

BLAST RADIUS ECONOMICS

One System, Two Instruments

Doug Merritt, Matt Snyder, Collin Weir, Willie Tejada

October 2026

An open framework. Free to adopt, adapt, and extend. No product, platform, or vendor engagement is required to use it.



Executive Summary

Security budgets are built on weakness metrics. Open vulnerabilities, patch time, control coverage, audit findings. Those matter, and none of them answers the question that decides whether an incident becomes a catastrophe, which is what the compromised thing can reach. The industry built rigorous measurement systems for the probability of compromise and the speed of detection. It never built one for consequence.

This paper introduces Blast Radius Economics, one system with two instruments. Reachable Value at Risk, RVaR, prices the exposure. It takes a critical system and asks what a compromise would cost, given what that system can reach and how long unauthorized access would persist. Durable Risk Reduction, DRR, ranks the work. It scores any project on any security roadmap by how much blast radius it removes and how long the removal holds against an adversary who can now iterate at compute speed.

$$RVaR = V \times R \times P \times T$$

what the system is worth $\times$ the share of it a compromise can reach $\times$ how likely compromise is this period $\times$ how long the access lasts

$$DRR \text{ Score} = (\text{Reduction} \times \text{Durability}) \div (\text{Effort} + 3 \times (\text{Debt} + \text{Cost}))$$

how much reach the work removes $\times$ how long the removal holds $\div$ what it costs once + three years of what it keeps costing

The two instruments are one system. A project's Durable Risk Reduction score approximates its RVaR reduction per dollar of total cost of ownership, with the two most contested variables in any risk conversation, business value and breach probability, already cancelled out of the arithmetic. Both instruments read from the same substrate, the observed reachability graph of the environment. Together they give a board an exposure number to track quarterly, and give the CISO and CIO a funding rule that lets structural work win in the same planning conversation where detection projects have always arrived carrying numbers.

Everything here is published to be adopted, adapted, and extended. The paper closes with the boundaries of the argument and the specific empirical result that would prove its central sequencing claim wrong.

1. The question the industry doesn't measure

Every workload in the cloud is born with two doors left open and one locked. The default posture from the major cloud vendors lets a workload reach the internet outbound and talk freely to its neighbors. AWS security groups allow all outbound traffic by default, and every managed Kubernetes service lets pods reach one another until someone writes a policy. The door the cloud does lock, unsolicited inbound access, is the one an attacker needs least.

A serious intrusion has three moves. Reach something, move to something valuable, and send the value back out. The open door between neighbors enables the second move, and the open door to the internet enables the third. The industry spent a decade buying better locks for the front door while leaving the two paths that decide consequence largely unmeasured.

The public record reflects it. Sort the breaches that made the news over the last fourteen months by what would have changed the outcome, and roughly four in five would have been prevented at the inbound edge or contained at the lateral and egress doors: the exposed database that should never have been reachable, the poisoned dependency whose only job was to call home, the foothold that walked to the crown jewels. Wiz found that 72 percent of cloud environments contain a publicly exposed PaaS database.

Yet most security plans still measure weaknesses. Not because anyone believes weakness counts decide outcomes, but because weaknesses are what the industry knows how to score, and scored work beats unscored work in every planning cycle regardless of which is more valuable.

The structure of the problem fits in one line.

Expected Damage = $P(\text{breach}) \times B(\text{architecture})$

thirty years of security investment went to P. This paper measures B.

Prevention tries to lower P. Detection tries to shorten how long an intruder operates unnoticed. Both remain essential, and both address only one factor. B, the set of systems, data stores, and external destinations reachable from the point of initial compromise, is set by architecture before any incident begins, and it is the only factor the defender controls completely, by design, independent of vulnerability count and detection speed.

The industry has never had a standard way to measure B, to price what B costs the business, or to score which projects reduce it durably. This paper supplies those three things.

