Multiverse Path Encryption for Securing Interpretation under Adversarial Access

Shaoyuan Wu [1 2]
[1] *Global AI Governance and Policy Research Center*
EPINOVA LLC
[2] *College of Computing*
Georgia Institute of Technology
Atlanta, USA
sywu@epinova.org

Yinning Zhang
College of Computing and Software Engineering
Kennesaw State University
Atlanta, USA
yzhang56@students.kennesaw.edu

***Abstract*—Multiverse Path Encryption (MPE) introduces an interpretation-control paradigm for post-compromise environments in which adversaries may observe ciphertext-derived outputs or queryable system states. Instead of mapping a ciphertext to a unique plaintext, MPE maps a ciphertext–key pair to a traversal policy over a semantic graph, generating a coherent world as output. Incorrect but admissible keys induce alternative traversal policies that produce internally consistent interpretations rather than obvious failure states. This paper formalizes interpretive security as protection against reliable interpretation after access has been obtained. Candidate worlds are generated from a shared semantic structure and constrained by global coherence, so that the intended world is not uniquely identifiable from the adversary's perspective. Experiments in a post-compromise cyber setting show that MPE sustains high target-selection error, high path reconstruction error, and substantial decision-ranking distortion under both heuristic and querying attackers.**

Keywords— Multiverse Path Encryption, interpretive security, cyber deception, post-compromise security, semantic graph, adversarial inference, decision uncertainty

I. Introduction

Access control becomes insufficient once systems are compromised [1], [2]. Traditional cryptography focuses on preventing plaintext recovery, but post-compromise attackers may observe ciphertext-derived outputs or queryable states, shifting the challenge from access prevention to interpretation control. In such settings, adversaries often interact with a system through symbolic outputs, textual traces, or query responses; security therefore depends not only on what information is exposed, but also on whether the intended interpretation can be resolved reliably.

Multiverse Path Encryption (MPE) addresses this problem by treating decryption as key-conditioned generation. A ciphertext-key pair induces a traversal policy over a semantic graph, producing a coherent world rather than a single plaintext. Incorrect but admissible keys generate alternative, internally consistent interpretations instead of obvious failure states.

Because generated worlds satisfy global consistency constraints, adversaries cannot reliably identify the intended interpretation from observable outputs alone without receiver-side authentication. MPE therefore separates adversarial access from reliable interpretation.

This paper introduces interpretive security, formalizes path-dependent world generation, defines world non-identifiability and decision indistinguishability, and evaluates MPE in adversarial cyber scenarios.

II. Related Work and Positioning

Deniable encryption and honey encryption allow incorrect keys to yield plausible plaintexts [3], [4]. Cyber-deception and honeypot systems misdirect adversaries through observation-level manipulation, rather than world-level semantic generation [5]–[8].

MPE does not replace conventional IND-CPA style confidentiality guarantees. It addresses post-compromise settings in which ciphertext-derived outputs or queryable states may become observable, adding an interpretation-control layer that separates access from reliable inference.

MPE differs from prior work in four respects: ambiguity is elevated from messages to coherent worlds; outputs are generated through path-dependent traversal rather than selection; global coherence constraints preserve structural consistency; and ambiguity functions as epistemic control rather than simple deception.

III. System Model

MPE is formalized as an interpretation-control framework for post-compromise settings in which adversaries may observe ciphertexts, generated outputs, or queryable system states. Rather than recovering a single hidden plaintext, decryption performs key-conditioned generation over a shared semantic structure. The same observable ciphertext context can therefore support multiple internally coherent interpretations, while only the legitimate receiver can validate the intended one.

A. System Definition

An MPE scheme is defined as:

$$\mathcal{S} = (\mathrm{G}, \mathrm{Enc}, \mathrm{Policy}, \mathrm{Traverse}, \psi, \Pi, \mathrm{Auth}) \quad (1)$$

where $\mathrm{G} = (V, E, \phi)$ is a semantic multiverse graph; Enc produces a ciphertext C; Policy derives a key-conditioned traversal policy τ_K; Traverse generates a path P_K; ψ maps the path to a coherent world W_K; Π projects query responses from W_K; and Auth enables validation of the intended interpretation by an authorized receiver. This separates representation, control, generation, interaction, and validation.

