Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales
Departamento de Computación

Implementación de una solución de seguridad de correo electrónico en el Sector Público Argentino

Tesis de Licenciatura en Ciencias de la Computación

Tomás Santiago Scally

Director: Dr. Diego Garbervetsky (diegog@dc.uba.ar)

Buenos Aires, 2025

IMPLEMENTACIÓN DE SOLUCIONES DE SEGURIDAD DE LA INFORMACIÓN EN EL SECTOR PÚBLICO ARGENTINO

El Estado Argentino emplea soluciones informáticas especializadas para salvaguardar sus activos digitales. El proceso estándar de adquisición sigue una metodología estructurada que incluye: identificación de necesidades, análisis comparativo de productos disponibles, elaboración de pliegos de especificaciones técnicas, publicación de licitaciones públicas e implementación de la solución seleccionada. Este procedimiento técnico-administrativo presenta una duración considerablemente extensa —frecuentemente varios años— desde la fase inicial hasta la puesta en funcionamiento efectiva de la solución. Esta demora se atribuye a múltiples instancias de revisión y control que involucran diversos organismos gubernamentales, generando un sistema de verificaciones cruzadas que, si bien aporta transparencia, impacta significativamente en los tiempos de ejecución. El presente trabajo analiza los principales obstáculos y desafíos que surgen durante el ciclo completo de implementación de proyectos de seguridad informática, tomando como caso de estudio específico la implementación de una solución antispam en un importante organismo nacional. Se examina desde la confección del pliego de especificaciones técnicas —que debe contemplar múltiples situaciones particulares como plazos de resolución de problemas, compatibilidad con sistemas existentes, definición de hitos de avance, criterios de aceptación y finalización del proyecto— hasta su puesta en funcionamiento. Simultáneamente, se proponen alternativas para acelerar los procesos de adquisición o incluso, evitar completamente estos procedimientos tradicionales.

Palabras claves: Seguridad de la información, Antispam, Administración Publica.

AGRADECIMIENTOS

A la educación pública, gratuita y de calidad.
A la educación pública, gratuita y de calidad.
A la educación pública, gratuita y de calidad.

A mis docentes, su compromiso fue excepcional. A Diego Garbervetsky, por obligarme a hacer esta tesis y guiarme. Al jurado, Pablo De Cristoforis y Pablo Brusco por su dedicación contrarreloj. A Matías López, por tantas horas de estudio y TPs realizados. A mis compañeros de cursada, fueron parte fundamental de esto. A Mari, por incentivarme y ser la persona más amada por mis hijos. A Eve, por el aliento.

A Felipe y Simón, que son mi vida.

Índice general

1.. Glosario 1

2.. Introducción 3

3.. Seguridad de la información en el Estado Nacional 5
3.1. ¿Qué es la seguridad de la información? 5
3.1.1. Alcance y Dimensiones 6
3.1.2. Estándares 6
3.1.3. Ingeniería de Seguridad 6
3.1.4. Gestión de Riesgos 7
3.1.5. Evolución Continua 7
3.2. Marcos Normativos para la Seguridad de la información en los organismo públicos 8
3.3. Desafíos en la adopción de seguridad en las organizaciones 9
3.4. Breve reseña de ataques informáticos 10
3.5. Cómo el Estado Argentino adquiere soluciones informáticas 11

4.. Implementación de un sistema antispam con sandboxing en un organismo público 15
4.1. Contexto y Justificación 15
4.1.1. El correo electrónico como vector de ataque predominante 15
4.1.2. Particularidades del sector público 15
4.2. Problemática Identificada 16
4.2.1. Limitaciones técnicas de las soluciones tradicionales 16
4.2.2. Obsolescencia de la infraestructura existente 18
4.3. Solución Propuesta 19
4.3.1. Implementación de Sandboxing 19
4.3.2. Arquitectura de Alta Disponibilidad 23
4.3.3. Diagrama de Flujo de Correo Electrónico 26
4.3.4. Desafíos de Implementación y Estrategia de Migración 28
4.3.5. Análisis detallado del riesgo de pérdida de conocimiento acumulado 28
4.3.6. Estrategia de migración gradual: diseño y justificación 29
4.3.7. Fases de implementación detalladas 31
4.4. Métricas y Resultados 37
4.4.1. Métricas del proceso de migración 37
4.4.2. Métricas de efectividad de detección 38

5.. Lecciones Aprendidas 41
5.1. Valor de las estrategias de migración gradual 41
5.2. Importancia de la documentación del conocimiento tácito 42
5.3. Necesidad de arquitecturas resilientes 42
5.4. Complementariedad de técnicas de detección 43
5.5. Replicabilidad y transferencia de conocimiento 44
5.6. Evolución continua de la seguridad 44

5.7. Síntesis de aprendizajes clave 45
5.8. Dos alternativas a la contratación fragmentada de soluciones comerciales . . 45
5.8.1. Unificación de una solución de antispam para toda la APN 45
5.8.2. Construcción de un antispam nacional 47

6.. Conclusiones 49
6.1. Viabilidad de actualización de infraestructuras críticas 49
6.2. Valor de la estrategia de migración gradual 49
6.3. Importancia del enfoque de defensa en profundidad 50
6.4. Replicabilidad del modelo en otras organizaciones 51
6.5. Evolución continua de la seguridad 51
6.6. Preservación y mejora del conocimiento organizacional 52
6.7. Síntesis: Un modelo replicable de transformación responsable 53

1. GLOSARIO

Antispam: Sistema de seguridad diseñado para detectar, filtrar y bloquear correos electrónicos no deseados, maliciosos o fraudulentos antes de que lleguen a la bandeja de entrada del usuario.

Vector de ataque: Método o ruta utilizada por los cibercriminales para acceder a sistemas informáticos, comprometer datos o ejecutar código malicioso. El correo electrónico es uno de los vectores más comunes.

Malware zero-day: Software malicioso que aprovecha vulnerabilidades desconocidas o recientemente descubiertas, para las cuales aún no existen parches de seguridad ni firmas de detección en las bases de datos.

Hash: Función matemática que convierte datos de cualquier tamaño en una cadena de caracteres de longitud fija. Se utiliza para identificar de manera única archivos y detectar modificaciones.

Algoritmos heurísticos: Técnicas de detección que analizan el comportamiento y características de archivos o programas para identificar posibles amenazas, incluso sin tener firmas específicas.

Sandboxing: Tecnología que ejecuta archivos o programas en un entorno aislado y controlado para analizar su comportamiento sin riesgo para el sistema principal.

Nodos de análisis: Componentes del sistema encargados de examinar y procesar los correos electrónicos antes de determinar su legitimidad.

Veredicto: Decisión automatizada del sistema sobre si un correo es seguro (permitir entrega) o malicioso (bloquear o cuarentenar).

Alta disponibilidad: Característica de un sistema que garantiza un funcionamiento continuo y confiable, minimizando los tiempos de inactividad mediante redundancia y tolerancia a fallos.

Nodo: Cada uno de los servidores o componentes individuales que forman parte de una arquitectura distribuida de sistemas.

Tráfico de correo: Volumen total de mensajes de correo electrónico que fluye a través de la infraestructura de comunicaciones de una organización.

IOC (Indicator of Compromise - Indicador de Compromiso): Evidencia observable que indica que un sistema ha sido comprometido o atacado, como direcciones IP maliciosas, hashes de archivos sospechosos, URLs de comando y control, o patrones de tráfico anómalos. Estos indicadores permiten detectar, investigar y prevenir incidentes de seguridad.

APN: Administración Publica Nacional

2. INTRODUCCIÓN

El correo electrónico constituye una de las herramientas de comunicación más consolidadas y ampliamente utilizadas en las organizaciones contemporáneas, habiendo evolucionado desde su concepción en la década de 1970 hasta convertirse en un pilar fundamental de la comunicación institucional. El Estado Nacional también ha experimentado una dependencia creciente de esta tecnología para el desarrollo de sus funciones administrativas, la coordinación interinstitucional y la comunicación con la ciudadanía. Sin embargo, esta omnipresencia del correo electrónico presenta un panorama complejo en términos de ciberseguridad. Los sistemas de correo electrónico representan uno de los vectores de ataque más efectivos y frecuentemente explotados por los delincuentes informáticos [1] [2], quienes aprovechan vulnerabilidades técnicas, así como factores humanos mediante técnicas de ingeniería social. Las amenazas incluyen desde ataques de phishing y malware hasta campañas de ransomware y espionaje cibernético dirigido, fenómenos que han experimentado un crecimiento exponencial en los últimos años. En este contexto, la protección integral de los servidores de correo electrónico y de las comunicaciones que fluyen a través de ellos se presenta como una prioridad estratégica ineludible para cualquier organización estatal. La criticidad de esta necesidad se acentúa considerando que las instituciones públicas manejan información sensible de carácter ciudadano, estratégico y operacional, cuyo compromiso podría tener consecuencias que trascienden el ámbito organizacional. El presente trabajo tiene como propósito examinar de manera integral los desafíos, procesos y oportunidades asociados a la implementación de sistemas de seguridad para correo electrónico en el ámbito del Estado Nacional. Específicamente, nos proponemos:

- Proporcionar un marco de referencia basado en experiencias prácticas para futuras implementaciones
- Identificar y analizar las principales oportunidades de mejora en los procesos actuales de protección del correo electrónico
- Describir estrategias alternativas e innovadoras que puedan dar respuesta más eficiente a estas necesidades de seguridad

Iniciaremos con una descripción comprehensiva de lo que actualmente se entiende por seguridad de la información en el contexto organizacional moderno, estableciendo los marcos teóricos y normativos que fundamentan las mejores prácticas en la materia. Detallaremos la experiencia concreta de implementación de un sistema integral de seguridad para la protección del correo electrónico en una organización de gran envergadura del Estado Nacional. Esta sección abarcará un análisis pormenorizado de todas las etapas del proceso, desde la identificación inicial de la necesidad hasta la finalización exitosa de la puesta en funcionamiento del sistema. Finalmente, presentaremos un conjunto de alternativas al proceso actual, fundamentadas en las lecciones aprendidas del caso de estudio, con el objetivo de contribuir a la optimización y eficiencia de todo el circuito técnico-administrativo involucrado en este tipo de implementaciones. A través de este enfoque integral, buscamos proporcionar no solo una descripción de la situación actual, sino también elementos propositivos que puedan contribuir al fortalecimiento de la ciberseguridad en el ámbito estatal y a la optimización de los recursos públicos destinados a este fin estratégico.

3. SEGURIDAD DE LA INFORMACIÓN EN EL ESTADO NACIONAL

3.1. ¿Qué es la seguridad de la información?

La seguridad de la información se fundamenta en tres pilares esenciales conocidos como la "tríada CIA"(por sus siglas en inglés: Confidentiality, Integrity, Availability), que constituyen los objetivos fundamentales que toda estrategia de seguridad debe cumplir. Estos principios interdependientes forman la base conceptual sobre la cual se construyen todas las políticas, procedimientos y controles de seguridad organizacional.

Confidencialidad

La confidencialidad garantiza que la información sensible permanezca protegida contra accesos no autorizados o divulgación indebida, asegurando que solo las personas explícitamente autorizadas puedan acceder a datos específicos. Este principio se implementa mediante múltiples mecanismos de control como sistemas de autenticación robustos (contraseñas fuertes, autenticación multifactor, biometría), cifrado de datos tanto en reposo como en tránsito, control de acceso basado en roles (RBAC), y clasificación de información según niveles de sensibilidad. La violación de la confidencialidad puede resultar en pérdida de ventaja competitiva, daños reputacionales severos, sanciones regulatorias y exposición de información personal que podría derivar en robo de identidad o espionaje industrial.

Integridad

La integridad asegura la exactitud, completitud y confiabilidad de la información durante todo su ciclo de vida, garantizando que los datos no sean alterados de manera no autorizada, ya sea accidental o maliciosamente. Este principio se mantiene mediante controles como sumas de verificación (checksums), firmas digitales, funciones hash criptográficas, control de versiones, pistas de auditoría detalladas, y procesos de validación de entrada. La integridad es crítica en contextos donde la precisión de los datos es fundamental, como transacciones financieras, registros médicos, o sistemas de control industrial. Su compromiso puede llevar a decisiones basadas en información incorrecta, pérdidas financieras, o incluso riesgos para la seguridad física en sistemas críticos.

Disponibilidad

La disponibilidad garantiza que los sistemas, servicios e información estén operativos y accesibles cuando los usuarios autorizados los requieran, manteniendo niveles de servicio acordados y minimizando interrupciones. Se logra mediante redundancia de sistemas, balanceo de carga, planes de continuidad del negocio, respaldos regulares, sistemas de recuperación ante desastres, protección contra ataques DDoS, y mantenimiento preventivo de infraestructura. La pérdida de disponibilidad puede paralizar operaciones críticas, generar pérdidas financieras significativas por tiempo de inactividad, afectar la satisfacción del cliente y dañar la reputación organizacional. En sectores como salud o servicios de

emergencia, la indisponibilidad puede tener consecuencias vitales.

Estos tres principios deben equilibrarse cuidadosamente, ya que enfocarse excesivamente en uno puede comprometer los otros, requiriendo un enfoque holístico y balanceado en la implementación de controles de seguridad.

3.1.1. Alcance y Dimensiones

La seguridad de la información abarca múltiples dimensiones que van más allá de la tecnología. Incluye aspectos físicos como la protección de instalaciones y equipos, elementos humanos como la capacitación del personal y la gestión de privilegios de acceso, aspectos tecnológicos como firewalls y sistemas de detección de intrusiones, y componentes organizacionales como políticas, procedimientos y marcos de gobierno. Esta disciplina no se limita únicamente a la información digital, sino que también protege documentos físicos, conversaciones verbales y cualquier forma en que la información pueda existir o transmitirse dentro de una organización.

3.1.2. Estándares

La seguridad de la información se estructura mediante marcos de referencia internacionales que proporcionan metodologías probadas y mejores prácticas. ISO 27001 establece un Sistema de Gestión de Seguridad de la Información (SGSI) completo, definiendo requisitos para implementar, mantener y mejorar continuamente la seguridad mediante un enfoque basado en procesos y mejora continua. El NIST Cybersecurity Framework ofrece un lenguaje común para gestionar riesgos de ciberseguridad, organizándose en cinco funciones principales: Identificar, Proteger, Detectar, Responder y Recuperar. COBIT proporciona un marco integral para el gobierno y gestión de TI empresarial, alineando objetivos de seguridad con metas de negocio, estableciendo métricas de desempeño y asegurando el cumplimiento regulatorio mediante principios de gobernanza efectiva.

3.1.3. Ingeniería de Seguridad

La seguridad informática constituye el núcleo técnico y operativo de la seguridad de la información, enfocándose específicamente en la protección de sistemas computacionales, redes y datos digitales mediante la implementación de controles técnicos y soluciones especializadas que conforman una arquitectura de defensa multicapa.

Si bien las organizaciones tradicionalmente comienzan con soluciones fundamentales como firewalls, antivirus y filtros antispam, el panorama actual de amenazas exige una estrategia de seguridad más sofisticada y comprehensiva. Las arquitecturas modernas de seguridad incorporan múltiples capas de protección especializadas: sistemas de mitigación contra ataques de denegación de servicio distribuido (DDoS) que protegen la disponibilidad de servicios críticos; sistemas de prevención de intrusos (IPS) que bloquean amenazas en tiempo real; plataformas de gestión de información y eventos de seguridad (SIEM) que correlacionan y analizan eventos de múltiples fuentes para detectar patrones anómalos; herramientas de análisis estático y dinámico de código (SAST/DAST) que identifican vulnerabilidades durante el ciclo de desarrollo; soluciones de prevención de pérdida de datos (DLP) que controlan el flujo de información sensible; firewalls de aplicaciones web

(WAF) que protegen contra ataques específicos a aplicaciones; y sistemas de gestión de acceso privilegiado (PAM) que aseguran el control sobre cuentas críticas.

La implementación efectiva de este ecosistema de seguridad presenta desafíos técnicos y organizacionales significativos. Los equipos especializados deben no solo deben dominar cada tecnología individualmente, sino también lograr su integración armoniosa con la infraestructura existente, garantizando interoperabilidad sin degradar el rendimiento o la experiencia del usuario. Este proceso requiere una planificación meticulosa, considerando aspectos como la compatibilidad entre soluciones, la gestión centralizada de políticas, la correlación de eventos entre plataformas heterogéneas y la optimización de recursos computacionales.

Además, la orquestación de estas herramientas demanda el desarrollo de procesos operativos maduros, capacitación continua del personal técnico y establecimiento de métricas claras para medir la efectividad de cada control implementado, asegurando así que la inversión en seguridad genere valor tangible para la organización.

