Conjunctive 키워드 검색 스킴에서의 취약점 분석+

*이현숙 **정익래 *변진욱 *임종인 *이동훈

고려대학교 정보경영공학전문대학원

*math33@cist.korea.ac.kr, jir@etri.re.kr, *{byunstar, jilim, donghlee,}@korea.ac.kr

요약

In a keyword search scheme a user stores encrypted data on an untrusted server and gives a database manager a capability for a keyword which enables a database manager to find encrypted data containing the keyword without revealing the keyword to the database manager. Conjunctive keyword search scheme enables a user to obtain data containing all of several keywords through only one query. One of the security requirements of conjunctive keyword search schemes is that a malicious adversary should not be able to generate new valid capabilities from the observed capabilities. In this paper we show that conjunctive keyword search schemes are not secure. In particular, given two capabilities corresponding two sets of keywords, an adversary is able to generate a new capability corresponding to the dierence set of two keywords sets.

1. 서론

인터넷에서 관리되고 저장되어지는 정보의 양이 급격하게 증가되어짐에 따라, 데이터베이스와 같은 저장 시스템의 중요성은 증가되었다. 그 결과로 저장시스템에 저장된 자료들에 대한 프라이버시를 보호하는 것은 데이터베이스 산업 분야에 가장 시급히 해결해야 할 과제이다. 최든 데이터베이스 보안에서의 가장 큰 이슈는 크게 두 가지로 압축된다. 첫 번째는 사용자가 저장시스템에 저장한 자료들을 외부 공격자로부터 보호하는 것이다. 이러한 문제를 간단히 해결하기 위해, 데이터베이스 관리 시스템 (DBMS)에 암호학적 암호 모듈을 심어, 모든 데이터를 암호화함으로써 해결할 수 있다. 실제적으로, Oracle 8i 와 MS Access는 DBMS 자체 내에 암호화 모듈을 제공한다. 그러므로 추가적인 키 관리 모듈에 대한 비용 없이 첫 번째 이슈는 해결될 수 있다. 두 번째 이슈는 시스템 관리자와 같은 내부 사용자들로 부터의 자료의 보호이다. 이를 위해 사용자들은 역시 그들의 자료를 암호화 하여 저장한다. 이와 더불어, 그 암호화키를 내부사용자들로부터 안전하게 사용자 측에서 관리한다면, 사용자는 내부사용자들로 부터의 자료 도청 및 관람을 방지할 수 있다. 이러한 암호화의 방법은 자신의 자료를 외부 및 내부 공격자로부터 안전하게 지켜주는 확실한 방법이지만, 이와 반대로 암호화는 자료를 랜덤하게 만들어 자료의 검색을 비효율적으로 만드는 단점도 존재한다. 최근에, 이러한 단점을 해결하기 위해 암호화된 문서상에서 효율적인 자료 검색 프로토콜들이 많이 제안되었다. 그 예가 암호화된 데이터에서의 검색어를 이용한 검색기술이다. 이러한 기술들은 데이터를 제공하고 있는 환경 별로 제안되었고, 이 기술 중에는 Conjunctive 검색 기술이 있다. Conjunctive 검색 기술은 검색자가 여러 개의 키워드를 선택하여 그 모든 키워드에 관련된 데이터를 검색하고자 할 때 사용하는 기술로 자신이 원하는 데이터에 대한 정보나 검색어에 대한 정보를 서버관리자나 제 3자가 알지 못하도록 검색하는 기술이다. 이 기술은 Golle et. al. 이 처음 제안하였으며 그후 환경별로 많은 연구가 진행되어졌다.

