CS427 Mini-Project Report: Cryptanalysis of two schemes of Baba et al. by linear algebra methods

B Siddharth Prabhu
200010003@iitdh.ac.in

Devesh Kumar
200030017@iitdh.ac.in

10 April 2023

1 Problem Statement

We show that the attacks based on the linear algebra-based method introduced by the author of this paper, and the span-method introduced by Tsaban allow one to find the transmitted message in the cryptosystem (and the exchanged key) in the protocol proposed in the paper 'A Non-Abelian Factorization Problem and an Associated Cryptosystem' by Baba et al. Effectively, in this project, we will:

- Explore the Cryptosystem and Key Exchange Protocol described in Baba et al., which is based on FACTOR, and motivate cryptanalysis.
- Solution 1: Conduct cryptanalysis of the system by Tsaban's Span Method.
- Solution 2: Conduct cryptanalysis of the Diffie-Hellman-Type Key Exchange Protocol itself, using the author's Linear Decomposition Method.
- Discuss the concepts of the course that helped us tackle the paper.
- Look at future scope of this work.

2 Cryptosystem and Key Exchange

In the paper, the cryptosystem based on FACTOR and the corresponding key exchange protocol are likened with the El-Gamal cryptosystem and the Diffie-Hellman Key Exchange Protocol respectively. Let us take a brief look at what each of these are, and then dive into cryptanalysis in the next section.

2.1 The El-Gamal Cryptosystem

Let G be a (public) finite cyclic group with generator g, and let $x \in \mathbb{Z}$ be Alice's private key. The element g^x is public. To send a message $m \in G$, Bob picks a random integer y and sends the cipher text $c = (g^y, g^{xy}m)$ to Alice. To decrypt this message, Alice calculates $(g^y)^x = g^{xy}$ and inverts it to retrieve m. Hence, a cryptosystem has been described.

2.2 The classical Diffie-Hellman Key Exchange

Let G be a public finite cyclic group with generator g, and let $x \in \mathbb{Z}$ be Alice's private key, as well as $y \in \mathbb{Z}$ be Bobs private key. Alice publishes g^x and Bob publishes g^y. Then each of them computes the exchanged key $g^{xy} = (g^x)^y = (g^y)^x$. Hence, key exchange has been done.

2.3 Cryptosystem based on FACTOR

Baba et al. consider the function FACTOR (which factors a number ab into a and b to be one-way, i.e., its inverse is easy to compute, while it, by itself, is computationally expensive. Let G be any public group. Let g, $h \in G$ be two private elements of Alice, and let $< g >$ and $< h >$ be the cyclic subgroups generated by these elements, respectively. In order to define the FACTOR problem one assume that $< g > \cap < h >= \{1\}$. Let $f :< g > \times < h > \rightarrow G$ be a function defined as follows: $f(g^x, h^y) = g^x h^y$, where $x, y \in \mathbb{Z}$. We can see that this f is injective. Then, we can define: $\texttt{FACTOR}(g^x h^y) = f^{-1}(g^x h^y)$.

Let us consider a case where Alice is the recipient of a message $m \in G$ that Bob is sending. Alice randomly picks integers x, y, and sets public key $(G, g, h, g^x h^y)$. She also has a private key (g^x, h^y) for decryption. When sending the message m, Bob picks arbitrary integers x', y' and sends ciphertext:

$$c = (g^{x+x'} h^{y+y'}, g^{x'} h^{y'} m)$$

2.4 Key Exchange Protocol based on FACTOR

Before exchanging messages, Alice and Bob must exchange keys. Suppose G, g, h are defined as earlier, in FACTOR. Let Alice pick a pair of integers (x_1, y_1), and Bob pick another pair (x_2, y_2). Then, Alice sends the element $g^{x_1} h^{y_1}$ to Bob, who also sends $g^{x_2} h^{y_2}$ to Alice. Then, they each can recover $K = g^{x_1+x_2} h^{y_1+y_2}$, which will be their private key.

3 Solutions and Methodology

Although the problem of factorization seems computationally hard, the above-described systems that are based on FACTOR can be broken using the linear algebra attacks described below. To show this, we will prove that an attacker can effeciently compute $g^{x'} h^{y'}$, and can hence retrieve m.

3.1 Cryptanalysis of the FACTOR-based system using Tsaban's Span Method

The paper first considers a finite group, G, which is presented as a matrix group over a finite field, i.e., the group is generated by the powers of a matrix, and the elements are invertible ($G =< g >$, where $g \in M_n(\mathbb{F}_q)$). Now, for $g \in M_n(\mathbb{F}_q)$, we consider V to be the linear subspace of $M_n(\mathbb{F}_q)$ generated by all matrices of the from g_i, $i \in \mathbb{Z}$. Then, the dimension of V is strictly less than n.

To Prove: *Dimension of V is at most* $(n-1)$.

