11

PART 11 · SECOND EDITION · SEPTEMBER 2026 (FIRST PUBLISHED JULY 2026)

Security

One fence line around nation-state-grade IP, allocation-locked silicon, and recognized war infrastructure. The first decision isn't which controls to buy — it's which adversary you intend to survive.

MARCH 2026 · UAE & BAHRAIN

2 + 1

Drones hit two UAE data centers and damaged a third in Bahrain; the IRGC claimed intent

The threat model stopped being hypothetical. A campus is now war infrastructure, and a facility built for opportunistic theft is designing against the wrong adversary.

DEFENSESCOOP, 2026-03-03 · AS-OF 2026-03

THE REFRAME

The crown-jewel asset is the one that loses without a sound.

Weight theft trips none of security's instincts: the original stays put, the cluster keeps running, a few terabytes leave looking like ordinary output. This can't be a perimeter bolted onto a colo.

ACT I

Pick the adversary, then derive the stack

Every isolation, attestation, and egress control taxes throughput. The security level you name in advance prices that tax — and decides which building you're actually constructing.

THE FRAMEWORK EVERYONE NOW CITES

RAND's Weights Security ladder, in three numbers

SL1–SL5 / OC1–OC5

Five security levels mapped to five adversary tiers — 38 attack vectors enumerated

RAND, SECURING AI MODEL WEIGHTS · AS-OF 2024-05-30

8 of 38

RAND 2024: eight vectors assessed likely infeasible for OC1–OC3; not impossible

RAND, SECURING AI MODEL WEIGHTS, 30 MAY 2024, PRINTED PAGE 18 · AS-OF 2024-05-30

SL1–SL5 map to OC1–OC5

The benchmark: a level is defined by thwarting weight theft within this horizon

RAND RRA2849-1 · AS-OF 2024-05-30

START WITH SERVICE INTAKE

Define the asset, adversary, control owner and failed-admission outcome before accepting the workload.

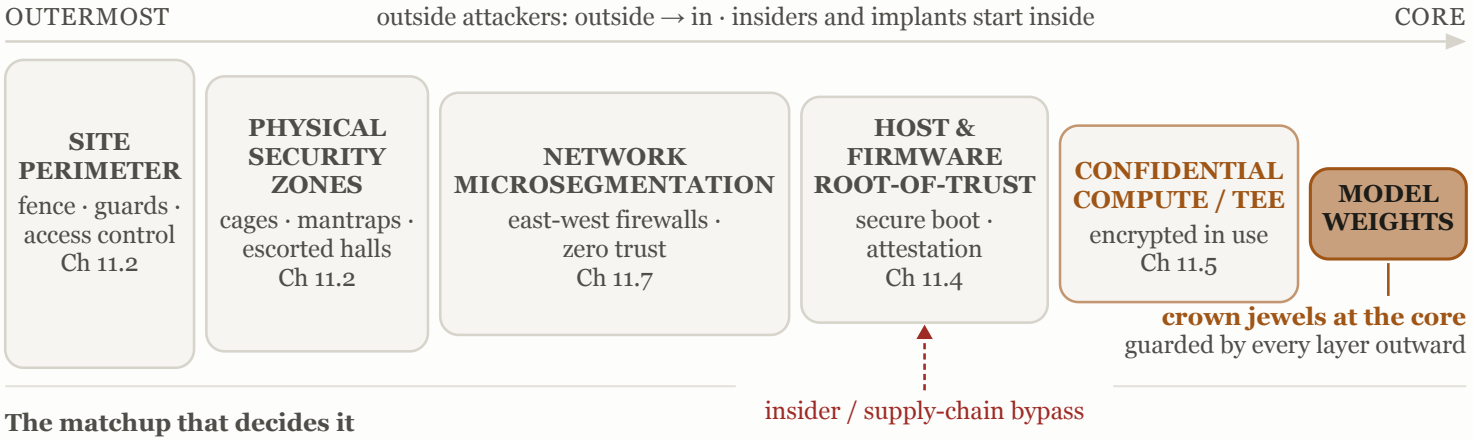
S11 is a constructed service case. Its missing workload-to-key binding keeps proprietary weights outside the cluster. A RAND framework does not certify the trace.

