핑거프린팅 기법을 이용한 부정 클릭의 식별

Fraud Click Identification Using Fingerprinting Method

홍영란(Young Ran Hong)*, 김동수(Dongsoo Kim)**

초 록

인터넷 광고에서 부정 클릭을 식별하기 위해 제시된 모델은 검색한 키워드, 광고 클릭 시간, 사이트 방문 시간, IP 등을 기준 변수로 한다. 이 방법은 클라이언트 IP를 기반으로 하고 사람이 부정 클릭을 행하는 것을 식별하는 것이기 때문에 자동화 도구 등을 이용하여 부정 클릭을 행하는 방법에 대한 식별 방법으로는 충분하지 않다. 본 논문에서는 자동화 도구를 이용한 부정 클릭을 보다 정확하게 식별하기 위해 각 정보의 조합을 핑거프린팅하여 그 값을 비교하는 방법을 제안하였다. 연구는 3단계로 나누어 진행되었다. 1단계에서는 IP 정보의 핑거프린팅, 2단계에서는 IP와 세션 정보의 핑거프린팅, 3단계에서는 세션 정보와 검색한 키워드의 핑거프린팅을 만든다. 결과 값을 통해 동일한 핑거프린팅 값을 갖는 것을 자동화 도구를 사용한 부정 클릭의 가능성이 높다고 판단한다. 본 연구에서 제시한 방법론은 부정 클릭 식별 모델을 만들기 위해 기존의 연구에서 사용하였던 개별 값의 비교에서 발전하여 DB에 저장된 각 로그의 값을 조합하여 사용함으로써 여러 가지 정보를 부정 클릭 식별을 위한 유의미한 값으로 사용할 수 있다는 점을 보여주고 있다.

ABSTRACT

To identify fraud clicks in the Internet advertisement, existing studies have considered keyword, visit time, and client IP as an independent variable for the standard. These methods have limitations in identifying the fraud clicks that utilize automation tools, for they are methods based on client IP and human activities on the Internet. This paper proposes that fingerprinting values of the variable combination should be used to identify fraud clicks. The proposed model is composed of 3 stages and the fingerprinting values are compared with the other input data at each stage; IP fingerprinting in the first stage, IP and session data fingerprinting in the second stage, and session data and keyword fingerprinting in the third stage. We showed that the proposed model of the fraud click identification is more correct than existing methods through experiments according to the proposed scheme.

키워드 : 부정 클릭, 인터넷 광고, 핑거프린팅, IP, 키워드, 세션 정보
Fraud Click, Internet Advertisement, Fingerprinting, IP, Keyword, Session

* 숭실대학교 산업정보시스템공학과 박사과정
** 교신저자, 숭실대학교 산업정보시스템공학과 부교수
2011년 07월 12일 접수, 2011년 08월 11일 심사완료 후 2011년 08월 18일 게재확정.

1. 서 론

한국의 인터넷 검색 광고 시장은 주로CPC (Cost Per Click) 방식의 요금 지불 체계를 따르고 있다. CPC 방식의 요금 지불 체계는 검색 키워드 입력 후 나타나는 검색결과 업체 중에서 소비자가 클릭하는 수에 따라 광고요금을 부과하는 방식이다. 그러나 CPC 방식은 클릭 후의 소비자 행동을 선혀 고려하지 않고 단순히 클릭 수에 따라 요금을 청구하는 방식이어서 광고 효과를 직접적으로 측정하기 어려워 광고주들의 불만이 매우 높다. 또한 CPC 방식은 무엇보다도 부정 클릭(fraud click)에 매우 취약하다는 문제점을 지니고 있다. 부정 클릭은 인터넷 상에서 경쟁업체 사이트를 집중적으로 클릭하여 경쟁업체가 포털에서 검색되지 않도록 하기나 검색광고의 경쟁을 유발시켜 많은 광고비를 발생시켜 손실을 가져오는 문제점을 야기시킨다[3, 6].

부정 클릭을 식별하기 위해 제시된 기존의 모델은 검색한 키워드, 광고 클릭 시간, 사이트 방문 시간, IP 등을 기준 변수로 한다. 이러한 방법은 클라이언트 IP를 기반으로 하고 사람이 부정 클릭을 행하는 것을 식별하는 것이기 때문에 자동화 도구 등을 이용한 부정 클릭을 식별하기가 어렵다. 이러한 배경에서 본 논문은 자동화 도구를 이용한 부정 클릭을 보다 정확히 식별하기 위해 각 정보의 조합을 핑거프린팅(fingerprinting)하여 그 값을 비교하는 방법을 제안한다.

