2001년 한국정보처리학회 춘계 학술발표논문집 제8권 제1호

인터넷 광고에서 방문 횟수를 측정하는 암호학적 방법에 관한 연구

신제용*, 김순석*, 김성권*
*중앙대학교 컴퓨터공학과
e-mail:sjy8282@alg.cse.cau.ac.kr

A Study on Cryptographical Metering Scheme for Advertisements on the Web

Je-Yong Shin*, Soon-Seok Kim*, Sung-Kwon Kim*
*Dept of Computer Science & Engineering, Chung-Ang University

요약

통신기술의 발전으로 많은 사람들이 인터넷에 접속하여 정보를 얻고 있다. 인터넷으로 서비스를 제공하는 회사들은 회원이나 방문자들에게 필요한 자료를 공급하고 인터넷 광고를 통해서 수입을 얻고 있다. 광고를 제공하는 서버에 고객들이 방문한 횟수 즉, 광고에 노출된 횟수에 비례해서 광고를 제공한 측에서 광고주에게 광고 수수료를 청구한다. 따라서 광고주와 광고를 직접 제공하는 서버측 모두에게 방문자 수의 측정은 중요하다. 현재 가장 많이 이용되는 웹 로그 분석 기법은 로그파일의 조작에 의해 방문자의 정확한 측정이 어렵고 또 정확한 통계자료로 보기도 어렵다. 따라서 본 논문에서는 이러한 단점을 극복하기 위해서 지금까지 제안된 방문자 측정 방법보다 효율성과 유연성을 가지면서 안전한 측정 방법을 제안한다.

1. 서론

인터넷이 생활에 밀접한 관련을 맺으면서 현재 우리는 인터넷 광고의 홍수 속에서 살고 있다. 어떤 사이트를 방문하든지 최소한 몇 개의 광고를 접하기 마련이다. 인터넷 광고는 한마디로 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 커뮤니케이션 활동을 말한다.

현재 인터넷 광고 분야에서 미해결분야로 남아있는 문제가 다름 아닌 '마케팅 효과의 측정', 다시 말해, 보다 자세하고 정확한 광고 효과의 측정분야이다. 그 효과를 측정하기 위해서는 무엇보다 광고가 게재된 사이트에 얼마만큼의 방문이 일어났느냐 하는 것이 중요하다. 웹사이트 발행자가 자신의 이익 즉, 보다 많은 광고를 자신의 사이트에 유치하기 위해서 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 믿을 수 있는 정확한 자료 즉, 방문객의 방문 횟수에 대한 기록을 웹사이트 발행자가 제시해야 한다.

2. 측정 방법에 따른 요구사항

인터넷에서 방문자 측정을 위해 갖추어야 할 요구사항들은 다음과 같다.[5]

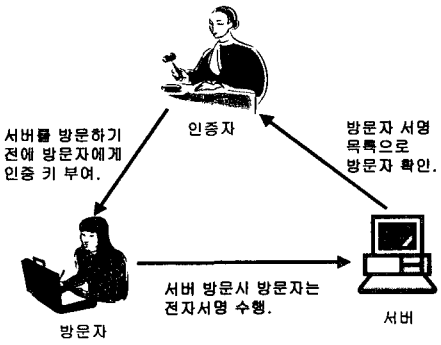
- 웹사이트 발행자가 제공하는 조회 횟수의 기록에 대한 정확성 및 신뢰성
- 방문객의 프라이버시 보호
- 조회 횟수 측정을 위한 메모리 사용 공간의 효율성
- 현 사용환경(예, 웹브라우저)과의 호환성 및 방문객의 편의성

위의 요구사항들 중 위의 두 가지가 문제를 해결하기 위한 방법적인 측면이라고 한다면 아래 둘은

주로 구현적인 측면이라고 볼 수 있다.