Category	The number it owns
Vulnerability management	CVEs, severity scores, patch latency
Posture and exposure management	Misconfigurations, exposure findings, attack paths

Category	The number it owns
Endpoint detection and response	Detections, dwell time, time to respond
Security operations	Alerts, incidents, time to acknowledge
Identity	Entitlements, privileged accounts, coverage
Cyber risk quantification	Annualized loss exposure
Consequence, what a compromise can reach	None before this paper. Here: RVaR and Durable Risk Reduction, read from reachability, blast radius, egress coverage, and time to containment

Every category of security spending reports a number it owns. The one that decides whether an incident becomes a catastrophe never had one.

2. Why the old assumptions broke

One of us spent much of a career helping enterprises build the detection model, so nothing here is written lightly. Prevention and detection continue to matter. Neither can carry the full burden anymore, and the evidence is now empirical rather than theoretical.

Exploitation arrives before remediation. In the first half of 2026, per the Zero Day Clock observatory's combined public catalogs, 484 vulnerabilities were newly identified as exploited and 133 were already under exploitation the day they were published. VulnCheck found 28.96 percent of its 2025 known-exploited entries exploited on or before CVE publication, up from 23.6 percent the year before. Attacks move faster than teams respond: CrowdStrike's 2026 Global Threat Report clocks average eCrime breakout at 29 minutes, the fastest at 27 seconds. And most intrusions produce little to detect. The same report finds 82 percent of intrusions were malware free, using valid credentials and built-in utilities through legitimate channels.

AI accelerates the part it touches hardest. In June 2026 Sophos documented an adversary using a frontier model to develop malware against commercial EDR products, building a variant, watching what fires, and adjusting at the speed developers test code. A month later Sophos found the inverse, legitimate AI coding agents tripping the same behavioral rules built to catch intruders. Both findings carry stated limits, one documented case and an early, small dataset. What they show together is that the gap between cheap to change and expensive to change, the gradient that made behavioral detection durable, can now close faster than at any point since David Bianco drew the Pyramid of Pain in 2013.

The cost asymmetry is measurable at the workload level. Defending one cloud workload for a year runs roughly \$1,268 in hard costs across endpoint protection, posture management, native detection, SIEM ingestion, automation, and a share of SOC labor, each component inside public list prices. Substitute your own contract rates, because the conclusion survives any reasonable substitution.

The other side of the ledger is measured in tens of dollars: Incalmo showed in July 2026 that a fully evasive malware variant can be generated for two to thirty five dollars with an open weight model, and 27 percent of the samples evaded all 75 VirusTotal scanners.

Much of the industry has concluded AI resolves this because the same capability reaches defenders. It reaches them, and it does not reach them symmetrically. Offense needs one path while defense needs all paths, and a search accelerator helps the side that needs a single success far more than the side that must eliminate every failure. Offense verifies for free, because an exploit announces success the moment it works, while a patch does not announce safety until production says so. Sergej Epp has formalized this as Verifier's Law, and we adopt his term with attribution because it names the asymmetry precisely. Offense accumulates globally, one exploit working everywhere the software runs, while one deployed fix protects one estate. The decisive asymmetry is consent. AI compresses the adversary's loop toward compute speed with nothing standing in the way. It compresses the defender's loop up to the edge of a human decision to change production, and there it stops, because a wrong defensive action is a self-inflicted outage that somebody has to own. AI generates patches. It does not grant permission to deploy them.

The consent gate now has direct measurements on both sides. Anthropic's June 2026 research on N-day exploitation found a frontier model produced its first working exploit against a Firefox patch within an hour of Mozilla shipping it, eight chains within roughly twelve hours, and against a set of Windows kernel bugs a first proof of concept in 31 minutes, at a total cost near \$15,700 in API credits, while a patch typically takes seven days to reach 90 percent of enrolled Windows devices. The model finished every exploit chain before any device had the fix. Defense shows the same shape on the generation side. OpenAI's Patch the Planet effort with Trail of Bits has AI finding flaws at scale, and as of this writing only about 13 percent of the issues found have patches merged upstream, because validation, triage, and merge approval are human gates that model throughput does not open. Chris Hughes' analysis of enterprise remediation timelines makes the arithmetic plain: writing the fix is roughly 10 to 15 percent of the end-to-end timeline, so even a tenfold improvement in fix-writing yields less than a doubling of remediation speed.

This is why raising remediation effort has not closed the gap. Qualys analysis of the CISA Known Exploited Vulnerabilities catalog, more than a billion remediation records across more than ten thousand organizations from 2022 through 2025, found organizations closed six and a half times more vulnerabilities while the share of critical vulnerabilities still unresolved at seven days worsened from 56 to 63 percent.

More effort produced a worse outcome because effort was never the binding constraint, permission was. And the catalog understates the field: Empirical Security's telemetry observed 16,116 CVEs with exploitation activity across roughly six years while the CISA catalog held 1,307.

The sequencing conclusion the rest of this paper rests on follows from the consent gate. Containment is the only defensive posture that can be bought in advance. The permission is granted once, in a normal change window, while the organization is calm, and it holds through every incident afterward. Nearly everything else on the roadmap has to be purchased during the emergency, at the moment permission is hardest to get. Section 8 states the empirical result that would weaken this conclusion.

3. The two time scales

Two time scales run through every security decision, and reading one as the other is the most common measurement error in the field. The field's most credible measurement project just demonstrated the point on itself. In 2026 the Zero Day Clock, whose headline time-to-exploit curve was among the most cited numbers in security, retired that metric and rebuilt itself as a multi-feed observatory, on the grounds that time-to-exploit ignores aging, saturates at zero once exploitation reaches disclosure day, and moves at both ends simultaneously. We cite the retirement itself: a single speed number could not carry the weight the industry put on it, because different time scales were being read as one.

The incident scale runs in minutes to days. It starts at initial compromise and measures how long unauthorized access persists and how losses grow while it does. Breakout in 29 minutes sits on this scale, and so does the industry's average time to identify and contain a breach, which IBM and Ponemon put at 247 days in 2026, a number that worsened last year and reversed five years of progress. The gap between those two readings is the operational case for architecture that does not wait for detection.

The decay scale runs in months to years, and most measurement systems ignore it entirely. It measures how long a deployed control keeps working before an adversary routes around it. A signature holds until a variant is generated, which now costs tens of dollars. A behavioral rule holds until the adversary learns the environment's normal. A hardness assumption holds until the mathematics falls. A network path that does not exist holds until someone changes the architecture. Every control decays on this scale, and only the rates differ.

Each instrument reads one scale. RVaR's containment term reads the incident scale. Durable Risk Reduction's Durability input reads the decay scale. Both report over the planning period, the quarters across which budgets are set and the probability of compromise is estimated.

4. The exposure instrument: Reachable Value at Risk

Boards have priced financial exposure as value at risk for decades. Cybersecurity never gave them a portable way to connect architecture to business loss, so the CISO arrives with a severity distribution while the board is trying to understand expected consequence.

$$RVaR = V \times R \times P \times T$$

V what the system is worth · R the share reachable from a compromise · P the probability of compromise this period · T how long access persists

V is the value at stake, what the system produces and what interruption or data loss would cost, built with the CFO rather than inferred from a scanner. R is reachability, the portion of that value accessible at runtime from an assumed point of compromise, derived from observed communication paths rather than an architecture diagram. P is the probability of compromise over the planning period, taken from published evidence and an openly stated assumption. T is a containment duration factor, how losses grow as unauthorized access persists, calibrated from exercises and incident economics, with Section 5's durability tiers as a second calibration input.

The output is not actuarial truth, and it does not pretend to be. It compares systems and ranks investments, and it has one property that makes it unusually useful in a budget conversation. Compare the current architecture with an improved one and take the ratio.

THE CANCELLATION

RVaR before ÷ RVaR after

= ($\Psi \times R \times P \times T$) before ÷ ($\Psi \times R \times P \times T$) after

= ($R \times T$) before ÷ ($R \times T$) after

Value and probability appear above and below the line, so they cancel. What remains, reachability and containment time, is what architecture controls, and both are measured in the environment.

The two most contested terms in any risk discussion drop out, and the improvement case rests entirely on measured changes to reachability and containment duration, the two variables architecture actually controls. A CISO does not have to win an argument about breach probability to justify the investment. The argument reduces to before and after readings of the environment's own communication graph.