3.1.4. Gestión de Riesgos

La gestión de riesgos es un proceso sistemático fundamental para proteger los activos organizacionales mediante la anticipación y respuesta efectiva a amenazas potenciales. La identificación de riesgos analiza exhaustivamente el panorama de amenazas, desde ciberataques sofisticados (ransomware o amenazas persistentes avanzadas) hasta desastres naturales, considerando tanto riesgos externos como internos (errores humanos, amenazas internas, vulnerabilidades en la cadena de suministro). La evaluación de vulnerabilidades examina debilidades en sistemas tecnológicos y procesos operativos mediante auditorías de seguridad, pruebas de penetración y análisis de configuraciones, identificando tanto vulnerabilidades técnicas como organizacionales. El análisis de impacto determina las consecuencias potenciales de incidentes de seguridad, evaluando pérdidas financieras, daños reputacionales, implicaciones legales y afectación a la continuidad del negocio mediante métricas cuantitativas y cualitativas. La implementación de controles adopta un enfoque de defensa en profundidad: controles preventivos (firewalls, autenticación robusta) que evitan incidentes; detectivos (monitoreo continuo) que identifican amenazas en tiempo real; y correctivos (planes de respuesta, backups) que facilitan la recuperación. El proceso incorpora frameworks reconocidos como ISO 31000 o NIST RMF, estableciendo criterios claros de aceptación de riesgos alineados con objetivos de negocio. La monitorización continua garantiza efectividad mediante actualización regular de evaluaciones, revisión de controles y análisis de lecciones aprendidas. Desarrollar una cultura de gestión de riesgos involucra todos los niveles organizacionales, promoviendo concienciación y responsabilidad compartida. La comunicación efectiva mediante reportes claros y métricas medibles asegura recursos necesarios para mantener un programa robusto que reduzca riesgos a niveles aceptables.

3.1.5. Evolución Continua

En el contexto actual, la seguridad de la información debe adaptarse constantemente a nuevas amenazas como el ransomware, la ingeniería social sofisticada y los ataques a la cadena de suministro, así como a nuevas tecnologías como la computación en la nube, Internet de las Cosas (IoT) y la inteligencia artificial. La seguridad de la información es, fundamentalmente, un proceso continuo de protección que integra aspectos estratégicos

(gestión de riesgos, marcos normativos) con aspectos tácticos-operativos (soluciones técnicas de seguridad informática), requiriendo vigilancia constante, actualización regular y una comprensión profunda tanto de los activos que se protegen como del panorama de amenazas en constante evolución.

3.2. Marcos Normativos para la Seguridad de la información en los organismo públicos

La gestión de la seguridad en entornos de sistemas constituye un desafío multidimensional que requiere la coordinación armoniosa de tareas técnicas, administrativas y organizacionales, cada una demandando competencias especializadas y enfoques diferenciados. Para abordar esta complejidad, la comunidad internacional ha desarrollado marcos normativos estandarizados que proporcionan metodologías estructuradas para la clasificación, organización e implementación sistemática de actividades de seguridad. Estos marcos normativos se articulan mediante catálogos comprehensivos de controles de seguridad que las organizaciones deben implementar según su contexto y nivel de madurez. El proceso de certificación requiere el cumplimiento integral de los requisitos establecidos o, alternativamente, la documentación formal de justificaciones técnicas o de negocio válidas para cualquier excepción identificada. Esta flexibilidad permite a las organizaciones adaptar los estándares a sus realidades operativas mientras mantienen la rigurosidad del marco.

Aunque no existe obligatoriedad para las organizaciones del Estado Nacional de certificación bajo estándares internacionales, varias entidades gubernamentales han adoptado voluntariamente estas certificaciones como estrategia proactiva para fortalecer su postura de seguridad y demostrar compromiso con las mejores prácticas internacionales. [3] [4] La Jefatura de Gabinete de Ministros establece un requisito fundamental: todas las organizaciones estatales deben desarrollar e implementar una política de seguridad de la información. Para facilitar este proceso, proporciona una política marco que sirve como punto de partida, invitando a las organizaciones a extenderla y personalizarla según sus características operativas, criticidad de activos y perfil de riesgo específico.[5] Es crucial reconocer que la política de seguridad nacional constituye únicamente un marco conceptual de alto nivel. Su aplicación directa, sin adaptación contextual, resulta insuficiente e inadecuada para garantizar niveles de seguridad efectivos. Esta limitación se manifiesta particularmente en la ausencia de especificaciones técnicas detalladas y controles operativos concretos. El marco se concentra predominantemente en definir qué aspectos deben controlarse, pero carece de directrices específicas sobre cómo implementar dichos controles. Esta brecha entre política y práctica genera desafíos significativos para los equipos técnicos responsables de la implementación. Por dar un ejemplo, en lo que refiere a *Seguridad de las comunicaciones* establece lo siguiente:

> ” El organismo adopta las medidas necesarias para proteger adecuadamente la información que se comunica por sus redes informáticas y para minimizar los riesgos que pudieran afectar la infraestructura de soporte. Toda información que se transfiere fuera del organismo, incluyendo la que se transmite a través de los servicios de correo electrónico es protegida de acuerdo a su nivel de criticidad. Se asignan cuentas institucionales a todos los empleados y funcionarios, quienes están obligados a utilizarlas para toda comunicación vinculada a sus funciones. Dicho personal es informado por sus respectivas autoridades sobre los riesgos de incumplir este requerimiento, y se les exige la firma de acuerdos

de confidencialidad y no divulgación, en los casos en los que el organismo lo considere necesario.”

Esta declaración, aunque válida conceptualmente, omite especificaciones técnicas críticas como: protocolos de cifrado requeridos, estándares de autenticación, configuraciones de seguridad de red, diseño de redes seguras, control de postura de dispositivos, mecanismos de prevención de fuga de datos, o requisitos específicos de logging, monitoreo de eventos de seguridad, etc.

La falta de especificidad técnica en el marco nacional genera una heterogeneidad significativa en los niveles de seguridad entre diferentes organismos públicos. Esta disparidad no solo aumenta el riesgo agregado del sector público, sino que también dificulta la interoperabilidad segura entre organizaciones y complica los esfuerzos de respuesta coordinada ante incidentes de seguridad.

Las organizaciones quedan obligadas a interpretar por si solas los requisitos generales y traducirlos en implementaciones técnicas, resultando frecuentemente en soluciones subóptimas y brechas de seguridad no identificadas. Ademas, las organizaciones cuentan con recursos presupuestarios y técnicos muy dispares, situación que colabora con la disparidad en la seguridad de los organismos. Esta situación demanda una evolución del marco normativo hacia directrices técnicas más específicas que, manteniendo flexibilidad para adaptación contextual, proporcionen una base técnica sólida y homogénea para la seguridad del sector público. [6]

3.3. Desafíos en la adopción de seguridad en las organizaciones

La implementación de proyectos de seguridad de la información en las organizaciones enfrenta desafíos estructurales que frecuentemente determinan el fracaso o éxito limitado de estas iniciativas, independientemente de la solidez técnica de las soluciones propuestas. El principal obstáculo radica en la desconexión fundamental entre la percepción del riesgo a nivel ejecutivo y la realidad operativa que enfrentan los equipos de seguridad, conocido como risk perception gap”[7]. Los proyectos de seguridad compiten por recursos con iniciativas que generan ingresos directos o mejoras operacionales tangibles, quedando relegados hasta que ocurre un incidente crítico. Esta naturaleza reactiva crea un ciclo vicioso donde los proyectos se aprueban bajo presión, con expectativas poco realistas de implementación inmediata, presupuestos insuficientes y sin el tiempo necesario para una planificación adecuada.

La obtención del patrocinio ejecutivo genuino representa un desafío crítico que trasciende la simple aprobación presupuestaria. Los líderes de seguridad deben traducir constantemente conceptos técnicos complejos y amenazas abstractas en términos de impacto empresarial, una tarea que se complica cuando los beneficios son principalmente preventivos. La métrica del retorno de inversión, fundamental para otros proyectos empresariales, resulta casi imposible de calcular con precisión para iniciativas de seguridad. Los CISOs enfrentan el dilema de justificar inversiones millonarias basándose en eventos que, idealmente, nunca deberían materializarse. Cuando los proyectos de seguridad son exitosos y no ocurren incidentes, paradójicamente se cuestiona la necesidad de mantener o expandir estas inversiones, creando una presión constante para demostrar valor en términos que el negocio comprenda.

La gestión del cambio organizacional emerge como uno de los aspectos más subestimados y complejos en los proyectos de seguridad. A diferencia de otras iniciativas tecnológicas

que pueden implementarse gradualmente o en grupos piloto, los proyectos de seguridad frecuentemente requieren adopción universal e inmediata para ser efectivos. Un proyecto de implementación de autenticación multifactor, por ejemplo, puede fracasar completamente si ciertos ejecutivos obtienen excepciones, creando vectores de ataque privilegiados. La resistencia no proviene solo de usuarios finales que ven complicados sus flujos de trabajo, sino también de mandos medios que perciben las nuevas políticas como cuestionamientos a su autoridad o impedimentos para cumplir objetivos de productividad. Los equipos de proyecto deben navegar políticas internas, territorios departamentales establecidos, y culturas organizacionales que pueden ser fundamentalmente incompatibles con los controles de seguridad necesarios.

La integración técnica con sistemas heredados presenta desafíos que frecuentemente se subestiman en la fase de planificación. Muchas organizaciones operan con aplicaciones críticas desarrolladas hace décadas, sin documentación adecuada, y fundamentalmente incompatibles con controles de seguridad modernos. Los proyectos que inicialmente parecían simples se convierten en iniciativas de transformación digital completas cuando se descubre que implementar seguridad adecuada requiere reemplazar o rediseñar sistemas core del negocio. Los fabricantes de soluciones de seguridad prometen integraciones sin inconvenientes que raramente se materializan, dejando a los equipos de proyecto enfrentando customizaciones costosas, workarounds complejos, o la difícil decisión de operar con brechas de seguridad conocidas.

La medición del progreso y el éxito constituye otro desafío fundamental que afecta negativamente a muchos proyectos de seguridad. Los KPIs tradicionales de gestión de proyectos como cumplimiento de cronograma o presupuesto no capturan la efectividad real de los controles implementados. Un proyecto puede completarse .exitosamente"según métricas tradicionales mientras deja a la organización vulnerable debido a configuraciones incorrectas, excepciones excesivas, o falta de adopción real por parte de los usuarios. La ausencia de métricas claras de éxito dificulta la corrección del rumbo durante la implementación y complica la justificación de fases posteriores o proyectos complementarios. Esta ambigüedad en la definición del éxito permite que proyectos inadecuados continúen consumiendo recursos mientras que iniciativas críticas permanecen sin financiamiento, perpetuando un ciclo de seguridad reactiva e incompleta que deja a las organizaciones perpetuamente vulnerables.

3.4. Breve reseña de ataques informáticos

La última década ha sido testigo de una escalada sin precedentes en la sofisticación, frecuencia e impacto de los ciberataques, tanto a nivel global como en Argentina. Estos incidentes han expuesto vulnerabilidades críticas en infraestructuras esenciales, cadenas de suministro de software, y sistemas gubernamentales, transformando fundamentalmente la manera en que las organizaciones abordan la seguridad de la información.

Entre los incidentes más devastadores a nivel mundial destaca el ransomware WannaCry de mayo 2017, que afectó más de 300.000 computadoras en 150 países, paralizando sistemas críticos como el Servicio Nacional de Salud del Reino Unido. NotPetya siguió en junio 2017, causando daños superiores a los $10 mil millones de dólares CNBC, afectando empresas como Maersk, Merck y FedEx [8]. La brecha de Equifax entre mayo y julio 2017 comprometió información de 147,9 millones de estadounidenses [9], exponiendo números de seguro social y datos financieros críticos debido a una vulnerabilidad sin parchear en Apache Struts.

El ataque a la cadena de suministro de SolarWinds, descubierto en diciembre 2020, representó un cambio paradigmático en las amenazas cibernéticas. Más de 18.000 clientes de SolarWinds instalaron actualizaciones maliciosas del software Orion [10], permitiendo a los atacantes, atribuidos a Rusia, acceder a redes gubernamentales y corporativas durante meses sin ser detectados. En 2021, el ataque de ransomware a Colonial Pipeline interrumpió el suministro de combustible a casi la mitad de la Costa Este de Estados Unidos [11], evidenciando la vulnerabilidad de la infraestructura crítica.

Más recientemente, el exploit de MOVEit Transfer en 2023 demostró la persistencia de las vulnerabilidades en herramientas de transferencia de archivos, afectando a más de 2.000 organizaciones y comprometiendo datos de más de 2,8 millones de registros solo de Amazon [12]. Los ataques coordinados contra MGM Resorts y Caesars Entertainment en septiembre 2023 son un ejemplo de la efectividad de la ingeniería social, donde una llamada telefónica de 10 minutos al help desk resultó en pérdidas de $100 millones para MGM [13].

En Argentina, el panorama no es menos preocupante. El CERT argentino registró 438 incidentes en 2024, un aumento del 15 % respecto a 2023 [14]. Los ataques han afectado sistemáticamente al sector público, incluyendo el Ministerio de Economía, la Dirección Nacional de Migraciones, PAMI, y más recientemente el portal *argentina.gob.ar* en diciembre 2024. El sector privado también ha sido duramente golpeado, con incidentes significativos en empresas como Aceitera General Deheza, el diario La Nación, y la logística OCASA. Existen varias compilaciones de incidentes relevantes de Argentina. [15]

Estos incidentes revelan patrones preocupantes: la persistencia del phishing como vector principal, la efectividad de la ingeniería social, la vulnerabilidad de las cadenas de suministro de software, y la brecha entre la velocidad de evolución de las amenazas y la capacidad de respuesta organizacional. La convergencia de grupos criminales especializados, la disponibilidad de ransomware-as-a-service, y el uso de inteligencia artificial por parte de los atacantes ha creado un ecosistema de amenazas sin precedentes que demanda una transformación fundamental en cómo las organizaciones implementan controles de seguridad.

3.5. Cómo el Estado Argentino adquiere soluciones informáticas

Aquí está el documento mejorado, manteniendo la estructura pero mejorando la fluidez, cohesión y precisión académica:

El Estado Argentino constituye una compleja red institucional distribuida en todo el territorio nacional para cumplir con sus múltiples responsabilidades. Esta estructura comprende 47 dependencias de la administración central, 66 organismos descentralizados, 59 universidades públicas y 122 entidades adicionales [16].

La práctica totalidad de estas organizaciones requiere infraestructura informática propia para garantizar su funcionamiento operativo, siendo el correo electrónico un componente crítico e indispensable. En el contexto actual de crecientes amenazas cibernéticas, todas las soluciones informáticas deben implementar múltiples capas de seguridad que abarcan desde marcos normativos hasta soluciones tecnológicas específicas. Los sistemas antispam constituyen herramientas de protección fundamentales para las infraestructuras de correo electrónico, resultando esenciales para prevenir un amplio espectro de ataques informáticos que pueden comprometer la seguridad institucional.

La Oficina Nacional de Tecnologías de la Información (ONTI) es el organismo res-

ponsable de asesorar e intervenir en la formulación de proyectos y en la implementación de desarrollos tecnológicos orientados a la transformación e innovación del Estado Nacional [17]. Entre sus funciones principales se encuentra la elaboración de los Estándares Tecnológicos de la Administración Pública (ETAP), documentos que establecen las características técnicas requeridas para soluciones informáticas y constituyen una referencia fundamental para la elaboración de especificaciones técnicas en procesos de adquisición tecnológica.

Si bien estos estándares resultan adecuados para organizaciones de tamaño pequeño o mediano con requerimientos computacionales moderados, presentan limitaciones significativas cuando se aplican a instituciones de alcance nacional. Las organizaciones con presencia en todo el territorio argentino, que gestionan miles de usuarios internos y externos y requieren disponibilidad continua 24x7x365, enfrentan demandas técnicas complejas que exceden considerablemente el alcance de los ETAP actuales. Además no existen ETAP´s para soluciones de seguridad como *firewalls*, *intrusion prevention systems*, *antispams*, *antivirus*, *web security*, entre otros. Esta brecha entre los estándares disponibles y los requerimientos reales de las grandes instituciones públicas representa un desafío crítico para la modernización tecnológica del Estado.

Aunque existen diversas modalidades de compra en el sector público —licitaciones públicas, licitaciones privadas y contratación directa, entre otras— cada una con procedimientos administrativos específicos, la mayoría de las soluciones informáticas se adquieren mediante licitaciones públicas nacionales. Estos procesos licitatorios siguen un protocolo estructurado con etapas claramente definidas que deben cumplirse rigurosamente para adherir a las normativas de compras vigentes [18].

