Universidad Distrital Francisco José de Caldas
Facultad de Ingeniería

Diagnóstico de Seguridad IP con Fuentes Abiertas en Colombia

Proyecto de Investigación

Investigadores:

Juan Manuel Serrano Rodríguez
Código: 20211020091
jumserranor@udistrital.edu.co

Nicolás Guevara Herrán
Código: 20211180026
nguevarah@udistrital.edu.co

Área de Investigación:
Ciberseguridad y Análisis de Redes

Grupo de Investigación:
Modelamiento en Ingeniería de Sistemas

Director:
Lilian Astrid Bejarano Garzón

1 de octubre de 2025

1. Selección y Definición del Tema de Investigación

1.1. Tema de Investigación

"Desarrollo de una Herramienta de Diagnóstico Unificado para el Análisis de Seguridad y Geolocalización de Direcciones IP mediante Integración de Fuentes Abiertas en el Contexto de Ciberseguridad Colombiana"

1.2. Título Provisional

"Diagnóstico de Seguridad IP con Fuentes Abiertas en Colombia"

1.3. Línea de Investigación

Este proyecto se enmarca dentro de la línea de investigación en **Inteligencia de Amenazas Cibernéticas**, específicamente en el área de **Análisis Automatizado de Infraestructuras de Red** y **Desarrollo de Herramientas de Ciberseguridad Basadas en Fuentes Abiertas**.

1.4. Área del Conocimiento

- **Área principal:** Ingeniería de Sistemas y Computación
- **Subárea:** Ciberseguridad y Redes de Computadores
- **Disciplina:** Inteligencia de Amenazas y Análisis de Vulnerabilidades

2. El Problema de Investigación

2.1. Planteamiento del Problema

2.1.1. Situación Actual (Diagnóstico)

En el contexto actual de amenazas cibernéticas sofisticadas, la evaluación de la legitimidad y seguridad de direcciones IP se ha convertido en una actividad fundamental para profesionales de ciberseguridad en Colombia. Según el reporte del Computer Emergency Response Team de Colombia (COLCERT), el país registró 36.000 millones de intentos de ciberataques en 2024, ubicándose como el cuarto país en América Latina con mayor exposición a amenazas cibernéticas [1].

El análisis detallado de la situación revela múltiples problemáticas interconectadas:

Fragmentación de la Información de Seguridad: Las organizaciones colombianas dependen de múltiples plataformas especializadas como VirusTotal, Shodan, AbuseIPDB, y servicios comerciales como IBM X-Force y ThreatConnect para obtener información sobre direcciones IP sospechosas. Esta fragmentación genera inconsistencias en los análisis, requiere múltiples subscripciones costosas, y consume tiempo significativo en la correlación manual de datos [2].

Limitaciones Económicas para Acceso a Herramientas Comerciales: Un estudio realizado por la Cámara Colombiana de Informática y Telecomunicaciones (CCIT) en 2024 reveló que el 73 % de las PYMES colombianas consideran prohibitivos los costos de herramientas especializadas de ciberseguridad, con precios que oscilan entre USD $5,000 y $50,000 anuales para soluciones empresariales [3].

Dependencia de APIs Comerciales con Restricciones: Los servicios gratuitos de análisis IP imponen limitaciones severas: VirusTotal permite 4 consultas por minuto para usuarios gratuitos, Shodan limita a 100 consultas mensuales, y AbuseIPDB restringe a 1,000 consultas diarias. Estas limitaciones hacen impracticable el análisis masivo requerido en contextos empresariales [4].

Carencia de Capacidades Técnicas Locales: El déficit de profesionales especializados en desarrollo de herramientas de ciberseguridad en Colombia es evidente. Según MinTIC, existe una brecha de 68,000 profesionales en ciberseguridad a nivel nacional, lo que limita las capacidades para desarrollar soluciones tecnológicas autóctonas [5].

Ausencia de Estándares Nacionales: Colombia carece de marcos estandarizados para la integración de fuentes abiertas en análisis de ciberseguridad, lo que resulta en implementaciones ad-hoc inconsistentes entre organizaciones [6].

2.1.2. Análisis de Causas

Las causas identificadas mediante análisis de Ishikawa incluyen:

Factores Tecnológicos:

- Dispersión de datos de inteligencia en servicios propietarios
- Falta de APIs unificadas para consulta de múltiples fuentes
- Limitaciones técnicas en la integración de datasets heterogéneos
- Ausencia de herramientas nacionales especializadas

Factores Económicos:

- Costos elevados de licencias para herramientas comerciales
- Limitaciones presupuestarias en organizaciones públicas y PYMES
- Dependencia de monedas extranjeras para servicios internacionales
- Falta de financiación para I+D en ciberseguridad nacional

Factores Humanos:

- Escasez de expertise técnico especializado
- Falta de programas de formación en desarrollo de herramientas de ciberseguridad
- Resistencia al cambio hacia soluciones open source
- Limitada cultura de colaboración entre organizaciones

Factores Regulatorios:

- Ausencia de políticas públicas para fomento de desarrollo tecnológico nacional
- Falta de estándares técnicos para herramientas de ciberseguridad
- Limitaciones en marcos legales para compartir información de amenazas

2.1.3. Pronóstico

Si persiste esta situación, se proyectan las siguientes consecuencias para el período 2025-2027:

Impacto Operacional:

- Incremento del 40 % en tiempos de respuesta ante incidentes de seguridad
- Reducción del 25 % en la efectividad de detección de amenazas
- Aumento del 60 % en costos operativos por múltiples licencias

Impacto Estratégico:

- Profundización de la brecha tecnológica con países desarrollados
- Mayor dependencia de proveedores extranjeros
- Limitación en capacidades de soberanía digital nacional

Impacto Económico:

- Pérdidas estimadas de USD $2.8 billones anuales por ciberataques exitosos
- Reducción de la competitividad de organizaciones colombianas
- Limitación en adopción de tecnologías emergentes por riesgos de seguridad

2.1.4. Control al Pronóstico

Para evitar este escenario negativo, es fundamental desarrollar una solución tecnológica que:

1. **Integre fuentes abiertas gratuitas** como Censys (disponible en Google BigQuery) y MaxMind GeoLite2 para proporcionar análisis comprehensivos sin costos de licenciamiento.
2. **Implemente algoritmos de correlación avanzados** que combinen reconocimiento pasivo con geolocalización para generar evaluaciones de riesgo contextualizadas.
3. **Proporcione interfaces intuitivas** que democraticen el acceso a capacidades avanzadas de análisis IP.
4. **Establezca metodologías reproducibles** que puedan ser adoptadas por organizaciones con diferentes niveles de madurez tecnológica.
5. **Genere capacidades técnicas locales** mediante documentación detallada y código abierto que facilite la transferencia de conocimiento.

2.2. Formulación del Problema

¿Cómo desarrollar e implementar una herramienta de diagnóstico unificado que integre eficientemente el análisis de seguridad y geolocalización de direcciones IP utilizando exclusivamente fuentes abiertas, para fortalecer las capacidades de inteligencia de amenazas cibernéticas en organizaciones colombianas, considerando las limitaciones presupuestarias, técnicas y operacionales del contexto nacional?

2.3. Sistematización del Problema

2.3.1. Preguntas Técnicas

1. ¿Qué características técnicas debe poseer una herramienta unificada de análisis de seguridad IP para maximizar la utilidad de fuentes abiertas disponibles como Censys BigQuery y GeoLite2?

2. ¿Cómo integrar efectivamente los conjuntos de datos de Censys en BigQuery con la base de datos GeoLite2 de MaxMind para generar análisis correlacionados?

3. ¿Qué algoritmos de machine learning son más efectivos para la clasificación automatizada de amenazas basada en patrones de comportamiento de IP?

4. ¿Cómo optimizar las consultas SQL en BigQuery para minimizar costos de procesamiento manteniendo tiempos de respuesta aceptables?

2.3.2. Preguntas Metodológicas

1. ¿Qué metodologías de presentación de resultados garantizan la reproducibilidad y comprensión de análisis de seguridad IP por parte de usuarios con diferentes niveles técnicos?

2. ¿Cómo establecer métricas de evaluación que permitan comparar objetivamente la efectividad de la herramienta desarrollada con soluciones comerciales?

3. ¿Qué protocolos de validación son necesarios para asegurar la precisión y confiabilidad de los resultados obtenidos?

2.3.3. Preguntas Contextuales

1. ¿Cuáles son las limitaciones inherentes del uso exclusivo de fuentes gratuitas en comparación con servicios comerciales en el contexto de amenazas específicas a Colombia?

2. ¿Cómo adaptar la herramienta a las particularidades del panorama de amenazas cibernéticas colombiano?

3. ¿Qué estrategias de implementación son más efectivas para facilitar la adopción de la herramienta en organizaciones con diferentes niveles de madurez tecnológica?

2.3.4. Preguntas de Evaluación

1. ¿Cómo validar la efectividad y precisión de la herramienta desarrollada en contextos reales de ciberseguridad colombiana?

2. ¿Qué indicadores de impacto permiten medir la contribución de la herramienta al fortalecimiento de capacidades nacionales de ciberseguridad?

3. ¿Cómo asegurar la sostenibilidad y evolución continua de la herramienta desarrollada?

3. Objetivos de la Investigación

3.1. Objetivo General

Desarrollar e implementar una herramienta de diagnóstico unificado que integre el análisis de seguridad y geolocalización de direcciones IP mediante el uso exclusivo de fuentes abiertas, para fortalecer las capacidades de inteligencia de amenazas cibernéticas en organizaciones colombianas, evaluando su efectividad comparativa con soluciones comerciales existentes.

3.2. Objetivos Específicos

3.2.1. Objetivos de Diseño y Desarrollo

1. **Diseñar la arquitectura técnica integral** de una plataforma que integre eficientemente los conjuntos de datos de Censys disponibles en Google BigQuery con la base de datos GeoLite2 de MaxMind, considerando aspectos de escalabilidad, rendimiento y mantenibilidad.

2. **Implementar consultas optimizadas en BigQuery** para la extracción eficiente de información sobre puertos, servicios, certificados SSL/TLS, y metadatos de red asociados a direcciones IP específicas, minimizando costos computacionales y tiempos de respuesta.

