Supplementary Material

ANCHOR: Adaptive Nash–Commitment Defense against Learning Attackers in Smart Environments

Mohammed Nurudeen, Godfred Yaw Koi-Akrofi, Hannah Ayaba Tanye, and Yvonne Gyasi Hayford
Department of Information Technology Studies, University of Professional Studies, Accra, Ghana
Corresponding author: mohammed.nurudeen@upsamail.edu.gh

Supplementary file for the Journal of Reliable Intelligent Environments

S1 Purpose and Relationship to the Main Manuscript

This supplement extends the system model, controller mechanics, attacker-learning protocol, development pathway, held-out evidence, and reproducibility safeguards reported in the main manuscript. It is organized as a technical continuation rather than a separate study. Sections S2–S4 provide the additional modeling and implementation detail needed to reproduce the simulated environment and the ANCHOR controller; Sections S5–S6 document how the controller was developed and then frozen before confirmation; and Sections S7–S9 provide expanded D1 and D2 evidence together with execution-integrity checks.

The central quantities are the same as in the main manuscript: effective risk R_t, deployed control intensity θ_t, latency L_t, energy E_t, switching burden S_t, total operational burden B_t^{op}, and defender utility U_t^D. D1 denotes the 3,300-bundle held-out confirmation campaign and D2 the independent 600-bundle mechanism-ablation campaign. This supplementary record presents the extended model, evaluation traceability, and mechanism-attribution evidence underlying the reported Pareto frontier.

S2 Extended Smart-Environment and Threat Model

S2.1 Operating Regimes

The operating regime is summarized by $z_t = (\lambda_t, \delta_t)$, where λ_t controls attack-effort cost and δ_t controls the penalty associated with detection. The three principal regimes are

$$z^{DF^\star} = (0.85, 0.70), \qquad z^{CT^\star} = (0.55, 0.50), \qquad z^{AF^\star} = (0.25, 0.10), \tag{S1}$$

representing defender-favorable, contested, and attacker-favorable operating conditions. Noisy and extreme variants broaden these regimes during robustness evaluation, while dedicated misspecification conditions perturb the regime information supplied to the controller. This design separates the strategic orchestration problem from online regime estimation and allows the evaluation to isolate how ANCHOR behaves when the strategic context changes.

S2.2 Tactic Families, Intensities, and Campaign Stages

The attacker chooses among six tactic families at three intensities (0.35, 0.65, and 1.00) plus abstention, producing 19 nominal actions. Stage validity restricts each tactic to campaign states in which it is operationally meaningful. Table S1 shows the resulting multi-stage structure.

Table S1: Multi-stage tactic structure used by the learning attacker.

Tactic family	Earliest valid stage	Progress target
Reconnaissance	External	External
Initial access	External	Foothold
Spoofing/manipulation	Foothold	Foothold
Lateral/persistence	Foothold	Privileged
Collection/exfiltration	Privileged	Impact-ready
Disruption/impact	Privileged	Impact-ready

The campaign stage is $\sigma_t \in \{\text{external}, \text{foothold}, \text{privileged}, \text{impact_ready}\}$. A successful undetected action can advance the campaign by at most one stage, so the attacker cannot jump directly from external access to impact. Detection can roll the campaign back one stage with probability

$$\Pr[\sigma_{t+1} = \max\{\sigma_t - 1, \text{external}\} \mid D_t = 1] = 0.12 + 0.25\theta_t. \tag{S2}$$

The rollback rule makes stronger defensive control useful in two ways: it raises the chance of detection and, conditional on detection, increases the chance of reversing campaign progress.

Table S2: Tactic-specific confidentiality–integrity–availability risk profiles.

Tactic	C	I	A
Reconnaissance	0.15	0.05	0.02
Initial access	0.45	0.40	0.25
Spoofing/manipulation	0.20	0.90	0.35
Lateral/persistence	0.40	0.55	0.35
Collection/exfiltration	1.00	0.20	0.05
Disruption/impact	0.10	0.40	1.00

S2.3 Detection and Attack Success

For tactic detectability χ_{m_t} and attack intensity x_t, the event probabilities are

$$p_t^{\text{det}} = \text{sigmoid}(-1.25 + 3.2\theta_t + 1.2\delta_t + 0.65\chi_{m_t} x_t), \tag{S3}$$

$$p_t^{\text{succ}} = \text{sigmoid}(0.55 + 1.55x_t + 0.22\,\text{ord}(\sigma_t) - 2.2\theta_t - 0.75\delta_t). \tag{S4}$$

These relationships encode the intended directional behavior of the environment: stronger defense increases detection and suppresses attack success; higher attack intensity increases the attacker's success pressure; and later campaign stages increase the attacker's ability to realize damaging actions. The detection-penalty component δ_t makes the same attack more difficult in defender-favorable conditions than in attacker-favorable conditions.

