Applying system of equations to factor semiprime numbers

Yonatan Zilpa

July 10, 2023

Abstract

This paper explores the use of a system of equations to factor semiprime numbers. Semiprime numbers are a special type of composite number that are the product of two prime numbers. Factoring semiprime numbers is important in cryptography and number theory. In this study, we present a method that applies a system of polynomial equations to factor semiprime number M. Where M can be any semiprime number. In fact, we build a family of systems where each system compose from three polynomial equations with three variables. The results of this study show that a solution for one system results with a complete factorization for a semiprime number. It may be possible to apply well known algorithms, such as Gröbner method [1], to solve one of those systems for a particular semiprime number M.

Keywords: Semiprime, factorization, system of equations

1 Introduction

Let s_1, s_2, and S be any integers such that $S = s_1 s_2$, then

$$(s_2 - s_1)^2 + 4S$$

is a perfect square. Indeed

$$(s_2 - s_1)^2 + 4S = (s_2 + s_1)^2.$$

Let M be a semiprime number and let p, q be its prime factors, where $q > p$. Let $d = q - p$ and let n and x be any integers, such that n divides $M - x$, then

$$\sqrt{\left(\frac{M-x}{n} - n\right)^2 + 4(M-x)} = \left|\frac{M-x}{n} + n\right| \tag{1.1}$$

is a positive integer. Thus, if $\left(\frac{M-x}{n} - n\right)^2 - 4x$ is a non-negative perfect square, then

$$\left(\frac{M-x}{n} - n\right)^2 - 4x = d^2. \tag{1.2}$$

Equation (1.2) implies that

$$\frac{M-x}{n} - n = \sqrt{4x + d^2}.$$

Hence, x must contain a factor t such that

$$\frac{x}{t} - t = d.$$

The number x must be of the form:

$$(d + j)j$$

where j is an integer. Let k be a positive integer less than p, then substituting x with $k(d + k)$ in equation (1.2) yields

$$\left(\frac{M - (d+k)k}{n} - n\right)^2 - 4(d+k)k = d^2 \tag{1.3}$$

Solving equation (1.3) for d we get the following two solutions

$$\begin{aligned} d_0 &= \tfrac{M-k^2+2kn-n^2}{k-n} = \tfrac{M-(k-n)^2}{k-n} \\ d_1 &= \tfrac{M-k^2-2kn-n^2}{k+n} = \tfrac{M-(k+n)^2}{k+n} \end{aligned} \tag{1.4}$$

Since d is a positive integer. The first equality of equation (1.4) implies that $|k-n|=1$ or

$$|n-k|=p \tag{1.5}$$

Substituting k with k_1 and n with $n+1$ in equation (1.3) yields $|k_1-(n+1)|=p$ and from this we get $k_1=k+1$. Similarly, substituting k with k_2 and n with $n+2$ in equation (1.3) yields $|k_2-(n+2)|=p$ which gives us $k_2=k+2$. This gives us the following system

$$\begin{aligned}
&\left(\tfrac{M-(d+k)k}{n}-n\right)^2-4k(d+k)=d^2\\
&\left(\tfrac{M-(d+(k+1))(k+1)}{n+1}-(n+1)\right)^2-4(k+1)(d+(k+1))=d^2\\
&\left(\tfrac{M-(d+(k+2))(k+2)}{n+2}-(n+2)\right)^2-4(k+2)(d+(k+2))=d^2
\end{aligned} \tag{1.6}$$

System (1.6) has three equations with three variables n,k,d, however this system is dependent. We may overcome this problem by trying other functions. Let $t:\mathbb{Z}\to\mathbb{Z}$ be any function, replace n with $t(n)$ and k with u in equation (1.3). Equality (1.5) implies that $u-t(n)=p$ (or $t(n)-u=p$) and $k-n=p$ (or $n-k=p$), which gives us a system of equations

$$\begin{aligned}
u-t(n)&=p\\
k-n&=p
\end{aligned}$$

from which we deduce $u-k-t(n)+n=0$ or equivalently $u=k+t(n)-n$. We get the following equality:

$$\begin{gathered}
\left(\frac{M-\Big(d+\big(k+t(n)-n\big)\Big)\big(k+t(n)-n\big)}{t(n)}-t(n)\right)^2+\\
-4\big(k+t(n)-n\big)\Big(d+\big(k+t(n)-n\big)\Big)=d^2
\end{gathered} \tag{1.7}$$

2 Building systems of equations with d_0

Based on equation (1.7) we can deduce a new system of three equations with three variables k,n and, d. We may find three functions $t_1,t_2,t_3:\mathbb{Z}\to\mathbb{Z}$ and replace $t(n)$ with $t_3(n)$ to get the third equation, $t(n)$ with $t_2(n)$ to get the second equation, and finally $t(n)$ with $t_1(n)$ to get the first equation. The key here is to select the functions t_1,t_2, and t_3 in such a way that our system has a unique solution, where $|n-k| \neq 1$. When moving d^2 to the left side of equality (1.7) and multiplying it with $t^2(n)$, the left side of this equality becomes:

$$
\begin{aligned}
\phi(t,n,k,d) := \quad & \Big(\Big(M - \Big(d + \big(k + t(n) - n \big) \Big) \big(k + t(n) - n \big) \Big) - t^2(n) \Big)^2 \\
& -4t^2(n) \big(k + t(n) - n \big) \Big(d + \big(k + t(n) - n \big) \Big) - t^2(n) d^2
\end{aligned}
\tag{2.1}
$$

