프라이버시 유출 방지 SNS 개발

김영아, 황천, 굴가, 윤원탁, 박두순
순천향대학교 컴퓨터소프트웨어공학과
e-mail : kya5452@naver.com

Development of SNS for Privacy Data Loss Prevention

Young-A Kim, Qian Huang, Ke Qu, Won-Tak Yoon, Doo-Soon Park
Dept. of Computer Software Engineering, Soonchunhyang University

요 약

데이터가 폭발적으로 증가함에 따라 필요한 정보들을 찾는 것은 더욱더 어려워지고 개인의 생각이나 많은 자료들을 SNS 공간을 통해 공유함으로써 프라이버시 유출도 많아지게 된다. 대부분의 SNS는 자신의 공간에 게재된 정보에 대한 접근권한 만을 설정할 수 있고 자신이 타인의 공간에 게재한 게시물에 대해서는 접근 권한 설정에 대한 자격을 부여하지 않는다. 이를 통해 원치 않은 사용자들에게 까지 자신의 개인 정보가 노출되는데 얼마든지 개인 정보의 유출로 인한 문제들이 일어날 수 있다.

따라서 본 논문에서는 서비스 제공자가 제 3자에게 SNS 그래프 데이터 제공시 개인 정보의 노출을 차단하기 위해 K-Means Clustering 기법을 사용한 방법을 보인다.

1. 서론

최근 인터넷과 스마트폰의 성장세를 통해 사회 전반적으로 많은 변화가 이루어 졌다. 세계에서 가장 많이 사용되는 SNS 는 페이스북으로 12년 10월 기준 10억이 넘는 사용자가 이용하고 있고 이들 중 60% 이상이 스마트폰을 통해 접속하고 있다[1]. 또한 SNS는 기존 커뮤니티 서비스가 주로 글과 댓글 위주로 소통에 제한이 있었던 것에 반해, 일기, 사진, 동영상, 즐겨찾기 정보, 게임 내에서의 성취나 상태, 실제 위치 정보에 이르기 까지 점점 복잡하고 다양한 정보를 공유하고 있다. 이렇게 SNS의 확산은 데이터들이 폭발적으로 증가하는 계기가 되었다.

SNS에서 프라이버시 유출 문제를 해결하기 위해 크게 두가지 방법으로 나누어 연구가 진행 중이다[2]. 이는 써드파티 어플리케이션이나 데이터 분석을 위해 제 3자에게 SNS의 데이터를 전달 할 경우 생기는 문제 하나와, SNS 이용 상황에서 발생하는 문제로 나누어 볼 수 있다.

본 논문에서는 SNS 이용 상황에서 발생하는 문제는 데이터 중심 접근제어라는 방법을 통해 기존에 정보에 대한 접근 권한을 독점했던 정보 보유자에서 정보 생성자나 정보 관련자에게도 권한을 전달하여 좀 더 개인 사생활이 제 3자에게 노출되지 않는 방법을 제안하였다. 또한 제 3자에게 데이터를 전달 시에 생기는 문제의 경우를 해결하기 위해 데이터가 가지는 고유의 성격은 유지하고, 제 3자가 그래프의 정점을 누구인지 파악할 수 없게 하기 위해 SNS의 자료를 제공하는 측에서 K-Means Clustering을 이용해 데이터를 군집화 시켜 보내주는 방법을 제안한다.

2. 프라이버시 보호 방법

기존의 SNS에서 사용자가 게시한 정보의 공개 범위는 정보의 보유자의 설정에 따라 노출이 되었다. 이는 정보를 생성한 자에게는 전혀 관리 할 수 없는 문제가 발생되어 문제점을 야기한다. 페이스북을 통해 실제로 일어날 수 있는 SNS 상에서의 프라이버시 침해 예시는 <그림 1>과 같다. A와 C가 친구가 아닐 때 댓글이 노출되는 상황을 보인다.

<그림 1> 소셜 네트워크 사이트에서 발생할 수 있는 프라이버시 침해 예시

소셜 네트워크는 소셜 네트워크에 속한 개인

$V=\{v_1, v_2, v_3, \dots, v_n\}$, 개인 간의 관계 집합 $E=\{e_1, e_2, e_3, \dots, e_n\}$의 그래프 $G(V, E)$로 나타낼 수 있다[3].

