인터넷 광고에서 안전하고 효율적인 측정방법
(A Secure and Efficient Metering Scheme for Internet Advertising)

김 순 석 † 신 제 용 † 김 성 권 ††
(Soon Seok Kim) (Je Yong Shin) (Sung Kwon Kim)

요 약 본 논문에서는 인터넷 광고에서 웹사이트를 방문하려는 고객과 서버들간에 상호작용을 측정하기 위한 안전하고도 효율적인 측정 방법에 관해 고려한다. 대개의 인터넷 광고는 많은 고객들과 서버들 그리고 서버들에 의해 제공되는 고객들의 수에 관한 측정 정보를 수집하는 감사 기관의 상호작용으로 이루어진다. 측정 방법은 항상 고객의 방문 횟수를 과장되게 조작하려는 서버의 악의 있는 시도라든가 측정 과정을 붕괴하려는 고객들의 시도로부터 항상 안전해야만 한다. 따라서, 본 논문에서는 여러 암호학적인 기법들에 기반하여, 안전하고 효율적이며 정확하고 강건한 측정 방법을 제안하고자 한다.

키워드 : 인터넷 광고, 암호 프로토콜

Abstract In this paper, we consider a secure and efficient metering scheme to measure the interaction between clients and servers in internet advertising. In most cases internet advertising is composed of clients, servers, and an audit agency who collects a metering information about the number of clients that were served by each server. The metering scheme should always be secure against fraud attempts by servers which maliciously try to inflate the number of their clients and against clients that attempt to disrupt the metering process. So we suggest secure and efficient metering schemes, based on some cryptographic techniques, which are also very accurate and robust.

Key words : Internet advertising, Cryptographic protocol

1. 서 론

인터넷 광고는 말 그대로 인터넷에 홈페이지를 만들어 두고 방문객들을 끌어들이기 위해 다른 대중 미디어 혹은 다른 인터넷상의 홈페이지에 광고를 통해 홍보나 부가적인 이벤트 또는 서비스를 하는 등 멀티미디어적인 특성을 살린 커뮤니케이션활동을 말한다. 인터넷이 광고미디어로서 주목받기 시작한 것은 월드와이드웹의 등장과 함께였는데 웹에서의 광고(web advertising)를 이용함으로써 갖는 여러 장점은 참고문헌 [1]에 나와 있다.

인터넷 광고가 지닌 많은 장점들에도 불구하고 현재까지 미해결분야로 남아있는 문제가 하나 있다. 이것은 '마케팅 효과의 측정', 다시 말해, 보다 자세하고 정확하게 광고 효과를 측정하는 것이다. 월드와이드웹을 이용하는 인터넷 광고에서 그 효과를 측정하기 위해서는 무엇보다 광고가 게재된 웹사이트에 방문객들이 얼마만큼 방문했는가가 중요한 이슈가 된다.

웹사이트의 방문객 수에 대한 측정은 광고 효과의 측정이란 면에서 광고주뿐만 아니라 광고를 게재할 수 있는 공간(웹사이트)을 제공해주는 웹사이트 발행자 모두에게 매우 중요한 의미를 지닌다. 기존 전통적인 광고에서 측정단위로 CPM(Cost Per Mile의 약자로 인쇄 미디어 광고에서 많이 사용되는 지표로 1000회 동안의 광고 노출을 기준으로 가격을 책정하는 단위이다)을 사용한 것과는 달리 인터넷 광고에서는 배너(banner)광고를 직접 클릭(click)한 횟수(이를 click-through라 부른다)를 측정함으로써 광고에 대한 효과를 측정할 수 있다. 그 이외의 방법들도 있긴 하지만, 본 논문에서는 여러 종류의 측정 단위보다는 이러한

· 본 연구는 정보통신부의 정보통신 우수 시범학교 지원(우수00-34)으로 수행되었음.
† 비 회 원 : 중앙대학교 컴퓨터공학과
sskim@alg.cse.cau.ac.kr
sly8282@alg.cse.cau.ac.kr
†† 종신회원 : 중앙대학교 컴퓨터공학과 교수
skkim@cau.ac.kr
논문접수 : 2001년 8월 3일
심사완료 : 2002년 1월 14일

측정 단위들에 맞게 얼마나 정확하고 안전하게 측정했는가를 중점적으로 연구하고자 한다. 따라서, 측정 단위가 클릭이든 페이지든 히트율이든 그것은 논외로 하려 한다.

그러나 이러한 웹사이트 측정은 현재로서는 한계를 가지고 있다. 예를 들어, 1,000명의 방문객이 배너광고를 각각 10번 보았는지, 아니면 10,000명의 방문객이 같은 광고를 한번씩만 보았는지를 정확히 구별할 수가 없다. 뿐만 아니라, 어느 방문객이 특정 광고를 보았는지에 대해서도 정확히 알 수 없다. 결국 웹사이트 발행자는 방문객이 등록한 통계적인 데이터, 즉 연령, 수입, 성별, 직업 등에 전적으로 의존할 수밖에 없다. 만일 방문객이 등록한 통계자료마저 없다거나 아니면 아예 방문객이 등록조차도 하지 않는다면 어떻게 할 것인가? 대개의 방문객들은 아마도 등록자체가 지루하거나 까다롭다는 이유로 아예 방문을 꺼려할는지도 모른다. 비단, 문제점은 여기서 그치지를 않는다. 만일 웹사이트 발행자가 자신의 이익 즉, 보다 많은 광고를 자신의 웹사이트에 유치할 목적으로 광고주를 속여 조회한 횟수에 대한 통계를 부풀린다면 어떻게 할 것인가? 과연 광고주는 이들 웹사이트 발행자를 얼마나 믿을 수 있을 것인가? 더군다나 직접 대면을 통해 의사를 주고받는 현실 세계와는 달리 간접 대면 형태의 인터넷상에서 이것은 받아들여지기가 힘들다. 따라서, 광고주가 믿을만한 정확한 통계자료 즉, 방문객의 조회 횟수에 대한 기록을 웹사이트 발행자가 제시해야 한다.

