

Bitogen Concept – A New Model for Data Representation, Compression, and Cryptography

1. Introduction

A bitogen is an abstract unit of information inspired by biological genes. It represents a predefined, fixed bit-pattern associated with a global bitogen map. A bitogen can replace raw bit sequences, introduce semantic structure, and enable more effective compression and new cryptographic methods.

2. Properties of a Bitogen

- It is a higher-level symbol compared to a bit.
- It is defined by a constant, globally known mapping to a bit-pattern.
- It can have variable length (e.g., 1–32 bits).
- It can encode recurring bit-patterns, text fragments, or structural units.

3. Bitogen Tokenization

Raw data (bit stream) is segmented into bitogens using a longest-match rule.

The result is a sequence of bitogen indices, replacing operations on raw bits.

4. Applications in Compression

- Bitogens reduce entropy by representing complex structures as symbols.
- They enable prediction at the symbol level rather than the bit level.
- They integrate naturally with AI-driven prediction models.
- They allow sub-bit coding through arithmetic or ANS-based compression.

5. Applications in Cryptography

5.1. Bitogen Layer

- Key-dependent permutation of bitogen indices.
- Optional stream transformations on indices (XOR, rotations).
- Ability to build high-level S-boxes operating on symbolic units.

5.2. Hybrid Bitogen Cipher

1. Bitogenization of plaintext.

2. Key-derived permutation of bitogen symbols.
3. Stream transformations on the permuted sequence.
4. Final encryption using a classical AEAD cipher such as AES-256-GCM or ChaCha20-Poly1305.

5.3. Advantages

- Structural obfuscation prior to classical encryption.
- Reduced statistical leakage.
- Combines traditional cryptographic strength with symbolic abstraction.

6. Global Bitogen Map

- Public and constant, similar to ASCII or the genetic code.
- Provides a semantic layer for binary data tokenization.
- Enables deterministic decoding and simplified symbolic modeling.

7. Research Directions

- AI-based learning of optimal bitogen maps.
- Integration into generative compression systems.
- Secure construction of bitogen permutations.
- Potential interaction with quantum data representation.

8. Use Cases

- New generations of compression codecs.
- Hybrid or layered cryptographic systems.
- Structural data analysis and classification.
- Obfuscation, watermarking, and metadata embedding.

9. Summary

Bitogens introduce a novel abstraction layer for data, merging ideas from compression, cryptography, and biological coding systems. They may serve as a foundation for innovative methods of data processing, modeling, and protection.