R is where this framework departs from every diagram-based method. Reachability is derived from observed runtime paths, what workloads actually reach today, and the gap between that and the intended architecture is itself a finding. In most environments it is wide.

RVaR does not replace FAIR or any other established risk quantification method. It isolates the two architecture-sensitive variables that risk teams may estimate but operating plans rarely require anyone to measure. Appendix A maps the relationship.

What a leadership team should be able to state today, for each crown jewel system: its reachable asset and dependency set, the machine identities holding standing access into it, the share of outbound paths governed by explicit communication policy, the measured time to contain a compromise, and whether recovery infrastructure sits inside or outside the same reachable environment. Walk into planning with those numbers, and you are negotiating what the business can afford to lose. Walk in with a better dashboard, and you are negotiating a set of tools.

5. The durability instrument: Durable Risk Reduction

Every security control has a half-life. Deploy a signature and a variant gets generated. Deploy a behavioral rule and the adversary learns the environment's normal. Most of what the industry funds decays faster than the contract renews, gets rebuilt on the same ground the next year, and is reported as progress. A smaller category of control does not decay that way, because what it defends is a fact about the architecture rather than a pattern an adversary can out-iterate. What the industry has never had is a way to tell the two categories apart inside a budget meeting. Durable Risk Reduction is that instrument, and the decay the half-life observation names is what its Durability input scores.

That gap, not a knowledge gap, is why blast radius reduction has been the correct prescription since John Kindervag published the zero trust model in 2010 and still has not been built at most organizations. Every CISO understands blast radius. The constraint is that structural work could not be scored, and unscored work loses to scored work regardless of which is more valuable. Detection got its scoreboard almost by accident. MITRE published ATT&CK in 2015 as a descriptive catalog of observed adversary behavior, and the industry built a measurement system on top of it: coverage percentages became posture, roadmaps became coverage plans, boards received heat maps. ATT&CK was never designed for that use, it grows with every intrusion written up, and coverage of a technique came to mean owning a tool that claims to detect it. The deeper limitation is that coverage measures detection breadth and says nothing about blast radius. An organization can instrument every technique in every tactic and still run a flat estate where one compromised credential reaches everything.

Three tiers of durability

Bianco's Pyramid of Pain ranked indicators by what abandoning them cost the adversary, and it worked because adversary R&D was expensive. Section 2 documented what happened to that gradient. Sorting controls by what actually sustains them, three tiers emerge.

Tier	What sustains it	What defeats it	Examples
Pattern-based	Recognizing the adversary	A frontier model iterating variants at compute speed	Signatures, indicators, generic behavioral rules
Assumption-based	A hardness assumption	The mathematics falling	Encryption, password length, hashing
Absence-based	Nothing is exposed to defeat	Architectural drift only, a door labeled closed that was standing open	A network path that does not exist, a revoked privilege, an expired credential

Pattern-based controls depend on recognizing the adversary, and Section 2 measured their decay. Assumption-based controls are durable by construction, but their durability rests on a hardness assumption, and assumptions have half-lives too. In July 2026 Anthropic disclosed that a Claude model independently derived a key recovery attack against HAWK, a lattice-based signature scheme in NIST's post-quantum process, in roughly sixty hours of model time, and the HAWK team withdrew the scheme a day later. No production system was affected, since the attack targets a challenge parameter, and the timeline is the point: a candidate built to survive computers that do not yet exist left the field in a day. Controls in this tier score high and must carry a re-score trigger. Absence-based controls carry no assumption at all. There is nothing to cryptanalyze and no pattern to out-iterate, and they fail only when the architecture itself has a gap.