THE TERRITORY

The weights sit at the center; every layer outward defends them

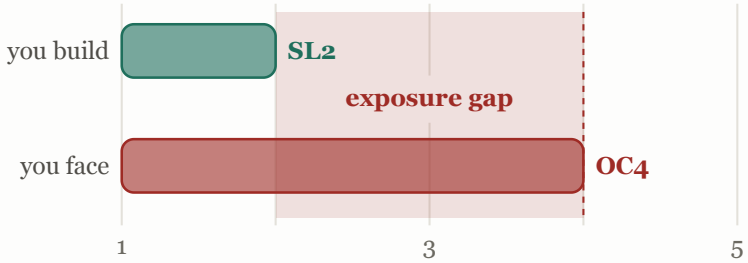
Defense in depth — every layer between the world and the weights

For a proprietary-model service: does your build level match the attacker you face?



The matchup that decides it

How high you build vs. who you face



example pair, not an assessed grade · SL1–5 build · OC1–5 attacker

Layers stop the attacker only if you out-build their tier.

The question to settle first: does the security level (SL) you build match the attacker tier (OC) you face?

Other service profiles put a different asset at the core.

Schematic · Source: RAND (Securing AI Model Weights) + Guide synthesis · as of August 2026

The real question is not how many layers you have — it's whether your build level matches the attacker you actually face, or one two tiers below him.

THE DESIGN-BASIS DECISION

Which asset and adversary does the service accept?

Enterprise inference service

— Define prompt, data and proprietary-weight ownership

— Qualify tenant, administrator and egress boundaries

— Price verifier availability, cleanup and support access

— Exclude higher-threat workloads until intake is reassessed

Frontier-weight service

— Select candidate controls against the named nation-state capabilities

— Hardware root of trust; attestation-gated key release

— Hard egress/near-air-gap; two-person supply-chain vetting

— Test attack paths; price performance overhead, operating drag and supply-chain cost

SO WHAT

Reserve the irreversible provisions the contracted workload needs. Underbuilding leaves an attack path open; overspecification adds operating friction and capital cost.

ACT II

The boundary starts at the fab and the fence

Physical security starts at siting. Supply-chain assurance joins silicon identity, firmware integrity and physical custody through receiving; none substitutes for the others.

CONSEQUENCES OUTSIDE THE FENCE

Protect the substation, fuel and fiber paths that can strand the hall.

Choose barriers, response and alternate-service capacity against the actual adversary. A generic drone price is not a campus protection budget.

THE LEGAL BOX YOU OPERATE IN

FY2026 NDAA

The first US statute letting certified law enforcement defeat a drone — private operators still cannot.

You may legally detect and track a UAS; jamming, spoofing, and kinetic downing stay federal-only. That forces a detect-and-respond posture built on a law-enforcement relationship, not a kinetic one.

SAFER SKIES ACT · AS-OF 2026

THE TARGET OUTSIDE THE FENCE

~1,500 MW

Load a single fault sequence stranded in Virginia — the prize a saboteur reaches without ever entering the building.

Map the actual protected boundary: substation, fuel farm and fiber vault may remain outside it. An interior zone count does not establish delay, response time or protection of these paths.

NERC, INCIDENT REVIEW: LARGE LOAD LOSS · AS-OF 2024-07

THE OTHER PERIMETER

From fab to rack: the attack at each link, the control that answers it

Every link in the chain needs its own evidence

No link's evidence proves the next one. Security starts at the fab, not at the cage.

Fab / OSAT	OEM assembly	Logistics	Integration	Rack	Decommission
die + HBM	board, node	freight, customs	staging, config	accepted fleet	drives + media
Attack at this link					
counterfeit die, re-marked bins	firmware implant at assembly	interdiction in transit	gray-market swap at staging	unverified unit enters the pool	resold drive leaks weights
Evidence that answers it					
component authentication + serial genealogy	signed firmware, measured boot	tamper seals, custody log	manifest vs signed platform certificate	acceptance test: verify or quarantine	Purge or Destroy media (NIST, Sept 2025)

748 suspect-part submissions logged with ERAI in 2025

~24% of suspect parts passed electrical test: a part that boots is not proven genuine

Missing or inconsistent evidence means unverified: quarantine and decide manually; it does not by itself prove counterfeit.

Schematic · Source: ERAI + NIST · as of August 2026

Genealogy, custody, tamper evidence, measured state, and authentication differ. Missing or inconsistent evidence means unverified: quarantine and document disposition; it does not prove counterfeit.

