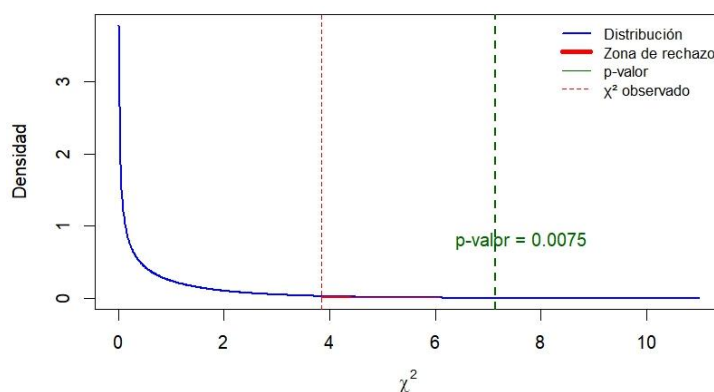


Figura 34. Regla de decisión HE₅

Fuente. El autor

Tal como se muestra en la Tabla 40, la prueba se realizó con un nivel de significancia de 0.05. Al comparar este valor con el p-valor obtenido en la prueba de McNemar (0,0075), se toma la decisión de rechazar la hipótesis nula (H_0), lo que además se comprueba a través de la Figura 34 donde se observa que el χ^2 calculado es mayor que el valor crítico, por lo que el p-valor recae sobre la zona de rechazo.

Esto indica que existe suficiente evidencia estadística para afirmar que hay diferencias significativas en las proporciones observadas antes y después de la integración de datos. Estos resultados permiten concluir que, efectivamente, el uso de una arquitectura de honeynet con integración de datos mejora la integridad de los datos en el análisis de evidencia digital, lo cual confirma la HE₅.

Por lo tanto, la capacidad de almacenamiento, consistencia y precisión técnica de la honeynet con herramienta de integración de datos mostró ventajas adicionales al momento de manejar la integridad de los datos en tránsito y durante el proceso de análisis. Esto es relevante porque no solo asegura que los datos almacenados sean consistentes, sino que también protege la veracidad y completitud de la información a lo largo de todo el ciclo de vida del dato: desde la captura inicial hasta la generación de reportes, como también lo señalan Okun and Barak (2004).

4.4.6 Hipótesis Específica HE₆

El uso de una arquitectura de honeynet con integración de datos optimiza la gestión de la información en el análisis de evidencia digital.

Tabla 41. Resultados para gestión de la información

Indicador	Honeynet sin integración de datos ANTES		Honeynet con integración de datos DESPUÉS	
	SI	NO	SI	NO
Usabilidad	0	4	4	0
	0	4	4	0
	0	4	4	0
Generación de reportes	2	2	4	0
	1	3	4	0
Total	3	17	20	0

Fuente. El autor.

A continuación, se muestran los resultados en la Tabla 42:

Tabla 42. Tabla de contingencia HE₆

Antes/Después HE₆	Con integración de datos Después (SI)	Con integración de datos Después (NO)
Sin integración de datos Antes (SI)	3	0
Sin integración de datos Antes (NO)	17	0

Fuente. El autor.

Se realiza la prueba de hipótesis de McNemar, para ello se plantea la siguiente prueba de hipótesis:

H₀: No hay cambios significativos en las proporciones sin integración y con integración de datos para la gestión de la información en el análisis de evidencia digital.

H₁: Hay cambios significativos en las proporciones sin integración y con integración de datos para la gestión de la información en el análisis de evidencia digital.

Se realiza la prueba mediante el uso del software R-Studio, cuyos resultados se muestran a continuación:

Tabla 43. Estadístico de prueba - Gestión de la información

McNemar's Chi-squared test	
χ^2	15,059
p-valor	0,0001042

Fuente. El autor.

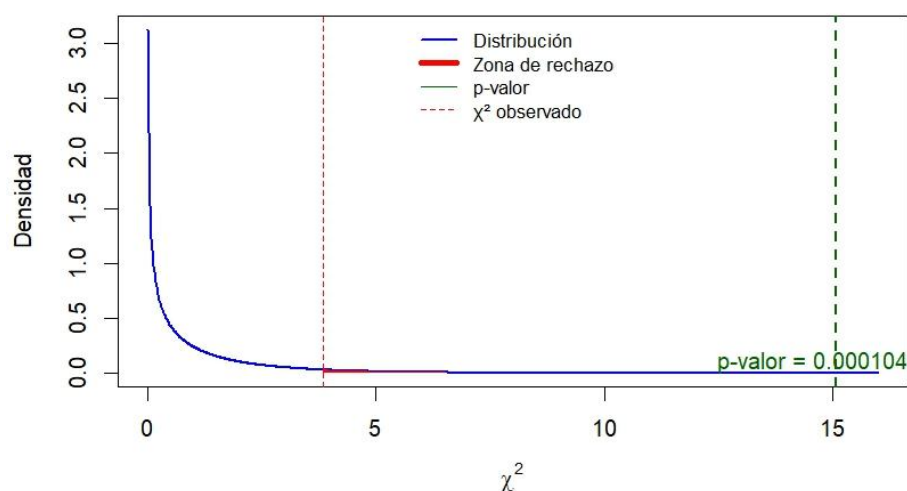


Figura 35. **Regla de decisión HE₆**

Fuente. El autor

La prueba se realizó con un nivel de significancia de 0.05. Al comparar este valor con el p-valor obtenido en la prueba de McNemar (0,0001042), mostrado en la Tabla 43, se toma la decisión de rechazar la hipótesis nula (H_0), lo que además se comprueba a través de la Figura 35 donde se observa que el χ^2 calculado es mayor que el valor crítico, por lo que el p-valor recae sobre la zona de rechazo. Esto indica que existe suficiente evidencia estadística para afirmar que hay diferencias significativas en las proporciones observadas antes y después de la integración de datos. En consecuencia, se concluye que el uso de una arquitectura de honeynet con integración de datos optimiza la gestión de la información, a través de una adecuada usabilidad y generación de reportes, en el análisis de evidencia digital, con estos datos se confirma que la Hipótesis HE₆ es correcta.

Es así, que la propuesta no solo captura evidencias digitales, sino que transforma esa información en conocimiento operativo accesible y confiable, fundamental para la defensa activa y la gestión integral de incidentes en redes informáticas, dentro del campo de la ciberseguridad.

4.4.7 Hipótesis General

Una vez verificadas cada hipótesis específica del estudio, se considera la Hipótesis General de la Investigación (H_0) y la Hipótesis Nula de la Investigación (H_1) para lo cual se realiza una nueva prueba de hipótesis utilizando todas las respuestas de las

encuestas aplicadas a los expertos antes y después de integrar los datos, se muestra a continuación en la Tabla 42 los resultados generales de ambos instrumentos.

Tabla 44. Resultados para la Hipótesis General

Indicador	Honeynet sin integración de datos ANTES		Honeynet con integración de datos DESPUÉS	
	SI	NO	SI	NO
Todos los indicadores				
Total	28	88	112	4

Fuente. El autor.

Para verificar si existen diferencias significativas en las respuestas de los expertos antes y después de aplicar la arquitectura honeynet con integración de datos, se realiza la prueba de McNemar sobre datos emparejados. La Tabla 43 muestra los datos resultantes antes de aplicar la prueba:

Tabla 45. Tabla de contingencia H_0

Antes/Después	Con integración de datos Después (SI)	Con integración de datos Después (NO)
Sin integración de datos Antes (SI)	28	0
Sin integración de datos Antes (NO)	88	4

Fuente. El autor

Se plantea la siguiente prueba de hipótesis:

H_0 : El uso de una arquitectura de Honeynet con integración de datos no mejora el análisis de evidencia digital producto de la heterogeneidad de los datos proporcionados por las herramientas utilizadas.

H_1 : El uso de una arquitectura de Honeynet con integración de datos mejora el análisis de evidencia digital producto de la heterogeneidad de los datos proporcionados por las herramientas utilizadas.

Se realiza la prueba mediante el uso del software R-Studio, cuyos resultados se muestran a continuación:

Tabla 46. Estadístico de prueba - Hipótesis General

McNemar's Chi-squared test	
χ^2	86,011
p-valor	1,79e-20

Fuente. El autor.

