社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2006－CSEC－33 (3)
2006／5／12

IDのリサイクルによる匿名化の提案

川崎 明彦† 佐藤 嘉則† 森田 豊久† 森藤 元†

(株)日立製作所 システム開発研究所†

個人情報保護法の施行に伴って，プライバシ保護の社会的な関心が高まっており，安心・安全を技術的に確保するシステムが求められるようになってきている．一方，今後も続くであろうIT技術の進歩は，個人に関して収集・蓄積される情報の種類や量を増加させ，悪意を持った人物が興味本位で覗いたり，本来とは別の目的で利用する機会の増加を助長する懸念がある．本稿では個人と情報の関連性に焦点を当て，新たなプライバシ保護技術としてリサイクルIDを提案する．リサイクルIDは匿名化手法の一種であり，その基本コンセプトはIDの交換である．個人が持つIDを他の個人が持つIDと随時交換して使用することにより，IDからの個人特定を困難にする．本稿ではリサイクルIDの基本コンセプト，適用例，評価実験について述べる．

The ID Recycle Management Technology for Anonymizing Service

Akihiko Kawasaki, Yoshinori Sato, Toyohisa Morita, Hajime Morito
Systems Development Laboratory, Hitachi Ltd.

The Personal Information Protection Act represents a big concern about privacy issues in our society. Because the quality and quantity of personal information will increase as progress of information technology, potential risk of privacy violation will increase. In this paper, we focus linkability between information and data subject as an important threat on privacy and propose a newly developed privacy enhancing technology, which is called "ID Recycled Management Technology". Our technology bases on a concept that users exchange their IDs each other to protect from behavior tracking by using ID. This paper describes the basic concept, application, and evaluation of recycle ID.

1. はじめに

個人情報保護法の施行に伴い，プライバシ保護の社会的な関心が高まっており，安心・安全を技術的に確保するシステムが求められるようになっている．

そうした要求に対して現状のシステムは，アクセス制御や暗号などのセキュリティ技術を用いて個人情報の機密性を高める方向に進みつつあるが，多くのビジネスでは個人情報を必要としているため，隠ぺいによってシステムの利便性が損われると，業務自体が成り立たなくなる恐れがある．

そこで我々は，プライバシ特有の脅威として個人と情報との関連性(リンカビリティ：linkability)に着目し，そのつながりを適切に隠ぺいすることで，プライバシ保護と個人情報の利用を両立させる技術の開発に取り組んでいる．

本稿では，IDによるリンカビリティに焦点を当て，企業などの組織が個人に対して割り当てたIDを介し，その個人に関連する様々な情報がつながる状況をプライバシの脅威として想定する．

現在の企業システムにおいて，顧客に割り当てられる管理用のIDは，いったん割り当てた後は変更されることのない固定IDであることが多い．このような固定IDが与えうる脅威としては，例えば，顧客がサービスを利用した日時や場所が蓄積されている履歴を用いて，IDから同一人の行動を不正に追跡する行為等が該当する．

このような背景を鑑み，本稿では新たなID匿名化手法として，我々はIDを複数のユーザで共有，再利用する，リサイクルIDシステムを提案する．

以降，第2章ではリサイクルIDの基本原理について述べ，第3章では適用例としてIPアドレスの匿名化について述べ，第4章以降ではプロトタイプ及びプロトタイプを用いた評価実験について報告する．

2. 従来手法

上記のプライバシ上の問題に対し，従来からIDを匿名化する手法が考えられてきた．ただし本稿における匿名とは，同一人物であることを特定できないデータを表し，典型的にはIDによる同一人の特定不可能性を意味する．

インターネットに接続されるネットワーク端末において使用されるIPアドレスをIDとみなすと，ID匿名化の従来手法としては，個人とサービス提供者の間に匿名プロキシを介在させる手法が挙げられる[1][2]．この方法では，個人からの要求を匿名プロキシが代わってサービス提供者に伝えることで，サービス提供者に対して個人が使用しているIPアドレスを隠ぺいする．しかしながら，送受信データが全て匿名プロキシを通過するため，匿名プロキシの信頼性が問題になる．

その他の匿名化手法には，使い捨てのIDを発行して利用するワンタイムIDがある[3]．この方法では，暗号等の手法を用いてユニークなIDを発行し，それを利用することでデータ送信時にIDからの同一人物が特定されることを困難にする．しかしながら，同じIDを使っている間のデータは同一人であることがわかってしまう．また，IDとユーザの関係が万が一露顕してしまうと，そのIDで蓄積された収集済みのデータ全てで匿名性が破綻するという問題がある．

3. リサイクルID

3.1. 基本コンセプト

リサイクルIDの基本コンセプトは，IDの交換である．

これにより，特定個人と ID のリンカビリティを断ち切りつつ，匿名プロキシで問題であった仲介者の信用性や，ワンタイム ID で問題であった ID から個人の履歴情報が一意に特定されるなどの問題を解決することが可能となる．

図 1 に示すように，リサイクル ID では，個人が ID プールの ID と自分が使用している ID を随時交換することで，ある瞬間においては全てのユーザに重複なく ID が割り当てられているが，一定の期間で見たときには ID が複数の個人に割り当てられている状態を作り出す．例えば，時刻 t0 において 01 という値を持つ ID は A 氏に割り当てられているが，時刻 t1 での交換によって ID プールへと移動し，時刻 t2 での交換で B 氏へ割り当たる．

これにより，ある瞬間の ID が悪意を持った第三者に知られたとしても，ID を用いて同一人物を追跡し続けることは困難となる．また，個人が ID を使用した際の記録がトランザクションとして蓄積されていた場合でも同様に，ID から同一人物の履歴が特定されることを困難にする．

以上のような特性から，ID 交換履歴さえ秘匿すれば，個人は ID を暗号等で秘匿せずに公開情報として利用でき，第三者に送受信データを覗かれることもない．

ID の交換には，複数のユーザが同時に行う場合も存在する．例えば，図 1 の時刻 t1 で A 氏と B 氏が同時に ID を交換している．2 人同時に ID を交換するには，ID プールに余剰 ID が最低でも 2 つ確保されている必要がある．余剰 ID が 1 つしかなければ，他のユーザの交換が終わるのを待たなければならない．

そこで，ID プールで確保すべき余剰 ID の数が問題となるが，ユーザの総数を N，同時に行われる交換の数を M とした時に，ユーザの総数よりも多い交換が同時に発生しないことから，M の最大値は N と等しくなる．したがって，ID プールでは余剰 ID を N 個用意すれば良いことが分かる．

ID の交換方式には，個人同士が直接 ID を交換する P2P 方式と，サーバが ID の交換を仲介する仲介方式がある．

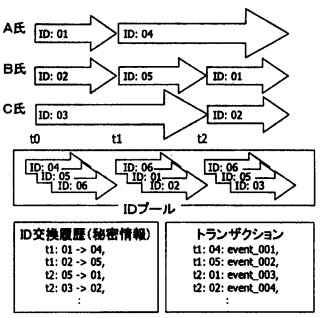

図 1. リサイクルの原理

3.2. P2P 方式

P2P(Peer to Peer)方式では，図 2 のように peer 同士が直接 ID を交換する．peer は ID を所有するエンティティで，端末や個人である．また，ID の交換相手はランダムに決定される．P2P 方式のメリットは以下の通りである．

- ID の交換に関わった当事者以外には ID を交換した事実を隠蔽でき，サービス提供者を含む，交換に関わらなかった者からの追跡を回避できる．
- 個人の数だけ ID を発行すれば良いため，ID プールを用意する必要がない．

一方，デメリットとしては，交換相手が常に存在するとは限らないため，必ずしも個人の好きな時に ID の交換が行えない点である．

なお，何らかの理由により後から ID の追跡可能とするために，各個人の ID の交換履歴を安全に保存しておく派生型も考えられる．

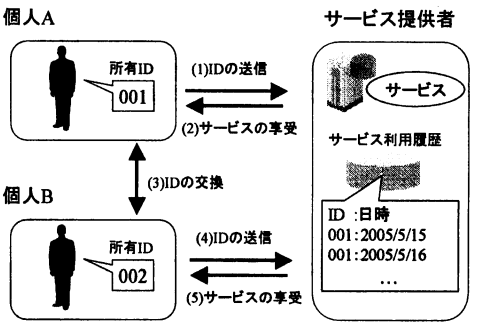

図 2. P2P 方式

3.3. 仲介方式

仲介方式では，図 3 のように仲介者を介して ID を交換する．個人の直接の交換相手は ID 交換サーバである．ID 交換サーバでは交換用の ID をあらかじめプールしておき，個人から要求があった際に ID を交換する．仲介方式のメリットは，以下の通りである．

- 交換相手が常時存在し，個人が好きなときに ID を変えられる．
- サービス提供者が個人の特定を行いたい場合には，ID 交換サーバを介さなければそれが分からない点である．

一方，デメリットは以下の通りである．

- ID 交換サーバとサービス提供者が共謀した場合に，P2P 方式と比較して個人の特定が容易に行える．
- 交換サーバの余剰 ID 数が少ない状況で同じ個人が繰り返し ID を交換すると，交換がワンパターンとなり ID の持ち主を特定され易くなる．そのため，余剰 ID 数に余裕を持たせる必要がある．

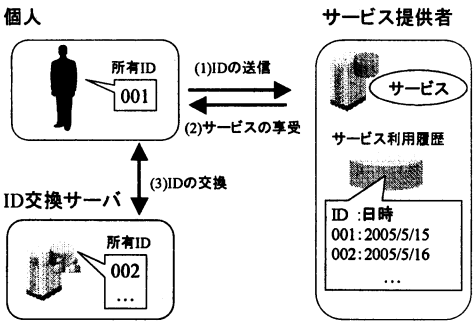

図 3. 仲介方式

4. P2P 方式による IP アドレス匿名化

4.1. 概要

固定 ID がプライバシ保護上，問題となる可能性がある分野の一つにネットワーク分野がある．

例えば，個人が Web ブラウザなどのクライアントアプリケーションを使用してサーバへアクセスする際に，ネットワーク端末が同じ IP アドレスを使用している間は，サーバに蓄積されるアクセスログを参照すれば，同一端末が閲覧した Web ページの履歴を追跡可能である．

個人ユースの多くでは，端末をネットワークに直接接続するか，宅内のルータを経由して接続する．前者の場合は当然として，後者の場合であっても，宅内ルータの IP アドレスは Web サーバから常に見えている．つまり，同一人物あるいは同一家庭の閲覧履歴を Web サーバ側で追跡することが可能である．

企業ユースの多くは組織内ネットワークとインターネットと通信路にファイヤウォールを設置しているため，ファイヤウォールが適切に設定されていれば，Web サーバに個々の端末の IP アドレスが通知されることはないが，ファイヤウォールのサーバの IP アドレスは見えている．そのため，同じ組織からのアクセスを Web サーバ側で追跡することが可能である．このような企業ユースにおける IP アドレスの追跡がプライバシ上の問題になるとは断言できないか，何らかの理由で追跡を防ぎたいならば，IP アドレスと組織のリンカビリティを秘匿するという側面では，問題の本質は同じである．

IP アドレス匿名化を想定したリサイクル ID の適用先としては，インターネットサービスプロバイダ(ISP)が提供する匿名化サービスや，インターネットマンションでの利用が考えられる．前者では，ISP の外部にあるサーバに対して個人の匿名性を確保でき，後者に適用した場合にはインターネットマンション内に設置されているルータに対して個人の匿名性を高めることができるようになる．

以下ではこのような IP アドレスを通じた追跡問題を想定し，P2P 方式による IP アドレス匿名化へのリサイクル ID 適用方式を説明する．

4.2. IP アドレス交換プロトコル

IP アドレスの交換は 2 台のネットワーク端末間で行う．図 4 のシーケンス図に示すように，左側のネットワーク端末が IP アドレスの交換の要求を出し，右側のネットワーク端末がその要求に応答する．

交換要求側端末で行われる処理は以下の通りである．

処理(a-1)：自分が IP アドレスを交換可能な状況であることをまず確認する．

処理(a-2)：サブネット内に交換募集通知を送信して，他のネットワーク端末から交換可能通知が返信されるのを待つ．

処理(a-3)：交換可能通知を送信してきたネットワーク端末の中から IP アドレスの交換を行う交換者を決定する．

処理(a-4)：交換者に対して交換依頼通知と IP アドレスを送信して交換受諾通知が返信されるのを待つ．

処理(a-5)：交換者に交換受諾確認通知を送信する．

処理(a-6)：交換した IP アドレスを自分の IP アドレスとして設定し，使用を開始する．

一方，応答側端末で行われる処理は以下の通りである．

処理(b-1)：自分が IP アドレスを交換可能な状況であることを確認する．

処理(b-2)：交換可能通知を交換要求側端末へ送信し，交換依頼通知が返信されるのを待つ．

処理(b-3)：交換要求側端末を交換者と決定し，交換受諾通知と IP アドレスを送信して交換受諾確認通知が返信されるのを待つ．

処理(b-4)：交換要求側端末から取得した IP アドレスを自分の IP アドレスとして設定し，使用を開始する．

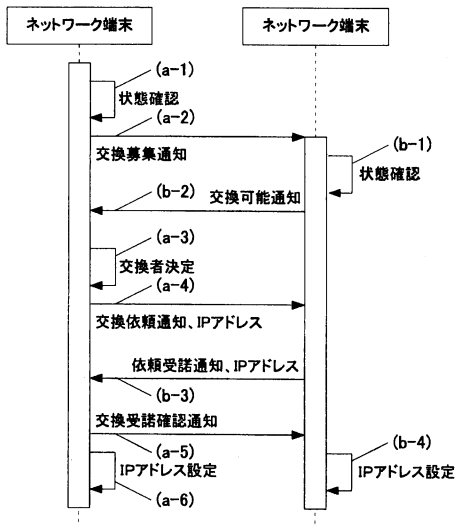

図 4. シーケンス図

4.3. IP アドレスの重複回避処理

本節では，前節の処理(a-6)および処理(b-4)の詳細を説明する．

処理(a-6)および処理(b-4)の中では，ネットワーク障害や端末自体のトラブルにより，どちらかの端末が IP ア

ドレスの設定に失敗する可能性を考慮し，IP アドレスの重複を回避する必要がある．

重複回避処理を行わなかった場合，例えば，処理(a-6)の IP アドレス設定に失敗して IP アドレスが変更されず，処理(b-4)の IP アドレス設定に成功して IP アドレスが変更され，どちらのネットワーク端末も同じ IP アドレスを持つような状態が生じる可能性がある．

重複回避処理は，処理(a-6)でも処理(b-4)でも同様である．本処理には正常に IP アドレスの設定を行える場合と，交換相手に障害が発生して正常に IP アドレスの設定を行えない場合とがある．以下，それぞれの場合における処理の流れを図 5 に示す．

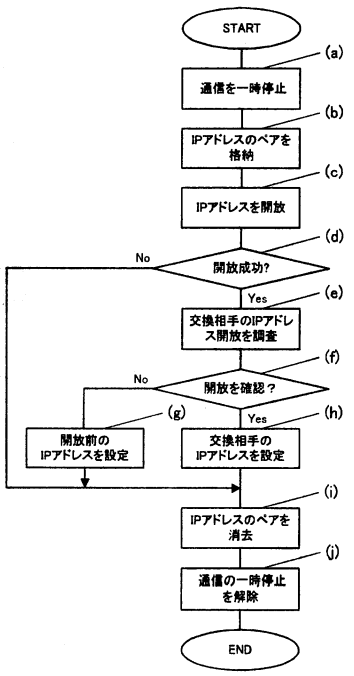

図 5. IP アドレスの重複回避処理

4.3.1. 正常に IP アドレスの設定を行える場合

前述のように，重複回避処理は IP アドレスの交換を行う両端末で同じように行われるが，ここでは説明の便宜上，交換要求側端末の処理を説明する．

まず，交換要求側端末が IP アドレスの設定処理を開始して処理(a)へ進む．

処理(a)：IP アドレスを用いる通信を一時停止し，処理(b)へ進む．

処理(b)：自分に付与されている IP アドレスと，応答側端末の IP アドレスとを一時領域に格納して，処理(c)へ進む．

処理(c)：自分に付与されている IP アドレスを開放し，処理(d)へ進む．

処理(d)：IP アドレスの開放状態をチェックする．成功した場合は処理(e)へ進む．失敗した場合は処理(i)以降に進み，元の IP アドレスを使い続けるよう設定して処理を終了する．

処理(e)：ARP (Address Resolution Protocol) を用いて，応答側端末の IP アドレスに対して一定期間繰り返して応答を求め，応答有無を調査する．

処理(f)：定期間中に応答側端末から応答がなくなったことを検知した場合には，応答側端末が IP アドレスを正常に開放したと判断して処理(h)へと進む．

処理(h)：応答側端末から受け取っていた IP アドレスを自分に設定して，処理(i)へ進む．

処理(i)：一時領域に格納していた IP アドレスを消去し，処理(j)へ進む．

処理(j)：一時停止させていた通信を再開させ，処理を終了する．

4.3.2. 正常に IP アドレスの設定を行えない場合

処理(a)から処理(e)までは，前節で説明した正常に行える場合と同じである．

処理(f)：一定期間経過しても交換相手から応答が有り続けた場合に，応答側端末が IP アドレスの開放に失敗したと判断して処理(g)へと進む．

処理(g)：自分が解放前に元々使用していた IP アドレスを設定して，処理(i)へ進む．

5. プロトタイプ

5.1. 概要

本章では実験的に構築したプロトタイプシステムについて述べる．我々はリサイクル ID の他，リンカビリティの制御によるプライバシ保護技術を提案してきてきた．プロトタイプ構築の狙いは，アプリケーション及び通信層で総合的に ID 追跡問題を捉え，提案技術の効果，問題点を検証することにある．

プロトタイプシステムで想定する環境は，インターネット上にあるショッピングサイトである．ここでは，単純化のため，顧客がショッピングサイトへアクセスする際にはファイヤウォールやルータを経由せず直接行うものとし，顧客が使用する端末にもあらかじめグローバル IP が割り当てられていることとする．

また，ショッピングサイト側では，誰が，いつ，何を注文したのかという購買情報の管理を全て顧客 ID で行う．顧客 ID はショッピングサイトがあらかじめ顧客に発行するアプリケーションレイヤの ID である．

以上の想定の下，本プロトタイプではアプリケーション層の顧客 ID の匿名化手法として，我々が開発した 3 レベル ID 制御技術[4]を実装した．3 レベル ID 制御は，偽名（同一人であることを保証），実名（身元を明らかにする）を暗号化して中に含んだ匿名を生成する技術である．3 レベル ID 制御技術については次節で概要を説明する．

ただし，3 レベル ID 制御技術を利用して匿名を通信した場合，例えば IP アドレスが同じであれば，せっかく毎回異なる匿名を使っていても IP 層では同一人の行動であることが露見する可能性があった．そこで，プロトタイプでは，この 3 レベル ID 制御を拡張する形とし，IP 層

にリサイクル ID 技術を適用し IP アドレスの匿名化もあわせて実現するプロトタイプを開発した．これにより，顧客 ID と IP アドレスのログをつき合わせても匿名性が保証できる．

本プロトタイプのリサイクル ID には P2P 方式をを採用し，基本的な動作を見るため，IP アドレス重複回避処理は実装せず，4．2 節の基本プロトコルのみを実装した．

5.2. 3 レベル ID 制御技術

3 レベル ID 制御技術は，ID に実名，偽名，匿名の異なる 3 つのレベルを持たせて使い分けることで，個人と紐付けられる情報の開示レベルをコントロールする．

ここで言う実名 ID とは，それが明らかになることで個人の身元が特定可能になる ID である．また，偽名 ID とは，同一人物によって行われた複数の行為が特定可能になる ID である．匿名 ID とは，一度限り使用される使い捨ての ID である．

3 レベル ID 制御技術では，レベルのコントロールを暗号技術で行う．実名にパラメータを加えて暗号化したものを偽名 ID とし，偽名 ID に乱数を加えて暗号化したもの匿名 ID とする．

このような構成により，復号鍵を全く持たない人物は，個人も同一人物も特定できないデータのみを利用できるようにする．匿名から偽名へ戻すための復号鍵を持つ人物は，個人を特定できないが同一人物を特的できるデータを利用できる．偽名から実名へ戻す復号鍵まで持つ人物は，個人を特定できるデータを利用できる．

5.3. システム構成

プロトタイプシステムは，図 6 に示すように，端末 A，端末 B，Web サーバで構成される．端末 A および端末 B は，オンラインショップの顧客が使用する PC であり，そこでは IP アドレス交換アプリケーション，匿名生成アプリケーション，Web ブラウザが動作する．

一方，Web サーバは，オンラインショップの従業員が使用する PC であり，そこではオンラインショップのホームページ，受注 DB，集計処理アプリケーション，ポイント管理アプリケーション，受注・配送管理アプリケーションが動作する．

上記に挙げたアプリケーションのうち，Web ブラウザは OS 付属のものである．また，オンラインショップのホームページは HTML 形式のテキストファイルであるが，Active Server Pages(ASP)を使用して顧客からの注文が受注 DB に蓄積されるようにしている．受注 DB は CSV 形式のテキストファイルである．それ以外のアプリケーションについては Java で実装を行った．

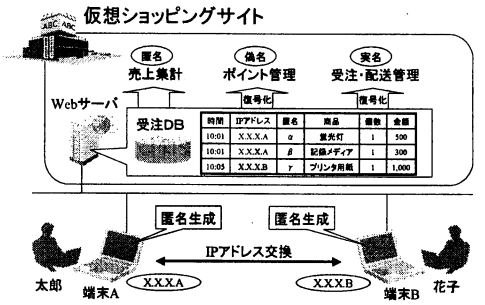

図 6. プロトタイプシステム構成

6. 評価実験

6.1. 実験条件

本章では，プロトタイプを用いた性能評価実験を説明する．プロトタイプでは，IP アドレスの交換要求を出す PC と，要求を受けて交換に応じる PC の 2 台を同一のネットワークセグメントに接続した．ただし，IP アドレスの交換要求を出して相手を決定する処理の時間と，自分の PC に交換相手の IP アドレスを設定する処理の時間を計測するため，処理時間の計測は IP アドレスの交換要求を出す PC 側で行う．交換される IP アドレスは，事前に各 PC へ割り振った固定 IP アドレスである．

実験に使用する各 PC のスペックは表 1 の通りである．

表 1. PC のスペック

	交換要求を出す PC	交換要求を受ける PC
CPU 周波数	1.13GHz	3.2GHz
メモリ容量	512MB	512MB
OS	Windows 2000 SP4	Windows XP SP2
J2SE バージョン	1.4.2_10	1.4.2_10
通信速度	100Mbps	100Mbps

また，IP アドレスの交換を行う実験用プログラムは Java で実装した．ただし，IP アドレスを PC に設定する処理については，ネットワークの設定を行うコマンドラインツールである netsh を Java の Runtime クラスから呼び出して使用している．

実験では，上記プログラムを用いて IP アドレスの交換に要する時間を計測した．

6.2. 実験結果

IP アドレスの交換を 10 回行った時の計測結果を表 2 に示す．表中の項目にある決定処理とは，4. 2 節で説明した処理(a-1)から処理(a-5)までの一連の処理を表し，設定処理とは処理(a-6)を表す．

表 2 から，IP アドレスの交換を開始後，凡そ 3 秒間は他の PC と通信が断絶することが分かった．決定処理に比べて設定処理の時間が 100 倍程度長く，処理全体の大部分を占めている．

表 2. IP アドレス交換に要する時間

	決定処理 (ms)	設定処理 (ms)	合計 (ms)
平均	19	3, 165	3, 184
分散	29	3, 922	3, 710

6.3. 考察

IP アドレスの交換に要する約 3 秒という時間は，ユーザが Web サイトを巡回しながら行う程度であれば充分実用的な時間であると考える．もし，自動的に IP アドレスの交換を行うのであれば，5 分間隔や 10 分間隔といったように，3 秒に対して充分長い期間を設定して通信のスループット低下を抑える必要がある．

また，処理時間の大部分を設定処理が占めているが，この処理は netsh コマンドを使用して IP アドレスの変更を行っている．設定処理の時間は OS の種類や，PC 上で動作するネットワークアプリケーションの数によっても変動すると思われるが，netsh コマンドの替わりに Windows Management Instrumentation(WMI)などを利用し，よりハードウェアに近いレイヤで IP アドレスを設定することにより，処理を高速化する余地はあると考える．

7. まとめ

本稿では，ID から個人の履歴情報が特定される問題の解決を目指し，ID の交換を基本コンセプトにした新しい ID 匿名化手法として，リサイクル ID を提案した．

リサイクル ID では，個人が交換によって手に入れた ID をサービス提供者へと渡すことにより，サービス提供者内に蓄積される ID を撹乱することができる．

また，リサイクル ID の適用例として TCP/IP ネットワークを想定し，IP アドレスを匿名化するための IP アドレス交換プロトコルを提案した．

オンラインショップを想定適用例としたプロトタイプによる性能評価では，ユーザからオンデマンドで匿名化を行うような使い方では実用性があるとの結論を得た．

参考文献

[1]Michael K. Reiter,Aviel D. Rubin,Crowds: Anonymity for Web Transactions,ACM Transactions on Information and System Security,Volume 1,Issue 1,pp.66-92,1998

[2]D. Chaum,Untraceable electronic mail, return addresses, and digital pseudonyms,Communications of the ACM,Vol.24,No.2,pp.84-88,1981

[3]木下，星野，小室，藤村，大久保，RFID プライバシー保護を実現する可変秘匿 ID 方式，CSS2003 論文集，2003

[4]森藤，川崎，森田，宝木，プライバシを考慮したアイデンティティ制御方法，情報処理学会「コンピュータセキュリティ研究会」，2005-CSEC-030，pp. 467-471，2005