ENFORCING PRIVACY BY WITHHOLDING PRIVATE INFORMATION

Frans Lategan
Department of Computer Science, Rand Afrikaans University, PO Box 524, Auckland Park, 2006, South Africa
fransl@discoveryhealth.co.za

Martin S. Olivier
Department of Computer Science, Rand Afrikaans University, PO Box 524, Auckland Park, 2006, South Africa
molivier@rkw.rau.ac.za

Abstract Privacy of information is becoming more and more important as we start trusting unknown computers, servers and organisations with more and more of our personal information. Thus far, no reliable and practical method to enforce privacy has been discovered. Often a set of private information has to be supplied simply to enable the recipient to verify that one member of the set is correct given the other methods. An income tax return is an example where such information has to be supplied simply to verify taxable income. The object of this paper is to consider mechanisms to safeguard our private information in cases where this information is required not for the contents, but as input to verify calculations. We shall present an encryption method to protect private information where the private information consists of a set of numeric values S on which some function G has to be applied and the result $\alpha = G(S)$ has to be supplied to a target organisation. The result α must be verifiable by the target organisation, without disclosing S. We apply this method to the specific case of protecting the privacy of electronic income tax returns, and discuss other possible applications.

1. INTRODUCTION

There are two sides to Privacy. It is something a lot of people are willing to sell very cheaply in certain instances. (We gladly pay for our purchases with credit cards for the convenience of not having to carry cash with us. The down side is that a complete record of our every purchase could be kept in a database somewhere and might be available for all to browse). In other cases, it is something that we are not willing to compromise on lightly (such as our exact

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: 10.1007/978-0-387-35515-3_53

S. Qing et al. (eds.), *Information Security for Global Information Infrastructures*
© IFIP International Federation for Information Processing 2000

earnings) or even not at all (such as the medical records of an AIDS sufferer or a past criminal record).

Privacy is an abstract concept. For our purposes, we shall define privacy as a state that exists when access to private information about a particular individual can be effectively controlled and managed by that individual even after a third party has collected such private information. The aim of privacy is not to prevent the use or collection of private information, but rather the misuse (intentional or not) thereof.

In our daily lives we constantly need to divulge part of this private information to do business. This is the way it has always been, and no-one has complained about privacy in the past, but with the advent of networked computerised files things have changed somewhat. When our private information is stored in a hard copy format where very few people have physical access to it, it is reasonably secure. However, the same private information can be several orders of a magnitude more vulnerable to unauthorised access when stored electronically, should it not be protected.

These days private information is stored in electronic files, and those files can be sold to advertisers, placed on the Internet for public viewing, hacked into by malicious hackers or even browsed by curious employees. The latter is particularly hard to prevent.

The purpose of this paper is to propose a method to safeguard our private information in cases where this information is required not for the contents, but as input to verify calculations. We shall present an encryption method which allows certain mathematical operations to be performed on a set of encrypted values and to cross check this result with the result obtained by performing the same operations on the unencrypted values.

More formally, this method is applicable to cases where the private information consists of a set of numeric values S on which some function G has to be applied and the result $\alpha = G(S)$ has to be supplied to a target organisation t. The result α must be verifiable by the target organisation t, but we would like to keep S private. A tax form submitted to the IRS is a very well known practical application hereof.

This method protects the private information against unauthorised access from outside the organisation it was sent to. It does not depend on the organisation to protect the data adequately and it also prevents people in the organisation with valid access to browse this information out of curiosity.

This paper is structured as follows: In Section 2 we shall supply some background information. This will be followed by a description of the problem in Section 3. Our proposed solution will be explained in Section 4. In Section 5 we shall discuss some further applications and the conclusion and areas for further study will be explained in Section 6.

2. BACKGROUND

There is no escaping the fact that sensitive information about every person is stored in one or more databases somewhere that that person would like to disappear, or not be accessible without his or her explicit permission. These databases are increasingly compromised, misused, sold, or even made freely available to the public over the Internet — even when the government controls them [8]. In [5] Samarati gives a good summary of the increasing concerns of privacy on the Internet; other agree — see also [4, 7, 14].