본 논문의 2장에서는 현재 웹 광고에서 안전하고 효율적인 조회 횟수 측정을 위한 기본 모델과 이에 따른 요구사항을 설명하고 3장에서는 이와 관련한 연구 동향을 4장에서는 새로운 측정 방법을 제안한 후에 5장을 끝으로 결론 및 향후 연구방향에 대해 논하고자 한다.

2. 측정 모델과 요구사항

인터넷을 통해 이루어지는 광고는 [그림 1]과 같이 네 구성원들의 상호작용으로 이루어진다. 먼저 광고주가 자신의 회사나 혹은 상품을 광고하기 위해 광고 대행사에 업무를 위탁하면(1) 광고 대행사는 웹사이트 제공자(이하, 서버라 부른다)에게 일정 기간(예, 하루 혹은 한달) 동안의 배너 광고(혹은 버튼이나 웹페이지 광고) 게재를 요청하고 서버의 행동을 감시한다(2). 그 후, 일정 기간 동안 서버가 고객(이하 방문자라 부른다)들로부터의 조회 횟수를 측정하여(3) 측정된 결과를 광고 대행사에 알린다(4). 이 때 광고 대행사는 서버로부터 받은 측정 결과가 정확한 것인지를 검증한 후에 이를 광고주에게 통보한다. 광고 대행사는 신뢰할 수 있는 제 3의 기관이어야 하며 서버가 보낸 측정결과가 정확한 지를 검증하는 역할을 주로 수행하므로, 앞으로는 이 광고 대행사를 감사기관(audit agency)이라 부르기로 한다.

인터넷 광고에서 방문자 측정방법이 갖추어야 할 요구사항들은 다음과 같다[2].

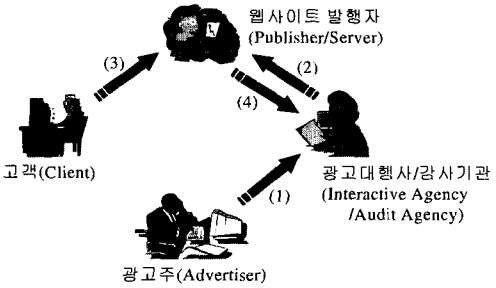

그림 1 측정 모델

안전성 : 서버가 방문자수를 부풀리는 것을 막을 수 있어야 한다. 즉, 서버는 자신을 방문한 방문자수를 입증 할 수 있어야 하고, 방문자가 서버에게 잘못된 정보를 제공함으로써 서버가 방문자수를 입증하는 일을 방해하는 것을 막을 수 있어야 하며, 또한 그런 문자를 찾아낼 수 있는 방법을 제공해야 한다.

효율성 : 감사기관과 방문자 사이에는 초기 통신 외에는 별다른 통신이 일어나지 않아야 하며 일정한 통신패턴을 유지해야한다. 혹은 각 구성원 측에서의 계산량과 메모리의 오버헤드를 최소화하여 수행 속도를 높일 수 있어야 한다.

정확성 : 대략적인 방문자수의 측정이 아니라 방문한 사용자의 정확한 측정 결과를 제공해야한다.

프라이버시 : 방문자의 프라이버시를 보호해야한다. 서버는 자신을 방문한 방문자로부터 얻은 정보를 가지고 그 방문자에 대한 최소한의 정보도 알아내지 못하게 해야한다.

3. 관련 연구 동향

3.1 기본적인 측정 방법

지금까지 제안된 여러 가지 방법들 중 가장 기본적인 측정 방법은 전자서명을 이용한 방법이다. 이 방법은 [그림 2]에서 보는 바와 같이 감사기관이 앞으로 방문할 방문자에게 인증된 서명키를 분배한다. 그 후 방문자가 해당 서버를 방문할 때마다 방문자는 전자 서명 프로토콜을 수행한다. 그리고 서버는 방문자들로부터 받은 서

명 목록을 저장하였다가 나중에 광고주에게 확인을 받는다. 이 방법은 각 방문자의 서명 목록을 확인할 수 있으므로 정확하게 사용자 수를 측정할 수 있는 장점이 있다. 하지만 방문자 수가 수 만 명 이상일 경우에는 전자서명을 수행하기 위한 비용이 커져서 효율적으로 측정하기가 힘들다. 또한 각 개인에 대한 서명을 확인하므로 개인의 프라이버시를 침해할 수 있는 단점이 있다.