WHY THE LOADING DOCK IS BLIND

The supply chain in two uncomfortable numbers

~24%

Suspect counterfeit parts that passed electrical test — the screen most builds rely on

ERAI 2025 REPORT · AS-OF 2025

42%

42% of 159 eBay used drives held data in one 2019 study, not a data-center rate

BLANCCO TECHNOLOGY GROUP STUDY · AS-OF 2019

ACT III

Below the OS, inside the GPU

Own the root of trust in silicon and you can prove what is running. Cede it — to the BMC, to a flat fabric, to a mis-configured container runtime — and every control you bought above it is defeated from beneath.

THE SECOND COMPUTER WITH GOD-MODE

CVSS 10.0

An auth-bypass in the AMI MegaRAC BMC, OEM'd across a dozen server vendors, added to CISA's KEV list.

The BMC is a separate SoC on standby power, able to remain live when the host is down. Host-memory and firmware reach depend on the platform; test those paths and recovery.

NVD (CVSS V4.0 10.0) · AS-OF 2025-06-25

WHAT A GPU TEE ACTUALLY HIDES

Confidential computing relocates trust to the silicon

~90%

Compute Protected Region: HBM is access-controlled on Hopper and Blackwell, not encrypted

ARXIV 2507.02770 · AS-OF 2025

~99.78%

Memory-mapped registers the BAR0 decoupler hides from the operator in confidential mode

NVIDIA WP-12554 · AS-OF 2025-07

~1–3% tuned serving

Paired B200: tuning changes serving overhead; one-host results are not fleet bounds

ARXIV, INDEPENDENT PAIRED-RUN B200 STUDY
ARXIV:2608.26575V2 · AS-OF 2026-09

THE FULL-STACK LESSON

9.0

A three-line container escape to host root — in the NVIDIA Container Toolkit.

A host escape exposes host-controlled secrets and paths. An independently qualified TEE may retain its boundary; verify the enabled mode, attestation and key-release policy.

NVIDIA PRODUCT SECURITY BULLETIN · AS-OF 2025-07

THE FLAT FABRIC PROBLEM

29 min

Average time from initial access to first lateral movement — on a fabric built with no interior walls.

An AI cluster is one east-west fabric built for bandwidth, not containment. Default-deny egress with an allow-listed proxy is the highest-leverage control for weight theft — and most often left open.

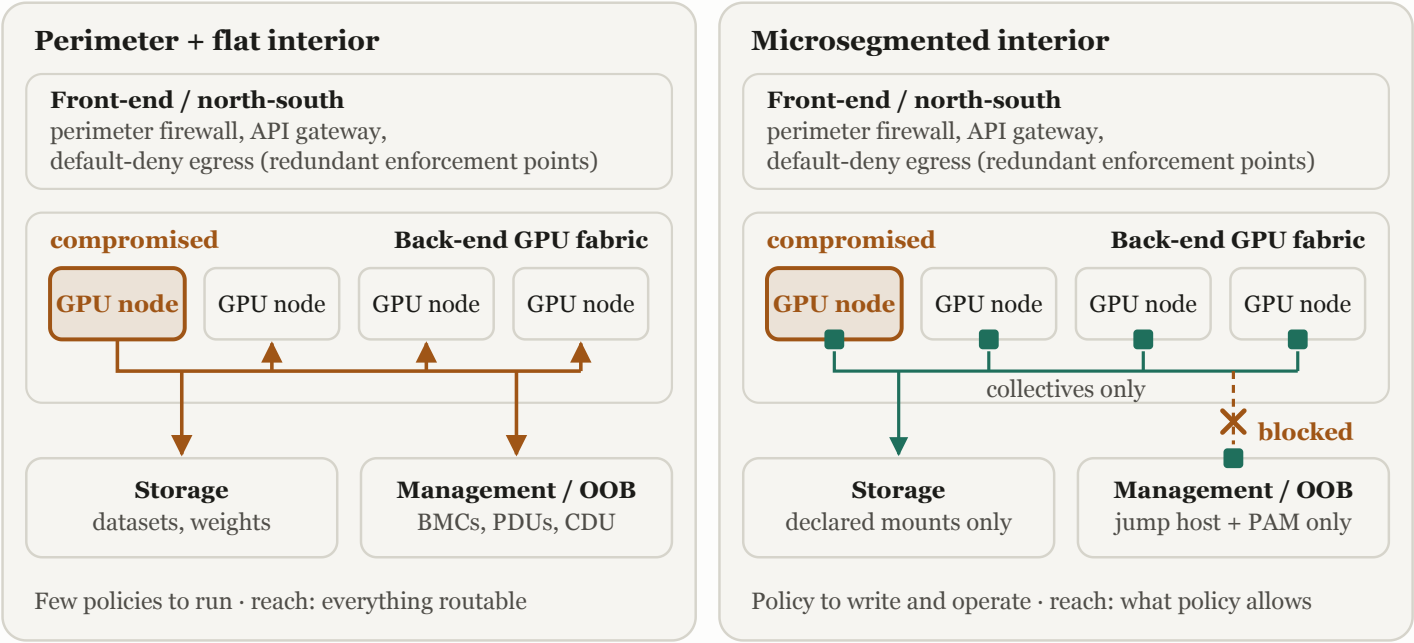
CROWDSTRIKE 2026 GLOBAL THREAT REPORT · AS-OF 2025