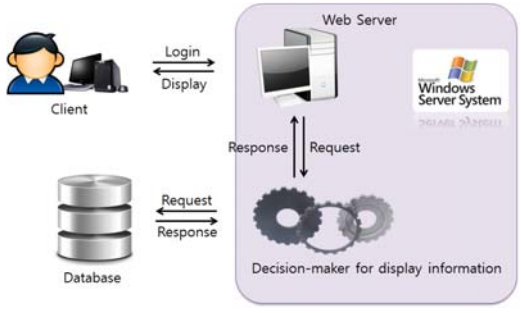
정점 v_a, v_b의 최단 경로 길이를 $dist(v_a, v_b)$라 정의한다. 또한 사용자 v_u의 n단계 관계집합을 $rs(v_u, n)$이라 정의한다. 그럼 이때 $rs(v_u, n)$는 $dist(v_u, v_t)=n$을 만족하는 원소들의 집합이 된다. 1 단계부터 k 단계 까지의 관계집합들의 모임을 $RS(v_u, r)$이라하고 이를 관계껍질이라 정의한다.

즉 정보 생성자 v_a가 공유하고자 하는 단계의 관계껍질과 정보 보유자 v_b가 공유하고자 하는 단계의 관계껍질 사이에 존재하는 사용자들에게 정보를 노출시킨다면 제 3자에게 정보가 노출되는 문제를 해결할 수 있다. 이를 수식으로 표현하면 $RS(v_a, x) \cap RS(v_b, y)$과 같다.

하지만 1차원적 정보인 프로필이나 텍스트, 사진 등의 정보에 대한 노출은 차단하였지만, 태그한 글이나 사진에 대한 접근 설정 권한을 부여하지 않아 사생활이나 그들이 공유하고자 하는 정보 유출 방지를 제공하지 못했다. 그러므로 2차원적 정보에 대한 유출 피해를 막는 장치를 제공하기 위해 태그 정보를 고려한 데이터 중심 접근제어를 이루어 냈다.

3. 프라이버시 유출 방지 SNS 설계

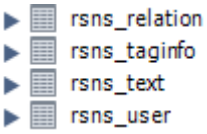
프라이버시 유출 방지 SNS의 구현에 앞서 SNS의 구성에 대해 살펴본다. 프라이버시 유출방지 SNS 구성도는 <그림 2>와 같다.



<그림 2> 프라이버시 유출방지 SNS 구성도

① 사용자는 Web Browser를 통해 SNS 홈페이지에 접속하게 된다. 사용자에게서 입력받은 id와 패스워드를 통해 로그인을 수락하고, ② 그 뒤에 페이지에 띄울 게시물들을 선택한다. 이는 ③ Decison-maker for display information(줄여서 Decision-maker)에서 수행한다. 먼저 Client의 친구들의 리스트를 구하고 친구 리스트에 포함된 친구들이 올린 최신 게시글 10개를 추출한다. 게시글에 대해 Display 해도 되는지에 대한 여부를 판단하는데 이는 위에서 살펴본 공식을 통해 정보 생성자, 정보 보유자, 정보 관련자의 관계껍질을 통한 계산을 통해 결정한다. 만약 Display의 판단 여부에서 접근 권한이 없는 것으로 확인이 될 경우 Decison-maker는 그 다음 친구가 게시한 최신 게시글을 호출한다. 위와 같은 방법으로 진행되어 모든 게시물들이 결정이 될 경우 사용자의 News Feed에 띄워준다. 사용자가 게시물을 남길 경우 사용자는 정보 생성자가 되고 게시물을 남기는 담벼락의 주인이 정보 보유자가 된다. 만일 정보 게재 시 태그 정보가 포함되어 있다면 정보 관련자 또한 다 같이 저장된다.

Decision-maker에서 직접 접근하는 Table 목록은 <그림 3>과 같다.



<그림 3> Decision-maker에서 사용하는 Table 목록

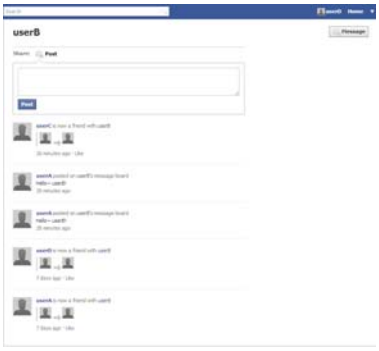
rsns_relation에는 사용자간의 관계를 저장하고 있다. rsns_text는 게시글의 정보를 저장하고 있고, rsns_taginfo는 게시물에 추가된 태그 정보를 저장하고 있다. rsns_user는 등록된 사용자들의 정보를 담고 있는 테이블이다.