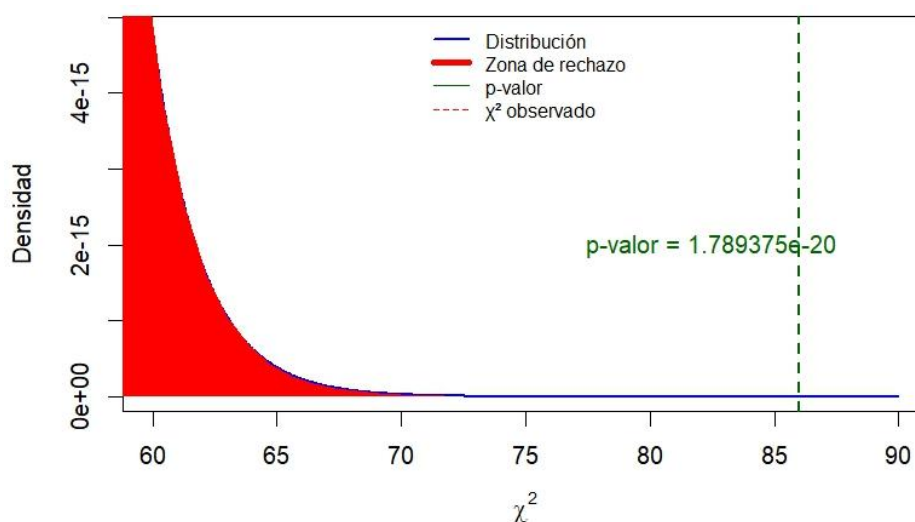


Figura 36. Regla de decisión H_0

Fuente. El autor

Dado que el p-valor es significativamente menor que 0.05, como se muestra en la Tabla 44, entonces se rechaza la hipótesis nula (H_0). Esto indica que sí existen diferencias estadísticamente significativas entre las respuestas antes y después de la integración de datos, esta conclusión se refuerza mediante la Figura 36 donde se observa claramente que el p-valor (1,79e-20) cae sobre la zona de rechazo en la distribución χ^2 . En consecuencia, se confirma que la arquitectura de Honeynet con integración de datos mejora sustancialmente la percepción de los expertos respecto a los indicadores evaluados, lo que respalda la hipótesis.

En función de los resultados estadísticos soportados por el análisis de los datos de cada experto; ello permite concluir que:

- a) La Hipótesis General de la Investigación (H_0): El uso de una arquitectura de honeynet con integración de datos mejora el análisis de evidencia digital producto de la heterogeneidad en los datos proporcionados por las herramientas utilizadas ha sido confirmada y es cierta.
- b) La Hipótesis Nula de la Investigación (H_1): El uso de una arquitectura de honeynet con integración de datos no mejora el análisis de evidencia digital producto de la heterogeneidad en los datos proporcionados por las herramientas utilizadas, se niega.

El análisis comparativo de las arquitecturas existentes realizado en el apartado 2.4 Comparación de arquitecturas honeynet, evidencia que, aunque arquitecturas como HoneyFactory, SIHA-SDN y Software Agents presentan fortalezas en adaptabilidad, escalabilidad, baja latencia y centralización de la data, sus resultados se limitan a la captura y almacenamiento de tráfico malicioso sin llegar a integrar ni correlacionar la información recolectada, estas arquitecturas dependen de procesos manuales para el análisis y gestión de incidentes, lo que incrementa la carga operativa y limita la capacidad de respuesta ante ataques complejos. En contraste, los resultados obtenidos con la arquitectura honeynet propuesta, que incorpora una herramienta de integración de datos, muestra una diferencia significativa en el desempeño global del sistema. La integración de datos soluciona los problemas de interoperabilidad detectados en los modelos previos y convierte a la honeynet en un centro unificado de análisis de evidencia digital, mejorando la eficiencia de la gestión de la seguridad en entornos digitales.

Las arquitecturas de honeynet sin integración de datos presentan resultados técnicamente confiables y generan evidencia válida para el estudio del comportamiento de los atacantes, ofreciendo una base sólida para la investigación en entornos controlados. Sin embargo, estas arquitecturas se ven limitadas en su capacidad de correlacionar y estructurar la información capturada, lo que restringe la profundidad del análisis y la velocidad de respuesta ante incidentes.

La arquitectura con integración de datos fortalece la capacidad de recopilación, intercambio y comparación de atributos, incrementando la efectividad en la identificación y clasificación de los ataques. A diferencia de las otras arquitecturas

analizadas, que solo registran eventos. La propuesta distingue de manera confiable el tráfico de red mediante correlación automatizada de fuentes, permitiendo un análisis más seguro y preciso.

De manera particular, los resultados comparativos muestran que la integración de datos reduce significativamente el tiempo de detección de los ciberataques, en vista que ninguna arquitectura realiza el proceso de integración de los datos, cada arquitectura realiza la captura de manera separada de acuerdo a las herramientas escogidas, además del proceso propio de la honeynet, la evidencia confirma que la arquitectura con integración de datos fortalece los procesos de respuesta temprana a incidentes, garantizando una defensa más eficiente, consistente y resiliente frente a ataques en infraestructuras críticas o entornos empresariales.

Las arquitecturas de honeynet garantizan que la información sea segura, íntegra y sin alteraciones, sin embargo, la propuesta permite no solo detectar ataques, sino que preserva su validez, así como su precisión y coherencia, lo que la convierte en una herramienta confiable para la toma de decisiones en la gestión de incidentes y procesos vinculados a delitos informáticos. Además, esta arquitectura facilita la operación del sistema ya que posee una adecuada usabilidad que reduce la complejidad de interpretación de logs a través la centralización de los datos sin necesidad que el usuario manipule herramientas independientes, lo que permite que los analistas gestionen la información, disminuyendo la posibilidad de errores humanos.

Luego del análisis de diversas arquitecturas honeynet presentadas en la literatura científica, se evidenció que ninguna de ellas es capaz de integrar datos. Así mismo, durante el desarrollo de la investigación se efectuó una comparación entre una arquitectura honeynet sin integración de datos y la propuesta con integración de datos, demostrando que esta última optimiza significativamente el análisis de evidencia digital generada, lo que permite un manejo más eficiente y coherente de la heterogeneidad de los datos, superando las limitaciones presentes en las arquitecturas tradicionales que no contemplan mecanismos de integración ni acceso unificado a las fuentes de información.

Finalmente, la arquitectura honeynet con integración de datos representa un avance significativo frente a los modelos tradicionales, al transformar la recolección aislada de información en un proceso centralizado, automatizado y confiable. Su capacidad para correlacionar múltiples fuentes, reducir tiempos de detección y facilitar la gestión operativa, optimiza tanto la eficiencia como la precisión del análisis de evidencia digital. Este enfoque no solo mejora la respuesta ante incidentes de seguridad, sino que también fortalece la toma de decisiones estratégicas en la gestión de ciberamenazas, consolidando a la propuesta como una solución integral, escalable y adaptable para entornos críticos de ciberseguridad.

CONCLUSIONES

La presente investigación demuestra que una arquitectura honeynet con integración de datos representa una solución efectiva para enfrentar los desafíos derivados de la heterogeneidad de la información en entornos de ciberseguridad, lo que mejora el análisis de la evidencia digital. Su implementación resuelve problemas de interoperabilidad entre herramientas de captura de tráfico de red, optimiza la consistencia en el proceso de análisis, reduce tiempos de detección de ataques y mejora la capacidad de respuesta ante incidentes. Además, la arquitectura propuesta no solo constituye un aporte técnico, sino que se configura como una herramienta práctica para analistas y administradores de red, facilitando la gestión centralizada de la información, la detección oportuna de ciberataques y la generación automática de reportes relevantes. Esta arquitectura, al integrar datos de distintas fuentes y aplicar criterios de análisis estructurado, fortalece la capacidad de tomar decisiones rápidas y precisas, manteniendo el entorno tecnológico bajo control y con mayor nivel de preparación ante amenazas. Así, se aporta una solución robusta, adaptable y escalable a las necesidades actuales de los sistemas de seguridad informática.

El diseño de la arquitectura física y lógica de la honeynet con integración de datos posibilita el acceso unificado a información proveniente de diversas herramientas y formatos empleados en la captura del tráfico de red. Esta integración soluciona los problemas de interoperabilidad mediante una estructura robusta, modular y funcional, gestionada de manera centralizada. Los resultados del análisis de expertos confirman su efectividad: el 100% de los evaluadores destacó mejoras significativas al comparar el uso de la aplicación con y sin integración de datos. Además, el resultado estadístico cuyo p-valor es 0.045 permite rechazar la hipótesis nula, evidenciando diferencias significativas en las proporciones de expertos que favorecen la arquitectura con integración. Estos hallazgos coinciden con la literatura científica, la cual respalda que

la integración de datos es un factor clave para resolver la interoperabilidad y garantizar un análisis confiable de evidencia digital, en contraste con otras tecnologías que carecen de esta capacidad.

La elaboración de la ontología honeynet estructura formalmente los conceptos, relaciones y atributos clave vinculados a la integración de datos en entornos de ciberseguridad, constituyéndose en la base fundamental para el diseño de la arquitectura propuesta. Esta ontología facilita la comprensión del funcionamiento interno de la honeynet y sus componentes, lo validaron los expertos que indicaron claramente que la integración de datos permite recopilar, intercambiar, comparar, usar y reusar la información afirmando con 100% todas las preguntas relacionadas con esta dimensión y cuya prueba de hipótesis tuvo un p-valor cercano a 0 lo que indica que hay suficiente evidencia estadística a favor del uso de esta arquitectura con integración de datos lo que potencia la eficiencia en la captura de evidencia digital, al proporcionar un modelo semántico que guíe la organización y clasificación coherente de los datos, como lo indica el análisis comparativo con otras tecnologías en ciberseguridad que no poseen estas ventajas.