본 연구에서 제안하는 부정 클릭 식별 방법은 3단계로 진행된다. 1단계에서는 기존 모델과 같은 IP 정보를 기준으로 하였다. 2단계에서는 로그의 세션 정보를 이용하였다. 3단계에서는 세션 정보와 키워드를 조합하였다. 본 연구는 단계별로 각 값을 핑거프린팅하고, 이 후 유입되는 로그 정보에 동일한 값이 나타날 경우, 이를 부정 클릭을 판단하는 기준으로 삼을 것을 제안한다.

본 연구에서 제안한 방법은 방문 기록 DB의 세션 정보를 분석하여 동일 패턴의 클릭 정보와 동일한 시간 간격으로 세션 정보가 처리될 경우, 혹은 이와 함께 동일한 키워드가 입력될 경우, 이를 핑거프린트 값과 비교함으로써 보다 정교한 부정 클릭 식별 모델을 만들 수 있는 효과를 가진다.

2. 관련 연구

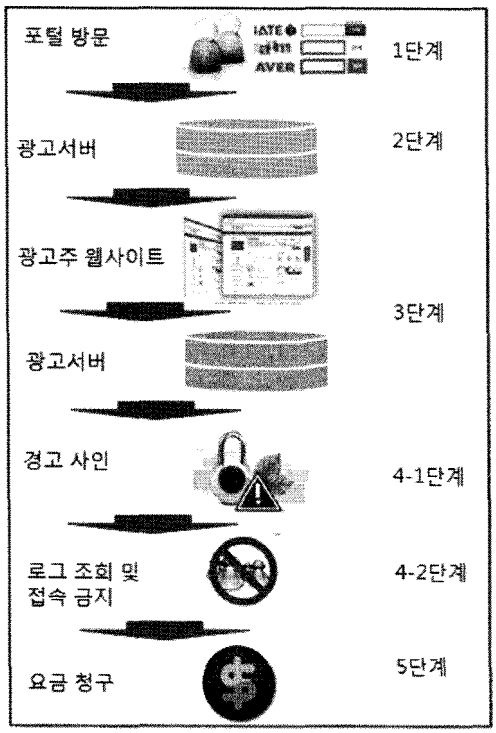
2.1 부정 클릭 식별을 위한 기존 방법론

검색광고 시장에서 사용하고 있는 CPC 방법은 부정 클릭에 취약하다는 문제점을 지니고 있다. 검색광고 광고주들이 가장 많이 경험하고 있는 부정 클릭 사례는 허위 클릭과 무효 클릭으로 나타나고 있다[3, 6].

따라서 CPC 방식의 키워드 검색 광고 및 그 안에서 발생하는 부정행위에 대한 대응 현황의 검토를 바탕으로 하여 CPC 방식의 키워드 검색 광고에서 발생하는 부정행위 방지 및 분쟁 해결에 대한 접근 방법을 제안하고 특히 '클릭'이라는 가장 기본적인 행동을 유효/무효, 사기의 고의성, 수행 방법(자동/수동) 등을 기준으로 하여 분류하는 방법이 계속 연구되고 있다[1, 7].

부정 클릭은 인터넷 광고비용과 밀접한 관

련을 갖는다. 다음 <그림 1>은 일반적인 CPC의 요금 청구 단계를 보이주고 있다.

