Calibrating PAIN Without Abandoning CVSS: High-Centered Normalization and Standards-Anchored Thresholds

Matthew Venne
Chief Technology Officer, stackArmor

July 2026

Abstract

The initial paper *A Deterministic, CVSS–Environmental Method for FedRAMP VDR/VER Vulnerability Prioritization* introduced a reproducible method for deriving Potential Agency Impact (PAIN) from a vulnerability's confidentiality, integrity, and availability impacts and an asset's CVSS Environmental Security Requirements. Operational backtesting exposed a calibration defect: the method normalized against the CVSS v3.1 Medium-requirement impact ceiling while High requirements could produce a larger pre-cap aggregate. A single High technical impact on a High-requirement dimension therefore normalized to approximately 0.918 and crossed directly into Debilitating. In one scan, approximately 40 percent of affected-resource observations were N4.

This paper documents the alternatives evaluated and the resulting revision. More asset sub-profiles could reduce severity but would multiply tagging choices. Lowering the High requirement multiplier to 1.25 approached the desired distribution but discarded a CVSS-defined constant. Boolean component gates expressed the intended compound-impact semantics but added a second classifier and discarded some continuous dimensional information. Raising a threshold alone treated the symptom while leaving the normalization axis centered on Medium requirements.

The selected method retains the CVSS impact weights, Security Requirement weights, and complement-product aggregation. It replaces the saturated 0.915 normalization with the maximum attainable pre-cap aggregate under High requirements, $D_H = 0.995904$. Rather than carrying forward the original lower thresholds, Narrow begins at the first complete technical loss on a Low-requirement dimension and Disruptive begins at a High impact on a Moderate-requirement dimension. The Debilitating threshold is set to 0.933, which lies strictly between the highest possible non-compound combination (0.931993) and the lowest possible compound catastrophic combination (0.933423). Exhaustive testing of all 729 discrete impact/requirement combinations and backtesting against two anonymized evaluation datasets support the result. All three boundaries are therefore anchored to stated consequence scenarios and empirically tested, not fitted to a desired histogram. The compound condition is a transparent proxy for the breadth of the vulnerability's potential effect, not a CWE-based vulnerability classifier. The resulting thresholds are the reference calibration used by this method; another CSP may adopt different governed values after defining its semantics, re-enumerating the state space, validating the consequences, and disclosing the policy version.

Contents

1 Status and relationship to the initial method **2**

2 The calibration problem **3**

3 Design criteria **4**

4 Alternatives evaluated **4**
4.1 Additional context modifiers 4
4.2 Changing the High multiplier from 1.50 to 1.25 5
4.3 Raising the original Debilitating threshold 6
4.4 Restoring round thresholds with another normalization 6
4.5 Strict two-dimension alignment gate 7
4.6 Compound component gate 7
4.7 CVSS v4 MacroVectors 7

5 High-centered normalization **7**

6 Deriving the thresholds **8**
6.1 Deriving the Minimal/Narrow and Narrow/Disruptive boundaries 8
6.2 Deriving Debilitating from the discrete state space 9
6.3 Reference calibration and provider variation 11

7 Ceiling-constrained reachability **12**
7.1 Low- and Moderate-impact ceilings 12
7.2 What a High ceiling does not prove 12

8 Worked scenarios **13**

9 Empirical backtest **14**
9.1 The threshold remains reachable 15
9.2 What the backtest does and does not prove 16

10 Why the final method is preferable **16**

11 Implementation and governance **16**
11.1 Normative parameter set 16
11.2 Precision and reproducibility 17
11.3 Migration 17

12 Limitations and future work **17**

13 Conclusion **18**

References **19**

1 Status and relationship to the initial method

This document is an informational calibration companion to the initial PAIN publication. It does not define a FedRAMP requirement, change a CVSS score, or claim that the resulting scalar is a conformant CVSS Environmental Score. It revises the PAIN-specific step that transforms CVSS Environmental impact inputs into a unit scalar and then into the words Minimal, Narrow, Disruptive, and Debilitating.

The initial method made three separations that remain intact:

- CVSS-derived impact and asset requirements determine the PAIN word.
- Agency scope maps the word to N1–N5.

- Exploitability and internet reachability select the remediation column; they do not change potential impact.

This calibration changes only the first of those steps. Remediation matrices, likely-exploitable logic, internet-reachability logic, multi-agency handling, VEX dispositions, and evidence-gated Modified impact reductions remain outside its scope.

The governed asset security-impact profile and any intended-use Security Requirements ceiling are also external inputs to this calibration. This paper does not derive either input. It holds the effective CR/IR/AR vector constant while comparing scoring methods. *Before the PAIN Equation* specifies the upstream derivation that caps a system-relative profile—whether derived from an archetype, direct per-objective assessment, or another governed method—by the confirmed federal information types in the affected agency scope.

IN PLAIN TERMS

The original model asked the right question but used the wrong ruler at the top of the scale. It measured High-requirement assets against a maximum created by Medium requirements. This paper replaces the ruler, preserves the CVSS inputs, and places the Debilitating line at the first combination that is both catastrophic and compound.

2 The calibration problem

Let C, I, A be the effective CVSS impact values for confidentiality, integrity, and availability:

$$C, I, A \in \{\, 0,\ 0.22,\ 0.56 \,\},$$

and let the corresponding asset requirements be

$$\mathrm{CR}, \mathrm{IR}, \mathrm{AR} \in \{\, 0.5,\ 1.0,\ 1.5 \,\}.$$

These are the CVSS v3.1 None/Low/High impact weights and Low/Medium/High Security Requirement weights. CVSS v3.1 defines the capped Modified Impact Sub-Score as

$$\mathrm{MISS} = \min(1 - (1 - C\mathrm{CR})(1 - I\mathrm{IR})(1 - A\mathrm{AR}),\ 0.915)\,. \tag{1}$$

The initial PAIN method normalized this value:

$$S_M = \frac{\mathrm{MISS}}{0.915}, \tag{2}$$

then used $(0.25, 0.55, 0.80)$ as the three word thresholds.