3. **Desarrollar algoritmos de correlación avanzados** que combinen datos de reconocimiento pasivo con información de geolocalización para generar evaluaciones comprehensivas de seguridad, incluyendo scoring de riesgo automatizado y detección de anomalías.

3.2.2. Objetivos de Interfaz y Usabilidad

4. **Crear una interfaz de usuario intuitiva y responsiva** que presente los resultados de análisis de forma clara, estructurada y reproducible, incluyendo visualizaciones interactivas, reportes exportables y dashboards personalizables.

5. **Implementar funcionalidades de análisis masivo** que permitan el procesamiento simultáneo de múltiples direcciones IP con capacidades de filtrado, agrupación y análisis comparativo de resultados.

3.2.3. Objetivos de Validación y Evaluación

6. **Validar la efectividad de la herramienta** mediante pruebas comparativas exhaustivas con servicios comerciales existentes (VirusTotal, Shodan, AbuseIPDB) utilizando métricas de precisión, recall, F1-score y tiempo de respuesta.

7. **Evaluar la aplicabilidad en contextos reales** mediante casos de uso específicos del panorama de ciberseguridad colombiano, incluyendo análisis de campañas de malware locales, investigación de infraestructuras de comando y control, y respuesta a incidentes.

3.2.4. Objetivos de Transferencia y Sostenibilidad

8. **Documentar metodologías y mejores prácticas** para la implementación, configuración y uso de la herramienta, facilitando su adopción por organizaciones con diferentes niveles de madurez tecnológica.

9. **Establecer un marco de evaluación de impacto** que permita medir la contribución de la herramienta al fortalecimiento de capacidades nacionales de ciberseguridad y la reducción de dependencia de soluciones comerciales extranjeras.

3.3. Objetivos Secundarios

3.3.1. Objetivos de Investigación Aplicada

- Caracterizar las limitaciones y ventajas del uso exclusivo de fuentes abiertas en análisis de seguridad IP
- Identificar patrones específicos de amenazas en el contexto colombiano mediante análisis de datos históricos
- Establecer benchmarks de rendimiento para herramientas de análisis IP basadas en fuentes abiertas

3.3.2. Objetivos de Contribución Académica

- Publicar resultados en conferencias nacionales e internacionales de ciberseguridad
- Generar artículos científicos sobre metodologías de integración de fuentes abiertas

- Contribuir al desarrollo de estándares nacionales para herramientas de ciberseguridad

3.3.3. Objetivos de Impacto Social

- Democratizar el acceso a herramientas avanzadas de análisis IP para organizaciones con limitaciones presupuestarias
- Fortalecer las capacidades técnicas locales mediante transferencia de conocimiento y código abierto
- Contribuir a la soberanía digital nacional mediante el desarrollo de capacidades tecnológicas autóctonas

4. Justificación de la Investigación

4.1. Justificación Teórica

4.1.1. Contribución al Conocimiento Científico

Esta investigación contribuye significativamente al avance del conocimiento en el campo de la inteligencia de amenazas cibernéticas mediante la exploración sistemática de metodologías de integración de fuentes abiertas para análisis de seguridad IP. Los marcos teóricos existentes sobre reconocimiento pasivo y geolocalización se han desarrollado principalmente en contextos de recursos ilimitados y acceso a herramientas comerciales premium, por lo que es fundamental contrastar estos enfoques con las limitaciones reales de organizaciones con presupuestos restringidos [7].

El estudio aborda una brecha significativa en la literatura científica relacionada con la **democratización de capacidades de ciberseguridad**. Mientras que investigaciones previas como las de Durumeric et al. (2015) y Hao et al. (2018) han explorado el uso de datasets masivos para análisis de seguridad, pocos estudios han abordado específicamente la viabilidad de implementar capacidades equivalentes utilizando exclusivamente recursos gratuitos [8][9].

4.1.2. Marcos Teóricos Emergentes

La investigación contribuye al desarrollo de marcos teóricos emergentes en:

Teoría de Fusión de Datos en Ciberseguridad: Expandiendo los modelos existentes de Llinas et al. (2004) para incluir fuentes heterogéneas de inteligencia de amenazas con diferentes niveles de confiabilidad y granularidad temporal [10].

Teoría de Análisis de Riesgos Contextualizados: Desarrollando extensiones a los frameworks de Kaplan y Garrick (1981) adaptados específicamente para amenazas cibernéticas geográficamente contextualizadas [11].

Teoría de Sistemas de Información Distribuidos para Ciberseguridad: Contribuyendo a los modelos de Tanenbaum y Van Steen (2016) mediante la implementación de arquitecturas que integran servicios cloud externos con procesamiento local [12].

4.1.3. Hipótesis Científicas a Contrastar

1. **Hipótesis de Equivalencia Funcional:** Las capacidades de análisis IP obtenidas mediante integración sistemática de fuentes abiertas pueden alcanzar niveles de efectividad estadísticamente equivalentes (p ¿0.05) a soluciones comerciales en detección de amenazas conocidas.

2. **Hipótesis de Correlación Espacial:** Existe una correlación positiva significativa (r ¿0.7) entre la precisión de geolocalización IP y la efectividad de detección de campañas de malware regionalmente específicas.

3. **Hipótesis de Optimización de Recursos:** La implementación de algoritmos de consulta optimizados puede reducir los costos computacionales en un factor de 10x comparado con enfoques naive de consulta a múltiples fuentes.

4.2. Justificación Metodológica

4.2.1. Innovación Metodológica

La investigación desarrolla una metodología innovadora para la integración sistemática de múltiples fuentes abiertas de datos de ciberseguridad, estableciendo protocolos estandarizados que abordan desafíos específicos no resueltos en la literatura existente:

Metodología de Sincronización Temporal: Desarrollo de algoritmos para correlacionar datos de fuentes con diferentes frecuencias de actualización (Censys: semanal, GeoLite2: mensual, feeds de amenazas: tiempo real) manteniendo coherencia temporal en los análisis [13].

Framework de Evaluación de Confiabilidad: Implementación de métricas bayesianas para asignar pesos dinámicos a diferentes fuentes basándose en su historial de precisión para tipos específicos de amenazas [14].

Metodología de Optimización de Consultas Distribuidas: Desarrollo de

algoritmos de optimización que minimizan el número de consultas necesarias a múltiples APIs manteniendo la completitud de la información [15].

4.2.2. Instrumentos Metodológicos Desarrollados

Los instrumentos metodológicos incluyen:

Algoritmos de Fusión de Datos Multidimensionales:

- Algoritmo de correlación espacial para integrar geolocalización con datos de red
- Algoritmo de scoring de riesgo que combina múltiples indicadores
- Algoritmo de detección de anomalías basado en patrones geográficos

Métricas de Evaluación Especializadas:

- Métrica de *Coverage Efficiency*: Relación entre amplitud de análisis y recursos computacionales
- Métrica de *Temporal Coherence*: Consistencia de resultados a lo largo del tiempo
- Métrica de *Cross-Source Reliability*: Concordancia entre diferentes fuentes de datos

Frameworks de Validación Comparativa:

- Protocolo de evaluación ciega con datasets de referencia
- Metodología de benchmarking contra soluciones comerciales
- Framework de evaluación de usabilidad para diferentes perfiles de usuario

4.2.3. Replicabilidad y Transferibilidad

La metodología desarrollada está diseñada para ser completamente replicable, con:

- Documentación detallada de todos los algoritmos implementados
- Código fuente abierto con comentarios exhaustivos
- Datasets de prueba estandarizados para validación
- Protocolos de instalación y configuración paso a paso

Esta contribución metodológica facilitará futuras investigaciones en el desarrollo de herramientas de ciberseguridad basadas en recursos abiertos, estableciendo un estándar de facto para la integración de fuentes heterogéneas.

4.3. Justificación Práctica

4.3.1. Impacto Inmediato en el Sector

Los resultados de esta investigación tendrán aplicación inmediata en la mejora de capacidades de ciberseguridad de organizaciones colombianas, particularmente aquellas con limitaciones presupuestarias para adquisición de herramientas comerciales especializadas. La herramienta desarrollada proporcionará una alternativa gratuita y efectiva para análisis de seguridad IP, con impacto directo en:

Sector Público:

- 156 entidades gubernamentales nivel nacional que podrán implementar capacidades avanzadas de análisis IP sin costos de licenciamiento
- 1,102 municipios colombianos que tendrán acceso a herramientas de ciberseguridad previamente inaccesibles por limitaciones presupuestarias
- Fortalecimiento del CERT Nacional (COLCERT) con capacidades de análisis automatizado

Sector Privado - PYMES:

- 2.7 millones de micro, pequeñas y medianas empresas en Colombia según datos del DANE 2024
- Reducción de costos operativos en ciberseguridad estimada en 70 %-90 % comparado con soluciones comerciales
- Mejora en capacidades de respuesta a incidentes para empresas sin departamentos especializados de seguridad

Sector Académico:

- 288 instituciones de educación superior que podrán incorporar herramientas reales en sus programas de ciberseguridad
- Laboratorios de investigación con acceso a capacidades de análisis de nivel empresarial
- Semilleros de investigación con herramientas para proyectos aplicados

4.3.2. Democratización del Acceso a Tecnología

La investigación contribuye significativamente a la democratización del acceso a tecnologías de ciberseguridad avanzadas mediante:

Eliminación de Barreras Económicas:

- Reducción de costos de entrada de USD $50,000 anuales (soluciones comerciales) a costos mínimos de infraestructura cloud (¡USD $100 mensuales)
- Eliminación de dependencia de licencias por usuario, permitiendo escalamiento sin costos proporcionales
- Reducción de costos de capacitación mediante interfaces intuitivas y documentación comprehensiva

Fortalecimiento de Capacidades Locales:

- Desarrollo de expertise técnico nacional mediante código abierto y documentación detallada
- Creación de una comunidad de práctica alrededor de la herramienta
- Generación de oportunidades de empleo especializado en el sector nacional

4.3.3. Impacto en Indicadores Nacionales

La implementación exitosa de la herramienta contribuirá a la mejora de indicadores nacionales de ciberseguridad:

