Proceedings of CONF-MLA 2024 Workshop: Securing the Future: Empowering Cyber Defense with Machine Learning and Deep Learning
DOI: 10.54254/2755-2721/110/2024MELB0088

Advancements and Prospects in Large Integer Factorization: A Comprehensive Review of the Number Field Sieve Method

Jiawei Gao

Computer science and technology, Macau University of Science and Technology, Macau, China

1220011984@student.must.edu.mo

Abstract. The Number Field Sieve (NFS) stands as one of the most effective algorithms for the factorization of large integers, playing a critical role in cryptographic security, particularly in breaking RSA encryption systems. This review paper provides an in-depth analysis of NFS, highlighting its structure, key processes, and computational challenges. NFS is notably effective for integers larger than 110 digits and is recognized for its sub-exponential computational complexity, making it superior to other classical factorization algorithms. The core processes of NFS—polynomial selection, number pair sieving, matrix solving, and square root extraction—are systematically examined to illustrate their roles in factorization. This study also reviews the latest advancements in NFS, including improvements in polynomial selection methods and optimizations in sieving and matrix-solving stages, which have significantly enhanced the algorithm's efficiency. Moreover, the paper discusses the future prospects of NFS, emphasizing the need for further optimization to reduce computational resource demands and increase practicality in large-scale applications. The ongoing evolution of NFS continues to push the boundaries of cryptographic analysis, driving future research towards even more efficient factorization methods.

Keywords: Number Field Sieve, RSA, algorithm.

1. Introduction

The secure transmission and storage of data are paramount in today's digital age, as society becomes increasingly reliant on information technology. Cryptography serves as the backbone of data security, safeguarding sensitive information across numerous applications. A fundamental challenge in cryptography is the factorization of large integers, which underpins the security of widely used encryption systems, such as RSA. Traditional factorization methods, including trial division and elliptic curve algorithms, have proven inefficient for extremely large numbers. However, with the continuous improvement of computational power and algorithmic sophistication, the demand for more effective factorization techniques has grown significantly [1, 2].

Among various algorithms, the Number Field Sieve (NFS) stands out as one of the most advanced and efficient methods for factorizing large integers, especially those exceeding 110 digits. Its sub-exponential computational complexity makes NFS particularly powerful in cryptographic contexts, such as breaking RSA encryption. The NFS method has been successfully employed in notable achievements, including the factorization of RSA-768, which required approximately 2000 core-years of

DOI: 10.54254/2755-2721/110/2024MELB0088

computational effort, highlighting its capability in tackling large-scale factorization problems [3]. The key processes within NFS—polynomial selection, sieving, matrix solving, and square root extraction—each contribute significantly to its overall efficiency. Recent advancements have focused on enhancing these processes, particularly through parallel computing techniques that reduce computational time and resource demands.

This paper provides a comprehensive review of the NFS algorithm, detailing its foundational principles, key components, and recent research developments. The study systematically examines the four main stages of NFS, highlighting both technical challenges and solutions that have emerged in recent years. By exploring advancements in polynomial selection, lattice sieving techniques, and matrix-solving optimizations, the paper aims to shed light on how NFS has evolved and continues to set new benchmarks in integer factorization. Additionally, the paper discusses future directions for improving the algorithm, focusing on reducing computational complexity and enhancing its practicality for cryptographic applications. This analysis not only underscores the ongoing relevance of NFS but also identifies critical areas for further research in large integer factorization [4-7].