The value 0.915 is the rounded Medium-requirement all-High impact ceiling:

$$1 - (1 - 0.56)^3 = 0.914816 \approx 0.915.$$

But one High impact on a High requirement contributes $0.56 \times 1.5 = 0.84$. Therefore a one-dimensional finding produces

$$S_M = \frac{0.84}{0.915} = 0.918033,$$

which exceeds the original Debilitating threshold of 0.80. This is the single-dimension cliff: a complete loss of one important security property becomes N4 on a single-agency asset before any second security property is affected.

The effect is operational, not cosmetic. For a Class C provider, an N4-to-N3 change moves the `LEV+NIRV` deadline from 8 days to 32 days and the `LEV+IRV` deadline from 4 days to 16 days. A large cluster at one arithmetic point can therefore create both remediation pressure and a foreseeable incentive to assign a genuinely High requirement as Moderate.

3 Design criteria

The review applied five criteria before accepting a change:

1. **Semantic first.** Each PAIN boundary must have a written meaning that can be tested with scenarios. A target distribution is validation evidence, not the definition.
2. **Preserve CVSS-derived inputs.** The official 0.56/0.22/0 impact weights, 1.5/1.0/0.5 Security Requirement weights, and complement-product form should remain unless evidence demonstrates that a particular input is unsuitable.
3. **Keep profile assignment usable.** A regular operator should resolve an asset's independently dimensional profile through a manageable governed workflow without navigating a cross-product of roles, modifiers, deployment modes, and impact exceptions.
4. **Preserve dimensional information.** Low impacts, Moderate requirements, and accumulation across C/I/A should continue to affect the scalar rather than disappear behind a boolean count.
5. **Remain deterministic and auditable.** Given the same effective impact vector and effective CR/IR/AR profile, every implementation must produce the same result, retain full-precision intermediate values, and expose the policy version.

4 Alternatives evaluated

4.1 Additional context modifiers

One approach was to retain broad role classes but add environment-specific variants, such as internal-only orchestration, administrator-only CI/CD, redundant control planes, or restricted security tooling. A primary role label could have been combined with a second context modifier, or the catalog could have introduced parallel role classes for each context.

Reviewing the underlying role taxonomy was still necessary: some broad High assignments warranted clearer, role-based treatment. But taxonomy correction and scoring calibration solve different problems. Reclassification is justified only when the asset's actual mission role supports a different CR/IR/AR vector. Even a well-governed catalog will contain genuine High requirements, and a scoring method that turns every single H/H alignment into Debilitating remains miscalibrated. Relying exclusively on re-categorizing assets would therefore move the calibration problem into operator tagging, reward under-classification, and leave the underlying arithmetic unchanged.

Illustrative two-label model. The two-label form would be straightforward at first:

```
asset-archetype: orchestrator
deployment-context: redundant-admin-only
```

The second label would alter the security-impact profile supplied by the first. Table 1 shows representative outcomes considered during the design discussion. They are illustrative rejected alternatives, not normative archetype assignments.

Primary role	Context modifier	CR	IR	AR	Distinction encoded
`orchestrator`	service-critical control plane	M	M	H	Sustained loss stops service delivery.
`orchestrator`	redundant, administrator-only	M	M	M	Failover limits the availability consequence.
`cicd-pipeline`	production release/signing	H	H	M	Disclosure or corruption reaches production.
`cicd-pipeline`	isolated internal test	M	M	L	Output is non-production and reproducible.
`security-tooling`	inline preventive enforcement	M	H	H	Corruption or outage removes an active control.
`security-tooling`	passive telemetry only	M	M	L	Loss reduces visibility but not service delivery.
`identity-secrets`	durable credentials	H	M	M	Disclosure is catastrophic; loss is recoverable.
`identity-secrets`	replicated ephemeral token cache	H	M	L	Replication and regeneration reduce availability need.

Table 1: Illustrative, non-normative two-label mappings considered and rejected.

Even this small example requires rules for whether every modifier is valid for every role, what an omitted modifier means, whether a modifier replaces or changes the base vector, and whether a direct CR/IR/AR override takes precedence. Eight roles and three modifiers expose up to 24 apparent choices before exceptions; adding a fourth modifier adds another eight combinations and their documentation, validation, tests, and migration behavior.

This could produce the desired numerical result because it moves more dimensions from High to Moderate or Low before scoring. It was not selected for three reasons:

- Two independent metadata axes create a cross-product of valid and invalid combinations and make precedence rules necessary.
- Parallel role classes cause catalog growth, ambiguous choices, and drift as teams interpret nearly identical labels differently.
- The approach moves calibration pressure into asset tagging. Operators can obtain a more favorable deadline by choosing a less severe profile even when the underlying system role is unchanged.

Environment-specific facts still matter, but they belong in evidence-backed CVSS Modified metrics or other governed control evidence, not in an unlimited taxonomy of scoring profiles.

4.2 Changing the High multiplier from 1.50 to 1.25

Reducing the High Security Requirement multiplier to 1.25 directly removes the single-dimension cliff:

$$\frac{0.56 \times 1.25}{0.915} = 0.765027,$$

which is Disruptive at the original thresholds. Two High impacts on two High requirements still produce

$$\frac{1 - (1 - 0.70)^2}{0.915} = 0.994536,$$

which remains Debilitating.

The distribution moved close to the desired range, but the alternative was not selected. The 1.5 value is the published CVSS v3.1 High requirement weight; 1.25 would be a PAIN-specific

replacement without an independent derivation. FIRST describes v3 calibration broadly as expert judgment checked against real vulnerabilities, not as a public parameter-specific derivation of 1.5. Retaining it preserves a standardized input rather than claiming unique correctness for PAIN. It would also alter every boundary, not only Debilitating. Finally, multiplier reduction alone does not express the desired semantics: three High impacts on an all-Moderate asset still produce $S \approx 1$ and therefore N4.

4.3 Raising the original Debilitating threshold

The review was not opposed in principle to changing the θ values. Unlike the CVSS metric weights, they are governed PAIN policy boundaries and should change when the intended meaning of a PAIN word changes. The review did, however, reject choosing thresholds primarily to manufacture a preferred distribution—for example, tuning until Minimal, Narrow, Disruptive, and Debilitating appeared near the 20th, 40th, 60th, and 80th percentiles of the observed vulnerability population.

