COMPANION PAPER 3

How to Score the 4D Cybersecurity Framework

A step-by-step method for scoring your own organization, a provider or a product



IQ4hire Research · November 2026
Proposed and not yet validated

COMPANION PAPER 3

How to Score the 4D Cybersecurity Framework

IQ4hire Research · November 2026 · Proposed and not yet validated

Abstract

The framework paper defines four dimensions and an index. This paper says how to produce the numbers. It covers what to gather, the ten parts to score, the questions to ask (each ends with "show me"), the evidence ladder that caps every score, the arithmetic, how two scorers work, and how to report the result. It ends with a worked example and a blank scoring sheet. The toolbox (framework paper, Appendix C) lists the kinds of tool that earn points and the evidence to ask for.

The one rule. Score what you can show, not what you have written. A claim without evidence is scored at the rung the evidence supports.

About the anchors. The anchors describe the standard IQ4hire sets. They point where attackers are going, so time to protection has no fixed number of minutes: the top score is protection that acts automatically, at machine speed. Facts about attackers are sourced in the evidence paper.

1. Before you start

1. **Set the scope.** Name the organization, the business unit or the product you are scoring. A provider is scored on the service you would buy, not on the company's brochure.
2. **Pick two scorers.** One scorer's judgment is an opinion. Two independent scorers give you a range. Neither should have built the controls being scored.
3. **Collect evidence first.** Gather logs, tickets, configurations, test reports, drill records, contracts and inventory lists before scoring. Scoring from memory produces the claimed level, which the ladder then cuts down.
4. **Use current evidence.** A test or drill counts only if it covers the systems as they are now. Anything from before a major change to those systems is stale. Use the current configuration for settings.
5. **Do not score from a policy alone.** A policy is evidence for rung 1 only.

2. The ten parts

The framework has four dimensions and ten parts. You score each part from 1 to 5, then combine.

Dimension	Parts scored
T, Tools	T (one part)
A, Authority & Seams	A1 edge to network, A2 outbound traffic, A3 behavior outside the playbook, A4 machine and AI identity, A5 handoffs between vendors and connected apps
S, Strike Speed & Offense	S1 time to protection, S2 offense
K, Keeps Adapting	K1 adapt now, K2 quantum readiness

3. The evidence ladder

Every part sits on a rung, and **its score can be no higher than its rung.**

Rung	What you can show	Highest score
1. Written	A policy, standard or procedure	1
2. Owned	A named person who answers for it	2
3. Operating	Evidence that it runs: a log, ticket, report or configuration	3
4. Tested	Someone has tried to break it, and the result is on record	4
5. Timed and rehearsed	A measured response time and a recent drill	5

Final score for a part = the lower of the level the answers claim and the rung the evidence supports.

Example: the answers describe a 5-level control (seam owner, short-lived credentials, automatic protection). The only evidence is a policy document. The part scores 1.

4. The questions

Ask each question, then say "show me." Record the answer, what was shown, and the rung. The columns say what the answer looks like at scores 1, 3 and 5. Scores 2 and 4 fall between.

T. Tools

Question	Score 1	Score 3	Score 5
T.1 Do you have a list of everything you run (servers, cloud, laptops, phones, SaaS, connected devices) with an owner for each? Show me.	No list, or a list nobody maintains	A list covering most of the estate	A trusted inventory, current, with owner, importance and exposure for each asset
T.2 Can you see all of it in one place? Show me.	Blind spots across cloud or endpoints	Several tools, partly connected	One view covering everything in the inventory
T.3 What internet-facing applications, APIs and test copies exist, and who owns each? Show me the list.	Unknown	Known, some without an owner	Complete, every item has an owner, unused ones are retired

T is one number from the three answers. Start from the average, never score it above the lowest answer plus one point, and write down why.

A. Authority & Seams (score each of the five seams)

Seam	Question (end each with "show me")	Score 1	Score 3	Score 5
A1 Edge to network	Who owns what happens between the website protection and internal monitoring?	Nobody; the two never meet	Partly mapped; manual handoff	Named owner; documented handoff; tested
A2 Outbound traffic	What leaves through cloud services and tunnels, and who watches it? Is egress default-deny at the network layer?	Not watched; open egress	Some monitoring; no default-deny	Default-deny egress at the network layer; named watcher; alerts tested
A3 Behavior outside the playbook	When activity matches no runbook, who looks and how fast?	No process	A person looks when someone notices	Named owner; automatic detection of unusual machine behavior; tested with an unscripted drill

Seam	Question (end each with "show me")	Score 1	Score 3	Score 5
A4 Machine and AI identity	How many AI agents and automated processes run here, what can each reach, what can each do without a person, and what credentials do they use? Do any agents write or deploy security controls? If so, who wrote each control, and can the host verify it before it runs?	Shared, long-lived keys; no inventory; no way to tell who or what wrote a control	Inventory exists; some scoped credentials; controls written by agents are reviewed by hand but not signed	Short-lived credential and separate identity per run; least privilege; monitored; every control an agent writes is signed, traceable to the agent that wrote it, and verified by the host before it runs
A5 Handoffs between vendors and connected apps	When one vendor's alert feeds another's, who is in charge? What permissions do connected third-party apps hold?	No owner; apps hold broad permissions	Some written service levels; app list exists	Written service levels and authority; every app cut to least privilege; reviewed

The A score is the average of A1 to A5. Report the lowest seam next to it.

