The BlackHart Risk Index: A Multiplicative Multi-Dimensional Framework for Continuous DeFi Protocol Risk Assessment

Thomas Hart

BlackHart Inc.
thom@blackhart.io

May 2026

Abstract

Decentralized finance protocols collectively secure over $200 billion in deposited assets, yet no standardized, continuously updated risk measurement framework exists for this asset class. This paper introduces the BlackHart Risk Index (BRI), a quantitative risk scoring framework that evaluates DeFi protocols across 12 independent dimensions using a weighted geometric mean. The multiplicative formulation ensures that catastrophic weakness in any single dimension propagates to the composite score, preventing the masking effects inherent in additive models. We describe the dimension taxonomy, the mathematical properties of the scoring function, the actuarial credibility model (Z-Factor), the Forge Scale ordinal rating system, and the calibration methodology anchored against 200+ historical exploit events. We further present the on-chain oracle implementation, which publishes BRI scores as a composable DeFi primitive, and discuss the emerging requirement for continuous risk assessment in an era of AI-accelerated offensive capability.

---

1 Introduction

The smart contract security industry relies on a model borrowed from traditional software assurance. A protocol engages an auditor, receives a report, remediates findings, and publishes a badge. This model treats security as a discrete event. A protocol is "audited" the way a building is "inspected" — at a single point in time, against a fixed set of criteria, by a team that moves on to the next engagement.

For static systems, this model is adequate. DeFi protocols are not static systems [1]. They ship code continuously, integrate new dependencies, modify governance parameters, and interact with a threat environment that evolves as novel vulnerability classes are discovered [2] and as AI-powered offensive tools reduce the cost of exploit discovery by orders of magnitude.

The inadequacy of point-in-time assessment is compounding. In April 2026, Anthropic's Mythos Preview demonstrated autonomous discovery of thousands of zero-day vulnerabilities across major operating systems and browsers, constructing working exploits without human guidance [3]. Five weeks later, OpenAI launched Daybreak, a cybersecurity platform built on GPT-5.5 that automates repository scanning, threat modeling, and vulnerability validation [4]. These developments signal a structural shift in offensive capability. When

the cost of finding exploitable flaws drops by two orders of magnitude, the interval between assessments becomes the dominant risk factor. Point-in-time audits, however thorough, cannot address a threat that operates continuously.

The industry requires a risk measurement framework with three properties that point-in-time audits lack: *continuity*, *dimensionality*, and *composability*. The BlackHart Risk Index is designed to provide all three.

2 Related Work

The dominant security assurance model in DeFi is the manual audit. A team of two to five researchers reviews a protocol's source code over two to eight weeks and produces a report classifying findings by severity [5]. While audits provide valuable signal, they suffer from three structural limitations: they capture risk at a single point in time, they produce qualitative rather than continuous quantitative scores, and their outputs are neither machine-readable nor composable.

Bug bounty programs address the continuity gap in theory but introduce their own structural weaknesses. Protocols define narrow scopes (specific contracts, specific conditions, specific severity definitions) and offer payment for vulnerabilities found within those boundaries. This creates blind spots by design. The most dangerous vulnerability classes (oracle manipulation chains, cross-contract assumption mismatches, multi-step governance exploits) frequently span the boundary between what is in scope and what is not. When no bounties are claimed, the absence of findings is indistinguishable from the absence of examination. The result is a false signal that protocols and their users treat as evidence of security.

Several existing risk scoring frameworks address parts of the measurement gap. DeFi Score (Codefi, 2020) proposed a weighted average across smart contract risk, financial risk, and centralization risk [6]. L2 Risk (L2Beat) provides qualitative categorization for rollups [7]. Gauntlet provides protocol-specific economic parameter modeling. Gudgeon et al. formalized the risk properties of DeFi lending protocols [8]. These approaches share common limitations: additive aggregation that permits masking, narrow dimension coverage, or lack of continuous updates.

The BRI draws conceptual precedent from credit rating methodologies (Moody's, S&P, Fitch) but departs in its use of a multiplicative rather than additive aggregation function. This departure reflects the distinct failure modes of smart contract systems, where a single critical weakness is sufficient for total loss of deposited assets.

3 Methodology

3.1 Design Principles

The BRI framework draws on multi-criteria decision theory [9], [10] and is constructed around five principles:

1. **Non-compensatory aggregation.** A catastrophic weakness in any single dimension must propagate to the composite score.
2. **Dimensional independence.** Each dimension captures a distinct axis of risk.
3. **Continuous and bounded scoring** on the interval $[300, 1000]$.
4. **Temporal awareness.** The Z-Factor distinguishes inherent quality from actuarial confidence.
5. **Machine composability.** Scores serve as primitives for downstream applications, including on-chain consumption by other smart contracts.