2. Relevant techniques

2.1. RSA

Encryption: Given the public key (e, N) and plaintext M, ciphertext C is calculated by the following formula:

$$C=E(M)=Me \bmod N \tag{1}$$

$$M=D(C)=Cd \bmod N \tag{2}$$

2.2. Advantages of the number field sieve

Large integer factorization is the process of factorizing a large integer into its prime factors. This has important applications in cryptography, especially in some encryption algorithms such as RSA. At present, large number decomposition has completely separated from the era of violent trial division, there are several common large number decomposition methods, such as trial division, ρ method, P-1 method, elliptic curve algorithm, random square method, quadratic sieve method and number field sieve method [4]. In addition, there are quantum algorithms that rely on quantum computers, which can greatly improve the efficiency of decomposition, but quantum computers are still under development. Zeng Yonghong, Liu Xinxing et al. and Yang Lili have made a very good review of large integer decomposition, pointing out that for integers greater than 110 bits, number field screening is the fastest algorithm at present, and its time complexity is as follows [5-7]:

$$0(\exp((\frac{64}{9}+0(1)^{1/3}(logN)^{1/3}(\log logN)^{2/3})) \tag{3}$$

The NFS is highly parallelizable, meaning that large-scale computations can be distributed across multiple machines. This feature makes it ideal for use in grid computing and cloud computing environments, where many processors can work together to factor very large numbers efficiently. By solving this congruence, we get the nontrivial factors of N, gcd (N, x + y) and gcd(N, x-y) giving the nontrivial factorization of N with a probability of 2/3. This process is particularly complex and requires a series of mathematical concepts and mappings to finally find the combination of x and y square numbers. The detailed process can be seen in references [8, 9]. These complex theories will eventually solidify into programs. From the point of view of computer science, the number field screening method is mainly divided into four parts, which are polynomial selection, number pair screening, matrix solving, and square root finding, as shown in Figure 1.

DOI: 10.54254/2755-2721/110/2024MELB0088

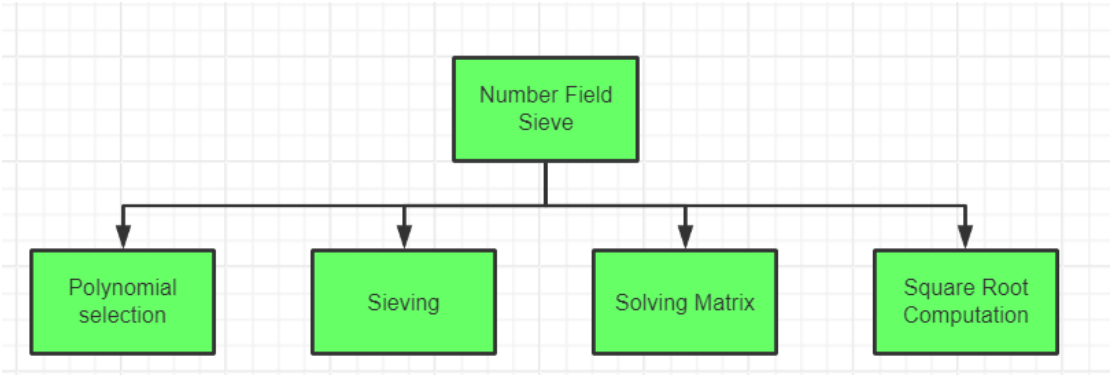

Figure 1. The main process of NFS (Photo credit: Original).

3. Research progress

The special number field sieve method deals specifically with integers that have a particular form or structure. The number field sieve is not an independent algorithm, but a collection of algorithms. The number field sieve method includes the four main calculation processes in the previous chapter, and each process contains one or more dependent algorithms [10].

Polynomial selection. Polynomial selection is the first and key step of the number field screening method, which makes the subsequent number pair screening more efficient. Polynomial selection is mainly the selection of two mutual-prime polynomials f(x), g(x), and the two polynomials have the same root in the module n. An m-based approach is usually used. Namely.

$$\mathbf{n} = \sum_{\mathbf{i=0}}^{\mathbf{d}} \mathbf{c_i}\, \mathbf{m}\mathrm{i}\ ,0 \leqslant \boldsymbol{c_i} < \mathrm{m} \tag{4}$$

Get further.

$$\mathrm{f(x)}=\sum_{\boldsymbol{i=0}}^{\boldsymbol{d}} \boldsymbol{c_i}\ \mathrm{xi,g(x)=x-m} \tag{5}$$

Finally.