That approach would produce a tidy histogram but make classification relative to the current dataset. The same impact and asset requirements could receive a different PAIN word when unrelated vulnerabilities entered or left the population, when a different tenant mix was scanned, or when asset-classification coverage improved. Percentile targets may be useful diagnostics, but they are not defensible definitions of consequence.

Moving θ_3 from 0.80 to 0.92 while retaining the 0.915 denominator demotes the 0.918033 single-dimension cluster. This is a legitimate PAIN policy change, and the scan distribution improved. Standing alone, however, it retains a Medium-centered axis and places the new line almost exactly above the cluster it is intended to demote. Without a scenario-derived boundary, the change can reasonably look like histogram fitting. The final method therefore changes θ_3 only after deriving a semantic boundary from the finite impact/requirement state space. The lower thresholds are likewise re-derived from stated Minimal, Narrow, and Disruptive anchor scenarios rather than inherited or percentile-fitted. The observed histogram is used to test the consequence, not select any threshold.

4.4 Restoring round thresholds with another normalization

The review also considered retaining visually simple thresholds such as $(0.25, 0.55, 0.80)$ by dividing the final score by an additional constant. No single constant can preserve the three adopted semantic boundaries. If $S' = S_H/k$, mapping the adopted boundaries to those three values would require, respectively,

$$k = \frac{0.2811515969}{0.25} \approx 1.125, \qquad k = \frac{0.5623031939}{0.55} \approx 1.022, \qquad k = \frac{0.933}{0.80} = 1.16625.$$

Because these are different constants, another uniform normalization cannot perform all three mappings. Selecting the top-boundary value alone would move the lower boundaries and compress the maximum possible score below one.

A piecewise or nonlinear monotone transformation could force all three display values while preserving the current classifications, but only by introducing another governed function whose sole purpose is cosmetic. Likewise, some rounded cut points happen to fall inside empty intervals in the present 729-state lattice, but they would no longer state the scenario anchors directly and could behave differently if a future method admitted intermediate impact values. The method therefore retains the directly derived S_H scale and thresholds. Implementations may round values for display, but must classify with the governed full-precision parameters.

4.5 Strict two-dimension alignment gate

A categorical alternative required two aligned pairs of High technical impact and High asset requirement before N4:

$$\#\{d : \text{impact}_d = H \ \wedge \ \text{requirement}_d = H\} \geq 2.$$

This made N4 rare, but it was too rigid. An archetype such as `orchestrator=MMH` could never reach N4, even under complete $C{:}H/I{:}H/A{:}H$ compromise, because only availability is High.

4.6 Compound component gate

The gate was broadened to require at least one catastrophic H/H alignment and at least two High technical impacts on Moderate-or-High requirement dimensions. It handled the orchestrator case and produced a credible distribution.

The method was not selected as the final implementation because it adds a second classifier after the scalar and loses dimensional information at the boundary. Low impacts cease to matter to eligibility, and vectors with materially different continuous scores collapse to the same pair of counts. Implementations must also explain why a raw scalar above the Debilitating threshold was subsequently capped at Disruptive.

The gate nevertheless supplied the semantic test used to derive the final arithmetic boundary.

4.7 CVSS v4 MacroVectors

CVSS v4 confirms that aligned impact and requirements can be treated categorically. Its EQ6 group enters the more-severe level when *any* of `VC:H/CR:H`, `VI:H/IR:H`, or `VA:H/AR:H` is true, and it evaluates EQ6 jointly with the vulnerable-system impact pattern in EQ3. CVSS v4 does not require two aligned dimensions or compound impact. Its full score also includes exploitability, subsequent-system impact, and threat metrics that PAIN deliberately routes elsewhere.

Accordingly, CVSS v4 informed the alignment analysis but was not copied as the PAIN classifier. The final method remains compatible with v4 vulnerable-system impact values by mapping `VC/VI/VA` onto the PAIN C/I/A inputs when a v3.1 vector is unavailable.

5 High-centered normalization

The final method separates the CVSS-defined cap from the underlying environmental impact aggregate. Define the pre-cap aggregate

$$U = 1 - (1 - C\text{CR})(1 - I\text{IR})(1 - A\text{AR}). \tag{3}$$

The symbol U is PAIN notation for the expression before the CVSS cap; it is not a new CVSS metric. CVSS v3.1 defines the conformant Modified Impact Sub-Score as

$$\text{MISS} = \min(U, 0.915). \tag{4}$$

When $U \leq 0.915$, U and MISS are numerically identical. When $U > 0.915$, MISS remains fixed at 0.915 regardless of how much additional weighted impact the other dimensions contribute. The cap belongs to the official CVSS 0–10 scoring machinery, where MISS subsequently interacts with scope, exploitability, and rounding rules.

PAIN does not reuse that downstream formula. Its severity axis is impact only; exploitability and internet reachability are routed separately to remediation timing. Consequently, collapsing every $U > 0.915$ to the same value removes information PAIN needs to distinguish different

compound-impact states. The revision branches immediately before the CVSS cap: the published CVSS calculation continues to use MISS, while the PAIN calculation retains U.

The maximum attainable U over the retained PAIN inputs occurs when all three impacts and all three requirements are High:

$$D_H = 1 - (1 - 0.56 \times 1.5)^3 = 1 - 0.16^3 = 0.995904. \tag{5}$$

The revised scalar is therefore

$$S_H = \frac{U}{D_H} = \frac{U}{0.995904} \in [0, 1]. \tag{6}$$

No cap is needed before division because D_H is the maximum attainable value in the defined input domain. An implementation may defensively evaluate $\min(U, D_H)/D_H$.

Table 2 makes the boundary concrete. It compares MISS, not a complete CVSS Environmental Score; a complete score would require the rest of the CVSS vector and formula.

Weighted-impact state	U	**CVSS MISS**	**PAIN** S_H	**Effect**
One H impact on one H requirement	0.840000	0.840000	0.843455	Below the cap; no information is clipped.
One H/H plus one H/M contribution	0.929600	0.915000	0.933423	CVSS clips 0.014600; PAIN retains the compound increment.
Three H impacts on three H requirements	0.995904	0.915000	1.000000	CVSS clips 0.080904; PAIN reaches its defined maximum.

Table 2: Where the CVSS MISS path and PAIN normalization path diverge.