La arquitectura honeynet con integración de datos mejora la consistencia del proceso de análisis de evidencia digital, los expertos coinciden en un 100% que esta arquitectura asegura que los datos recopilados sean precisos y exactos en los procesos de captura y análisis, y un 75% de los expertos indicaron que los resultados obtenidos son confiables, además la prueba de hipótesis confirma con un p-valor de 0.0201 a favor de la arquitectura con integración de datos pues se incrementa la confianza en los resultados y en la toma de decisiones, consolidando así un entorno de trabajo más seguro.

La implementación de la arquitectura honeynet con integración de datos evidencia una reducción significativa en los tiempos de detección de ataques relacionados con el análisis de evidencia digital, optimizando el flujo de información desde la captura hasta la interpretación de resultados de forma centralizada. En el análisis cualitativo, el 100% de los expertos coincidió en que la herramienta integrada permite detectar una mayor variedad de ataques en comparación con la versión sin integración de datos. De igual forma, el análisis cuantitativo reveló un tiempo promedio de detección de 7

minutos para la arquitectura con integración, frente a 15.375 minutos en la herramienta sin integración; lo que implica que, a gran escala, el tiempo de detección en esta última es 2.19 veces mayor. La prueba de hipótesis, con un nivel de confianza del 95%, confirmó que los tiempos de detección de la arquitectura con integración son significativamente menores. Este factor temporal resulta crítico en los entornos de ciberseguridad, donde la rapidez en la detección y análisis de incidentes es determinante para la toma de decisiones inmediatas y la ejecución de acciones correctivas que preserven la integridad y continuidad del entorno de red. En consecuencia, al minimizar los tiempos de respuesta, la arquitectura honeynet fortalece las capacidades de vigilancia activa, mitigación de riesgos y gestión eficiente de la seguridad informática.

La arquitectura honeynet con integración de datos fortalece la integridad de la información generada durante el análisis de evidencia digital, garantizando que los resultados se fundamenten en datos íntegros, correctamente almacenados y técnicamente precisos. En esta dimensión, el 100% de los expertos coincidió en que la herramienta con integración de datos asegura el almacenamiento adecuado, así como la integridad, consistencia y precisión técnica de la información. No obstante, solo un 25% de los evaluadores consideró que la herramienta facilita una interpretación unívoca de los datos, lo cual se atribuye a que esta etapa depende en gran medida de la experiencia y criterio del analista. A pesar de ello, la prueba de hipótesis arrojó un p-valor de 0.0075, evidenciando una diferencia estadísticamente significativa a favor del uso de la arquitectura honeynet con integración de datos. Este resultado confirma que el sistema preserva la integridad de los datos a lo largo de todo el proceso, desde la captura hasta el almacenamiento, evitando pérdidas, alteraciones o duplicaciones que puedan comprometer su confiabilidad y validez técnica.

La arquitectura honeynet con integración de datos optimiza la gestión de la información, esto lo confirman los expertos que coinciden en un 100% que al usar esta herramienta se facilitan la interacción efectiva entre el usuario y el sistema mediante una adecuada usabilidad a través de una interfaz amigable, intuitiva y funcional; con la generación de reportes claros y completos, que reduce la carga técnica, a diferencia de la herramienta sin integración que no posee esta ventaja, además esto se corrobora a través de la prueba de hipótesis cuyo p-valor es aproximadamente 0 lo que indica

que la herramienta con integración de datos permite una gestión eficiente de las actividades de monitoreo y análisis, que garantiza la toma de decisiones adecuadas.

RECOMENDACIONES

Se recomienda que las organizaciones y equipos de ciberseguridad implementen la arquitectura honeynet con integración de datos desarrollada en esta investigación para optimizar la interoperabilidad y mejorar la eficiencia en la captura y análisis de evidencia digital, adaptándola a sus necesidades específicas.

Es fundamental capacitar a los analistas y técnicos en el uso de la arquitectura y la ontología diseñadas, enfatizando en la interpretación correcta de los datos integrados y en la gestión eficiente de las herramientas para maximizar la confiabilidad y rapidez en el análisis.

Se recomienda integrar técnicas avanzadas de inteligencia artificial y analítica de datos en la arquitectura honeynet con integración de datos, para permitir una integración automatizada, flexible y escalable de información proveniente de cualquier herramienta de captura, sin importar su tipo, formato o cantidad. Esta integración independiente de la herramienta facilitará no solo la consolidación y correlación eficiente de datos heterogéneos, sino también el análisis en tiempo real para la detección precisa y oportuna de ataques cibernéticos, implementando respuestas rápidas que mitiguen o neutralicen los ataques sin intervención humana inmediata. Esta integración debe incluir análisis de datos orientado a la detección de patrones de ataques y amenazas en tiempo real, identificando y clasificando automáticamente las amenazas más comunes, al hacerlo se mejora significativamente la capacidad de anticipar, detectar y responder a incidentes de seguridad, optimizando la eficacia del sistema y reduciendo el impacto de los ataques.

REFERENCIAS BIBLIOGRÁFICAS

- Acosta, J. C., Basak, A., Kiekintveld, C., & Kamhoua, C. (2022). Lightweight On-Demand Honeypot Deployment for Cyber Deception. *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, 441 LNICST, 294–312. https://doi.org/10.1007/978-3-031-06365-7_18
- Ahsan, M. S., Islam, S., & Shatabda, S. (2025). A systematic review of metaheuristics-based and machine learning-driven intrusion detection systems in IoT. *Swarm and Evolutionary Computation*, 96, 101984. <https://doi.org/10.1016/j.swevo.2025.101984>
- Albarrak, A. M. (2024). Integration of Cybersecurity, Usability, and Human-Computer Interaction for Securing Energy Management Systems. *Sustainability (Switzerland)*, 16(18). <https://doi.org/10.3390/SU16188144>
- Alshaya, R., & Khediri, S. E. L. (2025). Optimizing cybercrime detection: A hybrid deep learning approach for enhanced intrusion detection systems. *Peer-to-Peer Networking and Applications*, 18(3), 145. <https://doi.org/10.1007/s12083-025-01933-w>
- Anand, A., Kallitsis, M., Sippe, J., & Dainotti, A. (2023). Aggressive Internet-Wide Scanners: Network Impact and Longitudinal Characterization. *CoNEXT Companion 2023 - Companion of the 19th International Conference on Emerging Networking EXperiments and Technologies*, 1–8. <https://doi.org/10.1145/3624354.3630583>
- Anqin, Z., & Yuwei, Q. (2025). Network Intrusion Detection Based on Time Heterogeneous Graph Representation. *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, 569 LNICST, 399–410. https://doi.org/10.1007/978-3-031-97352-9_24

- Anuradha, N., M., S., Marry, P., Sai, D. M., Ramesh, P., & Reddy, S. L. (2025). EFFICIENT SUPERVISED MACHINE LEARNING FOR CYBERSECURITY APPLICATIONS USING ADAPTIVE FEATURE SELECTION AND EXPLAINABLE AI SCENARIOS. *Journal of Theoretical and Applied Information Technology*, 103(6), 2458–2467.
- Aslan, Ö., Aktuğ, S. S., Ozkan-Okay, M., Yilmaz, A. A., & Akin, E. (2023). A Comprehensive Review of Cyber Security Vulnerabilities, Threats, Attacks, and Solutions. *Electronics* (Switzerland), 12(6).
<https://doi.org/10.3390/ELECTRONICS12061333>
- Bada, M., & Nurse, J. R. C. (2021). Profiling the Cybercriminal: A Systematic Review of Research. 2021 International Conference on Cyber Situational Awareness, Data Analytics and Assessment, CyberSA 2021.
<https://doi.org/10.1109/CYBERSA52016.2021.9478246>
- Balas, E., & Viecco, C. (2005). Towards a third generation data capture architecture for honeynets. *Proceedings from the 6th Annual IEEE System, Man and Cybernetics Information Assurance Workshop, SMC 2005*, 2005, 21–28.
<https://doi.org/10.1109/IAW.2005.1495929>
- Balogh, A., Ersok, M., Banati, A., & Erdodi, L. (2024). Concept for real time attacker profiling with honeypots, by skill based attacker maturity model. 2024 IEEE 22nd World Symposium on Applied Machine Intelligence and Informatics, SAMI 2024 - Proceedings, 175–180.
<https://doi.org/10.1109/SAMI60510.2024.10432876>
- Baral, S., Saha, S., & Haque, A. (2025). Autonomous Cyber Incident Response Using Reasoning and Action. 21st International Wireless Communications and Mobile Computing Conference, IWCMC 2025, 1392–1397.
<https://doi.org/10.1109/IWCMC65282.2025.11059476>