Índice Global de Ciberseguridad (GCI) de la ITU:

- Colombia ocupa actualmente la posición 87 de 194 países
- La herramienta contribuirá específicamente al pilar de Çapacidades Técnicas"del índice
- Meta: Mejora de 10 posiciones en el ranking global para 2027

Objetivos de Desarrollo Sostenible (ODS):

- ODS 9: Industria, Innovación e Infraestructura - Fortalecimiento de capacidades tecnológicas nacionales
- ODS 16: Paz, Justicia e Instituciones Sólidas - Mejora en capacidades de seguridad nacional
- ODS 17: Alianzas para lograr los Objetivos - Promoción de colaboración tecnológica

4.3.4. Sostenibilidad y Escalabilidad

El proyecto está diseñado para asegurar sostenibilidad a largo plazo mediante:
Modelo de Desarrollo Abierto:

- Código fuente disponible bajo licencia open source
- Documentación técnica comprehensiva para facilitar contribuciones externas
- Arquitectura modular que permite extensiones y mejoras incrementales

Independencia Tecnológica:

- Uso exclusivo de fuentes de datos gratuitas y permanentes
- Arquitectura cloud-native compatible con múltiples proveedores
- Diseño que minimiza dependencias de servicios comerciales específicos

Transferencia de Conocimiento:

- Capacitación de equipos técnicos locales para mantenimiento y evolución
- Establecimiento de alianzas con universidades para investigación continua
- Creación de certificaciones técnicas específicas para la herramienta

5. Marco de Referencia

5.1. Marco Teórico

5.1.1. Teorías de Reconocimiento Pasivo en Ciberseguridad

Fundamentos Conceptuales: El reconocimiento pasivo, definido formalmente por Gordon y Loeb (2002) como .[el] proceso de recolección de información sobre sistemas objetivo sin establecer comunicación directa que pueda ser detectada por sistemas de monitoreo", representa una metodología fundamental en ciberseguridad moderna [16]. Esta técnica se diferencia conceptualmente del reconocimiento activo en tres dimensiones críticas:

1. **Detectabilidad:** El reconocimiento pasivo no genera tráfico hacia el objetivo, eliminando rastros en logs de sistemas monitoreados
2. **Legalidad:** Utiliza información públicamente disponible, evitando implicaciones legales asociadas con scanning activo

3. **Escala:** Permite análisis masivos sin limitaciones de ancho de banda o restricciones de rate limiting

Evolución Tecnológica: Las técnicas modernas han evolucionado significativamente desde los enfoques manuales de los años 90 hasta sistemas automatizados impulsados por inteligencia artificial. Shodan (2009), Censys (2015), y Binary Edge (2018) han democratizado el acceso a datos de reconocimiento masivo, mientras que frameworks como Maltego, SpiderFoot y Recon-ng han automatizado la correlación de información OSINT [17][18][19].

La integración de machine learning ha introducido capacidades de:

- **Análisis predictivo** para identificar infraestructuras de comando y control emergentes
- **Clustering automático** de activos relacionados basándose en patrones de certificados y configuraciones
- **Detección de anomalías** en comportamientos de red y configuraciones de servicios

Modelos Teóricos Contemporáneos: El modelo de Ïnternet Scanning as a Service"propuesto por Durumeric et al. (2015) establece un framework donde organizaciones pueden acceder a capacidades de reconocimiento a escala de Internet sin operar infraestructura propia [20]. Este modelo se basa en tres pilares:

Centralización de Recursos: Concentración de capacidades de scanning en infraestructuras especializadas

Democratización de Acceso: APIs y interfaces que permiten acceso programático a datos

Estandarización de Datos: Formatos consistentes que facilitan análisis automatizado

5.1.2. Teorías de Geolocalización de Direcciones IP

Fundamentos Matemáticos: La geolocalización IP se basa en modelos matemáticos que correlacionan propiedades de red observables con ubicaciones geográficas. Padmanabhan y Subramanian (2001) establecieron los fundamentos teóricos mediante el modelo de çonstraining geography"que utiliza mediciones de latencia RTT (Round Trip Time) para establecer límites geográficos [21].

El modelo fundamental se expresa como:

$$D_{geo} \leq \frac{RTT \times c}{2n} \tag{1}$$

Donde D_{geo} es la distancia geográfica máxima, RTT es el tiempo de ida y vuelta medido, c es la velocidad de la luz, y n es el índice de refracción del medio (típicamente 1.5 para fibra óptica).

Limitaciones y Desafíos: Investigaciones posteriores han identificado limitaciones significativas en enfoques basados únicamente en latencia:

- **Routing Indirection:** El tráfico puede seguir rutas geográficamente subóptimas debido a políticas de BGP [22]
- **Infrastructure Asymmetry:** Diferencias en infraestructura de red entre regiones afectan significativamente las mediciones [23]
- **CDN and Anycast:** Tecnologías modernas de distribución de contenido invalidan asunciones básicas de correlación geográfica [24]

Enfoques Multifuente Contemporáneos: Los marcos modernos combinan múltiples fuentes de información para mejorar precisión:

Bases de Datos Comerciales: MaxMind GeoIP2, IP2Location, y Neustar utilizan combinaciones de:

- Registros WHOIS y asignaciones de bloques IP
- Información proporcionada por ISPs y operadores
- Datos crowdsourced de aplicaciones móviles y servicios web
- Correlaciones con información de DNS y servicios públicos

Algoritmos de Machine Learning: Wang et al. (2020) propusieron marcos que utilizan:

- **Redes Convolucionales de Grafos (GCNs)** para modelar topologías de red
- **Optimización de función de energía** con muestreo de Monte Carlo para abordar incertidumbre
- **Ensemble methods** que combinan predicciones de múltiples algoritmos especializados [25]

5.1.3. Marcos de Inteligencia de Amenazas Cibernéticas

Evolución Conceptual: La inteligencia de amenazas ha evolucionado desde análisis manuales reactivos hacia sistemas automatizados predictivos. El modelo de Ïntelligence-Driven Defense"propuesto por Hutchins et al. (2011) establece que la efectividad de la ciberseguridad es proporcional a la calidad y velocidad de la inteligencia de amenazas disponible [26].

Frameworks Contemporáneos:

NIST Cybersecurity Framework 2.0 (2024):

- **Identify:** Catalogación y priorización de activos y amenazas
- **Protect:** Implementación de salvaguardas basadas en inteligencia
- **Detect:** Monitoreo continuo informado por indicadores de amenazas
- **Respond:** Respuesta rápida basada en inteligencia contextual
- **Recover:** Restauración informada por análisis post-incidente
- **Govern:** Gestión estratégica de programas de ciberseguridad [27]

MITRE ATT&CK Framework: Taxonomía global de técnicas de adversarios basada en observaciones del mundo real, organizada en:

- 14 tácticas principales (Initial Access, Execution, Persistence, etc.)
- 193 técnicas específicas con sub-técnicas detalladas
- Matrices especializadas para Enterprise, Mobile, y ICS/SCADA [28]

D3FEND Framework: Complemento defensivo de ATT&CK que categoriza contramedidas técnicas:

- **Harden:** Técnicas para reducir superficie de ataque
- **Detect:** Metodologías de identificación de actividad maliciosa
- **Isolate:** Estrategias de contención de amenazas
- **Deceive:** Técnicas de engaño y honeypots [29]

5.2. Marco Conceptual

5.2.1. Definiciones Operacionales

Reconocimiento Pasivo: Proceso sistemático de recolección de información sobre sistemas objetivo mediante el uso de fuentes públicas y datasets pre-existentes, sin establecer comunicación directa que pueda ser detectada por sistemas de monitoreo del objetivo. Incluye análisis de certificados SSL/TLS, registros DNS históricos, y datos de scanning de terceros.

Geolocalización IP: Técnica computacional para determinar la ubicación geográfica aproximada de una dirección IP mediante el análisis correlacionado de bases de datos especializadas, mediciones de red, y algoritmos de inferencia espacial. La precisión típica varía entre 50-100 km para ubicaciones urbanas y 100-500 km para ubicaciones rurales.

Inteligencia de Amenazas: Conocimiento basado en evidencia, incluyendo contexto, mecanismos, indicadores, implicaciones y consejos orientados a la acción sobre amenazas existentes o emergentes. Se categoriza en cuatro niveles: Estratégica (tendencias a largo plazo), Táctica (TTPs específicos), Operacional (campañas activas), y Técnica (IOCs específicos).

Fuentes Abiertas (OSINT): Información disponible públicamente que puede ser recolectada, analizada y utilizada para propósitos de inteligencia sin restricciones legales, éticas o contractuales. Incluye datos de motores de búsqueda especializados, bases de datos académicas, registros públicos, y datasets gubernamentales.

BigQuery: Servicio de almacén de datos completamente administrado y sin servidor de Google Cloud Platform que permite consultas SQL escalables sobre petabytes de datos. Utiliza arquitectura distribuida columnar optimizada para análisis de grandes volúmenes de datos estructurados y semi-estructurados.

Censys: Plataforma de mapeo de Internet que proporciona visibilidad sobre dispositivos, servicios e infraestructuras conectadas mediante escaneos automatizados regulares. Mantiene datos históricos de puertos, certificados, y configuraciones de servicios accesibles desde 2015.

Threat Intelligence Platform (TIP): Sistema integrado que agrega, correlaciona, y analiza datos de múltiples fuentes para generar inteligencia accionable. Incluye capacidades de enriquecimiento automático, scoring de confianza, y distribución de indicadores.

Indicator of Compromise (IOC): Artefacto digital observado en redes o sistemas operativos que indica con alta confianza una intrusión o actividad maliciosa. Incluye direcciones IP, dominios, hashes de archivos, y patrones de tráfico específicos.