The field evidence for the top tier comes from sources with no commercial interest in providing it. In August 2026 CrowdStrike published the architecture it runs to contain its own offensive AI agents in production: seven nested layers, with layer one a network boundary running default-deny egress, so that when every inner layer fails the agent still has no route out. Their governing principle is to contain by construction rather than by trust, and the author matters, because the industry's benchmark for detection engineering put a network boundary ahead of its own platform. Anthropic's May 2026 disclosure on containing Claude across products makes the same separation, constraining what an agent is able to do rather than only supervising what it chooses to do, and reports a red-team exercise in which a malicious prompt induced credential exfiltration in 24 of 25 retries and only environmental containment held. In July 2026 Unit 42 documented an autonomous attack system that scanned twenty five thousand instances of a single platform, retargeted itself after evaluating its own failure, and never triggered an alert. Both of its exploitation attempts failed because the configuration each exploit required was absent. Against a human adversary, difficulty is a real cost because patience is finite. Against a system sampling thousands of targets without fatigue, difficulty costs nothing, and absence is a different property from difficulty.

The score

Five inputs, scored one to five by whoever owns the work, put every project on common ground regardless of which team it came from. Reduction is how much of the blast radius the project removes, a single narrow path at the low end, the enterprise estate at the top. Durability is how long the result holds before an adversary routes around it, five for structural controls that match no pattern, three for behavioral detection anchored to a validated baseline, one for a hash, an indicator, or a static rule. Effort is the build cost, paid once. Debt is the ongoing labor left behind, tuning, triage, vendor management. Cost is the ongoing dollars, licensing, subscriptions, infrastructure.

$$\text{DRR Score} = (\text{Reduction} \times \text{Durability}) \div (\text{Effort} + 3 \times (\text{Debt} + \text{Cost}))$$

how much reach the work removes × how long the removal holds ÷ what it costs once + three years of what it keeps costing

Debt and Cost recur and Effort does not. Most security tooling stays in production about three years before replacement, so a point of recurring burden is worth roughly three points of one-time effort. Three is a default rather than a law, and any team should substitute its own refresh cycle. Report the result in bands rather than decimals, because five integer inputs cannot support two decimal places of meaning. Below 0.15 is a sandcastle. Between 0.15 and 0.50 is weak. Between 0.50 and 1.50 is durable. Above 1.50 is a seawall.

Two inputs need protection from optimism. Durability is a forecast about adversary behavior and Reduction is a claim about how much the work removes, and both are easier to round up than an invoice is. Security architecture signs off on Reduction and Durability while the requesting team owns Effort, Debt, and Cost, so the number that decides the ranking is not graded by the person whose project is being ranked.

One contrasting pair shows the instrument working. Both projects are networking-owned, multi-quarter, and expensive, so neither wins by being cheaper. Rd is Reduction, D Durability, E Effort, Dt Debt, C Cost.

Project	Rd	D	E	Dt	C	Score	Band
East-west visibility platform, detection only	1	2	4	4	5	0.06	Sandcastle
Microsegment the crown-jewel production estate	4	5	5	3	2	1.00	Durable

One produces alerts that decay under adversary pressure and the other produces a structural fact that does not. The segmentation program carries a Debt of three deliberately, because any team that has run one knows the policy drift, the exceptions, and the review cycles. The framework does not hand structural controls an artificially low burden. It rewards them because Reduction and Durability are both high, and that advantage survives honest accounting. Note also which project tops the full reference set in Appendix B: identity work, revoking standing admin access in favor of just-in-time elevation, at 2.50. Blast radius has network, workload, and identity components, and the instrument is indifferent to which team's work reduces it.

A design note on the numerator

An earlier version scored only durability against cost of ownership, and it inverted its own thesis. Closing a single open firewall rule outscored microsegmenting the crown-jewel estate by roughly three to one, because both are structural while one is trivially cheap. A CIO following that version faithfully would book the cheap structural wins and defer the segmentation program another year. The missing variable was how much a control contains, and a framework built for blast radius reduction cannot omit blast radius from its own arithmetic. Reduction fixes it. Cheap structural wins remain worth taking first when available, and they are not a substitute for the architecture. We keep this history in the paper because a framework that caught itself being wrong once, and published the correction, is more trustworthy than one that claims it was born finished.

6. The identity that binds them

The two instruments are one system, and the connection is arithmetic rather than thematic. RVaR's improvement ratio rests on measured changes to reachability and containment duration, with value and probability cancelled. Durable Risk Reduction's numerator, Reduction times Durability, is that same reduction carried across the decay scale, and its denominator is total cost of ownership. So a project's score approximates its RVaR reduction per dollar, with the contested variables already out of the equation. On the balance sheet, RVaR is the stock of exposure and Durable Risk Reduction is the rate at which a funded project pays it down. One instrument states the liability, the other is the capital allocation rule.

