CRYPTOGRAPHIC ALGORITHM USING MATRIX INVERSION AS DATA PROTECTION

Peter B. Zirra

Department of Mathematical Sciences
Adamawa State University
Mubi, Nigeria

zirrapeter@yahoo.com

G.M. Wajiga

Department of Mathematics and Computer Science,
Federal University of Technology
Yola, Nigeria

gwajiga@gmail.com

ABSTRACT

This study aimed at providing algorithms to secure sensitive information in store or in transit via unsecure channels from the hands of the Internet criminals by scrambling the sensitive information into a set of linear equations in matrix form and deciphering by solving the systems of the linear equations in conjunction with the principles of the delta encoding scheme, a formula and a lookup alphanumeric position table. The algorithms were tested with samples of real data and the result practically demonstrated how systems of non-linear equations turned into linear equations are indeed of relevance in providing strong and complex encryption and decryption keys to protect sensitive information in store or in transit via unsecure channels. The proposed algorithms ensured that unauthorized users do not have access to sensitive and classified documents stored or transmitted to an intended recipient without a secret key. This means that, there is confidentiality, non-repudiation and integrity of our sensitive and classified information from the hands of unauthorized users on the Internet because of the robustness of the proposed algorithms.

Keywords: Cryptographic algorithms, Delta encoding, Newton’s method, Inversion matrix, Non-linear equations, Decryption key, Cryptanalyst.
INTRODUCTION

Information security has become a very critical aspect of modern computing systems. With the global acceptance of the Internet, virtually every computer in the world today is connected to every other. While this has created tremendous productivity and unprecedented opportunities in the world we live in, it has also created new risks and burden for the users of these computers. The users, business, education and organizations worldwide have to live with constant fear of threats from intruders (also known as terrorists, attackers, hackers, crackers, unauthorized users), who use a variety of techniques and tools to break into computer systems, steal information, change data and cause havoc (Kahate, 2008). If no security measures are taken, there is no doubt that such data and other sensitive information will continuously be faked, polished and formatted by the hackers.

There have been several high-profile incidents of loss of individual privacy and confidentiality of data in unsecure communication channels for example, the Russian attacker (called Maxim) managed to intrude into a merchant Internet site and obtained 300,000 credit card numbers from its database (Kahate, 2008); a Swedish hacker broke into Microsoft’s Hotmail Website and created a mirror site. This site allowed anyone to enter any Hotmail user’s email id and read their emails. These attributed to successful attacks on security (Kahate, 2008). An Australian long-time computer hacker called Julian Assange hacked into the master terminal that the Canadian telecom company Nortel, maintained in Melbourne, Australia. In 2006, he created a WikiLeaks, a website that published confidential government documents and images (Discoverthenetwork.org, 2010; The Times of India, 2011). There were also recent reports from the United States and Canada that indicate that ‘hackers’ penetrated the information database of the American Military System in 2009 to wreak monumental havoc (Olaoye, 2011).

Therefore, cryptographic algorithms which help prevent interception and enhance message security are now of primary importance but the challenging part of cryptography today is how to create a key that would prevent an unauthorized decryption of message. A key is simply a parameter to the algorithm that allows the encryption and the decryption processes to occur.

This paper provides cryptographic algorithms using high mathematical formulations that enciphered message into a form of systems of non-linear equations that are deciphered through the Newton’s method of solving the systems of non-linear equations, and the principles of delta encoding scheme together with the help of a lookup alphanumeric position table which is notoriously difficult.
The idea behind these proposed cryptographic algorithms is to provide protection access to sensitive resources in store or transit, by ensuring that the resources required for revealing the secret information are strong and complex enough through a high mathematical formulation of the secret key. The more illegal trials attempted, the more difficult it is for the criminals to unveil the sensitive information. This may discourage the criminals based on the complexity of the decryption key.

