L'écart de promotion. Pourquoi une mitigation publique reste un artefact tant qu'elle n'a pas été éprouvée sur la cible réelle"

Article VI. Le cas CVE-2026-31431 (Copy Fail) sur kernel WSL2 6.6.87.2 comme terrain empirique du port de promotion entre artefact technique et dépendance normative.

Une recommandation publique de sécurité est un artefact dont le périmètre éprouvé ne coïncide pas mécaniquement avec le périmètre des cibles industrielles auxquelles elle est prescrite. Le franchissement du port de promotion vers la dépendance normative ne peut pas être délégué à l'autorité émettrice. Démonstration sur quatre mouvements à partir d'un cas observé le 30 avril 2026.

Points saillants :

- "Une mitigation CERT-EU est un artefact. Sa promotion en dépendance normative passe par une validation comportementale sur la cible exacte."
- "L'écart de promotion désigne la distance mesurable entre le périmètre éprouvé par l'autorité et le périmètre prescrit à l'industrie."
- "Sur kernel WSL2 6.6.87.2, quatre mitigations publiques ont été comportementalement invalidées dans la même session opérationnelle."
- "La trace effective au sens kernel, accessible via /proc/<pid>/status, est le seul invariant qui distingue mitigation effective et mitigation cosmétique."

Introduction

L'Article V de cette série a posé l'architecture hexagonale comme cadre de gouvernance pour les phases de training en IA appliquée aux domaines régulés. Il y introduisait trois concepts opératoires :

- Le *port de promotion* comme frontière entre artefact expérimental et dépendance normative,
- La *délibération traçable* comme procédure documentée du franchissement,
- La *dette généalogique* comme passif que tout artefact emporte avec lui quand on le promeut sans avoir tracé les conditions du passage.

Ces concepts ont été formulés à partir du terrain d’expérimentation ToxTwin, dans une trajectoire interne, sous contrôle de l'auteur. Ils méritaient un test sur un terrain externe, où l'autorité émettrice n'est pas l'auteur, où la cadence d'application est imposée par la communauté, et où les conditions de validation ne sont pas négociables.

Ce terrain s'est présenté le 30 avril 2026, sous la forme de la CVE-2026-31431 dite Copy Fail, une élévation de privilège locale dans le sous-système crypto Linux exploitable par 732 octets de code Python depuis tout compte non privilégié.

La thèse tient en une phrase. ***Une mitigation publique de sécurité reste un artefact tant qu'elle n'a pas été éprouvée comportementalement sur la cible exacte où elle est prescrite, et le franchissement du port de promotion vers la dépendance normative ne peut pas être délégué à l'autorité émettrice.***

Le domaine de validité est circonscrit. L'argument concerne les recommandations de mitigation issues d'autorités publiques de cybersécurité appliquées à des environnements de production hétérogènes, et il est démontré sur le cas particulier du kernel WSL2 6.6.87.2-microsoft-standard-WSL2 face aux recommandations CERT-EU et openwall pour Copy Fail.

Il se généralise aux mitigations publiques appliquées à toute variante kernel non explicitement testée par l'autorité émettrice. Il ne concerne ni la conduite générale de la sécurité informatique, ni la doctrine de patch management, qui ont leurs propres cadres établis. La démonstration suit quatre mouvements : l'artefact CERT-EU dans son périmètre revendiqué, la preuve empirique de son inopérance hors de ce périmètre, la mitigation alternative validée sur la cible réelle, le critère de promotion qui en découle.

Mouvement 1. L'artefact CERT-EU et son périmètre éprouvé

Le 29 avril 2026, l'équipe Theori publiait la disclosure coordonnée de CVE-2026-31431. Une optimisation introduite en 2017 dans le module algif_aead.c du sous-système crypto Linux permettait à un utilisateur non privilégié d'écrire quatre octets contrôlés dans le page cache d'un fichier readable, en combinant la socket AF_ALG, la primitive splice(), et les particularités du template AEAD authencesn. La modification du page cache d'un binaire setuid suffit à obtenir root sur la prochaine invocation. Le proof-of-concept public, distribué sous forme d'un script Python de 732 octets ne requérant aucun offset kernel ni race window, fonctionnait sur Ubuntu 24.04 LTS, Amazon Linux 2023, RHEL 10.1 et SUSE 16. Le score CVSS de 7.8 sous-estimait probablement l'impact opérationnel, dont le marché gris reconnaît habituellement la valeur d'un appartement parisien.