Currently privacy can be safeguarded in one of three ways, according to Cranor in [7]:

1 Private information is not disclosed at all.

2 The source of the private information is hidden, that is, anonymity is preserved.

3 Privacy policies are in effect that promise the responsible usage of private information.

The use of privacy policies is becoming more popular on Internet sites these days. In some countries such as the United Kingdom legislation has already been passed regulating the use of personal data a company might possess about an individual, and action plans are being drawn up to comply with it (see [6]). The European Union's data protection directive goes as far as to empower national EU data regulators to halt exports of personal data to countries which they judge not to have adequate protection requirements, and this might start affecting trade with the United States, as discussed in [10]. The United States is still lagging behind the European Union as far as legislation is concerned, but is making progress (see [1], [2] and [3]).

The adoption of privacy policies might seem effective, but it still does not guarantee the safe keeping of private information; it is just the first step. What is required is an algorithmic solution, or a model, to ensure privacy.

We propose a way to apply the first of Cranor's options to a class of applications where previously the third option was the best alternative, by using homomorphic encryption algorithms. Although many researchers have mentioned or devised homomorphic encryption algorithms [11, 12, 13], we believe that there is still scope to combine such algorithms and apply them to protect privacy.

3. PROBLEM DESCRIPTION

Let's start by restating our definition of privacy.

Definition 3.1 *Privacy is defined as the state that exists when access to private information about a particular individual can be effectively controlled and*

managed by that individual even after a third party has collected such private information.

3.1. PROBLEM STATEMENT

In certain cases private information pertaining to a particular individual is not used directly; it is collated or joined in some way, giving an end result which is then used. More formally, information attributes $x_1, x_2, x_3, \ldots, x_n$ about a subject s is collected from a number of sources to which some function G is applied to give a result α where $\alpha = G(x_1, x_2, x_3, \ldots, x_n)$. The result α is then passed on to a target t.

What makes problems of this type interesting, and indeed what allows us to protect the private information, is the fact that although the target t needs to verify the calculation of α, t does not strictly require access to $x_1, x_2, x_3, \ldots, x_n$.

Our problem is then to find a way of proving to t that $\alpha = G(x_1, x_2, x_3, \ldots, x_n)$ without disclosing $x_1, x_2, x_3, \ldots, x_n$. This would enable us to retain exclusive control over our private information.

3.2. THE IRS EXAMPLE

Every year we supply the IRS with a lot of very private information about ourselves. The sole purpose of supplying them with this information is to allow them to verify that the amount of tax payable by us as stated on our tax returns has been correctly calculated.

In the terminology of our problem statement, the IRS is t and the taxable income is represented by α. The taxable income α is calculated by adding various amounts representing income, and subtracting other amounts representing allowable deductions. We shall use $x_1, x_2, x_3, \ldots, x_n$ to represent the various amounts that are used in the tax calculation, which we shall denote with the function $G(x_1, x_2, x_3, \ldots, x_n)$.

In the traditional or unsecured tax return, the IRS is now supplied with α as well $x_1, x_2, x_3, \ldots, x_n$ to enable them to verify α using G which is known to all parties.

4. PROPOSED SOLUTION

In order to keep our private information secure, we propose to encrypt $x_1, x_2, x_3, \ldots, x_n$ using a function f such that $y_i = f(\epsilon, x_i)$ where ϵ is the encryption key. The result of the encryption will give us $y_1, y_2, y_3, \ldots, y_n$. Note that most of these values are usually supplied by third parties such as employers or banks, and that they can use these encrypted values on the tax certificates issued by them instead of the original values. This prevents the tax payer from modifying these values.

If we submit our tax returns using $y_1, y_2, y_3, \ldots, y_n$ instead of $x_1, x_2, x_3, \ldots, x_n$ we can protect a lot of sensitive information, but the IRS loses the ability to verify α.

What we now require is a function G' such that $G'(y_1, y_2, y_3, \ldots, y_n) = \alpha'$ where $\alpha = f(\epsilon, \alpha')$. This would allow the IRS to verify the correctness of α using the encrypted values $y_1, y_2, y_3, \ldots, y_n$ without ever knowing $x_1, x_2, x_3, \ldots, x_n$ by applying the function G' to $y_1, y_2, y_3, \ldots, y_n$ to obtain α' and then apply function f to α' and comparing the results. If $f(\alpha') = \alpha, \alpha$ has been correctly calculated from $x_1, x_2, x_3, \ldots, x_n$.

