



Centro Universitario de la Defensa en la Escuela Naval Militar

TRABAJO FIN DE MÁSTER

Aplicación de Operaciones Red Team en Ciberejercicios: Metodología, Simulación y Propuesta de Marco para la Defensa Nacional

Máster Universitario en Dirección TIC para la Defensa

ALUMNO: José Ángel Cibeira Urtiaga

DIRECTORES: Rubén Nocelo López

Carlos Zamorano Pinal

CURSO ACADÉMICO: 2025-2026

UniversidadeVigo



Centro Universitario de la Defensa en la Escuela Naval Militar

TRABAJO FIN DE MÁSTER

*Aplicación de Operaciones Red Team en Ciberejercicios:
Metodología, Simulación y Propuesta de Marco para la Defensa
Nacional*

Máster Universitario en Dirección TIC para la Defensa
Especialidad de Sistemas y Tecnologías de Telecomunicación

UniversidadeVigo

RESUMEN

El Trabajo Fin de Máster analiza cómo integrar de forma sistemática las operaciones de Red Team dentro de ciberejercicios orientados a la Defensa Nacional, con el objetivo de mejorar la resiliencia operativa del Estado frente a amenazas avanzadas en un entorno caracterizado por la hiperconectividad, las operaciones multidominio y la denominada zona gris. El trabajo parte de la premisa de que las auditorías tradicionales y los pentests resultan insuficientes para evaluar la capacidad real de detección, respuesta y recuperación ante adversarios persistentes, por lo que propone un enfoque proactivo, realista y basado en la emulación del adversario.

Tras una revisión del estado del arte (auditoría, pentesting, Red Team y Purple Team), de los ciberejercicios y de los marcos internacionales de referencia (TIBER-EU, CBEST y MITRE ATT&CK), el estudio diseña una metodología integral para planificar, ejecutar y evaluar ejercicios Red Team en entornos autorizados. Dicha metodología se estructura en fases (planificación, ejecución, control y evaluación), incorpora inteligencia de amenazas para seleccionar TTPs realistas, y define escenarios representativos (infraestructuras críticas, redes gubernamentales, sistemas de mando y control y cadena logística digital). Asimismo, establece reglas de enfrentamiento (RoE), salvaguardas legales y éticas, y requisitos de instrumentación (EDR, SIEM, registros y telemetría) para asegurar trazabilidad, reproducibilidad y validez de los resultados.

El trabajo propone métricas y KPIs para medir la eficacia defensiva (detección, tiempo de respuesta, impacto, disponibilidad y madurez), y articula un ciclo de mejora continua basado en lecciones aprendidas, validación y retest. Como resultado, se presenta un marco nacional que integra a CSIRTs, operadores críticos y organismos públicos, con entregables estandarizados (playbooks, evidencias con cadena de custodia, informes técnicos y ejecutivos, y planes de mitigación priorizados) que facilitan la transferencia de conocimiento y la sostenibilidad de capacidades. En conjunto, el TFM demuestra que la emulación controlada de adversarios transforma hallazgos técnicos en mejoras doctrinales, tecnológicas y de gobernanza, fortaleciendo la resiliencia nacional frente a ciberamenazas sofisticadas.

PALABRAS CLAVE

Seguridad Nacional, Red Team, Ciberseguridad, Operaciones Ofensivas, Marco de Seguridad

AGRADECIMIENTOS

Como siempre, mi más sincero agradecimiento a mi familia, que siempre me ha apoyado en toda nueva aventura emprendida a lo largo de mi carrera profesional, y a aquellos que ya no pueden ver mis logros, y que siempre formarán parte de ellos.

CONTENIDO

Contenido	1
Índice de Figuras	3
Índice de Tablas.....	4
1 Introducción y objetivos	5
1.1 Justificación del trabajo	6
1.2 Objetivos generales y específicos	7
1.3 Metodología empleada en el estudio.....	7
1.4 Alcance y contexto de aplicación.....	8
1.5 Exclusiones	9
1.6 Limitaciones metodológicas	9
1.7 Contexto de aplicación.....	10
1.8 Supuestos operativos.....	10
1.9 Riesgos principales y medidas mitigadoras	10
1.10 Criterios de aplicabilidad y transferibilidad.....	11
1.11 Entregables y criterios de cierre.....	11
2 Estado del arte	12
2.1 Conceptos fundamentales: Auditoría de seguridad, Pentesting, Red Teaming y Purple Teaming	12
2.2 Ciberejercicios: definición, tipos y objetivos.....	15
2.2.1 Tipos de ciberejercicios	16
2.2.2 Objetivos.....	17
2.3 Marcos internacionales de referencia (TIBER-EU, CBEST, MITRE ATT&CK)	18
2.3.1 TIBER-EU: estándar europeo para red-teaming intelligence-led.....	19
2.3.2 CBEST: enfoque británico.....	20
2.3.3 MITRE ATT&CK	22
2.4 Herramientas y frameworks para ejercicios red team en contextos nacionales	23
2.4.1 Cyberkill Chain.....	24
2.4.2 Herramientas empleadas en ciberejercicios Red Team	26
2.4.3 Experiencias internacionales aplicadas a la defensa nacional	28
3 PROPUESTA.....	33
3.1 Metodología propuesta para operaciones Red Team en ciberejercicios	33
3.1.1 Fases de planeamiento, ejecución y evaluación	34
3.1.2 Diseño de escenarios representativos (infraestructura crítica, sistemas militares, redes gubernamentales).....	41

3.1.3 Técnicas, tácticas y procedimientos (TTPs) adaptados al contexto nacional español.....	43
3.1.4 Criterios de éxito y medición del impacto defensivo	44
3.2 Planificación y ejecución de ciberejercicios en entornos autorizados	48
3.2.1 Requerimientos técnicos y operativos	49
3.2.2 Ejecución de escenarios seleccionados, recolección y análisis de datos durante los ejercicios.....	53
3.2.3 Lecciones aprendidas y ciclo de mejora (validación, retest y transferencia).....	56
3.3 Propuesta de marco para la Defensa Nacional.....	57
3.3.1 Elementos clave del marco nacional para ejercicios Red Team.....	58
3.3.2 Integración con capacidades nacionales de ciberdefensa (CSIRTs y operadores críticos)	59
3.3.3 Rol de los ciberejercicios en la mejora continua de la resiliencia nacional.....	60
4 Resultados / Validación / Prueba.....	62
4.1 Análisis teórico de resultados	62
4.1.1 Evaluación de las capacidades defensivas nacionales frente a los ejercicios (enfoque proyectado).....	62
4.1.2 Propuesta de métricas de desempeño: definición, cálculo y criterios de interpretación...	64
4.1.3 Identificación de gaps y oportunidades de mejora.....	67
4.1.4 Impacto estratégico y operacional del uso de Red Team en ejercicios nacionales.....	68
5 Conclusiones y líneas futuras	70
5.1 Conclusiones generales	70
5.2 Limitaciones del trabajo.....	70
5.3 Líneas de investigación y desarrollo futuro	71
6 Bibliografía.....	73
Anexo A: Colección de métricas propuestas, incluyendo hollis, 2024.....	77
Anexo B: Check list de preejecución de un ciberejercicio ofensivo u operación de Red team	79
Anexo C: Glosario de términos y expresiones en inglés.....	82

ÍNDICE DE FIGURAS

Figura 1-1 Ataques DDoS en Europa. Fuente: ENISA	6
Figura 2-1 Ciberejercicios propuestos por ENISA.....	16
Figura 2-2 Factores para crear un ejercicio tabletop. Fuente: Red Goat Cyber Security	18
Figura 2-3 Fase de preparación de un ejercicio de Red Team. TIBER-EU	20
Figura 2-4 Fases de recolección de inteligencia. Fuente: Bank of England.....	21
Figura 2-5 Modelo de Cyber Kill Chain [24].	24
Figura 2-6 Modelo de Cyber Kill Chain de 18 fases [27].	26
Figura 2-7 Interface de operación de Cobalt Strike durante un ejercicio de Red Team.....	28
Figura 2-8 Personal implicado en el ejercicio Locked Shields 25 en su sede en Estonia. Fuente: CCdCOE.	30
Figura 2-9 Personal involucrado en la ejecución del ejercicio Crossed Swords. Fuente: CCdCOE.	32
Figura 3-1 Ejemplo de TTP's mapeadas en la matriz de MITRE. Se diferencian las tácticas por el código TA de las técnicas por el código T	36
Figura 3-2 Ejemplo de infraestructura levantada para un ejercicio de Red Team [48]......	39
Figura 3-3 Flujo de trabajo de un ejercicio de Red Team [49].	42
Figura 3-4 Diagrama de flujo de conversión de elementos de ciberinteligencia en TTP's [50].	44
Figura 3-5 Diseño de infraestructura ofensiva basada en la nube para ejercicios de Red Team.	53
Figura 3-6 Organigrama de un ejercicio Red Team.	55
Figura 3-7 Integrantes del Consejo Nacional de Ciberseguridad. Fuente: dsn.gob.es	61

ÍNDICE DE TABLAS

Tabla 2-1 Conceptos fundamentales de seguridad	15
Tabla 2-2 Tabla comparativa entre los diferentes marcos de referencia	23
Tabla 3-1 Valores asignados en función de la criticidad de la organización.	46
Tabla 3-2 Valores asignados en función del nivel de madurez de la ciberseguridad de la organización.	47
Tabla 3-3 Valores asignados en función del grado de disponibilidad de los servicios.	47
Tabla 3-4 Valores asignados en función del número de empleados.	48
Tabla 3-5 Medidas recomendadas en el aspecto de identidad y accesos según las guías STIC.....	51
Tabla 3-6 Medidas recomendadas en el aspecto de gestión de credenciales y accesos según las guías STIC.....	51
Tabla 3-7 Medidas recomendadas en el aspecto de gestión de puertos y servicios según las guías STIC.	52
Tabla 4-1 Matriz de cálculo del valor de Readiness o disponibilidad.....	64

1 INTRODUCCIÓN Y OBJETIVOS

Los sistemas TIC (Tecnologías de la Información y las comunicaciones) han experimentado un desarrollo exponencial en las últimas décadas. Actualmente vivimos en un mundo hiperconectado que ha transformado la superficie de exposición de los estados y sus Fuerzas Armadas. La creciente complejidad operativa, los entornos multidominio y las operaciones en Zona Gris han ampliado la probabilidad de sufrir ciberataques dirigidos contra infraestructuras críticas, sistemas de mando y control, ataques a la cadena logística, así como servicios esenciales. Bajo este nuevo paradigma, la ciberseguridad debe enfocarse en medidas proactivas, dejando atrás las reactivas. Un enfoque proactivo y realista que permita validar las capacidades defensivas mediante acciones activas, yendo más allá de la identificación de vulnerabilidades, en aras de una mejor resiliencia operativa en escenarios de amenaza avanzada, especialmente si se tiene en cuenta la cantidad de incidentes de ciberseguridad que suceden cada día (figura 1-1).

Los test de estrés en el ámbito de la ciberseguridad, particularmente las operaciones de Red Team [1], ofrecen un marco práctico para probar sistemas, procedimientos, equipos y personas bajo condiciones realistas que emulan de una forma muy veraz a los adversarios reales. Si bien las acciones llevadas a cabo por los Blue Team [1] se focalizan en la detección, respuesta y mitigación (acciones meramente reactivas), un Red Team permite mirar a través del prisma del atacante (emulando sus TTP's tácticas, técnicas y procedimientos) forzando a las organizaciones a enfrentar escenarios complejos más allá de las auditorías tradicionales. Existe también la posibilidad de una fusión de los dos enfoques. Esta metodología se conoce como Purple Team [2], donde los equipos defensivos y atacantes trabajan de forma coordinada, lo que favorece el hallazgo y las mejoras operacionales, procedimentales, doctrinales y tecnológicas.

El presente trabajo de fin de máster se diseña, simula y evalúan operaciones de Red Team en el marco de ejercicios orientados a la defensa nacional. Para ello, se plantea una metodología adaptada al contexto nacional, considerando sus particularidades para plantear escenarios representativos (ajustados a las infraestructuras críticas, sistemas de defensa y redes gubernamentales) en donde se ejecutarán simulaciones controladas con el fin de generar un marco operativo que haga viable la incorporación sostenida en el tiempo de estas acciones a la planificación de la ciberseguridad nacional. Con todo esto, se plantea contribuir de forma práctica y aplicable a la mejora de la resiliencia del país frente ciberamenazas avanzadas.

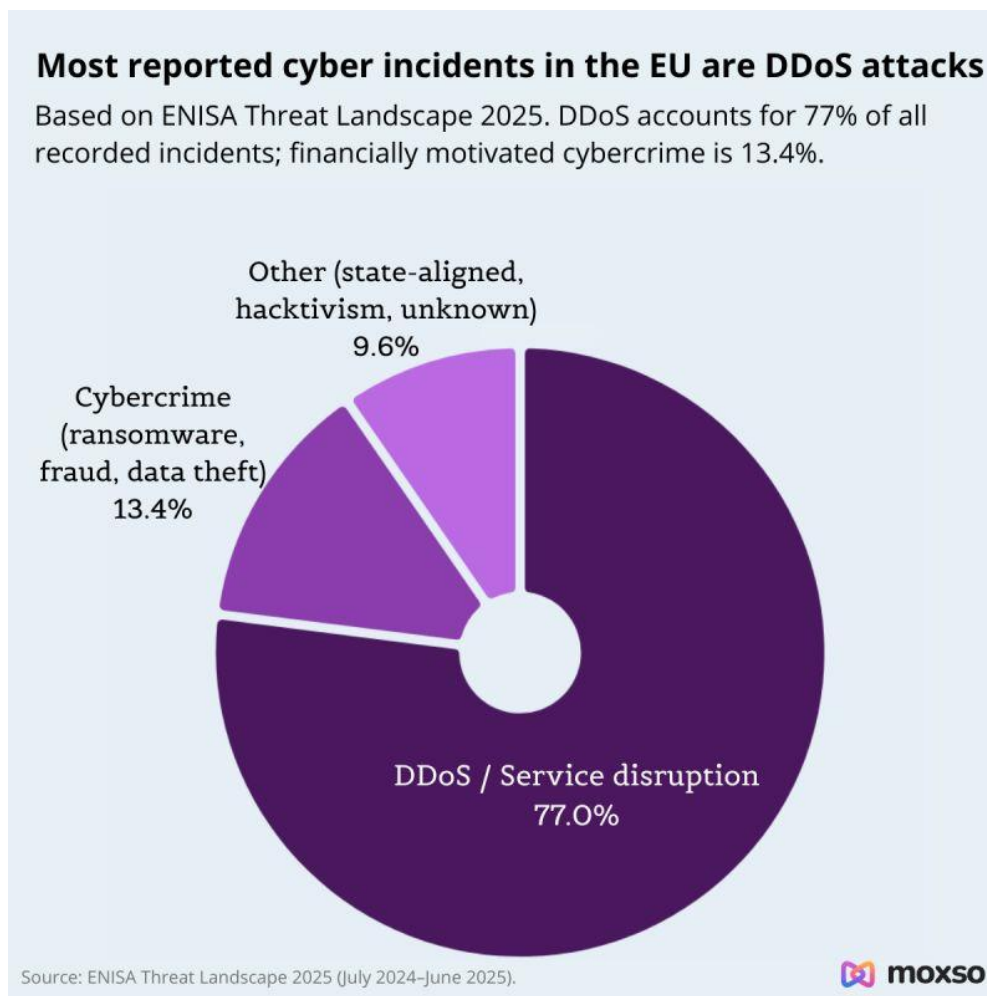


Figura 1-1 Ataques DDoS en Europa. Fuente: ENISA

1.1 Justificación del trabajo

Cada vez más los servicios estatales están digitalizados, favoreciendo su interconexión entre ellos y otros sistemas críticos, modificando de manera radical el escenario de la seguridad nacional. Las amenazas a la ciberseguridad son algo más que incidentes aislados. Han mutado en vectores de ataque capaces de degradar las capacidades operativas, cortar cadenas logísticas y comprometer los sistemas de mando y control. Bajo este escenario, las auditorías clásicas y los controles pasivos son manifiestamente insuficientes. Es necesario someter a las defensas a simulaciones de adversarios que los mecanismos tradicionales no detectan. Las operaciones Red Team ofrecen ese marco realista, controlado y orientado a la mejora de los sistemas defensivos, permitiendo evaluar su madurez en condiciones similares a una acción maliciosa real.

En un escenario de simulación de una amenaza avanzada se identifican aquellas debilidades técnicas que generan información accionable para la mejora en la toma de decisiones, aportando métricas de detección y respuesta, detectando fallos en la coordinación entre instituciones y generando una base empírica que permita priorizar inversiones y detectar la mejor formación para los operadores. Este trabajo plantea aportar en este apartado, proponiendo una metodología que combine rigor técnico con salvaguardas legales y éticas y que pueda integrarse en los ciclos de mejora continua de los diferentes organismos.

Por último, las operaciones de Red Team y la formalización de los resultados en playbooks, planes de mejora continua, así como la propuesta de un marco operativo, permiten la transferencia de conocimiento a la sostenibilidad de las capacidades. Teniendo en cuenta la alta rotación del personal, la interoperabilidad entre distintas entidades, así como la disparidad de herramientas de ciberseguridad

defensiva, disponer de procedimientos estandarizados y validados mediante la experiencia, aumenta la resiliencia operativa. Es por ello que, diseñar, simular y evaluar operaciones ofensivas orientadas a la defensa nacional no solo es ventajoso desde el punto de vista técnico, sino estratégico: transforma hallazgos derivado de la emulación de adversarios en mejoras doctrinales, tecnológicas y de gobernanza que reducen el riesgo operativo y fortalecen la capacidad de respuesta del Estado.

1.2 Objetivos generales y específicos

El objetivo general de este trabajo es desarrollar y validar una metodología integral para la realización de ciberejercicios, enfocados en operaciones de Red Team en el ámbito de la Defensa Nacional, que permita evaluar y mejorar las capacidades de los sistemas defensivos empleando para ello las tácticas, técnicas y procedimientos asociados a diferentes grupos adversarios. Con esto se plantea generar unas métricas objetivas y un marco operativo realista de implementar y sostenible y mejorable en el tiempo.

Los objetivos específicos se centran en una revisión y síntesis del estado del arte sobre las operaciones Red Team, ciberejercicios y marcos de referencia internacionales (e.g. MITRE ATT&CK, <https://attack.mitre.org>) con el fin último de extraer prácticas, limitaciones y lecciones que puedan ser extrapolables al marco de Seguridad Nacional. Para ello, se definirá una metodología operativa que además sea escalable para operaciones de Red Team que incorpore una serie de fases, a saber, planificación, ejecución, control y evaluación que permita generar procedimientos reproducibles aplicables a varias acciones. Se plantea también el diseño de una serie de catálogos de escenarios representativos (infraestructuras críticas, redes gubernamentales, etc...) donde se plantearán requisitos técnicos y de seguridad. Revisión de TTP's empleadas por diferentes grupos persistentes (APT's advanced persistent threat) para su posterior empleo por parte de los equipos de Red Team. Analizar los resultados de las acciones para identificar posibles puntos de mejora en los procedimientos, tecnologías y organización de la defensa nacional desde una perspectiva de ciberdefensa. Establecer un roadmap y líneas futuras de investigación y desarrollo para la continua evolución y mejora del marco propuesto. Evaluar las implicaciones éticas, legales y de seguridad asociadas a la metodología, para lo que se propondrán medidas de mitigación como la definición del alcance de cada operación, salvaguardas de protección de datos y la difusión de los resultados.

1.3 Metodología empleada en el estudio

La metodología empleada en este trabajo plantea diseñar, simular y validar una propuesta de operaciones de Red Team en ciberejercicios con un enfoque orientado a la mejora de la Defensa Nacional. Se plantea un enfoque que combina métodos mixtos cualitativo-cuantitativo, investigación aplicada al estudio y desarrollo de TTP's y la generación de una colección de procedimientos operativos que permitan reproducir las operaciones para así garantizar su validez, trazabilidad y aplicabilidad práctica. Se definen una serie de pasos a seguir:

1. Este trabajo adopta un enfoque basado en la evidencia, por lo que se realizará una revisión de la literatura y los marcos prácticos
2. Se diseñarán procedimientos operativos
3. Se simularán pruebas de impacto y se hará un análisis cuantitativo y cualitativo de los resultados para así poder extraer lecciones aprendidas y recomendaciones de mejora.

A continuación, se diseñará el estudio experimental, con el desarrollo de un entorno de simulación. Se definirán los escenarios representativos con unas condiciones iniciales, objetivos y alcance del Red Team, así como los supuestos de defensa (SIEM, EDR y antivirus). Cada escenario incorporará una serie de variables independientes (e.g. segmentación de red, sondas monitorizando, etc...) así como unas variables dependientes que se medirán según una valoración predefinida. Entre estas variables podemos contar los parámetros de detección, tiempo de respuesta o el grado de impacto.

Esta metodología se articulada en una serie de fases, que contemplan entregables y criterios de aceptación:

1. Revisión y síntesis del estado del arte. Búsqueda y recopilación de bibliografía para su posterior análisis.
2. Definición de la metodología operativa formalizando las diferentes fases (planificación y ejecución según la cyber kill chain), procedimientos y plantillas (NDA, ROE's, matriz de riesgos).
3. Diseño y selección de casos de uso, definición de supuestos, activos críticos y estudio de vectores de entrada.
4. Ejecución de campañas simuladas de Red Team los escenarios diseñados ajustándose a las reglas de enfrentamiento.
5. Análisis de los resultados obtenidos y validación de la metodología empleando un análisis cuantitativo de las métricas obtenidas. Se hará también una evaluación cualitativa mediante la revisión de procedimientos.
6. Formalización del marco desarrollado y consolidación de los resultados materializándose en recomendaciones doctrinales y técnicas, diseño del roadmap y propuesta de KPI's para su seguimiento.
7. Recolección de datos con el objetivo de obtener la telemetría suficiente, que permita caracterizar las acciones del Red Team para la materialización de la respuesta defensiva, facilitando la actuación de los equipos de análisis forense, así como la evaluación de KPIs y reproducción del ejercicio.
8. Métricas y criterios de evaluación que permitan disponer de indicadores medibles y cuantificables que midan la efectividad de los sistemas defensivos, los efectos de las acciones ofensivas con la vista puesta en la mejora a lo largo del ciclo de ejercicios, adaptándose a la evolución de los sistemas.
9. Técnicas analíticas y herramientas para poder aplicar métodos de análisis que conviertan la telemetría en hallazgos accionables y validar hipótesis sobre comportamiento atacante y respuesta defensiva. Por ejemplo, el modelado de la amenaza para la obtención de las TTP's en base a la matriz de MITRE ATT&CK.
10. Validación y verificación asegurándose de que los hallazgos, métricas y conclusiones inferidas de los ejercicios sean correctos, reproducibles y útiles para la toma de decisiones de mejora de sistemas en base a un criterio sustentado en la experiencia.
11. Criterios de reproducibilidad y transferencia que faciliten que los ejercicios puedan reproducirse con fidelidad y que el conocimiento generado pueda transferirse a otras unidades u organismos.
12. Limitaciones metodológicas. Definir las limitaciones de la propia metodología, generadas principalmente por la artificialidad de los ejercicios en sí, y el hecho de que se emulan técnicas empleados por los adversarios que ya han sido modeladas.

1.4 Alcance y contexto de aplicación

Este trabajo se centra en el diseño, simulación y evaluación de operaciones Red Team aplicadas a ciberejercicios para su empleo en el ámbito de la defensa nacional en España. Estos ciberejercicios sería realizados en entornos controlado, previa autorización y acuerdo NDA. El alcance operativo y técnico incluye:

1. Ámbitos tecnológicos: servicios digitales de la Administración pública, infraestructuras OT/ICS emuladas (p. ej. subestaciones, SCADA), sistemas de mando y control (C2) simulados, y elementos de la cadena logística digital que afectasen a la continuidad operativa.

-
2. Tipos de ejercicios: propuesta de ejercicios Red Team controlados sobre entornos autorizados, impresionándose como actores maliciosos, ejercicios híbridos que incorporen el factor humano como la intrusión física, y ejercicios de validación de procedimientos de detección y respuesta (purple teaming).
 3. Productos entregables: playbooks para la emulación de adversario, reglas de enfrentamiento (RoE por sus siglas en inglés), planes y procedimientos, registros de telemetría y evidencias con cadena de custodia (enfocado a la respuesta a incidentes con validez legal), informes ejecutivos y técnicos (hallazgos, indicadores de compromiso (IOCs), KPIs) y un plan de mitigación priorizado con roadmap con sus correspondientes hitos de implementación.
 4. Actores involucrados: unidades y organismos de ciberdefensa y ciberseguridad (militares y civiles), CERTs, operadores de infraestructuras críticas, departamentos de TI/seguridad de la administración pública y equipos operativos (Red/Blue/Purple).

1.5 Exclusiones

Por la naturaleza del presente trabajo, y dado que no existe ningún acuerdo con instituciones públicas, en aras de garantizar la legalidad y ajustarse a la normativa vigente, quedan fuera del alcance del trabajo:

1. Ejecución de operaciones ofensivas reales o pruebas que afecten a servicios públicos e infraestructuras críticas en producción.
2. Pruebas que impliquen el uso de datos personales no anonimizados, ni se incluirá en la metodología técnicas agresivas empleadas por actores reales tales como harassment.
3. Desarrollo o distribución de herramientas enfocadas a la explotación real de la infraestructura nacional; los playbooks contienen niveles de TTP conceptual y su modelado en la matriz de MITRE ATT&CK, en ningún caso se incluirán las instrucciones ni los comandos paso a paso para explotación.

1.6 Limitaciones metodológicas

Se identifican las siguientes limitaciones inherentes al enfoque y al alcance definido:

1. Realismo del entorno: no se podrá validar la metodología al no contar con acuerdos en entornos reales, por lo que la validación solo podrá realizarse de forma simulada, o mediante la réplica de la red objetivo o en su caso mediante el acuerdo con la organización pertinente.
2. Restricciones legales y éticas: la legislación penal y de protección de datos (GDPR) y políticas internas limitarán ciertas pruebas, especialmente las que impliquen datos personales o interacción con terceros proveedores.
3. Cobertura para la obtención de la telemetría: la disponibilidad y calidad de sondas, EDR y SIEM condiciona la validez de las métricas.
4. Sesgo del repertorio de TTPs: los ejercicios simulan un conjunto limitado de tácticas, técnicas y procedimientos de aquellos adversarios que han sido previamente modelados. Esto presenta una limitación desde el punto de vista del realismo, ya que existen TTP's ajenas al conocimiento de este trabajo.
5. Recursos y coste: no se disponen de un entorno tipo cyber range que permita un despliegue de una infraestructura que emule una organización real.
6. Reproducibilidad en sistemas dinámicos: la casuística dentro de las diferentes redes objetivo es muy amplia, lo que llevara a que este trabajo lleve a una metodología genérica, lo más amplia y laxa posible.

Cada limitación se aborda en el capítulo de metodología mediante mitigaciones (uso de datos anonimizados, plantillas de instrumentación mínima, pruebas de control y repetición, validación por expertos externos y documentación exhaustiva).

1.7 Contexto de aplicación

El presente trabajo se va a focalizar en aquellos contextos donde la defensa nacional y la continuidad de servicios públicos y críticos tienen prioridad. Se materializará en base a los siguientes puntos:

1. Todo ejercicio Red Team se planificará con un documento de alcance en el que se incluirán los objetivos buscados, así como un documento de RoE's, siendo ambos aprobados por la organización evaluada y por el equipo responsable de la intrusión. Identificación clara de activos incluidos/excluidos (scope), definición de métricas/KPIs de evaluación y un plan de comunicación con el responsable del organismo auditado, que incluirá un mecanismo de comunicación urgente en caso de detección de algún evento de gravedad (i.e. identificación de vulnerabilidades críticas, fallos de configuración que puedan llevar a ejecución remota de código, detección de comunicaciones con ip's vinculadas a grupos APT, etc...).
2. Para que los resultados sean válidos, se exige una instrumentación mínima acordada (registros de autenticación, logs de servicios críticos, puntos de captura de tráfico entre otros) y la implicación de los responsables del activo que van a ser objeto de la intrusión.
3. Plan de propuesta de mejora y remediación de las políticas de ciberseguridad.

En síntesis, el alcance del trabajo delimita los ejercicios Red Team como instrumentos ofensivos controlados en los que se emulan los comportamientos y las TTP's de adversarios reales, en los que se busca evaluar la eficacia defensiva, priorizando siempre la autorización, la protección de activos y datos, los sistemas en producción, para transformar los hallazgos en medidas de mejora concretas para la resiliencia nacional.

1.8 Supuestos operativos

El desarrollo y la aplicación del marco asumen, como condición mínima:

1. Existencia de autorización formal y de reglas de compromiso claras para cada ejercicio.
2. Disponibilidad de infraestructura de laboratorio aislada o entornos de prueba replicados.
3. Acceso a personal con competencias mínimas (operadores Red/Blue/Purple) y, cuando proceda, a asesoría jurídica.
4. Instrumentación mínima para recolección de telemetría en los activos incluidos (EDR/agent, logs, capturas de tráfico entre otros).

1.9 Riesgos principales y medidas mitigadoras

1. Riesgo de impacto accidental en sistemas reales. En efecto, y a pesar de que se tomen las debidas precauciones, nunca se puede garantizar que una operación de Red Team sea inocua en el objetivo. Existe el riesgo de producir caídas en algunos de los servicios evaluados. Para ello se proponen las siguientes medidas de mitigación: aislamiento lógico/físico, ventanas de mantenimiento, backups/snapshots almacenados en redes ajenas a la intrusión. Es muy importante que las pruebas sean acordadas y coordinadas con el dueño del sistema y debe existir un canal ágil de comunicación con el ejecutor de las pruebas para que se solicite pararlas en caso de impacto.

-
2. Riesgo legal por uso de datos personales. Una mala praxis de los empleados de las instituciones involucradas puede llevar al descubrimiento de información personal (e.g. registros médicos, datos bancarios, etc...). En este caso, no existe una medida de mitigación concreta, limitándose a la no utilización ni almacenamiento de ninguno de estos datos en caso de ser descubiertos. Por ello, es imprescindible contar con un asesor legal desde la fase de diseño.
 3. Riesgo reputacional. Aquel derivado de la exposición pública de un compromiso. Por ello, los resultados de este tipo de operaciones deben ser confidenciales y así debe quedar reflejado en el NDA.
 4. Riesgo de pérdida de evidencia. Aquel riesgo materializado en la pérdida de la información exfiltrada o los informes generados. Para minimizar este riesgo debe establecerse una cadena de custodia, “hashear” toda evidencia digital para evitar su manipulación, y un almacenamiento seguro y cifrado.

1.10 Criterios de aplicabilidad y transferibilidad

La metodología y los productos generados tras cada acción están pensados para ser transferibles a otras organizaciones y a ejercicios a nivel autonómico o sectorial. Esto es así buscando una mejora de las capacidades defensivas a nivel nacional. Para que esto pueda llevarse a cabo, son necesarios una serie de parámetros como una topología similar, similitud de activos críticos, y requisitos legales. Para facilitar la transferencia se proporcionan plantillas en los anexos.

1.11 Entregables y criterios de cierre

Un ejercicio o proyecto basado en este marco se considerará completado cuando se entreguen y validen los siguientes ítems por todas las partes, tanto por el responsable del equipo de intrusión como por el responsable de la institución evaluada:

1. Documento de alcance, NDA y RoE's firmados.
2. Snapshot de configuración.
3. Repositorio de evidencias y evidencia con cadena de custodia.
4. Informes de cada acción, que se estructurarán en un resumen ejecutivo junto con un informe técnico con hallazgos que incluya los IOCs y los KPIs.
5. Plan de mitigación priorizado por la criticidad de los hallazgos.
6. Material de transferencia para facilitar la mejora en su conjunto de la seguridad nacional.

2 ESTADO DEL ARTE

2.1 Conceptos fundamentales: Auditoría de seguridad, Pentesting, Red Teaming y Purple Teaming

La práctica sistemática de evaluación de la ciberseguridad (o como era conocida en la década de los 2000, seguridad informática) se articula alrededor de instrumentos metodológicos diferenciados por su propósito, alcance y enfoque operativo. Es indispensable comprender en profundidad estas diferencias para diseñar de forma eficiente y eficaz ejercicios ofensivos que aporten evidencia útil a los responsables de la toma de decisiones. En términos generales, puede distinguirse entre actividades orientadas al cumplimiento y la verificación técnica puntual (auditoría de seguridad y pruebas de penetración pentesting) y actividades diseñadas para emular la conducta de un adversario con objetivos estratégicos (como sería el caso de un Red Team), complementadas por procesos de aprendizaje colaborativo entre atacantes y defensores (Purple Teaming) donde el objetivo principal es una mejora de los sistemas defensivos gracias a las sinergias generadas en dicho entorno colaborativo. A continuación, se van a definir cada uno de los diferentes conceptos, matizando sus contribuciones y limitaciones desde una óptica aplicada a la ciberdefensa.

