RPT-02 // INFORME DE METODOLOGÍA

> DE IOC CRUDO A PUNTUACIÓN DE RIESGO

Cómo el pipeline News CTI convierte un indicador en bruto en una decisión accionable: un modelo de confianza en dos dimensiones — cuánto fiarse del indicador y cuánto daño haría si es real — con validación externa, decaimiento temporal y un caso real del propio panel en producción.

News CTI	OP-01	TLP:CLEAR	IOCs defangueados
SISTEMA	OPERACIÓN	DIFUSIÓN	SEGURIDAD

00 // PLANTEAMIENTO

> POR QUÉ UNA CIFRA NO BASTA

Los feeds públicos entregan miles de indicadores al día, pero un indicador aislado no dice al analista lo que necesita: **¿me lo creo?** y **¿me debería preocupar?** Son dos preguntas distintas y colapsarlas en un solo número pierde información.

Un ejemplo lo deja claro: una URL de *payload* vista por tres fuentes y confirmada por VirusTotal es **muy fiable**, pero si solo sirve un troyano genérico su impacto es medio. En cambio, una única mención antigua de un servidor de *Cobalt Strike* es **poco fiable** como dato, pero si resulta cierta el impacto es crítico. Ordenar ambos por el mismo número los confunde.

Por eso News CTI puntúa cada IoC en dos ejes independientes:

- **Score de confianza (0–100)** — mide *cuánto fiarse del indicador*, combinando la señal del propio feed con validación externa y frescura.
- **Gravedad (crítica / alta / media / baja)** — mide *el impacto de la amenaza si es real*, según la familia de malware.

01 // EJE 1 — FIABILIDAD

> ANATOMÍA DEL SCORE DE CONFIANZA

Cada IoC parte de la señal de su feed de origen y suma puntos por cada validación independiente que lo corrobora. La lógica de fondo: **una sola fuente es una pista; varias fuentes de acuerdo son una señal**.

COMPOSICIÓN DEL SCORE (0–100)

SEÑAL	APORTE	RACIONAL
Fuente base	`40 / 25`	ThreatFox y URLhaus aportan 40; OTX 25. Se toma el máximo, no la suma: refleja la fiabilidad del feed más sólido que vio el indicador.
Multi-fuente	`+10 × extra`	+10 por cada fuente adicional que corrobora, hasta +20. La corroboración independiente es la señal más fuerte de que el IoC es real.
Confianza del feed	`hasta +15`	El campo `confidence` del feed × 0,15. Aprovecha la valoración que el propio proveedor ya asigna.
AbuseIPDB	`hasta +20`	`abuseConfidenceScore` × 0,20. Reputación de IP validada por una comunidad externa.
VirusTotal	`hasta +25`	Ratio de detecciones × 25. El consenso de decenas de motores antivirus es la validación de mayor peso.

Sobre ese total se aplica un **decaimiento por antigüedad**: un indicador reciente pesa más que uno de hace semanas, porque la infraestructura de los atacantes rota rápido. El resultado se traduce a tres niveles operativos:

- `ALTA` score ≥ 70 — actuar / bloquear con confianza.
- `MEDIA` 40–69 — contexto útil, requiere correlación.
- `BAJA` < 40 — ruido probable, vigilar sin priorizar.

Ejemplo resuelto

Un IoC ilustrativo — una IP:puerto de C2 vista por ThreatFox y OTX, con reputación confirmada en AbuseIPDB y detecciones en VirusTotal — se puntuaría así:

45[.]134[.]26[.]11:8443	EJEMPLO ILUSTRATIVO	
Fuente base — ThreatFox (máx. sobre OTX)		+40
Multi-fuente — 1 fuente adicional (OTX)		+10
Confianza del feed — 75 × 0,15		+11
AbuseIPDB — 90 × 0,20		+18
VirusTotal — ratio 0,32 × 25		+8
SCORE DE CONFIANZA		87 · alta

Las cifras del ejemplo son ilustrativas para mostrar la mecánica; los pesos exactos viven en `enrich.py` y los resultados reales se recalculan cada 6 h en el pipeline.