사실 공통의 키워드 검색은 하나의 키워드 검색의 단순한 확장이라고 생각하기 쉽다. 즉, 하나의 키워드 검색 프로토콜을 여러 번 시행 시킨 결과들에 대해서 사용자가 공통의 부분을 찾으면 된다. 하지만 이러한 접근 방법은 사용자들에게 많은 통신량과 연산량을 요하게 되며, 특히, 서버가 공통된 키워드를 포함한 문서를 제외한 추가적인 문서들과 키워드들의 연관관계를 파악하게 되므로 효율성 및 프라이버시 보호 측면에서 옳지 못하다. 또 사용자가 검색어를 이용하여 검색하기 위해서 암호화된 검색어 형태의 Capability를 데이터 베이스 관리자에게 보내게 될 것이다. 이 경우 데이터베이스 관리자는 Capability로부터 직접적으로 검색어에 대한 정보를 알 수 없어야 하고 Capability와 데이터와의 관련성을 알 수 없어야 한다. 다음의 예를 보자.

예제 1. SIi(Alice) 와 SIj(Bob)는 검색어 "Alice"와 "Bob" 각각에 해당하는 암호화된 검색을 위한 정보들이다. Dk는 암호화되어 저장된 데이터이다. 데이터베이스 관리자는 다음의 정보들을 데이터베이스에 저장한다. K1 = {Alice, Bob} 와 K2 = {Alice}는 검색어들의 집합이다. C_1 =Cap(K1) 과 C_2 = Cap(K2)는 K1 과 K2 각각의 Capabilities 이다. C_1 =Cap(K1) 과 C_2 = Cap(K2)를 전송받는 데이터베이스의 관지라는 C_1 과 C_2로부터 계산 결과가 $\{D_5\}$와 $\{D_1,D_3,D_5\}$라는 사실을 알 수 있다. 이 경우, 데이터베이스 관리자는 D5 가 공통의 결과라는 것을 알 수 있는데 같은 검색어 필드에서 검색어가 다르다면 같은 결과를 포함할 수 없기 때문에 같은 결과가 있다면 검색에 사용된 검색어 집합들 사

+ 본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음" (IITA-2006-(C1090-0603-0025))

이에 부분집합 관계가 있다는 사실을 유추하는 것이 가능하다. 따라서 데이터 베이스 관리자는 다음의 결과를 얻게 된다.

(1) K1 이 {D_5}과 연관되어진다.
(2) K2 이 {D_1,D_3,D_5}과 연관되어진다.
(3) K2 이 K1 의 부분집합니다.

Conjunctive 검색 기술에서는 데이터베이스 관리자가 앞에서 언급한 정보를 제외한 어떠한 정보도 얻을 수 없어야 한다. 그러나 차집합인 K1 - K2 에 관련된 데이터가 {D2,D4}라는 사실을 데이터베이스 관리자가 아는 것이 가능하다. 즉, 안전성 측면에서 데이터의 비연결성(unlinkability)이 깨지게 된다. 본 논문에서는 conjunctive 키워드 검색 스킴들[13,19, 22]이 비연결성을 만족하지 못한다는 것을 다음과 같이 "Difference Set 공격"을 새롭게 정의하고 "Difference Set 공격"을 이용하여 설명하였다.

정의 1. A를 conjunctive 키워드 검색 스킴의 악의적인 데이터베이스 메니져라고 하고 $(K_1,...,K_l)$를 키워드들의 집합들이라고 하자. 또, $(C_1,...,C_l)$를 A에게 주어진 $(K_1,...,K_l)$에 대한 capability들이라고 하자. 이때, 악의적인 데이터베이스 메니져 A가 $K \notin \{K_1,...,K_l\}, K_j \subset K_i, 1 \le i,j \le l (i \ne j)$를 만족하는 키워드의 집합 $K = K_i - K_j$에 대한 capability를 만들 수 있다면 Conjunctive 키워드 검색 스킴은 Difference Set 공격에 대하여 안전하지 않다.

다음 절에서는 Golle et. al.의 Conjunctive 키워드 검색 스킴이 difference set 공격에 대하여 안전하지 않은 것을 보일 것이다. 또, 이와 동일한 방법으로 지금까지 제안되진 Conjunctive 키워드 검색 스킴들은 difference set 공격에 대하여 안전하지 않다.