IMPORTANT BOUNDARY

CVSS and PAIN use the same impact inputs, but they use them for different purposes. CVSS stops the aggregate at 0.915 and continues calculating the official CVSS score. PAIN keeps the additional impact above 0.915 and divides by 0.995904 so that strongly weighted dimensions remain distinguishable. This changes only the PAIN classification. It does not recalculate or replace the published CVSS score or vector. U and S_H must therefore be labeled as PAIN values, not CVSS scores.

6 Deriving the thresholds

6.1 Deriving the Minimal/Narrow and Narrow/Disruptive boundaries

The original 0.25 and 0.55 values were governed PAIN policy, not CVSS constants. Merely multiplying them by $0.915/D_H$ would preserve their old raw-U locations, but it would not independently explain why those locations separate the PAIN words. The revision instead applies the same scenario-first method used for Debilitating.

There is an important limit to this derivation. The C/I/A and CR/IR/AR lattice does not encode the number of affected users, outage duration, or amount and type of affected data. It therefore cannot reproduce every part of the FedRAMP word definitions by arithmetic alone. The boundaries below are governed, standards-informed proxies: CVSS supplies the discrete impact and requirement values; FIPS 199 supplies the distinction between limited, serious, and severe or catastrophic adverse effect; PAIN assigns those states to its operational words.

For a single dimension, the shorthand H/L means High technical impact on a Low Security Requirement; L/H and H/M are read the same way. When terms are joined by a plus sign, they occur on different C/I/A dimensions.

Narrow begins with a complete loss on a Low-requirement dimension. A High technical impact on a Low Security Requirement contributes $0.56 \times 0.5 = 0.28$. Although the asset owner has determined that loss of the property has limited adverse effect, the technical loss itself is complete. The revision treats that state as Narrow rather than Minimal:

$$\boxed{\theta_1 = \frac{0.28}{0.995904} \approx 0.2811515969}. \tag{7}$$

An isolated Low impact on a Moderate requirement remains below the line: $0.22/0.995904 = 0.2209048262$. Exhaustive enumeration confirms that no attainable state lies between 0.2209048262 and 0.2811515969. With an inclusive comparison, the complete-loss anchor itself is Narrow.

Disruptive begins with a High impact on a Moderate requirement. A Moderate Security Requirement represents a property whose loss can have a serious adverse effect. A High technical impact directly aligned to that property therefore provides the first unambiguous Disruptive anchor:

$$\boxed{\theta_2 = \frac{0.56}{0.995904} \approx 0.5623031939}. \tag{8}$$

The two closest disputed patterns remain Narrow. One H/L contribution combined with one L/H contribution yields $S_H = 0.5197288092$; two L/H contributions yield $S_H = 0.5533665896$. Neither contains a High technical impact on a Moderate-or-High requirement. The latter is the greatest attainable state below the H/M anchor, and no state lies between 0.5533665896 and 0.5623031939.

The boundary does not forbid lower-level contributions from becoming Disruptive. Because the complement product remains continuous over all three dimensions, any compound state whose aggregate reaches the H/M-equivalent value also crosses the line. The rule is an arithmetic threshold, not a categorical H/M gate.

6.2 Deriving Debilitating from the discrete state space

For derivation only, define a *compound catastrophic* state as one containing:

1. at least one High technical impact aligned with a High requirement; and
2. at least one additional High technical impact aligned with a Moderate-or-High requirement.

What the compound predicate approximates. The first condition preserves the required dimensional alignment: at least one security property must combine a complete technical loss with a catastrophic agency requirement. The second condition is a breadth test. It asks whether the same vulnerability can also cause complete technical loss on another property whose loss is at least serious. The reference calibration treats that multidimensional shape as a reasonable proxy for a broader and generally more consequential class of vulnerability.

External consistency check. This breadth requirement is directionally consistent with CISA Vulnrichment data. In a reproducible random sample of 15,000 CVE records, 3,439 contained both a CISA SSVC Technical Impact decision and CISA-provided CVSS v3 impact metrics. Of the 1,475 records labeled `Technical Impact: total`, 1,376 (93.3%) had at least two High C/I/A impacts. Conversely, only 25 of 1,006 records (2.5%) with exactly one High impact were labeled `total`. This comparison is corroborative rather than definitional: SSVC Technical Impact characterizes control or disclosure relative to the vulnerable component, whereas PAIN also evaluates dimensional alignment with the asset's security requirements [CISA-SSVC] [CISA-VULN].

This is not an arbitrary rule that High/High alignment is unimportant. A lone High/High pair is the strongest possible one-objective loss and remains Disruptive under the reference calibration. Likely exploitability, internet reachability, and Certification Class can still place it on a demanding remediation clock. The additional component is required only before the method uses the much narrower word *Debilitating*.

Why CWE is not an input. The method deliberately does not infer this breadth from a CWE identifier. CWE is a root-cause taxonomy: entries occur at different abstraction levels, mappings can be absent or imprecise, and MITRE's own mapping guidance calls for careful, specific review rather than assuming accurate mapping at scale [CWE-RCM]. A weakness category does not determine which C/I/A properties a particular exploit can affect in a particular deployment. Turning CWE IDs into PAIN multipliers or gates would therefore require a large provider-authored lookup table with uncertain precedence for multiple, missing, or differently abstracted mappings.

CWE remains useful outside the PAIN scalar for root-cause analysis, remediation ownership, secure-development trends, detection strategy, and finding grouping. PAIN instead uses the CVSS-derived impact dimensions already attached to the vulnerability and combines them with the effective requirements of the affected asset. The approximation is imperfect, but its inputs and failure modes are visible and reproducible.

This definition is not implemented as a gate. It partitions the finite input space so that a scalar boundary can be derived.

There are $3^3 = 27$ possible C/I/A impact patterns and $3^3 = 27$ possible CR/IR/AR patterns, for 729 combinations. Exhaustive enumeration establishes two extrema.

