非代替性トークンにおける真正性保証とセキュリティ脅威への対策

田仲 巧磨[1]　櫻井 幸一[1,*]

概要： NFT(Non-Fungible Token/非代替性トークン)はデジタル資産の唯一性を保証する技術として注目されるが、識別不能状態やセキュリティ上の脅威が課題である。発表者らは、先に NFT ID の埋め込みによる真正性保証を基に、識別不能状態を 9 分類し、それを引き起こす 6 つの攻撃を解析した。　さらに NFT 取引の各フェーズにおける 12 種の攻撃の発生傾向に関する既存最新の分析結果を加味し、本研究では、非代替性トークン（NFT）の信頼性を脅かす主要な技術的課題として、(1) メタデータ参照 URI の失効 (2)メタデータの改ざん、(3) 第三者による偽造 NFT の生成、の三点に着目し、それぞれに対する実効性の高い対策枠組みを提案する。

キーワード：非代替性トークン，デジタル資産，サイバー攻撃，スマートコントラクト

Ensuring Authenticity and Mitigating Security Threats in Non-Fungible Tokens

Takuma TANAKA[1,*]　Kouichi SAKURAI[1,2,†1]

Abstract: Non-Fungible Tokens (NFTs) have garnered significant attention as a technology capable of guaranteeing the uniqueness of digital assets. However, they face critical challenges related to indistinguishability and security threats. Building upon prior work in which the authors proposed authenticity assurance through embedded NFT identifiers, this study classifies nine types of indistinguishable states and analyzes six corresponding attack vectors. Furthermore, incorporating the latest analytical findings on twelve types of attacks occurring across various phases of NFT transactions, this research focuses on three major technical threats to NFT reliability: (1) expiration of metadata reference URIs, (2) tampering of metadata itself, and (3) generation of counterfeit NFTs by third parties. For each of these issues, the study proposes a robust and practical countermeasure framework aimed at enhancing long-term trust in NFT ecosystems.

Keywords: NFT (Non-Fungible Token), digital assets, Cyberattacks, smart contracts

1. はじめに

NFT（非代替性トークン）[1] は、ブロックチェーンを活用し、デジタル資産に唯一性を付与する技術であり、所有権の透明性や改ざん耐性を提供し、クリエイターが中間業者を介さずに直接収益を得られる点でも注目されている。

先行研究「非代替性トークン(NFT)における一意識別可能性の問題点に対する考察」[2]では、NFT コンテンツが一意に識別できなくなる原因、そしてその対策を NFT トークンとデジタルコンテンツの関係性で分類し考察した．さらに、田仲ら[3]は、メタデータを加味した管理システムにおいて、一意に識別できない状態を 9 つに分類し、これらの状態引き起こす 6 つ攻撃を解析した。

最近の研究[4]では、NFT がフィッシングや DoS 攻撃をはじめとした 12 種類に大別されるインシデントに分類されるセキュリティ課題を抱えていることが報告されているその一方で、対策としては、NFT ID をデジタル作品に埋め込み、作成者署名を用いることで真正性を保証する仕組みを提案し、セキュリティ課題に技術的解決策が提示されるなど防御手段[5]も研究さている。

本研究では 、発表者らの"一意に識別できない状態分類と解析"に関する一連の研究[2,3]に、上記の 12 種の攻撃分類[4]を加味し、NFT サイクルで起こる取引をフェーズごとに分析し、12 種類の攻撃と発表者ら[3]の 6 つの攻撃との対応関係を考察する。

また、本研究は、非代替性トークン（NFT）の信頼性を脅かす主要な技術的課題として、(1)メタデータ参照 URI の失効、(2)メタデータの改ざん、(3)第三者による偽造 NFT の生成、の三点に着目し、それぞれに対する実効性の高い対策枠組みを提案する。具体的には、(i)IPFS[6]やおよび Arweave[7]などの分散型ストレージ技術を活用した恒久的リンクの保持機構、(ii) メタデータのハッシュ値をブロックチェーン上に記録することで改ざん耐性を確保するプロトコル、(iii) さらにコンテンツ指紋と発行者署名を組み合わせた真正性検証ワークフローの設計を通じて、NFT の可用性と整合性を両立させるアーキテクチャを提示する。

1 九州大学
Kyushu University
* sakurai@inf.kyushu-u.ac.jp

2. NFT の一意識別（不）可能性の問題点

2.1 大城ら[2]の研究

大城ら[2]は、NFT コンテンツが一意に識別不可能となる要因を明らかにし、またブローカーおよびブロックチェーン参加者としてどのような対策が必要であるかについて、NFT トークンと関連付けられたデジタルコンテンツとの関係をもとに分類・考察したものである。さらに、購入者がデジタルコンテンツを一意に識別できなくなる攻撃に焦点を当て、SCIS2022 において発表された既存の 3 つの研究[2,3,4]を再検討し、それらの観点から脆弱性の存在について分析を行った。