5.2.2. Taxonomía de Amenazas IP

Categorización por Origen:

- **Malware C&C:** Direcciones IP utilizadas como servidores de comando y control
- **Botnet Infrastructure:** IPs comprometidas formando parte de redes bot
- **Phishing Hosting:** Servidores que alojan contenido de phishing
- **Scanning Sources:** IPs que realizan reconocimiento automatizado
- **Exploit Delivery:** Servidores que distribuyen exploits y payloads

Categorización por Comportamiento:

- **Persistent Threats:** Amenazas de larga duración con infraestructura estable
- **Fast-Flux Networks:** Infraestructuras que cambian rápidamente de IP
- **Domain Generation Algorithms (DGA):** Amenazas que generan dominios algorítmicamente
- **Bulletproof Hosting:** Servicios de hosting tolerantes a actividad maliciosa

5.3. Marco Espacial

5.3.1. Contexto Geográfico Nacional

La investigación se desarrolla en el territorio colombiano, considerando las particularidades geográficas, demográficas y tecnológicas que influyen en el panorama de ciberseguridad:

Distribución Poblacional y Digital:

- 50.4 millones de habitantes distribuidos en 1,142 municipios
- 36.2 millones de usuarios de Internet (71.5 % de penetración)
- Concentración del 60 % de infraestructura tecnológica en Bogotá, Medellín y Cali

- 2.1 millones de direcciones IPv4 asignadas a Colombia según LACNIC

Infraestructura de Conectividad:

- 15 cables submarinos internacionales
- 127 ISPs registrados con diferentes niveles de cobertura
- Red Nacional de Fibra Óptica con 28,000 km desplegados
- 4 Internet Exchange Points (IXPs) principales

5.3.2. Panorama de Amenazas Específico

Amenazas Predominantes en Colombia (2024):

- **Banking Trojans:** 34 % de los ataques detectados
- **Ransomware:** 28 % con tendencia creciente
- **Cryptomining Malware:** 18 % aprovechando infraestructura nacional
- **Mobile Malware:** 12 % dirigido específicamente a usuarios colombianos
- **Supply Chain Attacks:** 8 % con impacto en múltiples organizaciones

Sectores Más Atacados:

1. Sector Financiero (31 % de incidentes reportados)
2. Gobierno y Administración Pública (24 %)
3. Telecomunicaciones (18 %)
4. Energía y Servicios Públicos (15 %)
5. Educación (12 %)

5.3.3. Ecosistema de Ciberseguridad Nacional

Actores Institucionales:

- **COLCERT:** Grupo de Respuesta a Emergencias Cibernéticas de Colombia
- **MinTIC:** Ministerio de Tecnologías de la Información y las Comunicaciones
- **DNE:** Dirección Nacional de Inteligencia

- **CCOC:** Comando Conjunto Cibernético
- **SIC:** Superintendencia de Industria y Comercio

Iniciativas Estratégicas:

- CONPES 3995: Política Nacional de Transformación Digital e Inteligencia Artificial
- CONPES 3854: Política Nacional de Ciberseguridad y Ciberdefensa
- Estrategia Nacional de Economía Digital 2025
- Plan Nacional de Infraestructuras de Datos

5.4. Marco Temporal

5.4.1. Ventana de Investigación

El estudio abarca el período **enero 2025 - diciembre 2025**, un momento crítico que coincide con:

Contexto Nacional:

- Implementación de la nueva Ley de Ciberseguridad (Ley 2273 de 2022)
- Despliegue de la infraestructura 5G nacional
- Finalización del primer quinquenio del Plan Nacional de Desarrollo Digital
- Evaluación intermedia de la Estrategia Nacional de Ciberseguridad 2022-2026

Contexto Internacional:

- Entrada en vigor de nuevas regulaciones de ciberseguridad de la UE (NIS2 Directive)
- Implementación del Cyber Resilience Act europeo
- Evolución post-quantum cryptography standards
- Maduración de tecnologías de IA generativa en ciberseguridad

5.4.2. Cronología de Amenazas Relevantes

Período de Análisis Histórico (2020-2024):

- **2020:** Incremento del 300 % en ataques durante pandemia COVID-19
- **2021:** Surgimiento de ransomware-as-a-service dirigido a LATAM
- **2022:** Proliferación de ataques a infraestructura crítica nacional
- **2023:** Emergencia de amenazas basadas en IA generativa
- **2024:** Sofisticación de ataques de cadena de suministro

Proyecciones Futuras (2025-2027):

- Integración de quantum computing en capacidades defensivas
- Automatización completa de respuesta a incidentes
- Implementación de zero-trust architecture a nivel nacional
- Desarrollo de capacidades de ciberdefensa autónoma

5.4.3. Ventana de Validación

La efectividad de la herramienta será evaluada considerando:

- **Datos históricos:** Análisis retrospectivo de amenazas 2020-2024
- **Datos contemporáneos:** Validación con amenazas activas durante 2025
- **Proyección futura:** Evaluación de adaptabilidad para amenazas emergentes

6. Aspectos Metodológicos

6.1. Tipo de Estudio

6.1.1. Clasificación Epistemológica

La investigación adopta un **paradigma post-positivista** que reconoce la objetividad aproximada del conocimiento científico mientras acepta las limitaciones inherentes en la medición de fenómenos complejos en ciberseguridad. Se fundamenta en el **realismo crítico** de Bhaskar (1975), que distingue entre dominios empírico, actual y real en la investigación de sistemas socio-técnicos [30].

6.1.2. Diseño de Investigación Mixto

El estudio implementa un **diseño explicativo secuencial** (QUAN → qual) donde:

Fase Cuantitativa Dominante:

- **Nivel descriptivo:** Caracterización estadística del estado actual de herramientas de análisis IP disponibles y necesidades específicas del contexto colombiano
- **Nivel explicativo:** Establecimiento de relaciones causales entre la integración de fuentes abiertas y la efectividad de análisis de seguridad
- **Nivel correlacional:** Análisis de asociaciones entre variables técnicas, económicas y operacionales

Fase Cualitativa Complementaria:

- Exploración fenomenológica de percepciones de usuarios
- Análisis interpretativo de contextos organizacionales
- Evaluación hermenéutica de usabilidad y aceptación

6.1.3. Componente Experimental

Diseño Cuasi-experimental con Grupos de Comparación:

- **Grupo Experimental:** Herramienta desarrollada (fuentes abiertas)
- **Grupos Control:** Soluciones comerciales (VirusTotal Premium, Shodan Enterprise, IBM X-Force)
- **Variables Dependientes:** Precisión, recall, tiempo de respuesta, cobertura
- **Variables Independientes:** Tipo de herramienta utilizada
- **Variables de Control:** Conjunto de datos de prueba, condiciones de red, configuración hardware

6.2. Método de Investigación

6.2.1. Enfoque Metodológico Integrado

Design Science Research (DSR): Siguiendo el framework de Hevner et al. (2004), la investigación combina rigor científico con relevancia práctica mediante ciclos iterativos de construcción y evaluación [31]:

1. **Identificación del Problema:** Análisis sistemático de limitaciones en herramientas existentes
2. **Definición de Objetivos:** Especificación de requisitos funcionales y no funcionales
3. **Diseño y Desarrollo:** Construcción iterativa del artefacto tecnológico
4. **Demostración:** Pruebas de concepto en entornos controlados
5. **Evaluación:** Validación empírica mediante métricas objetivas
6. **Comunicación:** Diseminación de resultados y artefactos

Método Científico Deductivo-Inductivo:

- **Deductivo:** Derivación de hipótesis específicas desde marcos teóricos establecidos
- **Inductivo:** Generalización de patrones observados hacia principios aplicables
- **Abductivo:** Inferencia de mejores explicaciones para fenómenos observados

Método Analítico-Sintético:

- **Análisis:** Descomposición del problema en componentes específicos (consultas BigQuery, integración GeoLite2, algoritmos de correlación, presentación de resultados)
- **Síntesis:** Integración holística de componentes en solución unificada

6.2.2. Estrategias Metodológicas Específicas

Benchmarking Comparativo: Implementación de metodología estandarizada para comparación objetiva con soluciones comerciales:

- Conjuntos de datos de referencia estandarizados

- Métricas de evaluación consistentes
- Protocolos de prueba reproducibles
- Análisis estadístico de significancia

Ingeniería de Software Ágil:

- Desarrollo iterativo con sprints de 2 semanas
- Integration y deployment continuos (CI/CD)
- Test-driven development (TDD)
- Refactoring continuo basado en feedback

Evaluación Centrada en Usuario:

- Design thinking para comprensión de necesidades
- Prototyping rápido con feedback iterativo
- Usability testing con métodos cuanti-cualitativos
- Accessibility evaluation según estándares WCAG 2.1

6.3. Fuentes y Técnicas para Recolección de Información

6.3.1. Fuentes Primarias

Datasets Técnicos:

- **Censys BigQuery Dataset:** Datos históricos y actuales de scanning de Internet (2015-2025)
  - Volumen: 50TB de datos estructurados
  - Frecuencia: Actualizaciones semanales
  - Cobertura: 4.3 billones de direcciones IPv4, 28 trillones IPv6
  - Granularidad: Puerto, servicio, certificado, geolocalización básica
- **MaxMind GeoLite2 Database:** Base de datos de geolocalización IP
  - Volumen: 500MB (CSV), 200MB (MMDB)
  - Frecuencia: Actualizaciones mensuales
  - Cobertura: Global con énfasis en países desarrollados

  - Precisión: 95 % a nivel país, 75 % a nivel ciudad

- **Threat Intelligence Feeds Públicos:**
  - SANS Internet Storm Center
  - Emerging Threats Open Ruleset
  - MISP Community Feeds
  - ThreatCrowd API

Datos Experimentales Generados:

- Métricas de rendimiento de consultas optimizadas vs. naive
- Tiempos de respuesta bajo diferentes cargas de trabajo
- Resultados de comparación con herramientas comerciales
- Logs de uso y comportamiento de usuarios durante testing

Datos de Campo:

- **Encuesta a profesionales de ciberseguridad (n=200):**
  - Población: Profesionales colombianos en ciberseguridad
  - Muestreo: Estratificado por sector (público, privado, académico)
  - Instrumento: Cuestionario estructurado (40 preguntas)
  - Distribución: Online mediante plataformas profesionales
- **Entrevistas semi-estructuradas (n=30):**
  - Participantes: CISOs, analistas SOC, investigadores
  - Duración: 45-60 minutos por entrevista
  - Modalidad: Virtual con grabación (previo consentimiento)
  - Enfoque: Fenomenológico-interpretativo

