Securing Harari's Authentication Scheme

Behzad Malek[1] and Ali Miri[2]
(Corresponding author: Behzad Malek)
School of Information Technology and Engineering, University of Ottawa[1]
800 King Edward Ave., Ottawa, ON, Canada[1]
(Email: bmalek@site.uottawa.ca)
Department of Computer Science, Ryerson University, Toronto, ON, Canada[2]
(Email: ali.miri@ryerson.ca)
(Received June 17, 2010; revised and accepted Dec. 17, 2010)

Abstract

Code-based authentication schemes are very fast and efficient for implementation in small devices. The security of these schemes and the size of underlying coding matrices are often the main barriers to their use in practice. In this paper, we remove the security flaw of a classical code-based authentication scheme and modify it to achieve a secure code-based authentication scheme. We provide a formal proof of security for the modified scheme. Moreover, we apply the recent improvement to reduce the size of the coding matrix and to gain an efficient authentication scheme suitable for small devices.

Keywords: Authentication, coding, Harari, security, zero-knowledge

1 Introduction

Code-based cryptography is very efficient and high speed in implementation when compared with algebraic cryptographic scheme. It is highly recommended to design authentication schemes based on code-based cryptography, where resources are scarce in hardware [3]. Nevertheless, the size of encoding matrices that have to be used to ensure a minimum level of security is usually very large, imposing a huge overload on a given device. Recently, Berger et al. [1] showed that it is possible to use smaller matrices in code-based cryptography and have a comfortable level of security. This reduction allows most of the code-based authentication schemes [4, 6, 8, 9, 11], which had been marginalized in the literature due to their computation/storage complexities, to be reconsidered for use in small devices. In this work, we revisit one of the classical works by Harari [9] that has inspired many other code-based authentication schemes [5, 6, 7, 11]. Harari's authentication scheme is a special authentication method, as it is a *zero-knowledge system*. Informally, in a zero-knowledge authentication scheme, the verifier can readily verify that the prover has the correct credentials to pass the authentication, but it does not learn anything about the prover's credentials. It will be shown later in this work that recent improvements in reducing the size of encoding matrices make Harari's scheme more efficient than other authentication schemes, e.g. [4, 10, 11, 13]. However, Harari's initial design is proved to be insecure due to a flaw in the scheme, i.e. checking conditions of the authentication process [12]. In this work, we modify Harari's scheme to remove the flaw and provide a formal proof of security for the proposed authentication scheme based on the difficulty of a known complexity assumption.

1.1 Complexity Assumption

Linear codes have to possess certain complexity properties before they can be used in cryptography. The security of our scheme is based on the complexity of an NP-Hard problem, referred to as the Syndrome Decoding (SD) problem.

Definition 1 (Syndrome Decoding Problem). *Let H be a parity-check matrix of a binary code $[N, k]$. Suppose d is the syndrome of a vector s. Suppose that p is a given integer in the space of all possible weights of vector s. The question is if one can easily find a vector s of length N, such that $Hs^T = d$ and $w(s) \leq p$, where $w(\cdot)$ denotes the weight of a vector.*

It has been proven in [2] that decoding an arbitrary linear code is an NP-Hard problem. The security of many coding-based cryptographic protocols are based on the NP-Hardness of the SD problem.

1.2 Harari's Identification Scheme

Harari's scheme [9] is given as follows: a trusted center chooses a random binary code of length N dimension k. It is recommended that $N \geq 2000$ and $k \geq 1000$. The generator of the code is denoted by a $k \times N$-matrix G (or equivalently by its parity-check $(N-k) \times N$-matrix H). Let s be a random *codeword* of weight μ, where μ

is a small, odd quantity chosen at random in the interval $[50, 100]$. Note that since s is a codeword, then $Hs^T = \mathbf{0}$. User A will use s as her secret key in the authentication rounds. Notations used in Harari's scheme [9] and their application are summarized in Table 1.

The trusted center publicizes the generator matrix G or the parity-check matrix H, as well as μ. Note that s is kept private. The following protocol would allow A to authenticate herself to B:

1) A sets $l \in [100, 200]$ and chooses l random binary vectors, $r_i \in \mathbb{F}_{2^N}$ for $i = 1, \cdots, l$, where the supports of the vectors are disjoint. We denote the weight of r_i by w_i. Then A computes the syndromes $t_i = Hr_i^T$, where r_i^T is the transpose of vector r_i.