BACKGROUND

Koblitz (1994) viewed cryptography as an algorithmic process of converting a plain text message to a cipher text message based on an algorithm that both the sender and the receiver know, so that the cipher text message can be returned to its original, plain text form. In its cipher form, a message cannot be read by anyone but the intended receiver. The act of converting a plain text message to its cipher text form is called enciphering (Potdar & Chang, 2004). Reversing that act (i.e., cipher text form to plain text message) is deciphering. Enciphering and deciphering are commonly based on a key referred to as encryption or decryption key, respectively.

There are a number of key-based algorithms for performing encryption and decryption (Buchmann, 2005; Stallings, 2006). Figg (2005) classified cryptography algorithm into two: symmetric and asymmetric key cryptography.

In symmetric key cryptography, the same key is used for both encryption and decryption (Yusuf, 2007). A shared key is vulnerable to fall into the wrong hands. Once a key becomes known to someone not authorized, then the security based on that key is compromised. In asymmetric key cryptography, one key is used for encryption and another, mathematically related key, is used for decryption (Beekman, 1999; Rudolf, 2000; Su, Zobel, & Iwu, 2003; Laudan & Traver, 2004). An asymmetric key can be much less secure than a symmetric key since it avoids the mechanism of key sharing. This study reports on a similar encryption technique that uses an asymmetric key due to its robustness.

Types of Key-based Encryption Algorithms

Currently, there are several kinds of key-based encryption algorithms categorized by their functions and target groups. The most common key based encryption techniques are given (Rudolf, 2000; Moore, 2001; Diaa, Hatem & Mohiy, 2010) but most of these algorithms were found to have weak keys or sometimes the keys could easily be guessed. There is large research literature

Obviously, as observed from the forgoing, it seems that, there appears to be no stronger ciphers that can guarantee the safety of our vital information from the hands of the enemy.

**Systems of Linear and Non-linear Equations**

This refers to problems involving the solution of systems of linear and non-linear equations, possibly with a very large number of variables. In recent years considerable interest in systems of non-linear equations has been stimulated due to their numerous applications in the area of science and engineering (Biazar & Ghanbary, 2009).

Crina and Ajith (2008) defined a non-linear system of equation as:

\[
f(x) = \begin{bmatrix} F_1(x) \\ F_2(x) \\ \vdots \\ F_n(x) \end{bmatrix} = 0
\]

where \( x = x_1, x_2, \ldots, x_n \) refers to \( n \) equations and \( n \) variables, and \( F_1, F_2, \ldots, F_n \) are non-linear functions in the space of all real-valued continuous functions on \( \Omega = \prod_{i=1}^{n} [a_i, b_i] \subset \mathbb{R}^n \). Some of the equations can be linear, but not all of them. Finding a solution for a non-linear system of equations \( f(x) \) involves finding a solution such that every equation in the non-linear system is zero i.e. \( F(x) = 0 \) (Burden & Faires, 2001). This system can be referred by:

\[
\begin{align*}
f_1(x_1, x_2, x_3, \ldots, x_n) &= 0 \\
f_2(x_1, x_2, x_3, \ldots, x_n) &= 0 \\
&\vdots \\
f_n(x_1, x_2, x_3, \ldots, x_n) &= 0
\end{align*}
\]

Biazar and Ghanbary (2008) assume that the system admits a unique solution.

**Newton’s Method**

There several ways to solve systems of non-linear equations. Probably the most popular techniques are: Newton-type method, Trust-region method,
Broyden method, Secant method, and Halley method (Crina & Ajith, 2008). The most known iterative method for solving systems of non-linear equations is the classical Newton’s method.

\[ J(x_n)\delta_x = -F(x_n) \]  

(3)

where \( \delta_x = x_n - x_{n-1} \) is the solution vector for the set of simultaneous linear equations and \( J(x_n) \in \mathbb{R}^{m \times n} \) denotes the Jacobian matrix to \( F \) evaluated at current approximation \( x_n = x_{n,1}, x_{n,2}, \ldots, x_{n,n} \) and \( x_{n+1} \) the next approximation. The Jacobian matrix is defined as:

\[ J = \begin{pmatrix}
\frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n}
\end{pmatrix} \]  

(4)

Explicit computation of \( \delta_x \) is avoided by performing the operation in a two-step manner.