この結果、ブロックチェーン参加者に対してピアツーピア取引サービスを提供するブローカーが、ブロックチェーン外部における中央集権的な管理体制を有している場合、ブローカーが信頼できない状況において新たな攻撃が発生する可能性があるという共通点を確認している。

2.2 田仲ら[3]の研究

先の大城ら[2] が発表した「非代替性トークン(NFT)における一意識別可能性の問題点に対する考察」では、NFT トークンとそれに紐づくデジタルコンテンツとの関係性を基に、NFT コンテンツが一意に識別不可能となる原因およびその対策について分類・考察を行なっていた。研究本来の目的は、これら二者の関係性を体系的に分析することで、消費者がコピー作品に関連付けられた偽の NFT を購入する可能性を高める一意識別性の妨害要因を明らかにすることであった。

しかしながら、この先行研究[2]ではメタデータを含めたシステム考察[8]が欠如していた。メタデータは Exif 情報に記述されており、画像のサイズや色、位置情報など、多岐にわたる属性情報を示す。この Exif 情報を NFT と結び付けることで、NFT の取引履歴を参照し、画像所有者の正当性を検証可能にする利用権管理システムが提案されている[8]。メタデータは NFT と密接に関連する要素であるため、田仲ら[3]は NFT とデジタルコンテンツの関係性にメタデータの視点を追加し、一意識別性が失われる場合について 9(=4＊2+1) つの分類を提示した。

具体的には、NFT トークンおよびデジタルコンテンツが単数または複数である場合に、メタデータが単数、複数、無、または重複している場合が想定された。さらに、これらのパターンに基づき、想定される攻撃の原因を次の 6 つに分類した。

i. スマートコントラクトの実装による改ざん
ii. NFT 化されたコピー作品を市場に流通
iii. コンテンツ管理による改ざんや削除が容易
iv. メタデータの改ざん、削除
v. メタデータの重複
vi. NFT 電子市場同士の相互運用性の欠如

加えて、電子署名や IPFS(InterPlanetary File System)、およびメタデータの利用を通じて、これらの原因に対する具体的な対策も提案した。

3. NFT セキュリティの知識体系化(SoK)

Ma らの研究[4]では NFT セキュリティに関連する文献の体系的レビューを行い、2024 年 5 月 1 日までに発表された 248 件のセキュリティレポートと 35 件の学術論文から 176 件のインシデントを特定している。より正確には調査の質を上げるべく、

A) 大手セキュリティ企業の技術レポート 213 件＋査読論文 18 本を対象にし、低品質情報を排除
B) キーワード検索で候補を集め、142 件に確定
C) 既に見つけた攻撃カテゴリ名を使って追加クロールし、漏れを最小化。
D) 3 レイヤー × 12 カテゴリ × 16 問題に手動マッピング
E) Contract／Market／Auxiliary(補助サービス)の階層へ全インシデントを割り振り、件数を可視化

までも、行なっている。結果として手作業で分析された、セキュリティインシデントの主要なカテゴリは、次の 12 個である：

1. Unsafe External Call: 外部コントラクト呼び出しの設計ミスにより再入可能な状態
2. Insufficient Access Control: mint 等の権限チェック不足による不正操作の可能性
3. Improper Business Logic: smart contract 側のロジック実装ミス・設計ミス
4. Bad Randomness: 予測可能な乱数等による NFT の衝突、不正操作
5. Wash trading: 自作自演で NFT を売買し価格を高騰させる(価値があるようにみせる)
6. Arbitrage: 複数マーケット間の価格差やロジックの違いを利用して利益を得る
7. Rug Pull: 開発者のプロジェクト放棄等による詐欺行為
8. Copyright Theft: 他人の著作物を無断で NFT 化・販売するコンテンツ登用
9. Website Exploitation: NFT 関連サイトの脆弱性を攻撃
10. 1NFT-themed Phishing: 偽 NFT サイトやトランザクション画面でユーザーを騙す

11. Asset-related Issue: メタデータや外部 asset の消失・改竄により一意性を壊す
12. Private Key Compromised: ユーザーの秘密鍵漏洩、古いバージョンの脆弱性

さらに Ma らは、潜在的な解決策や緩和戦略を検討し、これらの分析に基づき NFT セキュリティ参照フレームを構築した。加えて、NFT セキュリティ問題の特徴として、蔓延性・深刻度・対処困難性、を抽出している。

4. 今回の関連性分析と対策

まず、田仲ら[3]が考察していた６つの攻撃原因は、Ma ら[4]の分類の(1,2,3,4,9,11)に対応している事が読み取れる。本研究では、レイヤーの内補助サービスレイヤー(Auxiliary Service Layer)の脆弱性に着目する。