2. Golle et al.의 Conjunctive 키워드 검색 스킴에 대한 분석

가. Golle I 스킴에서의 Difference Set 공격

본 절에서는 Golle I을 간단히 살표보도록 하겠다.
사용자는 데이터베이스에 다음의 데이터를 저장한다.
$E(m_i) \| CSI_i = \{I_i, CSI_{i,1}(W_{i1},..,K),..., CSI_{i,m}(W_{i1},..,K)\}$
이때, m_i는 메시지이고, E 는 안전한 암호화 알고리즘이고 m은 키워드 필드의 수이다. CSI_i는 데이터베이스 메니저나 제 삼자에게 메시지에 대한 어떠한 정보도 들어내지 않으면서,
정당한 사용자가 검색어들을 이용하여 capability를 만들어서 데이터베이스 메니저에게 주었을 때 검색하는 것이 가능하도록 한다. 구체적인 프로토콜의 구성은 다음과 같다.

- **Param**(1^k) : security 파라메터 k를 입력받고, 다음의 파라메터를 출력한다.
$$\rho = (G, g, f(\cdot,\cdot), h(\cdot))$$
이 때, G는 decisional Die-Hellman (DDH) 문제가 어렵고 grk 생성원(generator)인 위수(order) q의 군(group)이다. 또, $f: \{0,1\}^k \times 0,1\}^* \to Z_q^*$ 는 비밀키를 갖는 함수이고 h는 해쉬 함수이다.

- KeyGen(1^k) : security 파라메터 k를 입력받고, 함수 f에 대한 비밀키 K를 출력한다.

- CSI$(\rho, K, W_{i,1},..., W_{i,m})$: K와 $W_{i,1},..., W_{i,m}$를 입력받은 후, 다음의 conjunctive 검색정보를 출력한다.
$$CSI_i = \{I_i, CSI_{i,1}(W_{i1},..,K),..., CSI_{i,m}(W_{i1},..,K)\}$$
$$= \{g^{a_i}, g^{a_i f_K(W_{i1})},..., g^{a_i f_K(W_{im})}\}$$
이 때, I_i는 conjunctive 검색을 위해서 요구되는 추가정보이고, $CSI_{i,j}(W_{ij},..,K)(1 \le j \le m)$는 검색어 W_{ij}에 대한 검색 정보이다.

- TCK$(,K,p_1,...,p_l,Q_l)$: $1 \le l \le m$ 인 l에 대하여, 비밀키 sk와 검색하려는 키워드들 $W_{p_1},..., W_{p_l}$와 그 키워드들이 해당되는 필드이름들을 입력 받아서 결과로 트랩도어 $T_l = \{Q, C, p_1,...,p_l\}$을 출력한다. 이 때,
$Q = (h(g^{a_1 s}),...,h(g^{a_n s}))$이고, $C = s + \sum_{w=1}^{t} f_K(W_{j_w})$이다.

- Test(CSI_i, T) : 검색정보 CSI_i 와 트랩도어 T를 입력받아서 다음을 수행한다.
만약 $(W_{i,p_1} = W_{p_1}) \wedge ... \wedge (W_{i,p_l} = W_{p_l})$이면, "1"을 출력하고, 성립하지 않으면 "0"이다.

나. Golle I 스킴에서의 Difference Set 공격

정리 1. Golle I 스킴은 difference set 공격에 대하여 안전하지 않다.
증명. 우리는 S_A를 악의적인 데이터베이스 관리자라 가정하고 S_A가 정당하게 생성된 두 개의 트랩도어 T_1과 T_2를 이용하여 새로운 트랩도어 T_3를 만드는 것이 가능한 것을 다음과 같이 보이도록 하겠다.

- 단계 1) 악의적인 데이터베이스 관리자인 공격자 S_A는 정당하게 생성되어진 다음과 같은 트랩도어 값 T_1과 T_2를 얻는다고 가정하고 $T_1 = \{Q_1, C_1, 1,2,3\}$ $T_2 = \{Q_2, C_2, 1,2\}$는 다음과 같다.
$$Q_1 = \{h(g^{a_1 s_1}),...h(g^{a_n s_1})\}$$
$$C_1 = s_1 + f_K(W_1) + f_K(W_2) + f_K(W_3)$$
$$Q_2 = \{h(g^{a_1 s_2}),...h(g^{a_n s_2})\}$$
$$C_2 = s_2 + f_K(W_1) + f_K(W_2)$$
물론 공격자는 정당한 사용자에 의해서 랜덤하게 선택되어진 s_1과 s_2의 값을 알 수 없지만 전송된 결과 값들 Q_1, C_1, Q_2, C_2 은 알 수 있다. 또, 두 개의 트랩도어들에 이용된 키워드들이 각각 {1,2,3}과 {1,2}이고 이들이 {1,2}⊂{1,2,3}인 부분집합 관계에 있다는 사실은 알 수 있다.