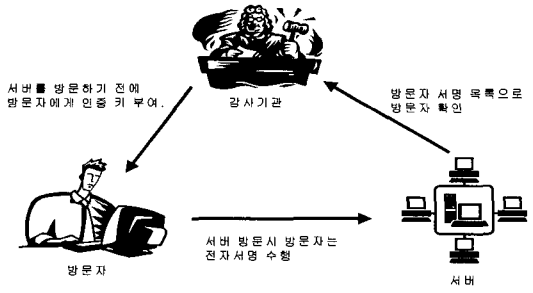

그림 2 전자서명을 이용한 기본적인 측정 방법

3.2 시스템의 부하를 줄인 측정 방법

이 방법은 기본적으로 시스템의 부하를 줄이기 위해 방문자에 대한 인증이나 제3의 신뢰기관을 필요로 하지 않는다. 즉, 방문자와 방문 확인자인 서버사이에서 프로토콜이 진행된다. 다만 일정 시간 동안 방문자는 주어진 계산을 수행하고 그 수행에 대한 결과 값만을 서버에게 돌려줌으로써 전체 시스템의 부하를 줄이는 방법이다. 이 방법은 특히 신뢰할 수 있는 제 3자를 가정하고 있지 않기 때문에 방문 확인자 측에서 언제든지 마음만 먹으면 방문자의 방문 횟수를 조작할 수 있다는 큰 단점이 있다. 그러나, 그 만큼 시스템의 부하가 줄어든다는 장점이 있기 때문에 인터넷 광고 이외의 응용분야(예를 들어, 방문자와 방문 확인자와의 1대 1 측정 등)에 적용할 수 있는 가능성이 있다.

Dwork와 Naor[3]가 제안한 방법은 방문자가 서버를 방문할 시에 일정한 계산량의 처리를 한 뒤에 방문자에게 해당 사이트를 방문할 수 있는 허가를 주는 방법을 제안하고 있다. 즉, 방문자가 처리해야 할 함수를 Fiat-Shamir[4]의 방법에 기반하여 유연성 있는 pricing function으로 정의하여 방문자에게 일정량의 함수를 처리한 후 그 결과 값만을 돌려 받는 방식이다. 이 방식을 간략히 설명하면 다음과 같다. 먼저 큰 소수 p, q를 임의로 선택하여 $N=p*q$와 $y_1(=x_1^2 \ mod \ N),...,y_k(=x_k^2 \ mod \ N)$를 계산한다. 이때 값 $x_1^2,...,x_k^2$은 서버로부터 사전에 부여받은 값이다. 일방향 해쉬함수를 h, 방문자가 임의로 선택한 랜덤 정수를 r, 그리고 $h(x, r^2)=b_1,...,b_k$라 하자. 이때 $b_1,...,b_k$ 또한 서버로부터 사전에 부여받은 값으로 k비트 길이를 갖는다. 방문자는 pricing function f_s를 이용하여 다음 식(1)을 만족하는 값 z와 r^2을 계산한 후 서버에게 전달한다.

$$f_s(x)=(z,\ r^2),\ z^2 = r^2\ x^2 \prod_{i=1}^{k} y_i^{\ b_i} \bmod N \qquad (1)$$

이 두 값을 전달받은 서버는 위의 준 식(1)이 만족하는지를 검증하여 만일 만족할 경우, 가격을 책정한 후 웹사이트 방문에 대한 허가를 부여한다. 또한 이 방법은 예를 들어 전자 메일을 보내고자 하는 송신자로 하여금 메시지의 양에 비례하여 그 만큼 위의 pricing function을 수행하도록 부담을 줌으로써 스팸메일을 방지할 수 있는 방법에도 이용할 수 있다. 단, 메일 수신자로부터 사전에 권한을 획득받은 합법적인 송신자로부터의 메일 송신에 있어서는 shortcut라 하여 간단한 계산만으로 메일을 보낼 수 있도록 하고 있다.

반면에, Franklin과 Malkhi[5]가 제안한 방법은 일정 시간동안 계산을 수행하는 시간함수(timing function)를 이용하여 방문자가 사이트를 방문했을 경우 결과를 얻기 위해서 일정 시간동안 계산을 한 후 이 결과 값으로서 방문자가 이 사이트에 접속했는지를 알 수 있는 방법이다. 이 방법은 앞서 언급한 방법을 개선하여 현재 사용하고 있는 브라우저에서 수행이 가능하고, 방문자의 등록단계를 제거하여 방문자들에게 불편함을 덜어 줄 수 있으며, 방문자가 획득한 콘텐츠들을 원래의 웹사이트에서 얻었는지 아니면 캐쉬나 프록시 서버에서 얻었는지를 알 수 있다. 그리고 가장 중요한 점은 사용자의 ID를 기반으로 하지 않기 때문에 방문자의 익명성을 보장할 수 있으며, 전자서명 방식을 사용하지 않기 때문에 보다 효율성을 갖는다는 특징이 있다. 그러나 이 방법 역시 인터넷 광고보다는 방문자와 서버간의 1:1 측정의 응용분야에 유용하다.

3.3 Naor와 Pinkas[2]가 제안한 방법

이 방법은 먼저 서버가 일정 기간 동안에 방문자의 방문을 받는다. 그 후, 방문자의 수가 감사 기관과 약속한 수와 같아지면 이 기관에 알려서 방문자 수를 확인받는 방법이다. 비밀공유기법(secret sharing)을 사용하여 여러 명의 방문자에게 미리 키를 주고 일정 시간(time frame, t)동안 방문자들이 서버에 접속할 경우에 이 키 값을 서버에게 주게된다. 그 후 서버가 일정 수 k명의 방문자들을 받게 되면 방문자들로부터 받은 키 값들을 가지고 방문자 수에 대한 증거를 생성하여 감사기관에 그 값을 넘겨준다. 이 때, 감사기관은 서버로부

터 넘겨받은 증거와 자신이 생성한 값이 일치하는지를 계산하여 증명하는 방법이다. 이 방법은 암호학적으로 안전함을 증명할 수 있고 원래의 통신 패턴을 유지 할 수 있다는 것이 장점이다. 그러나, 기본적으로 비밀공유 기법을 이용하기 때문에, 만약 1만 명의 방문자가 있어야만 확인 받을 수 있다고 할 경우, 1만 명에 가까운, 예를 들어 9990명이 방문한 사실도 확인 받을 수 없다는 단점이 있다. 따라서, 실제 적용을 위해서는 이 부분에 수정이 요구된다.