DRR Score $\approx$ RVaR reduction per dollar of total cost of ownership

THE SYSTEM ON ONE PAGE

THE LIABILITY	THE PAYDOWN
Reachable Value at Risk. Priced per crown-jewel system. What a compromise would cost, given what the system can reach and how long access persists. The board tracks the trajectory quarterly.	Durable Risk Reduction. Scored per project across the funded plan. How much blast radius the work removes and how long the removal holds, against its full cost. The teams fund by the distribution.

One substrate under both: the observed reachability graph of the environment. Neither instrument reads from a diagram.

The identity produces the operating workflow. Score the entire backlog with Durable Risk Reduction, which requires no argument about probability or business value and can be done by the teams who own the work in a single planning cycle. Then, for the top two or three projects, multiply value and probability back in and present the full RVaR before and after to the board. The board sees dollars of exposure and its trajectory. The operating teams see a ranking they can defend without a risk- modeling debate.

The substrate is shared literally, and it is the framework's maturity path. On day one, Reduction is a one-to-five judgment signed off by architecture. As the practice matures, the same graph data that feeds RVaR replaces the judgment with a measured change in reachable value, and the rubric becomes a measurement. The scale is designed to be replaced by the thing it estimates.

One design decision deserves its reasoning stated, because the obvious move is the wrong one. The formulas stay separate. Folding Durable Risk Reduction into $V \times R \times P \times T$ would re-import value and probability into the funding decision, and excluding them is the instrument's chief operating virtue, the reason teams can score a backlog without the two arguments that stall every risk conversation. What is unified is the thesis, the vocabulary, the substrate, and the reporting. That is Blast Radius Economics.

7. Who holds the number

Blast radius is not a security team decision. Network segmentation is built by IT operations, identity systems are owned by IT operations, and cloud infrastructure is provisioned by engineering. A CISO can advocate in every planning cycle and still lose, because the CISO does not control the decisions that determine blast radius. That is an accountability mismatch, and a measurement system alone does not repair one. Four roles have to hold the same number.

Role	Owns	The move
CISO	The metric	Score the existing plan now, everything in flight, queued, and up for renewal. Reprioritize by score regardless of team ownership. Require architecture sign-off on Reduction and Durability. Report RVaR and the score distribution to the board alongside detection metrics, not in place of them.
CIO	The architecture	Blast radius becomes a standing item at architecture review, beside performance, cost, and availability. Every workload provisioned without segmentation defaults, every service account with standing privilege, every AI tool deployed with broad access is a scored decision, made visible.
CEO	The investment thesis	Require scores on the security portfolio first, then extend the standard to adjacent infrastructure investment as the rubric proves itself. A project that cannot be scored is a project whose durability has not been examined.
Board	The question	Ask for the exposure number and the score distribution every quarter, and track the trajectory rather than the absolute level. Four quarters of direction is a better leading indicator than any single reading.

The reporting standard: RVaR for each crown-jewel system, the Durable Risk Reduction score distribution of the funded plan, the percentage of the environment reachable from one compromise across both network and identity paths, and the measured time to isolate. Produce the reachability number through workload attack path assessment, automated graph analysis, or a tabletop exercise starting from the most probable initial access scenario. The method matters less than establishing a baseline tracked quarter over quarter.

8. Honest boundaries, and what would prove this wrong

A measurement framework earns trust by stating what it cannot do.

The scores are ordinal. Durable Risk Reduction multiplies and divides one-to-five judgments, and the risk quantification community has spent a decade documenting how ordinal arithmetic misleads when dressed up as measurement. That criticism is correct, and this framework is designed around it rather than against it. The output is reported in four bands, not decimals. Two of the five inputs are signed off by a party other than the requester, calibration is by worked example, and cross-organizational comparison is not meaningful without shared calibration, so the score is an internal prioritization instrument rather than a benchmark. The maturity path in Section 6 exists to retire the ordinal scale where it matters most.

