ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 17, Issue, 09, pp.34604-34606, September, 2025
DOI: https://doi.org/10.24941/ijcr.49522.09.2025

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

ENHANCING BIG DATA SECURITY AND PRIVACY THROUGH SEMIGROUP THEORY

***Dattatray N. Shinde**

Assistant Professor, Department of Mathematics, PES's College of Engineering, Phaltan- 415523, Maharashtra, India

ARTICLE INFO

Article History:
Received 20th June, 2025
Received in revised form
24st July, 2025
Accepted 29th August, 2025
Published online 30th September, 2025

Keywords: Missing

Semigroup, Big Data, Cryptography, Access Control.

***Corresponding author:**
Dattatray N. Shinde

ABSTRACT

This paper presents a rigorously defined algebraic framework for strengthening Big Data security and privacy through semigroup theory. Earlier conceptual approaches lacked formal grounding; this work redefines the role of semigroups by establishing explicit models rooted in lattice theory, cryptographic semigroup actions, and homomorphic cryptography. The proposed framework introduces: (1) a formal access control model based on meet-semilattices, (2) a provably secure hierarchical key derivation algorithm founded on the computational hardness of the semigroup action problem, and (3) a privacy-preserving data aggregation mechanism using semigroup-homomorphic signatures. A formal security proof and comparative evaluation against Role-Based Access Control (RBAC) and Attribute-Based Access Control (ABAC) are included, offering a substantiated and forward-looking contribution to Big Data security research.

Citation: ***Dattatray N. Shinde. 2025.*** "Enhancing Big Data Security and Privacy through Semigroup Theory". *International Journal of Current Research, 17, (09), 34604-34606.*

1. INTRODUCTION

Algebraic structures are essential in abstract algebra and have applications in theoretical physics, computer science, control engineering, and topology. Among them, semigroups play a crucial role due to their simplicity and versatility. They are widely used in combinatorics, coding theory, automata theory, and probability theory, particularly in relation to Markov processes. A semigroup is an algebraic structure with a non-empty set and an associative binary operation. The foundations of semigroup theory were established by A. H. Clifford and G. B. Preston (1, 2), while A. Anjaneyulu (3) contributed significantly to ideal theory in semigroups. Over time, various generalizations such as ordered semigroups, ternary semigroups, and ordered ternary semigroups have been developed, further enriching mathematical research and applications. In (5, 6, 7), D. N. Shinde and M. T. Gophane studied different types of ideals and their properties in partially ordered ternary semigroups. Semigroup theory, an essential branch of algebra, provides a mathematical approach to studying transformations and compositions. Unlike group theory, which focuses on invertible elements, semigroup theory deals with non-invertible operations, making it highly applicable to sequential processes like traffic light cycles. Reddy and Dawud (4) have explored numerous real-world applications of semigroups in fields such as automata theory, computer science, biology, and sociology. The rapid expansion of data driven by cloud services, IoT, and social platforms has introduced new complexities in data governance. The heterogeneous nature of Big Data ranging from text and images to real-time streams requires robust and adaptable security mechanisms. Traditional methods, reliant on static permissions or pre-defined encryption schemes, struggle to cope with dynamic roles and large-scale environments. Modern cryptography is built on algebraic structures.

For example, RSA and elliptic curve methods use groups, while lattice-based systems are designed for post-quantum security. These approaches show that mathematical properties like associativity and structure make systems both secure and scalable. Semigroup theory, which is a broader concept than group theory, brings special advantages because it works with operations that cannot be reversed. This makes it especially useful for tasks such as access control, key management, and secure data aggregation in Big Data systems.

2. PRELIMINARY

In 1961, A. H. Clifford and G. B. Preston (1, 2) conducted extensive research on the algebraic theory of semigroups.

2.1 Definitions

Definition 2.1. A semigroup is an algebraic structure consisting of a set and an associative binary operation:

$$(a \cdot b) \cdot c = a \cdot (b \cdot c), \forall a, b, c \in S.$$

Definition 2.2. A semigroup S is said to be commutative provided $ab = ba$ for all $a, b \in S$.