If t is a polynomial function in $\mathbb{R}$ with integral coefficients, then ϕ can be viewed as a polynomial function from $\mathbb{R}^3$ to $\mathbb{R}$. In this case we also denote the function $\phi(t,n,k,d)$ with $\phi_t(x,y,z)$. We thus get a system of polynomial equations:

$$
\begin{aligned}
\phi_{t_1}(x,y,z) = 0 \\
\phi_{t_2}(x,y,z) = 0 \\
\phi_{t_3}(x,y,z) = 0
\end{aligned}
\tag{2.2}
$$

3 Building systems of equations with d_1

The problem with d_0 is that the variant of system (1.7) is infinite, any integer n,k such that $|n-k| = 1$ satisfying this system. However, applying solution d_1 in equality (1.4) and requiring that n,k be positive integers implies that

$$
k + n = p. \tag{3.1}
$$

Replacing n with $t(n)$ and k with u in equation (1.3) we get the following system

$$
\begin{aligned}
u + t(n) = p \\
k + n = p
\end{aligned}
\tag{3.2}
$$

from which we deduce $u + t(n) - k - n = 0$ or equivalently $u = n + k - t(n)$. Now we can replace k with $n + k - t(n)$ and n with $t(n)$ and d with d_1 in equation (1.3) to get

$$
\begin{aligned}
& \left(\frac{M - \Big(d + \big(n+k-t(n) \big) \Big) \big(n+k-t(n) \big)}{t(n)} - t(n) \right)^2 \\
& -4\big(n + k - t(n) \big) \Big(d + \big(n + k - t(n) \big) \Big) = d^2
\end{aligned}
\tag{3.3}
$$

Since $t(n)$ relies on the second equality of (1.4) and since $t(n)$ differs from n, the first solution in (1.4) won't solve equality (3.3). Hence, by replacing $t(n)$ with polynomial $t_1(n)$ with positive coefficients we get two independent polynomials.

Let us denote

$$\psi_t(n,k,d) := \quad \Big(\tfrac{M-(d+n+k-t(n))(n+k-t(n))}{t(n)} - t(n)\Big)^2$$

$$- 4\Big(d + (n+k-t(n))\Big)\Big(n+k-t(n)\Big) - d^2$$

then equality (3.3) becomes

$$\psi_t(n,k,d) = 0. \tag{3.4}$$

If we set $t(n) = n$, then equation (3.4) is equivalent to (1.3). However, if polynomial $t(n)$ differs from n, then solution d_0 is lost. Hence, for any polynomial $t_2(n)$ with positive integers that differs from n, polynomials ψ_n and $\psi_{t_2(n)}$ are independent.

We can repeatedly use the result $u = n+k-t(n)$, obtained from system (3.2), to get the following system of three polynomial equations with three variables:

$$\begin{aligned}
&\psi_{t_1}(n,k,d) = 0\\
&\lambda_{t_2}(n,k,d) = \psi_{t_1}(t_2(n), n+k-t(n), d) = 0 \qquad (3.5)\\
&\psi_{t_1}(t_3(n), n+k-t(n), d) + \lambda_{t_2}(t_3(n), n+k-t(n), d) = 0.
\end{aligned}$$

If polynomials t_1, t_2, and t_3 differ in pairs and having non-negative integers and if none of these polynomial is zero, then none of the polynomial in system (3.5) depends on the other.

4 Conclusions

The RSA cryptosystem [4] as well as all public key cryptography implementations rely on the complexity of semiprime factorization. Mathematical attacks based on known relations, such as Pythagorean primes [3] or the use of a polynomial of third degree order [6] have been recently proposed for potential methods for factoring semiprimes numbers. When it comes to factoring large semiprime numbers, well known existing algorithms may consume too much memory and running time. Other algorithms, such as the firefly algorithm [5], may address some of these issues [2].

In this article the problem of semiprime factorization has been attacked by exploiting relationships between M and two different numbers, that are less than M. We have used only quadratic relationships to construct a family of systems, where each of the system has three polynomial equations with three variables. Finding a solution for one of these systems may lead to a complete factorization of the semiprime number M.

Acknowledgements:

I would like to express my sincere gratitude to Professor Shai Haran, who provided invaluable feedback on the manuscript. His insightful comments and suggestions greatly improved the clarity and rigor of the research findings. I am grateful for his time and expertise, which helped me to refine my research questions and the methodology used to address them. Without his guidance, this paper would not have been possible.

References

[1] Bruno Buchberger. Ein algorithmus zum auffinden der basiselemente des restklassenrings nach einem nulldimensionalen polynomideal. *Universitat Innsbruck, Austria, Ph. D. Thesis*, 1965.

[2] Mohit Mishra, Utkarsh Chaturvedi, and Saibal K Pal. A multithreaded bound varying chaotic firefly algorithm for prime factorization. In *2014 IEEE International Advance Computing Conference (IACC)*, pages 1322–1325. IEEE, 2014.

[3] Anthony Overmars and Sitalakshmi Venkatraman. New semi-prime factorization and application in large rsa key attacks. *Journal of Cybersecurity and Privacy*, 1(4):660–674, 2021.

[4] Ronald L Rivest, Adi Shamir, and Leonard Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2):120–126, 1978.

[5] Xin-She Yang and Xingshi He. Firefly algorithm: recent advances and applications. *International journal of swarm intelligence*, 1(1):36–50, 2013.

[6] Y Zilpa. About efficient algorithm for factoring semiprime number. *J Theor Comput Sci Open Access*, 7:p053, 2021.