제36회 한국정보처리학회 추계학술발표대회 논문집 제18권 제2호 (2011. 11)

스마트카드를 이용한 프라이버시보호 다자간 교집합 연산 프로토콜*

김민구, 강주성, 이옥연
국민대학교 수학과, 정보보안연구소
{kmnine,jskang,oyyi}@kookmin.ac.kr

Privacy-preserving Set Intersection Multi-party Protocol using Smart Cards

Mim-Ku Kim, Ju-Sung Kang, Okyeon Yi
Dept. of Mathematics and CISI, Kookmin University

요 약

다자간 프라이버시보호 교집합 연산은 둘 이상의 참여자들이 서로 자신이 가지고 있는 데이터를 노출시키지 않으면서 교집합을 구하는 문제이다. 다자간 프라이버시보호 교집합 연산은 보험사기 방지시스템, 항공기 탑승 금지자 목록 검색, 의료 정보 검색, 전자투표 등에서 이용될 수 있다. 2009년 Hazay와 Lindell[1]은 스마트카드를 이용한 양자간 프라이버시보호 교집합 연산을 하는 프로토콜을 제안하였다. 이 프로토콜은 신뢰할 수 있는 제 3자를 설정할 수 없는 상황에서 스마트카드의 보안 요소를 사용하여 양자간 프라이버시보호 교집합 연산을 할 수 있다. 또한 이론적으로는 안전하나 실제로 구현이 어려운 일방향함수를 기반으로 한 모델의 단점을 의사난수치환을 사용하여 현실적인 모델로 보완하였다. 본 논문에서는 기존의 Hazay와 Lindell의 양자간 프로토콜에 Commodity Server를 도입하여, 다자간 프라이버시보호 교집합 연산을 할 수 있는 프로토콜을 제안한다.

1. 서론

특정 데이터 집합을 가지고 있는 둘 이상의 참여자들이 서로 자신이 가지고 있는 데이터 집합을 노출하지 않으면서 공통의 데이터를 찾는 일을 프라이버시보호 다자간 교집합 연산이라 한다. 즉, 프라이버시보호 다자간 교집합 연산은 참여자들의 프라이버시를 보호하며, 여러 참여자들의 데이터 집합으로부터 교집합을 연산할 수 있게 해준다.

프라이버시보호 다자간 교집합 연산의 문제는 최근 학계뿐 아니라 개인정보를 다루는 기업 혹은 정부기관에서 많은 관심을 가지고 있으며, 다양한 분야에서 응용할 수 있는 중요한 문제이다. 예를 들어 두 보험회사가 보험사기를 적발하기 위하여 두 보험사에 공통으로 가입한 고객을 찾고 싶을 때, 혹은 항공회사와 경찰청이 항공기 탑승자 명단에서 서로의 데이터베이스를 노출시키지 않으면서 탑승 금지자를 찾고 싶을 때 사용할 수 있다.

프로토콜 참여자들이 교집합을 얻기 위한 이상적인 방법은 완전히 신뢰할 수 있는 제 3자(Trusted Third Party, TTP)의 도움을 받는 것이다. 그러나 TTP를 사용하는 방법은 TTP에게 교집합의 결과가 노출이 되는 문제가 발생되며, 참여자들이 TTP를 신뢰할 수 없는 상황이 발생될 수 있다. 또 TTP를 설정할 수 없는 상황에서는 프라이버시보호 다자간 교집합 연산이 불가능하다.

TTP가 없는 다자간 교집합 연산은 Freedman 등[2]에 의해 공개키 기반의 준동형 암호시스템(homomorphic cryptosystem)을 사용하여 참여자의 데이터를 다항식으로 표현하여 전달하는 방법이 제안되었고, Kissner과 Song에[3] 의해 일반적인 집합 연산과 다자간 모델로 확장되었다. Kissner과 Song은 Freedman 등의 결과를 확장하기 위해 참여자의 데이터를 표현한 다항식에 차수가 같거나 큰 랜덤 다항식을 곱하여 이들을 더하는 방법을 사용하였다. 하지만 이들의 연구는 낮은 확률로 부정확한 결과를 낼 수 있는 단점이 있었으며, 전처리와 같은 연산과 공개키 방식을 사용하여 많은 연산량을 필요로 한다.