3. 관련 연구 동향

3.1 기본적인 측정 방법



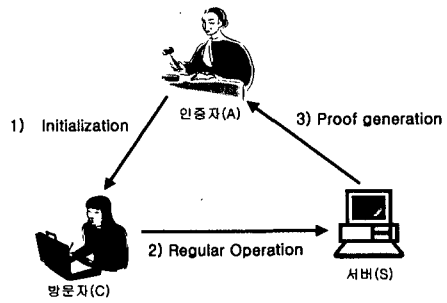
[그림 1] 전자서명을 이용한 기본적인 측정 방법

지금까지 제안된 여러 가지 방법들 중 가장 기본적인 측정 방법은 전자서명을 이용한 방법이다. 이 방법은 위의 그림 1에서 보는 바와 같이 제 3의 신뢰할 수 있는 기관이 앞으로 방문할 고객(이하 방문자라 부른다.)에게 인증된 서명키를 분배한다. 그 후 방문자가 해당 웹사이트(이하 서버라 부른다.)를 방문 할 때에 방문자는 전자 서명 프로토콜을 수행한다. 그리고 서버는 방문자들로부터 받은 서명 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 확인함으로 개인의 프라이버시를 침해할 수 있는 단점이 있다.

3.2 개선된 측정 방법들

3.2.1 Naor와 Pinkas[5]가 제안한 방법

이 방법은 위 그림 2에서 보는 바와 같이 먼저 서버(S)가 일정 기간 동안에 방문자(C)의 방문을 받는다. 그 후, 방문자의 수가 제3의 신뢰할 수 있는 기관(A)과 약속한 수와 같아지면 이 기관에 알려서 방문자 수를 확인 받는 방법이다. 즉, 비밀공유기법을 사용하여 여러 명의 방문자에게 미리 키를 주고 일정 기간(time frame: t)동안 방문자들이 서버에 접속할 경우에 이 키 값을 서버에게 준다. 그 후 서버가



[그림 2] Naor와 Pinkas가 제안한 방법

일정 수, k명의 방문자들을 받게 되면 방문자들로부터 받은 키 값들을 가지고 방문자 수에 대한 증거를 생성하여 감사 기관에 그 값을 넘겨준다. 이 때, 감사기관은 서버로부터 넘겨받은 증거와 자신이 생성한 값이 일치하는지를 계산하여 증명하는 방법이다.

이 방법은 암호학적으로 안전함을 증명할 수 있고 원래의 통신 패턴을 유지할 수 있다는 것이 장점이다. 그러나, 기본적으로 비밀공유기법을 이용하기 때문에 만약 1만 명의 방문자가 있어야만 확인 받을 수 있다고 했을 경우에 만약 1만 명에 가까운 즉 9990명의 방문이 발생했을 경우에 서버는 방문 사실을 확인 받을 수 없다는 단점이 있다. 따라서, 실제 적용을 위해서는 이 부분에 약간의 수정이 요구된다.

3.2.2 Masucci와 Stinson[4]이 제안한 방법

이 방법은 앞서 3.3.1에서 Naor와 Pinkas[5]가 제안한 방법의 단점 즉, 정해진 수의 방문보다 약간 적은 수의 방문이 발생했을 경우에 감사 기구는 서버의 방문자 수를 확인 받을 수가 없었다. 이를 개선하기 위해서 Massucci와 Stinson[4]은 최소 임계 방문자 수의 개념을 도입했다. 즉, 정해진 방문자 수보다는 작고 최소 임계 방문자 수보다는 큰 수의 방문이 이루어졌을 경우에 서버는 자신의 방문자 수를 확인 받을 수 있는 방법을 제안했다. 이 방법은 3.3.1에서 제안한 방법에 비해 방문자 수의 측정상 훨씬 유연성을 갖는다는 장점이 있다. 그러나, 이러한 유연성을 제공하는 만큼 서버 측에 부담을 준다.

4. 새로운 측정 방법에 대한 제안

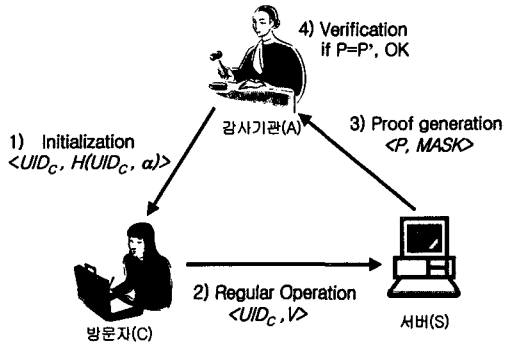
본 논문에서 제안하는 방법을 개괄적으로 설명하면 다음과 같다. 먼저, 감사기관이 사전에 각 방문자들에게 서로 다른 비밀정보와 임시 ID인 UID를 생