WHERE THE WALLS GO

Three fabrics, three trust zones — prove each enforcement boundary

On a flat interior, one compromised node can reach everything routable

Same cluster, same fabrics. Node-edge enforcement limits reach to what policy allows.



compromised node's reach — allowed path — enforcement point (NIC, DPU or switch; jump host for management)

Test each boundary with controls on, including bypass, maintenance and failure states.

DPU enforcement depends on firmware, control plane, configuration, shared resources + adversarial tests.

For egress, declare fail-open/closed, bypass + maintenance behavior before you rely on it.

Schematic · Source: Guide synthesis · as of August 2026

Keep collective-path controls lightweight; enforce at node edges and policy boundaries. Default-deny egress can use redundant paths and enforcement points without creating one failure domain.

ACT IV

The keyholder and the kill-switch

An insider authorized to decrypt can copy plaintext. A compromised control plane can damage equipment when its command path reaches the asset and defeats credited protection.

DEPARTURE IS A RESOURCE-LEVEL TEST

Disabling SSO does not terminate every open session or revoke every local credential.

S11 closes IdP, SSH, storage and BMC paths separately. Keep exit open until the final denial test passes; route suspicious intervening activity to cyber response.

YOU DON'T NEED AN EXPLOIT, YOU NEED A SETPOINT

>600 apartment buildings / ~2 days

FrostyGoop used Modbus process access; distinguish that malware from the controller downgrade.

FrostyGoop issued Modbus commands over TCP port 502. For a BMS or CDU, establish the actual command path, authorization and independent protective functions before assigning damage.

DRAGOS, FROSTYGOOP ICS MALWARE INCIDENT ANALYSIS, JULY 2024 · AS-OF 2024-07-23

THE DUAL-USE FAILURE MODE

>600 apartment buildings / ~2 days

FrostyGoop: 600 buildings lost heat for about two days; test the corresponding OT access path.

Random faults and attacks may share a top event but need distinct evidence and controls. Give protective trips the independence and bounded, monitored reset/override required by the safety analysis.

DRAGOS, FROSTYGOOP ICS MALWARE INCIDENT ANALYSIS, JULY 2024 · AS-OF 2024-07-23

THE DESIGN CONSTRAINT, NOT THE STAMP

Map obligations to the entity, service and data

LAYER	FRAMEWORKS	WHAT IT FORCES ON THE BUILD
Commercial baseline	SOC 2; ISO/IEC 27001	Attestation report and ISMS certification have different scopes
AI governance	ISO/IEC 42001; EU AI Act	Assign provider/deployer roles and the applicable obligations
Federal service	FedRAMP; CMMC	Use the authorization route and actual contract clauses
Sector or operator	NIS2; DORA; HIPAA; PCI DSS	Test legal/service applicability and assign each control owner

THE FALSIFICATION TEST

Exercise each control against a declared attack path. A silent SOC proves neither safety nor blindness; tested coverage, containment and retained evidence decide whether the control works.

TAKE IT HOME

What to do Monday

01 Define service, assets and adversaries; test controls against those paths before claiming protection.

02 Reserve the physical, firmware and key provisions the contracted workload needs; price operating friction.

03 Harden the targets outside the fence — substation, fuel, fiber, water. A saboteur strands the most megawatts for the least effort.