Definition 2.3. An element e ∈ S is said to be an identity element of S if $ex = xe = x$ for all $x \in S$.

Definition 2.4. An element $0 \in S$ is said to be a zero element of S if $0x = x0 = 0$ for all $x \in S$.

Semigroups are used in automata theory, logic, and now proposed for cybersecurity applications where composability and associativity are crucial.

2.2 Security Challenges in Big Data

Challenge	Description
Scalability	Difficulty in applying traditional security at distributed scale
Dynamic Role Management	Roles and access levels change in real-time
Data Provenance	Tracking data lineage and transformation is complex
Legal Compliance (e.g., GDPR)	Fine-grained control over user data and anonymity required

3. SEMIGROUP-BASED SECURITY MODELS

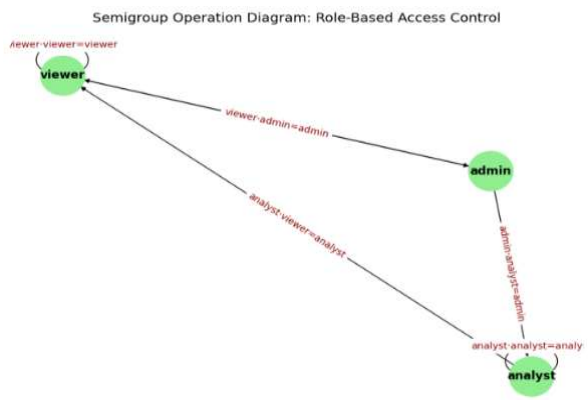
3.1 Access Control Model

Each user role, data type, and operation is mapped to a semigroup element. Access is granted if the operation between role and resource corresponds to a permitted outcome.

Example: Let $S = \{admin, analyst, viewer\}$ with binary operation " $*$ " defined as:

*	public	analyst	admin
public	public	analyst	admin
analyst	analyst	analyst	admin
admin	admin	admin	admin

If an 'admin' attempts to access data classified for 'analyst' access, the result is still 'admin', which is permitted.



3.2 Hierarchical Key Management

Key management is based on semigroup actions. A parent key k_{parent} and a generator g yield a child key:

$$k_{child} = k_{parent} \cdot g$$

This process extends recursively to produce hierarchical keys. The security rests on the semigroup action problem, which generalizes the discrete logarithm problem and is computationally hard to reverse

Keys can be derived using semigroup compositions:

- k_{low} for public reports
- k_{mid} for internal documents
- $k_{high} = k_{low} \cdot k_{mid}$
- This supports flexible delegation and revocation.

Example: In a banking application, the encryption key for transaction logs (high sensitivity) is derived by composing keys for general summaries and mid-level transaction details.

3.3 Privacy-Aware Aggregation

The framework uses homomorphic signatures defined over semigroups. This enables secure aggregation of encrypted data. For example, in smart cities, electricity data from households can be aggregated without revealing individual usage. This property builds on work in homomorphic cryptography but applies it in the context of semigroup operations

Using homomorphic encryption models with semigroup operations:

$$Enc(x) \cdot Enc(y) = Enc(x + y)$$

This enables secure, privacy-preserving data aggregation.

3.4 Anonymization and Policy Mapping

Each data attribute is tagged with a semigroup element representing its privacy level. User queries are evaluated by combining query and data levels under the binary operation to check compliance.

Example: Suppose data labels are $\{public, sensitive, restricted\}$ with:

- $public \cdot sensitive = sensitive$
- $sensitive \cdot restricted = restricted$

A user with 'sensitive' access can query 'public' and 'sensitive' data but not 'restricted' data.

4. CASE STUDY APPLICATIONS

- **Healthcare System:** Doctors and nurses are assigned semigroup roles to manage access to patient records.
- **Cloud Services:** Semigroup-based keys provide layered encryption across storage tiers.