S2.4 CIA-Specific Risk Profiles

Table S2 specifies the confidentiality, integrity, and availability impact profile associated with each tactic family. The profiles distinguish, for example, collection/exfiltration as confidentiality-dominant and disruption/impact as availability-dominant, while spoofing/manipulation primarily stresses integrity.

The aggregate risk is $\bar{r}_t = 0.30r_t^C + 0.35r_t^I + 0.35r_t^A$. The defender uses the stored value $\bar{r}_t$ before the transition. After detection, attack success, and the risk-state update, the policy-comparison outcome is $R_t = (\bar{r}_t + \bar{r}_{t+1})/2$. Thus the measured outcome includes the realized transition, while the online control does not depend on the still-unobserved next state.

S2.5 Operational Burden and Defender Utility

The three operational burden components are

$$L_t = 0.035 + 0.20\theta_t + 0.12R_t + 0.06x_t, \tag{S5}$$

$$E_t = 0.025 + 0.35\theta_t^2 + 0.04x_t, \tag{S6}$$

$$S_t = |\theta_t - \theta_{t-1}|, \tag{S7}$$

with total burden

$$B_t^{\text{op}} = L_t + E_t + S_t. \tag{S8}$$

Latency therefore grows with sustained control, realized risk, and attack intensity; energy grows convexly with control intensity; and switching directly measures reconfiguration magnitude. The complementary defender utility is

$$U_t^D = 1 - 1.35R_t - 0.32L_t - 0.27E_t - 0.24S_t - 0.10x_t(1 - D_t). \tag{S9}$$

The main analysis keeps effective risk and total burden as the co-primary interpretable outcomes, while utility provides an integrated systems measure that rewards secure operation without ignoring latency, energy, switching, and undetected attack pressure.

S3 Complete ANCHOR Controller Specification

S3.1 Strategic Anchors

ANCHOR uses a small set of interpretable operating anchors rather than continuously optimizing a new unconstrained action. The Static anchor is $\theta^S = 0.42$. The Nash anchor θ_t^N is computed on the finite defender grid $\Theta_G = \{0, 0.1, \ldots, 1.0\}$ using the attacker actions valid at the current stage. When a pure mutual best response exists, that pair is selected; otherwise the minimum-total-regret grid pair is used. The commitment anchor θ_t^C anticipates the attacker's utility-maximizing response for each defender candidate and selects the defender-favorable candidate. The emergency safety posture is

$$\theta_t^{\text{safe}} = \max\{\theta^S, \theta_t^N, \theta_t^C\}. \tag{S10}$$

This construction gives ANCHOR a low-cost operating region, a stable Static reference, a strategically responsive Nash region, and a commitment-derived safety ceiling that can be invoked when the campaign reaches its most consequential state.

S3.2 Frozen Controller Parameters and Stage Logic

Table S3 reports the controller values used in all held-out D1 and D2 evaluations. These values were fixed before confirmatory execution and were not re-tuned on held-out seeds.

Table S3: Frozen ANCHOR controller parameters used in held-out evaluation.

Parameter	Value
Base low-control posture	0.35
External Static threshold	0.20
External Nash threshold	0.27
Foothold Nash threshold	0.19
Recent-effectiveness weight ω	0.18
Emergency-risk threshold	0.36
De-escalation cooldown	5 steps
Minimum de-escalation change	0.05
Static anchor	0.42

The temporal risk signal is $q_t = \mathrm{clip}(\bar{r}_t + \omega\eta_t, 0, 1)$, where η_t is recent attacker effectiveness. At external stage, low-cost, Static, and Nash-level operation are all available, with thresholds determining the requested posture. Foothold removes the low-cost option and selects between Static and Nash-level operation. Privileged stage maintains at least Static/Nash protection, while impact-ready or emergency-risk conditions invoke the safety posture. This progression makes campaign stage an admissibility constraint: stronger campaign states progressively remove economical but insufficiently protective choices.

