複数の SNS を対象にしたアカウント到達可能性算出モデルの検討

理学専攻 情報科学コース
吉國綺乃

1 はじめに

近年，ソーシャルネットワークサービス（以下 SNS とする）が普及し，それに伴い世界中で利用者が増加している．SNS は多くの人とコミュニケーションをとる手段として有効である一方，その利用によるトラブルが原因となり利用者の個人情報が取得されるという事例が発生している．これらの個人情報は利用者自身が気がつかないうちに，投稿内容やプロファイルを公開していることが多い．

SNS の利用において個人情報が取得されることを防ぐためにも，自身が投稿した内容やプロファイルがどの程度のプライバシリスクになっているかを把握する必要性が増している．本研究では SNS におけるプライバシリスクの指標としてアカウント到達可能性を提案し，攻撃モデルを元にアカウント到達可能性の算出手法の検討を行う．また，利用者が直感的にプライバシリスクに気付くために，アカウント到達可能性算出により得られた情報がわかりやすく提示される必要がある．そこでそれらの情報の提示手法の検討を行う．

2 アカウント到達可能性

アカウント到達可能性 (Account Reachability) とは，SNS において攻撃者が対象ユーザの既知のアカウントから別の所有アカウントを見つけ出す可能性を表す．たとえば，ある利用者が二つの異なる SNS のアカウント s_1 ， s_2 をそれぞれ持っているとする．また攻撃者は利用者の SNS アカウントのうち， s_1 のみしか知らないとする．攻撃者は s_1 の情報を元にして，まだ知らないアカウントである s_2 をさまざまな手法を通して見つけ出そうとする．ここで攻撃者は s_1 のプロファイルや投稿内容から s_1 のキーワードを抽出し，検索エンジンなどを用いて検索を行い，s_2 になりうるアカウントの候補を取得する手法をとったとする．このとき，取得した候補アカウントそれぞれと s_1 から取得したキーワードを元に s_1 との類似度を測り，s_2 が s_1 のアカウントであると特定していく．これらの一連の過程を経て，未知のアカウントが発見される可能性がアカウント到達可能性である．アカウント到達可能性は以下の式で求められる．

$$AR(s_1 \to s_2) = max_{q \subseteq Q}\,(AR(s_1, s_2, q))$$

$$Q = GenQueries(s_1.prof, s_1.msg).$$

$$AR(s_1 \to s_2, q) =$$

$$Match(s_2, Cand(q)) * \frac{Score(s_1, s_2)}{\sum_{c \in Cand(q)} Score(s_1, c)}$$

$$Match(s_2) = \begin{cases} 1\ if\ s_2 \in Cand(q) \\ 0\ else \end{cases}.$$

ここで，$GenQueries\ (s_1.prof, s_1.msg)$ は s_1 のプロファイルや投稿内容から， s_2 のアカウントを見つけ出すためのクエリを生成する式である．Q は生成されたクエリの集合 $(q \in Q)$ であり，$Cand(q)$ はクエリ q で得られた s_1 の別アカウントの候補アカウントの集合である．$Score(s_1, c)$ は s_1 と候補アカウント c との類似度を表す．

3 アカウント到達可能性算出モデルの検討

アカウント到達可能性を求めるために用いられる関数 $GenQueries(s_1.prof, s_1.msg)$ と $Score(s_1, c)$ は，攻撃者が入手できるデータや利用出来る技術に基づいて実装される．考えられる攻撃モデルを元に，本研究では，表 1 に示す手法を用いてアカウント到達可能性算出式の検討を行った．

表 1: $GenQueries(s_1.prof, s_1.msg)$ と $Score(s_1, c)$ の代表例

$GenQueries(s_1.prof, s_1.msg)$	$Score(s_1, c)$
プロファイルから生成	検索エンジンでのランキング
プロファイル及び投稿内容から生成	s_1の投稿内容から取得した地域情報と cから取得した地域情報を元に類似度を算出
プロファイル及び投稿内容から生成	s_1の投稿内容から取得した興味と cから取得した興味を元に類似度を算出

3.1 攻撃モデル

攻撃者が技術知識を持っていなくてもできる最も単純な攻撃手法を元に，以下の手法を実装した．

$GenQueries(s_1.prof, s_1.msg)$ **：**
$KeywordSearch(ks, engine)$ という関数を導入した．検索エンジン $engine$ を用いてキーワード ks を検索する関数である．アカウント s_1 のプロファイルから名前，所属，誕生日といったキーワードを抽出する．それぞれのキーワードに対して，関数 $KeywordSearch(ks, engine)$ を適用する．この $KeywordSearch$ をクエリとする．

$Score(s_1, c)$ **：**
検索エンジンを用いた検索の結果は，検索順位順に結果ページに出力される．これを用い $Score$ を以下のように定義した．

$$Score(s_1, c) = \frac{1}{Rank(ks, c)}$$

$$Rank(ks, c) = \quad \{ks \text{ を検索した} \\ \text{結果ページにおける } c \text{ の順位}\}$$

3.2 検証

対象 SNS を Twitter と Facebook とし，両方の SNS アカウントを持つ被験者 50 名に対して検証を行った．本検証は，Facebook のアカウントから Twitter アカウントへの到達可能性を算出する．被験者 50 名のうち 26 名はふたつのアカウントが関連付けられても構わな