3.4 Masucci와 Stinson[6]이 제안한 방법

Naor와 Pinkas 방법의 단점은 정해진 수의 방문보다 약간 적은 수의 방문이 발생했을 경우에 감사 기관은 서버의 방문자 수를 확인 받을 수가 없었다는 것이다. 이를 개선하기 위해서 Masucci와 Stinson은 최소 임계 방문자 수의 개념을 도입했다. 즉, 정해진 방문자 수보다는 작고 최소 임계 방문자 수보다는 클 경우에 서버는 자신의 방문자 수를 확인 받을 수 있는 방법을 제안했다. 이 방법은 Naor와 Pinkas 방법에 비해 방문자 수의 측정상 훨씬 유연성을 갖는다는 장점이 있다. 그러나, 그러한 유연성을 제공하는 만큼의 서버측 부담이 있다. 따라서, 앞으로 이 부분에 보다 효율성을 개선하는 방향으로의 연구가 필요하다.

이외에 최근 W. Shin과 K. H. Rhee가 제안한 참고 논문 [7]에서도 관련한 연구가 진행된 바 있다. 이 방법의 특징은 본 논문에서 제안하고 있는 방법과 유사하게 일방향 해쉬함수와 XOR 연산을 증거 생성시에 이용하고 있다. 그러나, 이 방법은 계산량적인 측면에서 효율은 있으나 유연성이 미비하다. 왜냐하면 전체 방문객의 수가 n명이라고 할 때, n명의 방문자들이 모두 방문했을 때에만 서버가 증거를 생성할 수 있으며 또한 감사기관이 이에 대한 검증을 할 수 있기 때문이다. 그러나 제안하는 방법은 n명 가운데 임의의 k명에 대해 증거생성과 검증이 가능하다.

4. 새로운 측정 방법 제안

본 논문에서 제안하는 방법을 개괄적으로 설명하면 다음과 같다. 먼저, 앞서 설명한 바 있는 Naor와 Pinkas의 방법대로 감사기관(A)과 방문자(C) 그리고 서버(S)를 가정하고, 감사기관이 사전에 각 방문자들에게 서로 다른 비밀정보와 임시 ID인 UID를 생성하여 n명의 방문자들에게 나눠준다. 그 후, 방문자가 임의의 시간에 서버를 방문하게 되면 감사기관으로부터 받은 자신의 비밀정보와 UID를 서버에 전달한다. 서버는 각 방문자들의 비밀정보와 UID들을 모아서 '증명서'와 '출석부'를 만들고 이들을 감사기관에 제출한다. 감사기관은 서버로부터 전달받은 '출석부'를 이용하여 '증명서'의 진위여부를 가린다.

4.1 제안하는 측정 방법에 대한 요소

- A : 신뢰할 수 있는 감사기관.
- S : 서버 또는 서버의 ID. 여러 개의 서버가 있을 수 있다.
- C : 방문자 또는 방문자 ID. 모두 n명의 방문자들이 있다고 가정한다.
- UID_c : 방문자 C의 임시 ID. n명의 방문자들은 모두 다른 UID_c를 가진다. UID_C는 최초에 감사기관이 생성하는 실제 ID가 아닌 임시 ID로, 4.2절에서 제안하고 있는 측정방법의 [단계 3]에서 서버가 $MASK$ 정보를 생성하는데 이용되며 또한, 제 3자는 이 값을 인지할 수 없어야 한다. 이를 위해 감사기관은 사전에 서버와 임시 ID에 대한 정보를 교환해야 한다. 예를 들어, 임시 ID를 생성하기 위해 교환되는 정보가 해쉬 함수 f와 $seed$값 s라고 가정하자. 이때, 최초로 생성되는 UID_C는 $f(s)$, 두 번째 UID_C는 $f(f(s))$, 세 번째 값은 $f(f(f(s)))$와 같은 식으로 하여 n명의 방문자들에 대한 UID_C를 생성할 수 있다. 또한, 이 UID_C는 해쉬 함수의 성질에 의해 랜덤하고 유일하며 서버가 사전에 미리 계산하여 정렬(sorting)한 후, 이후에 사용될 $MASK$ 정보 생성을 위해 보관한다.
- t time frame. 측정기간으로 서버 S와 사전에 결정한다. 예를 들어, 하루나 혹은 한달 단위로 다양한 측정이 가능하다.
- $MASK$: 서버가 생성하는 n비트 벡터(초기에는 모두 0이다). 방문자들의 UID_c들을 모두 크기 순으로 정렬하여 정렬 리스트를 만든다. $rank\ (UID_c)$는 정렬 리스트에서 UID_c의 등수(rank)를 나타낸다. 방문자 C가 방문하면 그의 UID_c를 정렬 리스트에서 이진탐색으로 찾아 $rank(UID_c)$를 정하고, $MASK[rank(UID_c)] \leftarrow 1$로 한다. 즉, $MASK$는 일종의 '출석부'에 해당한다. n비트 사이즈이나, 만일 방문객들이 많은 경우는 비트 사이즈가 커질 수 있으므로 실제 전송 시에 효율성을 위해 압축코드를 활용할 수도 있다.
- H : 충돌회피 일방향 해쉬함수.