Escalation is immediate. Downward movement is accepted only after the cooldown has expired and the requested decrease exceeds the minimum-change threshold. The asymmetric rule avoids delaying stronger protection while damping rapid oscillation back toward lower-control states. In this way, the stage logic determines *where* the controller may operate and hysteresis determines *how quickly* it may move downward after escalation.

S3.3 Ex-ante Strategic Payoffs and Cooldown State

For each candidate defender intensity θ and tactic m of intensity x, let I_m be the CIA-weighted tactic impact and v_m the attack-surface multiplier. The strategic grid uses the predicted risk $\widehat{R}_t = \mathrm{clip}(0.66\bar{r}_t + 0.42 I_m v_m x - 0.36\theta + 0.08, 0, 1)$. Its predicted operating costs are $\widehat{L}_t = 0.035 + 0.20\theta + 0.12\widehat{R}_t + 0.06x$ and $\widehat{E}_t = 0.025 + 0.35\theta^2 + 0.04x$. The predicted defender utility is $\widehat{U}_t^D = 1 - 1.35\widehat{R}_t - 0.32\widehat{L}_t - 0.27\widehat{E}_t - 0.24|\theta - \theta_{t-1}| - 0.10x$. For the attacker, the anticipated utility is $\widehat{U}_t^A = I_m v_m x\, p_t^{\mathrm{succ}} - \lambda_t \kappa_m x^2 - \delta_t p_t^{\mathrm{det}}(0.35 + 0.65x)$. The distinction between these anticipated payoffs and the realized quantities in Section S2 underlies strategic selection and learning updates, respectively.

The executed simulator stores k_t, the number of decisions since the last control change, initially $k_0 = 999$. It resets $k_{t+1} = 0$ when $\theta_t \neq \theta_{t-1}$ and otherwise increments $k_{t+1} = k_t + 1$. Only the condition $k_t < 5$ controls de-escalation; values greater than or equal to five are equivalent for that decision. This implements immediate escalation with a five-decision cooldown for decreases.

S3.4 Contextual Regime Modulation

The frozen controller also applies small regime-family offsets to the external/foothold thresholds and low-control posture. These offsets were retained to test whether contextual modulation contributes beyond the three structural mechanisms. The D2 ablation results in Section S8 show that removing the explicit regime adjustment leaves aggregate burden essentially unchanged and slightly changes risk and utility. The dominant held-out performance therefore comes from Nash anchoring, stage awareness, and hysteresis, while regime modulation acts as a secondary contextual refinement.

S4 Attacker Learning and Nonstationarity Protocol

The attacker's state is

$$s_t = (b_t^\theta, b_t^\eta, b_t^p, \sigma_t, \mu_{t-1}), \tag{S11}$$

where the first three variables are three-level buckets for defender control, recent attacker effectiveness, and recent defensive pressure; σ_t is campaign stage; and μ_{t-1} is the previous tactic code. The Cartesian state representation

Table S4: Development rounds preceding held-out confirmation.

Round	Seeds	Design role and outcome
Round 1	1000–1009	Tested a stringent pointwise security constraint. Security was strong, but operation remained concentrated near the higher-burden strategic region.
Round 2	1100–1109	Replaced the pointwise constraint with an interpolated Nash–Static risk budget. Burden fell modestly while the solution remained close to costly strategic operation.
Stage-aware redesign	1200–1209	Reorganized control around campaign-stage admissibility, strategic anchors, and hysteresis. This produced the selected ANCHOR operating point used for confirmation.