- Barangi, H., Rahimi, S. K., Zamani, B., & Moradi, H. (2025). HarmonyCAS: A model-driven framework for facilitating interoperability in context-aware systems. *Journal of Systems and Software*, 231. <https://doi.org/10.1016/j.jss.2025.112611>
- Bidollahkhany, F., & Basiri, M. (2025). Developing a National-Level Strategic Management Model for Cybersecurity. *Communications in Computer and Information Science*, 2444 CCIS, 67–87. https://doi.org/10.1007/978-3-032-01948-6_5
- Borst, W. N. (1997). Construction of Engineering Ontologies for Knowledge Sharing and Reuse. <https://research.utwente.nl/en/publications/construction-of-engineering-ontologies-for-knowledge-sharing-and->
- Chamkar, S. A., Maleh, Y., & Gherabi, N. (2022). THE HUMAN FACTOR CAPABILITIES IN SECURITY OPERATION CENTER (SOC). *EDPACS*, 66(1), 1–14. <https://doi.org/10.1080/07366981.2021.1977026>
- Chen, X., Jiang, Y., Lin, X., & Zhou, W. (2024). Overview on industrial Internet honeynet technology[工业互联网蜜网技术综述]. *Chinese Journal of Network and Information Security*, 10(5), 23–38. <https://doi.org/10.11959/j.issn2096-109x.2024064>
- Cimmino, A., Poveda-Villalón, M., & García-Castro, R. (2024). Ontologies and Semantic Interoperability. *Springer Handbooks*, Part F3575, 407–425. https://doi.org/10.1007/978-3-031-39650-2_17
- Corcho, O., Fernández-López, M., Gómez-Pérez, A., & López-Cima, A. (2005). Building legal ontologies with METHONTOLOGY and WebODE. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3369 LNAI, 142–157. https://doi.org/10.1007/978-3-540-32253-5_9

- De Pace, G., Lin, H., & Sun, Y. L. (2023). A Real Time Physics-Based Industrial Control System Honeynet Architecture for the Smart Grid. *Proceedings - 2023 8th International Conference on Information and Network Technologies, ICINT 2023*, 33–37. <https://doi.org/10.1109/ICINT58947.2023.00014>
- Dhanvijay, D. M., Dhanvijay, M. M., & Kamble, V. H. (2025). Cyber intrusion detection using ensemble of deep learning with prediction scoring based optimized feature sets for IOT networks. *Cyber Security and Applications*, 3. <https://doi.org/10.1016/J.CSA.2025.100088>
- Diana, L., Dini, P., & Paolini, D. (2025). Overview on Intrusion Detection Systems for Computers Networking Security. *Computers*, 14(3), 87. <https://doi.org/10.3390/computers14030087>
- Djenna, A., Barka, E., Benchikh, A., & Khadir, K. (2023). Unmasking Cybercrime with Artificial-Intelligence-Driven Cybersecurity Analytics. *Sensors*, 23(14), 6302. <https://doi.org/10.3390/s23146302>
- Fan, W., Du, Z., & Fernandez, D. (2015). Taxonomy of honeynet solutions. *IntelliSys 2015 - Proceedings of 2015 SAI Intelligent Systems Conference*, 1002–1009. <https://doi.org/10.1109/INTELLISYS.2015.7361266>
- Fernández-López, M., & Gómez-Pérez, A. (2002). Overview and analysis of methodologies for building ontologies. *Knowledge Engineering Review*, 17(2), 129–156. <https://doi.org/10.1017/S0269888902000462>
- Floridi, Luciano. (2019). *The logic of information: a theory of philosophy as conceptual design*. Oxford University Press.
- Francelin, M. M. (2018). Epistemology of Information Science: Evolution of research and its referencial bases[Epistemologia da Ciência da Informação: Evolução da pesquisa e suas bases referenciais]. *Perspectivas Em Ciencia Da Informacao*, 23(3), 89–103. <https://doi.org/10.1590/1981-5344/3135>

- Gallardo, I., & Bazán, P. (2022). Estrategia de Ciberseguridad Distribuida, aplicando el concepto de Operación de Inteligencia [Universidad Nacional de La Plata]. <https://doi.org/10.35537/10915/147421>
- Ganesarathinam, R., Prabakar, M. A., Singaravelu, M., & Fernandez, A. L. (2020). A Detailed Analysis of Intruders' Activities in the Network Through the Real-Time Virtual Honeynet Experimentation. *Advances in Intelligent Systems and Computing*, 1056, 39–53. https://doi.org/10.1007/978-981-15-0199-9_4
- Ghazi, D. S., Hamid, H. S., Zaiter, M. J., & Behadili, A. S. G. (2024). Performance and efficacy of Snort versus Suricata in intrusion detection: A benchmark analysis. *AIP Conference Proceedings*, 3232(1). <https://doi.org/10.1063/5.0236936>
- Godbole, A., Dhabliya, R., Deshpande, V., Sivakumar, S. A., Shankar, B. M., & Khetani, V. (2024). Ethical hacking and penetration testing strengthening cybersecurity posture through offensive security measures. *Journal of Discrete Mathematical Sciences and Cryptography*, 27(4), 1295–1305. <https://doi.org/10.47974/JDMSC-1983>
- Gönültaş, Y. C., & Yayli, H. (2025). A soft power instrument in public diplomacy: Digital diplomacy. In *Digital Transformation in Public Administration: Issues and Perspectives* (pp. 377–391). Nova Science Publishers, Inc.
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199–220. <https://doi.org/10.1006/KNAC.1993.1008>
- Husáková, M., & Bureš, V. (2020). Formal ontologies in information systems development: A systematic review. *Information (Switzerland)*, 11(2). <https://doi.org/10.3390/INFO11020066>

- Jacobs, C., & Van Der Merwe, A. (2025). Artificial Intelligence for Enhanced Master Data Quality Management in Enterprise Resource Planning Systems. *Communications in Computer and Information Science*, 2583 CCIS, 164–185. https://doi.org/10.1007/978-3-031-96262-2_12
- Jain, R., Johnson, M., Albizri, A., & Elias, G. M. (2022). An open system architecture framework for interoperability. *International Journal of Business Information Systems*, 41(4), 423–452. <https://doi.org/10.1504/IJBIS.2022.127573>
- Javadpour, A., Ja’fari, F., Taleb, T., & Benzaid, C. (2023). A Mathematical Model for Analyzing Honeynets and Their Cyber Deception Techniques. *Proceedings of the IEEE International Conference on Engineering of Complex Computer Systems*, ICECCS, 81–88. <https://doi.org/10.1109/ICECCS59891.2023.00019>
- Javadpour, A., Ja’fari, F., Taleb, T., Shojafar, M., & Benzaïd, C. (2024). A comprehensive survey on cyber deception techniques to improve honeypot performance. *Computers and Security*, 140. <https://doi.org/10.1016/J.COSE.2024.103792>
- Karthigha, M., Indira Priyadharshini, M., Rohini, M., Reema Nafeesa, J., ShreeAkshaya, B., & Yagavi, R. (2025). Honeypot-based IDS for Cyber Attack Detection. *AIP Conference Proceedings*, 3204(1), 040018. <https://doi.org/10.1063/5.0245939>
- Kashtalian, A., Savenko, O., & Sachenko, A. (2021). Agglomerative Clustering of Data Collected by Honeypots. *Proceedings of the 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, IDAACS 2021, 1, 504–509. <https://doi.org/10.1109/IDAACS53288.2021.9661027>
- Kaur, J., & Ramkumar, K. R. (2022). The recent trends in cyber security: A review. *Journal of King Saud University - Computer and Information Sciences*, 34(8), 5766–5781. <https://doi.org/10.1016/J.JKSUCI.2021.01.018>

- Kaur, P., & Kaur, H. (2025). Honeypots and honeynets: investigating attack vectors. In *Emerging Trends in Computer Science and Its Application* (pp. 192–197). CRC Press. <https://doi.org/10.1201/9781003606635-30>
- Kitchener, R. F. (2004). Bertrand Russell's flirtation with behaviorism. *Behavior and Philosophy*, 32(2), 273–291.
- Kocaogullar, Y., Cetin, O., Arief, B., Brierley, C., Pont, J., & Hernandez-Castro, J. (2025). Hunting High or Low: Evaluating the Effectiveness of High-Interaction and Low-Interaction Honeypots. *Lecture Notes in Computer Science*, 13855 LNCS, 14–30. https://doi.org/10.1007/978-3-031-83072-3_2
- Kohli, N., Kumar, S., & Sehgal, R. K. (2023). SDN-based Intelligent Honeynet Architecture for Averting Attacks. 2023 14th International Conference on Computing Communication and Networking Technologies, ICCCNT 2023. <https://doi.org/10.1109/ICCCNT56998.2023.10307595>
- Koumar, J., Hynek, K., & Cejka, T. (2023). Network Traffic Classification Based on Single Flow Time Series Analysis. 2023 19th International Conference on Network and Service Management, CNSM 2023. <https://doi.org/10.23919/CNSM59352.2023.10327876>
- Kouper, I., & Stone, S. (2024). Data Sharing and Use in Cybersecurity Research. *Data Science Journal*, 23(1). <https://doi.org/10.5334/DSJ-2024-003>
- Kulle, L. (2020). *Intrusion Attack & Anomaly Detection in IoT Using Honeypots*. Malmö University Spring.
- Kumar, T., Minakshi, & Attri, S. H. (2025). Exploiting machine learning for robust security. In *Exploiting Machine Learning for Robust Security* (pp. 1–23). IGI Global. <https://doi.org/10.4018/979-8-3693-7758-1.ch001>

