ISSN: 2710-0588

A Survey on the Applications of Artificial Intelligence in Cryptanalysis and Cryptographic Design

Shuangjin Wu[1,*], Wenbo Wang[1]

[1]Information Engineering College, Henan University of Science and Technology, Luoyang 471000, China.

[*] Corresponding Author

Abstract

Artificial Intelligence (AI) is profoundly transforming cryptography by significantly enhancing cryptanalysis techniques and informing innovative cryptographic design approaches. This survey reviews recent advancements in applying deep learning methods to side-channel and differential fault analyses, demonstrating substantial improvements over traditional methods in attack efficiency, accuracy, and resilience. Additionally, it highlights breakthroughs such as neural differential cryptanalysis, which expand classical cryptanalytic boundaries. In cryptographic design, Generative Adversarial Networks (GANs) have successfully automated the creation of high-quality cryptographic primitives, particularly S-boxes. Furthermore, AI shows promise in post-quantum cryptography (PQC) by uncovering potential vulnerabilities and optimizing cryptographic parameters. Despite these advancements, challenges persist regarding data dependency, model generalization, and interpretability. Future research directions emphasize enhancing AI model explainability, creating standardized benchmarks, and integrating AI with emerging technologies such as quantum computing and zero-knowledge proofs.

Keywords

Cryptography, Cryptanalysis, Side-Channel Analysis (SCA), Differential Fault Analysis (DFA), Neural Differential Cryptanalysis, Generative Adversarial Networks (GANs), Lightweight Cryptography.

1. Introduction

Cryptography plays a pivotal role in safeguarding information security, forming the foundation of mechanisms such as data encryption and identity authentication, all of which rely on robust cryptographic algorithms. With the continuous evolution of attack techniques, artificial intelligence (AI) is increasingly demonstrating its potential in the field of cryptography. As early as 1991, Rivest identified significant parallels between machine learning and cryptanalysis, referring to them as "sister disciplines" [1].

In recent years, AI technologies, particularly deep learning, have achieved remarkable breakthroughs in both the compromise of cryptographic implementations and the design of novel cryptographic structures. For instance, researchers have successfully leveraged deep neural networks to exploit hardware side-channel leakage, effectively defeating AES implementations fortified with masking and jitter countermeasures [2,3,4]. Furthermore, the incorporation of neural networks into differential cryptanalysis has significantly reduced the complexity of key recovery attacks on an 11-round Speck32/64 instance, reaching the best-known results at the time . Even in cryptographic design, AI-driven models such as Generative Adversarial Networks (GANs) have been utilized to autonomously generate S-boxes with superior cryptographic properties [5].

ISSN: 2710-0588

The growing intersection of AI and cryptography has attracted widespread academic attention. This paper provides a comprehensive survey of recent advancements in AI applications within cryptanalysis and cryptographic design, covering areas such as side-channel analysis, fault injection attacks, statistical and algebraic cryptanalysis, lightweight cryptographic design, post-quantum cryptography, and cryptographic protocol verification. Additionally, we discuss prevailing challenges and potential future directions in the field.

2. AI Applications in Cryptanalysis

2.1. Side-Channel Analysis (SCA)

Side-Channel Analysis (SCA) leverages inadvertent physical emissions—such as electromagnetic radiation, power consumption, and execution timing—from cryptographic devices to deduce secret keys. Traditional SCA methods, notably Differential Power Analysis (DPA) and Template Attacks, necessitate manual feature extraction and statistical modeling processes. However, deep learning techniques have significantly transformed this field. Spectrum-based deep learning methods eliminate the reliance on Gaussian distribution assumptions and facilitate end-to-end modeling of misaligned power traces [6]. Compared to traditional template attacks, deep learning models demonstrate enhanced resistance to noise and timing jitter, negating extensive preprocessing steps and outperforming classical techniques against common implementation countermeasures, such as random delays and masking [5].

For instance, Cagli et al. introduced Convolutional Neural Networks (CNNs) combined with data augmentation, successfully breaching AES hardware implementations secured by random jitter, thereby achieving unprecedented attack efficiency compared to traditional methods [6]. Similarly, Maghrebi et al. showcased that Multi-Layer Perceptron (MLP) models could compromise first-order masked AES implementations using significantly fewer power traces than required by conventional methods [6].