성하여 n명의 방문자들에게 나눠준다. 그 후, 방문자가 임의의 시간에 서버를 방문하게 되면 감사기관으로부터 받은 자신의 비밀정보와 UID를 서버에 전달한다. 서버는 각 방문자들의 비밀정보와 UID들을 모아서 일종의 '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 : 감사기관(Audit Agency)
- S : 서버(Server) 또는 서버의 ID(여러 개의 서버가 있을 수 있다)
- C : 방문자 또는 방문자 ID(모두 n명의 방문자들이 있다고 가정)
- UID_c : 방문자 C의 유일한 임시 ID [1]
- t : time frame(측정을 위한 기간으로 서버 S와 사전에 결정한다.)
- f_{pr} : proof 함수로 입력 값들에 대한 bit-wise XOR 연산
- $MASK$[2] : 서버가 생성하는 n비트 벡터로, 미리 정렬(sorting)해둔 UID_C값과 방문자들로부터 전달받은 UID_C값을 비교하여 이 값이 존재하면 해당하는 인덱스에다 한 비트인 1로 표기하고 그렇지 않으면 0으로 표기한다.
- H : 일방향 해쉬함수

4.2 제안하는 측정 방법



[단계 1] *Initialization*

감사기관(A)은 n명의 방문자(C)들에 대한 UID_C 값과 자신만이 알고 있는 비밀정보인 임의의 α를 생성한 다음 이 두 값을 연접하여 일방향 해쉬함수 H를 수행, 그 결과 값($=H(UID_C, \alpha)$)과 UID_C 값을 각 방문자들에게 전달한다.

[단계 2] *Regular Operation*

방문자(C)가 t 시간에 서버(S)를 방문할 때, $V(=H(S, t, H(UID_C, \alpha))$값을 계산하여 감사기관으로부터 전달받은 UID_C값과 함께 서버에게 보낸다.

[단계 3] *Proof Generation*

서버(S)는 방문자(C)들로부터 받은 V값들을 입력으로 하여 f_{pr} 연산을 수행한 다음, 그 결과 값 P와 앞서 정의한 UID_C값들을 이용하여 $MASK$ 정보를 생성하여 이 두 값을 감사기관에 보낸다.

[단계 4] *Verification*

감사기관(A)은 서버(S)로부터 전달받은 $MASK$ 정보를 이용하여 서버가 수행한 것과 동일하게 P'를 생성하여[3], 서버로부터 전달받은 P값과 동일한지를 검사한다. 만일 이 두 값이 같다면 서버가 보낸 일종의 '증명서'인 P값을 증거로 받아들인다.

4.3 제안하는 측정 방법에 대한 분석

[안전성]

제안하는 방법의 안전성은 해쉬함수의 일방향성과 충돌회피성에 기반한다. 즉, $y=H(x)$에서 y의 값을 안다하더라도 x값을 알기란 계산적으로 거의 불가능

1) UID_C 값은 최초에 감사기관이 생성하는 실제 ID가 아닌 임시 ID로, [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 seed값 s라고 가정하자. 이때, 최초로 생성되는 UID값은 f(s), 두 번째 UID값은 f(f(s)), 세 번째 값은 f(f(f(s)))와 같은 식으로 하여 n명의 방문자들에 대한 UID를 생성할 수 있다. 또한, 이 UID값은 해쉬 함수의 성질에 의해 랜덤하고 유일하며 서버가 사전에 미리 계산하여 정렬(sorting)한 후, 이후에 사용될 MASK 정보의 생성을 위해 보관한다.

2) n비트 사이즈이나, 만일 방문객들이 많은 경우는 구현시에 효율성을 위해 압축코드를 활용할 수도 있다.

3) 서버로부터 전달받은 $MASK$ 정보를 보면 어느 방문자가 방문했는지를 계산을 통해 감사기관이 알 수 있으며 또한, 이를 이용하여 서버가 수행한 것과 동일한 정보인 V값들을 계산할 수 있다. 또한 계산의 효율성을 위해 이 값은 서버와 사전에 협의한 t 시간에 따라 미리 계산될 수 있다.

하다. 서버가 방문 횟수를 과장되게 조작하기 위해서는 방문하지 않은 방문자의 정보, V값을 새로이 생성하여 증거 P를 만들어야 한다. V값을 생성하기 위해서는 기타 정보들(S, t, UID_C)을 안다하더라도 α 값을 모르기 때문에 계산할 수 없다. 만일, α 값이 u비트 길이라면 적어도 $1/2^u$의 확률에 대해 안전하다.

