一般社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report

マイクロブログにおけるアカウントのなりすまし判定の試み

中才 恵太朗† 角田 雅照†

†近畿大学理工学部 〒577-0818 大阪府東大阪市小若江 3-4-1
E-mail: †n4keitaro@gmail.com, tsunoda@info.kindai.ac.jp

あらまし 近年，代表的なマイクロブログサービスである Twitter が広く利用されている．Twitter 社はなりすましを防止するため，サービス上で本人であることを保証する，認証済みアカウントを設けている．ただし，全ての著名人が認証済みアカウントを利用しているわけではなく，著名人のアカウントをフォローしようと試みるユーザは，アカウントが本物かどうかを十分に検討する必要がある．なりすましアカウントをフォローすることにより，誤った情報を入手してしまう可能性もある．本稿では，Twitter アカウントのなりすましを判定するアルゴリズムを提案するとともに，アルゴリズムの有効性を確かめるために，簡易な実験を行った．その結果，本人であるか疑わしいアカウント 2 個と，認証済みでないが本人であるアカウント 2 個を正しく判定できた．

キーワード 偽装アカウント，Twitter，フォロワー，判別

Preliminary Study of Identifying Spoofing Accounts on Microblog

Keitaro Nakasai† Masateru Tsunoda†

† Faculty of Science and Engineering, Kindai University 3-4-1 Kowakae, Higashiosaka City, Osaka, 577-8502 Japan
E-mail: †n4keitaro@gmail.com, tsunoda@info.kindai.ac.jp

Abstract Recently, microblog services such as Twitter are widely used. To prevent spoofing, Twitter provides verified account service which assures the account is not spoofing. However, not every famous people use verified account, and hence when a user tries to follow an account of a famous person, he/she should check whether the account is spoofing or not. If he/she follows spoofing account, he/she may believe fake information. This study proposes new algorithm which discriminate spoofing accounts, and performed a case study to evaluate the algorithm. As a result, the algorithm discriminated spoofing accounts and true accounts.

Keywords Imitation account, Twitter, follower, discrimination

1. はじめに

近年，代表的なマイクロブログサービスである Twitter が広く利用されている．個人に加え，企業や著名人も Twitter を利用しており，宣伝のために効果的に利用している．ただし，Twitter の普及に伴い，著名人のアカウントのなりすまし（他者があたかも本人のようなアカウントを作成して公開すること）も増加しており，問題となっている．女性アイドルや，女性声優，政治家の偽アカウントが特に多い傾向がある．これらのなりすましアカウントは，Google の検索結果でも上位に現れる場合が多くある．Twitter 社はなりすましを防止するため，サービス上で本人であることを保証する，認証済みアカウントを設けている．

ただし，全ての著名人が認証済みアカウントを利用しているわけではなく，著名人であっても通常のアカウントで Twitter を利用している場合もある．この場合，著名人自身の公式サイトやブログに自身のアカウントをリンクすることで本物であることを証明している場合もある[10]．この場合でも，全ての著名人が自身の公式サイトを持っているわけではなく，公式サイトを持っている場合でもアカウントをリンクしているとは限らない．

このため，著名人のアカウントをフォローしようと試みるユーザは，アカウントが本物かどうかを十分に検討する必要がある．Twitter には SNS (Social Networking Service)の機能が実装されており，アカウント同士でつながりを持つことができる．これがフォロー機能と呼ばれるものであり，あるアカウントをフォローしている他アカウントをフォロワーと呼ぶ．偽物のアカウントをフォローすることにより，誤った情報を入手してしまう可能性もある．

そこで本研究では，マイクロブログで最も普及している Twitter に着目し，著名人のアカウントがなりすましであるかどうかの判定を支援することを目的とする．そのために，Twitter アカウントのなりすましを判定するアルゴリズムを提案するとともに，アルゴリズムの

図 1 標準タイプのアカウントの一例 (https://twitter.com/potus, 2016 年 2 月 9 日閲覧)

図 2 フォロー返しタイプのアカウントの一例(https://twitter.com/yokoono, 2016 年 2 月 9 日閲覧)

図 3 フォロー無しタイプのアカウントの一例 (https://twitter.com/masuzoeyoichi, 2016 年 2 月 9 日閲覧)

