攻撃者が存在する場合の多地域 Crowdsensing のワーカ最適サンプリング法の性能評価

Performance Evaluation of Optimum Worker Sampling in Crowdsensing with Multiple Areas under Attacks

松浦千紘[1]　上山憲昭[2]

立命館大学 情報理工学研究科[1]

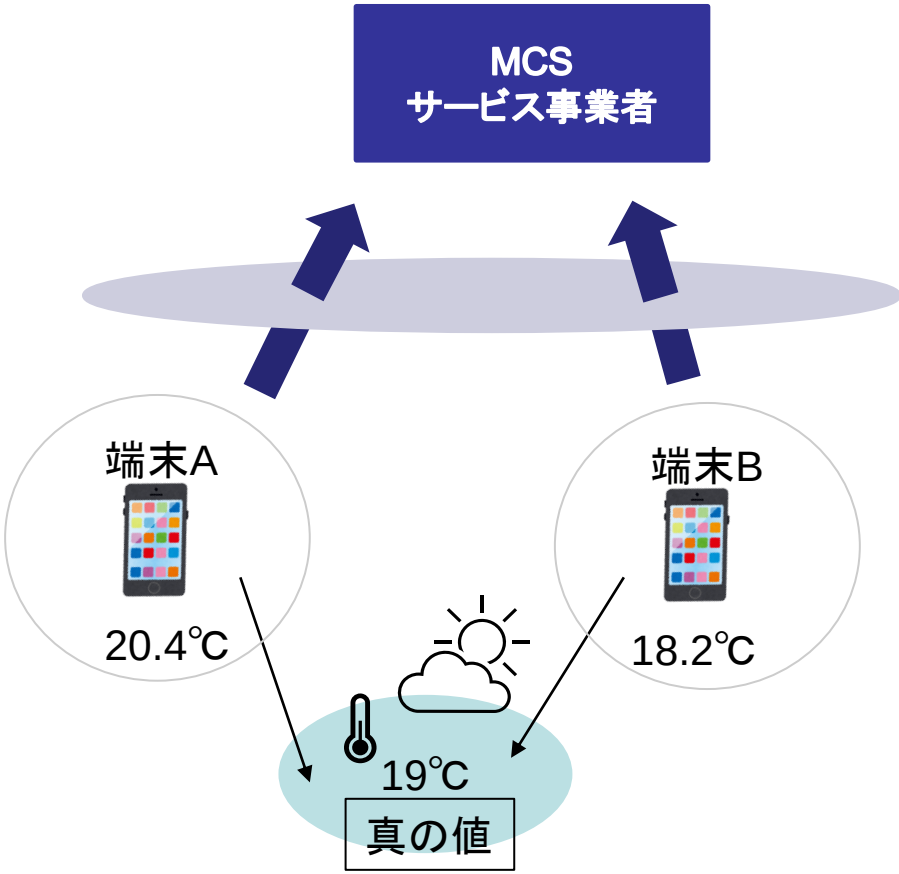
立命館大学 情報理工学部[2]

研究の背景

- モバイルクラウドセンシング（MCS：mobile crowdsensing）
  - モバイル端末をIoTデバイスとして活用

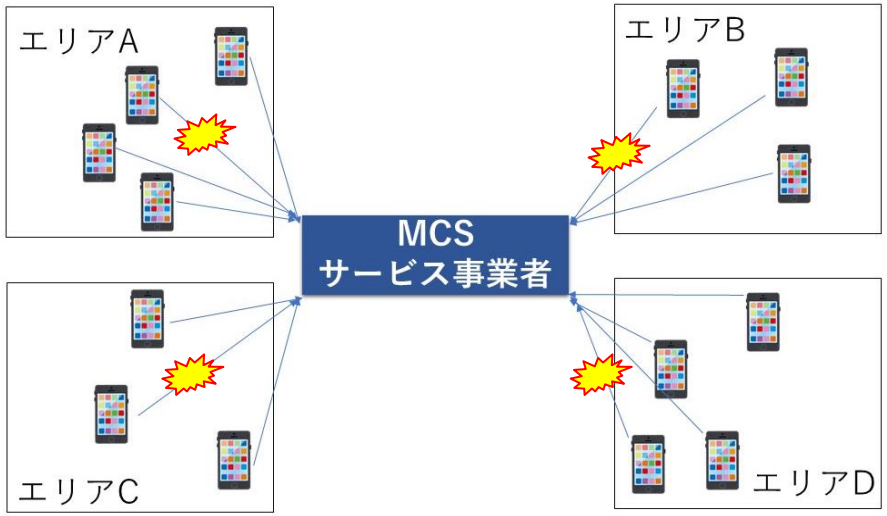
- MCSの利点
  - インフラの新規構築が不要なため低コスト
  - 高機能なセンサを搭載
  - 膨大なデータを収集可能

- MCSの問題点
  - 計測時の誤差の発生
    - センサの故障やヒューマンエラー
  - 悪意のあるワーカによる誤ったデータの送信
    → 推定値が歪むデータポイズニング攻撃の発生



関連研究

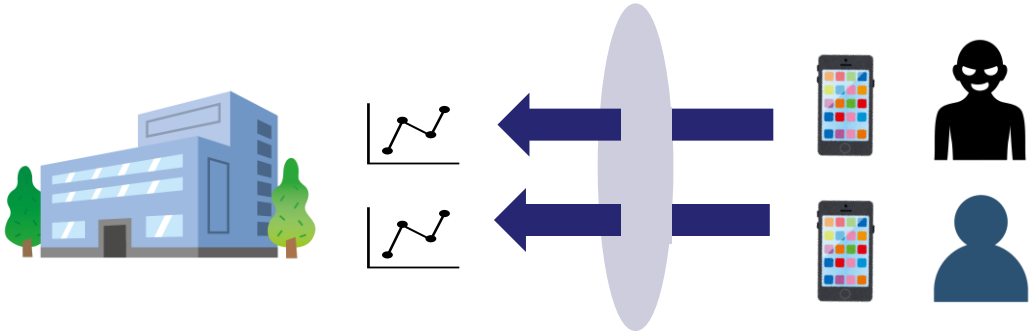
- ワーカ推定誤差を最小化するよう，各ワーカの測定値を重みづけした重みづけ平均で推定する CRH（Conflict Resolution on Heterogeneous data）法の提案 [1].
- 複数エリアごとに複数ワーカから測定値を推定する MCS において，攻撃ワーカが推定誤差を最大化するように各エリアの配置攻撃者数を最適化する方式の提案 [2].
- 複数エリアのMCSにおいて，全エリアの誤差総和の最小化を目的とする各エリアの最適サンプル数設定法の提案 [3].



[1] Q. Li, et al., Conflicts to Harmony: A Framework for Resolving Conflicts in Heterogeneous Data by Truth Discovery, IEEE Trans. Know. Data Eng., 28 (8), Aug. 2016
[2] R. Fujimoto and N. Kamiyama, Poisoning Attacks in Crowdsensing Over Multiple Areas, IEEE GLOBECOM 2022
[3] 松浦 千紘，上山 憲昭, 多地域 Crowdsensing におけるワー カ最適サンプリング, 信学会 NS 研究会, NS2022-218, 2023 年 3 月

研究の目的

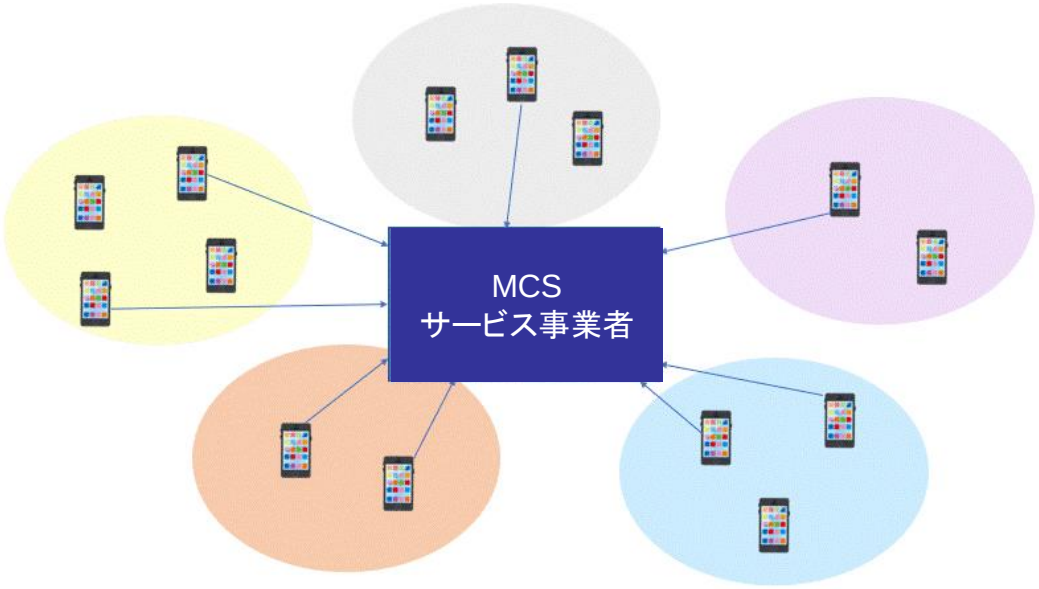
- 着目する課題
  - 既存の研究では，正常ワーカのみが存在する状態でのサンプリング方法を考案
  - データポイズニング攻撃は統計的に外れない値を設定するという性質を持つため，攻撃ワーカを特定することは現段階では難しい
    → ワーカの判別が行えない状態での防御法を検討する必要がある



攻撃者が存在する複数エリアのMCSにおいて，
全エリアの誤差を最小化するような各エリアの最適サンプル数設定法の提案

複数エリアMCS

- データ収集領域は複数の K 個のエリアから構成
- 各エリアにおいてセンシングデータをモバイル端末のユーザから収集し，エリアごとに測定値を推定
  - 市街を 100 メートルの正方領域に分割し，各エリアの二酸化炭素濃度を推定
- 攻撃ワーカは各エリアに一定数含まれている（母集団）
- サービス事業者が収集できる総サンプルワーカ数をNとしたときの，各エリアのサンプルワーカ数を決定する問題を考える



CRHアルゴリズム(Conflict Resolution on Heterogeneous data)

- 目的
  - 複数の測定値から真の値を推測する

- 概要
  - 真の値と測定値との差異が小さいワーカの信頼性は高く，大きいワーカの信頼性は低くなるように各ワーカの信頼性を設定
  - 信頼性を重みとした測定値の加重平均を推定値として用いる
  - 推定値と各ワーカの重みを交互に更新し，収束した値を推定値として扱う

- アルゴリズム
  1. 各ユーザ k の信頼性（重み）w_k を 1 に初期化
  2. 式(1)で，各ユーザ k の測定値 v_k と w_k から推定値を計算
  3. 式(2)でユーザ毎の信頼性 w_k を更新
  4. 推定値及び信頼性が収束するまで step 2, 3 を反復

$$w_k = -\log \frac{(v_k - \tilde{v})^2}{\sum_{k \in N \cup A} (v_k - \tilde{v})^2} \quad (1)$$

$$\tilde{v} = \frac{\sum_{k \in N \cup A} v_k w_k}{\sum_{k \in N \cup A} w_k} \quad (2)$$

（正常ワーカの集合を N，攻撃ワーカの集合を A とする）

DPAアルゴリズム(Data Poisoning Attack)

- 目的
  - 誤差を最大化するよう攻撃ワーカの報告値を更新

- 概要
  - 単一エリアでのみ適用可能なアルゴリズム
  - CRH 法による推定値及びユーザごとの信頼性の更新ステップ，攻撃者の報告値計算を交互に行い，収束した値を報告値として扱う

- アルゴリズム
  1. 各攻撃ワーカ k の報告値 v_k の初期化
  2. 正常ワーカのみでCRH法を用いて推定値を算出
  3. 全ワーカを対象にCRH法を用いて推定値を算出
  4. 各攻撃ワーカ k に対し，式(3)で報告値 v_k を更新
  5. v_k が収束するまで step 3, 4 を反復

$$v_k = v_k + 2 \times (\hat{v} - \tilde{v}) \times \frac{w_k}{\sum_{k \in N \cup A} w_k} \qquad (3)$$

提案方式(1/2)

- 目的
  - 総サンプルワーカ数の上限 N を制約条件として考慮し，攻撃ワーカが存在している場合に総誤差 E が最小となるよう各エリア i のサンプルワーカ数 u_i を最適設計

- 最適化問題
  - 正常ワーカのみで計算された推定値を $\tilde{v}$ ，攻撃ワーカ混入後に算出された推定値を $\hat{v}$ とし，各エリアの誤差を $|\tilde{v} - \hat{v}|$ と表す
    目的関数を（1）式のように立式（K:エリア数）

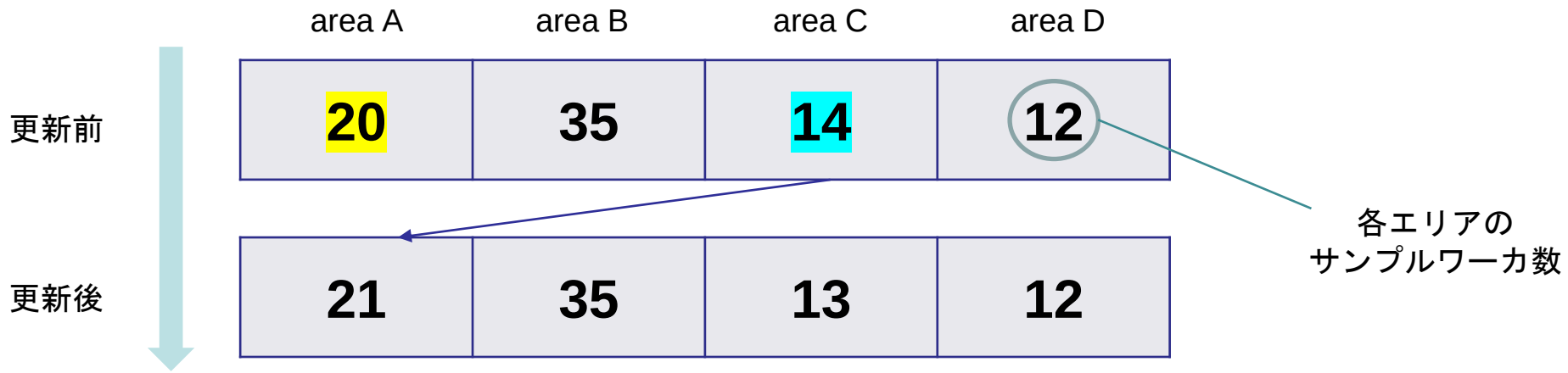
$$\min E(u_1, u_2, \cdots, u_K) = \sum_{i=1}^{K} (\tilde{v}_i - \hat{v}_i)^2 \qquad (1)$$

  - 制約条件を（2）式に示す

$$\sum_{i=1}^{K} u_i = N \qquad (2)$$

提案方式(2/2)

- アルゴリズムの概要
  1. 各エリアのサンプルワーカ数の初期値を $u_i = N/K$ に初期化し，このときの総誤差 E_{ini} を算出
  2. ランダムにサンプルワーカ数を与えたときの平均推定誤差を算出し，各エリアのサンプルワーカ数に対する平均推定誤差 $e_{i,ui}$ の近似解を得る（DBに格納）
  3. 各エリア i のサンプル人数をインクリメント（デクリメント）し，推定誤差の減少量 e_{dec} (増加量 e_{inc}) を算出
  4. 減少量が最大であるエリアのサンプル人数をインクリメント，増加量が最小であるエリアのサンプル人数をデクリメント
  5. 総誤差の変化量 $|E_{post} - E_{pre}|$ が閾値 η を下回るまでこれを反復し，このときの総誤差 E_{conv} を算出



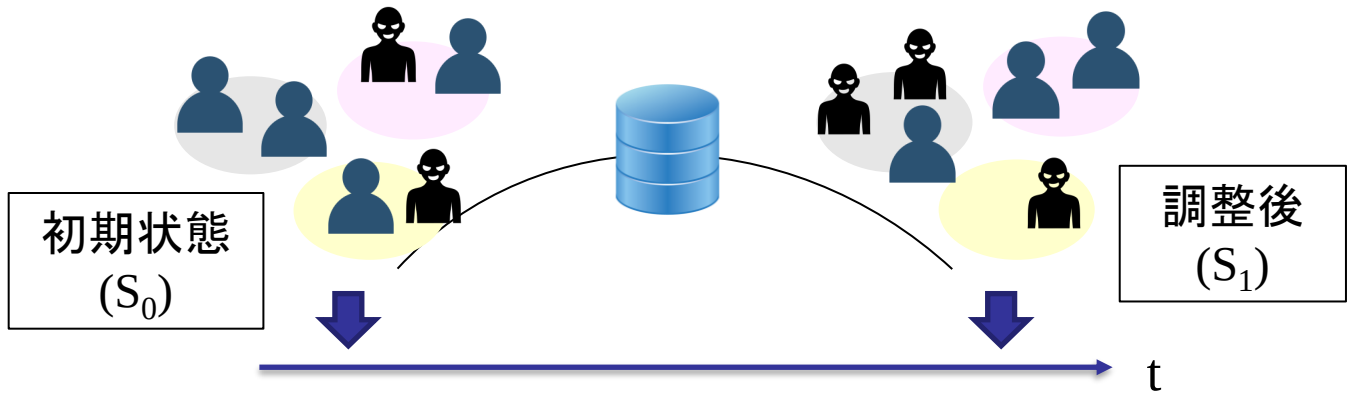
実験条件

数値条件

記号	定義	値
K	エリア数	10
N	総サンプルワーカ数	400
μ_i	各エリア i のワーカ報告値の平均値	50
σ_i	各エリア i のワーカ報告値の標準偏差	2, 3, 4, 5, 6, 7, 8, 9, 10, 11
m	攻撃ワーカの初期報告値	50
p	母集団の攻撃ワーカ割合	0.05
η	総誤差の差分の判定に用いる閾値	10^{-5}

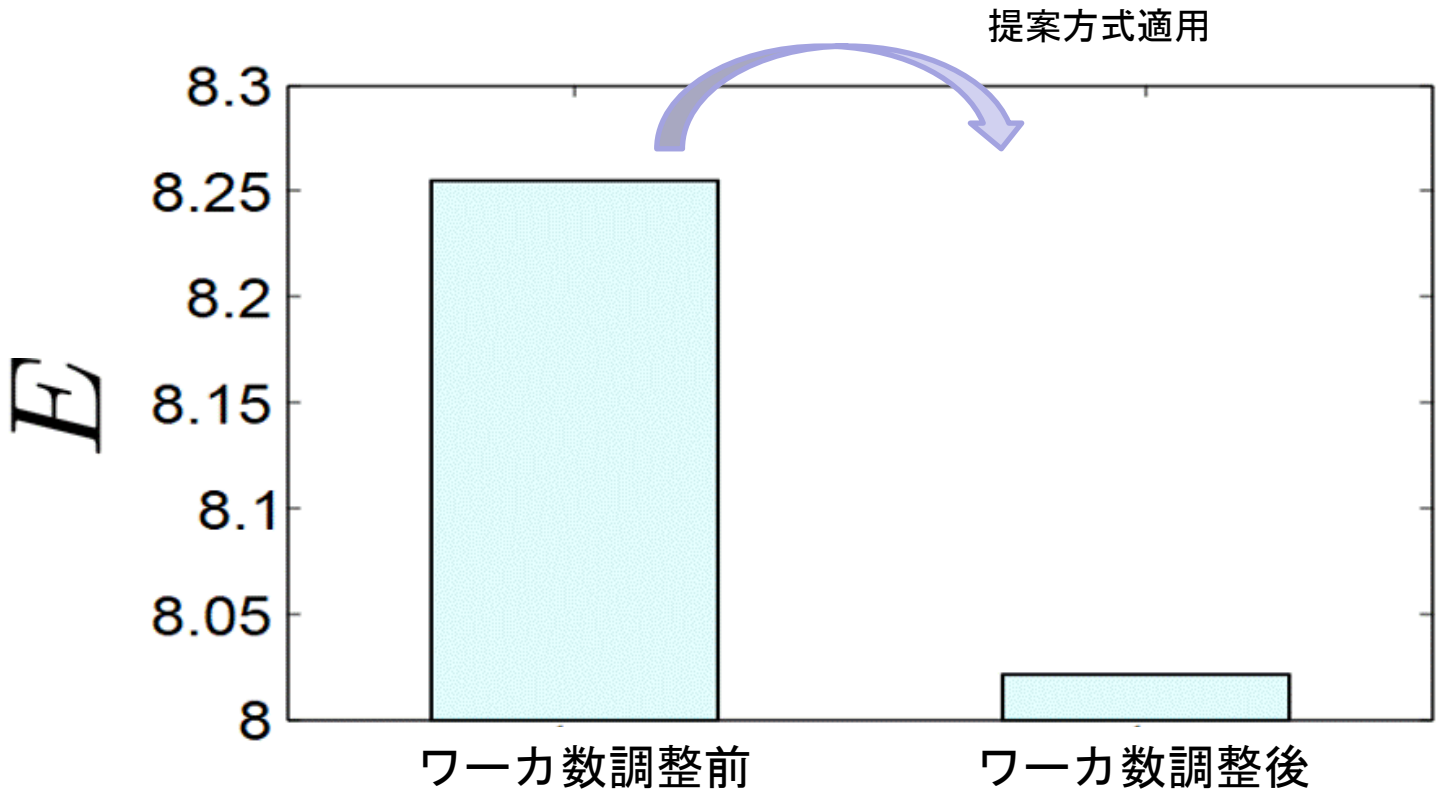
シミュレーション条件

- サービス事業者は時刻 t_n で推定値を計算



サービス事業者が推定値計算を行う時刻

評価結果



area	σ_i	u_i（状態S_1）
1	2	17
2	3	23
3	4	29
4	5	34
5	6	38
6	7	44
7	8	47
8	9	52
9	10	56
10	11	60

σ_i : 各エリア i のワーカ報告値の標準偏差
u_i : 各エリア i のサンプルワーカ数

- 時刻 t_0 から t_1 の区間にかけて，総誤差の減少を確認
- サービス事業者は収集データの区別ができない条件のもとでサンプリングを行っていたが，一定の割合で攻撃ワーカが混在している場合においても本提案方式は有効である

まとめ

- 本研究では，総推定誤差の最小化を目的として各エリアの最適サンプル数設定法を提案
  - 総サンプルワーカ数が固定であるという条件のもと複数エリアからワーカのデータを取集
  - 攻撃者が DPA 法により攻撃を行った場合においても，推定精度の劣化を回避できることを示した

- 今後の方針
  - 正常ワーカの持つ値やアルゴリズムが既知であるために，本研究で想定しているDPA法を用いた攻撃が発生
    → データ保護の観点でシステム構築を目指す