Uso malicioso de Deepfakes em ambientes IoT: Um mapeamento sistemático sobre ameaças, vetores e estratégias de mitigação

Almeida Italo M. Santos[1], **Marcos Vinícius de S. Santos**[2], **Rubens de Souza M. Junior**[3]

[1]Departamento de Computação (DCOMP) – Universidade Federal de Sergipe (UFS)
Av. Marechal Rondon, s/n – Jardim Rosa Elze – CEP 49100-000
São Cristóvão – SE – Brazil

{italoalmeida733, oficialviniciussantana, rubens.matos}@gmail.com

Abstract. *Context: The growing integration of AI and IoT technologies fuels smart environments but also brings new cybersecurity threats like* ***deepfakes****. These synthetic media can deceive facial recognition, voice authentication, and surveillance in connected spaces. Objective: This work systematically maps the malicious use of deepfakes in IoT, pinpointing* ***attack vectors****,* ***vulnerable sensors****, and* ***mitigation techniques****. Method: We conducted a systematic mapping, searching IEEE Xplore and Scopus. From 85 initial studies, we selected 16, analyzing them based on three research questions concerning threats, vulnerabilities, and defense mechanisms. Results: Our analysis shows* ***facial and voice authentication systems*** *are most affected. Key attack vectors include* ***face spoofing****,* ***voice cloning****, and* ***live video manipulation****. Effective detection and mitigation techniques in IoT scenarios involve* ***CNNs****,* ***blockchain****,* ***GNNs****, and* ***multimodal detection (Wi-Fi + video)****. Conclusion: Deepfakes are a concrete and evolving threat to IoT security.* ***Robust, lightweight detection models****, alongside* ***biometric authentication frameworks*** *and* ***cryptographic protection****, are crucial for real-time defense in resource-constrained IoT environments.*

Resumo. *Contexto: A crescente integração de IA e IoT tem viabilizado ambientes inteligentes, mas também expandido a superfície de ataque cibernético com ameaças como os* ***deepfakes****. Esses conteúdos sintéticos conseguem enganar sistemas de reconhecimento facial, autenticação por voz e vigilância em tempo real. Objetivo: Este trabalho mapeia sistematicamente o uso malicioso de deepfakes em ambientes IoT, identificando* ***vetores de ataque****,* ***sensores vulneráveis*** *e* ***técnicas de mitigação*** *presentes na literatura. Método: A pesquisa seguiu um mapeamento sistemático com buscas nas bases IEEE Xplore e Scopus. De 85 estudos iniciais, 16 foram selecionados e analisados com base em três questões de pesquisa sobre ameaças, vulnerabilidades e soluções. Resultados: A análise revelou que* ***sistemas de autenticação facial e vocal*** *são os mais impactados. Os principais vetores envolvem* ***falsificação de rosto****,* ***clonagem de voz*** *e* ***manipulação de vídeo ao vivo****. Técnicas como* ***CNN****,* ***blockchain****,* ***Redes Neurais Gráficas (GNN)*** *e* ***detecção multimodal (Wi-Fi + vídeo)*** *foram identificadas como eficazes na mitigação em ambientes IoT. Conclusões: Deepfakes representam uma ameaça real e crescente à segurança de dispositivos IoT.* ***Modelos leves e robustos de detecção****, aliados a* ***estruturas biométricas***

*e **mecanismos criptográficos**, são essenciais para enfrentar essas ameaças em tempo real e em dispositivos com recursos limitados.*

1. Introdução

O avanço da **Inteligência Artificial (IA)**, notadamente **Redes Generativas Adversariais (GANs)**, elevou os **deepfakes** ao centro do debate sobre segurança digital. Inicialmente associados à manipulação de áudio e vídeo em entretenimento, deepfakes agora permeiam ambientes críticos como a **Internet das Coisas (IoT)**. A falsificação de identidades via mídia sintética ameaça a confiabilidade de sistemas automatizados e a segurança em ambientes conectados [Javed et al. 2022].