Dans la journée du 29 avril, plusieurs autorités publiaient leurs recommandations de mitigation en attendant les patches kernel des distributions. CERT-EU, openwall et la communauté kernel.org convergeaient sur la même instruction : neutraliser le module algif_aead par une directive modprobe install algif_aead /bin/false, accompagnée d'un rmmod algif_aead immédiat pour vider la mémoire. La recommandation était techniquement sensée. Elle visait à rendre indisponible le handler crypto qui exposait le primitif d'écriture, sans toucher aux services userspace utilisant les couches supérieures (dm-crypt, kTLS, IPsec, OpenSSL).

Le périmètre éprouvé de cette recommandation, tel qu'il pouvait être reconstitué à partir des avis publiés, comprenait les distributions explicitement listées dans la disclosure Theori et un ensemble vraisemblable de configurations standard sur kernels distribution officiels. L'autorité ne revendiquait pas une validation universelle. Elle ne pouvait pas le faire, et c'est rationnel. Aucune autorité publique de cybersécurité n'a les moyens de tester une recommandation sur l'ensemble combinatoire des variantes kernel, distribution, version, architecture, déployées dans l'industrie. Les autorités testent sur les cibles dominantes, publient sur les cibles dominantes, et laissent à chaque organisation le soin de vérifier la portabilité sur les variantes périphériques.

Le périmètre prescrit, lui, était universel. La recommandation CERT-EU était formulée sans qualification, applicable à toute distribution Linux livrée depuis 2017. Cette asymétrie entre périmètre éprouvé borné et périmètre prescrit illimité est la condition même de l'artefact public. Elle n'est pas un défaut de l'autorité émettrice, c'est sa condition de fonctionnement à l'échelle. Mais elle a une conséquence opérationnelle directe : la recommandation prescrite à un parc industriel hétérogène contient nécessairement un *écart de promotion* non nul, dont l'amplitude n'est connue de personne au moment de l'application. La doctrine implicite veut que l'autorité publie, que l'industrie applique, et que le silence sur les variantes signifie qu'elles tiennent. Sur ce dernier point, la doctrine implicite a tort plus souvent qu'on ne le croit.

Mouvement 2. La preuve empirique d'inopérance sur kernel WSL2 6.6.87.2

L'environnement d'un Windows Server 2025 avec WSL2 sur lequel j'ai conduit l'investigation fait tourner Ubuntu 24.04 sur kernel WSL2 6.6.87.2-microsoft-standard-WSL2, hébergeant des données cliniques pseudonymisées et une stack ML active. Sept comptes locaux y coexistent, dont six non administratifs. ***Le profil correspond exactement à la cible que la recommandation CERT-EU est censée couvrir : Ubuntu 24.04 LTS, multi-utilisateurs, données régulées, urgence d'application avant disponibilité d'un patch kernel.***

La recommandation a été appliquée selon les instructions publiées : création de /etc/modprobe.d/disable-algif-aead.conf avec la directive d'install, déchargement du module si chargé. Une validation comportementale a été conduite en parallèle, consistant à reproduire le primitif d'amorçage de l'exploit dans un script Python contrôlé : ouverture d'une socket AF_ALG, bind sur le template authencesn(hmac(sha256),cbc(aes)), mesure de la valeur retournée par le kernel. Le test a invalidé la mitigation. Le bind réussissait. Le primitif d'attaque restait disponible.

La cause technique a été identifiée par instrumentation. Le kernel WSL2, sur ce build précis, utilise un chemin de chargement de modules qui fait appel à request_module() côté noyau lorsqu'un userspace demande un service crypto AF_ALG, et ce chemin

contourne la résolution modprobe userspace dans laquelle la directive d'install avait été inscrite. Le module est rechargé à la demande, malgré la blacklist.

Quatre variantes successives de la recommandation ont été testées dans la même session opérationnelle, chacune plus restrictive que la précédente.

1. a blacklist du framework af_alg complet, qui couvrait af_alg, algif_aead, algif_skcipher, algif_hash, algif_rng, a échoué pour la même raison.
2. La désactivation kernel-level du chargement de modules par sysctl kernel.modules_disabled=1 a produit un piège opérationnel particulier : le verrou interdit également le déchargement, ce qui a gelé en mémoire les modules déjà chargés par les tests précédents.
3. La suppression physique des fichiers .ko a été bloquée par le verrou sysctl en cours d'opération.
4. Une cinquième tentative, qui aurait nécessité un redémarrage de la machine virtuelle WSL2 par commande PowerShell sur l'hôte Windows, a échoué pour des raisons humaines triviales que je m'épargne de détailler.