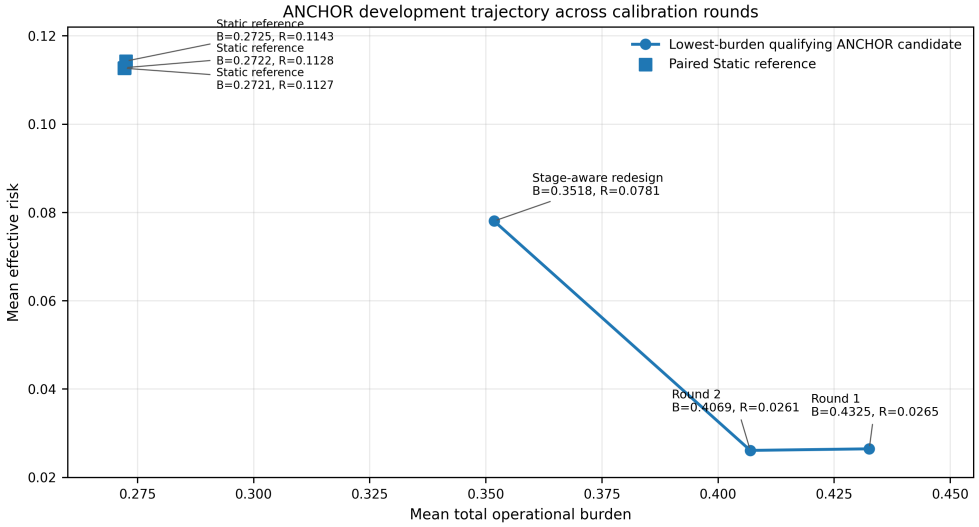

Figure S1: Development trajectory across three calibration rounds. The first two rounds occupy the high-security, high-burden region. The stage-aware structural redesign shifts the selected ANCHOR candidate downward in burden while preserving a substantial security advantage over the paired Static reference.

contains $3 \times 3 \times 3 \times 4 \times 7 = 756$ nominal states. Control bucket boundaries are 0.34 and 0.67, recent-effectiveness thresholds are 0.20 and 0.55, and defensive-pressure thresholds are 0.25 and 0.60. Both recent signals use a rolling window of five interactions.

The primary learner uses learning rate $\alpha = 0.15$, discount factor $\gamma = 0.90$, Boltzmann temperature $\tau = 0.25$, and a 0.05 uniform exploration floor. During drift, the visited state receives per-update forgetting $\phi = 0.03$; stationary and transfer conditions use $\phi = 0$. The realized attacker reward is $g_t = \mathrm{clip}(b_t^A - \lambda_t \kappa_{m_t} x_t^2 - \delta_t D_t(0.25 + 0.75x_t), -1, 1)$, where $b_t^A = I_{m_t} v_{m_t} x_t(0.35 + 0.65Y_t)(1 - 0.20D_t)(0.55 + 0.45R_t)$. Here I_{m_t} is CIA-weighted tactic impact, v_{m_t} the attack-surface multiplier, Y_t the realized attack-success flag, D_t the detection flag, and κ_{m_t} the tactic-specific effort multiplier. The strategic controller evaluates anticipated payoffs before the transition; the learner updates from this realized reward afterward.

The evaluation stresses history dependence in three complementary ways. In stationary learning, the attacker improves while the operating regime remains fixed. In transfer, it trains for 20 episodes in a source regime and warm-starts a 10-episode target phase, covering all six directed transitions among defender-favorable, contested, and attacker-favorable regimes. In drift, the learner remains active over a 600-step trajectory with abrupt, gradual, or recurring regime change. SARSA and Double Q-learning provide additional robustness checks against dependence on the primary Q-learning update. Together, these conditions test whether the defender retains its operating advantage when the attacker learns within a regime, carries knowledge across regimes, or adapts while the environment itself changes.

S5 Development Path and Controller Selection

The final ANCHOR architecture emerged through three calibration rounds that progressively shifted the design objective from strong pointwise security toward a stage-aware security–burden balance. All development used seed sets disjoint from D1 and D2. Figure S1 summarizes the trajectory, while Table S4 records the role of each round.

The trajectory in Figure S1 is important because the movement toward the final operating point is structural rather than a small coefficient perturbation. Round 1 produced mean risk of 0.026467 at burden 0.432544, compared with Static risk 0.112683 and burden 0.272114. Round 2 preserved similarly low risk (0.026108) while