El proceso se inicia con la identificación formal de la necesidad del bien o servicio. En el caso particular analizado en este trabajo, este requisito surge naturalmente debido al vencimiento inminente de las garantías de la solución de protección de correo electrónico existente. Una vez establecida la necesidad, se procede a elaborar un documento detallado que especifique las características técnicas de la solución requerida. Esta etapa demanda un análisis exhaustivo para enumerar todas las especificaciones técnicas necesarias que garanticen el cumplimiento de los objetivos institucionales. Simultáneamente, resulta fundamental equilibrar la rigurosidad técnica con la promoción de la competencia, evitando especificaciones excesivamente restrictivas que pudieran limitar innecesariamente la participación de potenciales oferentes.

La siguiente fase comprende la elaboración de las condiciones generales de contratación. Durante esta etapa deben definirse con precisión aspectos críticos como los acuerdos de nivel de servicio (SLA), los tiempos máximos de respuesta ante incidentes, las garantías del equipamiento, los plazos de reposición ante fallas críticas, la acreditación del oferente como representante autorizado del fabricante, y las certificaciones técnicas del personal que proporcionará el soporte. Paralelamente, deben establecerse todos los requisitos administrativos y legales que deberán cumplir los participantes del proceso licitatorio.

Completada la documentación, el pliego de bases y condiciones debe remitirse a la ONTI para su evaluación y verificación de conformidad con los estándares tecnológicos estatales vigentes. Aprobado el pliego, se procede a su publicación oficial, estableciendo un período determinado para la recepción de consultas aclaratorias por parte de los potenciales oferentes y fijando una fecha límite para la presentación de las propuestas.

La evaluación de las propuestas constituye una fase crítica que requiere análisis tanto técnico como administrativo-legal. Las áreas especializadas deben emitir dictámenes fun-

dados sobre el cumplimiento de cada oferta respecto a los requisitos establecidos. Con base en estas evaluaciones, se elabora el orden de mérito, clasificando las ofertas que cumplen satisfactoriamente todos los requisitos según su propuesta económica, de menor a mayor costo.

Publicado el orden de mérito, se abre un período reglamentario para la recepción de posibles impugnaciones por parte de los oferentes. Transcurrido este plazo sin objeciones válidas, o resueltas las que se hubieran presentado, se procede a adjudicar el contrato a la oferta más conveniente, momento en el cual finalmente puede iniciarse la ejecución del proyecto. Este proceso evidencia una complejidad considerable, involucrando la coordinación de áreas técnicas, legales y administrativas, la intervención de organismos de control externos como la ONTI, y la participación de múltiples oferentes del sector privado. El tiempo transcurrido desde la identificación de la necesidad hasta la adjudicación efectiva del contrato típicamente alcanza los dos años, un plazo que frecuentemente resulta excesivo para responder oportunamente a las necesidades tecnológicas dinámicas de los organismos públicos, especialmente en un contexto de rápida evolución de las amenazas cibernéticas.

4. IMPLEMENTACIÓN DE UN SISTEMA ANTISPAM CON SANDBOXING EN UN ORGANISMO PÚBLICO

4.1. Contexto y Justificación

4.1.1. El correo electrónico como vector de ataque predominante

El correo electrónico representa actualmente el principal vector de ataques cibernéticos en organizaciones de todo tipo, constituyendo aproximadamente el 27 % de las infecciones de malware [19]. Esta prevalencia se explica por múltiples factores que convergen para hacer del correo electrónico un objetivo atractivo para los actores maliciosos:

En primer lugar, el correo electrónico es una herramienta de comunicación universal y omnipresente en prácticamente todas las organizaciones, independientemente de su tamaño, sector o ubicación geográfica. Esta característica garantiza que cualquier ataque lanzado a través de este medio tendrá un alcance potencialmente masivo.

En segundo lugar, el correo electrónico permite la ingeniería social, técnica mediante la cual los atacantes explotan aspectos psicológicos y comportamentales de los usuarios para lograr que ejecuten acciones que comprometan la seguridad. Los ataques de phishing, spear-phishing, y whaling son manifestaciones concretas de esta estrategia, donde mensajes aparentemente legítimos engañan a los usuarios para que divulguen credenciales, ejecuten archivos maliciosos o realicen transferencias financieras fraudulentas.

En tercer lugar, el correo electrónico facilita la distribución masiva y automatizada de amenazas. Un solo atacante puede enviar millones de correos maliciosos utilizando botnets o servicios especializados. Finalmente, la naturaleza del protocolo SMTP (Simple Mail Transfer Protocol) y sus extensiones históricas no fueron diseñados con la seguridad como prioridad fundamental, lo que ha generado vulnerabilidades estructurales que persisten hasta la actualidad.

4.1.2. Particularidades del sector público

En el contexto específico de organismos públicos, la criticidad del correo electrónico se amplifica considerablemente debido a características propias de estas instituciones:

- Manejo de información sensible y clasificada: Los organismos gubernamentales procesan regularmente información que afecta la seguridad nacional, datos personales de ciudadanos protegidos por normativas de privacidad (como leyes de protección de datos personales), información tributaria, registros judiciales, y comunicaciones oficiales que podrían tener implicaciones económicas y/o legales significativas. La exposición o compromiso de esta información puede tener consecuencias que trascienden lo meramente económico, afectando la confianza ciudadana, la gobernabilidad y en casos extremos, la seguridad del Estado.

- Prestación de servicios críticos a la ciudadanía: Muchos organismos públicos son responsables de servicios esenciales como salud pública, seguridad ciudadana, educación, infraestructura crítica y servicios de emergencia. Una interrupción del servicio de correo electrónico puede paralizar la coordinación interinstitucional, demorar

respuestas a situaciones de emergencia, y afectar la capacidad de la administración pública para cumplir con sus funciones constitucionales.

- Objetivo atractivo para actores de amenazas persistentes avanzadas (APT): Los organismos gubernamentales son objetivos prioritarios para grupos de hackers patrocinados por estados, organizaciones criminales sofisticadas, y activistas políticos. Estos actores frecuentemente disponen de recursos considerables, capacidades técnicas avanzadas, y motivaciones que van desde el espionaje político y económico hasta el sabotaje y la desestabilización institucional.

- Restricciones presupuestarias y procedimentales: A diferencia del sector privado, los organismos públicos enfrentan limitaciones presupuestarias más rígidas, procesos de adquisición más complejos y prolongados, y requisitos de transparencia que pueden dificultar la implementación ágil de soluciones de seguridad. Estas restricciones hacen especialmente crítico diseñar soluciones que maximicen la efectividad con recursos limitados.

4.2. Problemática Identificada

4.2.1. Limitaciones técnicas de las soluciones tradicionales

Las soluciones antispam convencionales, desarrolladas principalmente durante las décadas de 2000 y 2010, presentan limitaciones arquitecturales fundamentales que reducen significativamente su efectividad frente al panorama de amenazas contemporáneo.

Detección basada en firmas: un enfoque reactivo

El método tradicional de detección de malware opera mediante la comparación de hashes criptográficos (típicamente MD5, SHA-1 o SHA-256) de archivos adjuntos contra bases de datos de firmas conocidas de malware. Este enfoque presenta varias limitaciones críticas:

Ventana de vulnerabilidad temporal: Existe inevitablemente un período de tiempo entre el momento en que un nuevo malware es creado y distribuido, y el momento en que es detectado, analizado, clasificado y su firma es incorporada a las bases de datos de los proveedores de seguridad. Durante esta ventana, que puede extenderse desde minutos hasta días o incluso semanas dependiendo de la sofisticación del malware, las organizaciones permanecen completamente vulnerables. En este período, conocido como "zero-day window", el malware puede propagarse libremente sin ser detectado por sistemas basados exclusivamente en firmas.

Técnicas de evasión triviales: Los atacantes han desarrollado múltiples técnicas para evadir la detección basada en firmas, muchas de las cuales son sorprendentemente simples de implementar. El polimorfismo, por ejemplo, permite que el malware modifique automáticamente su código con cada infección, generando un hash diferente que no coincide con las firmas conocidas, aunque la funcionalidad maliciosa permanece idéntica. El cifrado de payloads es otra técnica común donde el código malicioso es cifrado y solo se descifra en tiempo de ejecución, haciendo imposible su detección mediante análisis estático de firmas. La ofuscación de código y la inserción de bytes aleatorios sin función son estrategias que invalidan la efectividad de la detección por firmas.

Crecimiento exponencial de variantes: El número de nuevas variantes de malware crece exponencialmente año tras año. AV-TEST Institute [20] reporta que se registran más de 450.000 nuevas muestras de malware diariamente. Este volumen hace físicamente imposible que los analistas humanos examinen manualmente cada muestra, y los sistemas automatizados de generación de firmas no siempre logran capturar adecuadamente las características esenciales de las amenazas, generando tanto falsos negativos (malware no detectado) como falsos positivos (archivos legítimos marcados incorrectamente como maliciosos).

Análisis heurístico: mejora insuficiente

Para complementar las limitaciones de la detección basada en firmas, los sistemas antispam incorporan motores de análisis heurístico. Estos motores intentan identificar características sospechosas en archivos y correos electrónicos sin depender de firmas específicas. El análisis heurístico examina aspectos como:

- Estructuras de código sospechosas o inusuales
- Intentos de ofuscación o empaquetado
- Solicitudes de permisos o privilegios inusuales
- Patrones de comportamiento potencialmente maliciosos inferidos del código estático

Sin embargo, el análisis heurístico enfrenta un dilema fundamental conocido como el "trade-off de sensibilidad": si el sistema se configura con alta sensibilidad para detectar la mayor cantidad posible de amenazas, inevitablemente generará un número elevado de falsos positivos, bloqueando correos legítimos y afectando la productividad organizacional. Esto genera frustración en los usuarios, erosiona la confianza en el sistema de seguridad, e incentiva la búsqueda de métodos alternativos para enviar archivos que eviten los controles de seguridad.

Por el contrario, si el sistema se configura con baja sensibilidad para minimizar los falsos positivos, aumentará la tasa de falsos negativos, permitiendo que malware sofisticado evada la detección. Este equilibrio es particularmente difícil de mantener considerando que el panorama de amenazas evoluciona constantemente, requiriendo ajustes continuos de los parámetros heurísticos.

Incapacidad para detectar amenazas por comportamiento

Una limitación fundamental de los sistemas tradicionales es su incapacidad para observar el comportamiento real de los archivos adjuntos y enlaces cuando se ejecutan en un sistema. Muchas amenazas modernas son "lógica bombs.º "time bombs"que permanecen inactivas hasta que se cumplen ciertas condiciones (una fecha específica, la presencia de ciertos archivos, conexión a determinadas redes, etc.). El análisis estático de estos archivos, incluso con heurísticos sofisticados, puede no revelar su naturaleza maliciosa.

Los ataques de phishing avanzados pueden incluir enlaces a sitios web que ejecutan exploits contra vulnerabilidades del navegador, descargan malware dinámicamente desde servidores remotos, o emplean técnicas de fingerprinting para determinar si están siendo analizados por sistemas de seguridad y modificar su comportamiento en consecuencia. El análisis estático de una URL no revela estas capacidades maliciosas.

Conclusión sobre limitaciones técnicas

En síntesis, las soluciones antispam tradicionales, basadas exclusivamente en detección por firmas y análisis heurístico estático, presentan brechas de seguridad significativas que son sistemáticamente explotadas por amenazas modernas, particularmente aquellas categorizadas como zero-day (vulnerabilidades no conocidas públicamente). Esta aproximación reactiva resulta fundamentalmente insuficiente frente al panorama de amenazas actual, caracterizado por su dinamismo y sofisticación.

4.2.2. Obsolescencia de la infraestructura existente

Paralelamente a las limitaciones técnicas inherentes a las soluciones tradicionales, el organismo enfrentaba una problemática específica de obsolescencia de su infraestructura de seguridad de correo electrónico existente.

Análisis de la situación inicial

En el momento de iniciar este proyecto, el organismo operaba una solución de seguridad de correo electrónico que había sido adquirida e implementada aproximadamente nueve años antes. Durante este período, la solución había cumplido adecuadamente su función, protegiendo el correo electrónico institucional y siendo refinada continuamente mediante la incorporación de reglas personalizadas, listas blancas y negras, y ajustes de políticas específicas para las necesidades del organismo.

Sin embargo, el fabricante de esta solución había anunciado oficialmente que el producto había alcanzado su EOL (End of Life) [21], una designación que en la industria tecnológica indica que el fabricante discontinúa definitivamente el soporte técnico, las actualizaciones de seguridad, y las correcciones de errores para ese producto específico. Esta situación planteaba riesgos significativos:

- Ausencia de actualizaciones de seguridad: Las vulnerabilidades de seguridad que se descubrieran en el software después de la fecha de EOL no serían corregidas mediante parches. Esto convertía al sistema en un objetivo cada vez más vulnerable, ya que los atacantes eventualmente descubrirían y explotarían estas vulnerabilidades conocidas pero no parcheadas.

- Incompatibilidades futuras: Sin actualizaciones, la solución podría volverse incompatible con actualizaciones de otros componentes del ecosistema tecnológico del organismo, como sistemas operativos, bases de datos, o protocolos de seguridad que evolucionaran con el tiempo.

- Ausencia de soporte técnico: Cualquier problema técnico, desde mal funcionamiento hasta dudas sobre configuración, ya no podría ser resuelto con asistencia del fabricante, dejando al equipo técnico del organismo sin un respaldo crítico.

La situación se agravó cuando el fabricante anunció que no solo ese modelo específico alcanzaba su EOL, sino que la línea completa de productos de seguridad de correo electrónico sería discontinuada.

Esta decisión del fabricante implicaba que no existía una ruta de actualización"dentro del mismo ecosistema tecnológico. No había un modelo superior o una versión actualizada

del mismo fabricante a la cual migrar. Cualquier solución de reemplazo debería necesariamente provenir de un fabricante diferente, con una arquitectura potencialmente distinta, una interfaz de administración diferente, y lo más crítico: un sistema de reglas y políticas incompatible con el existente.

Esta combinación de factores (EOL del producto específico y discontinuación de la línea completa) convertía el reemplazo en una necesidad inevitable. No se trataba de una actualización preventiva o una mejora deseada, sino de un requisito crítico para mantener un nivel mínimo aceptable de seguridad institucional. Continuar operando con una solución en EOL y sin fabricante habría constituido negligencia técnica y potencial responsabilidad institucional en caso de un incidente de seguridad.

El organismo enfrentaba así un dilema complejo: la necesidad urgente de reemplazar una infraestructura crítica de seguridad, pero sin poder aprovechar el conocimiento acumulado (en forma de reglas, políticas y configuraciones) en la solución existente, ya que este conocimiento estaba codificado en un formato propietario incompatible con cualquier solución de reemplazo.

4.3. Solución Propuesta

Para abordar simultáneamente las limitaciones técnicas de las soluciones tradicionales y las restricciones específicas del proceso de migración, se diseñó una arquitectura integral que incorporaba tecnologías avanzadas de detección y una metodología de transición que minimizara los riesgos operacionales.

4.3.1. Implementación de Sandboxing

La incorporación de capacidades de sandboxing constituyó el componente técnico más innovador de la solución propuesta, representando un cambio paradigmático desde la detección estática hacia el análisis de comportamiento dinámico.

Fundamento técnico del sandboxing

Un sandbox es un entorno de ejecución aislado que permite ejecutar código potencialmente malicioso de manera controlada, observando su comportamiento sin exponer sistemas productivos a riesgo alguno. El concepto se fundamenta en principios de contención y observación: al ejecutar un archivo sospechoso en un ambiente completamente aislado de la red de producción y monitoreado exhaustivamente, es posible identificar comportamientos maliciosos que no serían evidentes mediante análisis estático.

El sandboxing resuelve fundamentalmente la limitación de las técnicas tradicionales al permitir que el malware ejecute su funcionalidad maliciosa de manera observable mediante su ejecución real. Incluso el malware más sofisticado, que utiliza técnicas avanzadas de ofuscación, cifrado, o polimorfismo, eventualmente debe ejecutar acciones maliciosas observables cuando se ejecuta en un sistema real. Estas acciones pueden incluir:

- Modificación no autorizada de archivos del sistema o del registro de Windows
- Intentos de establecer comunicaciones de red hacia servidores de comando y control (C&C)
- Escalación de privilegios

- Creación de persistencia mediante tareas programadas o servicios
- Evasión o deshabilitación de mecanismos de seguridad del sistema operativo
- Exfiltración de información sensible
- Descarga de payloads adicionales desde servidores remotos

Arquitectura del componente de sandboxing

La solución de sandboxing implementada en este proyecto consistió en un cluster de dos servidores físicos dedicados exclusivamente a esta función, completamente segregados de la red de producción del organismo mediante controles de red estrictos (VLANs dedicadas, firewalls con políticas deny-by-default, y monitoreo exhaustivo de tráfico).