Sistemas IoT, vastos e heterogêneos, são vulneráveis devido à ubiquidade, limitação de recursos e, muitas vezes, mecanismos de segurança insuficientes. A convergência deepfake-IoT cria uma nova fronteira cibernética, onde a manipulação de dados sensoriais, falsificação de identidades e subversão de decisões autônomas podem ter consequências catastróficas. Exemplos incluem deepfakes de áudio controlando dispositivos médicos ou deepfakes de dados de temperatura desregulando sistemas industriais.

Este mapeamento sistemático busca desvendar essa paisagem de ameaças, identificando e categorizando ameaças de deepfakes em IoT, vetores de ataque, estratégias de mitigação e lacunas de pesquisa, visando garantir a segurança e integridade dos ecossistemas IoT.

2. Metodologia

Este estudo emprega a metodologia de Mapeamento Sistemático da Literatura (MSL) [Petersen et al. 2015] para proporcionar uma visão estruturada do estado da arte sobre deepfakes e IoT. A metodologia foi delineada para garantir abrangência, transparência e reprodutibilidade.

2.1. Questões de Pesquisa (QPs)

As QPs abaixo guiaram a seleção e análise da literatura, focando nos aspectos críticos da intersecção deepfakes-IoT:

- **QP1**: Quais são os tipos e manifestações das ameaças de segurança que emergem do uso malicioso de deepfakes em ambientes IoT, e como essas ameaças se diferenciam entre os diversos domínios da IoT?
- **QP2**: Quais vetores de ataque e pontos de vulnerabilidade são explorados por agentes maliciosos para a implantação, propagação ou utilização de deepfakes em sistemas IoT?
- **QP3**: Quais estratégias de mitigação e contramedidas têm sido propostas para combater o uso malicioso de deepfakes em contextos IoT?
- **QP4**: Quais lacunas de pesquisa persistem na literatura acadêmica sobre deepfakes em IoT, e quais são os desafios futuros?

2.2. Estratégia de Busca e Artigos Selecionados

A estratégia de busca utilizou uma combinação de termos-chave e operadores booleanos: **(”deepfake”AND ”IoT”) AND (”security”OR ”threats”OR ”vulnerabilities”OR ”attacks”OR ”malicious use”OR ”cybersecurity”OR ”mitigation”OR ”detection”OR ”defense”)**. As bases de dados consultadas foram **IEEE Xplore** e **Scopus**. Os resultados foram gerenciados para remover duplicatas e facilitar a triagem.

2.3. Critérios de Seleção (Inclusão e Exclusão)

Artigos foram selecionados com base em rigorosos critérios:

2.3.1. Critérios de Inclusão

- **Relevância Temática Direta**: Interseção entre deepfakes (ou técnicas maliciosas de síntese de mídia) e IoT.
- **Foco em Segurança Cibernética**: Discussão de ameaças, vetores, vulnerabilidades ou mitigação de deepfakes em IoT.
- **Tipo de Publicação**: Periódicos científicos, anais de conferências internacionais e workshops especializados revisados por pares. Teses e dissertações podem ser consideradas.
- **Disponibilidade**: Acesso ao texto completo.
- **Período de Publicação**: A partir de 2018 até Maio de 2025.

2.3.2. Critérios de Exclusão

- **Irrelevância Temática**: Foco exclusivo em deepfakes sem IoT, ou segurança IoT sem deepfakes.
- **Tipo de Publicação**: Não revisadas por pares (blogs, notícias, white papers não revisados, pré-prints não validados).
- **Duplicatas**.
- **Escopo Excessivamente Limitado**: Estudos de caso sem generalizações aplicáveis.

2.4. Processo de Extração de Dados

Informações cruciais foram extraídas de cada estudo: **Dados Bibliográficos** (título, autores, ano, tipo), **Contexto da IoT** (domínio/aplicação), **Tipos de Deepfakes** (áudio, vídeo, imagem, sensoriais, rede, biometria), **Ameaças de Segurança**, **Vetores de Ataque**, **Estratégias de Mitigação Propostas**, e **Lacunas de Pesquisa/Desafios Futuros**. Os dados foram organizados em uma planilha para análise.