La auditoría de seguridad es un proceso estructurado de verificación que persigue comprobar el cumplimiento de controles, políticas y estándares. Suelen basarse en listas de verificación estandarizadas (checklists) y en procedimientos reproducibles que permiten establecer un estado de conformidad respecto a marcos normativos o a buenas prácticas. Por su naturaleza, la auditoría aporta trazabilidad documental y es especialmente útil para la gobernanza y la toma de decisiones en materia de inversión y certificación, aunque su alcance operativo es limitado frente a amenazas avanzadas, que explotan vulnerabilidades en cadena, zerodays o hacen una intrusión compleja y elaborada, no contemplada en las checklists.

El pentesting profundiza en el análisis técnico: se trata de ejercicios de carácter ofensivo-controlado cuyo objetivo es identificar y explotar vulnerabilidades concretas en sistemas, aplicaciones o redes, con un alcance y una duración acotados. Los pentesters utilizan herramientas similares a las de los atacantes maliciosos. Como indica el nombre, un pentesting requiere intrusión (si es posible). Una prueba que prohíbe la intrusión no es un pentest. En ocasiones los objetivos solicitan una prueba de penetración, pero luego no permiten explotación, exfiltración de credenciales u otras actividades que permitirían a los evaluadores verificar fallos y encontrar vulnerabilidades de segundo orden. Estas pruebas más constreñidas se asemejan más a una evaluación de vulnerabilidades.

El principal activo del pentesting reside en la validación práctica de controles técnicos y en la generación de pruebas reproducibles de explotación de vulnerabilidades. Sin embargo, por lo general

opera sobre vectores y márgenes temporales restringidos (motivado por lo costoso de estas acciones), por lo que no siempre revela deficiencias organizativas, de detección o de gestión del incidente que sólo emergen ante campañas prolongadas con una perspectiva de APT.

El Red Teaming, en contraste, plantea la emulación deliberada y sostenida de un adversario a modo de campaña maliciosa: no se limita a explotar una vulnerabilidad aislada, sino que persigue objetivos estratégicos que pueden requerir inteligencia previa, persistencia, movimientos laterales y encadenamiento de TTPs. En el ámbito de la defensa nacional, un ejercicio Red Team tiene como propósito evaluar la capacidad real de la organización para detectar, coordinar y tomar decisiones bajo condiciones adversas, incluyendo la respuesta interinstitucional y la continuidad operativa. Por ello, su diseño debe incorporar una narrativa de amenaza avanzada, reglas de compromiso claras y métricas que midan no sólo la explotación sino la eficacia defensiva, impacto funcional y capacidad de remediación. Para ello emulan las TTP's de adversarios del mundo real con el objetivo de medir debilidades en la defensa, la detección y la respuesta. Con el fin de lograr este objetivo, se emplean algunas de las herramientas usadas en un pentest, sin embargo, su enfoque es distinto. Un Red Team se concentra en identificar deficiencias en la capacidad del del Blue Team (i.e. equipo responsable de la defensa) para detectar, responder y erradicar una amenaza. Un Red Team utiliza únicamente aquellas vulnerabilidades que le ayudan a cumplir su objetivo, mientras que a los pentesters se les encarga, con mayor frecuencia, encontrar el mayor número de estas vulnerabilidades.

El Purple Teaming emerge como un mecanismo para cerrar el ciclo entre emulación y aprendizaje: consiste en actividades de colaboración estructurada entre los equipos que ejercen la agresión (Red) y los equipos defensivos (Blue) con el fin de traducir hallazgos en reglas de detección. Este tipo de evaluación incluye métricas sobre la capacidad del Blue Team para identificar, detectar y responder a un ataque. Las pruebas deben documentarse para mostrar avances en las capacidades de detección y respuesta. Estas métricas pueden utilizarse para justificar inversiones en herramientas y formación. Lejos de ser un mero taller técnico, el Purple Teaming facilita la transferencia de conocimiento y acelera la maduración de capacidades defensivas.

Finalmente, y para contextualizar el tipo de acciones a las que se enfocan los términos anteriores, se va a hacer referencia a la RFC 4949 [3] en la que se hace referencia a distintos elementos que se explotan tanto en las pruebas de intrusión como en los Red Team:

1. Vulnerabilidad: La RFC 4949 la define como: “Un fallo o debilidad en el diseño, la implementación, la operación o la gestión de un sistema que podría ser explotado para violar la política de seguridad del sistema”. En términos sencillos, la podemos definir como un fallo o debilidad que puede ser aprovechada por un actor malicioso. La mayoría de la gente asocia las vulnerabilidades con fallos de corrupción de memoria (e.g. desbordamientos de búfer o buffer overflow por sus siglas en inglés BOF), pero eso es solo una parte de todas las vulnerabilidades de seguridad (en concreto, los BOF se han visto muy mitigados tras la introducción de medidas como la aleatorización de memoria ASLR, prevención de ejecución en memoria DEP etc). Las vulnerabilidades incluyen, pero no se limitan a:
 1. Mala configuración (misconfiguration)
 2. Fallos de diseño y arquitectura
 3. Falta de funciones de auditoría
 4. Contraseñas débiles/fáciles de adivinar
 5. Credenciales por defecto

Existen muchas otras categorías de vulnerabilidades que no figuran en la lista anterior. MITRE (MITRE Corporation. n.d.) creó el *Common Weaknesses Enumeration* (CWE), que es un

listado de tipos de debilidades en hardware y software. Esto no debe confundirse con CVE (*Common Vulnerabilities and Exposures*), que lista fallos específicos en un determinado software.

2. Exploit: Un exploit es un código o técnica que aprovecha una vulnerabilidad [4, 5]. Algunos ejemplos son:
 1. Explotar una vulnerabilidad en la gestión de memoria para permitir que el atacante obtenga acceso remoto al sistema.
 2. Usar funcionalidad existente para subir código y ejecutarlo, por ejemplo, subiendo una web shell a un servidor web mediante una característica que permite a los usuarios subir un currículum o un avatar.
 3. Un atacante cambia su nombre por código JavaScript, aprovechando un filtrado débil de entradas en el sitio web.
 4. Ejecutar un binario que realiza operaciones privilegiadas en el sistema, escalando así los privilegios.
3. Amenaza: Una amenaza es un agente o actor que puede causar un daño [6, 7] En sentido más amplio, es cualquier cosa que pueda causar perjuicio. Una amenaza puede ser un atacante de la delincuencia organizada que intenta robar números de tarjeta de crédito, pero también puede ser el usuario que pulsa el botón equivocado por accidente. Una amenaza ni siquiera tiene que ser humana: un gusano (worm) es una amenaza, dado que es el código el que se propaga y causa daños. De manera similar, nuestras propias herramientas de seguridad pueden considerarse una amenaza en ciertos contextos. Un escáner de vulnerabilidades para aplicaciones web rastreará el sitio y enviará datos. Si el escáner apunta a un sitio de producción con datos reales, ¿qué ocurriría si el escáner “pulsara” el botón de eliminar? Los escáneres intentan evitar acciones sensibles, pero este tipo de incidentes sucede. Esta es una de las razones por las que es importante disponer de copias de seguridad y de entornos de prueba.
4. Riesgo: El riesgo se define como el potencial de pérdida o daño [8, 9, 10]. El riesgo suele calcularse como $(\text{Probabilidad de Ocurrencia}) \times (\text{Impacto Potencial})$. El impacto potencial es un valor que podemos estimar con un grado razonable de certeza; sin embargo, la probabilidad es un número mucho más debatible. A menudo, la probabilidad (y, en última instancia, el propio valor del riesgo) se reduce a la opinión experta del equipo de seguridad. El equipo utiliza su conocimiento y experiencia colectivos para determinar el riesgo y, finalmente, ofrecer orientación al negocio. Equilibrar el riesgo con las necesidades organizativas es uno de los objetivos fundamentales del equipo de seguridad. Desafortunadamente, muchos equipos de seguridad son percibidos como obstáculos, que frenan a la organización en lugar de permitir que se desarrolle y progrese de forma segura.

A continuación, en la tabla 2-1 se presenta un resumen de los términos definidos anteriormente.

Término	Definición y ejemplos
Vulnerabilidad	«Un fallo o debilidad en el diseño, la implementación, la operación o la gestión de un sistema que puede ser explotado para violar su política de seguridad.» (RFC 4949). Ejemplos: BOF, mala configuración, fallos de diseño/arquitectura, ausencia de auditoría, contraseñas débiles, credenciales por defecto.
Exploit	Código o técnica que aprovecha una vulnerabilidad para provocar un comportamiento no previsto (p. ej. ejecución remota, escalado de privilegios). Ejemplos: explotación de gestión de memoria, subida de web-shells vía carga de ficheros, XSS/JS para manipular entradas, ejecución de binarios con privilegios. [4, 5]
Amenaza	Agente, condición o evento con capacidad de causar daño a los activos de información; puede ser humano, accidental o no humano (malware). Ejemplos: atacante organizado, usuario que pulsa un botón por error, gusano autorreplicante, herramientas mal configuradas. Consecuencia práctica: las propias herramientas pueden comportarse como amenaza; por eso son necesarias copias de seguridad y entornos de prueba controlados. [3, 6, 11]
Riesgo	Potencial de pérdida o daño; frecuentemente modelado como $\text{Riesgo} = \text{Probabilidad} \times \text{Impacto}$. Notas: el impacto puede estimarse con cierto grado de certeza; la probabilidad suele requerir juicio experto y métodos cualitativos/estadísticos. El equipo de seguridad debe equilibrar riesgo y necesidades del negocio [8, 9, 10]

Tabla 2-1 Conceptos fundamentales de seguridad

2.2 Ciberejercicios: definición, tipos y objetivos

En el mundo hiperconectado en el que vivimos, la dependencia de los sistemas conectados a Internet es cada día mayor. Las administraciones públicas han tenido que digitalizarse con gran celeridad para adaptarse a los nuevos requisitos de la ciudadanía. Es en este contexto que la protección de los sistemas IT se vuelve clave para el correcto funcionamiento de los servicios prestados al ciudadano. La digitalización de las diferentes entidades públicas no siempre ha ido pareja a una correcta securización y bastionado de los sistemas. Igualmente, la normativa de ciberseguridad ha ido un paso por detrás de los actores maliciosos, por lo que se hace necesaria la creación de una figura que permita realizar test de estrés en redes tan sensibles para el correcto funcionamiento de los servicios públicos.

Los ciberejercicios surgen como una forma de prepararse para un posible ataque. Emulando técnicas adversarias, permiten evaluar una amplia gama de controles técnicos, más allá de los puntuales, generando la capacidad organizativa de detectar, coordinar y responder adecuadamente ante una campaña adversaria real. Según [12], los ejercicios permiten identificar interdependencias y validar planes de continuidad ante ciberincidentes.

De una forma más concreta, y citando de nuevo [12], los ciberejercicios son actividades planificadas y estructuradas cuyo propósito es reproducir, de forma controlada, ciberataques, aspectos relevantes de incidentes o actuaciones de grupos patrocinados por estados, para evaluar y mejorar la preparación de organizaciones y colectivos frente a crisis digitales. Estas actividades pueden perseguir fines diversos (e.g. formación, verificación de procesos, comprobación de controles técnicos, validación de coordinación interinstitucional o cumplimiento regulatorio) pero, para lograr el objetivo deseado, necesitan de unas ROE's (reglas de compromiso) y criterios de evaluación que permitan

convertir la experiencia en mejoras operativas y de gobernanza, plasmándose en una parte del informe final como puntos de mejora y lecciones aprendidas.

En este trabajo, se tratará un ciberejercicio como una actividad planeada, con metas claras y reglas fijas, que busca probar si las organizaciones pueden detectar, responder y recuperarse de ciberataques, no solo hackear sistemas. Estos ejercicios también sirven para de elemento verificador de cómo se comunican en caso de crisis, ver cómo se afectan los distintos sectores entre sí y encontrar puntos débiles para mejorar la seguridad. Según la ENISA, en Europa estos ejercicios son de vital importancia para que el sector público y privado trabajen juntos y para que cada sector sea más fuerte ante incidentes.

2.2.1 Tipos de ciberejercicios

Existe una amplia variedad de ejercicios aportando cada uno de ellos una perspectiva diferente de la ciberseguridad. No existe una taxonomía internacional que estandarice estos ejercicios, a pesar de parte de la terminología empleada es común.

La tipología de ejercicios es extensa, no limitándose únicamente a aquellos de carácter técnico (como pueden ser los Red Team Blue Team engagements). Existen ejercicios de menor nivel enfocados al estudio y validación de procedimientos, desde un punto de vista teórico, sin llegar a poner a poner en práctica lo propuesto. También existen otro tipo de ejercicios enfocados a validar los procedimientos con el fin último de comprobar que todos los participantes cumplen los requisitos organizativos. Evidentemente los más complejos son aquellos que involucran un ciclo entero de planeamiento, y que se llevan a cabo, materializándose en una emulación de actores hostiles, como son los ejercicios de Red Team. ENISA propone una serie de ejercicios que evolucionan según su nivel de complejidad, como se muestra en la figura 2-1.

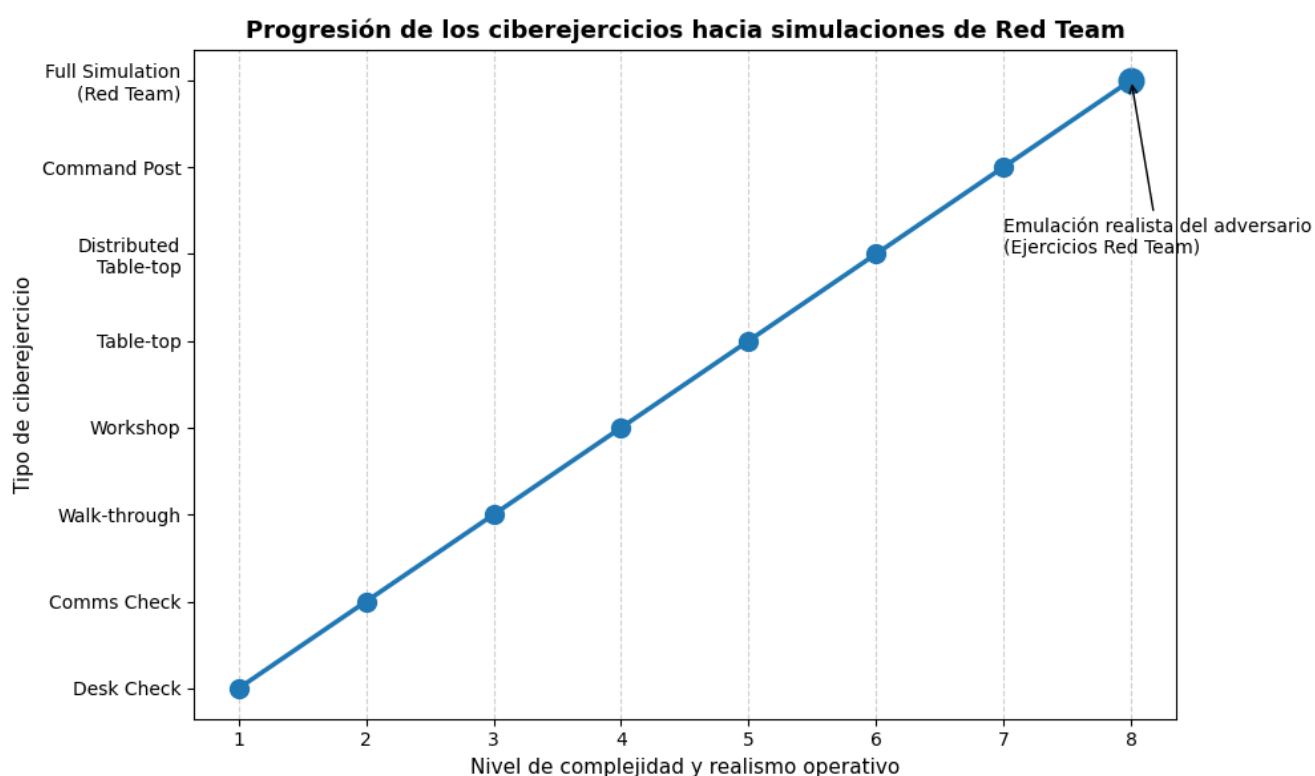


Figura 2-1 Ciberejercicios propuestos por ENISA.

Es en este último apartado donde se incluyen los ejercicios de Red Team. Como ya se ha mencionado previamente, son ejercicios completos y complejos, con un ciclo de preparación que necesita de unos tiempos (en la figura 2-1 se dan los puntos clave para realizar un ejercicio table top).

También, son necesarias reuniones de coordinación para la determinación del alcance ROE's, límite temporal, etc...

Los ejercicios Red Team aportan una perspectiva adversaria al cuestionar las suposiciones realizadas por una organización y sus defensores. Suposiciones como “estamos seguros porque aplicamos parches”, “solo X personas pueden acceder a ese sistema” y “la tecnología Y lo impediría” son peligrosas y, a menudo, no resisten un análisis minucioso. Llevando estas suposiciones al punto de ruptura, un Red Team puede identificar áreas de mejora significativas en la defensa operativa de una organización.

Los Red Team tienen un objetivo claro definido por la organización, ya sea obtener acceso a un sistema concreto, una cuenta de correo electrónico, una base de datos o un recurso compartido de archivos. Al fin y al cabo, comprometer la confidencialidad, integridad y/o disponibilidad de una organización representa un riesgo tangible, ya sea financiero o de reputación. Un Red Team también emulará una amenaza real para la organización. Por ejemplo, una empresa financiera (o un banco estatal, como podría ser el Banco de España) puede estar en riesgo por parte de grupos cibercriminales conocidos, especializados en el sector bancario. Un Red Team estudiará y reutilizará (cuando sea apropiado) las TTP de la amenaza que está emulando. Esto permite a la organización crear detecciones y procesos diseñados para combatir las amenazas a las que espera enfrentarse. Los Red Team también analizarán de forma holística la postura de seguridad general de una organización y no se centrarán exclusivamente en un área específica, lo que incluye tanto a las personas y los procesos como a la tecnología.

2.2.2 Objetivos

Los ciberejercicios son un mecanismo para probar y mejorar la preparación y respuesta ante incidentes, así como para reunir datos empíricos para desarrollar políticas y capacidades [13]. Por encima de todo, el principal objetivo de cualquier tipo de ejercicio es la práctica de procedimientos y la mejora continua. Si bien estos son objetivos a muy alto nivel, para el éxito de un ciberejercicio, la definición clara de las metas y objetivos específicos es fundamental. Las metas deben reflejar el propósito general del ejercicio, alineándose con la estrategia de ciberseguridad y el perfil de riesgo de la organización. Los objetivos específicos deben focalizarse en las áreas más relevantes y críticas. Por ejemplo, poner a prueba las capacidades de respuesta a incidentes de una determinada organización, con el fin de evaluar su capacidad de resiliencia real. Otro ejemplo sería evaluar los procesos de comunicación e identificar brechas en las políticas de seguridad.

Para que los ciberejercicios produzcan unos resultados evaluables, es necesario asegurarse de que los objetivos sean específicos, medibles, alcanzables, relevantes y con plazos definidos para facilitar una evaluación eficaz. Para ello, antes de comenzar el proceso es necesario establecer una serie de restricciones [14], entre las que están la limitación presupuestaria, el tiempo y los recursos, y que deben ser comunicadas a todos los implicados.

No hay que olvidar que para que los ejercicios sean provechosos y se puedan extraer lecciones y mejoras, los escenarios han de ser realistas, y deben poner a prueba los conocimientos técnicos, habilidades y experiencia del personal involucrado. Por ello, ha de llegarse a un compromiso entre dificultad y realismo, de forma que el ejercicio sea provechoso. Mantener un nivel de estrés positivo es fundamental para que los participantes se mantengan comprometidos y concentrados. Independientemente del tamaño del ejercicio, los escenarios deben diseñarse para poner a prueba y perfeccionar las habilidades de los participantes desde una perspectiva realista, reflejando retos reales, similares a los que puedan enfrentarse en su entorno profesional.

Un punto fundamental de un ciberejercicio es el debriefing. Se trata de una reunión con la organización y los participantes en la que se recaba su opinión acerca de la experiencia. El análisis de los resultados del ejercicio basándose en criterios de evaluación previamente definidos, la identificación de los puntos fuertes y débiles, así como las áreas de mejora redundan en la capacidad

de la organización para detectar y responder a los ataques reales. Todo esto se materializa en una actualización de las políticas y los procesos de seguridad, inversión en tecnologías de defensa proactiva, concienciación del personal y planes de mejora.

Por último, los objetivos buscados con la realización de ciberejercicios podrían dividirse en diferentes clases. En primer lugar, estarían los técnico-operativos enfocados a la evaluación de la detección, contención y remediación frente a TTP's simuladas. Después estarían los objetivos organizativos, cuya finalidad es comprobar roles, escalado y comunicación entre los diferentes implicados en el proceso de gestión de incidentes. En último lugar de definen los objetivos estratégicos, que verifican el cumplimiento, continuidad de negocio y preparación entre instituciones. Para que estos objetivos se traspongan en procedimientos operativos, es necesario definir antes del ejercicio una serie de KPI's y métricas [15].



Figura 2-3 Factores para crear un ejercicio tabletop. Fuente: Red Goat Cyber Security

2.3 Marcos internacionales de referencia (TIBER-EU, CBEST, MITRE ATT&CK)

Existen una serie de principios comunes, como son la recolección de inteligencia, reglas de enfrentamiento (ROE), métricas reproducibles, entre otras, que comparten los principales marcos internacionales de referencia en el entorno del Red Team y la simulación de adversarios. Entre los frameworks (que es como se conoce a los marcos de referencia en inglés) se encuentran TIBER-EU (estándar europeo para red teaming), CBEST (Bank of England), así como otros pensado con un

enfoque operativo, empleados para el modelado de amenazas en el ciberespacio, como es el framework MITRE ATT&CK. A continuación, se detallarán cada uno de ellos.

2.3.1 TIBER-EU: estándar europeo para red-teaming intelligence-led

Con la finalidad de desarrollar un entorno lo más realista que posible, y que redundará en una mayor calidad del ejercicio, TIBER-EU [16] propone pruebas realistas basadas en Threat Intelligence (proceso de recopilar y analizar información sobre ciberamenazas para prevenir y responder a ataques). Proporciona un estándar uniforme para implementar pruebas lo más realistas posible, empleando los IOC's obtenidos tras el ciclo de threat intelligence.

Bajo este framework, los ejercicios de Red Team se realizan en un entorno controlado y basado en la inteligencia previamente adquirida, lo que permite personalizar en las funciones críticas o importantes (CIF) de las entidades, así como de los sistemas subyacentes que las respaldan (i.e. personas, procesos y tecnologías) e imita las tácticas, técnicas y procedimientos (TTP) de los actores maliciosos reales, como pueden ser APT's (Advanced Persistent Threat), grupos patrocinados por estados o ciberdelinquentes. TIBER-EU permite a las autoridades europeas y nacionales colaborar con entidades financieras y de otro tipo para poner en marcha un programa destinado a probar y mejorar su resiliencia frente a ciberataques sofisticados. Permitiendo en última instancia ayudar a las entidades participantes a evaluar sus capacidades de protección, detección y respuesta.

TIBER-EU es un framework muy completo, que abarca todos los aspectos del ejercicio. Partiendo de un diseño centrado en un núcleo operativo que tiene como objeto garantizar que la aproximación sea lo más realista posible. Este núcleo se articula en 4 piezas: la recopilación y el procesado de la información de inteligencia de amenazas previa, que se formalizará en un producto. En segundo lugar, la parte que atañe a la gobernanza y la separación tácita de roles articulada en la figura de un Control Team. En tercer lugar, el acotamiento de la extensión del ejercicio donde se acotan, entre otros, el alcance y las ROE's. Por último, aquellas medidas técnicas y legales destinadas a la mitigación del riesgo, que minimizan la probabilidad de impacto en los sistemas en producción [16]. Este diseño favorece que los ejercicios de Red Team se traduzcan en intervenciones bien planificadas y trazables, que permitan su evaluación por parte de terceras partes, como puede ser la figura de supervisores o de direcciones ejecutivas.

Como ya se ha mencionado, la fase de threat intelligence (TTI, Target Threat Intelligence) es el core sobre el que se articula el framework TIBER-EU. Esta fase se materializa en un documento técnico-operativo que convierte la inteligencia recabada de actores maliciosos y campañas identificadas para generar un catálogo priorizado de TTP que serán el objeto para emular. Bajo este enfoque, el TTI es un requisito que guía la prueba. Por ello debe incorporar fuentes, criterios de justificación y un desarrollo preliminar de las tácticas y técnicas a emular (empleándose para ello la matriz de MITRE ATT&CK, por ejemplo), así como las limitaciones y exclusiones [16].

En la práctica, TIBER-EU facilita que el plan de ejecución de las pruebas de Red Team (en la figura 2-2 se muestran las fases de preparación de un ejercicio de Red Team) no solo la entrada de la matriz de MITRE, sino que además incluya la evidencia esperada tras la implementación de la técnica, así como un criterio para la detección o no de la misma y en qué capa de la defensa (i.e. en el host o el SIEM). Además de la reproducibilidad del ejercicio, este enfoque permite trasponer aquellos hallazgos críticos en una colección de medidas de mitigación concretas como pueden ser reglas de SIEM.

En último lugar, este framework estandariza los entregables. Estos entregables deben ser definidos junto con el alcance y formalizados mediante sendos NDA (Non Disclosure Agreement). Entre los entregables, uno de los que toma especial relevancia es el paquete de evidencias tras el compromiso, así como su cadena de custodia (al ser evidencias digitales deben "hashear"). La importancia de este paquete reside en que es una práctica de lo que sucedería en un escenario real para dar validez legal a las evidencias recogidas por los equipos de respuesta a incidentes. Otros entregables definidos en este marco de referencia incluyen el informe técnico, las ROE's que delimitan los límites y procedimientos

de emergencia (e.g. comunicación urgente en caso de detectar un compromiso real por parte de un actor malicioso) o el informe ejecutivo dirigido a la capa directiva.

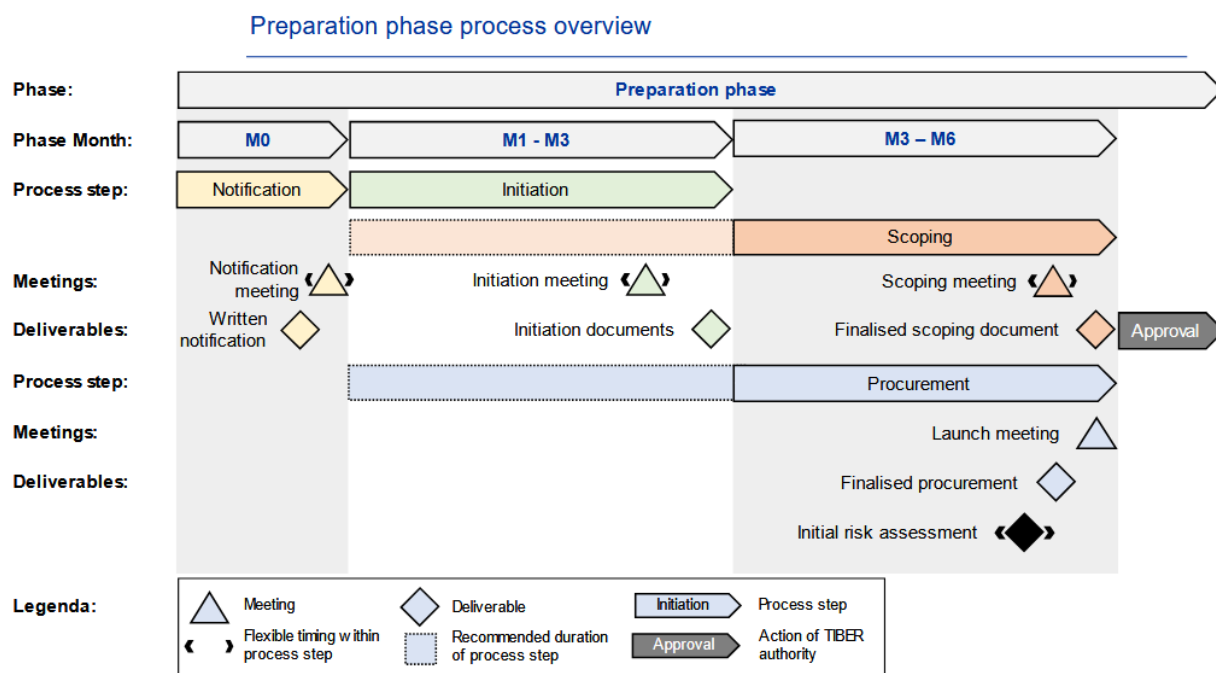


Figura 2-4 Fase de preparación de un ejercicio de Red Team. TIBER-EU

2.3.2 CBEST: enfoque británico

Al igual que TIBER-EU, CBEST es un modelo intelligence-led (figura 2-3), desarrollado por el Bank of England junto con la PARA (Prudential REGulation Authority), donde la emulación se basa en la TTP's reales empleadas por grupos maliciosos. El objetivo perseguido es mantener las capacidades críticas en entornos degradados bajo ciberataques. Para ello, no se busca evaluar vulnerabilidades puntuales, sino la capacidad de la entidad auditada para proseguir con sus funcionalidades bajo campañas ofensivas avanzadas. A pesar de que el marco de referencia trabaja con el concepto de "pentesting", un análisis en profundidad del mismo, evidencia de que se tratan de acciones de Red Team, mucho más elaboradas y complejas que las derivadas de un test de intrusión. Es por ello que se incluye en el presente trabajo este marco de referencia.

El propio marco de referencia dictamina cuándo debe realizarse el ejercicio, que es al cumplir uno de los siguientes criterios.

1. El regulador estatal indica que la organización debe realizar como parte del ciclo de supervisión.
2. La organización solicita realizar el ejercicio como parte de su programa de resiliencia bajo ciberataque previa consulta al regulador estatal.
3. Se han producido un incidente u otros eventos de seguridad que llevan al regulador a solicitar un ejercicio para validar los procedimientos de remediación.

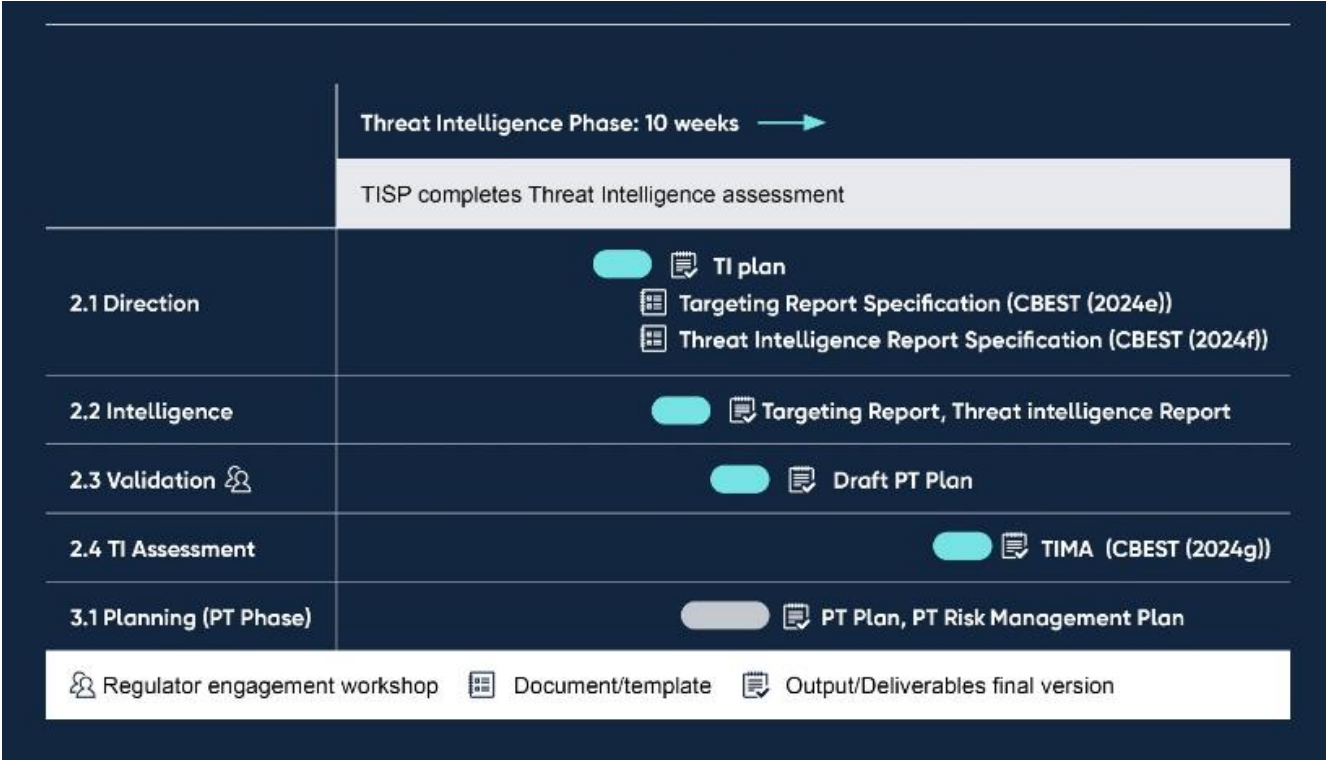


Figura 2-5 Fases de recolección de inteligencia. Fuente: Bank of England

Este framework propone una serie de fases que permiten definir perfectamente todos los apartados del ejercicio. Estas fases son 4:

4. Fase 1: Fase inicial donde se van a definir los puntos clave del ejercicio. Estos puntos son la duración del “engagement”, alcance, redes y sistemas objetivo, medidas de coordinación, etc...
5. Fase 2: Es la fase dedicada a la recolección de ciberinteligencia e inteligencia de amenazas. A partir de aquí se generan los entregables que incluyen los escenarios que se llevarán a cabo durante el ejercicio, y que se facilitarán al equipo de Red Team.
6. Fase 3: La parte ofensiva/intrusión. Durante esta fase se planifica, ejecuta y revisa la simulación de adversario basada en la inteligencia recogida en la fase anterior. Además de la intrusión misma, otra finalidad es evaluar la madurez de la inteligencia recabada, así como las capacidades de detección y respuesta a incidentes.
7. Fase 4: Fase de cierre durante la cual se finaliza el Plan de Remediación (uno de los entregables contemplados en el marco de referencia) y se supervisa la ejecución del plan de remediación por parte de la entidad.

