군사분야 비밀자료 관리를 위한 암호 알고리즘★

김홍태* · 이문식* · 강순부*

요 약

군사분야에는 개인의 신상정보 수준에서부터 국가의 존망에 영향을 미치는 수준의 정보까지 다양한 정보들이 존재한다. 암호 알고리즘은 이러한 정보들을 안전하게 관리하는데 해결책을 제시해 줄 수 있다. 우리는 온라인 상에서 비밀스럽게 처리되어야 하는 군사분야 자료를 효과적으로 관리할 수 있는 검색 가능 암호 알고리즘을 제안한다. 추가적으로, 상황에 따라 군사분야 자료 관리에 효과적인 알고리즘을 모색한다.

Cryptographic Algorithms for the Military Secret Data Management

HongTae Kim* · MoonSik Lee* · Sunbu Kang*

ABSTRACT

There exists vast amount of information in the military, both at the personal level and at the national security level. Cryptographic algorithms can present a solution to managing this information safely. We suggest searchable encryption algorithms for efficiently managing military data, which should be treated securely in an online environment. Furthermore, we investigate algorithms that are effective for managing military data under various conditions.

Key words : Military, Cryptography, Algorithm, Secret Data, Searchable Encryption

접수일(2014년 10월 7일), 수정일(1차: 2014년 10월 17일), 게재확정일(2014년 10월 29일)

★ 본 논문은 2011년도 공군사관학교 국고지원연구비에 의하여 연구되었음(KAFA 11-13).

* 공군사관학교/기초과학과

1. 서 론

군사분야에는 엄정하게 관리해야 하는 각종 비밀자료는 물론이고 이에 준하는 수준으로 관리해야 하는 자료도 다수 존재한다. 구성원 개인의 신상 정보의 경우에는 이를 저장하고 검색하는 것이 각 개인 뿐 아니라 인사담당자와 같이 관련된 업무를 수행하는 사람에게도 권한이 주어져야 하는 경우가 있다. 이러한 정보들을 안전하게 관리하기 위해서는 정보의 저장과 관리 등에 대한 통제가 필요하며 이를 해결하기 위한 방법들 중 검색 가능 암호가 하나의 방법이 될 수 있다.

검색 가능 암호 알고리즘(searchable encryption algorithm)은 특정 키워드에 대한 정보 노출 없이 이를 포함하는 정보를 검색할 수 있도록 고안된 알고리즘이다. 검색 가능 암호 알고리즘은 2000년대 이후 들어 많은 연구들이 진행되어 왔다.

Song, Wagner, Perrig[1]은 키워드를 이용하여 암호화된 데이터에서 원하는 키워드를 평문의 정보를 유출하지 않으면서 효율적으로 검색할 수 있도록 하였으며, 각 문서의 내용을 워드(word) 단위로 나누어 암호화하는 알고리즘을 제시하였다. 하지만 기존의 암호 시스템과 호환이 불가능하다는 점과 통계적인 공격 방법(statistical attack)에 취약하다는 점은 단점으로 지적되고 있다. 2003년 Goh[2]가 해시함수(hash function)와 블룸 필터(bloom filter)라는 특수 함수를 이용하여 알고리즘을 설계하였다. 처음으로 검색 가능 암호시스템에 대한 명확한 안전성 정의를 제시하였으며, 이를 이용하여 제안한 검색 가능 암호 시스템이 안전함을 증명하였다. 알고리즘에서 사용한 특수 함수의 성질로부터, 검색한 키워드가 없음에도 불구하고 그 키워드를 포함하고 있다는 오류 정보를 제공하는 문제가 발생하기도 한다. 2005년 Chang, Mitzenmacher[3]은 현실적인 요구를 반영하여 작은 트랩도어(trapdoor)를 사용하여 검색 가능 암호를 고안하였다. 기존의 검색 가능 암호 시스템들이 각각의 키워드에 대한 인덱스(index)를 각각 저장했던 것에 비해 이 시스템은 키워드마다 자료가 주어진 키워드를 포함하는지의 여부만을 1비트(bit)로 저장하였다. 또한 해시 테이블(hash table)을 이용한 방식으로 저장량을 최소화하였지만 사용자와 서버간의 통신 횟수가 증가하는 단점이 발생하였다. Curtmola 등이 제안한 방식[4]은 대칭키 기반 검색 가능 암호 시스템에 대한 기존의 능동 안전성(adaptive security) 정의를 새롭게 수정하였으며, 이를 바탕으로 새로운 검색 가능 암호 시스템을 제안하였다. 시스템에 사용된 주요 기술은 해시 테이블과 링크드 리스트(linked list)이다. 기존에 제안된 모든 검색 가능 암호 기술이 서버에 저장된 모든 자료의 인덱스에 대해 검색을 수행했다면, 이들이 제안한 알고리즘은 주어진 키워드에 대응하는 자료의 인덱스 만을 검사하여 검색 속도 면에서 장점을 갖는다. 하지만 하나의 자료가 여러 키워드와 관련성을 지닐 때, 중복 저장이 발생하여 저장 공간 면에서는 비효율적이다.