B. Decryption as World Generation

Decryption is defined as:

$$\tau_K = \mathrm{Policy}(C, K),\ P_K = \mathrm{Traverse}(G, \tau_K),\\ W_K = \psi(P_K),\ M_K(q) = \Pi(W_K, q) \quad (2)$$

Unlike conventional decryption, which recovers a unique plaintext under the correct key, MPE generates a queryable

world. Different keys may generate different interpretations from the same ciphertext context, while each generated world remains internally coherent and stable under repeated querying. The adversary's task is therefore transformed from plaintext recovery to hypothesis selection under structured ambiguity.

C. Semantic Multiverse Graph

The semantic graph G encodes structured domain knowledge, consistent with ontology and knowledge-representation frameworks [9]–[11]. Nodes represent semantic elements such as time, units, locations, actions, resources, dependencies, and operational states. Edges encode admissible relations, including temporal ordering, spatial feasibility, causal dependency, role compatibility, and command alignment.

A traversal path induces a world $W_K = \psi(P_K)$, where ψ is a deterministic closure function that enforces global consistency. A world is admissible only if:

$$\mathrm{Coh}(W_K) = 1 \tag{3}$$

This predicate enforces temporal consistency, role compatibility, spatial feasibility, causal compatibility, and resource constraints. These constraints ensure that generated interpretations are globally coherent, preventing adversaries from eliminating decoy worlds through contradiction-based reasoning or cross-query consistency checks.

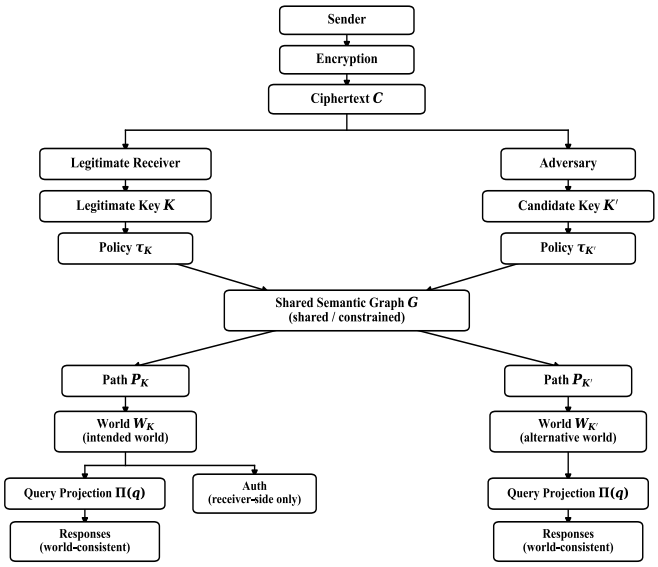

Fig. 1. MPE system architecture under post-compromise access. Ciphertext C defines an observable semantic context. The legitimate receiver uses key K to derive policy τ_K, generating path P_K and intended world W_K. A candidate key K' induces a different policy $\tau_{K'}$ and an alternative coherent world $W_{K'}$. Query projection Π returns world-consistent responses, while Auth enables receiver-side validation of the intended interpretation.

D. Query Projection and Authentication

The projection function Π, defined in (2), supports interactive interpretation by mapping a generated world W_K and query q to the response $M_K(q)$. Because all responses are projected from the same generated world, answer sets remain internally consistent across repeated or related queries.

Since multiple plausible worlds may be generated from the same ciphertext context, MPE includes receiver-side validation:

$$\mathrm{Auth}(W, \sigma_R) \rightarrow \{0,1\} \tag{4}$$

where σ_R denotes receiver-side secret validation material that is not exposed through the adversarial query interface. The validation tag and result are never returned through Π, exposed to the adversary, or made queryable as an oracle. Otherwise, validation would become a distinguishing signal and collapse world non-identifiability.

E. System-Level Interpretation

MPE extends encryption-oriented security into a controlled interpretation architecture: the ciphertext defines an observable semantic context; the key defines a traversal policy; decryption executes constrained traversal; and the output is a coherent, queryable world. Conventional encryption follows: $K \rightarrow M$, whereas MPE follows:

$$K \rightarrow \tau_K \rightarrow P_K \rightarrow W_K \rightarrow M_K(q).$$

Thus, ambiguity is elevated from isolated messages to coherent worlds, shifting the security objective from preventing access alone to preventing reliable interpretation under access.

IV. Keyed Policy Construction

In MPE, the cryptographic key does not directly encode a plaintext or select from a predefined output set. Instead, it functions as a policy seed that induces a traversal strategy over the shared semantic graph. Interpretations are therefore generated dynamically rather than retrieved from fixed storage.