RVaR's output is not actuarial truth. V is built with the CFO and is as good as that conversation. P is an openly stated assumption. T is the weakest term, calibrated from exercises and incident economics, and the durability tiers improve it without perfecting it. The framework's defense is the cancellation property, which removes the two softest terms from the investment case entirely. What remains is measurable in the environment.

Containment itself has boundaries. It does not prevent initial compromise. Its connectivity floor is greater than zero, because workloads have to talk to something, and it concentrates risk in hub workloads that must communicate widely, which need the tightest discipline. It carries real operational complexity, which progressive deployment mitigates without eliminating: monitor mode first, validation against real traffic, enforcement activated per workload group, rollback always available. Nothing here argues for cutting detection. Pattern-based controls still catch most phishing, most stolen credential use, and most exploitation of known CVEs. Compliance regimes and insurers will keep requiring them. The argument governs where the next dollar and the next engineering hour go.

One empirical question decides whether the sequencing claim in Section 2 holds. Does AI compress the time to deploy a change safely, or only the time to generate one. Everything rests on the claim that the human consent gate on production change is the real remediation ceiling and that no model lifts it.

The measure to watch is the median time from patch availability to verified production deployment in regulated and uptime-sensitive environments. If that number falls sharply over the next four quarters, the case for funding structural work ahead of remediation weakens considerably, and whoever uses this framework should re-score in public, ourselves first. The argument would not collapse, because the 82 percent of intrusions that run on valid credentials have nothing to patch, but the ranking would change, and we state the test in advance so readers can hold us to it.

9. Adoption

Nothing here requires a product. The formulas, the inputs, the rubric, and the reporting standard are published to be adopted, adapted, and extended by any organization, consultant, or vendor, the way Bianco's pyramid was, rather than controlled by anyone.

The first ninety days are the same in any organization. Calibrate on two or three projects everyone already agrees on before scoring anything contested. Score the existing plan, everything in flight, queued, and up for renewal. Build V for the top crown-jewel systems with the CFO, and derive R from observed paths rather than the diagram. Report the distribution and the exposure numbers. Fund the highest scores first regardless of which team owns them or what they cost up front.

The detection era asked how fast an organization could find the adversary, and it built a real discipline around a real question. The question that decides outcomes now is how much the adversary can reach, and that is set by architecture in advance rather than measured in damage afterward. The industry has answered hard moments before: it built the SOC when nothing watched the watchers, and it built a common language for adversary behavior when every team spoke its own. What has been missing is a way to price consequence and score durability in the same conversation where everything else already has a number. Both are solvable, and this paper is offered as the solution to take, adapt, or improve.

Appendix A. Relationship to FAIR and cyber risk quantification

FAIR quantifies risk as the frequency and magnitude of loss events, built from calibrated estimates across a full taxonomy, and nothing in this paper competes with it. In FAIR terms, RVaR's V corresponds to loss magnitude, P to loss event frequency, and R and T decompose the portion of magnitude that architecture determines, which FAIR handles inside its magnitude and secondary loss estimates but does not require anyone to measure in the environment. RVaR's contribution is isolating those two architecture-sensitive variables and insisting they be measured rather than estimated, because they are the two terms a defender can change by design and verify by observation. The cancellation property then removes the estimated terms from the investment case, which a full probabilistic model, correctly, does not do, because its purpose is stating total risk rather than ranking architectural interventions. A team running FAIR should keep running it. RVaR feeds it better inputs on the architecture side, and Durable Risk Reduction ranks the projects that change those inputs. Neither instrument produces loss distributions, and neither pretends to.

Appendix B. Calibration reference set

Score these, or local equivalents, before scoring anything contested. They anchor the scale because teams already agree on them. Rd is Reduction, D Durability, E Effort, Dt Debt, C Cost.