최근에는 Hazay와 Lindell[1]에 의해 TTP가 없는 상황에서 스마트카드 사용하여 양자간 교집합 연산을 할 수 있는 모델이 제안되었다. 이 모델은 TTP를 설정할 수 없는 상황에서 스마트카드의 보안 요소를 사용하여 양자간 교집합 연산을 할 수 있다. 또한 이론적으로는 안전하나 실제로 구현이 어려운 일방향함수를 기반으로 한 모델의 단점을 의사난수치환(Pseudorandom permutations, PRP)을 사용한 실질적인 모델로 보완하였다. 하지만 단순히 Hazay와 Lindell[1]의 양자간 교집합 연산을 다자간 교집합 연산으로 확장하게 되면 중간과정에서 프로토콜 참여자들의 개인정보가 노출되는 문제점이 발생된다.

* 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임 (20100024870)

본 논문에서는 별도의 스마트카드를 이용한 양자간 교집합 연산을 하는 Hazay와 Lindell[1]의 프로토콜을 이용하여 Commodity Server(CS)가 있는 다자간 교집합정보를 계산하는 프로토콜을 제안한다.

2. 프라이버시보호 교집합 연산

2.1 프라이버시보호 교집합 연산의 목적

양자간 프라이버시보호 교집합 연산은 두 참여자인 P_1과 P_2가 각각 자신의 개인정보 데이터로 구성된 집합 $X = \{x_1, \cdots, x_{n_1}\}$, $Y = \{y_1, \cdots, y_{n_2}\}$를 가지고 있을 때, 프로토콜을 수행한 후 어느 한쪽 집합에만 존재하는 원소는 노출시키지 않으면서 $X \cap Y$을 알아내는 것을 말한다.

다자간 프라이버시 보호 교집합 연산은 여러 참여자인 $P_1, P_2, \cdots, P_m$ $(m \geq 3)$들이 각각 자신의 개인정보 데이터로 구성된 집합인 $X_1=\{x_{1,1}, \cdots, x_{1,n_1}\}$, $X_2=\{x_{2,1}, \cdots, x_{2,n_2}\}$, $\cdots$, $X_m=\{x_{m,1}, \cdots, x_{m,n_m}\}$을 가지고 있을 때, 프로토콜을 수행한 후에 어느 한쪽 집합에만 존재하는 원소를 노출시키지 않으면서 $X_1 \cap X_2 \cap \cdots \cap X_m$을 알아내는 것을 말한다.

2.2 용어의 정의

프로토콜에서 사용되는 PRP의 입력 값의 집합과 출력 값의 집합 사이에는 일대일 대응의 성질을 만족한다. 이 성질을 이용하여 PRP인 f에 대해 입력 값 a, b에 대해 두 출력 값 $f(a) = f(b)$이면, $a = b$임을 알 수 있다. 마스터키가 고정된 암호알고리즘은 PRP로 간주 할 수 있으며 암호알고리즘은 AES[4]와 같이 안전성이 검증된 것을 사용한다.

스마트카드는 마이크로프로세서와 메모리를 내장하고 있어서 카드 내에서 정보의 저장과 처리가 가능하다. 프로토콜에서 사용되는 스마트카드는 외부프로그램의 접근에 대해 완벽한 보안체계를 제공한다. 스마트카드는 물리적으로 칩에 직접적인 침해를 가하는 공격 등으로 보안에 이상이 생기지 않으며, FIPS 140-2[5] 레벨3 또는 4수준의 인증을 통과한 것을 사용한다.

Semi-Honest 모델은 프로토콜에 참여하는 참여자들이 모두 프로토콜의 모든 규칙을 준수하되 각자 상대방이 보내는 메시지와 프로토콜의 중간 결과를 저장하여 두었다가 가능하다면 그 정보를 이용하여 상대방의 데이터를 알아내려는 시도를 할 수 있는 상황을 의미한다.