4.1. RESTRICTIONS

It is clear that we have to supply t with f and ϵ, which rules out symmetric encryption as a possibility for f. Furthermore for a given mapping $f(x_1) = y$ it should be hard or impossible to find x_2 such that $f(x_2) = y$. The function f would still protect the private information without this restriction, but then a possibility of fraud exists: assume $(\exists x_1, x_2)(x_1 \neq x_2)(x_1, x_2 \neq 0)$ such that $f(x_1) = f(x_2) = y$ with x_2 easy to find; then the sender could replace x_1 in all calculations with x_2 without affecting the result. To prevent this, we therefore assume that f has to map one and only one x to each y. We can summarise these restrictions more formally.

Suitable functions for f, G and G' would be functions that have the following properties:

Property 4.1 *The function $f(\epsilon, x) = y$ should obscure x so that there is no easy way to determine x given f, ϵ and y.*

Property 4.2 *If $f(\epsilon, x_1) = f(\epsilon, x_2) = y$, then $x_1 = x_2$*

Property 4.3 $G'(f(\epsilon, x_1), f(\epsilon, x_2)) = f(\epsilon, G(x_1, x_2))$

It would also be convenient, but not a requirement, if G and G' can be the same functions.

4.2. FUNCTION DEFINITIONS

We shall define suitable functions for f based on the type of operation that we want to perform on x and then prove that each function satisfies the properties stated above.

Multiplication. If the function G to be performed is normal multiplication, we define f to be f_* where f_* is RSA encryption as defined in [11]. We use p and q to denote the primes, and N to be their product. We shall define ϵ to be the public key, and k to be the private key. Thus $f_*(x) = x^\epsilon \bmod N$ and α' is obtained by $\alpha' = y_1 * y_2 * y_3 * \ldots * y_n \bmod N$.

Theorem 1 *f_* satisfies the properties defined in Section 4.1 where G = is multiplication and G' is multiplication mod N.*

Proof:

1 This is trivial, from the usage of RSA for f.

2 This will hold if the modulus N used for the RSA encryption is larger than the largest possible x to be encrypted.

3 This follows from the homomorphic property of RSA:
$((x_1{}^{\epsilon} \bmod N) * (x_2{}^{\epsilon} \bmod N)) \bmod N$
$= (x_1{}^{\epsilon} * x_2{}^{\epsilon}) \bmod N$
$= (x_1 * x_2)^{\epsilon} \bmod N$

◇

Note that the final product α of all the unencrypted values has to be less than N. In practice N would be at least a 768 bit number, the current minimum value for reasonably secure RSA encryption.

Addition. If the function G to be performed is normal addition, we use a new public key cryptosystem as described in [12] for f. We define f to be f_+ where $f_+(\epsilon, x) = \epsilon^x \bmod N$ and ϵ is a generator for GF(N). N is the product of two large primes p and q as for RSA. For a full description the reader is referred to [12].

Theorem 2 *f_+ satisfies the properties defined in Section 4.1 where G = is addition and G' is multiplication mod N.*

Proof:

1 If we examine $f = \epsilon^x \bmod N$ it is clear that the encryption can easily be reversed if N can be factored to create the decryption key, or if x can be recovered by solving the discrete logarithm. Since factoring large numbers and solving large discrete logarithms are unpractical using current technology, we argue that f_+ satisfies the first condition.

2 This will hold if N is larger than the largest possible x to be encrypted.

3 This follows from the homomorphic property of the cryptosystem:
$((\epsilon^{x_1} \bmod N) * (\epsilon^{x_2} \bmod N)) \bmod N$
$= (\epsilon^{x_1} * \epsilon^{x_2}) \bmod N$
$= \epsilon^{x_1+x_2} \bmod N.$

◇

Note that the final sum α of all the unencrypted values has to be less than N. In practice N would be at least a 768 bit number, the current minimum value for reasonably secure RSA encryption.