---

02 // EJE 2 — IMPACTO

> GRAVEDAD: EL DAÑO SI ES REAL

La gravedad es **independiente del score**. No pregunta si el indicador es fiable, sino qué ocurriría si lo es. Se deriva de la familia de malware asociada, con listas que se amplían conforme aparecen familias nuevas en los feeds.

NIVELES DE GRAVEDAD Y CRITERIO

NIVEL	CRITERIO
CRÍTICA	Ransomware (LockBit, Akira, RansomHub…) y frameworks C2 (Cobalt Strike, Sliver, Havoc, AdaptixC2…).
ALTA	RATs, stealers, loaders y botnets — familia conocida o categoría genérica en el nombre («X Stealer», «Unknown RAT»…).
MEDIA	Resto de amenazas identificadas, o desconocidas con ratio de detecciones VT ≥ 0,3.
BAJA	Sin familia identificada ni señal externa.

Las listas se mantienen en `SEV_CRITICA`, `SEV_ALTA` y `SEV_ALTA_GENERICAS` dentro de `enrich.py`. Separar la clasificación en constantes mantiene el criterio auditable y fácil de extender.

03 // LOS DOS EJES EN ACCIÓN

> DOS IOCS REALES DEL PANEL

El valor de separar los ejes se ve en indicadores reales del dashboard. Estos dos, capturados en una actualización del pipeline, se ordenarían al revés según qué eje mires:

SCORE ALTO ≠ GRAVEDAD ALTA

IOC	AMENAZA	SCORE	GRAVEDAD
`130[.]12[.]180[.]51:43782`	RedTail	80 · ALTA	MEDIA
`36[.]140[.]162[.]173:12443`	Cobalt Strike	59 · MEDIA	CRÍTICA

Por **fiabilidad**, el minero RedTail va primero: mejor corroborado, score 80. Por **impacto**, el servidor de Cobalt Strike es el que quita el sueño pese a su score 59, porque un framework de C2 en tu red es un compromiso activo. Un único número habría enterrado el segundo bajo el primero. Con dos ejes, el analista puede filtrar por *«alta gravedad aunque el score sea medio»* — justo la cola de indicadores que más importa no pasar por alto.

04 // VALIDACIÓN EXTERNA

> ENRIQUECIMIENTO Y SUS LÍMITES

El score sólo es tan bueno como sus fuentes de validación, pero las APIs gratuitas tienen cuotas. El diseño reparte un presupuesto por ejecución y cachea resultados para no malgastarlo:

- **GeoIP** — base DB-IP Country Lite, *offline* y sin clave: cubre el 100 % de las IPs sin límite.
- **VirusTotal** — 4 req/min · 500/día: 40 *lookups* por ejecución con pausa de 15,5 s para respetar el límite.
- **AbuseIPDB** — 1.000 *checks*/día: 150 IPs por ejecución.
- **Caché de 7 días** en `ioc_state.json`, priorizando siempre lo nuevo.

Todo el enriquecimiento ocurre en el *backend*: el dashboard solo lee `iocs_latest.json` y **nunca hace llamadas externas**. Y el sistema degrada con elegancia — sin claves opcionales, sigue puntuando con fusión multi-fuente + confianza del feed + frescura + gravedad + GeoIP. Los presupuestos son ajustables por variables de entorno (`VT_BUDGET`, `ABUSEIPDB_BUDGET`, `CTI_RECHECK_DAYS`…).

> **Decisión de arquitectura:** enriquecer en el *backend* y publicar solo JSON estático significa que el panel no expone ninguna clave de API ni depende de que el navegador del visitante alcance servicios de terceros. Rendimiento, privacidad y superficie de ataque mínima.

---

05 // HONESTIDAD TÉCNICA

> QUÉ MEJORARÍA

Un modelo útil no es un modelo terminado. Las líneas abiertas:

- **Calibración de pesos** — los coeficientes actuales son razonados, no ajustados con datos etiquetados. Contrastarlos contra verdad-terreno afinaría los umbrales.
- **Gravedad por comportamiento** — hoy depende del nombre de familia; mapear a técnicas MITRE ATT&CK daría una señal de impacto independiente del *naming* del feed.
- **Exportador a detección** — convertir los IoCs de mayor score en reglas listas para SIEM cierra el ciclo de CTI a defensa (evolución ya prevista).