참여자들은 개인정보 데이터 집합을 가지고 프로토콜에 참여하며, CS는 참여자들의 암호화된 데이터 집합을 받아 교집합 연산 후 계산결과를 참여자들에게 준다. CS는 프로토콜의 모든 규칙을 준수한다. CS는 병렬 컴퓨팅 구성요소를 사용하여 연산을 하며, 하나의 컴퓨팅 요소에 문제가 발생하더라도 전체적인 연산의 신뢰성을 유지할 수 있다. 또한 Semi-Honest모델에서 모든 참여자들이 잠재적인 공격자가 될 수 있으므로 CS는 어느 참여자와도 협잡하지 않는다.

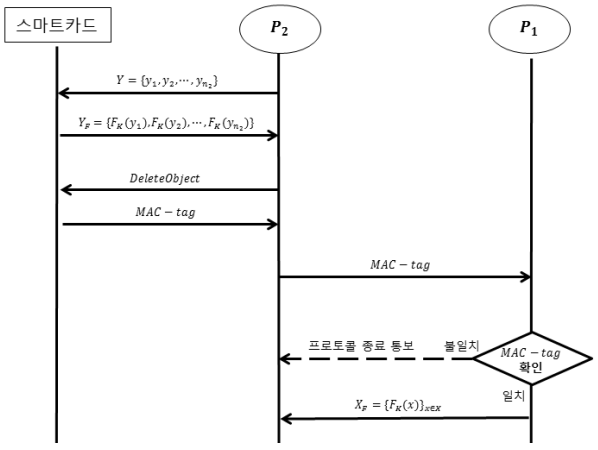
2.3 Hazay와 Lindell의 프라이버시보호 교집합 연산

Hazay와 Lindell[1]의 스마트카드를 이용한 프라이버시보호 교집합 연산은 프로토콜에 참여하는 두 참여자 P_1, P_2가 TTP가 없는 상황에서 각각이 가지고 있는 개인정보 데이터집합에서 어느 한쪽 집합에만 존재하는 원소는 노출시키지 않으면서 P_1, P_2가 각각 가지고 있는 개인정보 데이터집합의 교집합을 P_2가 스마트카드를 사용하여 계산한다.

P_1, P_2는 각각 n_1, n_2개의 개인정보 데이터로 구성된 집합 $X = \{x_1, \cdots, x_{n_1}\}$, $Y = \{y_1, \cdots, y_{n_2}\}$를 갖고 있으며, P_1은 프로토콜에서 사용되는 스마트카드를 P_2에게 발급해 주고, 스마트카드를 받은 P_2는 P_1과 자신의 개인정보 데이터집합에서 교집합을 계산한다. 이때, P_2는 Y에는 포함되지 않는 X의 원소인 P_1의 개인정보들을 알아낼 수 없다.

프로토콜은 P_1이 P_2가 가지고 있는 개인정보 데이터집합 Y의 원소의 개수인 n_2를 정확하게 확인할 수 있다는 가정을 하고 시작한다.

프로토콜을 시작하기 전 P_1은 P_2에게 보낼 스마트카드에 키 K, K_{init}와 사용횟수 n_2를 설정하여 입력한다. K는 스마트카드에는 PRP로 간주될 암호알고리즘에 사용된다. n_2는 P_2가 스마트카드에 있는 PRP를 사용할 수 있는 횟수를 나타내며, n_2회 이후에는 PRP를 사용하지 못한다. K_{init}는 P_2가 PRP를 n_2회 사용한 후 PRP에 내장된 K를 삭제하고 올바르게 삭제되었는가를 확인하기 위한 메시지 생성 용도로 사용된다.



(그림 1) Hazay와 Lindell의 프로토콜

스마트카드를 발급 받은 P_2는 스마트카드에 K와 K_{init}가 입력되어 있는지 확인한다. P_2는 K와 K_{init}의 구체적인

정보는 알 수 없으며, 입력되어 있는지 여부만 확인가능하다. 만약 K와 K_{init}가 입력되어 있지 않으면 프로토콜은 중단 된다.