3. Resultados e Discussão

Este mapeamento sistemático confirmou que deepfakes representam uma ameaça complexa e crescente em ambientes IoT. A detecção isolada não é suficiente; mitigação deve abranger autenticação de proveniência e integridade de dados/dispositivos, segurança da comunicação e resiliência de sistemas IA/ML. As implicações são vastas, variando de interrupções em infraestruturas críticas a ameaças à segurança física. A manipulação convincente de dados sensoriais e a falsificação de identidades minam a confiança, essencial para a adoção da IoT.

3.1. Resultados

Das 85 publicações iniciais do IEEE Xplore e Scopus, 16 artigos foram selecionados após triagem por título, palavras-chave, resumos, introdução e conclusão. A Figura **??** ilustra o processo.

Uma assimetria de recursos entre atacantes (com IA generativa) e defensores (dispositivos IoT limitados) exige soluções de segurança eficientes e leves para edge computing. A dependência de ML para detecção levanta preocupações sobre ataques adversariais. Este cenário indica uma corrida armamentista contínua, demandando soluções de segurança adaptativas e evolutivas.

4. Análise dos Resultados

Esta seção sintetiza e analisa os dados para responder às QPs.

4.1. Visão Geral dos Estudos Selecionados

A pesquisa mostra crescente preocupação com deepfakes e segurança IoT desde 2020. A maior parte foca em Cidades Inteligentes e IIoT devido ao alto impacto. Smart Homes e Veículos Conectados também são relevantes. As modalidades de deepfakes incluem áudio, vídeo, e agora, dados sensoriais (temperatura, pressão, biometria) e de rede. As mitigação predominam em ML para detecção e DLT (Blockchain) para integridade. Desafios de recursos em dispositivos limitados são comuns.

A seguir, uma análise concisa dos artigos selecionados:

- [Bethu and Erukala 2025]: Arquitetura de segurança para reconhecimento facial em IoT (Deep Q-Network e Blockchain) para detecção de deepfakes em tempo real.
- [Karathanasis et al. 2025]: Técnicas de compressão e transferência de aprendizado para modelos de detecção de DeepFake em dispositivos de borda IoT.
- [Samrouth et al. 2025]: Modelo de detecção de deepfake em retratos (redes siamesas) para dispositivos com recursos limitados.
- [Liu et al. 2025]: Sistema WiSil de detecção de falsificação de vídeo cross-modal (visual + Wi-Fi) para vigilância IoT.
- [Eidmum et al. 2025]: Modelos de detecção de deepfake para reconhecimento facial (aprendizado profundo, empilhamento, meta-aprendizagem) em IoT.
- [Sisson and Puckett 2025]: Autenticação de imagens/vídeos IoT via assinatura digital com coprocessadores criptográficos para integridade legal.
- [Zhou et al. 2024]: Técnica leve de detecção de deepfake (inferência bayesiana) para sistemas IA embarcados em IoT.
- [Bethu et al. 2024]: Estrutura de segurança para vigilância IA-IoT (Deep Q Networks) para detecção de deepfakes.
- [Zhang et al. 2024]: Modelo de detecção de deepfake para CIoT (Transformador Espectral com Atenção Piramidal) para manipulações faciais.
- [Xu et al. 2024]: Modelo de detecção de deepfake em tempo real para câmeras IoT (GNNs) contra spoofing facial.
- [Fang et al. 2023]: Sistema WiSil (visão + WiFi) para detectar falsificações em vídeos de vigilância com alta precisão.
- [Javed et al. 2022]: Framework anti-spoofing de voz (padrões de coocorrência acústico-ternária e coeficientes cepstrais Gammatone) contra deepfakes em IoT controlada por voz.
- [Frolov et al. 2022]: Análise de deepfake como ameaça à segurança da informação, com soluções para IoT.