有効性を確かめるために，簡易な実験を行う．

2. 著名人の Twitter アカウントの分析

著名人の Twitter アカウントは 3 個のタイプに分けることができる．以下，3 個のタイプについて説明する．

2.1. 標準タイプ

多くの著名人の Twitter アカウントでは，フォロー数は少ないが，フォロワー数は多いという特徴がある．図 1 にこのタイプの代表例を示す．フォロー数が 71 であるのに対し，フォロワー数が 6,226,946 と非常に多い．フォローするアカウントには類似の職業に従事している直接の知り合いが多く，相互にフォローしていることが多い．Twitter サービスでは，数多くのアカウントをフォローすると，タイムラインと呼ばれる画面にフォローしているアカウント全てのツイート（入力内容）が表示される．このため，著名人のアカウントでない場合でもフォロー数は抑えられることが多い．従って，このタイプのアカウントにフォローされたアカウントは厳選されたアカウントである可能性が高く，信頼できるといえる．これに対し，著名人のフォロワー数は特に多い傾向がある．これは著名人の情報を知りたい多くのユーザがフォローするためである．

2.2. フォロー返しタイプ

著名人の中には一般の人の Twitter アカウントをフォロー返しする人も存在する．図 2 にこのタイプの代表例を示す．フォロー数が 955,264 と非常に多い．このタイプはフォロー数が多すぎるため，フォローする人を精査していないと考えられる．自身のアカウントにフォローがあったものを定期的にフォローしている．また，リプライと呼ばれるユーザ宛にツイートがあったものを定期的にフォローしていることが考えられる．従って，このタイプのアカウントにフォローされているからと言って信頼できるとは言えない．

2.3. フォロー無しタイプ

著名人の中には何もフォローしないタイプも存在する． 図 3 にこのタイプの代表例を示す．このタイプはフォロー数が 0 のため，本研究の提案手法であるフォローリストを用いるアルゴリズムは適用できない．

3. 提案方法

3.1. 提案アルゴリズム

本研究では，著名人の Twitter アカウントが本人であるかどうかを確認するために，対象アカウントのフォローとフォロワーの関係を利用する．標準タイプの著名人の Twitter アカウントのフォローは 2.1 で述べたとおり，類似の職業に従事している知り合いが多く，相互フォローが多い．また，著名人は Twitter アカウントのフォローに慎重であり，認証済みアカウントを持った著名人がフォローするアカウントは信頼性が比較的高いと考えられる．しかし，認証済みアカウントであっても 2.2 で述べたとおり，フォロー返しタイプのアカウントにフォローされていても信頼性が高いとは言えない．

そこで，本手法では，標準タイプの認証済みアカウントに相互フォローされているかどうかに基づき，Twitter アカウントのなりすましを判定する．具体的には，提案するアルゴリズムでは，以下の手順でアカウントの偽装を判断する．フローチャートを図 4 に示す．