Figure 1 compares the attack success rates between deep learning-based techniques and traditional approaches, clearly demonstrating that deep learning consistently achieves superior success rates under equivalent sampling conditions. Although optimally configured template attacks may occasionally approach deep learning performance in strictly controlled scenarios, deep learning generally excels in practical, complex environments affected by trace misalignment and timing jitter [6]. Consequently, AI-based side-channel analysis has become indispensable for evaluating hardware security.

2.2. Differential Fault Analysis (DFA)

Differential Fault Analysis (DFA) capitalizes on computational faults intentionally induced during cryptographic operations, comparing erroneous ciphertexts with correct ones to derive secret keys. Classical DFA techniques typically require precise fault models and multiple injections to achieve successful key recovery. Recent integration of deep learning into fault analysis substantially reduces reliance on prior fault knowledge and manual analysis [7].Cheng et al. proposed a Deep Learning-based Fault Analysis (DLFA) framework, wherein neural networks autonomously extract key-related information from faulty ciphertext datasets [7]. Empirical evaluations illustrate that, in AES scenarios, DLFA requires only 1,488 faulty ciphertexts and a single fault injection, achieving complete key recovery in an average computational duration of merely 0.12 seconds. This signifies a marked advancement in efficiency and data utilization compared to traditional Statistical Fault Analysis (SFA), which typically demands hundreds of fault injections and nearly an hour for equivalent key recovery success [7].Further research has employed machine learning methodologies to enhance fault localization accuracy and fault model inference, significantly automating DFA attacks [20]. As

ISSN: 2710-0588

shown in Figure 2, deep learning-based DFA methods consistently achieve higher key recovery success rates with fewer fault injections relative to traditional DFA techniques, which exhibit markedly slower progression toward successful outcomes.

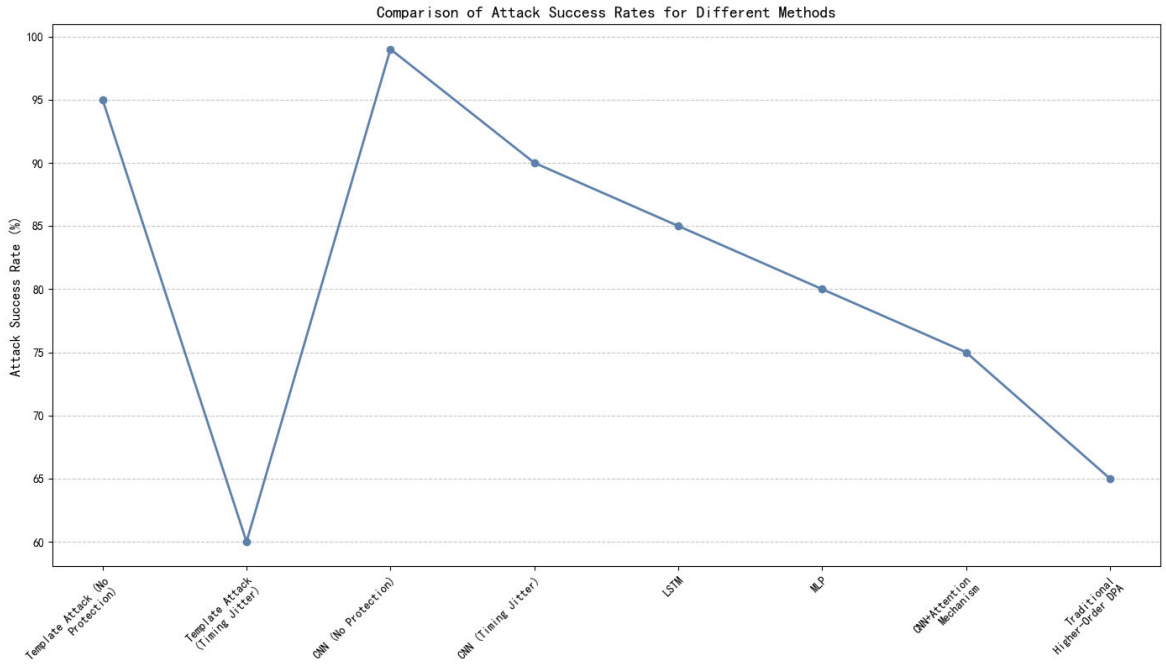

Figure 1. Comparison of Attack Success Rates for Different Methods

Overall, incorporating AI transforms DFA from a knowledge-intensive heuristic technique into a data-centric pattern recognition task, notably elevating both efficiency and attack success rates. This evolution presents enhanced security challenges for embedded device manufacturers, necessitating more robust defensive measures against fault-based attacks.