A. Policy-*Induced Traversal*

Given ciphertext C and key K, the system derives a traversal policy τ_K, which governs node selection during graph traversal. The policy can be instantiated using a pseudorandom scoring function:

$$\mathrm{score}(v, S_t) = \mathrm{H}(K \parallel v \parallel S_t) \tag{5}$$

where v is a candidate node; S_t is the current partial world state at step t; H is modeled as a pseudorandom function [12]. At each step, the next node is selected as:

$$v_t = \arg \max_{v \in \mathrm{Candidates}(S_t)} \mathrm{score}(v, S_t) \tag{6}$$

Then, the system state is updated as:

$$S_{t+1} = \mathrm{Update}(S_t, v_t) \tag{7}$$

This process continues until a complete path P_K is constructed and mapped to a coherent world W_K.

B. State Evolution and Constraint Enforcement

Traversal is constrained by the semantic graph and by the current partial state. $\mathrm{Candidates}(S_t)$ filters nodes according to local feasibility, including valid transitions, compatible roles, and admissible dependencies. $\mathrm{Update}(\cdot)$ integrates selected nodes while preserving consistency, and the closure function ψ enforces global coherence.

Thus, incorrect but admissible keys need not produce random failure states. They can induce alternative worlds that remain structurally valid and query-consistent, transforming decryption into policy execution rather than lookup.

C. Distinction from Multi-Output Encryption

MPE is not equivalent to a conventional multi-key or multi-plaintext system. In traditional multi-output schemes, each key corresponds to a predefined output, such as $K_i \rightarrow M_i$.

MPE does not preassign separate plaintexts to different keys. Instead, each key conditions a traversal policy over a shared semantic graph, so the recovered interpretation is generated through constrained traversal rather than selected from a fixed set of stored outputs.

This distinction avoids explicit storage or enumeration of all possible plaintexts. Outputs are generated from a common semantic substrate, and variation emerges from path-dependent decisions rather than discrete selection among predefined alternatives. Different keys therefore induce different trajectories within a unified semantic space.

D. Implementation Considerations

A practical implementation can use PRF-based scoring, constraint-aware candidate filtering, deterministic traversal under a fixed key, and bounded-depth or goal-conditioned traversal. This design ensures that identical keys reproduce identical worlds, different keys generate diverse but valid interpretations, and the system scales without enumerating all alternatives.

The implementation must also prevent authentication leakage. Receiver-side validation must not be exposed as an adversarial oracle, and query responses must not reveal whether a world is intended or decoy. This separation preserves legitimate recoverability without weakening world non-identifiability.

V. EXPERIMENTAL EVALUATION

This section evaluates whether MPE prevents reliable adversarial interpretation under post-compromise access, measuring target-selection error, path-reconstruction error, and decoy-world ranking distortion when ciphertext-defined contexts and candidate worlds are observable.

A. Experimental Setup and Adversary Model

The evaluation considers a post-compromise cyber-defense scenario in which a ciphertext C defines an observable semantic context. The legitimate key K induces the traversal policy τ_K, and the resulting traversal path is closed into the intended world W_K. Candidate keys K' induce alternative traversal policies and coherent decoy worlds $W_{K'}$ rather than random failure states. Auth is used only for receiver-side validation and is not exposed as an adversarial oracle.

The semantic graph contains network segments, hosts, actions, targets, privilege levels, impact levels, dependency edges, and time states. Edges encode privilege transitions, temporal ordering, dependency support, and path feasibility. Each candidate world is represented as a structured tuple over target, host, privilege level, attack action, dependency edge, time state, and impact value. All trials use fixed scoring weights, deterministic traversal under a fixed key, controlled random seeds, and the same adversarial scoring function.

The adversary operates under post-compromise conditions. It can observe candidate worlds, compare internal consistency, evaluate projected responses, rank worlds according to a priority model, and select a target and attack path. Its objective is to identify the intended world and derive the correct operational decision, corresponding here to correct target selection and path reconstruction. A stronger querying attacker is also evaluated; it issues repeated queries to candidate worlds and uses the resulting responses to refine its ranking, while still having no access to Auth.

B. Baselines and Metrics

Two baselines are used. The local perturbation baseline modifies target attributes, path segments, or priority values while preserving local plausibility. The plausibility-optimized baseline generates semantically realistic decoys without key-conditioned traversal. All methods operate under the same observable context and produce query-consistent outputs.

Let T_K denote the intended target associated with the intended world W_K, and let R_K denote the intended attack path associated with W_K. Let $\hat{T}$ and $\hat{R}$ denote the adversary's inferred target and inferred attack path, respectively.