La arquitectura específica comprendía los siguientes elementos:

Nodos de ejecución (sandbox workers): Dos servidores que ejecutaban las máquinas virtuales donde se analizaba el contenido sospechoso. Cada nodo podía ejecutar simultáneamente múltiples instancias de máquinas virtuales, permitiendo procesar varios archivos en paralelo. Estas máquinas virtuales estaban configuradas con diferentes versiones de sistemas operativos (Windows 7, Windows 10, Windows Server 2016, Windows Server 2019) y diferentes conjuntos de software comúnmente instalado (Microsoft Office en diferentes versiones, Adobe Reader, navegadores web diversos, etc.) para simular ambientes de usuario final realistas. Sistema de monitoreo y análisis comportamental: Software especializado que instrumentaba cada máquina virtual, capturando exhaustivamente toda actividad del sistema incluyendo: llamadas al sistema operativo (syscalls), accesos al sistema de archivos, modificaciones del registro de Windows, creación de procesos, conexiones de red, y cualquier otra acción observable. Este sistema generaba un registro detallado de la "historia comportamental"de cada archivo analizado.

Motor de correlación y detección: Un componente de inteligencia artificial y aprendizaje automático que analizaba los datos comportamentales recopilados, correlacionándolos con patrones conocidos de comportamiento malicioso, y generando un "score"de maliciosidad para cada muestra analizada. Este motor se actualizaba continuamente con inteligencia de amenazas proporcionada por el fabricante de la solución de sandboxing.

Infraestructura de snapshot y recuperación rápida: Dado que algunas muestras de malware podían comprometer o corromper la máquina virtual donde se ejecutaban, era crítico poder resetearrápidamente cada VM a un estado limpio conocido antes de analizar la siguiente muestra. Esto se implementaba mediante tecnología de snapshots de virtualización, permitiendo revertir una VM completa a su estado original en cuestión de segundos.

Flujo de análisis en el sandbox

El proceso de análisis de un archivo adjunto o enlace sospechoso en el sandbox seguía el siguiente flujo detallado:

- Paso 1: Recepción y cola:

  Cuando un nodo antispam identificaba un archivo adjunto o enlace que requería análisis adicional (ya sea porque no coincidía con ninguna firma conocida, o porque

ciertas heurísticas indicaban sospecha), enviaba el archivo o URL al cluster del sandbox. Este registro incluía metadatos como: hash del archivo, tipo MIME, nombre del archivo, dirección de correo del remitente y destinatario, y un identificador único de la solicitud para rastreo.

- Paso 2: Selección de entorno de análisis:

  El nodo del cluster que recibía la solicitud de análisis seleccionaba el entorno de ejecución más apropiado basándose en características del archivo. Por ejemplo:

  - Documentos de Microsoft Office se analizaban en una VM con Microsoft Office instalado

  - Archivos ejecutables se analizaban en VMs con diferentes versiones de Windows

  - Enlaces (URLs) se analizaban abriendo navegadores web en las VMs

  - PDFs se analizaban con Adobe Reader

  En muchos casos, el mismo archivo se analizaba en múltiples entornos diferentes, dado que cierto malware exhibía comportamientos diferentes dependiendo del sistema operativo o software instalado, o incluso permanecía inactivo en algunos entornos.

- Paso 3: Ejecución controlada:

  Luego, el archivo se ejecutaba. Para documentos, esto significaba abrirlos con la aplicación correspondiente. Para ejecutables, se iniciaba el proceso. Para enlaces, se navegaba a la URL. El sistema de monitoreo comenzaba inmediatamente a registrar toda actividad.

- Paso 4: Monitoreo y registro comportamental:

  Durante la ejecución, el sistema capturaba exhaustivamente:

  - Actividad del sistema de archivos: Creación, modificación, eliminación, o lectura de archivos. Esto incluía identificar si se accedía a directorios sensibles del sistema operativo, o si se creaban archivos ejecutables en ubicaciones de inicio automático.
  - Modificaciones del registro de Windows: Especialmente críticas porque el registro controla la configuración del sistema, el inicio automático de programas, y políticas de seguridad. Malware frecuentemente modifica el registro para establecer persistencia o deshabilitar antivirus.
  - Actividad de red: Todas las conexiones salientes, incluyendo dirección IP de destino, puerto, protocolo, y contenido transmitido. Esto permitía identificar comunicación con servidores de comando y control, descarga de payloads adicionales, o exfiltración de datos.
  - Intentos de escalación de privilegios: Intentos de obtener permisos administrativos o de sistema que no correspondían con la naturaleza aparente del archivo.
  - Manipulación de servicios del sistema: Creación, modificación o detención de servicios de Windows, una técnica común para establecer persistencia o deshabilitar herramientas de seguridad.

- Paso 5: Análisis y correlación:

  Una vez finalizada la ejecución (ya sea porque el malware completó sus acciones, alcanzó el límite de tiempo, o se detectó comportamiento definitivamente malicioso que justificaba terminar el análisis inmediatamente), el sistema de correlación analizaba y evaluaba el comportamiento registrado.

  El resultado era un puntaje numérico (típicamente de 0 a 100) indicando la probabilidad de maliciosidad, junto con un veredicto categórico: malicioso, sospechoso, o benigno.

- Paso 6: Generación de reporte y veredicto:

  El sistema generaba un reporte detallado que incluía:

  - El veredicto final (malicioso/sospechoso/benigno).
  - El score de confianza.
  - Un resumen de los comportamientos observados más relevantes.
  - Capturas de pantalla de la VM durante momentos clave de la ejecución.
  - Indicadores de compromiso (IoCs) extraídos, como direcciones IP contactadas, URLs visitadas, hashes de archivos creados, etc.
  - Clasificación de la familia de malware, si era identificable.

  Este veredicto se enviaba al nodo antispam que había iniciado la solicitud de análisis.

- Paso 7: Limpieza y preparación para siguiente análisis:

  Inmediatamente después de completar el análisis, la VM se revertía a su snapshot inicial limpio, eliminando cualquier modificación que el malware hubiera realizado, y quedaba disponible para el análisis de la siguiente muestra. Este proceso de limpieza se completaba en menos de 30 segundos mediante tecnología de snapshots.

Integración con los nodos antispam

Los nodos antispam integraban la funcionalidad de sandboxing mediante un protocolo de comunicación API REST. El flujo de integración era el siguiente:

- El nodo antispam recibe un correo electrónico desde Internet.
- Ejecuta sus análisis tradicionales (firmas, heurísticos, análisis de reputación del remitente, etc.).
- Si el correo contiene archivos adjuntos o enlaces que no pueden ser definitivamente clasificados como benignos mediante los análisis tradicionales, se envía una solicitud de análisis al sandbox.
- El nodo antispam coloca el correo en una çola de cuarentena temporal"mientras espera el veredicto del sandbox.
- El sandbox procesa la solicitud y devuelve el veredicto.

- Basándose en el veredicto combinado (análisis tradicional + sandbox), el nodo antispam toma la decisión final:

  Si el veredicto es definitivamente malicioso: el correo se bloquea permanentemente, se registra en logs de seguridad, y se genera una alerta para el equipo de seguridad.

  Si el veredicto es sospechoso pero no concluyente: el correo se envía a cuarentena, se da aviso al destinatario, se registra en logs de seguridad.

  Si el veredicto es benigno: el correo se entrega normalmente al buzón del destinatario.

Beneficios específicos del sandboxing

La implementación de sandboxing proporcionó múltiples beneficios concretos:

- Detección de zero-day: El análisis de comportamiento podía identificar malware completamente nuevo, sin firmas conocidas, basándose exclusivamente en su comportamiento observable.
- Inmunidad a técnicas de evasión basadas en ofuscación: Las técnicas de polimorfismo, cifrado de payload, y otras formas de ofuscación que invalidan la detección por firmas, no son efectivas contra el análisis de sandbox, ya que el código eventualmente debe descifrarse y ejecutarse de manera observable.
- Inteligencia de amenazas enriquecida: Los reportes detallados de sandbox proporcionaban inteligencia valiosa sobre las tácticas, técnicas y procedimientos (TTPs) de los atacantes, los indicadores de compromiso asociados, y las familias de malware prevalentes atacando al organismo. Esta información alimentaba mejoras en otras capas de defensa.
- Reducción de falsos positivos: Al poder observar el comportamiento real de archivos sospechosos, muchos archivos que generaban alertas heurísticas (pero que eran en realidad benignos) podían ser correctamente clasificados, reduciendo las molestias a usuarios finales.

4.3.2. Arquitectura de Alta Disponibilidad

La criticidad del servicio de correo electrónico para las operaciones del organismo exigía una arquitectura que garantizara disponibilidad continua, incluso ante fallos de hardware, tareas de mantenimiento programado, o picos inusuales de carga. Para abordar este requisito, se diseñó una arquitectura redundante con alta disponibilidad en todos los componentes críticos.

Principios de diseño para alta disponibilidad

El diseño se fundamentó en varios principios arquitectónicos establecidos:

- Eliminación de puntos únicos de fallo (Single Points of Failure - SPOF): Cada componente crítico del sistema debía estar duplicado, de manera que el fallo de cualquier componente individual no resultara en pérdida de servicio. Esto aplicaba no solo a

servidores, sino también a componentes de red, fuentes de alimentación y enlaces de comunicación.

- Redundancia activo-activo: A diferencia de configuraciones activo-pasivo donde un nodo secundario permanece inactivo hasta que el primario falla, se optó por configuraciones activo-activo donde ambos nodos procesaban tráfico simultáneamente. Esto maximizaba la utilización de recursos y eliminaba el "tiempo de activación"del nodo secundario en caso de fallo del primario.

- Detección proactiva de fallos y recuperación automática (failover): El sistema debía detectar fallos de componentes y redirigir automáticamente el tráfico a componentes funcionales sin intervención humana.

Componentes de la arquitectura de alta disponibilidad

La implementación de alta disponibilidad se estructuró en múltiples capas, cada una con sus propios mecanismos de redundancia:

Capa de entrada: Clúster de dos nodos antispam

Se implementaron dos nodos antispam en modo clúster activo-activo:

- Capacidad de procesamiento: Cada nodo estaba dimensionado para procesar aproximadamente el 200 % del tráfico total promedio. Esto significaba que:
  - En operación normal con dos nodos activos, cada uno operaba al 25 % de su capacidad, proporcionando margen para picos de tráfico
  - Con un solo nodo activo, el nodo podría procesar el tráfico sin inconvenientes

- Mantenimiento sin interrupción: La configuración de dos nodos permitía desconectar temporalmente un nodo para tareas de mantenimiento programado (actualizaciones de software, reinicios para aplicar parches, cambios de configuración que requerían pruebas, o incluso reemplazo de hardware) sin impacto perceptible para los usuarios. El nodo restante absorbía la carga del nodo en mantenimiento sin esfuerzo.

Cada nodo antispam ejecutaba:

- Motor de análisis de firmas de malware (actualizado cada hora desde servidores del fabricante)

- Motor de análisis heurístico

- Análisis de reputación del remitente (consultando bases de datos de reputación de IPs y dominios)

- Análisis de contenido del mensaje (detección de phishing, análisis de enlaces mediante seguimiento de redirecciones, detección de técnicas de ingeniería social)

- Cliente de integración con sandbox para envío de muestras sospechosas

- Sistemas de cuarentena temporal para correos pendientes de veredicto del sandbox

- Interfaces de administración y monitoreo

Capa de análisis de comportamiento: Cluster de sandbox redundante
La infraestructura de sandboxing también se diseñó con alta disponibilidad:

- Cluster de nodos de ejecución: Dos servidores físicos actuaban como nodos de ejecución (sandbox workers). La carga se distribuía dinámicamente entre ellos.
- Sistema de colas distribuidas: Las solicitudes de análisis se almacenaban en una cola distribuida resistente al fallo de componentes individuales. Si el análisis de una muestra fallaba a mitad de camino (por ejemplo, por un crash de la VM o del nodo), la solicitud volvía automáticamente a la cola para ser reintentada en otro nodo.

Monitoreo y alertas

Para gestionar efectivamente una arquitectura de alta disponibilidad compleja, se implementó un sistema integral de monitoreo: Métricas monitoreadas en tiempo real:

- Disponibilidad de servicios: Estado up/down de cada componente, con alertas inmediatas ante fallos
- Métricas de rendimiento: Latencia de procesamiento de correos, throughput (mensajes procesados por minuto), utilización de CPU/memoria/disco en cada nodo
- Métricas de seguridad: Tasa de detección de malware, proporción de correos bloqueados vs permitidos, tipos de amenazas detectadas
- Métricas de cola: Cantidad de correos en cola esperando procesamiento, cantidad de muestras esperando análisis en sandbox
- Métricas de integración: Latencia de comunicación entre componentes, tasa de timeouts, errores de API

Sistema de alertas multinivel.
Las alertas se categorizaban en varios niveles de severidad:

- Critical: Caída de múltiples componentes, riesgo de pérdida total de servicio, requería respuesta inmediata 24/7
- High: Caída de un componente individual, degradación significativa de rendimiento, requería atención dentro de 1 hora
- Medium: Problemas de rendimiento menores, anomalías en patrones de tráfico, revisión necesaria durante horario laboral

Las alertas se distribuían mediante múltiples canales:

- Envio de registros SNMP al NOC (Network Operation Center)
- Llamado telefónico desde el NOC al integrante del equipo de seguridad que se encuentre de guardia
- Integración con sistema de ticketing para seguimiento formal
- Dashboard visual en tiempo real en el centro de operaciones

Análisis de tendencias y capacidad Además del monitoreo en tiempo real, se realizaba análisis histórico de tendencias:

- Crecimiento del volumen de correo electrónico
- Evolución de los tipos de amenazas detectadas
- Patrones temporales de carga (horarios pico, días de mayor actividad)
- Proyecciones de capacidad futura

4.3.3. Diagrama de Flujo de Correo Electrónico

La arquitectura completa del sistema implementado puede visualizarse en el siguiente flujo, que ilustra el recorrido de un correo electrónico desde su origen en Internet hasta su entrega final (o bloqueo) en el buzón del usuario:

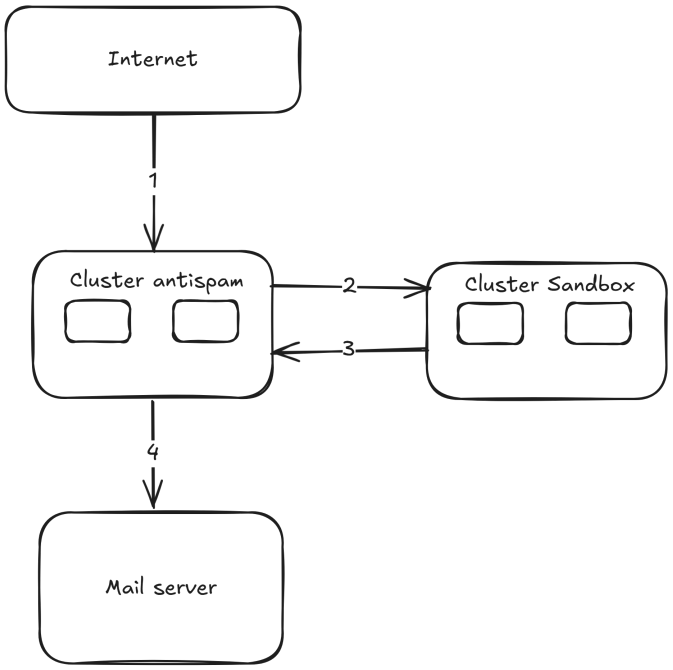

Fig. 4.1: Flujo de análisis de correo

Explicación detallada del diagrama:

El flujo comienza cuando un servidor de correo remoto en Internet que envía un mensaje al dominio del organismo. El clúster antispam recibe la conexión SMTP y la distribuyen a uno de los nodos antispam disponibles. Esta distribución se realiza en tiempo real considerando la carga actual de cada nodo.

El nodo antispam seleccionado ejecuta múltiples capas de análisis en secuencia:

- Autenticación del remitente: Verifica registros SPF (Sender Policy Framework), DKIM (DomainKeys Identified Mail), y DMARC (Domain-based Message Authentication, Reporting & Conformance) para validar que el servidor remitente está autorizado para enviar correos en nombre del dominio declarado.

- Análisis de reputación: Consulta bases de datos en tiempo real sobre la reputación de la dirección IP y dominio del remitente. Estas bases de datos, mantenidas por proveedores especializados, rastrean el historial de comportamiento de millones de servidores de correo.
- Análisis de malware conocido: Extrae y calcula hashes de todos los archivos adjuntos, comparándolos contra bases de datos de firmas de malware conocido. También descomprime archivos comprimidos (ZIP, RAR, etc.) para analizar su contenido.
- Análisis heurístico: Aplica reglas heurísticas para identificar características sospechosas como: nombres de archivo que intentan ocultar la extensión real (.pdf.exe), macros de Office sospechosas, scripts incrustados, técnicas de ofuscación, etc.
- Detección de phishing: Analiza el contenido del mensaje buscando indicadores de phishing como: discrepancias entre dominios mostrados y URLs reales, solicitudes urgentes de credenciales, etc.
- Análisis de enlaces: Extrae todas las URLs del mensaje, las sigue a través de redirecciones, y verifica su reputación en bases de datos de URLs maliciosas. También identifica técnicas de acortamiento de URLs que podrían ocultar destinos maliciosos.

