Post-Alphabet-Dictionary Cryptography

Symbol Identity, Dictionary Distribution, and Cryptographic Structure

Alexey A. Nekludoff

AstraVerge Research

E-mail: an@astraverge.org
ORCID: 0009-0002-7724-5762

June 23, 2026

Abstract

Modern cryptography and cryptanalysis normally assume that the analytical domain has already been established. Plaintext and ciphertext are treated as sequences of stable units drawn from identifiable alphabets, while recurrent units, object boundaries, and cross-session identities remain available for statistical and structural comparison.

This paper introduces *Post-Alphabet-Dictionary Cryptography* (PADC), a framework in which the formation of the observable alphabet, dictionary, statistical profile, and cryptographic-object structure becomes a controllable part of the encoding architecture. A source symbol or dictionary unit may be represented by multiple externally distinct code units, while observable frequencies and higher-order transitions may be shaped according to a selected artificial, uniform, technical, or language-like target process.

PADC distinguishes three levels of controlled non-equivalence: symbol identity, dictionary distribution, and cryptographic structure. Repeated source units need not preserve a stable external identity; the observable dictionary need not preserve the statistical organization of the source; and visible carrier segmentation need not determine the internal units of a cipher, hash, signature, or other cryptographic transformation.

The complete architecture is represented as

$$M \xrightarrow{F_\Theta} X \xrightarrow{E_K} Y \xrightarrow{R_\Phi} O,$$

where F_Θ forms an alternative alphabet-dictionary representation and R_Φ forms the externally observable carrier structure.

A central result is that one observable sequence may remain compatible with multiple mutually incompatible cryptographic-object reconstructions. The observer may therefore lack not only the key or the plaintext, but also a uniquely determined ciphertext object and ciphertext space in which conventional cryptanalysis could begin. Language Segment Encoding (LSEG) is identified as a concrete candidate for the representation function R_Φ because it can preserve recoverable carrier segmentation while leaving higher-level cryptographic-object assembly dependent on a hidden structural state.

The framework is further extended across sessions. If each session forms a distinct observable alphabet and dictionary, observations from different sessions cannot be pooled into one statistical corpus until the adversary reconstructs the cross-session identification maps connecting their analytical

units. More observations therefore do not automatically produce one larger statistical sample.

The resulting security objective is not the elimination of observable regularity, but the prevention of a unique unauthorized stabilization of the alphabet, dictionary, cross-session identity relations, and cryptographic-object structure required to interpret the observed stream. Successful compromise of one transformation reveals only an adjacent representation layer and does not necessarily recover the source message.

Keywords: post-alphabet cryptography; dictionary formation; cryptanalysis; statistical analysis; frequency shaping; non-canonical encoding; observable structure; LSEG; cryptographic architecture.

Contents

1 Introduction

Modern cryptography is usually formulated as a transformation between already established message spaces. A plaintext message is assumed to exist as a sequence of elements drawn from a defined alphabet, after which an encryption transformation produces a ciphertext represented in another defined alphabet. Cryptanalysis then attempts to recover the key, the plaintext, or some structural property of the transformation.

This conventional formulation can be represented as

$$M \in \mathcal{A} * P^* \xrightarrow{,E*K,} C \in \mathcal{A}_C^*, \tag{1}$$

where $\mathcal{A} * P$ is the plaintext alphabet, $\mathcal{A} * C$ is the ciphertext alphabet, and E_K is an encryption transformation determined by a key K.

The formulation appears technically neutral. It nevertheless contains a strong epistemic assumption: the elementary units belonging to both message spaces are already sufficiently stable to be distinguished, compared, counted, and recombined. The alphabet is therefore not merely a notation used after the message has been discovered. It is a condition under which a message can become an object of analysis at all.

Previous work examined the boundary case in which a stable ciphertext alphabet fails to form for an external observer [1]. In that case, standard analytical operations such as frequency counting, pattern comparison, entropy estimation, and known-plaintext matching cease to possess a stable symbolic domain.

The present paper extends this argument from alphabet formation to dictionary formation.

A cryptanalyst does not operate only on isolated symbols. Statistical analysis may be performed on morphemes, words, word pairs, word triples, phrases, headers, protocol sequences, instruction patterns, and other persistent constructions. The relevant analytical object is therefore not only an alphabet but an extended dictionary of recurrent forms.

The dependence of alphabets on dictionaries and of messages on both was examined in *Before Information: Languages, Dictionaries, Alphabets, and Messages* [2]. The dynamics through which language constructions are coordinated, canonized, and separated into diverging dictionaries were considered in *Dynamics of Linguistic Constructions: Coordination, Canonization, and Divergence of Dictionaries* [3].

Building on these results, this paper introduces *Post-Alphabet-Dictionary Cryptography* (PADC). PADC does not assume that an external observer must receive a stable symbolic and lexical projection of the source message. Instead, the formation of the observable alphabet and dictionary is treated as a controllable component of encoding.

The central transformation is therefore not merely

$$M \xrightarrow{,E_K,} C, \tag{2}$$

but

$$M \xrightarrow{,F_\Theta,} X \xrightarrow{,E_K,} Y \xrightarrow{,R_\Phi,} O, \tag{3}$$

where:

- F_Θ forms an alternative alphabet-dictionary representation;
- X is a message expressed in the alternative code dictionary;
- E_K is a cryptographic transformation;
- R_Φ forms the externally observable representation;
- O is the physical or digital sequence available to the external observer.

The observer does not obtain M, X, or Y directly. The observer obtains O and must reconstruct the transformations that connect the observable sequence to the source message.

The proposed framework distinguishes three levels:

1. **symbol identity**;
2. **dictionary distribution**;
3. **cryptographic structure**.

These levels correspond to three different assumptions normally available to cryptanalysis:

$$\begin{gathered}\text{source symbol} \leftrightarrow \text{observable symbol},\\ \text{source dictionary} \leftrightarrow \text{observable distribution},\\ \text{cryptographic unit} \leftrightarrow \text{observable segment}.\end{gathered}$$

PADC replaces these correspondences with controlled non-equivalence:

$$\begin{gathered}\text{source symbol} \not\leftrightarrow \text{single observable symbol},\\ \text{source dictionary} \not\leftrightarrow \text{observable statistical profile},\\ \text{cryptographic unit} \not\leftrightarrow \text{observable segmentation}.\end{gathered}$$

The purpose is not to claim that physical observations cease to exist. An external observer necessarily receives variations, samples, bytes, or other measurement outputs. However, the alphabet of the measuring system must not be confused with the alphabet of the source. A radio telescope may produce a binary data stream without thereby establishing that the observed source contains binary symbols, a dictionary, a language, or a message.

Likewise, the existence of an observable byte stream does not imply that its bytes correspond directly to source symbols, dictionary units, cipher blocks, hash boundaries, or message structures.

The central thesis of this paper is therefore:

> Cryptanalysis requires not only access to observable data, but also the stabilization of analytical units whose identity, distribution, and structural relations can be connected to the source message. PADC makes that stabilization a controllable part of the encoding architecture.

2 Alphabet and Dictionary as Conditions of Analysis

2.1 Alphabet

Let

$$\mathcal{A} = a_1, a_2, \ldots, a_n \tag{4}$$

denote an alphabet of elementary units.

For an external observer, the alphabet is not established merely because physical variations exist. Alphabet formation requires that observations can be divided into stable equivalence classes:

$$o_i \sim o_j \quad \Longleftrightarrow \quad o_i \text{ and } o_j \text{ are treated as occurrences of the same unit.} \tag{5}$$

Only after such equivalence classes have stabilized can the observer count occurrences, identify repetition, estimate frequencies, and construct higher-order sequences.

2.2 Extended dictionary

In this paper, a dictionary is not restricted to a list of conventional words. It is defined as a system of recurrent and stably recognizable composite units.

Let

$$\mathcal{D} = \mathcal{D} * 1 \cup \mathcal{D} * 2 \cup \mathcal{D}_3 \cup \cdots, \tag{6}$$

where:

- $\mathcal{D} * 1$ contains morphemes, words, tokens, or other elementary dictionary units;
- $\mathcal{D} * 2$ contains recurrent pairs of units;
- $\mathcal{D}_3$ contains recurrent triples;
- higher-order sets contain persistent phrases, sequences, templates, or protocol constructions.

The statistical profile of the dictionary includes not only unigram probabilities

$$P(w_i), \tag{7}$$

but also higher-order relations such as

$$P(w_i, w_j), \qquad P(w_i, w_j, w_k), \tag{8}$$

and conditional transitions

$$P(w_j \mid w_i). \tag{9}$$

Consequently, the elimination of elementary-symbol frequency alone does not eliminate dictionary-level statistical regularity.

2.3 The hidden assumption

Cryptanalysis does not require that the alphabet and dictionary be known in advance. It requires that they can be stabilized from observations.

The conventional analytical chain is

$$\begin{aligned} \text{observations} &\to \text{recurrent units} \\ &\to \text{alphabet} \\ &\to \text{dictionary} \\ &\to \text{statistical model} \\ &\to \text{interpretation.} \end{aligned}$$

PADC intervenes before the final interpretation by controlling the formation of the recurrent units and their observable statistical relations.

3 Level I: Symbol Identity Diversification

Let $a_i \in \mathcal{A}_S$ be a source symbol.

Canonical encoding normally assigns one stable code representation:

$$a_i \mapsto c_i. \tag{10}$$

Repeated source symbols therefore produce repeated observable units:

$$a_i = a_j \quad \Longrightarrow \quad c_i = c_j. \tag{11}$$

This preserves symbol identity across occurrences.

PADC replaces the canonical mapping with a family of admissible representations:

$$a_i \mapsto C_i = c_{i1}, c_{i2}, \ldots, c_{im_i}. \tag{12}$$

A particular occurrence is encoded by selecting

$$c_{ij} \in C_i \tag{13}$$

according to an encoding state, context, key, distribution, or selection rule.

Thus,

$$a_i = a_j \not\Longrightarrow \text{Enc}(a_i) = \text{Enc}(a_j). \tag{14}$$

The source identity remains recoverable for the legitimate decoder because

$$c_{ij} \in C_i \quad \Longrightarrow \quad \mathrm{Dec}(c_{ij}) = a_i. \tag{15}$$

For the external observer, however, different representations need not stabilize as occurrences of one source unit.

This level was examined in detail in the earlier analysis of alphabet formation and the epistemic limits of cryptographic analysis [1].

4 Level II: Dictionary Distribution Shaping

4.1 From frequency elimination to distribution formation

Uniform frequency is the simplest special case of dictionary distribution shaping.

Let the source dictionary be

$$\mathcal{D} * S = w * 1, w_2, \ldots, w_n \tag{16}$$

with source distribution

$$p_i = P_S(w_i). \tag{17}$$

Let the observable code dictionary be

$$\mathcal{D} * O = c * 1, c_2, \ldots, c_m. \tag{18}$$

Encoding is described by the conditional matrix

$$T_{ij} = P(c_j \mid w_i), \tag{19}$$

subject to

$$T_{ij} \geq 0, \qquad \sum_{j=1}^{m} T_{ij} = 1. \tag{20}$$

The observable distribution is

$$q_j = P_O(c_j) = \sum_{i=1}^{n} p_i T_{ij}. \tag{21}$$

In vector form,

$$\mathbf{q} = \mathbf{p}T. \tag{22}$$

The encoding matrix T can be selected so that the observable distribution **q** differs from the source distribution **p**.

4.2 Uniform shaping

The simplest target is a uniform observable distribution:

$$q_j = \frac{1}{m}. \tag{23}$$

In this mode, all observable dictionary units occur with approximately equal frequency.

The natural lexical hierarchy of the source language is therefore not preserved in the observable stream.

4.3 Target-profile shaping

Uniformity is not the general objective. In the general PADC framework, the encoder may specify an arbitrary target distribution

$$\mathbf{q}^* = (q_1^*, q_2^*, \ldots, q_m^*)\,. \tag{24}$$

Whether the selected profile can be realized exactly depends on the admissible encoding class.