- [Mitra et al. 2022]: Prova de conceito de sistema de identidade digital (autenticação facial) para cidades inteligentes, contra deepfake e ataques de apresentação.
- [Mitra et al. 2021a]: Método eficiente de detecção de deepfake (CNN) para dispositivos IoT em cidades inteligentes.
- [Mitra et al. 2021b]: Sistema de identidade digital biométrica seguro contra deepfakes (CNN para características faciais e biochave criptográfica) em IoT de cidades inteligentes.

4.1.1. RQP1: Ameaças de Deepfakes em IoT

O uso malicioso de deepfakes em IoT introduz ameaças que comprometem a confiança em dados e identidades, incluindo:

- **Manipulação de Dados Sensoriais**: Geração de leituras falsas (temperatura, GPS) para enganar sistemas autônomos, com impactos em Indústria 4.0 (falhas catastróficas) e Cidades Inteligentes (desorganização de tráfego).
- **Falsificação de Identidade**: Criação de identidades digitais fraudulentas para dispositivos/usuários (e.g., deepfakes de voz para smart homes, enganando autenticação biométrica).
- **Injeção de Comandos Maliciosos**: Geração de comandos falsos mas convincentes para controlar dispositivos (e.g., deepfakes de áudio para alarmes ou veículos conectados).
- **Desinformação em Larga Escala**: Propagação de informações falsas para causar pânico ou interrupção de serviços críticos em cidades inteligentes.

A diferenciação reside na **criticidade e escala**: falhas em IoT industrial e veículos conectados podem ser físicas e letais; em Smart Homes, impactam privacidade e patrimônio; em Cidades Inteligentes, afetam serviços públicos e segurança coletiva.

4.1.2. RQP2: Vetores de Ataque e Vulnerabilidades

A implantação de deepfakes em IoT explora vulnerabilidades em várias camadas:

- **Comprometimento de Sensores e Atuadores**: Injeção direta de dados falsificados via acesso físico/remoto, explorando firmware desatualizado ou falta de validação na origem.
- **Canais de Comunicação Inseguros**: Interceptação e substituição de dados legítimos por deepfakes em redes sem criptografia robusta (ataques Man-in-the-Middle).
- **Vulnerabilidades em Gateways e Plataformas IoT**: Exploração de falhas para injetar/distribuir deepfakes em larga escala, frequentemente por falta de validação de dados na entrada.
- **Sistemas de IA e Aprendizado de Máquina Suscetíveis**: Envenenamento de dados de treinamento (data poisoning) ou criação de deepfakes que enganam modelos de IA/ML, explorando sua fragilidade adversarial.
- **Vulnerabilidades em Firmware e Software**: Falhas no código ou em atualizações inseguras permitindo injeção de deepfakes ou software malicioso gerador de deepfakes.

- **Engenharia Social e Credenciais Fracas**: Acesso inicial via senhas padrão ou manipulação humana, servindo como porta de entrada para ataques de deepfake.

A defesa deve ser multifacetada.

4.1.3. RQP3: Estratégias de Mitigação e Contramedidas

Para combater deepfakes em IoT, propõem-se diversas estratégias para uma defesa em profundidade:

- **Mecanismos de Detecção de Deepfakes**: Uso de ML para identificar artefatos e anomalias em dados (visuais, áudio, sensoriais), forense digital e análise de metadados. O desafio é a detecção leve para dispositivos com recursos limitados.
- **Autenticação e Verificação Robustas**: Emprego de Blockchain/DLT para garantir proveniência/integridade de dados; autenticação multifator e biometria segura com ”liveness detection”.
- **Criptografia e Comunicações Seguras**: Implementação de criptografia ponta a ponta (TLS/DTLS) e gerenciamento seguro de chaves.
- **Hardware Trust Anchors e Enclaves Seguros**: Uso de TPMs, HSMs e enclaves seguros para proteger chaves e processos críticos no hardware.
- **Educação e Conscientização**: Treinamento de desenvolvedores, operadores e usuários sobre ameaças de deepfakes e melhores práticas de segurança.
- **Padronização e Regulamentação**: Desenvolvimento de normas de segurança e leis para responsabilização, promovendo conformidade.