Song, Wagner, Perrig[1]이 제안한 알고리즘의 단점을 보완하기 위한 연구도 제안되었다. Shen, Shi, Waters[5]은 Song, Wagner, Perrig[1]의 알고리즘에서 결정론적(deterministic) 암호를 사용하여 문제가 되었던 부분을 페어링(pairing)을 사용하여 확률적 암호로 바꾸었다. 하지만 페어링을 사용하면서 비효율적으로 계산량이 늘어나는 새로운 문제를 야기하였다. 2011년 Yoshino, Naganuma, Satoh[9]은 확률적 암호를 쓰면서도 효율적으로 계산 가능한 알고리즘을 제시하였다. 이외에도 검색 가능 암호에 대한 다수의 결과들이 소개되었다[7,8,9,10,11,12,13,14].

본 논문에서는 다자간 환경이 필요한 군사분야 비밀자료 관리를 위한 알고리즘으로, 다수 사용자가 자료를 검색할 수 있는 검색 가능 암호 알고리즘을 공개키 기반 시스템과 비밀키 기반 시스템에 대해 각각 소개하고 알고리즘들이 어느 상황에 더욱 적합한지에 대해 설명한다. 암호학적으로 안전한 알고리즘을 소개하기 보다는 개념적인 이해를 돕기 위한 수준으로 설명을 전개하여 정확한 안전성 분석 등에 관한 내용은 배제한다.

2. 암호 알고리즘과 해시 함수

2.1 공개키/비밀키 암호 알고리즘

암호화는 특정 정보를 의미를 알 수 없는 형식으로

변경하는 과정이며, 복호화는 변경된 정보를 암호화하기 전의 정보로 되돌리는 것을 말한다. 공개키 암호 알고리즘은 암호화 및 복호화를 수행하는데 누구나 알 수 있는 공개키와 그에 대응하는 소유자만이 아는 비밀키를 이용하는 암호 알고리즘이다. 비밀키 암호 알고리즘은 송신자와 수신자가 동일한 키를 이용하여 암호화 및 복호화를 수행하는 암호 알고리즘을 말한다.

2.2 해시 함수

해시 함수는 임의의 길이를 갖는 입력 자료를 고정된 길이로 변환해 주는 함수로 다음과 같이 표현할 수 있다.

$$H : \{0,1\}^{*} \rightarrow \{0,1\}^{y}$$

결국, 해시 함수는 *비트 입력을 y비트 출력으로 바꾸어주는 함수이다. 이 결과를 흔히 해시 값(hash value)이라 하며, 이 값은 결정론적으로 정해져야 한다. 이렇게 정의한 함수가 현재 암호학적으로 안전하기 위해 출력은 160비트이며 이는 해시 함수의 역상 저항성(preimage resistance), 제2역상 저항성(second preimage resistance), 충돌 저항성(collision resistance)을 보장하기 위해 설정된 수치이다.