4.2 제안하는 기본 측정 방법

본 논문에서 제안하는 기본 측정 방법은 아래와 같다.([그림 3] 참조)

[단계 1] *Initialization*

감사기관은 n명의 방문자 C들에 대한 UID_C와 자

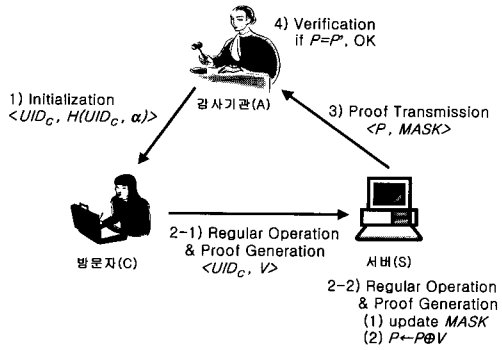

그림 3 제안하는 기본 측정 방법

신만이 알고 있는 비밀정보인 임의의 난수 α를 생성한 다음 이 두 값을 연접하여 일방향 해쉬함수 H를 수행, 그 결과 $H(UID_C, \alpha)$와 UID_C를 각 방문자들에게 랜덤 순서로 전달한다.

[단계 2] *Regular Operation & Proof Generation*

방문자 C가 기간 t내에 서버 S를 방문하면 $V(=H(S, t, H(UID_C, \alpha))$를 계산하여 감사기관으로부터 전달받은 UID_C와 함께 서버에게 보낸다. 그리고 이 값을 전달받은 서버는 다음 두 가지 작업을 수행한다.

(1) 방문자 C로부터 전달받은 UID_C를 이용하여 $MASK$를 갱신한다. 즉, 앞서 설명한 대로 방문자들의 UID_c들을 모두 크기 순으로 정렬하여 정렬 리스트를 만든 다음, 방문자가 방문하면 C의 UID_c를 정렬 리스트에서 이진탐색으로 찾아 등수인 $rank(UID_c)$를 정하고 $MASK[rank(UID_c)] \leftarrow 1$로 한다.

(2) 현재의 증명서 P와 V를 XOR하여 새 증명서를 만든다. 즉, $P \leftarrow P \oplus V$를 계산한다. P의 초기값은 0이다.

[단계 3] *Proof Transmission*

측정기간인 t가 만료되면, 서버는 증명서 P와 현재까지 생성한 $MASK$를 감사기관에 보낸다.

[단계 4] *Verification*

감사기관은 서버로부터 전달받은 $MASK$를 이용하여 서버가 수행한 것과 동일하게 증명서 P'를 생성하여, 서버로부터 전달받은 P와 동일한지를 검사한다. 만일 이 두 값이 같다면 서버가 보낸 증명서 P를 증거로 받아들인다. 이때 P'는 서버로부터 전달받은 $MASK$를 보면 어느 방문자가 방문했는지를 계산을 통해 감사기관이 알 수 있으며, 이를 이용하여 서버가 수행한 것과 동일한 정보인 V들을 계산할 수 있다. 또한 계산의 효율성을 위해 이 값은 서버와 사전에 협의한 시간 t에 따라 미리 계산될 수 있다.

4.2.1 제안하는 기본 측정 방법에 대한 분석

[안전성]

제안하는 방법의 안전성은 해쉬함수의 일방향성과 충돌회피성에 기반한다. 즉, $y=H(x)$에서 y를 안다하더라도 x를 알기란 계산적으로 거의 불가능하다. 서버가 방문 횟수를 과장되게 조작하기 위해서는 방문하지 않은 방문자의 정보, V를 새로이 생성하여 증명서 P를 만들어야 한다. V를 생성하기 위해서는 기타 정보들(S, t, UID_C)을 안다하더라도 α를 모르기 때문에 계산할 수 없다. 따라서, $1/\alpha$의 확률에 대해 안전하다.

표 1 제안된 측정 방법들과의 비교(* : 감사기관의 초기(initialization)단계 계산량이 아닌 검증(verification)단계에서의 계산량임)

		전자서명법	Dwork & Naor[3]	Franklin & Malkhi[5]	Naor & Pinkas[2]	Masucci& Stinson[6]	제안한 방법
측정방법		공개키를 이용한 서명	Pricing Function	해쉬함수	다항식을 이용한 비밀분산기법	다항식을 이용한 비밀분산기법	해쉬함수와 MASK 정보
계산량	감사기관	방문자의 서명 인증	없음	없음	1번의 d-1차 다항식 계산*	1번의 d-1차 다항식 계산*	k-1번의 XOR 연산*
	서버	방문자의 서명 확인	방문자의 결과 저장	방문자의 결과 저장	1번의 k-1차 다항식 보간	1번의 k-1차 다항식 보간	k-1번의 XOR 연산과 k번의 이진탐색
	방문자	서명	Pricing 함수 수행	해쉬함수 수행	1번의 d-1차 다항식 계산	1번의 d-1차 다항식 계산	1번의 해쉬함수 수행
장점		정확성 안정성	효율성	효율성 익명성	정확성 안전성	정확성 안전성 유연성	정확성 안전성 효율성 유연성
단점		효율성	정확성	정확성	유연성	추가적인 네트워크 오버헤드	익명성