補助サービスレイヤーでは、(攻撃者がスクリプトを埋め込んだ NFT を鋳造／出品する)悪性スクリプトアップロード、(説明欄や画像にフィッシング URL などを添付)するスパム NFT、さらにはメタデータ改ざんなどの問題がある

今回は、特に URI 失効・メタデータ改ざん、偽の NFT(コピー)に焦点を当て、その解決案を議論する

[I] リンク切れ、メタデータ改ざんの対策の提案

(I-a) コントラクト設計で “物理的に書き換え不可能” にする、そのためには固定 のトークン URI を用いる。

(1-b) また、重要メタデータはスマートコントラクト内にハッシュを埋め込み、トークン URI が死んでも同一性を検証できるようにする。

(1-c) ストレージ構成で “差し替えても意味がない” 状態を作る、そのためには、Arweave[7] を用いた永久保存や完全オンチェーン化[9]を活用する。

[II] コピー／偽 NFT の対策の提案

基本的には、既存の NFT 埋め込みによる分類に基づく対策[10]で可能である。また、メタデータがコピーされた場合、デジタルコンテンツを参照することで本物を見分けることが可能となる。

5. おわりに

5.1 今後の展望

今後の展望としては、こうした対策が実際のマーケットプレイスやパブリックチェーン上でどの程度うまく機能するかを確かめるために、テストネット上で検証フレームワークを構築し、リンク切れ・改ざん・コピー生成といった多様なシナリオでストレステストを実施する必要がある。あわせて、導入コストやガス消費量、ユーザビリティへの影響を定量的に測定し、コミュニティ主導でアップグレードしやすい標準仕様にまとめる課題も出てくる。

With Smart Contract: さらに、NFT を支える基盤レイヤーであるスマートコントラクトにも脆弱性は多くある[11,12,13]。今回の関連性分析では、スマートコントラクトに関係する部分は、未着手のままとなっている。再入可能性攻撃、整数オーバーフロー、アクセス制御の不備といった典型的な欠陥を体系的に洗い出し、静的解析ツールと手動監査を組み合わせてリスクプロファイルを作成する事が今後の作業課題である。

5.2 NFT with AI-use の光と影

NFT と AI の融合による新しい創作モデル[14]: AI が生成したアートや音楽を NFT 化することで、著作権や収益化の新しい枠組みが生まれている。たとえば、AI が生成したイラストや音楽を NFT として販売する事例が増加しており、クリエイターの支援や新しいビジネスモデルの構築が期待されている。

NFT マーケットにおける AI の活用[15]: AI によるレコメンド機能や価格予測、トレンド分析などが NFT マーケットプレイスで導入されており、ユーザー体験の向上や取引の効率化に貢献している。

しかし

AI による偽造 NFT の生成と詐欺リスク[16]:AI が既存のアート作品を模倣・改変して NFT 化することで、著作権侵害や偽造品の流通が懸念されている。特に、AI が生成したコンテンツが本物と区別しづらい場合、購入者が誤認するリスクが高まる。

AI を悪用した NFT 詐欺の手口[17,18]: AI チャットボットを使ったフィッシング詐欺や、AI による偽のマーケットプレイスの構築など、NFT 関連の詐欺に AI が利用されるケースが報告されている 。

その上で、それら AI による NFT 偽造や NFT 詐欺に対する対策にも、また AI が積極的に活用されているというイタチごっこ現状にある[19,20]。最新の研究[21]でも、この同時進行の現状と今後の政策提言[22]までを論じている。

「生成 AI × NFT」: 現在、生成 AI が主流の時代に入っている。しかし、「生成 AI × NFT」の悪用を明示的に扱った学術論文はほとんどない現状にある。ただし、生成 AI の汎用研究を基礎に、NFT への応用やリスクを論じることは十分可能であろう。Zhou ら[23]は、GPT-3 による偽情報生成の特性と検出困難性を実証的に分析し、AI 生成コンテンツが信頼性・透明性・説得力を持つため、既存の検出モデルが機能しにくいと主張している。これは NFT 化された偽情報コンテンツ（例えば、偽の歴史的文書やアート）への適用可能性/危険性を示唆していると言える。

謝辞 参考文献[4]に論文[24]も加えた、脆弱性と対策に関する比較を調査してくれた研究室の北島琉斗君に感謝します。また、本研究の調査には、主要 AI として、MS-

copilot を活用しました。

参考文献

[1]”ERC-721 NON-FUNGIBLE TOKEN STANDARD”, https://ethereum.org/ja/developers/docs/standards/tokens/erc-721/