- Lennartsson, M., Kävrestad, J., & Nohlberg, M. (2021). Exploring the meaning of usable security – a literature review. *Information and Computer Security*, 29(4), 647–663. <https://doi.org/10.1108/ICS-10-2020-0167>
- Lucchese, M., & Merro, M. (2024). Design, implementation and evaluation of a physics-aware honeynet for Industrial Control Systems. *Universita degli Studi di Verona*.
- Ma, L., Chen, M., & Liu, L. (2024). VDH: a dynamic honeynet technology based on game theory. 43. <https://doi.org/10.1117/12.3029002>
- Mamat, N. F. A., Noor, N. M. M., Mohamad, R., Mat, N. A. C., Sardorbek, D., Marxabo, F., & Vidanage, K. (2024). DEVELOPMENT OF ONTOLOGY FOR RUBRIC ASSESSMENT USING METHONTOLOGY. *Journal of Theoretical and Applied Information Technology*, 102(24), 8838–8850.
- Markert, J. (2020). Honeypot for Wireless Sensor Networks. In *School of Engineering, Computing and Mathematics Theses*.
- Martínez, C. J., & Moreno, H. O. (2024). Learning models to detect personality traits of cyber attackers: a combined approach using honeypot and surveys. <https://doi.org/10.1007/978-3>
- Masmoudi, M., Ben Abdallah Ben Lamine, S., Karray, M. H., Archimede, B., & Baazaoui Zghal, H. (2024). Semantic Data Integration and Querying: A Survey and Challenges. *ACM Computing Surveys*, 56(8). <https://doi.org/10.1145/3653317>
- McCarthy, J. (1980). Circumscription-A form of non-monotonic reasoning. *Artificial Intelligence*, 13(1–2), 27–39. [https://doi.org/10.1016/0004-3702\(80\)90011-9](https://doi.org/10.1016/0004-3702(80)90011-9)
- McCarty, B. (2003). The honeynet arms race. *IEEE Security and Privacy*, 1(6), 79–82. <https://doi.org/10.1109/MSECP.2003.1253575>

- Mesbah, M., Elsayed, M. S., Jurcut, A. D., & Azer, M. (2023). Analysis of ICS and SCADA Systems Attacks Using Honeypots. *Future Internet*, 15(7), 241. <https://doi.org/10.3390/fi15070241>
- Minjiao, Z., Yufeng, M., Bo, W., & Zhang, Q. (2022). A Dynamic Deceptive Honeynet System with A Hybrid of Virtual and Real Devices. 2022 5th International Conference on Computing and Big Data, ICCBD 2022, 113–117. <https://doi.org/10.1109/ICCBD56965.2022.10080304>
- Mohamed, K. F., Abdelbaki, N., & Shosha, A. (2023). Clipboard Data Attacks and Detection via Remote Desktop Protocol. 5th Novel Intelligent and Leading Emerging Sciences Conference, NILES 2023 - Proceedings, 98–102. <https://doi.org/10.1109/NILES59815.2023.10296672>
- Möller, D. P. F. (2023). Cybersecurity in Digital Transformation. *Advances in Information Security*, 103, 1–70. https://doi.org/10.1007/978-3-031-26845-8_1
- Mundt, M., & Baier, H. (2023). Threat-Based Simulation of Data Exfiltration Toward Mitigating Multiple Ransomware Extortions. *Digital Threats: Research and Practice*, 4(4), 54. <https://doi.org/10.1145/3568993>
- Murphy, K. R. (2022). Data, methods and theory in the organizational sciences: A new synthesis. In *Data, Methods and Theory in the Organizational Sciences: A New Synthesis*. Taylor and Francis. <https://doi.org/10.4324/9781003015000>
- Murti, R. K., Joshi, V. V., Wagh, R., Jangale, M. S., Chaudhari, A. C., & Wagh, M. (2024). Stochastic Models for Cyber Attack Detection and Response: A Mathematical Approach to Intrusion Detection Systems. *Communications on Applied Nonlinear Analysis*, 31(3S), 487–502. <https://doi.org/10.52783/cana.v31.803>
- Nayeem, A. A. R., Elshambakey, M., Dobbs, T., Lee, H., Crichton, D., Zhu, Y., Chokwitthaya, C., Tolone, W. J., & Cho, I. (2021). A Visual Analytics

- Framework for Distributed Data Analysis Systems. Proceedings - 2021 IEEE International Conference on Big Data, Big Data 2021, 229–240. <https://doi.org/10.1109/BigData52589.2021.9671768>
- Neches, R., Fikes, R. E., Finin, T., Gruber, T., Patil, R., Senator, T., & Swartout, W. R. (1991). Enabling Technology for Knowledge Sharing. *AI Magazine*, 12(3), 36–36. <https://doi.org/10.1609/AIMAG.V12I3.902>
- Nimanth, R., & Saxena, D. (2025). A Self-adaptive Multi-modal Cyber Threat Intelligence Framework for Securing Industrial Communication Networks. *Lecture Notes in Computer Science*, 15707 LNAI, 103–115. https://doi.org/10.1007/978-981-96-8892-0_9
- Noaman, A., Abdel-Hamid, A., & Eskaf, K. (2019). A novel honeynet architecture using software agents. 2019 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies, 3ICT 2019. <https://doi.org/10.1109/3ICT.2019.8910299>
- Okun, M., & Barak, A. (2004). Atomic Writes for data integrity and consistency in shared storage devices for clusters. *Future Generation Computer Systems*, 20(4), 539–547. [https://doi.org/10.1016/S0167-739X\(03\)00172-9](https://doi.org/10.1016/S0167-739X(03)00172-9)
- Park, J., Sung, K., & Moon, S. (2008). Developing graduation screen ontology based on the METHONTOLOGY approach. Proceedings - 4th International Conference on Networked Computing and Advanced Information Management, NCM 2008, 2, 375–380. <https://doi.org/10.1109/NCM.2008.215>
- Patel, N. (2024). AI-Powered Intrusion Detection and Prevention Systems in 5G Networks. Proceedings of the 9th International Conference on Communication and Electronics Systems, ICCES 2024, 834–841. <https://doi.org/10.1109/ICCES63552.2024.10859892>

- Paz, M. (2003). INVESTIGACIÓN CUALITATIVA EN EDUCACIÓN FUNDAMENTOS Y TRADICIONES. In N° (Vol. 18). McGraw Hill.
- Perkins, R. C., & Howell, C. J. (2021). Honeypots for Cybercrime Research. *Researching Cybercrimes: Methodologies, Ethics, and Critical Approaches*, 233–261. https://doi.org/10.1007/978-3-030-74837-1_12
- Peterson, J. M., Leevy, J. L., & Khoshgoftaar, T. M. (2021). A Review and Analysis of the Bot-IoT Dataset. *Proceedings - 15th IEEE International Conference on Service-Oriented System Engineering, SOSE 2021*, 20–27. <https://doi.org/10.1109/SOSE52839.2021.00007>
- Pieterse, H., Barbour, G., McDonald, A., Badenhorst, D., & Gertenbach, W. (2024). Utilisation of a Virtual Honeynet to Proactively Secure the South African National Research and Education Network Against Cyberattacks. *Communications in Computer and Information Science*, 2159 CCIS, 404–420. https://doi.org/10.1007/978-3-031-64881-6_24
- Qin, X., Jiang, F., Cen, M., & Doss, R. (2023). Hybrid cyber defense strategies using Honey-X: A survey. *Computer Networks*, 230. <https://doi.org/10.1016/J.COMNET.2023.109776>
- Rabhi, A., & Fissoune, R. (2024). A Parallel Processing Architecture for Querying Distributed and Heterogeneous Data Sources. *Communications in Computer and Information Science*, 1728 CCIS, 183–194. https://doi.org/10.1007/978-3-031-47366-1_14
- Redekar, R., & Bharati, R. (2025). DDoS Attacks: Detection Techniques, Challenges, and Modern Practices. *2025 IEEE International Students' Conference on Electrical, Electronics and Computer Science, SCEECS 2025*. <https://doi.org/10.1109/SCEECS64059.2025.10940527>

- Rogers, M. K. (2013). Analysis of Digital Evidence. In *Encyclopedia of Forensic Sciences: Second Edition* (pp. 455–460). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-382165-2.00325-1>
- Rohan, R., Pal, D., Hautamäki, J., Funilkul, S., Chutimaskul, W., & Thapliyal, H. (2023). A systematic literature review of cybersecurity scales assessing information security awareness. *Heliyon*, 9(3), e14234. <https://doi.org/10.1016/J.HELİYON.2023.E14234>
- Sabillon, R. (2020). Cyber security auditing, assurance, and awareness through CSAM and CATRAM. In *Cyber Security Auditing, Assurance, and Awareness Through CSAM and CATRAM*. IGI Global. <https://doi.org/10.4018/978-1-7998-4162-3>
- Sadeghi, M., Carenini, A., Corcho, O., Rossi, M., Santoro, R., & Vogelsang, A. (2023). Interoperability of Heterogeneous Systems of Systems: Review of Challenges, Emerging Requirements and Options. *Proceedings of the ACM Symposium on Applied Computing*, 741–750. <https://doi.org/10.1145/3555776.3577692>
- Salihu, M., Wu, J., & Yusuf, I. (2024). Performance estimation of honeynet system for network security enhancement via copula linguistic. *Sigma Journal of Engineering and Natural Sciences*, 42(4), 1169–1182. <https://doi.org/10.14744/sigma.2024.00095>
- Saravana, K. D., Biju, A., Muthalagu, R., Pawar, P. M., & Rajak, A. (2025). Design and analysis of a honeypot Pi for intrusion detection system. In *Hardware Security: Challenges and Solutions* (pp. 125–138). Springer Nature. https://doi.org/10.1007/978-3-031-81213-2_6
- Schildhauer, M. (2017). Data integration: Principles and practice. *Ecological Informatics: Data Management and Knowledge Discovery: Third Edition*, 129–157. https://doi.org/10.1007/978-3-319-59928-1_8

- Seo, S., & Kim, J. (2019). Heterogeneous data integration using confidence estimation of unseen visual data for zero-shot learning. *ACM International Conference Proceeding Series*, 171–174. <https://doi.org/10.1145/3305160.3305216>
- Silva, D. V., & Rafael, G. R. (2017). Ontologies for network security and future challenges. *Proceedings of the 12th International Conference on Cyber Warfare and Security, ICCWS 2017*, 541–547.
- Sinan, M., Shahin, M., & Gondal, I. (2025). Implementing and integrating security controls: A practitioners' perspective. *Computers and Security*, 156. <https://doi.org/10.1016/J.COSE.2025.104516>
- Singla, A., K, D., K., M., Kantak, G. K., Singh, A. P., & Rajesh, A. (2023). Improving Real-Time Data Analysis Accuracy with Machine Learning Tools. 2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), 1–6. <https://doi.org/10.1109/SMARTGENCON60755.2023.10442981>
- Spitzner, L. (2002). Honeypots: Tracking Hackers. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) : Vol. 15057 LNCS*. Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-981-97-8801-9_13
- Sreemathy, J., Infant Joseph, V., Nisha, S., Chauri Prabha, I., & Gokula Priya, R. M. (2020). Data Integration in ETL Using TALEND. 2020 6th International Conference on Advanced Computing and Communication Systems, ICACCS 2020, 1444–1448. <https://doi.org/10.1109/ICACCS48705.2020.9074186>
- Sreemathy, J., Naveen Durai, K., Lakshmi Priya, E., Deebika, R., Suganthi, K., & Aishwarya, P. T. (2021). Data Integration and ETL: A Theoretical Perspective. 2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS 2021, 1655–1660. <https://doi.org/10.1109/ICACCS51430.2021.9441997>

- Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge Engineering: Principles and methods. *Data and Knowledge Engineering*, 25(1–2), 161–197. [https://doi.org/10.1016/S0169-023X\(97\)00056-6](https://doi.org/10.1016/S0169-023X(97)00056-6)
- Suleiman, A. K., Kayed, A., Shamat, R. A., Jagani, V., Obaid, I., & Awad, A. (2022). A Passive OS-Fingerprinting framework using HoneyPot. 2022 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems, ICETIS 2022, 178–185. <https://doi.org/10.1109/ICETIS55481.2022.9888932>
- Taddeo, M., McCutcheon, T., & Floridi, L. (2021). Trusting Artificial Intelligence in Cybersecurity Is a Double-Edged Sword. In *Philosophical Studies Series* (Vol. 144, pp. 289–297). Springer Nature. https://doi.org/10.1007/978-3-030-81907-1_15
- Uschold, M., & King, M. (1995). Towards a methodology for building ontologies. Artificial Intelligence Applications Institute, University of Edinburgh.
- Vega, R., Jove, E., & Quintián, H. (2022). Análisis y detección de ataques informáticos mediante sistemas inteligentes de reducción dimensional [Universidad de Coruña]. <https://ruc.udc.es/dspace/handle/2183/30571>
- Venugopal, S., Arun, A., Ghildyal, A., Seshadri, P. S., & Singh, D. (2021). HoneyNet Deploying a Connected System of Diverse HoneyPots Using Open-Source Tools. In *Implementing Enterprise Cybersecurity with Open-source Software and Standard Architecture* (pp. 5–50). River Publishers.
- Vidhyasagar, B. S., & Kartheek, R. (2023). A Review on Effective Detection and Prevention for DDOS Attack in Cloud Virtual Environment Using HoneyPot. *Proceedings - 4th IEEE 2023 International Conference on Computing, Communication, and Intelligent Systems, ICCIS 2023*, 939–946. <https://doi.org/10.1109/ICCIS60361.2023.10425464>

- Wang, S., Peng, Y. W., Lan, H., Luo, Q. W., & Peng, Z. Y. (2020). Survey and Prospect: Data Integration Methodologies. *Ruan Jian Xue Bao/Journal of Software*, 31(3), 893–908. <https://doi.org/10.13328/J.CNKI.JOS.005911>
- Watson, D., & Riden, J. (2008). The honeynet project: Data collection tools, infrastructure, archives and analysis. *Proceedings - WOMBAT Workshop on Information Security Threats Data Collection and Sharing, WISTDCS 2008*, 24–30. <https://doi.org/10.1109/WISTDCS.2008.11>
- Williams, T. D. (2020). Epistemological Questions for Cybersecurity. *International Conference on Cyber Security and Protection of Digital Services, Cyber Security 2020*. <https://doi.org/10.1109/CYBERSECURITY49315.2020.9138884>
- Wise, J., & Hoque, M. T. (2025). Lyric-based passwords: Enhancing security and recall with AI. *Cyber Security and Applications*, 3. <https://doi.org/10.1016/j.csa.2025.100108>
- Wu, Y. (2022). Weighted McNemar's test for the comparison of two screening tests in the presence of verification bias. *Statistics in Medicine*, 41(16), 3149–3163. <https://doi.org/10.1002/sim.9409>
- Xie, M., Yu, Y., Chen, R., Zhao, Y., Ning, J., Yang, X., & Jiang, Z. L. (2025). Multiuser data integrity auditing atop blockchain with secure user revocation for cognitive IoT networks. *Computer Standards and Interfaces*, 95. <https://doi.org/10.1016/j.csi.2025.104042>
- Xu, H. (2023). Cybersecurity and Data Quality in Cloud Computing: A Research Framework. *Lecture Notes in Business Information Processing*, 464 LNBIP, 201–208. https://doi.org/10.1007/978-3-031-30694-5_15

- Yan, H., Wang, Y., Liu, G., & Zhao, J. (2025). Efficient edge-based data integrity auditing in cloud storage. *Future Generation Computer Systems*, 172. <https://doi.org/10.1016/J.FUTURE.2025.107899>
- Yang, Z., Li, H., Yang, Z., & Ye, L. (2023). Honeynet optimization strategy based on attack graph. 36. <https://doi.org/10.1117/12.2679157>
- Yu, T., Xin, Y., & Zhang, C. (2024). HoneyFactory: Container-Based Comprehensive Cyber Deception Honeynet Architecture. *Electronics (Switzerland)*, 13(2). <https://doi.org/10.3390/ELECTRONICS13020361>
- Zhang, Y., Qu, T., Hong, Z., Zhang, Z., & Huang, G. Q. (2025). Digital twin driven opti-state control approach for smart warehousing in the synchronous operating environment. *Robotics and Computer-Integrated Manufacturing*, 97. <https://doi.org/10.1016/j.rcim.2025.103099>
- Zhang, Y., Xia, X., Pan, Z., Dai, C., Chen, L., & Chen, Y. (2024). Honeynet Module Simulation and Dynamic Configuration of Honeypots Based on Online Learning. *Communications in Computer and Information Science*, 1994 CCIS, 115–130. https://doi.org/10.1007/978-981-97-1010-2_9

ANEXO A



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Dirección General de Estudios de Posgrado

Facultad de Ingeniería de Sistemas e Informática

Unidad de Posgrado

ACTA DE VALIDACIÓN

Quien suscribe, **Amílcar Antonio Arenas Arredondo**, con cédula de identidad: **V-15667074**, actuando como experto en el área de: **Tecnologías de la Información y la Comunicación**, hago constar que he leído el instrumento de recolección de datos diseñado para el estudio **“ARQUITECTURA DE HONEYNET CON INTEGRACIÓN DE DATOS PARA EL ANÁLISIS DE EVIDENCIA DIGITAL”** y, una vez validado el mismo, procedo a emitir el siguiente Juicio de Experto:

Valido sin modificaciones, puede ser aplicado: ()

Válido con modificaciones de forma, tal y como se señala en la tabla de validación: (**X**)

No es válido, por ello, no debe ser aplicado: ()

En la ciudad de Maracay, a los 30 días del mes de agosto del 2023.