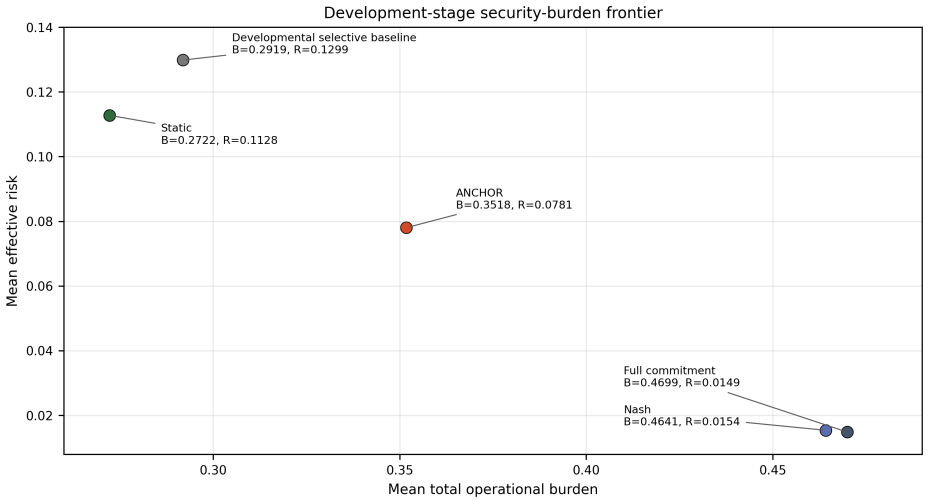

Figure S2: Development-stage security–burden frontier. ANCHOR occupies an intermediate non-dominated point between the low-burden Static policy and the high-security Nash/full-commitment policies. The developmental selective baseline is dominated by Static.

Table S5: Same-seed development frontier used to verify the selected ANCHOR operating point before confirmation.

Method	Risk	Burden	Pareto-efficient
Full commitment	0.014876	0.469900	Yes
Nash	0.015426	0.464131	Yes
ANCHOR	0.078068	0.351763	Yes
Static	0.112800	0.272207	Yes
Developmental selective baseline	0.129884	0.291882	No

Table S6: Seed separation and role of each experiment phase.

Phase	Seeds	Role
Exploratory foundation	Independent production campaign	Established the broader design space and motivated selective strategic control; not used for confirmatory inference.
Development Round 1	1000–1009	Initial constrained controller calibration.
Development Round 2	1100–1109	Alternative security-budget calibration.
Stage-aware redesign	1200–1209	Final controller selection and same-seed development frontier.
D1 confirmation	2000–2029	Held-out policy comparison over stationary, transfer, drift, and extreme conditions.
D2 ablation	3000–3019	Independent held-out mechanism attribution.

reducing burden to 0.406929, but it still operated close to Nash/full-commitment cost levels. The stage-aware redesign moved mean risk to 0.078068 and burden to 0.351763. This created the intended middle operating region: substantially safer than Static while requiring materially less sustained burden than the strongest strategic policies.

The same-seed development frontier in Figure S2 makes this geometry explicit. ANCHOR is non-dominated between Static and Nash/full commitment, whereas the developmental selective baseline lies above and to the right of Static and is therefore dominated on both co-primary dimensions.

S6 Development–Confirmation Separation and Reproducibility Controls

The evaluation protocol separates controller development from confirmatory inference. Before any D1 or D2 seed was executed, the selected ANCHOR configuration, controller source, confirmatory manifests, execution protocol, and D1 analysis code were cryptographically hash-locked and implementation-equivalence checks were passed. The cryptographic records are archived with the reproducibility materials and are intended for public release with the code and data upon publication; these materials document the separation between model selection and confirmatory evaluation.

Table S7: Full block-balanced held-out D1 results.

Policy	Risk	Burden	Utility	Latency	Energy	Switching
Full commitment	.019136	.458142	.833294	.212146	.242387	.003609
Nash	.019665	.452563	.834007	.210772	.238080	.003712
ANCHOR	.078240	.362917	.775088	.173583	.148156	.041177
Static	.118829	.271763	.741980	.163268	.106745	.001750
Dev. selective baseline	.132285	.291755	.718194	.161792	.103484	.026480

Table S8: Primary D1 confirmatory tests.