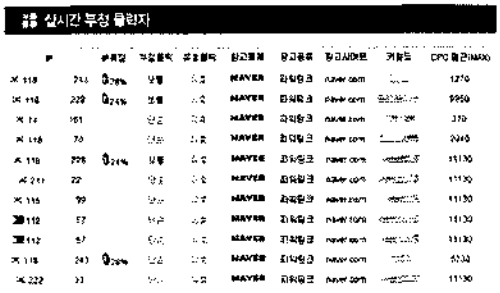


〈그림 1〉 CPC 요금 청구 단계

<그림 1>에서 보는 것과 같이 일반적인 CPC 과금 체계는 다음 다섯 가지 단계로 구별된다. 첫 번째는 검색 엔진의 키워드를 통하여 부정 클릭을 발생시키는 단계이다. 두 번째 단계는 사용자 정의를 통한 부정 클릭 판단 기준을 통해 부정 클릭의 발생을 인지하는 단계이다. 세 번째 단계는 대부분 광고 관리 등 각 부정 클릭 방지 업체들의 솔루션이 적용되는 단계이다. 네 번째 단계는 설정에 따라 랜딩 페이지를 별도 지정 페이지로 변경하여 부정 클릭을 제지하거나 경고하는 단계로서 처음에는 경고 사인을 띄우고, 계속 접속을 시도하는 경우 접속을 막는 단계이다. 다섯 번째는 마지막 단계로서 검색 엔진 광고의 클릭이 발생할 때 오버추어와 각 포털별 요금을 실제 업체의 클릭 수와 비교하여 검색 광고의 요금으로 청구하는 단계이다. 이 다섯 단계로 구성되는 부정 클릭 분석 방법은 대부분 로그 분석을 통한 IP 추적이 그 중심 기술이 된다[2, 6]. 이 방법은 IP 변환 시스템의 적발과 동일 IP가 지속적으로 특정 사이트를 클릭하거나 검색할 때 이를 부정 클릭으로 간주하여 해당 IP를 막는 형식을 취한다.

2.2 기존 방법론의 실제 활용 예시

현재 많이 사용되는 부정 클릭 방지 시스템은 IP와 접속횟수를 비교하는 방법이다. 동일한 IP로 일정 시간 동안 일정횟수 이상 클릭한 행위를 부정 클릭으로 간주하는 방법과 IP 변경 프로그램을 추적하여 부정 클릭으로 간주하는 방법을 들 수 있다[8].



〈그림 2〉 동일 IP 접속을 부정 클릭으로 간주하는 시스템 예

<그림 2>에서 보는 것처럼 부정 클릭을 적발하기 위하여 특정 시간대에 동일한 IP가

여러 번 같은 상품 관련 키워드를 검색하거나 특정 사이트의 특정 광고를 클릭한 경우, 최초 방문 정보 중 IP와 키워드를 UUID로 만들고 이후의 방문 정보를 계속 UUID로 만든 후, 같은 UUID를 가진 경우가 발생하면 이를 부정 클릭으로 간주하는 방법이다.

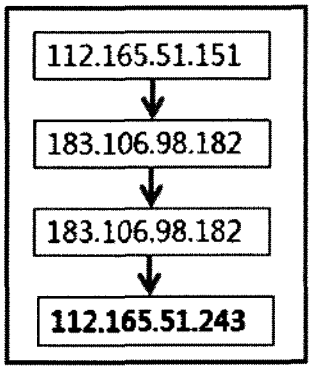
그 다음으로는 IP 변경 프로그램을 추적하여 IP를 변화시켜 클릭하는 움직임을 부정 클릭으로 간주하는 방법이 있다. <그림 3>은 IP 변경 추적 시스템을 이용하여 원래 IP를 추출해냄으로써 부정 클릭을 식별하는 부정 클릭 방지 시스템의 예이다[9].

IP	검색어	검색엔진
121.146.194.126(22) URL 정보	구구절	naver.com
121.146.194.92(21) URL 정보	구구절	naver.com
121.146.194.64(20) URL 정보	웃사마	naver.com
121.146.194.183(19) URL 정보	짱카절	naver.com
121.146.194.50(18) URL 정보	웃사마	naver.com
121.146.196.132(17) URL 정보	바가지머리	naver.com
121.146.194.193(16) URL 정보	베비드레스	naver.com
121.146.194.181(15) URL 정보	스타일드레싱	naver.com
121.146.194.130(14) URL 정보	웃사마	naver.com
121.146.194.151(13) URL 정보	짱코렌에마	naver.com
121.146.194.151(12) URL 정보	스타일드레싱	naver.com
121.146.196.189(11) URL 정보	짱커절	naver.com
121.146.196.214(10) URL 정보	웃사마	naver.com
121.146.194.224(9) URL 정보	짱카절	naver.com
121.146.194.224(9) URL 정보	스타일드레싱	naver.com

〈그림 3〉 네이버 Click Choice 접속 목록으로 본 IP 정보

IP는 여러 가지로 들어오지만 이것에 대해 IP 변경 추적 프로그램을 돌려보면 아래 <그림 4>와 같이 IP가 변경된 과정을 추적해냄으로써 동일 IP가 자동 변경 프로그램을 사용하여 접속하였음을 알 수 있다.