A combinação dessas estratégias é essencial para sistemas IoT resilientes.

4.2. RQP4: Lacunas de Pesquisa e Desafios Futuros

A pesquisa sobre deepfakes em IoT é incipiente, com desafios cruciais:

- **Detecção em Tempo Real com Recursos Limitados**: Desenvolver algoritmos leves e eficientes para edge computing, lidando com volume massivo de dados IoT.
- **Resistência a Ataques Adversariais**: Criar modelos de detecção robustos e adaptáveis a novas técnicas de falsificação.
- **Integração de Soluções Cross-Layer**: Desenvolver abordagens de segurança holísticas que integrem detecção e mitigação em todas as camadas (dispositivo, rede, nuvem).
- **Impacto Legal, Ético e Social**: Abordar questões de responsabilidade, implicações éticas e a necessidade de arcabouços legais/regulatórios adequados.

Superar esses desafios exige colaboração multidisciplinar.

4.3. Discussões dos Principais Trabalhos Selecionados

Os trabalhos analisados concordam que deepfakes são uma ameaça perigosa em IoT. [Samrouth et al. 2025] foca em redes siamesas para detectar falsificação facial, enquanto [Javed et al. 2022] propõe um framework anti-spoofing de voz para identificar padrões acústicos manipulados.

Para vídeo ao vivo, [Fang et al. 2023] e [Liu et al. 2025] combinam dados de câmeras e sinais Wi-Fi para detecção multimodal. [Bethu and Erukala 2025] enfatizam a integridade de dados via blockchain e aprendizado por reforço. [Zhang et al. 2024] abordam detecção sequencial de manipulações faciais, e [Xu et al. 2024] utilizam GNNs para detecção em tempo real em câmeras IoT.

[Zhou et al. 2024] discutem inferência Bayesiana leve para detecção eficiente em dispositivos IoT. Por fim, [Mitra et al. 2021b] demonstram um sistema de identidade digital robusto contra manipulação facial usando CNNs e biochaves criptográficas.

Esses trabalhos indicam que uma solução única não existe; uma combinação de abordagens pode aumentar a robustez da IoT contra deepfakes. A necessidade de detecção em tempo real e em larga escala, sem comprometer o desempenho, é um desafio central. A integração de segurança em todas as camadas da arquitetura IoT é crucial. A complexidade do problema exige uma abordagem multidisciplinar, envolvendo academia, indústria e governos para o desenvolvimento de padrões e regulamentações.

5. Conclusão

Este mapeamento sistemático forneceu uma visão abrangente sobre o uso malicioso de deepfakes em ambientes IoT. Foram delineadas ameaças emergentes, desde a manipulação de dados sensoriais e falsificação de identidades até a injeção de comandos maliciosos, com potenciais impactos catastróficos. Identificamos os principais vetores de ataque, destacando vulnerabilidades em dispositivos, canais de comunicação e plataformas IoT.

As estratégias de mitigação propostas na literatura, incluindo detecção baseada em IA, autenticação robusta (com blockchain), criptografia avançada e ancoragens de confiança de hardware, são promissoras. Contudo, o mapeamento também evidenciou lacunas de pesquisa e desafios futuros significativos. A capacidade de implementar soluções de detecção e mitigação em tempo real em dispositivos IoT com recursos limitados, a resiliência contra ataques adversariais e a importância de abordagens de segurança integradas e cross-layer são áreas que demandam mais investigação.

A ameaça dos deepfakes na IoT transcende o técnico, com implicações éticas, legais e sociais que exigem colaboração multidisciplinar. À medida que os ecossistemas IoT se expandem, a urgência de construir ambientes seguros e resilientes contra essa forma sofisticada de manipulação de dados só aumenta. Pesquisa futura e inovação em cibersegurança serão cruciais para garantir a confiança, integridade e confiabilidade da próxima geração de sistemas inteligentes e conectados.