Por último, existen diferentes entes involucrados en el ejercicio propuesto por el marco CBEST. Todas estas figuras tienen un rol definido en el marco de referencia y sus cometidos son claves para el correcto desarrollo de este.

1. Regulador: Los reguladores proporcionan orientación y supervisión durante toda la evaluación, verificando que el ejercicio se ejecute de acuerdo con el marco de evaluación CBEST. Los equipos reguladores incluirán personal supervisor y especialistas en ciberseguridad. El regulador es responsable de utilizar los resultados de la evaluación CBEST para formarse una idea de la postura de ciberseguridad del participante.
2. Grupo de control de la organización auditada: El organismo auditado designa un equipo responsable de la gestión y supervisión de la evaluación [17].
3. Threat Intelligence service provider: El TISP debe estar acreditado por CBEST. El TISP implementará el análisis de Threat Intelligence siguiendo las mejores prácticas descritas en la

Guía de Evaluación de Servicios de CBEST [18]. Este equipo tiene asignadas unas tareas que deben cumplir unos criterios mínimos, a saber, proporcionar una evaluación de inteligencia de amenazas externas de la organización, que incluya perfiles respaldados por evidencias de actores de ciberamenazas susceptibles de atacar a la organización, proporcionar información que los posibles actores maliciosos podrían descubrir sobre los sistemas clave identificados dentro del alcance de CBEST, crear escenarios de amenaza basados en los resultados de la evaluación de objetivos y la inteligencia de amenazas, entre otros [19].

4. Penetration tester/Red Team: es una empresa independiente, contratada por la organización para planificar y ejecutar la actividad de test de intrusión/Red Team con base en los escenarios de amenaza identificados durante la fase de Threat Intelligence. Este equipo trabajará de forma colaborativa tanto con la organización objetivo como con el equipo de Threat Intelligence durante el ejercicio [20].

2.3.3 MITRE ATT&CK

MITRE es una organización sin ánimo de lucro ubicada en los Estados Unidos que gestiona Centros de Investigación y Desarrollo Financiados Federalmente (FFRDCs). En esencia, MITRE está financiada por el gobierno estadounidense y trabaja en varios proyectos importantes, varios de los cuales están orientados a necesidades de seguridad de la información.

El MITRE ATT&CK Framework (Adversarial Tactics, Techniques and Common Knowledge) es una base de conocimiento gratuita que mapea tácticas, técnicas y procedimientos (TTPs) conocidos de adversarios a partir de observaciones del mundo real. ATT&CK tiene como objetivo caracterizar las técnicas de los adversarios, desde el acceso inicial hasta la exfiltración de datos y el impacto organizacional, identificando los grupos de adversarios que han implementado exitosamente dichas técnicas.

ATT&CK también es útil como un ejemplo real de lenguaje y nomenclatura consistente para los ataques. ATT&CK integra una nomenclatura estándar para los términos y definiciones que describen ataques de una manera clara y coherente.

Fundamentalmente, caracterizar los TTPs de los diversos grupos atacantes resulta complicado. El MITRE ATT&CK Framework proporciona una descomposición accesible de técnicas de ataque y conceptos que pueden utilizarse para obtener una mejor comprensión de cómo operan los atacantes y, a partir de ese conocimiento, cómo defender mejor nuestras redes (en la tabla 2-2 se presenta una comparativa de los tres marcos).

El proyecto del MITRE ATT&CK Framework está disponible en <https://attack.mitre.org>.

	TIBER-EU	CBEST	MITRE ATT&CK
Tipo de marco	Marco europeo de Red Team basado en inteligencia	Marco británico de Red Team basado en inteligencia	Base de conocimiento de TTPs de adversarios
Propósito	Evaluar la resiliencia de funciones críticas del sector financiero a nivel nacional	Evaluar la resiliencia de entidades financieras reguladas en UK	Estandarizar tácticas y técnicas para planificación, ejecución y detección
Alcance geográfico	Unión Europea (aplicable nacionalmente)	Reino Unido	Global, sector-agnóstico
Sector objetivo	Fundamentalmente financiero e infraestructuras críticas asociadas	Exclusivamente financiero	Todos los sectores
Gobernanza y supervisión	Autoridad financiera nacional y TIBER Cyber Team	Bank of England/PRA/FCA	No tiene supervisión regulatoria
Uso de TTPs	Mapeo habitual a ATT&CK para definir cobertura técnica	Mapeo habitual a ATT&CK	Fuente primaria de TTPs y taxonomía
Obligatoriedad	Voluntario, aunque puede exigirse según criticidad	Requerido para entidades sistémicas	No obligatorio
Resultados esperados	Informe técnico y ejecutivo, plan de remediación, validación	Informes supervisables y remediación	Mapas de cobertura, detecciones y mitigaciones

Tabla 2-2 Tabla comparativa entre los diferentes marcos de referencia.

2.4 Herramientas y frameworks para ejercicios red team en contextos nacionales

Existe una amplia variedad de herramientas diseñadas para la ciberseguridad ofensiva. La comunidad dedicada a esta parte de la ciberseguridad mantiene en constante desarrollo estas herramientas. No se trata solo de exploits, existen desarrollos enormemente complejos e incluso desarrollos ad hoc para explotar un determinado sistema o servicio. Este tipo de herramientas avanzadas suelen ser de pago bajo licenciamiento y limitadas al uso legítimo en entornos corporativos, siempre enfocados a la mejora de la ciberseguridad de las entidades auditadas. Herramientas como Cobalt Strike, Nighthawk o Brute Ratel [21] en malas manos, se convierten en auténticas armas capaces de comprometer sistemas de gran complejidad.

En general, durante una acción de Red Team se sigue una serie de pasos preestablecidos, que, si bien no son de obligatorio cumplimiento, suelen emplearse como base durante los enfrentamientos o “engagements” en inglés. A este conjunto de pasos se le conoce como Cyberkill Chain [22, 23]. La Cyberkill chain es un modelo estratégico que describe las etapas de un ciberataque, desde el reconocimiento inicial hasta la consecución del objetivo final. Este marco se refleja a menudo en los test de intrusión (pentest) y en los Red team, donde se representan las distintas fases de un ataque ciberataque autorizado (consensuado mediante un NDA entre la empresa de Red Team y la entidad atacada) a un sistema informático con objeto de evaluar su seguridad. Al comprender los pasos de la Cyberkill Chain, los operadores ofensivos pueden imitar las estrategias de los atacantes maliciosos, explorando las vulnerabilidades en cada etapa de la cadena. En los siguientes apartados se definirán en profundidad herramientas y frameworks más usados en los enfrentamientos de Red Team.

2.4.1 Cyberkill Chain

Originalmente propuesta por Lockheed Martin (a pesar de que otras empresas han propuesto sus versiones, la más comúnmente empleada es esta), la Cyber Kill Chain (mostrada en figura 2-4) es un framework que forma parte del modelo de defensa activa basado en inteligencia para la identificación y prevención de intrusiones maliciosas. El modelo identifica los puntos necesarios que un adversario debe completar para la consecución de sus objetivos.

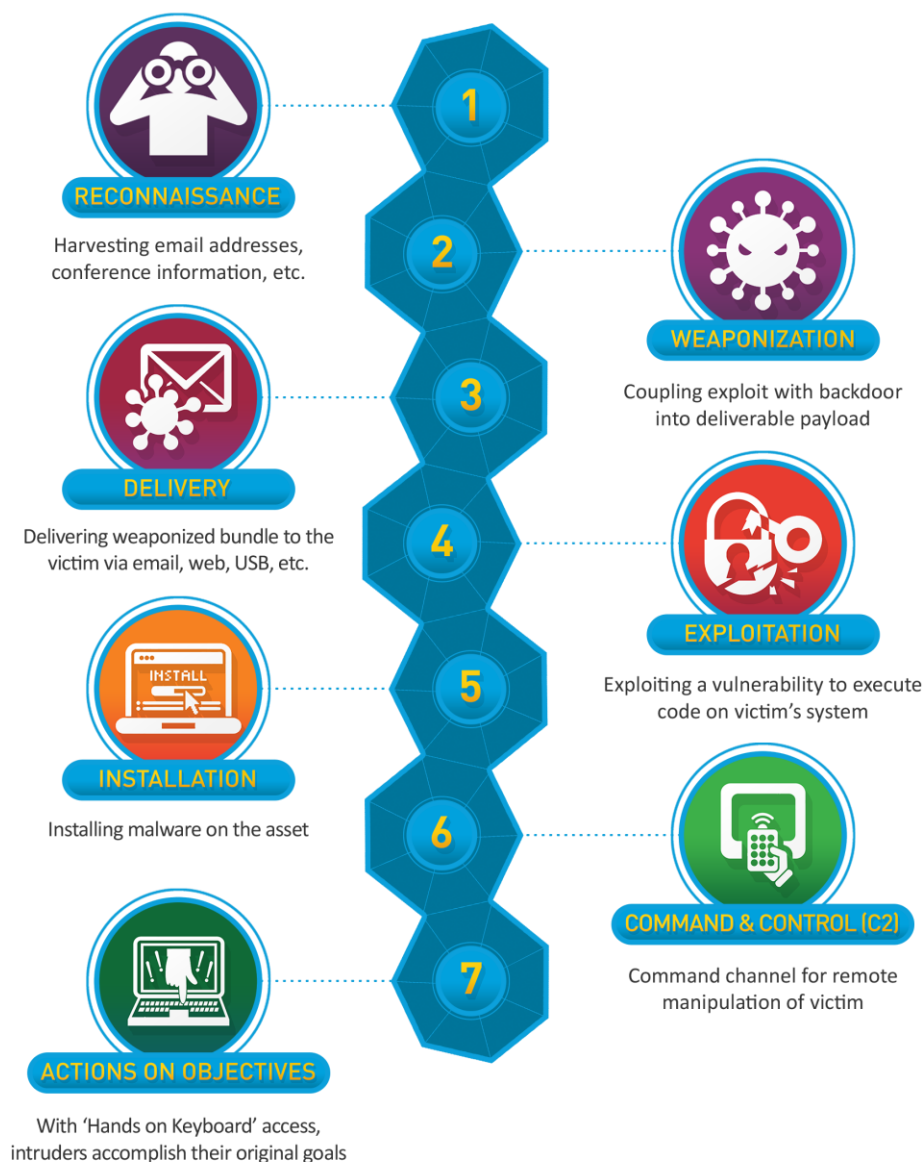


Figura 2-6 Modelo de Cyber Kill Chain [24].

En última instancia la Cyber Kill Chain tiene como objetivo ayudar a los expertos en la materia a planificar, identificar y ejecutar medidas contra las actividades de intrusión en el ciberespacio, no solo en los sistemas IT [25], describiendo las siete etapas necesarias para que los adversarios materialicen un ataque, ejecutando ataques sofisticados para explotar las vulnerabilidades del sistema [26]. Para conseguirlo, se siguen 7 pasos que permiten mejorar la visibilidad de cómo se conforma un ciberataque, así como enriquecer al analista a la hora de comprender las tácticas, técnicas y procedimientos adversarios. A continuación, se definen las 7 fases propuestas por Lockheed Martin.

-
5. Reconocimiento: Los ciberataques más exitosos comienzan con una gran recopilación de información. El reconocimiento es el primer paso en la Cyber Kill Chain y utiliza muchas técnicas y herramientas diferentes, así como funciones de navegación web de uso común. Este reconocimiento puede ser pasivo o activo. La principal diferencia radica en la interacción o no con el objetivo. En el reconocimiento pasivo emplea técnicas que buscan la exposición del objetivo en la red, como pueden ser OSINT (Open Source Intelligence) y SOCMINT (Social Media Intelligence). La información recolectada ayuda a trazar los posibles vectores de entrada, así como el diseño de la campaña en general. En el caso de la Administración Pública la inteligencia disponible en fuentes abiertas es muy extensa, debido a la idiosincrasia de la contratación y licitación, que en la mayoría de los casos está expuesta. Por otra parte, está el reconocimiento activo, que implica interactuar con el objetivo. Un ejemplo típico es un escaneo activo contra el direccionamiento expuesto, en aras de buscar vulnerabilidades y servicios expuestos. Esta última parte permite modelar la superficie de exposición del objetivo, alimentando nuevamente las siguientes fases.

Los atacantes utilizan una amplia gama de herramientas y técnicas para recopilar información sobre sus objetivos, cada una de las cuales expone diferentes datos que pueden utilizarse para conseguir el acceso a sus aplicaciones, redes y bases de datos, que cada vez más se basan en la nube.

6. Armado (weaponization): Una vez que un atacante ha recopilado suficiente información sobre su objetivo, elegirá uno o varios vectores de ataque para comenzar la intrusión en el sistema objetivo. Un vector de ataque es un medio que utiliza un actor malicioso para obtener acceso no autorizado a sus sistemas de información. Los vectores de ataque van desde los más básicos hasta los más técnicos, pero lo que hay que tener en cuenta es que, para un atacante, los objetivos suelen elegirse evaluando el coste frente al retorno de la inversión, ya que, en última instancia, un grupo de cibercriminales es una “empresa” que se busca el rédito económico en sus actividades ilícitas. Todo, desde la potencia de procesamiento hasta el tiempo de amortización, es un factor que los atacantes tienen en cuenta. En esta parte de la Cyber Kill Chain está una de las partes de mayor importancia para conseguir el éxito en la intrusión, la evasión de las defensas. El artefacto que se diseñe ya sea una beacon de un C&C o un malware desarrollado ad hoc para el objetivo, tiene que ser capaz de desplegarse sin ser detectado por antivirus ni EDR’s. Sin una capacidad sólida de evasión, la intrusión no tendrá éxito.
7. Delivery: Una vez identificado y definido el vector de entrada, el actor malicioso lo empleará para hacer llegar el “payload” al objetivo y así comprometerlo. Entre los vectores más comunes de acceso se encuentran el phishing, acceso a través de una aplicación vulnerable, o incluso en acciones más sofisticadas la intrusión física (e.g. despliegue de USB’s con software malicioso más conocido como malware).
8. Explotación: Entregado el “payload” en el objetivo, comienza la explotación del sistema, en función del tipo de ataque. Una vez explotado el sistema, se reconoce el sistema, se buscan elementos valiosos como controladores de dominio, se busca persistencia (i.e. la permanencia en el objetivo a pesar de que se reinicie), movimientos laterales etc...
9. Instalación: De cara a ampliar el ataque, su siguiente paso que realizará un actor malicioso es instalar una puerta trasera para tener acceso constante a los sistemas del objetivo. De esta manera, pueden entrar y salir de la red del objetivo sin correr el riesgo de ser detectados al volver a entrar a través de otros vectores de ataque. Este tipo de puertas traseras se pueden establecer mediante rootkits u otros elementos más burdos, como el uso de credenciales débiles, y siempre que su comportamiento no levante sospechas entre el equipo de seguridad (como horarios de inicio de sesión inusuales o grandes movimientos de datos).
10. Mando y control (C&C): En este punto, y con las persistencias establecidas, el atacante establece comunicación con el/los hosts comprometidos, permitiendo su control en remoto y la interacción con el objetivo. Se hace esto para que los sistemas comprometidos reciban

instrucciones. Se suelen emplear protocolos de comunicación estándar, para enmascararse con el tráfico legítimo (e.g. HTTP, HTTPS, y en algunos casos hasta DNS).

11. Acciones sobre el objetivo: Esta es la fase de ejecución continua en la que un atacante actúa sobre el objetivo y realiza diferentes acciones maliciosas. Puede cifrar sus datos para pedir un rescate, filtrar sus datos para obtener beneficios económicos, derribar su red mediante un ataque de denegación de servicio o supervisar el comportamiento de su sistema en busca de cualquier otra vulnerabilidad mediante spyware, por nombrar solo algunos posibles resultados de la intrusión. El control de los sistemas objetivo es la meta en este último paso de la cadena de ataque, en el que los atacantes mantienen un perfil bajo y persisten. Aquí es donde resulta crucial la supervisión en tiempo real del movimiento de datos y la detección de comportamientos anómalos, ya que los atacantes actuarán lo más rápido posible para lograr sus objetivos.

La Cyber Kill Chain constituye una serie de fases que no representan un dogma a seguir (en la figura 2-5 se muestra una variación de la Cyber Kill Chain). Si bien en el planeamiento de una acción de Red Team son los pasos recomendados a seguir, en función del escenario la Kill Chain puede variar. De hecho, no es única, existiendo modificaciones adaptadas por otras empresas como Netskope y StormShield donde el número de fases puede no solo disminuir, sino también aumentar. Al final, de lo que se trata es de secuenciar el ataque, para de esta forma establecer unos procedimientos operativos, que los equipos de Red Team (aunque también es aplicable a los actores maliciosos) conduzcan este tipo de acciones de la forma más efectiva posible, maximizando resultados y reduciendo costes.

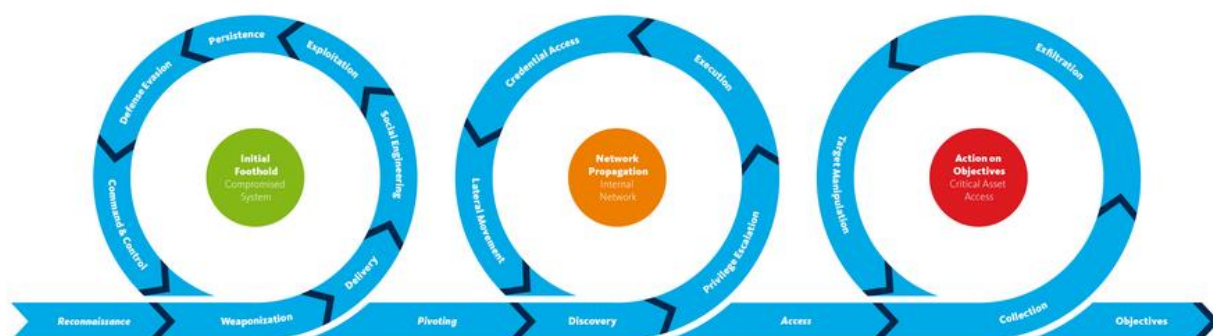


Figura 2-7 Modelo de Cyber Kill Chain de 18 fases [27].

2.4.2 Herramientas empleadas en ciberejercicios Red Team

Existe una amplia variedad de herramientas empleadas en ciberseguridad ofensiva y más específicamente para Red Team [28, 29]. Algunas de ellas incluso son empleadas en otras capas de la ciberseguridad. Por ejemplo, una herramienta de descubrimiento de vulnerabilidades puede ser empleada tanto por el equipo Blue (i.e. los responsables de la parte defensiva) como por un equipo de auditorías o de pentesting. También, las herramientas pueden agruparse por cada una de las fases de la Cyber Kill Chain en donde se emplean (aunque también existe la posibilidad de que sean empleadas en varias). A continuación, se procederá a enumerar alguna de las más comunes empleadas en el mundo de la ciberseguridad ofensiva.

12. Kali Linux: Se trata de una distribución GNU/Linux [30, 31] que más que una herramienta, es un compendio de herramientas enfocadas a la ciberseguridad ofensiva y auditoría [32, 33]. Kali Linux en su concepción, nace como un toolkit integrado, que condensa en una única plataforma una multitud de herramientas tanto de reconocimiento, explotación, post-explotación, análisis forense y recolección de telemetría, que no solo son útiles en el ámbito ofensivo, sino que también pueden ser empleadas por la parte defensiva. El empleo de Kali presenta riesgos que

deben mitigarse. Ningún test de intrusión ni Red Team puede garantizar que su actividad sea inocua en el sistema auditado. Por otra parte, existe también el riesgo reputacional. Este riesgo deriva de la filtración o pérdida de datos sensibles (e.g. vulnerabilidades, credenciales o fallos de configuración) de la entidad objetivo. Por eso es recomendable el uso de máquinas virtuales con una distribución de Kali Linux para cada acción, siendo eliminadas tras cada ejercicio tras la redacción del informe y el registro de evidencias.

13. Metasploit: Es un framework (comúnmente se conoce por su abreviatura MSF) de explotación libre (<https://www.metasploit.com/>) y de código abierto [34]. Existen herramientas comerciales con capacidades similares metasploit Pro de Rapid7, IMPACT de Core Security Technologies y CANVAS de Immunity. Pero incluso los operadores que disponen de esas herramientas comerciales suelen complementar su arsenal con la versión gratuita de Metasploit, y que viene de serie disponible en la distribución de Kali Linux. Es un entorno en el que se pueden emplear exploits para comprometer objetivos y en el que pueden desarrollarse nuevos exploits [35]. Un framework de explotación completo como Metasploit [36], ofrece numerosos componentes reutilizables y bibliotecas de código que simplifican y aceleran el proceso de creación de nuevos exploits. Metasploit funciona en Linux, macOS y Windows. Sin embargo, según la documentación de Metasploit para algunas versiones de la herramienta, “El Metasploit Framework solo tiene soporte parcial en la plataforma Windows. Si desea acceder a la mayoría de las funcionalidades del Framework desde Windows, recomendamos usar un entorno de virtualización, como VMware, con una distribución de Linux compatible...”. No existe documentación detallada sobre qué característica puede fallar en Windows, por lo que lo más prudente es instalarlo en otro tipo de sistema, como Linux. Si alguna funcionalidad no funciona correctamente en Metasploit sobre Windows durante un pentesting o Red Team, puede resultar en un falso positivo sobre una determinada vulnerabilidad de la máquina objetivo, haciendo que el ejercicio resulte mucho menos productivo.
14. Cobalt Strike: Se trata de una herramienta de post-explotación concebida para ejercicios de Red Team y emular campañas adversarias [37]. La herramienta es muy maleable y configurable, permitiendo modificar múltiples aspectos para enmascararla como malware conocido o servicios legítimos. Cuenta con capacidades nativas que permiten realizar múltiples TTP's [38, 21], incluyendo el compromiso inicial, escalada de privilegios y movimientos laterales (en la figura 2-6 se muestra una progresión en una red empleando esta herramienta). También cuenta con informes integrados y mapeo de matriz de MITRE. Cobalt Strike consta de tres componentes principales:
 1. Beacon: También conocido como BOF (Beacon Object File), es el agente de postexplotación de Cobalt Strike. Una vez que se ejecuta en el objetivo comprometido, se encarga de comunicarse con el Team Server, recuperar las tareas de postexplotación, ejecutarlas y enviar los resultados. Beacon se implementa como una DLL de Windows, pero se puede empaquetar en payloads de varios formatos, incluidos ejecutables y scripts de PowerShell.
 2. Team Server: sistema central de control y registro. Almacena la información recopilada durante una intervención y crea varios modelos de datos que pueden utilizarse para generar informes y en algunos flujos de trabajo de Beacon.
 3. Cliente: utilizado por los operadores del Red Team para conectarse a uno o más TEam Servers con el fin de gestionar la operación e interactuar con los payloads (i.e. el código malicioso embebido y que indica lo que hace el sistema objetivo una vez comprometido) del Beacon. Cuando se conecta a varios Team Server al mismo tiempo, el cliente recopila y agrega los datos de todos ellos para proporcionar una visión única y coherente de la intervención.

El patrón de diseño de Cobalt Strike permite operar de forma distribuida, estableciendo varios Team Server dedicados para cada fase de una operación/ejercicio, como el acceso inicial, la post-explotación y la persistencia. La razón es que, si se descubre y bloquea parte de una operación, se

pueden utilizar los demás canales para mantener el acceso. Cobalt Strike cuenta con múltiples funciones de consolidación de infraestructura que facilitan este trabajo.

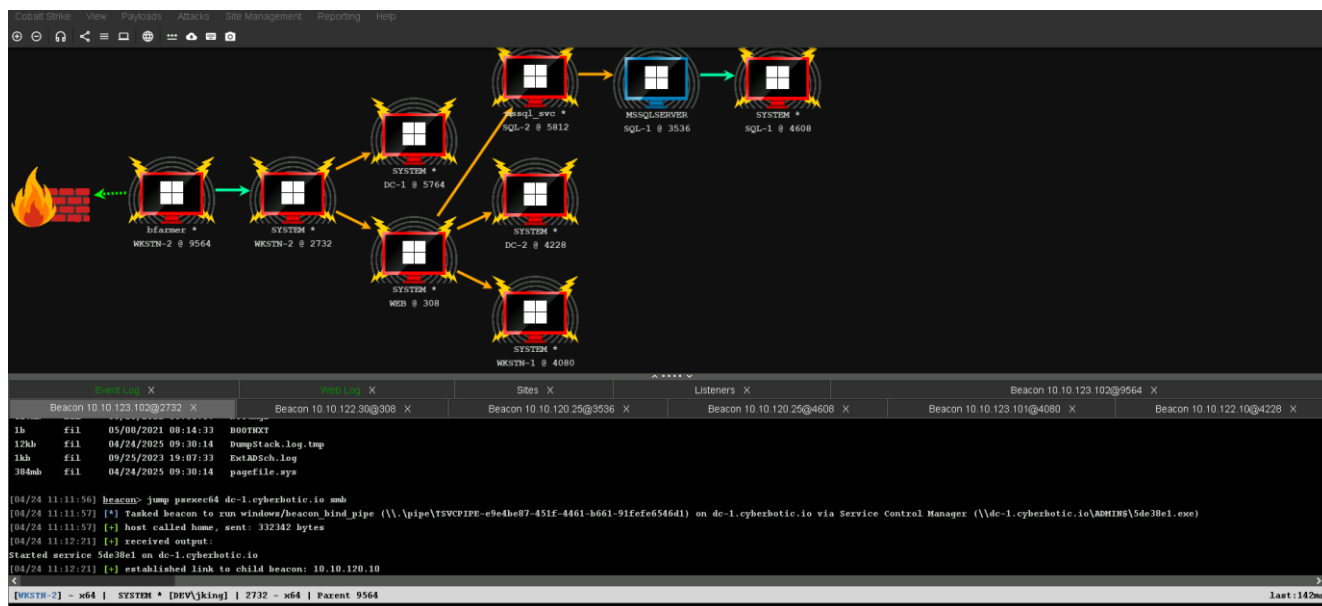


Figura 2-8 Interface de operación de Cobalt Strike durante un ejercicio de Red Team.

Se han detallado algunas de las herramientas más comunes empleadas en ejercicios de Red Team, aunque existen muchas más, tanto comerciales como de código abierto. Gran parte de ellas vienen instaladas por defecto en la distribución Kali Linux. Por mencionar algunas de las más relevantes, se encuentran NMAP, ampliamente empleada en la fase de reconocimiento de la Cyber Kill Chain. Entre sus principales funcionalidades están la de descubrir hosts en una red, identificar puertos y servicios abiertos, detectar versiones de software y sistemas operativos, y ejecutar scripts de comprobación adicionales mediante su motor de scripts (NSE, Nmap Scripting Engine). Otro ejemplo de herramienta extensamente utilizada es Nikto. Se trata de un escáner de servidores web de código abierto (escrito en Perl) diseñado para encontrar problemas de seguridad comunes en aplicaciones y servidores HTTP/HTTPS. Está orientado a pentesting y auditorías: detecta configuraciones inseguras, archivos y scripts peligrosos, versiones de servidores y módulos con vulnerabilidades conocidas, cabeceras inseguras y otros problemas que podrían facilitar un compromiso.

En definitiva, la lista de herramientas y utilidades enfocadas al Red Team es amplia y variada y cubre todas las fases de la Cyber Kill Chain. Aunque es importante tener presente que los grupos de cibercriminales y APT's desarrollan sus propias herramientas, lo que les ofrece una ventaja táctica a la hora de enfrentarse a los Blue Teams. Hay que tener en cuenta, aunque se emulen patrones adversarios, no siempre se puede conseguir una acción lo suficientemente similar al modo de actuación de los actores maliciosos, por lo que, siendo necesarios, los ejercicios de Red Team no garantizan una emulación completamente realista.

2.4.3 Experiencias internacionales aplicadas a la defensa nacional

La OTAN cuenta con centros de excelencia de distintas disciplinas repartidos por todos los estados miembros. En Tallín, Estonia, se encuentra el centro de excelencia de la ciberseguridad de la OTAN (CCdCOE por sus siglas en inglés). Como se refleja en su página web www.ccdcoe.org sus principales cometidos son “apoyar a los países miembros y a la OTAN con conocimientos interdisciplinarios únicos en el campo de la investigación, la formación y los ejercicios de ciberdefensa, que abarcan las áreas de tecnología, estrategia, operaciones y legislación”. Este centro organiza dos ejercicios anuales, que suponen un reto tecnológico llevando a los países y entidades participantes al límite. Estos dos ejercicios son Locked Shields (a partir de ahora LS) y Crossed Swords (a partir de ahora XS).

2.4.3.1 *Locked Shields*

Desde su primera edición en 2010, LS permite a los expertos en ciberseguridad mejorar sus habilidades en la defensa de los sistemas informáticos nacionales y las infraestructuras críticas frente a ataques en tiempo real. Se centra en escenarios realistas, tecnologías de vanguardia y la simulación de toda la complejidad de un ciberincidente a gran escala, incluyendo la toma de decisiones estratégicas y los aspectos legales y de comunicación.

Se trata de un ejercicio que enfrenta a Red Teams y Blue Teams, estos últimos formados por los países miembros del CCDCOE. Según consta en su página web oficial, en 2021 participaron 22 Blue Teams con una media de 40 expertos en cada uno. Los equipos asumen el papel de equipos nacionales de reacción rápida ante incidentes de ciberseguridad que se despliegan para ayudar a un país ficticio a gestionar un incidente en el ciberespacio a gran escala con todas sus implicaciones. El ejercicio de 2021 contó con unos 5000 sistemas virtualizados que fueron objeto de más de 4000 ataques. Los equipos deben ser eficaces a la hora de informar de los incidentes, ejecutar decisiones estratégicas y resolver retos forenses, jurídicos y mediáticos. Para mantenerse al día con los avances tecnológicos, Locked Shields se centra en escenarios realistas y tecnologías de vanguardia, redes relevantes y métodos de ataque.

En su última edición, en 2025, el objetivo principal del ejercicio fue proporcionar una plataforma multinacional para el entrenamiento en actividades de ciberdefensa en un entorno de crisis de ritmo acelerado. Este concepto facilita el desarrollo y la experimentación de futuras tecnologías, métodos, conceptos operativos y mejoras tecnológicas, al tiempo que facilita la cooperación y el intercambio de información dentro y entre equipos técnicos y estratégicos.