Target Selection Error Rate (TSER) measures the probability that the adversary selects an incorrect target:

$$\mathrm{TSER} = \Pr[\hat{T} \neq T_K] \tag{8}$$

Path Reconstruction Error (PRE) measures the probability that the adversary reconstructs an attack path different from the intended attack path:

$$\mathrm{PRE} = \Pr[\hat{R} \neq R_K] \tag{9}$$

Attack Success Rate (ASR) measures the probability that the adversary correctly identifies both the intended target and the intended attack path:

$$\mathrm{ASR} = \Pr[\hat{T} = T_K \wedge \hat{R} = R_K] \tag{10}$$

Decision Ranking Distortion (ΔS) measures the difference between the priority score of the highest-ranked decoy world and that of the intended world:

$$\Delta S = \max_{W_j \in \mathcal{W} \setminus \{W_K\}} S(W_j) - S(W_K) \tag{11}$$

Here, $S(W_j)$ denotes the adversary's priority score for a decoy world W_j, and $S(W_K)$ denotes the score of the intended world. A positive ΔS indicates that at least one decoy world is ranked above the intended world.

Rank of the Intended World measures the position of W_K in the adversary's ranking over all candidate worlds. Together, TSER, PRE, and ASR capture outcome-level accuracy, while ΔS and rank capture structural distortion in adversarial reasoning.

C. Main Results and Scaling Behavior

Four configurations are evaluated: **low-scale** (100 trials / 10 decoys), **medium-scale** (300 trials / 20 decoys), **high-scale** (1000 trials / 30 decoys), and high-density stress-test (1000 trials / 40 decoys). The first three configurations evaluate scaling behavior, while the final configuration tests robustness under increased hypothesis density. **Table I** reports the main results.

TABLE I. ADVERSARIAL PERFORMANCE ACROSS SCALING AND STRESS-TEST CONFIGURATIONS

Trials / Decoys	TSER (±95% CI)	PRE (±95% CI)	ASR (±95% CI)	Target Accuracy	ΔS	Avg. Rank
100/10	0.660 ± 0.093	0.840 ± 0.072	0.160 ± 0.072	0.34	113.8	5.72
300/20	0.733 ± 0.050	0.903 ± 0.033	0.097 ± 0.033	0.267	134.2	10.96
1000/30	0.780 ± 0.026	0.925 ± 0.016	0.075 ± 0.016	0.22	141.9	16.48
1000/40	0.805 ± 0.025	0.926 ± 0.016	0.074 ± 0.016	0.195	144.4	21.65

Notes: TSER denotes Target Selection Error Rate; PRE denotes Path Reconstruction Error; ASR denotes joint target-and-path attack success; Target Accuracy is reported as $(\mathbf{1-TSER})$; and ΔS denotes Decision Ranking Distortion. Confidence interval s are 95% trial-level estimates.

As shown in **Table I** and **Fig. 2**, MPE maintains high adversarial misdirection as the number of candidate worlds increases. TSER rises from 0.660 under the 100/10 configuration to 0.805 under the 1000/40 high-density setting, while target accuracy falls from 0.340 to 0.195. PRE remains high, increasing from 0.840 to 0.926, indicating that adversarial path reconstruction remains unreliable across configurations.

Scaling produces a structural effect rather than a purely combinatorial effect. ΔS increases from 113.8 to 144.4, and the average rank of the intended world increases from 5.72 to 21.65. Thus, additional candidate worlds primarily distort adversarial ranking and bury the intended world deeper in the hypothesis space. ASR decreases from 0.160 to 0.074, but ASR should be interpreted cautiously because path reconstruction is difficult across all methods. MPE's main security effect lies in amplifying target misdirection and decision-ranking distortion.