3. 제안 검색 가능 알고리즘

본 논문에서는 다자간 환경에서 가능한 공개키 기반 검색 가능 알고리즘과 비밀키 기반 검색 가능 알고리즘을 제안한다. 공개키 기반 검색 가능 알고리즘은 우선 복호화 할 수 있는 사용자가 한 명일 때의 알고리즘을 제시한 후에 복호화 할 수 있는 사용자가 여러 명일 때 검색 가능한 알고리즘을 제시한다. 두 알고리즘 모두 여러 개의 검색어를 한 번에 검색 가능하다. 비밀키 기반 검색 가능 알고리즘은 Song, Wagner, Perrig[1]이 제안한 알고리즘을 변형한 것으로 여러 명이 검색 가능하도록 하는 알고리즘을 제시한다.

3.1 검색 가능 암호 알고리즘 구조(structure)

검색 가능 암호 알고리즘은 키 생성, 암호화, 트랩도어 생성, 검색을 위한 테스트 과정으로 구성되어 있으며, 세부 구조는 다음과 같다.

1) 키 생성(key generation): 보안 파라미터 1^{k}를 입력 받아 사용자 A의 공개키 A_{pub}와 비밀키 A_{priv}를 생성한다.

2) 암호화(encryption with keyword search): 공개키 A_{pub}와 키워드 W를 입력 받아 W의 검색 가능 암호 S를 생성한다.

3) 트랩도어 생성(trapdoor generation): 비밀키 A_{priv}와 키워드 W를 입력 받아 트랩도어 T_{W}를 생성한다.

4) 검색을 위한 테스트(test for search): 공개키 A_{pub}와 검색 가능 암호 S, 그리고 트랩도어 T_{W}를 입력 받아 조건식을 만족하면 '예'를 만족하지 않으면 '아니요'를 출력한다.

위 구조는 공개키 기반 검색 가능 알고리즘에 관한 세부 구조이며 유사한 구조를 가지는 비밀키 기반 검색 가능 알고리즘은 3.3에서 구체적인 예제를 통해 살펴보기로 한다.

3.2 공개키 기반 검색 가능 알고리즘

3.2.1 제안 알고리즘의 안전성 기반

G는 소수 위수 p를 갖는 곱셈 순환군이고 g는 G의 생성원이다. 제안된 공개키 암호 알고리즘의 안전성은 다음의 DDH 가정에 기반한다.

• Decisional Diffie-Hellman Assumption(DDH 가정) : $a, b, c \in Z_{p}^{*}$에 대해, (g^{a}, g^{b}, g^{ab})와 (g^{a}, g^{b}, g^{c})가 주어질 때 다항식 시간에 무시할 수 없는 확률로 둘을 구별할 수 있는 공격자는 없다.

3.2.2 사용자가 한 명인 경우

검색 가능 알고리즘의 이해를 돕기 위해 우선 복호화 할 수 있는 사용자가 한 명인 경우를 소개한다.

1) 키 생성

보안 파라미터 1^k를 입력 받아 사용자의 공개키 pk와 비밀키 sk를 생성한다. 무작위 값 $x \in Z_p^*$를 선택하고 $y = g^x$를 구한다. 이때, 공개키와 비밀키는 $(pk, sk) = (y, x)$로 주어진다.

2) 암호화

$H : \{0,1\}^* \rightarrow \{0,1\}^{\log p}$라 하자. 사용자는 키워드 집합 $W = \{w_1, \ldots, w_n\}$의 모든 원소 $w_1, \ldots, w_n$에 대해 $v_1 = H(w_1), \ldots, v_n = H(w_n)$을 계산한다. W의 검색 가능 암호 S를 다음과 같이 만든다.

$$S = \{g^x, g^{xv_1}, g^{xv_2}, \ldots, g^{xv_n}\}$$

3) 트랩도어 생성

비밀키 x와 집합 J를 입력 받아 트랩도어 T_J를 생성한다. 집합 J는 다음과 같이 주어진다.

$$J = \{j_1, \ldots, j_r\} \subseteq \{1, \ldots, n\}$$

무작위 값 $u \in Z_p^*$에 대해 트랩도어 T_J는 다음과 같이 생성한다.

$$T_J = \{g^u, j_1, j_2, \ldots, j_r, H(g^{S_J})\}$$

여기서, $u = S_J + v_{j_1} + v_{j_2} + \ldots + v_{j_r}$을 만족한다.