6.3.2. Fuentes Secundarias

Literatura Científica:

- Bases de datos académicas: IEEE Xplore, ACM Digital Library, SpringerLink, ScienceDirect

- Repositorios especializados: arXiv, IACR ePrint Archive, SANS Reading Room
- Conferencias relevantes: S&P, CCS, NDSS, USENIX Security, BlackHat, DEF CON
- Journals especializados: TDSC, TIFS, Computer & Security, JCS

Documentación Técnica y Estándares:

- APIs y documentación de Censys, MaxMind, Google BigQuery
- Estándares ISO/IEC 27000 series
- NIST Cybersecurity Framework documentation
- MITRE ATT&CK Framework knowledge base
- RFC documents relacionados con protocolos de Internet

Reportes Institucionales:

- Reportes de amenazas de COLCERT
- Estadísticas de MinTIC sobre infraestructura nacional
- Informes de organismos internacionales (ITU, ENISA, CISA)
- Análisis de mercado de firmas consultoras (Gartner, Forrester)

6.3.3. Técnicas de Recolección de Datos

Técnicas Automatizadas:

- **Web Scraping Ético:** Extracción de datos públicos de threat intelligence
- **API Integration:** Consultas programáticas a servicios públicos
- **Database Queries:** Extracción optimizada desde BigQuery
- **Log Mining:** Análisis de patrones en logs de sistema durante pruebas

Técnicas de Medición Experimental:

- **Benchmarking Automatizado:** Scripts para evaluación consistente de rendimiento
- **A/B Testing:** Comparación de interfaces y algoritmos alternativos

- **Load Testing:** Evaluación de escalabilidad con herramientas como JMeter
- **Security Testing:** Vulnerability assessment de la aplicación desarrollada

Técnicas de Investigación Social:

- **Encuestas Online:** Cuestionarios estructurados con escalas Likert
- **Focus Groups:** Sesiones de 6-8 participantes por perfil de usuario
- **User Journey Mapping:** Documentación de procesos de uso
- **Think-Aloud Protocol:** Evaluación de usabilidad con verbalización

6.4. Población y Muestra

6.4.1. Definición de Poblaciones

Población 1: Profesionales de Ciberseguridad en Colombia

- **Tamaño estimado:** 15,000 profesionales (MinTIC, 2024)
- **Características:** Profesionales activos en roles de seguridad informática
- **Distribución geográfica:** 60 % Bogotá, 20 % Medellín, 10 % Cali, 10 % otras ciudades
- **Distribución sectorial:** 35 % privado, 25 % público, 20 % consultoría, 20 % académico

Población 2: Organizaciones con Capacidades de Ciberseguridad

- **Tamaño estimado:** 2,500 organizaciones
- **Criterios de inclusión:** Presupuesto IT ¿$50,000 USD anuales, personal dedicado a seguridad
- **Sectores:** Financiero, gobierno, telecomunicaciones, energía, educación superior

Población 3: Direcciones IP para Análisis Técnico

- **Conjunto de referencia:** 1 millón de direcciones IP categorizadas
- **Distribución:** 40 % legítimas, 30 % sospechosas, 20 % maliciosas confirmadas, 10 % neutrales
- **Fuentes:** Threat intelligence feeds, honeypots, reportes de incidentes

6.4.2. Estrategia de Muestreo

Muestreo Estratificado Proporcional: Para la encuesta a profesionales:

- **Estratos:** Sector (público/privado), experiencia (¡5 años, 5-10 años, ¿10 años), rol (analista, manager, director)
- **Tamaño muestral:** n = 200 (error muestral ±6.9 %, confianza 95 %)
- **Asignación:** Proporcional al tamaño del estrato en la población

Muestreo Intencional: Para entrevistas cualitativas:

- **Criterio:** Máxima variación en experiencia y contexto organizacional
- **Tamaño:** n = 30 (saturación teórica esperada en 25-30 casos)
- **Distribución:** 10 sector público, 10 sector privado, 10 académico/consultoría

Muestreo Aleatorio Simple: Para evaluación técnica:

- **Dataset de prueba:** 10,000 direcciones IP seleccionadas aleatoriamente
- **Dataset de validación:** 5,000 direcciones IP adicionales
- **Criterio:** Representatividad de distribución geográfica y sectorial

6.5. Tratamiento de la Información

6.5.1. Análisis Estadístico Cuantitativo

Estadística Descriptiva:

- Medidas de tendencia central (media, mediana, moda)
- Medidas de dispersión (desviación estándar, rango intercuartílico)
- Distribuciones de frecuencia y percentiles
- Análisis de normalidad (Shapiro-Wilk, Kolmogorov-Smirnov)

Estadística Inferencial:

- **Pruebas paramétricas:** t-test, ANOVA, regresión lineal múltiple
- **Pruebas no paramétricas:** Mann-Whitney U, Kruskal-Wallis, Spearman correlation

- **Análisis multivariado:** PCA, clustering jerárquico, análisis discriminante
- **Machine Learning:** Random Forest, SVM, Neural Networks para clasificación

Métricas Especializadas de Ciberseguridad:

- **Precision:** $P = \frac{TP}{TP+FP}$
- **Recall:** $R = \frac{TP}{TP+FN}$
- **F1-Score:** $F1 = 2 \times \frac{P \times R}{P+R}$
- **Accuracy:** $Acc = \frac{TP+TN}{TP+TN+FP+FN}$
- **AUC-ROC:** Área bajo la curva ROC
- **Mean Time to Detection (MTTD):** Tiempo promedio de detección
- **Coverage Efficiency:** $CE = \frac{Indicadores\_Detectados}{Recursos\_Computacionales}$

6.5.2. Análisis Cualitativo

Codificación Temática:

1. **Codificación abierta:** Identificación inductiva de conceptos emergentes
2. **Codificación axial:** Establecimiento de relaciones entre categorías
3. **Codificación selectiva:** Integración alrededor de categorías centrales

Análisis de Contenido:

- Análisis léxico con herramientas NLP (frequency analysis, sentiment analysis)
- Identificación de patrones discursivos y marcos interpretativos
- Análisis de narrativas sobre experiencias con herramientas de ciberseguridad

Triangulación de Fuentes:

- Validación cruzada entre datos cuantitativos y cualitativos
- Comparación de perspectivas entre diferentes stakeholders
- Contraste entre datos auto-reportados y observaciones directas

6.5.3. Herramientas y Software

Análisis Estadístico:

- **R:** Análisis estadístico avanzado y visualización
- **Python:** Machine learning con scikit-learn, pandas, numpy
- **SPSS:** Análisis estadístico para ciencias sociales
- **Tableau:** Visualización de datos y dashboards interactivos

Análisis Cualitativo:

- **NVivo:** Análisis de datos cualitativos y codificación
- **ATLAS.ti:** Análisis de contenido y teoría fundamentada
- **MaxQDA:** Análisis de métodos mixtos

Desarrollo y Testing:

- **Google Cloud Platform:** Infraestructura y BigQuery
- **Docker:** Containerización y deployment
- **Kubernetes:** Orquestación y escalamiento
- **Git/GitHub:** Control de versiones y colaboración
- **JMeter:** Load testing y performance evaluation
- **SonarQube:** Code quality y security analysis

6.6. Consideraciones Éticas

6.6.1. Aspectos Éticos Técnicos

- **Uso Responsable de Datos:** Cumplimiento con regulaciones de privacidad
- **Transparencia Algorítmica:** Documentación detallada de decisiones de diseño
- **Prevención de Uso Malicioso:** Controles de acceso y audit trails
- **Responsabilidad Social:** Consideración de impactos societales

6.6.2. Protección de Participantes

- Consentimiento informado para todas las actividades de investigación
- Anonimización de datos personales y organizacionales
- Derecho de retiro sin penalización
- Confidencialidad de información sensible empresarial

6.6.3. Consideraciones de Seguridad

- Implementación de security by design en la herramienta
- Evaluación de vulnerabilidades antes del deployment
- Protección de datos en tránsito y en reposo
- Compliance con estándares de seguridad relevantes

7. Cronograma de Trabajo

7.1. Metodología de Planificación

El cronograma se desarrolla utilizando la metodología **Critical Path Method (CPM)** combinada con principios de gestión ágil de proyectos. Se implementa un enfoque iterativo con sprints de 2 semanas para el desarrollo de software y hitos trimestrales para evaluación de progreso. La planificación considera dependencias críticas, recursos limitados y ventanas de oportunidad para actividades de transferencia.

7.2. Estructura Temporal

7.2.1. Duración Total

Duración: 12 meses (52 semanas) **Esfuerzo:** 2,080 horas-persona **Fases principales:** 5 fases secuenciales con solapamiento controlado

7.2.2. Fase 1: Investigación y Diseño (Meses 1-3)

identifican dependencias críticas, hitos clave, y buffers de riesgo para asegurar la entrega exitosa dentro del marco temporal establecido.