$$\mathrm{f(m)} \equiv 0(\mathrm{mod\ n}),\ \mathrm{g(m)} \equiv 0(\mathrm{mod\ n}) \tag{6}$$

f(m) and g(m) modules n have the same root m. In 1999, Murthy conducted a systematic study on polynomial selection in his doctoral thesis, and believed that good polynomials could not only generate more number pairs, but also reduce the size of the matrix [11]. Based on the base-M method, the techniques of transformation and rotation are introduced to generate skew and non-skew polynomials. The combination of nonlinear polynomial and non-first linear polynomial is a commonly used technique at present, and nonlinear polynomial is also developing [12, 13]. Prest et al. have extended the nonlinear polynomial method and increased the degree of polynomial generated by the original algorithm, but the practical application is limited at present. Yang et al. analyzed polynomial roots and found that polynomials with more small primes in their tail coefficients could generate more number pairs, and its engineering practicability needs to be further confirmed [14].

It is mainly to decompose the expression of logarithmic pairs (a, b) on the factor basis to find smooth numbers that can be decomposed into the product of small prime numbers. For each a and b, calculate two values [15]:

$$\mathrm{F(a,b)=bd \cdot f}\left(\frac{\boldsymbol{a}}{\boldsymbol{b}}\right),\mathrm{G(a,b)=b \cdot g}\left(\frac{\boldsymbol{a}}{\boldsymbol{b}}\right) \tag{7}$$

$$\mathrm{F(a,b)}=\prod_{\boldsymbol{i}} \boldsymbol{p_i^{e_i}}\ ,\ \mathrm{G(a,b)}=\prod_{\boldsymbol{j}} \boldsymbol{q_j^{f_j}} \tag{8}$$

Where pi and qj are small prime numbers.

The line sieve is an earlier method proposed within the Number Field Sieve . It works by performing sieving in a fixed linear space to find integers that satisfy specific conditions. While the line sieve is relatively simple, its efficiency is low when dealing with very large integers. The lattice sieve is a more modern method in the NFS, which increases efficiency by performing sieving on a lattice structure. The lattice sieve is more effective in handling large integers and shows significant advantages in practical applications. In 1994, Golliver et al. introduced the operation of trial division to the lattice sieve, further

DOI: 10.54254/2755-2721/110/2024MELB0088

improving its efficiency [16]. In 2005, Franke et al. proposed using continued fractions to enhance the sieving process in the lattice sieve [17]. With the development of computer technology, L.T. Yang et al. and Gad et al. continuously optimized the sieving algorithms in the NFS and significantly accelerated the sieving process by utilizing parallel computing techniques [18-22].

Matrix Solving. Matrix solving refers to the process of constructing linear relations from the smooth numbers obtained during sieving and finding quadratic congruences under modulus N. Initially, Gaussian elimination was used, but the Block Wiedemann algorithm or Block Lanczos algorithm are now predominantly used to solve linear algebra systems over binary fields to obtain several perfect squares, as shown in Figure 2. Yang et al. designed methods to reduce synchronization and communication for parallel Block Lanczos algorithms, obtaining approximately 7 times speedup on 16 cores and 11 times speedup on 30 cores [23-25]. Zhou proposed a custom parallel architecture for the Block Wiedemann algorithm, which can be categorized under Field Programmable Gate Array (FPGA) acceleration [26].

Solving for Square Roots. Solving for square roots refers to the process of obtaining the final factorization result based on the perfect squares obtained from the matrix solving phase. Lenstra et al. introduced the steps for solving square roots in 1993, and several methods have since emerged. Thome reviewed square root-solving methods, which primarily include the direct method, Couveignes' algorithm, and Montgomery's algorithm, as shown in Figure 3.