Contrast	Mean diff.	95% paired-bootstrap CI	Decision
ANCHOR risk − Static	-0.040589	[-0.041391,-0.039844]	Supported
ANCHOR burden − Nash	-0.089647	[-0.091073,-0.088248]	Supported
ANCHOR burden − Full commitment	-0.095225	[-0.096638,-0.093842]	Supported

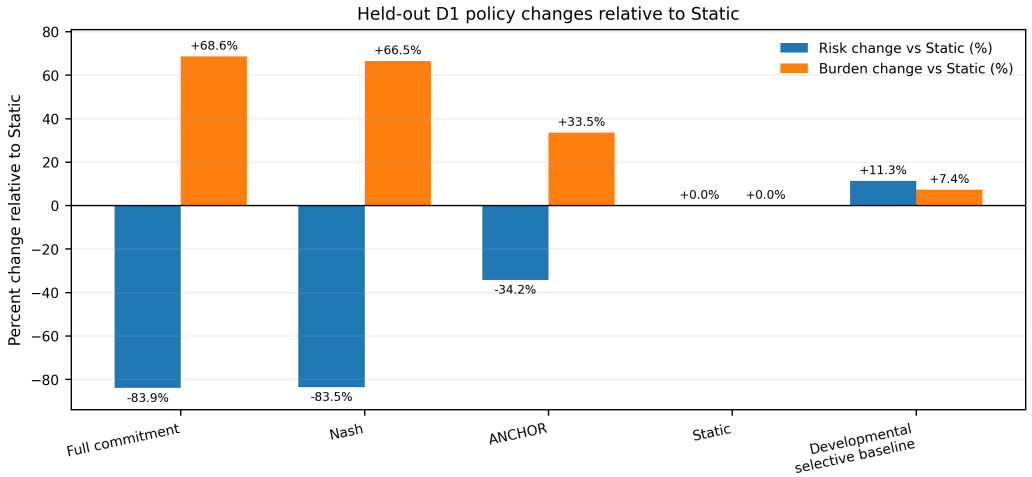

Figure S3: Held-out D1 policy changes relative to Static. Negative risk change denotes improved security; positive burden change denotes additional operational cost. ANCHOR produces a substantial risk reduction with a markedly smaller burden increase than Nash or full commitment.

The seed partitions are fully disjoint. Common random numbers are used only within matched policy comparisons so that policies face the same stochastic realization for a given scenario and seed. This pairing increases precision without treating timesteps or bundles as independent inferential units. Once the controller was frozen, the same configuration was carried unchanged into D1 and D2, preserving a clean boundary between design selection and evaluation.

S7 Expanded D1 Held-Out Confirmation Evidence

D1 contains 3,300 policy–scenario–seed bundles distributed across stationary learning, cross-regime transfer, regime drift, and extreme conditions. Table S7 gives the complete block-balanced means reported in the main analysis. The ordering shows the expected security–operations gradient: full commitment and Nash achieve the lowest risk at the highest sustained burden; Static minimizes burden at higher risk; and ANCHOR occupies a non-dominated middle position with lower latency and energy than the strategic benchmarks.

The three pre-specified directional tests in Table S8 directly evaluate the operating-point claim. ANCHOR's effective risk is lower than Static by 0.040589, while its total burden is lower than Nash by 0.089647 and lower than full commitment by 0.095225. The paired-bootstrap intervals exclude zero in every case, and all three Wilcoxon comparisons remain significant after Holm correction.

Figure S3 normalizes each policy to Static and therefore makes the trade-off particularly easy to interpret. Full commitment and Nash reduce risk by approximately 83.9% and 83.5%, respectively, but increase burden by about 68.6% and 66.5%. ANCHOR reduces risk by about 34.2% while increasing burden by about 33.5% relative to Static. The developmental selective baseline moves in the wrong direction on both axes, increasing risk and burden relative to Static. The relative view therefore complements the absolute frontier in the main manuscript by showing that ANCHOR converts part of the security benefit of strong strategic control into a substantially lower-burden operating point.

Table S9: D2 overall ablation means.

