A Statistical and Topological Analysis of Private Key Distributions

in secp256k1 via Topological Projection

Abstract

This report presents an empirical and mathematical analysis of the distribution of *topologies* induced by private keys on the elliptic curve secp256k1.
A *topology* is defined as a structural invariant of the scalar multiplication algorithm, describing the ordered configuration of point doublings and additions used to compute a public key from a private scalar.

Using two independent datasets of randomly generated private keys—one produced in a local environment and one in a remote cloud environment (Kaggle)—we analyze the distribution of the top 5,000 most frequent topologies. Despite differences in operating systems, execution environments, and entropy sources, we observe a strong and reproducible concentration phenomenon: approximately 97% of generated private keys map into a small subset of ≈5,000 topologies out of a theoretical space of 86,709,504 possible topologies.

Moreover, a direct comparison of the two top-5,000 topology sets reveals a large and statistically significant intersection (≈77%), indicating the existence of an invariant, high-measure core of topologies. We argue that this phenomenon is not attributable to weaknesses in random number generation, but rather to intrinsic structural properties of the topological projection itself.

Finally, we explain why this topological formulation of the Elliptic Curve Discrete Logarithm Problem (ECDLP) is fundamentally different from classical brute-force approaches and why it opens a distinct research direction without contradicting the assumed hardness of ECDLP.

---

1. Introduction

The Elliptic Curve Discrete Logarithm Problem (ECDLP) is classically formulated as follows:
given a base point G on an elliptic curve and a public key $Q = d \cdot G$, recover the private scalar d.

In standard cryptographic analysis, the private key $d \in Z_n$ is treated as an atomic, structureless object, and the scalar multiplication $d \cdot G$ is viewed as a black-box group operation. Under this model, exhaustive search (or its algorithmic equivalents) is unavoidable, leading to exponential complexity.

The work underlying this report departs from this abstraction by introducing a *topological representation* of private keys, derived from the internal structure of the scalar multiplication algorithm itself.

---

2. Definition of Topology

2.1 Scalar Multiplication as a Structured Process

Scalar multiplication on elliptic curves is typically implemented via a sequence of:

- point doublings,
- conditional point additions.

For a fixed curve and implementation, the doubling points

$$D_i = 2^i G$$

form a strictly ordered sequence. A private key d induces a specific pattern of additions over this ordered set.

2.2 Topology as a Structural Invariant

A *topology* is defined as the ordered structural configuration of the scalar multiplication process, characterized by:

1. the indices of doubling points involved in additions,
2. the total number of additions,
3. the entry point (first addition),
4. the exit point (last addition, yielding the public key),
5. the non-permutability of indices due to fixed doubling order.

Crucially, many distinct private keys may induce the same topology. Thus, topology defines an equivalence class over the space of private keys.

---

3. Experimental Setup and Data

Two datasets were analyzed:

- **Dataset A (Local)**:
  Random private keys generated on a personal laptop.
- **Dataset B (Remote)**:
  Random private keys generated in a Kaggle environment.

In both cases:

- the curve is secp256k1,
- approximately 100,000 private keys were generated,
- an enhanced entropy generation method was used to minimize environment-dependent bias,
- topologies were computed and counted.

From each dataset, the **top 5,000 most frequent topologies** were extracted from the section “Grouped key count by conditions”.

---

4. Empirical Results

4.1 Concentration of Measure

For both datasets, the following empirical fact holds:

- ≈97% of all generated private keys map into the top 5,000 topologies.

This must be contrasted with the total number of theoretically possible topologies for secp256k1:

$$N_{\text{total}} = 86{,}709{,}504.$$

Thus, approximately

$$\frac{5000}{86{,}709{,}504} \approx 5.8 \times 10^{-5}$$

of the topology space accounts for almost all observed keys.

Under any reasonable assumption of near-uniformity, such concentration would be impossible.

4.2 Cross-Environment Intersection

Let:

- T_{local}^{5000} be the set of top 5,000 topologies from Dataset A,
- T_{remote}^{5000} be the corresponding set from Dataset B.

Empirical comparison yields:

$$|T_{\text{local}}^{5000} \cap T_{\text{remote}}^{5000}| = 3854,$$

which corresponds to:

$$\frac{3854}{5000} \approx 77.1\%.$$

The Jaccard similarity coefficient is:

$$J = \frac{|T_{\text{local}}^{5000} \cap T_{\text{remote}}^{5000}|}{|T_{\text{local}}^{5000} \cup T_{\text{remote}}^{5000}|} \approx 0.63.$$

For a space of 86 million possibilities, the expected intersection under independence would be below 1. The observed value exceeds this by **four orders of magnitude**.

4.3 Rank Instability vs Membership Stability

While individual topology rankings differ between environments (average rank displacement ≈528), the *membership* in the top-5,000 set is remarkably stable.

This indicates:

- local frequency fluctuations,
- but a globally stable high-measure subset.

---

5. Interpretation of the Distribution

5.1 Excluding Trivial Explanations

The persistence of the same core topologies across:

- different operating systems,
- different execution environments,
- different entropy sources,

rules out:

- weak RNG artifacts,
- environment-specific bias,
- accidental sampling effects.

5.2 Structural Explanation

The only remaining explanation is structural:

> The mapping
>
> $$d \longmapsto \text{topology}(d)$$
>
> is **not measure-preserving**.

Even if d is uniformly distributed in Z_n, its image under the topological projection is highly non-uniform.

This implies the existence of:

- attractor regions in topology space,
- a small, high-measure core,
- a vast low-measure periphery.

---

6. Implications for ECDLP

6.1 Why This Is Not Classical Brute Force

Classical brute force treats each private key as independent and equally likely. Partial information does not reduce the search space structurally.

In contrast, the topological formulation introduces:

- equivalence classes of keys,
- ordered, non-permutable trajectories,
- entry and exit constraints,
- monotonic reduction of combinatorial space upon partial recovery.

Partial reconstruction of a trajectory *eliminates entire index intervals* from consideration, producing a cascading reduction in complexity—something impossible in the classical model.

6.2 What Is (and Is Not) Claimed

This work does **not** claim:

- a break of secp256k1,
- an efficient algorithm for solving ECDLP,
- a practical attack.

What it does establish is:

> The *effective* space explored by real private keys, when viewed through a topological lens, is dramatically smaller and highly structured.

This changes the geometry of the problem, even if the worst-case hardness remains unchanged.

---

7. Scientific Significance

The key contribution of this work is not an attack, but a **new representation**:

- it reveals structure invisible to the classical formulation,
- it enables incremental progress rather than all-or-nothing search,
- it opens the door to statistical, combinatorial, and learning-based analysis.

Any cryptographic hardness argument that ignores this structure is, at minimum, incomplete with respect to real-world key distributions.

---

8. Conclusion

The empirical evidence demonstrates a strong, reproducible, and environment-independent concentration of private keys into a small, stable subset of topologies for secp256k1.

This phenomenon:

- cannot be explained by randomness artifacts,
- arises from intrinsic properties of the scalar-multiplication topology,
- fundamentally alters the effective landscape of ECDLP when viewed through this lens.

The topological approach does not contradict the established hardness of ECDLP, but it **reframes the problem** in a way that exposes new structure, new invariants, and new avenues for rigorous research.

This justifies further investigation of topological methods as a distinct and scientifically valid direction in elliptic-curve cryptography research.