Non-linear problems are often treated numerically by reducing them to a sequence of linear problems (Burden & Faires, 2001).

Matrix Inversion of Square Matrix

A linear system of equations is a set of \( n \) linear equations in \( k \) variables (sometimes called “unknowns”). Linear systems can be represented in matrix form as the matrix equation.

\[ Ax = b \]  

(5)

Where \( A = \begin{pmatrix}
a_{11} & \cdots & a_{1n} \\
\vdots & \ddots & \vdots \\
a_{n1} & \cdots & a_{nn}
\end{pmatrix} \) is the matrix of coefficients, \( x = \begin{pmatrix}
x_1 \\
\vdots \\
x_n
\end{pmatrix} \) is the column vector of variables, and \( b = \begin{pmatrix}
b_1 \\
\vdots \\
b_n
\end{pmatrix} \) is the column vector of solutions.

If \( k<n \), then the system is (in general) over-determined and there is no solution.

If \( k=n \) and the matrix \( A \) is non-singular, then the system has a unique solution in the \( n \) variables. In particular, as shown by Cramer’s rule, there is a unique solution if \( A \) has a matrix inverse \( A^{-1} \). In this case,
\[ x = A^{-1}b \]  \hspace{1cm} (6)

If \( b = 0 \), then the solution is simply \( x = 0 \). If \( A \) has no matrix inverse, then the solution set is the translate of a subspace of dimension less than \( n \) or the empty set (Stroud and Dexter, 2007).

**Delta Encoding**

Delta encoding \( \delta \) refers to several techniques that store data as the difference between successive samples (or characters), rather than directly storing the samples themselves (Steven, 2007) as shown Table 1.

<table>
<thead>
<tr>
<th>Original data stream</th>
<th>05 10 05 12 20 15 15 01 23 52 45 ...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta encoded (( \delta ))</td>
<td>05 -05 05 -07 08 05 00 14 -22 -29 07 ...</td>
</tr>
</tbody>
</table>

The first value in the delta encoded file is the same as the first value in the original data. All the following values in the encoded file are equal to the difference (delta) between the corresponding value in the input file, and the previous value in the input file.

More precisely,

\[ \delta_x = (x_n - x_{n-1}), \quad \text{where } n = 1, 2, 3, \ldots \text{ and } x_0 = 0 \]  \hspace{1cm} (7)

**EXPERIMENTAL DESIGN**

The proposed encryption and decryption algorithms consist of multilevel enciphers.

**Words Compression Process**

The first level is realized through the compression of words. The plain text to be ciphered is first compressed word by word as follows:
Algorithm Words_Compression(input, output)

Step 1: Read plain text (T)
Step 2: Split plain text (T) into words
Step 3: Determine the word length (L) from split words
Step 4: Initialize variable K = 0
Step 5: Set K = K + 1
Step 6: If K = L then
    a: set the first word = word token (K)
    b: go to Step 7
else Step 6
    output the compressed text
end of Step 6
Step 7: Initialize variable J = 0
Step 8: a: append K to index of words
    b: set J = J+1
Step 9: Check if J < L then
    a: set second word = word token (J)
    b: repeat Step 10
else Step 9
    output the index of the words
    return to Step 6
end of Step 9
Step 10: Compare the first word and the second word
Step 11: If first word = second word then
    a: check first word against any empty value
    b: if check is false then
        i: append content of first word to compressed word
        ii: set the word token (J) to empty
        iii: append J to the trailing of the index of words
        iv: go to Step 12
    else Step 11
        return Step 8(b)
    end Step 11(b)
else of Step 11
    c: check second word against an empty value
    d: if check is false
        i: append content of second word to compressed text
        ii: return to Step 8(b)
    else
        return to Step 8(b)
    end of Step 11(d)
end of Step 11(d)
Step 12: append content of compressed word to compressed text
return to Step 8(b)
End Algorithm Word_Compression
Transformation of Words to Linear and Non-linear Equation

The second level is achieved by transforming the compressed words (the number of words that were left from Algorithm (1) into the system of non-linear equations (2). The proposed mathematical formulation steps are as given below.