El diseño operativo de LS, donde el principal valor a evaluar es conocido como availability (i.e. disponibilidad de los servicios levantados para el ejercicio en un entorno degradado), incorpora elementos que resultan directamente transferibles a programas nacionales de resiliencia. Cuenta con un scoring multidimensional que combina métricas técnicas (detección, contención, recuperación) con indicadores organizativos (coordinación interinstitucional, gestión de crisis, comunicación en prensa y redes sociales). También define un paquete de telemetría y evidencia bajo unos estándares preestablecidos que permite realizar análisis forense reproducible. Además, la presencia de evaluadores independientes facilita la validación en el cumplimiento de reglas, procedimientos y resultados. El conjunto de estas características eleva la categoría de LS a algo superior y de más valor que una prueba técnica. Lo convierten en un ejercicio de gobernanza operativa diseñado para exponer y corregir fricciones organizativas que, en situación real, afectan la continuidad de servicios críticos y a la disponibilidad de los mismos.

Aunque el ejercicio se basa en el componente técnico, este se complementa con escenarios que introducen objetivos políticos y de comunicación que en la vida real se traduciría en presiones de stakeholders y el estrés en la gestión de incidentes con impacto público para forzar decisiones reales bajo incertidumbre. La combinación de acciones técnicas parametrizables junto con escenarios socio-políticos complejos, convierte los hallazgos en lecciones operativas útiles y punto de mejora accionables para responsables de seguridad nacional, responsables de continuidad y responsables políticos.

Locked Shields permite extraer una serie de lecciones que tienen tres elementos de aplicación directa sobre la seguridad nacional. Primero, se fundamenta en la capa de instrumentación y herramientas. LS demuestra que sin elementos de motorización (EDR, PCAP a nivel perimetral y de segmentación, registros de autenticación normalizados, SIEM con correlación) no es posible cuantificar el desempeño de los Blue Team y equipos de respuesta, así como verificar los mecanismos de remediación.

En segundo, en lo referente a la capa de gobernanza y procedimientos, LS fuerza la coordinación entre actores heterogéneos (defensa, operadores de infraestructura crítica, fuerzas de seguridad,

autoridades reguladoras y comunicación pública). La experiencia indica que la falta de coordinación en el intercambio de información, la ausencia de canales de escalado claros y la falta de roles definidos son causas recurrentes de fallo de transmisión de la información. Por ello, la extrapolación de las lecciones aprendidas en LS a nivel nacional exige protocolos de coordinación predefinidos (e.g. comités ejecutivos, puntos de contacto interinstitucionales y playbooks de comunicación) y la inclusión de procedimientos legales previos.

El tercer y último elemento se centra en el apartado de la formación y capital humano. Locked Shields combina equipos técnicos de diferentes procedencias y administraciones con responsables de toma de decisiones. Esto pone de manifiesto la necesidad de programas de formación multidisciplinar y ejercicios integrados que integren capacitación técnica (detección, forense, remediación) con capacidades de mando y gestión. LS podría equipararse a una escuela de softkills bajo condiciones de estrés (priorización de recursos y gestión de reputación), competencias críticas para la resiliencia nacional.

Locked Shields (en la figura 2-7 se muestra la fase de preparación del ejercicio en su sede en Tallín, Estonia) supone un claro ejemplo de los beneficios de ejercicios de Red Team. Siendo una propuesta mucho más ambiciosa, en el fondo es un ejercicio que complementa la parte ofensiva con la coordinación multinivel durante la resolución de una crisis motivada por un entorno degradado bajo un ciberataque masivo. Entrar a formar parte de los equipos de Red Team en LS supone un reto al alcance de muy pocos, ya que los operadores ofensivos seleccionados deben poseer unos extensos conocimientos en varias disciplinas, además de una coordinación entre los miembros de su equipo.



Figura 2-9 Personal implicado en el ejercicio Locked Shields 25 en su sede en Estonia. Fuente: CCdCOE.

2.4.3.2 Crossed Swords

Crossed Swords es un ejercicio integral de ciberdefensa cuyo objetivo es mejorar las capacidades tanto de los elementos de mando militar a nivel operativo en la gestión de las capacidades ofensivas y defensivas del ciberespacio, como de los especialistas en ciberdefensa a nivel táctico en la ejecución de operaciones en el ciberespacio defensivas y ofensivas de amplio espectro en un entorno simulado de crisis y conflicto. Además de la formación en guerra en el quinto dominio, el ejercicio promueve el desarrollo de competencias operativas multidominio y otras habilidades tácticas relacionadas con el ciberespacio, como las operaciones de guerra de información, las perspectivas jurídicas y la cooperación con las fuerzas de operaciones especiales.

Desde la edición de 2023, se cuenta con un CHQ (cyber head quarter) que coordina las acciones de la OCU (offensive cyber unit), compuesta únicamente por personal militar de diferentes países. A partir de 2025, el ejercicio se amplía para incluir dos CHQ con personal de planificación al completo. Entre los participantes se encontrarán operadores militares, analistas forenses digitales, threat hunters y otros especialistas de las Mandos del Ciberespacio nacionales o unidades que colaboran estrechamente con ellas. Cabe destacar que en la edición de 2025 el ejercicio amplió su alcance al incorporar a participantes del sector público y privado en la planificación de la respuesta del CHQ, con especial atención a la protección de las infraestructuras críticas nacionales. El realismo y la eficacia del ejercicio han sido significativamente mejorados gracias a la colaboración con socios académicos e industriales, que colaboraron para garantizar la autenticidad y relevancia. Estas asociaciones también contribuyeron a la integración de tecnologías avanzadas, incluidas herramientas de inteligencia artificial, en aspectos clave del ejercicio.

Para entender la aplicabilidad de un ejercicio del tipo XS a la seguridad nacional es necesario entender sus elementos diferenciadores. En primer lugar, ofrece un enfoque integral, donde se combinan los perfiles técnicos (integrados en las unidades ofensivas) como de mando y fuerzas cinéticas. Esta amalgama de elementos hace indispensable un nivel de coordinación muy elevado para la correcta consecución del ejercicio. En segundo lugar, es un ejercicio orientado al perfeccionamiento del ciclo de mando (MDMP, ROE, LEGAD/INFOOPS) a través de un CHQ multinacional con battle rhythm realista. En tercer y último lugar, la componente de experimentación y aprendizaje, materializada en el Yellow Team, monitoriza la actividad de las unidades de ciberseguridad ofensiva para así poder retroalimentar y formar a los elementos de threat hunting, factor clave para su extrapolación a un entorno de ciberseguridad activa en el marco nacional.

En términos pedagógicos y de preparación operativa, XS se orchestra en cuatro fases:

1. Formación fundamental
2. Planificación de respuesta del CHQ (CHQ response planning)
3. Ejecución, fase en la que los equipos se desplazan a Tallin, preparan las operaciones y las ejecutan
4. Análisis y reporting

Este ciclo permite validar no sólo habilidades técnicas (aquellas que se pueden encajar en alguna fase de la Cyber Kill Chain como son explotación, persistencia, movimiento lateral) sino también procedimientos operativos, empleo de inteligencia como producto final de los equipos de análisis y coordinación con fuerzas cinéticas, convirtiendo a XS en un instrumento idóneo para ejercitar la gobernanza militar en el ciberespacio.

Los objetivos de XS subrayan su doble propósito: (a) capacitar al personal técnico en la ejecución de OCO completos y (b) entrenar al personal de mando en la planificación, priorización y coordinación de campañas ofensivas (tasking, C2, LEGAD/INFOOPS). Además, XS busca facilitar intercambio de buenas prácticas y confianza multilateral entre países aliados, lo que contribuye a la construcción de redes operativas y doctrinales de carácter multinacional.

Crossed Swords (en la figura 2-8 se muestra a personal participante durante la ejecución del ejercicio en Estonia) aporta valor en diferentes aspectos para la seguridad nacional, porque permite ensayar una acción ofensiva a gran escala en el ciberespacio bajo parámetros militares controlados, como ya sucedió en el mundo real con anterioridad, como es el caso del ciberataque masivo a Estonia en 2007 [39, 40] o las acciones en el quinto dominio llevadas a cabo por los combatientes en la

invasión rusa a Ucrania. Este ejercicio genera evidencias operativas sobre procedimientos, capacidades y vacíos organizativos, y permite adiestrar al factor humano en la interacción de dominios (ciber, legal, comunicación y cinética). Crossed Swords es el perfecto ejemplo de preparación para un escenario de conflicto multidominio, que eleve la preparación del país para enfrentar crisis de estas características.



Figura 2-9 Personal involucrado en la ejecución del ejercicio Crossed Swords.
Fuente: CCdCOE.

3 PROPUESTA

3.1 Metodología propuesta para operaciones Red Team en ciberejercicios

En el presente capítulo se plasmará una propuesta que plantea articular el marco operativo y metodológico para la planificación, ejecución y evaluación de las operaciones de Red Team, encuadrados en ciberejercicios cuya finalidad última es la mejora de la defensa nacional. Se va a partir de la premisa de que las pruebas donde se emulan comportamientos adversarios posibilitan la mejora de la respuesta a incidentes y resiliencia nacional, al generar información accionable para la dirección, los supervisores y los equipos de respuesta. Por lo tanto, en este capítulo no se busca describir las principales tácticas, técnicas y procedimientos empleados en el mundo del Red Team, ni tampoco una explicación técnica sobre desarrollo de exploits, se busca presentar un método reproducible, escalable y verificable que replique amenazas plausibles y de especial relevancia para el país. Además, también se busca medir la capacidad defensiva sometiéndola al estrés de acciones ofensivas complejas, para así transformar los hallazgos realizados en mejoras permanentes.

Para el diseño se han tomado como referencia los marcos internacionales TIBER-EU y CBEST, adoptando principios ya consolidados. En ambos marcos la inteligencia previa juega un papel clave a la hora de seleccionar las TTP's de los grupos a emular. También se establecen una serie de roles y responsabilidades claras, ROE's que limitan el alcance y por lo tanto limitan el riesgo (si bien hay que reseñar que este nunca es nulo). Estos son los elementos principales sobre los que se sustentará el enfoque a desarrollar: se da credibilidad a la operación de Red Team gracias al empleo de inteligencia previa, la gobernanza garantiza la legalidad y control, las ROE's establecen límites que ayudan a proteger a los activos y los entregables generados permiten una auditoría y análisis exhaustivo de los resultados.

Una de las características más destacadas de esta propuesta es su enfoque en la medición verificable. Para que un ejercicio realmente aporte valor en la toma de decisiones, es fundamental que los resultados se expresen en términos cuantificables. Esto incluye KPIs técnicos, así como métricas de impacto que abordan el tiempo de degradación de funciones críticas y la pérdida de disponibilidad. También se consideran medidas organizativas, como los tiempos de escalado y la calidad del reporte interinstitucional. Todo esto debe ir acompañado de criterios claros para la evaluación y la reproducibilidad de la operación. Para lograrlo, se recomienda el uso sistemático de una ontología operativa, como MITRE ATT&CK, que facilite el mapeo de TTPs, la especificación de la telemetría mínima por técnica y la definición de un scoring que sea reproducible.

Todo esto tiene como finalidad transformar los resultados técnicos en indicadores que la dirección y los supervisores puedan entender y comparar a lo largo del tiempo (i.e. establecer un estado de referencia o baseline en inglés que permita la continuidad). En la práctica, la propuesta se estructura

como un proceso por fases: scoping, TTI, diseño del Red Team Test Plan, test-run, ejecución controlada, reporte y retest (nueva ejecución tras la corrección de las deficiencias detectadas). Cada una de estas fases incluye controles de calidad y seguridad en cada hito. Además, cada fase cuenta con productos obligatorios que siguen un formato estandarizado: el informe TTI que justifica la selección de técnicas; las RoE que establecen límites operativos; el plan de pruebas técnicas con mapeo ATT&CK y requisitos de telemetría; el informe técnico del Red Team que incluye un manifiesto de evidencias; y el AAR ejecutivo que se orienta a la toma de decisiones. También se introduce la figura del Evidence Custodian y prácticas de cadena de custodia, como hashing, manifiesto y un custodio designado, para garantizar la integridad de la información utilizada por supervisores y auditores. La propuesta también tiene en cuenta criterios de proporcionalidad y escalado: no todos los ejercicios necesitan realizarse en producción ni con la misma intensidad. Sugerimos una hoja de ruta escalonada que comience con pilotos controlados, abarcando un servicio único y entornos representativos, y que progrese hacia ejercicios multi-actor y de mayor complejidad.

3.1.1 Fases de planeamiento, ejecución y evaluación

Una operación de Red Team implica modelar las técnicas utilizadas por actores maliciosos reales al nivel que sea, bien grupos cibercriminales o APT's, para encontrar vulnerabilidades y, bajo circunstancias controladas, emplear esos fallos de seguridad para explotarlos de manera profesional y segura, de acuerdo con un alcance cuidadosamente diseñado y ROE's, con el fin de determinar el riesgo para la operativa de la institución y el impacto potencial, todo ello con el objetivo de ayudar a la organización a mejorar sus prácticas de seguridad.

Al contrario de un pentesting, donde la clave es que la identificación y explotación de todas las vulnerabilidades posibles, un Red Team se enfoca en explotar aquellas que permiten el acceso al sistema. Por otra parte, no hay que perder de vista que el objetivo del ejercicio debe estar centrado en el servicio prestado por la organización auditada, o lo que en el entorno empresarial sería el negocio. Los hallazgos de una y lecciones derivadas de estos ejercicios deben centrarse siempre en el riesgo organizacional, no solo en el riesgo técnico.

Muchas organizaciones utilizan hacking ético, pruebas de penetración y ejercicios de Red Team para encontrar fallos de seguridad antes que los atacantes maliciosos. Después de aplicar sus políticas, procedimientos y tecnología de seguridad, una organización gubernamental debería realizar un ciberejercicio complejo para evaluar ver cuán efectiva es su seguridad frente a un ataque real, aunque sea realizado por atacantes amistosos.

Un beneficio adicional de los ciberejercicios, al mostrar vulnerabilidades reales e indicar lo que un atacante malicioso podría ser capaz de lograr, pueden captar la atención de las capas directivas. Los responsables de la toma de decisiones, cuando se les presentan los resultados cuidadosamente formulados de una prueba en términos empresariales, tienen más probabilidades de proporcionar recursos y atención para mejorar la postura de seguridad de una organización.

Para el correcto desarrollo de un ciberejercicio, y la maximización de los resultados de las operaciones de Red Team, se hace necesario un ciclo operativo completo y reproducible. El objetivo es definir fases claras, productos obligatorios derivados de los resultados de las acciones ofensivas, responsabilidades y criterios de aceptación que garanticen seguridad, trazabilidad y valor operativo. Cada fase incluye una lista de resultado mínimo y productos exigibles, que deben recogerse en las *Rules of Engagement* (RoE) y los NDA's.

3.1.1.1 Fase de Planeamiento

El propósito principal de esta fase es fijar el alcance, garantizar la gobernanza y generar inteligencia accionable que será la base sobre la que se fundamentará el diseño técnico del ejercicio. Según esto, podemos diferenciar los siguientes apartados:

3.1.1.1.1 Definición del alcance funcional y el riesgo

Las partes que participan en el ejercicio deben firmar un acuerdo de confidencialidad (NDA), especialmente si se ven involucradas organizaciones externas, como pueden ser empresas de ciberseguridad ofensiva. Después, los evaluadores y el personal del objetivo discuten las preocupaciones más importantes de la organización objetivo. En este momento se responden a preguntas del tipo: ¿Cuáles son las mayores amenazas? ¿Qué sistemas son los más sensibles? ¿Qué tipo de información es la más valiosa? También se acuerdan las Reglas de Enfrentamiento que describen cómo se llevará a cabo la operación de Red Team. Una definición precisa del alcance es imprescindible. Uno de los elementos más importantes que incluir en el alcance del proyecto es una declaración concisa de lo que será probado. Para ello, se expondrán explícitamente aquellos nombres de dominio, rangos de direcciones de red, hosts individuales y aplicaciones específicas que están incluidas en el ciberejercicio. También deberán registrarse todos aquellos elementos fuera de scope.

Durante la realización del ejercicio existe la posibilidad del descubrimiento de sistemas adicionales dentro de rangos de direcciones de red o nombres de dominio que no fueron considerados en la formulación original del alcance. Es preciso que las Reglas de Enfrentamiento indiquen a los operadores que consulten con la dirección del ejercicio antes de probar cualquier otra máquina fuera del alcance original que se descubra mediante reconocimiento o escaneo después de iniciado el ejercicio.

Si se incluyen en el alcance sistemas propiedad de terceros o gestionados por ellos, es necesario obtener autorización documentada de esos terceros antes de comenzar la prueba. Es responsabilidad de la organización objetivo obtener este permiso. Muchas organizaciones importantes hoy en día utilizan al menos un tercero para gestionar parte de su infraestructura informática y de red. Algunas empresas subcontratan esto completamente, con su infraestructura realmente perteneciendo al propio tercero. Ejemplos incluyen proveedores de servicios de internet, cuyos routers, switches, servidores de correo y servidores DNS pueden entrar dentro del alcance de una prueba.

Muchos proveedores de servicios en la nube ofrecen sistemas virtualizados subyacentes, en un modelo de plataforma como servicio (PaaS) o infraestructura como servicio (IaaS), que permiten a los clientes realizar pruebas a nivel de aplicación (como pruebas de aplicaciones web) pero prohíben realizar pruebas sobre vulnerabilidades subyacentes de red y servicios en escucha.

3.1.1.1.2 Gobernanza y roles

La gobernanza de un ciberejercicio *Red Team* tiene como objetivo definir tanto las responsabilidades como el marco operativo en aras de una ejecución segura del ejercicio, ajustándose al marco legal y normativo, y que sea de utilidad para la posterior toma de decisiones. Para evitar posibles incidentes no deseados se hace imprescindible una gobernanza fuerte, que permita añadir valor tras el ejercicio. Sus principales cometidos son garantizar que las decisiones críticas se tomen por parte de la autoridad adecuada, que la custodia y la integridad de las evidencias generadas sean las adecuadas, así como controlar los riesgos legales antes, durante y tras el ejercicio. Para conseguirlo, se definen una serie de roles:

1. Responsable ejecutivo de la entidad auditada: Responsable de aceptar el alcance del ejercicio, firmar las ROE's y aprobar el plan de remediación y los plazos.
2. Equipo de supervisión operativa: Encargado de supervisar en tiempo real la ejecución del ciberejercicio. También es el responsable del registro de decisiones operativas y comunica incidentes reales al responsable ejecutivo.

3. Red Team: Es el equipo responsable de realizar las operaciones ofensivas en base al modelado de TTP's acordado (en la figura 3-1 se muestra un ejemplo de modelado de TTP's).
4. Responsable de evidencias: Es el personal a cargo de la custodia de las evidencias digitales obtenidas por el Red Team.

Estos son los roles mínimos exigibles para la correcta consecución del ejercicio. No obstante, se pueden definir más roles en función de cada ejercicio concreto. Por ejemplo, si se tratase de un Red Team perteneciente a la administración pública, y existiese el interés en evaluar su OPSEC (i.e. cómo de sigilosas son sus acciones para evitar la detección), podría sopesarse el rol de Yellow Team, grupo encargado para la monitorización de red y que puede aportar valor para la mejora del desempeño del Red Team.

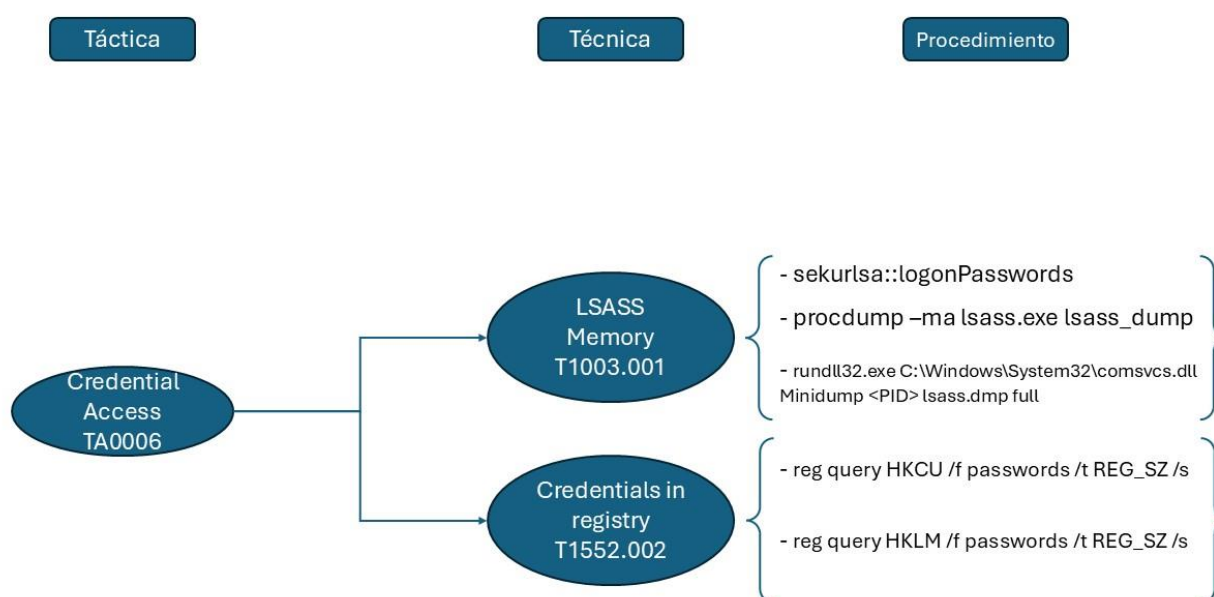


Figura 3-1 Ejemplo de TTP's mapeadas en la matriz de MITRE. Se diferencian las tácticas por el código TA de las técnicas por el código T

3.1.1.1.3 Informe de inteligencia del objetivo

De forma similar al marco TIBER-EU (que contempla el Target Threat Intelligence), para que el ejercicio emule de manera correcta las amenazas reales que pueden afectar a la organización objetivo [41, 42, 43], es necesario contar con un informe técnico y operativo que documenta, justifica y prioriza las amenazas y TTPs relevantes para el objetivo del ejercicio. Su propósito es obtener inteligencia accionable, que marque el patrón de selección de aquellas técnicas objeto de la emulación. Para ello es preciso contar con feeds de inteligencia proporcionados por empresas punteras (como puede ser el caso de Recorded Future), un análisis de campañas llevadas a cabo por actores maliciosos en sectores similares a los que desarrolla su actividad la organización objetivo y un mapeo preliminar a MITRE ATT&CK que facilite la trazabilidad técnica. Además, incorpora una metodología de priorización (e.g. una puntuación compuesta basada en probabilidad de ocurrencia, explotabilidad e impacto sobre funciones críticas) que permite justificar por qué ciertas técnicas se autorizan y otras quedan fuera del alcance. También se debe especificar para cada técnica la telemetría mínima requerida (EDR, sysmon, PCAP, logs de autenticación, NetFlow) y las métricas que se emplearán para su scoring (detección automática, detección humana, tiempo a detección).

3.1.1.1.4 Reglas de enfrentamiento

Las reglas de enfrentamiento, o ROE, dictan el cómo se va a llevar a cabo la evaluación de lo que se definió en el alcance una vez completada la fase de planeamiento [44]. La legitimidad y legalidad de todas las acciones que lleva a cabo el Red Team mientras realiza la evaluación están ratificadas por las ROE. Como ya se ha mencionado, un documento de ROE bien establecido y acordado debe ser reconocido y firmado por ambas partes, tanto los operadores ofensivos como el responsable ejecutivo de la organización auditada. De lo contrario, la participación ofensiva en materia de ciberseguridad por parte de hackers éticos, pentesters o red teamers puede constituir la violación de la ley establecida, con las consiguientes responsabilidades penales.

Por ende, en última instancia la finalidad de las ROE's es establecer las responsabilidades, relaciones y guían de enfrentamiento entre el Red Team y los responsables de la toma de decisiones de la organización auditada.

3.1.1.1.5 Plan de evidencia y cadena de custodia

El plan de evidencia y la cadena de custodia supones dos elementos críticos que, a partir de los resultados recopilados por el personal técnico, materializan pruebas admisibles y verificables, útiles tanto para la toma de decisiones como para la supervisión. Este plan tiene como finalidad asegurar la confidencialidad, disponibilidad, integridad y trazabilidad de todos los productos generados durante el ciberejercicio, como pueden ser capturas de tráfico (PCAP), logs, IOC's, capturas de pantalla o cualquier otro resultado de la acción ofensiva. El plan define quién produce cada una de las evidencias, el responsable de su registro y transferencia en cada paso y el personal que puede acceder a ellos. La ausencia de la cadena de custodia invalida el valor probatorio de la misma, incrementando el riesgo legal y reputacional. Esto último se escenifica con un ejemplo claro. En el caso de que durante la acción se encuentren evidencias de un ciberataque real en curso, para que tengan validez legal han de cumplirse una serie de requisitos, entre ellos la cadena de custodia. Debido a la importancia de este plan, este apartado debe figurar como producto obligatorio en las Rules of Engagement y en el NDA.

El proceso de custodia y transferencia debe incluir formularios de registro y firmas en cada salto de custodia. Desde el momento de generación de la evidencia se hace una copia inicial, que se plasma en el registro junto con su hash (prueba digital de la no alteración de esta). Una vez iniciado el registro, se procede al almacenamiento seguro, accesible para que los investigadores puedan consultarlas, o sean entregadas a la autoridad pertinente. Cada transferencia debe quedar reflejada en la bitácora indicando quién, qué, cuándo y por qué se accede a la misma, siendo la firma del responsable de la custodia el sello que certifica la integridad del paquete en cada hito. Las normas sobre retención y eliminación se pactan contractualmente y deben cumplir con la normativa aplicable (e.g. GDPR para datos personales): incluir plazos, criterios de anonimización, condiciones para retención extendida por requerimiento legal y procedimiento documentado de destrucción segura.

Poniendo el foco en el punto de vista legal y de privacidad, el plan debe enlazar con cláusulas de confidencialidad y con los protocolos de notificación a autoridades responsables. Si en el ejercicio se hallan datos personales reales (e.g. registros médicos por el uso privado de un terminal de trabajo, suponiendo un mal uso del mismo), la evidencia debe sufrir un proceso de anonimización que evite la identificación del usuario, así como la propia información en sí. En caso de que la evidencia deba presentarse a una autoridad judicial, el responsable de la cadena de custodia debe poder producir un paquete con firma digital verificable y un registro de la cadena de custodia que permita validar la integridad desde la creación hasta la entrega [45, 46, 47].

3.1.1.2 Fase de ejecución

Una vez que se han definido tanto alcance como ROE's, y se ha firmado el NDA, es el momento de realizar el ejercicio. Es recomendable, previo inicio del ejercicio realizar un test-run. Aquí radica una de las diferencias principales con un Red Team realizado por una empresa contratada, donde este punto se obvia. El test-run tiene como objeto la validación, antes del inicio del ejercicio

real, de que todas las herramientas, procedimientos, flujos de notificación y control funcionan tal y como se han diseñado. Una de las principales ventajas que plantea el test-run es la disminución del riesgo de impactos inesperados, retrasos imprevistos y asegurar que las evidencias podrán recogerse con calidad, sirviendo también como criterio para el go/no-go del ejercicio. Otra recomendación es la confección de una check list (se ha reflejado un modelo de check list en el Anexo B) para cada una de las áreas involucradas en el ejercicio, para poder contar así con un criterio objetivo para continuar con el ejercicio, en base a criterios previamente establecidos.

Desde el punto de vista operacional, en un ciberejercicio de seguridad ofensivo la parte que involucra al Red Team es el núcleo principal. Este equipo es el responsable de emular todas aquellas tácticas, técnicas y procedimientos empleados por los actores adversarios. Además, se presenta la peculiaridad de que no debe cubrir sólo la parte técnica. Este equipo ha de ceñirse al marco controlado preestablecido que garantice la trazabilidad de la evidencia y la integridad de los activos, de forma que se puedan articular decisiones rápidas ante riesgos reales. Por ello, aunque no imprescindible, sí que es recomendable la realización del test run que permita la verificación final de sistemas, ROE's, alcance y aquellas salvaguardas que permitan realizar el ejercicio de una forma segura y controlada. Así mismo, debe producirse una "kick off call". Este término se refiere a la reunión final de coordinación donde se da el go al ejercicio.

Durante la ejecución han de aplicarse todas aquellas salvaguardas preestablecidas que garanticen la integridad de los sistemas auditados, parando el ejercicio si llegase a ser necesario. Como se ha mencionado anteriormente, un posible criterio de parada del ejercicio es el hallazgo de un compromiso real. Si durante el ejercicio se descubriese algún artefacto que estuviese realizando actividades maliciosas en algún host, se hace imprescindible notificarlo a la mayor brevedad para que intervengan los equipos de respuesta a incidentes.

Otro factor esencial durante la ejecución es la documentación en tiempo real. No se trata de la redacción del informe en tiempo real, sino el registro de toda actividad en el momento, para que la información se recopile con la mayor precisión posible. Para ello, existen múltiples herramientas de código abierto para este cometido. Entre ellas están Cherry Tree (<https://www.giuspen.net/cherrytree/>) u Obsidian (<https://obsidian.md/>), herramientas muy prácticas a la hora de ir registrando toda la actividad llevada a cabo durante el enfrentamiento, y que servirá como base para la redacción del informe final. Es importante registrar de manera exhaustiva todas y cada una de las acciones. En el caso en que alguna acción produjese una evidencia, debe realizarse una copia de la misma y registrarse su hash para garantizar su integridad.

Para que tenga una aplicación práctica en el mundo real, los procedimientos establecidos para la gestión de incidentes deben ser lo más fieles posible a la realidad. Dentro del registro de actividad que se va llevando han de anotarse con exactitud todas las decisiones tomadas antes una incidencia, así como los tiempos de escalado desde que el operador detecta la amenaza hasta que el team lead toma la decisión oportuna. Uno de los valores a evaluar lo compone el tiempo desde la ejecución de la actividad maliciosa hasta la detección de los efectos de esta. Este valor temporal será una de las métricas incluidas en el informe final y dará una idea de la capacidad de respuesta de la organización frente a un incidente real.

Una vez concluida la ventana de ejecución, comienza la desactivación. Este es un proceso en el que el Red Team debe destruir todos los artefactos empleados en el compromiso, así como la infraestructura empleada (en la figura 3-2 se muestra un ejemplo de infraestructura ofensiva empleada en una campaña de Red Team). Por artefactos se entiende todas aquellas balizas generadas por los C2 (e.g. Cobalt Strike o Brute Ratel) que se encuentran dentro de los hosts comprometidos. De la misma forma, han de eliminarse todas aquellas persistencias implantadas derivadas del compromiso de otros elementos de la red objetivo, tras los pertinentes movimientos laterales. La infraestructura hace referencia a la red de VPS empleados tanto como redirectores como aquellos donde se alojó el Team Server (en el caso de Cobalt Strike), o los empleados para el envío de phishing.

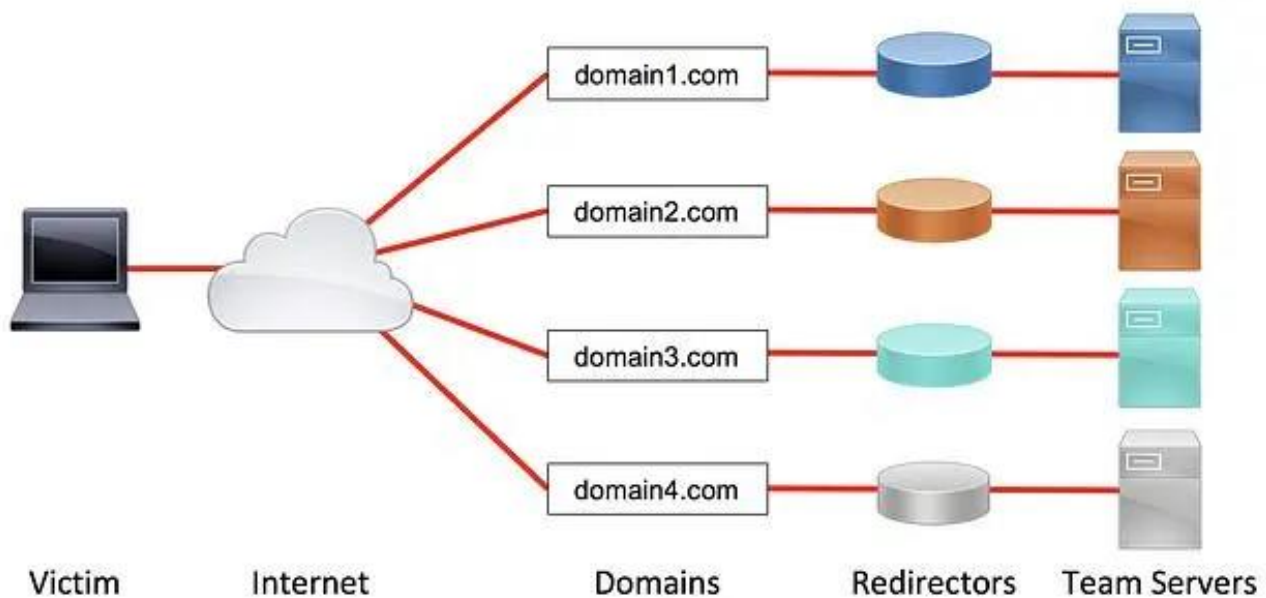


Figura 3-2 Ejemplo de infraestructura levantada para un ejercicio de Red Team [48].