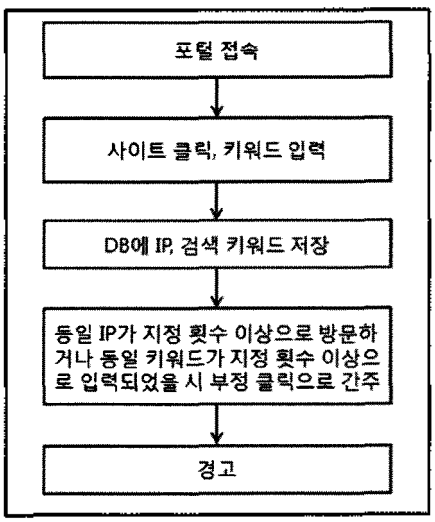
그러나 기존의 연구 방법은 사람이 접속하는 IP 정보만을 기반으로 하기 때문에 자동화 도구를 이용하여 부정 클릭을 실시한 경우에는 그 식별에 한계점을 갖는다. 여러 개



〈그림 4〉 IP 변경 추적 프로그램으로 분석한 IP 변경상태

의 가상 머신(VM Ware)에 여러 개의 IP를 등록한 후, 자동화 도구로 부정 클릭을 순서를 바꾸어 가면서 실시하게 되면 IP 기반의 부정 클릭 식별 방법은 식별 능력이 떨어질 수 밖에 없다.

기존에 IP를 이용한 부정 클릭 연구 모델은 다음 <그림 5>와 같은 흐름으로 부정 클릭을 식별한다.

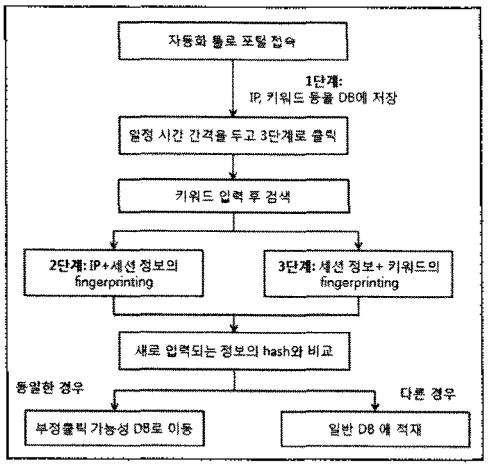


〈그림 5〉 기존 연구에서 사용하는 IP 기반의 부정 클릭 식별 프레임워크

본 연구에서 제안하는 세션 정보, 혹은 세션 정보와 키워드를 조합하여 핑거프린팅하고 이 값을 비교하는 방식은 동일한 정보를 가진 로그의 핑거프린트 값이 같다는 점을 이용하여 부정 클릭을 식별한다. 이를 통해 자동화 도구를 이용한 부정 클릭을 식별하는데 있어 기존 기법이 가지는 한계점을 극복할 수 있다.

3. 제안 방법론 프레임워크

최근 자동화 도구를 이용한 부정 클릭이 늘어나고 있기 때문에 본 연구에서는 자동화 도구를 이용한 부정 클릭 식별을 높이는 방법을 제안한다.



〈그림 6〉 전체적 제안 방법론 흐름도

<그림 6>은 추가된 변수와 비교 방법을 적용한 자동 부정 클릭 식별 방법에 대한 전체적인 제안 방법론 프레임워크를 보여 주고 있으며, 부정 클릭의 식별을 3단계로 식별하고 있다. 1단계는 <그림 5>에서 보는 것과 동일한 IP를 기반으로 하는 부정 클릭 식별 모델이다. 여기서는 변수를 IP 하나로 지정하고, IP 비교 방법을 사용한다. 2단계에서는 세션 정보를 변수로 추가한다. 세션 정보를 핑거프린팅하여 세션 정보 값을 비교함으로써 자동화 도구를 사용한 부정 클릭 탐지 능력을 높일 수 있다. 3단계에서는 세션 정보와 키워드라는 변수를 추가한다. 동일한 키워드를 입력하는 행위를 핑거프린팅 하여 고유 값을 만들고 이후 입력되는 정보의 핑거프린팅 값과 비교하여 부정 클릭 탐지 능력을 높인다.