Si estos análisis tradicionales generan un veredicto concluyente (definitivamente malicioso o definitivamente benigno), el nodo antispam procede directamente con la acción correspondiente. Sin embargo, si el veredicto es inconcluso (el archivo o enlace es sospechoso pero no coincide con patrones conocidos), el nodo envía el contenido sospechoso al sistema de sandboxing.

El clúster de sandbox recibe la solicitud, la registra en la cola distribuida de análisis, y selecciona un sandbox worker apropiado que tenga capacidad disponible. El worker escogido prepara una máquina virtual con el ambiente adecuado, ejecuta el contenido sospechoso, y monitorea exhaustivamente su comportamiento.

El comportamiento observado se analiza mediante modelos de machine learning que clasifican la muestra como maliciosa o benigna, generando un score de confianza y extrayendo indicadores de compromiso. Este veredicto se envía de vuelta al nodo antispam original que había solicitado el análisis.

Con el veredicto completo (análisis tradicional + sandbox), el nodo antispam toma la decisión final:

- Si el correo es malicioso: Se bloquea permanentemente, se coloca en cuarentena para análisis forense potencial, se registra detalladamente en logs de seguridad, y se genera una alerta al SOC (Security Operations Center) para conocimiento del equipo de seguridad.
- Si el correo es sospechoso: Se envía el correo a la cuarentena y se da aviso al destinatario de tal acción.
- Si el correo es benigno: Se entrega al servidor de correo interno del organismo, que lo almacena en el buzón del destinatario donde puede ser accedido mediante clientes de correo (Outlook, webmail, móviles, etc.).

Todo este proceso, desde la recepción inicial hasta la decisión final, típicamente se completa en menos de 10 segundos para correos que no requieren sandboxing, y entre

3-5 minutos para correos que sí lo requieren (el tiempo de análisis en sandbox siendo el componente más lento del proceso).

4.3.4. Desafíos de Implementación y Estrategia de Migración

La implementación de la nueva solución enfrentaba un desafío fundamental: ¿cómo reemplazar un sistema crítico de seguridad que había estado en producción durante siete años, conteniendo un conocimiento organizacional enorme codificado en cientos de reglas personalizadas, sin comprometer la seguridad durante el período de transición ni causar interrupciones operacionales?

4.3.5. Análisis detallado del riesgo de pérdida de conocimiento acumulado

Naturaleza del conocimiento en la solución legacy

La solución antispam existente contenía 147 reglas personalizadas que habían sido creadas, refinadas y ajustadas durante nueve años de operación continua. Estas reglas representaban varios tipos de conocimiento organizacional:

Reglas específicas del dominio de negocio del organismo: Por ejemplo, el organismo había identificado que ciertos términos o frases específicas del sector público (nombres de programas gubernamentales, acrónimos administrativos, terminología legal específica) eran frecuentemente utilizados en ataques de spear-phishing dirigidos específicamente contra empleados públicos. Las reglas personalizadas permitían análisis más exhaustivo de correos que contenían estos términos en ciertos contextos.

Excepciones legítimas a reglas generales: Algunos remitentes legítimos (proveedores de servicios, entidades gubernamentales de otros países, sistemas automatizados internos) generaban correos que coincidían con patrones normalmente asociados a spam, pero que en el contexto específico de este organismo eran comunicaciones válidas y necesarias. Se habían creado reglas para permitir estos casos específicos sin comprometer la seguridad general.

Respuestas a campañas de ataque históricas: A lo largo de los años, el organismo había sido blanco de diversas campañas de ataque (por ejemplo, intentos masivos de phishing). Cada una de estas campañas había resultado en la creación de reglas específicas para bloquear ataques similares en el futuro.

Ajustes culturales y de idioma: Dado que el organismo opera en un contexto de habla hispana, se habían creado reglas específicas para detectar intentos de phishing en español (que tienen características diferentes a los intentos en inglés) y ajustes para reducir falsos positivos causados por particularidades del español como idioma.

Incompatibilidad técnica entre plataformas

La migración a una plataforma de un fabricante diferente implicaba incompatibilidades técnicas fundamentales:

Sintaxis de reglas propietaria: Cada fabricante de soluciones antispam utiliza su propio lenguaje o sintaxis para definir reglas.

No existía una herramienta de conversión automática entre estos formatos, porque las diferencias no eran meramente sintácticas sino también semánticas (diferentes capacidades, diferentes funciones disponibles, diferentes formas de organizar la lógica).

Modelos de datos diferentes: Las dos plataformas representaban internamente la información de formas diferentes. Por ejemplo, la plataforma original podría almacenar listas blancas como tablas relacionales en una base de datos SQL, mientras que la nueva plataforma utilizaba archivos de configuración en formato JSON. Los metadatos asociados a cada entrada (comentarios explicando por qué se añadió, fecha de creación, usuario responsable, casos de uso asociados) podrían no tener equivalente directo en la nueva plataforma.

Documentación inadecuada del conocimiento existente

Un hallazgo preocupante durante la fase de análisis previo a la migración fue que muchas de las reglas personalizadas en la solución original carecían de documentación adecuada. Examinando la configuración existente, se encontraron situaciones como reglas creadas por administradores que ya no trabajaban en el organismo, sin documentación sobre su justificación, y comentarios en el código de reglas que eran ambiguos o se referían a .[el] problema de la semana pasada"sin más contexto. Esta falta de documentación significaba que no era posible interpretar directamente las reglas existentes y reimplementarlas en la nueva plataforma. Era necesario primero comprender qué hacía cada regla, por qué existía, y si seguía siendo relevante en el contexto actual (algunas reglas podrían haber sido creadas para amenazas que ya no existían o que ahora eran detectadas automáticamente por capacidades mejoradas de la nueva plataforma).

4.3.6. Estrategia de migración gradual: diseño y justificación

Para abordar estos desafíos sin comprometer la seguridad durante la transición, se diseñó una estrategia de migración gradual que permitiera:

- Mantener la protección completa existente durante todo el proceso de migración
- Construir progresivamente las capacidades de la nueva plataforma sin presión de tiempo
- Validar experimentalmente que la nueva configuración era equivalente (o superior) a la original
- Proporcionar un mecanismo de rollback seguro en caso de problemas críticos

Concepto fundamental: Arquitectura en paralelo

La idea central era operar ambas soluciones en pipeline durante el período de transición, con la nueva solución posicionada "delante"de la original pero inicialmente configurada en modo permisivo (registrando pero no bloqueando de manera definitiva). Esto creaba una arquitectura temporal que lucía así:

En esta configuración:

- La Nueva Solución recibía todo el tráfico de correo electrónico entrante, realizaba su análisis completo (incluyendo sandboxing), pero en lugar de tomar decisiones definitivas de bloqueo, simplemente "marcaba"los correos con sus veredictos y los reenviaba todos (tanto los que consideraba maliciosos como los benignos) hacia la solución original.

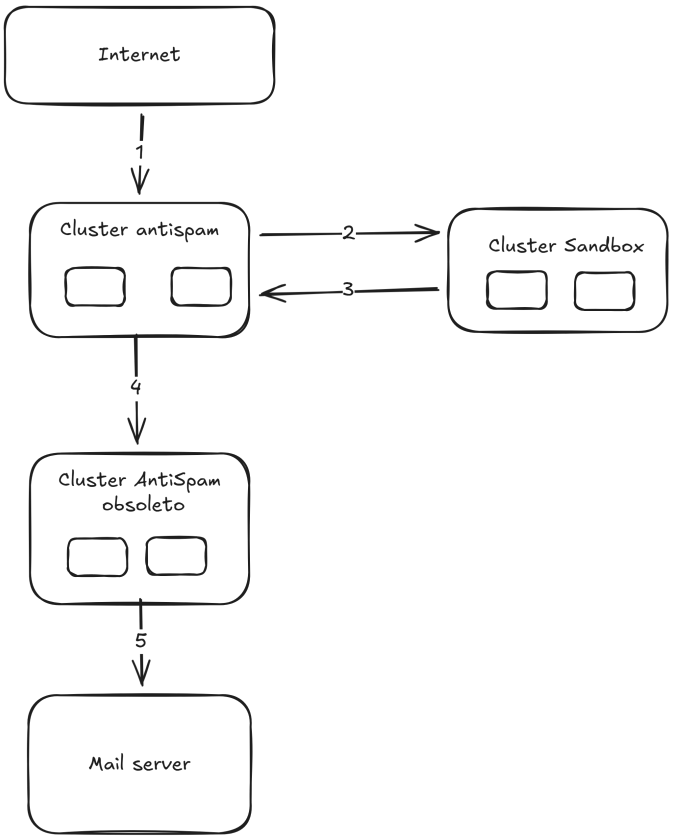

Fig. 4.2: Flujo de correo durante la migración

- La Solución Original continuaba operando exactamente como lo había hecho durante años, aplicando todas sus reglas refinadas y tomando las decisiones finales de bloqueo o entrega. Esta solución permanecía como la .ªutoridad final"de seguridad.

- El análisis comparativo se realizaba comparando continuamente las decisiones de ambos sistemas: ¿qué correos bloqueó la solución original? ¿Los había detectado también la nueva solución? ¿Qué correos marco como maliciosos la nueva solución que la original permitió? ¿Era correcto bloquearlos?

Objetivos medibles de la estrategia

La estrategia de migración se diseñó con criterios de éxito claros y medibles:

Objetivo primario: Alcanzar una situación donde la solución original no bloqueara ningún correo adicional durante un período sostenido de al menos dos semanas consecutivas. Esto indicaría que la nueva solución había alcanzado (o superado) las capacidades de detección de la original.

Objetivo secundario: Reducir la tasa de falsos positivos respecto a la solución original, aprovechando las capacidades mejoradas de la nueva plataforma (particularmente el sandboxing) para clasificar más precisamente archivos sospechosos.

Objetivo de seguridad: Zero compromisos de seguridad durante la transición. Ningún correo malicioso debería llegar a usuarios finales que habría sido bloqueado por la solución

original. Objetivo operacional: Completar la migración en un plazo máximo de 20 semanas, considerando las restricciones de personal y presupuesto.

4.3.7. Fases de implementación detalladas

La implementación se estructuró en cuatro fases secuenciales, cada una con objetivos, actividades y criterios de éxito específicos.

Fase 1: Instalación en modo observación (Semanas 1-2)

Objetivos de la fase:

- Instalar y configurar completamente la infraestructura de la nueva solución
- Verificar funcionalidad básica de todos los componentes
- Establecer el flujo de datos en paralelo con la solución original
- Iniciar la recolección de datos comparativos

Actividades realizadas:
Semana 1 - Instalación física e infraestructura:

- Instalación física de los servidores en racks del datacenter
- Configuración de conectividad de red (VLANs, direccionamiento IP, rutas)
- Actualización a la ultima versión de software estable
- Instalación y configuración del cluster de sandboxing
- Preparación de las máquinas virtuales del sandbox con diferentes sistemas operativos y aplicaciones
- Configuración de certificados SSL/TLS para comunicaciones cifradas entre componentes
- Implementación de los sistemas de monitoreo y alertamiento

Semana 2 - Configuración básica y puesta en marcha:

- Aplicación de la configuración inicial "de fábrica"de la nueva solución, sin personalizaciones
- Habilitación de las actualizaciones automáticas de bases de datos de firmas de malware
- Configuración de la integración entre los nodos antispam y el cluster de sandbox
- Implementación de la configuración de "bridge mode"donde la nueva solución reenvía todo el tráfico (sin bloquear) hacia la solución original
- Validación de que el flujo de correo end-to-end funciona correctamente

- Configuración de scripts de recolección de logs y métricas comparativas

La configuración implementada resultó fundamental para la estrategia de migración. Técnicamente, se implementó de la siguiente manera: La nueva solución analizaba los correos electrónicos y los enviaba al sandbox cuando correspondía. Luego, agregaba un encabezado (header) a cada mensaje que indicaba si hubiera sido bloqueado, enviado a cuarentena o si se trataba de un correo limpio. Finalmente, la solución reenviaba todos los mensajes a la solución original, que ejecutaba los controles utilizando la configuración original. La nueva solución generaba un registro de los mensajes procesados incluyendo el header mencionado para un análisis y control posterior.

Verificaciones de funcionalidad:
Durante esta fase se ejecutaron múltiples pruebas de validación:

- *Pruebas de conectividad:* Envío de correos de prueba desde servidores externos verificando que llegaban correctamente a través de toda la cadena (nueva solución → solución original → buzón de usuario)
- *Pruebas de rendimiento:* Envío de volúmenes elevados de correo de prueba para verificar que la infraestructura podía manejar la carga esperada sin cuellos de botella
- *Pruebas de failover:* Simulación de fallos de componentes individuales (apagado de un nodo antispam, un balanceador de carga, un worker de sandbox) para verificar que el sistema continuaba operando
- *Pruebas de sandbox:* Envío de muestras conocidas de malware (desde un repositorio de muestras para investigación) para verificar que el sandbox las detectaba correctamente
- *Pruebas de latencia:* Medición de los tiempos de procesamiento para asegurar que la adición de la nueva capa no incrementaba excesivamente el tiempo de entrega de correos legítimos

Métricas iniciales recolectadas:
Al final de esta fase, se habían recolectado métricas baseline:

- Volumen diario de correo electrónico: 45.000 mensajes/día
- Tasa de correos con archivos adjuntos: 18 % del total
- Tasa de correos que requerían análisis de sandbox: 2.3 % del total (aproximadamente 1.000 correos/día)
- Tiempo promedio de procesamiento sin sandbox: 1.2 segundos
- Tiempo promedio de procesamiento con sandbox: 3.8 minutos
- Correos bloqueados diariamente por la solución original: 850 mensajes/día (1.9 % del total)

Criterios de éxito de la fase:
Para considerar exitosa esta fase y proceder a la siguiente, se verificó:

- Cero interrupciones del servicio de correo electrónico durante la instalación
- Todos los componentes funcionando correctamente según monitoreo
- 100 % de los correos de prueba entregados exitosamente
- Sistema de recolección de logs comparativos operando correctamente
- Latencia introducida por la nueva capa aceptable (menor a 5 segundos para correos sin sandbox, menor a 5 minutos para correos con sandbox)

Fase 2: Análisis comparativo y migración de reglas (Semanas 3-12)

Esta fue la fase más extensa y técnicamente compleja del proceso de migración, consumiendo diez semanas de trabajo intensivo.

Objetivos de la fase:

- Analizar sistemáticamente las discrepancias entre las decisiones de ambos sistemas
- Identificar y documentar todas las reglas personalizadas de la solución original
- Recrear reglas equivalentes en la nueva plataforma
- Validar la efectividad de las reglas migradas
- Alcanzar convergencia progresiva entre ambos sistemas

Metodología de trabajo:

Se estableció un ciclo de trabajo diario:

Cada mañana, el equipo técnico revisaba los reportes generados automáticamente durante las 24 horas previas. Estos reportes identificaban: correos bloqueados por la solución original pero permitidos por la nueva, correos bloqueados por la nueva solución pero permitidos por la original, correos bloqueados por ambas soluciones.

```
========================================
REPORTE DE DIFERENCIAS - SEMANA 4
========================================

Total de correos procesados: 47.234
Correos bloqueados por solución original: 891
Correos bloqueados por nueva solución: 856
Coincidencias (ambas bloquearon): 823

ATENCIÓN REQUERIDA:
-------------------
68 correos bloqueados SOLO por solución original
33 correos bloqueados SOLO por nueva solución
```

Luego, para cada caso identificado como "bloqueado SOLO por la solución original", el equipo realizaba un análisis detallado de la causa y de corresponder replicaba la regla que había causado el bloqueo en la nueva solución documentando el propósito de la misma.

El progreso de la migración se medía semanalmente:

```
REPORTE DE PROGRESO - SEMANA 4 (Semana 2 de Fase 2)
==================================================

MÉTRICAS DE CONVERGENCIA:
- Reglas totales identificadas en solución original: 147
- Reglas migradas a nueva plataforma: 43 (29.3%)
- Reglas en proceso de validación: 12 (8.2%)
- Reglas pendientes de análisis: 92 (62.6%)

EFECTIVIDAD DE DETECCIÓN:
- Correos bloqueados solo por original: 34/día (↓ desde 68/día semana anterior)
- Correos bloqueados solo por nueva: 28/día (↑ desde 33/día)
- Tasa de convergencia: 94.2% (↑ desde 91.5%)

NUEVAS CAPACIDADES:
La nueva solución ha detectado 28 amenazas diarias que la original no detectaba:
- 18 casos: detecciones mediante sandbox (malware zero-day)
- 6 casos: mejoras en análisis de URLs (sandbox de navegación)
- 4 casos: detecciones basadas en reputación mejorada

INCIDENTES:
- 2 falsos positivos detectados en reglas nuevas, corregidos en menos de 24h
- 0 compromisos de seguridad
- 0 interrupciones de servicio
```

Desafíos específicos encontrados durante esta fase:

Desafío 1: Reglas con lógica compleja difícil de replicar

Algunas reglas de la solución original utilizaban lógica extremadamente compleja que no tenía equivalente directo en la nueva plataforma. Por ejemplo, alertar en caso de detectar un patrón de comportamiento que indique que una cuenta de correo interna envió mas de 20 correos en menos de 7 días hacia una misma casilla de correo externa con adjuntos de mas de 5MB.