Proof: The powers of g (i.e., g^k for $k = 0, 1, 2, ..., n-1$) can be thought of as representing n different linear transformations of $M_n(\mathbb{F}_q)$ given by left multiplication by g^k. In other words, the matrix g^k acts on a vector x in $M_n(\mathbb{F}_q)$ by computing $g^k x$. Since g is an $n \times n$ matrix, note that its characteristic equation is of degree n. This can be written as:

$$c_0 I + c_1 g + c_2 g^2 + ... + c_{n-1} g^{n-1} = 0 \ ,$$

where the coefficients $c_i \in \mathbb{F}_q \ \forall \ i$, and I is the identity matrix of $M_n(\mathbb{F}_q)$. Note that $\{1, g, g^2, ..., g^n\}$ will be linearly dependent, since g is a root of the above-mentioned equation. Hence, any power of g beyond g^{n-1} can be expressed as a linear combination of the previous powers.

So, we can take any set of $n-1$ linearly independent powers of g (for example, $\{1, g, g^2, ..., g^{n-1}\}$), and any additional power of g can be expressed as a linear combination of those. Therefore, the dimension of the subspace generated by the powers of g is at most $n-1$. Hence, $dim(V) \leq n-1$.

If g^{k+1} lies in $Lin_{\mathbb{F}_q}(1, g, g^2, ..., g^k)$, then, evidently, $g^{k+t}, g^{1-t} \in Lin_{\mathbb{F}_q}(1, g, g^2, ..., g^k)$ (for every $t = 2, 3, ...$). We can efficiently construct a basis $1, g, g^2, ..., g^k$, by checking for $l = 1, 2, ...$ if g^{l+1} lies in $Lin_{\mathbb{F}_q}(1, g, g^2, ..., g^l)$ or not. The smallest l that does not lie in the set will be chosen as the value of k. Gauss Elimination can be used to efficiently do this.

Consider the equation $f(g^x h^y)h = hf(g^x h^y) \sim fg^x h = hfg^x$. that is linear with respect to n^2 unknown entries of matrix f. We will seek an f in V, i.e., f of the form:

$$f = \sum_{i=0}^{k} \alpha_i g^i$$

We know that there is a non-degenerate solution $f = g^{-x}$. Also, we can efficiently construct a basis $e_1, e_2, ..., e_p$ of the subspace of all solutions in V. Then, we shall use the *Invertibility Lemma*.

Invertibility Lemma: *Let $a_1, a_2, ...a_m \in M_n(\mathbb{F})$ be such that $span\{a_1, a_2, ...a_m\} \cap GL_n(\mathbb{F}) \neq \emptyset$. Let S be a finite subset of $\mathbb{F}$. If $\alpha_1, \alpha_2, ..., \alpha_m$ are chosen uniformly and independently from S, then the probability that $\alpha_1 a_1 + ... + \alpha_m a_m$ is invertible is at least $1 - \frac{n}{|S|}$.*

Proof: Let $f(t_1, ..., t_m) = det(t_1 a_1 + ... + t_m a_m) \in \mathbb{F}[t_1, ..., t_m]$, where $t_1, ..., t_m$ are scalar variables. This is a determinant of a matrix whose coefficients are linear in the variables.
By the definition of determinant as a sum of products of n elements, f is a polynomial of degree n. As $span\{a_1, a_2, ...a_m\} \cap GL_n(\mathbb{F}) \neq \emptyset$, f is non-zero. This proof will be completed using the Schwartz-Zippel Lemma, which will be proven after this.

Schwartz-Zippel Lemma: *Let $f(t_1, ..., t_m) \in \mathbb{F}[t_1, ..., t_m]$ be a nonzero multivariate polynomial of degree n. Let S be a finite subset of $\mathbb{F}$. If $\alpha_1, ..., \alpha_m$ are chosen uniformly and independently from S, then the probability that $f(\alpha_1, ..., \alpha_m) \neq 0$ is at least $1 - \frac{n}{|S|}$.*

Proof: Let us prove this lemma by induction on m. If $m = 1$, then f is a univariate polynomial of degree n, and thus has at most n roots. For the inductive step, assume that $m > 1$ and write:

$$f(t_1, ..., t_m) = f_0(t_2, ..., t_m) + f_1(t_2, ..., t_m)t_1 + f_2(t_2, ..., t_m)t_1^2 + ... + f_k(t_2, ..., t_m)t_1^k$$

with $k \leq n$ maximal such that $f_k(t_2, ..., t_m)$ is nonzero. The degree of $f_k(t_2, ..., t_k)$ is at most $m - k$. For each choice of $\alpha_2, ..., \alpha_m \in \mathbb{F}$ with $f_k(t_2, ..., t_m) \neq 0$, $f(t_1, \alpha_2, ..., \alpha_m)$ is a univariate polynomial of degree k in the variable t_1. By the induction hypothesis (for $m = 1$), for random $\alpha_1 \in S$, $f(\alpha_1, \alpha_2, ..., \alpha_m)$ is nonzero with probability at least $1 - \frac{n}{|S|}$.
By the induction hypothesis,

$$\begin{aligned} Pr[f(\alpha_1, ..., \alpha_m) \neq 0] &\geq Pr[f_k(\alpha_2, ..., \alpha_m) \neq 0] \cdot Pr[f(\alpha_1, ..., \alpha_m) \neq 0 \mid f_k(\alpha_2, ..., \alpha_m) \neq 0] \\ &\geq \left(1 - \frac{n-k}{|S|}\right)\left(1 - \frac{k}{|S|}\right) \\ &\geq \left(1 - \frac{n}{|S|}\right) \end{aligned}$$

Let the required f be found (best case probability consideration). Then, $f(g^x h^y) = h(fg^x)$, and further, $f(g^{x+x'} h^{y+y'}) = (g^{x'} h^{y'})h^y(fg^x)$. Also,

$$(g^{x'} h^{y'})h^y(fg^x)f^{-1}(g^x h^y)^{-1} = g^{x'} h^{y'}$$

So, we have:

$$(g^{x'} h^{y'})^{-1}(g^{x'} h^{y'} m) = m$$

Hence, the message m is recovered.