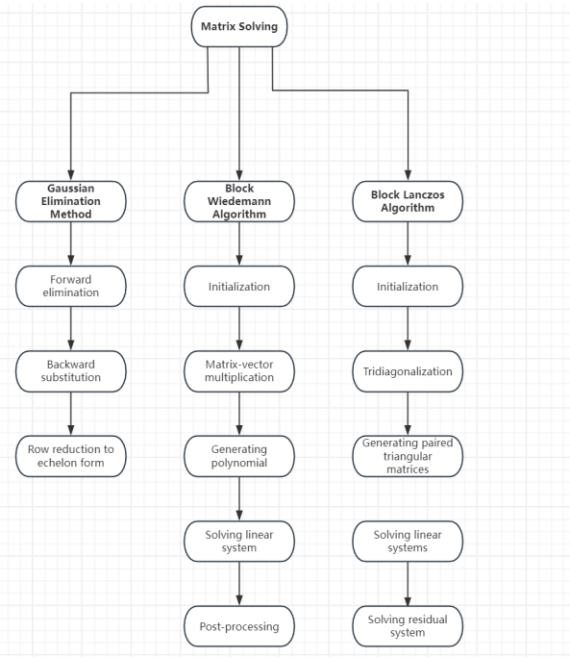

Figure 2. Three common methods of matrix solving (Photo credit: Original).

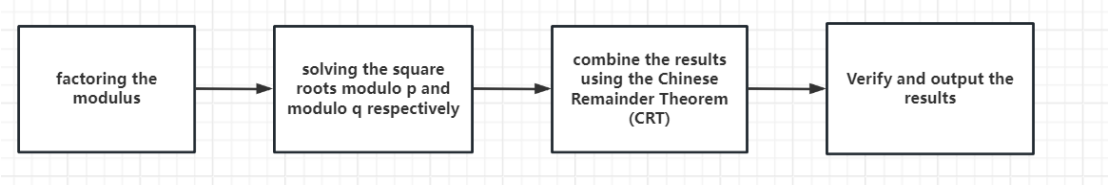

Figure 3. A new algorithm idea based on CRT (Photo credit: Original).

Since the Number Field Sieve (NFS) was proposed in 1988, it has continuously broken factorization records, setting one milestone after another in the factorization of large numbers [27]. From the

DOI: 10.54254/2755-2721/110/2024MELB0088

successful factorization of RSA-100 in 1991 to the factorization of RSA-768 in 2009, NFS has demonstrated itself as one of the best algorithms for large integer factorization to date [28].

4. Future research directions and prospects

High Algorithmic Complexity: Will Simpler Algorithms Emerge? Currently, the NFS has the lowest complexity among large integer factorization algorithms, but it is also highly complex, involving multiple steps and a great deal of mathematical theory. Each step requires specialized knowledge and sophisticated algorithmic implementation. Whether there will be a more efficient algorithm with lower complexity is still an open question.

High Computational Resource Requirements and Large Memory Consumption. The NFS demands substantial computational resources, especially when factoring larger integers. Both the sieving and matrix-solving phases require significant computational power and memory. The high computational cost makes the NFS impractical for extremely large integers.

Inefficiency of the Sieving Stage. The relation collection (sieving) stage is the most resource-intensive part of the NFS, as it requires the collection of a large number of smooth numbers. Although this process can be parallelized, it still consumes a vast amount of computational time and resources, accounting for 90% of the total computation. Currently, there is limited research focused on improving the efficiency of the sieving process.

These issues highlight areas for potential improvement and further research to make the NFS more feasible and efficient for large-scale integer factorization tasks.

5. Conclusion

This paper provides a comprehensive review of the Number Field Sieve (NFS) algorithm, highlighting its critical role in the factorization of large integers, particularly within cryptographic applications like RSA encryption. The discussion encompasses the main components of NFS, including polynomial selection, number pair sieving, matrix solving, and square root extraction, each of which contributes to the overall efficiency of the algorithm. Recent advancements in NFS, such as improvements in polynomial selection techniques and optimizations in sieving and matrix-solving stages, have significantly enhanced the algorithm's performance. The review also emphasizes the importance of NFS in cryptography, showcasing its effectiveness in breaking large integers and its superiority over other classical factorization methods. Looking ahead, the future research directions for NFS are centered on addressing its current limitations. One major challenge is the high computational resource requirement, especially in the sieving and matrix-solving stages, which account for the majority of the computational effort. Reducing the complexity and resource demands of these stages through innovative algorithms and enhanced parallel computing techniques is crucial for making NFS more practical for large-scale applications. Additionally, exploring new methods to simplify the overall structure of NFS could lead to the development of even more efficient algorithms with lower complexity. As computational capabilities continue to advance, further optimization and adaptation of NFS will play a vital role in the ongoing evolution of cryptographic security, driving the need for continuous research in large integer factorization methods.