본 연구에서 적용하는 핑거프린팅 기법은 해쉬 기법을 사용하기 때문에 다음과 같은 장점을 갖는다. 첫째, 해쉬값은 고유값을 갖기 때문에 다른 값들과 비교를 하는 연구에서 값의 대표성 및 고유성을 유지할 수 있다. 둘째, 여러 가지 값을 빠르게 비교할 수 있는데, 이는 부정 클릭을 탐지해내는데 있어 중요한 요소가 된다. 데이터 관리에서 이 핑거프린트를 사용할 경우 해쉬 테이블을 이용하여 다차원 데이터의 저장 레코드 순서를 빠르게 찾아 저장함으로써 데이터 생성 속도가 향상된다. 또한 해쉬 테이블 만을 유지하면 되므로 메모리 사용량이 감소한다. 따라서 해쉬 테이블의 사용으로 데이터의 빠른 검색과 데이터 생성 요청에 빠른 응답이 가능하다는 장점을 가지고 있다[1].

4. 부정 클릭 식별 실험

본 연구는 자동화 도구마다의 특성을 고려하여 초기에 'Test Complete'와 'Win Auto

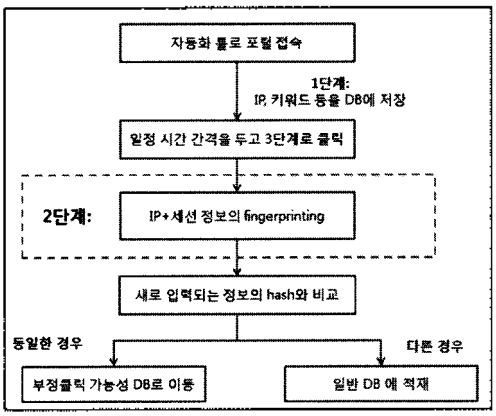
mation'이라는 자동화 도구를 사용하였다. 두 가지 도구를 비교하여 실험한 결과 다양한 접속 기법을 사용할 수 있는 'Test Complete'을 도입하여 1~3단계까지의 실험을 재실시하였다. 특정 웹 페이지에 자동으로 접속하여 일정시간의 세션을 유지하게 하면서 3단계까지 클릭을 유도하였다. 실험은 하루에 6회씩 총 14일간 배치 작업을 통해 실험을 실시했다.

제4.1절과 제4.2절에서 <그림 6>에서 제시한 연구 방법론의 2단계와 3단계의 부정 클릭 식별 방법과 각 단계별 방법에 따른 부정 클릭의 식별 실험 내용을 설명한다.

4.1 IP와 세션 정보 조합의 핑거프린팅

본 연구에서는 부정 클릭 식별을 위한 변수에 IP 이외에 세션 정보를 추가함으로써 자동화 도구를 이용한 부정 클릭의 식별을 강화하는 방법을 제안한다. 자동화 도구는 그 특성상 매크로를 이용하기 때문에 다음 단계로 이동하면서 'Depth'있게 클릭을 할 경우, 반드시 일정한 시간 간격으로 동일한 위치 정보를 클릭하는 성격을 가진다. 따라서 부정 클릭의 식별 방법을 강화하는 1단계로써 'Depth'를 한 단계에서 다음 단계로 넘어가면서 클릭을 할 때 대기하는 정보인 세션 정보를 부정 클릭 식별의 중요한 기준으로 활용할 수 있다. 특정 웹 페이지에 자동으로 접속하여 일정시간의 세션을 유지하게 하면서 3단계까지 클릭을 유도하였다. 첫 번째로는 매크로를 이용하여 일정한 패턴으로 웹 페이지에 접속하게 하였다.

2단계의 실험 프레임워크는 아래 <그림 7>과 같이 구성된다.



〈그림 7〉 IP와 세션 정보의 핑거프린팅을 통한 부정 클릭의 식별 흐름도

단계별 클릭 정보는 <표 1>과 같다. 접속 페이지는 포털의 경우 1단계의 IP 정보를 참조하여 부정 클릭으로 간주하기 때문에 포털이 아닌 국내의 모 회사 사이트를 참조하였다.

〈표 1〉 접속 단계

구 분	1단계	2단계	3단계
Depth	홈페이지 접속	링크	국내보안사이트/국외보안사이트
		소개	고객/연락처/연혁/세휴사/회사소개
		검색	디비 아이
		소만사 홈	
		제품	메일아이/웹키퍼

<표 1>에서 보는 것처럼 3단계에 걸쳐 'Test Complete'로 사용하여 지속적으로 클릭을 시도하였다. 이 때 1~3단계까지의 세션 유지 시간은 다음과 같다.