When the target profile is realizable by the conditional encoding matrix T, it satisfies

$$\mathbf{p}T = \mathbf{q}^*. \tag{25}$$

For a fixed-length, stateless, and disjoint mapping, the realizable target profiles are constrained by the probability mass assigned to the source equivalence classes. More general target profiles may require variable-length sequences, auxiliary units, context-dependent encoding, overlapping representations, or additional encoder and decoder state.

The target distribution may represent:

- an artificial statistical profile;
- a uniform distribution;
- a technical corpus;
- a natural-language corpus;
- a selected genre;
- a protocol or telemetry profile;
- a deliberately misleading dictionary.

The observable stream may therefore possess clear, stable, and highly convincing statistics while those statistics remain unrelated to the source language.

The relevant condition is not

$$P_O(c_j) \text{ does not exist,} \tag{26}$$

but

$$P_O(c_j) \not\Rightarrow P_S(w_i). \tag{27}$$

The observer receives a statistical structure, but the structure belongs to the encoding architecture rather than to the source dictionary.

4.4 Higher-order shaping

If only unigram frequencies are shaped, higher-order structures may remain available:

$$P(c_i, c_j), \qquad P(c_i, c_j, c_k). \tag{28}$$

The general PADC model therefore extends distribution shaping to recurrent sequences.

For an order-r dictionary model, the target condition becomes

$$P_O\left(c_t \mid c_{t-r+1}, \ldots, c_{t-1}\right) = Q_r\left(c_t \mid c_{t-r+1}, \ldots, c_{t-1}\right), \tag{29}$$

where Q_r is a selected target transition model.

Thus, PADC may shape not only the frequency of individual code units but also the observable dynamics of the code dictionary.

5 Level III: Cryptographic Structure Decoupling

5.1 Observable and internal units

Conventional cryptographic representation often exposes units that correspond directly to the architecture of the transformation.

For a block cipher, the ciphertext is commonly represented as

$$Y = B_1 \| B_2 \| \cdots \| B_r, \tag{30}$$

where the block boundaries and block size are externally recoverable.

Similar structural units may occur in:

- stream-cipher synchronization intervals;
- hash and digest boundaries;
- message authentication codes;

- digital signatures;
- key containers;
- protocol records;
- serialized cryptographic objects.

The general structural assumption is

$$\text{observable unit} \cong \text{cryptographic unit}. \tag{31}$$

PADC removes this assumption.

5.2 Structural representation layer

Let Y be the output of a cryptographic transformation. A representation layer R_Φ maps it into the observable sequence O:

$$R_\Phi : Y \to O. \tag{32}$$

The mapping may divide internal units into fragments, reorder fragments, combine fragments from different units, insert auxiliary segments, and route fragments through different interpretation spaces.

The legitimate receiver applies

$$R_\Phi^{-1} : O \to Y. \tag{33}$$

The external segmentation of O therefore need not reveal the internal segmentation of Y:

$$\text{Seg}(O) \ncong \text{Struct}(Y). \tag{34}$$

5.3 LSEG as a structural carrier

Language Segment Encoding (LSEG) provides a natural carrier for such a representation layer because it separates visible stream segmentation from the interpretation of segment contents [4].

A basic LSEG segment has the form

$$\texttt{0x00} \| \texttt{LANG\_ID} \| \texttt{DATA}. \tag{35}$$

The visible segment boundary identifies a region of data, while `LANG_ID` selects an interpreter. The internal meaning and composition of `DATA` are determined by the selected interpretation rule rather than by the visible segment boundary alone.

In a PADC architecture, fragments belonging to one cryptographic unit may be distributed across multiple LSEG segments, while one visible sequence of segments may contain fragments from multiple cryptographic units.

Therefore,

$$\text{visible LSEG segment} \neq \text{cipher block}, \tag{36}$$

and, more generally,

$$\text{visible segmentation} \neq \text{cryptographic structure}. \tag{37}$$

Block ciphers provide the clearest example because their fixed blocks produce particularly visible analytical units. However, structural decoupling is not restricted to block ciphers.

6 Layered PADC Architecture

The complete transformation can be represented as

$$M \xrightarrow{,F_\Theta,} X \xrightarrow{,E_K,} Y \xrightarrow{,R_\Phi,} O. \tag{38}$$

The legitimate receiver performs

$$O \xrightarrow{,R_\Phi^{-1},} Y \xrightarrow{,E_K^{-1},} X \xrightarrow{,F_\Theta^{-1},} M. \tag{39}$$

The security architecture therefore contains at least three distinct inverse problems:

1. reconstruction of the internal cryptographic structure;
2. recovery or bypass of the cryptographic transformation;
3. reconstruction of the source alphabet-dictionary.

Successful solution of one inverse problem does not automatically solve the others.

For example, suppose an attacker recovers the key K or otherwise inverts the cryptographic layer:

$$E_K^{-1}(Y) = X. \tag{40}$$

The result X is not necessarily the source message M. It is a message expressed in an alternative alphabet-dictionary.

The attacker must still reconstruct

$$F_\Theta^{-1} : X \to M. \tag{41}$$

If the observable distribution of X was formed independently of the source language, conventional linguistic statistics do not directly provide that inverse mapping.

This yields the general principle

$$\text{compromise of cryptographic transformation} \not\Longrightarrow \text{recovery of source language}. \tag{42}$$

7 Illustrative Encoding Case

This section provides a minimal example of the complete PADC transformation

$$M \xrightarrow{,F_{\Theta},} X \xrightarrow{,E_{K},} Y \xrightarrow{,R_{\Phi},} O. \tag{43}$$

The purpose of the example is not to propose a production-ready cryptographic scheme. It demonstrates how symbol identity, dictionary distribution, and cryptographic structure may be modified at distinct architectural levels.

7.1 Source message and dictionary

Consider the source message

$$M = (\mathrm{A}, \mathrm{A}, \mathrm{A}, \mathrm{B}, \mathrm{B}, \mathrm{C}). \tag{44}$$

Its source dictionary is

$$\mathcal{D}_S = \mathrm{A}, \mathrm{B}, \mathrm{C}, \tag{45}$$

with empirical distribution

$$\mathbf{p} = \left(\frac{3}{6}, \frac{2}{6}, \frac{1}{6}\right). \tag{46}$$

The source units therefore have visibly different frequencies:

$$P_S(\mathrm{A}) > P_S(\mathrm{B}) > P_S(\mathrm{C}). \tag{47}$$

7.2 Alternative dictionary formation

Let the observable code dictionary be

$$\mathcal{D} * X = x * 1, x_2, x_3, x_4, x_5, x_6. \tag{48}$$

The source units are assigned multiple admissible representations:

$$\mathrm{A} \mapsto x_1, x_2, x_3, \ \mathrm{B} \qquad \mapsto x_4, x_5, \ \mathrm{C} \mapsto x_6. \tag{49}$$

The message may then be encoded as

$$X = (x_1, x_2, x_3, x_4, x_5, x_6). \tag{50}$$

Although the source message contains repeated units, the observable code sequence contains no re-

peated unit. Its empirical distribution is uniform:

$$P_X(x_j) = \frac{1}{6}, \qquad j = 1, \ldots, 6. \tag{51}$$

The source frequency profile has therefore not been preserved:

$$\mathbf{p} \neq \mathbf{q}, \tag{52}$$

where

$$\mathbf{q} = \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6} \right). \tag{53}$$

The legitimate decoder retains the inverse class relation

$$x_1, x_2, x_3 \mapsto \mathrm{A}, \; x_4, x_5 \qquad \mapsto \mathrm{B}, \; x_6 \mapsto \mathrm{C}. \tag{54}$$

The observer, however, sees six apparently independent code units.

7.3 Target-profile formation

Uniformity is only the simplest case. Suppose that the encoder must produce a selected target distribution

$$\mathbf{q}^* = (0.30, 0.25, 0.20, 0.15, 0.10). \tag{55}$$

Let

$$T_{ij} = P(c_j \mid w_i) \tag{56}$$

be the conditional encoding matrix between source dictionary units w_i and observable units c_j.

The matrix is selected subject to

$$\sum_j T_{ij} = 1 \tag{57}$$

and

$$\mathbf{p}T = \mathbf{q}^*. \tag{58}$$

The resulting observable distribution is therefore determined by the encoding architecture rather than by the source language:

$$P_O(c_j) = q_j^*. \tag{59}$$

The target profile may be uniform, artificial, or derived from a selected natural-language corpus. In the latter case, the observer may recover a stable and plausible statistical profile that nevertheless belongs to the code dictionary rather than to the source dictionary.

7.3.1 Numerical Example of Distribution Shaping

Consider a source dictionary with three units and empirical distribution

$$\mathbf{p} = \left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right). \tag{60}$$

Let the observable code dictionary contain five units:

$$\mathcal{D} * O = c * 1, c_2, c_3, c_4, c_5. \tag{61}$$

Define the conditional encoding matrix

$$T = \begin{pmatrix} \frac{3}{5} & \frac{2}{5} & 0 & 0 & 0 \\ 0 & 0 & \frac{3}{5} & \frac{2}{5} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{62}$$

Each row describes the distribution of observable representations available for one source unit. Thus,

$$w_1 \mapsto c_1, c_2,\ w_2 \ \mapsto c_3, c_4,\ w_3 \mapsto c_5.$$

The observable distribution is

$$\mathbf{q} = \mathbf{p}T. \tag{63}$$

Direct multiplication gives

$$\mathbf{q} = \mathbf{p}T = \left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right) \begin{pmatrix} \frac{3}{5} & \frac{2}{5} & 0 & 0 & 0 \\ 0 & 0 & \frac{3}{5} & \frac{2}{5} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

$$\mathbf{q} = \left(\frac{3}{10}, \frac{1}{5}, \frac{1}{5}, \frac{2}{15}, \frac{1}{6}\right).$$

Therefore,

$$\mathbf{q} = \left(0.30, 0.20, 0.20, 0.133\overline{3}, 0.166\overline{6}\right). \tag{64}$$

The observable profile differs from the source profile even though the mapping remains reversibly partitioned: each observable unit belongs to exactly one source equivalence class.

The example also shows that distribution shaping is not limited to uniformization. The encoder divides the probability mass of each source unit among several observable representations and thereby constructs

a new statistical profile.

For disjoint reversible representation classes, the total probability mass assigned to the class of source unit w_i must satisfy

$$\sum_{c_j \in C_i} q_j = p_i. \tag{65}$$

Consequently, not every arbitrary target vector can be produced by a single fixed-length, stateless, and disjoint mapping. More general target profiles may require one or more of the following mechanisms:

- variable-length code sequences;
- auxiliary or null dictionary units;
- context-dependent encoding;
- session state;
- overlapping observable representations resolved through additional decoder state;
- higher-order sequence shaping.

This distinction separates simple probability-mass redistribution from the more general problem of synthesizing an arbitrary observable dictionary process.

7.4 Cryptographic transformation

The alternative dictionary message X is then transformed by a conventional cryptographic operation:

$$Y = E_K(X). \tag{66}$$

For illustration, suppose that Y consists of three internal units:

$$Y = B_1 \| B_2 \| B_3. \tag{67}$$

A conventional representation would expose the sequence of these units directly. PADC introduces an additional representation layer.

7.5 Structural decoupling

Let each internal unit be divided into fragments:

$$B_1 = f_{11} \| f_{12},\ B_2 \qquad = f_{21} \| f_{22},\ B_3 = f_{31} \| f_{32}. \tag{68}$$

The representation layer may emit the fragments in a different observable organization:

$$O = (f_{21}, f_{11}, f_{31}, f_{22}, f_{12}, f_{32}). \tag{69}$$

An LSEG-based representation may further place these fragments in independently interpreted segments:

$$O = \texttt{0x00}\|\ell_4\|f_{21}\|\texttt{0x00}\|\ell_1\|f_{11} \quad \|\texttt{0x00}\|\ell_7\|f_{31}\|\texttt{0x00}\|\ell_2\|f_{22}\ \|\texttt{0x00}\|\ell_5\|f_{12}\|\texttt{0x00}\|\ell_3\|f_{32}. \tag{70}$$

The visible segmentation remains valid, but it does not directly expose the internal sequence

$$B_1\|B_2\|B_3. \tag{71}$$

The legitimate receiver applies the inverse representation rule

$$R_{\Phi}^{-1}(O) = B_1\|B_2\|B_3, \tag{72}$$

then the inverse cryptographic transformation

$$E_K^{-1}(Y) = X, \tag{73}$$

and finally the inverse dictionary mapping

$$F_{\Theta}^{-1}(X) = M. \tag{74}$$

The example shows that the three layers protect different relations:

$$\begin{gathered}\text{source identity} \not\leftrightarrow \text{observable identity},\\ \text{source frequency} \not\leftrightarrow \text{observable frequency},\\ \text{internal unit} \not\leftrightarrow \text{observable segment}.\end{gathered}$$

Recovery of one layer does not automatically invert the other two.