7.3. Estructura de Desglose del Trabajo (WBS)

7.3.1. Fase 1: Investigación y Diseño (Meses 1-3)

Duración: 12 semanas
Esfuerzo: 480 horas-persona
Entregables principales: Arquitectura técnica, especificaciones funcionales, revisión de literatura

1.1 Investigación Preliminar (Semanas 1-4)

- **1.1.1** Revisión sistemática de literatura (40 horas)
- **1.1.2** Análisis de herramientas comerciales existentes (32 horas)
- **1.1.3** Caracterización del panorama nacional de amenazas (24 horas)
- **1.1.4** Evaluación de datasets disponibles (Censys, GeoLite2) (40 horas)
- **1.1.5** Definición de requerimientos funcionales y no funcionales (24 horas)

1.2 Diseño de Arquitectura (Semanas 5-8)

- **1.2.1** Diseño de arquitectura de sistema (48 horas)
- **1.2.2** Modelado de datos y esquemas de integración (40 horas)
- **1.2.3** Diseño de APIs y interfaces (32 horas)
- **1.2.4** Especificación de algoritmos de correlación (36 horas)
- **1.2.5** Diseño de seguridad y compliance (24 horas)

1.3 Prototipado Inicial (Semanas 9-12)

- **1.3.1** Desarrollo de proof-of-concept para consultas BigQuery (40 horas)
- **1.3.2** Prototipo de integración GeoLite2 (32 horas)
- **1.3.3** Validación técnica de factibilidad (24 horas)
- **1.3.4** Evaluación de rendimiento preliminar (32 horas)
- **1.3.5** Documentación de arquitectura (32 horas)

7.3.2. Fase 2: Desarrollo del Backend (Meses 3-6)

Duración: 14 semanas
Esfuerzo: 672 horas-persona
Entregables principales: API funcional, algoritmos de correlación, integración de datos

2.1 Infraestructura y Configuración (Semanas 13-16)

- **2.1.1** Setup de infraestructura Google Cloud Platform (24 horas)
- **2.1.2** Configuración de BigQuery y optimización de costos (32 horas)
- **2.1.3** Implementación de CI/CD pipeline (40 horas)
- **2.1.4** Setup de monitoreo y logging (24 horas)
- **2.1.5** Configuración de seguridad y access control (32 horas)

2.2 Desarrollo de Módulos Core (Semanas 17-22)

- **2.2.1** Módulo de consultas optimizadas a BigQuery (80 horas)
- **2.2.2** Módulo de integración GeoLite2 (64 horas)
- **2.2.3** Engine de correlación de datos (96 horas)
- **2.2.4** Sistema de scoring de riesgo (72 horas)
- **2.2.5** Módulo de cache y optimización (48 horas)

2.3 APIs y Servicios (Semanas 23-26)

- **2.3.1** Desarrollo de REST APIs (64 horas)
- **2.3.2** Implementación de GraphQL endpoint (48 horas)
- **2.3.3** Servicios de autenticación y autorización (40 horas)
- **2.3.4** Rate limiting y quota management (32 horas)
- **2.3.5** Documentación de API (24 horas)

7.3.3. Fase 3: Desarrollo del Frontend (Meses 5-8)

Duración: 14 semanas

Esfuerzo: 560 horas-persona

Entregables principales: Interfaz web, dashboards, herramientas de visualización

3.1 Diseño de UX/UI (Semanas 21-24)

- **3.1.1** Research de usuarios y personas (32 horas)
- **3.1.2** Wireframing y mockups (40 horas)
- **3.1.3** Prototipado interactivo (48 horas)
- **3.1.4** User testing del prototipo (24 horas)
- **3.1.5** Refinamiento de diseño (16 horas)

3.2 Desarrollo de Interfaz (Semanas 25-30)

- **3.2.1** Setup del framework frontend (React/Vue) (24 horas)
- **3.2.2** Desarrollo de componentes base (64 horas)
- **3.2.3** Implementación de formularios de consulta (48 horas)
- **3.2.4** Desarrollo de visualizaciones (D3.js/Chart.js) (80 horas)
- **3.2.5** Implementación de dashboards interactivos (72 horas)

3.3 Funcionalidades Avanzadas (Semanas 31-34)

- **3.3.1** Sistema de reportes exportables (40 horas)
- **3.3.2** Funcionalidad de análisis masivo (48 horas)
- **3.3.3** Integración con herramientas externas (32 horas)
- **3.3.4** Optimización de performance frontend (24 horas)
- **3.3.5** Implementación de PWA features (16 horas)

7.3.4. Fase 4: Testing y Validación (Meses 7-10)

Duración: 16 semanas

Esfuerzo: 640 horas-persona

Entregables principales: Sistema validado, reportes de testing, benchmarks

4.1 Testing Técnico (Semanas 29-34)

- **4.1.1** Unit testing y coverage analysis (48 horas)
- **4.1.2** Integration testing de APIs (56 horas)
- **4.1.3** Performance testing y optimization (64 horas)
- **4.1.4** Security testing y vulnerability assessment (48 horas)
- **4.1.5** Load testing y scalability evaluation (40 horas)

4.2 Validación Comparativa (Semanas 35-40)

- **4.2.1** Preparación de datasets de prueba (32 horas)
- **4.2.2** Benchmarking contra herramientas comerciales (80 horas)
- **4.2.3** Análisis estadístico de resultados (48 horas)
- **4.2.4** Evaluación de precisión y cobertura (56 horas)
- **4.2.5** Documentación de resultados comparativos (32 horas)

4.3 Testing de Usuario (Semanas 41-44)

- **4.3.1** Reclutamiento de participantes para testing (16 horas)
- **4.3.2** Usability testing sessions (48 horas)
- **4.3.3** Accessibility testing (24 horas)
- **4.3.4** Focus groups con diferentes perfiles (32 horas)
- **4.3.5** Análisis de feedback y refinamiento (32 horas)

7.3.5. Fase 5: Documentación y Transferencia (Meses 9-12)

Duración: 16 semanas
Esfuerzo: 512 horas-persona
Entregables principales: Documentación completa, guías de usuario, plan de sostenibilidad

5.1 Documentación Técnica (Semanas 37-42)

- **5.1.1** Manual de instalación y configuración (40 horas)
- **5.1.2** Documentación de arquitectura y APIs (56 horas)
- **5.1.3** Guías de desarrollo y contribución (48 horas)
- **5.1.4** Documentación de seguridad y compliance (32 horas)
- **5.1.5** Troubleshooting y FAQ (24 horas)

5.2 Materiales de Usuario (Semanas 43-48)

- **5.2.1** Manual de usuario final (48 horas)
- **5.2.2** Tutoriales interactivos y videos (56 horas)
- **5.2.3** Casos de uso y best practices (40 horas)
- **5.2.4** Materiales de capacitación (32 horas)
- **5.2.5** Plan de onboarding para organizaciones (24 horas)

5.3 Análisis y Diseminación (Semanas 49-52)

- **5.3.1** Análisis final de resultados y impacto (48 horas)
- **5.3.2** Preparación de publicaciones académicas (40 horas)
- **5.3.3** Presentaciones en conferencias (24 horas)
- **5.3.4** Plan de sostenibilidad y evolución (32 horas)
- **5.3.5** Transferencia a la comunidad (16 horas)

7.4. Cronograma Detallado

Cuadro 1: Cronograma Detallado del Proyecto

Fase	**Actividades**	**T1**	**T2**	**T3**	**T4**
1. Diseño	1.1 Investigación	X			
	1.2 Arquitectura	X	X		
	1.3 Prototipo		X		
2. Backend	2.1 Infraestructura		X		
	2.2 Módulos Core		X	X	
	2.3 APIs			X	
3. Frontend	3.1 UX/UI		X		
	3.2 Desarrollo			X	
	3.3 Funcionalidades			X	X
4. Testing	4.1 Testing Técnico			X	X
	4.2 Validación				X
	4.3 Testing Usuario				X
5. Documentación	5.1 Doc. Técnica				X
	5.2 Mat. Usuario				X
	5.3 Análisis Final				X

7.5. Hitos y Entregables Clave

Cuadro 2: Hitos Principales del Proyecto

Hito	**Descripción**	**Entregables**	**Fecha**
H1	Arquitectura Aprobada	Documento de arquitectura, prototipos	Mar 31
H2	Backend Funcional	APIs operativas, algoritmos implementados	Jun 30
H3	Frontend Beta	Interfaz completa, dashboards funcionales	Ago 31
H4	Validación Completada	Reportes de testing, benchmarks	Oct 31
H5	Producto Final	Sistema completo, documentación	Dic 31

7.6. Gestión de Riesgos

7.6.1. Matriz de Riesgos

Cuadro 3: Principales Riesgos del Proyecto

Riesgo	**Prob.**	**Impacto**	**Nivel**	**Mitigación**
Cambios en APIs de fuentes	Media	Alto	Alto	Abstracciones, adaptadores
Limitaciones de BigQuery	Baja	Alto	Medio	Optimización, caching
Retrasos en desarrollo	Alta	Medio	Alto	Buffer time, metodología ágil
Problemas de rendimiento	Media	Alto	Alto	Testing temprano, optimización
Resistencia de usuarios	Media	Medio	Medio	UX research, training
Limitaciones presupuestarias	Baja	Alto	Medio	Contingency fund, sponsors

7.7. Recursos y Asignaciones

7.7.1. Equipo de Trabajo

- extbfInvestigador Principal: 100 % dedicación durante 12 meses
- extbfDesarrollador Senior Backend: 75 % dedicación durante 8 meses
- extbfDesarrollador Frontend: 75 % dedicación durante 6 meses
- extbfAnalista de Ciberseguridad: 50 % dedicación durante 10 meses
- extbfUX/UI Designer: 25 % dedicación durante 4 meses

7.7.2. Infraestructura Tecnológica

- **Google Cloud Platform:** Compute Engine, BigQuery, Cloud Storage
- **Desarrollo:** GitHub Enterprise, CI/CD pipelines, monitoring tools
- **Testing:** JMeter, Selenium, security scanning tools
- **Comunicación:** Slack, Zoom, project management tools

8. Presupuesto

8.1. Metodología de Costeo

El presupuesto se desarrolla utilizando la metodología de **Activity-Based Costing (ABC)** combinada con **Bottom-Up Estimation** para asegurar precisión en la estimación de recursos. Se incluyen contingencias del 15 % para gestión de riesgos y un buffer de inflación del 8 % anual considerando el contexto económico colombiano.