[효율성]

[표 1]에서 살펴보는 바와 같이, Naor와 Pinkas가 제안한 방법은 공개키 연산을 수행하는 전자서명 방식에 비해 비밀공유기법을 사용하므로 훨씬 효율적이다. 그러나 이 방법 역시 실제로 방문한 방문자의 수를 k라 할 때, 서버 측의 증거 생성에 있어 다항식에 대해 최소한 $k-1$차수만큼의 보간법(interpolation)을 수행해야한다. 이 경우 시간 복잡도는 $O(klog^2k)$로 알려져 있다 [8]. 위 [표 1]에서 d는 time frame의 수를 나타낸다. 그러나 본 논문의 경우는 보간법 대신 증거 생성에 있어 $k-1$번의 XOR 연산과 k번의 이진탐색을 수행하기 때문에 훨씬 효율적이다. 이때 시간 복잡도는 XOR 연산에 있어 $k-1$번의 약 150비트 정도의 벡터 수행과 이진탐색에 있어 전체 n명의 방문 대상자들에 대해 $O(klogn)$ 정도가 소요된다.

[프라이버시]

제안한 방법은 기본적으로 유일한 임시 아이디인 $UIDc$를 이용함으로써 방문자의 프라이버시를 보호한다. 이 $UIDc$는 감사기관이 랜덤 순서로 각 방문자들에게 제공하며, 서버는 단지 방문자가 임시로 사용한 $UIDc$만을 알 수 있을 뿐 실제 방문자가 누구인지는 알 수 없다. 임시 아이디는 별명(alias)과 같은 개념으로 사용자의 실제 아이디가 아닌 방문시에만 사용되는 별칭이다.

[유연성]

Naor와 Pinkas가 제안한 방법은 앞서 살펴본 바와 같이 방문자 수가 미리 정해진 임계값에 정확히 도달할 때에만 증거를 생성할 수 있었다. 그러나 본 논문의 경우는 이러한 임계값을 필요로 하지 않으므로 어떠한 수의 방문자들에 대해서도 자유로이 측정이 가능하다. 이것은 방문자들에 대한 일종의 출석부라 할 수 있는 $MASK$와 방문자들이 보낸 정보들에 대한 XOR 연산의 결과만을 보내기 때문에 가능하다. 만일 제안한 방법에서 time frame인 t를 제거하고 프로토콜을 진행할 경우 본 방법은 시간의 제약없이 무제한 이용될 수도 있다.

4.3 강건성을 위한 개선된 측정 방법

앞서 제안한 기본 방법에서 드물기는 하지만 방문자가 거짓 정보를 서버에 전달했을 경우에 이를 서버가 확인할 길이 없으므로 잘못된 증거를 산출할 수 있다. 즉, [단계 2]에서 $P \leftarrow P \oplus V$를 계산할 경우 어떤 V가 거짓 정보라면 올바른 P를 계산할 수 없다. 따라서 서버는 P를 생성하기 이전에 방문자로부터 받은 $<UID_C, V>$가 올바른 값인지를 알아 낼 수 있는 방법을 제공해야한다.

제안한 기본 방법에서 이러한 강건성을 만족시키기 위해서는 약간의 변형이 필요하다. 강건성의 주목적은 [단계 2]에서 방문자가 생성한 정보인 $V(=H(S, t, H(UID_C, \alpha))$를 변형없이 올바르게 서버에게 전달했는가를 검증하는 것이다. 만일 이 단계에서 검증된 결과가 잘못되었을 경우는 더 이상의 수행을 멈추고 해당 방문자의 측정에 대한 정보 생성을 취소해야 한다.

제안하는 방법을 간략히 설명하면 다음과 같다. 먼저 [단계 1]에서 감사기관이 자신만이 알고 있는 비밀 정보인 β를 생성한 다음, $w=H(\beta, V)$를 계산한다. 이때 V는 앞서 말한 $H(S, t, H(UID_C, \alpha))$로 만일 정해진 기간 t'에 특정 서버 S'를 방문자가 방문한다고 가정할 경우, 이 값은 미리 계산할 수 있다. V는 원래 기본 방법에서는 방문자가 임의의 기간 t에 임의의 서버 S를 방문할 경우에 계산된 값이다. 만일 특정 기간 t'동안에 특정 서버 S'을 방문한다고 가정할 경우 강건성을 구현하기가 훨씬 효율적인데, 왜냐하면 V를 사전에 감사기관이 고정적으로 계산할 수 있기 때문이다. 이 경우는 좀더 보안 강도를 높이기 위한 목적으로 특정 응용(예를 들어, 회원제로 운영하는 특정 사이트에서 임의의 회원들이 특정 시간에 특정 사이트를 방문한 횟수를 측정할 경우)에 이용될 수 있다. 만일, 특정 기간동안 임의의 서버를 방문하고자 한다면, 감사기관이 해당 서버들마다 각기 다른 w들을 생성하여 방문자들에게 부여함으로써 선택적으로 임의를 서버를 방문하도록 할 수 있다. 그러나 이 방법은 전자에 비해 약간 비효율적이다.

그런 다음 β를 서버에게, w를 방문자들에게 각각 전달한다. 그 후 [단계 2]에서 방문자는 자신이 계산한 V에 감사기관에게서 받은 w를 연접하여 서버에게 전달한다. [단계 3]에서 서버는 식 $w=H(\beta, V)$를 계산하여 방문자가 전달한 V가 올바른지를 검증한다.