04 Contract for device identity, signed firmware and custody; reconcile all three at receiving and retest the installed build.

05 Anchor trust in silicon; segment and patch the BMC; set egress default-deny with an allow-listed proxy as the anti-exfil linchpin.

06 Shrink the population with standing weight access, and make every grant just-in-time, peer-approved, time-boxed, and recorded.

07 Allocate protection by hazard analysis; verify independence, integrity, bypass and proof tests; converge cyber-physical response.

THE DEFINITIVE GUIDE TO

AI Data *Centers*

This deck is the map. The guide is the territory — every number here is dated, sourced, and kept current in the full chapters.

Ch 11.1 Threat model, assets, and setting a target Security Level

Ch 11.5 GPU confidential computing and trusted execution

Ch 11.8 Model and weight protection at-rest, in-transit, in-use

Ch 11.10 Cyber-physical and destructive attacks on OT

READ THE FULL PART, WITH EVERY SOURCE AND DERIVATION, AT AIDATACENTERGUIDE.COM

SOURCES

Every figure in this deck traces to the guide's dated register (1/2)

DefenseScoop, 2026-03-03; AWS status update reported by Reuters, 2026-09-16 · as-of 2026-03 · https://defensescoop.com/2026/03/03/commercial-data-centers-drone-warfare-amazon-aws/

RAND, Securing AI Model Weights (RRA2849-1) · as-of 2024-05-30 · https://www.rand.org/pubs/research_reports/RRA2849-1.html

RAND, Securing AI Model Weights, 30 May 2024, printed page 18 (PDF page 28). · as-of 2024-05-30 · https://www.rand.org/pubs/research_reports/RRA2849-1.html

RAND RRA2849-1 · as-of 2024-05-30 · https://www.rand.org/pubs/research_reports/RRA2849-1.html

SAFER SKIES Act (FY2026 NDAA, Pub. L. 119-60); DOJ/DHS interim final rule, 91 FR 41466 (2026-07-06) · as-of 2026 · https://www.govinfo.gov/content/pkg/FR-2026-07-06/pdf/2026-13609.pdf

NERC, Incident Review: Large Load Loss (2025-01-08), pp.7–8; NERC comments to FERC, Docket EL25-49 (2025-04-23), Attachment A (April 17 presentation), PDF p.17 · as-of 2024-07 · https://www.nerc.com/globalassets/our-work/reports/event-reports/incident_review_large_load_loss.pdf

ERAI 2025 report · as-of 2025 · https://www.erai.com/erai_blog/3192/2025_annual_report

Blancco Technology Group study · as-of 2019 · https://blancco.com/blancco-reveals-42-of-used-drives-sold-on-ebay-are-holding-sensitive-data/

SOURCES

Every figure in this deck traces to the guide's dated register (2/2)

NVD (CVSS v4.0 10.0) / CISA KEV alert 25 Jun 2025 / Eclypsium · as-of 2025-06-25 · https://nvd.nist.gov/vuln/detail/CVE-2024-54085

arXiv 2507.02770 (GPU CC Demystified); NVIDIA WP-12554 · as-of 2025 · https://arxiv.org/html/2507.02770v1

NVIDIA WP-12554 / arXiv 2507.02770 · as-of 2025-07 · https://docs.nvidia.com/nvidia-secure-ai-with-blackwell-and-hopper-gpus-whitepaper.pdf

arXiv, independent paired-run B200 study arXiv:2608.26575v2; submitted 2026-08-27, revised 2026-09-01 · as-of 2026-09 · https://arxiv.org/abs/2608.26575

NVIDIA Product Security bulletin (Wiz Research found the flaw) · as-of 2025-07 · https://nvidia.custhelp.com/app/answers/detail/a_id/5659

CrowdStrike 2026 Global Threat Report · as-of 2025 · https://www.crowdstrike.com/en-us/blog/crowdstrike-2026-global-threat-report-findings/

Dragos, FrostyGoop ICS malware incident analysis, July 2024; January 2024 Lviv district-heating incident. · as-of 2024-07-23 · https://hub.dragos.com/report/frostygoop-ics-malware-impacting-operational-technology

SOURCES

About these figures

Figures are resolved by reference from the guide's claim register at build time (Second edition · September 2026 (first published July 2026)) — a number appears here only as it appears, dated and sourced, at aidatacenterguide.com. Newer data may exist: check the live register.