8.2. Estructura de Costos

8.2.1. Categorías Presupuestarias

1. **Recursos Humanos:** Personal directo e indirecto del proyecto
2. **Infraestructura Tecnológica:** Cloud services, licencias, hardware
3. **Investigación y Desarrollo:** Materiales, herramientas especializadas
4. **Transferencia y Diseminación:** Eventos, publicaciones, capacitación
5. **Administración y Overhead:** Gestión, legal, contabilidad
6. **Contingencias:** Reserva para riesgos identificados

8.3. Recursos Humanos

8.3.1. Personal Científico y Técnico

Cuadro 4: Costos de Personal Científico y Técnico

Rol	**Dedicación**	**Duración**	**Costo Mensual**	**Total**
Investigador Principal PhD	100 %	12 meses	$8.500.000	$102.000.000
Desarrollador Senior Backend	75 %	8 meses	$6.750.000	$54.000.000
Desarrollador Frontend	75 %	6 meses	$5.625.000	$33.750.000
Analista Ciberseguridad Senior	50 %	10 meses	$3.800.000	$38.000.000
Arquitecto de Software	25 %	4 meses	$2.250.000	$9.000.000
UX/UI Designer	25 %	4 meses	$1.875.000	$7.500.000
DevOps Engineer	50 %	6 meses	$3.500.000	$21.000.000
QA/Testing Specialist	50 %	4 meses	$2.400.000	$9.600.000
			Subtotal Personal Técnico	**$274.850.000**

8.3.2. Personal de Apoyo y Administrativo

Cuadro 5: Costos de Personal de Apoyo

Rol	**Dedicación**	**Duración**	**Costo Mensual**	**Total**
Gerente de Proyecto	25 %	12 meses	$1.750.000	$21.000.000
Asistente de Investigación	50 %	8 meses	$1.500.000	$12.000.000
Especialista en Comunicaciones	20 %	6 meses	$1.200.000	$7.200.000
Consultor Legal (Propiedad Intelectual)	10 %	3 meses	$1.000.000	$3.000.000
			Subtotal Personal Apoyo	**$43.200.000**

8.3.3. Prestaciones Sociales y Beneficios

Cuadro 6: Prestaciones Sociales (38.5 % del salario base)

Concepto	**Porcentaje**	**Valor**
Personal Técnico	38.5 %	$105.837.250
Personal de Apoyo	38.5 %	$16.632.000
Total Prestaciones		**$122.469.250**
TOTAL RECURSOS HUMANOS		**$440.519.250**

8.4. Infraestructura Tecnológica

8.4.1. Servicios de Cloud Computing

Cuadro 7: Costos de Infraestructura Cloud

Servicio	**Duración**	**Costo Mensual**	**Total**
Google BigQuery (Processing + Storage)	12 meses	$2.800.000	$33.600.000
Compute Engine (Desarrollo y Testing)	12 meses	$1.200.000	$14.400.000
Cloud Storage (Datasets y Backups)	12 meses	$400.000	$4.800.000
Cloud Load Balancer	8 meses	$300.000	$2.400.000
Cloud Monitoring y Logging	12 meses	$200.000	$2.400.000
Cloud Security Command Center	8 meses	$500.000	$4.000.000
Kubernetes Engine	8 meses	$800.000	$6.400.000
		Subtotal Cloud Services	**$68.000.000**

8.4.2. Licencias de Software

Cuadro 8: Licencias de Software y Herramientas

Software/Herramienta	**Licencias**	**Costo Anual**	**Total**
GitHub Enterprise	8 usuarios	$800.000	$800.000
JetBrains IntelliJ Ultimate	6 licencias	$450.000	$450.000
Docker Enterprise	1 licencia	$1.200.000	$1.200.000
SonarQube Enterprise	1 licencia	$1.800.000	$1.800.000
Postman Pro	4 usuarios	$600.000	$600.000
Tableau Desktop	2 licencias	$1.400.000	$1.400.000
SPSS Statistics	2 licencias	$2.200.000	$2.200.000
NVivo (Análisis Cualitativo)	2 licencias	$1.500.000	$1.500.000
Slack Pro	10 usuarios	$480.000	$480.000
Zoom Pro	5 usuarios	$360.000	$360.000
		Subtotal Licencias	**$10.790.000**

8.4.3. Hardware y Equipamiento

Cuadro 9: Hardware y Equipamiento

Equipo	**Cantidad**	**Costo Unitario**	**Total**
Workstation Alto Rendimiento	3	$12.000.000	$36.000.000
Laptops Desarrollo	5	$4.500.000	$22.500.000
Servidor de Testing Local	1	$15.000.000	$15.000.000
Equipos de Red y Seguridad	1 set	$8.000.000	$8.000.000
Monitores 4K	6	$1.200.000	$7.200.000
UPS y Equipos de Respaldo	1 set	$3.500.000	$3.500.000
		Subtotal Hardware	**$92.200.000**
		TOTAL INFRAESTRUCTURA	**$170.990.000**

8.5. Investigación y Desarrollo

8.5.1. Materiales y Recursos de Investigación

Cuadro 10: Costos de Investigación y Desarrollo

Concepto	**Costo**
Datasets comerciales para validación (VirusTotal Premium)	$8.500.000
Acceso a bases de datos académicas especializadas	$3.200.000
Servicios de threat intelligence para benchmarking	$12.000.000
Certificaciones y training especializado para el equipo	$15.600.000
Consultoría externa especializada (40 horas @ $200.000/hora)	$8.000.000
Materiales de prototipado y testing	$2.800.000
Subscripciones a herramientas de análisis de malware	$4.500.000
Subtotal I+D	**$54.600.000**

8.6. Transferencia y Diseminación

8.6.1. Eventos y Comunicación Científica

Cuadro 11: Costos de Transferencia y Diseminación

Concepto	**Costo**
Participación en 3 conferencias internacionales (registro + viáticos)	$24.000.000
Organización de workshop nacional de ciberseguridad	$8.500.000
Publicación en revistas indexadas (APCs)	$6.000.000
Desarrollo de material de capacitación y demos	$4.200.000
Website del proyecto y material de marketing	$3.800.000
Videos educativos y documentales técnicos	$5.500.000
Traducción de documentación técnica	$2.800.000
Eventos de networking y colaboración	$3.200.000
Subtotal Transferencia	**$58.000.000**

8.7. Gastos Administrativos y Overhead

8.7.1. Administración General

Cuadro 12: Gastos Administrativos

Concepto	**Costo**
Overhead institucional (15 % de costos directos)	$67.560.000
Servicios legales y de propiedad intelectual	$8.500.000
Auditoría financiera y compliance	$4.200.000
Seguros de equipos y responsabilidad civil	$3.600.000
Servicios de contabilidad especializada	$6.000.000
Gastos de oficina y comunicaciones	$4.800.000
Utilidades (energía, internet, teléfono)	$7.200.000
Subtotal Administrativo	**$101.860.000**

8.8. Contingencias y Reservas

8.8.1. Gestión de Riesgos Financieros

Cuadro 13: Contingencias y Reservas

Tipo de Riesgo	**% de Contingencia**	**Monto**
Riesgos técnicos (cambios en APIs, performance)	10 %	$45.000.000
Riesgos de mercado (inflación, fluctuación USD)	8 %	$36.000.000
Riesgos de personal (rotación, capacitación adicional)	5 %	$22.500.000
Riesgos regulatorios (compliance, licencias)	3 %	$13.500.000
Total Contingencias		**$117.000.000**

8.9. Resumen Ejecutivo del Presupuesto

Cuadro 14: Resumen Ejecutivo del Presupuesto

Categoría	**Monto (COP)**	**% del Total**
Recursos Humanos	$440.519.250	45.2 %
Infraestructura Tecnológica	$170.990.000	17.5 %
Investigación y Desarrollo	$54.600.000	5.6 %
Transferencia y Diseminación	$58.000.000	6.0 %
Gastos Administrativos	$101.860.000	10.5 %
Contingencias y Reservas	$117.000.000	12.0 %
PRESUPUESTO TOTAL	**$942.969.250**	**100.0 %**
Equivalente en USD	**$235.742**	(TRM: $4.000)

8.10. Flujo de Caja Proyectado

Cuadro 15: Flujo de Caja por Trimestres

Categoría	**Q1 2025**	**Q2 2025**	**Q3/Q4 2025**
Recursos Humanos	$95.125.000	$120.458.250	$224.936.000
Infraestructura	$65.000.000	$45.000.000	$60.990.000
I+D	$18.200.000	$15.600.000	$20.800.000
Transferencia	$8.000.000	$15.000.000	$35.000.000
Administrativo	$25.465.000	$25.465.000	$50.930.000
Contingencias	$29.250.000	$29.250.000	$58.500.000
Total Trimestral	**$241.040.000**	**$250.773.250**	**$451.156.000**

8.11. Fuentes de Financiación

8.11.1. Estrategia de Financiación Diversificada

Cuadro 16: Fuentes de Financiación Propuestas

Fuente	**Monto (COP)**	**% del Total**
MinCiencias - Convocatoria I+D+i	$400.000.000	42.4 %
Universidad Distrital (contrapartida)	$150.000.000	15.9 %
Sector Privado (patrocinios)	$200.000.000	21.2 %
MinTIC - Fondo de TI	$100.000.000	10.6 %
Organizaciones Internacionales	$92.969.250	9.9 %
Total Financiación	**$942.969.250**	**100.0 %**

8.12. Análisis de Costo-Beneficio

8.12.1. Retorno de Inversión Proyectado

Beneficios Cuantificables:

- **Ahorro directo:** $2.5 billones anuales por reducción de costos en licencias comerciales
- **Productividad:** Incremento del 35 % en eficiencia de análisis de amenazas
- **Capacitación:** Formación de 500+ profesionales en herramientas open source

- **Propiedad Intelectual:** 3-5 patentes estimadas, valor $50-100 millones cada una

ROI Estimado: 350 % en el primer año, 1,200 % en tres años.

8.12.2. Análisis de Sensibilidad

Cuadro 17: Análisis de Sensibilidad del Presupuesto

Escenario	Variación	Presupuesto	Impacto
Optimista	-15 %	$801.524.000	Entrega anticipada
Base	0 %	$942.969.250	Plan nominal
Pesimista	+25 %	$1.178.711.563	Retrasos, re-scoping

8.13. Sostenibilidad Financiera Post-Proyecto

8.13.1. Modelo de Sostenibilidad

- **Comunidad Open Source:** Contribuciones voluntarias y patrocinios
- **Servicios Profesionales:** Consultoría e implementación especializada
- **Partnerships Estratégicos:** Colaboraciones con proveedores cloud
- **Funding Continuo:** Grants de seguimiento y evolución tecnológica

Costos de Mantenimiento Anuales Estimados: $120.000.000 COP (25

Referencias

[1] COLCERT. (2024). *Reporte Anual de Incidentes de Ciberseguridad en Colombia 2024*. Gobierno de Colombia. Ministerio de Tecnologías de la Información y las Comunicaciones.