The highest non-compound combination is one High impact on a High requirement plus two Low impacts on High requirements:

$$U_{\text{max,noncompound}} = 1 - (1 - 0.84)(1 - 0.33)^2 = 0.928176, \tag{9}$$

$$S_{\text{max,noncompound}} = \frac{0.928176}{0.995904} = 0.9319934452. \tag{10}$$

The lowest compound catastrophic combination is one High impact on a High requirement plus one High impact on a Moderate requirement, with no third impact:

$$U_{\text{min,compound}} = 1 - (1 - 0.84)(1 - 0.56) = 0.929600, \tag{11}$$

$$S_{\text{min,compound}} = \frac{0.929600}{0.995904} = 0.9334233018. \tag{12}$$

The open interval between those values is non-empty. The selected threshold is

$$\boxed{\theta_3 = 0.933}, \tag{13}$$

which satisfies

$$0.9319934452 < 0.933 < 0.9334233018.$$

Across all 729 discrete combinations, $S_H \geq 0.933$ if and only if the combination meets the compound catastrophic definition above. Seventy combinations qualify; none of the 659 non-compound combinations cross the threshold, and no compound combination falls below it.

FINAL CALIBRATION

The final PAIN scalar and thresholds are:

$$U = 1 - (1 - C\text{CR})(1 - I\text{IR})(1 - A\text{AR}), \qquad S_H = \frac{U}{0.995904},$$

$$(\theta_1, \theta_2, \theta_3) = \left(\frac{0.28}{0.995904},\ \frac{0.56}{0.995904},\ 0.933\right) \approx (0.2811515969,\ 0.5623031939,\ 0.933).$$

Minimal is $S_H < \theta_1$; Narrow is $\theta_1 \leq S_H < \theta_2$; Disruptive is $\theta_2 \leq S_H < \theta_3$; and Debilitating is $S_H \geq \theta_3$. Implementations classify with full-precision S_H and thresholds. Rounded values are for display only. The parameter set must be versioned as governed PAIN policy.

6.3 Reference calibration and provider variation

FedRAMP defines the PAIN words and requires providers to estimate them, but it does not prescribe this scalar or these cut points. The values above are the governed reference calibration for this method, not a claim that 0.933 is the only defensible risk policy.

Table 3 illustrates other top-boundary viewpoints while holding the CVSS-derived weights, denominator, and lower thresholds constant:

Policy lens	θ_3	**Meaning in the 729-state lattice**	**Top-band states**
High-objective dominant	0.919	Every qualifying state contains at least one H/H alignment, but compound High impact is not required.	88
Reference compound catastrophic	0.933	At least one H/H alignment plus another High impact on a Moderate-or-High requirement.	70
Strict two-objective catastrophic	0.976	At least two H/H dimensional alignments.	25
Complete CIA catastrophic	1.000	Three H/H dimensional alignments.	1

Table 3: Illustrative, non-normative Debilitating lenses.

An exact policy that makes *every* isolated H/H state Debilitating cannot be represented by one scalar threshold on this aggregation. The weakest isolated H/H state scores 0.843455, while complete High technical impact against all-Moderate requirements scores 0.918578 without any H/H pair. A provider that needs an exact single-alignment rule must add a categorical predicate rather than disguise it as a normalization constant.

A CSP MAY adopt a different threshold set based on its own documented risk evaluation, but it SHOULD:

- define the intended meaning of each word before examining the resulting distribution;
- enumerate the full supported input lattice and test the boundary cases;
- validate the operational consequences against rated scenarios or findings;
- govern and version the denominator and thresholds as one calibration set;
- disclose the active policy to affected agencies and retain it in finding audit output; and
- avoid changing thresholds per finding, customer, or Certification Class after observing the deadline outcome.

Certification Class belongs downstream: it determines the operational timeframe for the already-derived PAIN result. Varying θ_3 by Class would make the same agency consequence change meaning merely because the provider selected a different assurance tier.

7 Ceiling-constrained reachability

The 729-state proof above describes the unconstrained Cartesian product of every CVSS impact vector and every effective Security Requirements vector. A particular deployment may expose only a subset of those states after its intended-use ceiling is applied.

Let the uncapped asset security-impact-profile requirement be $R_x(o)$ and the intended-use ceiling be $E_x(o)$. The PAIN calculation receives:

$$R_x^{\text{eff}}(o) = \min\big(R_x(o), E_x(o)\big).$$

The ceiling derivation is upstream of this calibration; see *Before the PAIN Equation*. This section examines only what the adopted PAIN arithmetic does with that input.

7.1 Low- and Moderate-impact ceilings

If a ceiling contains no High objective, then every effective requirement is at most Moderate, regardless of the uncapped profile. The largest possible pre-cap aggregate occurs with High technical impact on all three Moderate requirements:

$$U_{\max,\le M} = 1 - (1 - 0.56)^3 = 0.914816, \tag{14}$$

$$S_{\max,\le M} = \frac{0.914816}{0.995904} = 0.918578\ldots \tag{15}$$

This is below the adopted Debilitating threshold:

$$0.918578\ldots < 0.933.$$

Therefore no state permitted by an all-Low/Moderate ceiling reaches Debilitating under this calibration. All 70 Debilitating states in the unconstrained lattice require at least one effective High requirement and are removed by such a ceiling.

IMPORTANT BOUNDARY

This is a consequence of the PAIN method's retained CVSS weights, High-centered denominator, and 0.933 threshold. NIST SP 800-60 and FIPS 199 do not independently define this reachability result. A different PAIN aggregation or threshold could classify the same effective requirements differently.

FIPS 199 separately establishes that any High system objective makes the information system High overall. Combining that categorization rule with this methodology-specific result means that a genuinely Low- or Moderate-impact information system cannot reach the Debilitating word *under the adopted PAIN calibration*. The dimensional vector must still be retained: an overall High system may be H/M/L, and the label High does not turn it into H/H/H.

7.2 What a High ceiling does not prove

A High objective makes Debilitating reachable in the deployment's state space; it does not make every vulnerability Debilitating. The finding still needs:

1. a High technical impact aligned with an effective High requirement; and
2. another High technical impact aligned with an effective Moderate-or-High requirement.

Certification Class does not change this reachability result. A Class C certification is not a Moderate-impact ceiling: current FedRAMP guidance allows a Class C package to provide adequate information for some High security objectives [FEDRAMP-CLASS]. For a Class B or Class C use,

however, any High objective handed to this calculation must be affirmatively identified and justified by the deploying agency; it must not be inferred from generic product capabilities or hypothetical content. If that agency-identified objective remains High after the asset profile and intended-use ceiling are intersected, the High requirement remains in scope for PAIN.