8 Weak Algorithms in a Layered Architecture

PADC does not alter the mathematical properties of an existing cipher or hash function. A short key remains short. A known collision remains a collision. A structurally weak transformation remains structurally weak.

However, the compromise of such a transformation need no longer be equivalent to compromise of the source message.

Consider

$$C = \mathrm{DES} *K\left(F * \Theta(M)\right). \tag{75}$$

Exhaustive recovery of K reveals

$$F_\Theta(M), \tag{76}$$

not necessarily M.

Similarly, consider

$$H = \text{MD5}\left(F_\Theta(w)\right). \tag{77}$$

Recovery of an encoded preimage reveals a code-dictionary unit rather than necessarily revealing the source unit w.

The claim is therefore not that PADC repairs the internal mathematics of DES or MD5. The claim is that algorithmic compromise ceases to be a sufficient condition for source-message recovery.

9 Observer, Measurement, and Structure

An observer necessarily receives physical or digital variations. Let

$$O = (o_1, o_2, \ldots, o_N) \tag{78}$$

be an observed sequence.

The representation used by the measuring apparatus is not automatically the representation used by the source.

For example,

$$s(t) \xrightarrow{\text{sampling}} (0, 1, 0, 1, \ldots) \tag{79}$$

does not demonstrate that the observed source possesses a binary alphabet.

The binary sequence belongs initially to the measurement architecture.

The interpretive chain is instead

$$\begin{aligned} \text{physical variation} &\to \text{measurement representation} \\ &\to \text{candidate units} \\ &\to \text{candidate dictionary} \\ &\to \text{candidate structure}. \end{aligned}$$

PADC acts on the conditions under which this chain stabilizes.

The objective is not necessarily to remove every regularity from the observable sequence. A target statistical profile may itself be stable. The objective is to prevent a stable inference from observable regularity to the source alphabet, source dictionary, and cryptographic architecture.

Thus,

$$\text{Regularity}(O) \not\Rightarrow \text{Structure}(M). \tag{80}$$

10 Security as Architectural Non-Equivalence

Traditional analysis often evaluates an encryption algorithm as the principal unit of security.

PADC instead evaluates the complete architecture:

$$\mathcal{S} = (F_\Theta, E_K, R_\Phi)\,. \tag{81}$$

Security depends on the relations among:

- the source language;
- the source dictionary;
- the alternative code dictionary;
- the cryptographic transformation;
- the structural representation;
- the observer's reconstruction model.

The central architectural property is non-equivalence:

$$O \ncong Y \ncong X \ncong M. \tag{82}$$

This does not mean that no mapping exists. The legitimate receiver must possess or reconstruct the required mappings.

It means that the mappings are not supplied merely by the observable organization of the transmitted stream.

The security objective can therefore be expressed as the preservation of decodability for the intended receiver together with the non-stabilization of a unique source reconstruction for an external observer.

11 Non-Given Representational Spaces and Cryptographic Reconstruction

11.1 Why PADC Is Not Merely an Encoding Layer

A natural objection to the proposed framework is that the complete PADC transformation

$$M \xrightarrow{,F_\Theta,} X \xrightarrow{,E_K,} Y \xrightarrow{,R_\Phi,} O \tag{83}$$

may be regarded as an ordinary composition of preprocessing, encryption, and postprocessing.

If all component transformations and representational spaces are known and stable, the objection is valid. One may define a composite transformation

$$\widetilde{E} * K = R \circ E * K \circ F \tag{84}$$

and write

$$O = \widetilde{E}_K(M). \tag{85}$$

Under these conditions, F and R are merely known encoding layers around a conventional cryptographic transformation. The message space, ciphertext space, elementary units, and object boundaries remain fixed and available to the adversary.

PADC addresses a different case.

11.2 The classical assumption of given spaces

A conventional cryptosystem is commonly defined over established message and ciphertext spaces:

$$E_K : \mathcal{M} \to \mathcal{C}. \tag{86}$$

The adversary may not know the message or the key, but the analytical domain is already specified.

In particular, the adversary is assumed to know or to be able to identify:

- what constitutes one ciphertext object;
- which elementary units belong to the ciphertext alphabet;
- where the boundaries of records, blocks, or messages occur;
- which observations are instances of the same type of unit;
- which mathematical operations are defined over those units.

The classical cryptanalytic problem can therefore be expressed as

$$\text{given } c \in \mathcal{C}, \quad \text{recover or distinguish information about } m \in \mathcal{M}. \tag{87}$$

The ciphertext object c is assumed to have already been formed as an element of a known analytical space.

11.3 Non-given representational spaces

In PADC, the external observer initially receives only an observable sequence

$$O = (o_1, o_2, \ldots, o_N). \tag{88}$$

The existence of this sequence does not by itself determine:

- the source alphabet;
- the alternative code alphabet;
- the source dictionary;
- the partition of observable units into source equivalence classes;
- the boundaries of internal cryptographic units;
- the ciphertext space over which the cryptographic transformation is defined.

The observer must first infer a mapping

$$O \longrightarrow \widehat{Y}, \tag{89}$$

where $\widehat{Y}$ is only a candidate reconstruction of the internal cryptographic representation.

The observable stream does not necessarily determine one unique ciphertext object:

$$O \not\Rightarrow Y. \tag{90}$$

More generally,

$$O \not\Rightarrow \mathcal{C}. \tag{91}$$

The external observer may therefore lack not only the key, but also the stable representational space in which the key-recovery problem would normally be formulated.

11.4 Formation of the analytical domain

The distinction between ordinary encoding and PADC lies in what the transformations F_Θ and R_Φ control.

In a conventional architecture, the transformations act on elements of already established spaces:

$$F : \mathcal{M} \to \mathcal{X}, \tag{92}$$

$$R : \mathcal{Y} \to \mathcal{O}, \tag{93}$$

where $\mathcal{M}$, $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{O}$ are known and stable.

In PADC, the parameters Θ and Φ participate in the formation of the observable analytical domain itself:

$$\mathcal{D} * \Theta, \qquad \mathcal{C} * \Theta, \Phi, \qquad \mathrm{Struct}_\Phi(O). \tag{94}$$

Here:

- $\mathcal{D} * \Theta$ is the alternative dictionary induced by the active dictionary-formation state;

- $\mathcal{C} * \Theta, \Phi$ is the candidate ciphertext space induced by dictionary formation and structural representation;
- $\text{Struct}_\Phi(O)$ is the internal organization reconstructed from the observable sequence.

The unknown state therefore does not merely select a function inside a fixed domain. It participates in determining which observable units, equivalence classes, and structural relations constitute the domain.

11.5 Multiple compatible cryptographic reconstructions

A single observable sequence may admit multiple candidate internal representations:

$$O = R_{\Phi_1}(Y_1) = R_{\Phi_2}(Y_2) = \cdots = R_{\Phi_r}(Y_r). \tag{95}$$

The candidate objects $Y_1, Y_2, \ldots, Y_r$ need not share the same segmentation or even belong to the same candidate ciphertext space:

$$Y_i \in \mathcal{C}_i. \tag{96}$$

Consequently,

$$\mathcal{C} * 1 \neq \mathcal{C} * 2 \neq \cdots \neq \mathcal{C}_r \tag{97}$$

may remain observationally compatible with the same sequence O.

In this case, the adversary cannot begin with the classical question

$$E_K^{-1}(c) = ? \tag{98}$$

because the object c has not yet been uniquely identified.

The preceding problem is instead

$$\text{Which candidate } Y_i \text{ constitutes the cryptographic object?} \tag{99}$$

Only after a representational hypothesis has stabilized does conventional cryptanalysis become applicable.

11.6 Constructive Non-Uniqueness of Cryptographic Reconstruction

The non-uniqueness of reconstruction must be distinguished from a merely alternative description of the same object.

To demonstrate genuine structural ambiguity, it is necessary to construct an observable sequence that is simultaneously valid under two mutually incompatible cryptographic object models.

Let

$$O = 000000110011 \tag{100}$$

be a twelve-bit observable sequence.

Consider first the candidate cryptographic space $\mathcal{C}_1$, whose elements are sequences of three four-bit blocks:

$$Y_1 = B_1 \| B_2 \| B_3, \tag{101}$$

where

$$|B_i| = 4. \tag{102}$$

A block

$$B_i = (b_{i1}, b_{i2}, b_{i3}, b_{i4}) \tag{103}$$

is valid if its fourth bit satisfies the parity condition

$$b_{i4} = b_{i1} \oplus b_{i2} \oplus b_{i3}. \tag{104}$$

Under this reconstruction, the observable sequence becomes

$$Y_1 = 0000 \| 0011 \| 0011. \tag{105}$$

The first block is valid because

$$0 = 0 \oplus 0 \oplus 0. \tag{106}$$

The second and third blocks are valid because

$$1 = 0 \oplus 0 \oplus 1. \tag{107}$$

Thus,

$$Y_1 \in \mathcal{C}_1. \tag{108}$$

Now consider a different candidate cryptographic space $\mathcal{C}_2$, whose elements are sequences of four three-bit units:

$$Y_2 = S_1 \| S_2 \| S_3 \| S_4, \tag{109}$$

where

$$|S_i| = 3. \tag{110}$$

A unit

$$S_i = (s_{i1}, s_{i2}, s_{i3}) \tag{111}$$

is valid if its third bit satisfies

$$s_{i3} = s_{i1} \oplus s_{i2}. \tag{112}$$

Under the second reconstruction, the same observable sequence becomes

$$Y_2 = 000\|000\|110\|011. \tag{113}$$

The first and second units are valid because

$$0 = 0 \oplus 0. \tag{114}$$

The third unit is valid because

$$0 = 1 \oplus 1, \tag{115}$$

and the fourth unit is valid because

$$1 = 0 \oplus 1. \tag{116}$$

Therefore,

$$Y_2 \in \mathcal{C}_2. \tag{117}$$

Both reconstructions generate the same observable sequence:

$$R_{\Phi_1}(Y_1) = R_{\Phi_2}(Y_2) = O. \tag{118}$$

However,

$$Y_1 \neq Y_2. \tag{119}$$

The difference is not merely notational. The two reconstructions have different:

- numbers of internal units;
- unit lengths;
- unit boundaries;

- validation equations;
- candidate cryptographic spaces.

In particular,

$$\mathcal{C} * 1 \neq \mathcal{C} * 2. \tag{120}$$

The observable sequence therefore belongs to the intersection

$$O \in R_{\Phi_1}(\mathcal{C} * 1) \cap R * \Phi_2(\mathcal{C}_2). \tag{121}$$

The ambiguity is not restricted to one specially selected sequence.

For twelve-bit sequences, define

$$\mathcal{C}_1^{(12)} = \left\{ x \in \{0,1\}^{12} \;\middle|\; x \text{ satisfies the three four-bit parity constraints} \right\}, \tag{122}$$

and

$$\mathcal{C}_2^{(12)} = \left\{ x \in \{0,1\}^{12} \;\middle|\; x \text{ satisfies the four three-bit parity constraints} \right\}. \tag{123}$$

Their intersection contains

$$\left| \mathcal{C} * 1^{(12)} \cap \mathcal{C} * 2^{(12)} \right| = 64 \tag{124}$$

distinct observable sequences.

Thus, structural ambiguity is not an accidental property of one stream. It is a stable property of two overlapping spaces of valid internal objects.

More generally, let

$$\mathfrak{R}(O) = \{ (\mathcal{C}, \Phi, Y) \mid Y \in \mathcal{C},\ R_\Phi(Y) = O \}. \tag{125}$$

be the set of cryptographic reconstructions compatible with observation O.

Reconstruction is unique only when

$$|\mathfrak{R}(O)| = 1. \tag{126}$$

Cryptographic-object ambiguity exists when

$$|\mathfrak{R}(O)| > 1. \tag{127}$$

A stronger condition holds when there exist two compatible reconstructions

$$(\mathcal{C} * i, \Phi * i, Y_i), \qquad (\mathcal{C} * j, \Phi * j, Y_j) \tag{128}$$

such that

$$\mathcal{C} * i \neq \mathcal{C} * j \tag{129}$$

and

$$R_{\Phi_i}(Y_i) = R_{\Phi_j}(Y_j) = O. \tag{130}$$

The observer then faces not merely several descriptions of one known ciphertext object, but several mutually incompatible candidates for what the cryptographic object is.

The initial cryptanalytic problem is therefore

$$\text{given } O, \quad \text{recover } (\mathcal{C}, \Phi, Y), \tag{131}$$

rather than immediately

$$\text{given } Y, \quad \text{recover } K \text{ or } M. \tag{132}$$

Only after the cryptographic object space, reconstruction state, and internal object have been selected does conventional key or message recovery become well-defined.

This construction establishes formal non-uniqueness of cryptographic reconstruction: the same observable sequence is a valid output of two structurally incompatible internal object models.

11.7 Protocol-Level Construction with Real Cryptographic Object Spaces

The preceding construction uses deliberately simplified parity-constrained spaces in order to establish the formal possibility of non-unique cryptographic reconstruction.

A stronger question is whether the same phenomenon can arise between object spaces associated with actual cryptographic algorithms.

AES, RSA, and LSEG do not automatically produce such ambiguity. The ambiguity must be created at the representation level by withholding or transforming the metadata that would ordinarily stabilize one unique cryptographic interpretation.

Consider an observable byte sequence

$$O = Q_1 \| Q_2 \| \cdots \| Q_m, \tag{133}$$

where every chunk Q_j contains 256 bytes:

$$|Q_j| = 256 \text{ bytes}. \tag{134}$$

Assume that the first byte of every Q_j is zero:

$$Q_j[1] = \texttt{0x00}. \tag{135}$$

This restriction ensures that the integer represented by Q_j is less than 2^{2040}.

Let n be a 2048-bit RSA modulus. Then

$$n \geq 2^{2047}, \tag{136}$$

and therefore

$$\text{int}(Q_j) < 2^{2040} < n. \tag{137}$$

Each chunk Q_j is consequently a valid element of the RSA ciphertext domain

$$\mathcal{C}_{\text{RSA}} = \{c \in \mathbb{Z} \mid 0 \leq c < n\}. \tag{138}$$

At the same time, every 256-byte chunk can be divided into sixteen 16-byte AES blocks:

$$Q_j = B_{j,1} \| B_{j,2} \| \cdots \| B_{j,16}, \tag{139}$$

where

$$|B_{j,k}| = 16 \text{ bytes}. \tag{140}$$

Each block belongs to the AES block space

$$\mathcal{C}_{\text{AES}} = 0, 1^{128}. \tag{141}$$

The same observable sequence therefore admits two stable reconstructions.

Under the first structural state Φ_{AES}, the internal cryptographic object sequence is

$$Y_{\text{AES}} = (B_{1,1}, \ldots, B_{1,16}, B_{2,1}, \ldots, B_{m,16}). \tag{142}$$

It contains

$$16m \tag{143}$$

candidate AES blocks.

Under the second structural state Φ_{RSA}, the internal object sequence is

$$Y_{\text{RSA}} = (Q_1, Q_2, \ldots, Q_m). \tag{144}$$

It contains

$$m \tag{145}$$

candidate RSA ciphertext objects.

Both reconstructions produce the same observable byte sequence:

$$R_{\Phi_{\text{AES}}}\left(Y_{\text{AES}}\right) = R_{\Phi_{\text{RSA}}}\left(Y_{\text{RSA}}\right) = O. \tag{146}$$

However,

$$Y_{\text{AES}} \neq Y_{\text{RSA}}. \tag{147}$$

The two reconstructions differ in:

- the number of cryptographic objects;
- the boundaries of those objects;
- the corresponding key spaces;
- the applicable decryption operations;
- the resulting plaintext spaces;
- the interpretation of the observable byte sequence.

The ambiguity is stable for arbitrary positive m. Increasing the stream length does not eliminate either reconstruction:

$$O \in R_{\Phi_{\text{AES}}}\left(\mathcal{C}_{\text{AES}}^{16m}\right) \tag{148}$$

and simultaneously

$$O \in R_{\Phi_{\text{RSA}}}\left(\mathcal{C}_{\text{RSA}}^{m}\right). \tag{149}$$

Hence,

$$O \in R_{\Phi_{\text{AES}}}\left(\mathcal{C} * \text{AES}^{16m}\right) \cap R * \Phi_{\text{RSA}}\left(\mathcal{C}_{\text{RSA}}^{m}\right). \tag{150}$$

This ambiguity does not result from AES or RSA alone.

In an ordinary protocol, external metadata normally specifies:

- the cryptographic algorithm;
- the cipher suite;
- the object type;
- the record boundaries;
- the block or modulus size;
- the interpretation of each transmitted field.

Such metadata stabilizes one reconstruction:

$$O + \text{protocol metadata} \longrightarrow Y. \tag{151}$$

A PADC representation instead seeks to preserve several candidate reconstructions:

$$O \longrightarrow \{Y_1, Y_2, \ldots, Y_N\}. \tag{152}$$

LSEG may serve as the carrier representation for this construction.

Suppose that the observable stream consists of a sequence of LSEG DATA fragments. The physical segment boundaries remain visible, but the rule that combines DATA fragments into cryptographic objects depends on the structural state Φ.

Under Φ_{AES}, the DATA octets are assembled into 16-byte units:

$$\text{Assemble} *\Phi * \text{AES}(O) = Y_{\mathrm{AES}}. \tag{153}$$

Under Φ_{RSA}, the same DATA octets are assembled into 256-byte units:

$$\text{Assemble} *\Phi * \text{RSA}(O) = Y_{\mathrm{RSA}}. \tag{154}$$

The observable LSEG sequence is unchanged:

$$O_{\mathrm{LSEG}} = R_{\Phi_{\mathrm{AES}}}(Y_{\mathrm{AES}}) = R_{\Phi_{\mathrm{RSA}}}(Y_{\mathrm{RSA}}). \tag{155}$$

The representation layer therefore does not conceal the existence of bytes or segments. It conceals which higher-level grouping of those segments constitutes the cryptographic object.

The result establishes a protocol-level form of reconstruction non-uniqueness involving real cryptographic object spaces:

$$\mathcal{C} * \text{AES}^{16m} \neq \mathcal{C} * \text{RSA}^{m}, \tag{156}$$

while both remain compatible with the same observation O.

This construction must not be interpreted as a claim that every AES, RSA, LSEG, or PADC implementation automatically possesses this property.

The claim is conditional:

> A PADC system creates cryptographic-object ambiguity when its representation function deliberately preserves compatibility between two or more stable and mutually incompatible cryptographic reconstructions.

The technical design problem is therefore to construct a family of representation functions satisfying

$$R_{\Phi_i}(\mathcal{C} * i) \cap R * \Phi_j(\mathcal{C}_j) \neq \varnothing \tag{157}$$

for real cryptographic object spaces $\mathcal{C}_i$ and $\mathcal{C}_j$, while preserving deterministic reconstruction for the authorized receiver.

11.8 LSEG as a Carrier of Structural Ambiguity

Within the PADC architecture, LSEG should not be understood primarily as an alternative network protocol.

Its more fundamental role is to provide a candidate implementation of the structural representation function

$$R_\Phi : Y \longrightarrow O. \tag{158}$$

The purpose of this function is not merely to serialize an already identified cryptographic object. Its purpose is to produce an observable stream whose physical segmentation does not uniquely determine the higher-level cryptographic structure from which it was formed.

An LSEG segment has the general form

$$\texttt{0x00} \| \texttt{LANG\_ID} \| \texttt{DATA}. \tag{159}$$

The marker and language identifier make the observable stream self-synchronizing at the segment level. However, segment-level synchronization does not imply unique reconstruction of the higher-level object architecture.

The observer may identify a sequence of physical segments

$$O = (s_1, s_2, \ldots, s_n), \tag{160}$$

while remaining unable to determine which subsets of those segments form the internal cryptographic units.

The distinction is therefore between

$$\mathrm{Seg}(O), \tag{161}$$

the observable segmentation of the carrier stream, and

$$\mathrm{Obj}_\Phi(O), \tag{162}$$

the state-dependent reconstruction of cryptographic objects.

The first may be visible, while the second remains dependent on the structural state Φ.

For example, under one admissible structural state,

$$\Phi_1, \tag{163}$$

the DATA fields of successive LSEG segments may be assembled into 16-byte units:

$$\text{Obj} *\Phi * 1(O) = (B_1, B_2, \ldots, B_r)\,, \tag{164}$$

where

$$|B_i| = 16 \text{ bytes}. \tag{165}$$

These units may be interpreted as candidate AES blocks.

Under another admissible structural state,

$$\Phi_2, \tag{166}$$

the same DATA fields may be assembled into 256-byte units:

$$\text{Obj} *\Phi * 2(O) = (C_1, C_2, \ldots, C_q)\,, \tag{167}$$

where

$$|C_j| = 256 \text{ bytes}. \tag{168}$$

These units may be interpreted as candidate RSA ciphertext objects.

The observable LSEG stream remains identical:

$$O = R_{\Phi_1}(Y_1) = R_{\Phi_2}(Y_2)\,, \tag{169}$$

while

$$Y_1 \neq Y_2. \tag{170}$$

Thus, the visible LSEG segments are not themselves the hidden cryptographic objects.

They are carrier units from which different higher-level object spaces may be reconstructed.

This distinction is essential. The visibility of bytes, segment markers, or DATA-field boundaries does not reveal:

- which segments belong to the same cryptographic object;
- whether one object spans several segments;
- whether one segment contains fragments of several objects;
- which interpreter applies to a reconstructed object;
- which internal object boundaries are cryptographically significant;
- which candidate cryptographic space is active.

Accordingly,

$$\mathrm{Seg}(O) \not\Rightarrow \mathrm{Obj}_{\Phi}(O). \tag{171}$$

LSEG therefore separates two forms of structure:

1. observable carrier structure;
2. state-dependent cryptographic object structure.

The first provides synchronization and transport recovery. The second determines the analytical units relevant to cryptanalysis.

This makes LSEG particularly suitable for PADC. A fully unstructured byte stream may conceal boundaries, but it also creates synchronization and recovery problems for the authorized receiver. A rigid conventional protocol solves synchronization by exposing the object architecture.

LSEG permits an intermediate construction:

$$\text{recoverable carrier segmentation} \quad + \quad \text{non-unique object reconstruction.} \tag{172}$$

The authorized receiver, possessing Φ, can reconstruct the intended cryptographic object sequence:

$$\mathrm{Obj}_{\Phi}(O) = Y. \tag{173}$$

An external observer without Φ instead obtains a set of compatible reconstructions:

$$\mathcal{Y}_{\mathrm{LSEG}}(O) = \{Y \mid \exists\, \Phi : R_{\Phi}(Y) = O\}\,. \tag{174}$$

The PADC objective is to maintain

$$|\mathcal{Y}_{\mathrm{LSEG}}(O)| > 1 \tag{175}$$

while preserving deterministic reconstruction for the authorized receiver.

LSEG is therefore not introduced here as a claim that a segment format alone provides cryptographic security.

The claim is narrower and more precise:

> LSEG is a candidate carrier architecture for constructing observable streams in which transport-level segmentation remains recoverable while cryptographic-object reconstruction remains dependent on a hidden structural state.

Under this interpretation, LSEG is not an external addition to PADC. It is the first concrete candidate for the representation layer R_{Φ}.

Its relevant security contribution is not encryption of DATA fields, but the controlled separation between the structure that must remain visible for reliable transmission and the structure that must remain ambiguous for cryptanalytic reconstruction.