- Count the words from Algorithm 1, i.e. the number of words that were left. Assuming we have $p$ words, it requires $p$ equations that must be within the permitted region of the alphabetic range coupled with the characters to be used.
- Turn each word into an equation whose variables must not be more than $p$, the number of equations allowed.
- Transform each word into the proposed equation starting with $x_1$ and next $x_2$, ..., $x_p$ and then adding or subtracting the ‘space’ of the current character from $x$ (i.e adding a quantity denoting the position of the alphabet from the position of the variable being used helps to ensure the accomplishment of this).
- Multiply each term obtained in (iii) by a factor, sum all and then equate to the sum of the variable positions in the word. These give nonlinear systems of equations $Ax=b$.

Use Newton’s Method

To further confuse the cryptanalyst, the system of non-linear equations is transformed again by using equation (3) as follows:

(a) set $n = 0$,
(b) guess an approximation solution $X_0$,
(c) calculate $J(x_n)$ and $F(x_n)$, where $J(x_n)_{i,j} = \frac{\partial f_i(x)}{\partial x_j}$, for $1 \leq i, j \leq n$,
(d) solve the linear system, $J(x_n)\delta x = -F(x_n)$,
(e) set, $\delta x = x_{n+1} - x_n$,
(f) set, $\delta x = \delta x + 1$, and
(g) test for convergence.

For the purpose of this paper, we consider the following:

1. Explicit computation of the inversion of Jacobi (i.e. $J(x)^{-1}$) is avoided as this will involve additional iteration for determining $J(x)^{-1}$.
2. Instead a new approximation $x_n$ is obtained by adding $\delta x$ to $x_{n-1}$ to as the next iterate.
3. The starting point is chosen carefully with an approximate idea of the guess of the function in mind as the difference between the sums of the coefficients of and divided by of a given system of equations.
Inversion of Matrix

To disguise the message further, the matrix $A^{-1}$ is obtained from matrix $A$, and $x = A^{-1}b$ is transmitted as cipher text.

Delta Encoding Scheme

This is the final level of the encryption:

1. We formulate a lookup table as shown in Table 2.

Table 2

<table>
<thead>
<tr>
<th>A Lookup Alphanumeric Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>D</td>
</tr>
<tr>
<td>E</td>
</tr>
<tr>
<td>F</td>
</tr>
</tbody>
</table>

2. Hide the positions of the variables that represent the original characters by encoding the distance of each variable solution of the variables of the text message from Table 2 as follows:
(a) for each variable index representing a character in a word, $x_n$, we write the difference between its actual position of the character it represents and its variable solutions obtained, and

(b) find

$$\delta_x = x_n - x_{n-1}$$

with $x_0 = 0$, $n = 1, 2, 3, \ldots$ Thus, is the current quantity representing the position of each character from its variable solution and $x_{n-1}$ is the previous quantity representing the position of each character from its variable solution.

This is also a step further to create more states of confusion for an unauthorized person.

Decryption Process

The decryption process is performed simply by:

1. Solving the systems of linear equations in Inversion of Matrix.
2. Using the variable solutions obtained in conjunction with the formula,

$$s_k = v_s + \sum_{n=1}^{k} \delta_{x_n}$$

where is the variable solution.
3. Using Table 2 as a lookup table to finally locate the position of the characters of each word.
4. Writing, the characters (in deciphering) as they appeared coupled with spaces between them and where there are index values in place of the words that previously occurred, the index values are interpreted and written in the places they appeared.

