Anomalous Operation Detection for Smart Home based on In-home Behaviors of Users

宅内のユーザ行動に基づくスマートホーム不正操作検出手法

情報科学研究科 情報ネットワーク学専攻
先進ネットワーク学講座 博士 3 年
山内 雅明

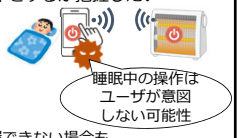
大阪大学 OSAKA UNIVERSITY

IoT 機器への不正操作攻撃

- **第三者による操作パケットの送信**
  - ユーザへの重大な被害
    - 例：ヒータの不正操作による火傷や火災
    - 例：スマートロックの解除操作による盗難被害
  - 高電力負荷家電の一斉操作による停電 [16]
- **不正操作パケットの検出が重要**

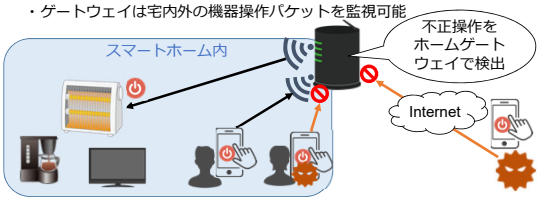
IoT 機器への不正操作パケット検出における課題

- **機器操作パケットは正常時も不正操作時も同様のトラヒック**
  - 乗っ取ったスマートフォン経由だと送信元 IP アドレスも同じ
  - ➢既存の侵入検知システムでは検知不能
- **ユーザが意図した操作かどうかを見分ける必要**
  - ユーザが普段どのようなタイミングで操作をするか把握したい
  - ➡宅内でのユーザの行動に着目
- **ユーザの行動をどうモデル化するか**
  - 複数のユーザに対応可能である必要
    - 多くの家庭にはユーザが複数居住
    - IP アドレスの情報から誰が操作したか把握できない場合も
      - AI スピーカや共有タブレットによる操作



研究目的とアプローチ

- [研究目的]
  - **IoT 機器に対する不正操作トラヒックの検出**
- [アプローチ]
  - **ユーザが機器を操作する際の行動パターンをもとに検出**
    - ユーザは家庭内の環境に応じて一定の行動
      - 例：部屋が寒い時、ヒータ ⇒ 加湿器と操作
  - **家庭内のゲートウェイでユーザの行動を学習し、学習内容と合致しない不正操作パケットを検出して遮断**
    - ゲートウェイは宅内外の機器操作パケットを監視可能



ユーザの行動学習におけるデータ量に関する問題

- **十分な行動データを集めるために時間がかかる**
  - 行動パターンの網羅にはデータ数が必要だが 1 家庭から得られる行動データは限られている
  - 時間帯、曜日、季節によって行動が変化
    - 例：季節の変わり目に新家電を導入すると新機器に関するデータが不足

対策

同じような生活習慣・行動をする他家庭が協力

新たな問題点

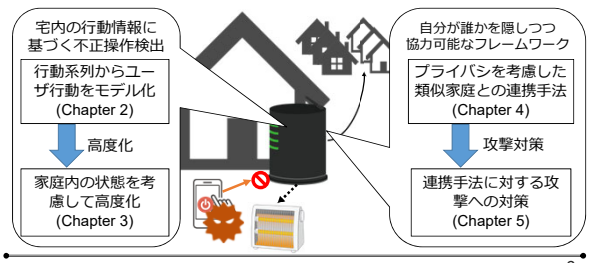
- **攻撃者が連携に参加して異常操作攻撃のヒントとなる情報を得ようとする**

対策

自分が誰かという情報を隠して、似た行動をとるユーザのデータを利用

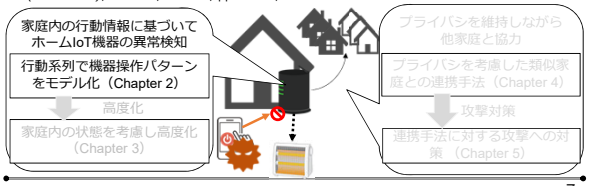
博士論文の構成

- **宅内の行動データを用いたホーム IoT 機器の不正操作検出**
  - データが十分集まれば手法の高度化
- **データ量が不十分で行動が網羅できなければ他家庭と連携**
  - プライバシを維持しながら似た行動をとる家庭と連携



Chapter 2: Anomaly Detection Method using User Behavior Sequences

1. Masaki Yamauchi, Yuichi Ohsita, Masayuki Murata, Kensuke Ueda, and Yoshiaki Kato, "Anomaly Detection in Smart Home Operation from User Behaviors and Home Conditions," *IEEE Transactions on Consumer Electronics*, vol. 66, no. 2, pp. 183-192, May 2020.
2. Masaki Yamauchi, Yuichi Ohsita, Masayuki Murata, Kensuke Ueda, and Yoshiaki Kato, "Anomaly Detection for Smart Home Based on User Behavior," in *Proceedings of 37th IEEE International Conference on Consumer Electronics 2019 (ICCE 2019)*, pp. 1-10, January 2019.
3. Masaki Yamauchi, Yuichi Ohsita, Masayuki Murata, Kensuke Ueda, and Yoshiaki Kato,"[Invited Talk] Anomaly Detection for Smart Home Based on User Behavior -ICCE2019 Report-," *Technical Reports of ITE (ME2019-46)*, vol. 43, no. 5, pp. 249-254, February 2019.
4. Masaki Yamauchi, Yuichi Ohsita, Masayuki Murata, Kensuke Ueda, and Yoshiaki Kato, "Anomaly Detection for Smart Home IoT based on Users' Behavior," *Technical Reports of IEICE (IN2017-58)*, vol. 117, no. 353, pp. 73-78, December 2017.