지금까지 설명한 아이디어를 토대로 강건성을 만족시키기 위한 측정 방법을 재구성하면 아래와 같다. ([그림 4]참조) 단, 앞서 말한 대로 방문자가 특정 기간 t'에 특정 서버 S'로 임의의 방문자들이 방문한다고 가정한다.

[단계 1] *Initialization*

감사기관은 먼저 임의의 값 β(비밀 정보로 α가 아닌 임의의 큰 수)를 생성한 다음 $w=H(\beta, V)$를 계산하여 β를 서버에게 전달한다. 그 후 n명의 방문자 C들에 대한 UID_C와 자신만이 알고 있는 비밀정보인 임의의 α를 생성한 다음, 이 두 값을 연접하여 일방향 해쉬함수 H를 수행, 그 결과 $H(UID_C, \alpha)$와 UID_C 그리고 w를 연접하여 각 방문자들에게 랜덤 순서로 전달한다.

표 2 Naor와 Pinkas의 방법과 제안하는 방법과의 강건성에 대한 효율성비교(* : 감사기관의 초기(initialization) 단계 계산량이 아닌 검증(verification)단계에서의 계산량임)

		Naor & Pinkas[2]	제안한 방법
방법		다항식 V(x,y)=A(x,y)*P(x,y)+B(y)를 이용	해쉬함수 w=H(β,V) 이용
계산량	감사기관	d-1차 다항식 계산*	k-1번의 XOR 연산*
	서버	k-1차 다항식 보간 **k번의 c_k+k-1차 다항식 계산**	k-1번의 XOR 연산과 k번의 이진탐색 **k번의 해쉬함수 수행**
	방문자	d-1차 다항식 계산 **c_d+d-1차 다항식 계산**	1번의 해쉬함수 수행

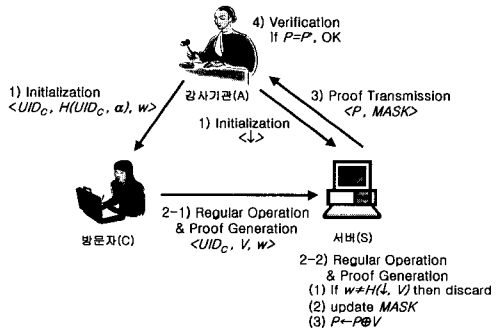

그림 4 강건성을 위한 개선된 측정 방법

[단계 2] *Regular Operation & Proof Generation*

방문자 C가 특정 기간 t'내에 서버 S'를 방문하면, 감사기관으로부터 전달받은 w, UID_C와 함께 $V(=H(S', t', H(UID_C, \alpha))$를 계산하여 서버에게 보낸다. 그리고 이 값을 전달받은 서버는 다음 세 가지 작업을 수행한다.

(1) 먼저 방문자들로부터 받은 w와 V를 이용하여 식 $w=H(\beta, V)$를 계산하여 V가 올바른지를 검증한다. 만일, 이 값이 올바르면 (2)를 계속해서 수행하고, 그렇지 않으면 현 방문자가

제공한 w와 V가 올바르지 않으므로 무시한다. 본 논문에서 이러한 경우의 방문자는 방문에는 성공했지만 방문 횟수에는 반영되지 않는 것으로 정의한다. 그러나 실제 응용에 따라서는 방문자에게 올바른 V를 제공하도록 요청할 수 있다.

(2) 방문자 C로부터 전달받은 UID_C를 이용하여 $MASK$를 갱신한다. 앞서 설명한대로 UID_C의 $rank$를 구하여 $MASK[rank(UID_C)]$←1로 한다.

(3) 현재의 증명서 P와 V를 XOR하여 새 증명서를 만든다. 즉, P←$P\oplus V$를 계산한다. P의 초기값은 0이다.

[단계 3] *Proof Transmission*

기본 측정 방법과 동일.

[단계 4] *Verification*

기본 측정 방법과 동일.

4.3.1 개선된 방법에 대한 분석

[강건성]

제안한 방법의 강건성은 해쉬함수의 충돌회피성과 일방향성에 기반한다. 즉, 방문자가 V를 조작하기 위해서는 β를 알아야 하는데, w와 V를 알더라도 유일한 β를 알아낼 가능성은 $1/\beta$이다.

[효율성]

[표 2]에서 살펴보는 바와 같이, 강건성을 위해 Naor와 Pinkas가 제안한 방법은 검증식이 $V(x,y)=A(x,y)*P(x,y)+B(y)$로서, P는 x, y를 변수로 갖는 최고차가 $k-1$, $d-1$인 이변수(bivariate) 다항식, A는 x, y를 변수로 갖는 최고차가 c_k, c_d인 이변수 다항식, 그리고 B는 y를 변수로 갖는 최고차가 c_d인 다항식이다. 이때, k는 방문자의 수이며 d는 time frame t의 수로, 이 경우 다항식 P는 d번의 time frame t가 경과될 때마다 감사기관에 의해 새로운 다항식으로 갱신된다. c_k와 c_d는 각각 감사기관이 어느 정도의 안전성을 부여하기 위해 임의로 생성한 값이다. 이 방법의 기본적인 아이디어는 다음과 같다. 먼저 감사기관이 초기단계에서 A, B, P를 임의로 생성한 다음 V를 계산하여 이중 A, B를 서버에게 전달하고 나머지 P, V를 각 방문자들에게 전달한다. 그 후 방문자가 서버를 방문할 때 일정 계산을 거쳐 P, V를 넘겨주면 서버는 자신이 가진 A, B를 이용하여 $V=A*P+B$인가를 확인함으로써 검증을 한다. 이때, k명의 방문자에 대해 서버가 검증을 하려면 기본 방법에서 서버가 최소 k번의 c_k+$k-1$차 다항식을 계산해야 하며, 각 방문자마다 c_d+$d-1$차 다항식 계산을 추가로 수행해야한다. 그러나 제안한 방법은 단지 서버가 k번의 해쉬함수 $H(\beta,V)$만을 추가로 수행하기 때문에 훨씬 효율적이다.