Comparison. Comparisons between encrypted values seem impossible, since ordering is lost through encryption, and if this were not the case, the encryption would be severely weakened. Assume an encryption function g preserves ordering, and that $y = g(x)$ is an encrypted value. Since we have to supply t with g, t can encrypt as many constants with g as needed, and a binary search will quickly determine the value of x. For a tax calculation, this can fortunately be circumvented by disclosing the intermediate values that require comparison. This will reduce the privacy of the tax return slightly, but this will still be an improvement over the unprotected return. See the appendix, where the total interest calculation is disclosed, but the various sources of interest are hidden for a concrete example.

The above then allows us to use this approach in a simplistic tax return, where various amounts are added and subtracted or multiplied and divided to calculate a taxable income. See the appendix for a (fictitious) real world example using South African tax law.

4.3. SECURITY AND OTHER ISSUES

Some issues need to be taken into consideration when applying this method. We mention the more prominent ones.

Exhaustive searches. A possible problem with this method is that the exhaustive searches required to decrypt $y_1, y_2, y_3, \ldots, y_n$ would not take too long using computers, since the values of $x_1, x_2, x_3, \ldots, x_n$ are reasonably small compared to the computing power available today (usually less than 10^6, and almost always less than 10^{10}). The magnitude of this problem can be reduced by using a very big number of significant digits after the decimal point, and randomly adding a very small number (less than 1 cent) to each amount before the encryption function f is applied. This will not modify the taxable amount in any significant way, but should protect the amounts.

Decimal amounts. Although it seems that decimals can not be used due to the use of the modulus function, this can easily be overcome by only using cents for amounts (or millicents, picocents, etc. as applicable to prevent exhaustive searches).

5. OTHER PROBLEMS TO WHICH THIS SOLUTION IS APPLICABLE

This solution is not just suitable to protect the privacy of electronic tax forms. Many other examples exist, such as the ability to modify encrypted database values without decryption, secret e-mail voting with verifiable results, verifying directors' remuneration, etc. We describe some in more detail.

In a secret, verifiable e-mail vote, a trusted party (the vote counter) generates the encryption and decryption keys. The encryption keys are distributed to the voters. Each voter then submits his encrypted vote. We assume a yes or a no vote, represented by 1 and 0 respectively. To prevent all the encrypted values from being the same when the vote is the same, we can increase the magnitude. (As a trivial example for 10 voters, a yes vote would be a random number between 9950 and 10050, an a no vote would be a random number less than 100).

In practice a yes vote can be defined a 1 followed by n 0's added to a random number m. A no vote would just be a random number m. Since n can be arbitrarily chosen, the vote counter can increase the security of the system by increasing n. If m is relatively small compared to n, the random numbers should not significantly change the overall vote. The vote counter adds all the encrypted votes together, and decrypts the result. The vote counter then publishes the encrypted votes and the decrypted result. Everyone can now verify the calculations and ensure that their vote was counted, without compromising anyone else's vote.

When a company is audited, the total directors' remuneration must be disclosed on the financial statements, although their individual remuneration may be kept secret (according to South African law). The problem is that the total amount has to be audited, and the directors have to disclose their individual remuneration to the auditors. This approach can be used to prove that the total has indeed been calculated correctly without disclosing the individual amounts.

6. CONCLUSION AND AREAS FOR FURTHER STUDY

In our paper we have presented a novel approach to the protection of electronic data privacy: that of algorithmic protection of privacy using encryption, which allowed us to use our private information (in this case details of income) but while still protecting our private information from prying eyes, including the party to which we sent our private information (at least in a simplified tax example).

More problems were this type of solution is applicable could be researched, and the solution could then be expanded to include more mathematical functions.

It would also be quite useful if a single encryption function that is homomorphic for both addition and multiplication could be designed. This does not have to be a public key function; a one way encryption function could also be suitable. Such a function would transform this solution from a theoretical solution into a practical, easily implementable solution, and could make a significant contribution to electronic privacy.

Appendix

An fictional example using the South African tax law. For this example we shall set $N = 19697446673$ and $\epsilon = 131$. All the **bold figures** can be kept private; the others, as well as all the encrypted figures are to be supplied to the IRS.