Le fait empirique brut est que sur le kernel WSL2 6.6.87.2, dans la session du 30 avril 2026, quatre variantes de la recommandation publique ont été comportementalement invalidées par le même test syscall. Aucune ne neutralisait le primitif d'attaque. Sans la validation comportementale conduite en parallèle de l'application, l'incident aurait été clos en quatre tickets de conformité déclarative cumulant chacun la satisfaction d'avoir appliqué la directive officielle. La distance entre *avoir appliqué* et *avoir mitigé* a été, dans ce cas particulier, de quatre itérations.

Cette distance n'est pas anecdotique. Elle est la mesure locale de l'écart de promotion entre le périmètre où la recommandation a été éprouvée par l'autorité et le périmètre où elle est prescrite à l'industrie. Microsoft annonçait en 2025 plus de quatre millions de WSL2 en usage entreprise. Aucun de ces environnements ne figure dans les avis officiels CERT-EU ou openwall. Tous sont, au moment de leur application de la recommandation, dans la situation où la conformité déclarative coïncide avec la vulnérabilité comportementale. La probabilité que tous présentent le même comportement non standard que celui observé sur 6.6.87.2 reste à établir. La probabilité qu'au moins une fraction non négligeable y présente un comportement comparable est, elle, très élevée. Si la doctrine de mitigation publique tenait silencieusement sur ces quatre millions de variantes, la convergence serait remarquable. Elle ne l'est pas. Elle est seulement supposée.

Mouvement 3. La mitigation alternative empirique et sa validation comportementale

Une fois acquis que la recommandation publique ne tenait pas sur la cible réelle, deux trajectoires étaient ouvertes. La première consistait à attendre le patch kernel WSL2

publié par Microsoft, donc à laisser la fenêtre d'exposition ouverte sur une durée non maîtrisée, peu compatible avec la nature des données hébergées. La seconde consistait à concevoir une mitigation alternative à partir des couches kernel disponibles, indépendamment de la doctrine modprobe défaillante.

La voie retenue mobilise le mécanisme seccomp BPF, intégré au kernel Linux depuis 3.5 et standardisé par l'industrie depuis Docker 1.10. Un filtre BPF chirurgical refuse le syscall socket(AF_ALG, ...) au niveau du kernel, en retournant EAFNOSUPPORT (errno 97), sans toucher à aucun autre vecteur. Le primitif d'amorçage de Copy Fail est mort à l'entrée. Le filtre est attaché aux processus de connexion des comptes non privilégiés via un shell-wrapper substitué dans /etc/passwd, ce qui rend l'application irrévocable au sein d'une session. Le compte administrateur est volontairement exempté pour les besoins d'investigation et de debug, créant une asymétrie kernel-level entre les comptes utilisateurs et les comptes opérationnels.

La validation comportementale finale repose sur une commande reproductible. Pour chaque compte non privilégié, exécution d'un script Python qui tente la création de la socket AF_ALG et capture la valeur retournée. Six comptes utilisateurs retournent PROTECTED (errno=97). Le compte administrateur retourne VULNERABLE, c'est-à-dire un socket valide. La preuve la plus directe de l'attachement du filtre passe par la lecture de /proc/<pid>/status, où l'invariant Seccomp: 2 indique le mode FILTER actif. Cette information n'est pas un message imprimé par le shell. C'est l'état du noyau qui se décrit lui-même. Elle ne peut pas mentir par défaut de quoting ou par redirection silencieuse, qui sont les pathologies habituelles des traces performatives produites par les scripts de validation cosmétique. *Le kernel ne ment pas. Le shell, parfois, oui.*

La mitigation tient. Elle a été éprouvée sur la cible exacte, par un test comportemental qui reproduit le primitif d'attaque, et la trace effective au sens kernel confirme l'attachement du filtre. Aucun avis officiel ne la mentionnait à la date de l'incident. Elle a été produite par opération empirique locale, en réponse à l'inopérance documentée des recommandations publiques. Elle constitue, par construction, un artefact dont le périmètre éprouvé est exactement le périmètre où elle est appliquée. Le port de promotion entre artefact et dépendance normative a été franchi par la même opération qui a produit l'artefact. C'est ce qui distingue cette mitigation des recommandations publiques : elle ne porte pas d'écart de promotion résiduel.