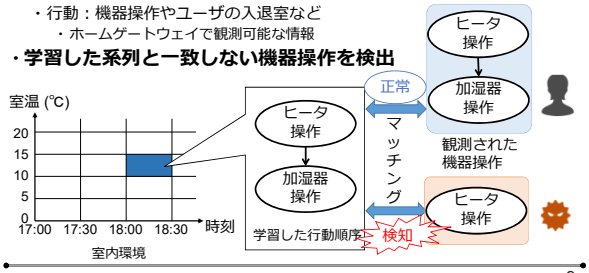
Chapter 2 の研究目的とアプローチ

- [研究目的]
  - **正常なユーザの行動パターンに基づいた不正操作攻撃の検出**
- [アプローチ]
  - **機器が操作される室内環境と行動順序を学習**
    - 室内環境：時刻や室温などの家庭内での観測値
      - 本章では時刻情報のみ利用
    - 行動順序：機器操作や入退室が行われる系列
    - 学習結果と一致しない機器操作を異常として検知

例：19 時に帰宅し室温が 10℃ 以下の室内環境では、
①ヒータ ⇒ ②加湿器という行動順序で操作

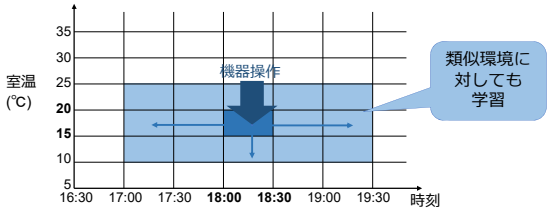
行動パターンの学習と検出

- **室内環境：時刻や室温の表をもとに定義**
  - 時刻やセンサの観測値を軸に設定
  - 一定間隔に分割し、セルを生成
    - 各セルが各環境に相当
- **行動順序：室内環境ごとに機器操作の順番を蓄積**
  - 行動：機器操作やユーザの入退室など
    - ホームゲートウェイで観測可能な情報
- **学習した系列と一致しない機器操作を検出**



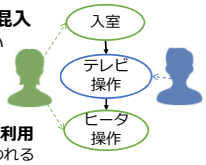
室内環境学習における問題点とアプローチ

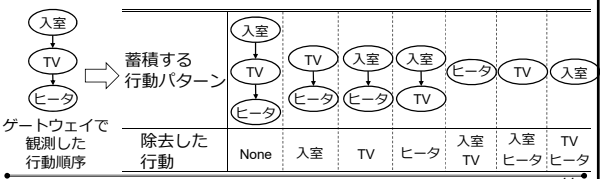
- [問題点]
  - **各室内環境における機器操作回数が非常に少ない**
    - ユーザが機器を操作する回数が限られている
      - 1 日に家で機器を操作することは、数十回程度
- [アプローチ]
  - **類似した室内環境に対しても同時に学習**
    - ユーザは似たような室内環境下で同じ行動を行う



行動順序学習における問題点とアプローチ

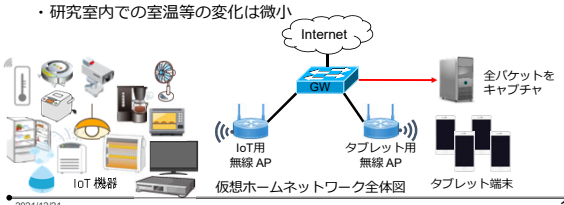
- [問題点]
  - **他のユーザによる操作がノイズとして混入**
    - 家庭内にはユーザが複数いることが多い
- [アプローチ]
  - **複数の系列を同時に蓄積**
    - ユーザが実際に行う行動パターンが蓄積可能
  - **複数回蓄積された行動パターンのみを検知に利用**
    - ユーザが実際に行う行動パターンは何度も行われる





実験環境

- **研究室内に仮想ホームネットワークを構築**
  - IoT 家電：16 種類
    - テレビ、冷蔵庫、自動掃除機、空気清浄機、扇風機、冷暖房機、加湿器、ヒータ、コーヒーメーカ、監視カメラ、可視化デバイス、HDDレコーダ、電子レンジ、自動調理器、移動型空気清浄機、照明器具
  - IoT センサ：10 種類
    - 外気温、室温、湿度、気圧、騒音、CO2濃度、VOC、PM2.5、雨量、風速
  - 実験期間：3 か月間
  - 被験者：学生 4 名
- **パケット情報から機器操作時刻と被験者の入退室時刻を記録**
- **室内環境は時刻の情報のみから分類**
  - 研究室内での室温等の変化は微小



評価方法

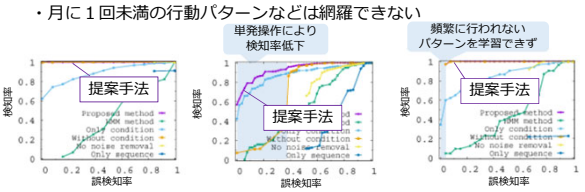
- **データセット**
  - 機器操作および入退室の履歴
  - 正常操作：被験者による機器操作履歴
  - 不正操作：偽の操作履歴をランダムな時刻に 100 回 (/ 日) 分混入
- **パラメータ**
  - 各パラメータを変化させ、それぞれ検知率と誤検知率を算出
- **評価手法**
  - LOO-CV (Leave-One-Out Cross-Validation)
    - 学習データ：特定の 1 日分を除くデータ
    - テストデータ：特定の 1 日分のデータに不正操作を混入
    - 各日の結果を合算し検知率、誤検知率を算出
- **評価指標**

$$\text{検知率} = \frac{\text{検知した不正操作数}}{\text{混入した不正操作数}}$$

$$\text{誤検知率} = \frac{\text{誤検知した正常操作数}}{\text{正常操作の総数}}$$

評価結果

- **行動順序の学習により不正操作を検出可能**
  - 誤検知 10% 未満で 90% 以上の不正操作を検知可能
- **単体で操作される機器操作（単発操作）の検出は困難**
  - 行動順序の情報が使えないため
    - 現在の室内環境の利用方法では不十分
- **頻繁に行われない機器操作パターンは検出できない**
  - 月に 1 回未満の行動パターンなどは網羅できない



テレビの不正操作の検知結果　コーヒーメーカの不正操作の検知結果　扇風機の不正操作の検知結果

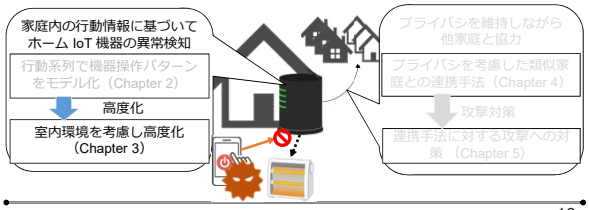
Chapter 2 のまとめ

- **ユーザの行動を基にホーム IoT 機器の不正操作を検知可能**
  - 誤検知 10% 未満で 90% 以上の不正操作を検知可能
  - **「室内環境」**と**「行動順序」**を用いてユーザの行動を学習
    - 特に、**「行動順序」**による影響が大きい
- **課題①：機器単体で操作される機器の異常検知は難しい**
  - 行動順序の情報が利用できないため
  - 「室内環境」を詳細に分析して対策 (Chapter 3)
- **課題②：頻繁に行われない機器操作パターンは検知できない**
  - 学習データが網羅できなかった
  - 似た特徴を持つ家庭と協力 (Chapter 4, 5)

日経新聞 1 面掲載、
招待講演 2 件、
Chester Sall Award

Chapter 3: Improving Anomaly Detection Method by Estimating In-home Situation

1. Masaaki Yamauchi, Masahiro Tanaka, Yuichi Ohsita, Masayuki Murata, Kensuke Ueda, and Yoshiaki Kato, "Smart-home anomaly detection using combination of in-home situation and user behavior," *submitted for publication*, pp. 1-13, September 2021.
2. Masaaki Yamauchi, Masahiro Tanaka, Yuichi Ohsita, Masayuki Murata, Kensuke Ueda, and Yoshiaki Kato, "Smart-home anomaly detection using combination of in-home situation and user behavior," CoRR, abs/2109.14348, pp. 1-13, September 2021.
3. Masaaki Yamauchi, Masahiro Tanaka, Yuichi Ohsita, Masayuki Murata, Kensuke Ueda, and Yoshiaki Kato, "Modeling Home IoT Traffic using Users' in-Home Activities for Detection of Anomalous Operations," in *Proceedings of 32nd International Teletraffic Congress (ITC32) - PhD workshop*, pp. 1-2, September 2020.



Chapter 3 の背景・目的・アプローチ

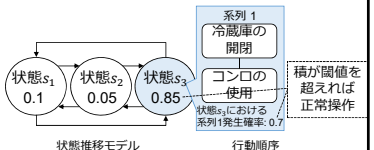
- [背景]
  - **機器単体で操作される機器操作（単発操作）の学習が困難**
    - 行動順序の情報が利用できないため
  - **機器操作がされそうかどうかを時刻のみでは決定できない場合も**
    - 例：コンロの操作
      - 家族が夕飯の準備中：頻繁に使用
      - 家族が全員外出中　：使用されない
- [目的]
  - **室内環境に代わり、宅内の状態を推定することで検知精度を向上**
- [アプローチ]
  - **機器操作履歴とセンサデータから宅内の状態の遷移を推定し、各状態において発生しやすい行動系列を学習**
    1. 宅内の状態遷移モデルを生成
    2. 各状態において行動系列が発生する確率を算出
    3. 発生する確率が低い機器操作を不正操作として検出

宅内の状態遷移推定を用いた異常検知手法

- **機器操作とセンサデータから推定する状態のラベル名を事前定義**
  - ユーザおよび機器の状態のかけあわせ
    - ユーザの状態（睡眠中、外出中、活動中）
      - 騒音やCO2、時間帯といったセンサデータ等から推測
    - 機器の状態（使用中、T_X 分以内に使用、使用後、その他）
      - 不正操作検出の対象機器やそれ以外の機器操作から推測
- **状態 s である確率 α と各状態における行動系列発生確率 b' を算出**
  - 状態遷移モデル
    - 時間帯に応じた状態遷移
    - 機器操作およびセンサの観測値を用いたベイズ推定による補正
  - 行動系列発生確率
    - 各状態において、ある行動系列が発生する確率
- **現在の状態確率 α と系列の発生確率 b' の積から、閾値ベースで不正操作かどうか判断**

宅内状態の表

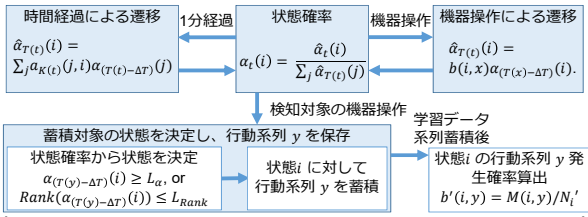
状態ラベル	活動中	睡眠中	外出中
機器使用前	s_1	s_2	s_3
機器使用中	s_4	---	---
機器使用後	s_5	s_6	s_7
その他	s_8	s_9	s_{10}



モデルの学習方法

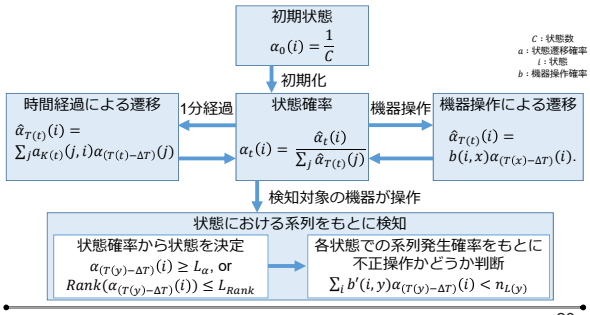
1. 状態遷移確率と各状態における機器操作確率を算出
- 学習データにラベル付けられた「状態」をもとに算出
  - 状態遷移確率 $a_t(s_j, s_i)$：時間帯 t において状態 s_i から s_j に遷移する確率
    - 時間帯ごとの行動変化を考慮し、周辺 T_z 分間の状態遷移を算出に利用
  - 機器操作確率 $b(i, x)$：状態 i において機器操作 x が行われる確率

2. 状態に対して行動順序を保存
- 状態は閾値 L_α 以上もしくは上位 L_{Rank} 以内の確率のものを利用
- 行動系列の蓄積後、状態 i における行動系列 y の発生確率 $b'(i, y)$ を算出



検知方法

・検知対象の機器操作が発生した時刻の状態確率と、行動系列発生確率を算出し、閾値をもとに不正操作かどうか判断



評価方法

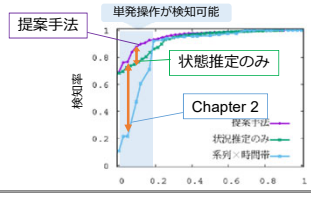
- **評価用データ**
  - 実家庭での行動履歴とセンサ観測値を収集
    - センサ：室温、湿度、騒音、気圧、CO2
  - 1 か月間のデータを 20 データセット
- **検知対象：コンロ**
  - 日常的に使われる機器で、不正に操作されると危険なもの
- **状態ラベル**
  - 機器の状態の定義：「料理中」「料理前」「料理後」「その他」
    - コンロ、電子レンジ、トースター、炊飯器が使用された場合も「料理中」と定義
- **評価方法**
  - テストデータに 100 回分 (/ 日) の不正操作を混入
  - LOO-CV (Leave-One-Out Cross-Validation)
    - 学習データ：特定の1日分を除くデータ
    - テストデータ：特定の 1 日分
    - 各日の結果を合算
- **評価指標**
  - 検知率 $= \frac{\text{検知した不正操作数}}{\text{混入した不正操作の総数}}$
  - 誤検知率 $= \frac{\text{誤検知した正常操作数}}{\text{正常操作の総数}}$

比較手法

使用する情報	提案手法	系列×時間帯 (Chapter 2)	状態推定のみ
行動順序	✓	✓	
状態推定	✓		✓
室内環境 (時間帯)		✓	

評価結果

・状態推定と行動系列を組み合わせることで、正常な単発操作を把握でき、より多くの不正操作を検知可能
- 時間帯の情報から室内環境を分割する手法よりも検知率が改善
  - 特に、誤検知率が 20% 未満の範囲で、検知率が大幅に改善
  - 状態推定により、中長期の家庭内での状態変化を把握
- 行動順序の学習によって検知精度が向上
  - 誤検知率 10% 未満での最高検知率が 74.7% から 87.7% に向上
  - 行動順序を利用することで、非調理機器との関連性を学習可能
    - 短期間の家庭内の状態変化は行動系列から把握



A_3 の検知結果

Chapter 3 のまとめ

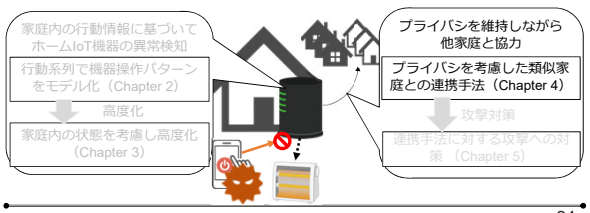
・宅内状態の推定と行動系列の学習を用いた異常検知手法
- 機器操作とセンサデータから宅内状態遷移をモデル化
- 各状態での行動順序発生確率を学習
- 現在の観測値から現在の宅内の状態を推定し、推定された宅内状態において行動系列が発生する確率から不正操作の検知

・状態推定と行動順序の学習を組み合わせることでより多くの不正操作を検知可能
- 中長期的なユーザの行動を状態推定によって把握可能
  - 状態推定を用いない方法と比較して、誤検知 10% 未満の範囲で最大検知率が 47.1% から 87.7% に検知率が向上
  - 前手法に比べて、正当とみなされる時間的範囲を狭めた
- 短期的なユーザの行動を行動系列から把握
  - 状態推定のみを利用した場合と比べ 74.7 %から 87.7% に検知率が向上

・状態推定により単発操作を検知可能

Chapter 4: Privacy-Preserved Cooperation Framework for Anomaly Detection without using Private Information

1. Masaaki Yamauchi, Yuichi Ohsita, and Masayuki Murata, "Platform Utilizing Similar Users' Data to Detect Anomalous Operation of Home IoT without Sharing Private Information," *IEEE Access*, vol. 9, pp. 130615-130626, September 2021.
2. Masaaki Yamauchi, Yuichi Ohsita, and Masayuki Murata, ``Platform Utilizing Others' Behavior Data to Detect Anomalous Operation Hiding Private Information,'' in *Proceedings of 7th IEEE International Conference on Consumer Electronics - Taiwan*, pp. 1-2, September 2020.
3. Masaaki Yamauchi, Yuichi Ohsita, and Masayuki Murata, "Framework to Utilize Others' Behavior without Sharing Privacy Information," *Technical Reports of IEICE (IN2019-114)*, vol. 119, no. 461, pp. 213-218, March 2020.



Chapter 4 の背景・目的

- **[背景]**
  - 機器操作の学習にデータが不足すると検知精度が低下
    - 行動パターンを網羅できない
  - 新たに導入された機器、使用頻度の低い機器の異常検知が困難

⇩

- 似た行動をとるユーザの行動データを利用
  ➡ただし、異常操作の攻撃者が連携パケットを監視することで、異常操作攻撃のヒントとなりうる

- **[目的]**
  - 自分が誰かという情報を隠しつつ、似たユーザの行動データを相互に利用可能なフレームワークを考案

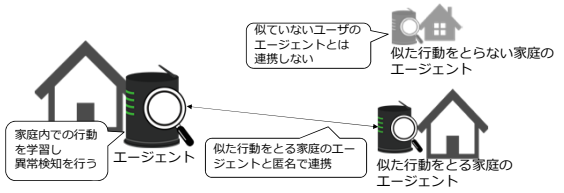
似た行動様式のユーザ

協力者のプライバシを保護

要件定義

1. ユーザ個人を特定する情報を使用しない
- 利用者とエージェントに識別子を割り当てない
  - エージェント：家庭内での行動を学習し異常検知を行う機器
- 利用者の個人情報を共有しない
  - 利用者の過去の全行動や，年齢，性別，職業など

2. 生活習慣の異なる家庭のエージェントとの連携を避ける
- 異常な操作の検出が不正確になるため



複数人のデータを活用する既存手法と今回の要件

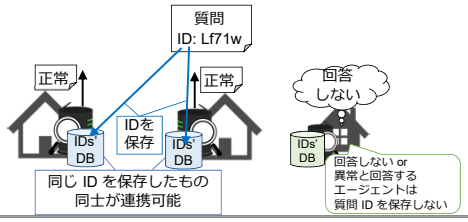
- **既存の手法では要件を満たすことは困難**
  - ユーザ個人を特定する情報を収集することはできない
  - 一般化されたモデルは各家庭のライフスタイルに一致しない

要件	クラウド集約 [27]	差分プライバシ [56] を用いたクラウド集約	連合学習 [26]
要件１：ユーザ個人を特定する情報を使用しない		✓	✓
要件２：生活習慣の異なるユーザのエージェントとの連携を避ける	✓		

特徴	クラウド集約 [27]	差分プライバシ [56] を用いたクラウド集約	連合学習 [26]
データの集約方法	クラウド集約	クラウド集約	分散
共有する情報	生データ	ノイズ付きデータ	学習モデルの勾配
生成されるモデルの特徴	個別化モデル	一般化モデル	一般化モデル

アプローチ

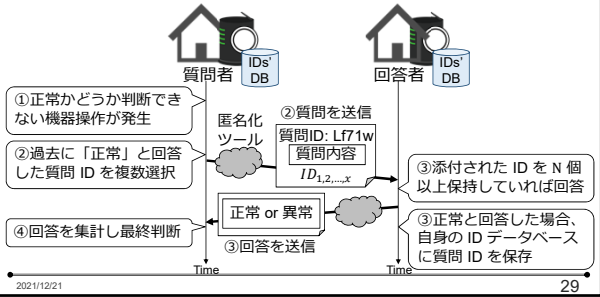
- **エージェントが得たい情報について匿名で質問および回答**
  - 「正常」か「異常」で回答可能な質問
  - 質問には識別子（ID）が発行
- **<u>過去に同じ質問に、同じ回答をしたもの同士が協力</u>**
  - ユーザを識別するのではなく、質問に識別子を付与して類似度を判定
  - 過去に「正常」と回答した質問の ID を保存
  - 質問時に保存した過去の質問の ID を添付
    - 質問に添付された ID を保存しているエージェントのみが回答



提案手法

① 家庭内のデータから正常かどうか判断できない機器操作が発生
② 過去に「正常」と回答した質問 ID を添付し、他家庭に質問
③ 自身も保存している ID が N 個以上添付されていれば回答
- 「正常」と回答した場合は、質問の ID を保存

④ 回答結果を集計して当該操作の正常／異常を判断



要件定義を満たすかどうか

1. ユーザ個人を特定する情報を使用しない
- 利用者とエージェントに識別子を割り当てない
- エージェント：家庭内での行動を学習し異常検知を行う機器

⇩

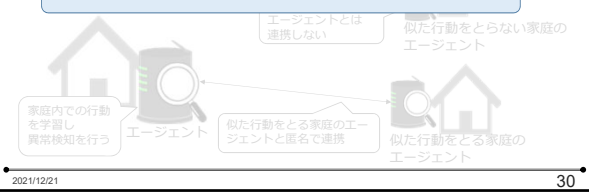
ユーザの識別子、宅内での行動を共有不要なため達成

2. 生活習慣の異なる家庭のエージェントとの連携を避ける
- 異常な操作の検出が不正確になるため

⇩

実際に似た生活習慣の家庭と連携できているかを確認する必要

評価が必要

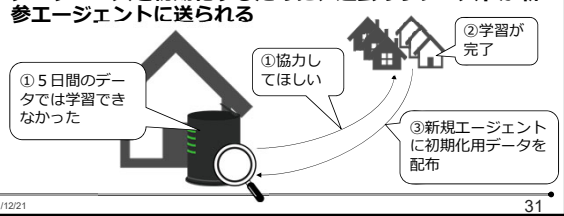


評価シナリオ

① **あるエージェントが 5 日間学習を行ったが、十分なデータを集められなかった**
- 不正操作検出のために、フレームワークを利用して他のユーザの行動データを利用したい

② **フレームワークにはすでに学習が完了しているエージェントが参加**

③ **フレームワークに初めて参加するエージェントには、ID データベースを初期化するために、過去のリクエストが新参エージェントに送られる**



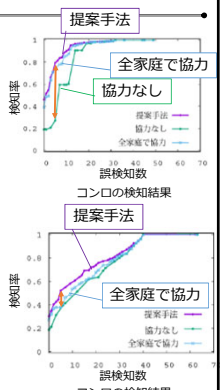
評価方法

- **実家庭の行動データ 1 か月分を 1 家庭分のデータとして利用**
  - 家庭Aで10 か月間、家庭Bで11 か月間データ収集
- **評価手順**
  1. 協力する20エージェント：実家庭データ1か月分を用いて各々学習
  2. 新規エージェント：月の最初の 5 日分のみ配布し学習
  3. 各月の前半2週間分のデータを利用して ID データベースを初期化
  4. 新規エージェント：月の後半 2 週間分のデータで検知テスト
  5. 各月の前後半のデータを入れ替えてクロスバリデーション
- **評価指標**
  - ユーザが実際に行った機器操作 (正常操作) の誤検知数を算出
  - 1 日あたり 100 回分の不正操作を混入し、検知率を算出
- **評価ケース**
  - 全家庭が系列情報を利用できるケース
  - 時間帯の情報のみが利用できるケース
- **パラメータを変えながら各結果を収集**
- **検知対象：コンロ、扇風機**

比較手法	提案手法	全家庭で協力	協力なし
他家庭と協力	✓	✓	
IDを用いた類似度の判定	✓		---

評価結果

- [行動順序の情報が利用可能なケース]
  - **他家庭との協力によって検知精度が向上**
    - 学習できていない行動順序を網羅
    - 誤検知 5 回の場合検知率が 27.0%→79.3%
  - **全家庭のデータを利用した場合と検知精度が同等**
    - 系列情報が非常に有用かつ同様の機器が実家庭A、B内に存在したため
- [時間帯の情報のみ利用可能なケース]
  - **各家庭によって行動時間帯が異なり、似た家庭間で連携することで検知精度が向上**
    - データを収集した実家庭A、B家間で機器操作をする時間帯の差が出ている
    - 本フレームワークによって似たもの同士が匿名で協力可能であったため

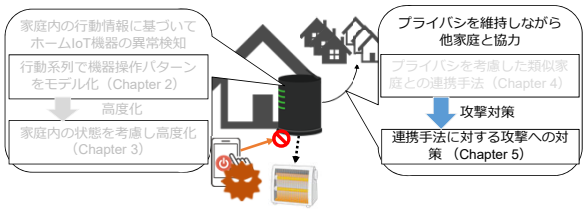


Chapter 4 のまとめ

- **似た特徴をもつユーザ同士が匿名で連携するフレームワーク**
  - 過去に同じ質問に「正常」と回答したエージェント同士が連携
    - 得たい情報を「正常」「異常」で回答可能な質問で匿名送信
    - 質問に過去に「正常」と回答した質問の ID を複数添付
    - 同様の ID を保存するものが匿名で回答
- **過去の回答情報をもとに匿名でも似たユーザの情報が利用可能**
  - 似た家庭と連携することによって異常検知精度が向上
    - 似ていない家庭と連携すると検知精度が悪化
- **課題：攻撃者が参画した場合の悪影響**
  - 全エージェントが提案手法に従うことが前提
  - 偽質問や偽回答によってエージェントが誤った情報を得る可能性
  - ➢似たユーザでないと分からない情報を利用し対策 （Chapter 5）

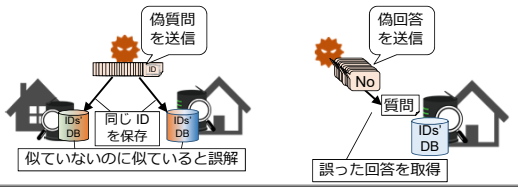
Chapter 5: Improving Attack Tolerance of Privacy-Preserved Cooperation Framework

1. Masaaki Yamauchi, Yuichi Ohsita, Masayuki Murata, “PPCwC: Improving attack tolerance method of privacy-preserved-cooperation framework,” *submitted for publication*, pp. 1—14, December 2021.



Chapter 5 の背景・目的

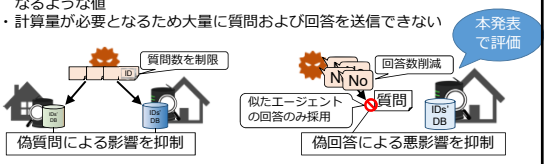
- [背景]
  - **大量に偽情報が送信されるとフレームワークから誤った情報が拡散**
    - 偽の質問が送信され、偽質問の ID が保存されると類似度判断に悪影響
    - 偽の回答が送信され、偽回答が集計されると集計結果が改ざん
      - 異常検知の判断結果を誤り、誤検知や検知漏れが発生
- [目的]
  - **偽質問、偽回答による悪影響を抑制**
    - ただし、匿名で類似エージェントと協力する利点を維持



アプローチ

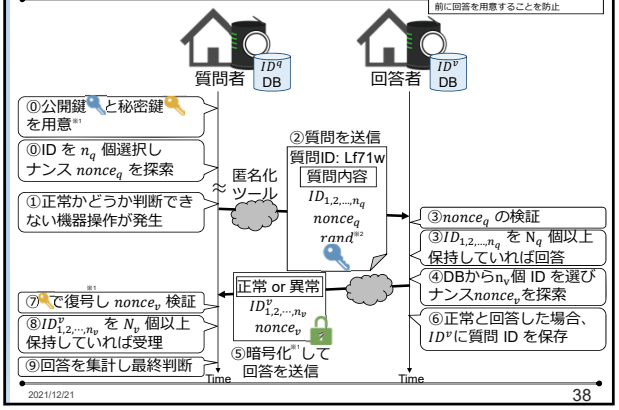
- **生活習慣が似た家庭のエージェントのみが答えられる問題に答えさせる**
  - 攻撃者による偽回答を採用しないようにする
    - 全ての質問、回答を受理してしまうと攻撃によって悪影響
  - 過去に「正常」と回答した質問 ID を回答にも利用
    - 似ているエージェントから送信されたものかどうかを判断

  ➢大量に送信されると偶然正解する可能性
- **質問および回答の送信数を制限**
  - 添付する ID に対するナンスの探索を要求
    - ナンス：添付する ID に追加されることで、そのハッシュ値が閾値以下となるような値
    - 計算量が必要となるため大量に質問および回答を送信できない



提案手法

※1：暗号化により他のエージェントの回答を見て真似た回答を送ることを防止
※2 $rand$：ランダム文字列を送信し回答時にハッシュ計算に含めさせることで事前に回答を用意することを防止



評価：受け取った回答のうち正規の回答が占める割合の導出

- t回のハッシュ計算と回答送信にかかる時間を t とする。
- **対策手法を導入した場合、**
  - 時間 t に関して受信する全回答数
    - $V(t, c_l, y) = V_{legitimate}(t, c_l) + V_{attack}(t, y)$
    - 正規エージェントによる回答数 $V_{legitimate}(t, c_l) = p_l \cdot c_l p_n \sum_{k=1}^{t}(1-p_n)^{k-1}$
      - 正規回答が受理される確率
      - 正規回答の総数
      - t 回目のハッシュ計算で初めてナンスが見つかる確率
    - 攻撃者 (B 個の質問 ID を保持し、y% が質問者と共通) による偽回答数 $V_{attack}(t, y) = \sum_{m=N_v}^{N_q} \frac{\binom{yB}{m}\binom{(1-y)B}{N_q-m}}{\binom{B}{N_q}} \cdot c_a \sum_{k=1}^{t} k\binom{t}{k} p_n^k (1-p_n)^{t-k}$
      - 回答が受理される確率
      - 偽回答の総数
      - 発見できるナンスの期待値
  - 回答数が閾値以上集まるまで受理するので、回答が集まる時間は $t^* = \mathrm{Min}(t \mid V(t, c_l, y) \geq V_{threshold})$
  - 受け取った回答のうち正規エージェントによる回答が占める割合 $R_v(c_l, y) = \frac{V_{legitimate}(t^*, c_l)}{V(t^*, c_l, y)}$
- **対策手法を利用しない場合、**
  - 時間 t に関して受信する全回答数
    - $\hat{V}(t, c_l, y) = \hat{V}_{legitimate}(t, c_l) + \hat{V}_{attack}(t)$
    - 正規エージェントはすぐに、 1回だけ回答するため $\hat{V}_{legitimate}(t^*, c_l) = \begin{cases} 0, & t = 0 \\ c_l, & t > 0 \end{cases}$
    - 攻撃者は何度も回答を送信するため $\hat{V}_{attack}(t) = c_a t$
  - 回答数が閾値以上集まるまで受理するので、回答が集まる時間は $t^* = \mathrm{Min}(t \mid \hat{V}(t, c_l, y) \geq \hat{V}_{threshold})$
  - 受け取った回答のうち正規エージェントによる回答が占める割合 $\hat{R}_v(c_l, y) = \frac{\hat{V}_{legitimate}(t^*, c_l)}{\hat{V}(t^*, c_l, y)}$

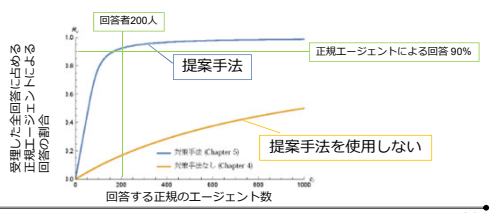
p_l: 受信した回答を受理する確率
c_l: 回答者数
p_n: 1回のハッシュ計算でナンスが見つかる確率
n_q: 回答に添付する質問 ID の個数
N_v: 回答に添付された ID が質問者と N_v 個以上一致すれば回答受理
$V_{threshold}$: 受理する回答数の閾値

数値例

- **フレームワークの規模感**
  - **40,000 家庭**がフレームワークに参加
    - 各家庭が平均 1 件 (/日) 質問
  - 平均 5% のエージェント同士が連携し、**約 2,000 件程度回答を受信**
    - 家庭を 24 種類に分類可能と仮定
      - 朝型 or 夜型、昼間に複数人在宅 or 一人 or 不在、子どもがいる or いない、外食が多い or 少ない
  - エージェントの質問に回答しようとする正規のエージェント数 c_l
- **回答について**
  - 回答には 100 個の質問 ID を添付し 30 個が一致すれば回答受理
  - **100 件の回答を受理したら集計終了**
  - 正規エージェントによる回答が受理される確率は 90%
    - もともと似たエージェントからしか回答は来ない
  - 回答に添付するナンス値の探索
    - 1 回のハッシュ計算でナンスが発見できる確率：1/10
- **攻撃者について**
  - **攻撃者は 1,000 台の端末**で攻撃
  - 回答に添付するために過去の質問から 20,000 個の質問 ID を保存
    - そのうち 20% が質問者が保存している質問 ID と共通

評価結果

- **提案手法によって、偽回答攻撃による悪影響を抑制可能**
  - 回答者が 200 人以上存在する場合に、全回答に占める正規エージェントによる回答の割合が 90% 以上
    - 提案手法を導入せずにフレームワークを利用した場合は、半数以上が偽回答となり、投票結果が改変される可能性
  - 過去に「正常」と回答した質問 ID の利用により、偽回答を破棄
  - 攻撃者の数が質問者より多くても偽回答の影響を抑制
    - 攻撃者：1,000 台のマシンを使用して攻撃



Chapter 5 のまとめ

- **似たエージェントのみが答えられる問題に答えさせることで、偽質問、偽回答を集計しない対策手法を提案**
  - 偽質問、偽回答によって誤ったデータが拡散されることを抑制
  - 大量送信によってたまたま正解することを防ぐために送信数も制限
- **匿名で似たユーザのエージェントと協力可能な点は維持**
- **数値例を用いて偽質問および偽回答による悪影響が抑制可能であることを確認**
  - 偽回答による集計結果への悪影響を抑制可能
  - 偽質問による類似度判定への悪影響を抑制可能
  - また、本提案手法が実時間内に計算可能であることを確認

  本発表では割愛
- **今後の課題：本フレームワークの他アプリケーションへの適用**
  - 似た特徴をもつユーザ同士が連携するサービスへ適用し、有効性を確認

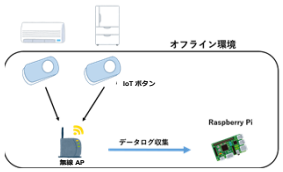
博士論文のまとめ

- **各家庭に十分なデータ量がなくても、ユーザーの行動に基づいて不正操作を検出可能**
  - 家庭内のエージェントが行動系列と宅内状態から行動パターンを学習
  - ホーム IoT 機器への操作パケットが行動パターンと一致するかどうか確認し、一致しない操作は不正操作として検出
  - 機器操作時間帯と行動系列から約 1 ～ 3 か月分のデータで学習可能
    - 行動系列が利用できない場合は、室温など詳細なデータを収集することで学習可能
    - 行動パターンが網羅できずに機器操作が正常か判断できない場合は匿名で協力可能なフレームワークを介して似た行動をとるユーザのエージェントと連携して判断可能
    - 行動が監視できるなど宅内の状態が分かる場合や、行動データが全くないような場合は不正操作の検出が困難
- **今後の課題：**
  - 家庭内のユーザ行動学習：家庭内の機器全般の異常検知、行動リコメンドへの応用
  - 匿名協力フレームワーク：別のエリアの課題への適用

Appendix

Chapter 3, 4: 実家庭でのデータ収集システム

- **行動履歴の収集**
  - 家電に IoT ボタンを取り付け
  - 家電使用時に、対応する IoT ボタンを押す
  - Raspberry Pi でボタンを押したタイミングを記録
- **センサデータの収集**
  - ネットワーク経由でデータを収集可能な IoT センサを宅内に設置



Chapter 3, 4: 実家庭で収集したデータリスト

- **センサデータ**
  - 室温、湿度、騒音、気圧、CO2濃度
- **行動履歴**

機器など	行動
ユーザの入退室	入室、退室
ライト	オン、オフ
エアコン	冷房、暖房、ドライ、温度上げる、温度下げる、オフ
扇風機	オン、オフ
ヒーター	オン、オフ
洗濯機	オン
冷蔵庫	開閉
テレビ	オン、オフ
コンロ	オン、オフ
電子レンジ	オン
トースター	オン
炊飯器	オン

各 Chapter における異常検知手法および評価方法のまとめ

	Chapter 2	Chapter 3	Chapter 4	Chapter 5
異常検知手法	行動系列×時間帯	行動系列×状態推定	行動系列×時間帯	なし
データ収集環境	仮想ホームネットワーク	実家庭	実家庭	---
検知対象	コーヒーメーカ、扇風機、テレビ、ヒータ、加湿器	コンロ	コンロ、扇風機	---
検知対象における学習データ数	約 3 か月分	約 1 か月分	5 日分	---
評価における比較対象	隠れマルコフモデル手法、時間帯のみの手法、行動系列のみの手法、ノイズを除去しない提案手法、ノイズを除去しない系列のみの手法	状態推定のみの手法、行動系列×時間帯	協力しない場合、全家庭のデータで協力する場合	Chapter 4

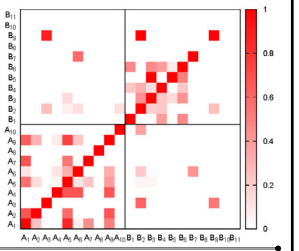
Chapter 3：コンロ以外の評価結果 [37]

- **[評価方法]**
  - 仮想ホームネットワークで得られたデータを利用
    - 1 ユーザの 1 か月分の機器操作を 1 家庭分として設定
    - 各家庭での検知結果を集計
- **[評価結果]**
  - 似た家庭と協力することで、 1 か月あたり 5 回分の正常操作を正しく判断可能に
    - ただし、少量の検知漏れが発生
  - そもそも同じ環境での行動のため、全家庭分のデータが類似

16 家庭分の扇風機の検知結果の合計	誤検知率	誤検知数 / 計	検知漏れ率	検知漏れ数 / 計
単一家庭のみ	23.6%	30/127	0.493%	224/45176
プラットフォーム (類似家庭データ)	19.7%	25/127	0.645%	293/45176
プラットフォーム (全家庭データ)	16.5%	21/127	0.819%	372/45176

Chapter 3: 本当に似た家庭と連携できているかどうか

- **各家庭が保存した ID の一致率をヒートマップにプロット**
  - 縦軸のエージェントに対して横軸のエージェントが何%同じIDを保持しているか
  - 実家庭 A, B から得られたデータで学習したエージェントをそれぞれ $A_{1,2,\ldots,10}$, $B_{1,2,\ldots,11}$ と表記
- **➢エージェント A 同士、B 同士が同じ質問 ID を保持**
  - 同一家庭のデータから学習したエージェント同士が連携した
    - 似た家庭 (同じ家庭でのデータ) 同士が連携した



Chapter 5: 偽質問による影響についての評価

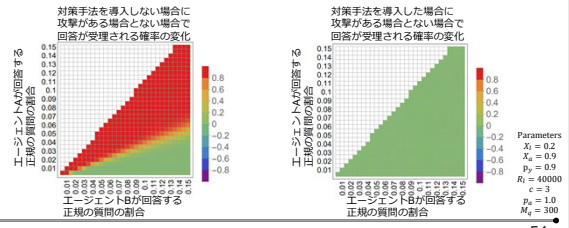
- **エージェント A の質問にエージェント B が回答する確率 $P_{receive}$ が、攻撃前後でどう変化するかを確認**
  - エージェント A の質問に B が回答する確率 $P_{question}$ と、エージェント B の回答がエージェント A に受理される確率 P_{vote} の積
    - $P_{receive} = P_{question}P_{vote}$
    - $P_{question}$：エージェントAが保持するID $S_A(M_q, r_{\{s,A\}})$ 個から n_q 個選択し、N_q 個以上がエージェントBも保持するID $S_{AB}(M_q, X_l r_{\{s,A\}}, X_a r_{\{s,A\}})$ 個と一致する確率
      - $P_{question} = \sum_{i=N_q}^{n_q} \binom{S_{AB}}{i}\binom{S_A - S_{AB}}{n_q - i} / \binom{S_A}{n_q}$
    - P_{vote}：エージェントBが保持するID $S_B(M_q, r_{\{s,A\}})$ 個から n_q 個選択し、N_q 個以上がエージェントAも保持するID $S_{AB}(M_q, X_l r_{\{s,A\}}, X_a r_{\{s,A\}})$ 個と一致する確率
      - $P_{vote} = \sum_{i=N_q}^{n_u} \binom{S_{AB}}{i}\binom{S_B - S_{AB}}{n_u - i} / \binom{S_B}{n_u}$
  - エージェント A および B が保持する ID 数
    - $S(M_q, r_s) = d(p_y r_s R_l + c p_a r_s R_a(M_q))$
  - エージェント A および B が共通して保持する ID の数
    - $S_{AB}(M_q, x_l, x_a) = d(p_y x_l R_l + c p_a x_a R_a(M_q))$
  - 攻撃者が1日に生成可能な偽質問数
    - $R_a(M_q) = \frac{86400}{M_q}$

正規質問　偽質問

文字	意味
M_q	1 つのナンスを探索するのに必要な平均秒数
$r_{\{s,A\}}$, $r_{\{s,B\}}$	エージェントA,Bが回答する正規の質問の割合
$X_l r_{\{s,A\}}$, $X_a r_{\{s,A\}}$	全正規質問,偽質問に対してエージェントAとBが共通して回答する質問の割合
p_y	回答しようとする似た質問に対して、Yesと回答する確率
r_s	エージェントが回答する正規の質問の割合
R_l	1日にフレームワークを流れる正規の質問数
$1/c$	エージェントAとBが共通保持しているIDを攻撃者が何%まで絞れているか
p_a	回答しようとする攻撃者の偽質問にYesと回答する確率
x_l	全質問のうちエージェントAとBが共通して回答する質問の割合
x_a	全偽質問のうちエージェントAとBが共通して回答してしまう質問の割合
d	有効期限の日数

Chapter 5: 偽質問による影響についての評価結果

- **似ていないもの同士を似ていると誤解させる攻撃を実装**
- **対策手法を導入することで、エージェントAの質問にBが回答して受理される確率は、攻撃がない場合と同様**
  - 逆に、対策手法が無い場合は、似ていないのに回答させられ受理
- **偽の質問による類似度判定への影響を削減可能**
  - Aの質問にBが回答し、受理される確率が変化しない



Chapter 5: 計算時間に関する評価

- **1 日あたりの計算時間 $F(u, r_s)$ (秒) から評価**
  - $F(u, r_s) = uM_q + R_l r_s M_v \leq 86400$

質問にかかる秒数　回答にかかる秒数

  - 質問時間：質問数 u × 1 回の質問用ナンス探索時間 M_q
  - 回答時間：回答数 $R_l r_s$ × 1 回の回答用ナンス探索時間 M_v
    - R_l：1日にフレームワークを流れる全質問数
    - r_s：エージェントが回答する質問の割合
    - M_v：ナンス発見確率が閾値 T_v 以上となるまでに必要な計算時間
      - $M_v = T_{nonce} * \mathrm{Min}(t | p_n \sum_{k=1}^{t}(1-p_n)^{k-1} > T_v)$
        - T_{nonce}：1 回のハッシュ計算にかかる時間
        - p_n：1回のハッシュ計算でナンスが発見される確率
        - $p_n \sum_{k=1}^{t}(1-p_n)^{k-1}$：k 回目のハッシュ計算でナンスが発見される確率

Chapter 5: 計算時間に関する評価結果

- **1 日あたり 20,100 秒で計算可能**
  - 1 日 1 件質問、1 質問に必要な計算時間 300 秒
  - 1 日 1,000 件回答
- **実時間内で計算可能**