P_2가 스마트카드를 발급받은 후 프로토콜 과정은 (그림 1)과 같다. P_2는 데이터들을 스마트카드에 내장되어 있는 PRP($F_K(\bullet)$)를 사용하여 $Y_F=\{(y,F_K(y))\}_{y\in Y}$를 구한다. P_2는 PRP를 n_2회만 사용할 수 있으며, n_2회 이상은 PRP를 사용할 수 없다.

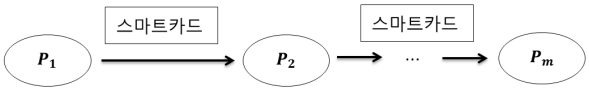
n_2회 PRP를 사용한 P_2는 스마트카드에 DeleteObject명령을 입력하여 PRP에 고정된 K를 삭제한다. 이 과정에서 스마트카드는 K_{init}를 사용하여 *MAC-tag*값을 생성한다. 이때 생성되는 *MAC-tag*값은 P_1이 PRP에 K가 삭제되었다는 것을 확인 하는 용도로 사용된다. P_2는 스마트카드에서 생성된 *MAC-tag*값을 P_1에게 보낸다.

P_1은 P_2에게 받은 *MAC-tag*값이 K_{init}를 사용하여 올바르게 만들어졌는지 확인을 한다. 올바르게 만들어졌으면 P_1은 스마트카드에 있는 K가 삭제되었음을 인증하며 프로토콜을 계속 진행하고, 그렇지 않으면 P_2에게 프로토콜 종료를 통보한다.

*MAC-tag*값을 확인한 P_1은 가지고 있는 PRP를 이용하여 자신의 개인정보데이터 X로 $X_F=\{F_K(x)\}_{x\in X}$를 계산하고 X_F를 P_2에게 전송한다.

X_F를 받은 P_2은 $X\cap Y = \{y|F_K(y)\in X_F\}$를 계산함으로써 프로토콜은 끝이 난다.

3. 다자간 교집합 연산 프로토콜



(그림 2) 양자간 교집합 연산을 여러 번 적용한 경우

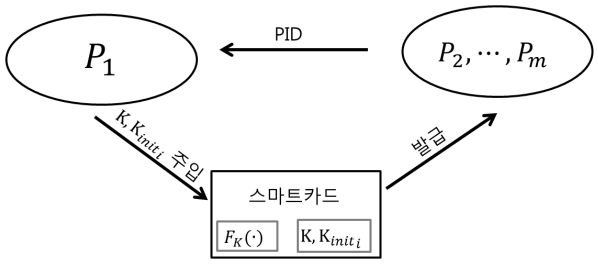
2장에서 설명한 Hazay와 Lindell의 프로토콜을 (그림 2)에서 보는 것과 같이 여러 번 적용하여 다자간 교집합 연산을 할 수 있다. m명의 프로토콜 참여자 $P_1, P_2, \cdots, P_m$들은 차례로 양자간 교집합 연산을 진행한다. P_1은 P_2에게 스마트카드를 발급해주며, 2장의 프로토콜을 진행하여 P_2는 P_1과 P_2 공통 정보를 구할 수 있다. 다시 P_2는 P_3에게 스마트카드를 발급해주며, P_2는 P_1과 P_2 공통 정보를 PRP에 입력하여 출력 값을 P_3에게 보냄으로써 P_3는 P_2과 P_3 공통 정보를 구할 수 있다. P_m까지 $m-1$회 반복하게 되면 P_m은 프로토콜 참여자들의 공통 정보를 구할 수 있다. 하지만 이 경우에는 다자간 교집합 연산의 목적과는 달리 P_2가 m명의 프로토콜 참여자들의 공통정보이외에 P_1과의 공통 정보를 알 수 있으며, P_3도 m명의 프로토콜 참여자들의 공통정보이외에 P_1과P_2의 공통 정보와 자신이 가지고 있는 정보의 공통부분을 알 수 있는 문제점이 발생한다.

본 논문에서는 앞에서 언급한 문제점들을 보완하기 위하여 CS의 도움을 받아 스마트카드를 이용한 프라이버시 보호 다자간 프로토콜에 대해 언급한다.