3.2 Dimension Taxonomy

The BRI evaluates protocols across 12 dimensions organized into three tiers by weight class. Ten dimensions are active in every assessment. Two are conditional, activated when the protocol's architecture requires them.

ID	Dimension	Weight	Category
D1	Access Control	0.18	Primary
D2	Economic Soundness	0.13	Primary
D3	Oracle Integrity	0.13	Primary
D6	Battle-Tested Maturity	0.12	Primary
D5	Governance & Upgradeability	0.10	Primary
D7	Adversarial Resilience	0.10	Primary
D11	Operational Security	0.10	Primary
D4	Compositional Risk	0.05	Secondary
D12	Cascade Exposure	0.05	Secondary
D8	Supply Chain	0.04	Secondary
D9	Liquidity & Market Structure	0^*	Conditional
D10	Cross-Chain Messaging	0^*	Conditional

Table 1: BRI dimension taxonomy. Dimensions are ordered by weight within category. *Conditional dimensions are activated based on protocol architecture; weight is drawn proportionally from active dimensions to maintain $\sum w_i = 1.0$.

Access Control (D1) receives the highest weight at 0.18. Access control failures are the leading exploit vector by dollar loss in DeFi history. Oracle Integrity (D3) and Economic Soundness (D2) each carry 0.13. Oracle manipulation remains one of the most reliable exploit vectors, while Economic Soundness captures whether a protocol's incentive structure creates profitable attack paths even when all code operates as intended.

Compositional Risk (D4) and Cascade Exposure (D12) represent complementary perspectives on systemic interconnection. D4 measures protocol-intrinsic composition: how many external dependencies exist, how deep the cross-contract interaction surface extends, and whether those integration boundaries carry implicit assumptions. D12 measures extrinsic systemic risk: the upstream propagation of dependency failures into the protocol, and the downstream blast radius if the protocol itself fails. The distinction matters because a protocol can have low compositional risk (few external dependencies) while still carrying high cascade exposure (if it serves as a dependency for many others).

3.3 Scoring Formula

The BRI composite score is computed as a weighted geometric mean, scaled and shifted to the range $[300, 1000]$:

$$\text{BRI} = 300 + 700 \cdot \prod_i \left(\frac{D_i}{100} \right)^{w_i}$$

The choice of geometric mean over arithmetic mean is deliberate. In an arithmetic mean, a protocol scoring 95 across eleven dimensions and 10 on Access Control receives a misleading composite of approximately 858, placing it in the MITHRIL tier. The multiplicative formula produces approximately 702, placing it in the TEMPERED tier. One catastrophically weak dimension drags the composite down in a way that mirrors the

mechanics of real exploitation. An attacker does not need eleven dimensions to be weak. A single exploitable flaw is sufficient.

3.3.1 Sensitivity Comparison

Degraded Dimension	Score	BRI (Multiplicative)	BRI (Additive)
D1 (Access Control)	20	639 FORGED	808 DAMASCUS
D2 (Economic)	20	713 TEMPERED	816 DAMASCUS
D8 (Supply Chain)	20	795 DAMASCUS	833 DAMASCUS
D1 + D2 (both)	20	504 CAST	774 DAMASCUS

Table 2: Sensitivity comparison. All other dimensions set to 95. The multiplicative formula produces appropriate tier degradation for catastrophic failures. The additive formula masks these failures within the DAMASCUS tier in every case examined.

3.4 Confidence Model

Each dimension score carries an associated confidence value $c_i \in [0, 1]$. The composite confidence C is computed as the weighted harmonic mean of dimension confidences. The harmonic mean is chosen because low confidence in any single high-weight dimension should dominate the composite, mirroring the multiplicative score formula's non-compensatory behavior.

3.5 Z-Factor and Actuarial Credibility

The Z-Factor encodes a principle drawn from actuarial science [11] and reliability theory [12]: time in production without catastrophic failure constitutes evidence of safety. The factor follows a saturating curve:

$$Z = \frac{T}{T + \tau} \qquad \text{where } \tau = 180 \text{ days}$$

At six months, $Z = 0.50$. At two years, $Z = 0.80$. The curve approaches but never reaches 1.0. This functional form is a special case of the Bühlmann-Straub credibility formula. The half-life of 180 days was chosen empirically: approximately 50% of protocols that suffer critical exploits do so within their first six months of deployment.