1. 対象アカウントのアカウント情報を取得し，フォロー数が 1 以下なら判定不能とする．3.2.1 で述べる API を使用する．

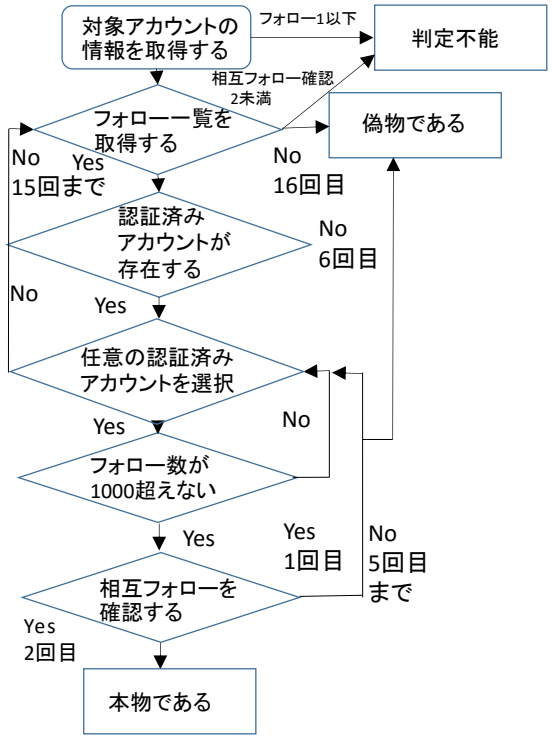

図 4 提案する判定アルゴリズム

2. 判別対象の Twitter アカウントの，フォロー一覧を取得する(最大 15 回まで)．3.2.4 で述べる API を使用する．16 回目で対象アカウントがなりすましである可能性が高いと判断する．また，4 を通り，かつ取得できるフォロー一覧がなくなった場合なりすましである可能性が高いと判断する．ただし，相互フォローを確認した数が 2 個未満の場合，判定不能とする．
3. フォロー一覧から認証済みアカウントがあるかどうかを確かめ，存在しない場合は 2 に戻る．
4. 取得した認証済みアカウントのフォロー数が 1000 以下のものと対象 Twitter アカウントが相互フォローであるかを調べる．3.2.5 で述べる API を使用する．
5. 相互フォローアカウントが 2 個存在するか確認するまで 3 に戻る．相互フォローアカウントが 2 個以上確認できた場合本物であると判断する．ただし，6 回確認して相互フォローアカウントが存在しない場合，偽物であると判断する．

3.2. システムの実装

本研究でも使用する代表的な Twitter API について説明する．Twitter API を利用することで Twitter アカウントの情報（認証済みアカウントかどうか，フォロワー数，フォロー数など）を簡単に取得することができる．また，時間あたりの使用回数が API ごとに決められており，更にユーザ単位及びアプリケーション単位ごとに使用回数が決められている．アプリケーション認証を行う場合は使用するユーザの認証は必要がないがアプリケーション全体で API の使用回数を共有することになる．ユーザ認証を行う場合はユーザ単位で API を使用することができる．本研究では提案アルゴリズムを使用するユーザがアプリケーション認証を行うことを前提としている．

3.2.1. GET users/show

指定したユーザのプロフィール情報を取得できる．15 分に 180 回まで使用できる．認証済みアカウントかどうか，フォロー数，フォロワー数を調べることができる．

3.2.2. GET friends/ids

指定したユーザのフォローリストを取得することができる．一回の取得件数は最大で 5000 件であり 15 分に 15 回まで使用できる．取得件数は多いがユーザ id しか取得できない．提案アルゴリズムでは使用しない．

3.2.3. GET followers/ids

指定したユーザのフォロワーリストを取得することができる．一回の取得件数は最大で 5,000 件であり 15 分に 15 回まで使用できる．取得件数は多いがユーザ id しか取得できない．提案アルゴリズムでは使用しない．

3.2.4. GET friends/list

指定したユーザのフォローリストを取得することができる．一回の取得件数は最大で 200 件であり 15 分に 15 回まで使用できる．また，個々のユーザの情報は認証済みアカウントであるかどうか，フォロー数，フォロワー数などを取得できる．提案アルゴリズムで使用する．

3.2.5. GET friendships/show

指定したユーザ同士が相互フォローであるかを調べることができる．15 分間に 180 件調べることができる．

3.3. 提案アルゴリズムの特徴

相互フォローを調べるツール[1]は存在するが，アカウントの偽装を判定する機能は実装されていない．このサービスでは，フォローまたはフォロワーが 10,000 人以上ある場合，正常に動作しない．これはこのサービスがアプリケーション認証を行っており，Twitter API の使用を抑えようとしているためと考えられる．

なお，提案手法では 3.2.4 の API を使用するため，対象アカウントのフォローリストのうち最新 3,000 アカウントまでが確認対象である．しかし，対象アカウントが 2.2 のフォロー返しタイプであってもフォロー

表 1 実験に用いたアカウント(2016/2/9 現在)

アカウント名	本人かどうか
@rino_sashihara	本人でない可能性が高い
@oshima_yuko_	本人でない可能性が高い
@tokui_sorangley	認証済みではないが本人
@mimori_suzuko	認証済みではないが本人

表 2 本人でないと思われるアカウントの概要（2016/2/9 現在）

アカウント名	フォロー	フォロワー	開設日
@rino_sashihara	14	278,137	2010/12
@oshima_yuko_	19	18,364	2010/10

図 5 なりすましの可能性があるアカウントの一例 (https://twitter.com/rino_sashihara, 2015 年 7 月 23 日閲覧)

リストに著名人の認証アカウントは多く存在し，また，フォロー返しタイプであれば相互フォロー率は極めて高いとため，本アルゴリズムでなりすましの判定が可能ではあるが，精度は落ちると考えられる．