[2]大城侑也, 池辺慶, 櫻井幸一. 非代替性トークン(NFT)における一意識別可能性の問題点に対する考察.SCIS2023 暗号と情報セキュリティシンポジウム,3C1-3, 2023

[3]田仲 巧磨, 畑 雄大, 櫻井 幸一, メタデータを用いた NFT 管理システムの脆弱性分類と対策, 2023, 第 198 回 DPS 研究発表会

[4] Kai Ma, Jintao Huang, Ningyu He, Zhuo Wang,Haoyu Wang, SoK: On the Security of Non-Fungible Tokens, 2023, arXiv:2312.08000. also in Blockchain: Research and Applications (2025)

[5] João B. Sousa, Gonçalo Marques. "Authentic Non Fungible Tokens". ICBTA '23: Proc. 2023 6th International Conference on Blockchain Technology and Applications Pages 42 - 47.

[6] Juan Benet, " IPFS - Content Addressed, Versioned, P2P File System" [https://arxiv.org/abs/1407.3561] (2014)

[7] Sam Williams ″Arweave: A Protocol for Archiving the Web Permanently″[https://www.arweave.org] (2018)

[8] 岡達也, 藤本真吾, 東角芳樹. NFT による利用権管理手法の提案. コンピュータセキュリティシンポジウム 2023, 2G4-2

[9] Shanmugam, U., Krisha, K.M., Nancy John Mari Vienni, S., Nandhini, D.U. (2025). Implementing an NFT by Developing a Smart Contract in OpenSea Marketplace. In: Luhach, A.K., Jat, D.S., Ghosh, U., Gao, XZ., Lingras, P. (eds) Advanced Informatics for Computing Research. ICAICR 2023. Communications in Computer and Information Science, vol 2073. Springer, Cham.

[10]高橋 悠太, 松崎 なつめ 電子透かしを用いた NFT 偽造問題対策の提案, コンピュータセキュリティシンポジウム 2023

[11] Xin Wang, Xiaoqi Li " AI-Based Vulnerability Analysis of NFT Smart Contracts " arXiv:2504.16113 (2025)

[12] Chetan Pathade, Shweta Hooli " Exposing Hidden Backdoors in NFT Smart Contracts: A Static Security Analysis of Rug Pull Patterns " arXiv:2506.07974 (2025)

[13] 矢内直人, NFTスマートコントラクトの脆弱性診断と傾向分析, CSS/BWS2022 [https://www.iwsec.org/bws/2022/pdf/yanai_open_bws2022.pdf] (2025.08.22 HP/確認済み)

[14]出井甫『AI 生成物に関する知的財産権の現状と課題』 情報の科学と技術 68 巻 12 号, 580〜585（2018) [https://www.jstage.jst.go.jp/article/jkg/68/12/68_580/_pdf] (2025.08.22 確認済み)

[15]Abdelkader Chebli & Sarra Cherbal, "Building a Secure and Efficient NFT Marketplace: AI Integration, Lazy Minting, and TrustGuard," International Journal of Information Security, Springer, 24, 140 (2025).

[16] Nahema Marchal, Rachel Xu, Rasmi Elasmar, Iason Gabriel, Beth Goldberg, William Isaac "Generative AI Misuse: A Taxonomy of Tactics and Insights from Real-World Data" arXiv:2406.13843 (2024)

[17]Jason Scharfman, "The Cryptocurrency and Digital Asset Fraud Casebook, Volume II" Springer (2024)

[18] Anastasiia Terekhova, "The Impact of AI-Generated NFT Art on the Global Art Market" Scientific Journal of Bielsko-Biala School of Finance and Law, Vol. 29 No. 1 (2025)

[19] 木村圭吾,今村光良, 面和成, NFT流通における深層学習を用いた分散型真正性検証プロトコルの提案, SCIS2022, 1D1-1 (2022)

[20] Eric Esposito " Building a Secure and Efficient NFT Marketplace: AI Integration, Lazy Minting, and TrustGuard" (2023) [https://www.nft.com/articles/ai-tools-can-help-detect-fake-nfts-heres-how] (2025.08.22 確認済み)

[21] Merlin Balamurugan "AI vs. AI: The Digital Duel Reshaping Fraud Detection," European Journal of Computer Science and Information Technology 12 (7), 12-20 (2024)

[22] Nitin Upadhyay & Shalini Upadhyay "The Dark Side of Non-Fungible Tokens: Understanding Risks in the NFT Marketplace, " *Financ Innov* **11**, 62 (2025).

[23] Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury, "Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions," In Proc. 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, (2023)

[24] Mi-Na Shim, “Classification of NFT Security Issues and Threats through Case Analysis,” *International Journal of Internet, Broadcasting and Communication*, vol. 15, no. 1 (2023)