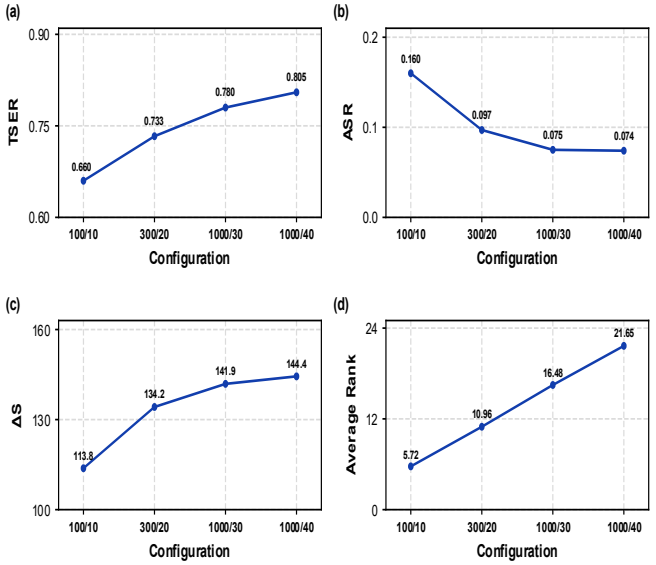

Fig. 2. Adversarial performance metrics across scaling and stress-test configurations. (a) Target Selection Error Rate (TSER). (b) Joint target-and-path Attack Success Rate (ASR). (c) Decision Ranking Distortion (ΔS). (d) Average Rank of Intended World.

D. Ablation and Baseline Comparison

Fig. 3 reports the ablation and baseline comparison under the high-density setting. The ablation study isolates the contribution of key-conditioned traversal and global coherence enforcement. Full MPE produces 41 coherent candidates on average, while removing global coherence reduces this number to 25.14. Full MPE achieves TSER = 0.805, compared with TSER = 0.788 without keyed policy and TSER = 0.781 without global coherence. These results indicate that keyed traversal links ambiguity to the cryptographic key, while global coherence preserves a dense set of structurally valid candidate worlds.

Compared with the local perturbation baseline, MPE produces a much larger coherent interpretation space. Local perturbation achieves comparable target misdirection (TSER = 0.790), but generates only 5.31 coherent candidates on average, compared with 41 under full MPE. Thus, local perturbation can misdirect individual target choices but does not provide the same scalable coherent-world generation mechanism.

The plausibility-optimized baseline is substantially weaker, with lower target misdirection (TSER = 0.337 versus 0.805), higher target accuracy (0.663 versus 0.195), and lower ranking distortion (ΔS = 104.5 versus 144.4). These results show that plausibility alone does not sufficiently disrupt adversarial target selection or priority ranking; MPE's advantage lies in key-conditioned generation of coherent worlds over a shared semantic graph.

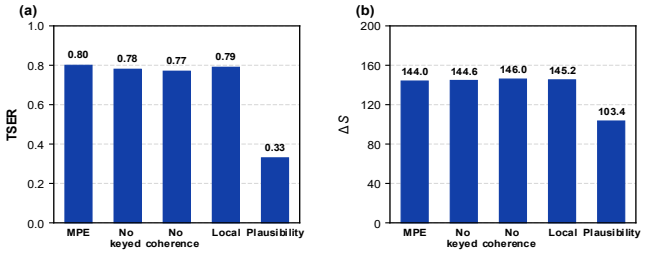

Fig. 3. Ablation and baseline comparison under the high-density setting. (a) Target Selection Error Rate (TSER). (b) Decision Ranking Distortion (ΔS).

E. Strong Attacker and Statistical Significance

A stronger querying attacker is evaluated to test whether repeated interaction collapses ambiguity. This attacker queries candidate worlds, compares projected responses, and updates its ranking based on inferred utility, while still having no access to Auth.

As shown in **Fig. 4**, MPE remains robust under adaptive querying. MPE achieves TSER = 0.810, PRE = 0.925, and ASR = 0.075, while the plausibility-optimized baseline achieves TSER = 0.352, PRE = 0.959, and ASR = 0.041. Although both methods keep joint target-and-path success low, MPE produces stronger target misdirection: target accuracy is 0.190 under MPE, compared with 0.648 under the plausibility baseline. MPE also produces stronger ranking distortion, with $\Delta S = 158.6$ compared with $\Delta S = 113.0$. Because MPE worlds are generated through coherent traversal and closed under global constraints, repeated queries do not expose simple contradictions.

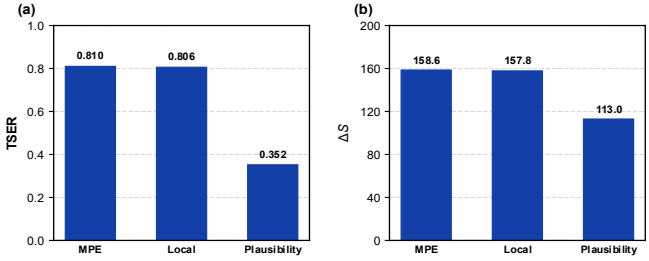

Fig. 4. Comparison with optimized baselines. (a) Target Selection Error Rate (TSER). (b) Decision Ranking Distortion (ΔS).