4) 검색을 위한 테스트

검색 가능 암호 S와 트랩도어 T_J를 입력 받고 공개키 y를 이용하여 다음을 테스트 한다.

$$H(g^u \, g^{-(x_i v_{j_1} + \ldots + x_i v_{j_r})}) = H(g^{S_J})$$

제안하는 알고리즘의 장점은 계산량이 굉장히 적어 기존에 제안된 알고리즘보다 속도 향상이 크다는 것이다. 한 번의 해시 함수 연산과 한 번의 지수승 연산을 통해 검색하고자 하는 키워드를 테스트하는 것이 가능해진다. 그렇지만, 이에 대한 안전성에 관해서는 조금 더 논의가 필요하다.

3.2.3 사용자가 여러 명인 경우

1) 키 생성

보안 파라미터 1^k를 입력 받아 사용자의 공개키 $pk_1, \ldots, pk_n$와 비밀키 $sk_1, \ldots, sk_n$을 생성한다. 무작위 값 $x_1, \ldots, x_n \in Z_p^*$를 선택하고 이에 대응하는 $y_i = g^{x_i}$를 구한다. 이때, 각 사용자의 공개키와 비밀키는 $(pk_i, sk_i) = (y_i, x_i)$로 주어진다.

2) 암호화

$H : \{0,1\}^* \rightarrow \{0,1\}^{\log p}$라 하자. 사용자는 키워드 집합 $W = \{w_1, \ldots, w_n\}$의 모든 원소 $w_1, \ldots, w_n$에 대해 $v_1 = H(w_1), \ldots, v_n = H(w_n)$을 계산한다. W의 검색 가능 암호 S를 다음과 같이 만든다.

$$S = \{g^{x_i}, g^{x_i v_1}, g^{x_i v_2}, \ldots, g^{x_i v_n}\}$$

3) 트랩도어 생성

비밀키 x_i와 집합 J를 입력 받아 트랩도어 T_J를 생성한다. 집합 J는 다음과 같이 주어진다.

$$J = \{j_1, \ldots, j_r\} \subseteq \{1, \ldots, n\}$$

무작위 값 $u \in Z_p^*$에 대해 트랩도어 T_J는 다음과 같이 생성한다.

$$T_J = \{g^u, j_1, j_2, \ldots, j_r, H(g^{S_J})\}$$

4) 검색을 위한 테스트

검색 가능 암호 S와 트랩도어 T_J를 입력 받고 공개키 y를 이용하여 다음을 테스트 한다.

$$H(g^{u-x_i}g^{x_i}g^{-(v_{j_1}+\ldots+v_{j_r})}) = H(g^{S_J})$$

여기서, $u = S_J + v_{j_1} + v_{j_2} + \ldots + v_{j_r}$을 만족한다.

전체 알고리즘을 고려하면 사용자가 한 명인 알고리즘에 비해 계산량이 크게 늘어나지 않은 상황에서 검색 가능하며, 각 개인별로 계산량을 생각하면 검색을 위한 테스트 과정에서 한 명의 사용자를 위한 알고리즘과 동일한 계산량을 갖는 장점을 가진다. 각 개인의 공개키/비밀키를 생성하는 시간이 사용자에 비례하여 늘어난다는 단점이 있으나 서버 수준의 계산량이므로 영향력이 미비하다.