Identificación del experto (Nombre, Apellido, Cédula, Firma):

Amílcar Antonio Arenas Arredondo

C.I. V-15667074





Universidad Nacional Mayor de San Marcos
Universidad del Perú. Decana de América

Dirección General de Estudios de Posgrado
Facultad de Ingeniería de Sistemas e Informática
Unidad de Posgrado

ACTA DE VALIDACIÓN

Quien suscribe, **Dr. José Rafael Rodríguez Requena**, con cédula de identidad: **V-8.672.283**, actuando como experto en el área de: Ingeniería en Informática, hago constar que he leído el instrumento de recolección de datos diseñado para el estudio **“ARQUITECTURA DE HONEYNET CON INTEGRACIÓN DE DATOS PARA EL ANÁLISIS DE EVIDENCIA DIGITAL”** y, una vez validado el mismo, procedo a emitir el siguiente juicio de experto:

Valido sin modificaciones, puede ser aplicado: (**X**)

Válido con modificaciones de forma, tal y como se señala en la tabla de validación: ()

No es válido, por ello, no debe ser aplicado: ()

En la ciudad de Maracay, a los 27 días del mes de agosto del 2023.

Identificación del experto (Nombre, Apellido, Cédula, Firma):

Dr. José Rafael Rodríguez Requena

C.I.: 8.672.283



Universidad Nacional Mayor de San Marcos
Universidad del Perú. Decana de América

Dirección General de Estudios de Posgrado
Facultad de Ingeniería de Sistemas e Informática
Unidad de Posgrado

ACTA DE VALIDACIÓN

Quien suscribe, **Gabriela Gardié Quintero**, con cédula de identidad: **V-9676689**, actuando como experto en el área de: **Tecnologías de la Información y la Comunicación**, hago constar que he leído los instrumentos de recolección de datos diseñado para el estudio **“ARQUITECTURA DE HONEYNET CON INTEGRACIÓN DE DATOS PARA EL ANÁLISIS DE EVIDENCIA DIGITAL”** y, una vez validados los mismos, procedo a emitir el siguiente Juicio de Experto:

Valido sin modificaciones, puede ser aplicado: ()

Válido con modificaciones de forma, tal y como se señala en la tabla de validación: (**X**)

No es válido, por ello, no debe ser aplicado: ()

En la ciudad de Maracay, a los 30 días del mes de agosto del 2023.

Identificación del experto (Nombre, Apellido, Cédula, Firma):

Gabriela Gardié Quintero

C.I. V-9676689

ANEXO B

VERSIÓN DEFINITIVA DE LOS INSTRUMENTOS DE RECOLECCIÓN DE DATOS PARA SER ENTREGADOS A LOS EXPERTOS

Arquitectura HoneyNet sin herramienta de integración de datos

Cuestionario #1

No	Interrogante	Resp Pos	Resp Neg	No Evid
1	¿La ontología de la HoneyNet, sin la herramienta de integración de datos, permite la recopilación de los atributos para la captura de data?			
2	¿La ontología de la HoneyNet, sin la herramienta de integración de datos, permite el intercambio de la información de los atributos para la captura de data?			
3	¿La ontología de la HoneyNet, sin la herramienta de integración de datos, permite la comparación de los atributos de la data?			
4	¿La ontología de la HoneyNet, sin la herramienta de integración de datos, permite la integración de la data?			
5	¿La ontología de la HoneyNet, sin la integración de datos, permite el uso de toda la información?			
6	¿La ontología de la HoneyNet, sin la integración de datos, permite el reuso de toda la información?			
7	¿La arquitectura de la HoneyNet, sin la integración de datos, permite el acceso a los datos desde diferentes fuentes de una sola vista?			

8	¿La arquitectura de la HoneyNet, sin la integración de datos, permite el acceso a los datos de diferentes fuentes desde un mismo dispositivo?			
9	¿La arquitectura de la HoneyNet, sin la integración de datos, mejora la exactitud del análisis?			
10	¿Los resultados del análisis con la arquitectura de la HoneyNet, sin la integración de datos, son exactos?			
11	¿La arquitectura de la HoneyNet, sin la integración de datos, genera confianza en el proceso de análisis?			
12	¿Es confiable el resultado del proceso de análisis con la arquitectura de la HoneyNet, sin la integración de datos?			
13	¿Con la arquitectura de la HoneyNet, sin la integración de datos, es aceptable el tiempo de detección de? Ataques de suplantación.			
	Inyecciones de SQL en el servidor.			
	Ataques de Fuerza bruta.			
	Ataques de Denegación del servicio.			
	Envenenamiento DNS.			
	Ataques de exploración de puertos.			
	Ataques del Internet Control Message Protocol (ICMP).			
	Ataque de Cross Site Scripting.			
14	¿La funcionalidad de la arquitectura la HoneyNet sin la herramienta de integración de datos, garantiza el almacenamiento de la información?			

15	¿La funcionalidad de la arquitectura la HoneyNet sin la herramienta de integración de datos, garantiza la integridad de los datos almacenados?			
16	¿Los datos manejados con la arquitectura la HoneyNet sin la herramienta de integración, son consistentes?			
17	¿La arquitectura de la HoneyNet, sin la herramienta de integración de datos, permite la interpretación univoca de la información?			
18	¿Existe precisión técnica en la data administrada con la arquitectura de la HoneyNet, sin la herramienta de integración de datos?			
19	¿La interfaz de usuario de la HoneyNet sin la herramienta de integración de datos facilita la interacción con el usuario?			
20	¿La HoneyNet, sin la herramienta de integración de datos, permite al usuario operar el sistema con facilidad?			
21	¿La interfaz de usuario de la HoneyNet, sin la herramienta de integración de datos, satisface la interacción con el usuario?			
22	¿La arquitectura de la HoneyNet, sin la herramienta de integración de datos, genera los reportes de incidentes suficientes?			
23	¿La arquitectura de la HoneyNet, sin la herramienta de integración de datos, genera los reportes de incidentes a tiempo?			

Arquitectura HoneyNet con herramienta de integración de datos

Cuestionario #2

No	Interrogante	Resp Pos	Resp Neg	No Evid
1	¿La ontología de la HoneyNet, con la herramienta de integración de datos, permite la recopilación de los atributos para la captura de data?			
2	¿La ontología de la HoneyNet, con la herramienta de integración de datos, permite el intercambio de la información de los atributos para la captura de data?			
3	¿La ontología de la HoneyNet, con la herramienta de integración de datos, permite la comparación de los atributos de la data?			
4	¿La ontología de la HoneyNet, con la herramienta de integración de datos, permite la integración de la data?			
5	¿La ontología de la HoneyNet, con la integración de datos, permite el uso de toda la información?			
6	¿La ontología de la HoneyNet, con la integración de datos, permite el reuso de toda la información?			
7	¿La arquitectura de la HoneyNet, con la integración de datos, permite el acceso a los datos desde diferentes fuentes de una sola vista?			
8	¿La arquitectura de la HoneyNet, con la integración de datos, permite el acceso a los datos de diferentes fuentes desde un mismo dispositivo?			

9	¿La arquitectura de la HoneyNet, con la integración de datos, mejora la exactitud del análisis?			
10	¿Los resultados del análisis con la arquitectura de la HoneyNet, con la integración de datos, son exactos?			
11	¿La arquitectura de la HoneyNet, con la integración de datos, genera confianza en el proceso de análisis?			
12	¿Es confiable el resultado del proceso de análisis con la arquitectura de la HoneyNet, con la integración de datos?			
13	¿Con la arquitectura de la HoneyNet, con la integración de datos, es aceptable el tiempo de detección de:...? Ataques de suplantación.			
	Inyecciones de SQL en el servidor.			
	Ataques de Fuerza bruta.			
	Ataques de Denegación del servicio.			
	Envenenamiento DNS.			
	Ataques de exploración de puertos.			
	Ataques del Internet Control Message Protocol (ICMP).			
	Ataque de Cross Site Scripting.			
14	¿La funcionalidad de la arquitectura la HoneyNet con la herramienta de integración de datos, garantiza el almacenamiento de la información?			
15	¿La funcionalidad de la arquitectura la HoneyNet con la herramienta de integración de datos, garantiza la integridad de los datos almacenados?			