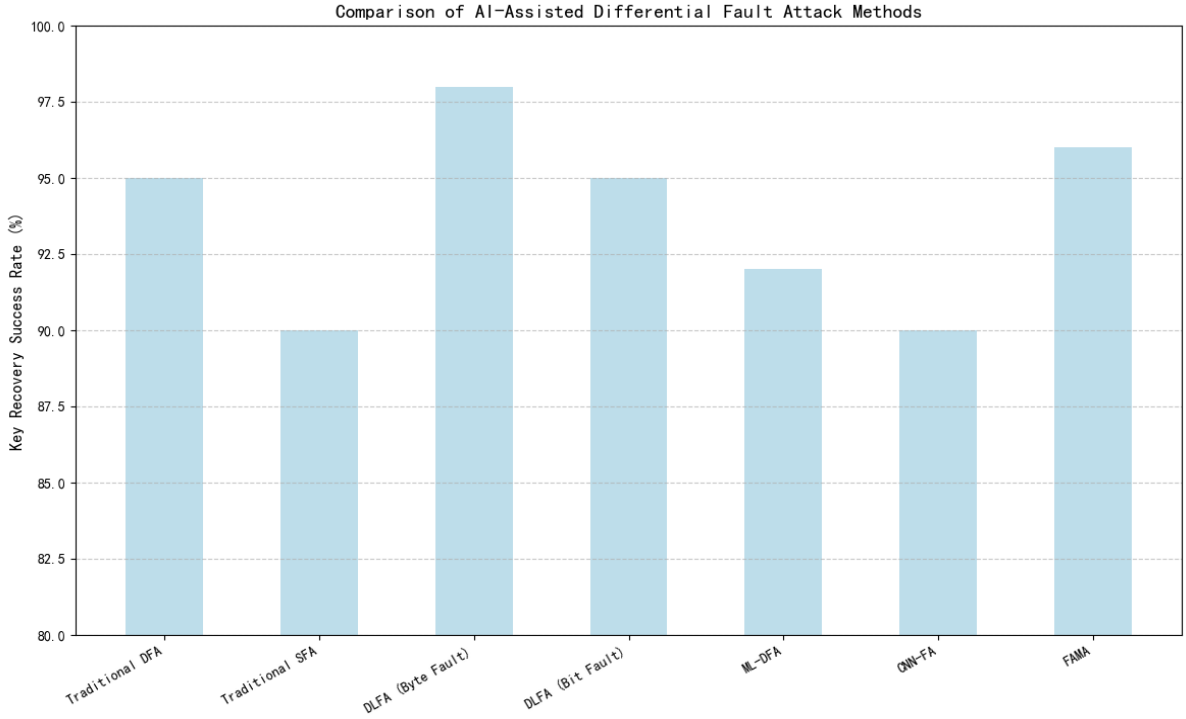

Figure 2. Comparison of AI-Assisted Differential Fault Attack Methods

ISSN: 2710-0588

2.3. Statistical and Algebraic Cryptanalysis

Statistical vulnerabilities have long underpinned classical cryptanalysis approaches such as Differential Cryptanalysis (DC), Linear Cryptanalysis (LC), and Algebraic Attacks. Recent research has increasingly explored AI-driven methodologies to automate the discovery of intricate statistical patterns, thus enhancing attack effectiveness.

A significant recent development is Neural Differential Cryptanalysis. In 2019, Gohr introduced a neural network-enhanced differential cryptanalysis framework, successfully launching effective attacks on the Speck32/64 cipher. His neural network model discriminated ciphertext pairs possessing specific input differentials from random pairs, enabling the attack to cover 11 rounds—surpassing previous manually derived results [11]. This groundbreaking work significantly stimulated further exploration of neural-enhanced cryptanalysis methods across diverse cryptographic primitives and architectures [11]. AI models can autonomously discern relationships between plaintext-ciphertext differentials and underlying key structures, eliminating manual crafting of differential characteristics. Similarly, linear cryptanalysis has benefited from AI techniques capable of identifying optimal linear approximations and effectively estimating key likelihood distributions, expediting linear attack convergence.