Project	Rd	D	E	Dt	C	Score	Band
SIEM correlation rule, one technique	1	1	2	4	2	0.05	Sandcastle
East-west visibility platform, detection only	1	2	4	4	5	0.06	Sandcastle

Project	Rd	D	E	Dt	C	Score	Band
Close one standing open firewall rule	1	5	1	1	1	0.71	Durable
Microsegment the crown-jewel production estate	4	5	5	3	2	1.00	Durable
Standing admin access to just-in-time elevation	4	5	2	1	1	2.50	Seawall

Expect two failure modes. Teams inflate Reduction on projects that touch many systems without removing any path, which is what separates a visibility platform from a segmentation program. And teams inflate Durability on behavioral detection anchored to a baseline nobody has validated. Architecture sign-off on those two inputs catches both.

Appendix C. The time scales on one page

The incident scale runs from compromise to containment, in minutes to days. RVar's T lives here. Reference readings: average eCrime breakout 29 minutes, fastest observed 27 seconds, average time to identify and contain 247 days.

The decay scale runs from a control's deployment to its circumvention, in months to years, and shorter every year. Durable Risk Reduction's Durability input lives here. Reference readings: a fully evasive malware variant for two to thirty five dollars, a post-quantum candidate withdrawn after sixty hours of model cryptanalysis, and a network path that does not exist, which this scale cannot touch.

The planning period runs in quarters, the horizon over which P is estimated and budgets are set. Both instruments report on it, RVar as an exposure trajectory and Durable Risk Reduction as the score distribution of the funded plan.

The rule: never let one scale answer for another. A control that shortens the incident scale may still decay quickly on the decay scale. Each number reported carries its scale with it.

Sources

The statistics in this paper are drawn from the following published sources. CrowdStrike, 2026 Global Threat Report, malware-free intrusion share, breakout times. CrowdStrike, Secure Development Practices for Offensive AI Agents, August 4, 2026. The Zero Day Clock project (Sergej Epp and signatories), observatory data on 2026 exploitation counts and the published rationale for retiring the time-to-exploit metric. VulnCheck, State of Exploitation 2026, known-exploited catalog counts and exploitation-at-publication rates. Empirical Security, How to Deal with Speed, exploitation telemetry versus catalog coverage. Sergej Epp, Verifier's Law. Qualys Threat Research Unit, The Broken Physics of Remediation, March 2026, analysis of more than one billion CISA Known Exploited Vulnerabilities remediation records across more than 10,000 organizations, 2022 through 2025, remediation volume and seven-day unresolved rates.

IBM Security and Ponemon Institute, Cost of a Data Breach Report 2026, time to identify and contain. Wiz, cloud environment exposure research, publicly exposed PaaS databases. Sophos, research reported June and July 2026, frontier-model iteration against EDR and coding-agent detection collisions. Incalmo, The danger frontier: low-cost, evasive, abundant malware, July 29, 2026. Anthropic, research on measuring LLM impact on N-day exploitation, June 2026. Anthropic, Investigating Incidents in Cybersecurity Evaluations, July 28, 2026. Anthropic, post-quantum cryptanalysis disclosures, July 28, 2026. Anthropic, How we contain Claude across products, May 25, 2026. OpenAI and Trail of Bits, Patch the Planet, upstream patch-merge rates as of September 2026. Chris Hughes, Resilient Cyber, analysis of enterprise remediation timelines and fix-writing share, 2026. Palo Alto Networks Unit 42, autonomous attack campaign analysis, July 30, 2026. MITRE ATT&CK, published 2015. David Bianco, The Pyramid of Pain, 2013. John Kindervag, No More Chewy Centers, Forrester, 2010. Per-workload defensive cost built from public vendor list pricing and United States SOC analyst compensation surveys, retrieved August 2026.

About Aviatrix

Aviatrix® is the network containment company. The Aviatrix Containment Platform sets what every cloud workload may reach before any attack, so a compromised workload does not become a company-wide breach. Built on the Cloud Native Security Fabric, it enforces that policy in the network, at every workload and in every cloud, with nothing to install and nothing to detect first, and encrypts the traffic it allows. It is simpler, safer, faster and less costly than the firewalls and agents it replaces. The Cloud Native Security Fabric is quantum safe. Trusted by more than 500 of the world's leading enterprises. For more information, visit aviatrix.ai.