5V-06

位置情報の匿名化の有用性向上に関する一検討

西山 賢志郎† 林 弘悦‡ 土井 洋†
情報セキュリティ大学院大学情報セキュリティ研究科† 中央大学大学院理工学研究科‡

1. はじめに

位置情報は有用性が高い情報であるが，その反面で高い個人識別性を有している．プライバシ保護の指標として k-匿名性[1]が知られており，位置情報の k-匿名性を満たす匿名化手法として Interval Cloak[2]や Casper[3]等がある．これらの手法では k-匿名性を満たすためにデータの加工を行うが，そのため情報の有用性が損なわれる．匿名化の限界[4]は存在するが，匿名性を満たしつつ有用性も有する効率よい加工方法が望まれている．本稿では，Interval Cloak のデータの加工処理を細かく制御することで情報の有用性を保ちつつ匿名性を満たす手法を提案し，評価結果を示す．

2. 既存手法とその問題点

本稿では、与えられた領域を再帰的に 4 分割し，その後データの加工を行う手法を扱う．図 1(i)及び図 1(ii)は，与えられた領域(最上位領域)を 2 度再帰的に 4 分割した例である．この例では 16 個ある最下位領域に含まれる位置情報の数は，最大で 32，最小で 0 である．位置情報の k-匿名性とは，データの加工の結果出力される各領域に位置情報が k 個以上存在することを保証する指標である．本稿ではデータの加工として一般化と削除を考える．一般化とは隣接する領域を複数結合することでより広い領域にすることであり，削除とはその領域を加工結果として出力しないことである．位置情報の k-匿名性を満たすように位置情報の一般化や削除を行うことを位置情報の k-匿名化という．

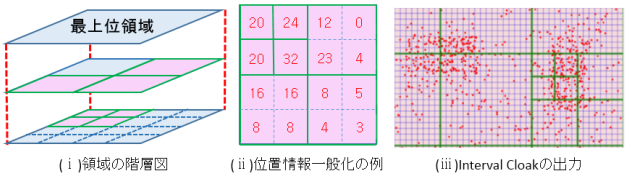


図 1 位置情報の一般化の例

Interval Cloak[2]では最下位領域から順に，領域内に位置情報が k 個以上存在するかどうかを判定する．その領域に位置情報が k 個以上存在しない場合は，1 つ上の階層で同じ処理を行う．そして，最終的に k 個以上の位置情報を持つ領域を出力する．図 1(ii)の実線で囲まれた領域は，k=20 とした場合の Interval Cloak の出力である．出力される各々の領域には 48, 39, 32, 24, 20, 20, 20 個の位置情報が含まれているが，7 個の領域(正方形)のみが出力される．図 1-(iii)は後述する人工データに対する Interval Cloak の出力結果の一部である．Interval Cloak は，一般化，すなわち階層が上がるたびに領域の面積が 4 倍の大きさになるので，これが原因で過度な一般化が発生する可能性が高い．

Casper[3]は，Interval Cloak と同様の処理を行うが，領域に位置情報が k 個以上存在しない場合，1 つ上の階層に上がる前に，隣接する領域と結合(面積が 2 倍)して位置情報数の評価を行う．それでも k 個以上存在しない場合は 1 つ上の階層で同様の処理を行う．この工夫により，Interval Cloak の過度な一般化の問題を緩和している．なお，Interval Cloak や Casper は一般化は行うが削除は行わない．

これらの方法では，処理対象の領域に位置情報が偏って存在する場合，出力される領域の面積が大きくなりすぎる可能性がある．なぜなら，位置情報数が少ない領域(疎な領域)を k-匿名性を満たすために他の領域と結合することになるが，結合された領域の位置情報数が十分大きい領域(密な領域)である可能性があるからである．そのため，情報の有用性が大きく損失する可能性がある．また，出力結果の利用方法にも依存するが，面積が大きすぎる領域が得られたとしても役に立たない場合も考えらえる．

3. 提案手法

我々は密な領域と疎な領域とを結合するよりも，密な領域はそのままにし，疎な領域を削除するアプローチを検討した．この結果として，面積が比較的小さい密な領域が出力されることが期待できる．なお，本研究では事前情報により，密になりやすい領域，疎になりやすい領域の情報を利用できることとした．

提案手法では，匿名化の前に位置情報が密な領域の情報を元に領域の一般化を停止する目印を用意する．この目印を元に Interval Cloak の一般化(領域の結合)処理を制御することで，情報の有用性の損失を防ぐ．

提案手法では，まず事前情報に基づき位置情報が密になりやすい領域 A_C を定める．その後，A_C を利用して最上位領域から再帰的にトップダウン方式で一般化停止フラグの設定処理を行う．初めに最上位領域 N を引数として渡す．N が A_C の一部と重複する場合，N のすぐ下の領域の集合 CHILDREN(N) 全てに対して一般化停止フラグ OSFlg を有

A study on the improvement of usefulness of location anonymization
† Kenshiro NISHIYAMA, Hiroshi DOI, Institute of Information Security University
‡ Hiroyoshi HAYASHI, Graduate school of Science and Engineering, Chuo University.