공개키 기반 검색 가능 알고리즘은 업로드는 누구나 가능하나 다운로드는 비밀키를 아는 사람만 할 수 있는 환경에 적합하다. 이를 활용하면 개인 신상정보 입력은 개인별로 누구나 하고 이를 확인하는 사람은 권한을 부여받아 비밀번호를 아는 사람만이 가능하도록 하는 환경이 조성된다.

3.3 대칭키 기반 검색 가능 알고리즘

공개키 기반 검색 가능 알고리즘에서의 네 개의 과정과 다른 방법으로 대칭키 기반 검색 가능 알고리즘 과정을 기술한다. Song, Wagner, Perrig[1]이 제안한 알고리즘의 변형으로 여러 사용자가 검색 가능한 알고리즘을 소개한다. 대칭키 기반 검색 가능 알고리즘은 업로드와 다운로드 모두 비밀키를 아는 사람이 하는 것으로 군사분야 대외비 수준 이상의 비밀자료 작업 시 활용하면 유용하다.

1) 키 생성

보안 파라미터 1^k를 입력 받아 사용자의 키 k_1, k_2를 생성한다. 키 k_1, k_2를 자료를 검색하려는 다른 사용자와 공유한다.

2) 암호화

$i = 1, \ldots, n$에 대해, 사용자는 키워드 집합 $W = \{w_1, \ldots, w_n\}$의 각 원소 w_i들을 대칭키 암호 E를 이용하여 $X_i = E_{k_1}(w_i)$를 계산한다. 각각의 X_i를 $X_i = L_i \| R_i$와 같이 나누어 이 중 L_i를 의사 난수 함수(pseudorandom function) f를 이용하여 각 워드에 사용될 새로운 비밀키 $k_3 = f_{k_2}(L_i)$를 생성한다. 사용자는 스트림 암호(stream cipher) F를 이용하여 난수열 S_i를 생성하고 이로부터 $Y_i = S_i \| F_{k_3}(S_i)$를 만든다. 암호문은 $C_i = X_i \oplus Y_i$로 만들어 낸다.

3) 검색을 위한 테스트

X_i와 k_3을 입력받아 암호문 C_i를 이용하여 다음을 테스트한다.

$$C_i \oplus X_i = S_i \| F_{k_3}(S_i)$$

이렇게 함으로써 암호문 C_i를 복호화하지 않고 평문의 특정 정보를 포함한 검색을 가능하게 한다. 소개한 알고리즘은 Song, Wagner, Perrig[1]이 제안한 알고리즘에 키를 공유하는 과정을 추가한 것으로, 키를 공유하는 과정이 사전 계산 가능하므로 실제 검색을 위한 테스트 시간은 Song, Wagner, Perrig[1]의 알고리즘과 동일하다. 대칭키 암호의 속도가 공개키 암호의 속도보다 대략 1000배 정도 빠르므로 구현 속도에서 큰 장점을 지닌다. 워드를 테스트하는 자료 보관자는 S_i를 사전에 인지하고 있어야 하는 단점이 있다.

여러 사용자가 검색할 수 있도록 하기 위해 키 생성 단계에서 생성한 비밀키는 검색을 하려는 여러 사용자와 공유하도록 해야 한다. 이에 대한 명확한 안전성 증명에 관한 연구도 추가적으로 필요하다.

4. 결 론

자료 관리의 중요성은 향후 온라인 환경의 보편화

와 더불어 더욱 증대될 것이므로 안전하고 효율적으로 이를 관리하는 기술에 대한 연구 필요성이 점차 커질 것이다. 본 논문에서는 특히 군사분야에서 다수가 업로드하고 비밀키를 아는 사람만이 복호화 할 수 있는 환경과 업로드와 다운로드가 모두 비밀키를 아는 사람만이 할 수 있는 제한적인 검색 가능한 암호 알고리즘을 소개하였다. 대칭키 기반 알고리즘이 속도 면에서 장점을 가지나 안전성 증명이 다소 취약한 반면 공개키 기반 알고리즘은 대칭키 기반에 비해 체계적으로 안전성 증명이 되어 있어 자료 관리의 안전성 측면에서 보다 효과적으로 판단된다. 업로드하는 사람들도 모두 비밀키를 알고 있어야 하는 상황이라면 비밀키 기반 검색 가능 알고리즘이 효과적이나 업로드하는 사람들은 비밀키를 몰라도 되는 상황이라면 공개키 기반 검색 가능 알고리즘이 적합하다.