S. Strike Speed & Offense

Part	Question (end with "show me")	Score 1	Score 3	Score 5
S1 Time to protection	For the last critical flaw, how long from learning of it to something in place that stopped the exploit? Did a person have to decide, or did a pre-approved rule act? Did the protection run in observe mode before it blocked, and what evidence turned blocking on? Did it expire when the patch landed? Who holds the kill switch at 3 a.m.?	Protection waits for the patch, which takes weeks	A person decides each time; containment takes hours; temporary controls have no expiry	Pre-approved protection fires automatically at machine speed; it observes first, enforces only on evidence that it is safe for the service, and expires when the patch lands; a human holds the kill switch; each temporary control has an owner and an expiry date; response is timed against the current attacker window
S2 Offense	When did someone last test you like an attacker, and what did it find? When did you last rehearse the 3 a.m. response? Do you hunt for intruders? Do you run traps on your own network?	Never tested; no rehearsal	Tested occasionally, not after major changes; informal rehearsal	Tested on a schedule and after every major change; rehearsal on record; traps or hunting in place

The S score is the average of S1 and S2.

K. Keeps Adapting

Part	Question (end with "show me")	Score 1	Score 3	Score 5
K1 Adapt now	If one system is compromised, what stops it spreading? How fast can you rotate credentials? After an incident, how do you confirm every added credential and permission is gone?	Hard-coded keys; single point of failure	Manual rotation; some segmentation	Rotation and containment in place and tested; full cleanup process on record

Part	Question (end with “show me”)	Score 1	Score 3	Score 5
K2 Quantum readiness	Where do you use public-key encryption? How long must each kind of data stay secret? How long would a migration take? Is there a plan to move to the NIST post-quantum standards, with the longest-lived secrets first?	No inventory; legacy encryption only	Inventory started; not ranked by how long data must stay secret; no migration plan	Inventory complete and ranked by data shelf life; abstraction layer around encryption libraries; hybrid post-quantum encryption protecting long-lived secrets

The K score is the average of K1 and K2.

Why K2 counts. Encrypted data can be copied today and read later. If how long the data must stay secret plus how long your migration takes runs longer than the time to a capable quantum computer, you are exposed now. Nobody can say when that computer arrives, so the score does not wait for a date.

Why A4 includes defender agents. An agent that writes or deploys security controls holds more power than most business agents, and a control with no record of its author cannot be trusted at machine speed. The questions are the same as for any agent (identity, scope, log), plus one: can the host check where a control came from before it runs?

5. The arithmetic

1. Score each of the ten parts as a whole number from 1 to 5 (apply the ladder cap).
2. Dimension scores: T = the T score. A = average of A1, A2, A3, A4, A5. S = average of S1 and S2. K = average of K1 and K2.
3. **4D Index = (T × A × S × K)^(1/4)**, rounded to one decimal.
4. **Weakest dimension:** the lowest of T, A, S and K. If two tie, report both.
5. Report the lowest single part as well, because an average can hide it.

On a calculator: multiply the four, take the square root, then the square root again.

6. Two scorers

1. Each scorer completes the sheet alone, with the same evidence.
2. Compare part by part. A difference of one point is normal. A difference of two or more points needs a conversation: each scorer says what evidence they relied on, and the lower rung wins unless the other produces more evidence.
3. Report both indices and their mean. The gap between them is the range, and a range above 0.5 means the evidence was thin or the questions were read differently.
4. Never average away a disagreement about evidence. If one scorer saw a test report and the other did not, find the report.

7. How to read the result

These bands are a starting proposal, to be calibrated by the pilot described in the framework paper (Appendix B).

4D Index	Reading
Below 2.0	Critical: basic protection is missing or unproven
2.0 to 3.4	Partial: controls exist, with gaps in speed, ownership or testing

4D Index	Reading
3.5 to 4.4	Strong: most parts tested; improve the weakest dimension
4.5 and above	Ready: tested and timed across all four dimensions

The index is a prompt for work, not a grade. The weakest dimension says where to start.

Report these with the index: time to protection and whether a person or a rule produced it, the share of critical flaws fixed within a set time, seams with an owner, the date of the last test, and the share of long-lived secrets under post-quantum protection.