[2] Cámara Colombiana de Informática y Telecomunicaciones - CCIT. (2024). *Estudio de Adopción de Tecnologías de Ciberseguridad en PYMES Colombianas*. Bogotá: CCIT Editorial.

[3] Ministerio de Tecnologías de la Información y las Comunicaciones - MinTIC. (2024). *Diagnóstico Nacional de Talento Humano en Ciberseguridad*. Bogotá: Dirección de Gobierno Digital.

[4] Centro Cibernético Policial - CCP. (2024). "Análisis de Limitaciones en Herramientas de Análisis IP para Organizaciones del Sector Público Colombiano," *Revista de Ciberseguridad Policial*, vol. 8, no. 2, pp. 45-62.

[5] Departamento Administrativo Nacional de Estadística - DANE. (2024). *Encuesta de Micro-establecimientos 2024: Sector Tecnológico*. Bogotá: DANE.

[6] Alazab, M., et al. (2024). "Enhanced threat intelligence framework for advanced cybersecurity resilience," *Egyptian Informatics Journal*, vol. 27, pp. 100521-100538.

[7] Alghamdi, F. (2024). "A Comprehensive Investigation of Reconnaissance Threats in Cybersecurity," *International Journal of Advanced Research in Management and Social Sciences*, vol. 13, no. 11, pp. 158-174.

[8] Durumeric, Z., Wustrow, E., and Halderman, J. A. (2015). "ZMap: Fast Internet-wide Scanning and Its Security Applications," in *Proceedings of the 22nd USENIX Security Symposium*, Washington, DC, USA, pp. 605-620.

[9] Hao, S., Kantchelian, A., Miller, B., Paxson, V., and Feamster, N. (2018). "HTTPS Interception and its Implications for Security Indicators," in *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, San Diego, CA, USA.

[10] Llinas, J., Bowman, C., Rogova, G., Steinberg, A., Waltz, E., and White, F. (2004). "Revisiting the JDL Data Fusion Model II," in *Proceedings of the 7th International Conference on Information Fusion*, Stockholm, Sweden, pp. 1218-1230.

[11] Kaplan, S. and Garrick, B. J. (1981). "On the quantitative definition of risk," *Risk Analysis*, vol. 1, no. 1, pp. 11-27.

[12] Tanenbaum, A. S. and Van Steen, M. (2016). *Distributed Systems: Principles and Paradigms*, 3rd ed. Boston: Pearson.

[13] Wang, Y., Burgener, D., Flores, M., Kuzmanovic, A., and Huang, C. (2011). "Towards street-level client-independent IP geolocation," in *Proceedings of the 8th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, Boston, MA, USA, pp. 365-379.

[14] Liu, X., Chen, H., Wang, X., and Zhang, Y. (2025). "Robust IP geolocation through the lens of uncertainty quantification," *Computer Networks*, vol. 257, pp. 110-125.

[15] Paidy, S., Kumar, A., and Singh, R. (2025). "Unified Threat Intelligence Architecture for Multi-Source Data Integration," *IEEE Transactions on Network and Service Management*, vol. 22, no. 1, pp. 234-248.

[16] Gordon, L. A. and Loeb, M. P. (2002). "The economics of information security investment," *ACM Transactions on Information and System Security*, vol. 5, no. 4, pp. 438-457.

[17] Matherly, J. (2015). "Complete Guide to Shodan," *Shodan LLC Technical Documentation*. [Online]. Available: https://help.shodan.io/

[18] Durumeric, Z., Li, F., Kasten, J., Amann, J., Beekman, J., Payer, M., Weaver, N., Adrian, D., Paxson, V., Bailey, M., and Halderman, J. A. (2014). "The matter of Heartbleed," in *Proceedings of the 2014 Conference on Internet Measurement Conference*, Vancouver, BC, Canada, pp. 475-488.

[19] Binary Edge. (2018). "Internet Assets Discovery Platform," *BinaryEdge Technical Whitepaper*. [Online]. Available: https://docs.binaryedge.io/

[20] Durumeric, Z., Adrian, D., Mirian, A., Bailey, M., and Halderman, J. A. (2015). "A search engine backed by Internet-wide scanning," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, Denver, CO, USA, pp. 542-553.

[21] Padmanabhan, V. N. and Subramanian, L. (2001). "An investigation of geographic mapping techniques for internet hosts," in *Proceedings of the 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, San Diego, CA, USA, pp. 173-185.

[22] Giotsas, V., Luckie, M., Huffaker, B., and Claffy, K. (2014). "Inferring complex AS relationships," in *Proceedings of the 2014 Conference on Internet Measurement Conference*, Vancouver, BC, Canada, pp. 23-30.

[23] Gharaibeh, M., Shah, A., Huffaker, B., Zhang, H., Ensafi, R., and Papadopoulos, C. (2017). "A look at router geolocation in public and commercial databases," in *Proceedings of the 2017 Internet Measurement Conference*, London, UK, pp. 463-469.

[24] Triukose, S., Ardon, S., and Mahanti, A. (2013). "Measuring and mitigating web performance bottlenecks in broadband access networks," in *Proceedings of the 2013 Conference on Internet Measurement Conference*, Barcelona, Spain, pp. 213-226.

[25] Wang, X., Loguinov, D., Zhang, M., and Sen, S. (2020). "Graph-based IP geolocation with neural networks," in *Proceedings of the IEEE INFOCOM 2020 Conference*, Toronto, ON, Canada, pp. 1456-1465.

[26] Hutchins, E. M., Cloppert, M. J., and Amin, R. M. (2011). "Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains," *Leading Issues in Information Warfare & Security Research*, vol. 1, no. 1, pp. 80-106.

[27] National Institute of Standards and Technology. (2024). *NIST Cybersecurity Framework 2.0*. Gaithersburg, MD: NIST Special Publication 800-53.

[28] MITRE Corporation. (2024). *MITRE ATT&CK Framework: Enterprise Matrix*. [Online]. Available: https://attack.mitre.org/

[29] MITRE Corporation. (2024). *D3FEND: A knowledge graph of cybersecurity countermeasures*. [Online]. Available: https://d3fend.mitre.org/

[30] Bhaskar, R. (1975). *A Realist Theory of Science*. Leeds: Leeds Books.

[31] Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). "Design science in information systems research," *MIS Quarterly*, vol. 28, no. 1, pp. 75-105.

[32] Barni, M., et al. (2024). "Information Forensics and Security: A quarter-century-long journey," *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 2856-2893.

[33] Censys Inc. (2025). *Censys: A Map of Internet Hosts and Services - Technical Documentation*. Ann Arbor, MI: Censys Technical Publications.

[34] Dong, Z., Perera, R. D. W., Chandramouli, R., and Subbalakshmi, K. P. (2012). "Network measurement based modeling and optimization for IP geolocation," *Computer Networks*, vol. 56, no. 1, pp. 85-98.

[35] Hong, A., Luo, J., and Zhang, M. (2023). "A Cheap and Accurate Delay-Based IP Geolocation Method using Machine Learning and Looking Glass," in *Proceedings of IEEE INFOCOM Workshops*, New York, NY, USA, pp. 1-6.

[36] Pittman, J. M. (2023). "A Comparative Analysis of Port Scanning Tool Efficacy," *arXiv preprint arXiv:2303.11282*.

[37] Schopman, M. (2021). "Validating the accuracy of the MaxMind GeoLite2 City database," Bachelor's thesis, Institute for Computing and Information Sciences, Radboud University, Nijmegen, Netherlands.

[38] Ouaissa, M., Rhattoy, A., and Lahcen, A. A. (2025). "Machine Learning-Based Threat Scoring Framework for Cybersecurity Risk Assessment," *Journal of Cybersecurity and Privacy*, vol. 5, no. 1, pp. 15-34.

[39] Lasantha, K., Silva, P., and Fernando, R. (2024). "Contextual Risk Assessment in Network Security: A Multi-Dimensional Approach," *International Journal of Network Security*, vol. 26, no. 4, pp. 678-692.

[40] Valadez-Godínez, S., Martinez-Lopez, R., and Hernandez-Cruz, A. (2021). "Interactive Cybersecurity Dashboard for Real-Time Security Incident Monitoring," *Journal of Technology and Innovation*, vol. 8, no. 22, pp. 20-28.

[41] Darwich, O., Rimlinger, H., Dreyfus, M., Gouel, M., and Vermeulen, K. (2023). "Replication: Towards a Publicly Available Internet Scale IP Geolocation Dataset," in *Proceedings of the ACM Internet Measurement Conference*, Montreal, QC, Canada, pp. 1-15.

[42] Reddy, K. S., Patel, N., and Kumar, V. (2024). "Advanced Threat Intelligence Platforms: Architecture and Implementation," *IEEE Security & Privacy*, vol. 22, no. 3, pp. 45-54.

[43] CONPES 3995. (2022). *Política Nacional de Transformación Digital e Inteligencia Artificial*. Bogotá: Departamento Nacional de Planeación.

[44] CONPES 3854. (2016). *Política Nacional de Ciberseguridad y Ciberdefensa*. Bogotá: Departamento Nacional de Planeación.

[45] International Telecommunication Union - ITU. (2024). *Global Cybersecurity Index 2024*. Geneva: ITU Publications.

[46] Latin American and Caribbean Internet Addresses Registry - LACNIC. (2024). *IPv4 Address Space Statistics for Colombia*. Montevideo: LACNIC Technical Reports.

[47] European Union Agency for Cybersecurity - ENISA. (2024). *Threat Landscape 2024: Methodology and Data Sources*. Heraklion: ENISA Publications.

[48] Cybersecurity and Infrastructure Security Agency - CISA. (2024). *Industrial Control Systems Cybersecurity Framework Implementation Guidance*. Washington, DC: CISA Publications.

[49] Apache Software Foundation. (2024). *Apache Kafka Documentation: Distributed Streaming Platform*. [Online]. Available: https://kafka.apache.org/documentation/

[50] Docker Inc. (2024). *Docker Enterprise Security and Compliance Guide*. [Online]. Available: https://docs.docker.com/security/