2) B receives the set $\{(w_i, t_i)|i = 1, \cdots, l\}$ from A and chooses an odd weight binary vector e of length l, where $\frac{l}{3} \leq w(e) \leq \frac{2l}{3}$. B returns e to A.

3) A randomly computes a permutation π of $\{1, \cdots, l\}$ and sends it back to B.

4) A and B both compute $t = \pi(e)$.

5) B randomly chooses $b \in \{t, \bar{t}\}$, where $\bar{t}$ is the binary complement of t. Each bit of b is denoted by b_i for $i = 1, \cdots, l$. Then, B sends b to A.

6) A computes $r = \sum_{i=1}^{l} b_i \cdot r_i$ and returns $y = r + s$ to B. Note that the notation $b_i \cdot r_i$ is used to denote the operation that the vector r_i is selected or deselected if $b_i = 1$ or $b_i = 0$, respectively.

7) B checks three conditions to verify A's identity:

   a. First condition on weights:

$$w(y) \neq \sum_{i=1}^{l} b_i \cdot w_i.$$

   b. Condition on syndromes:

$$Hy^T = \sum_{i=1}^{l} b_i \cdot t_i.$$

   c. Second condition on weights:

$$-\mu + \sum_{i=1}^{l} b_i \cdot w_i \leq w(y) \leq \mu + \sum_{i=1}^{l} b_i \cdot w_i.$$

Harari's security analysis only provides a quantitative proof for the security of the proposed authentication scheme [9]. However, this proof was based on the attacks known at the time, and also it did not offer any formal proof of security. Later Véron [12] showed that Harari's initial scheme is insecure. We briefly review the attack in the rest of this section and refer the reader to [12] for full analysis.

1.3 Cryptanalysis of Harari's Identification Scheme

Véron [12] breaks Harari's scheme by finding a flaw in the design of the authentication scheme. The attack is laid out as follows: Let's suppose that $\pi(e) = (1, \cdots, 1, 0, 0, \cdots, 0)$ and $w(e) = \frac{l}{3} + q$ is odd and $0 \leq q \leq \frac{l}{3}$. It is possible for a dishonest prover to build three sequences $(w_i \in \mathbb{Z})_{1 \leq i \leq l}, (t_i \in \mathbb{F}_{2^{(N-k)}})_{1 \leq i \leq l}$ and $(x_i \in \mathbb{F}_{2^N})_{1 \leq i \leq l/3}$, where x_i's are vectors of length N such that:

$$\begin{cases} w(x_j) & = & \sum_{i=1}^{\frac{l}{3}+q} w_i \\ w(x_j) & \equiv & \frac{l}{3} + j \mod 2 \\ Hx_j^T & = & x + \sum_{i=1}^{\frac{l}{3}+j} t_i. \end{cases} \quad (1)$$

We have set x in Equation (1) as $x = Hz^T$ for a random vector z of weight μ. Note that μ, w_i's and t_i's are publicly known. Now let's suppose that the dishonest prover sends $y = x_q + z$ to B. It can be shown that this vector satisfies all three conditions of authentication, since:

1) First condition on weights is satisfied because:

$$w(y) \neq \sum_{i=1}^{\frac{l}{3}+q} w_i = \sum_{i=1}^{l} b_i w_i.$$

2) Condition on syndromes is satisfied because:

$$Hy^T = Hz^T + Hx_q^T = x + (x + \sum_{i=1}^{\frac{l}{3}+q} \cdot t_i) = \sum_{i=1}^{l} b_i \cdot t_i).$$

3) Second condition on weights is satisfied because:

$$-\mu + w(x_q) \leq w(y) \leq \mu + w(x_q)$$

and $w(x_q) = \sum_{i=1}^{\frac{l}{3}+q} w_i = \sum_{i=1}^{l} b_i \cdot w_i$.

Knowing $w(e)$, the adversary can easily build such x_i vector that satisfy the above conditions. For more details on how to find x_i's, the reader is referred to [12]. Therefore, the dishonest prover could pass the identification protocol on average every second time that the protocol is executed. As it can be seen from Equation (1), the breach in Harari's identification scheme is the result of using a codeword as the secret vector s. In Harari's original scheme, since $s \in \mathcal{C}$, we have $Hs^T = 0$. This helps the adversary (the dishonest prover) in Véron's attack to pick any random vector z of weight μ instead of s, as its syndrome (x) will be canceled by the syndrome of x_q when checking the condition on syndromes.