〈표 2〉 접속 단계 세션 유지 시간

구 분	1단계	2단계	3단계
세션 유지 시간	5초	3초	10초

처음 클릭하여 페이지에 머무는 시간을 5초로 하고, 다음 단계로 이동하여 머무는 시간을 3초, 마지막인 세 번째 단계에서는 페이지에 머무는 시간을 10초로 설정하였다. 이는 일반적으로 사람들이 웹 페이지에 접속을 하여 'Depth'있게 클릭을 하는 행위 패턴을 참고한다. <표 2>에 제시된 세션 유지 시간은 세션 지연(delay) 시간을 포함하며, 이 시간이 지나도 접속이 되지 않는 경우는 로그를 사용하여 접속을 끊도록 유도하였다. 결과적으로 접속 기록 DB에 다음의 <그림 8>에 제시된 것과 같은 테이블이 생성된다.

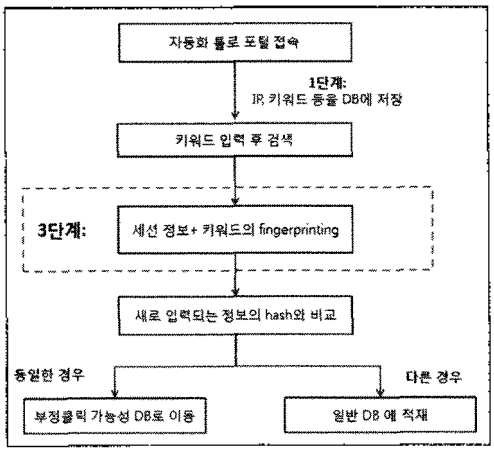
번호	Depth1	Depth2	Depth3	세부경로 URL	클라이언트 IP
1	소만사			www.somansa.com	192.168.1.11
2		펌프	국내보안사이	www.somansa.com/korean/link/L	192.168.1.11
3			국외보안사이	www.somansa.com/korean/link/L	192.168.1.11
4		사이트맵		www.somansa.com/korean/link/s	192.168.1.11
5		소개	고객	www.somansa.com/korean/abou	192.168.1.11
6			연락처	www.somansa.com/korean/abou	192.168.1.11
7			연혁	www.somansa.com/korean/abou	192.168.1.11
8			제품사	www.somansa.com/korean/abou	192.168.1.11
9			회사소개	www.somansa.com/korean/abou	192.168.1.11
10		소만사 통		www.somansa.com/korean/index	192.168.1.11
11		제품	메일아이	www.somansa.com/korean/prodi	192.168.1.11
12			웹키퍼	www.somansa.com/korean/prodi	192.168.1.11
13		검색	디비아이	www.somansa.com/korean/searc	192.168.1.11

〈그림 8〉 단계별 DB 테이블

DB 로그를 보면 각 세션 별로 방문 경로는 하나이므로 방문목록 테이블(VisitList table)에 데이터가 적재되게 되므로 테이블 단위로 핑거프린팅을 하여 동일한 결과값을 비교할 수 있다.

4.2 세션 정보와 키워드 정보 조합의 핑거프린팅

본 연구에서는 클릭 이외에 키워드 입력 행위가 있을 것을 가정하여 변수에 세션 정보 이외에 키워드를 추가한 것을 3단계의 부정 클릭 식별 방법으로 제안한다. 3단계의 실험 프레임 워크는 <그림 9>와 같다.



〈그림 9〉 세션 정보와 키워드의 핑거프린팅을 통한 부정 클릭 식별 흐름도

자동화 도구를 통해 사이트에 접속을 하는 세션의 패턴을 분석하기 위해 'Click Mind'라는 로그 분석 도구를 사용하였다. 동일한 시

번호	Depth1	Depth2	Depth3	세부경로 URL	클라이언트 IP
1	소만사	검색	디비아이	www.somansa.com/korean/searc	192.168.1.11
2			디비아이	www.somansa.com/korean/searc	192.168.1.11
3			디비아이	www.somansa.com/korean/searc	192.168.1.11
4			디비아이	www.somansa.com/korean/searc	192.168.1.11
5			디비아이	www.somansa.com/korean/searc	192.168.1.11
6			디비아이	www.somansa.com/korean/searc	192.168.1.11
7			디비아이	www.somansa.com/korean/searc	192.168.1.11

〈그림 10〉 키워드에 따른 DB 테이블

간 간격을 갖는 세션 정보에 동일한 키워드를 반복적으로 검색하였다. 이에 대한 DB 테이블은 다음의 <그림 10>과 같이 생성된다.