In algebraic cryptanalysis, AI primarily assists in simplifying multivariate polynomial equation systems. For instance, machine learning approaches predict key-dependent variables most likely conducive to solution convergence, while reinforcement learning strategies guide SAT solvers to expedite analysis of reduced-round block ciphers [10].

Nevertheless, AI has not yet reached comparable success in algebraic cryptanalysis due to inherent computational complexities associated with large Boolean equation systems, limiting AI models' capabilities in approximating viable solution spaces. Still, preliminary studies indicate potential improvements; for example, neural networks have successfully identified subtle structural weaknesses in reduced-round DES ciphertext distributions, previously undetectable by conventional differential methodologies [10].

Overall, AI applications in differential, linear, and algebraic cryptanalysis remain nascent yet promising. AI's automated pattern-recognition abilities significantly augment traditional analytical methods, providing new insights into cryptographic vulnerabilities.

3. AI Applications in Cryptographic Design

3.1. Optimization of Lightweight Cryptography

Lightweight cryptography addresses the challenge of providing robust security solutions tailored specifically for resource-constrained devices, which demands a careful balance between security, performance, and resource utilization. Recently, artificial intelligence methods have played an increasingly prominent role in optimizing essential cryptographic components, notably S-boxes and nonlinear transformation elements. Historically, S-boxes were designed either by manual mathematical construction or exhaustive search, methods challenged by the vast search spaces (approximately 2168} possibilities for 8-bit S-boxes) and multi-dimensional optimization criteria [12]. Intelligent search algorithms, including genetic algorithms, hill climbing, neural networks, and cellular automata, have proven their global search efficiency, discovering S-box configurations with enhanced nonlinearity and improved differential uniformity.

Further advancing this field, Zhang et al. employed Generative Adversarial Networks (GANs) to automatically derive cryptographic primitives with optimal characteristics. Specifically, they proposed integrating cryptographically tailored loss functions, such as differential uniformity and nonlinearity losses, into the GAN training process, resulting in S-boxes that simultaneously optimize multiple cryptographic metrics [12]. Experimental evaluations demonstrated

ISSN: 2710-0588

promising outcomes, where the AI-generated S-boxes attained minimum differential uniformity of 8 and maximum nonlinearity of 104. While marginally lower than the AES standard (differential uniformity: 4, nonlinearity: 112), these outcomes notably surpassed randomly generated counterparts [12].

Further comparative analysis indicated the improved WGAN-GP model, combined with the customized WGP-IM loss function, achieved significant advantages in differential uniformity, differential probability, and boomerang uniformity metrics compared to traditional approaches and alternative GAN models [12]. For instance, the WGP-IM-generated S-boxes exhibited the lowest differential uniformity (achieving values as low as 8), enhancing resilience against differential cryptanalysis, with the maximum differential probability as low as 0.0313 [12].

Figure 3 illustrates the progressive optimization of S-box nonlinearity across different GAN training configurations, demonstrating rapid convergence toward nonlinearity scores exceeding 100 with adjusted loss weights (e.g., L3 configuration) [12].

Ultimately, the increasing integration of intelligent algorithms within lightweight cryptographic design is driving the automated discovery of highly secure yet resource-efficient cryptographic components.