Variant	Risk	Burden	Utility	Switching	Mean control
Full ANCHOR	.077396	.349459	.779416	.037725	.478407
No hysteresis	.098909	.344657	.751217	.066457	.420879
No Nash anchor	.111223	.292542	.746796	.016463	.415957
No regime adjustment	.076275	.349594	.780937	.038516	.478931
No stage awareness	.093644	.336761	.759885	.034639	.456607

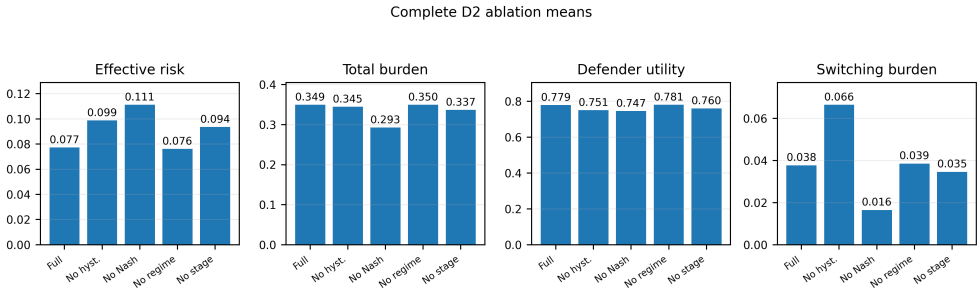

Figure S4: Complete D2 ablation means. The four panels show how removing each mechanism changes effective risk, total burden, defender utility, and switching burden.

Table S10: Paired D2 mechanism tests. Delta is ablation minus full ANCHOR.

Variant	Metric	Delta	95% CI low	95% CI high	Holm p
No hysteresis	Risk	.021513	.021108	.021900	3.05×10^{-5}
No hysteresis	Burden	-.004802	-.006096	-.003481	3.05×10^{-5}
No hysteresis	Utility	-.028199	-.028890	-.027497	3.05×10^{-5}
No hysteresis	Switching	.028733	.027647	.029879	3.05×10^{-5}
No Nash anchor	Risk	.033827	.032863	.034747	3.05×10^{-5}
No Nash anchor	Burden	-.056917	-.058548	-.055178	3.05×10^{-5}
No Nash anchor	Utility	-.032620	-.033587	-.031620	3.05×10^{-5}
No Nash anchor	Switching	-.021261	-.021817	-.020651	3.05×10^{-5}
No regime adjustment	Risk	-.001121	-.001411	-.000833	3.05×10^{-5}
No regime adjustment	Burden	.000135	-.000882	.001271	.674223
No regime adjustment	Utility	.001521	.001171	.001873	3.05×10^{-5}
No regime adjustment	Switching	.000792	.000494	.001132	5.34×10^{-5}
No stage awareness	Risk	.016248	.015878	.016632	3.05×10^{-5}
No stage awareness	Burden	-.012698	-.014316	-.011029	3.05×10^{-5}
No stage awareness	Utility	-.019531	-.020205	-.018917	3.05×10^{-5}
No stage awareness	Switching	-.003085	-.003715	-.002449	3.05×10^{-5}

S8 Expanded D2 Mechanism Evidence

D2 isolates the contribution of four controller components using 600 held-out bundles and paired seeds. Table S9 reports the mean profile of full ANCHOR and each ablation. Figure S4 presents the same quantities graphically so that changes in risk, total burden, defender utility, and switching can be read together rather than as separate one-dimensional effects.

The four panels in Figure S4 reveal distinct functional roles. Removing the Nash anchor produces the largest deterioration in security: risk rises from 0.077396 to 0.111223 and utility falls to 0.746796, even though mean control and total burden fall. Removing stage awareness likewise lowers mean control and burden but raises risk to 0.093644 and reduces utility, showing that campaign progression contributes information not captured by the scalar risk signal alone. Removing hysteresis raises risk to 0.098909 and switching to 0.066457, compared with 0.037725 under full ANCHOR; this combination identifies hysteresis as a security-stability mechanism rather than merely a cost-control device. In contrast, removing the explicit regime adjustment leaves the aggregate profile very close to full ANCHOR, which localizes the dominant performance gains in the three structural mechanisms.

Table S10 and Figure S5 add paired uncertainty estimates. The no-hysteresis, no-Nash, and no-stage-awareness contrasts are all significant across the principal mechanism metrics after the 16-test Holm correction. The only nonsignificant total-burden contrast is the no-regime-adjustment ablation ($p_{\text{Holm}} = 0.674223$). The confidence-interval plot therefore provides the statistical counterpart to the mean profiles: large security losses appear when the Nash anchor, stage awareness, or hysteresis is removed, while the explicit contextual offset has a much smaller aggregate effect.