また，単に，3.2.2 及び 3.2.2 の API を使って対象アカウントのフォローとフォロワーの全てを抜き出して同じ ID が存在するかで相互フォローを確認する方法で相互フォロー率を算出することは可能ではあるが，ユーザ認証で相互フォローの確認を行ってもフォロワー数が 75,000 を超えていると 15 分では相互フォローを確認することができない．著名人の Twitter アカウントはフォロワー数が 75,000 を超えることがしばしば発生するため，Twitter API を用いてアカウントをアカウントがなりすましであると判定する際，フォロワーを用いることは困難である．また，そのアカウントが認証済みであることを確認することは困難である．

提案アルゴリズムでは相互フォローを確認する際，信頼できるアカウントだけに絞り込むことで，アカウントがなりすましであると判断している．

4. 実験

提案したアルゴリズムが有効であるか確かめるための，簡単な評価実験を行った．2 個の本人でない可能性が高いアカウント（本人である認証済みアカウントが別に存在しているため偽アカウントと考えられる）と，2 個の認証済みアカウントではない，本人であるアカウント（公式ブログからリンクが存在する）に対してアルゴリズムを適用した．各アカウントを表 1 に示す．

実験で用いたなりすましの可能性が高いアカウントの情報を表 2 に示す．どちらのアカウントも開設日が 4 年以上前である（長い間，偽アカウントだと知られていない）であるに関わらず，フォロワーが 1 万人を超えており，まるで本人であるかのように見える．また，Google の検索結果でもかなり上位に表れる．ただし，実際には偽アカウントであるため，なりすましの判別が必要となる．

以降では，各アカウントに対する判定方法とその結果を述べる．

@rino_sashihara（図 **5**，なりすまし）に対する判定：

1. @rino_sashihara のフォロー数は 14 であり，フォロー数が 1 以上であるため判定可能である．
2. フォローリストを取得する．
3. フォローリストに認証済みアカウントが 12 個存在する．
4. 最初に表示される認証済みアカウント @AyakaUmeda はフォロー数が 460 である．
5. @AyakaUmeda は@rino_sashihara をフォローしていない（1 回目）．
   3. フォローリストに認証済みアカウントが 11 個存在する．
   4. 次に表示される認証済みアカウント @Kobayashikana48 はフォロー数が 152 である．
   5. @Kobayashikana48 は@rino_sashihara をフォローしていない（2 回目）．
   3. フォローリストに認証済みアカウントが 10 個存在する．
   4. 次に表示される認証済みアカウント @ohori_megumi はフォロー数が 125 である．
   5. @ohori_megumi は@rino_sashihara をフォローしていない．（3 回目）．
   3. フォローリストに認証済みアカウントが 9 個存在する．
   4. 次に表示される認証済みアカウント @yuka_masuda はフォロー数が 41 である．

5. @yuka_masuda は@rino_sashihara をフォローしていない．(4 回目)．
3. フォローリストに認証済みアカウントが 8 個存在する．
4. 次に表示される認証済みアカウント@akimotooo726 はフォロー数が 199 である．
5. @akimotooo726 は@rino_sashihara をフォローしていない．(5 回目)．
3. フォローリストに認証済みアカウントが 7 個存在する．
4. 次に表示される認証済みアカウント@sumire_princess はフォロー数が 381 である．
5. @sumire_princess は@rino_sashihara をフォローしていない．(6 回目)．

よって，提案アルゴリズムに基づき，@rino_sashihara はなりすましであると判定できた．同様に，提案アルゴリズムに基づき，@oshima_yuko_もなりすましであると判定できた．