Class is nevertheless operationally consequential. It states the provider's reusable level of assurance and selects the standard vulnerability-response and incident-communication commitments supplied for the already-derived consequence. A Class C service supporting a High objective therefore supplies Class C commitments unless a tighter contract or higher Class applies. The agency and CSP must evaluate and record whether those commitments, service design, supplemental controls, and any agency-managed safeguards are adequate. They must not express an assurance shortfall by lowering the effective requirement or the PAIN result.

IN PLAIN TERMS

A genuinely FIPS 199 Low- or Moderate-impact system cannot produce Debilitating under the reference calibration because it has no effective High requirement. Class C is not a synonym for FIPS Moderate. If an agency-identified High objective remains High on the affected asset, Debilitating becomes mathematically possible, while the Class C level of assurance still governs the standard operational commitment. Possibility, classification, and assurance are separate judgments.

These layered conditions make a Class B or Class C Debilitating result exceptional by construction, not impossible. The calibration establishes the necessary reachability conditions; it does not claim a universal empirical frequency. Certification Class selects the applicable remediation timeframe only after the PAIN word and agency scope have been determined; the upstream paper specifies how the agency and CSP document the resulting assurance fit or divergence.

The asset security-impact profile can also keep a dimension below the ceiling because the ceiling is only a cap. Finally, multi-agency scope maps an already derived customer-effect word to the N-level; it does not create the Debilitating word.

8 Worked scenarios

Tables 4 and 5 apply the final arithmetic to boundary and ordinary cases. Requirement vectors are explicit effective inputs: the governed asset assessment supplies the uncapped security-impact profile and any intended-use ceiling caps it before these calculations. Results assume a single-agency resource; the separate agency-scope mapping can promote qualifying words to N4 or N5.

Boundary scenario	**C/I/A**	**CR/IR/AR**	U	S_H	**Result and reason**
Isolated limited effect	L/N/N	M/M/M	0.220000	0.220905	N1. Low impact on a Moderate requirement remains Minimal.
Complete loss, Low requirement	H/N/N	L/M/M	0.280000	0.281152	N2. First Narrow anchor.
Crossed mixed pair	H/L/N	L/H/M	0.517600	0.519729	N2. No direct serious alignment.
Two limited effects on High requirements	L/L/N	H/H/M	0.551100	0.553367	N2. Greatest attainable state below H/M.
Direct serious alignment	H/N/N	M/M/M	0.560000	0.562303	N3. First Disruptive anchor.

Table 4: Strict lower-boundary scenarios. Threshold comparisons are inclusive.

Scenario	C/I/A	CR/IR/AR	U	S_H	Result and reason
Availability-only orchestrator	N/N/H	M/M/H	0.840000	0.843455	N3. Catastrophic availability is isolated.
Integrity and availability on orchestrator	N/H/H	M/M/H	0.929600	0.933423	N4. Minimum compound catastrophic case.
Complete orchestrator compromise	H/H/H	M/M/H	0.969024	0.973010	N4. Catastrophic availability plus serious C/I impact.
Confidentiality and integrity on data backbone	H/H/N	H/H/M	0.974400	0.978408	N4. Two catastrophic dimensions.
Complete app-tier compromise	H/H/H	M/M/M	0.914816	0.918578	N3. Broad serious impact, but no catastrophic requirement alignment.
Worst non-compound boundary	H/L/L	H/H/H	0.928176	0.931993	N3. Low-impact accumulation remains below the line.
Low availability on orchestrator	N/N/L	M/M/H	0.330000	0.331358	N2. Partial availability impact remains Narrow.

Table 5: Ordinary and Debilitating-boundary behavior under High-centered normalization.

The first two rows of Table 5 are the central upper-tier separation. A single total loss on a catastrophic dimension is Disruptive. Adding a High impact on a second serious dimension crosses the derived line. The fifth row is an explicit policy decision: complete CIA compromise of an all-Moderate asset remains N3 because the governed requirements describe every loss as serious rather than catastrophic.

Concrete finding example. Dataset A contains a denial-of-service finding with $C{:}N/I{:}N/A{:}H$ on a resource governed by an `MMH` requirement vector. $U = 0.84$ and $S_H = 0.843455$, producing N3. The asset can retain the correct High Availability Requirement without creating an N4 deadline from a one-dimensional impact.

9 Empirical backtest

The candidate methods were replayed against two anonymized evaluation reports generated on July 21, 2026. Dataset A contained 2,120 findings and 10,231 affected-resource observations. Dataset B contained 3,691 findings and 60,817 affected-resource observations. The same CVE may affect multiple resources; the two units are therefore reported separately.

All candidate runs used the same adopted archetype catalog and the existing engine's impact-source precedence, including CVSS v3/v4 vector parsing, qualitative severity fallback, and technical-impact lifting where recorded. The stored-current rows retain the reports' original policy and provide operational context; they are not an isolated single-variable control.

Method	Dataset A N4	Dataset B N4
Stored scan output	40.56%	63.95%
Adopted catalog baseline; original $D = 0.915$, $\theta_3 = 0.80$	26.62%	62.22%
Raise θ_3 to 0.92 only	15.66%	29.54%
Change High multiplier to 1.25 only	17.49%	29.76%
Strict two-H/H gate	6.61%	20.39%
Compound gate	11.60%	21.22%
Final: $D_H = 0.995904$, $\theta_3 = 0.933$	**11.60%**	**21.22%**

Table 6: N4 share of affected-resource observations.

The final arithmetic selected exactly the same N4 affected-resource observations as the compound gate in both reports. With the lower thresholds independently anchored to the H/L and H/M scenarios, the complete distributions are shown in Table 7.

Scan and policy	N1	N2	N3	N4
Dataset A stored current	11.37%	28.47%	19.60%	40.56%
Dataset A final	16.12%	27.08%	45.20%	**11.60%**
Dataset B stored current	1.50%	28.29%	6.25%	63.95%
Dataset B final	2.21%	30.41%	46.16%	**21.22%**

Table 7: Affected-resource PAIN distributions under the final calibration.

At the finding level, taking the worst affected resource per finding, N4 moved from 55.80% to 18.16% in Dataset A and from 50.50% to 17.83% in Dataset B.