Statistical tests use two-proportion z-tests for TSER and ASR, Mann–Whitney U tests for ΔS and rank, and Holm correction for multiple comparisons. Under the heuristic attacker, MPE achieves significantly higher TSER than the plausibility-optimized baseline (0.805 versus 0.337, $z = 21.14$, $p = 3.14 \times 10^{-99}$) and greater ranking distortion ($\Delta S = 144.4$ versus 104.5; standardized *U-score* = 13.05, $p = 6.25 \times 10^{-39}$). Under the querying attacker, MPE maintains higher TSER (0.810 versus 0.352, $z = 20.76$, $p = 1.07 \times 10^{-95}$) and stronger ranking distortion ($\Delta S = 158.6$ versus 113.0; standardized *U-score* = 14.49, $p = 1.39 \times 10^{-47}$). These effects remain significant after Holm correction.

Overall, the evaluation shows that MPE transforms adversarial reasoning from deterministic state identification into hypothesis selection under structured ambiguity. Its main effect is structural: coherent decoy worlds can outrank the intended world, burying it deeper in the adversary's ranking and reducing target accuracy. By generating multiple coherent worlds rather than perturbing observations within a single assumed reality, MPE supports the central claim that access need not imply reliable interpretation.

VI. Mechanism of Deception

The results in *Section V* show that MPE does not derive its effectiveness solely from the number of decoy worlds. Rather, its effectiveness arises from a structural change in the adversary's inference problem: observable outputs no longer map uniquely to a single intended interpretation. This section formalizes why MPE preserves ambiguity under post-compromise access, where observation may be extensive but interpretation remains unreliable.

A. Non-identifiability and Decision Uncertainty

MPE separates observability from unique interpretability. Let O denote the adversary's observable information, including the ciphertext-defined context, projected query responses, and visible candidate worlds. Let $\mathcal{W}(O)$ denote the set of coherent worlds consistent with O.

Proposition 1. Non-identifiability under MPE. Under the stated observation model and assuming that Auth is not exposed as an oracle, MPE induces a non-singleton set of coherent candidate worlds $\mathcal{W}(O)$ for an observation O. The intended world belongs to this set, and all candidate worlds are consistent with the observation:

$$W_K \in \mathcal{W}(O), \forall W_i \in \mathcal{W}(O), W_i \vDash O;\ |\mathcal{W}(O)| > 1 \quad (12)$$

Therefore, the observation O alone does not uniquely determine the intended world W_K.

Corollary 1. Decision uncertainty under MPE. Under the same conditions, an adversary cannot reduce the interpretation problem to a single uniquely valid world using observations alone. Any decision rule must therefore select among multiple coherent alternatives rather than recover a uniquely determined state:

$$\mathrm{Adv}_{\mathrm{MPE}} = \Pr\left[\widehat{W} = W_K\right] - \frac{1}{|\mathcal{W}(O)|} \quad (13)$$

For non-uniform priors, the uniform baseline can be replaced by the adversary's prior probability of the intended world. MPE aims to keep this advantage small by ensuring that candidate worlds remain coherent and query-consistent. Thus, even adaptive observation does not necessarily provide a unique mapping from observable outputs to the intended world.

B. Hypothesis Selection and Coherence

In conventional systems, the adversary attempts to reconstruct a hidden system state, and sufficient observations are expected to converge toward a unique interpretation. Under MPE, this assumption no longer holds. Instead, the system exposes a set of mutually incompatible but internally coherent worlds:

$$\mathcal{W} = \left\{W_K, W_{K_1}, W_{K_2}, \dots, W_{K_n}\right\} \quad (14)$$

where W_K is the intended world and $W_{K_1}, \dots, W_{K_n}$ are coherent decoy worlds. Each world is consistent with the observable context, but different worlds support different target selections, paths, and operational interpretations. The inference problem therefore shifts from state recovery to hypothesis selection under structured ambiguity.

A key property of MPE is that candidate worlds satisfy the global coherence predicate defined in (3). Admissible candidate worlds must be globally coherent, not merely locally plausible. As a result, contradiction-based elimination becomes less effective: candidate worlds satisfy feasibility constraints, remain stable across projected queries, and preserve internal consistency. This prevents adversaries from easily pruning the hypothesis space through local inconsistency detection.