@tokui_sorangley に対する判定：

1. @tokui_sorangley のフォロー数は 544 である，フォロー数が 1 以上であるため判定可能である．
2. フォローリストを取得する．544 中最新の 200 を取得する．
3. フォローリストに認証済みアカウントが 27 個存在する．
4. 最初に表示される認証済みアカウントである@atsushilonboo のフォロー数は 6 個である．
5. @atsushilonboo は@tokui_sorangley をフォローしていない．(1 回目)．
3. フォローリストに認証済みアカウントが 26 個存在する．
4. 次に表示される認証済みアカウントである@nhk_rhousoku のフォロー数は 10,746 であるため，次の認証済みアカウントを見る．
4. 次に表示される認証済みアカウントである@TOKYOMX のフォロー数は 45 である．
5. @TOKYOMX は@tokui_sorangley をフォローしていない．(2 回目)．
3. フォローリストに認証済みアカウントが 24 個存在する．
4. 次に表示される認証済みアカウントである@iRis_y_saki のフォロー数は 234 である．
5. @iRis_y_saki は@tokui_sorangley をフォローしている．(1 回目)．
3. フォローリストに認証済みアカウントが 23 個存在する．
4. 次に表示される認証済みアカウントである@pripara_PR のフォロー数は 6 である．
5. @pripara_PR は@tokui_sorangley をフォローしていない．(3 回目)．
3. フォローリストに認証済みアカウントが 22 個存在する．
4. 次に表示される認証済みアカウントである@iRis_w_yuki のフォロー数は 283 である．
5. @iRis_w_yuki は@tokui_sorangley をフォローしている．(2 回目)．

よって，提案アルゴリズムに基づき，@tokui_sorangley は本物であると判定できた．同様に，提案アルゴリズムに基づき，@mimori_suzuko も本物であると判定した．

5. 考察

4 の実験結果では，本人であるか怪しいアカウント 2 個と，認証済みでないが本人であるアカウント 2 個を正しく判定できたため，提案するアルゴリズムはある程度の妥当性を持つと考えられる．

しかし，@tokui_sorangley のアルゴリズム適用結果では 5 個の認証済みアカウントのうち，3 個のアカウントが相互フォローしていないことが確認された．今後，サンプルを増やすと，誤判定してしまう可能性がある．提案アルゴリズムでは 6 個以上のアカウントにフォローされていないと判断する場合なりすましであると判断しているが，このパラメーターについては吟味する必要がある．

また，相互フォローしていない 3 個のアカウントについて確認すると@tokui_sorangley の職業である声優ではないことがわかった．2.1 では，類似の職業のものが多く，相互フォローが多いと述べたが，それ以外の職業のものをフォローしている場合については同様といえないためである．また，@TOKYOMX のような企業アカウントは個人をフォローする場合は極めて少ないので除外すると，判定精度が上がると考えられる．これらは，3.2.1 と 3.2.4 の API を用いると同時にプロフィール情報を取得できるため，そこから職業や，企業アカウントなどを判別することができれば判定精度が上がると考えられる．

また，今回は判定したいアカウントと類似の職業ではなかったが，@atsushilonboo はフォロー数が 6 個と非常に少なく，たとえ同じ職業であると確認できたとしても相互フォローしている可能性は低いと考えられる．そのため，フォロー数が極端に少ないアカウントは除外するのが良いと考えられる．また，2.3 のフォロー無しタイプも当然除外すべきである．

提案アルゴリズムでは認証済みアカウントにフォローされていないと本物であっても本物でないと判定されてしまう問題点がある．この問題を解決するため

には認証済みアカウントではなくとも，信頼できるアカウントを特定する必要がある．信頼できるアカウントを特定するには，本人の Web サイトに Twitter アカウントが記載されている．本アルゴリズムで本物であると確認できた．リプライと呼ばれる Twitter 上での会話を元に本人であるかを確認する．などが考えられる．信頼できるアカウントを本アルゴリズムで認証済みアカウントとして処理することで，精度は向上すると考えられる．しかし，信頼できるアカウントを選ぶ際，慎重にならなければ精度は落ちてしまう問題点もある．

6. 関連研究

これまで，Twitter を中心とするマイクロブログに関する研究が数多く発表されている．例えば，Twitter 上のスパムに関する分析[3], Twitter においてハッシュタグがどのように使われているかの分析[6]，重要な Tweet を推薦する方法の提案[12]，Twitter の各ユーザが興味のあるトピックの特定[8]，ハリケーン被害の偽の写真の拡散について分析[4]などが行われている．また，Twitter 上でのアカウント間の繋がりに着目した研究がいくつか存在する[7][9]．ただし Twitter アカウントが著名人本人であるかをフォロワーに基づいて判定するアルゴリズムは，我々の知る限り提案されていない．