3.1.1.3 Fase de evaluación

El elemento principal para la fase de evaluación es el informe generado por el equipo de seguridad ofensiva, confeccionado en las semanas posteriores al ejercicio. El formato del informe puede cambiar, pero como mínimo se recomiendan una serie de apartados: resumen ejecutivo, introducción, hallazgos, metodología y anexos.

3.1.1.3.1 Resumen ejecutivo

Como su propio nombre indica, es una síntesis global del informe y está enfocado a la parte ejecutiva de la organización, que será la responsable de tomar acción sobre los hallazgos encontrados. El Resumen Ejecutivo debe ser bastante breve. Idealmente, ocupará menos de una página y no deberá superar una página y media. No debe ser muy técnico ni detallar cada hallazgo. Debe estar redactado de manera que pueda leerse por sí solo, ofreciendo conclusiones firmes y recomendaciones para alguien que no necesariamente leerá el resto del documento. El Resumen Ejecutivo comienza con una breve descripción del proyecto en tres o menos frases, describiendo la fecha y los objetivos del proyecto, así como quién participó en él. También proporciona una breve visión general de lo que se hizo.

3.1.1.3.2 Introducción

Esta sección, de entre una y tres páginas, proporciona una visión general del proyecto para que el lector entienda cuándo ocurrió el proyecto, qué se incluyó en el alcance (y posiblemente elementos que se dejaron fuera deliberadamente, si corresponde) y quién participó en el ejercicio. Incluye una lista de

personas, organizada en una tabla, que identifique a los operadores y a las personas asociadas con el entorno objetivo que apoyaron a los operadores ofensivos. La tabla debe incluir el nombre de cada individuo, su rol y su información de contacto, como número de teléfono y dirección de correo electrónico. Luego, proporciona un breve resumen de los hallazgos más relevantes, utilizando un lenguaje similar al del Resumen Ejecutivo, pero posiblemente con más detalle técnico.

Estos componentes de la Introducción en los informes son inmensamente útiles seis meses después del proyecto o más tarde cuando se analiza qué fue probado. Muchas organizaciones suelen consultar estas Introducciones para determinar qué más en su entorno necesita ser probado y qué necesita volverse a probar.

3.1.1.3.3 Hallazgos

En este apartado se describen los resultados técnicos del ejercicio. Es recomendable iniciar esta sección con los hallazgos de alto riesgo, priorizados de modo que los de mayor impacto aparezcan primero. Se continúa con los de riesgo medio y bajo, priorizando nuevamente cada uno dentro de su subsección. Para cada hallazgo, se enumeran los sistemas que presentan el fallo, identificando la máquina por dirección IP y nombre. Incluye el nivel de riesgo y una estimación de cuán fácilmente podría explotarse el problema. El nivel de riesgo suele clasificarse en Alto, Medio y Bajo, pero cada organización puede utilizar otra forma de categorizar riesgos. Algunos operadores también añaden “Crítico” como categoría por encima de “Alto”, e “Informativo” como categoría por debajo de “Bajo”. Luego se proporciona un resumen de dos frases del hallazgo, nuevamente centrado en la medida en que el problema afecta a la operativa de la organización. Para algunos hallazgos de menor riesgo, el impacto operacional puede ser bastante pequeño.

A continuación, se entra en los detalles técnicos. Aquí se proporciona una descripción de qué es la vulnerabilidad y cómo un atacante podría explotarla y causar un impacto en la operatividad de la organización objetivo. La explicación se describe en términos técnicos, pero nuevamente, vinculándola con los riesgos operacionales, pudiendo incluirse otros como el reputacional.

Finalmente, en este apartado debe incluirse una sección de recomendaciones, que debe proporcionar una o más formas de remediar la vulnerabilidad descubierta. Si hay múltiples maneras de solucionar el problema, es recomendable incluirlas todas, pero destacando la que tenga más probabilidad de ajustarse a las necesidades operacionales y al entorno técnico de la organización objetivo.

Es importante no incluir en este apartado aquella información sensible obtenida durante el ejercicio, como, por ejemplo, aquellas credenciales que han podido ser exfiltradas. Este tipo de documento suele compartirse entre muchas personas de la organización objetivo. Incluir contraseñas en un informe que podría reutilizarse en otro lugar puede constituir un riesgo de seguridad, así que es necesario verificar con la organización objetivo si tal información debiera incluirse. Llegado el caso, es preferible describir las características de las contraseñas crackeadas (como longitud o conjunto de caracteres) pero sin incluir las contraseñas reales en el informe. Otro tipo de elementos como diagramas de red o capturas de pantallas son incluidos en este apartado y dan valor al informe final.

3.1.1.3.4 Metodología

En esta sección se proporciona una descripción del proceso paso a paso utilizado, y resume brevemente los hallazgos descubiertos en cada fase. Dependiendo de cómo se conduzca el ejercicio, esta sección puede incluir detalles de las fases de reconocimiento, escaneo y obtención de acceso. También se puede incluir una lista de herramientas usadas en el proyecto. En esta parte se describirán las TTP's empleadas en todas las fases del ejercicio, incluyendo también aquellos desarrollos que se hallan hecho para lograr la intrusión (e.g. evasión de antivirus y EDR).

3.1.1.3.5 Anexos

En este apartado se incluyen los resultados extensos que serían demasiado voluminosos para colocarlos dentro del cuerpo principal del documento. Un caso de resultado extenso sería, por ejemplo, la salida detallada de los escaneos de vulnerabilidades. Sin embargo, es necesario asegurarse de que esta información sea realmente útil para la evaluación del ejercicio, y pueda materializarse en una mejora de la seguridad de la organización objetivo. En este apartado pueden incluirse las reglas de enfrentamiento y el alcance.

A partir de este informe que recopila toda la información generada en el ejercicio desde la perspectiva ofensiva, hay que transmutar la evidencia en crudo recogida durante la ejecución para materializarla en conclusiones que deriven en un plan accionable para la dirección y en un plan de remediación. La parte de la evaluación tiene como finalidad cuantificar según criterios objetivos los patrones de comportamiento defensivos frente a las TTP's emuladas, identificar aquellos vacíos técnicos y organizativos que permitieron la intrusión, priorizar acciones según el riesgo y el impacto previsto. Es recomendable que esta parte se inicie justo después del fin de las operaciones ofensivas, articulándose en dos niveles: análisis técnico y síntesis ejecutiva.

El análisis técnico validará la integridad de las evidencias recolectadas, reconstruirá el timeline por cada una de las técnicas utilizadas y extraerá los IOC's (Indicators of Compromise). Toda conclusión técnica debe poder ser trazada a la inversa, para llegar sin ningún tipo de ambigüedad al elemento que lo generó (de ahí la importancia de hashear las evidencias). El análisis cuantitativo de los resultados se hará en función de la tabla de puntuación del ejercicio que haya sido establecida en base a los criterios previamente definidos. La priorización de los hallazgos técnicos enlazará con el impacto y la criticidad de la operatividad de la organización auditada. Como ejemplo, la exfiltración de credenciales de un determinado sistema clasificado. Para una correcta evaluación es recomendable la generación de una matriz de priorización que incluya, por ejemplo, riesgo residual, recomendación técnica, responsable y criterio de validación y verificación (e.g. regla en el SIEM que alerte de un determinado patrón con una tasa de falsos positivos inferior al 8%).

La síntesis ejecutiva, o resumen ejecutivo, es la parte destinada a la alta dirección y responsables de la toma de decisiones. Debe ser conciso, y con un claro enfoque a la toma de decisiones, incluyendo aquellos elementos que ya se han descrito anteriormente. Esta es quizás la parte más importante de la evaluación, ya que debe tener elementos y argumentos de peso para que los responsables de mayor rango tomen las medidas necesarias enfocadas a la remediación de todos aquellos defectos que sean relevantes para la organización.

3.1.2 Diseño de escenarios representativos (infraestructura crítica, sistemas militares, redes gubernamentales)

Un correcto diseño del escenario (figura 3-3) es crucial para dar coherencia operativa a una operación de Red Team, especialmente cuando se trata de mejorar la seguridad nacional. Un buen escenario trasciende más allá de un story line que introduzca el ejercicio. Debe contener el diseño que cubra los aspectos tecnológicos y pedagógicos, que convierta los objetivos estratégicos en condiciones reproducibles y medibles cuantitativamente, ya que el scoring es fundamental a la hora de evaluar la reacción de la organización puesta a prueba. Un punto de especial relevancia en el diseño es la confluencia de tres aspectos críticos: realismo, seguridad y trazabilidad. Realismo, para que emule con la mayor fiabilidad posible los sistemas implementados en el mundo real. Seguridad, con el foco puesto en evitar incidentes de seguridad real. Trazabilidad, en el sentido de la generación de evidencias digitales verificables. Para aunar estos tres factores, se propone un principio de diseño de escenario de ciberejercicio basado en los siguientes puntos:

1. Importancia por amenaza y activo: en una fase inicial de diseño, ha de realizarse un completo análisis desde el punto de vista de ciberinteligencia (que genere un producto similar al TTI en el caso del framework TIBER-EU), identificando posibles actores maliciosos y las TTP's que

- emplean en su operativa habitual. Así mismo, debe hacerse un análisis de los activos de la organización objetivo, de forma que se permitan levantar la superficie de exposición de este e identificar posibles vectores de entrada.
2. Objetivos cuantificables: el objetivo principal del ejercicio es la evaluación del nivel de madurez de la ciberseguridad de la organización auditada. Para ello, deberán definirse una serie de elementos medibles (como el tiempo de detección de una intrusión o el nivel de disponibilidad de servicios en caso de ciberataque), que permitan en base a criterios objetivos, determinar los puntos de débiles y trazar un plan de mejora.
3. Escalabilidad y modularidad: Una práctica muy recomendable es el diseño modular de los escenarios, que permita la ejecución por fases, pudiendo ser fácilmente escalable, añadiendo nuevos módulos al escenario. Por ejemplo, se puede dividir un ciberejercicio en fases que se asimilen a la Cyber Kill Chain, o por hitos que se ajusten a los parámetros que mejor se ajusten a los requisitos de la organización de la organización (i.e. acceso inicial, escalada de privilegios, persistencia, etc..).
4. Seguridad por diseño: Este principio permite que la seguridad operacional de la organización objetivo no se vea comprometida. Para ello se deben contemplar medidas que eviten el daño real, como el uso de datasets generados ad hoc para el ejercicio, ventanas de baja actividad entre otros.
5. Reproducibilidad: la reproducibilidad habilita a la organización a realizar de forma periódica el ejercicio, permitiendo la comparación con el paso del tiempo, y evaluando el nivel de mejora de la ciberseguridad y resiliencia. Así mismo, hace que el ejercicio sea transferible a otros organismos, adaptando a las necesidades y casuística de cada entidad gracias al modularidad.

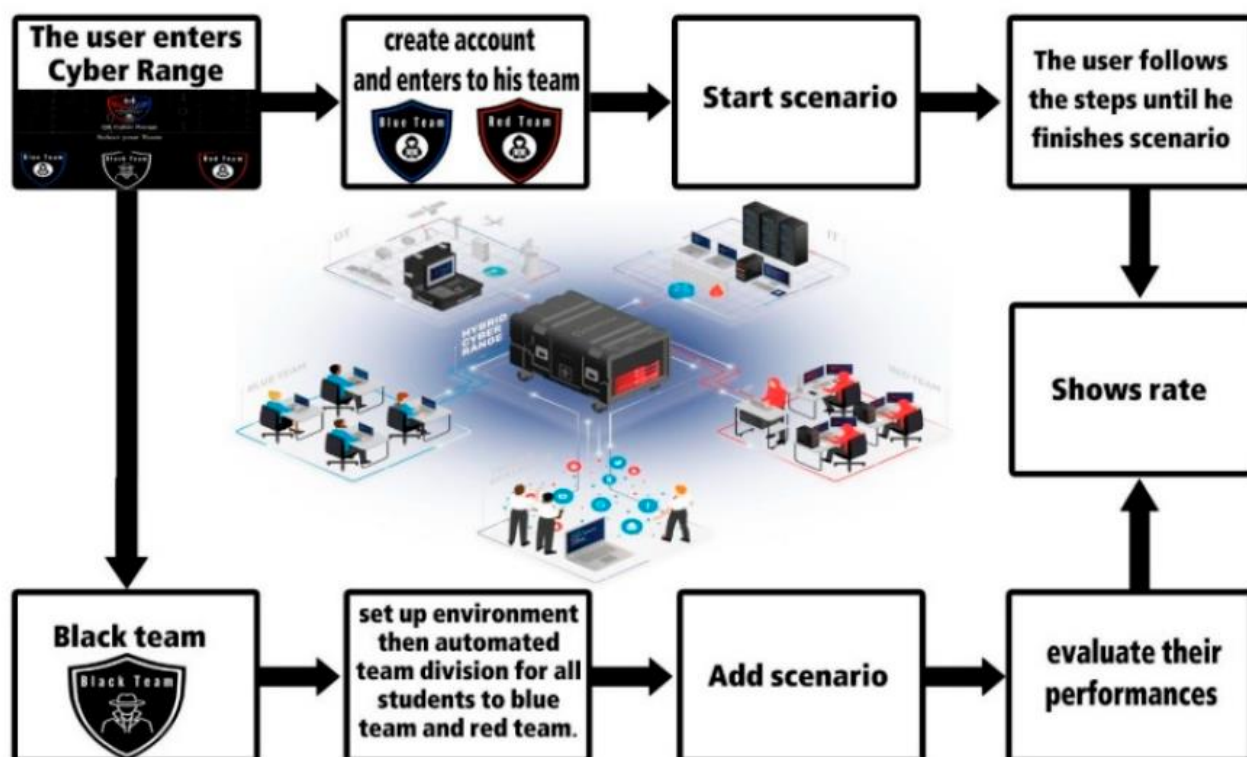


Figura 3-3 Flujo de trabajo de un ejercicio de Red Team [49].

Como última consideración, aquellos escenarios que afectan a infraestructuras críticas o redes y sistemas militares plantean riesgos regulatorios. Para salvaguardar a los equipos operativos que forman parte de los ejercicios, se considera necesario contar con la validación legal explícita, acuerdos que

definan responsabilidades, autorización del máximo nivel de las autoridades competentes, así como medidas de protección de datos. En el caso de que el ejercicio involucre organizaciones de diferentes organismos, es necesario la creación de un grupo de coordinación que habilite línea directa entre las capas directivas.

3.1.3 Técnicas, tácticas y procedimientos (TTPs) adaptados al contexto nacional español

Todo ciberejercicio que aporte valor operativo, debe desarrollarse en base a unas TTP's que a su vez deben cumplir el marco legal y organizativo del país en el que se realizan, ajustándose explícitamente al contexto nacional. La principal implicación de esto último es la selección y priorización de las TTP's para que reflejen amenazas reales sobre los organismos objeto de los ejercicios. Así mismo, se debe contar con la idiosincrasia propia de las infraestructuras nacionales, como por ejemplo los despliegues masivos de Directorios Activos en Administraciones Públicas, servicios en la nube pública gestionados por empresas foráneas, así como las particularidades desde el punto de vista regulatorio (e.g. NIS, Esquema Nacional de Seguridad, etc...).

Para que los ejercicios tengan una utilidad práctica, deben emular cadenas de ataque realistas muy similares a las observadas en el mundo real. Un buen ejemplo serían los ataques de spear phishing, donde en lugar de emplear una campaña masiva de envío de correos, la acción se limita a un número de usuarios muy limitado, resultado de un estudio exhaustivo de los trabajadores de la organización objetivo. Otro ejemplo claro es la técnica conocida como "living-off-the-land". Esta es una técnica que consiste en el uso de ejecutables y scripts que forman parte de la distribución legítima, y emplearlos para ejecutar binarios maliciosos (e.g. MSBuild). Con esto presente, una forma eficaz de presentar las TTP's a emplear en el ciberejercicio es su mapeo en MITRE ATT&CK. Además, la adaptación nacional exige combinar técnicas con escenarios de gobernanza (incidentes que requieren coordinación entre los tres CERTS de referencia, a saber, Ministerio de Defensa a través del MCCE, CCN-CERT, INCIBE-CERT) y respetar límites legales sobre emulación ofensiva. Un ejemplo claro es el no hostigamiento ni amenazas al personal miembro de la organización objetivo. Cada TTP autorizada debe ir acompañada de una justificación de su empleo, en base a la labor de investigación de ciberamenazas realizada en las fases previas al ejercicio, y que refleje su utilidad para la mejora de la detección y respuesta en el ecosistema español.

Para una correcta identificación de las TTP's de interés para la mejora del marco de seguridad nacional, la fase previa de ciberinteligencia es de vital importancia (en la figura 3-4 se muestra el flujo de conversión de ciberinteligencia en TTP's). Los hallazgos derivados de las labores de ciberinteligencia deben constituir la base para decidir las TTP's a emplear. Por otra parte, debe adoptarse una mentalidad proactiva, adelantándose a escenarios futuros. Es por ello, que la labor de recopilación de ciberinteligencia debe ir acompañado de un estudio de TTP's relacionadas en otros incidentes de ciberseguridad, y que puedan suponer una evolución de una acción maliciosa. Con las TTP's identificadas, se acometerá un mapeo de las mismas en la matriz de MITRE ATT&CK, así como la telemetría asociada a su detección (e.g. correlación en el SIEM de la actividad de una baliza de un C2 que establece comunicaciones con una dirección IP previamente relacionada en una lista negra). Esto facilitará la labor de scoring del ejercicio y permitirá establecer una comparativa entre el ejercicio y el caso de uso real.

No hay duda de que las posibles fuentes de información para identificar y adaptar TTP's empleadas por actores maliciosos con los requisitos de seguridad nacional son muchas y muy variadas. Este particular hace necesario establecer una ponderación de las fuentes, para asignar un valor alto a aquellas de mayor confianza. Entre las fuentes clásicas se hallan los informes de INCIBE/CCN, OSINT elaborado por equipos propios o mediante la subcontratación de terceros o feeds comerciales (informes de ciberinteligencia proporcionado por empresas como Recorded Future).

Una vez están identificadas las TTP's, hay que materializarlas en forma de campaña ofensiva. No hay que olvidar que la correcta realización del ciberejercicio depende en gran medida de la emulación del adversario. Solo será beneficioso para los intereses nacionales si el ejercicio es realista, y de él se pueden desprender lecciones aprendidas y elementos de mejora de los sistemas defensivos implementados. Por ello, es recomendable trazar un playbook con los ataques que se van a realizar, recogiendo indicadores de la respuesta iniciada ante cada uno de ellos. Esto facilitará enormemente la labor de scoring.

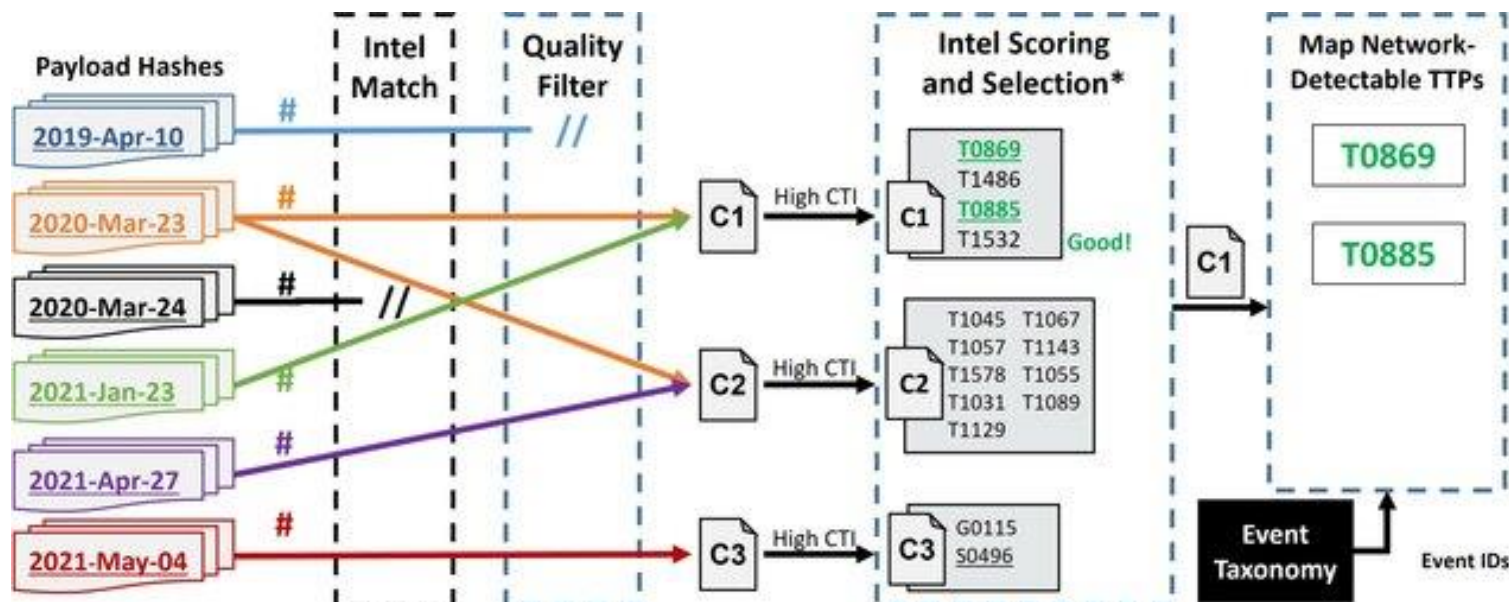


Figura 3-4 Diagrama de flujo de conversión de elementos de ciberinteligencia en TTP's [50].

3.1.4 Criterios de éxito y medición del impacto defensivo

El éxito del ciberejercicio radica en la capacidad de la organización de extraer lecciones accionables que ayuden a mejorar las capacidades de ciberseguridad. Por ello, la medición eficaz de los resultados en base al scoring es un elemento clave que ayudará a comprender los puntos de mejora. Por ello se proponen una serie de métricas que ayudarán en esta labor. Estas métricas permiten afirmar, con el respaldo de la evidencia digital, si los sistemas defensivos han mejorado, si los mecanismos de control funcionan y en caso de fallar, cuáles son las medidas de remediación y su prioridad. Algunas de las métricas de mayor interés son:

1. Tiempo de identificación de una amenaza: Una detección temprana y adecuada es esencial para la correcta mitigación de los riesgos. Esta métrica es una medida del tiempo entre el inicio de un ataque y el momento en el que el Blue Team o alguno de sus sistemas detectan el intento de intrusión. Esta métrica se puede subdividir en etapas, como el número de alertas generadas, cuántas de ellas han supuesto un falso positivo o cuánto tiempo demoró investigarla.
2. Tiempo de respuesta ante un ciberataque: Con el registro del tiempo de detección, ya se tiene una idea de la capacidad de detección de la organización. El siguiente paso es conocer el tiempo de respuesta (i.e. tiempo entre la detección y el inicio de las acciones de contención por parte del equipo de respuesta a incidentes). El principal objetivo es minimizar el impacto del ciberataque, evitando su expansión por la red y los sistemas. Esta métrica ayuda a dilucidar la eficacia de las capacidades de respuesta, y como lecciones accionables se extraen puntos de mejora de los procesos que minimizan el impacto de un ciberataque.
3. Tiempo de resolución de un incidente de ciberseguridad: Una vez producido el ataque, el tiempo medio de remediación (i.e. deshacer los efectos producidos por el ciberataque) es otra

métrica de gran valor. Algunas de las submétricas subyacentes son factores de colaboración y coordinación entre los diferentes equipos de la organización, comunicación y disponibilidad de los servicios afectados. La capacidad de un equipo para recuperarse eficazmente de un ciberataque está íntimamente relacionada con la continuidad de negocio (o de servicios públicos) y la resiliencia (factor este medido en el ejercicio Locked Shields). Esta métrica, tiene una clara relación con la preparación en materia de ciberseguridad del organismo auditado.

4. Eficacia de las políticas de seguridad: Con una imagen clara de las capacidades de detección y respuesta ante ciberataques tanto del Blue Team como de los sistemas defensivos (e.g. EDR y SIEM), las políticas de seguridad son otro factor clave a evaluar. Es necesario verificar si se trata de políticas ceñidas a la realidad y casan con la realidad operativa de la organización. La revisión de políticas de seguridad debe realizarse de forma periódica para evaluar si los miembros de la organización están en condiciones de cumplirlas. Al contar con mediciones precisas, puede identificar las áreas clave de mejora y asignar recursos para garantizar que en el futuro se disponga de políticas de seguridad más avanzadas y eficaces. Algunos ejemplos de políticas de seguridad son el mínimo privilegio de usuarios, políticas de contraseñas seguras, inhabilitación de los puertos USB de los terminales de trabajo, entre otras.
5. Coste estimado del incidente: No hay que perder de vista que el impacto económico es fundamental a la hora de evaluar un incidente de ciberseguridad. Esta métrica permite medir el éxito del ciberejercicio a escala organizativa y de coste. Este parámetro da cuenta del daño financiero potencial que un ciberataque puede causar a la organización. Este valor, además es un elemento de justificación para el presupuesto de ciberseguridad de la organización, y que permite evaluar el retorno de inversión del gasto en esta área. Los costes reales asociados a los ciberataques no son solo el coste del tiempo, las herramientas, los procedimientos de mitigación y recuperación o las posibles multas por incumplimiento de la normativa vigente, sino también la pérdida potencial de productividad y el daño reputacional.
6. Medida del factor humano: Esta métrica es una de las más difíciles de evaluar. Sin embargo, en un amplio porcentaje, el éxito de un ciberataque es debido a un fallo humano. Por norma general, las personas son el eslabón más débil en la cadena de seguridad. Esta métrica dependerá en gran medida de la propia organización. Aquellas con procesos automatizados tendrán un menor impacto de este factor. Un ejemplo de métrica asociada al factor humano sería el porcentaje medio de empleados que se exponen con relativa facilidad a los ataques de phishing. Esto incluye hacer click en enlaces no verificados o descargar archivos dañinos. Una medida derivada de un valor negativo en esta métrica sería, por ejemplo, la realización de jornadas de concienciación de ciberseguridad.

Para poder vincular estas métricas a criterios de éxito es necesario definir cada uno de estos criterios desde una perspectiva operacional. Estos criterios deben ser objetivos, materializándolo en medidas concretas (e.g. IOC's verificados, medidas de tiempo medio, etc...). Además, será necesario dar una descripción de cómo se calculan a partir de datos normalizados. Hay otro factor para tener en cuenta: el peso específico que cada métrica tiene en el valor final del resultado del ejercicio. Dicho de otra forma, un valor corrector que relacione importancia del hallazgo con su criticidad dentro del organismo objeto del ejercicio. Un resultado no tiene el mismo impacto en función del sistema afectado. Por ejemplo, una detección tardía de un movimiento lateral no tiene la misma gravedad si se produce entre dos hosts con bajos privilegios que si se hace contra un con permisos de administración. Estos pesos específicos deben ser calculados durante la preparación del ejercicio, teniendo en cuenta múltiples factores, entre los que se encuentran el nivel de criticidad de la organización involucrada, nivel de madurez de sus elementos de ciberseguridad, grado de disponibilidad de sus servicios o volumen de personal del organismo.

Los criterios de éxito representados en el scoring final, serán el punto de partida para el desarrollo de nuevos ciberejercicios y cuyo fin será la mejora de los anteriores. Los criterios que no se cumplan serán los primeros puntos de mejora para la organización. A partir de ahí, y cuando se hayan solventado los problemas detectados durante el ejercicio, comienza el planeamiento del siguiente. Este

tipo de actividad debe realizarse de forma periódica, para poder así medir el nivel de evolución de las soluciones de seguridad implementadas por la organización. Los resultados de pruebas anteriores han de conservarse para su posterior comparación con los nuevos resultados. Además, puede servir como elemento comparador entre organizaciones que provean servicios similares, como pueden ser los servicios autonómicos. También, es recomendable que las lecciones aprendidas tras cada ciberejercicio se difundan entre entidades similares (en especial las autonómicas y locales), para contribuir de forma activa a la defensa colaborativa de las instituciones públicas, algo que redundará en la mejora de la seguridad nacional. A continuación, se propone un modelo de ponderación para ciberejercicios:

$$\text{Peso Total } (W) = (C \times 0.4) + (M \times 0.25) + (D \times 0.2) + (T \times 0.15)$$

1. C representa la criticidad de la organización y se le asigna un valor del 40%
2. M representa el grado de madurez de la organización y se le ha asignado un valor del 25%
3. D representa el grado de disponibilidad de los servicios y se le asigna un valor del 20%
4. T representa el volumen de trabajadores de la empresa, y se pondera con un 15%

Cada uno de los criterios representados tendrá un valor de 1 a 5, proponiéndose las siguientes tablas

3.1.4.1 Criticidad

Mide el nivel de importancia de la organización objetivo en la seguridad nacional.

Valor	Descripción	Ejemplo
5	Infraestructura crítica nacional, defensa, comunicaciones estratégicas o sector energético	Ministerio de Defensa, Red Eléctrica, RENFE
4	Servicios públicos esenciales o sector financiero	Agencia Tributaria, Banco de España
3	Entidades públicas de apoyo u operadores TIC relevantes	Ayuntamientos, universidades
2	PYMEs con datos sensibles, pero bajo impacto sistémico	Proveedores TIC de tamaño medio
1	Organizaciones sin funciones críticas nacionales	Centros formativos

Tabla 3-1 Valores asignados en función de la criticidad de la organización.

3.1.4.2 Nivel de madurez de la ciberseguridad

Evalúa la implantación de políticas, controles y capacidades defensivas. La siguiente tabla se basa en la NIST CSF Maturity Levels, guía que ayuda a evaluar la eficacia con la que se integra la ciberseguridad en una organización, desde las respuestas reactivas (nivel 1) hasta los procesos totalmente integrados y adaptativos (nivel 4) (<https://drata.com/blog/nist-maturity-model>).

Valor	Nivel	Características
5	Optimizado	SOC 24/7, SOAR, respuesta automatizada, auditorías y Red/Purple Team regulares
4	Gestionado	SIEM centralizado, políticas formalizadas, gestión de incidentes madura
3	Definido	Políticas establecidas, detección y respuesta manuales
2	Inicial	Controles parciales, sin correlación centralizada
1	Ad hoc	Ausencia de políticas o controles técnicos

Tabla 3-2 Valores asignados en función del nivel de madurez de la ciberseguridad de la organización.

3.1.4.3 Grado de disponibilidad de los servicios

Representa la necesidad de continuidad operativa dentro del marco de seguridad nacional. Cuanto más crítico sea el tiempo de inactividad permitido debido a un incidente de ciberseguridad, mayor será el peso específico.

Valor	Descripción	Ejemplo
5	Alta disponibilidad 24/7, interrupción inaceptable	Centros de control, sistemas SCADA, telecomunicaciones
4	Servicios con ventanas limitadas	Servicios financieros, plataformas de transporte
3	Continuidad diurna o en horario laboral	Administración pública general
2	Operación no continua	Oficinas, educación
1	Baja dependencia del servicio	Áreas de soporte o formación interna

Tabla 3-3 Valores asignados en función del grado de disponibilidad de los servicios.

3.1.4.4 Volumen de empleados

Criterio cuyo valor se asigna en función del número de empleados, que se materializa en el número de usuarios. Por lo tanto, cuanto mayor sea este valor, mayor será la superficie de exposición de la organización y mayor será la probabilidad de ataque

Valor	Rango de empleados	Tipo de organización
5	> 10.000	Grandes organismos del Estado o industria nacional
4	5.000–10.000	Ministerios, universidades grandes
3	1.000–5.000	Agencias o PYMEs grandes
2	250–999	Entidades medianas
1	< 250	PYMEs o startups

Tabla 3-4 Valores asignados en función del número de empleados.