1. CALCULATION

	Description	*Clear*	*Encrypted*	**A***dd* **M***ultiply*
	Salary	**110000**	18182403837	A
less	Pension Fund Contribution	**-9000**	2608534625	A
plus	Car Allowance	**40000**	7122548528	A
plus	Taxable Interest	7365	9973525707	A *From 2*
less	Travel expenses	**-26136**	847550577	A *From 3*
	Total taxable income	122229	6922828699	

Note that $18182403837 * 2608534625 * 7122548528 * 9973525707 * 847550577 \bmod 19697446673 = 6922828699$

2. INTEREST SUB-CALCULATION

	Description	*Clear*	*Encrypted*	**A***dd* **M***ultiply*
	Unit Trusts	**3556**	5403644383	A
plus	Bank account	**345**	15350080410	A
plus	Fixed Deposit	**5432**	11953675544	A
plus	Share trading account	**32**	11140409637	A
	Total	9365	4224346997	

Note that the product of the encrypted values mod $N = 19697446673$ is 4224346997, the same as the encrypted value for 9365. We now disclose the total amount, since tax is not payable on the first 2000.

3. TRAVEL EXPENSES SUB-CALCULATION

	Description	*Clear*	*Encrypted*	**A***dd* **M***ultiply*
	Fixed allowance/1000 km	**1022**	4201531621	A
plus	Fuel allowance/1000 km	**226**	7357125748	A
plus	Maintenance allowance/1000 km	**204**	15157866258	A
	Total per 1000 km	1452	18998024939	
	Total per 1000 km	1452	7632560095	M
times	Business km travelled (in 1000)	**18**	10205764354	M
	Total deduction	26136	12475071519	M

Note that the totals can be verified as above.

References

[1] J. Canavan, "Information Privacy: It's Your Business", *Telecommunications*, November 1998, 38.

[2] C. J. Dorobek, "Gore proposes an Electronic Bill of Rights to protect the privacy of personal data.", *Government Computer News*, August 24, 1998 v17 n27, 16.

[3] J. L. Guerra, W. Jackson, C. J. Dorobek and F. Olsen, "Clinton signs privacy order (President Clinton signs executive order to ensure that new technology does not erode privacy of US citizens)", *Government Computer News*, June 15, 1998 v17 n17, 3.

[4] H. Wang, M. K. O. Lee and C. Wang, "Consumer privacy concerns about Internet marketing.", *Communications of the ACM*, March 1998, 63–70.

[5] B. Thuraisingham, S. Jajodia, P. Samarati, J. Dobson and M. Olivier, "Security and Privacy Issues for the World Wide Web: Panel Discussion", *Database Security XII: Status and Prospects*, Kluwer, To Appear.

[6] C. Pounder, "An Initial Plan of Action for Compliance with the UK's Data Protection Act 1998", *Computers & Security*, February 1999, 23–27.

[7] L. F. Cranor, "Internet Privacy", *Communications of the ACM*, February 1999, 29–31.

[8] W. Madsen, "IRS Fails to Provide Adequate Security, Privacy Controls", *Computer Fraud & Security*, February 1999, 8.

[9] K. Q. Nguyen, V. Varadharajan and Y. Mu, "Batching Proofs of Knowledge and its Applications", DEXA99 International Workshop on Electronic Commerce and Security, Florence, September 1999.

[10] "EU, USA make little progress with data protection", *Computer Fraud & Security*, February 1999, 3.

[11] J. Seberry and J. Pieprzyk, "Cryptography: An Introduction to Computer Security", Prentice Hall, 1989.

[12] D. Naccache and J. Stern, "A New Public Key Cryptosystem Based on Higher Residues.", *Proceedings of the 5th ACM conference on Computer and Communications Security*, 1998, 59–66.

[13] N. Ahituv, Y. Lapid and S. Neumann, "Processing Encrypted Data.", *Communications of the ACM*, September 1987, 777–780.

[14] R. Clarke, "Internet Privacy Concerns Confirm the Case for Intervention.", *Communications of the ACM*, February 1999, 60–67.