Referências

Bethu, S. and Erukala, S. B. (2025). Blockchain-enhanced secure guard: a deep q network framework for robust iot surveillance person detection. *Cluster Computing*, 28(4). Cited by: 0.

Bethu, S., Trupthi, M., Mandala, S. K., Karimunnisa, S., and Banu, A. (2024). Ai-iot enabled surveillance security: Deepfake detection and person re-identification strategies. *International Journal of Advanced Computer Science and Applications*, 15(7):1013 – 1022. Cited by: 1; All Open Access, Gold Open Access.

Eidmum, M. Z. A., Muiz, B., and Saha, S. (2025). Enhancing fake face detection with meta-learning and ensemble methods. *International Journal of Computing and Digital Systems*, 18(1). Cited by: 0.

Fang, X., Liu, J., Chen, Y., Han, J., Ren, K., and Chen, G. (2023). Nowhere to hide: Detecting live video forgery via vision-wifi silhouette correspondence. volume 2023-May. Cited by: 3.

Frolov, D. B., Makhaev, D. D., and Shishkarev, V. V. (2022). Deepfakes and information security issues. In *2022 International Conference on Quality Management, Transport and Information Security, Information Technologies (ITQMIS)*, pages 147–150.

Javed, A., Malik, K. M., Malik, H., and Irtaza, A. (2022). Voice spoofing detector: A unified anti-spoofing framework. *Expert Systems with Applications*, 198. Cited by: 29; All Open Access, Bronze Open Access.

Karathanasis, A., Violos, J., and Kompatsiaris, I. (2025). A comparative analysis of compression and transfer learning techniques in deepfake detection models. *Mathematics*, 13(5). Cited by: 0; All Open Access, Gold Open Access.

Liu, J., Fang, X., Chen, Y., Yuan, J., Yu, G., and Han, J. (2025). Real-time video forgery detection via vision-wifi silhouette correspondence. *IEEE Transactions on Mobile Computing*, 24(3):1585–1601.

Mitra, A., Bigioi, D., Mohanty, S. P., Corcoran, P., and Kougianos, E. (2022). Iface 1.1: A proof-of-concept of a facial authentication based digital id for smart cities. *IEEE Access*, 10:71791 – 71804. Cited by: 11; All Open Access, Gold Open Access.

Mitra, A., Mohanty, S. P., Corcoran, P., and Kougianos, E. (2021a). Detection of deepmorphed deepfake images to make robust automatic facial recognition systems. page 149 – 154. Cited by: 12.

Mitra, A., Mohanty, S. P., Corcoran, P., and Kougianos, E. (2021b). iface: A deepfake resilient digital identification framework for smart cities. In *2021 IEEE International Symposium on Smart Electronic Systems (iSES)*, pages 361–366.

Petersen, K., Vakkalanka, S., and Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64:1–18.

Samrouth, K., El Housseini, P., and Deforges, O. (2025). Siamese network-based detection of deepfake impersonation attacks with a person of interest approach. *ACM Transactions on Multimedia Computing, Communications and Applications*, 21(3). Cited by: 0.

Sisson, E. S. and Puckett, S. C. (2025). Securing digital media in the age of ai and deepfakes: A hardware-based solution. page 1171 – 1176. Cited by: 0.

Xu, J., Lin, W., Fan, W., Chen, J., Li, K., Liu, X., Xu, G., Yi, S., and Gan, J. (2024). A graph neural network model for live face anti-spoofing detection camera systems. *IEEE Internet of Things Journal*, 11(15):25720–25730.

Zhang, G., Gao, M., Li, Q., Guo, S., and Jeon, G. (2024). Detecting sequential deepfake manipulation via spectral transformer with pyramid attention in consumer iot. *IEEE Transactions on Consumer Electronics*, pages 1–1.

Zhou, L., Ma, C., Wang, Z., Zhang, Y., Shi, X., and Wu, L. (2024). Robust frame-level detection for deepfake videos with lightweight bayesian inference weighting. *IEEE Internet of Things Journal*, 11(7):13018–13028.