11.9 Cross-Session Alphabet Rotation and the Failure of Statistical Accumulation

A central objection to representational ambiguity is that an adversary may accumulate observations across multiple sessions.

In a conventional cryptographic setting, this objection is well founded because the observable alphabet normally remains stable.

Let

$$O^{(1)}, O^{(2)}, \ldots, O^{(t)} \tag{176}$$

be observable streams collected from t sessions.

Conventional accumulation assumes that all streams belong to the same observable alphabet:

$$O^{(s)} \in (\mathcal{A}_O)^*, \qquad s = 1, \ldots, t. \tag{177}$$

The observer may therefore concatenate the sessions into a common corpus:

$$O^{(1)} \| O^{(2)} \| \cdots \| O^{(t)}, \tag{178}$$

and estimate increasingly accurate frequencies, transition probabilities, recurrent sequences, and structural relations.

This procedure contains an assumption that is normally left implicit:

> An observable unit identified in one session remains the same analytical unit in every other session.

PADC does not preserve this assumption.

For session s, let the active alphabet-dictionary state be

$$\Theta_s, \tag{179}$$

and let the corresponding observable alphabet be

$$\mathcal{A}_O^{(s)}. \tag{180}$$

The observable streams then satisfy

$$O^{(s)} \in \left(\mathcal{A}_O^{(s)}\right)^*, \tag{181}$$

rather than belonging to one fixed alphabet shared by all sessions.

Thus,

$$\mathcal{A}O^{(1)}, \mathcal{A}O^{(2)}, \ldots, \mathcal{A}_O^{(t)} \tag{182}$$

need not possess a known common identity relation.

An observable unit

$$o_i^{(r)} \in \mathcal{A}_O^{(r)} \tag{183}$$

cannot be combined statistically with

$$o_j^{(s)} \in \mathcal{A}_O^{(s)} \tag{184}$$

unless the observer first establishes that both units represent the same cross-session analytical class:

$$o_i^{(r)} \sim o_j^{(s)}. \tag{185}$$

Without this relation, the operation

$$\text{Count}\left(o_i^{(r)}, o_j^{(s)}\right) \tag{186}$$

does not estimate the frequency of one stable unit. It combines occurrences whose common identity has not been established.

Accordingly, the obstacle to cross-session accumulation is not merely noise in the estimated distribution.

The obstacle is the absence of a common statistical domain.

11.9.1 Session-specific formation

Let the session-specific PADC transformation be

$$O^{(s)} = R_{\Phi_s}\left(E_{K_s}\left(F_{\Theta_s}\left(M^{(s)}\right)\right)\right). \tag{187}$$

The states

$$\Theta_s, \qquad \Phi_s, \qquad K_s \tag{188}$$

may evolve independently or be derived from a synchronized session state.

The relevant requirement is not simply that different visible labels are used in different sessions.

A substitution of identifiers alone would leave the underlying statistical structure recoverable through frequency alignment.

Instead, the session state may modify:

- the observable alphabet;
- the partition of observable units into decoding classes;
- the number of representations assigned to each source unit;

- the conditional encoding matrix T_s;
- the target distribution $\mathbf{q}s^*$;
- higher-order transition rules;
- auxiliary and null units;
- the segmentation and assembly state Φs;
- the candidate cryptographic object spaces compatible with the observation.

Thus, the relation between sessions is not necessarily a simple permutation

$$\pi_s : \mathcal{A}O^{(1)} \to \mathcal{A}O^{(s)}. \tag{189}$$

The alphabets may differ in cardinality, equivalence classes, distributions, transition models, and structural use.

11.9.2 The cross-session identification problem

For two sessions r and s, define a candidate cross-session identification map as

$$J_{r,s} : \mathcal{A}O^{(r)} \to \mathcal{A}O^{(s)}. \tag{190}$$

Conventional statistical accumulation presupposes that $J_{r,s}$ is known, trivial, or efficiently recoverable.

PADC seeks to prevent this stabilization.

Let

$$\mathcal{J}_{r,s} = \{J \mid J \text{ is compatible with the observations from sessions } r \text{ and } s\}. \tag{191}$$

be the set of cross-session identification maps compatible with the available data.

Cross-session alignment is unique only if

$$|\mathcal{J}_{r,s}| = 1. \tag{192}$$

Cross-session alphabet ambiguity exists when

$$|\mathcal{J}_{r,s}| > 1. \tag{193}$$

In that case, the observer cannot determine which units from different sessions should be counted as occurrences of the same analytical object.

The conventional accumulated corpus

$$O^{(1)} \| O^{(2)} \| \cdots \| O^{(t)} \tag{194}$$

therefore does not automatically define one enlarged statistical sample.

It defines a collection of samples over potentially different and non-aligned alphabets.

11.9.3 Failure of direct distribution pooling

Let

$$\mathbf{q}^{(s)} \tag{195}$$

be the observable distribution in session s.

In a fixed-alphabet system, an accumulated estimator may be formed as

$$\widehat{\mathbf{q}}t = \frac{\sum s = 1^{t} n_s \mathbf{q}^{(s)}}{\sum_{s=1}^{t} n_s}, \tag{196}$$

where n_s is the number of observations in session s.

This expression is meaningful only if all vectors $\mathbf{q}^{(s)}$ refer to the same ordered system of observable units.

Under session-specific alphabet formation, one instead has

$$\mathbf{q}^{(s)} \in \Delta\left(\mathcal{A}_O^{(s)}\right), \tag{197}$$

where

$$\Delta\left(\mathcal{A}_O^{(s)}\right) \tag{198}$$

is the probability simplex over the alphabet of session s.

Unless a valid identification map $J_{r,s}$ has been recovered, the vectors belong to different coordinate systems.

Consequently,

$$\mathbf{q}^{(r)} + \mathbf{q}^{(s)} \tag{199}$$

is not an invariantly defined statistical operation.

The attacker must solve the alphabet-alignment problem before statistical accumulation becomes possible.

11.9.4 Why symbol relabeling alone is insufficient

The use of different visible symbols in each session does not by itself guarantee protection.

Suppose that every session preserves the same source-to-observable probability structure up to a permutation:

$$\mathbf{q}^{(s)} = \mathbf{q}^{(1)} P_s, \tag{200}$$

where P_s is a permutation matrix.

If the entries of $\mathbf{q}^{(1)}$ are sufficiently distinct, the observer may recover the permutation by ranking frequencies.

Likewise, stable higher-order transitions, fixed unit lengths, recurring positions, persistent LSEG identifiers, constant fragment sizes, or unchanged cryptographic boundaries may supply cross-session alignment invariants.

Therefore,

$$\mathcal{A}O^{(r)} \neq \mathcal{A}O^{(s)} \tag{201}$$

is necessary but not sufficient.

The stronger requirement is

$$\text{no efficiently recoverable } J_{r,s} \tag{202}$$

from the observable data available to the adversary.

Session rotation must therefore alter not only unit labels but also the observable relations by which those labels could be aligned.

11.9.5 Session independence of representational state

A strong session-rotation condition may be expressed as

$$\Theta_{s+1} = G_{\Theta}\left(Z_{s+1}\right), \tag{203}$$

and

$$\Phi_{s+1} = G_{\Phi}\left(Z_{s+1}\right), \tag{204}$$

where Z_{s+1} is a fresh synchronized session state.

The observable architecture of session $s+1$ should not be derivable from that of session s without knowledge of Z_{s+1}:

$$\left(\mathcal{A}O^{(s)}, T_s, \Phi_s\right) \nRightarrow \left(\mathcal{A}O^{(s+1)}, T_{s+1}, \Phi_{s+1}\right). \tag{205}$$

The authorized receiver reconstructs the active session state through the shared derivation mechanism.

The external observer receives only the resulting sequence $O^{(s)}$.

11.9.6 Cross-session accumulation as a prior reconstruction problem

Under PADC, the attacker cannot begin by accumulating symbol statistics across sessions.

The analytical sequence is instead

$$\begin{aligned}\text{multiple observed sessions} &\to \text{candidate session alphabets}\\ &\to \text{candidate cross-session identification maps}\\ &\to \text{candidate common dictionary}\\ &\to \text{accumulated statistical model}\\ &\to \text{source reconstruction.}\end{aligned}$$

Thus, cross-session accumulation is not an operation available before alphabet reconstruction.

It is a result that becomes available only after successful cross-session alphabet alignment.

This reverses the conventional assumption.

The conventional model states:

$$\text{more sessions} \longrightarrow \text{better statistics.} \tag{206}$$

The PADC model states:

$$\text{more non-aligned sessions} \not\Longrightarrow \text{one larger statistical sample.} \tag{207}$$

The observations become cumulatively useful only if the adversary can recover the hidden relations among their session-specific alphabets, dictionaries, and structural states.

11.9.7 Cross-session PADC condition

Let

$$\mathfrak{G}_t \tag{208}$$

denote the set of global reconstructions compatible with the first t sessions:

$$\mathfrak{G}_t = \left\{ \mathcal{G} \;\middle|\; \mathcal{G} \text{ jointly explains } O^{(1)}, \ldots, O^{(t)} \right\}. \tag{209}$$

A global reconstruction includes:

- one candidate alphabet-dictionary state for each session;
- candidate structural states;
- candidate cross-session identification maps;
- a candidate common source interpretation.

Long-term ambiguity is not measured by the number of independent per-session descriptions alone.

It requires multiple inequivalent global reconstructions:

$$|\mathfrak{G}_t/! \sim| > 1, \tag{210}$$

where $\sim$ removes reconstructions that differ only by irrelevant renaming or notation.

The PADC objective is therefore not merely to preserve ambiguity inside one session.

It is to prevent the stabilization of one unique family of cross-session identification maps.

This condition can be stated as follows:

> A session-rotating PADC system resists statistical accumulation when observations from different sessions cannot be efficiently embedded into one common and correctly aligned alphabet-dictionary space without the authorized session state.

Under this condition, the conventional growth of statistical confidence with corpus size does not directly apply, because the corpus itself has not yet been constructed as one coherent analytical object.

11.10 A Distinct Class of Cryptanalytic Problems

PADC therefore introduces a class of problems that can be described as

> cryptanalysis under non-given representational spaces.

The complete analytical sequence becomes

$$\begin{aligned} \text{observable sequence} &\to \text{candidate units} \\ &\to \text{candidate alphabet} \\ &\to \text{candidate dictionary} \\ &\to \text{candidate cryptographic structure} \\ &\to \text{cipher or key analysis} \\ &\to \text{source-message reconstruction.} \end{aligned}$$

Conventional cryptography typically begins after the first four stages have already been resolved or fixed by the protocol specification.

PADC moves the security boundary to those earlier stages.

The adversary must solve not only

$$\text{key recovery}$$

or

$$\text{message recovery},$$

but also

representation recovery,

dictionary recovery,

and

cryptographic-object recovery.

11.11 Criterion of reduction to a classical cryptosystem

A PADC construction is reducible to an ordinary classical cryptosystem when the following conditions hold for the adversary:

1. the transformations F_Θ and R_Φ are known or effectively fixed;
2. their active states Θ and Φ are known or directly recoverable;
3. the induced message and ciphertext spaces are stable;
4. observable segmentation uniquely determines internal cryptographic units;
5. the alternative dictionary is known or uniquely recoverable.

Under these conditions, one may define

$$\widetilde{E} = R_\Phi \circ E_K \circ F_\Theta \tag{211}$$

and treat the complete architecture as one conventional keyed transformation.

The reduction is not justified when the observable sequence remains compatible with multiple alphabets, dictionaries, segmentations, or ciphertext spaces.

This criterion may be summarized as follows:

> PADC is reducible to a classical cryptosystem only when its alphabet-dictionary formation and structural representation induce a known and stable cryptographic object space for the adversary.

11.12 Encoding secrecy and architectural secrecy

The distinction does not depend on keeping the general design secret.

The adversary may know that the system uses:

- symbol identity diversification;

- target-profile dictionary shaping;
- a conventional cipher;
- LSEG-based structural representation.

Knowledge of these principles does not necessarily reveal:

- the active equivalence classes C_i;
- the active matrix T;
- the session-specific dictionary state Θ;
- the structural reconstruction state Φ;
- the internal cryptographic units represented by the observable segments.