効に設定する．更に，これらの領域を引数として，同処理を再帰的に行う．アルゴリズム 1 は擬似コードであり，図 2 は一般化停止フラグ設定の例である．

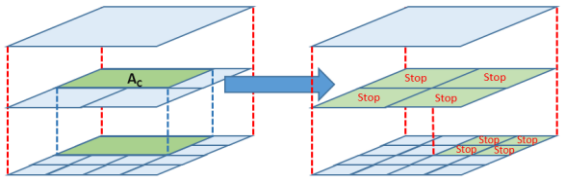


図 2 一般化停止フラグの設定例

アルゴリズム 1　一般化停止フラグの設定

```
1.  Function Set_OrganizationStopFlg(N)
2.  if  (N ⊂ A_c) or (N ⊂ Ā_c)  then
3.    return
4.  else
5.    for each N_C ∈ CHILDREN(N) do
6.      if |CHILDREN(N_C)| = 0 then
7.        N_C.OSFlg = true
8.        return
9.      else
10.       N_C.OSFlg = true
11.       Set_OrganizationStopFlg(N_C)
12.     end if
13.   end for
14. end if
```

一般化停止フラグに対応した Interval Cloak では，ある領域 N に存在する位置情報数 locs が k 未満である場合は，親の領域を参照する前にその領域の一般化停止フラグが有効か確認する．有効であれば N の親の領域 PARENT(N)を参照せず，その領域は k-匿名性を満たせないとして削除された領域 Suppression(N)を返す．

アルゴリズム 2　一般化停止フラグに対応した Interval Cloak

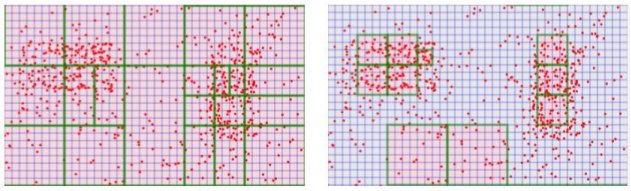
```
1.  Function IntervalCloak_OSP (k, N)
2.  if  (N.locs ≥ k) then
3.    return N
4.  else
5.    if (N.OSflg = true) then
6.      return Suppression(N)
7.    else
8.      IntervalCloak_OSP (k, PARENT(N))
9.    end if
10. end if
```

4. 評価結果

既存手法と検討手法を Java 8 Update 60 で実装した．実装環境は CPU が Inter® Core™ i7-4790 CPU(3.60Ghz)，メモリが 24GB である．また，一部の領域に位置情報が密集しやすいようにした人工データを作成し，評価を行った．

k=20 とした場合の Interval Cloak の出力結果が前述の図 1(iii)である．図 3 (i)は Casper，図 3 (ii)は提案手法の出力結果の一部である．従来法(図 1(iii)や図 3 (i))の出力結果の左側で，既存手法の問題点である過度な一般化が生じていることを確認できる．提案手法(図 3 (ii))の同箇所では過度な一般化が抑制され，面積の小さな領域が出力されている．しかし， 出力結果の右側のように，従来手法の方が面積の小さな領域を多く出力する場合もある．

図 3　既存手法，検討手法の実装結果



(i)Casperの出力　　(ii)提案手法の出力

表 1 は，最下位領域の面積を 1 とし，匿名化後の面積が 16 以下の領域の総数を記したものである．k=20 とし，位置情報数を 1750 程度に設定した場合，提案手法の方が面積の小さな(16 以下の)領域の数が多い．しかし，位置情報数を 3500 程度に設定すると，Casper のほうが面積の小さい領域が発生する割合が高くなる．

表 1　面積が 16 以下の領域数

	k=20					
	位置情報数≒1750			位置情報数≒3500		
	IC	Casper	提案手法	IC	Casper	提案手法
0	0	4	8	12	38	33
1	0	4	7	15	42	41
2	4	5	10	16	50	43
3	0	6	7	12	51	46
4	0	7	8	16	55	45

5. おわりに

本稿では，事前に位置情報が密な領域を設定し、これをもとに削除及び一般化を細かく制御する位置情報の匿名化手法を提案した．位置情報の分布に偏りがある場合，領域面積が小さく有用な結果が出力されることも確認した．

参考文献

[1] Sweeney, L.: k-Anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, Vol.10, No.5, pp.557–570 (2002).

[2] Gruteser, M., Grunwald, D.: Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking, Mobisys 2003, pp.31-42 (2003).

[3] Mokbel, M. K., Chow, C. Y., and Aref, W. G.: The New Casper: Query Processing for Location Services without Compromising Privacy, VLDB'06, pp.763-774 (2006).

[4] 高橋克巳，正木彰伍，濱田浩気，個人データの匿名化とその限界，電子情報通信学会誌，Vol.98, No.3, pp. 193-201 (2015).