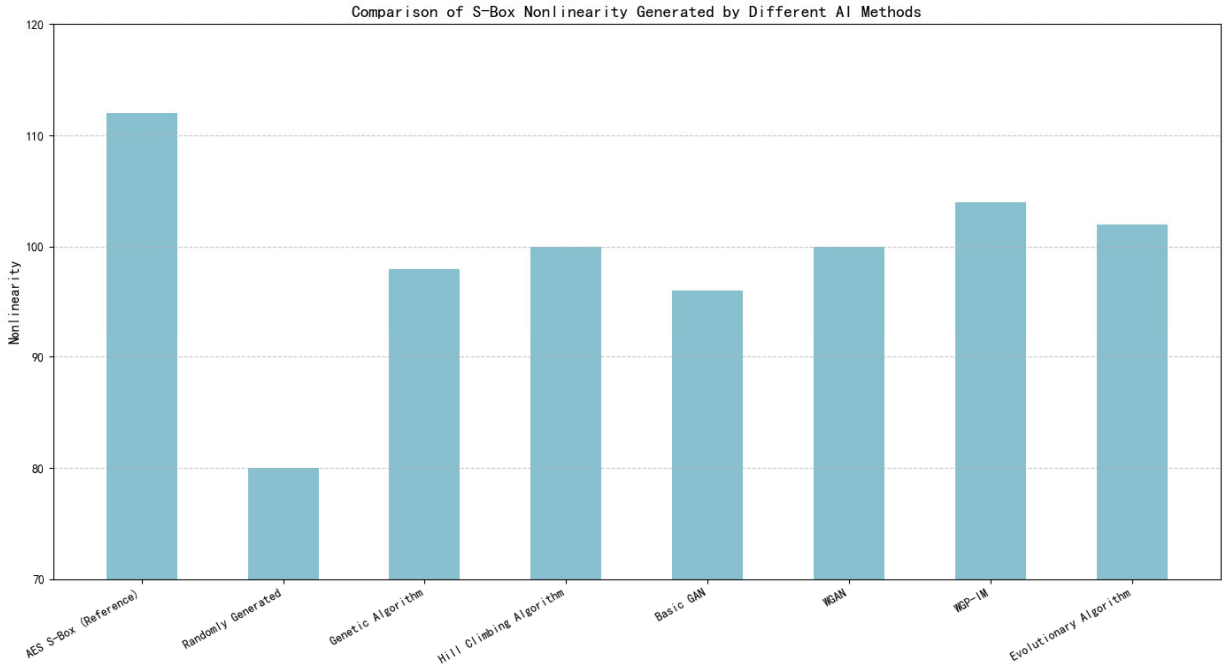

Figure 3. Comparison of S-Box Nonlinearity Generated by Different AI Methods

3.2. Applications in Post-Quantum Cryptography (PQC)

Post-quantum cryptography (PQC), relying on computationally intensive mathematical problems such as Learning With Errors (LWE), Short Integer Solution (SIS), and code-based cryptography, is designed to withstand quantum computing threats. These computational complexities provide fertile ground for AI-driven analysis.

AI techniques have notably impacted security evaluations of PQC schemes. For instance, Lauter et al. leveraged deep learning to exploit LWE instances employing sparse binary secret keys intended for computational efficiency. Initial attacks, such as “Salsa,” partially recovered secret keys from LWE instances (dimension $n \leq 128$, Hamming weight $h \leq 4$) using millions of ciphertext samples [15]. The subsequent “Picante” attack further scaled this approach to higher dimensions (up to $n=350$), successfully recovering most secret bits even at a relatively high sparsity ($h \approx n/10$), outperforming conventional lattice-based solvers in these specific scenarios

ISSN: 2710-0588

[16]. Nonetheless, such AI-based methods currently pose minimal immediate threats to standardized PQC schemes but highlight potential structural vulnerabilities in non-standardized implementations [14]. Additionally, neural networks have been explored in code-based cryptanalysis, such as syndrome decoding acceleration in the McEliece cryptosystem, and reinforcement learning approaches have been utilized for fine-tuning cryptographic parameters, optimizing performance-security trade-offs. While no AI-driven attacks have fully compromised standardized PQC algorithms to date, early research offers valuable insights into potential vulnerabilities and strengthens parameter selection strategies. Looking forward, AI is anticipated to assume a dual role—enhancing PQC security while concurrently presenting novel threats—necessitating sustained vigilance and research.

