Journal of Natural Sciences Research
ISSN 2224-3186 (Paper) ISSN 2225-0921 (Online)
Vol.3, No.1, 2013

A Probabilistic Data Encryption scheme (PDES)

Aldrin W. Wanambisi[1*] Cleophas Maende[2] , Geoffrey Muchiri Muketha[3] ,Shem Aywa[4]

1. School of Pure and Applied Science, Mount Kenya University, P.O box 342-00100, Thika, Kenya.
2. School of Post graduate studies, Mount Kenya University, P.O box 342-00100, Thika, Kenya.
3. Dept of Computer Science, Masinde Muliro University of Science and Technology, P.O Box 150-50100, Kakamega, Kenya.
4. Dept of Mathematics, Masinde Muliro University of Science and Technology, P.O Box 150-50100, Kakamega, Kenya.

* E-mail of the corresponding author: wawanambisi@gmail.com

Abstract

In this paper the author presents a probabilistic encryption scheme that is polynomially secure and has the efficiency of deterministic schemes. From the theoretical construction of Brands and Gill (1996), it is clear that the proof of Pseudo randomness of the quadratic residue generator is complete if it can be shown that there exists a one-way function under the possible assumption that it is infeasible to solve the quadratic residuacity problem provided the factorization of the composite integer is unknown.

Key words: *Quadratic residuacity, pseudorandom number generator, one-way function*

1. Introduction

Encryption schemes were the first central area of interest in cryptography (Diffie and Hellman 1976). They deal with providing means to enable private communication over an insecure channel. A *sender* wishes to transmit information to a receiver over an insecure channel that is a channel which may be tapped by an *adversary*. The information to be communicated, which we call the plaintext derived from an alphabet (a field, F), must be put into a special code (encrypt) to a cipher text (encoded information). The authorized person must be given some way to convert the cipher text back to the original message (decrypt), while this must not be possible for an unauthorized person. The authorized person is considered to have a key at his disposal, enabling him to recover the actual message. Probabilistic encryption, discovered by Goldwasser and Micali (1984), is a design approach for encryption where a message is encrypted into one of many possible cipher texts (not just a single cipher text as in deterministic encryption), in such a way that it is provably as hard to obtain partial information about the message from the cipher text, as it is to solve some hard problem. In previous approaches to encryption, even though it was not always known whether one could obtain such partial information, neither was it proved that one could not do so. The scheme had substantial message expansion due to the bit-by-bit encryption of the message which in general makes the scheme not practical. (Fuchsbauer 2006).

In this paper the authors seek to develop a practical encryption scheme that combines the security criteria of the Goldwasser and Micali probabilistic scheme and the efficiency of the deterministic schemes by use of one-way function with a predicate, hence PDES.

2. Encryption Design Concepts

The first public-key Data Encryption Schemes were deterministic algorithms based on trapdoor functions. According to Diffie and Hellman (1976), the two main drawbacks of encryption schemes based on trapdoor functions are: Inverting may be easy for plaintexts for some special form, like always encrypting the messages 1 and 0 to themselves and it could be easy to compute at least partial information of the plaintext. Furthermore, for a deterministic scheme it is easy to detect if a message is sent twice.

In this section, we discuss some of the trapdoor functions employed in data encryption schemes. So far it's not known whether these functions are indeed one way but research has shown that there is no efficient inverting algorithm for any of them unless one has partial information or the trapdoor.

2.1 Multiplication and factoring

The function *f* takes as inputs two prime numbers *p* and *q* in binary notation and returns their product. This function can be computed in $O(n^2)$ time where *n* is the total length (number of digits) of the inputs. Inverting this function requires finding the factors of a given integer *N*. The best factoring algorithms known run in $O\left(2^{(\log N)^{1/3}(\log\log N)^{2/3}}\right)$ time, which is only Pseudo-polynomial in $\log N$, the number of bits needed to represent *N*. This function can be generalized by allowing *p* and *q* to range over a suitable set of semi-primes. Note that *f* is not one-way for arbitrary *p,q*>1, since the product will have 2 as a factor with probability 3/4.

2.2 RSA function (Modular exponentiation)

RSA is a public key algorithm invented by Rivest, Shamir and Adleman (1978). The key used for encryption is different from (but related to) the key used for decryption.

The algorithm is based on modular exponentiation. Numbers *e*, *d* and *N* are chosen with the property that if *A* is a number less than *N*, then $(Ae \bmod N)d \bmod N = A$.

2.3 The Rabin function (modular squaring)

The Rabin function, or squaring modulo $N = pq$, where p and q are primes is believed to be a collection of one-way functions. We write $Rabin_N(x) \triangleq x^2 \bmod N$ to denote squaring modulo N: a specific member of the Rabin collection. It can be shown that extracting square roots, i.e. inverting the Rabin function, is computationally equivalent to factoring N. Hence it can be proven that the Rabin collection is one-way if and only if factoring is hard. This also holds for the special case in which p and q are of the same bit length. The Rabin Cryptosystem is based on the assumption that this Rabin function is one-way (1979).