Esta regla intentaba detectar posibles fugas de datos identificando patrones anómalos de correo saliente de un empleado interno hacia destinatarios externos. Requería:

- Mantener estado histórico de patrones de envío por usuario
- Análisis estadístico de tamaños de archivos
- Análisis temporal de distribución horaria

La nueva plataforma no tenía capacidades nativas para este tipo de análisis estadístico. La solución requirió:

- Exportar logs de correo a un sistema SIEM (Security Information and Event Management) externo
- Implementar las reglas de correlación complejas en el SIEM

- Configurar el SIEM para enviar alertas al equipo de seguridad
- Mantener las reglas básicas de bloqueo en la plataforma antispam

Este proceso tomó dos semanas adicionales de coordinación con el equipo responsable del SIEM.

Desafío 2: Reglas obsoletas o redundantes

Durante el análisis, se descubrió que aproximadamente el 22 % de las reglas personalizadas (32 de 147) eran obsoletas o redundantes:

- 15 reglas bloqueaban dominios o IPs que ya estaban en listas negras públicas actualizadas
- 8 reglas detectaban patrones de malware que ahora eran detectados automáticamente por las firmas actualizadas
- 5 reglas respondían a campañas de ataque que habían cesado años atrás
- 4 reglas eran duplicados funcionales de otras reglas con diferente sintaxis

Esto presentaba un dilema: ¿migrar literalmente todas las reglas (incluyendo las obsoletas) para garantizar idéntica funcionalidad, o aprovechar la migración para optimizar el conjunto de reglas eliminando las innecesarias?

Después de análisis, se decidió:

- Documentar todas las reglas, incluyendo las obsoletas
- Migrar inicialmente solo las reglas que seguían siendo necesarias (115 reglas)
- Mantener documentación de las 32 reglas obsoletas por si en el futuro surgía necesidad de reactivarlas
- Validar que las reglas eliminadas efectivamente no detectaban amenazas actuales mediante análisis de falsos negativos

Desafío 3: Diferencias en capacidades de análisis de archivos

La plataforma original y la nueva tenían diferentes capacidades de análisis de tipos de archivo. Por ejemplo:

- Configurar la nueva plataforma para análisis de 4 niveles de profundidad en archivos comprimidos (superando a la original)
- Implementar límites de tamaño descomprimido para prevenir "zip bombs"(archivos comprimidos pequeños que se expanden masivamente)
- Ajustar los timeouts de análisis para acomodar el procesamiento adicional

Criterios de éxito de la fase:

Para considerar exitosa esta fase y proceder a la siguiente, se verificó:

- Al menos el 75 % de las reglas personalizadas migradas exitosamente
- Tasa de convergencia de detección ¿98 %

- Brecha de detección residual menor a 5 casos/día de amenazas no críticas
- Tasa de falsos positivos aceptable (menor a 5/semana)
- Documentación completa de todas las reglas

Fase 3: Desactivación de la solución legacy (Semana 13)

La fase final consistió en la desactivación controlada y documentada de la solución original, manteniendo capacidad de rollback como medida de contingencia.

Objetivos de la fase:

- Desactivar la solución original de manera ordenada y documentada
- Archivar configuraciones y datos históricos para referencia futura
- Mantener capacidad de reactivación rápida durante período de contingencia
- Documentar lecciones aprendidas y completar la transición

Actividades de la semana 13:

- Backup completo de todas las configuraciones de la solución original
- Exportación de todas las reglas personalizadas (incluso las no migradas)
- Extracción de logs históricos y datos de cuarentena
- Documentación del estado final del sistema
- Preparación de procedimiento de rollback de emergencia
- El miércoles a las 2:00 AM (horario de menor actividad de correo), se modifico el ruteo del trafico desde la nueva solución hacia los servidores de correo, en vez de enviar el flujo de mensajes a la antigua solución de antispam. De esta manera la vieja solución quedo desafectada.
- Monitoreo cada 30 minutos durante las primeras 24 horas
- Monitoreo cada 2 horas durante los siguientes 4 días
- Solución original en standby, lista para su reactivación en menos de 15 minutos si fuera necesario

Al finalizar la semana no se detectaron incidentes ni indisponibilidades. El procesamiento se mantuvo estable dentro de los tiempos esperados y no fue necesario reactivar la solución original. Esta se conservó en modo stand-by durante tres semanas adicionales; dado que no se requirió su utilización, se procedió al desmantelamiento definitivo.

El proceso de desmantelamiento incluyó las siguientes tareas:

- Restablecimiento de configuración de fábrica del sistema
- Registro y liberación de las direcciones IP asignadas

- Extracción física de las unidades de disco
- Borrado seguro y certificado de los discos
- Liberación de las unidades de rack con su correspondiente documentación en el registro de infraestructura

Criterios de éxito de la fase:
Para considerar exitosa esta fase final, se verificó:

- Operación estable durante 30 días continuos sin la solución original
- Sin necesidad de reactivar la solución original
- Sin incidentes de seguridad atribuibles a la migración
- Toda la información histórica correctamente archivada

4.4. Métricas y Resultados

La implementación del nuevo sistema antispam con capacidades de sandboxing se consideró exitosa basándose en múltiples indicadores cuantitativos y cualitativos medidos durante y después del proceso de migración.

4.4.1. Métricas del proceso de migración

Temporalidad

- Duración total del proyecto: 13 semanas desde la instalación inicial hasta la desactivación de la solución original
- Fase de instalación: 2 semanas
- Fase de análisis y migración de reglas: 10 semanas
- Fase de desactivación: 1 semana
- Período adicional de validación: 3 semanas antes del desmantelamiento físico

Continuidad operacional

Durante todo el proceso de migración se logró mantener una disponibilidad del servicio de correo del 100 %, sin registrarse incidentes críticos ni interrupciones de ningún tipo. Esta métrica demuestra que la estrategia de implementación en paralelo cumplió exitosamente su objetivo de mantener continuidad operacional durante un proceso de reemplazo de infraestructura crítica.

Seguridad durante la transición

- Correos maliciosos que llegaron a usuarios finales durante la migración: 0 (cero confirmados)
- Incidentes de seguridad atribuibles a la migración: 0 (cero)
- Compromisos de sistemas durante el período de transición: 0 (cero)

Estos números validan que la estrategia de mantener la solución original como autoridad final durante las fases iniciales fue acertada, garantizando que no se sacrificara seguridad por velocidad de implementación.

4.4.2. Métricas de efectividad de detección

Mejora en detección de amenazas

Comparando el mes anterior a la migración con el primer mes de operación completa de la nueva solución:

```
COMPARATIVA DE DETECCIÓN DE AMENAZAS
====================================

                                Pre-migración  Post-migración  Mejora
                                (Mes -1)       (Mes +1)
-------------------------------------------------------------------
Amenazas zero-day detectadas    -              47              N/A*
Ransomware bloqueado            -              19              N/A*
Phishing avanzado detectado     89             167             +88%
Malware total bloqueado         178            312             +75%
URLs maliciosas en emails       234            398             +70%
Documentos con macros maliciosas 45            71              +58%

* Funcionalidad nueva, no existente en solución original
```

Análisis de la mejora:

La mejora más dramática se observa en la detección de amenazas zero-day y ransomware. Esto se atribuye directamente a las capacidades de sandboxing, que pueden identificar malware completamente nuevo basándose en su comportamiento, sin depender de firmas conocidas.

El incremento en detección de phishing avanzado (88 %) refleja las capacidades mejoradas de análisis de URLs de la nueva plataforma con *IOC's* actualizados, que puede seguir cadenas de redirección, detectar ofuscación de enlaces, y analizar el contenido real de sitios web sospechosos mediante navegación automatizada en el sandbox.

Reducción de falsos positivos

```
COMPARATIVA DE FALSOS POSITIVOS
===============================
```

Métrica	Pre-migración	Post-migración	Mejora
Falsos positivos por semana	12-15	3-5	-70%
Tiempo promedio de resolución	8.5 horas	2.3 horas	-73%
Quejas de usuarios	25/mes	7/mes	-72%

La reducción significativa en falsos positivos se atribuye a dos factores:

- Análisis de comportamiento: El sandbox permite verificar si archivos sospechosos realmente exhiben comportamiento malicioso, reduciendo bloqueos basados en heurísticas imprecisas.
- Inteligencia de amenazas actualizada: Las bases de datos de reputación más modernas de la nueva plataforma reducen bloqueos incorrectos basados en información desactualizada.

5. LECCIONES APRENDIDAS

La experiencia de este proyecto de migración generó aprendizajes significativos que pueden ser valiosos para otras organizaciones enfrentando desafíos similares de actualización tecnológica con preservación de conocimiento organizacional. Este capítulo sintetiza las principales lecciones derivadas del proceso, analizando tanto los aspectos técnicos como los organizacionales que resultaron determinantes para el éxito del proyecto.

5.1. Valor de las estrategias de migración gradual

Una de las decisiones más críticas del proyecto fue optar por una implementación gradual en lugar de un reemplazo en un solo paso. Esta elección, aunque requirió significativamente más tiempo y recursos, demostró ser fundamental para mitigar riesgos operacionales y de seguridad. La lección principal que emerge de esta experiencia es que en sistemas críticos de seguridad, el tiempo adicional invertido en una migración gradual representa una inversión en mitigación de riesgos que se justifica ampliamente frente a las consecuencias potenciales de un fallo.

La justificación empírica de esta conclusión surge directamente de los resultados del proyecto. La implementación completa requirió aproximadamente dieciséis semanas, equivalentes a cuatro meses de trabajo continuo. En contraste, un enfoque de reemplazo instantáneo habría completado la transición en apenas dos o tres semanas. Sin embargo, esta diferencia temporal debe evaluarse en el contexto de los resultados obtenidos. Durante todo el proceso de migración gradual, el proyecto experimentó cero incidentes críticos y cero tiempo de inactividad no planificado.

Más allá de la simple reducción de incidentes, la estrategia gradual habilitó un aprendizaje continuo que resultó valioso. Cada fase del proyecto generó conocimiento que se aplico para optimizar las fases subsiguientes. Este proceso de aprendizaje iterativo permitió al equipo identificar y corregir problemas menores antes de que afectaran a un número significativo de usuarios, construyendo progresivamente la confianza tanto del equipo técnico como de los usuarios finales en el nuevo sistema.

Las recomendaciones derivadas de esta experiencia es que para sistemas críticos de seguridad, siempre que sea posible debe preferirse una migración gradual sobre un reemplazo instantáneo. La planificación del proyecto debe contemplar desde su inicio la capacidad de operación en paralelo, aunque esto implique costos adicionales temporales. Es fundamental definir métricas claras de convergencia que indiquen objetivamente cuándo es seguro proceder a la siguiente fase del proceso. Finalmente, resulta preferible extender el cronograma del proyecto dos o cuatro semanas adicionales antes que arriesgar un incidente de seguridad que podría comprometer la operación crítica del organismo.

Esta lección trasciende el contexto específico de sistemas antispam y aplica a cualquier infraestructura crítica. Los mismos principios son aplicables a migraciones de sistemas de detección y prevención de intrusiones, plataformas de autenticación como Active Directory o Single Sign-On, sistemas de backup y disaster recovery, y bases de datos críticas. En cada uno de estos contextos, la naturaleza crítica del servicio justifica ampliamente la inversión adicional en una estrategia de migración gradual.

5.2. Importancia de la documentación del conocimiento tácito

El conocimiento organizacional almacenado en sistemas técnicos frecuentemente carece de documentación adecuada sobre su propósito y contexto, y esta falta de documentación representa un riesgo significativo durante procesos de migración o transición de personal. La experiencia con las ciento cuarenta y siete reglas personalizadas del sistema original ilustra vívidamente esta problemática.

Durante el proceso de análisis, se descubrió que aproximadamente el sesenta por ciento de las reglas personalizadas del sistema original carecían completamente de documentación sobre su propósito, contexto de creación o criterios de activación. Esta ausencia de documentación no es un problema trivial: muchas de estas reglas fueron creadas hace años por personal que ya no trabaja en el organismo. En algunos casos, las reglas respondían a amenazas o situaciones específicas cuyo contexto se había perdido con el tiempo. La comprensión de estas reglas requirió de análisis, examinando patrones de activación y correlacionando con logs históricos.

Este proceso de reconstrucción retrospectiva del conocimiento organizacional consumió aproximadamente el treinta por ciento del tiempo total del proyecto, un costo significativo que podría haberse reducido drásticamente con documentación adecuada desde el inicio. Más preocupante aún, durante el análisis se identificaron aproximadamente doce reglas cuyo propósito no pudo reconstruirse con certeza razonable, y otras ocho reglas que, tras análisis detallado, se determinaron completamente obsoletas pero que continuaban activas y consumiendo recursos computacionales sin aportar valor alguno.

Toda configuración personalizada en sistemas de producción debe documentarse en el momento de su creación, no retrospectivamente. Esta documentación debe incluir al menos cinco elementos esenciales: el propósito o problema que la configuración intenta resolver, el contexto organizacional o de amenaza que motivó su creación, los criterios técnicos de activación y comportamiento esperado, el responsable de la creación con información de contacto, y la fecha de creación e historial de modificaciones. Además, es fundamental establecer procesos de revisión periódica, idealmente trimestral o semestral, para validar que las configuraciones personalizadas continúan siendo necesarias y relevantes.

Más allá de las prácticas de documentación, resulta crítico desarrollar una cultura organizacional que valore el conocimiento documentado. Esto requiere asignar tiempo específico para documentación en las cargas de trabajo del personal técnico y establecer la documentación como prerequisito para el cierre de tickets o proyectos.

5.3. Necesidad de arquitecturas resilientes

La arquitectura de alta disponibilidad implementada en el proyecto no fue un lujo sino una necesidad imperativa para servicios críticos como el correo electrónico institucional. La experiencia del proyecto validó convincentemente el valor de invertir en redundancia y capacidad de recuperación ante fallos, demostrando que el costo adicional de estas capacidades se justifica ampliamente frente a las consecuencias de indisponibilidad.

Durante la operación post-implementación, el sistema enfrentó múltiples eventos que habrían causado interrupciones significativas en una arquitectura no resiliente. Se registraron tres fallos de hardware en nodos individuales durante los primeros seis meses de operación, ninguno de los cuales causó interrupción del servicio gracias a la redundancia de la arquitectura. El sistema experimentó dos interrupciones de conectividad con pro-

veedores externos debido a problemas del proveedor de internet, pero la redundancia de conexiones permitió mantener la operación sin impacto para los usuarios. Se identificó un error de software en una actualización del sistema de sandboxing que afectó a uno de los nodos, pero el sistema continuó operando normalmente en capacidad reducida mientras se remediaba el problema.

El valor cuantificable de la arquitectura resiliente es significativo. Sin alta disponibilidad, estos eventos habrían causado interrupción del servicio durante los primeros seis meses. El costo reputacional y operacional de tales interrupciones justifico la inversión en una solución con alta disponibilidad.

Las lecciones sobre arquitecturas resilientes van más allá de la simple redundancia de componentes. La resiliencia verdadera requiere redundancia en múltiples niveles: hardware mediante configuraciones de alta disponibilidad con failover automático, conectividad con múltiples enlaces de internet de proveedores diferentes, energía con sistemas de alimentación ininterrumpida y generadores de respaldo y procesos operacionales con procedimientos claros de respuesta ante incidentes y capacitación cruzada del personal.

5.4. Complementariedad de técnicas de detección

Otra lección técnica del proyecto fue la validación empírica de que ninguna técnica individual de detección de amenazas es suficiente en el panorama actual de ciberseguridad. La arquitectura implementada combinó varios enfoques complementarios: detección basada en firmas y patrones conocidos, análisis heurístico y comportamental, y análisis en entorno aislado mediante sandboxing. La evidencia del proyecto demostró que esta combinación proporciona una defensa en profundidad significativamente más robusta que cualquier técnica individual.