Deployment Age	Z-Factor	Interpretation
1 week	0.04	Minimal credibility
1 month	0.14	Low credibility
3 months	0.36	Emerging credibility
6 months	0.50	Half-life; moderate credibility
1 year	0.67	Substantial credibility
2 years	0.80	High credibility
5 years	0.91	Very high credibility

Table 3: Z-Factor values by deployment age ($\tau = 180$ days).

The Z-Factor deliberately does not modify the BRI score itself. A protocol deployed yesterday with sound design may score BRI = 850 with $Z = 0.02$. The BRI reflects the assessed quality of the code and architecture. The Z-Factor reflects how much actuarial weight that assessment carries. Separating these signals preserves

meaningful comparisons between protocols at different maturity stages and avoids conflating code quality with deployment history.

3.6 The Forge Scale

The Forge Scale maps continuous BRI scores to a 7-tier ordinal rating system designed for rapid communication. The metallurgical naming convention encodes a metaphor for the relationship between security and process: raw metal becomes progressively stronger through heat, pressure, and refinement.

Tier	BRI Range	Description
ADAMANTINE	950–1000	Near-zero adversarial surface. Formally verified, battle-tested, minimal complexity.
MITHRIL	850–949	Extremely robust. Deep layered defenses, minimal attack surface.
DAMASCUS	750–849	Strong security posture with well-managed risks.
TEMPERED	650–749	Acceptable baseline. Identifiable risks with active mitigation.
FORGED	550–649	Material risks present. Additional monitoring recommended.
CAST	450–549	Significant vulnerabilities likely. Not recommended for large exposure.
RAW	300–449	High probability of exploitable flaws. Immediate remediation recommended.

Table 4: Forge Scale tier definitions.

4 On-Chain Oracle

The BRI is published on-chain through a two-contract architecture. The `BROOracle` contract stores and updates protocol scores. The `BRORegistry` contract provides a stable consumer-facing address that routes reads to the active oracle implementation, enabling upgrades without breaking downstream integrations.

Each on-chain score record is storage-optimized into two slots. The first slot packs the composite BRI (16 bits), Forge Scale tier (8 bits), last-updated timestamp (64 bits), sequence number (64 bits), and existence flag (8 bits) into 160 bits. The second slot stores all 12 dimension scores as a packed array of 16-bit values (192 bits total). An evidence hash provides a verifiable link to the off-chain assessment data.

Consumer contracts interact through three read functions. `getScore()` returns the composite BRI, Forge Scale tier, timestamp, and a staleness flag. `getFullScore()` returns the complete struct including all 12 dimension scores and the evidence hash. `getDimension()` returns a single dimension score by index. A batch read function supports multi-protocol queries in a single call.

On-chain publication transforms the BRI from an informational product into a composable DeFi primitive. A lending protocol can set minimum BRI thresholds for accepted collateral. A DEX aggregator can route preferentially through higher-scoring protocols. A vault strategy can automatically reduce exposure when a dependency protocol's score degrades. These integrations are permissionless and require no relationship with BlackHart.

5 Calibration and Validation

The BRI was calibrated against three categories of ground truth: a historical exploit database of over 200 DeFi exploit events (2020–2026) with documented root causes [13], 15 well-characterized protocol anchors

with established security reputations, and cross-validation using an 80/20 train/holdout partition of the exploit database.

Validation across 100 DeFi protocols representing over $200 billion in combined TVL produced a score distribution approximately normal with mean 829 and standard deviation 57. By tier, 27% scored MITHRIL, 61% DAMASCUS, and 12% TEMPERED. Retrospective exploit prediction showed that 10 of 12 historically exploited protocols (83%) would have received dimension scores below 60 on the dimension corresponding to their exploit's root cause.

All 15 calibration anchor protocols scored within one Forge Scale tier of their expected rating. No anchor deviated by more than 75 BRI points from its expected value.

6 The Case for Continuous Assessment

The traditional security lifecycle treats risk assessment as a series of discrete events. A protocol is audited before launch, reviewed periodically thereafter, and monitored through a bug bounty program that incentivizes external researchers to report findings. Between these events, the protocol's risk posture is assumed to be stable.

This assumption was tenable when threats moved at human speed. A skilled researcher might spend weeks to months analyzing a complex protocol before developing a viable exploit. The interval between assessments, while never ideal, was tolerable because the interval between vulnerability introduction and exploitation was long enough to permit detection and response.