3.1 스마트카드 발급 과정

스마트카드 발급자인 P_1은 자신을 제외한 $m-1$명의 프로토콜 참여자 $P_2, \cdots, P_m$들이 각각 가지고 있는 개인정보로 구성된 집합 $X_2, \cdots, X_m$의 원소의 개수($n_2, \cdots, n_m$)를 정확하게 확인할 수 있다는 가정을 한다.

스마트카드 발급 과정은(그림 3)과 같다. 프로토콜을 시작하기 전 참여자 $P_2, \cdots, P_m$들은 P_1에게 각자의 사용자 영구신원(Permanent ID, PID)를 보낸다. P_1은 PID를 스마트카드 발급대장에 입력한 뒤 P_2에게는 스마트카드에 키 K, K_{init_2}와 사용횟수 n_2를 입력하여 발급하고, P_3에게는 스마트카드에 키 K, K_{init_3}와 사용횟수 n_3를 입력하여 발급하고, …, P_m에게는 키 K, K_{init_m}와 사용횟수 n_m를 입력하여 발급한다. 입력된 파라미터들의 사용은 2장에서와 같다.



(그림 3) 스마트카드 발급 과정

스마트카드에 입력되는 키 K는 PRP을 결정하므로 CS에서 교집합 연산을 하기 위해서는 모든 참여자에게 같은 값이 입력되어야 한다. 반면에 PRP에 고정된 K를 삭제하고 확인 메시지를 생성하는데 사용되는 키 K_{init_i}는 참여자들이 동일한 값을 사용하면, 모든 참여자들의 MAC_i-*tag* 값이 같아지는 보안상의 문제가 발생되어 모두 다른 값을 스마트카드에 입력한다.

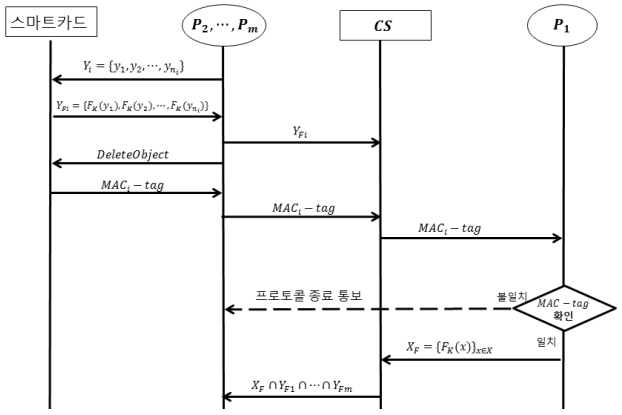
3.2 다자간 교집합 연산 프로토콜

프로토콜은 참여자 $P_2, \cdots, P_m$들이 P_1으로부터 스마트카드를 발급받은 이후 시작된다. 참여자들은 발급받은 스마트카드에 K와 K_{init_i}가 입력되어 있는지 확인하며, 참여자들은 K와 K_{init_i}의 구체적인 정보는 알 수 없으며 입력되어 있는지 여부만 확인가능하다. 만약 K와 K_{init_i}가 입

력되어 있지 않으면 프로토콜은 중단 된다. 프로토콜을 계속하기를 원하면 P_1에게 키가 입력되어 있지 않다는 것을 알려주고 카드를 재발급 받는다.

K와 K_{init_i}가 입력되었다는 것이 확인이 되면 P_1에게 프로토콜 참여를 알리고 프로토콜을 진행 한다.

참여자들이 스마트카드를 발급받은 후의 프로토콜의 과정은 (그림 4)과 같다. 참여자들은 각자의 데이터들을 스마트카드에 내장되어 있는 PRP($F_K(\bullet)$)를 이용하여 각각 $Y_{Fi}=\{(y,F_K(y))\}_{y\in Y_i}$를 구한다. 참여자들은 PRP를 정해진 횟수만큼 만을 사용할 수 있으며, 그 이상은 PRP를 이용할 수 없다. Y_{Fi}를 구한 참여자들은 CS에게 그 값을 보낸다.