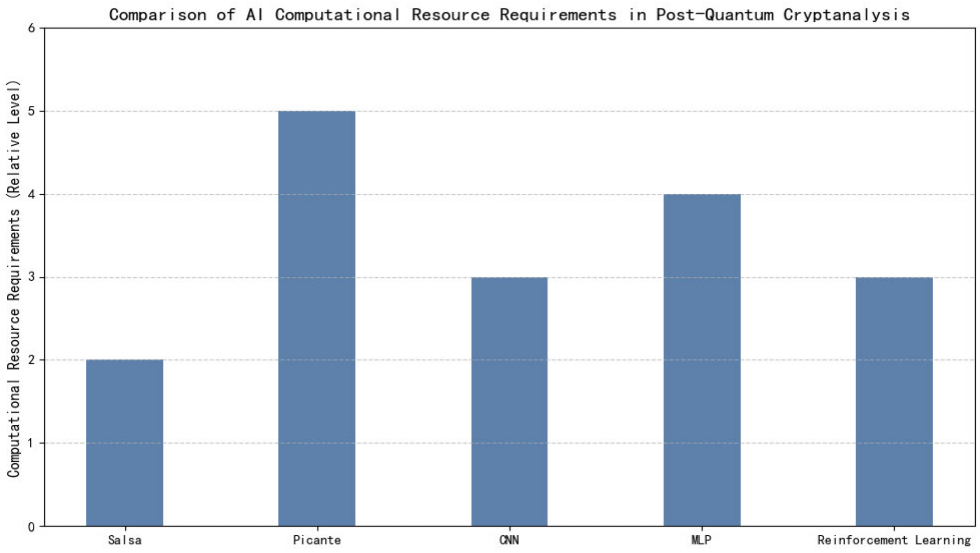

Figure 4. Comparison of AI Computational Resource Requirements in Post-Quantum Cryptanalysis

3.3. Cryptographic Protocol Verification

Verifying the security of complex cryptographic protocols (e.g., SSL/TLS, blockchain consensus protocols) is an inherently challenging task due to state-space explosion and inherent undecidability, often encountered by formal verification tools like ProVerif or Tamarin. Consequently, this process typically demands extensive computational resources and expert oversight.

Recent advances explore AI-assisted protocol verification methodologies. Ohno and Nakabayashi developed a deep learning-based framework that automatically classified cryptographic protocol security. They generated extensive random protocol datasets labeled as secure or insecure via traditional formal verification tools, subsequently training neural networks capable of quickly predicting the security status of unseen protocols in linear computational time [15]. This approach demonstrated considerable efficiency and successfully identified known vulnerabilities within complex systems, including SSH and electronic voting protocols. Beyond verification, AI has facilitated the analysis of cryptographic protocols via automated fuzz testing, where reinforcement learning efficiently uncovers vulnerabilities by exploring protocol state spaces. Additionally, large language models (LLMs) have been employed effectively in smart contract auditing, rapidly identifying common security flaws and risks. Moreover, AI has contributed to the optimization of Zero-Knowledge Proof (ZKP)

ISSN: 2710-0588

protocols, aiding selection of optimal proving parameters and predicting bottlenecks within ZK circuits to streamline computational overhead.

These advancements underscore AI's growing integration into protocol verification and cryptographic design, significantly enhancing both security assurance and development efficiency. As protocols continue to grow in complexity, AI-driven methodologies are expected to become increasingly crucial for ensuring their trustworthiness.

4. Challenges and Future Directions

4.1. Data Dependency and Generalization Ability

The efficacy of deep learning models critically depends on large-scale, high-quality training datasets, which poses substantial challenges within cryptographic contexts. Collecting sufficient data for cryptanalysis can be exceedingly demanding and resource-intensive. For example, robust training of side-channel attack models typically requires capturing hundreds of thousands to millions of power traces [17]. Google researchers noted that even extensive datasets—such as a masked AES dataset comprising 65,000 compressed power traces totaling 7GB—may remain inadequate for training reliable models against advanced security countermeasures, requiring datasets in excess of 1TB [17]. Such extensive data requirements present practical difficulties concerning data storage, labeling accuracy, and computational overhead. Moreover, many deep learning models assume attackers possess unrestricted access to cryptographic devices to gather extensive samples, a scenario often unrealistic in passive or limited-access attack environments.

Additionally, existing datasets lacking diversity frequently lead to model overfitting, thereby severely impairing their effectiveness when confronted with variations in device-specific noise patterns or cryptographic implementations. Addressing these issues necessitates future exploration of specialized machine learning techniques such as few-shot learning, domain adaptation, and transfer learning, tailored specifically to cryptographic contexts to mitigate heavy data-dependency challenges.