Véron [12] recommended to repeat Harari's scheme n times, in order to reduce the probability of success of the previous attack to 2^{-n}. However, this dramatically increase the number of bits exchanged (i.e. transmission rate) between A and B. The prover and verifier's computational overheads are also increased linearly with the number of identification rounds.

Table 1: Summary of notations used in Harari's identification scheme

Notation	Description
A	The prover
B	The verifier
k	Dimension of the binary codes
N	Length of the binary codes
G	Generator matrix of the underlying code
$w(\cdot)$	Weight function
H	Publicly known parity-check $(N-k) \times N$-matrix
μ	Weight of the secret key of the prover
s	Secret key of A, as a codeword of weight μ
r_i	Private, random vector that is kept hidden from B
l	Number of random vectors r_i
w_i	Weight of r_i, i.e. $w_i = w(r_i)$
t_i	Syndrome of r_i, i.e. $t_i = Hr_i^T$
e	Random binary selection-vector of length l
π	A random permutation on e
t	The permutation of e, i.e. $t = \pi(e)$
$\bar{t}$	The binary complement of t
b	The challenge selection-vector $b \in \{t, \bar{t}\}$
b_i	Each bit of b
r	The masking, random vector, s.t. $r = \sum_{i=1}^{l} b_i \cdot r_i$
y	The response vector, s.t. $y = r + s$

2 Securing Harari's Identification Scheme

We propose an adaptation of Harari's scheme that is secure against Véron's attack without increasing the transmission rate or computational overheads. In our scheme, the trusted center publicizes the random parity check matrix H, as before. The notations used in our adaptation are the same as in Harar's original scheme as given in Table 1, but there are a few changes; as the private key of A, we choose a *random vector* $s \in \mathbb{F}_{2^N}$ of weight μ instead of a codeword. Since, s is a random vector, instead of codeword, its has a non-zero syndrome that is denoted by d, where $d = Hs^T$. The trusted center publishes $\{d, \mu\}$ as A's public parameters. The following identification protocol would allow A to authenticate herself to B:

- A chooses l random binary vectors, $r_i \in \mathbb{F}_{2^N}$ for $i = 1, \cdots, l$, where the supports of the vectors are disjoint. Then A computes the syndromes $t_i = Hr_i^T$, where r_i^T is the transpose of the vector r_i.
- B receives the set $\{(w_i, t_i) | i = 1, \cdots, l\}$ and chooses an odd weight binary vector e of length l, where $\frac{l}{3} \leq w(e) \leq \frac{2l}{3}$. B sends e to A. Each bit of e is denoted by e_i for $i = 1, \cdots, l$.
- A computes $r = \sum_{i=1}^{l} e_i \cdot r_i$ and returns $y = r + s$ to B. Note that the notation $e_i \cdot r_i$ is used to denote the operation that the vector r_i is selected/deselected by the bit e_i.
- B checks three conditions to verify A's identity

1) First condition on weights:

$$w(y) \neq \sum_{i=1}^{l} e_i \cdot w_i.$$

2) Condition on syndromes:

$$Hy^T = d + \sum_{i=1}^{l} e_i \cdot t_i.$$

3) Second condition on weights:

$$-\mu + \sum_{i=1}^{l} e_i \cdot w_i \leq w(y) \leq \mu + \sum_{i=1}^{l} e_i \cdot w_i.$$

Note that in checking the condition on syndromes, d is added for verification. An attacker can no longer send arbitrary vectors to the victim as described in the Véron's attack, as adding a random vector x of weight μ might satisfy the conditions on weights, but it would fail the condition on the syndromes. It is shown in the next section that addition of d in the condition on syndromes forces the adversary to send the correct s. We will prove that if the adversary's attack algorithm $\mathcal{A}$ can successfully compromise the proposed authentication mechanism, then the same algorithm can be used to solve the SD problem.

2.1 Security Analysis

Let's suppose that there is an algorithm $\mathcal{A}$ that generates vectors to pass the checking conditions in the proposed

scheme. There is a simulator algorithm $\mathcal{B}$ that tries to find a vector of syndrome d^* and weight less than p, i.e. an instance of the SD problem.

The algorithm $\mathcal{B}$ challenges $\mathcal{A}$ for authentication in the proposed scheme once with d and another time with d', such that $d+d'=d^*$, and it sets the weights respectively to μ and μ', such that $\mu+\mu' \leq p/2$.