16	¿Los datos manejados con la arquitectura la HoneyNet con la herramienta de integración, son consistentes?			
17	¿La arquitectura de la HoneyNet, con la herramienta de integración de datos, permite la interpretación univoca de la información?			
18	¿Existe precisión técnica en la data administrada con la arquitectura de la HoneyNet, con la herramienta de integración de datos?			
19	¿La interfaz de usuario de la HoneyNet con la herramienta de integración de datos facilita la interacción con el usuario?			
20	¿La HoneyNet, con la herramienta de integración de datos, permite al usuario operar el sistema con facilidad?			
21	¿La interfaz de usuario de la HoneyNet, con la herramienta de integración de datos, satisface la interacción con el usuario?			
22	¿La arquitectura de la HoneyNet, con la herramienta de integración de datos, genera los reportes de incidentes suficientes?			
23	¿La arquitectura de la HoneyNet, con la herramienta de integración de datos, genera los reportes de incidentes a tiempo?			

ANEXO C

MATRIZ DE DATOS

Arquitectura HoneyNet sin herramienta de integración de datos:

Cuestionario #1

N °	Interrogante	Resp Pos	Resp Neg	No Evid
1	¿La ontología de la HoneyNet, sin la herramienta de integración de datos, permite la recopilación de los atributos para la captura de data?		4 100%	
2	¿La ontología de la HoneyNet, sin la herramienta de integración de datos, permite el intercambio de la información de los atributos para la captura de data?		4 100%	
3	¿La ontología de la HoneyNet, sin la herramienta de integración de datos, permite la comparación de los atributos de la data?		4 100%	
4	¿La ontología de la HoneyNet, sin la herramienta de integración de datos, permite la integración de la data?		4 100%	
5	¿La ontología de la HoneyNet, sin la integración de datos, permite el uso de toda la información?		4 100%	
6	¿La ontología de la HoneyNet, sin la integración de datos, permite el reuso de toda la información?		4 100%	

7	¿La arquitectura de la HoneyNet, sin la integración de datos, permite el acceso a los datos desde diferentes fuentes de una sola vista?		4 100%	
8	¿La arquitectura de la HoneyNet, sin la integración de datos, permite el acceso a los datos de diferentes fuentes desde un mismo dispositivo?		4 100%	
9	¿La arquitectura de la HoneyNet, sin la integración de datos, mejora la exactitud del análisis?	1 25%	3 75%	
10	¿Los resultados del análisis con la arquitectura de la HoneyNet, sin la integración de datos, son exactos?	4 100%		
11	¿La arquitectura de la HoneyNet, sin la integración de datos, genera confianza en el proceso de análisis?	3 75%	1 25%	
12	¿Es confiable el resultado del proceso de análisis con la arquitectura de la HoneyNet, sin la integración de datos?	4 100%		
13	¿Con la arquitectura de la HoneyNet, sin la integración de datos, es aceptable el tiempo de detección de? Ataques de suplantación.		4 100%	
	Inyecciones de SQL en el servidor.		4 100%	
	Ataques de Fuerza bruta.		4	

			100%	
	Ataques de Denegación del servicio.		4 100%	
	Envenenamiento DNS.		4 100%	
	Ataques de exploración de puertos.		4 100%	
	Ataques del Internet Control Message Protocol (ICMP).		4 100%	
	Ataque de Cross Site Scripting.		4 100%	
14	¿La funcionalidad de la arquitectura la HoneyNet sin la herramienta de integración de datos, garantiza el almacenamiento de la información?	3 75%	1 25%	
15	¿La funcionalidad de la arquitectura la HoneyNet sin la herramienta de integración de datos, garantiza la integridad de los datos almacenados?	4 100%		
16	¿Los datos manejados con la arquitectura la HoneyNet sin la herramienta de integración, son consistentes?	2 50%	2 50%	
17	¿La arquitectura de la HoneyNet, sin la herramienta de integración de datos, permite la interpretación univoca de la información?	2 50%	2 50%	
18	¿Existe precisión técnica en la data administrada con la	2 50%	2 50%	

	arquitectura de la HoneyNet, sin la herramienta de integración de datos?			
19	¿La interfaz de usuario de la HoneyNet sin la herramienta de integración de datos facilita la interacción con el usuario?		4 100%	
20	¿La HoneyNet, sin la herramienta de integración de datos, permite al usuario operar el sistema con facilidad?		4 100%	
21	¿La interfaz de usuario de la HoneyNet, sin la herramienta de integración de datos, satisface la interacción con el usuario?		4 100%	
22	¿La arquitectura de la HoneyNet, sin la herramienta de integración de datos, genera los reportes de incidentes suficientes?	2 50%	2 50%	
23	¿La arquitectura de la HoneyNet, sin la herramienta de integración de datos, genera los reportes de incidentes a tiempo?	1 25%	3 75%	

Arquitectura HoneyNet con herramienta de integración de datos:

Cuestionario #2

No	Interrogante	Resp Pos	Resp Neg	No Evid
1	¿La ontología de la HoneyNet, con la herramienta de integración de datos, permite la recopilación de los atributos para la captura de data?	4 100%		
2	¿La ontología de la HoneyNet, con la herramienta de integración de datos, permite el intercambio de la información de los atributos para la captura de data?	4 100%		
3	¿La ontología de la HoneyNet, con la herramienta de integración de datos, permite la comparación de los atributos de la data?	4 100%		
4	¿La ontología de la HoneyNet, con la herramienta de integración de datos, permite la integración de la data?	4 100%		
5	¿La ontología de la HoneyNet, con la integración de datos, permite el uso de toda la información?	4 100%		
6	¿La ontología de la HoneyNet, con la integración de datos, permite el reuso de toda la información?	4 100%		
7	¿La arquitectura de la HoneyNet, con la integración de datos, permite el acceso a los datos	4 100%		

	desde diferentes fuentes de una sola vista?			
8	¿La arquitectura de la HoneyNet, con la integración de datos, permite el acceso a los datos de diferentes fuentes desde un mismo dispositivo?	4 100%		
9	¿La arquitectura de la HoneyNet, con la integración de datos, mejora la exactitud del análisis?	4 100%		
10	¿Los resultados del análisis con la arquitectura de la HoneyNet, con la integración de datos, son exactos?	4 100%		
11	¿La arquitectura de la HoneyNet, con la integración de datos, genera confianza en el proceso de análisis?	4 100%		
12	¿Es confiable el resultado del proceso de análisis con la arquitectura de la HoneyNet, con la integración de datos?	3 75%	1 25%	
13	¿Con la arquitectura de la HoneyNet, con la integración de datos, es aceptable el tiempo de detección de:...? Ataques de suplantación.	4 100%		
	Inyecciones de SQL en el servidor.	4 100%		
	Ataques de Fuerza bruta.	4 100%		
	Ataques de Denegación del servicio.	4 100%		

	Envenenamiento DNS.	4 100%		
	Ataques de exploración de puertos.	4 100%		
	Ataques del Internet Control Message Protocol (ICMP).	4 100%		
	Ataque de Cross Site Scripting.	4 100%		
14	¿La funcionalidad de la arquitectura la HoneyNet con la herramienta de integración de datos, garantiza el almacenamiento de la información?	4 100%		
15	¿La funcionalidad de la arquitectura la HoneyNet con la herramienta de integración de datos, garantiza la integridad de los datos almacenados?	4 100%		
16	¿Los datos manejados con la arquitectura la HoneyNet con la herramienta de integración, son consistentes?	4 100%		
17	¿La arquitectura de la HoneyNet, con la herramienta de integración de datos, permite la interpretación univoca de la información?	1 25%	3 75%	
18	¿Existe precisión técnica en la data administrada con la arquitectura de la HoneyNet, con la herramienta de integración de datos?	4 100%		

19	¿La interfaz de usuario de la HoneyNet con la herramienta de integración de datos facilita la interacción con el usuario?	4 100%		
20	¿La HoneyNet, con la herramienta de integración de datos, permite al usuario operar el sistema con facilidad?	4 100%		
21	¿La interfaz de usuario de la HoneyNet, con la herramienta de integración de datos, satisface la interacción con el usuario?	4 100%		
22	¿La arquitectura de la HoneyNet, con la herramienta de integración de datos, genera los reportes de incidentes suficientes?	4 100%		
23	¿La arquitectura de la HoneyNet, con la herramienta de integración de datos, genera los reportes de incidentes a tiempo?	4 100%		