The relevant security property is therefore not obscurity of the general algorithm, but uncertainty of the active representational architecture.

The distinction is

$$\text{knowledge of construction principles} \not\Longrightarrow \text{recovery of active representational spaces.} \tag{212}$$

11.13 Consequences for the definition of cryptography

If cryptography is defined narrowly as the study of keyed transformations between fixed message and ciphertext spaces, PADC may appear to be an external encoding architecture.

If cryptography is defined as the construction of communication systems that prevent unauthorized source-message recovery, then the formation of the analytical spaces available to the adversary is itself cryptographically relevant.

PADC adopts the latter position.

Its object of protection is not only the value of the source message inside a known symbolic system. It also includes the correspondence between:

$$\text{physical observations} \leftrightarrow \text{symbolic units}, \ \text{symbolic units} \leftrightarrow \text{dictionary structures}, \ \text{observable segments} \leftrightarrow \text{cryptograph}$$

Accordingly, the defining PADC problem is not merely

$$M \to C, \tag{213}$$

but

$$O \longrightarrow \{\mathcal{A}, \mathcal{D}, \mathcal{C}, F_\Theta, E_K, R_\Phi, M\}.$$

Preventing the unique stabilization of this reconstruction is not an external convenience added to cryptography. It is the principal security objective of the PADC architecture.

12 Computational and Storage Complexity

Let the source message contain n dictionary units, and let the source dictionary contain d distinct units.

For each source unit w_i, let

$$C_i = c_{i1}, c_{i2}, \ldots, c_{ir_i} \tag{214}$$

be the set of admissible observable representations.

If the encoder selects one representation through a direct lookup table, the encoding cost per source unit is

$$O(1), \tag{215}$$

and the complete symbol- or word-level transformation requires

$$O(n) \tag{216}$$

operations.

The storage cost of the representation classes is

$$O\left(\sum_{i=1}^{d} r_i\right). \tag{217}$$

If every source unit has at most r admissible representations, this becomes

$$O(dr). \tag{218}$$

For a dense conditional encoding matrix

$$T \in \mathbb{R}^{d \times m}, \tag{219}$$

where m is the size of the observable dictionary, the storage cost is

$$O(dm). \tag{220}$$

In practical implementations, T may be sparse because each source unit normally maps only to a limited subset of observable units. If each row contains at most r non-zero values, the storage cost is reduced to

$$O(dr). \tag{221}$$

The computational cost of sampling from a row of T depends on the representation used. With cumulative probability tables, selection requires

$$O(\log r) \tag{222}$$

operations. With precomputed alias tables, the expected selection cost may be reduced to

$$O(1). \tag{223}$$

Higher-order shaping introduces additional state. If the observable dictionary contains m units and the encoder reproduces a context of order k, a dense model may contain up to

$$O(m^k) \tag{224}$$

contexts.

This exponential growth is the principal theoretical cost of direct higher-order shaping. In practice, sparse transition tables, bounded context models, finite-state encoders, and corpus-specific pruning can substantially reduce the effective state space.

Let

$$\rho_F = \frac{|X|}{|M|} \tag{225}$$

denote the expansion introduced by alphabet-dictionary formation, and let

$$\rho_R = \frac{|O|}{|Y|} \tag{226}$$

denote the expansion introduced by structural representation.

The total representation expansion, excluding the internal expansion of the cryptographic algorithm, is

$$\rho_{\text{PADC}} = \rho_F \rho_R. \tag{227}$$

Thus, the main practical trade-off is not necessarily asymptotic processing time, which may remain linear, but the balance among:

the number of admissible representations,
the order of the shaped dictionary process,
the amount of encoder and decoder state,
the expansion of the transmitted stream,
the remaining reconstructability for an observer.

For bounded representation classes and bounded-order models, the complete PADC transformation can remain linear in the length of the source message:

$$T_{\text{PADC}}(n) = O(n) + T_E(n), \tag{228}$$

where $T_E(n)$ is the complexity of the underlying cryptographic transformation.

13 Adversarial Knowledge Model

The strength of a PADC construction depends not only on the observable stream, but also on which architectural components are known to the adversary.

Let the complete transformation be

$$O = R_\Phi \left(E_K \left(F_\Theta(M) \right) \right). \tag{229}$$

The adversary may possess knowledge of some or all of the following components:

1. the target observable distribution $\mathbf{q}^*$;
2. the observable dictionary $\mathcal{D} * O$;
3. the general form of the dictionary transformation $F * \Theta$;
4. the conditional encoding matrix T;
5. the underlying cryptographic algorithm E;
6. the cryptographic key K;
7. the structural representation mechanism R_Φ;
8. the exact representation state Φ;
9. one or more source-message fragments;
10. multiple streams generated under the same dictionary or structural state.

These knowledge levels must be distinguished.

Knowledge of the target distribution

$$\mathbf{q}^* \tag{230}$$

does not by itself reveal the source distribution

$$\mathbf{p}, \tag{231}$$

because the relation

$$\mathbf{p}T = \mathbf{q}^* \tag{232}$$

may admit multiple combinations of $\mathbf{p}$ and T.

Likewise, knowledge of the observable dictionary

$$\mathcal{D}_O \tag{233}$$

does not necessarily reveal its partition into source equivalence classes:

$$\mathcal{D}_O = C_1 \cup C_2 \cup \cdots \cup C_d. \tag{234}$$

An adversary may know the general mechanism of F_Θ while lacking the actual session-specific mapping between observable units and source units.

Similarly, knowledge of the structural representation algorithm R_Φ does not necessarily imply knowledge of the state Φ used to assemble internal cryptographic units from observable segments.

The adversarial model can therefore be divided into several cases.

13.1 Observable-only adversary

The adversary receives only the observable stream O and may estimate:

- frequencies of observable units;
- higher-order transition statistics;
- physical segmentation;
- stream length and timing;
- cross-stream correlations.

The adversary does not know F_Θ, K, or Φ.

13.2 Target-profile-aware adversary

The adversary knows $\mathbf{q}^*$ and may know which language, corpus, or artificial process the observable distribution is intended to imitate.

This knowledge identifies the target profile, but not necessarily the source dictionary or the inverse mapping.

13.3 Encoding-aware adversary

The adversary knows the architecture of F_Θ and may know the family of admissible matrices T, but not the actual matrix or session-specific dictionary partition.

This corresponds to the usual cryptographic principle that the system architecture may be public while its state remains secret.

13.4 Cipher-compromise adversary

The adversary knows or recovers the key K and obtains

$$X = E_K^{-1}(Y). \tag{235}$$

The recovered object is the alternative dictionary representation X, not necessarily the source message M.

The remaining problem is

$$F_\Theta^{-1}(X) = M. \tag{236}$$

Thus,

$$\text{recovery of } K \not\Longrightarrow \text{recovery of } M. \tag{237}$$

13.5 Representation-aware adversary

The adversary knows the general form of R_Φ but not the active state Φ.

The visible segmentation of O may therefore remain insufficient to reconstruct the internal organization of Y.

13.6 Full-architecture adversary

The strongest adversary knows:

$$\mathbf{q}^*, \quad \mathcal{D} * O, \quad F * \Theta, \quad T, \quad E, \quad K, \quad R_\Phi, \quad \Phi. \tag{238}$$

Under this condition, the architecture is fully compromised and source recovery is expected.

PADC therefore does not claim security against disclosure of every mapping and state. Its claim is that compromise of one layer does not automatically compromise the remaining layers.

The relevant implication chain is

$$\text{knowledge of } \mathbf{q}^* \not\Longrightarrow \text{knowledge of } T, \ \text{knowledge of } T \not\Longrightarrow \text{knowledge of } \Phi, \ \text{knowledge of } \Phi \not\Longrightarrow \text{knowledge of } K,$$

Accordingly, PADC security must be evaluated relative to an explicitly declared adversarial knowledge set:

$$\mathcal{K} * \mathcal{A} \subseteq \mathbf{q}^*, \mathcal{D} * O, F_\Theta, T, E, K, R_\Phi, \Phi. \tag{239}$$

A security statement without such a declaration is incomplete because different knowledge sets correspond to different reconstruction problems.

14 Research Questions

The present paper establishes the architectural basis of PADC, including symbol-identity diversification, dictionary-distribution shaping, cryptographic-object ambiguity, LSEG-based structural representation, and session-specific alphabet rotation.

Several technical questions nevertheless remain open.

1. Under what necessary and sufficient conditions does a target observable distribution $\mathbf{q}^*$ admit a reversible encoding process with bounded expansion?
2. Which classes of target distributions can be realized by fixed-length disjoint mappings, and which require variable-length, stateful, overlapping, or auxiliary-unit constructions?
3. How can higher-order dictionary distributions be shaped without exponential growth of the encoder and decoder state?
4. What is the minimum secret session state required to reconstruct the intended alphabet, dictionary, and cryptographic-object structure?
5. Which session-state derivation and rotation mechanisms prevent efficient recovery of cross-session identification maps while preserving deterministic synchronization for the authorized receiver?
6. Which observable invariants remain usable for aligning independently formed session alphabets despite symbol, dictionary, and structural rotation?
7. Under what conditions can an LSEG-based representation preserve multiple stable cryptographic-object reconstructions without exposing the active assembly state Φ?
8. How should the ambiguity of competing cryptographic-object spaces be measured after quotienting out reconstructions that differ only by renaming, permutation, or other analytically irrelevant transformations?
9. How rapidly does the set of compatible global reconstructions shrink as the adversary accumulates sessions, timing information, known-message fragments, or active probes?
10. Which formal adversarial model best captures architectural reconstruction, including alphabet alignment, dictionary recovery, object-space identification, and conventional key recovery?
11. How should alphabet stabilization, dictionary stabilization, cross-session alignment, and cryptographic-object stabilization be measured experimentally?
12. What computational, storage, bandwidth, and synchronization costs arise in practical implementations of session-rotating PADC and LSEG-based structural representation?

15 Conclusion

Modern cryptography normally begins after the analytical object has already been formed.

The plaintext alphabet, the ciphertext alphabet, the boundaries of cryptographic objects, and the relation between observable units are usually treated as established before cryptanalysis begins. The attacker may not know the message or the key, but the spaces in which those unknowns are defined are assumed to be stable.

Post-Alphabet-Dictionary Cryptography moves the security boundary to an earlier stage.

It treats the formation of observable symbols, dictionaries, distributions, segment relations, and cryptographic-object boundaries as part of the cryptographic architecture itself.

The resulting transformation is

$$M \xrightarrow{F_\Theta} X \xrightarrow{E_K} Y \xrightarrow{R_\Phi} O. \tag{240}$$

Here, F_Θ does not merely replace one known symbol by another. It forms an alternative alphabet-dictionary representation whose observable identity and statistical organization need not preserve those of the source.

Likewise, R_Φ does not merely serialize an already established ciphertext object. It forms the observable carrier representation from which the internal cryptographic structure must be reconstructed.

Three distinct levels of controlled non-equivalence have been identified.

First, symbol identity diversification removes the requirement that repeated source units produce repeated observable units:

$$a_i = a_j \not\Longrightarrow \mathrm{Enc}(a_i) = \mathrm{Enc}(a_j). \tag{241}$$

Second, dictionary-distribution shaping separates the source distribution from the observable distribution:

$$\mathbf{p} \not\Longrightarrow \mathbf{q}. \tag{242}$$

The observable stream may exhibit a uniform, artificial, technical, or language-like statistical profile even when that profile is unrelated to the source dictionary.

Third, cryptographic-structure decoupling separates visible carrier segmentation from the internal organization of cryptographic objects:

$$\mathrm{Seg}(O) \not\Longrightarrow \mathrm{Obj}_\Phi(O). \tag{243}$$

This third result is essential.

The observable stream does not necessarily identify one unique ciphertext object or even one unique ciphertext space. A single observation may remain compatible with several mutually incompatible reconstructions:

$$R_{\Phi_1}(Y_1) = R_{\Phi_2}(Y_2) = O, \tag{244}$$

while

$$Y_1 \neq Y_2 \tag{245}$$

and

$$\mathcal{C} * 1 \neq \mathcal{C} * 2. \tag{246}$$

The attacker therefore faces a problem that precedes conventional key recovery:

$$\text{given } O, \quad \text{recover } (\mathcal{C}, \Phi, Y)\,. \tag{247}$$

Only after the cryptographic-object space, the structural state, and the internal object have been stabilized does the conventional problem

$$\text{given } Y, \quad \text{recover } K \text{ or } M \tag{248}$$

become well-defined.