4. 프라이버시 유출 방지 SNS 구현

userB가 자신의 게시판에 사진을 올린 화면은 <그림 4>와 같다. 기존의 SNS에서는 userB가 올린 게시물이 친구인 userA, userC, userD에게 모두 보여야 하겠지만 이는 userD와 친구 관계가 아닌 userA, userC에게는 달가운 상황이 아닐 수 있다. 그렇기에 userA와 userC가 태그가 된 점을 감안하여 userA, userC 두 사용자의 관계껍질을 계산하여 userD가 포함되지 않는 것을 확인 후 노출시키지 않는다.

<그림 4> userB가 자신의 담벼락에 사진을 올린 화면

userD가 로그인 한 후 userB의 페이지로 간 화면은 <그림 5>와 같다. userA가 userB에게 남긴 메시지는 확인할 수 있으나 앞에서 확인했던 Presentation.png 파일과 userC의 comment는 확인할 수 없다.



<그림 5> userD가 userB의 담벼락을 조회

게시물에 태그 정보가 존재하는지 rsns_taginfo 테이블의 조회를 통해 정보 관련자를 알아낸다. 이와 같은 과정을 통해 알게된 정보 생성자와 정보 보유자, 정보 관계자들의 관계 꼽질에 해당하는 사용자들을 rsns_relation 테이블에서 찾아내고 그들의 교집합에 존재하는 사용자들에게 게시 글을 노출시킨다.

RobustSNS에서 제 3자에게 데이터 제공시에 개인 정보를 은폐하기 위해 데이터를 가공하기 위한 알고리즘은 <표 1>와 같다.

<표 1> RobustSNS의 데이터 제공 기법 알고리즘

RobustSNS의 데이터 제공 기법 알고리즘
1 : INPUT : U_k : 사용자 개인 성향 정보, R_n : 사용자 간의 관계 정보
2 : OUTPUT : CD : 클러스터링 된 사용자간의 관계정보
3 : PROCEDURE : 제 3자의 SNS 데이터 정보 요청
4 : 개인 성향 정보를 제외한 이름, 연락처 정보 삭제
5 : U_1부터 U_k까지의 개인 성향 정보를 이용한 클러스터링 = C_k
6 : C_k의 분류에 따라 R_n의 정보 또한 클러스터링 된 형태로 변환
7 : 클러스터링 된 사용자 정보와 관계 정보 제공

위와 같은 데이터를 얻은 이후 사용자들의 K-Means Clustering을 통한 군집화를 진행하여 개개 사용자의 관계 정보가 아닌 클러스터링과 클러스터링간의 관계 정보를 제공하고, 이는 개개인을 식별할 수 없게 한다.

K-means Clustering 기법을 이용하여 SNS 그래프 데이터를 가공하고 제 3자에게 제공한다면, 분석을 위한 SNS 그래프의 특성은 그대로 살리면서 특정 개인의 정보를 노출시키지 않는 방법으로 데이터를 제공할 수 있다.

5. 결론

마이스페이스를 시작으로 SNS는 거대한 인기를 가지고 왔다. 하지만 화려한 발전과 동시에 SNS가 가져오는 사생활 침해는 이미 위험수위를 넘어서고 있다.

본 논문에서는 SNS상에서 자주 사용되는 태그 정보를 고려한 데이터 중심 접근 제어 모델을 구현하였다. 기존에 연구되었던 데이터 중심 접근제어는 1차원적인 게시글, 사진 등과 같은 정보만을 고려한 방식 이였고 태그 정보와 같은 2차원적인 문제에 대해 접근하지 못했다.

이러한 문제를 해결하기 위해 기존에 고려했던 정보 생성자와 정보 보유자와 함께, 정보 관련자에게도 정보에 대한 접근 권한 설정을 주고 이를 통해 2차적인 문제의 발생을 막을 수 있게 하는게 본 연구의 목적이다.

참고문헌

[1]http://www.phonearena.com/news/Facebook-has-1-billion-active-users-60-mobile_id35183

[2] 성민경, 정연돈, “SNS에서 프라이버시 문제 및 보호방안”, 한국통신학회논문지, 제 29권, 제 5호, pp.92-97, 2012.

[3] D.M. Boyd and N.B. Ellison, "Social network sites: definition, history, and scholarship," Journal of Computer-Mediated Communication, vol.13, no.1, Article 11, 2007.