Algorithm $\mathcal{A}$ sends $\{(w_i,t_i)|i=1,\cdots,l\}$ in the first challenge and $\{(w'_i,t'_i)|i=1,\cdots,l'\}$ in the second challenge. Algorithm $\mathcal{B}$ sets e and e' such that $\sum_{i=1}^{l} e_i \cdot t_i = \sum_{i=1}^{l'} e'_i \cdot t'_i$ and also $|\sum_{i=1}^{l} e_i \cdot w_i - \sum_{i=1}^{l'} e'_i \cdot w'_i| \leq p/2$. Note that since the vectors r_i and r'_i have disjoint supports, we should have: $\sum_{i=1}^{l} e_i \cdot w_i \leq N$ and also $\sum_{i=1}^{l'} e'_i \cdot w'_i \leq N$. Therefore, the two series of w_i and w'_i are bounded by N, and it is possible to find two close enough series that satisfy:

$$|\sum_{i=1}^{l} e_i \cdot w_i - \sum_{i=1}^{l'} e'_i \cdot w'_i| \leq p/2.$$

The algorithm $\mathcal{A}$, using e and e', sets $r = \sum_{i=1}^{l} e_i \cdot r_i$ and $r' = \sum_{i=1}^{l'} e'_i \cdot r'_i$, and then returns $y = r+s$ and $y' = s'+r'$, respectively. It can easily be shown by following the inequality properties that for the weight series we have:

$$|w(y+y')|-|w(r)-w(r')| \leq |y-w(r)|+|y'-w(r')| \leq \mu+\mu'.$$

Since we have $\mu+\mu' \leq p/2$ and the algorithm $\mathcal{B}$ has set the weights such that $|w(r)-w(r')| \leq p/2$, we can conclude that:

$$|w(y+y')| \leq p.$$

On the other hand, the condition on the syndromes returns:

$$\begin{aligned} Hy^T + Hy'^T &= H(y^T + y'^T) \\ &= (d+d') + (\sum_{i=1}^{l} e_i \cdot t_i + \sum_{i=1}^{l'} e'_i \cdot t'_i) \\ &= (d+d') = d^*. \end{aligned}$$

That is $y+y'$ is a binary vector of syndrome d^* and weight less than p. Thus, the vector $y+y'$ is a solution to the SD problem instance.

2.2 Performance Analysis

In this section, we provide an estimation on the resources that would be required to implement the proposed protocol. Using the recommendations on the size of the parity check matrix in [1], we use a 225×450 parity-check matrix. Let's pick $l \in [50, 100]$ the same as in [9]. In Table 2, we have compared the performance of the proposed scheme to results of Véron [12], pages 266-267, and the newest implementation [4] of Stern's scheme [11]. Note that the prover's computational overhead in Table 2 is approximated by $\mathcal{O}(N^2)$, where N is the size of the codewords.

As it can be seen in Table 2, the proposed scheme outperforms other similar schemes with the same security level. It has a significantly better transmission rate than the newest implementation [4] of Stern's scheme [11] with a comparable matrix size and computational overheads. In Table 2, we did not consider the prover's work to compute a cryptographic hash function. Despite the fact that Stern's scheme [4, 11] has a lower memory complexity, it requires implementation of hash functions that will further increase the complexity of the scheme.

3 Conclusions

Code-based cryptosystems are very promising for small devices mostly due to their simplicity of operations and high speed performance. In this paper, we have removed the security flaw from Hararis scheme and provided a formal proof of security based on the hardness of the SD problem. We have also shown that our proposed scheme outperforms other code-based authentication schemes, while providing the same level of security. The proposed protocol is very practical for resource-constrained devices. It is based on an asymmetric-key algorithm and simplifies the key management in the system. It also provides a superior transmission overhead as compared to other work reported in the literature.

References

[1] T. P. Berger, P. L. Cayrel, P. Gaborit, and A. Otmani, "Reducing key kength of the McEliece cryptosystem," *Advances in Cryptology*, LNCS 5580, pp. 77-97, Springer-Verlag, 2009.

[2] E. R. Berlekamp, R. McEliece, and H. C. A. V. TilBorg, "On the inherent intractability of certain coding problems," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 384-386, 1978.

[3] B. Biswas and N. Sendrier, "McEliece cryptosystem implementation: Theory and practice," *Second International Workshop in Post-Quantum Cryptography: PQCrypto'08*, LNCS 5299, pp. 47-62, Springer-Verlag, 2008.