'Click Mind'에는 세션 별 정보는 기록되지만 동일한 세션과 키워드의 패턴을 찾아내는 기능이 없기 때문에 이는 추가적으로 다음과 같은 스크립트를 이용하여 세션 정보와 키워드를 조합하여 핑거프린팅 하도록 구현하였다.

```
www.somansa.com/korean/search.asp?
Keyword id=**&level=**
-> 이름: mapping rule 1
1st variable : Keyword id=
2nd variable :              : level=
3rd fingerprint: SHA-2 =
```

〈그림 11〉 동일한 세션과 키워드를 식별하여 핑거프린팅하는 스크립트 예시(일부)

로그에 대해 세션 정보와 키워드 정보를 결합하여 SHA-2를 사용하여 핑거프린팅 값을 만든 후, 이후 입력되는 로그에 대해 모두 핑거프린팅 값을 만들었다. 이후 14일간 DB 내에 쌓인 핑거프린팅 값의 비교를 위해 'Click Mind'의 DB 내에 Hash 값 비교 테이블을 만들어 배치 작업을 실시하였다. 배치 작업은 1일 1회 오전 1시에 동작되었고, 동일한 값이 존재하면 이를 부정 클릭 가능성이 있는 테이블인 이상 탐지 로그로 이동시켰다.

14일간의 실험 결과는 다음과 같다. 세 개의 단계로 구성된 자동화 도구의 자동 클릭 행위는 모두 DB에 저장되었다. 저장된 정보 중 방문 시각을 제외한 세 단계의 세션 정보만을 핑거프린팅 한 결과, 자동화 도구가 발생시킨 세션 정보 핑거프린팅 값은 모두 동일하게 나타났다. 3단계에서 동일한 키워드를 입력한 경우에도 세션정보와 동일 키워드의 핑거프린팅 값은 동일하였다. 이를 바탕으로 로그 분석 도구를 이용하여 같은 핑거프린팅 값을 갖는 로그의 IP 값을 추출하여 비교해 준 결과, 모두 'Test Complete'을 이용하여 부정 클릭을 발생시킨 IP이었음을 확인하였다. Hash 비교는 다음과 같은 스크립트를 사용하였다.

```
저장된 DB내 Hash 비교 스크립트:
이름: mapping rule
2단계: Baseline(SHA-2 = 2nd
variable :              : level=
3단계: 3rd fingerprint SHA-2 =

웹 페이지 접속 시 부정클릭의
Hash 생성스크립트:
2단계: Baseline(SHA-2 = 2nd
variable :              ; level=
3단계: 3rd fingerprint SHA-2 =
```

〈그림 12〉 DB에 저장된 핑거프린팅 값과 웹 페이지 접속 시 클릭 핑거프린팅 값의 비교 스크립트(일부)

<그림 12>에서 보는 것처럼 접속하는 세션정보와 키워드 값은 IP 정보와 함께 핑거프린팅되어 DB에 저장된다. 웹 페이지에 접속을 시도하는 움직임은 DB에 저장되기 전에 핑거프린팅 값으로 변환되어 DB에 저장된 값과 비교된다. 양쪽의 값을 비교하는 비교 엔진은 로그 분석 툴인 'Click Mind'에 있는 것을 사용하였다. 핑거프린트 값을 만들 때는 해쉬 알고리즘으로 SHA-2만을 사용했지만, 두 값의 비교 시에는 SHA-2의 결과값을 유일한 값으로 지정하기 위해 baseline이라는 함수를 사용하였다.

5. 결 론

본 논문에서는 부정 클릭의 식별 능력을 향상시킬 수 있는 새로운 방법론을 제안하고 이에 대한 간단한 실험을 통해 모델의 유효성을 검증하였다. 연구에서 제안한 부정 클릭의 식별 방법은 3단계로 나누어 진행되었다. 1단계에서는 IP 정보의 평거프린팅, 2단계에서는 IP와 세션 정보의 평거프린팅, 3단계에서는 세션 정보와 검색한 키워드의 평거프린팅 값을 만든다. 결과 값을 통해 동일한 평거프린팅 값을 갖는 것을 자동화 도구를 사용한 부정 클릭의 가능성이 높다고 판단함으로써 보다 정교한 부정 클릭 식별 모델을 제시하였다.