References

[1] Kudelić, R. (2024). On the Theory of Quantum and Towards Practical Computation. arXiv preprint arXiv:2403.09682.

[2] Cheng, C., Xu, F., Pan, X. F., Wang, C., Fan, J., Yang, Y., ... & Zheng, J. S. Genetic Mapping of Serum Metabolome to Chronic Diseases Among Han Chinese.

[3] Wang, Z., Li, Z., Ran, B., Liu, S., Wu, W., Ye, Y., ... & Li, J. (2023). Mid-cretaceous rapid denudation of Eastern Tibetan plateau: Insights from detrital records at the Southwestern corner of Sichuan basin. Frontiers in Earth Science, 11, 1113377.

DOI: 10.54254/2755-2721/110/2024MELB0088

[4] Dong, P., Wang, L., Chen, Y., Wang, L., Liang, W., Wang, H., ... & Guo, F. (2024). Germplasm Resources and Genetic Breeding of Huang-Qi (Astragali Radix): A Systematic Review. Biology, 13(8).

[5] Ahmed, N. (2024). Quantum Computing Algorithms for Integer Factorization: A Comparative Analysis. Modern Dynamics: Mathematical Progressions, 1(1), 6-9.

[6] Zhu, X., Zhang, Y., Zhao, Z., & Zuo, J. (2019). Radio frequency sensing based environmental monitoring technology. In Proceedings of the Fourth International Workshop on Pattern Recognition (pp. 187-191). Springer.

[7] Qureshi, F., Yusuf, M., Ahmed, S., Haq, M., Alraih, A. M., Hidouri, T., ... & Ibrahim, H. (2024). Advancements in sorption-based materials for hydrogen storage and utilization: A comprehensive review. Energy, 132855.

[8] Zhang, Y., Xu, H., & Zhu, X. (2020). Detection and quantization technique of optical distributed acoustic coupling based on φ-OTDR. Journal of Shanghai Jiaotong University (Science), 25, 208-213.

[9] Ho, H. T., Nguyen, L. V., Le, T. H. T., & Lee, O. J. (2024). Face Detection Using Eigenfaces: A Comprehensive Review. IEEE Access.

[10] Zhang, Y., Zhao, H., Zhu, X., Zhao, Z., & Zuo, J. (2019). Strain measurement quantization technology based on DAS system. In 2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC) (pp. 214-218). IEEE.

[11] Letchumanan, I., Yunus, R. M., Mastar, M. S., & Karim, N. A. (2024). Advancements in electrocatalyst architecture for enhanced oxygen reduction reaction in anion exchange membrane fuel cells: A comprehensive review. International Journal of Hydrogen Energy.

[12] Zhao, Z., Peng, Y., Zhu, X., Wei, X., & Wang, X. (2020). Research on prediction of electricity consumption in smart parks based on multiple linear regression. In 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC) (pp. 812-816). IEEE.

[13] Kumar, S. L., Sureka, P., Sowmitha, A., Sentisenla, J., & Swathy, M. (2024). Recent advancements of hydroxyapatite and polyethylene glycol (PEG) composites for tissue engineering applications–A comprehensive review. European Polymer Journal, 113226.

[14] Xu, H., Zhu, X., Zhao, Z., Wei, X., & Wang, X. (2020). Research of pipeline leak detection technology and application prospect of petrochemical wharf. In 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC) (pp. 263-271). IEEE.