This establishes the central distinction between PADC and an ordinary encoding layer.

If F_Θ and R_Φ act between known, stable, and externally recoverable spaces, then the complete architecture may be reduced to a single composite transformation:

$$\widetilde{E} * K = R * \Phi \circ E_K \circ F_\Theta. \tag{249}$$

Under those conditions, PADC reduces to conventional preprocessing, encryption, and postprocessing.

The reduction is not justified when the active alphabet, dictionary, cross-session identity relations, object boundaries, and ciphertext space remain dependent on hidden representational states.

PADC is therefore not defined merely by the presence of additional transformations. It is defined by the absence of a uniquely given analytical domain for the external observer.

LSEG acquires a precise role within this architecture.

It is not used merely as an alternative transport protocol. It provides the first concrete candidate for the representation function R_Φ.

Its visible segments support synchronization and recovery at the carrier level, while the assembly of those segments into cryptographic objects remains dependent on the hidden structural state:

$$\text{recoverable carrier segmentation} \quad + \quad \text{non-unique cryptographic-object reconstruction.}$$

Thus, the security contribution of LSEG is not the concealment of bytes or physical segments. It is the controlled separation between the structure that must remain visible for reliable transmission and the structure that must remain unavailable for unique cryptanalytic reconstruction.

The same reasoning applies across sessions.

Conventional statistical accumulation assumes that observations from different sessions belong to one

stable alphabet:

$$O^{(s)} \in \mathcal{A}_O^*. \tag{250}$$

Under session-specific PADC formation, one instead has

$$O^{(s)} \in \left(\mathcal{A}_O^{(s)}\right)^*. \tag{251}$$

Statistics from different sessions cannot be pooled as observations of the same units unless the adversary first reconstructs cross-session identification maps:

$$J_{r,s} : \mathcal{A} * O^{(r)} \to \mathcal{A} * O^{(s)}. \tag{252}$$

Consequently,

$$\text{more observed sessions} \not\Longrightarrow \text{one larger statistical sample.} \tag{253}$$

The attacker must first establish that units observed in different sessions belong to the same analytical classes.

If the alphabet, dictionary partition, target distribution, transition model, auxiliary units, and structural state change between sessions, then cross-session accumulation becomes a prior reconstruction problem rather than an immediately available analytical operation.

The compromise of the conventional cryptographic transformation therefore does not necessarily recover the source message.

If

$$E_K^{-1}(Y) = X, \tag{254}$$

the attacker has recovered an alternative dictionary stream, not necessarily the source message.

The remaining inverse problem is

$$F_\Theta^{-1}(X) = M. \tag{255}$$

Likewise, recovery of visible carrier segmentation does not necessarily recover Y, and recovery of one session alphabet does not necessarily identify the alphabets of later sessions.

The security architecture is therefore layered:

$$\begin{aligned} \text{carrier reconstruction} &\to \text{cryptographic-object reconstruction} \\ &\to \text{cryptographic transformation recovery} \\ &\to \text{alternative dictionary reconstruction} \\ &\to \text{source-message recovery}. \end{aligned}$$

Success at one stage does not automatically imply success at the next.

The central conclusion of this paper is therefore:

> Cryptographic security is not only the difficulty of inverting a transformation inside a predefined symbolic space. It is also the difficulty of stabilizing the alphabet, dictionary, cross-session identity relations, and cryptographic-object structure required to form that space as an object of analysis.

PADC does not replace conventional encryption.

It extends the domain of cryptography from keyed transformations within predefined message and ciphertext spaces to the controlled formation of the symbolic, dictionary, statistical, and structural spaces available to an external observer.

Its principal security objective is not the elimination of all observable regularity.

It is the prevention of a unique and unauthorized stabilization of the relations

$$\begin{aligned} \text{observation} &\leftrightarrow \text{symbol}, \\ \text{symbol} &\leftrightarrow \text{dictionary unit}, \\ \text{session unit} &\leftrightarrow \text{cross-session identity}, \\ \text{carrier segment} &\leftrightarrow \text{cryptographic object}, \\ \text{alternative dictionary} &\leftrightarrow \text{source message}. \end{aligned}$$

Under this formulation, the formation of the analytical object is not an external preliminary to cryptography.

It is itself a cryptographic problem.

A Architectural Security Properties of PADC

The preceding sections introduced Post-Alphabet-Dictionary Cryptography as an architecture in which the alphabet, dictionary, statistical domain, cryptographic-object structure, and cross-session identity relations available to an observer need not be given in advance.

This appendix summarizes the corresponding architectural security properties.

These properties do not replace conventional security properties of the internal cryptographic transformation E_K. A block cipher, stream cipher, hash function, signature scheme, or other cryptographic primitive must still be evaluated according to the security criteria appropriate to that primitive.

The properties introduced here concern a prior analytical level:

> Can the observer uniquely stabilize the alphabet, dictionary, message space, cryptographic-object space, and representational mappings required to formulate the conventional cryptanalytic problem?

Let

$$\Omega \tag{256}$$

denote a declared class of admissible PADC architectures.

Let

$$O \tag{257}$$

be an observable sequence.

A complete reconstruction is represented by

$$\rho = (\mathcal{A}, \mathcal{D}, \mathcal{M}, \mathcal{X}, \mathcal{C}, F, E, R, M, X, Y)\,, \tag{258}$$

subject to

$$M \in \mathcal{M}, \tag{259}$$

$$X \in \mathcal{X}, \tag{260}$$

$$Y \in \mathcal{C}, \tag{261}$$

and

$$X = F(M), \tag{262}$$

$$Y = E(X), \tag{263}$$

$$O = R(Y). \tag{264}$$

The reconstruction may additionally include hidden dictionary, distribution, segmentation, and session states such as

$$\Theta, \qquad \Phi, \qquad K. \tag{265}$$

Define the set of reconstructions compatible with observation O as

$$\mathfrak{R}_{\Omega}(O) = \{\rho \in \Omega; |; \rho \text{ generates or explains } O\}\,. \tag{266}$$

Not every difference between two reconstructions is analytically significant.

Two reconstructions may differ only by:

- renaming of symbols;
- permutation of identifiers;
- reordering of equivalent internal labels;
- another transformation that preserves all relevant analytical relations.

Such reconstructions are treated as equivalent under a relation

$$\sim\,. \tag{267}$$

The relevant object is therefore the quotient set

$$\mathfrak{R}_{\Omega}(O)/\!\sim. \tag{268}$$

A.1 Representational Space Non-Uniqueness

A PADC architecture possesses *Representational Space Non-Uniqueness* (RSNU) relative to Ω and observation O if

$$|\mathfrak{R}_{\Omega}(O)/\!\sim| \geq 2. \tag{269}$$

RSNU requires that at least two analytically inequivalent reconstructions remain compatible with the same observation.

The competing reconstructions may differ in one or more of the following:

- observable alphabet;

- source alphabet;
- observable dictionary;
- source dictionary;
- partition of observable units into equivalence classes;
- source-message space;
- intermediate representation space;
- cryptographic-object space;
- segmentation and assembly rules;
- relation between the observable stream and the source message.

RSNU therefore concerns a multiplicity of compatible analytical worlds, rather than several notational descriptions of one already established cryptographic object.

The property may be summarized as

$$O \not\Rightarrow \text{one unique representational space.} \tag{270}$$

A.2 Structural Ambiguity Preservation

Let

$$\text{Struct}_{\Omega}(O) \tag{271}$$

denote the set of cryptographic-object reconstructions compatible with O:

$$\text{Struct} * \Omega(O) = \{(\mathcal{C}, \Phi, Y)\,;|;Y \in \mathcal{C},;R * \Phi(Y) = O\}\,. \tag{272}$$

A PADC representation possesses *Structural Ambiguity Preservation* (SAP) if

$$|\text{Struct}_{\Omega}(O)/\!\sim| \geq 2. \tag{273}$$

SAP is a structural specialization of RSNU.

It requires that the observable sequence remain compatible with at least two inequivalent cryptographic-object reconstructions.

A stronger form holds when there exist

$$(\mathcal{C} * 1, \Phi * 1, Y_1) \tag{274}$$

and

$$(\mathcal{C} * 2, \Phi * 2, Y_2) \tag{275}$$

such that

$$\mathcal{C} * 1 \neq \mathcal{C} * 2, \tag{276}$$

$$Y_1 \neq Y_2, \tag{277}$$

and

$$R_{\Phi_1}(Y_1) = R_{\Phi_2}(Y_2) = O. \tag{278}$$

In this case, the ambiguity concerns not merely the boundaries of one known object type, but the identity of the cryptographic-object space itself.

Thus,

$$\mathrm{Seg}(O) \not\Rightarrow \mathrm{Obj}_{\Phi}(O). \tag{279}$$

The visible carrier structure may remain recoverable while the higher-level cryptographic-object structure remains non-unique.

A.3 Architectural Non-Stabilizability

Logical non-uniqueness alone is insufficient as a security criterion.

For a finite observation, several artificial reconstructions can often be constructed even for conventional systems. The stronger PADC property concerns whether one reconstruction can be selected as uniquely supported by the available observation.

A PADC architecture possesses *Architectural Non-Stabilizability* (ANS) relative to an adversary $\mathcal{A}$ and an admissible architecture class Ω if the observable data do not permit $\mathcal{A}$ to efficiently select one canonical equivalence class

$$[\rho] \in \mathfrak{R}_{\Omega}(O)/\sim \tag{280}$$

as the uniquely supported architecture.

ANS requires both:

1. representational non-uniqueness;
2. absence of an efficiently recoverable rule that selects the intended reconstruction without the hidden representational state.

Informally,

$$\text{ANS} = \text{RSNU} + \text{non-selectability of one intended reconstruction.} \tag{281}$$

The hidden object in ANS is not only a cryptographic key.
It may include:

- the active alphabet state;
- the active dictionary partition;
- the distribution-formation state;
- the conditional encoding matrix;
- the structural assembly state;
- the cryptographic-object space;
- the cross-session identification state;
- the mappings connecting observation to source interpretation.

Accordingly, the protected uncertainty is not merely

$$K \in \mathcal{K}, \tag{282}$$

inside a known cryptographic domain.
The protected uncertainty includes the domain itself:

$$(\mathcal{A}, \mathcal{D}, \mathcal{M}, \mathcal{C}, F, R)\,. \tag{283}$$

The PADC security problem is therefore not merely

$$\text{recover } K \text{ inside a known space,} \tag{284}$$

but

$$\text{stabilize the space in which } K \text{ and its associated operations are defined.} \tag{285}$$

ANS may be expressed informally as

$$O \not\Rightarrow \text{one uniquely stabilized architecture.} \tag{286}$$

A.4 Cross-Session Analytical Decorrelation

Let

$$O^{(1)}, O^{(2)}, \ldots, O^{(t)} \tag{287}$$

be observations collected from different sessions.

Let the corresponding observable alphabets and dictionaries be

$$\mathcal{A}^{(s)} * O, \qquad \mathcal{D}^{(s)} * O, \qquad s = 1, \ldots, t. \tag{288}$$

A cross-session alphabet alignment consists of mappings

$$J^A * s : \mathcal{A}^{(s)} * O \to \mathcal{A}^*, \tag{289}$$

where $\mathcal{A}^*$ is a candidate common alphabet.

A cross-session dictionary alignment consists of mappings

$$J^D * s : \mathcal{D}^{(s)} * O \to \mathcal{D}^*, \tag{290}$$

where $\mathcal{D}^*$ is a candidate common dictionary.

A session-rotating PADC architecture possesses *Cross-Session Analytical Decorrelation* (CSAD) if an observer without the authorized session state cannot efficiently recover a unique family of alignment maps

$$\left\{ J_s^A, J_s^D \right\}_{s=1}^{t}. \tag{291}$$

that embeds the observed sessions into one correctly aligned alphabet and dictionary space.