Los datos de operación post-implementación demuestran que durante los primeros seis meses, el sistema detectó y bloqueó más de doscientas cincuenta amenazas únicas. El análisis de estas detecciones mostró patrones claros de complementariedad entre las técnicas. Las firmas tradicionales detectaron el cincuenta y dos por ciento de las amenazas, principalmente variantes conocidas de malware y campañas de phishing establecidas. El análisis heurístico identificó un treinta y uno por ciento adicional, capturando variantes nuevas de amenazas conocidas y patrones sospechosos que aún no tenían firmas específicas. El sandboxing resultó crítico para el diecisiete por ciento restante, detectando amenazas de día cero y ataques dirigidos sofisticados que evadieron tanto firmas como heurísticas.

La implementación efectiva de defensa en profundidad requiere más que simplemente activar múltiples tecnologías. Es necesario configurar las técnicas para operar en secuencia, con cada capa proporcionando análisis progresivamente más profundo. El sistema debe implementar umbrales de confianza adecuados para cada técnica, reconociendo sus fortalezas y limitaciones específicas. Es fundamental establecer mecanismos de correlación que permitan combinar señales de múltiples fuentes para decisiones más informadas. Finalmente, resulta crítico contar con procesos de actualización constante de todas las capas de detección, ya que las amenazas evolucionan continuamente.

Las lecciones sobre complementariedad de técnicas van más allá del dominio específico del correo electrónico y antispam. Los mismos principios aplican a otras áreas de ciberseguridad. La detección de intrusiones en redes se beneficia de combinar análisis de firmas, detección de anomalías y análisis de comportamiento. La protección de endpoints requiere integrar antivirus tradicional, detección de comportamiento y capacidades de respues-

ta ante incidentes. La seguridad de aplicaciones web necesita combinar web application firewalls, análisis de código estático y pruebas dinámicas. En todos estos contextos, la diversidad de técnicas de detección proporciona resiliencia frente a técnicas de evasión y reduce el riesgo de puntos únicos de verificación.

5.5. Replicabilidad y transferencia de conocimiento

Una reflexión importante sobre el proyecto es su replicabilidad y la transferencia del conocimiento generado a otros contextos. Los proyectos técnicos tienen valor no solo por sus resultados específicos sino también por el conocimiento y las metodologías que generan y que pueden aplicarse en otros contextos.

La experiencia de este proyecto es altamente replicable, con adaptaciones apropiadas, en organizaciones similares. La metodología de migración gradual con operación en paralelo es aplicable a cualquier actualización de infraestructura crítica. Los principios de arquitectura resiliente y alta disponibilidad son universales para servicios críticos. Las métricas de éxito definidas, incluyendo convergencia de detección, tasas de falsos positivos y negativos y disponibilidad, proporcionan un modelo medible para proyectos similares.

Sin embargo, la replicación no es literal sino que requiere adaptación al contexto específico. Organizaciones de diferentes tamaños requerirán diferentes escalas de infraestructura. El volumen de tráfico de correo dictará el dimensionamiento necesario de hardware y capacidad. Diferentes organizaciones enfrentan diferentes perfiles de amenaza, por lo que las reglas personalizadas deben reflejar el contexto específico de riesgo. Las restricciones presupuestarias variarán, requiriendo diferentes balances entre capacidades técnicas y costos.

Aunque este proyecto se ejecutó en un organismo público, las lecciones aprendidas son igualmente valiosas para el sector privado. Empresas de sectores regulados como servicios financieros, salud o energía enfrentan desafíos similares de alta criticidad y requisitos estrictos de disponibilidad. Organizaciones de cualquier sector acumulan conocimiento en sistemas legacy que necesitan preservar durante migraciones. Los principios de defensa en profundidad, arquitecturas resilientes y gestión rigurosa del cambio son universales, trascendiendo las diferencias entre sector público y privado.

5.6. Evolución continua de la seguridad

Una conclusión fundamental es que la seguridad no es un estado final sino un proceso continuo de evolución y mejora. El proyecto no resolvió permanentemente la seguridad del correo electrónico del organismo, sino que estableció una plataforma moderna y capaz sobre la cual se continuará construyendo.

Durante los primeros seis meses post-migración, se implementaron varias mejoras. El organismo se suscribió a feeds adicionales de threat intelligence. Se implementó integración con el sistema SIEM del organismo para correlación avanzada de eventos de seguridad. Se amplió el programa de capacitación en concientización de seguridad para usuarios finales basado en amenazas reales observadas.

Esta visión de la seguridad como proceso continuo tiene implicaciones importantes para la planificación y presupuestación. Los presupuestos de seguridad no deben verse como gastos únicos sino como inversiones continuas necesarias para mantener y mejorar la postura de seguridad. Las arquitecturas técnicas deben diseñarse con flexibilidad y ex-

tensibilidad para acomodar evoluciones futuras sin requerimientos de rediseño completo. Los equipos técnicos necesitan tiempo dedicado no solo a operación sino también a investigación, aprendizaje e implementación de mejoras. Las métricas de éxito deben incluir no solo estabilidad operacional sino también capacidad de evolución y adopción de nuevas capacidades.

5.7. Síntesis de aprendizajes clave

En síntesis, las lecciones más valiosas del proyecto pueden resumirse en varios principios fundamentales que trascienden el contexto específico y ofrecen guía para futuros proyectos similares.

El tiempo invertido en mitigación de riesgos mediante estrategias graduales es siempre tiempo bien invertido en sistemas críticos. La documentación del conocimiento organizacional no es opcional sino esencial para la continuidad operacional a largo plazo. La inversión en alta disponibilidad se justifica por los costos evitados de interrupciones y la confiabilidad del servicio. La defensa efectiva requiere múltiples capas complementarias, sin depender de ninguna técnica única. Los aspectos humanos y organizacionales son tan críticos como los aspectos técnicos para el éxito del proyecto.

Estos principios, validados empíricamente en el contexto de este proyecto, representan aprendizajes valiosos que pueden guiar no solo proyectos similares de seguridad en correo electrónico, sino cualquier iniciativa de actualización o modernización de infraestructura crítica en organizaciones de cualquier tipo y tamaño.

5.8. Dos alternativas a la contratación fragmentada de soluciones comerciales

Como se expuso anteriormente, la administración pública argentina está conformada por cientos de organismos que requieren soluciones de protección de correo electrónico. Para dimensionar el impacto económico, los costos de soluciones de seguridad de correo para organizaciones chicas puede estimarse en 50.000 dolares anuales, mientras que las grandes pueden alcanzar los 300.000. Esta realidad plantea tanto una oportunidad como una necesidad imperiosa de repensar el modelo actual mediante el cual el Estado aborda estos requerimientos de seguridad informática. La fragmentación existente en la gestión de estas soluciones sugiere la posibilidad de implementar alternativas que permitirían optimizar significativamente el uso de los recursos nacionales, tanto en términos presupuestarios como en el aprovechamiento del capital humano especializado disponible en el sector público. A continuación se describen brevemente algunas alternativas a la contratación fragmentada de soluciones de seguridad de correo. Si bien para implementar cualquiera de estas alternativas sería necesaria una adecuación normativa y legal, es importante destacar la factibilidad técnica de las propuestas.

5.8.1. Unificación de una solución de antispam para toda la APN

Una alternativa a la realidad actual consiste en la construcción de una solución única de antispam que analice todo el tráfico de correo electrónico de la administración pública. Desde el punto de vista técnico, es factible realizar una instalación centralizada que brinde servicio a todos los organismos, procesando sus correos electrónicos de manera unificada. Esta alternativa presenta varias ventajas significativas:

- **Unificación de políticas:** Permite establecer criterios homogéneos de aceptación y rechazo de correos en toda la administración pública.

- **Ahorro presupuestario:** La adquisición de equipamiento para un servicio centralizado resultaría órdenes de magnitud más económica que la contratación de la misma solución por parte de cientos de organismos de forma independiente.

- **Mejora en la capacidad de filtrado:** Al analizar todo el tráfico de correo de la administración pública, la detección de spam y correos maliciosos se vuelve significativamente más efectiva. Actualmente, si un organismo A detecta un correo malicioso y lo bloquea, un organismo B podría recibirlo posteriormente sin detectarlo. Con una solución unificada esta situación se elimina, ya que al identificar un correo malicioso se bloquearían automáticamente los siguientes envíos similares para toda la *APN*.

- **Especialización del personal:** Al centralizarse la administración de esta solución tecnológica, se podría conformar un equipo de expertos altamente especializados en esta materia. Esta situación contrasta con la realidad actual en la mayoría de los organismos, donde el mismo personal es responsable de administrar múltiples soluciones de seguridad simultáneamente, perdiendo profundidad y especificidad en cada área.

- **Visibilidad de amenazas a nivel nacional:** Permite tener una vista consolidada de las campañas de phishing y ataques dirigidos contra el Estado, facilitando la identificación de patrones de ataque coordinados contra múltiples organismos.

- **Respuesta rápida ante incidentes:** Ante la detección de una campaña maliciosa dirigida a la administración pública, se puede implementar una respuesta coordinada y simultánea en todos los organismos, reduciendo significativamente la ventana de exposición.

- **Inteligencia de amenazas compartida:** Los indicadores de compromiso identificados en cualquier organismo se aplican automáticamente a toda la APN, creando un sistema de defensa colaborativo donde el conocimiento de amenazas se distribuye instantáneamente.

- **Métricas y reportes unificados:** Facilita la generación de estadísticas consolidadas sobre amenazas, intentos de ataque y efectividad de las medidas de seguridad a nivel nacional, información valiosa para la toma de decisiones estratégicas y la asignación de recursos.

- **Simplificación del cumplimiento normativo:** Facilita la implementación y auditoría de estándares de seguridad y normativas de protección de datos de manera homogénea en toda la administración, reduciendo los costos y complejidad asociados al cumplimiento regulatorio.

- **Actualizaciones:** Las mejoras, parches de seguridad y actualizaciones de la solución se implementan una sola vez y benefician simultáneamente a todos los organismos, garantizando que toda la administración cuente siempre con las últimas protecciones disponibles.

- **Reducción de la superficie de ataque:** Al contar con un punto centralizado de control, resulta más sencillo mantener configuraciones seguras y evitar inconsistencias que puedan ser explotadas por atacantes, minimizando las vulnerabilidades derivadas de configuraciones heterogéneas.

5.8.2. Construcción de un antispam nacional

Otra opción, considerablemente más ambiciosa pero técnicamente factible, es la construcción completa de una solución de seguridad de correo electrónico por parte del Estado Nacional. Este camino es viable gracias a la altísima calidad de los profesionales argentinos en el campo de las ciencias de la computación. Para lograr resultados exitosos resulta imprescindible conformar un equipo altamente calificado con salarios competitivos a nivel internacional, ya que de otra manera no sería posible la retención del talento necesario.

Las ventajas de esta opción son las siguientes:

- **Retorno de la inversión:** Si bien inicialmente se requeriría invertir en el desarrollo de la solución, posteriormente se evitaría la adquisición de cientos de licencias comerciales de seguridad de correo, generando un ahorro sostenido a largo plazo.
- **Escalabilidad del conocimiento:** Una vez finalizada la construcción de la solución, el equipo de ingeniería podría abordar otros proyectos con características similares, tales como soluciones de correo electrónico, herramientas de seguridad adicionales u otros desarrollos estratégicos.
- **Soberanía tecnológica:** Representaría un aporte sustancial a la independencia tecnológica del país, reduciendo la dependencia de proveedores extranjeros para infraestructura crítica.
- **Retención de talento:** Se ofrecería una opción de proyecto desafiante y de alto impacto para los profesionales que habitualmente emigran en busca de desafíos técnicos más complejos y mejores condiciones salariales, contribuyendo así a frenar la fuga de cerebros en el sector tecnológico.
- **Personalización a necesidades específicas:** Permite adaptar la solución a las particularidades y requerimientos específicos de la administración pública argentina, incluyendo regulaciones locales, características del español rioplatense y particularidades de las amenazas regionales.
- **Independencia de proveedores extranjeros:** Elimina la dependencia de empresas internacionales, evitando riesgos asociados a cambios en políticas comerciales, discontinuación de productos, aumentos abusivos de precios o restricciones geopolíticas.
- **Control total sobre datos sensibles:** Garantiza que toda la información y metadatos del correo gubernamental permanezcan bajo control nacional, sin riesgo de acceso por parte de terceros, jurisdicciones extranjeras o cumplimiento de legislaciones foráneas como el CLOUD Act.
- **Generación de conocimiento técnico local:** El desarrollo crea capacidades técnicas especializadas dentro del Estado que pueden aplicarse a otros proyectos críticos de ciberseguridad, transformación digital e infraestructura tecnológica nacional.

- **Ecosistema de innovación:** Puede servir como base para el desarrollo de startups o spin-offs que ofrezcan servicios complementarios, generando un ecosistema tecnológico local y oportunidades de empleo calificado en el sector.
- **Flexibilidad y tiempos de respuesta:** Ante nuevas amenazas o necesidades específicas, el equipo local puede desarrollar respuestas personalizadas y actualizaciones urgentes sin depender de roadmaps comerciales o prioridades de proveedores externos.
- **Integración optimizada:** Facilita la integración profunda con otros sistemas del Estado, tales como plataformas SIEM, sistemas de gestión de identidades, herramientas de respuesta a incidentes y otros componentes de la infraestructura de ciberseguridad nacional, sin las limitaciones típicas de soluciones comerciales cerradas.
- **Valor estratégico a largo plazo:** El conocimiento y las capacidades desarrolladas permanecen en el país y pueden evolucionar continuamente según las necesidades nacionales, sin quedar obsoletas por decisiones comerciales de terceros.
- **Posibilidad de comercialización regional:** Una vez madura y probada, la solución podría ofrecerse a otros países de la región enfrentando problemáticas similares, generando ingresos que financien su mantenimiento, evolución y posicionando a Argentina como referente tecnológico regional.
- **Transparencia y confianza institucional:** Al ser una solución desarrollada localmente, puede auditarse públicamente si se opta por un modelo de código abierto, generando mayor confianza en instituciones y ciudadanos respecto al manejo de las comunicaciones gubernamentales.

6. CONCLUSIONES

La implementación exitosa de un sistema antispam con capacidades avanzadas de sandboxing y alta disponibilidad en un organismo público, incluyendo la migración completa desde una solución legacy obsoleta, demuestra múltiples conclusiones de valor tanto técnico como organizacional.

6.1. Viabilidad de actualización de infraestructuras críticas

El proyecto valida empíricamente que es posible actualizar infraestructuras críticas de seguridad sin comprometer la continuidad operacional ni la postura de seguridad durante la transición, siempre que se apliquen metodologías apropiadas.

Los resultados cuantitativos respaldan contundentemente esta conclusión. Durante las dieciseis semanas de migración activa, el proyecto experimentó cero interrupciones de servicio y cero incidentes de seguridad atribuibles al proceso de migración. La disponibilidad del sistema se mantuvo en un noventa y nueve punto noventa y ocho por ciento durante y después de la migración, superando incluso los objetivos originales del proyecto. Más significativamente aún, la detección de amenazas mejoró en un setenta y cinco por ciento en comparación con el sistema original, demostrando que la migración no solo preservó sino que activamente mejoró la postura de seguridad del organismo.

Esta demostración es particularmente relevante para el sector público, donde frecuentemente existen percepciones arraigadas de que actualizar sistemas en producción es demasiado riesgoso y que lo que funciona no se debe tocar. Estas percepciones, aunque comprensibles dado el contexto de restricciones presupuestarias y la criticidad de los servicios públicos, pueden llevar a una inercia tecnológica peligrosa. El proyecto demuestra que estas percepciones, sin ser infundadas, pueden superarse mediante planificación adecuada, metodología rigurosa y ejecución disciplinada.

El éxito del proyecto establece que la modernización de infraestructura crítica, lejos de ser imposible o prohibitivamente riesgosa, es no solo factible sino también imperativa. Los sistemas tecnológicos inevitablemente envejecen y eventualmente alcanzan su fin de vida útil. La capacidad organizacional de ejecutar migraciones sin interrupciones es, por lo tanto, una competencia estratégica fundamental para cualquier organismo que dependa de infraestructura tecnológica crítica, lo cual en el contexto actual incluye prácticamente a todas las organizaciones públicas y privadas.

6.2. Valor de la estrategia de migración gradual

La estrategia de migración gradual con operación en paralelo se valida como superior a enfoques de reemplazo en un solo paso para sistemas críticos, a pesar de su mayor duración y complejidad aparente. Esta conclusión tiene implicaciones importantes para la planificación de proyectos similares en otras organizaciones.

Los beneficios observados de la estrategia gradual fueron múltiples y significativos. La detección y corrección temprana de problemas, antes de que afectaran a todos los usuarios, resultó invaluable. Durante las primeras fases de la migración, se identificaron y resolvieron aproximadamente treinta y dos problemas de configuración o ajuste que, en un enfoque de

reemplazo instantáneo, habrían impactado a la totalidad de los usuarios simultáneamente. La validación continua de que la nueva solución alcanzaba y superaba las capacidades de la original proporcionó confianza progresiva a todos los integrantes del equipo involucrado, desde el equipo técnico hasta la dirección.