C. Priority Distortion and Multi-step Reasoning

Even when multiple worlds remain plausible, an adversary may rank them based on perceived utility. Let $S(W_i)$ denote the adversary's priority score for candidate world W_i, consistent with models of decision-making under uncertainty and strategic trust [13]–[15]. MPE introduces priority distortion by increasing the likelihood that a coherent decoy world dominates the adversary's ranking.

When $\Delta S > 0$, at least one decoy world is ranked above the intended world, leading to incorrect selection. The experimental results in *Section V* show that ΔS remains consistently large under MPE, indicating that MPE alters the adversary's decision function rather than merely expanding the candidate set.

MPE also degrades multi-step reasoning. Let path inference be represented as:

$$\pi = (x_1, x_2, \dots, x_n) \quad (15)$$

where each step depends on prior assumptions. Under MPE, the initial world selection is already uncertain. As reasoning proceeds, early interpretive errors propagate across subsequent steps, leading to divergence in reconstructed paths. Thus, MPE induces systemic instability in sequential reasoning by making the starting interpretation itself ambiguous.

D. Information-Theoretic View and Design Implication

From an information-theoretic perspective [16], [17], MPE preserves a high-entropy interpretation space. Let W denote the random variable over candidate worlds, and let O denote the adversary's observable outputs, including projected query responses. The objective is to keep the information gained from observation small relative to the uncertainty over candidate worlds:

$$I(W; O) \ll H(W) \quad (16)$$

This means that observations do not substantially collapse uncertainty over the candidate-world space. Because candidate worlds are generated through structured traversal and closed under global coherence constraints, repeated queries may increase local detail without eliminating the broader interpretation space or making the intended world uniquely identifiable.

The resulting design implication is that post-compromise security need not rely only on preventing access or hiding information. Even when relevant information is visible, the absence of a unique mapping from observation to the intended world preserves uncertainty. Security can therefore be

maintained by structuring ambiguity so that access does not imply reliable interpretation.

VII. Discussion

MPE shifts post-compromise security from preventing exposure to constraining interpretation. In classical cryptography, the correct key uniquely determines the plaintext, while an incorrect key typically yields failure or unusable output. In MPE, different keys may induce different traversal policies and produce coherent interpretations. The security problem is therefore not only whether access succeeds, but whether the intended interpretation can be reliably identified.

This also distinguishes MPE from conventional deception. Traditional deception manipulates signals within a single assumed reality, and decoys may often be eliminated through inconsistency detection. MPE instead constructs multiple internally coherent worlds over a shared semantic structure. These worlds remain stable under projected querying and are not easily removed through contradiction-based reasoning, forcing the adversary to select among coherent alternatives.

This uncertainty-based view frames MPE security as the preservation of residual uncertainty after observation. Even when candidate worlds are visible and query-consistent, observation does not uniquely determine which world corresponds to the intended interpretation. Coherent decoys may also outrank the intended world, producing misalignment between observation, belief, and action. Thus, MPE preserves world non-identifiability while degrading decision reliability under post-compromise access.

More broadly, MPE reframes security as control over the mapping from observation to interpretation. In access-control systems, the central question is whether an adversary can obtain information. In MPE, the central question is whether an adversary can determine what the information means. After access occurs, the system still shapes whether the intended interpretation remains identifiable.

VIII. Implications, Limitations, and Future Work

MPE reframes post-compromise security from information acquisition to belief resolution: when access is obtained but interpretation remains unstable, multiple coherent worlds fragment adversarial belief and reduce decision reliability.

This work is limited by a constrained semantic graph, a bounded cyber-domain setting, and specific heuristic and querying adversary models. The proposed notions of non-identifiability and decision uncertainty should therefore be read as operational security conditions for interpretation control, not as reduction-based cryptographic guarantees.

Future work will extend MPE to richer domains, larger candidate-world spaces, more adaptive adversaries, and human-in-the-loop evaluation. It will also develop receiver-side authentication mechanisms, formalize indistinguishability games and information-theoretic bounds, and evaluate scalability under larger semantic graphs and multi-turn adversarial querying.

IX. Conclusion

Multiverse Path Encryption (MPE) provides an interpretation-control layer for post-compromise settings in which access cannot always be prevented. Rather than mapping ciphertexts to a single plaintext, MPE uses key-conditioned traversal over a semantic structure to generate coherent, query-consistent candidate worlds. By separating observability from interpretability, MPE prevents observable outputs from becoming reliable evidence of the intended world. Its core security effect is to preserve world non-identifiability and decision uncertainty, transforming adversarial reasoning from state recovery into hypothesis selection under structured ambiguity. These results suggest that post-compromise security can be strengthened not only by limiting visibility, but also by shaping inference conditions so that access does not imply reliable interpretation.