3.2 Planificación y ejecución de ciberejercicios en entornos autorizados

La planificación de un ciberejercicio es un proceso complejo, que debe articularse documentalmente en base a unas especificaciones del ejercicio (EXSPEC por su acrónimo en inglés) incluso aunque se trate de un contrato técnico entre ambas partes. Como norma general, y con el enfoque planteado, las EXSPEC deben ser un compendio de documentación requerida para la ejecución del ejercicio. Punto clave como las ROE's, el informe generado durante la fase de ciberinteligencia (homólogo al TTI en el marco TIBER-EU), los criterios de scoring, así como las métricas requeridas. Este conjunto de documentos supone la base legal y técnica que auspicia toda acción ofensiva y plantea las líneas rojas que no se deben traspasar (cabe destacar los posibles datos personales de trabajadores hallados durante la intrusión o las acciones fuera de las ventanas temporales establecidas). Se considera una buena práctica el normalizar el formato de la documentación del ejercicio. Esto facilitar la reproducibilidad a lo largo del tiempo, permitiendo también que se compartan con otros organismos, redundando esto es un ahorro en los tiempos de preparación. También es necesario recalcar que los responsables ejecutivos y legales de la entidad auditada deben mostrar su conformidad a la ejecución del ejercicio mediante la firma del NDA.

El entorno en el que va a realizar el ciberejercicio es de una importancia crítica. Para que sea de provecho, es necesario que exista un equilibrio entre realismo y seguridad operativa. Los ejercicios en entornos OTAN como Locked Shields y Crossed Swords, se emplean entornos simulados muy realistas que permiten el desarrollo del ejercicio sin generar un impacto real en los sistemas en producción. Aun así, antes de la ejecución es necesario garantizar la operatividad del escenario generando copias de seguridad de cada uno de los elementos del escenario. Desde el punto de vista técnico, la instrumentación debe desplegarse y validarse mediante un test-run, que en condiciones similares a las que se va a desarrollar el ejercicio permita verificar que todo está listo para el go.

La ejecución del ejercicio exige una separación de funciones entre los distintos equipos involucrados, así como un equipo de supervisión y otro de soporte. El control del ejercicio debe llevarse de forma independiente, con una autoridad responsable con nivel suficiente como para parar el ejercicio en caso de necesidad. La dinámica del ejercicio debe estar marcada por un plan preestablecido, en el que el Red Team tenga un flujo de trabajo claro, que permita generar las evidencias necesarias para su posterior evaluación, sin olvidar el mapeo sobre la matriz de MITRE. Para aportar realismo al ejercicio y acercarlo al mundo real, es recomendable introducir injects (píldora de información que permiten contextualizar una nueva acción) de contexto (mediante la simulación de publicación de noticias en prensa, llamadas de responsables directivos o peticiones administrativas). Estos injects permiten evaluar la capa no técnica de respuesta y coordinación, algo muy practicado en Locked Shields.

3.2.1 *Requerimientos técnicos y operativos*

En la fase de planeamiento hay que tener en cuenta diversas consideraciones. Estas engloban principalmente a los requisitos técnicos que se necesitan para la preparación del ejercicio y para su ejecución. Desde el punto de vista de la seguridad de las entidades implicadas en el ejercicio, es muy recomendable el uso de máquinas virtuales con una distribución ofensiva (e.g. Kali Linux o Parrot), debiendo, tras la finalización del ejercicio, eliminar las máquinas virtuales empleadas, evitando su reutilización. A pesar de que la máquina se limpie, siempre es posible que quede información en ella, lo que puede suponer un riesgo de seguridad para la entidad auditada. El ejemplo más claro lo supone el hecho de que, por ejemplo, no se borre por un descuido un fichero con credenciales exfiltradas del objetivo. Si este tipo de información sensible se distribuyese por error, supondría un serio daño reputacional.

Trabajar con un entorno virtualizado permite al personal ofensivo levantar, probar, revertir y clonar diferentes máquinas virtuales de una forma muy rápida. Además, facilita la creación de un laboratorio de pruebas. Este paso es fundamental para probar nuevas herramientas en un entorno de pruebas, con el fin de asegurarse que se comprende su funcionamiento y cuáles son sus posibles impactos en una máquina objetivo. Este tipo de pruebas y análisis en laboratorio es especialmente vital para herramientas gratuitas descargadas de internet, debido a la preocupación por la calidad, el riesgo de provocar un fallo en el sistema objetivo y funcionalidades ocultas que podrían comprometer los sistemas de prueba. Es importante contar con un buen laboratorio que incluya muchos de los mismos sistemas y software que se encontrarán en el objetivo. Obviamente, el laboratorio no puede contener todas las combinaciones posibles ni todo el software, pero una muestra representativa es suficiente para nuestros propósitos.

Es importante disponer de sistemas que se asemejen a una amplia variedad de redes objetivo. La mayoría de las redes utilizan un Dominio de Windows y, por lo tanto, un Controlador de Dominio. De manera similar, la mayoría de estas redes incluyen servidores de archivos Windows y servidores web Windows. Se pueden combinar estos dos roles en un solo sistema para ahorrar recursos físicos y licencias, en aras de una optimización de recursos. Por supuesto, el sistema de usuario final más utilizado en las instituciones públicas es Windows, así que es muy recomendable uno de esos sistemas en el entorno de laboratorio.

No hay que olvidar los sistemas Linux, que, aunque en menor medida, también son empleados en administraciones públicas, y que presentan la ventaja de ser gratuitos. Por ejemplo, Ubuntu es una distribución Linux común y gratuita que se puede encontrar en diferentes organismos públicos. Aunque la mayoría de las instalaciones y actualizaciones en Linux son gratuitas, algunas no lo son. Por ejemplo, las actualizaciones de Red Hat no son gratuitas, pero es muy similar a Fedora.

Como se ha comentado anteriormente, es altamente recomendable utilizar sistemas dedicados para cada ejercicio y que sean únicos para cada ejercicio. Además, para un correcto desarrollo del ejercicio, los sistemas propios, hay que tener en cuenta una serie de consideraciones:

1. **Defensas deshabilitadas:** Los sistemas de ataque no tendrán software defensivo como AV/EDR, y a menudo tendrán el firewall deshabilitado. Estas modificaciones disminuyen la capacidad defensiva del sistema en caso de que sea atacado por un adversario. Evidentemente, esta operativa no es aconsejable para la operativa del equipo de trabajo diario (donde se redacte el informe, por ejemplo).
2. **Tiempo de actividad:** Los sistemas de prueba a menudo necesitarán estar disponibles y conectados a internet durante semanas o meses. Si ejecutas el ataque desde un portátil (o incluso una VM corriendo en un equipo portátil), la prueba podría verse afectada si traslada el equipo a otro lugar (cambio de red) o se realizan tus tareas diarias (instalar software, reiniciar, aplicar parches).

3. Prevención de fuga de datos: empleando sistemas únicos para cada ejercicio, es menos probable que los datos de un ejercicio terminen mezclándose con los de otro. Esto es especialmente importante cuando en el ejercicio se contratan empresas externas, donde sí se contienen datos de pentesting de otras empresas, pudiendo considerarse una violación de la confidencialidad. Esta práctica podría derivar en una contaminación cruzada con la posibilidad de que derive en situaciones comprometidas, como pérdida de confianza e incluso acciones legales.

Existe otra consideración a tener en cuenta. Cuando se realiza el ejercicio, toda la infraestructura ofensiva puede ser levantada on-premise o en la nube. Ambas opciones presentan sus ventajas e inconvenientes, sin embargo, por seguridad operativa la opción cloud es la más recomendable. Exponer una infraestructura propia, podría generar problemas con el proveedor de servicios de Internet, ya que alguna de las actividades realizada podría interpretarse como maliciosas. Por otra parte, también existe la posibilidad de que se incluya la Ip atacante en listas negras, dificultando acciones futuras. Es por ello, que la infraestructura en la nube para ejercicios de Red Team es la más recomendable. El levantamiento de la infraestructura en la nube es fácilmente automatizable, permitiendo el uso de plantillas genéricas que se adaptarían a cada acción en concreto. Además, presenta la ventaja de que, si la Ip del VPS empleado en la infraestructura como redirector es incluida en alguna lista negra, o bloqueada durante la acción por parte del Blue Team, se solucionaría levantando un nuevo VPS (si bien esto supondría modificaciones más profundas en la propia infraestructura o herramientas, como la modificación de algún artefacto).

Un punto de gran importancia y que no puede perderse de vista es la seguridad de la propia infraestructura. Al estar expuesta a Internet, está expuesta a posibles atacantes. El compromiso de la infraestructura propia por parte de terceros supone un gran riesgo tanto para el Red Team como para la organización participante en el ejercicio. Para minimizar el riesgo de que esto ocurra es necesario bastionar la infraestructura ofensiva. El Centro Criptológico Nacional (CCN) promulga una serie de guías que definen cómo configurar de forma segura diferentes tecnologías, sistemas operativos, componentes de red, aplicaciones y servicios. Estas guías, entre la que se encuentra la STIC-600 (configuración segura, bastionado o hardening en inglés) son la que se emplean en la securización de las administraciones públicas. En base a las guías STIC se proponen las siguientes medidas de bastionado para garantizar el acceso no deseado a la infraestructura.

Medida de Bastionado	Acción Concreta para Redirectores	Justificación STIC-600 / ENS
MFA fuerte	Habilitar MFA en el acceso al proveedor cloud y consola de gestión.	Robustez del control de acceso (STIC-600-1, ENS OP.ACC-01).
Principio de mínimo privilegio	Crear rol <i>redirector-admin</i> con permisos exclusivos. No usar cuentas root del cloud.	Minimización de permisos – STIC-600.
Acceso just-in-time	Activar cuentas y accesos solo durante el ejercicio	Exposición mínima temporal.
Restricción de IPs de administración	Limitar SSH a IPs de operador/VPN mediante SGs/NACLs.	Control de acceso ENS.
Bloqueo de login root	Deshabilitar PermitRootLogin en SSH. Acceso solo mediante usuario de mínimo privilegio	STIC-610 (Linux).
Acceso mediante VPN	Uso de VPN para la administración del VPS.	Comunicaciones seguras ENS.

Tabla 3-5 Medidas recomendadas en el aspecto de identidad y accesos según las guías STIC.

Así mismo, uno de los aspectos más críticos es la gestión de credenciales. En este aspecto las guías STIC también ofrecen recomendaciones para su correcta implementación y uso, reflejándose algunas de ellas en la siguiente tabla.

Medida de Bastionado	Acción Concreta para Redirectores	Justificación STIC-600 / ENS
Uso de Secret Manager del cloud	Guardar claves, tokens y certificados en AWS SM / Azure KV / GCP SM.	Protección de secretos ENS OP.ACC-03.
Evitar claves persistentes	Usar claves SSH efímeras o rotarlas automáticamente.	Reducción de impacto en fugas.
Claves SSH fuertes	RSA 4096 o ED25519. Prohibido login por contraseña.	Bastionado STIC-600.
Rotación periódica	Rotación programada de claves SSH y certificados del redirector.	Mejora continua ENS.
No almacenar claves en el VPS	No dejar claves en <i>home</i> , scripts, contenedores o temporales.	Restricción crítica ante compromiso.
Control y protección de logs	Cifrar o centralizar logs SSH y de auditoría.	Trazabilidad ENS.

Tabla 3-6 Medidas recomendadas en el aspecto de gestión de credenciales y accesos según las guías STIC.

Otro aspecto crítico es la gestión de los puertos y servicios, ya que son estos elementos expuestos, que conforman gran parte de la superficie de exposición. Las guías STIC también contemplan su bastionado y seguridad. En la siguiente tabla se muestran algunas de las medidas recopiladas en dichas guías.

Medida Crítica	Acción Concreta en Redirectores	Justificación STIC-600 / ENS
Política “deny-all” en firewall	Configurar iptables/nftables/SG bloqueando todo por defecto; solo abrir los puertos estrictamente necesarios (comunicación con el C2).	Reducción drástica de la superficie expuesta
Exposición mínima de servicios	Mantener únicamente el servicio de redirección. Restringir SSH únicamente a VPN o IPs en lista blanca.	Limita vectores de ataque — STIC-600.
Filtrado granular de puertos (ingress)	Permitir solo tráfico entrante al puerto del redirector. Bloquear puertos comunes asociados a fingerprinting (22, 8000, 8080...).	Control de acceso estricto ENS.
Filtrado granular de puertos (egress)	Permitir únicamente las conexiones necesarias hacia el C2 interno; bloquear DNS externos, SMTP y cualquier salida no autorizada.	Previene exfiltración y canales laterales.
Cifrado y protocolos seguros	Forzar TLS cuando aplique; deshabilitar SSLv2/3 y TLS 1.0/1.1. Configurar cifrados robustos.	Cumplimiento cifrado ENS OP.CIF-02.
Protección frente a fingerprinting	Ocultar versiones (server_tokens off), quitar banners, limitar headers y activar rate-limit básico.	Reduce la probabilidad de detección y clasificación hostil.

Tabla 3-7 Medidas recomendadas en el aspecto de gestión de puertos y servicios según las guías STIC.

En la tabla 3-5 quedan reflejadas una serie de medidas de bastionado con el objetivo de que sólo los operadores autorizados y participantes en el ejercicio tengan acceso a la infraestructura levantada para la acción ofensiva. Algunos de los puntos fuertes son el principio de mínimo privilegio (que permite que un acceso no autorizado no tenga, a priori y de inicio, acceso como administrador) o un múltiple factor de autenticación que añade robustez al acceso a la infraestructura. Por su parte, la tabla 3-6 lidia con uno de los aspectos más críticos de cualquier infraestructura expuesta a Internet: la gestión de claves. Como medida de buenas prácticas no deben almacenarse credenciales en el propio VPS y además deben emplearse claves criptográficas robustas. Por último, la tabla 3-7 busca minimizar la superficie de ataque de los redirectores. Como su propio nombre indica, su función es la de redirigir el tráfico de red durante la operación. Por ello, no debe ser accesible ningún puerto que no se vaya a usar durante la operación. Además, los servicios empleados deben estar configurados en puertos no convencionales, para así dificultar una posible intrusión. Con las medidas reflejadas en las tres últimas tablas se busca que la infraestructura ofensiva desplegada sea agresiva hacia fuera, pero extremadamente defensiva hacia adentro.

Los redirectores (infraestructura de redirectores mostrada en figura 3-5) son una pieza fundamental en la operativa de una operación de Red Team, ya que desligan la ubicación física del C2 de la telemetría registrada en el objetivo, reduciendo la exposición de la estructura de mando y control. En términos de OPSEC esta aumenta considerablemente al emplear una metodología que involucre a redirectores. En este aspecto es necesario tener en cuenta una última recomendación. El uso de redirectores permite el uso de infraestructura “desechable”. Puesto que esta infraestructura va a ser de un solo uso, es muy recomendable facilitar la IP del último salto (i.e. el último VPS, que será el que interactúe con el objetivo) a la organización auditada. Puede darse el caso de que durante la realización del ejercicio la entidad sufra un ataque real, por lo que tener conocimiento del origen de la actividad emulada es de vital importancia para discernir entre actividad maliciosa o simulada.

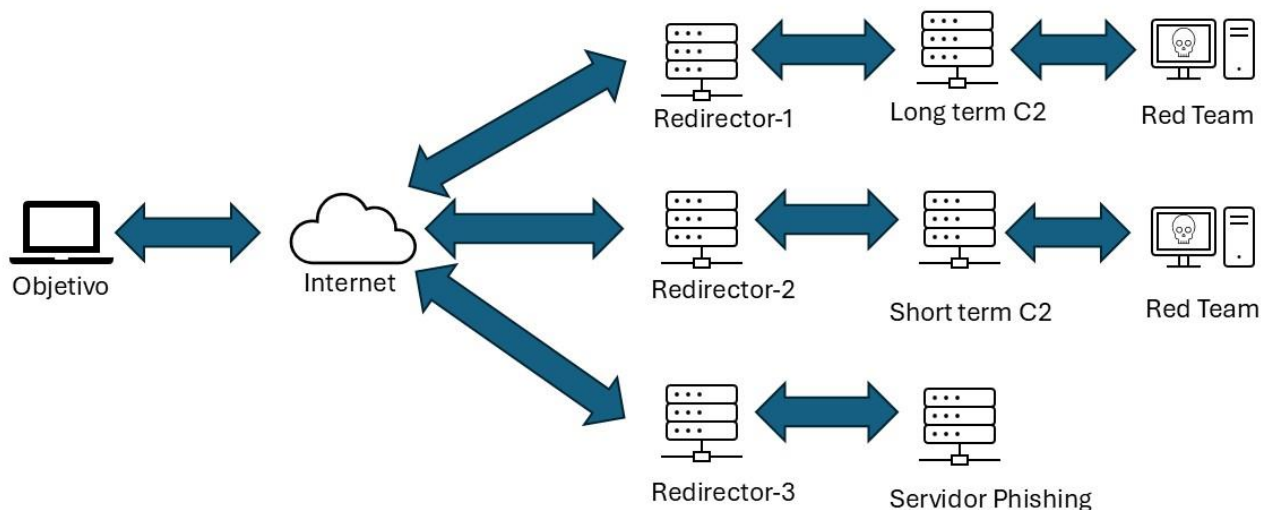


Figura 3-5 Diseño de infraestructura ofensiva basada en la nube para ejercicios de Red Team.

3.2.2 Ejecución de escenarios seleccionados, recolección y análisis de datos durante los ejercicios

La ejecución de los escenarios seleccionados comprende la secuencia operativa que materializa la ciberamenaza en una serie de acciones realizadas en un entorno controlado sobre objetivos concretos, previamente delimitados en el alcance del ejercicio. Así mismo, y dado que la finalidad del mismo es la evaluación de la ciberseguridad del objetivo, la recolección y posterior análisis de los datos recabados es otro de los pilares fundamentales que se producen durante la ejecución del ejercicio. Es necesario dar un enfoque integral a estas dos vertientes, de modo que la planificación, metodología y criterios de documentación se diseñen en conjunto facilitando la trazabilidad, reproducibilidad y máximo aprovechamiento de los resultados obtenidos.

El escenario del ejercicio debe recoger con precisión la estructura del objetivo, la hipótesis en la que se sustenta la amenaza (principalmente la colección de TTP's derivadas de la fase de recolección de inteligencia), los criterios de éxito y los sistemas que bajo ningún concepto deben ser alcanzados (algo que como se ha dicho quedará perfectamente reflejado en el alcance). Un escenario

bien estructurado facilita la identificación de actividad anómala y por lo tanto ayuda a recopilar la telemetría derivada de las diferentes acciones ofensivas que se lleven a cabo, y que permitirán juzgar los resultados. Es muy recomendable que cada escenario se valide en el test-run, verificando que las técnicas que se van a emplear son las adecuadas, los sistemas defensivos están operativos y funcionando correctamente y que la ventana de ejecución está clara por todas las partes.

Bajo una perspectiva operativa, la ejecución de cada escenario consta de un número de fases bien delimitadas que facilitan el control y el registro de evidencias (la figura 3-6 da cuenta de un organigrama para la ejecución de un ciberejercicio ofensivo). Las fases, pueden ajustarse a cada escenario en concreto, ajustándose a la cada casuística. Por ejemplo, se pueden plantear las fases de preparación (en la que se realizaría la fase de recolección de ciberinteligencia y el reconocimiento de la superficie de exposición del objetivo), el acceso inicial (en el caso de la infraestructura reflejada en la figura 3-5 podría plantearse mediante un phishing, ya que se contempla un servidor dedicado a esos efectos), escalada de privilegios, movimientos laterales, acciones sobre el objetivo (como exfiltración de documentos o despliegue de un ransomware, si bien este último se aconseja sea una acción simulada), cierre y limpieza (que incluirá la destrucción de la infraestructura creada para el ejercicio). Contar con las fases bien delimitadas ofrece la posibilidad de introducir hitos en cada una de ellas, que mejoran el control de la ejecución y el control de los resultados obtenidos.

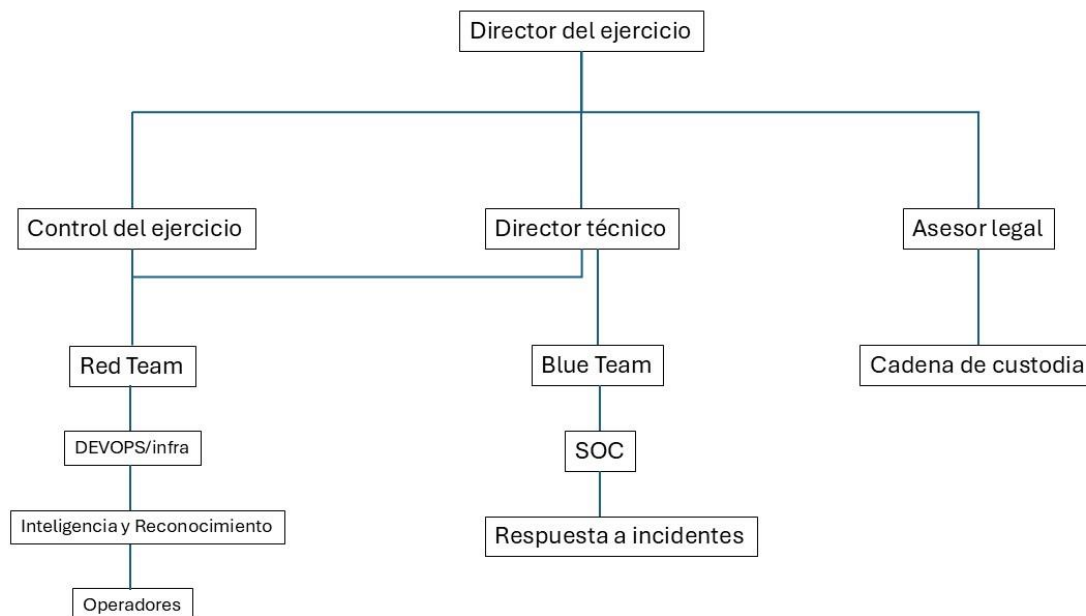
La documentación de los resultados tiene como derivada la generación de productos accesorios. Nuevamente, estos productos dependerán en gran medida del escenario y de los objetivos del ejercicio. Si el acceso inicial se consigue mediante la detonación de un beacon (i.e. un agente que se implanta en memoria en un sistema comprometido y permite al operador ofensivo ejecutar acciones y comunicarse con el C2) de un C2 y esta es detectada por el SIEM, se generará una evidencia forense, así como una reacción para la contención de la posible infección. Todo esto se materializa en un informe que incorporará la trazabilidad completa sustentada en las métricas que han permitido la identificación del compromiso inicial. Esto conllevará otros subproductos, como un extracto de la captura de tráfico acotado al margen horario de la acción, logs asociados, eventos del EDR entre otros. Así mismo, la modelización de la amenaza es otro subproducto generado, y que será de gran ayuda en la reproducibilidad del ejercicio. Las recomendaciones y el plan de retest con sus criterios de aceptación son otros de los productos generados.

Es aspecto de la recolección y el análisis de datos será la piedra angular del ciberejercicio. Para que la calidad de la información sea la adecuada, deben exigirse unos protocolos técnicos y operativos estrictos. Una correcta forma de proceder comprende la división de la telemetría por capas complementarias (e.g. a nivel host, nivel de red o de aplicación) donde cada sonda o elemento que recolecte datos debe haber sido verificado previamente, bien en el test-run o mediante un protocolo de verificación previo. El timestamp (o la marca temporal de registro de las evidencias) debe estar referido a un patrón común, para que la comparación entre acción detección sea lo más precisa posible. Esto se puede conseguir mediante la sincronización vía NTP. Junto con el timestamp deben registrarse otros elementos identificadores asociados a la métrica, como el identificador del elemento que recoge la evidencia, el host, el identificador del ataque, identificador del operador entre otras.

El análisis de los datos debe realizarse con el máximo rigor, garantizando la utilidad del ejercicio. Este proceso debe ser reproducible y auditable. Cada evento registrado puede ser reconstruido mediante la correlación de diferentes fuentes, como son el EDR (Endpoint Defense and Response), el SIEM (Security Incident Event Manager), capturas de tráfico de red, registros en el DNS o cualquier otro elemento de seguridad. Esta correlación del evento permitirá a los equipos defensivos determinar qué técnicas fueron empleadas, cómo fue detectado el incidente de seguridad y en base a qué elementos.

Por último, tendrá lugar el proceso de verificación y cierre, donde se certifica la integridad de las evidencias generadas (mediante la creación de su correspondiente hash), la validación de que el ejercicio se ha desarrollado conforme a las reglas de enfrentamiento y se ha ceñido al alcance determinado en la fase de preparación. En el caso de que detectasen deficiencias graves que requiriesen de una remediación inmediata, deberán fijarse los criterios de aceptación y proceder a la planificación de un retest. En caso de que el ciberejercicio se realizase en forma modular, el retest podría limitarse exclusivamente a una serie de pruebas enfocadas a la verificación de la remediación de la deficiencia detectada. Finalmente, por razones éticas y legales, todo proceso de recolección y análisis debe contemplar la minimización de datos personales (anonimizando aquellos que hayan sido detectados) y canales claros para la notificación a autoridades competentes (INCIBE-CERT, CCN-CERT) en los supuestos previstos por la normativa aplicable.

Esta aproximación aborda el ejercicio con un enfoque integrado que asegura el procedimiento según el método científico, generando evidencias verificables basadas en métricas consensuadas y productos finales que permiten una mejora sostenida en el tiempo y un upgrade de la resiliencia en materia defensiva, yendo más allá de la simple prueba técnica.



Soportes transversales: Comunicación, enlace de prensa, logística, soporte IT

Figura 3-6 Organigrama de un ejercicio Red Team.

3.2.3 Lecciones aprendidas y ciclo de mejora (validación, retest y transferencia)

Las lecciones aprendidas son el último producto obtenido en un ciberejercicio de Red Team y suponen el principal elemento sobre el que se articulará la mejora sostenida de las capacidades de defensa y resiliencia de las organizaciones participantes. Transforman el ejercicio en la cadena de transmisión de la experiencia técnica, en mejoras operacionales y normativas. Las lecciones aprendidas deben extenderse a todo el ciclo del ejercicio, no limitándose exclusivamente a las medidas de mejora en el campo defensivo. Por ejemplo, durante la fase de inteligencia, la aplicación de nuevas técnicas que mejoren el proceso de reconocimiento debe ser reflejadas para su aplicación en ejercicios futuros (e.g. Google dorks que hayan sido identificados como elementos valiosos durante el reconocimiento).

El proceso de creación de las lecciones aprendidas implica que todo hallazgo o campo de mejora debe materializarse en una serie de medidas correctivas que puedan ser implementadas en unos plazos razonables, para la mejora de la calidad de la defensa y la respuesta a incidentes de los organismos implicados. Cada lección aprendida debe incluir, como mínimo, la descripción del hallazgo, la evidencia que la motiva (referenciado al id que la identifica unívocamente), el trazado en la matriz de MITRE, una estimación del impacto operacional, recomendaciones técnicas de remediación con un responsable o responsables de la misma, así como un plazo para su resolución. El resultado de la aplicación de las lecciones aprendidas debe ser trazable para que su remediación pueda ser verificable, pudiendo remontarse al artefacto o técnica que generó la detección. Por ejemplo, un acceso inicial por el uso de credenciales por defecto de una aplicación generará una evidencia mediante el acceso a la propia aplicación. La lección identificada es el cambio de las credenciales por defecto. En este caso, la verificación de la remediación es el cambio de la contraseña a una que cumpla con la una política robusta de credenciales.

La validación técnica de las lecciones aprendidas engloba dos acciones fundamentales y que a su vez se complementan. En primer lugar, la revisión técnica que verifica elementos como la integridad de la evidencia que genera la acción de remediación, su congruencia temporal entre acción ejecutada y telemetría de detección y su hash. Por otra parte, una verificación que garantice el cumplimiento de las ROE durante el ejercicio y que la evidencia se generó dentro de los parámetros establecidos para el ejercicio. Solo cuando se haya validado la evidencia, se procederá a las acciones de mitigación, y se planificará un nuevo ejercicio (que podrá ser de menor entidad y podrá ceñirse a los sistemas afectados). En este punto deberán fijarse los criterios de aceptación que darán por buena la remediación.

El retest es una prueba, generalmente acotada, cuyo propósito es comprobar que las remediaciones cumplen los criterios de aceptación previamente acordados. Buscando la eficiencia operacional y presupuestaria, el retest se diseñará como una réplica a pequeña escala del ejercicio limitada a las acciones de remediación en los sistemas afectados, con una ventana temporal más ajustada. No se trata de una nueva prueba, sino de una réplica de la acción que derivó en el hallazgo que motiva la evidencia. Por lo tanto, debe ajustarse a los mismos parámetros y TTP's que en el ejercicio inicial. En el caso de que es retest fallase, es decir, de nuevo se generase una evidencia de éxito de un ataque, la remediación deberá priorizarse, con una nueva asignación de responsables para su mitigación con la vista puesta en un nuevo retest.

Con el enfoque puesto en la mejora del marco de Seguridad Nacional, y con la motivación de que el conocimiento técnico adquirido por una institución se convierta en una capacidad operacional permanente, debe producirse una transferencia operacional entre organismos. Esta transferencia se producirá en dos frentes. En primer lugar, el técnico representado por IOC's, firmas, reglas de SIEM, playbooks de respuesta y cualquier otra capacidad técnica que suponga una mejora defensiva y de la

resiliencia. En segundo lugar, la transferencia de conocimiento organizativo, materializado en procedimientos de detección y respuesta. La transferencia operacional debe incluir aquellos artefactos como scripts de detección, expresiones de correlación o ejemplos de telemetría que permitan validar un entorno en preproducción.

Finalmente, todo ciclo de mejora eficiente incorpora actividades de aprendizaje institucional. Se trabajarán sinergias en el plano de la ciberinteligencia, sobre las que se modelizarán las TTP's emuladas. También se compartirán los playbooks de respuesta a incidentes generados por cada organización. Este proceder favorece la defensa cooperativa y redundante en una mejora de las capacidades defensivas a nivel nacional. Por otra parte, la formación continua es imprescindible para mantener el conocimiento del personal actualizado. La realización de formaciones conjuntas permitirá la optimización de los recursos económicos destinados a este apartado. Esta colaboración también lleva a la identificación de barreras comunes, como son las limitaciones presupuestarias o la reticencia al cambio, así como medidas de mitigación. En este aspecto, son elementos comunes la priorización basada en la criticidad de los sistemas, las métricas de impacto operativo o la comunicación a nivel ejecutivo centrado en beneficios medibles. El cierre formal de cada ejercicio incluye la programación de la próxima iteración de evaluación, garantizando así la continuidad de la mejora defensiva.

3.3 Propuesta de marco para la Defensa Nacional

A lo largo de esta sección se va a proponer un marco integrador con el objetivo de incorporar de forma sistemática ejercicios de Red Team en la arquitectura de la Seguridad Nacional. Para ello, se partirá de la base de que la resiliencia nacional ante ciberataques está íntimamente ligada tanto a las capacidades técnicas como a marcos organizativos y jurídicos sólidos. Se va a plantear un marco cuyos objetivos son articular la gobernanza, dar criterios metodológicos, requerimientos operativos y aquellos elementos que ligan la cooperación público-privada, para lograr el máximo aprovechamiento de ciberejercicios ofensivos con un carácter periódico y escalable. Con esto se busca consolidar un ciclo de mejora continua con resultados medibles en la capa de detección, en la de contención y en la recuperación de servicios críticos a escala nacional ante una crisis motivada por un ciberataque a gran escala.

El alcance del marco propuesto se articulará en tres capas complementarias. En primer lugar, la capa normativa, en la que también se incluye el marco procedimental. En este nivel se definen los roles, responsabilidades y las reglas de enfrentamiento a las que se ceñirán los ejercicios. En segundo lugar, la capa operativa en donde se establecerán qué perfiles son necesarios para la correcta realización de la emulación. En este punto se establecen los equipos mínimos de personal, los de infraestructura (herramientas que se van a utilizar, VPS como redirectores o repositorios compartidos) y los procedimientos de custodia de evidencias. Por último, la capa de evaluación y mejora, donde se homogeneizarán las métricas, criterios de éxito del ejercicio, procedimientos de repetición de pruebas y la transferencia del conocimiento para que los hallazgos se compartan con otras entidades y organizaciones de la manera más productiva posible.

Finalmente, el marco sugerirá la integración de aquellos actores fundamentales (e.g. entidades estatales, operadores de infraestructuras críticas y sector privado entre otros) para la Seguridad Nacional, articulando canales operativos que deriven en ejercicios coordinados y escalables. La creación de mecanismos de coordinación centralizada, así como modelos descentralizados, son necesarios para la coordinación y planificación. Esto es necesario para adaptar los requisitos de sectores concretos a su criticidad operativa. El fin último será la institucionalización de los ejercicios de Red Team en la estrategia nacional de ciberdefensa, convirtiéndolos en elementos verificables y alineados con los objetivos estratégicos de protección de los intereses públicos.