2.4 Discrete exponential and logarithm (Elgamal)

The function *f* takes a prime number *p* and an integer *x* between 0 and *p*−1; and returns the remainder of 2^x divided

by p. This discrete exponential function can be easily computed in time $O(n^3)$ where n is the number of bits in p. Inverting this function requires computing the discrete logarithm modulo p; namely, given a prime p and an integer y between 0 and $p-1$, find x such that $2^x = y$. There is no published algorithm for this problem that runs in polynomial time up to date. The Elgamal Data encryption scheme is based on this function (Elgamal 1984)

2.5 Hash functions

There are a number of Cryptographic hash functions that are fast to compute like MD5. MD5 is a hashing algorithm that takes a message of up to 264 bits and reduces it to a digest of 128 bits (16 bytes).The algorithm is a development of the MD4 algorithm. Unfortunately, MD4 was flawed, so Rivest made some revisions, and the resulting algorithm was named MD5.Any hashing algorithm should be such that, given a digest and the corresponding message from which it was derived, it should be computationally infeasible to construct a different message with the same digest. Some of the simpler versions have fallen to sophisticated analysis, but the strongest versions continue to offer fast, practical solutions for one-way computation (Rivest et al 2009).

2.6 Subset sum problem (Naccache-Stern Knapsack encryption scheme).

This was an early suggestion but it turned out to be unsuitable. Other one-way functions have been based on the hardness of the decoding of random linear codes (Martello et al, 1990)

3. Probabilistic approach

Probabilistic public-key data encryption scheme was invented by Goldwasser and Micali. They used the predicate "is quadratic residue modulo composite *n"*. In their scheme, every message had many possible encodings and every bit of a message is encrypted independently. Due to this last property, this scheme is not workable according to Georg J. Fuchsbauer. If k is the security parameter (e.g. the size of the modulus in the RSA encryption function) then each bit is encoded individually by a k-bit long string and even worse, resulting in at least a k-bit data expansion factor.

In 1999, T. Okamoto, S. Uchiyama and E. Fujisaki of Nippon Telegraph and Telephone (NTT) in Japan, developed EPOC (Efficient probabilistic public key scheme) based on the random oracle (a theoretic black box), a mapping of every possible query to a random response from its output domain. The primitive encryption is the OU (Okamoto-Uchiyama) function, in which to invert the OU function is proven to be hard as factoring a composite integer.

4. The quadratic residuacity problem

Given a composite integer $n = pq$ and $a \in Z_n^*$ with $\left(\frac{a}{n}\right) = 1$, decide whether or not a is a quadratic residue modulo n. There is no efficient procedure known for solving the quadratic residuacity problem if the factorization of n is unknown. This problem is based on the Quadratic residuacity assumption which states that for sufficiently large primes p and q for real-life algorithm it is infeasible to solve Quadratic Residuacity Problem, but if the factorization of $n = pq$ is known , it is easy to solve QRP by computing $\left(\frac{a}{p}\right)$, since a is a pseudo square if and only if

$\left(\frac{a}{p}\right) = \left(\frac{a}{q}\right) = -1$. Our encryption scheme is based on the function that maps elements of Z_n^* to quadratic residues modulo n. (Hall, 2003)

6.1 Results: The Probabilistic Data Encryption Scheme

The Quadratic-Residue generator (function) therefore is an efficient pseudorandom number generator. This allows for the construction of an efficient probabilistic encryption scheme as follows:

The encryption algorithm (E_n);

1. Set $E_n \leftarrow f_n$
2. E_n is operating in the message space Z_n^* where $n = pq$ is a *Blum* integer where p, q are kept secret.
3. The encryption of $x \in Z_n^*$ of binary length $1 + \lfloor \lg x \rfloor$ bits is $E_n(x) = x^2 \pmod n$, where x is a quadratic residue. Thus the cipher text $C=E_n(x)$, the plaintext x is random.

The decryption algorithm (D_n);

Given an encrypted message, $C = E_n(x)$, the plain text is reconstructed as follows:

1. Two integers $a, b \in Z_n^*$ are picked at random.
2. Apply to $ax \pmod n$ and $bx \pmod n$ (the parity algorithm is assumed at this point to give correct answers). Even though neither $ax \pmod n$ nor $bx \pmod n$ is clearly known, we can manipulate them via their encryption.
3. When gcd procedure terminates , we get a representation of gcd ($ax \pmod n$, $bx \pmod n$) in the form of d and E_n ($dx \pmod n$)
4. If $ax \pmod n$ and $bx \pmod n$ are relatively prime, then $dx \pmod n = 1$. Since $E_n(x)=1$
5. $x \leftarrow \pm d^1 \pmod n$
6. return x

The key generator (K);

1. Select two large primes p and q both congruent to 3 modulo 4
2. Set $n \leftarrow pq$, *a Blum* integer
3. Let f_n be one-way function
4. The public key is (n, f_n), the private key is (p,q)

6.2 Example

Let $p = 11$, $q = 19$ and $m = 3$ (where m is the message.) We can expect to get a large cycle length for those small numbers, because $gcd(\varphi(p-1, \varphi(q-1)) = 2$ for the output (in bits). The generator starts to evaluate x_0 by using $x_{-1} = m$ and creates the sequence $x_0, x_1, x_2, \ldots x_5 = 9, 81, 82, 36, 42, 92$. The following table shows different bit selection to determine the output.