- 단계 2) 공격자는 앞 단계에서 얻은 정보들을 이용하여 Test과정을 수행하여 다음의 결과를 얻는다.

1. 첫 번째 트랩도어 T_1의 Test과정 수행 결과는 $\{D_1\}$이라는 사실을 얻고, T_2 의 Test과정 수행 결과는 $\{D_1, D_2\}$이라는 사실을 얻는다.
2. 두 개의 결과에서 모두 $\{D_1\}$이 포함된다는 사실은 {1,2,3}과 {1,2}가 $\{1,2\} \subset \{1,2,3\}$인 부분집합 관계에 있기 때문에 두 개의 트랩도어를 생성할 때 첫 번째 필드와 2번째 필드에 당하는 키워드로 같은 키워드를 이용해야지만 같은 결과를 얻을 수 있다.
   따라서, 공격자는 $C_1 - C_2 = s_1 - s_2 + f_K(W_3)$ 가 키워드 값을 정확히 알 수는 없지만 3번째 필드에 있는 키워드에 대한 $f_K(W_3)$과 어떤 적당한 랜덤 값의 합으로 표현 가능하다는 것을 알 수 있다.
3. 공격자는 데이터베이스에 저장 되어져 있는 검색 정보들을 이용하여 다음의 R_i와 R_i'을 계산한다.

$$R_i = \frac{g^{a_i C_1}}{g^{a_i V_{i,1}} \cdot g^{a_i V_{i,2}} \cdot g^{a_i V_{i,3}}}$$

$$R_i' = \frac{g^{a_i C_2}}{g^{a_i V_{i,1}} \cdot g^{a_i V_{i,2}}}$$

4. 공격자는 위에서 계산한 결과들을 이용하여 다음의 T_3 값을 얻을 수 있다.

$$T_3 = \{Q_3, C_3, 3\}$$

$$Q_3 = \left\{ h(\frac{R_1}{R_{1'}}), \ldots, h(\frac{R_n}{R_{n'}}) \right\}$$

$$C_3 = C_1 - C_2$$

여기서, $h(\frac{R_i}{R_{i'}}) = h(g^{a_i(s_1 - s_2)})$ 가 성립되는 이유는 $g^{a_i C_3} / g^{a_i V_{i,3}} = g^{a_i(s_1 - s_2)}$가 만족하기 때문이다.

5. 앞 단계에서 얻은 트랩도어 T_3를 이용하여 Test과정을 수행하게 되면 공격자는 트랩도어를 정당하지 않게 위조하여 데이터 검색이 가능하게 된다.

따라서 Golle I 스킴은 difference set 공격에 대하여 안전하지 않다. □

2. 결론

본 논문에서는 Golle I 스킴은 difference set 공격에 대하여 안전하지 않다는 것을 보였다. 또한 앞에서 언급한 공격과 유사한 방법으로 지금까지 제안되진 Conjunctive 키워드 검색 스킴들은 difference set 공격에 대하여 안전하지 않다. 앞으로 difference set 공격을 고려하여 안전성 모델을 정의하고 difference set 공격에 대하여 안전한 Conjunctive 키워드 검색 스킴을 제안하는 것을 향후 연구할 가치가 있다.

참고문헌

1. Michel Abdalla, Mihir Bellare, Dario Catalano, Eike Catalano, Tanja Lange, John Malone-Lee, Gregory Neven, Pascal Paillier, Haixia Shi,"Searchable Encryption Revisited: Consistency Properties, Relation to Anonymous IBE, and Extions.", Crypto'05, LNCS Vol3621. , pp205-222 , 2005.