The D2 evidence provides a layered interpretation of the controller. The Nash anchor supplies strategic escalation strength; campaign-stage awareness determines when stronger operating regions become admissible; and hysteresis governs a stable return toward lower-cost operation. Because the components act at different points in the control loop, their effects are complementary rather than redundant. This mechanism-level result explains why

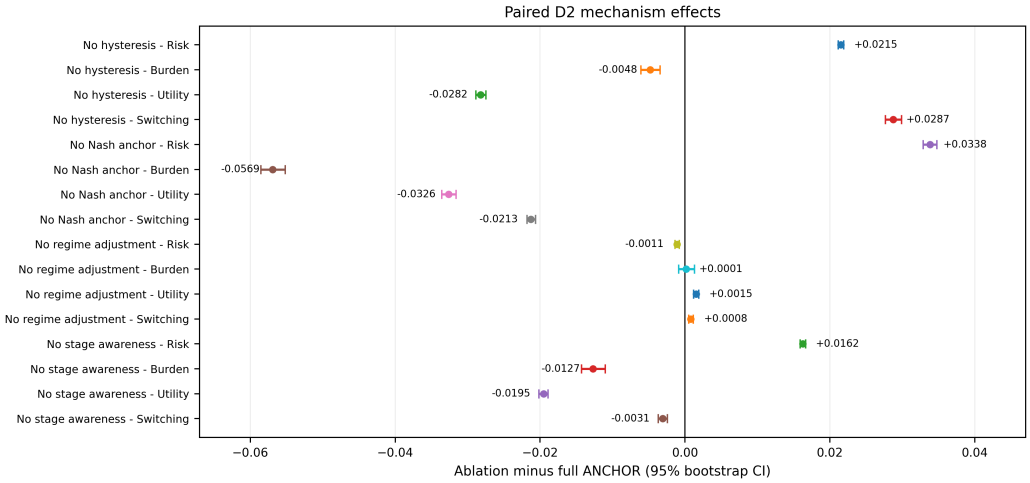

Figure S5: Paired D2 mechanism effects with 95% bootstrap confidence intervals. Each point is the seed-level ablation-minus-full-ANCHOR mean difference. Intervals displaced from zero identify systematic mechanism effects, while the no-regime-adjustment burden effect remains centered near zero.

the complete controller can preserve a favorable risk–burden position even though individual ablations sometimes reduce mean control or burden.

S9 Execution Integrity and Reproducibility

The broader exploratory production campaign contained 19,050 unique bundles and passed release validation before ANCHOR development and confirmation. The held-out extension then completed all 3,300 D1 bundles and all 600 D2 bundles. Re-running the D1 driver after completion reported 3,300 skipped and zero pending runs, confirming resume-safe execution rather than silent duplication.

Policy comparisons use common random numbers within matched scenario/seed pairs. The validation suite checks deterministic replay, campaign-stage prerequisites, stage-valid actions, transfer warm starts, drift routing, CIA-specific risk consequences, learner execution, and manifest uniqueness. Development outputs and held-out outputs are separated so that confirmatory runs do not overwrite or retroactively modify the development record.

The data, source code, experiment configurations, manifests, analysis scripts, lock records, and publication figures have been archived. Consistent with the main manuscript, these materials are intended for public release upon publication. The archived structure therefore supports both computational reproduction of the reported figures and verification of the development/confirmation separation described above.

S10 Supplementary Synthesis

The additional evidence in this supplement connects the controller specification to the held-out results in three steps. First, the extended model shows how campaign stage, tactic validity, strategic regime, detection, success, CIA risk, and operational burden interact within each decision step. Second, the development record shows how stage-aware orchestration moved the selected controller from a high-security/high-burden region toward the intermediate Pareto region before any confirmatory seed was evaluated. Third, the independent D1 and D2 campaigns verify that this operating position persists under stationary learning, transfer, drift, extreme conditions, and targeted mechanism removal.

Taken together, the supplementary results reinforce the main manuscript's interpretation of ANCHOR as a multi-objective adaptive defense architecture. Its held-out advantage is carried primarily by the combination of Nash anchoring, campaign-stage awareness, and asymmetric hysteresis: strategic strength is available when escalation is needed, while the controller avoids sustaining the full latency and energy burden of the strongest posture throughout lower-risk operation. The comparatively small aggregate contribution of explicit regime-threshold modulation further sharpens the architecture by distinguishing the dominant structural mechanisms from secondary contextual refinement.