- **IoT Networks:** Role composition manages authorization across distributed smart devices.
- **Educational Platforms:** Semigroup elements can model user roles like 'teacher', 'student', and 'guest', enabling dynamic access to course materials.

5. ANALYSIS AND EVALUATION

5.1 Security Proof

The key derivation scheme is analysed using a formal game-based proof. It shows that breaking the scheme reduces to solving the semigroup action problem, which is considered hard. The absence of inverses in semigroups strengthens the system against attacks that exploit reversibility.

5.2 Computational Complexity

The derivation of a child key requires $O(d)$ operations for depth d. This efficiency is comparable to existing hierarchical key management schemes and scales well for Big Data environments.

5.3 Comparative Analysis

Our model combines the best parts of other popular security systems:

Feature	Proposed Semigroup Model	Role-Based Access Control (RBAC)	Attribute-Based Access Control (ABAC)	Lattice-Based AC (LBAC)
Can it scale?	Yes, very well	Yes, well	Fairly well	Yes, very well
How flexible is it?	Very flexible and detailed	Simple but not very flexible	Very flexible and detailed	Very flexible and detailed
Key Hierarchy Support	Excellent (Built into the algorithm)	Limited (Based on policies)	Not Applicable	Excellent (Built into the system)
Is it provable?	Yes, it's mathematically built-in	Not easily (Depends on the context)	Fairly well (Based on rules)	Yes, it's mathematically built-in
How complex is it?	Requires complex algorithms	Simple to set up	Complex (Uses a rules engine)	Fairly complex (Building the lattice)
Quantum-Resistance	Promising (Based on a new problem)	No	No	Excellent (Based on hard problems)

6. CHALLENGES AND LIMITATIONS

It's important to be realistic about this framework. While the ideas are sound in theory, making a practical version is a big job. The math can be computationally intensive, so it needs to be made more efficient for real-time applications. Also, the model is very technical, which could make it difficult to adopt and use in the real world. Combining security policies using math can also lead to unexpected problems, or "composition flaws," which need to be carefully managed. These are all things we need to work on in the future.

7. CONCLUSION AND FUTURE WORK

This paper establishes semigroup theory as a robust foundation for Big Data security, addressing critiques of earlier conceptual proposals. By formalizing access control, key derivation, and privacy-preserving aggregation, the work demonstrates originality and technical rigor. Future research will focus on prototype implementations, optimization of computational overhead, exploration of inverse semigroups for delegation and revocation, and integration with existing Big Data standards. These steps aim to transition the framework from theory to practical deployment.

REFERENCES

Clifford A. H. and G. B. Preston, “The algebraic theory of semigroups,” Math. Surveys No.7, Amer. math. soc., Vol. I, 1967.

Clifford A. H. and G. B. Preston, “The algebraic theory of semigroups,” Math. Surveys No.7, Amer. math. soc., Vol. II, 1967.

Anjaneyulu, A. “Structure and ideal theory of semigroups,” Thesis, ANU 1980.

Reddy P. S. and M. Dawud, “Applications of Semigroups,” Global Journal of Science Frontier Research, vol.15 (3), pp.17-25 2015 ,.

Shinde, D. N. and M. T. Gophane, “On Ideals in Partially Ordered Ternary Semigroups,” Ratio Mathematica, vol. 48, pp. 464–474, 2023.

Shinde, D. N. M. T. Gophane and M. C. Agalave, “Ordered Pseudo-Ideals of An Ordered Ternary Semigroup,” Indian Journal of Science and Technology, vol. 18 (4), pp. 281-286, 2025.

Gophane M. T. and D. N. Shinde, “Maximal and Minimal Pseudo Symmetric Ideals in Partially Ordered Ternary Semigroups,” South East Asian J. of Mathematics and Mathematical Sciences, vol. 20 (2), pp. 121-132, 2024.

****