References

[1] C. E. Shannon, "Communication theory of secrecy systems," Bell System Technical Journal, vol. 28, no. 4, pp. 656–715, Oct. 1949, doi: 10.1002/j.1538-7305.1949.tb00928.x.

[2] J. Katz and Y. Lindell, Introduction to Modern Cryptography, 2nd ed. Boca Raton, FL, USA: CRC Press, 2014, ISBN: 978-1-4665-7026-9.

[3] R. Canetti, C. Dwork, M. Naor, and R. Ostrovsky, "Deniable encryption," in Advances in Cryptology—CRYPTO'97, B. S. Kaliski, Ed., Lecture Notes in Computer Science, vol. 1294. Berlin, Germany: Springer, 1997, pp. 90–104, doi: 10.1007/BFb0052229.

[4] A. Juels and T. Ristenpart, "Honey encryption: Security beyond the brute-force bound," in Advances in Cryptology—EUROCRYPT 2014, P. Q. Nguyen and E. Oswald, Eds., Lecture Notes in Computer Science, vol. 8441. Berlin, Germany: Springer, 2014, pp. 293–310, doi: 10.1007/978-3-642-55220-5_17.

[5] N. Provos and T. Holz, Virtual Honeypots: From Botnet Tracking to Intrusion Detection. Boston, MA, USA: Addison-Wesley Professional, 2007, ISBN: 978-0-321-33632-3.

[6] M. H. Almeshekah and E. H. Spafford, "Cyber security deception," in Cyber Deception: Building the Scientific Foundation, S. Jajodia, V. S. Subrahmanian, V. Swarup, and C. Wang, Eds. Cham, Switzerland: Springer, 2016, pp. 23–50, doi: 10.1007/978-3-319-32699-3_2.

[7] J. Pawlick, E. Colbert, and Q. Zhu, "A game-theoretic taxonomy and survey of defensive deception for cybersecurity and privacy," ACM Computing Surveys, vol. 52, no. 4, Art. no. 82, pp. 1–36, Aug. 2019, doi: 10.1145/3337772.

[8] K. J. Ferguson-Walter, D. LaFon, and T. Shade, "The science of cyber deception," IEEE Security & Privacy, vol. 19, no. 2, pp. 36–44, Mar.–Apr. 2021, doi: 10.1109/MSEC.2020.3045875.

[9] T. R. Gruber, "A translation approach to portable ontology specifications," Knowledge Acquisition, vol. 5, no. 2, pp. 199–220, Jun. 1993, doi: 10.1006/knac.1993.1008.

[10] D. L. McGuinness and F. van Harmelen, "OWL Web Ontology Language Overview," W3C Recommendation, Feb. 10, 2004. [Online]. Available: https://www.w3.org/TR/owl-features/

[11] R. Navigli and S. P. Ponzetto, "BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network," Artificial Intelligence, vol. 193, pp. 217–250, Dec. 2012, doi: 10.1016/j.artint.2012.07.001.

[12] D. Boneh and V. Shoup, A Graduate Course in Applied Cryptography, version 0.6, Jan. 2023. [Online]. Available: https://toc.cryptobook.us/

[13] D. Kahneman, Thinking, Fast and Slow. New York, NY, USA: Farrar, Straus and Giroux, 2011, ISBN: 978-0-374-27563-1.

[14] R. Hastie and R. M. Dawes, Rational Choice in an Uncertain World: The Psychology of Judgment and Decision Making. Thousand Oaks, CA, USA: SAGE Publications, 2001, ISBN: 978-0-7619-2275-9.

[15] J. Pawlick and Q. Zhu, "Modeling and analysis of leaky deception using signaling games with evidence," IEEE Transactions on Information Forensics and Security, vol. 14, no. 7, pp. 1871–1886, Jul. 2019, doi: 10.1109/TIFS.2018.2886472.

[16] T. M. Cover and J. A. Thomas, Elements of Information Theory, 2nd ed. Hoboken, NJ, USA: Wiley-Interscience, 2006, ISBN: 978-0-471-24195-9, doi: 10.1002/047174882X.

[17] I. Csiszár and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems, 2nd ed. Cambridge, U.K.: Cambridge University Press, 2011, ISBN: 978-0-521-19681-9.