본 연구는 실제로 다양한 대형 포털의 사이트에 적용되어 실제 유효성을 검증하지 못했다는 한계를 갖는다. 또한 실험의 배치 작업의 결과값을 메일 통계치로 받아들여 시각적으로 제공하는 엔드 포인트 단계의 완성된 서비스까지는 확장되지 못한 프로토 타입의 한계를 갖는다. 그러나 기존의 연구가 부정 클릭을 식별해내는 변수를 IP, 키워드 등으로 정하고 개별 정보를 비교하는 한계를 가진 것에 비해, 본 연구에서 제안한 방법과 같이 각각의 개별 변수를 조합하여 만든 평거프린팅 값을 비교하여 부정 클릭의 식별에 활용한 경우 부정 클릭의 식별 방법은 한층 더 정교화 될 수 있음을 확인하였다. 본 연구를 바탕으로 이후 DB에 저장되는 일정한 패턴들에 대해 고유 값을 갖는 조합 정보로서 만들어낸다면 한층 더 정교화 된 부정 클릭 식별 방법을 만들어 낼 수 있을 것으로 기대한다.

참 고 문 헌

[1] 김형신, 유병섭, 이재동, 배해영, "데이터웨어하우스에서 해쉬 테이블을 이용한 효율적인 데이터 큐브 생성 기법", 한국정보과학회 2005년도 가을 학술발표논문집, 제32권, 제2호, pp. 211-213, 2005.

[2] 이경전, 이현석, 전정호, "CPC 방식의 키워드 검색광고에서의 사기 클릭의 정의와 대응방안 평가", 한국경영정보학회, 2008 추계 한국경영정보학회 학술대회논문집, pp. 111-117, 2008.

[3] 오창우, "인터넷 검색광고 요금체계의 특징 및 무성 클릭 유형에 관한 연구, 광고학 연구", 한국광고학회, 광고학연구, 제19권, 4호, pp. 7-28, 2008년.

[4] 홍영란, 김동수, "부정 클릭의 식별 방법을 높이는 방법에 대한 연구", 한국전자거래학회 2011 춘계학술대회논문집, 연세대학교, 서울, 2011.

[5] 박태윤, 최한주, "네모 도리, 최유진 외 'D.I.Y 액세서리 쇼핑몰 절대로 하지 마", 정보문화사, 2007.

[6] 방송통신 위원회, 한국전파진흥원, "방송통신 융합환경에서의 시스템적인 광고유통방안에 관한 연구 : 인터넷 광고를 중심으로", 방송통신 위원회, pp. 172-174, 2009.

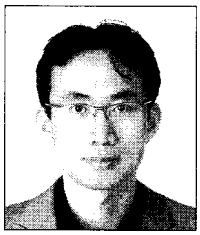
[7] 한국 인터넷 진흥원, "인터넷 광고 관련 국내·외 법·제도 동향 조사 분석 위탁용역", 한국인터넷 진흥원, pp. 28-39, 2009.

[8] www.cpcgaurd.com.

[9] www.logger.co.kr.

저 자 소 개

홍영란	(E mail : yrhong@hotmail.com)
1994년	서울대학교 경영학과 (학사)
1996년	서울대학교 대학원 경영학과 (석사)
2010년 ~ 현재	숭실대학교 산업정보시스템공학과 (박사과정)
관심분야	정보보호, DLP)Data Loss Prevention), Network and Endpoint Security



김동수	(E-mail : dskim@ssu.ac.kr)
1994년	서울대학교 산업공학과 (학사)
1996년	서울대학교 산업공학과 (석사)
2001년	서울대학교 산업공학과 (박사)
2001년 ~ 2003년	한국정보사회진흥원 전자거래연구부 e-Biz 표준팀장
2003년 ~ 2006년	가톨릭대학교 의료경영대학원 전임강사, 조교수
2006년 ~ 현재	숭실대학교 산업 · 정보시스템공학과 조교수, 부교수
관심분야	BPM, e-Business 정책 및 기술, 기업정보시스템, e-Health, 정보보호