That interval is collapsing. The emergence of AI systems capable of autonomous vulnerability discovery, demonstrated by Mythos [3] and operationalized by Daybreak [4], compresses the time from code change to working exploit from weeks to hours. When offensive capability operates continuously, defensive assessment must do the same. A quarterly audit becomes a liability rather than an asset, because the badge of assurance it provides applies to code that may have changed materially since the assessment was performed.

The BRI framework was designed for this environment. Scores update on a six-hour cycle. Each update reflects the current state of the protocol's codebase, its dependency graph, its governance configuration, and any newly disclosed vulnerabilities that affect its dimension scores. The on-chain oracle ensures that downstream consumers always read the most recent assessment, and the staleness flag allows them to detect and handle delayed updates gracefully.

Continuous risk assessment is not an incremental improvement over point-in-time auditing. It is a qualitatively different capability, made necessary by a threat environment that no longer pauses between attacks.

7 Applications

The BRI serves four primary application domains:

- **Insurance pricing.** A continuous, 12-dimensional risk signal for actuarial models with dynamic premium adjustment.
- **Institutional due diligence.** DeFi-native credit ratings enabling score thresholds, dimension-level analysis, and trend monitoring.
- **On-chain risk gating.** The oracle contract allows protocols to condition operations on real-time risk data from their dependencies.
- **Protocol self-improvement.** The dimension-level breakdown identifies specific risk areas and creates a measurable feedback loop between security investment and risk reduction.

8 Limitations and Future Work

The current framework has several known limitations. The initial scored set exhibits selection bias toward established, high-TVL protocols. The dimension taxonomy may prove incomplete as novel vulnerability classes emerge. Weight stability may be affected if attacker methodology shifts significantly. Boundary effects exist at Forge Scale tier transitions.

Planned extensions include temporal derivative signals ($d\text{BRI}/dt$ as a first-class output), systemic risk products (cross-protocol correlation matrices, cascade simulation models, and ecosystem-wide stress tests), and expansion of the scored protocol set beyond the initial 100. These extensions represent a shift from individual protocol assessment to ecosystem-level risk intelligence.

9 Conclusion

The BlackHart Risk Index provides a continuous, multi-dimensional, multiplicative risk scoring framework for DeFi protocols. Its core innovation is non-compensatory aggregation: the weighted geometric mean ensures that catastrophic weakness in any single dimension propagates to the composite score, preventing the masking effects that render additive frameworks unreliable for systems where a single critical flaw enables total loss.

The framework is live, scoring 100 protocols across 12 dimensions with on-chain publication via the `BROOracle` and `BRORegistry` contracts. Scores update every six hours. Full methodology, dimension definitions, and protocol scores are available at blackhart.io/oracle.

References

[1] S. M. Werner, D. Perez, L. Gudgeon, A. Klages-Mundt, D. Harz, and W. J. Knottenbelt, "SoK: Decentralized Finance (DeFi)," *ACM Computing Surveys*, 2022.

[2] N. Atzei, M. Bartoletti, and T. Cimoli, "A Survey of Attacks on Ethereum Smart Contracts," in *Proceedings of POST 2017*, 2017.

[3] Anthropic, "Mythos Preview: Autonomous Vulnerability Discovery at Scale." [Online]. Available: https://red.anthropic.com/2026/mythos-preview/

[4] OpenAI, "Daybreak: AI-Powered Cybersecurity Platform." [Online]. Available: https://openai.com/daybreak/

[5] D. Perez and B. Livshits, "Smart Contract Vulnerabilities: Vulnerable Does Not Imply Exploited," in *USENIX Security Symposium*, 2021.

[6] Codefi, "DeFi Score Methodology." [Online]. Available: https://defiscore.io/

[7] L2Beat, "L2Beat Risk Framework." [Online]. Available: https://l2beat.com/

[8] L. Gudgeon, S. Werner, D. Perez, and W. J. Knottenbelt, "DeFi Protocols for Loanable Funds: Interest Rates, Liquidity and Market Efficiency," in *ACM DeFi Workshop*, 2020.

[9] T. L. Saaty, *The Analytic Hierarchy Process*. McGraw-Hill, 1980.

[10] E. Triantaphyllou, *Multi-Criteria Decision Making Methods: A Comparative Study*. Springer, 2000.

[11] H. Bühlmann and A. Gisler, *A Course in Credibility Theory and Its Applications*. Springer, 2005.

[12] M. W. Bridger, *Reliability: An Introduction to Results and Methods*. Springer, 2003.

[13] L. Zhou *et al.*, "SoK: Decentralized Finance (DeFi) Attacks," in *IEEE Symposium on Security and Privacy (S&P)*, 2023.