資料４

Innovative R&D by NTT

位置情報の匿名化について

2014.3.18

NTTセキュアプラットフォーム研究所
高橋克巳

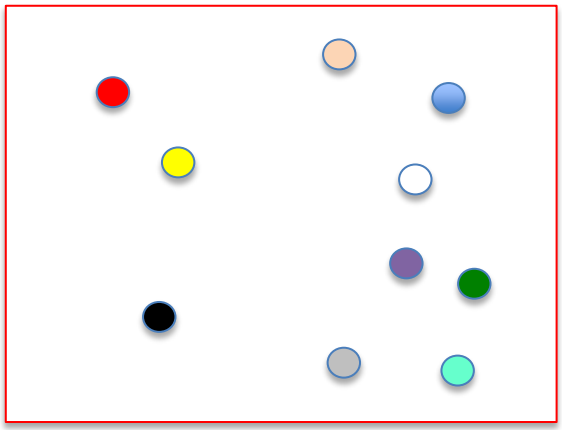
本資料の目的と位置付け

- 位置情報に限定して、どのような加工をすれば、非個人情報として取り扱えるのかの検討
  - 位置情報の「非識別・非特定情報への加工」に関する技術情報の提供を行う
- 位置情報は、正しい手続きで取得されていることを前提とする
- 位置情報の匿名化による保護は、技術的に難しい対象であり、一般的に「こうすればよい」という学術的合意は形成されていない
- 本資料が指摘するリスクは制約を意図せず、むしろ位置情報の取扱いのプロセスでリスクを共有し、健全な流通を行うための道具として参照されることを期待する
  - あらゆるリスクに対応しようとすると、そのデータはあたかも「市町村別人口統計」のような粒度になってしまうかもしれない
  - （仮称）法第23条1項適用除外情報による取り扱いを含め、ケースに応じて柔軟に合意形成することが重要
  - 匿名化手法云々ではなく、どのようなデータを作成するかが重要

位置情報：ケース１

- 識別情報（ID）がなく、時間が単一のケース

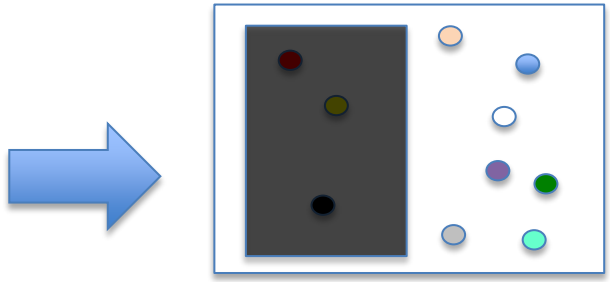
時刻	位置
08:01	位置1
08:01	位置2
08:01	位置3
08:01	位置4
08:01	位置5
08:01	位置6
08:01	位置7
08:01	位置8
08:01	位置9
08:01	位置10



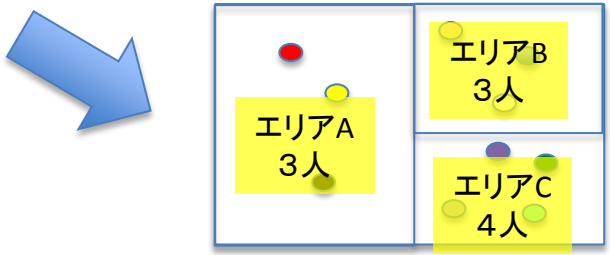
【論点】
これだけでは問題がないのではないか？
（個人特定は起きない）

《想定されるリスク》
- 居る/居ない/行ったことが知られては困る、たとえば自宅

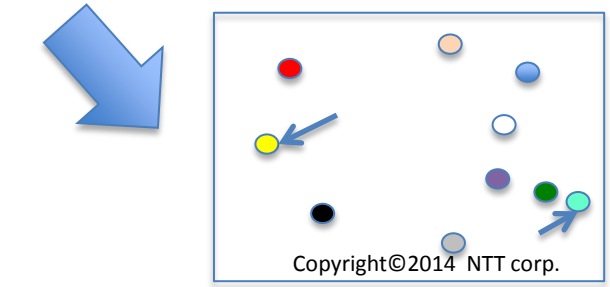
考え方a：センシティブ位置の削除



考え方b：位置のまとめ（エリア化）



考え方c：位置のランダム化



位置情報：ケース１の考え方 (1/3)

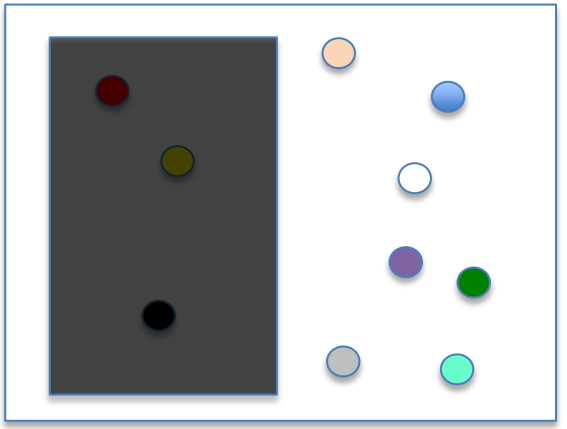
- 考え方a:センシティブ位置の削除

時刻	位置
08:01	位置1
08:01	位置2
08:01	位置3
08:01	位置4
08:01	位置5
08:01	位置6
08:01	位置7
08:01	位置8
08:01	位置9
08:01	位置10

時刻	位置
08:01	位置1
08:01	位置2
08:01	位置3
08:01	位置4
08:01	位置5
08:01	位置6
08:01	位置7
08:01	位置8
08:01	位置9
08:01	位置10

（注．黒塗りは削除を示す）

考え方a:センシティブ位置の削除
（自宅は削除・商業地域を利用）



※センシティブ位置としては
病院などの施設も考えられる
（分析目的による）

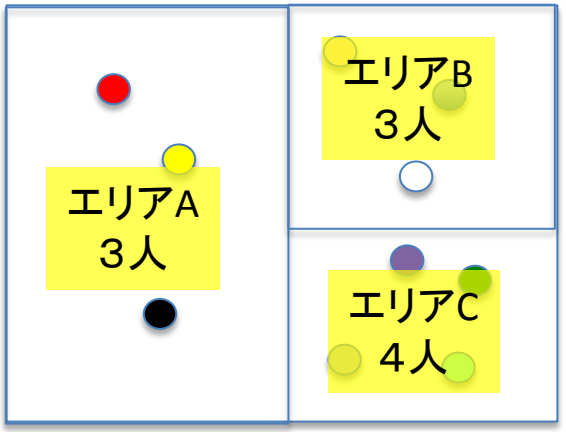
位置情報：ケース１の考え方 (2/3)

- 考え方b：位置のまとめ（エリア化）

時刻	位置
08:01	位置1
08:01	位置2
08:01	位置3
08:01	位置4
08:01	位置5
08:01	位置6
08:01	位置7
08:01	位置8
08:01	位置9
08:01	位置10

時刻	位置
08:01	エリアA
08:01	エリアA
08:01	エリアA
08:01	エリアB
08:01	エリアB
08:01	エリアB
08:01	エリアC
08:01	エリアC
08:01	エリアC
08:01	エリアC

考え方b：位置のまとめ（エリア化）



※位置を適切なサイズのエリアに拡大させて、エリアごとに十分な人数を確保する

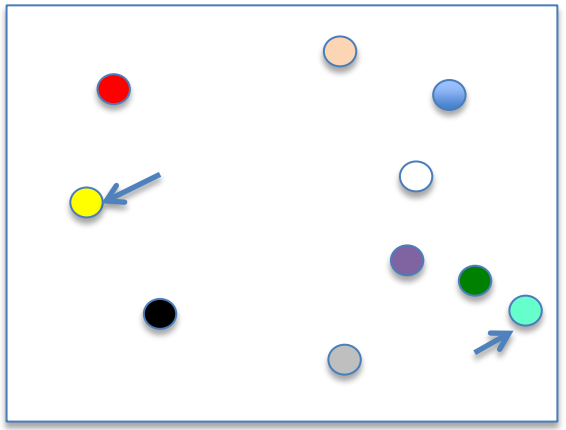
位置情報：ケース１の考え方 (3/3)

- 考え方C：位置のランダム化

時刻	位置
08:01	位置1
08:01	位置2
08:01	位置3
08:01	位置4
08:01	位置5
08:01	位置6
08:01	位置7
08:01	位置8
08:01	位置9
08:01	位置10

時刻	位置
08:01	位置1
08:01	位置2’
08:01	位置3
08:01	位置4
08:01	位置5
08:01	位置6
08:01	位置7
08:01	位置8
08:01	位置9
08:01	位置10’

考え方C：位置のランダム化

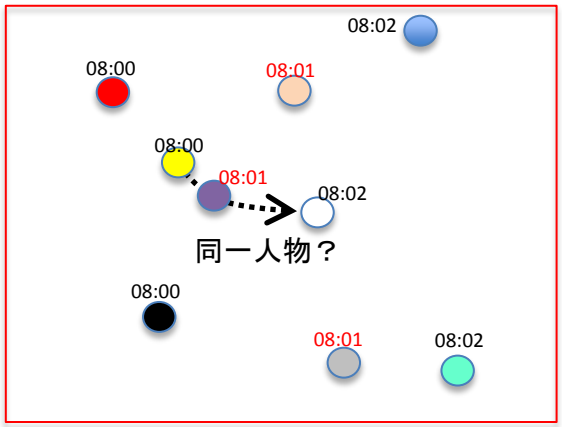


※いくつかの位置の場所をランダム化（不規則にずらす、置き換える）することにより知られたくない位置が確定的にもれることを防ぐ

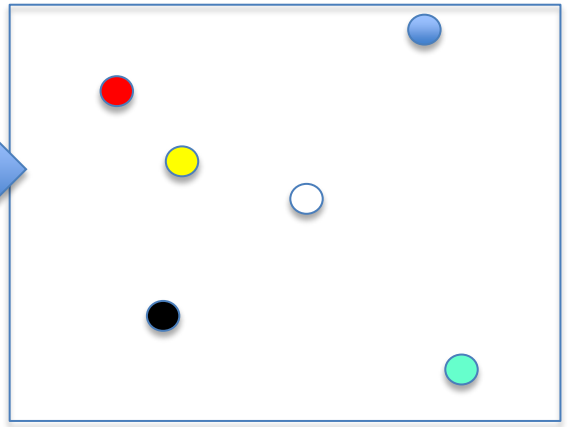
位置情報：ケース２

- 識別情報（ID）がなく、時間が複数のケース

時刻	位置
08:00	位置1
08:00	位置2
08:00	位置3
08:01	位置4
08:01	位置5
08:01	位置6
08:02	位置7
08:02	位置8
08:02	位置9



考え方d：更新間隔の間引き



【論点】
これだけでは問題がないのではないか？

《想定されるリスク》
- 複数の位置が同一人物のものと推定されないか？

※位置の更新間隔が極端に短い場合

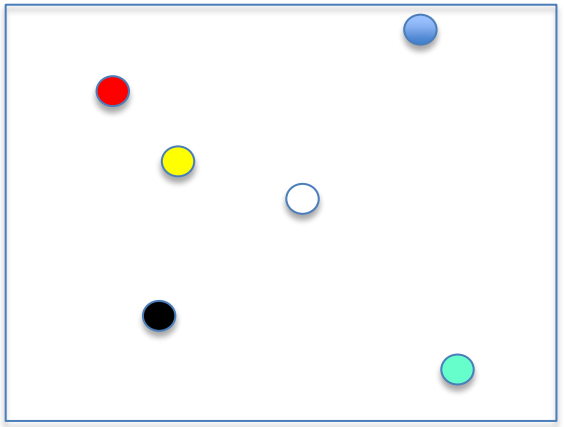
位置情報：ケース２の考え方

- 考え方d:位置の更新間隔の間引き

時刻	位置
08:00	位置1
08:00	位置2
08:00	位置3
08:01	位置4
08:01	位置5
08:01	位置6
08:02	位置7
08:02	位置8
08:02	位置9

時刻	位置
08:00	位置1
08:00	位置2
08:00	位置3
08:01	位置4
08:01	位置5
08:01	位置6
08:02	位置7
08:02	位置8
08:02	位置9

考え方d:更新間隔の間引き

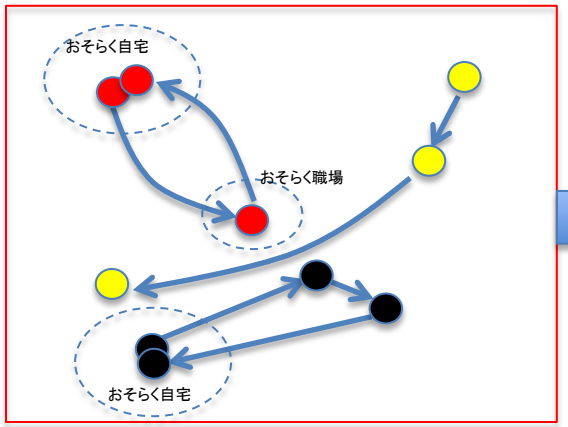


※ 時間的に近接した位置を間引いて、奇跡の推定を防止する

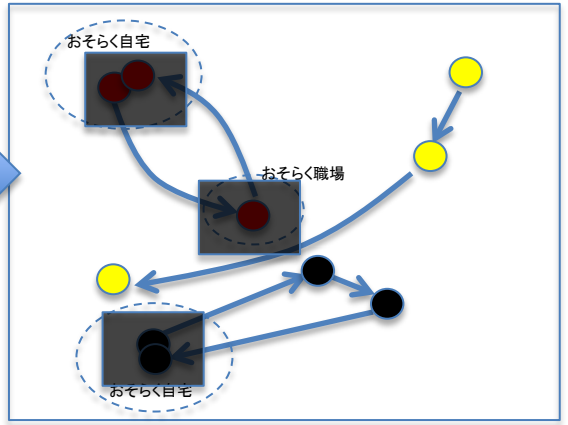
位置情報：ケース３

- 識別情報（ID）があり、位置が複数のケース

ID	時刻	位置
ID01	08:00	位置１
ID01	09:00	位置２
ID01	20:00	位置１
ID02	08:00	位置３
ID02	10:00	位置４
ID02	15:00	位置５
ID02	21:00	位置３
ID03	08:00	位置６
ID03	09:00	位置７
ID03	20:00	位置８



考え方a：センシティブ位置の削除

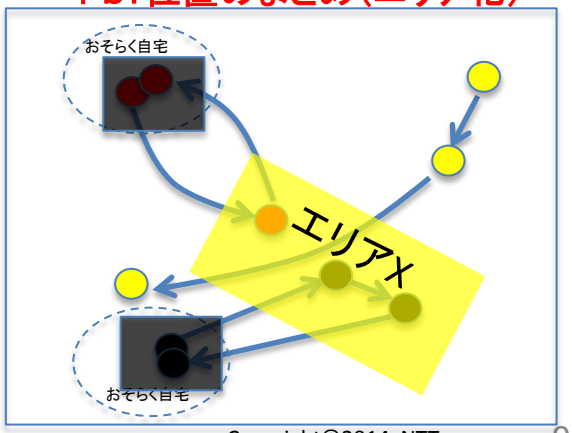


【論点】

《想定されるリスク》

- 移動軌跡から知られたくない位置の「意味」がわかってしまわないか？
  （始点→自宅、終点→職場）

考え方a：センシティブ位置の削除
＋b：位置のまとめ（エリア化）



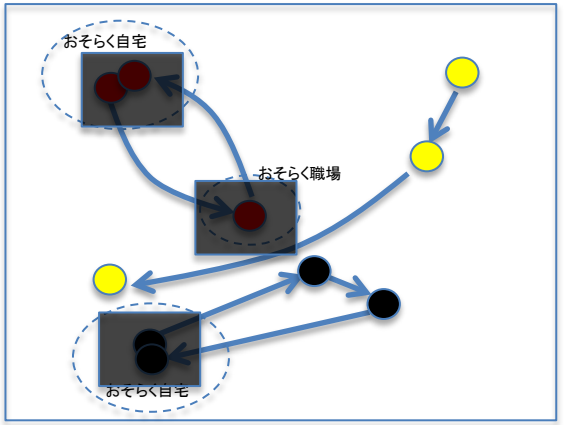
位置情報：ケース３の考え方 (1/2)

- 考え方a:センシティブ位置の削除

ID	時刻	位置
ID01	08:00	位置1
ID01	09:00	位置2
ID01	20:00	位置1
ID02	08:00	位置3
ID02	10:00	位置4
ID02	15:00	位置5
ID02	21:00	位置3
ID03	08:00	位置6
ID03	09:00	位置7
ID03	20:00	位置8

ID	時刻	位置
仮名01	08:00	位置1
仮名01	09:00	位置2
仮名01	20:00	位置1
仮名02	08:00	位置3
仮名02	10:00	位置4
仮名02	15:00	位置5
仮名02	21:00	位置3
仮名03	08:00	位置6
仮名03	09:00	位置7
仮名03	20:00	位置8

考え方a:センシティブ位置の削除



※ IDは仮名に変換
※ 仮名の再利用はしない

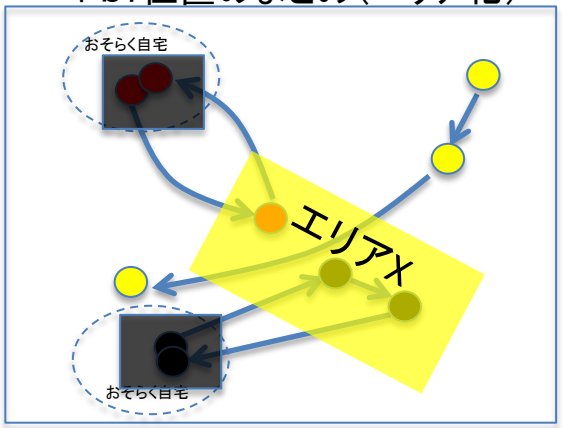
位置情報：ケース３の考え方 (2/2)

- 考え方a:センシティブ位置の削除
  ＋b:位置のまとめ(エリア化)

ID	時刻	位置
ID01	08:00	位置1
ID01	09:00	位置2
ID01	20:00	位置1
ID02	08:00	位置3
ID02	10:00	位置4
ID02	15:00	位置5
ID02	21:00	位置3
ID03	08:00	位置6
ID03	09:00	位置7
ID03	20:00	位置8

ID	時刻	位置
仮名01	08:00	位置1
仮名01	09:00	エリアX
仮名01	20:00	位置1
仮名02	08:00	位置3
仮名02	10:00	エリアX
仮名02	15:00	エリアX
仮名02	21:00	位置3
仮名03	08:00	位置6
仮名03	09:00	位置7
仮名03	20:00	位置8

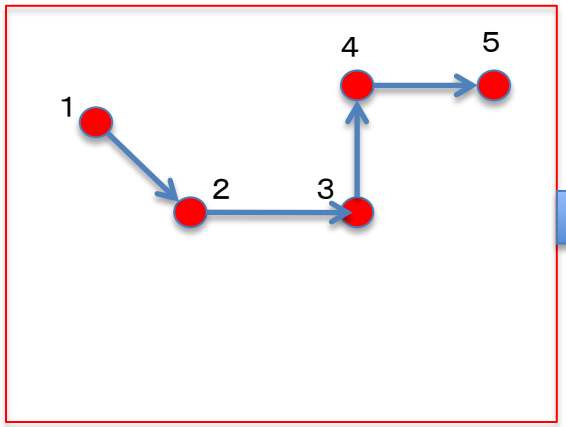
考え方a:センシティブ位置の削除
＋b:位置のまとめ(エリア化)



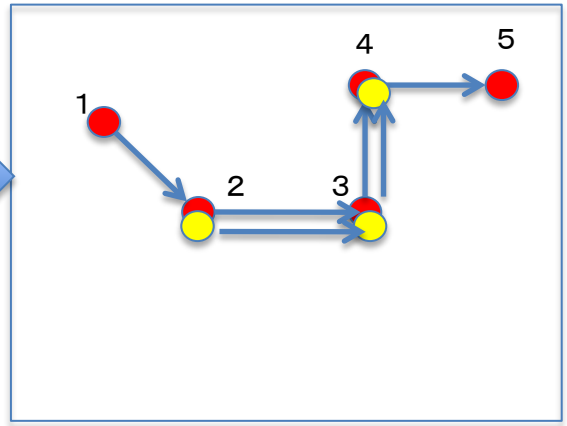
位置情報：ケース3’

- 識別情報（ID）があり、位置が複数でかつ系列が長い

ID	時刻	位置
ID01	08:00	位置1
ID01	09:00	位置2
ID01	10:00	位置3
ID01	11:00	位置4
ID01	12:00	位置5
ID02	08:00	位置X
ID02	09:00	位置2
ID02	10:00	位置3
ID02	11:00	位置4



考え方e：同じ行動の人がいる軌跡を使う

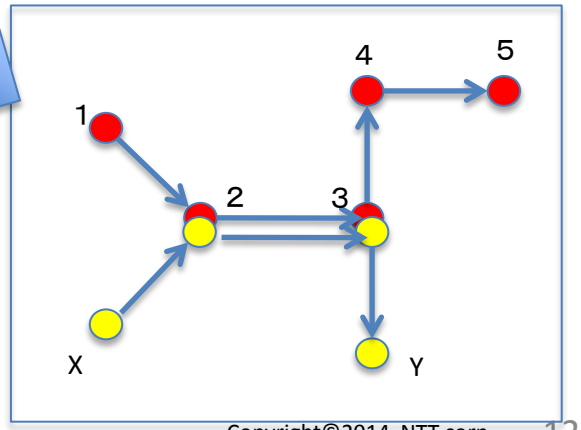


【論点】

《想定されるリスク》

- 誰の軌跡であるのかわかってしまうのではないか？（個人特定）

考え方f：他人の軌跡と混ぜて使う



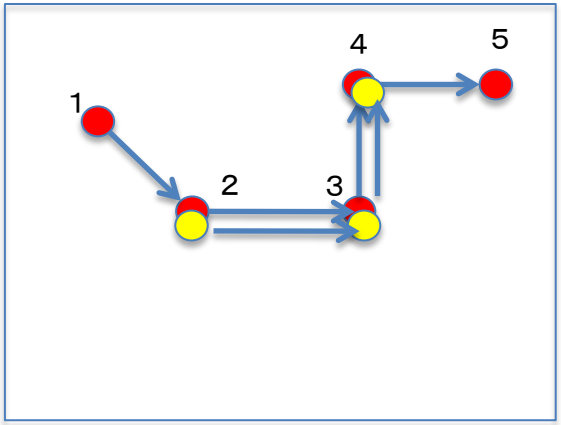
位置情報：ケース3'の考え方(1/2)

- 考え方e:同じ行動の人がいる軌跡を使う

ID	時刻	位置
ID01	08:00	位置1
ID01	09:00	位置2
ID01	10:00	位置3
ID01	11:00	位置4
ID01	12:00	位置5
ID02	08:00	位置X
ID02	09:00	位置2
ID02	10:00	位置3
ID02	11:00	位置4

ID	時刻	位置
仮名01	09:00	位置2
仮名02	09:00	位置2
仮名01	10:00	位置3
仮名02	10:00	位置3
仮名01	11:00	位置4
仮名02	14:00	位置4

考え方e:同じ行動の人がいる軌跡を使う



《同時間・同位置》の人は現実には存在しないため、位置のまとめ・時間のまとめ、および系列長の制限を行う

同じ行動の人が多数いるほど個人識別が起こりにくい

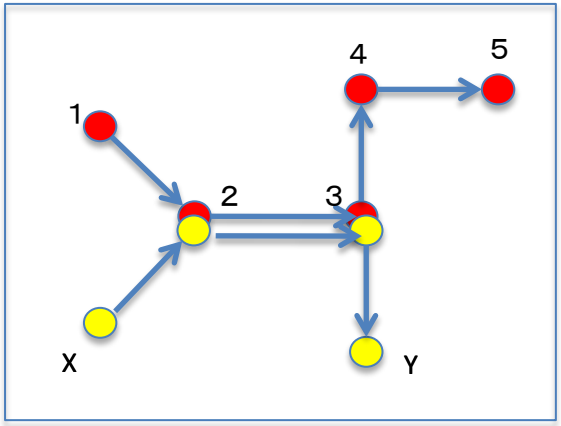
位置情報：ケース3'の考え方(2/2)

- 考え方f:他人の軌跡と混ぜて使う

ID	時刻	位置
ID01	08:00	位置1
ID01	09:00	位置2
ID01	10:00	位置3
ID01	11:00	位置4
ID01	12:00	位置5
ID02	08:00	位置X
ID02	09:00	位置2
ID02	10:00	位置3
ID02	11:00	位置Y

ID	位置
仮名01	位置1→2
仮名02	位置X→2
仮名03	位置2→3
仮名04	位置2→3
仮名05	位置3→4
仮名06	位置3→Y

考え方f:他人の軌跡と混ぜて使う

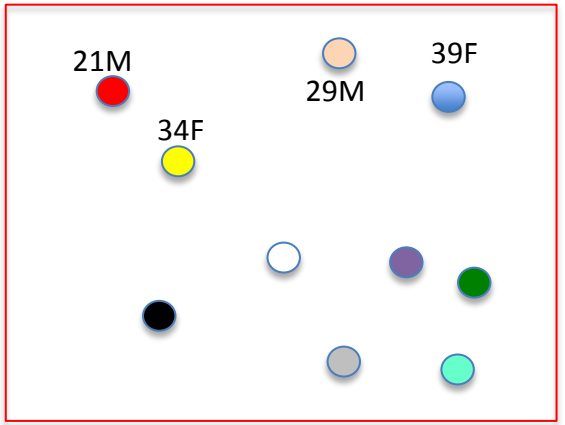


同じ点に同時に出入りする人が多数いるほど個人識別が起こりにくい

位置情報：ケース４

- 単一時間で、位置・年齢・性別の情報がある

年齢	性別	時刻	位置
21	男	08:00	位置1
34	女	08:00	位置2
29	男	08:00	位置3
39	女	08:00	位置4
57	男	08:00	位置5
49	女	08:00	位置6
55	男	08:00	位置7
.....	.....	.....	.....



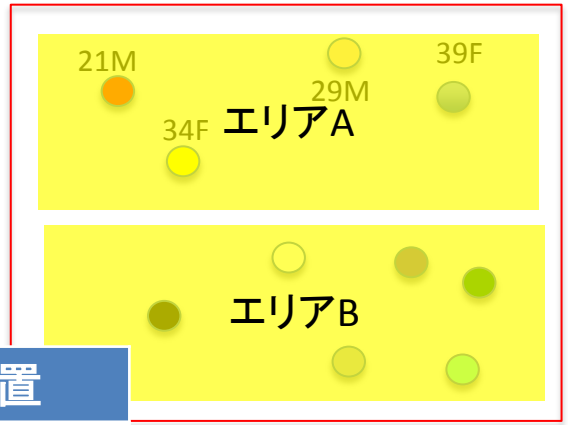
【論点】
- どのような条件下で個人特定が可能になるか？

《考え方》
- 一般には「ある位置に8時に21歳男性がいた」ことだけを知っても個人特定をすることは困難
- 「どこかの位置で8時に通過滞留する者を見張っている」人は個人特定が可能になる
- 位置に関するリスクを回避するためにはケース1から3までで紹介した手法が有効である
- IDあり軌跡ありのケースも同様の考え方で拡張可能

位置情報：ケース４の考え方

- 位置の加工をしながら全ての属性も加工して、総合的に保護された状況を作る

年齢	性別	時刻	位置
21	男	08:00	位置1
34	女	08:00	位置2
29	男	08:00	位置3
39	女	08:00	位置4
57	男	08:00	位置5
49	女	08:00	位置6
55	男	08:00	位置7
.....	.....	.....	.....



年齢	性別	時刻	位置
20代	男	08:00	エリアA
20代	男	08:00	エリアA
30代	女	08:00	エリアA
30代	女	08:00	エリアA
50代	男	08:00	エリアB
50代	男	08:00	エリアB
40代	女	08:00	エリアB
.....	.....	.....	.....

たとえば、位置のまとめをしながら全ての属性も加工して、複数人いる状況を作る

加工されたデータの望ましい性質

- センシティブな位置の保護ができているか
  - 削除されているか
  - またはエリアで扱われているか
  - またはランダムに扱われているか
- 位置の軌跡の保護ができているか
  - 同じ行動の人がいる軌跡になっているか
  - または他人と混ぜて使える軌跡になっているか
- 位置以外の属性がある場合は、
  - 直接個人が特定できる属性が取り除かれているか
  - 他の属性は位置との組み合わせた上で保護された状況ができているか

位置情報の表現形式と加工のテクニック

- 《位置情報を上位情報に置き換えて、カバーする人や事象を多くする》テクニック
  - 緯度経度 → メッシュやエリアに置き換え
  - メッシュ・エリア → 上位のメッシュやエリアに置き換え
  - 行政区界（住所）→ 上位の住所界への置き換え
  - 郵便番号 → 下4桁の削除
  - 店舗・施設名→ 施設名や地域名、あるいは包含される住所への置き換え
  - 建物内の位置（1F-XX会議室）→ 同上

まとめ

- 単独の位置でも自宅等のセンシティブであれば保護が必要
- 位置の軌跡（履歴）は様々なリスクを持つので保護が必要
- 上２項を中心とした要件を満たすことで、非識別・非特定情報が作成可能
- 保護手法には様々なものがあり、データと利用目的に応じて、適切なものを選ぶ
- 位置情報を保護と活用で最大限の成果を得るためには、匿名化の程度や規律をケースに応じて柔軟に合意形成することが合理的