[효율성]

아래 표 1에서 살펴보는 바와 같이, Naor와 Pinkas[5]가 제안한 방법은 공개키 연산을 수행하는 전자서명 방식에 비해 비밀공유기법을 사용하므로 훨씬 효율적이다. 그러나 이 방법 역시 방문자의 수를 k라 할 때, 다항식에 대해 최소한 k차수만큼의 보간법(interpolation)을 수행해야한다. 그러나 본 논문의 경우는 보간법 대신 해쉬 연산과 증거 생성에 있어 한 번의 *XOR* 연산을 수행하기 때문에 훨씬 효율적이다.

[유연성]

Naor와 Pinkas[5]가 제안한 방법은 앞서 3절에서 살펴본 바와 같이 방문자 수가 k명이 되었을 때에만 증거를 생성할 수 있었다. 그러나 본 논문의 경우는 k명과 같은 임계값을 필요로 하지 않으므로 어떠한 수의 방문자들에 대해서도 자유로이 측정이 가능하다. 이것은 방문자들에 대한 일종의 출석부라 할 수 있는 *MASK* 정보와 방문자들이 보낸 정보들에 대한 *proof* 함수의 결과만을 보내기 때문에 가능하다.

제안한 측정 방법과 앞서 언급한 다른 측정 방법을 비교해 보면 오른쪽 표 1과 같다.

[표 1] 제안된 측정 방법들과의 비교

		전자서명법	Naor와 Pinkas의 방법[5]	제안한 방법
특징		공개키를 이용한 전자서명	다항식을 이용한 비밀공유(Secret Sharing)	해쉬함수와 *MASK* 정보
계산량	감사기관	방문자의 서명 인증	다항식 생성	단계 1에서 해쉬함수 수행과 단계 4에서 f_{pr}수행
	서버	방문자의 서명 확인	다항식에 대한 보간법(k명의 방문자에 대해 $O(k(\log k)^2)$)수행	1회 f_{pr}수행, *MASK* 정보 생성
	방문자	서명	다항식 연산	1회 해쉬함수 수행
장점		정확성, 안전성	정확성, 안전성	정확성, 안전성, 효율성, 유연성

5. 결론

본 논문에서는 인터넷 광고에서 안전하면서도 효율적이고 유연성을 가지는 방문자 수에 대한 측정 방법을 제안하였다. 특히 Naor와 Pinkas[5]가 제안한 방법에 비해 효율성과 유연성을 높일 수 있는 방향으로 개선하였다. 방문자의 계산량을 줄이고 또 방문자 수를 확인받는 과정의 계산량과 확인하는 과정의 계산량을 기존의 방법들에 비해서 개선하였으며, 방문자 수에 대한 임계치를 부여해야하는 기존의 제약 없이 자유로이 방문자 수를 측정할 수 있게끔 개선하였다.

향후 연구과제로는 본 논문에 제안한 방법에 약간의 변형을 통해서 강건성과 익명성을 부여하는 것이다. 악의를 가진 방문자가 거짓 정보를 부여하는 경우에 방문서버는 그 값을 가지고서는 방문자 수를 인증 받을 수 없다. 이럴 경우 방문 서버가 미리 이 거짓 정보를 알아 낼 수 있는 방법을 제공해야 하는데 이것이 강건성(Robustness)이다. 그리고 최근 인터넷에서 중요시되는 방문자의 프라이버시를 보장할 수 있는 익명성(anonymity)을 가진 측정 방법을 연구 중이다.

참고문헌

[1] V. Anupam, A. Mayer, K. Nissim, B. Pinkas and M. K. Reiter, On the Security of Pay-Per-Click and Other Web Advertising Schemes, *8th International WWW Conference*, pp. 12-22 May 1999.

[2] C. Dwork and M. Naor, Pricing via Processing or Combating Junk Mail, *Crypto '92*, LNCS 576, pp. 114-128 1992.

[3] M. K. Franklin and D. Malkhi, Auditable metering with lightweight security, *Financial Cryptography '97*, LNCS 1318 pp. 151-160 1997.

[4] B. Masucci and D. R. Stinson, Efficient Metering Schemes with Pricing, *Technical Report CORR*, 2000.

[5] M. Naor and B. Pinkas, Secure and Efficient Metering, *EuroCrypt '98*, LNCS 1403, pp. 576-590, 1998.

[6] J. Pitkow, In search of reliable usage data on the WWW, *6th International WWW Conference.*, pp. 451-463 Apr. 1997.