Mouvement 4. Le critère de promotion. L'écart de promotion comme concept opératoire

Le cas Copy Fail rend explicite ce qui restait implicite dans l'Article V. Le port de promotion n'est pas seulement une procédure de gouvernance interne pour les artefacts produits dans l'organisation. Il est un dispositif vérificationnel dont la nécessité s'étend à tout artefact reçu de l'extérieur, y compris quand l'extérieur est une autorité publique

de référence. Cette extension introduit un concept nouveau, dans le prolongement direct du port de promotion, que je propose d'appeler *l'écart de promotion*.

L'écart de promotion désigne la distance mesurable entre le périmètre où un artefact a été éprouvé par son autorité émettrice et le périmètre où il est prescrit comme dépendance normative. Cet écart n'est pas un défaut de doctrine. Il est sa condition d'existence. Toute recommandation publique a un écart de promotion non nul, parce qu'aucune autorité ne peut éprouver son artefact sur l'ensemble combinatoire des cibles industrielles. La question opérationnelle n'est pas de réduire l'écart à zéro, ce qui est impossible, mais de le rendre *mesurable* et *attribuable*. Mesurable signifie que l'organisation qui applique la recommandation sait préciser la distance entre sa propre cible et les cibles éprouvées par l'autorité, et qu'elle conduit, pour les variantes non couvertes, une validation comportementale dont le résultat est documenté. Attribuable signifie que la responsabilité du franchissement du port de promotion est assignée à un acteur qui a les moyens techniques et juridiques de l'assumer. Cette responsabilité ne peut pas être déléguée à l'autorité émettrice, qui n'a ni accès à la cible ni mandat pour la couvrir. Elle ne peut pas non plus être diluée dans la chaîne d'audit, qui ne mesure que la conformité déclarative. Elle reste à la charge de l'organisation qui exploite l'environnement, et seulement d'elle.

Le critère de promotion se formule alors ainsi. *Un artefact reçu de l'extérieur n'est promu en dépendance normative pour une cible donnée qu'après production d'une trace de validation comportementale sur cette cible exacte, attestée par un invariant kernel ou un équivalent observable de niveau équivalent.* Cette formulation est exigeante. Elle l'est délibérément. Elle reconnaît que la conformité déclarative reste utile pour l'audit, sans accepter qu'elle se substitue à la sécurité. Elle reconnaît également que l'établissement systématique d'un dossier de validation comportementale par variante représente une dépense d'ingénierie qui n'est pas négligeable, et qui doit être proportionnée à l'écart de promotion estimé. Pour les organisations dont le parc informatique est homogène et standard, l'écart est faible et la conformité déclarative reste suffisante. Pour les domaines régulés exploitant des stacks hétérogènes, l'écart est structurellement élevé et le port de promotion devient une obligation opérationnelle, indissociable de l'obligation réglementaire.

L'objection prévisible à ce critère est qu'il déplace la charge depuis les autorités émettrices vers les organisations utilisatrices, en transformant une recommandation publique en travail local. L'objection est juste, et c'est précisément le point. La charge n'a jamais été ailleurs. L'illusion qu'elle l'était venait du silence sur les variantes, silence que la doctrine implicite interprétait comme couverture. Le critère de promotion ne déplace pas la charge, il rend visible la charge qui était déjà là.

Conclusion

Quatre mitigations publiques inopérantes, une mitigation empirique validée, un concept opératoire dégagé. La séquence n'est pas accidentelle. Elle reproduit, sur un terrain externe et sous contrainte temporelle, ce que l'Article V a posé en théorie pour les phases de training. Le port de promotion entre artefact et dépendance normative est un dispositif que l'industrie ne peut pas externaliser. Il n'a pas de propriétaire structurel. Il est à la charge de l'organisation qui exploite la cible, parce que c'est elle qui en porte les conséquences. Cette répartition n'est pas un défaut de l'écosystème de cybersécurité publique. C'est sa condition de fonctionnement à l'échelle.

Pour les domaines régulés, où l'écart de promotion mal géré devient un risque réglementaire avant d'être un risque technique, cette condition n'est pas optionnelle. Elle est l'expression opérationnelle de ce que la gouvernance demande à la sécurité. *Une recommandation publique est un artefact jusqu'à preuve qu'elle tient sur la cible exacte. La preuve n'incombe pas à l'émetteur. Elle incombe à celui qui exploite.*