8. Worked example (illustrative)

A mid-size company, scored by two scorers who agreed on every row. The company is invented to show the method. Final = the lower of claimed and rung.

Part	Claimed	What was shown	Rung	Final
T	4	Inventory spreadsheet, last updated 14 months ago, in use	3	3
A1	3	Owner named in a policy; no handoff record	2	2
A2	2	Monitoring on cloud services; egress left open	3	2
A3	3	Procedure for unknown activity; never used	1	1
A4	2	Shared admin key in build scripts; inventory of agents started; no agents write controls	3	2
A5	3	Connected-app list; two apps with broad permissions	3	3
S1	3	Last critical flaw took 9 days; a person added a firewall rule on day 2, with no record of who approved it	2	2
S2	2	One outside test, run before the last major system change; no current evidence	1	1
K1	3	Credentials rotated by hand once a year; a named owner, no records	2	2
K2	1	No inventory	1	1

Dimension scores: T = 3. A = (2 + 2 + 1 + 2 + 3) / 5 = 2.0. S = (2 + 1) / 2 = 1.5. K = (2 + 1) / 2 = 1.5.

4D Index = (3 × 2.0 × 1.5 × 1.5)^(1/4) = (13.5)^(1/4) = **1.9**.

Reading: Critical, just under the Partial band. Weakest dimensions: S and K (tied at 1.5). Lowest parts: A3, S2 and K2 (1 each). First moves: run a one-hour 3 a.m. rehearsal (lifts A3 and S2), pre-approve one protection rule and name the kill-switch holder (S1), and start the encryption inventory ranked by data shelf life (K2). Without the ladder, the claimed levels would have produced a higher index.

9. Common mistakes

- **Scoring the policy.** If the only evidence is a document, the part is 1.
- **Scoring the tool you own, not the tool you run.** A security product that is bought and not switched on, or set to log and not block, earns the rung of what runs.
- **Counting a stale test.** A test from before a major change to the systems no longer describes them.
- **Scoring the provider's brochure.** Ask for the last real incident and how long protection took, and whether a person or a rule produced it. A provider's "within minutes" is a claim until shown.

- **Treating a human response as the top score.** However fast, a person who reads, decides and approves is slower than automatic protection, and the gap grows as attackers speed up.
- **Trusting an AI-written control because it arrived fast.** A control with no signature and no record of its author scores at the lowest rung, and a shield that blocks before it has been tested can break the service it protects.
- **Leaving quantum for later.** K2 starts with an inventory, and the exposure of long-lived secrets is already running.
- **One scorer.** One scorer is a guess. Use two.
- **Forgetting the weakest dimension.** An index of 3.3 with K at 1 is not a 3.3 organization.

10. Scoring a provider or product

Use the same questions, with these changes.

1. Ask the provider for the evidence, not a description. “Show me the last drill record and the last time-to-protection measurement.”
2. Score only what the provider delivers under the contract. A capability that costs extra scores as absent until it is in the order.
3. Compare providers on the weakest dimension first, then the index.
4. For a managed security provider, A5 (handoffs) and S1 (time to protection) usually decide the outcome: who is in charge when the provider’s alert reaches your team, and does something stop the attack automatically or wait for a person?
5. Run the framework against the provider’s own contract: written service levels, who holds the kill switch, how credentials for its agents are issued, what it does about post-quantum encryption, and, if its agents write controls, how those controls are signed and verified.

11. Blank scoring sheet

Organization or provider: _____ Date: _____ Scorers: _____

Part	Claimed	Evidence shown	Rung	Final	Scorer 1	Scorer 2
T						
A1						
A2						
A3						
A4						
A5						
S1						
S2						
K1						
K2						

T = ____ A = ____ S = ____ K = ____ 4D Index = ____ Weakest dimension = ____ Lowest part = ____

12. Limitations

- The anchors and the bands in section 7 are the IQ4hire standard. They have not been tested against real organizations.
- The question set is one version. A pilot may add, merge or drop questions.
- The T rule (three answers into one number) is a judgment call. A pilot may split T into parts.
- How far two scorers differ on real organizations has not been measured. The range rule in section 6 is a proposal.
- The time-to-protection anchor is reviewed whenever new attacker-speed data is published.
- The A4 signing and S1 observe-then-enforce requirements have not been tested in practice.

Notices

IQ4hire is not affiliated with, endorsed by or sponsored by NIST, the Center for Internet Security, MITRE, Cisco, OpenAI, Microsoft, Verizon, GreyNoise, CERT-In or any other organization named in this paper. Product, framework and company names belong to their owners and are used only to identify the works discussed. Quotations are taken from the published documents, with sources cited. The scoring in this paper is IQ4hire's own method and opinion, and it is open to anyone to repeat or dispute.