Table 6.1

Even parity bit	Odd parity bit	Least significant bit
011010	100101	110000

7. Conclusion and Recommendations

The Pseudorandomness of the Quadratic-Residue generator arising from computational complexity of random numbers allows the construction of an efficient encryption scheme. Assuming the hardness of the quadratic residuacity problem, this proposed scheme is semantically secure as the Goldwasser-Micali encryption: $x \in \mathbb{Z}_n^*$ is picked at random, x^2 is a random quadratic residue and x^2 is a random pseudosquare modulo n. So in order to decrypt a block of bits of the cipher text, an attacker would have to solve the quadratic residuacity problem.

The authors suggests that further investigations be done on how many bits of the integers produced at each iteration and what length of the parameter of the Quadratic -Residue generator can be out put such that all the statistical tests are passed. This is because statistical analysis shows that if the number of bits is too small the cipher text may be vulnerable to attacks. One such attack involves simple frequency analysis of cipher text blocks. This may thwarted by use of mode operation. However, choosing too large a value of bits may create difficulties during implementation but this is solved by the Pseudo randomness of the Quadratic-Residue generator.

Authors' contributions

All authors contributed to the conceptualisation of the paper. Wanambisi A.W. did the initial review, the selection of abstracts, and the identification of papers to be included in the final review. All authors contributed to the assessment of papers. All authors reviewed the results of the analysis. Wanambisi drafted the manuscript, and all authors contributed to its completion.

Acknowledgements

Thanks to those who have been instrumental in the success of this research: The Masinde Muliro University of Science and Technology, the adviser, for participating in this research study and for their support of this study.

References

S.GOLDWASSSER, S. MICALI, *Probabilistic Encryption*, Journal of Computer and Systems Sciences, **28** PP.270-279, 1984.

W. DIFFIE, M. HELLMAN, *New Directions in Cryptography*, IEEE Transactions on Information Theory, IT-22(6),PP.664-654, 1976

A.J. MENEZES, P.C. VAN OORSCHOT, S.A. VANSTONE, *Handbook of Applied Cryptography*, CRC Press, Boca Raton, FL, USA. 1997

R. RIVEST, A. SHAMIR, L.ADLEMAN, *A method for Obtaining Digital Signature and public key cryptosystems*, communications of the ACM, **21(2)**, PP. 120-126, 1978

A.C YAO, *Theory and applications of trapdoor junctions*, proc. 23rd IEEE symposium. Computer science,1982, PP. 458-463.

S. BRANDS, R. GILL, *Cryptography, Statistics and Pseudorandomness* II probability and mathematics statistics, volume **16**. Fasc. 1 (1996), PP. 1-17.

GEORG J. FUCHSBAUER, *An Introduction to Probabilistic Encryption*, Osjecki matematicki list **6** (2000), PP. 37-44.

M. BEN-OR CHOR and A. SHAMIR, *On the cryptographic security of single RSA bits*, Proc. 15th ACM Symp. Theory of Comp., 1983, PP. 421-430.

MARTELLO, SILVANO; TOTH, PAOLO (1990). "4 Subset-sum problem". *Knapsack problems: Algorithms and computer interpretations*. Wiley-Interscience. pp. 105–136. ISBN 0-471-92420-2. MR 1086874

CORMEN,RIVEST, RONALD L.; STEIN, CLIFFORD (2009). *Introduction to Algorithms* (third edition ed.). MIT Press. ISBN 0-262-03384-4.

This academic article was published by The International Institute for Science, Technology and Education (IISTE). The IISTE is a pioneer in the Open Access Publishing service based in the U.S. and Europe. The aim of the institute is Accelerating Global Knowledge Sharing.

More information about the publisher can be found in the IISTE's homepage: http://www.iiste.org

CALL FOR PAPERS

The IISTE is currently hosting more than 30 peer-reviewed academic journals and collaborating with academic institutions around the world. There's no deadline for submission. **Prospective authors of IISTE journals can find the submission instruction on the following page:** http://www.iiste.org/Journals/

The IISTE editorial team promises to the review and publish all the qualified submissions in a **fast** manner. All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Printed version of the journals is also available upon request of readers and authors.

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar