RO-005

データマイニングによるプライバシー侵害を防ぐデータベース構築
Database construction which prevents data mining to protect privacy

金森 祥子† 川口 嘉奈子‡ 田中 秀磨††
Sachiko Kanamori Kanako Kawaguchi Hidema Tanaka

1. はじめに

データマイニングによるプライバシーの侵害とは、様々なデータベースから部分的な情報を抽出することで個人情報を復元したり、私生活や個人的趣向を明らかにすることである。最近では、スマートフォンに代表される携帯端末の機能が高度化したことにより、位置情報の利用、カメラ画像、音声や動画など様々な情報をリアルタイムでやり取りできるソーシャル・ネットワーキング・サービス(social networking service,以下 SNS) が充実している。これらの情報はサービス毎のデータベースとして生成され、広く全世界に向けて、もしくは閲覧許可を与えた相手に向けて公開している。このようなサービスは、グルメ情報や写真、ゲームなど特定の分野に特化したものや、Twitter や Facebook のように汎用的なものまで広く存在し、単一のユーザが複数のサービスを同時に利用することも一般的である。SNS の最大の特徴は個人が実名で情報を発信する点である。1990 年代までは個人のホームページなどでプライベートに関わる情報を公開することのないよう啓発活動がなされていたが、このような変遷は、ネットワーク技術に対する社会の対応や関心の変化を顕著に表している。しかしながら予想されていたことではあるが、プライバシー侵害に関するトラブルは頻発している。SNS におけるデータマイニングの典型的な例は、匿名で登録しているサービス X のデータベースと実名で登録しているサービス Y のデータベースから、投稿した時間と場所の一致性を見出し、個人を特定する、また累積した情報から 1 日のタイムスケジュールや利用する駅など、生活習慣の傾向が割り出されることなどである。

データベースの管理者視点としては、このようなデータマイニングに対し、データベースをサービス単位で閉じることにより、情報の取得を限定的にする防止策がある。例えば公的機関が運営しているネットワークサービスの場合、個人情報の漏洩を防ぐためにアクセス制御や、暗号化によるデータ管理など情報セキュリティ技術が多用されている。最近は利便性を高めたプライバシーを考慮した検索及び統計情報の算出[13], [15] や、必要なプライバシー機能を呼び出して使えるソフトウェア(Privacy As A Service) [4] などが盛んに研究されている。また、利便性の意味から疑問視もされるが、サービス毎に独立したアカウントが割り振られたデータベースを構築されている点は、データマイニングによる個人情報の復元やプライバシーの侵害を防ぐ手段とも見なせる。さらに、このような公的ネットワークサービスの場合は扱う情報の範囲が決められているため、違反行為や罰則の設定も行いやすいという特徴もある。一方で、サービス約款については、ユーザが同意した範囲内での変更に留まっていると考えられるが、サービス運営会社の買収や合併などによりデータベースの統合や管理方針の変更が行われる可能性がある。また、データベースに記録された情報はユーザが自発的に発信したものであり、その内容についてはサービス運営者には責務を負うことはできない。このようにデータベースの設定や管理によって、データマイニングによる情報の取得のしやすさは異なるが、SNS のデータベースが橋渡しのような役割を果たすことにより実効が容易になることがある。本論文では SNS によって生成されるデータベースに注目し、性質の異なる複数のデータベースに跨ってデータマイニングを行う問題について考察する。第 2 節では現状の解決策とその限界について述べる。

ユーザ視点のデータマイニングの防止策としては、特定の話題については閲覧者を制限する、サービス毎に話題を限定する、など情報の扱い方をユーザ自身で決定することが挙げられる。これは情報をどのように扱うかという問題であるが、同時にプライバシーに対する扱い方である。本来プライバシーは主観的価値に基づくためユーザ毎に尊重する情報が異なる。そのため「プライバシー」として一般化し定義することが不可能である。さらに同じ情報であっても時、場所、相手によってプライバシーとして扱うか否かが異なることが普通であり、普遍化した扱いが難しい。このようなプライバシーの扱いについて第 3 節で述べる。ここでは能動的プライバシー権と受動的プライバシーを取り上げ、プライバシーを確立するためには両者が成立していることが不可欠であるが、現時点のシステムは能動的プライバシー権のみに立脚していることを示す。

第 4 節において、受動的プライバシーによるデータマイニングの対策とユーザ自身による情報の扱いの仕組みについて考察する。既存の暗号技術である、秘密分散、属性暗号などを用いた階層型アクセス制御技術が、このような要求に対して解決可能な手段であるかの解析を行い、データベース構築のための要件をまとめる。さらに提案プロトコルを示し、安全性評価及び運用的な評価を述べる。第 5 節においてまとめを述べる。

2. データマイニングの問題と現状の解決策の限界

2.1 準備

本論文では以下のエンティティを用いる。

ユーザ:SNS に設定したプライバシー設定に従って自分のアカウント情報を公開し、SNS の内容に即した情報を発信する。発信した情報はデータベースに記録される。

管理者:ネットワークサービスを運営しデータベースを管理する。プライバシーポリシーを策定する。

† 独立行政法人情報通信研究機構, NICT
‡ 千葉大学, Chiba University
†† 防衛大学校, National Defense Academy of Japan

検索者/閲覧者:SNS に公開された情報を閲覧し、データベースを検索する許可を持つ閲覧者。

ここで、プライバシーポリシーとは個人情報の定義、個人情報管理責任者の所在、情報の取得と利用、免責事項などに関するもので、管理者側の運営規定である。一方、プライバシー設定は、ユーザのアカウント情報の何を公開し、どれを非公開とするかの設定である。多くの場合、名前、年齢、生年月日、住所、電話番号、メールアドレスなどであるが、例えば生年月日において年を省略し月日のみ公開する、住所は都道府県までの記述とするなど、公開する範囲まで設定できるのが一般的である。

本論文では、これらエンティティによるシステムに不正侵入する攻撃や、アカウント詐称などの不正行為は想定しないが、特定人物の情報取得には関心があり、許可された範囲でデータマイニングを行うことは制限しない。また、1 度得た情報を再掲し、拡散することで意図的に特定のユーザの情報を広める行為は行わないこととする。本論文で注目する問題は、セキュリティ技術が有効に機能している状態であっても、プライバシー侵害の問題が生じる状況である。

2.2 データベースの種類

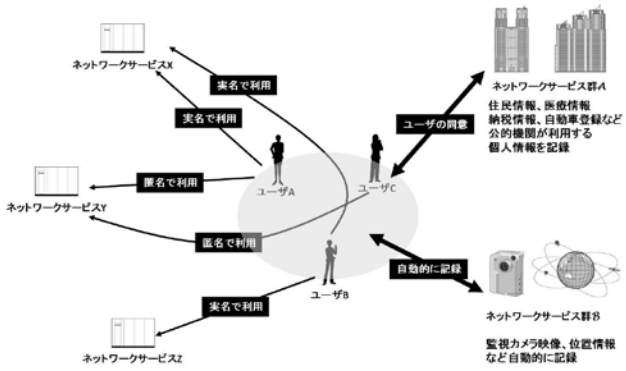

図1　ネットワークサービスの概略

図 1 にユーザから見た各ネットワークサービスへの情報入力の概念を示す。ユーザ A,B,C はそれぞれネットワークサービス X,Y,Z を利用している。各ユーザは各ネットワークサービスに設定したプライバシーポリシーに従って情報を発信し、それぞれのデータベースに情報が記録される。

一方で、A を公的機関の運営するネットワークサービス群とし、住民情報、医療情報、納税情報、自動車登録情報などとする。これらのデータベースはお互いに独立しており、相互のアクセスは制限されていると仮定する。ネットワークサービスA は、扱う情報が決められており、その入力範囲も決定されている。さらにユーザが閲覧できるのは自分の情報のみであり、自由に検索を実行することはできないが、管理者は検索が許可されている。

また、Bを M2M(Machine-to-Management) によって生成されるデータベース群とする。これはネットワークに接続された監視カメラなどのセンサーや機械同士がオペレータである人間などの介在無しに自動的に相互通信を行った結果生成される情報を記録している。

利用例として、自動販売機の在庫管理や電気・水道・ガスの検針などが挙げられる。データベースへの検索の権限は管理者のみにある。ユーザは自分の情報がどのように記録されているか否かを明示的に知ることができない。自分の情報を知る場合、もしくは情報が公開される場合は法的手続きなど何らかの手続きを必要とする。B もAと同様にお互いのデータベースは独立しており、相互のアクセスは制限されている。

このように、ユーザが自発的に発信して生成されるデータベース、行政サービスなど業務の効率化を目的としてユーザの情報が入力されたデータベース、M2M などにより自動的に生成されるデータベースの 3 種類が想定できる。

2.3 データマイニングの問題

ネットワークとデータベースの拡充により注目されているデータマイニングの問題とは、部分的な情報を様々なデータベースから抽出することで個人情報を復元し、私生活や個人的趣向を明らかにすることである。典型的なものは、ネットワークサービス X,Y,Z を検索することで実行できる。例えばユーザ A は X では実名登録、Y では匿名利用をしていたとすると、データベース X とデータベース Y の共通情報から、匿名利用の Y における身元が A であることが閲覧者に知られることがある。また、サービス Y を匿名利用していたユーザ C の情報を、友人であるユーザ B がサービス X で実名発信し、ユーザ C の身元が明らかになるケースもある。さらに発言だけではなく、投稿したデジタルカメラの写真データに位置情報が記載されていて住所が特定できることもある。例えば、ユーザ C の自宅を訪問したユーザ A が、リビングでの食事メニューをデジタルカメラデータで公開することで、ユーザ C の自宅が突き止められるなどの事例も知られている。

さらに異なる内容のデータを紐づけることで新たな意味を生成する問題もある。顕著な例はゲイビソンによって示された司祭の例であるが[3] (注 1)、複数のユーザの不特定多数に向けた発言から特定のユーザの情報を浮かび上がらせることも広い意味ではデータマイニングの問題と言える。例えば、繁華街にてグルメ情報を匿名でネットワークサービス Z で発信したユーザ B の姿が店前の監視カメラによりデータベースB に記録された場合である。これらは独立した行為であるはずだが、両方のデータベースにアクセスが許される検索者には匿名のはずのユーザ B の容姿が得られることになる。

検索が許されるのは管理者だけであるが、データベースA 及びB と SNS の情報を利用することでユーザのプライバシーを詳細に知ることができる。例えばユーザ A が公開している自家用車のデジタルカメラの写真データから車両番号を読み取り、それを元にデータベース A から自宅住所を割り出すことや、SNS に公開された自宅住所と電気の使用量の時間変化を記録しているデータベース B から留守になりやすい時間帯を明らかにすることなどである。実際に電車の自動改札を通過した情報から特定の

（注 1）：「私が司祭として初めて受けた告白は殺人に関するものだった」と「私はあの司祭に告白した最初の人間だ」という発言があったとする。それぞれは独立した情報であるが、それぞれが組み合わされることによって、全く新しい情報に転化する例。

人物を追尾したり[17]、企業に応募してきた就職活動のエントリーシートと SNS のデータベースのデータマイニングを行い、素行調査を行った例などの事例が報告されている[5]。

図 2 にデータマイニングによって、個人の情報が集積された結果構築されたデータベースの概念図を示す。中心がプライバシーであり、周辺に行くにしたがって個人情報、公開可能な情報へとプライバシーの範囲から離れる。また、実線で囲まれた範囲は暗号プロトコル等で保護された情報であり、許可された者のみ検索/閲覧できる。

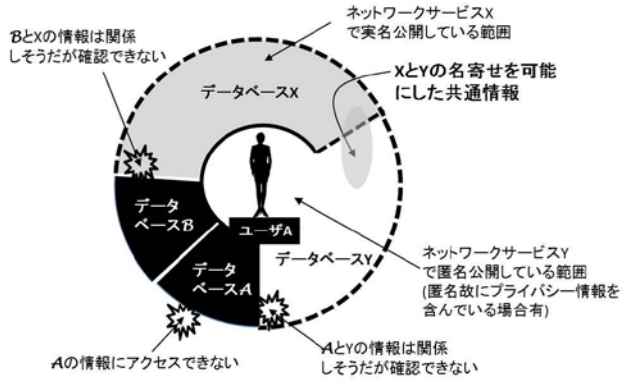

図 2 データマイニングによって生成されるデータベース

破線はだれでも検索を実行できる領域である。例えばユーザ A の場合、サービス X,Y は破線であるので、誰でも検索が可能である。サービス X と Y の境界も破線であり、相互にデータマイニングが実行可能であるため、サービス Y からサービス X を通じて個人名の特定が可能であり、個人情報へ到達することが可能である。また、サービス X,Y からの情報セキュリティ対策が有効なため、データベース *A* 及び *B* の情報へたどり着くことはできない。

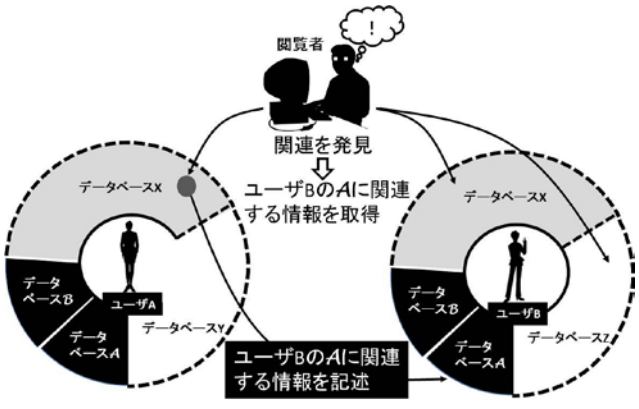

図 3　データマイニングによる情報の流出

しかしながら、図 3 のようにユーザ A がサービス X で公開した情報からユーザ B のデータベース *A* の情報へ間接的にアクセスしている状態を作ることができる(前述の例と同様に、B の自家用車写真を A が公開することで B の住所や納税情報を公開したことと同等のような状況)。これは情報セキュリティ対策が機能していても、プライバシー侵害の問題が生じる状況の一例である。このように、既存のセキュリティ技術は、データベースへの不正アクセスや改竄など悪意ある行為を防ぐことはできるが、自由な発言を許す SNS のデータベースを組み合わせた場合、実質的なプライバシー情報の入手や推測を許すため、プライバシー侵害への対策としては機能していないと結論できる。

2.4 現状の解決策

前節で示したデータマイニングによる情報取得に対する対策は、現状では技術的解決ではなく法律的解決に依存している。特にデータベース *A* 及び *B* と SNS を組み合わせたデータマイニングは、管理者が個人情報保護法に違反したケースであり罰則が適用される。しかしながら複数の SNS のデータベースからデータマイニングした事例に対しては、個人情報保護法の適用は困難である。現在、日本で定められている法律としては、10 年前に制定された個人情報保護法が存在するが、プライバシーに関しては法律では明文化されていない。日本国内の法解釈においては、プライバシーは日本国憲法第 3 章第 13 条で解釈されるのが現状である。

最近ではプライバシーを尊重した判例も見られるようになってきたが、プライバシー侵害は、名誉棄損や機会損失など何らかの被害を訴えることで解決されるケースが多い。ここから法律で罰則を定めるには、違反に相当する情報の種類を定めることが必要であることがわかるが、そもそもプライバシーを定義することは困難である。一方、個人情報は、氏名、性別、生年月日、住所、住民票コード、勤務場所、職業、収入、家族構成、身長体重、指紋などの生体情報など、と個人を特定する情報として定義されている[16]。プライバシーは思想信条や宗教観なども含まれる主観的価値を重視したものであるため、同じ内容であっても人によってプライバシーと考えるか否かは異なる。そのため前掲の個人情報であっても、プライバシーを主張する場合と自ら公開する場合とがあり、自ら公開した場合はプライバシーとはならない。このようにプライバシーと個人情報の関係は、プライバシーの一部を個人情報として定義し扱い方について法律的に明示したという関係であることが分かる。

3. プライバシーの分類とシステム構築の視点

3.1 プライバシーの概要

一般にプライバシーは私生活上の事柄をみだりに公開されたり詮索されないことであり「プライバシー権」として権利として主張されたり、主観的価値観であり尊重されるべきものであるが権利ではないと解釈されるなど、様々な定義がある[2], [7]。プライバシーの概念は 1890 年代にウォーレンとブランダイスにより主張され、人間が独立した自律的行動を行うためには、一定の私的領域確保が必要不可欠であるとされた[9]。また、一度ネットワークに流出した情報は時場所の制限が無く、過去の情報と現在の情報に意味としての区別は無い。そのため「能動的プライバシー権(積極的プライバシー権)」として、他

者が管理する自己の情報について、どの情報を公開・非公開とするかを決定できる権利が認められるものへ変化してきている[10], [11], [14]。

一方で前述の「私生活上の事柄をみだりに公開されたり詮索されないこと」の意味でのプライバシーは古典的プライバシーと呼ばれ区別されている。古典的プライバシーは他者の価値観への尊重が根本にあり、周囲全体としての協力体制を必要とする。そのため、特にネットワークにおいてはユーザ個人で実現することが難しい。本論文ではこのような性質から「受動的プライバシー」と呼ぶ。現時点では明示的に受動的プライバシーが実現されているネットワークサービスは存在しないが、そもそも自律的行動のためには必要不可欠とされていることから本来は提供されるべきものである[1], [7]。特に SNS においては、ネットワークを介して新たな人間関係の構築を期待するという自律的行動を利用動機に挙げることが多い。ここで他人との接触において、ある程度の個人情報やプライベートな話題が明らかにされることは自然であり、このような話題が可能となる場には受動的プライバシーが必要である。

3.2 既存 SNS におけるプライバシーの考え方と限界

自分の情報を制御できるとするプライバシーは、制御できる対象の情報が明示的であることが前提なので権利として定義可能である。従って能動的プライバシー権として扱うことができ、権利として定義されれば遵守義務が発生するので罰則規定を設けることができ、法的解決との親和性が高い。つまり、第 2.4 節で示したように個人情報保護法の制定の基本的な考え方と合致している。このように、多くの SNS におけるプライバシーポリシーは能動的プライバシー権の視点で策定されている。これはユーザと管理者の双方にとって都合の良い設定である。すなわちユーザは能動的プライバシー権を行使し自分の情報管理や公開の範囲の設定を可能にし、管理者にとっては情報の管理範囲が制限されることでユーザの権利として遵守しやすい環境の構築となっているからである。

しかしながら能動的プライバシー権では、情報の意味(含み,implicature) までは制御できない。第 2.3 節で示したデータマイニングによるプライバシー侵害の例では、個々の情報単体ではそれが示す事象としての情報以外の意味は生成されないが、他の情報と組み合わせることで別の意味が生成される。これは特に M2M によって生成されるデータベース B において顕著と言える。機械は能動的プライバシー権で設定された通りに動作するのみで、取得したデータの意味までは判断できない。また、ユーザの能動的プライバシーは運営会社の方針変更や新機能追加の影響も受ける。例えばネットショッピングの購入履歴が SNS と連携して、SNS からユーザ趣向を紹介できるような機能追加である。この場合、プライバシーポリシーに購入履歴が加わり、機能としてショッピングサイトとの連携が加わったことを意味する。このようなプライバシーポリシー設定項目の変更や新技術がもたらすサービスの性質の変化への対応など、ユーザに要求するリテラシーが年々高度化しているのは良く知られたことである。能動的プライバシー権の設定に必要なユーザのスキルやリテラシーが高度になるにつれて、かえって権利の行使を難しくするという状況にある。さらに、能動的プライバシー権の設定に求められるユーザに高度なスキルやリテラシーは、ユーザ本人だけではなく、関係する人全てのスキルやリテラシーが高いことも求められる。

このように、一見すると SNS は能動的プライバシー権の視点からはユーザの自律的行動を保障しているかのように見えるが、実際は管理者がユーザの権利の遵守を容易に達成できる仕組みであって、本質的にユーザのプライバシーを保護するシステムやデータベース構築になっていないことが分かる。

3.3 受動的プライバシーを実現する要件

受動的プライバシーは第 3.1 節で示したように「私生活上の事柄をみだりに公開されたり詮索されないこと」を実現することであり、能動的プライバシー権のようにどの情報を対象とするかを設定することができない。これは受動的プライバシーが、その情報を受け取った相手の解釈に依存した、意味に関連した性質を持つためである。また能動的プライバシー権が、事前に公開する情報を選択するという性質であるのに対し、受動的プライバシーは既に起こった事に対する他者の解釈へ期待する性質であるという違いもある。ユーザが SNS で情報を公開する状況を例として挙げる。内容、表現、公開範囲の設定という事前の行為については能動的プライバシー権によってユーザ自身が決定可能である。しかし、その情報がどのように入手され解釈されるかは閲覧者次第である。従ってユーザは、情報を公開した後は他者の解釈を期待する以上のことはできない。逆に言えば、ユーザは事後の解釈や情報の扱いを期待して、能動的プライバシー権の行使たる自律的行動を可能にしていると言える(注 2)。このように受動的プライバシーを権利として定義することが不適切であり、能動的プライバシー権の視点からのシステム構築とは区別して考える必要がある。一方で、データマイニングとは部分的な情報を様々なデータベースから抽出することで意味を生成する行為であるから、このような行為を防ぐ観点から能動的プライバシー権よりも受動的プライバシーの視点が適切である。さらに「詮索されないこと」は、データベースにおいては検索を制限することと等価と言える。注意しなければならないのは、ユーザは自分の情報を検索されず秘匿した状態に置きたいのではなく、対象の相手に対してはある程度のプライベートな情報も伝えたい意思がある点である。また、場合によっては、プライベートな情報を全公開したいこともあり得る。現時点の SNS の利用状況を考えると、特定の相手とのプライベートなやり取りを設定した利用目的と全公開を目的とした場合で、異なるアカウントを利用する、各々で独立した別個のサービスを利用するなどしているのが一般的であるが、これは第 2.3 節で示したように根本的な解決ではなくデータマイニングによって侵害されるケースもある。従って、SNS でのデータベースの構築そのものにおいて対策をしなければならない。

（注 2）：これも SNS の例で言えば、ユーザは閲覧者からの反応を予想しており、これまでその予想にある程度合致してきたという理由によって、これから発信する情報の扱いを期待している。そのため、SNS で情報を発信するという自律的行動を促す動機となっている。

以上より要件としては、検索者(閲覧者) に対して得られる情報の範囲を設定できることであり、許可されていない検索者には情報が得られないことである。これは階層型アクセス制御を用いることで実現可能と考えられる。階層型アクセス制御において、閲覧者に対してアクセス範囲を設定できることはユーザにとって事前の選択であり、能動的プライバシー権の行使と言える。このように閲覧者への設定がされている状況で、許可された範囲を超えたアクセスを可能にするためには、正規の権限を取得するためのユーザに対する交渉としてのソーシャルコストか、暗号解読などの不正に取得するための計算量的コストを支払う必要が生じる。このようなコストを理由に階層型アクセス制御は、受動的プライバシーを実現していると見なせる(注 3)。

しかしながらデータベースには情報が記録されているため、SNS 運営方針の変更や新機能の追加により、常にユーザの設定の思惑通りになるとは言えない。第 3.2 節で挙げたネットショッピングの購入履歴と SNS の連携のように、SNS の発言や振る舞いと異なる趣向がネットショッピングの購入履歴から明らかになったり、動画サイトのお気に入り情報が SNS と連携することで想定していなかった情報を公開されるようなトラブルが現実に生じている。そこで、SNS のデータベースの統合やサービスの変更、新機能追加に影響されないことも要件として挙げられるため、単純に階層型アクセス制御を導入するだけでは受動的プライバシーを実現できない。

受動的プライバシーを実現する SNS 及びそのデータベースに対する要件を以下にまとめる。

要件 1:ユーザは検索者(閲覧者) に対して得られる情報の範囲を設定できる

要件 2:検索者(閲覧者) は許可された範囲外の情報を得られない

要件 3:データベースの統合やサービスの変更、新機能追加に影響されない

要件 1 はユーザの能動的プライバシー権であり、要件 2 はそれに必要な受動的プライバシーである。両者は相補的関係にある。しかしながらデータマイニング防止という意味では要件 2 がより重要である。要件 3 は受動的プライバシーである。従って、次節ではこれら要件を満足するデータベース構築について考察する。

4. 受動的プライバシーに基づくデータベース構築

要件 1 と 2 に注目すると、許可された者が情報を復元できるデータベースの構築が解決策として考えられる。例えば SNS が複数のデータベースサーバを管理していると仮定し、それらにデータを分散させる。正規の閲覧者は情報を復元するためのデータが記録されているデータベースサーバ群を知っているが、許可されていない閲覧者はそれらを知らないため情報が得られない、という仕組みである。別の解決策としては、情報を暗号化してデータベースに登録し、許可を与えた者に復号鍵を配布する方法である。これは一般的な解決と考えられるが、特定のユーザのみが情報を得られるので SNS の用途にはそのままでは適さないと考えられる。そこで属性暗号を用いて、ある属性を持つ者は設定された範囲で情報を得られる仕組みを導入する。前者は秘密分散を用いた手法、後者は属性暗号を用いた手法として考察する。

4.1 秘密分散による階層型アクセス制御と信頼できる第三者機関

秘密分散とは情報を n 個に分散し、そのうち特定の数以上のデータを集めないと情報が復元できない手法である。代表的なものは(k, n) 閾値型秘密分散であり、k − 1 個以下のデータからは情報が復元できないものである[8]。SNS のデータベースへの適用として考えると、N(≫n) のデータベースサーバを用意し、ユーザはその中から n 個のサーバを選び出しデータを記録する。閲覧を許可したものにはそのうち k 個のサーバを伝え、閲覧するにはそれらのサーバにアクセスし情報を復元するというものである。この手法には、ユーザが n − k 個以上のデータを削除することで情報が復元されることを防げるという利点もある。また(k,L,n) ランプ型秘密分散法[12] を応用することで、k 個以上のデータで情報が完全に復元できるが (k–l)(ただし 1≦l≦L−1) 個のデータからは部分的な情報が復元できるように設定することで、親しい友人には完全な情報を、他の友人には一部分については公開しても良いという利用も可能である。

しかしながら、ユーザ自身が秘密分散処理を行い、閲覧者の設定に応じて分散したデータベースサーバの所在を通知することは SNS の利用から考えて現実的ではない。そのため、この手法には秘密分散を行うディーラーが必要であり、閲覧者とデータベースを SNS サービスを通じて連結する必要がある。このディーラーは信頼できる第三者機関(Trusted Third Party, TTP) として機能しなければならない。

(k, L, n) ランプ型秘密分散と TTP を用いた手法を手法 1 として以下にまとめる。

準備: SNS はデータベースサーバを N 個用意する。ユーザは閲覧者の設定を行う。情報の完全な復元を許可する場合は k 個のサーバの通知、一部公開情報の場合は (k–l) 個のサーバの通知とする。ただし N ≫n, k≦n, 1≦ l≦ L − 1。

Step-1: ユーザは情報を TTP に送信する。

Step-2: TTP は n 個に情報を分散する。ランダムに n 個のデータベースサーバを選び、データを記録する。

Step-3: TTP は SNS にアクセスし、ユーザの閲覧者設定を得る。完全な復元を許可する場合はデータを記録した n 個のサーバから、k 個のサーバをランダムに選んで復元情報を作成する。同様に一部公開情報の場合もランダムに (k − l) 個のサーバを選んで復元情報を作成する。

Step-4:TTP は SNS に各閲覧者用の復元情報を送信する。

Step-5:SNS は各閲覧者に復元情報を送信し、復元情報を得た閲覧者は情報を閲覧者設定に応じた情報を得る。

図 4 に手法 1 の概略図を示す。

（注 3）：現実社会に置き換えると、ある人の情報を知るためには知り合いになり直接尋ねるか、(犯罪であるが) ストーキング行為により強引に知る方法がある。どちらも対人コミュニケーションというコストか、体力・時間というコストを支払っている。このように、受動的プライバシーを超えて情報を得るためには、コストを支払う必要が生じる。

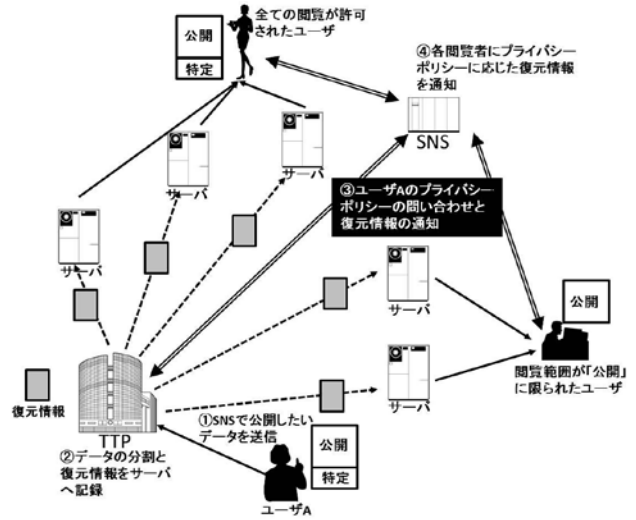

図 4　手法 1 の概要

4.2 属性暗号を用いた階層型アクセス制御

属性暗号は公開鍵暗号の一種であり、さまざまなパラメータを暗号文もしくは復号鍵に導入することで、属性情報を与え、それに対応する条件式で復号できる権限を制限し、復号できる範囲を設定できるものである[6]。復号鍵に属性情報を与え暗号文に条件式を与えた場合、ある属性を持つ閲覧者しか暗号文を復号できないため、アクセス制御として利用できる。逆に暗号文に属性情報を与え復号鍵に条件式を与えた場合、閲覧者は自分に与えられた範囲でしか復号できない。前者を手法 2、後者を手法 3 と呼ぶ。

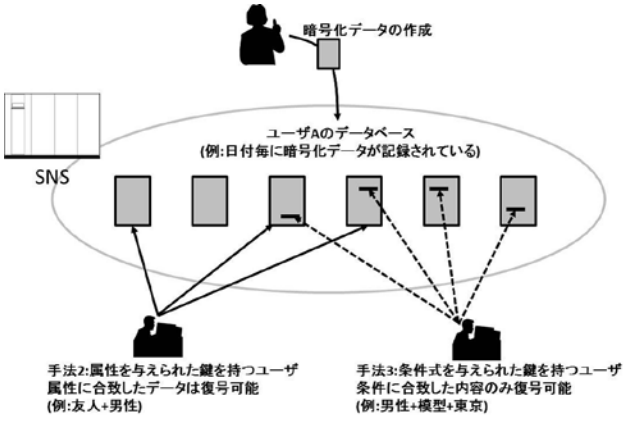

図 5　手法 2 及び 3 の概要

手法 2 の場合、ユーザは閲覧者の属性に応じた復号鍵を生成する。属性として例えば性別、親族、友人、趣味、居住地域などを設定し、それに応じて閲覧者に復号鍵を配布する。属性は複数与えることができるので、「性別＋趣味」や「性別＋友人＋居住地域」などの設定も可能である。ユーザが発信する情報の属性に応じて暗号化を行い SNS 上で発信し、閲覧者は既に配布されている復号鍵で復号する。例えば「女性+コミック+名古屋」という属性の情報をユーザが生成した場合、これらの属性全てに該当する閲覧者は情報を全て復号できるが、「女性+名古屋」の属性しか持たない場合は、一部分の情報しか復号できない。また「男性」の属性を持つ閲覧者は、復号が許可されていないため、情報は全く得られない。

手法 3 の場合は、ユーザは閲覧者の条件式に応じた復号鍵を生成する。条件式として例えば性別、価格、趣味、地域の組み合わせを与えるとする。条件式として「趣味 AND (価格 OR 地域)」を閲覧者に復号鍵として与えたとすると、ユーザが発信した情報のうち閲覧者が復号できる情報としては、例えば「アニメという趣味でかつ、値段が 3000 円または東京に関連する内容」のような情報となる。条件が緩い鍵を与えられた閲覧者は全ての情報を復号可能であり、条件を細かく設定された閲覧者は内容が制限される。これは現在の SNS の閲覧者設定に近い考え方である。

手法 2 及び 3 は閲覧者の増減や条件の追加に対しては、復号鍵の再発行を行う必要があり、それに伴い過去のデータに対する処理が課題である。また、ユーザは復号鍵の不正利用を知る術がないため、閲覧者の復号鍵管理を信頼しなければならない。図 5 に手法 2 と手法 3 の概略を示す。

4.3 受動的プライバシーから見た評価

手法 1,2 及び 3 に関して第 3.3 節で示した要件に基づいて受動的プライバシーの視点から考察を行う。

全ての手法において閲覧者が設定されているため、権限を持たない検索者には情報が与えられず、また、閲覧者に対しても得られる情報が制限されている。従って、全ての手法において、要件 1 及び要件 2 は満足されていると結論できる。さらに、許可された閲覧者であってもデータマイニングによって集めることのできる部分情報は制限されることも可能であるが、個人情報の復元に対して既存の SNS やデータベースと比較してコストは上昇している。また、閲覧のための許可を得るためには、ユーザと直接交渉しなければならない。特に手法 2 及び 3 においては新たな復号鍵の発行を必要とするため、検索者には高いコストが必要となる。

要件 1 を拡大し、閲覧者が情報へのアクセスを禁じることも含めるならば、第 4.1 節で示したように手法 1 は分散データを削除することで、この要求を満たす。既存の SNS は情報の消去を指定しても確実に消去されたか否かをユーザが確認できる方法がないため、手法 1 は既存の能動的プライバシー権よりも優れている。一方、手法 2 及び 3 は、第 4.2 節に示したように閲覧者の鍵管理を信用するしかなく、情報の消去も既存の SNS と同等である。

SNS 側のサービスの変更、新機能追加に対しては、全ての手法が公開を前提とした情報以外は暗号化処理が施されているため影響を受けないことは明らかである。従ってユーザが気づかないうちに公開されていたというトラブルや、新機能に付随して過去の情報が公開されるということは生じない(注 4)。しかしながら、過去の情報に対しても SNS の新機能を追加したいという要求に対しては課題が大きい。このような要求に対しては、提案手法

(注 4)：ここでは第 2.1 節に示したように各エンティティの悪意ある行為は仮定していないことに注意。悪意ある行為に対する対策は今後の課題である。

ではデータベースを作り直すしかない。また、手法 2 と 3 は暗号化データが単体として存在しているため、データベースの統合の影響を受けないが、手法 1 の場合は統合によって分散数が減少する問題も発生し得る。このため、要件 3 に対しては部分的な実現に留まる。

ところで全ての手法において、閲覧者同士が結託した場合、もしくは閲覧者が有する情報が漏洩した場合はユーザの受動的プライバシーは侵害される。特に手法 2 及び 3 における復号鍵の漏洩は致命的である。本論文では悪意ある行為を仮定していないが、解決すべき問題である。

実現性に対しては、全ての手法が計算コストとデータ量が増加するため、利便性は大きく損なわれることとなる。特に手法 1 の場合は、データベースサーバの管理数が N 倍になるため、サービス運営コストは単純に N 倍程度増加と思われる。さらに信頼できる第三者機関(TTP) の設置と運営が必要となるが、これは次節で述べる。また閲覧者は複数のサーバからデータを受信するため、現在のようなスマートフォンなど携帯デバイスでの用途においては通信環境の問題も大きい。手法 2 と 3 は新たに公開鍵基盤の導入を必要とし、閲覧者に対して復号の処理を負わせるため手法 1 と同様にスマートフォンなど携帯デバイスでの用途においては高い計算機能力を要求することとなる。しかしながらこれらの欠点は、受動的プライバシーを実現するために計算コスト及びソーシャルコストに基づいた手法を採用したことからトレードオフの関係になっている。逆に言えば、利便性を高め、低コストで検索が可能な現在の SNS は結果的にデータマイニングによるプライバシー侵害を許す環境を提供していると言える。

4.4 受動的プライバシーと信頼できる第三者機関の設定

手法 1 では秘密分散を行うディーラーを設定している。SNS 運営側をディーラーとしないのは、要件 3 を満足するためであるが、ディーラーはユーザからの情報を扱うため、ユーザのプライバシー情報を知る立場にある。また、分散したデータの記録先であるデータベースサーバの情報を持つ。従って、手法 1 によるプライバシー保護はディーラーがその要であるから、SNS 運営側の統廃合に影響されないために本論文では信頼できる第三者機関(TTP) とした。

しかしながら、プライバシーの扱いを第三者機関に委ねることの妥当性の議論が新たに必要となる。欧米にはプライバシー問題を裁定する独立の委員会が組織されていることがある。このような中立の立場でディーラーを運営することが望ましいと考えられるが、扱う問題の性質が委員会とディーラーでは異なる。委員会ではある事例がプライバシー侵害と認められるかどうかの判断が中心であるのに対し、ディーラーは生のユーザの情報を扱う。この点では、電子投票におけるオニオンルーティングと同様な運営姿勢が求められる。しかしながら電子投票は守るべき情報が定義できるのに対し、プライバシーでは主観的価値観に基づくものであり、保護する範囲が決定できないことは第 3.3 節で示した通りである。さらに前述のように TTP とは言えユーザの情報を全て知る立場にありながら、ユーザの受動的プライバシーを実現するために全情報を保護しつつ、かつユーザの能動的プライバシー権を尊重して、ユーザが自律的に情報公開を可能にする必要がある。

このように手法 1 には技術的解決だけでなく運営的な問題と、受動的プライバシーにおける TTP の是非に関する議論が必要となる。しかしながら、この問題は手法 1 と手法 2 の組み合わせ、もしくは手法 1 と手法 3 の組み合わせで解決できる。しかし、暗号化されたデータを秘密分散する解決方法は、前節で示したように通信量だけでなく計算コストがさらに増加する。また、要件 3 に対しては、手法 1 と同様にデータベースの統合の影響を受ける場合がある。これらは今後の課題である。

5. まとめと今後の課題

本論文では SNS のデータベースを利用したデータマイニングによるプライバシー侵害について考察し、その対策手法について提案した。プライバシーは能動的プライバシー権と受動的プライバシーに分類され、既存手法が能動的プライバシー権に基づいて構築され、法的解決との親和性が高い理由を示した。一方で本来プライバシーは受動的プライバシーが確立された上で能動的プライバシー権を行使できるので、本来であれば能動的プライバシーの前提となるべき、受動的プライバシーを実現するための要件を 3 つ導出した。これを実現するため、秘密分散と属性暗号に注目し 3 種類の手法を提案し、それぞれが受動的プライバシーの要件を満足するか評価した。現時点では受動的プライバシーから見た信頼できる第三者機関(TTP) の妥当性に関する判断が未成熟であるが、3 種類の手法は現実的運用において十分受動的プライバシーを実現できると考えられる。TTP を設置しない手法についても提案手法を述べたが、計算コストと通信量の増加が現実的な SNS の運用を可能にするかの評価が必要である。本論文では、各エンティティは悪意ある行為をしないと仮定したが、悪意ある行為に対する対策及びその評価も重要な課題である。これは今後の課題である。

プライバシー侵害の問題は、ユーザが被害を認識して初めて成立する。被害を自覚してからでは根本的解決は難しい問題であるから、受動的プライバシーを提供するデータベースを構築し、予防的な仕組みを実現することは重要性が増すと考えられる。

謝辞

本論文作成にあたり、ご協力いただいた財団法人未来工学研究所笠井祥氏に感謝いたします。

参考文献

[1]E.J.Blowstein, “Privacy as an Aspect of human Dignity An Answer to Dean Prosser,” New York University Law Review 39, pp.962-1007, 1964.
[2]C.Fried, “Privacy [a moral analysis],” Philosopical Dimensions of Privacy: An Anthology, Cambridge University Press, 1984.
[3]R.Gaivison,“Privacy and the Limits of Laws”,Yale Law Journal,vol.89,no 3,1980.
[4]W. Itani, A. Kayssi, and A. Chehab, “Privacy as a Service: Privacy-Aware Data Storage and Processingin Cloud Computing Architectures,” 2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, pp.711-716, 2009.

[5]D. Lyon “Surveillance Society: Monitoring Everyday Life,” Open University Press, 2001.
[6]T.Okamoto, K.Takashima, “Adaptively Attribute-Hiding (Hierarchical) Inner Product Encryption,” Advances in Cryptology - EUROCRYPT 2012-,Lecture Notes in Computer Science 7327, pp.591-608 (2012)
[7]J. Rachels, “Why privacy is important,” Philosophy & Public Affairs 4(4), 1975. (rpt.) F. D. Schoeman (ed.), Philosophical Dimensions of Privacy, Cambridge University Press, pp.291-299, 1984
[8]A. Shamir, “How to share a secret,” Communications of the ACM, Vol. 22, Issue 11, pp. 612-613, 1979.
[9]S. D. Warren and L. D. Brandeis, “The Right to Privacy,” Harvard Law Review, Vol. Ⅳ, No.5, 1890.
[10] A.F.Westin, “Privacy and Freedom”, New York:Atheneum,1967.
[11] A.F.Westin, “The origins of modern claims to privacy”, Philosopical Dimensions of Privacy : An Anthology, Cambridge University Press, 1984.
[12] H.Yamamoto, “On Secret Sharing System Using (k,L,n) Threshold Scheme,” Electronics and Communications in Japan, Part I, vol.69, no.9, pp.46-54,1986.
[13] 伊藤孝一、牛田芽生恵、山岡裕司、小櫻文彦、津田宏、属性値に紐づく数値のプライバシ保護集計方式」, 暗号と情報セキュリティシンポジウム(SCIS2012) 予稿 CDROM,3D2-3, 2012
[14] 辻井重男、山口浩、五太子政史、只木孝太郎、藤田亮、井堀幹夫、土居範久、「電子行政・医療介護ネットワークにおける個人情報の保護と活用の両立のための情報連携システム-第 2 報」、暗号と情報セキュリティシンポジウム(SCIS2012) 予稿 CDROM 2D1-4, 2012
[15] 安田雅哉、矢嶋純、下山武司、小暮淳、「複数企業が持つ購買履歴データのクラウド秘匿集計」, 暗号と情報セキュリティシンポジウム(SCIS2012) 予稿 CDROM 3D2-5,2012
[16] 財団法人日本規格協会「JIS Q 15001(個人情報保護マネジメントシステム-要求事項)」
[17] 日本経済新聞　2012 年 4 月 17 日記事
http://www.nikkei.com/article/DGXNASDG1700Z_X10C12A4CC0000