4.2. Explainability and Security Risks

AI models, particularly deep neural networks, often function as "black boxes," which introduces two critical issues within cryptographic applications:

Interpretability in Cryptanalysis: When a neural network model, such as those used in neural differential cryptanalysis, identifies correct key guesses, it is frequently unclear which cryptographic weaknesses the model exploits. This lack of interpretability restricts cryptographers' ability to identify underlying vulnerabilities and thus hinders targeted improvements in cryptographic algorithms. Furthermore, it increases the risk of false positives, as AI models may inadvertently rely on coincidental statistical correlations unrelated to actual cryptographic weaknesses. Recent efforts to visualize and analyze intermediate neural network layers offer preliminary insights into model behavior, but substantial advancements in interpretability remain urgently needed [10].

AI systems themselves can become security vulnerabilities if subjected to adversarial attacks or data poisoning during training phases. If attackers introduce carefully engineered malicious data into training datasets, AI models could produce intentionally misleading outcomes, potentially compromising the integrity of AI-driven cryptographic designs. The growing adoption of AI further expands potential attack surfaces, as adversaries might attempt to reverse-engineer AI-based cryptographic components to uncover sensitive embedded data or hidden vulnerabilities. Therefore, ensuring the integrity and robustness of AI training procedures is crucial. Future research must emphasize the development of verifiable and

ISSN: 2710-0588

secure AI methodologies to bolster confidence in AI-assisted cryptographic assessments and designs.

4.3. Integration with Emerging Technologies

The intersection of AI with emerging technologies presents both compelling opportunities and considerable challenges for cryptographic research. For example, quantum computing poses existential threats to classical cryptographic schemes. This raises critical questions: Could AI techniques mitigate threats posed by quantum cryptanalysis? Conversely, can quantum-enhanced machine learning introduce fundamentally new cryptanalytic methods? Current research remains exploratory, as existing AI-driven attacks are inherently classical and may not directly translate to quantum contexts, warranting further exploration.

Similarly, integrating AI with zero-knowledge proofs (ZKPs) has led to the emerging field of Zero-Knowledge Machine Learning (ZKML). ZKML techniques aim to verify AI model predictions cryptographically without revealing sensitive information. With AI models increasingly embedded within cryptographic protocols, such as blockchain smart contracts, ensuring the transparency, verifiability, and security of AI-driven decisions is paramount. Zero-knowledge proofs could offer robust solutions for cryptographically validating AI outcomes, thus securely embedding AI into cryptographic infrastructures.

Additionally, AI could significantly optimize post-quantum digital signature schemes by fine-tuning key parameters, signature sizes, and verification efficiency. However, careful attention is necessary to ensure that such optimizations do not inadvertently compromise foundational cryptographic assumptions. Consequently, rigorous evaluation and testing of these integrated methodologies must accompany technological advances.

In summary, the convergence of AI and emerging cryptographic technologies promises significant security and performance breakthroughs. Nonetheless, it also demands comprehensive, rigorous investigation to manage and mitigate potential security risks effectively.

5. Conclusion

Artificial intelligence has profoundly reshaped the landscape of cryptanalysis and cryptographic design. AI-driven methodologies—from automated pattern recognition in side-channel analyses and efficient key recovery in differential fault analyses, to the intelligent optimization of cryptographic primitives and accelerated vulnerability detection in protocol verification—demonstrate unprecedented effectiveness. This survey has systematically explored recent advances in AI-driven cryptography, underscoring how AI has become a dual-edged tool capable of both undermining existing cryptographic implementations and inspiring the development of more secure cryptographic systems. However, AI's reliance on large datasets, limited generalization capabilities, and interpretability constraints highlight essential limitations, emphasizing the necessity for cautious and controlled integration into cryptographic contexts.

Future research should prioritize the following directions:

1. Developing AI models capable of achieving reliable performance under limited data conditions, thus lowering barriers associated with data acquisition.

2. Enhancing the transparency and interpretability of AI-driven cryptanalysis, integrating symbolic reasoning methods to facilitate verifiable AI outcomes.

3. Exploring interdisciplinary integration of AI with quantum computing, zero-knowledge proofs, and other emergent technologies, fostering comprehensive cryptographic security solutions.