APPLICATION OF THE EXPERIMENTAL DESIGN

The Encryption Process

Taking the text “who is promising who”, it can be encrypted using the Algorithm 1 and the procedures as follows:

1. If the plain text “who is promising who” is input into Algorithm 1, it reduces the words to “who is promising” and one index value that indicated the re-occurrence of a word. In other words, it means that this algorithm produces $n = 3$ number of words. This number will now serve as a guide to formulate the number of equations whose variable index must not also exceed the number of words.
2. Using the procedure in section Transformation of Word to Linear and Non-linear Equation, result of:

76
who: \[ x_1^2 + x_1x_2 + x_1x_3 = 6 \]
is: \[ x_1^2 + x_1x_2 = 3 \]
promising: \[ 4x_1^2 + 2x_1x_2 + 3x_1x_3 = 17 \]

This above representation can be written as:

\[
\begin{align*}
& x_1^2 + x_1x_2 + x_1x_3 - 6 = 0 \\
& x_1^2 + x_1x_2 + 0x_1x_3 - 3 = 0 \\
& 4x_1^2 + 2x_1x_2 + 3x_1x_3 - 17 = 0 \\
\end{align*}
\]

3. Apply the principle of solving equation (3) on equation (9) to obtain:

\[
J(x_n)\delta x = -F(x_n) \Rightarrow \begin{bmatrix} 2x_1 + x_2 + x_3, & x_1 & x_1 \\ 2x_1 + x_2, & x_1 & x_1 \\ 8x_1 + 2x_2 + 3x_3, & 2x_1 & 3x_1 \end{bmatrix} \delta x = \\
- \begin{bmatrix} x_1^2 + x_1x_2 + x_1x_3 - 6 \\ x_1^2 + x_1x_2 - 3 \\ 4x_1^2 + 2x_1x_2 + 3x_1x_3 - 17 \end{bmatrix}
\]

4. We take an initial guess of \( x_0 = (1, 1, 1)^T \) and substitute in equation (10), to obtain \( J(x_0) \) and \( f(x_0) \) as in equation (11) in matrix notation.

Hence, from equation (10) we obtain:

\[
\begin{bmatrix} 4 & 1 & 1 \\ 3 & 1 & 0 \\ 13 & 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \\ x_3 - 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 8 \end{bmatrix}
\]

where \( A = \begin{bmatrix} 4 & 1 & 1 \\ 3 & 1 & 0 \\ 13 & 2 & 3 \end{bmatrix} \), \( x = \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \\ x_3 - 1 \end{bmatrix} \) and \( b = \begin{bmatrix} 3 \\ 1 \\ 8 \end{bmatrix} \)

5. Transform equation (11) to \( x = A^{-1}b \) as follows:

(a) compute \( |A| \) as \(-\frac{1}{4}\).
Equation (12) is now transmitted as cipher text to the intended recipient.

The Decryption Process

1. Apply the steps in Decryption Process to equation (12), to obtain the approximate solution $x_1 = 1$, $x_2 = 2$ and $x_3 = 3$.
2. Obtain the lookup alphanumeric position in Table 2.
3. Find the $\delta_x$ values associated with each solution of the variable index of the formulated systems of non-linear equations calculated using equation (7) in conjunction with Table 2. The result is depicted in Table 3.

<table>
<thead>
<tr>
<th>Virtual Position of Characters ($x_v$)</th>
<th>76 66 6c 68 71 6d 71 6c 6b 68 71 68 6b 6b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solution of Equations ($v_x$)</td>
<td>1 2 3 1 2 3 1 3 2 1 2 1 3 1</td>
</tr>
<tr>
<td>$\delta_x$</td>
<td>76-10 6 68 9 6d 04-05-01-03 09-09 03-05</td>
</tr>
</tbody>
</table>

4. Use the formula in conjunction with Tables 2 and 3. The effect of the formula is shown in Table 4.

RESULT

Table 4 shows the result of the deciphered text from the enciphered text obtained from the proposed algorithms.
Table 4

