

複数地域の Crowdsensing への Data Poisoning 攻撃法

1. 研究背景

- MCS(モバイルクラウドセンシング)
 - モバイル機器をIoTデバイスとして活用
 - 今日のモバイル機器は様々なセンシング機能が搭載
 - 多くのユーザが取得したデータを収集可能

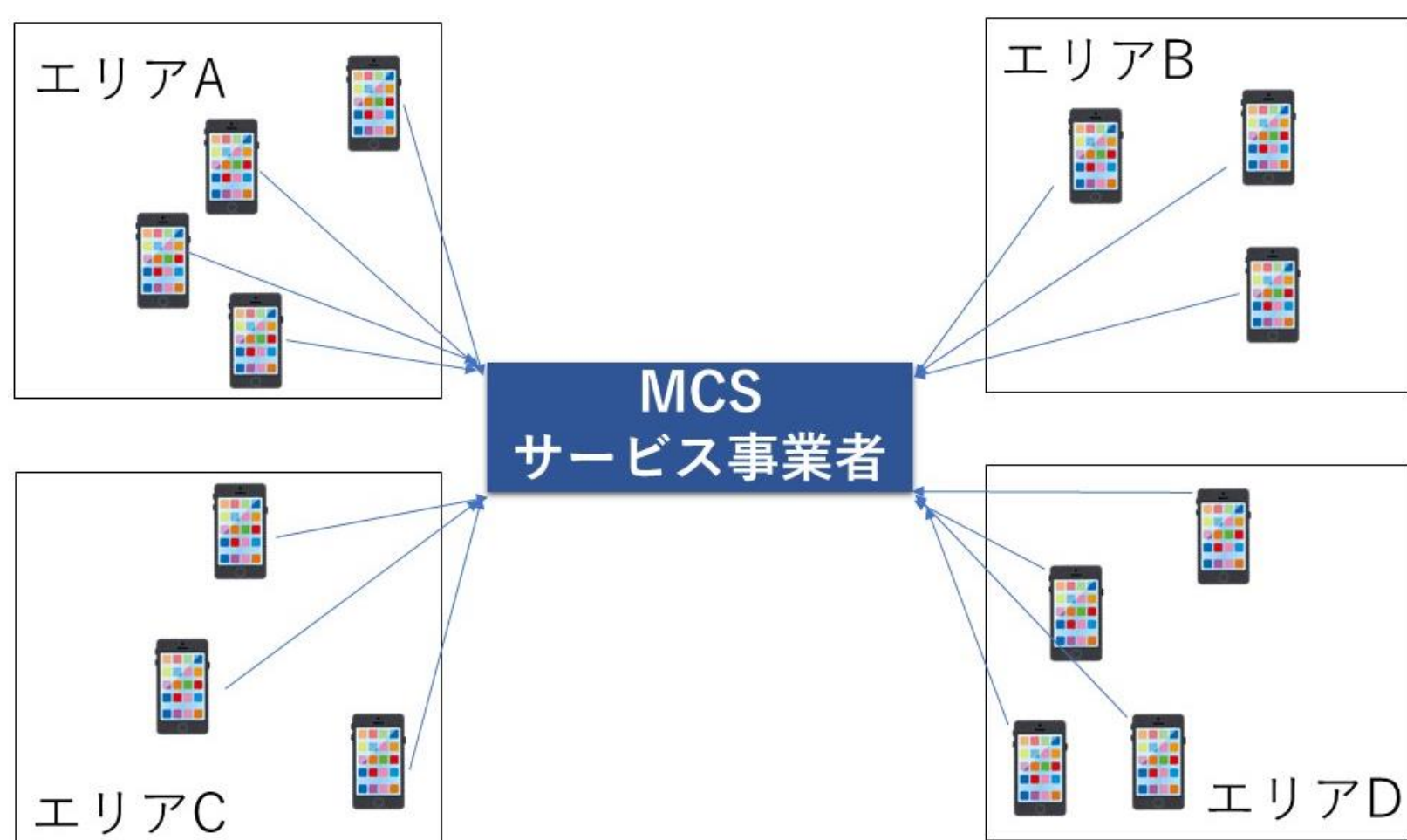


- MCSのメリット
 - インフラの新規構築が不要で低コスト
 - 従来のIoTデバイスより高機能
 - 普及率が高く膨大なデータを収集可能
- MCSのデメリット
 - センサーの不具合やヒューマンエラーによった誤ったデータの発生



意図的に誤差の大きなデータを送信することで推定値を歪ませる**データポイズニング攻撃**の問題が指摘

- 単一エリアで適用可能なアルゴリズム
 - MCSシステム側: CRH (Conflict Resolution on Heterogeneous data) 法
誤差が最小となるように推定
 - 攻撃者側: DPA (Data Poisoning Attack) 法
推定誤差を最大化するよう攻撃者の測定値を設定
- 複数エリアを対象としたCrowdsensing
 - 複数の地点からセンシングデータを集め、エリアごとに値を推定するサービス 例: 気象情報



2. 研究の目的

- 単一地点のMCSの攻撃法は考察されているが、複数地点の場合は未検討
- 本研究では複数地点のMCSに対する攻撃方式を提案
 - M人の攻撃者をN個のエリアに配置する問題を考察
 - **攻撃の効果を最大化する攻撃者の最適配置法を提案**

3. CRH法/DPA法

- CRH (Conflict Resolution on Heterogeneous data)
 - 目的: 複数のソースから得られた値から真値を推定
 - アルゴリズム

- (i) 各ユーザ k の信頼性 (重み) w_k を 1 に初期化
- (ii) 式 (1) で、各ユーザ k の報告値 v_k と w_k から、推定値 $\tilde{v}$ を計算

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

- (iii) 式 (2) でユーザごとの信頼性 w_k を更新

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

- (iv) $\tilde{v}$ 及び w_k が収束するまで (ii)(iii) を反復

- DPA (Data Poisoning Attack)

- 目的: 誤差を最大化するよう測定値を更新
- アルゴリズム

- (i) 各攻撃者 k の報告値 v_k の初期化
- (ii) 正常ユーザのみで CRH 法を用いて推定値 $\tilde{v}$ を算出
- (iii) 全ユーザを対象に CRH 法を用いて推定値 $\hat{v}$ を算出
- (iv) 各攻撃者 K に対し、式 (3) で報告値 v_k を更新

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

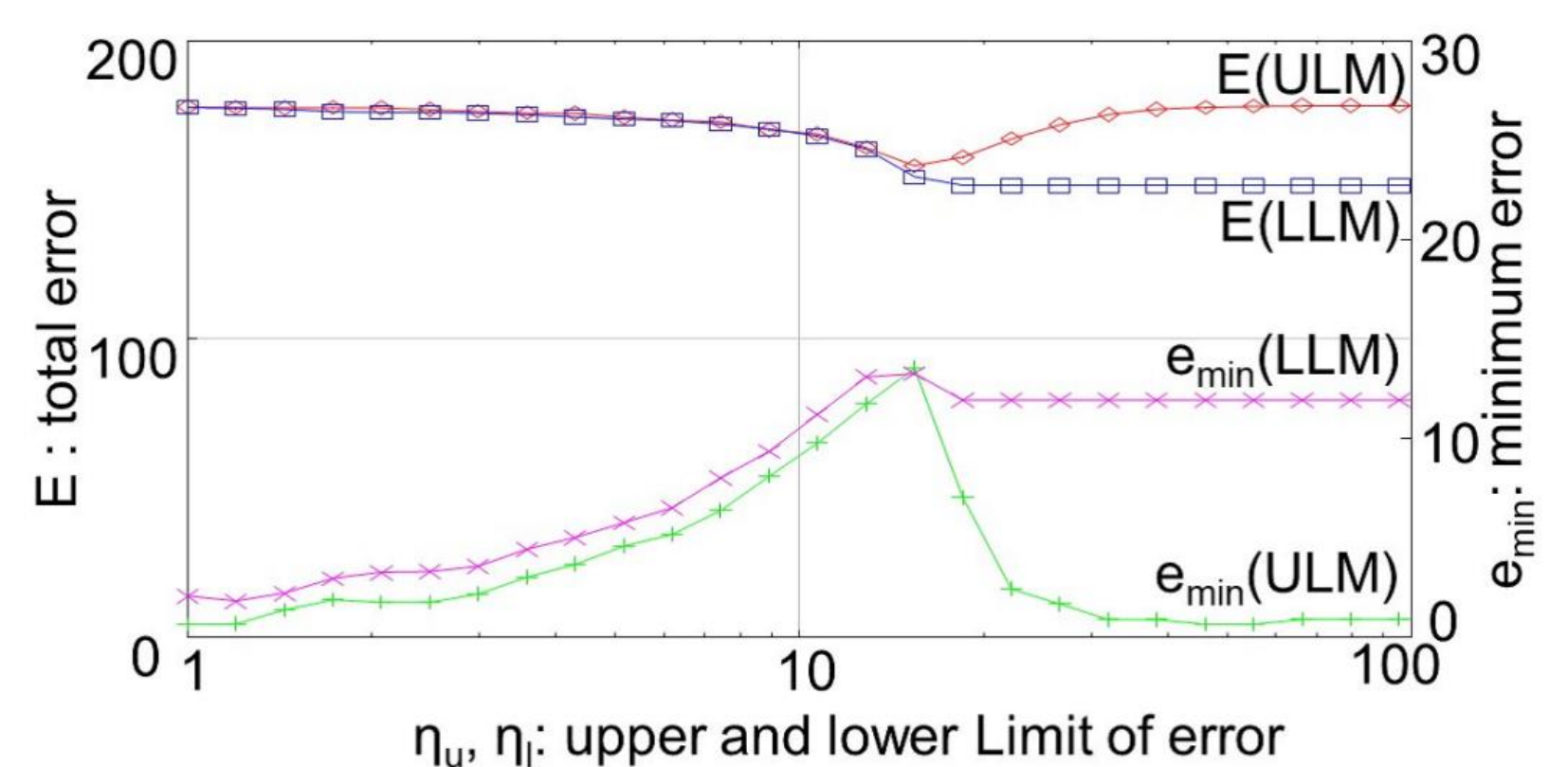
- (v) v_k が収束するまで (ii)~(iv) を反復

N: 正常ユーザの集合

A: 攻撃者の集合

4. 提案方式と評価結果

- 攻撃者の最適化目標
 - ① 全エリアの誤差の総和 E の最大化
 - ② 各エリアの誤差の最小値 $e_{\min}$ の最大化
- DPAアルゴリズムをベースに閾値を設けた2つの方式を提案
 - ① **誤差上限法 ULM (upper limit method)**
 - ・ 誤差の上限値を設定
 - ・ 誤差が上限値未満のエリアから**誤差増加量が最大**のエリアに攻撃者を一人ずつ配置
 - ・ 全エリアで上限値以上の場合は全エリアを対象
 - ② **誤差下限法 LLM (lower limit method)**
 - ・ 誤差の下限値を設定
 - ・ 誤差が下限値未満のエリアから**誤差が最小**のエリアに攻撃者を一人ずつ配置
 - ・ 全エリアで下限値以上の場合、全エリアを対象に誤差増加量が最大のエリアに配置
- 誤差の上限・下限値に対する総誤差と最小誤差
例: エリアの特性が同一である場合の評価結果



- ・ 最適化目標②の観点から下限値12のLLMが攻撃効果を最大化
- ・ エリアの特性が異なる場合も同様の傾向