9.1 The threshold remains reachable

The purpose of 0.933 is not to make N4 exceptional or unattainable. It is crossed by the common two-High impact shapes $H/H/N$, $H/N/H$, and $N/H/H$, as well as $H/H/H$, whenever the requirements provide the necessary catastrophic and serious alignment.

IN PLAIN TERMS

0.933 does not mean that a vulnerability must cause "93.3 percent damage," nor does it represent the 93.3rd percentile of observed findings. The PAIN scalar is an ordered index over discrete CVSS-derived states. The first qualifying compound state lands at 0.933423, immediately above the line, while ordinary two-High and three-High scenarios land from that point through 1.0. A vulnerability does not need three High impacts, three High requirements, multi-agency scope, or a specially constructed vector to reach Debilitating.

Dataset A retained 1,187 N4 affected-resource observations across multiple archetypes: 701 *data-backbone*, 177 *orchestrator*, 125 *config-actuation*, 102 *identity-secrets*, 49 *security-tooling*, and 33 across other roles. Its most frequent qualifying patterns included HHH/HHM, HHH/MMH, HHN/HHM, HHH/MHM, and HHH/HMM.

Dataset B retained 12,903 N4 observations, but 11,622 of them were `unclassified` and therefore used the conservative HHH fail-safe. This is an important interpretation constraint: the remaining Dataset B N4 concentration is primarily a classification-coverage issue, not evidence that 0.933 is too low.

At the worst-resource-per-finding level, 385 Dataset A findings (18.16%) and 658 Dataset B findings (17.83%) remained N4. A top tier reached by approximately one in six findings in both snapshots is plainly operational rather than theoretical. The method nevertheless removes the one-dimensional cases that made N4 cease to signal the worst compound consequences.

9.2 What the backtest does and does not prove

The backtest demonstrates that the threshold is operationally reachable, does not collapse N4 to a handful of exotic vectors, and produces the same top-tier set as the independently stated compound predicate. It does not prove that 11.60% or 21.22% is an ideal universal N4 rate. The reports are two deployment snapshots, not a representative sample of all federal cloud systems. The mathematical state boundary supplies the policy justification; the scans test its consequences.

10 Why the final method is preferable

Approach	**CVSS constants**	**Tagging burden**	**Dimensionality**	**Boundary rationale**
Asset sub-profiles	Preserved	High	Preserved	Hidden in classification choices
1.25 multiplier	Changed	Low	Preserved	Indirect
Threshold-only	Preserved	Low	Preserved	Weak unless separately derived
Strict gate	Preserved	Low	Reduced to H/H count	Too restrictive
Compound gate	Preserved	Low	Reduced at the final branch	Explicit but adds a second classifier
High-centered arithmetic	Preserved	Low	Preserved continuously	Derived from the full discrete state space

Table 8: Decision summary. "CVSS constants preserved" does not mean the final PAIN scalar is a conformant CVSS score; the normalization is explicitly PAIN-specific.

The final method is preferable because it obtains the compound-gate result without executing a gate. It keeps Low impacts and all three requirement levels inside one continuous calculation, retains the official metric weights, and makes the top threshold reproducible from the input lattice. It also keeps the operator-facing model simple: one independently dimensional asset security-impact profile supplies the reusable architectural vector, and one scope-resolved intended-use ceiling supplies the federal consequence cap. A role-based archetype catalog is one optional way to derive the first input. Neither input changes per finding.

11 Implementation and governance

11.1 Normative parameter set

An implementation of this calibration should govern the following values together:

```
impactWeights:
  high: 0.56
  low: 0.22
  none: 0.0

requirementWeights:
  high: 1.5
  medium: 1.0
  low: 0.5
```

```
normalizationDenominator: 0.995904

wordThresholds:
  narrow: 0.28115159694107
  disruptive: 0.56230319388214
  debilitating: 0.933
```

The denominator and thresholds are one calibration set. The first two thresholds should be constructed from the exact expressions $0.28/D_H$ and $0.56/D_H$ when the implementation permits it. The decimal serializations above are deliberately truncated rather than rounded upward so that the H/L and H/M anchors remain inside their intended bands under an inclusive comparison. Changing a metric weight, adding an impact level, or changing the aggregation invalidates the boundary proof and requires re-enumeration.

11.2 Precision and reproducibility

The separation between the highest non-compound and lowest compound states is approximately 0.00143. Implementations therefore MUST:

- classify using unrounded internal values;
- construct θ_1 and θ_2 from $0.28/D_H$ and $0.56/D_H$, or use decimal serializations that do not round above those exact anchors;
- compare against the configured full-precision threshold;
- display enough decimal places to explain a boundary decision;
- record U, D_H, S_H, the thresholds, resolved C/I/A, CR/IR/AR, and the policy version in audit output; and
- include boundary unit tests for the predecessor/anchor pairs 0.2209048262/0.2811515969, 0.5533665896/0.5623031939, and 0.9319934452/0.9334233018.

11.3 Migration

The change should be released as a new PAIN methodology version. Historical findings should retain the policy version under which their deadline was assigned. Active findings should be recomputed under an explicitly approved transition rule; the organization must acknowledge that N4-to-N3 changes extend remediation deadlines. The scoring calibration should be versioned and reviewed independently from any asset-catalog revision.

The default for an unclassified asset remains conservative HHH. The final method reduces single-dimension findings on that default to N3, but two- and three-High impact findings readily remain N4. Classification coverage should be monitored so that the fail-safe does not become the dominant production profile.

12 Limitations and future work

- The exact 0.933 separation depends on the discrete $\{N, L, H\} \times \{L, M, H\}$ domain and the retained CVSS v3.1 weights. It must be re-derived if those inputs change.
- The finding that an all-Low/Moderate ceiling cannot reach Debilitating is a consequence of this calibration, not an independent NIST or FedRAMP rule. It must be re-evaluated if the aggregation, weights, denominator, or thresholds change.

- Native CVSS v4 subsequent-system impacts and Safety are not included in the three-dimension proof. A future PAIN version may add them, but doing so requires a new denominator and threshold analysis rather than silently extending this one.
- The two backtests establish observed consequences, not statistical generalizability. Additional production datasets and analyst-rated scenario sets should be evaluated before broad standardization.
- A scalar cannot replace architecture review. Security-impact-profile assignment and its derivation evidence, Modified impact evidence, intended-use ceiling, agency scope, and vulnerability applicability remain governed inputs.
- The threshold's narrow mathematical gap is safe for deterministic software using full precision but unsuitable for manual classification from rounded dashboard values.