ISSN: 2710-0588

4. Establishing standardized benchmarks and open datasets, promoting uniformity and reproducibility within AI-driven cryptographic research.

As the fields of AI and cryptography increasingly converge, they will mutually reinforce each other's theoretical foundations and practical methodologies. This symbiosis promises to drive next-generation security innovations, reinforcing our collective capacity to defend the digital landscape robustly.

Ultimately, careful and considered integration of AI technologies into cryptographic practice will be critical to unlocking their full potential, thus establishing a more resilient and trustworthy cybersecurity infrastructure for the future digital world.

References

[1] R. Rivest: Machine Learning and Cryptanalysis, ASIACRYPT'91 (1991), p. 427–439.

[2] E. Cagli, C. Dumas and E. Prouff: Convolutional Neural Networks with Data Augmentation against Jitter-Based Countermeasures, CHES 2017 (2017), p. 45–68.

[3] H. Maghrebi, T. Portigliatti and E. Prouff: Breaking Cryptographic Implementations Using Deep Learning, SPACE 2016 (2016), p. 3–26.

[4] R. Benadjila, E. Prouff, R. Strullu, E. Cagli and C. Dumas: Deep Learning for Side-Channel Analysis and Introduction to ASCAD Database, Journal of Cryptographic Engineering, Vol. 10 (2020) No. 2, p. 163–188.

[5] L. Wu et al.: Ranking Loss: Maximizing the Success Rate in Deep Learning Side-Channel Analysis, IACR Transactions on Cryptographic Hardware and Embedded Systems, Vol. 2021 (2021) No. 1, p. 25–55.

[6] Y. Zotkin et al.: Deep Learning vs. Template Attacks: Experimental Study, IACR ePrint Archive, Report 2018/1213 (2018).

[7] Y. Cheng et al.: DLFA: Deep Learning-based Fault Analysis against Block Ciphers, IACR ePrint Archive, Report 2023/021 (2023).

[8] A. Heuser et al.: Side-Channel Analysis of Lightweight Ciphers: Does Lightweight Equal Easy?, RFIDSec 2016 (2016), p. 91–104.

[9] J. Kim et al.: Make Some Noise: Unleashing the Power of CNNs for Profiled Side-Channel Analysis, IACR Transactions on Cryptographic Hardware and Embedded Systems, Vol. 2019 (2019) No. 3, p. 148–179.

[10] A. Gohr: Improving Differential Cryptanalysis with Deep Learning, CRYPTO 2019 (2019), p. 3–24.

[11] D. Gerault et al.: SoK: Five Years of Neural Differential Cryptanalysis, IACR ePrint Archive, Report 2024/1300 (2024).

[12] R. Zhang et al.: A Novel S-Box Generation Methodology Based on the Optimized GAN Model, Computers, Materials & Continua, Vol. 76 (2023) No. 2, p. 1911–1927.

[13] L. Lerman et al.: A Machine Learning Approach Against a Masked AES, CARDIS 2013 (2013), p. 61–75.

[14] Y. Li et al.: Salsa Picante: A Machine Learning Attack on LWE with Binary Secrets, Proceedings of the 2023 ACM CCS (2023), p. 112–125.

[15] K. Ohno and M. Nakabayashi: A Security Verification Framework of Cryptographic Protocols Using Machine Learning, arXiv preprint arXiv:2304.13249 (2023).

[16] Y. Liu et al.: Machine Learning Assisted Differential Cryptanalysis, IEEE Access, Vol. 7 (2019), p. 76547–76556.

[17] E. Bursztein: Hacker's Guide to Deep Learning Side-Channel Attacks (Part 1) (2021).

[18] S. Picek et al.: When Theory Meets Practice: Profiled Side-Channel Analysis with Hyperparameter Tuning, IACR Transactions on Cryptographic Hardware and Embedded Systems, Vol. 2021 (2021) No. 3, p. 677–707.

ISSN: 2710-0588

[19] J. Zhang et al.: Fault Template Attacks on AES and Their Countermeasures, IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Vol. 28 (2020) No. 12, p. 2638–2650.

[20] A. Saha et al.: Improved Fault Analysis on LED Block Cipher Using DFA and Machine Learning, Security and Communication Networks, Vol. 2018 (2018), p. 8165294.