3.2 Cryptanalysis of the FACTOR-based key exchange protocol using the Linear Decomposition Method

Here, the paper applies and describes only the author's linear decomposition method, by first considering $G \leq M_n\mathbb{F}$ as a matrix group over an arbitrary field $\mathbb{F}$. From this, we consider $V = Lin_{\mathbb{F}}(< g >< h >)$, the linear subspace of $M_n(\mathbb{F})$, generated by all matrices of the form $g^i h^j$, $i, j \in \mathbb{Z}$. Then, $dim(V) \leq (n-1)^2$. (Recall that our goal here is to obtain the private key $K = g^{x_1+x_2} h^{y_1+y_2}$.)

Let $e_1, e_2, ..., e_r$ be a basis of V that can be efficiently obtained, in the same way as earlier. Let us consider $e_i = g^{u_i}h^{v_i}$, for $u_i, v_i \in \mathbb{Z}, i = 1, ..., r$. Then, since $g^{x_1}h^{y_1} \in V$, we can efficiently find a representation of this vector in terms of the basis vectors obtained above, as follows:

$$g^{x_1}h^{y_1} = \sum_{i=1}^{r} \alpha_i e_i \ , \ \ \alpha_i \in \mathbb{F} \ , \ i = 1, ..., r$$

Then,

$$\sum_{i=1}^{r} \alpha_i g^{u_i}(g^{x_2}h^{y_2})h^{v_i} = g^{x_2}\left(\sum_{i=1}^{r} \alpha_i e_i\right) h^{y_2} = K$$

Hence, we have succeeded in finding K. (Note that the paper also talked about cryptanalysis of the cryptosystem using this Linear Decomposition method, but it hasn't been detailed in this report due to its similarity with this part.) Hence, we have found linear algebraic ways to attack cryptosystems based on the given Non-Abelian Factorization problem (FACTOR), to get the plaintext message back from the ciphertext, and to get the key itself!

4 Code Simulation

Although this is a theory-heavy project, we have coded a basic simulation of the cryptosystem in `factor.py` based on the non-abelian factorization problem. The code has some sample inputs, and it can also take plaintext input from the user. The encrypted and decrypted text are displayed as output. While this code may seem simple, it is just to illustrate the ease of implementation of the cryptosystem. Run the code using `bash run.sh` , and enter input as prompted.

5 From CS427: Contributions by the course itself

This course has helped us in different ways with respect to this project – including, but not limited to:

- Basic Field Theory in the direction of Linear Algebra (in addition to whatever ideas of Field Theory we had already)
- Linear Algebra: Linear Independence, Spans, Vector Spaces, Subspaces, Determinants

While this project may not illustrate the second half of CS427 (Optimization), we hope it could establish how this course serves a multi-level purpose. While the primary goal is optimization, the math we've learned is immeasurably useful in many domains - both related and unrelated. This intersectionality is something that we have increasingly come to appreciate during the course of this project.

6 Future Scope of this work

- Improvements in the Cryptographic schemes of the referenced paper, keeping in mind the vulnerabilities. This could include larger key size, alternative one-way functions (that are more resistant to linear algebra attacks), while preserving the efficiency and ease of implementation.
- This paper highlights the importance of rigorously analyzing the security of cryptographic schemes, and provides valuable insights into the limitations of the non-abelian factorization-based specific techniques, in the context of public-key (asymmetric) encryption.
- Creating new cryptosystems based on new mathematical structures, like higher-dimensional subspaces
- Advanced linear algebra techniques, like tensor algebra and multilinear maps, in the design of more secure cryptographic protocols. This would be useful in some areas of cryptography like homomorphic encryption and zero-knowledge proofs.
- If we include optimization, then possible improvements that could be made to cryptosystems include: Efficient Cryptanalysis, Secure Key Exchange, Quantum-safe cryptography, and Privacy-preserving machine learning.

7 Related Works

- (Primary Reference of the Paper): Baba et al. (2011). A non-Abelian factorization problem and an associated cryptosystem. IACR Cryptology ePrint Archive. 2011. 48.
- Ben-Zvi, Adi et al. "Cryptanalysis via Algebraic Spans." Annual International Cryptology Conference (2018).
- Tsaban, Boaz. 'Polynomial Time Cryptanalysis of Noncommutative-Algebraic Key Exchange Protocols'. CoRR, vol. abs/1210.8114, 2012, http://arxiv.org/abs/1210.8114.