13 Conclusion

The original PAIN method correctly joined vulnerability impact with asset-specific security requirements, but its Medium-centered denominator saturated too early for High-requirement assets. The resulting single-dimension cliff made Debilitating too common and created pressure to solve a scoring problem through asset underclassification.

The review considered richer tagging, a lower High multiplier, threshold-only tuning, and categorical gates. Each could reduce N4, but each either increased operator complexity, abandoned a CVSS-defined value, obscured the semantic reason for the threshold, or discarded dimensional information.

High-centered normalization resolves the underlying mismatch. The new denominator is derived directly from the retained CVSS weights. Narrow begins at a complete technical loss on a Low-requirement dimension; Disruptive begins at a High impact on a Moderate-requirement dimension. The Debilitating threshold lies between two provable extrema in the complete discrete state space and remains readily reachable by ordinary two-High impact scenarios. The empirical scans then validate the operational consequences rather than dictate the boundaries.

The resulting rule is concise:

$$\boxed{S_H = \frac{1 - (1 - C\mathrm{CR})(1 - I\mathrm{IR})(1 - A\mathrm{AR})}{0.995904}}$$

with

$$\boxed{(\theta_1, \theta_2, \theta_3) = \left(\frac{0.28}{0.995904},\ \frac{0.56}{0.995904},\ 0.933 \right).}$$

This preserves the CVSS-derived evidence chain while giving PAIN a more defensible meaning at its most consequential boundary: N4 and N5 are reserved for impact that is not only severe, but compound and catastrophic in the affected environment. When an intended-use ceiling has no High objective, the adopted formula makes that boundary unreachable; this is a transparent property of the PAIN calibration, not a categorization rule imported from NIST or FedRAMP.

References

[PAIN] M. Venne, *A Deterministic, CVSS–Environmental Method for FedRAMP VDR/VER Vulnerability Prioritization*, stackArmor, July 2026. https://stackarmor.github.io/rfc-fedramp-vdr/vdr-pain-cvss.pdf.

[CEILING] M. Venne, *Before the PAIN Equation: Deriving Security-Requirements Ceilings from Intended Federal Information Types*, stackArmor, July 2026. https://stackarmor.github.io/rfc-fedramp-vdr/vdr-pain-security-requirements.pdf.

[CVSS31] FIRST, *Common Vulnerability Scoring System v3.1: Specification Document*. Environmental Security Requirements, MISS formula, and metric values. https://www.first.org/cvss/v3.1/specification-document.

[CVSS40] FIRST, *Common Vulnerability Scoring System v4.0: Specification Document*. Environmental metrics and EQ3/EQ6 MacroVectors. https://www.first.org/cvss/v4.0/specification-document.

[CVSS40-UG] FIRST, *CVSS v4.0 User Guide*. Expert comparison, equivalence-set, and interpolation methodology. https://www.first.org/cvss/user-guide.html.

[FIPS199] National Institute of Standards and Technology, *Standards for Security Categorization of Federal Information and Information Systems*, FIPS PUB 199. https://csrc.nist.gov/pubs/fips/199/final.

[SP800-60] National Institute of Standards and Technology, *Guide for Mapping Types of Information and Information Systems to Security Categories*, SP 800-60 Volumes I and II, Revision 1. https://csrc.nist.gov/pubs/sp/800/60/r1/final.

[FEDRAMP] FedRAMP, *Potential Agency Impact Definitions*, Notice 0012, and *Vulnerability Evaluation and Reporting*, Consolidated Rules 2026. https://www.fedramp.gov/notices/0012/; https://www.fedramp.gov/2026/providers/20x/rules/vulnerability-evaluation-and-reporting/.

[FEDRAMP-CLASS] FedRAMP, *Choosing a Certification Class*, Consolidated Rules 2026. https://www.fedramp.gov/2026/providers/start/class/#alignment-with-agency-use-cases.

[CWE-RCM] MITRE, *CVE to CWE Root Cause Mapping Guidance*. https://cwe.mitre.org/documents/cwe_usage/guidance.html.

[CISA-SSVC] Cybersecurity and Infrastructure Security Agency, *Stakeholder-Specific Vulnerability Categorization Guide*. Technical Impact decision values and affected-component scope. https://www.cisa.gov/resources-tools/resources/stakeholder-specific-vulnerability-categorization-ssvc.

[CISA-VULN] Cybersecurity and Infrastructure Security Agency, *Vulnrichment*, repository snapshot, July 31, 2026. GitHub commit `6c7a125`.

Reproducibility record

The discrete-state verification enumerates the Cartesian product of three impact values for each of C/I/A and three requirement values for each of CR/IR/AR. A conforming test run over the resulting 729 states must return:

Check	Expected result
D_H	0.995904000000
Predecessor to Narrow	0.22090482616798
Narrow anchor (H/L)	0.28115159694107
Predecessor to Disruptive	0.55336658955080
Disruptive anchor (H/M)	0.56230319388214
Maximum non-compound S_H	0.93199344515134
Minimum compound S_H	0.93342330184435
States at or above 0.933	70
State counts Minimal/Narrow/Disruptive/Debilitating	90 / 218 / 351 / 70
Predicate/threshold mismatches	0
Maximum S_H with every requirement $\leq M$	0.918578497526
Debilitating states after an all-Low/Moderate ceiling	0

The empirical run used the two July 21, 2026 evaluation exports described in Section 9: 2,120 findings and 10,231 affected-resource observations for Dataset A; 3,691 findings and 60,817 observations for Dataset B. Dataset names, resource names, and source identifiers are intentionally omitted from this public paper. Authorized reviewers may reproduce the run using the controlled internal evidence package.

The external CISA consistency check sampled 15,000 unique `CVE-*.json` records without replacement from the [CISA-VULN] snapshot using seed `20260801`. Eligibility required a CISA-ADP SSVC Technical Impact value of `total` or `partial` and CISA-ADP CVSS v3.0 or v3.1 Scope and C/I/A impact metrics. The filter yielded 3,439 records: 1,475 `total` and 1,964 `partial`.