[15] Guo, R., Wang, F., Rhamdhani, M. A., Xu, Y., & Shen, W. (2024). Managing the surge: a comprehensive review of the entire disposal framework for retired lithium-ion batteries from electric vehicles. Journal of Energy Chemistry.

[16] Zhu, X., Zhao, Z., Wei, X., & Wang, X. (2021). Action recognition method based on wavelet transform and neural network in wireless network. In Proceedings of the 2021 5th International Conference on Digital Signal Processing (pp. 60-65). IEEE.

[17] Wang, R., Zhu, J., Wang, S., Wang, T., Huang, J., & Zhu, X. (2024). Multi-modal emotion recognition using tensor decomposition fusion and self-supervised multi-tasking. International Journal of Multimedia Information Retrieval, 13(4), 39.Gad I and Daoud S 2014 Int. J. Adv. Comput. Sci. Appl. 5(7) 178-185.

[18] Raza, S., Hayat, A., Bashir, T., Chen, C., Shen, L., Orooji, Y., & Lin, H. (2024). Electrochemistry of 2D-materials for the remediation of environmental pollutants and alternative energy storage/conversion materials and devices, a comprehensive review. Sustainable Materials and Technologies, e00963.

[19] Ji, W., Huang, X., Wang, S., & He, X. (2023). A Comprehensive Review of the Research of the "Eye–Brain–Hand" Harvesting System in Smart Agriculture. Agronomy, 13(9), 2237.

DOI: 10.54254/2755-2721/110/2024MELB0088

[20] Zheng, Z., Zhou, J., & Zhu, Y. (2024). Computational approach inspired advancements of solid-state electrolytes for lithium secondary batteries: from first-principles to machine learning. Chemical Society Reviews.

[21] Kong, F., & Chen, W. (2024). Carbon Dioxide Capture and Conversion Using Metal–Organic Framework (MOF) Materials: A Comprehensive Review. Nanomaterials, 14(16).Thomé E 2012 Proc. of the 2012 Int. Workshop on the Arithmetic of Finite Fields LNCS 7369 (Berlin: Springer) 208-224.

[22] Tewari, C., Tatrari, G., Kumar, S., Pathak, M., Rawat, K. S., Kim, Y. N., ... & Sahoo, N. G. (2023). Can graphene-based composites and membranes solve current water purification challenges-a comprehensive review. Desalination, 116952.

[23] Waqas, M., Hashim, S., Humphries, U. W., Ahmad, S., Noor, R., Shoaib, M., ... & Lin, H. A. (2023). Composting processes for agricultural waste management: a comprehensive review. Processes, 11(3), 731.

[24] Tawonezvi, T., Nomnqa, M., Petrik, L., & Bladergroen, B. J. (2023). Recovery and recycling of valuable metals from spent lithium-ion batteries: A comprehensive review and analysis. Energies, 16(3), 1365.

[25] Azami, M., Kazemi, Z., Moazen, S., Dubé, M., Potvin, M. J., & Skonieczny, K. (2024). A Comprehensive Review of Lunar-based Manufacturing and Construction. arXiv preprint arXiv:2408.05823.

[26] Bavdekar, R., Chopde, E. J., Agrawal, A., Bhatia, A., & Tiwari, K. (2023, January). Post quantum cryptography: A review of techniques, challenges and standardizations. In 2023 International Conference on Information Networking (ICOIN)(pp. 146-151). IEEE.

[27] Stocco, T. D., Zhang, T., Dimitrov, E., Ghosh, A., da Silva, A. M. H., Melo, W. C., ... & Lobo, A. O. (2023). Carbon nanomaterial-based hydrogels as scaffolds in tissue engineering: a comprehensive review. International Journal of Nanomedicine, 6153-6183.

[28] Singh, R., Priya, H., Kumar, S. R., Trivedi, D., Prasad, N., Ahmad, F., ... & Rana, S. S. (2024). Gum Ghatti: A Comprehensive Review on Production, Processing, Remarkable Properties, and Diverse Applications. ACS omega, 9(9), 9974-9990.