カウントが通常のユーザであるかどうかを判定する方法については，いくつか提案または実装されている．例えば，Twitter アカウントのフォロワーが，（ボットなどではない）通常のユーザであるかどうかを確認する Web サービスがいくつか公開されている[2][11]．また，Gurajala ら[5]は，twitter アカウントのプロフィールに基づいて，スパム用のユーザかどうかを判別している．Zangerle ら[13]は Tweet の内容に基づき，SVM によりアカウントがハッキングされたかどうかを判別している．ただし，これらを用いて，あるアカウントが著名人のなりすましかどうかを判定することはできない．

7. おわりに

本研究では，マイクロブログで広く普及している Twitter に着目し，アカウントのなりすましを判定するアルゴリズムを提案した．提案アルゴリズムでは，実際の Twitter API の制限を考慮して，なりすましを判定しているため，実装が容易である．今後は，実際にシステムを構築し，多くの著名人の Twitter アカウントを利用して偽装を判断し，本手法の信頼性を確かめることである．また，現在のアルゴリズムでは判定できないアカウントも多く存在すると考えられるため，Twitter API の制限を考慮してアルゴリズムを改善し，判別精度を高めることも課題のひとつである．

謝辞 本研究の一部は，文部科学省科学研究補助費（挑戦的萌芽：課題番号 26540029）による助成を受けた．

参 考 文 献

[1] 058.jp，twitter 片思いチェッカー，http://kataomoi.058.jp/，参照 Feb.11，2016.

[2] @davc and @grossnasty, Twitter Audit, https://www.twitteraudit.com/，参照 Feb.11，2016.

[3] C. Grier, K. Thomas, V. Paxson, and M. Zhang, “@spam: the underground on 140 characters or less,” In Proc. of the 17th ACM conference on Computer and communications security (CCS '10), pp.27-37, Chicago, USA, Oct. 2010.

[4] A. Gupta, H. Lamba, P. Kumaraguru, and A. Joshi, “Faking Sandy: characterizing and identifying fake images on Twitter during Hurricane Sandy,” In Proc. of the 22nd International Conference on World Wide Web (WWW '13 Companion), pp.729-736, Rio de Janeiro, Brazil, May. 2013.

[5] S. Gurajala, J. S. White, B. Hudson, and J. N. Matthews, “Fake Twitter accounts: profile characteristics obtained using an activity-based pattern detection approach,” In Proc. of the 2015 International Conference on Social Media & Society (SMSociety '15), Article no.9, pp.1-7, Toronto, Canada, Jul. 2015.

[6] J. Huang, K. M. Thornton, and E. N. Efthimiadis, “Conversational tagging in twitter,” In Proc. of the 21st ACM conference on Hypertext and hypermedia (HT '10), pp.173-178, Toronto, Canada, Jun. 2010.

[7] 小出明弘，斉藤和巳，風間一洋，鳥海不二夫，“ネットワーク分析による Twitter ユーザのフォロー形成に関する一考察，” 情報処理学会論文誌 数理モデル化と応用，No.6，Vol.2，pp.164-173 (2013).

[8] M. Michelson and S. A. Macskassy, “Discovering users' topics of interest on twitter: a first look,” In Proc. of the fourth workshop on Analytics for noisy unstructured text data (AND '10), pp.73-80, Toronto, Canada, Oct. 2010.

[9] 西村章宏，土方嘉徳，三輪祥太郎，西田正吾，“一般ユーザの観点に基づく Twitter からの人物関係の可視化と事例の考察，” 情報処理学会論文誌，Vol,56，No.3，pp.972-982 (2015).

[10] OSCARPROMOTION CO., LTD.，剛力彩芽 official web site，http://www.oscarpro.co.jp/talent/goriki/，参照 Feb.11，2016.

[11] StatusPeople.com, Fake Follower Check, http://fakers.statuspeople.com/，参照 Feb.11，2016.

[12] I. Uysal and W. B. Croft, “User oriented tweet ranking: a filtering approach to microblogs,” In Proc. of the 20th ACM international conference on Information and knowledge management (CIKM '11), pp.2261-2264, Glasgow, UK, Oct. 2011.

[13] E. Zangerle and G. Specht, “Sorry, I was hacked": a classification of compromised twitter accounts,” In Proc. of the 29th Annual ACM Symposium on Applied Computing (SAC '14), pp.587-593, Gyeongju, Republic of Korea, Mar. 2014.