(그림 4) 다자간 교집합 연산 프로토콜

스마트카드에서 정해진 횟수만큼 PRP를 이용한 참여자들은 각각 가지고 있는 스마트카드에 DeleteObject명령을 입력하여 PRP에 고정된 K를 삭제한다. 스마트카드는 K_{init_i}를 사용하여 MAC_i-tag값을 생성한다. MAC_i-tag값은 P_1이 참여자들의 스마트카드에 있는 PRP에 고정된 K가 삭제되었다는 것을 확인하는데 사용된다. MAC_i-tag값을 받은 참여자들은 이 값을 CS에게 보낸다.

CS는 P_1을 제외한 모든 참여자들에게서 MAC_i-tag값을 받으면 P_1에게 MAC_i-tag값들을 보낸다. P_1은 MAC_i-tag값들이 K_{init_i}들을 사용하여 올바르게 만들어졌는지 확인을 한다. 만약 올바르게 만들어지지 않았으면 프로토콜 종료를 참여자들에게 통보하며, 올바르게 만들어졌으면 프로토콜을 계속 진행한다.

MAC_i-tag값들을 확인한 P_1은 가지고 있는 PRP를 이용하여 $X_F=\{F_K(x)\}_{x\in X}$를 계산하고 CS에게 전송한다.

X_F를 받은 CS는 모든 참여자들로부터 받은 집합을 가지고 교집합 $Z=X_F\cap Y_{F2}\cap\cdots\cap Y_{Fm}$을 계산하고 모든 참여자들에게 교집합의 값을 보낸다.

Z값을 받은 참여자들은 $\{y|F_K(y)\in Z\}_{y\in Y_i}$를 계산함으로써 프로토콜은 끝이 난다.

3.3 프로토콜의 안전성 분석 및 계산 효율성

본 논문에서 제시한 다자간 교집합 연산 프로토콜은 Semi-Honest모델에서 안전하다. 참여자들은 비밀키 K가 내장된 암호알고리즘인 PRP의 결과 값만을 얻을 수 있다.

또한 참여자들은 비밀키 K의 값을 알 수 없고, 다자간 교집합 연산결과만을 알 수 있으므로 다른 참여자들의 정보는 알 수 없다.

공개키 기반의 준동형 암호시스템 대신 대칭키 기반의 암호시스템을 사용한 PRP를 사용하여 연산의 계산량이 줄어 효율적이다.

4. 결론

본 논문에서는 스마트카드를 이용하여 다자간 교집합 연산을 하는 프로토콜을 제시하였다. 제시한 프로토콜은 TTP의 설정 없이 스마트카드의 잘 갖추어진 보안요소를 사용하여 양자간 교집합 연산뿐만 아니라 다자간 교집합 연산 프로토콜에 적용할 수 있다는 것을 보여준다.

제시한 모델은 일방향함수를 사용한 모델보다 구현이 쉬우며, 더 현실적이다. 또한 PRP를 사용하여 공개키 기반의 준동형 암호시스템보다 계산이 효율적인 장점을 가지고 있다.

본 논문에서 제시한 다자간 프라이버시보호 교집합 연산은 보험사기 방지시스템, 항공기 탑승 금지자 목록 검색, 의료 정보 검색, 전자투표 등에서 유용하게 적용할 수 있을 것이라고 판단된다.

현재 프로토콜이 공격자 모델이 Semi-Honest일때 안전성을 분석하였는데 Malicious모델로 공격자 모델을 확장하는 것은 이후 좋은 연구 주제가 될 것이다.

참고문헌

[1] C Hazay, Y Lindell "Constructions of Truly Practical Secure Protocols using Standard Smartcards" 15th ACMCCS 2008.

[2] Michael Freedman, Kobi Nissim, and Benny Pinkas, Efficient private matching and set intersection. In Proceedings of Advances in Cryptology (Eurocrypt'04), page 1-19, 2004 .

[3] Les Kissner and Dawn Song. Privacy-preserving set operations. In Advances in Cryptology (Crypto'05), page 241 - 257, 2005

[4] "Report on the Development of the Advanced Encryption Standard(AES)", http://www.csrc.nist.gov/encryption/aes/.

[5] "Security Requirements for Cryptographic Modules", http://csrc.nist.gov/groups/STM/cmvp/standards.html#02