Plain Text Recovery

<table>
<thead>
<tr>
<th>Position of Word</th>
<th>From Table 2 (R.C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 ) = 1 + 76 = 77</td>
<td>77 = w</td>
</tr>
<tr>
<td>( x_2 ) = 2 + 76 - 10 = 68</td>
<td>68 = h</td>
</tr>
<tr>
<td>( x_3 ) = 3 + 76 - 10 = 6f</td>
<td>6f = o</td>
</tr>
<tr>
<td>( x_1 ) = 1 + 68 = 69</td>
<td>69 = i</td>
</tr>
<tr>
<td>( x_2 ) = 2 + 68 + 9 = 73</td>
<td>73 = s</td>
</tr>
<tr>
<td>( x_3 ) = 3 + 6d = 70</td>
<td>70 = p</td>
</tr>
<tr>
<td>( x_1 ) = 1 + 6d + 4 = 72</td>
<td>72 = r</td>
</tr>
<tr>
<td>( x_3 ) = 3 + 6d + 4 - 5 = 6f</td>
<td>6f = o</td>
</tr>
<tr>
<td>( x_2 ) = 2 + 6d + 4 - 5 - 1 = 6d</td>
<td>6d = m</td>
</tr>
<tr>
<td>( x_1 ) = 1 + 6d + 4 - 5 - 1 - 3 = 69</td>
<td>69 = i</td>
</tr>
<tr>
<td>( x_2 ) = 2 + 6d + 4 - 5 - 1 - 3 + 9 = 73</td>
<td>73 = s</td>
</tr>
<tr>
<td>( x_3 ) = 3 + 6d + 4 - 5 - 1 - 3 + 9 - 9 = 69</td>
<td>69 = i</td>
</tr>
<tr>
<td>( x_1 ) = 3 + 6d + 4 - 5 - 1 - 3 + 9 - 9 + 3 = 6e</td>
<td>6e = n</td>
</tr>
<tr>
<td>( x_1 ) = 1 + 6d + 4 - 5 - 1 - 3 + 9 - 9 + 3 - 5 = 67</td>
<td>67 = g</td>
</tr>
</tbody>
</table>

Hence the enciphered words are fully recovered.

DISCUSSION

It is clearly observed from the result of the proposed algorithm that any method of attack to find the decryption key by the unauthorized user required: solving the set of linear equations in equation (11), finding the values associated with
the variable index in Table 3, formulating the equation and understanding the position of the alphanumeric characters in Table 2. This is notoriously difficult. Hence, the proposed method provides strong protection of data from unauthorized access in storage or transmission. This is in agreement with an earlier study by Ritter (2007) who revealed that resources required for revealing a secret message should be strong and complex enough through a hiding key.

The result in Table 4 was obtained through the use of a public key (asymmetric cryptography). This avoided the risk of sharing keys (key falling into the wrong hands) associated with the symmetric cryptography. The findings are in accordance with the study conducted by Murdoch (2001) who found out that there is less risk associated with a public key than the symmetric key and the security based on that key is not compromised (Talbot & Welsh, 2006).

From the finding, the decrypted messages were not visible to an unauthorized person without the decryption keys. This proved how the proposed algorithms can fortify the security of our sensitive and classified information both in store and transit from the hands of Internet crooks. According to Wang (2002) it means that there is confidentiality, non-repudiation and integrity of our sensitive and classified information over the Internet from the hands of Internet terrorists. This is because of the robustness of the proposed algorithms. This view was also highlighted by Kahate (2008), and Kessler (2010).

The result in Table 4 has proved the validity, reliability and practicality of the proposed application of the systems of non-linear equations to encryption and decryption algorithms in handling real life situations in an unsecured environment. It was for this reason that Biazar and Ghanbary (2008) observed that in recent years, considerable interest in systems of non-linear equations has been stimulated due to its numerous applications in the area of Science and Engineering.

**CONCLUSION**

The proposed theoretical design provided strong protection of sensitive resources from illegal access by hackers while in store or transit, by ensuring that the resources required to reveal the secret information was strong and complex enough through a complex decryption key. The more illegal trials attempted, the more difficult it was for the criminals to unveil the sensitive information.
The findings of this study are consistent with the claims of researchers such as Biazar and Ghanbary (2009), and Kahate (2008) that the systems of non-linear equations are indeed of relevance in providing security to secure our sensitive information in store or in transit via unsecure channels from the hands of the eavesdroppers through the use of strong and complex encryption and decryption keys that are notoriously difficult to break.

Thus, these algorithms have proved their practicality in handling real life problems in an unsecured environment.

REFERENCES