3.3.1 Elementos clave del marco nacional para ejercicios Red Team

Un marco nacional que incluya la realización de ejercicios de Red Team tendrá como fin último la mejora continua de las capacidades de un país en un entorno degradado del ciberespacio. Para ello, debe constituirse como un ecosistema regulado e interoperable. El diseño del ejercicio debe ir más allá de los procedimientos técnicos, ya que en sí supondría un sesgo a la hora de analizar las capacidades defensivas. Por lo tanto, deberá integrar gobernanza, obligaciones legales, requisitos operativos, criterios objetivos de evaluación ajustados a cada casuística y articular los canales de cooperación público-privada que habiliten la ejecución de emulaciones realistas sin que ello suponga un compromiso de la seguridad de los sistemas en producción ni la continuidad de los servicios prestados, siendo ejemplos claros los ejercicios OTAN como Crossed Swords [51].

En primer lugar, es necesario definir las responsabilidades durante el ejercicio y los mecanismos que aseguren el compromiso de todos los actores implicados en el ejercicio. En este punto, entra en juego la gobernanza y el marco legal [52, 53]. Tanto la normativa de cada sector como sus obligaciones derivadas (e.g. NIS o ENS) deben integrarse explícitamente en los requisitos de la ejecución, minimizando el riesgo normativo para la organización auditada. Así mismo, deberán incluirse los esquemas de acreditación de equipos y operadores, punto importante en la definición del alcance y la definición de las reglas de enfrentamiento.

En segundo lugar, se hace indispensable contar con una documentación técnica y contractual. A nivel legal, el NDA y las ROE's representa el contrato operativo que establece el alcance, sistemas y personas excluidas del ejercicio, objetivos, ventanas temporales y límites de impacto. Por ello, estos documentos deben estar rubricados por las más altas instancias de los organismos involucrados, constatando el conocimiento no solo del ejercicio, sino también de las posibles implicaciones de este. Estos documentos deben estar disponibles para una auditoría externa, en caso de que se solicite. La experiencia validada en ejercicios como Locked Shields y Crossed Swords demuestra la necesidad de un documento técnico-operativo único que articule tanto parámetros tácticos (listas de activos incluidos/excluidos, TTPs autorizadas mapeadas a ATT&CK) como requisitos de instrumentación mínima.

En tercer lugar, todo ejercicio de Red Team necesita una infraestructura para operar. Sobre esta infraestructura se levantarán los servicios necesarios para la realización del ejercicio desde el punto de vista ofensivo. Es necesario regular el uso de herramientas comerciales (como herramientas de post-explotación tipo Cobalt Strike) y la seguridad operacional definiendo los requisitos de seguridad y bastionado de la infraestructura, para evitar riesgos legales y mitigar posibles malas praxis de algún operador. Estos controles técnicos permiten que la emulación sea lo más realista posible y adaptada a las acciones de APT's sin trasladar riesgos inaceptables a infraestructuras o sistemas en producción [5, 4].

Cuarto, el marco deberá especificar la telemetría necesaria por cada tipo de técnica empleada en la acción ofensiva. Un formato estándar de las métricas generadas favorecerá la reproducibilidad del ejercicio y su homogeneidad de cara a comparación entre ejercicios [15], posibilitando la integración y compartición de los resultados en sistemas nacionales de inteligencia y alerta, e incluso internacionales como MISP (Malware Intelligence Sharing Platform). Además, esta estandarización facilita la generación de KPI's (propuesta de KPI's representada en Anexo A).

El quinto punto hace hincapié en la gestión de riesgos y el control del propio ejercicio. La realización de un test-run garantiza que la planificación ha sido correcta, y permite verificar el correcto funcionamiento de la infraestructura operativa y de los servicios. En este punto se decide si se cumplen con los criterios necesarios para la ejecución del ejercicio con todas las garantías de seguridad. La figura del director del ejercicio es fundamental, ya que bajo su criterio se podría parar el ejercicio, ya sea por causas propias, como por alcanzar sistemas fuera del scope, como por motivos ajenos (e.g. la detección de un compromiso de seguridad por parte de actores maliciosos reales que ponga en serio riesgo la organización involucrada en el ejercicio).

En sexto lugar se debe hacer hincapié en la cooperación público-privada, especialmente en aquellos mecanismos de intercambio de inteligencia de amenazas y compartición de IOC's. El marco debe incluir acuerdos de cooperación, cláusulas de compartición de IOCs con garantías legales y procedimientos para la intersectoriales cuando se detecten riesgos sistémicos [54, 11].

En séptimo lugar, los resultados obtenidos deben ser validados y materializados en un plan de mitigación, cuyos componentes estarán priorizados por orden de criticidad. Llegado el caso, si los resultados no son satisfactorios o la criticidad del sistema afectado lo requiere, podrá acordarse un retest acotado. El marco plantea también los procesos de transferencia de las lecciones aprendidas, no solo a la propia organización auditada, sino que también extendido a otros organismos públicos. Se busca fomentar una cultura de defensa cooperativa, no solo entre instituciones, sino también entre los elementos ofensivos y defensivos [55, 56].

Por último, es preciso definir el gobierno del dato y un comportamiento ético. El marco debe incorporar una política clara sobre el tratamiento de los datos personales, especialmente aquellos sensibles hallados durante las acciones ofensivas. Asimismo, se debe incluir una colección de principios éticos que limiten el uso de técnicas o herramientas destructivas (e.g wipers o ransomware) en aras de la protección de los servicios esenciales y un uso legítimo de las simulaciones.

En suma, se busca generar una seguridad jurídica que garantice la realización de ciberejercicios ofensivos reproducibles, que generen evidencias trazables, que establezca una metodología que permita transformar las deficiencias o defectos operativos hallados, en mejoras operativas extrapolables a los distintos organismos públicos, cuya criticidad los convierte en pilares básicos de la resiliencia nacional.

3.3.2 Integración con capacidades nacionales de ciberdefensa (CSIRTs y operadores críticos)

Los ciberejercicios no deben circunscribirse al mero intercambio de hallazgos. La integración con las capacidades nacionales de ciberdefensa, y en particular con los CERT's de referencia (a saber, los CERT de INCIBE, CCN y MCCE) debe materializarse en un proceso de alineación no solo en el aspecto técnico, sino también en el procedimental y normativo. Para que la mejora de la Seguridad Nacional en el plano ciber sea real, deben adoptarse mecanismos permanentes que permitan que los resultados de los ejercicios se incorporen en un plazo corto de tiempo a los sistemas defensivos de los organismos nacionales (IOC's, playbooks de respuesta a incidentes, resolución de fallos procedimentales, etc...). Para lograr este objetivo, se hace imprescindible formatos estándar para el intercambio de ciberinteligencia, procedimientos claros que definan roles, plazos y protocolo de escalado y transferencia operativa que transformen hallazgos en reglas efectivas en el SOC [12, 54].

Desde un punto de vista técnico, la consecución de la interoperabilidad para por la exigencia de acuerdos previos de compartición de hallazgos para su integración y validación en los SIEM de otros organismos. Con esto lo que se busca es una defensa cooperativa que proporcionará una mejor respuesta ante amenazas futuras. Si se practican, mediante la ejecución de ciberejercicios ofensivos, el intercambio no solo de hallazgos, sino también de procedimientos de respuesta, se mejorará enormemente la capacidad de respuesta de los diferentes organismos. No hay que olvidar que, en el caso de España, que cuenta con una administración fuertemente deslocalizada, procedimientos de respuesta comunes por sectores permitirán una mejora continua de la resiliencia.

La integración, bajo una perspectiva regulatoria y de procedimientos, debe considerar los requisitos de notificación y protección derivados de marcos como NIS2 y del Esquema Nacional de Seguridad. Es recomendable formalizar las rutas de escalado en caso de incidente de seguridad, definiendo obligaciones como quién notifica al CERT de referencia, cuándo y cómo. Se trata de reducir incertidumbre en un momento de alta tensión (i.e. durante una crisis derivada de un ciberataque), trazando unos caminos claros de ejecución de la respuesta para un rápido control del

evento. Este aspecto debe ser ensayado durante un ciberejercicio ofensivo con el fin de automatizar respuestas y generar mecanismos de gran celeridad. Por ello, la realización de sesiones conjuntas de adiestramiento y compartición de medidas de respuesta entre diferentes organismos es algo que se debe incluir en las ventanas temporales entre ejercicios.

Finalmente, la integración debe atender a criterios de protección de datos, proporcionalidad y minimización: todo intercambio entre Red Team, CERT y operadores ha de contemplar evaluaciones de impacto en la protección de datos, empleando técnicas de anonimización para evitar exposición innecesaria de datos personales. Asimismo, la práctica regular de ejercicios conjuntos y la estandarización de formatos y procedimientos facilitan la construcción de una capacidad nacional proactiva, que trasponga la experiencia de ejercicios ofensivos en mejoras defensivas operativas y en la resiliencia de los diferentes sectores y servicios críticos [12, 54].

3.3.3 Rol de los ciberejercicios en la mejora continua de la resiliencia nacional

Los ejercicios ofensivos son un pilar fundamental en la estrategia de mejora de las capacidades defensivas de cualquier organización. Mediante la emulación de la operativa de adversarios reales, contribuyen a la mejora continua de la resiliencia de las entidades porque permiten, en un entorno controlado, validar la capacidad de detección y respuesta a incidentes de ciberseguridad. Un ejercicio de Red Team va más allá del plano técnico, midiendo la capacidad de decisión del personal del Blue Team, sus procesos, responsabilidades y coordinación con los CERT's de referencia, que, en una crisis real, determinan la capacidad de contención y recuperación. Un ejercicio bien ejecutado, generará evidencias accionables que se incorporarán al ciclo de adiestramiento de los agentes defensivos, y que permitirán la mejora de la defensa en líneas generales.

Ejercicios internacionales como Locked Shields ponen de manifiesto la necesidad de mejora continua. Escenarios tan complejos, en el que se integran diferentes sectores críticos, son el banco de pruebas perfecto para evaluar las capacidades defensivas de una organización. A mayor complejidad y realismo del ejercicio, mayor parecido con la realidad, y los objetivos conseguidos serán más beneficiosos para mejorar la resiliencia nacional. No es solo la parte de la respuesta al ataque, sino también los procedimientos mejorados, derivados de la respuesta a una acción ofensiva que emula con gran realismo a las realizadas por agentes maliciosos reales. Como se ha mencionado, un ejercicio de Red Team de estas características busca convertir hallazgos puntuales en capacidades permanentes, cuantificando las mejoras mediante KPI's, permitiendo monitorizar la evolución en el tiempo de las capacidades de defensa [15].

Un elemento de gran aportación por parte de los ciberejercicios es su contribución a la armonización de formatos y la estandarización de procedimientos técnicos. La adopción de formatos y protocolos comunes facilita enormemente la interoperabilidad, reduciendo el tiempo de reacción en un escenario real, al enfrentarse a una ciberamenaza. Por otra parte, la información de ciberinteligencia recolectada en la fase inicial del ejercicio, es una información de gran ayuda que permite trazar con mayor exactitud la superficie de exposición, así como las posibles brechas de seguridad que hayan sido expuestas, por ejemplo, en la dark web. Como se ha visto, el marco TIBER-EU, donde se articulan ciberejercicios en base a la inteligencia recabada, respetando requisitos jurídicos en la recolección de evidencias y cadena de custodia, es un caso de éxito de marco de referencia aplicado a la ciberseguridad.

No hay que desdeñar el proceso de mejora de comunicación entre roles atacantes y defensores que se produce durante los ejercicios de Red Team. La capa defensiva participa de la mentalidad proactiva de la ciberseguridad, donde la prueba recurrente y el testeo continuo con claves en una resiliencia eficiente. Además, permiten explorar el concepto de Purple Team, muy recomendable en

los periodos entre ejercicios, donde la colaboración entre los equipos atacantes y defensivos es plena, ejecutándose las acciones de forma coordinada.

Se puede concluir que, para que un ciberejercicio ofensivo cumpla su papel en la mejora continua de las capacidades de detección y gestión de incidentes, es imprescindible que se integren en una estrategia nacional coherente. España ya cuenta con un órgano específico de ciberseguridad dentro de su Departamento de Seguridad Nacional, el Consejo de Ciberseguridad Nacional. Se trata de *un órgano colegiado de apoyo al Consejo de Seguridad Nacional en su condición de Comisión Delegada del Gobierno para la Seguridad Nacional, en el marco de la Ley 50/1997, de 27 de noviembre, del Gobierno*. Su misión es la de reforzar las relaciones de coordinación, colaboración y cooperación entre las distintas Administraciones Públicas, y que tengan competencias en materia de ciberseguridad, así



Figura 3-7 Integrantes del Consejo Nacional de Ciberseguridad. Fuente: dsn.gob.es

como entre los sectores públicos y privados. También *facilitará la toma de decisiones del propio Consejo mediante el análisis, estudio y propuesta de iniciativas tanto en el ámbito nacional como en el internacional*. Entre sus cometidos, el Consejo Nacional de Ciberseguridad (en la figura 3-7 se muestran sus integrantes) prestará apoyo al Consejo de Seguridad Nacional verificando que se cumpla la Estrategia de Seguridad Nacional en lo concerniente a la ciberseguridad, promoviendo e impulsando sus revisiones. Por lo tanto, se trata de un órgano idóneo para el impulso de este tipo de ejercicios a nivel nacional.

4 RESULTADOS / VALIDACIÓN / PRUEBA

El presente capítulo reúne y sistematiza los resultados derivados de la propuesta teórico-metodológica y se expone el marco de validación concebido para garantizar la coherencia, la reproducibilidad y la utilidad práctica de las propuestas, sin que ello implique la realización de simulaciones u operaciones de Red Team reales, dados los recursos que se necesitarían para ello. Por lo tanto, se expondrán, siempre desde el punto de vista teórico, los artefactos propuestos, cómo deben emplearse en un entorno real, y la definición de protocolos de validación.

4.1 Análisis teórico de resultados

A continuación, se presenta un análisis desde un punto de vista teórico, basado en la experiencia. Con ello se ofrece un marco robusto, que permita medir el impacto de un ejercicio ofensivo bien estructurado y planificado en los organismos involucrados. Asimismo, se busca también proporcionar instrumentos (en forma de métricas) que permitan que el ejercicio sea reproducible en otros entornos similares.

4.1.1 Evaluación de las capacidades defensivas nacionales frente a los ejercicios (enfoque proyectado)

Como punto de partida, es necesario conocer el estado de las capacidades defensivas, tanto activas como pasivas, de cada organismo. El resultado de esta evaluación inicial será el establecimiento de un baseline que servirá de elemento comparador tras la implantación de las lecciones aprendidas (i.e. la implementación de las medidas correctoras y de mitigación de vulnerabilidades detectadas durante la realización del ejercicio ofensivo) derivadas del ciberejercicio. Debido a la ausencia de ensayos prácticos reales, la evaluación de capacidades se planteará como un proceso de diagnóstico y verificación. El procedimiento de verificación propuesto debe detallar los siguientes puntos:

1. Inventario y clasificación de activos y servicios críticos (que según la tabla 3-1 se les ha asignado un valor que va del 1 al 5), registro de telemetría disponible y evaluación de la madurez de la organización.
2. Cuestionario estructurado dirigido al SOC, equipos de respuesta y responsables de TI para identificar el nivel de conocimiento de los procedimientos propios, así como la cadena de escalado en caso de incidente de ciberseguridad.
3. Matriz de estado operativo que cruza criticidad de los servicios con la madurez de la organización y la disponibilidad de telemetría, generando un valor de disponibilidad. Este valor orienta la selección de escenarios, casos de uso y la intensidad del ejercicio, que puede

comprender desde la simple formación de operadores y concienciación de usuarios hasta una emulación completa de actores adversarios avanzados.

De la matriz de estado operativo se derivará la recomendación o no de proceder con el ejercicio. La matriz arrojará un valor de “readiness” o disponibilidad, que dará cuenta del estado de preparación de la organización para enfrentar un ejercicio ofensivo complejo. El valor de disponibilidad se calcula a partir de tres factores, a saber: criticidad (C), madurez de la organización (M) y disponibilidad de telemetría (T) (logs, eventos registrados en el SIEM, alertas de EDR entre otras). En consonancia con la definición de valores que se han descrito en la metodología, se asigna a estas variables un valor de 1 a 5. En base a esto, se propone la siguiente fórmula para el cálculo del valor de disponibilidad:

$$readiness = \frac{0,25C + 0,4M + 0,35T}{5} \times 100$$

La ponderación realizada en la fórmula se justifica de la siguiente manera. Tanto madurez como telemetría determinan la capacidad real del organismo para realizar y medir un ejercicio ofensivo sin riesgo para la operativa, de ahí que sus valores sean mayores. La criticidad de los servicios ofrecidos actúa como factor que prioriza y condiciona la intensidad, pero no sustituye a la preparación técnica. El valor arrojado como resultado del readiness dará cuenta de la intensidad con la que se puede realizar el ejercicio. Como propuesta, se presentan las siguientes acciones en función del valor obtenido:

1. Readiness < 40 implica que la organización necesita una mejor preparación para afrontar un ejercicio de entidad. Se recomienda, antes de la realización del ejercicio, la formación de los operadores y responsables.
2. 40 < Readiness < 60 la organización está preparada para ejercicios de baja intensidad en entornos replicados, y puede afrontar pruebas limitadas y controladas. Un buen escenario sería la ejecución de ejercicios previamente maquetados y encapsulados, con playbooks definidos.
3. 60 < Readiness < 80 la organización tiene la capacidad de enfrentar a un Red Team en un escenario limitado, sin acceso a los sistemas en producción. En este escenario, la realización de un Purple Team es aconsejable.
4. Readiness > 80 la organización está preparada para afrontar un escenario de emulación avanzada en un marco autorizado, con registro de evidencia completa, gestión de incidentes y escalado de responsabilidades, posibilitando en ejercicio con un realismo alto.

Existe un caso a tener en cuenta de manera separada. Si la criticidad de la organización es muy alta, con un valor de 5, pero Readiness es bajo, el resultado debe interpretarse como que existe un riesgo alto para la realización del ejercicio, y por lo tanto la recomendación será no ejecutar simulaciones en producción hasta mitigar carencias.

C	M	T	Readiness (%)	Recomendación
5	1	1	40%	No ejecutar; formación + preparación telemetría
4	2	2	50%	Ejercicios de baja intensidad
3	3	3	60%	RT controlado en preproducción
2	4	4	70%	RT limitado. Recomendado Purple Team
5	5	5	100%	Emulación avanzada (APT-like) autorizada

Tabla 4-1 Matriz de cálculo del valor de Readiness o disponibilidad.

El resultado de esta evaluación es el nivel de capacitación que identifica el nivel de realismo que la organización puede enfrentar. De nada sirve ejecutar un ejercicio que emula actores avanzados contra una institución que no está preparada para hacer frente a una amenaza de ese calibre. Siempre es necesario adaptar el ejercicio al nivel del organismo objetivo, puesto que se trata de mejorar capacidades, identificar defectos de ciberseguridad a todos los niveles, y estructurar una defensa cooperativa que permita la mejora de la Seguridad Nacional.

4.1.2 Propuesta de métricas de desempeño: definición, cálculo y criterios de interpretación

Las métricas y los KPI's vertebran el núcleo del marco propuesto para evaluar la efectividad defensiva de las organizaciones durante la realización de ciberejercicios ofensivos. El propósito de la estandarización de estos valores es que cualquier auditor, analista o equipo técnico pueda reproducir los cálculos simulados sobre datos reales que permitan obtener indicadores comparables. Es por esto que todas las métricas incluyen una definición, método de cálculo si fuese necesario, la fuente de donde se obtienen los datos requeridos y umbrales orientativos que podrán ser ajustados en función de la organización objetivo.

Así pues, desde el punto de vista teórico, se proponen las siguientes métricas y KPI's, incluyendo aquellos propuestos por [15] y que se considera recomendable incluir en el marco.

1. Detección por técnica (DT): nivel de detección de una determinada técnica o TTP empleada por el Red Team. Esta variable tendrá los siguientes valores
 1. 0 no detectado
 2. 1 detección por parte de un analista humano
 3. 2 detección automática por algún sistema defensivo

Para su cálculo, se empleará el log de registro de actividades de los operadores del Red Team y se cruzará con las alertas del EDR o los eventos del SIEM. Este valor se empleará como dato para el cálculo del valor de cobertura.

2. Cobertura: se trata de un valor no ponderado que da cuenta de la proporción normalizada de detección sobre técnicas ejecutadas por los operadores del Red Team.

Para su cálculo se propone la siguiente fórmula

$$Cobertura = \frac{\sum_{i=1}^N DT_i}{2N}$$

N es el número de técnicas ejecutadas (acciones evaluables) y $DT \in \{0, 1, 2\}$. Los valores propuestos para la cobertura son aceptables para valores mayores del 75% y excelentes a partir

del 90% de cobertura de detección. Estos valores, pueden ser ajustados en función de la criticidad del organismo o de los servicios prestados.

3. Cobertura ponderada: se trata de la métrica anterior ponderada por el valor de criticidad.

$$Cobertura_{ponderada} = \frac{\sum_{i=1}^N (DT_i \times w_i)}{2 \sum_{i=1}^N w_i}$$

w es el valor de ponderación de cada activo por su criticidad con un valor de 1 a 5 según el criterio escogido.

4. MTTD (Median Time to Detect): mediana del tiempo transcurrido entre la ejecución de la técnica y la primera detección/alerta. Para su cálculo se emplea la mediada del tiempo de ejecución por parte de los operadores del Red Team, como el tiempo de detección.

$$MTTD_{mediana} = mediana(\Delta t_i = t_{detección,i} - t_{ejecución,i} | i = 1..N_{detectadas})$$

La unidad de medida recomendada son minutos. En esta métrica cabe preguntar qué pasa con las acciones no detectadas. Dichas acciones, desde el punto de vista estadístico se registran como censuradas (i.e. no se puede saber si el evento ocurrirá o no, solo se sabe que hasta ahora no ha ocurrido). Para evitar un sesgo, estas medidas no deben incluirse en el cálculo. En caso de que haya muchas observaciones censuradas, se recomienda el uso de un análisis de supervivencia como el de Kaplan-Meier, donde el caso de análisis no es cuántos eventos ocurren, sino el tiempo que pasa hasta que ocurre el evento. Por lo tanto, el resultado de este análisis es la probabilidad de que el evento no haya sucedido pasado un cierto tiempo.

5. MTTR (Median Time to Response/Content): mediana del tiempo entre detección y contención/mitigación efectiva. En este caso se trata del tiempo medio desde la detección de una determinada TTP hasta la respuesta inicial de la organización para realizar el triaje.

$$MTTR_{mediana} = mediana(\Delta t_i = t_{contenidos,i} - t_{detectados,i} | i = 1..N_{contenidos})$$

Como en el caso anterior es necesario tener en cuenta a los casos no detectados. Por ejemplo, si la contención no se produjo durante el periodo de observación el caso debería ser etiquetado como censurado, y como la métrica anterior se recomienda el uso de un análisis de supervivencia. También es necesario discernir entre dos conceptos que suelen mezclarse: contención (i.e. detener la propagación del ataque) y remediación completa (que suele implicar parcheo definitivo de la vulnerabilidad que permite la intrusión).

6. MTIA (Mean Time to Initial Access): Tiempo medio empleado por el Red Team en lograr acceso inicial no autorizado tras iniciar el ataque. El acceso inicial puede lograrse explotando diferentes vectores, sin tener relevancia esto a la hora del cálculo de esta métrica. Podría considerarse introducir un elemento de ponderación en función de la complejidad de la técnica, ya que no es lo mismo la explotación de una vulnerabilidad ya reportada mediante un exploit público, que la detonación de un BOF generado por algún C2 y enviado a través de un phishing.

$$MTIA = mediana\{t_{acceso_inicial,i} - t_{inicio_acción,i}\}$$

7. MTTA (Mean Time To Action): Tiempo medio entre la respuesta inicial tras la detección de una determinada técnica o TTP y la implementación de una solución o acción mitigadora. Dicha acción deberá tener un efecto real sobre la técnica de ataque empleada, por lo que en el

retest debe ser verificado. Si la acción mitigadora se demuestra ineficaz, esta métrica deberá ser calculada de nuevo.

$$MTTA = mediana\{t_{implementación,i} - t_{detección,i}\}$$

8. TTFR (Time to Fix Remediation): Tiempo entre el cierre del ejercicio y la verificación de la efectividad del control implementado para remediar la vulnerabilidad. Se trata de una de las métricas de más valor, ya que da cuenta de la preparación de la organización para de forma efectiva, corregir deficiencias que han permitido una intrusión.

$$TTFR_i = t_{verificación,i} - t_{cierre}$$

En el caso de que no sea posible la verificación dentro de un margen temporal razonable, se recomienda marcar esta métrica como censurada y usar un análisis de supervivencia.

9. Porcentaje de técnicas con ingesta mínima: porcentaje de técnicas ejecutadas que contaban con la telemetría mínima definida para su evaluación (host, red, app). Esta métrica se calcula sobre el plan de ejecución que tendrá previsto el Red Team. El objetivo de esta métrica debe estar próximo al 90% para así dar fe de que todos los casos previstos por el RT son implementados. Un porcentaje menor no garantiza que toda la superficie de ataque haya sido puesta a prueba.

$$Ingest = 100 \times \frac{N_{ingest\ mínima}}{N_{ejecutadas}}$$

N representa el número de técnicas.

10. Tasa de falsos positivos: porcentaje de alertas generadas durante el período que no correspondían a actividad maliciosa llevada a cabo por el Red Team. La tasa de falsos positivos también es una variable de gran importancia, ya que da cuenta del nivel de madurez del Blue Team. Una tasa alta de falsos positivos evidencia que los sistemas están mal configurados y es necesario un mejor ajuste para que las alertas se correspondan con incidentes de seguridad reales. La tasa de falsos positivos se calcula de la siguiente forma:

$$TFP = 100 \times \frac{N_{falsos_positivos}}{N_{alertas_totales}}$$

Si bien no es uno de los objetivos de los ejercicios de Red Team, el altamente recomendable que la experiencia adquirida sirva como catalizador para mejorar esta métrica. Un elevado número de falsos positivos puede derivar en una mala praxis a la hora de gestionar las alertas.

11. Tasa de éxito tras retest: porcentaje de mitigaciones que superaron un retest (i.e. tras la fase de lecciones aprendidas el nuevo test de intrusión que verifique que las mitigaciones implementadas son efectivas ante ataques similares) según los criterios de aceptación previamente definidos.

$$TER = 100 \times \frac{N_{mitigaciones_aprobadas}}{N_{mitigaciones_probadas}}$$

Nuevamente, lo deseado en esta métrica son valores cercanos al 90%.

Todas las métricas definidas anteriormente dan cuenta del estado defensivo de la entidad. Sin embargo, están condicionadas al correcto registro por parte del Red Team de todas las acciones emuladas. Como consecuencia de esto, es recomendable contar con un log de las acciones llevadas a cabo por el equipo ofensivo, y referenciadas al mismo huso horario, y con sincronía, ya que los valores temporales se registran en minutos. Para la correcta interpretación de las métricas es necesaria la trazabilidad de los artefactos, ya que de lo contrario se introducirían valores espurios. Por último, los umbrales de aceptación de las métricas deberán adaptarse al sector objetivo (defensa, energía, salud).

4.1.3 Identificación de gaps y oportunidades de mejora

Al no contar con datos reales derivados de ciberejercicios, la identificación de los gaps y puntos de mejora se planta como un protocolo de análisis que el evaluador aplicará a las evidencias operativas generadas durante un ejercicio de Red Team. Como parte de la propuesta, se propone un protocolo que tenga en cuenta la información recabada durante la fase de inteligencia y reconocimiento, que permita transformar los hallazgos plausibles en un registro priorizado de mitigaciones. Este registro deberá contar con un listado de las acciones, los responsables, los recursos los plazos y los criterios de verificación. Con esto se busca que, en el momento en que se dispongan de evidencias operativas tras un ejercicio, el evaluador pueda ejecutar el proceso maximizando la trazabilidad y minimizando la ambigüedad.

El protocolo de análisis contempla una serie de pasos, que a continuación se pasan a definir.

12. Mapeo de las TTP's: durante la fase de inteligencia y reconocimiento se identificarán no solo la superficie de exposición, sino también las TTP's empleadas por los principales grupos APT y que centren sus actividades en el sector de la organización a auditar. Por cada técnica que va a ser empleada por el equipo ofensivo deberán implementarse los controles tanto desde el punto de vista técnico como de procedimiento para su detección en condiciones ideales. Así mismo, también se preparará el plan de gestión de incidentes, para que la reacción ante cada una de las técnicas empleadas esté preparada y pueda implementarse a la mayor brevedad. El conjunto de TTP's recopilado se puede considerar como una lista de requisitos, contra la que se evaluará la cobertura real o teórica de la organización. En el caso de que falte alguna técnica, se considerará un potencial punto ciego de la entidad.
13. Matriz de gap: se trata de una matriz resultado de cruzar la métrica de detección (como se definió en el apartado anterior tendrá un valor de 0, 1 o 2 en función de si no se detecta o la detección es automática o manual) con la criticidad y el tiempo de mitigación, dando como resultado una priorización que se puede escalar en urgente, alta, media o baja. De esta forma, se obtiene un resultado único por gap, que permite asignar la prioridad de forma transparente y reproducible.
14. Categorías de oportunidad: el objetivo de este punto es facilitar la toma de decisiones de cara a orientar la respuesta al incidente. Se trata de identificar si las reglas que existen en los sistemas defensivos son suficientes, hay deficiencias en los procesos (e.g. SLA's con tiempos no realistas, playbooks de respuesta a incidentes desactualizados, etc...). En este punto también se detectarán deficiencias formativas. Por ejemplo, si se cuenta con un SIEM Splunk, todos los operadores de este deberán contar con la formación necesaria para su correcto uso. Cada gap detectado ayuda a seleccionar la solución más adecuada y asignar responsables para su corrección.
15. Plan de mitigación estructurado: Por cada uno de los posibles gaps identificados es necesario definir una acción de remediación, un responsable de llevarla a cabo, una asignación de recursos, así como un criterio de éxito verificable y realista. Esto último es lo más importante, ya que el posterior retest pondrá a prueba la eficacia de la remediación. También se puede plantear un plan de retest acotados, que abarque exclusivamente los puntos que se han

mejorado. Esta atomización de las pruebas es más trabajosa, pero supone un excelente mecanismo de comprobación de la remediación de gaps.

Para gestionar el flujo operativo se recomienda mantener un registro formalizado (podría ser en un formato de ticketing, similar al usado por herramientas del CCN como Lucía), con campos estandarizados que faciliten su uso. Esos campos deberán incluir el identificador del gap, la TTP vinculada, activo afectado y su criticidad, la prioridad calculada, la acción de mitigación recomendada, responsable de su remediación, la fecha de retest y los criterios de aceptación. Este registro es en última instancia un documento operativo que permite transformar hallazgos en actuaciones concretas y facilita el seguimiento.

4.1.4 Impacto estratégico y operacional del uso de Red Team en ejercicios nacionales

En este apartado se evaluará de forma prospectiva y basada en la literatura y las experiencias documentadas, los beneficio estratégicos y operacionales que conllevan la aplicación sistemática de ciberejercicios ofensivos encuadrados dentro de un catálogo de ejercicios nacionales. Se identifican 4 campos donde redundarán estos beneficios:

1. **Beneficios estratégicos:** de cara a una mejora de la defensa estratégica de las organizaciones pública y críticas, la realización de ejercicios Red Team facilita el aumento de la madurez de detección y respuesta. Esta mejora va de la mano del empleo de la emulación de adversarios reales. Evidentemente, nunca será una emulación completa, pero sí permite mejorar las estrategias defensivas contra adversarios avanzados, algo fundamental en un mundo hiperconectado y tan dependiente de la tecnología. También permite la identificación de riesgos sistémicos llevándolos al extremo en ejercicios tan demandantes. Otro factor clave es la mejora de la coordinación interinstitucional, punto este clave en la mejora de la resiliencia como país. Los ejercicios informados por recolección de inteligencia previa (e.g. TIBER-EU) favorecen la priorización de inversiones, así como la estandarización normativa.
2. **Impacto operacional:** desde el punto de vista operacional, la implementación estructurada de ejercicios de Red Team se materializa en mejoras concretas y medibles. Además, cuando se baja al nivel táctico, en los primeros escalones de la defensa, de las lecciones aprendidas tras ciberejercicios ofensivos se traducen en mejoras en la primera línea de defensa, refinando, por ejemplo, reglas de SIEM, evolucionando los playbooks de respuesta a incidentes. El impacto operacional debe plasmarse en la mejora de las métricas operativas, de ahí la importancia de la periodicidad de estos ejercicios. Su ejecución recurrente permitirá la comparación de métricas entre diferentes años, y será un criterio cuantificable que permitirá medir la eficacia de las mitigaciones implementadas y la evolución de la defensa en primera línea. Como producto secundario se obtienen una metodología de estandarización de los procesos de verificación de mitigaciones, lo que acorta los ciclos de corrección e incrementa la probabilidad de que los hallazgos muten en mejoras operativas reales.
3. **Costes y riesgos:** en rasgos generales, los ejercicios de Red Team conlleva un coste elevado, que va parejo a riesgo inherente de su realización. Ningún ejercicio ofensivo garantiza que sus acciones sean inocuas. Por lo tanto, es necesario que el gasto generado sea justificado en forma de mejoras a todos los niveles de la capa de seguridad. Por eso es tan importante contar con métricas trazables y reproducibles, que sirvan de argumento justificativos de la inversión. Respecto a los riesgos implícitos, la probabilidad de afectar a algún servicio en producción siempre está presente. Debido a esto, la evaluación del readiness debe ser lo más escrupulosa posible, dando un criterio de no go si fuese necesario para prevenir incidentes no deseados derivados de una mala preparación del ejercicio.