[4] P. L. Cayrel, P. Gaborit, and E. Prouff, "Secure implementation of the Stern authentication and signature schemes for low-resource devices," *Eighth Smart Card Research and Advanced Application Conference*, LNCS 5189, pp. 91-205, Springer-Verlag, 2008.

[5] F. Chabaud and J. Stern, "The cryptographic security of the syndrome decoding problem for rank distance codes," *Advances in Cryptology*, LNCS 1163, pp. 368-381, Springer-Verlag, 1996.

[6] K. Chen, "A new identification algorithm," *Cryptography Policy and Algorithms*, LNCS 1029, pp. 244-249, Springer-Verlag, 1996.

[7] N. T. Courtois, "Efficient zero-knowledge authentication based on a linear algebra problem minrank,"

Table 2: Comparison of code-based authentication schemes with the same security level

Scheme	H $k \times N$	Memory (bits)	Prover's computation	Transmission rate (bits)
Harari[9]	1000×2000	1,002,000	$2^{28.2}$	153,900
Véron[12]	1000×2000	1,002,000	$2^{32.2}$	3,078,000
Stern [4]	256×512	1,718	2^{18}	40,000
This work	225×450	6,750	$2^{17.7}$	17,100

Advances in Cryptology, LNCS 2248, pp. 402-421, Springer-Verlag, 2001.

[8] M. Girault, "A (non practical) three-pass identification protocol using coding theory," *Advances in Cryptology*, LNCS 453, pp. 265-272, Springer-Verlag, 1990.

[9] S. Harari, "A new authentication algorithm," *3rd International Colloquium on Coding Theory and Applications*, pp. 91-105, Springer-Verlag, 1989.

[10] N. T. Hoa, K. Naoe, and Y. Takefuji, "Simplified IPSec protocol stack for micro server," *International Journal of Network Security*, vol. 11, no. 2, pp. 65-73, 2010.

[11] J. Stern, "A new identification scheme based on syndrome decoding," *Advances in Cryptology*, pp. 13-21, 1993.

[12] P. Véron, "Cryptanalysis of Harari's identification scheme," *Proceedings of the 5th IMA Conference on Cryptography and Coding*, LNCS 1025, pp. 264-269, Springer-Verlag, 1995.

[13] X. Tian, R. W. Zhu and D. S. Wong, "Improved efficient remote user authentication schemes," *International Journal of Network Security*, vol. 4, no. 2, pp. 149-154, 2007.

Behzad Malek received his Master's degree from University of Ottawa, Canada in Electrical Engineering in 2005. He is currently a PhD Candidate in Electrical and Computer Engineering at the University of Ottawa, Canada. His research interests are communications security and privacy. He is specialization is in applied cryptography and design of cryptographic security/privacy systems. He has been a student IEEE member and also a member of Computational Laboratory in Coding and Cryptography (CLiCC) at the University of Ottawa since 2003.

Ali Miri has been a Full Professor at School of Computer Science, Ryerson University, Toronto since 2009. He has also been with the School of Information Technology and Engineering and Department of Mathematics and Statistics since 2001 as an Assistant Professor, and later as an Associate Professor in 2005 and a Full Professor at 2008. He is the founder and the director of Computational Laboratory in Coding and Cryptography (CLiCC) at the University of Ottawa. He has held visiting positions at the Fields Institute for Research in Mathematical Sciences, Toronto in 2006, and Universite de Cergy-Pontoise, France in 2007, and Alicante and Albecete Universities in Spain in 2008. His research interests include computer networks, digital communication, and security and privacy technologies and their applications, in which he has authored and co-authored more than 120 peer-reviewed papers in international conference and journals. Dr. Miri has served in more than 50 organizing and technical program committees of international conferences and workshops. He has chaired/co-chaired The 14th workshop on Selected Areas in Cryptography (SAC), 2007, The Fields workshop on New Direction on Cryptography, 2008, IFIP Conference in Wireless Sensor and Actor Networks (WSAN), 2008, The Canadian Workshop in Information Theory (CWIT), 2009, The Eighth International Conference on Privacy, Security, and Trust (PST), 2010, IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMoB), 2010, and he will co-chairing The 18th workshop on Selected Areas in Cryptography (SAC), 2011. He has served as a guest editor for Journal of Ad Hoc and Sensor Wireless Networks, Journal of Telecommunications Systems, and is currently serving on the editorial board of International Journal On Advances in Internet Technology.

He is a member of Professional Engineers Ontario, ACM and a senior member of IEEE.