いユーザ (“NotCare” users) ，24 名は関連付けられたくないユーザ (“NotWant” users) であり，各被験者のアカウント到達可能性を算出した．*GenQueries* で生成したキーワードは Facebook のプロファイルからそれぞれ名前，所属，誕生日，居住地などを取得した．検索エンジンは Google を用いている．結果を図 1 に示す．

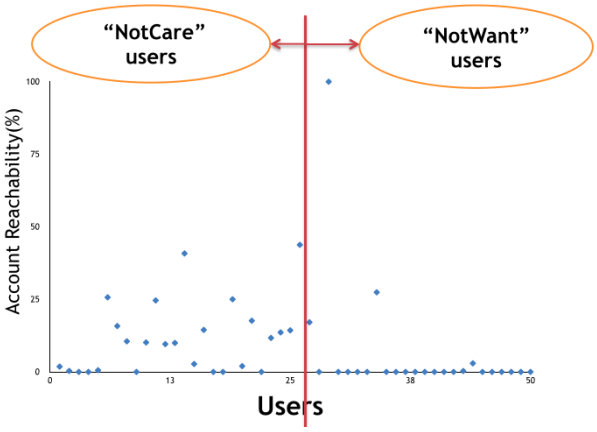

図 1: アカウント到達可能性の値

関連付けられたくないと考えていてもアカウント到達可能性が 100%の被験者が存在した．この被験者は自身が想定している以上の情報をそれぞれの SNS に公開していた．このように自身が気付くことのなかった危険性にアカウント到達可能性を求めることで気付くことができた．これより，プライバシリスクを示す指標としてアカウント到達可能性が有効であると考えることができる．

4 アカウント類似性の可視化

アカウント到達可能性は未知のアカウントを見つけ出す可能性であり，利用者は得られた数値が危険な値であるのか判断が難しく，また数値だけでは今後どのように対策すべきかも分かりづらいことが課題となった．

そこで，アカウント到達可能性によりアカウントが特定されたと仮定し，それぞれのアカウント同士の類似性を可視化し利用者に提示することを検討する．類似性の可視化には *Score* の算出過程で得られた情報に着目した．それぞれの SNS で得られた情報を可視化することで，ふたつの SNS に共通しているところ，また異なるところの比較が可能となる．本節では地域情報に着目して類似性の可視化の検討を行う．地域情報は個人を特定するために重要な情報のひとつである．各 SNS で得られた地域名を地図上にマッピングすることで視覚的に行動範囲が見られるようになり，文字情報より直感的に理解できると考えられる．

Facebook ではプロファイルが決められたフォームで登録されており，地域情報は利用者がプロファイルにて公開している場合抽出が可能である．一方，Twitter ではプロファイルは自由記述であり，決められたフォームが存在しない．そこで地域情報は利用者のツイートから出現頻度の高い地名を取得することとした．

4.1 地域情報を用いた類似性の可視化の具体例

被験者の SNS 情報から得られた地域を地図上にマッピングした結果を元に，提案手法がプライバシリスクの意識向上に有効であるか考察する．地図は Google Maps を用いた．取得する地域情報は，Twitter では被験者の最新ツイート 2000 件から抽出した地域名のうち出現頻度上位 5 件を取得する．地域名の抽出には形態素解析器 MeCab を用いた．Facebook からはプロファイルに公開している地域名（居住地，出身地，スポット等）を取得した．それぞれ取得した地域を各 SNS のアイコンと Google Maps API で公開されているサークルを描画する機能を用いて行動範囲を描画する．ここでは 1 名の被験者の結果を図 2, 3 に示す．

図 2: 地域情報提示結果

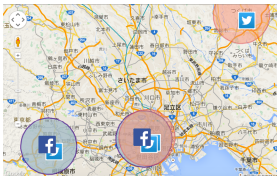

図 3: 結果 詳細

この被験者のアカウント到達可能性は 0.17 である．値を見るとあまり特定される危険性は感じられない．しかし図 2，3 を見ると類似性は高く，特定される危険性の高さがわかる．図 2 をみると，それぞれの SNS で似た分布が得られた．関東近辺を拡大した図が図 3 である．それぞれ重複しており，ふたつの SNS の使い分けができていないことがわかる．また地図上にマッピングすることで行動範囲がわかりやすくなった．

5 まとめと今後の課題

SNS の普及により，利用者の個人情報が取得される事例が多く発生している．これらのプライバシリスクに利用者が気付くためにアカウント到達可能性を提案した．また，さまざまな攻撃モデルを元に算出手法の検討を行った．最も単純な攻撃モデルを元に検討した算出式で検証を行った結果，想定以上の値を得た被験者が多く，この指標はプライバシリスクに気付く有効な指標であると考えられる．また，ふたつの SNS から得られた情報をわかりやすく提示することで，利用者のプライバシリスクに対する意識が向上することを目的とし，その提示手法を検討した．本研究では地域情報を提示する手法の検討を行った．地図上に情報をマッピングすることで文字ではわからなかった行動範囲などがわかりやすくなり，自身では気付かなかった情報を得ることができた．

今後，被験者を増やし検討と考察を行っていく．また，地域情報以外の情報の提示手法も検討を行う．具体的には，SNS 上での活動時間の可視化，それぞれの SNS の友人の可視化など検討している．

参考文献

[1] Narayanan Arvind and Shmatikov Vitaly ”De-anonymizing Social Networks”, Proceedings of the 2009 30th IEEE Sym- posium on Security and Privacy, 2009.

[2] Mislove Alan and Viswanath Bimal and Gummadi Krishna P. and Druschel Peter ”You are who you know: inferring user profiles in online social networks”, WSDM, 2010.