[안전성] [프라이버시] [유연성]
기본 측정 방법과 동일하므로 4.2.1을 참조하시오.

5. 결론 및 향후 연구 방향

본 논문에서는 인터넷 광고에서 안전하면서도 효율적이고 유연성을 가지는 방문자 수에 대한 측정 방법을 제안하였다. 특히 Naor와 Pinkas가 제안한 방법에 비해 효율성과 유연성을 높일 수 있는 방향으로 개선하였다. 방문자의 계산량을 줄이고 또 방문자 수를 확인 받는 과정의 계산량과 확인하는 과정의 계산량을 기존의 방법들에 비해서 개선하였으며, 방문자 수에 대한 임계치를 부여해야하는 기존의 제약 없이 자유로이 방문자 수를 측정할 수 있게끔 개선하였다. 또한 좀더 특수한 경우이긴 하지만 제안한 기본 방법에서 강건성을 부여하고자 기본 방법에서 보다 개선된 방법을 제안하였다.

현재 연구중인 문제점은 방문자에 대한 익명성을 강화하는 문제이다. 만일 어느 방문자가 여러 다른 서버들을 방문할 경우 각각 동일한 UID를 사용한다면 서버들 간의 공모에 의해서 방문자의 정보를 알아 낼 수 있다. 즉 서버는 정확히 방문자가 누구인지는 모르지만 어느 서버를 언제 방문했는지 그리고 얼마나 자주 방문하였는지 등의 정보를 알아내는 것이 가능하다. 따라서 방문자의 의도와는 다르게 서버들에게 자신의 정보가 노출될 수가 있다. 기본적인 해결 방법은 방문자들이 서로 다른 서버들을 방문할 때 서로 다른 UID를 보내도록 감사기관이 방문자에게 각 서버마다 다른 UID를 임의로 생성하여 [단계 1]의 과정에서 보내는 방법을 생각해 볼 수 있다. 하지만 이 경우 방문자 측에서 각 서버들에 따른 서로 다른 UID를 알고 있어야 한다는 부담이 따른다. 따라서 이 문제에서 방문자측의 부담을 최소화하면서도 즉, 최소의 계산만으로 앞서 말한 요구사항들을 만족시킬 수 있는 보다 나은 방법을 모색하는 중이다.

그밖에 향후 연구과제로는 본 논문에서 제안한 방법들을 실제 실험을 통해 그 이론적인 결과를 확인하는 것이다.

참 고 문 헌

[1] http://www.powerpage.co.kr/powerzine/
[2] M. Naor and B. Pinkas, Secure and Efficient Metering, *EuroCrypt '98*, LNCS, Vol. 1403, pp. 576-590, 1998.
[3] C. Dwork and M. Naor, Pricing via Processing or Combating Junk Mail, *Crypto '92*, LNCS, Vol. 576, pp. 114-128, 1992.
[4] A. Fiat and A. Shamir, How to Prove Yourself. *Crypto '86*, LNCS, Vol. 263, pp. 181-187, 1986.
[5] M. K. Franklin and D. Malkhi, Auditable Metering with Lightweight Security, *Financial Cryptography '97*, LNCS, Vol. 1318, pp. 151-160, 1997.
[6] B. Masucci and D. R. Stinson, Metering Schemes for General Access Structures, *ESORICS 2000*, LNCS, Vol. 1895, pp. 72-87, 2000.
[7] W. Shin and K. H. Rhee, A WWW Metering Scheme using a Secure Primitive, Proceedings of *WISA 2000*, Vol. 1, No. 1, pp. 182-191, 2000.
[8] A. Aho, J. Hopcroft, and J. Ullman, The Design and Analysis of Computer Algorithms, Addison-Wesley, pp. 299, 1974.

김 순 석
1997년 2월 진주대학교 컴퓨터공학과 학사. 1999년 2월 중앙대학교 컴퓨터공학과 석사. 1999년 3월 ~ 현재 중앙대학교 대학원 컴퓨터 공학과 박사과정. 관심분야는 암호 프로토콜, 이동통신 보안, 정보보호

신 재 용
2000년 2월 중앙대 컴퓨터공학과 졸업(학사). 2002년 2월 중앙대 컴퓨터공학과 졸업(석사)

김 성 권
1981년 2월 서울대학교 계산통계학과 학사. 1983년 2월 한국과학기술원 전산학과 석사. 1990년 8월 University of Washington 전산학 박사. 1991년 3월 ~ 1996년 2월 경성대학교 전산통계학과 조교수. 1996년 3월 ~ 현재 중앙대학교 컴퓨터공학과 부교수. 관심분야는 계산기하학, 암호 응용 및 정보보호, 생물정보학