Under CSAD, the formal concatenation

$$O^{(1)} \| O^{(2)} \| \cdots \| O^{(t)} \tag{292}$$

does not automatically constitute one enlarged statistical sample.

The observations become jointly usable only after the cross-session identity relations have been reconstructed.

CSAD does not require the absence of statistics within individual sessions.

Each session may possess a stable and measurable local distribution

$$\mathbf{q}^{(s)} \in \Delta\left(\mathcal{A}_O^{(s)}\right). \tag{293}$$

However, two distributions

$$\mathbf{q}^{(r)} \tag{294}$$

and

$$\mathbf{q}^{(s)} \tag{295}$$

belong to different coordinate systems unless a valid cross-session alignment has been established.

Consequently,

$$\mathbf{q}^{(r)} + \mathbf{q}^{(s)} \tag{296}$$

is not an invariantly defined statistical operation before alignment.

The relevant condition is therefore not the elimination of local statistics, but the prevention of their canonical cross-session pooling.

This may be summarized as

$$\text{more non-aligned sessions} \not\Longrightarrow \text{one larger statistical sample.} \tag{297}$$

A.5 Why Symbol Relabeling Alone Does Not Establish CSAD

Different visible labels in different sessions are insufficient to establish CSAD.

Suppose that session alphabets differ only by a permutation

$$\pi_s : \mathcal{A}^{(1)} * O \to \mathcal{A}^{(s)} * O. \tag{298}$$

Suppose further that the observable distributions satisfy

$$\mathbf{q}^{(s)} = \mathbf{q}^{(1)} P_s, \tag{299}$$

where P_s is the permutation matrix induced by π_s.

If the entries of $\mathbf{q}^{(1)}$ are sufficiently distinct, the permutation may be reconstructed by frequency ranking.

Likewise, cross-session alignment may be supported by persistent invariants such as:

- stable higher-order transitions;
- fixed unit lengths;
- recurring positions;
- persistent LSEG identifiers;
- constant fragment sizes;
- stable timing relations;
- unchanged cryptographic-object boundaries;
- repeated auxiliary structures.

Therefore,

$$\mathcal{A}^{(r)} * O \neq \mathcal{A}^{(s)} * O \tag{300}$$

is necessary but not sufficient.

The stronger condition is

$$\text{no efficiently recoverable } J_{r,s} \tag{301}$$

from the observable data available to the adversary.

Session rotation must alter not only visible labels, but also the relations by which those labels could be aligned.

The session state may therefore modify:

- alphabet cardinality;
- symbol-equivalence classes;
- dictionary partition;
- conditional encoding matrix;
- target distribution;
- higher-order transition model;
- auxiliary and null units;
- segment assembly rules;
- cryptographic-object hypotheses compatible with observation.

A.6 Long-Term Architectural Non-Stabilizability

Let

$$\mathfrak{G}_t \tag{302}$$

denote the set of global reconstructions compatible with the first t sessions:

$$\mathfrak{G}_t = \left\{ \mathcal{G} \;\middle|\; \mathcal{G} \text{ jointly explains } O^{(1)}, \ldots, O^{(t)} \right\}. \tag{303}$$

A global reconstruction includes:

- one candidate alphabet state for each session;
- one candidate dictionary state for each session;

- candidate structural states;
- candidate cryptographic-object spaces;
- candidate cross-session identification maps;
- a candidate common source interpretation.

Long-term ANS requires

$$|\mathfrak{G}_t/\sim| \geq 2 \tag{304}$$

and the absence of an efficient canonical selector for the intended global reconstruction.

The condition is not satisfied merely because each session independently admits several descriptions.

It requires multiple inequivalent explanations of the entire sequence of sessions.

CSAD is therefore a cross-session mechanism supporting long-term ANS:

$$\text{ANS} * 1 + \text{CSAD} \longrightarrow \text{ANS} * t, \tag{305}$$

provided that no persistent observable invariant independently identifies the session-specific representational states.

A.7 Relation Among the Architectural Properties

The proposed properties form a hierarchy.

SAP concerns non-uniqueness at the level of cryptographic-object structure.

RSNU concerns non-uniqueness of the broader representational space.

ANS strengthens RSNU by requiring that the intended reconstruction cannot be efficiently and canonically selected from the compatible alternatives.

CSAD extends this condition across sessions by preventing the immediate formation of one shared alphabet-dictionary space.

Thus,

$$\text{SAP} \subseteq \text{RSNU} \subseteq \text{ANS}, \tag{306}$$

while

$$\text{CSAD} \tag{307}$$

is a temporal condition intended to preserve ANS under repeated observation.

The relations should not be interpreted as claims that every instance of RSNU automatically satisfies ANS, or that every session-rotating system automatically satisfies CSAD.

They identify progressively stronger architectural requirements.

A.8 Why Conventional Security Games Presuppose the Protected Object

Conventional cryptographic games are formulated over a common analytical domain.

For example, message-indistinguishability games assume that

$$m_0, m_1 \in \mathcal{M}, \tag{308}$$

where $\mathcal{M}$ is one established message space.

The competing messages are therefore already comparable as objects of the same type.

This assumption extends beyond the message set.

A conventional security game normally presupposes a shared:

- alphabet;
- dictionary;
- message space;
- ciphertext space;
- cryptographic-object type;
- interpretation of the challenge interface;
- relation of identity among the objects being compared.

The game varies a message, key, nonce, challenge bit, or oracle response inside this common domain.

It does not normally test whether the domain itself can be formed uniquely from observation.

PADC addresses the prior case in which competing reconstructions may belong to different dictionaries and different representational spaces:

$$M_0 \in \mathcal{M}_0, \tag{309}$$

$$M_1 \in \mathcal{M}_1, \tag{310}$$

with

$$\mathcal{D} * 0 \neq \mathcal{D} * 1, \tag{311}$$

and potentially

$$\mathcal{M} * 0 \neq \mathcal{M} * 1. \tag{312}$$

Without a shared dictionary, the objects are not yet established as two alternatives inside one analytical space.

To place such reconstructions into one conventional challenge game, one must first introduce a common metarepresentational space

$$\mathcal{U} \tag{313}$$

and embeddings

$$\iota_0 : \mathcal{M}_0 \to \mathcal{U}, \tag{314}$$

$$\iota_1 : \mathcal{M}_1 \to \mathcal{U}. \tag{315}$$

These embeddings establish which objects from the different dictionaries are to be treated as comparable alternatives.

However, the construction of such a common space already performs an analytical stabilization:

$$\mathcal{D} * 0, \mathcal{D} * 1 \longrightarrow \mathcal{D}^*. \tag{316}$$

It establishes a common dictionary in which the challenge alternatives can be identified and compared.

This is precisely the operation whose unique recoverability PADC places under protection.

Consequently, a conventional game cannot be treated as a neutral foundational definition of PADC security when the game presupposes a common dictionary that is not available to the external observer.

The prior PADC question is not

$$\text{which of two messages in } \mathcal{M} \text{ was selected?} \tag{317}$$

It is

$$\text{which alphabet, dictionary, message space, and object structure make the observation a message at all?} \tag{318}$$

Only after these spaces have been stabilized does a conventional indistinguishability game become well-defined.

Thus,

$$\text{standard game-based security} \tag{319}$$

begins after the central PADC reconstruction problem has already been resolved.

A.9 Why a Metatheoretical Hypothesis Set Does Not Supply the Observer's Dictionary

An analyst may define a metatheoretical class

$$\Omega \tag{320}$$

containing several candidate architectures.

The existence of this class for the author of the model does not imply that the external observer possesses a shared dictionary in which all members of Ω are already aligned.

The model author may write

$$\rho_0, \rho_1 \in \Omega, \tag{321}$$

because both reconstructions are represented in the author's metatheoretical language.

The observer, however, receives only

$$O. \tag{322}$$

The observer is not automatically given:

- the correspondence between the alphabets of ρ_0 and ρ_1;
- the correspondence between their dictionaries;
- the relation between their message spaces;
- the equivalence relation under which their objects become comparable;
- the intended embedding into the author's metatheoretical space.

Therefore,

$$\rho_0, \rho_1 \in \Omega \tag{323}$$

does not imply

$$\text{the observer can formulate } \rho_0 \text{ and } \rho_1 \text{ as two alternatives of one type.} \tag{324}$$

The metatheoretical hypothesis set belongs to the description of the security problem.

It does not automatically belong to the adversary's stabilized analytical domain.

A.10 PADC as a Different Epistemic Starting Point

Classical cryptography normally assumes that the following are given:

- the message space;
- the ciphertext space;
- the alphabet;

- the dictionary;
- the cryptographic-object type;
- the object boundaries;
- the relation of identity across observations;
- the relation of identity across sessions.

The remaining central uncertainty is usually the key, the message, or a random choice made inside those established spaces.

PADC begins from a different condition.

For the external observer, the following may remain non-given:

- the source-message space;
- the intermediate representation space;
- the ciphertext space;
- the source alphabet;
- the observable alphabet;
- the source dictionary;
- the observable dictionary;
- the partition into equivalence classes;
- the cryptographic-object structure;
- the cross-session identity relations;
- the mappings connecting observation to source interpretation.

The cryptographic key is therefore only one component of a broader hidden architectural state.

The difference may be summarized as follows.

In conventional cryptography,

$$\text{analytical space} + \text{unknown key} \longrightarrow \text{cryptanalytic problem.}$$

In PADC,

$$\text{observation} \longrightarrow \text{candidate analytical spaces} \longrightarrow \text{candidate cryptographic problems.}$$

The security objective is therefore not only to prevent inversion inside one established space.

It is to prevent the unique unauthorized stabilization of the space in which inversion would be defined.

A.11 Architectural Security Before Algorithmic Security

The architectural properties introduced in this appendix operate before conventional algorithm-level security properties.

They can be ordered as follows:

$$\begin{aligned}\text{observation} &\rightarrow \text{alphabet stabilization}\\ &\rightarrow \text{dictionary stabilization}\\ &\rightarrow \text{cross-session alignment}\\ &\rightarrow \text{cryptographic-object stabilization}\\ &\rightarrow \text{message-space stabilization}\\ &\rightarrow \text{conventional cryptographic analysis.}\end{aligned}$$

RSNU, SAP, ANS, and CSAD concern the earlier stages of this sequence.

Conventional cryptographic definitions concern the later stage, after the analytical domain has already been established.

The two levels are therefore complementary but not interchangeable.

A PADC architecture may preserve ANS while using an internally weak cryptographic primitive.

Conversely, an internally strong cryptographic primitive may operate inside a representation that exposes a unique and stable analytical architecture.

A complete security evaluation must therefore distinguish:

1. algorithmic security of E_K;
2. architectural non-stabilizability of (F_Θ, R_Φ);
3. cross-session preservation of that non-stabilizability.

The central architectural principle is:

> Before an observer can attack a key, message, or cryptographic transformation, the observer must possess a stable alphabet, dictionary, message space, and cryptographic-object space in which those objects can be identified.

PADC places the formation of that analytical space within the protected architecture.

References

[1] A. A. Nekludoff, “Alphabet formation and the epistemic limits of cryptographic analysis,” AstraVerge Research, Research Paper, Dec. 31, 2025. DOI: 10.5281/zenodo.18126229 [Online]. Available: https://doi.org/10.5281/zenodo.18126229

[2] A. A. Nekludoff, “Before information: Languages, dictionaries, alphabets, and messages,” AstraVerge Research, Research Paper, 2026. DOI: 10.5281/zenodo.20771404 [Online]. Available: https://doi.org/10.5281/zenodo.20771404

[3] A. A. Nekludoff, “Dynamics of linguistic constructions: Coordination, canonization, and divergence of dictionaries,” AstraVerge Research, Research Paper, 2026, In Russian. DOI: 10.5281/zenodo.20727647 [Online]. Available: https://doi.org/10.5281/zenodo.20727647

[4] A. A. Nekludoff, “Lseg: A segment-based protocol for data interpretation,” AstraVerge Research Lab, Research Paper, 2025, In Russian. DOI: 10.5281/zenodo.17786342 [Online]. Available: https://doi.org/10.5281/zenodo.17786342