Igualmente importante fue la capacidad de rollback mantenida durante todo el proceso. En cualquier punto de la migración, si se hubiera detectado un problema crítico, el sistema podía revertir a la configuración anterior sin pérdida de seguridad ni interrupción del servicio. Esta opción, aunque nunca fue necesaria, proporcionó tranquilidad crucial al equipo del proyecto.

El costo adicional en tiempo debe evaluarse en su contexto completo. La migración gradual requirió dieciseis semanas en comparación con las dos o tres semanas estimadas para un enfoque de reemplazo instantáneo. Un solo incidente de seguridad significativo durante una migración mal ejecutada habría costado más, en términos económicos, reputacionales y de confianza organizacional, que las catorce semanas adicionales invertidas en un enfoque gradual. Los costos directos de un incidente mayor, incluyendo respuesta técnica, recuperación de sistemas, comunicación de crisis y potenciales implicaciones legales, habrían excedido largamente cualquier ahorro en tiempo del proyecto.

La lección estratégica es clara: para sistemas críticos, el tiempo invertido en mitigación de riesgos mediante estrategias graduales es siempre tiempo bien invertido. La velocidad de implementación nunca debe priorizarse sobre la mitigación de riesgos en infraestructuras críticas. Las organizaciones deben resistir presiones para acelerar artificialmente procesos de migración cuando la naturaleza crítica del sistema justifica un enfoque más conservador y controlado.

6.3. Importancia del enfoque de defensa en profundidad

El proyecto confirma empíricamente que la combinación de múltiples técnicas de detección proporciona una defensa significativamente más robusta que cualquier técnica individual, por avanzada que sea. Esta validación de la filosofía de defensa en profundidad tiene implicaciones importantes para el diseño de arquitecturas de seguridad.

Los datos operacionales demuestran la complementariedad de las diferentes técnicas. Durante los primeros seis meses post-implementación, el 15,2 % de las amenazas fueron detectadas únicamente por sandboxing, lo que significa que habrían evadido completamente tanto las firmas tradicionales como el análisis heurístico. El 44,9 % por ciento de las amenazas fueron detectadas eficientemente por firmas tradicionales mientras que el 39,9 % restante fue detectado por análisis heurístico, demostrando que las técnicas establecidas continúan siendo relevantes y efectivas para amenazas conocidas.

Estos datos refutan la idea de que técnicas modernas como el sandboxing reemplazan completamente a técnicas tradicionales como la detección por firmas. En realidad, cada técnica cubre las debilidades inherentes de las otras, y todas contribuyen al resultado de seguridad final de manera complementaria. Las firmas proporcionan detección eficiente y de baja latencia para amenazas conocidas. El análisis heurístico identifica variantes nuevas de amenazas conocidas mediante características sospechosas. El sandboxing captura amenazas completamente nuevas y ataques dirigidos sofisticados diseñados específicamente para evadir otras técnicas de detección.

El proyecto demuestra que la inversión en arquitecturas de defensa en profundidad, combinando múltiples técnicas de detección, representa un imperativo de seguridad en el

contexto actual de amenazas cibernéticas cada vez más sofisticadas.

La filosofía de defensa en profundidad reconoce que ninguna técnica de detección es perfecta y que todas tienen limitaciones y puntos ciegos. Al combinar múltiples técnicas con fortalezas y debilidades diferentes, se crea una defensa donde el fallo de una capa no compromete la seguridad total del sistema. Esta redundancia defensiva, lejos de ser innecesaria, es esencial en un panorama de amenazas donde los atacantes diseñan activamente técnicas de evasión específicas para cada tipo de defensa.

6.4. Replicabilidad del modelo en otras organizaciones

La estrategia desarrollada en este proyecto es altamente replicable en otras instituciones que enfrenten desafíos similares de actualización tecnológica con preservación de conocimiento organizacional. Sin embargo, la replicabilidad no implica una aplicación literal sino una adaptación inteligente al contexto específico de cada organización.

Los elementos transferibles del proyecto incluyen su metodología fundamental de migración gradual. La secuencia de fases desarrollada, comenzando con la instalación en paralelo, continuando con el análisis y migración de reglas y culminando con la desactivación del sistema original, es aplicable a diversos contextos de migración de sistemas críticos más allá del dominio específico del correo electrónico. El mecanismo de análisis comparativo, que involucra comparar continuamente las decisiones de ambos sistemas, analizar discrepancias y refinar configuraciones, proporciona un mecanismo robusto para validar la equivalencia o superioridad funcional del nuevo sistema.

Las métricas de éxito definidas en el proyecto proporcionan un modelo medible y objetivo para evaluar el éxito en proyectos similares. Los indicadores clave de rendimiento utilizados, incluyendo convergencia de detección entre sistemas, tasas de falsos positivos y falsos negativos, disponibilidad del servicio y experiencia de usuario, ofrecen un framework para medir tanto el éxito técnico como el organizacional de iniciativas de migración.

Diferentes organizaciones enfrentan diferentes perfiles de amenaza basados en su sector, visibilidad pública y valor de sus activos informáticos. Las reglas personalizadas y configuraciones de seguridad deben reflejar el contexto específico de riesgo de cada organización. Las restricciones presupuestarias variarán significativamente entre organizaciones, requiriendo diferentes balances entre capacidades técnicas deseables y costos realistas. El marco regulatorio específico de cada sector o jurisdicción puede imponer requisitos de compliance adicionales que necesiten configuraciones o controles específicos.

Aunque este proyecto se ejecutó en un organismo público, las lecciones aprendidas son igualmente valiosas para el sector privado. Empresas de sectores regulados como servicios financieros, salud o energía enfrentan desafíos similares de alta criticidad de servicios y requisitos estrictos de disponibilidad. Organizaciones de cualquier sector acumulan conocimiento organizacional en sistemas legacy que necesitan preservar durante migraciones tecnológicas. Los principios de defensa en profundidad, arquitecturas resilientes y gestión rigurosa del cambio son universales, trascendiendo las diferencias entre sector público y privado.

6.5. Evolución continua de la seguridad

Una conclusión fundamental del proyecto es que la seguridad no es un estado final que se alcanza y se mantiene, sino un proceso continuo de evolución y mejora. Esta perspectiva

tiene implicaciones importantes para cómo las organizaciones conceptualizan, planifican y presupuestan sus iniciativas de seguridad.

El proyecto no resolvió permanentemente la seguridad del correo electrónico del organismo en ningún sentido final o definitivo. En lugar de eso, estableció una plataforma moderna, capaz y flexible sobre la cual se continuará construyendo y mejorando. Esta distinción es crítica: el objetivo realista de proyectos de seguridad no es alcanzar un estado de seguridad perfecta y permanente, sino establecer capacidades robustas que puedan evolucionar continuamente para enfrentar amenazas cambiantes.

La planificación de mejoras futuras contempló desarrollos ambiciosos que continuarían expandiendo las capacidades del sistema. Se propuso la implementación de capacidades de machine learning para detección más avanzada de anomalías y patrones sutiles que podrían indicar ataques sofisticados. Se planeó expandir las capacidades de sandboxing para incluir análisis de documentos y archivos más sofisticados, incluyendo macros ofuscadas y técnicas avanzadas de empaquetamiento.

Para el largo plazo, el organismo está considerando la participación en iniciativas de compartición de threat intelligence con otros organismos públicos, reconociendo que las amenazas contra el sector público frecuentemente son campañas coordinadas que afectan a múltiples instituciones. Se están explorando posibles expansiones del sistema para cubrir otros vectores de amenaza más allá del correo electrónico, como navegación web, mensajería colaborativa y transferencia de archivos.

Esta visión de la seguridad como proceso continuo tiene implicaciones importantes para la planificación organizacional y presupuestación. Los presupuestos de seguridad no deben verse como gastos únicos asociados con proyectos específicos, sino como inversiones continuas necesarias para mantener y mejorar la postura de seguridad a lo largo del tiempo. Las arquitecturas de soluciones deben diseñarse con flexibilidad y extensibilidad incorporadas desde el inicio, para acomodar evoluciones futuras sin requerir rediseños completos. Los equipos técnicos necesitan tiempo dedicado no solo a operación y mantenimiento sino también a investigación, aprendizaje e implementación de mejoras continuas. Las métricas de éxito organizacional deben incluir no solo estabilidad operacional y ausencia de incidentes sino también capacidad demostrada de evolución y adopción de nuevas capacidades.

6.6. Preservación y mejora del conocimiento organizacional

Un descubrimiento significativo del proyecto fue el estado del conocimiento organizacional almacenado en el sistema original y las lecciones sobre su preservación. Las organizaciones acumulan conocimiento invaluable en sus sistemas técnicos a lo largo de años de operación y refinamiento, pero frecuentemente este conocimiento carece de documentación adecuada sobre su propósito, contexto y lógica.

El sistema original contenía ciento cuarenta y siete reglas personalizadas, representando años de refinamiento y adaptación a las necesidades específicas del organismo. Sin embargo, aproximadamente el 60 % de estas reglas carecían completamente de documentación sobre su propósito o contexto de creación. La reconstrucción del conocimiento detrás de estas reglas requirió análisis forense extensivo, consumiendo aproximadamente el 30 % del tiempo total del proyecto. Doce reglas permanecieron parcialmente opacas incluso después de análisis detallado, y ocho se determinaron completamente obsoletas pero habían continuado activas consumiendo recursos.

El proyecto no solo preservó el conocimiento organizacional existente sino que ac-

tivamente lo mejoró. De las 147 reglas originales, 115 se migraron exitosamente, 32 se determinaron obsoletas y se retiraron, y 18 reglas completamente nuevas se crearon durante el proceso de migración, resultando en un total de 133 reglas activas en el nuevo sistema. Más importante aún, todas estas reglas están ahora completamente documentadas con información sobre propósito, contexto, criterios técnicos, responsables y historial de modificaciones.

La lección crítica es que la documentación del conocimiento organizacional debe ser un proceso continuo integrado en la operación diaria, no un ejercicio retrospectivo realizado durante migraciones. Toda configuración personalizada en sistemas de producción debe documentarse en el momento de su creación, capturando el contexto y la lógica mientras están frescos en la mente de los creadores. Esta documentación debe incluir elementos esenciales como el propósito o problema que la configuración intenta resolver, el contexto organizacional o de amenaza que motivó su creación, los criterios técnicos de activación, el responsable de la creación y el historial de modificaciones.

6.7. Síntesis: Un modelo replicable de transformación responsable

En síntesis final, este proyecto demuestra que es posible transformar infraestructuras críticas de seguridad, preservando y mejorando capacidades existentes, sin comprometer la seguridad o la operación continua, incluso en entornos con restricciones significativas como el sector público. Los resultados cuantitativos son contundentes: mejora del 75 % en detección de amenazas, reducción del 70 % en falsos positivos, cero interrupciones de servicio y cero compromisos de seguridad durante diecisiete semanas de migración activa.

Más importante que las métricas específicas es la demostración de que la transformación responsable y bien ejecutada es posible, deseable y alcanzable. En un contexto donde las amenazas cibernéticas continúan evolucionando en sofisticación y donde la infraestructura tecnológica existente inevitablemente envejece, la capacidad organizacional de modernizar sin romper es una competencia crítica.

Los principios fundamentales validados por el proyecto trascienden su contexto específico y ofrecen guía para cualquier organización enfrentando desafíos similares. El tiempo invertido en mitigación de riesgos mediante estrategias graduales es siempre tiempo bien invertido en sistemas críticos. La documentación del conocimiento organizacional no es opcional sino esencial para la continuidad operacional a largo plazo. La inversión en alta disponibilidad se justifica por los costos evitados de interrupciones y por la confiabilidad del servicio. La defensa efectiva requiere múltiples capas complementarias sin depender de ninguna técnica única. Los aspectos humanos y organizacionales son tan críticos como los aspectos técnicos para el éxito del proyecto. La seguridad es un viaje continuo de mejora, no un destino final que se alcanza y se mantiene estático.

El legado más valioso de este proyecto no son los sistemas específicos implementados, que eventualmente también requerirán actualización, sino la demostración de que la transformación responsable de infraestructura crítica es alcanzable cuando se combinan metodología rigurosa, ejecución disciplinada, gestión efectiva del cambio y un compromiso genuino con la preservación de capacidades durante la modernización. Este modelo de transformación responsable es replicable y puede servir como referencia para otras organizaciones, tanto en el sector público como privado, que enfrenten desafíos similares de actualización tecnológica en los años venideros.

Bibliografía

[1] Jefe de Gabinete de Ministros de la Nación Argentina. (2024). Informe anual de gestión de incidentes de ciberseguridad 2024. https://www.argentina.gob.ar/sites/default/files/2025/07/informe_cert-ar_2024.pdf

[2] Instituto Nacional de Ciberseguridad de España. (2022). Los 10 vectores de ataque más utilizados por los ciberdelincuentes. https://www.incibe.es/empresas/blog/los-10-vectores-ataque-mas-utilizados-los-ciberdelincuentes

[3] Datacenterdynamics. (2015). PAMI certifica su data center con la ISO 27001. https://www.datacenterdynamics.com/es/noticias/pami-certifica-su-data-center-con-la-iso-27001

[4] Secretaría de Innovación, Ciencia y Tecnología. (2019). Recibimos la certificación ISO 27001 que fortalece el Sistema de Gestión de Seguridad Informática. https://www.argentina.gob.ar/noticias/recibimos-la-certificacion-iso-27001-que-fortalece-el-sistema-de-gestion-de-seguridad

[5] Secretaría de Innovación, Ciencia y Tecnología. (2022). Modelo Referencial de Política de Seguridad de la Información. https://www.argentina.gob.ar/noticias/elaboran-modelo-de-politica-de-seguridad-de-la-informacion-para-organismos-publicos

[6] Rosario Marina. (2024). Filtración de datos personales en el Renaper: ¿qué es y qué consecuencias puede tener? https://chequeado.com/el-explicador/filtracion-de-datos-personales-en-el-renaper-que-es-y-que-consecuencias-puede-tener/

[7] Keith Barker. (2014). The gap between real and perceived security risks. https://www.sciencedirect.com/science/article/pii/S1361372314704786

[8] Kate Fazzini. 2019. In a decade of cybersecurity alarms, these are the breaches that actually mattered. https://www.cnbc.com/2019/12/23/stuxnet-target-equifax-worst-breaches-of-2010s.html

[9] Breachsense. (2024). Equifax Data Breach Explained: A Case Study https://www.breachsense.com/blog/equifax-data-breach/

[10] Saheed Oladimeji, Sean Michael Kerner. (2023). SolarWinds hack explained: Everything you need to know https://www.techtarget.com/whatis/feature/SolarWinds-hack-explained-Everything-you-need-to-know

[11] Terry Thompson. (2021). The Colonial Pipeline ransomware attack and the SolarWinds hack were all but inevitable – why national cyber defense is a 'wicked' problem. https://theconversation.com/the-colonial-pipeline-ransomware-attack-and-the-solarwinds-hack-were-all-but-inevitable-why-national-cyber-defense-is-a-wicked-problem-160661

[12] Alex Scroxton. (2024). More data stolen in 2023 MOVEit attacks comes to light. https://www.computerweekly.com/news/366615522/More-data-stolen-in-2023-MOVEit-attacks-comes-to-light

[13] Dirk Schrader. (2025). An Overview of the MGM Cyber Attack. https://netwrix.com/en/resources/blog/mgm-cyber-attack/

[14] Dirección Nacional de Ciberseguridad. (2024). Informe anual de gestión de incidentes de ciberseguridad. https://www.argentina.gob.ar/sites/default/files/2025/07/informe_cert-ar_2024.pdf

[15] Marcela Pallero. (2025). Incidentes de ciberseguridad relevantes de Argentina desde 2017. https://time.graphics/es/line/630567

[16] Boletin Oficial de la Republica Argentina. (2023). https://www.boletinoficial.gob.ar/detalleAviso/primera/278850/20230102 (Anexo 1)

[17] Oficina Nacional de Tecnologías de la Información. https://www.argentina.gob.ar/jefatura/innovacion-ciencia-y-tecnologia/onti/elaboracion-de-estandares-y-dictamenes-tecnologicos

[18] Jefatura de gabinete de ministros. Oficina Nacional de Contrataciones. https://comprar.gob.ar/BuscarAvanzado.aspx

[19] Verizon. (2025). Data Breach Investigations Report. https://www.verizon.com/business/resources/reports/dbir/

[20] AV-TEST. (2024). Software malicioso. https://www.av-test.org/es/estadisticas/software-malicioso/

[21] SecurityWeek. 2015. Intel Announces EoL for McAfee Email Security Products https://www.securityweek.com/intel-announces-eol-mcafee-email-security-products/