2. R. Agrawal and R. Srikant,"Privacy-Preserving Data Mining", In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, pp. 439-450, 2000.

3. M. Balze, "A Cryptographic file system for UNIX.", Processings of 1st ACM Conference on Communications and Computing Security, 1993.

4. S. Bellovin , W. Cheswick, "Privacy-enhanced searches using encrypted bloom filters", Cryptology ePrint Archive, Report 2004/022, Feb 2004.

5. D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key Encryption with Keyword Search", EUROCRYPT'04, 2004.

6. D. Brassard, C. Crepeau, and J. M. Robert, "All-or-Nithing Disclosure of Secrets", Crypto'86, Springer-Verlag, 1987, pp. 234-238.

7. M. Bellare, C. Namprempre, and D. Pioncheval,"The Power of RSA Onversion Oracles and the Security of Chaum's RSA-Based Blind Signature Scheme", Proc. of Finandcial Cryptography 2001, LNCS vol. 2339, pp. 319-338.

8. Y. C. Chang, M. Mitzenmacher, "privacy preserving keyword searches on remote encrypted data",ePrint, October 7th 2003.

9. G. Cattaneo, G. Persiano, A. Del Sorbo, A. Cozzolino, E. Mauriello, and R. Pisapia, "Design and implementation of a transparent cryptographic file system for UNIX", Techincal Report, University of Salerno, 1997.

10. S. Even, O. Goldreich, and A. Lempel, "A Randomized Protocol for Signing Contracts", comm. of ACM, 28:637-647, 1985.

11. E. J. Goh, "secure index", ePrint, October 7th 2003.

12. P. Golle, M. Jakobsson, A. Juels, and Paul Syverson, "Universal Re-encryption for Mixnets", In proceedings of CT-RSA 2004, 2004.

13. P. Golle, J. Staddon and B.Waters, "Secure Conjunctive Keyword Search Over Encrypted Data", Proceedings of the Second International Conference on ACNS:Applied Cryptography and Network Security, 2004.

14. J. Hughes and D. Corcoran, "A nuiversal access, smart-card-based, secure fiel system.", Atlanta Linux Showcase , October 1999.

15. A. John , R. Peter, "Electric Communication Development", Communications of the ACM ,40,May 1997, pp. 71-79. 48-63, 2002.

16. K. Kurosawa, "Multi-recipient Public-Key Encryption with Shortened Ciphertext", In proceedings of PKC 2002, LNCS 2274, pp. 48-63, 2002.

17. M. Noar and B. Pinkas, "Ecient Oblivious trnasfer protocols", 12th Annual Symposium on Discrete Algorithms(SODA), pp 448-457(2001).

18. W. Ogata and K. Kurosawa, "Oblivious Keyword Search", Journal of complexity'04, Vol 20. April/Jun 2004.

19. D. Park, K. Kim, and P. Lee, "Public key Encryption with Conjunctive Field Keyword Search", WISA'04, LNCS 3325, pp73-86, 2004.

20. D. Pointcheval and J. P. Stern, "Provably secure blind signature schemes", Proc.of Asiacrypt'96, LNCS Vol. 1163, pp 252-265, 1996.

21. M. Rabin, "How to exchange secrets by oblivious transfer", Technical Report TR81, Aiken computation Lab, Harvard University.

22. H.S. Rhee, J. W. Byun, D. H. Lee, J. I. Lim, "Oblivious Conjunctive Keyword Search", WISA 2005, LNCS Vol3786. , pp318-327 , 2005.

23. D. Song, D. Wagner, and A. Perrige, "Practical Techniques for searches on Encrypted Data", In Proc. of the 2000 IEEE Security and Privacy Symposium, May 2000.

24. B. R. Waters, D. Balfanz, G. Durfee, and D. K. Smetters, "Building an Encrypted and Searchable Audit Log",11th Annual Network and Distributed Security Symposium (NDSS '04) 2004.

25. E. Zadok, I. Badulescu, and A. Shender, "Cryptfs : A stackable vnode level encryption fiel system.", Technical Report UCS-021-98 : 1998.