-
4. **Efecto en la resiliencia nacional:** los ejercicios recurrentes, sistemáticos y acreditados, constituyen un elemento de gran valor de cara a la mejora de la resiliencia nacional. La mentalidad debe ir más allá de la preparación para el ataque, es necesario asumir la brecha inicial y el compromiso, para generar una mentalidad proactiva que facilite la defensa activa. Se trata, en última instancia, de estar preparados y prevenidos para una respuesta rápida y efectiva ante incidentes, algo que solo se logra a través de la preparación constante tanto del personal como de los sistemas.

5 CONCLUSIONES Y LÍNEAS FUTURAS

A continuación, se presentan las conclusiones a las que se ha llegado en el presente trabajo. Durante la realización de este se han presentado una serie de limitaciones que también serán presentadas en este capítulo. Principalmente aquellas derivadas del planteamiento teórico del trabajo. Así mismo se presentan las futuras líneas de investigación y posibles desarrollos futuros.

5.1 Conclusiones generales

El presente Trabajo de Fin de Máster tiene como objetivo el desarrollo de una metodología integral y reproducible para el planeamiento de ciberejercicios ofensivos realizados por equipos de Red Team, orientados a la mejora de las capacidades defensivas de la Defensa Nacional. Como principales resultados, se proporciona un catálogo de KPI's materializados en forma de métricas (Anexo A). Así mismo también se ha definido una matriz de readiness, que permite relacionar valores como criticidad, madurez de la organización y disponibilidad de telemetría para definir un criterio de intensidad del ejercicio, o incluso dar el “no go” si la organización no está lo suficientemente preparada. Estas aportaciones permiten traducir hallazgos técnicos en procesos de mejora medibles, priorizables y trazables, facilitando la transferencia práctica de resultados (procesos de mejora de los procedimientos del SOC, refinamiento de reglas de SIEM, actualización de procedimientos de gestión de incidentes, etc...) y la posterior verificación por terceros.

Para la capa estratégica, la metodología propuesta ofrece criterios objetivos para la priorización de las inversiones en ciberseguridad, armonizando criterios entre los diferentes actores (CERT's, operadores críticos, reguladores) gracias a la normalización de formatos y procedimientos, compatibles con enfoques basados en la recolección de inteligencia previa, como ya se aplica en marcos como TIBER-EU. Bajando al plano operacional, las métricas desarrolladas facilitan instrumentos que cuantifican la eficacia defensiva, lo que conlleva una optimización de los ciclos de corrección gracias a procesos estándar de mitigación y retest. Asimismo, el protocolo de identificación y priorización de gaps ofrece una guía reproducible para convertir vulnerabilidades detectadas (ya sean reales o plausibles en base a la inteligencia recolectada) en un registro operativo con responsables, recursos y criterios de aceptación medibles.

En resumen, la propuesta constituye una base metodológica para institucionalizar ejercicios Red Team en el ámbito nacional. Esto se logra gracias a la aportación de métricas y trazabilidad, habilitando que los resultados de la emulación ofensiva se traduzcan en una capacidad operativa real y permanente que permite la toma de decisiones a nivel estratégico.

5.2 Limitaciones del trabajo

Durante la realización del presente trabajo se han presentado una serie de limitaciones:

-
1. No se ha podido realizar la validación empírica mediante ejercicios reales. Los recursos de los que se han dispuesto no permitieron la ejecución de simulaciones, ni se ha contado con un Red Team. En consecuencia, muchas conclusiones sobre impacto operacional son prospectivas y requieren validación práctica (pilotos, test-runs o ejercicios controlados) para confirmar su aplicabilidad.
 2. Las métricas propuestas se han diseñado para un máximo de validez y reproducibilidad. Sin embargo, su utilidad práctica depende en gran medida con la existencia de colectores adecuados, bien configurados y operativos, como son EDR o SIEM. En entornos con menor capacidad, la limitada cobertura ofrecida podría limitar las métricas introduciendo posibles sesgos.
 3. La ponderación y los pesos asignados en la matriz de readiness se han definidos sobre criterios de buenas prácticas y en base a la experiencia del autor. Sin embargo, dichos pesos deben ser ajustados en base a pruebas empíricas por sectores y contexto, para evitar la toma de decisiones en base a resultados espurios.
 4. A pesar de que la metodología se ha pensado con un factor de escalabilidad pensando en sectores específicos. Cada uno de los diferentes sectores presentan sus propias particularidades técnicas, regulatorias y operativas, por lo que el catálogo de TTP's a aplicar varía significativamente. No es lo mismo escenarios OT/ICS que el sector sanitario, por poner un ejemplo. Así mismo, este particular también puede incidir en la decisión de las reglas de enfrentamiento a emplear.
 5. El trabajo incorpora recomendaciones sobre formación y concienciación de un factor clave, el humano, pero no han medido empíricamente este factor. Las reacciones bajo estrés, toma de decisión en un entorno muy demandante o la experiencia de los operadores y personal que tome las decisiones influyen decisivamente en la eficacia operacional y que requieren estudios complementarios.

5.3 Líneas de investigación y desarrollo futuro

Las líneas de investigación y los desarrollos futuros se centran en la validación de la metodología mediante ejercicios pilotos controlados. Una opción es replicar la infraestructura objetivo (si no en su totalidad al menos en un modelo a escala) en un cyber range que permita una emulación realista y así poder validar las métricas definidas. Otra opción sería la generación de datasets sintéticos, que permitan a los auditores probar la metodología sin comprometer datos reales, permitiendo la evaluación previa al despliegue. Así mismo, en lugar de usar datos sintéticos, existe la posibilidad de anonimizar datos reales de organismos del mismo sector, para lo que habría que buscar mecanismos de colaboración.

Por otra parte, es necesario priorizar los sectores con un impacto mayor debido a su criticidad. Esto hace necesario adaptar y validar la metodología para sectores específicos (infraestructuras críticas, OT), con unos requisitos donde la instrumentación, los riesgos y las reglas de intervención difieren significativamente del sector IT convencional. Su alta especialización hace necesario también el desarrollo de métricas ad hoc para estos casos concretos (e.g. registros de consignas maliciosas desde un sistema de control SCADA a diferentes PLC's).

Asimismo, dada la importancia del factor humano, clave en la gestión de la ciberseguridad, se hace necesario el diseño de experimentos y estudios cualitativos que midan el impacto de los ejercicios en la toma de decisiones bajo ambientes muy demandantes, el escalado de responsabilidades y la coordinación interinstitucional. Se trata de que el diferente personal involucrado, a todos los niveles, sea capaz de gestionar sus cometidos con la mayor eficacia, sacando el máximo provecho de los ciberejercicios ofensivos.

Por último, la creciente dependencia entre los diferentes ecosistemas digitales, así como su propia idiosincrasia, hace que se las ciberamenazas se extiendan más allá de nuestras fronteras. Así pues, en aras de una integración de la metodología con otros marcos nacionales y europeos, se propone el

estudio de mecanismos que habiliten dicha integración. La integración con otros marcos busca dos objetivos complementarios: la evaluación de riesgos sistémicos que se pudiesen manifestar en cadenas de suministro y servicios críticos distribuidos a lo largo de varios países, y la potenciación del aprendizaje colaborativo, así como la convergencia de buenas prácticas y lecciones aprendidas derivadas de ciberejercicios realizados en otros entornos ajenos al nacional. Para lograr una integración de facto y jurídicamente segura es necesario desarrollar aquellos mecanismos que permitan la interoperabilidad entre marcos nacionales y europeos (e.g. entre las adopciones nacionales de TIBER-EU y los instrumentos y guías promovidos por ENISA) abordando las capas técnica, procedimental y legal.

6 BIBLIOGRAFÍA

- [1] Kotwani, B., Sawant, R., & Chopra, S. (2023). *Red Teaming vs. Blue Teaming: A Comparative Analysis of Cyber Security Strategies in the Digital Battlefield*. IJSREM.
- [2] INSSIDE. (2024). *¿Qué es el Purple Team en Ciberseguridad?* <https://www.insside.net/purple-team-en-ciberseguridad/>
- [3] Shirey, R. (2007). *Internet security glossary, version 2* (RFC 4949). <https://doi.org/10.17487/RFC4949>
- [4] Erickson, J. (2003). *Hacking: The art of exploitation* (2nd ed.). No Starch Press.
- [5] Allsopp, W. (2017). *Advanced penetration testing: Hacking the world's most secure networks*. Wiley.
- [6] National Institute of Standards and Technology. (2010). *Contingency planning guide for federal information systems* (SP 800-34 Rev. 1). U.S. Department of Commerce
- [7] International Organization for Standardization. (2022). *ISO/IEC 27001:2022 Information security management systems—Requirements*.
- [8] National Institute of Standards and Technology. (2012). *Guide for conducting risk assessments* (SP 800-30 Rev. 1).
- [9] European Union Agency for Cybersecurity. (2017). *Risk management: Implementing the NIS Directive*.
- [10] Hopkin, P. (2018). *Fundamentals of risk management: Understanding, Evaluating and Implementing Effective Risk Management* (5th ed.). Kogan Page.
- [11] International Organization for Standardization & International Electrotechnical Commission. (2022). *ISO/IEC 27000:2022. Information security, cybersecurity and privacy protection. Information security management systems. Overview and vocabulary*.
- [12] European Union Agency for Cybersecurity. (2012). *On national and international cyber exercises: Survey, analysis and recommendations*.
- [13] Seker, E. (2018). *The concept of Cyber Defence Exercises (CDX): Planning, execution, evaluation*. CCdCOE.
- [14] European Cyber Security Organisation, Working Group 5 (Skills & Human Factors). (2024,

- October 2). *White paper: Cyber exercise scenario development*. ECSO
- [15] Hollis, R. (2024). *Red Team testing: Essential KPIs and metrics*. *Cyber Security: A Peer-Reviewed Journal*, 7(4), 323–332. <https://doi.org/10.69554/KMOV5511>
- [16] European Central Bank. (2025). *TIBER-EU control team guidance*.
- [17] CBEST (2024k), *Supplementary Guidance on Outsourcing and Third-Party Scenarios in CBEST*, PRA, Bank of England
- [18] CBEST. (2024b). *CBEST Services Assessment Guide*. PRA, Bank of England
- [19] CREST (2024a), *CREST Certified Threat Intelligence Manager (CCTIM)*
- [20] CREST (2024b), *CREST Certified Simulated Attack Manager (CCSAM)*
- [21] Alharthi, K., Mani, J., Padmanabha, V., & Kumar, B. (2025). *Analysis and exploitation of Active Directory vulnerabilities using Cobalt Strike: A practical approach*. https://doi.org/10.1007/978-3-031-88874-8_29
- [22] Zhao, Letao. (2024). Navigating the Cyber Kill Chain: A modern approach to pentesting. *Applied and Computational Engineering*. 38. 170-175. 10.54254/2755-2721/38/20230549.
- [23] Tuo, Wenxin & Wang, Cheng & Lee, Ping & Relafos, Dinazam & Duan, Huai. (2025). *Investigation and Study on the Cyber Kill Chain Framework in Accordance with Railway Signalling Systems of the Suburban and Urban Lines*.
- [24] <https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html>
- [25] Presekal, Alfán & Ștefanov, Alexandru & Rajkumar, Vetrivel & Semertzis, Ioannis & Palensky, Peter. (2024). *Advanced Persistent Threat Kill Chain for Cyber-Physical Power Systems*. *IEEE Access*. PP. 1-1. 10.1109/ACCESS.2024.3507386.
- [26] Kour, Ravdeep & Karim, Ramin & Dersin, Pierre. (2025). *Modelling cybersecurity strategies with game theory and cyber kill chain*. *International Journal of System Assurance Engineering and Management*. 10.1007/s13198-025-02733-4.
- [27] <https://blog.fox-it.com>
- [28] Landauer, M., Mayer, K., Skopik, F., Wurzenberger, M., & Kern, M. (2024). *Red team redemption: A structured comparison of open-source tools for adversary emulation*. <https://doi.org/10.48550/arXiv.2408.15645>
- [29] Shaffer, Alan & Hembree, Dwain & Singh, Gurminder. (2025). *Obfuscation, Stealth, and Non-Attribution in Automated Red Team Tools*. *International Conference on Cyber Warfare and Security*. 20. 132-141. 10.34190/iccws.20.1.3290.
- [30] <https://www.offsec.com/blog/category/kali-linux/>
- [31] Tanner, N. (2019). *Kali Linux*.
- [32] Surarapu, Pavani & Baddam, Parikshith. (2023). *An Overview of Kali Linux: Empowering Ethical Hackers with Unparalleled Features*. 1. 171-180.
- [33] Simovic, Vladimir & Varga, Matija & Bevanda, Kristina. (2021). *How to use the Linux Kali operating system and ethical hacking*.
- [34] Tandon, Aradhya & Pandey, Siddharth. (2025). *A Study on Penetration Testing Using Metasploit*

Framework. International Journal of Research and Review in Applied Science, Humanities, and Technology. 10.71143/y4sten11.

- [35] Malkapurapu, Sivamanikanta & Abbas, Mohamed & Das, Pranjit. (2023). Exploring the Capabilities of the Metasploit Framework for Effective Penetration Testing. 10.1007/978-981-99-6755-1_35.
- [36] Athauda, I. (2023). *Penetration testing report for Metasploit*.
- [37] Shaikhanova, Aigul & Kadyrov, D. (2023). A Deep Dive into Cobalt Strike Tool. Bulletin of Shakarim University. Technical Sciences. 1. 46-52. 10.53360/2788-7995-2023-4(12)-7.
- [38] Khaver, A. (2023). Principles of operation and detection on the target system of the dual-purpose tool Cobalt Strike. Modern Information Security. 54. 10.31673/2409-7292.2023.020002.
- [39] Ottis, Rain. (2008). Analysis of the 2007 cyber attacks against Estonia from the information warfare perspective. 7th European Conference on Information Warfare and Security 2008, ECIW 2008. 163-168.
- [40] Delfrate, V. (2021). *2007 cyber attacks in Estonia: A case study*.
- [41] Arslan, H., & Canay, Ö. (2023). *Cyber threat intelligence systems and applications*.
- [42] Raval, Jinit & Pandya, Nitin & Soni, Jigar. (2025). Cyber Threat Intelligence (CTI): A Comprehensive Review of Automated Threat Intelligence Platforms, Dark Web Monitoring, and Threat Hunting. SPU Journal of Science Technology and Management Research. 2. 20-35. 10.63766/spujstmr.24.000029.
- [43] Kotenko, Igor & Saenko, Igor & Mityakov, Evgeniy. (2025). Multi-criteria assessment of information security threats based on the technologies of digital twins and threat intelligence. Journal Of Applied Informatics. 20. 100-120. 10.37791/2687-0649-2025-20-5-100-120.
- [44] Oakley, J. (2019). *Rules of engagement: Conducting successful cybersecurity engagements*. https://doi.org/10.1007/978-1-4842-4309-1_5
- [45] Cosic, Jasmin & Baca, Miroslav & Grd, Petra. (2018). *The Necessity of Developing a Standard for Exchanging a Chain of Custody of Digital Evidence Data (A DEMF STORY)*. International Journal of Computer Science and Information Security,. 15.
- [46] Tiwari, Neelesh & Baroniya, Yash & Bharadwaj, Vishwas & Kumar, Akash & Srivastava, Ankit. (2023). Cyber and Digital Forensic. 10.1007/978-981-99-1377-0_25.
- [47] Bharati, R. (2025). *Digital forensics and electronic evidence*.
- [48] <https://specterops.io/blog/2017/12/04/designing-effective-covert-red-team-attack-infrastructure/>
- [49] Alothman, B., Alhajraf, A., Alajmi, R., Farraj, R. A., Alshareef, N., & Khan, M. (2022). *Developing a cyber incident exercises model to educate security teams*. Electronics, 11(10), 1575. <https://doi.org/10.3390/electronics11101575>
- [50] Leite, Cristoffer & Hartog, Jerry & dos Santos, Daniel & Costante, Elisa. (2023). Actionable Cyber Threat Intelligence for Automated Incident Response. 10.1007/978-3-031-22295-5_20.
- [51] Blumbergs, B., Ottis, R., & Vaarandi, R. (2019). *Crossed swords: A cyber red team oriented technical exercise*. In Proceedings of the European Conference on Cyber Warfare and Security (pp. 37–44).

- [52] European Parliament & Council. (2022). Directive (EU) 2022/2555 on measures for a high common level of cybersecurity across the Union (NIS2). Official Journal of the European Union.
- [53] Ministerio de Asuntos Económicos y Transformación Digital. (2022). *Real Decreto 311/2022, de 3 de mayo, por el que se regula el Esquema Nacional de Seguridad*. Boletín Oficial del Estado
- [54] European Central Bank. (2018). *Threat intelligence-based ethical red teaming (TIBER-EU) framework*.
- [55] Zenko, M. (2015). *Red team: How to succeed by thinking like the enemy*. Basic Books.

ANEXO A: COLECCIÓN DE MÉTRICAS PROPUESTAS, INLUYENDO HOLLIS, 2024

KPI / Métrica	Abreviatura	Definición	Cálculo	Importancia
Detección por técnica	DT	Nivel de detección de una técnica o TTP específica empleada por el Red Team (0 no detectado, 1 detección humana, 2 detección automática).	$DT \in \{0,1,2\}$ a partir del cruce entre logs del Red Team y alertas del EDR/SIEM.	Base para evaluar la capacidad de detección y para el cálculo de la cobertura.
Cobertura de detección	Cobertura	Proporción normalizada de técnicas ejecutadas que fueron detectadas.	$Cobertura = (\sum DT_i) / (2N)$	Indicador sintético de eficacia global de detección.
Cobertura ponderada	$Cobertura_p$	Cobertura de detección ponderada según la criticidad del activo afectado.	$Cobertura_p = (\sum (DT_i \times w_i)) / (2 \sum w_i)$	Introduce contexto de riesgo priorizando activos críticos.
Median Time to Detect	MTTD	Mediana del tiempo entre la ejecución de una técnica y su detección.	$MTTD = \text{mediana}(t_{\text{detección}} - t_{\text{ejecución}})$	Mide la rapidez de detección del Blue Team.
Median Time to Respond / Contain	MTTR	Mediana del tiempo entre la detección y la contención inicial.	$MTTR = \text{mediana}(t_{\text{contención}} - t_{\text{detección}})$	Evalúa la capacidad operativa de respuesta.
Mean Time to Initial Access	MTTIA	Tiempo medio que tarda el Red Team en lograr acceso inicial.	$MTTIA = \text{mediana}(t_{\text{acceso_inicial}} - t_{\text{inicio}})$	Refleja la fortaleza de controles perimetrales.

Mean Time To Action	MTTA	Tiempo medio entre la detección y la implementación de una acción mitigadora efectiva.	$MTTA = \text{mediana}(t_{\text{implementación}} - t_{\text{detección}})$	Mide la capacidad de convertir detección en mitigación.
Time to Fix Remediation	TTFR	Tiempo entre el cierre del ejercicio y la verificación de la remediación.	$TTFR = t_{\text{verificación}} - t_{\text{cierre}}$	Indicador de madurez organizativa.
Porcentaje de técnicas con ingest mínima	Ingest %	Porcentaje de técnicas ejecutadas con telemetría mínima disponible.	$\text{Ingest} = 100 \times \frac{N_{\text{ingest_mínima}}}{N_{\text{ejecutadas}}}$	Garantiza cobertura real de la superficie de ataque.
Tasa de falsos positivos	TFP	Porcentaje de alertas que no corresponden a actividad maliciosa real.	$TFP = 100 \times \frac{N_{\text{falsos_positivos}}}{N_{\text{alertas_totales}}}$	Indica el nivel de ajuste y madurez del Blue Team.
Tasa de éxito tras retest	TER	Porcentaje de mitigaciones que superan un retest posterior.	$TER = 100 \times \frac{N_{\text{mitigaciones_aprobadas}}}{N_{\text{mitigaciones_probadas}}}$	Mide la efectividad real de las mejoras implementadas.

ANEXO B: CHECK LIST DE PREEJECUCIÓN DE UN CIBEREJERCICIO OFENSIVO U OPERACIÓN DE RED TEAM

A. Gobernanza, autorizaciones y legal

1. Autorización formal firmada (Documento de aprobación con alcance, fechas, responsables y firmas).
2. Reglas de compromiso (RoE) definidas y aprobadas (activos incluidos/excluidos, límites de impacto, técnicas prohibidas).
3. Análisis legal realizado y aprobado por asesoría jurídica (delitos informáticos, protección de datos, responsabilidades).
4. Acuerdos de confidencialidad (NDA) firmados por todos los participantes y observadores.
5. Notificación/Coordinación con CSIRT(s) relevantes (nacional/regional/proveedores) si procede.
6. Permisos de proveedores/terceros obtenidos cuando se vaya a interactuar con servicios externos.

B. Alcance, objetivos y planificación

1. Documento de alcance finalizado (escenario, objetivos operativos, criterios de éxito).
2. Listado de activos críticos incluidos y excluidos, con identificación única (IP/hostname/assetID).
3. Plan de pruebas y timelines (fechas/hora de inicio y fin, ventanas de mantenimiento).
4. Definición de métricas y KPIs a recoger durante el ejercicio.
5. Plan de rollback / restauración detallado y validado (snapshots, backups, puntos de restauración).

C. Roles, responsabilidades y comunicación

1. Matriz de responsabilidades publicada (Red, Blue, Purple, Dirección, Legal, Soporte).
2. Lista de contactos operativos y de escalado (nombre, rol, teléfono móvil, correo, horario).
3. Canal de comunicaciones principal y alternativo (war room, teléfono, mensajería segura).
4. Briefing pre-ejecución programado para todos los participantes.
5. Asignación de observadores/auditores externos (si aplica) y reglas para su intervención.

D. Seguridad física y aislamiento del entorno

1. Entorno de laboratorio aislado (segmentación lógica y/o física; confirmación de «no cross-connect» con producción).
2. Revisión de rutas de red y acceso a Internet controlado / bloqueado según RoE.
3. Dispositivos y accesos físicos controlados (salas, racks, credenciales temporales).

E. Preparación técnica e infraestructura

1. Snapshots / Backup completo de todos los sistemas objetivo antes de la prueba (con verificación de integridad).
2. Mecanismos de reversión probados (restauración de snapshot en un entorno de prueba).
3. Sincronización temporal (NTP) verificada en todos los equipos; zona horaria uniforme.
4. Inventario y versiones de herramientas, agentes y software (EDR, SIEM, OS, firmware).

5. Espacio de almacenamiento suficiente para pcaps, logs y artefactos forenses.
6. EDR / logging / agentes desplegados en hosts seleccionados y en modo reportado (no bloqueante salvo si se prueba respuesta activa).
7. SIEM / colector central configurado y recibiendo eventos (prueba de ingestión y parsers OK).
8. Monitorización del estado del laboratorio (CPUs, discos, estado red) lista y operativa.

F. Recolección de datos y evidencias

1. Política de retención y almacenamiento definida y aplicada (incluida custodia de evidencias).
2. Hashing y firma inicial de artifacts críticos (snapshots, imágenes, pcaps).
3. Plan de etiquetado y metadatos (escenario, ejecución, fase, actor) aplicado a cada archivo.
4. Prueba de integridad (hashes conciliados) tras captura de muestras de prueba.

G. Seguridad de datos y cumplimiento (GDPR / protección de datos)

1. Evaluación de datos personales: identificación y minimización/anonimización de cualquier dato real.
2. Consentimientos o bases legales para tratamiento si se usan datos personales reales.
3. Procedimiento de eliminación/anónimo para datasets tras el ejercicio.

H. Diseño de pruebas y casos de control

1. Plan de casos de prueba aprobado (lista de TTPs a ejecutar, objetivos concretos por TTP).
2. Test-cases de control (check smoke tests) definidos para verificar pipelines de recolección y detección.
3. Lista de IOCs de prueba y su momento de inyección (para validar detección).
4. Lista de técnicas prohibidas y umbrales de impacto nunca superar.

I. Seguridad operacional durante la ejecución

1. Interrupción de emergencia implementado y probado (procedimiento para detener la ejecución completa).
2. Procedimiento de escalado ante incidente real detectado fuera del ejercicio.
3. Ventana de mantenimiento confirmada con servicios dependientes (si aplica).
4. Plan de notificación externa preparado (qué comunicar, a quién y cuándo).

J. Herramientas, scripts y reproducibilidad

1. Repositorios (Git) con scripts y artefactos listos y versionados; tag de release para el ejercicio.
2. Documentación de ejecución (playbook, comandos, parámetros) revisada y disponible.
3. Prueba previa (dry-run) de scripts críticos en entorno no productivo.

K. Formación y salud del equipo

1. Comprobación de competencia de operadores (años/experiencia, certificaciones mínimas o formación interna).
2. Briefing en seguridad y ética realizado (resume RoE y límites).
3. Roles de sustitución asignados en caso de ausencia.

L. Post-ejecución y cierre planificado

1. Agenda de debriefing y cronograma de entregables (informes, KPIs, recomendaciones).
2. Plan de limpieza (borrado de artefactos, cierre de cuentas temporales, eliminación de persistencia).

-
3. Verificación de restauración (comprobación de que servicios han vuelto al estado pre-prueba).
 4. Plan de revisión y lecciones aprendidas programado (incluye validación por terceros/expert review).

M. Go / No-Go — Lista mínima para autorización final (checklist resumen)

1. Autorización legal y RoE firmados.
2. Snapshots/Backups realizados y verificados.
3. NTP sincronizado y logs preparados.
4. SIEM y sensores recibiendo eventos (prueba OK).
5. Kill-switch definido y probado.
6. Canales de comunicación y contactos de emergencia confirmados.
7. Plan de anonimización y tratamiento de datos aprobado.

ANEXO C: GLOSARIO DE TÉRMINOS Y EXPRESIONES EN INGLÉS

Acrónimos

1. **APT (Advanced Persistent Threat)**: Amenaza persistente avanzada; actores sofisticados, a menudo patrocinados por estados, que mantienen accesos prolongados y sigilosos.
2. **ASLR (Address Space Layout Randomization)**: Técnica de mitigación que aleatoriza direcciones de memoria para dificultar exploits.
3. **ATT&CK (Adversarial Tactics, Techniques and Common Knowledge)**: Base de conocimiento de MITRE que cataloga tácticas y técnicas de adversarios.
4. **BOF (Buffer Overflow)**: Desbordamiento de búfer; vulnerabilidad de memoria.
5. **CBEST**: Marco británico de red teaming basado en inteligencia, promovido por el Bank of England.
6. **C2 / C&C (Command and Control)**: Infraestructura de mando y control usada por atacantes para gestionar sistemas comprometidos.
7. **CDX (Cyber Defence Exercise)**: Ejercicio de ciberdefensa.
8. **CERT / CSIRT (Computer Security Incident Response Team)**: Equipo de respuesta a incidentes de seguridad.
9. **CIF (Critical Important Function)**: Función crítica o importante de una organización.
10. **CWE (Common Weakness Enumeration)**: Clasificación de debilidades de software y hardware.
11. **CVE (Common Vulnerabilities and Exposures)**: Identificador de vulnerabilidades específicas.
12. **DEP (Data Execution Prevention)**: Mitigación que evita ejecución de código en regiones de memoria no ejecutables.
13. **DNS (Domain Name System)**: Sistema de nombres de dominio.
14. **EDR (Endpoint Detection and Response)**: Herramienta de detección y respuesta en endpoints.
15. **ENISA (European Union Agency for Cybersecurity)**: Agencia europea de ciberseguridad.
16. **FFRDC (Federally Funded Research and Development Center)**: Centros de I+D financiados por el gobierno de EE. UU.
17. **GDPR (General Data Protection Regulation)**: Reglamento General de Protección de Datos.
18. **HTTP / HTTPS (Hypertext Transfer Protocol / Secure)**: Protocolos de comunicación web.
19. **ICS (Industrial Control Systems)**: Sistemas de control industrial.
20. **IOC / IOCs (Indicator(s) of Compromise)**: Indicadores de compromiso.
21. **ISO/IEC**: Organismos internacionales de normalización.
22. **IT / TIC (Information Technology / Tecnologías de la Información y las Comunicaciones)**.
23. **KPI / KPIs (Key Performance Indicator(s))**: Indicadores clave de desempeño.
24. **MITRE**: Organización sin ánimo de lucro que mantiene marcos como ATT&CK.

-
25. **MSF (Metasploit Framework)**: Framework de explotación.
 26. **NDA (Non-Disclosure Agreement)**: Acuerdo de confidencialidad.
 27. **NIS**: Directiva europea de seguridad de redes y sistemas de información.
 28. **OSINT (Open Source Intelligence)**: Inteligencia de fuentes abiertas.
 29. **OT (Operational Technology)**: Tecnología operacional.
 30. **PRA (Prudential Regulation Authority)**: Autoridad reguladora británica.
 31. **RoE / ROE (Rules of Engagement)**: Reglas de enfrentamiento/compromiso.
 32. **SCADA (Supervisory Control and Data Acquisition)**: Sistemas de supervisión y control industrial.
 33. **SIEM (Security Information and Event Management)**: Gestión de eventos e información de seguridad.
 34. **SOCMINT (Social Media Intelligence)**: Inteligencia de redes sociales.
 35. **STIC**: Guías y estándares técnicos del CCN (España).
 36. **TIBER-EU (Threat Intelligence-Based Ethical Red Teaming – EU)**: Marco europeo de red teaming basado en inteligencia.
 37. **TISP (Threat Intelligence Service Provider)**: Proveedor de servicios de inteligencia de amenazas.
 38. **TTI (Target Threat Intelligence)**: Inteligencia de amenazas orientada al objetivo.
 39. **TTP / TTPs (Tactics, Techniques and Procedures)**: Tácticas, técnicas y procedimientos.
 40. **VPS (Virtual Private Server)**: Servidor privado virtual.

Expresiones y términos en inglés

41. **Blue Team**: Equipo defensivo encargado de detección, respuesta y mitigación.
42. **Cyber Kill Chain**: Modelo que describe las fases de un ciberataque.
43. **Desk Check / Comms Check / Walk-through / Table-top / Full Simulation**: Tipos de ejercicios de ciberseguridad.
44. **Engagement**: Enfrentamiento o ejecución de un ejercicio.
45. **Exploit**: Código o técnica que aprovecha una vulnerabilidad para obtener un comportamiento no autorizado del sistema (por ejemplo, ejecución de código o escalada de privilegios).
46. **Framework**: Marco de referencia metodológico.
47. **Kick-off call**: Reunión inicial/final de coordinación previa al ejercicio.
48. **Pentest / Penetration Testing**: Prueba de penetración.
49. **Playbook**: Guía operativa estandarizada.
50. **Purple Team / Purple Teaming**: Colaboración estructurada entre Red y Blue Team.
51. **Red Team / Red Teaming**: Equipo y metodología de emulación de adversarios.
52. **Retest**: Nueva prueba tras aplicar medidas de remediación.
53. **Roadmap**: Hoja de ruta.
54. **Scoring**: Sistema de puntuación del ejercicio.
55. **Storyline**: Narrativa del escenario del ejercicio.

- 56. **Threat Intelligence:** Inteligencia de amenazas.
- 57. **Timeline:** Línea temporal de eventos.
- 58. **Tooling / Toolkit:** Conjunto de herramientas.
- 59. **Unified Kill Chain:** Variante extendida de la Cyber Kill Chain.
- 60. **Zero Day:** Vulnerabilidad desconocida para el fabricante y sin parche disponible en el momento de su explotación.