추후 연구과제로 제안한 알고리즘을 실제로 활용하여 모의상황에 적합하게 적용해보는 것이 필요하며, 안전성 분석에 관해서는 조금 더 엄밀한 검증이 요구된다.

참고문헌

[1] D. Song, D. Wagner and A. Perrig, "Practical Techniques for Searching on Encrypted Data", IEEE Symposium on Security and Privacy, 2000.

[2] E.J. Goh, "Secure Indexes", Technical Report 2003/216, IACR ePrint Cryptography Archive, 2003.

[3] Y.C. Chang and M. Mitzenmacher, "Privacy Preserving Keyword Searches on Remote Encrypted Data", Proceedings of the Third International Conference on Applied Cryptography and Network Security, 2005.

[4] R. Curtmola, J. Garay, S. Kamara and R. Ostrovsky, "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions", Proceedings of the ACMCCS, 2006.

[5] E. Shen, E. Shi and B. Waters, "Predicate Privacy in Encryption Systems", Proceedings of the Sixth Theory of Cryptography Conference, 2009.

[6] M. Yoshino, K. Naganuma and H. Satoh, "Symmetric Searchable Encryption for Database Applications", International Conference on Network-Based Information Systems, 2011.

[7] B. Waters, D. Balfanz, G. Durfee and D. Smetters, "Building an Encrypted and Searchable Auditlog", Proceedings of the Network and Distributed System Security Symposium, 2004.

[8] D. Boneh and B. Waters, "Conjunctive, Subset and Range Queries on Encrypted Data", Proceedings of the Fourth Theory of Cryptography Conference, 2007.

[9] D. Boneh, G. Crescenzo, R. Ostrovsky and G. Persiano, "Public Key Encryption with Keyword Search", Proceedings of the Eurocrypt, 2004.

[10] F. Bao, R.H. Deng, X. Ding and Y. Yang, "Private query on encrypted data in multi-user setting", Proceedings of the ISPEC, 2008.

[11] M. Abdalla, M. Bellare, D. Catalano, E. Kiltz, T. Kohno, T. Lange, J. Malone-Lee, G. Neven, P. Pailier and H. Shi, "Searchable Encryption Re visited: Consistency Properties, Relation to Anon ymous IBE and Extensions", Proceedings of the Crypto, 2005.

[12] P. Golle, J. Staddon and B. Waters, "Secure Conjunctive Keyword Search over Encrypted Data", Proceedings of the Second International Conference on Applied Cryptography and Network Security, 2004.

[13] R. Ostrovsky and W. Skeith, "Private Searching on Streaming Data", Proceedings of the Crypto, 2005.

[14] Y. H. Hwang and P. J. Lee, "Public Key encryption with Conjunctive Keyword Search and its Extension to a Multi-User System", Proceedings of the Pairing, 2007.

[저 자 소 개]

김 홍 태 (HongTae Kim)

2003년 2월 서울대 수리과학부 학사
2006년 2월 서울대 수리과학부 석사
2013년 2월 서울대 수리과학부 박사
2013년 2월 ~ 현재 공군사관학교
기초과학과 수학교수

email : yeskafa@naver.com

이 문 식 (MoonSik Lee)

2001년 2월 서울대 수리과학부 학사
2004년 2월 서울대 수리과학부 석사
2010년 2월 서울대 수리과학부 박사
2010년 2월 ~ 현재 공군사관학교
기초과학과 수학교수

email : kafa0443@gmail.com

강 순 부 (Sunbu Kang)

1989년 2월 공학사
1996년 2월 이학사
1999년 2월 이학석사
2004년 8월 이학박사

email : sbkang@postech.ac.kr