Multicollision attacks on tree-based hash functions

UDC 004.056.55:004.032.26
Vitalii Kazmirevskyi[1], Yurii Baryshev[2]
Vinnytsia National Technical University,
[1]kazmirevskiy1999@gmail.com, [2]yuriy.baryshev@vntu.edu.ua

Tee-based hash functions are increasingly adopted across a wide range of practical applications: ensuring and verifying data integrity in blockchain systems, optimization of electronic document workflow, structured hashing of complex objects (such as JSON or XML), digital signature schemes etc. This approach allows reduced complexity of hash values updating due to ability of processing only an updated part of the hashed message. However, tree structure exhibits internal non-uniformity, which creates new attack vectors such as subtrees manipulation.

The aim of this research is to improve tree-based hash functions infeasibility be analyzing the structural vulnerabilities of tree-based hash functions in the context of multicollision attacks.

One of the most important attacks is the method proposed by Antoine Joux in 2004 [1], which is schematically illustrated in Figure 1.

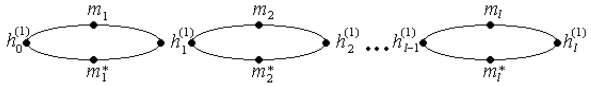

Fig. 1. Schematic representation of Joux's multicollision.

It is based on the ability to combine collisions of the compression function into an exponentially growing set of messages that yield the same hash value [1]. Another significant method is the Colliding subtree attack, which relies on the repeated usage of collisions to construct tree nodes those yield to identical hash values. In the absence of domain separation or positional information within the tree structure, this attack enables the generation of 2^k collisions with the same computational cost as finding k collisions [2]. Another case of efficient attack is the Herding-based tree collision attack [3]. Due to the hierarchical structure of the hash tree, this attack allows an adversary to manipulate not only individual nodes but entire subtrees.

Hash tree grafting attack, which exploits the ability to replace a subtree of the hash tree with an alternative, structurally compatible subtree containing different content, without altering the final hash. This attack is based on the existence of collisions between subtrees that differ in input data but produce the same intermediate hashes [4]. The position-blind multicollision attack constitutes a separate class of attacks those arise when the compression function of the hash does not rely on the position of input blocks within the tree [3]. A further critical threat is posed by the Expand-compress tree attack [5], which is based on alternating phases of tree expansion and compression to create multicollisions between trees of differing topology but with the same resulting hash. A summarized comparative analysis of these attacks is presented in Table 1.

Table 1

Comparative analyses of the multicollision attacks

Attack	Threat	Complexity
Joux's attack	Generation of an exponentially growing number of multicollisions	Decreases with tree height
Colliding Subtree	Possibility of subtree manipulation	Decreases with tree height
Herding-based Tree Collision	Possibility of manipulating subtrees	High complexity; requires manipulation of tree structure
Hash Tree Grafting Attack	Ability to alter the tree structure	Medium complexity; requires finding subtrees collisions
Position-Blind Multicollision	Generation of an exponentially growing number of multicollisions	Decreases with tree height
Expand-Compress Tree Attack	Construction of different trees with the same final hash	Technically complex; effective when structural protections are absent

Tree-based hash functions, despite their advantages, are vulnerable to multicollision attacks. The research findings confirm the critical impact of tree height and the structure of the compression function on the overall resilience of the system. The analyzed attacks highlight the necessity of integrating contextual information, positional encoding, and unique identifiers for nodes and levels in the design of tree-based hash functions.

1. Joux A. Multicollisions in Iterated Hash Functions. Application to Cascaded Constructions. *Advances in Cryptology - CRYPTO 2004*, 2004. p. 306-316. URL: https://link.springer.com/chapter/10.1007/978-3-540-28628-8_19 (application date 23.04.2025).

2. Andreeva E., Bouillaguet C., Dunkelman O., Fouque P., Hoch J., Kelsey J., Shamir A., Zimmer S. New Second-Preimage Attacks on Hash Functions. *Journal of Cryptology*. – 2015. – V. 29 – p. 657-696. (application date 23.04.2025).

3. Andreeva E., Bouillaguet C., Dunkelman O., Kelsey J. Herding, Second Preimage and Trojan Message Attacks Beyond Merkle-Damgard. Selected Areas in Cryptography, 2009. p. 393-414. URL: https://link.springer.com/chapter/10.1007/978-3-642-05445-7_25 (application date 23.04.2025).

4. Castelnovi L., Martinelli A., Prest T. Grafting Trees: A Fault Attack Against the SPHINCS Framework. *Post-Quantum Cryptography*, 2018. p. 165–184. URL: https://eprint.iacr.org/2018/102.pdf (application date 23.04.2025).

5. Blackburn S., Stinson D., Upadhyay J.. On the complexity of the herding attack and some related attacks on hash functions. *Designs, Codes and Cryptography*,

2011. p. 171-193. URL: https://eprint.iacr.org/2010/030.pdf (application date 23.04.2025).

Дослідження ролі машинного навчання та глибоких нейронних мереж у боротьбі з фінансовими злочинами

УДК 004.8 Богдан Калинюк[1], Ірина Замрій[2]
Державний університет інформаційно-комунікаційних технологій,
[1]*b.kalyniuk@duikt.edu.ua*, [2]*i.zamrii@duikt.edu.ua,*

У сучасному фінансовому середовищі кількість та складність шахрайських схем стрімко зростає, що зумовлює необхідність впровадження інтелектуальних систем виявлення загроз.

Класичні методи, що базуються на правилах або статистичних порогах, виявилися обмеженими у виявленні складних і нових форм фінансових злочинів. З огляду на це, особливу увагу привертають підходи машинного навчання (ML) та глибоких нейронних мереж (DNN), які здатні аналізувати великі обсяги даних транзакцій, виявляючи аномалії навіть у складних часових і структурних шаблонах [1].

Мета дослідження — дослідити потенціал сучасних методів машинного та глибокого навчання для виявлення фінансових злочинів, оцінити їхню точність, адаптивність, продуктивність, а також проаналізувати можливості їх практичного впровадження у фінансових установах.

Актуальність дослідження зумовлена необхідністю підвищення ефективності систем протидії шахрайству в умовах динамічного зростання обсягу електронних транзакцій, зниження довіри користувачів до фінансових послуг через часті інциденти шахрайства, а також появи більш витончених атак, які не фіксуються традиційними інструментами [2].

Наукова новизна роботи полягає у порівняльному аналізі продуктивності кількох архітектур DNN — зокрема згорткових нейронних мереж (CNN), довготривалої короткочасної пам'яті (LSTM), автоенкодерів (AE) та графових нейронних мереж (GNN) — у задачах детекції аномалій у транзакційних потоках.

Додатково проаналізовано гібридні моделі, які поєднують навчання з підкріпленням (Reinforcement Learning, RL) з моделями класифікації, що дає змогу адаптуватися до нових шаблонів поведінки моделі в режимі реального часу [3].

Дослідження базується на обробці двох відкритих наборів даних:

1) Credit Card Fraud Detection Dataset (розміщений на Kaggle, містить анонімізовані дані реальних транзакцій з високою диспропорцією між класами) [4];

2) PaySim — симульований набір мобільних грошових переказів на основі даних з M-Pesa (опублікований у Harvard Dataverse) [5].

Було здійснено повний цикл обробки даних: очищення, нормалізація, зменшення дисбалансу класів через SMOTE, побудова та тренування моделей. Нижче подано порівняльну таблицю з результатами основних метрик: