Computer Security Symposium 2017
23 - 25 October 2017

ネットワーク家電の使用状況データからのプライバシ分析
冷蔵庫を例にして

樋口 尚宏[1]　石原 靖哲[1]　矢内 直人[1]　上田 健介[2]　藤原 融[1]

概要：近年，さまざまな家電がネットワークに接続されるようになってきている．それらの家電が使用状況に応じて送出するパケットは宅内の快適さや利便性の向上，節電といった目的で利用される．その一方で，悪意を持った第三者が宅内のトラフィックを取得し，分析すると，使用者を特定したり，趣味嗜好などの情報を得られたりする可能性がある．本稿では，研究室に構築した模擬環境において，冷蔵庫が使用状況に応じて送出するパケットに注目して，個人識別をどの程度行えるかを実験的に検証した結果について報告する．

キーワード：ネットワーク家電，プライバシ，パケット収集

Privacy analysis of usage data sent by network-connected home appliances
Focusing on refrigerators

Takahiro Higuchi[1]　Yasunori Ishihara[1]　Naoto Yanai[1]　Kensuke Ueda[2]　Toru Fujiwara[1]

Abstract: In recent years, various home appliances have been connected to the network. Such network-connected home appliances send packets according to their usage. The packets are used for making lives of people at home more comfortable and more convenient, for saving power, etc. On the other hand, if a malicious third party acquires and analyzes the in-house traffic, there is a possibility of identifying the user and obtaining personal information such as hobbies and preferences. In this paper, we report an experimental verification result of individual identification, focusing on packets sent by a refrigerator in a simulation environment built in our laboratory.

Keywords: network-connected home appliances, privacy, packet collection

1. はじめに

近年，世の中に存在する様々な「モノ」に通信機能を持たせる IoT（Internet of Things）の普及が急速に進んでいる．IoT により各種センサー，家電機器，自動車などのあらゆる「モノ」がネットワークでつながることで，クラウドにビッグデータとしてデータが集積され，人工知能による分析，最適化，自律化などにより，新しい価値を生み出すことが期待されている．文献 [1] には，IHS Technology 社による IoT 普及の現状と予測が紹介されている．それによれば，2015 年時点でインターネットにつながるモノ（IoT 機器）の数は 154 億台であり，2020 年にまでに 304 億台まで増大するとされている．その中でも特に IoT 機器の普及が進むとみられているのは家電機器などのコンシューマ分野であり，2020 年には 124 億台まで増大する予測である．

一方，「モノ」がネットワークにつながることにより，サイバー攻撃による新たな脅威が増大することが懸念されて

[1] 大阪大学
Osaka University, 1-5 Yamadaoka, Suita, Osaka, 565-0871, Japan
[2] 三菱電機株式会社
Mitsubishi Electric Corporation,1 Zusho Baba Nagaokakyo City Kyoto 617-8550, Japan

いる．特に，HEMS (Home Energy Management System) に代表される宅内 IoT システムがサイバー攻撃の標的になると，家電機器だけでなくその利用者自身が危険にさらされたり，個人情報や趣味嗜好などのプライバシ情報が流出するなど一般利用者の生活が直接脅かされる可能性があり，いかに利用者のプライバシ情報を保護するかが課題となっている．

我々は，ネットワーク家電が使用状況に応じて送出するパケットにおいて，宅内における在室人数や性別，年齢層などのプライバシ情報がどれくらい漏れているのかをそれぞれのネットワーク家電ごとに定量化することを目的としている．プライバシ情報漏洩度が高い機器にパケットデータを暗号化するなどの対策を適用することで上記の課題を解決することができると考えている．

我々は，一般家庭におけるネットワーク家電の利用を模倣できる環境を大学の研究室内に構築し，1 年近くにわたってパケットを収集してきた．本稿では，模倣環境内のさまざまなネットワーク家電のうち冷蔵庫に着目して，冷蔵庫が使用状況に応じて送出するパケットを収集し分析した結果について述べる．本稿の貢献は，この分析結果により冷蔵庫からどれくらいプライバシ情報が漏れているのかや使用者を特定することが可能であるのかなどを検証したことである．

本稿の構成は以下のようになっている．2 章では本稿に関連する研究を挙げる．3 章では，研究室にどのように模擬環境を構築したのかについて解説する．4 章では，模擬環境内に設置した冷蔵庫のパケットを分析した結果を報告し考察を示す．最後に，5 章に本稿のまとめと今後の展望を述べる．

2. 関連研究

文献 [2] では，実在する 5 つの家庭に対して，18 ヶ月間パケットキャプチャを行い，ホームネットワーク外からの攻撃と思われる通信の頻度を調査している．パソコンやスマートフォン，プリンタなどの従来からネットワーク接続することができた機器を対象としている．これに対し本稿では，ネットワーク接続することができる家電全てを対象に，ホームネットワーク内から攻撃を受けた場合にどれくらいのプライバシ情報が漏れるのかを検証しているため，対象機器および観測する観点が大きく異なっている．しかし，本稿の成果と文献 [2] の成果とを合わせることで，ホームネットワーク内外の脅威分析が示せるようになると考えられる．

文献 [3] [4] [5] では，センサ機器を対象として，暗号化された状態のパケットを取得し続け，パケットサイズに注目した分析を行っている．この分析により，被験者となるユーザ 1 人がトラフィックをみてネットワーク内にいるかどうかを判定する．なお，分析に使用する情報はパケットサイズと到着間隔のみに着目している．このような分析方法であったとしても，ネットワーク内にユーザがいるかどうかを 8 割以上の確率で正しく判定できたという興味深い結果を報告している．本稿では，センサ機器以外の生活家電を対象に，パケットサイズと到着間隔のみでなく，パケットデータにも着目しているので，より高度な分析を行う攻撃者を対象としているといえる．

3. 実験環境

我々の目的は，それぞれのネットワーク家電ごとにプライバシ情報がどれくらい漏れているのかを定量化することである．その目的に向けて本稿では，次節で述べるように，攻撃者が家庭内のネットワーク家電が送受信するパケットを収集した場合，プライバシに関してどのような危険が生じうるかを検討する．なお，攻撃者がパケットをどのように収集するかに関して，特定のシナリオは想定しない．

上述の目的達成のために，一般家庭におけるネットワーク家電の利用を模倣できる，以下のような環境を大学の研究室内に構築した．環境内にはテレビや冷蔵庫，扇風機，コーヒーメーカー，ヒーターなどのネットワーク家電が設置されている．被験者は 1 ヶ月あたり 4 人であり，それぞれにスマートフォンが貸与される．被験者は環境内のネットワーク家電を本来の付属している機器 (テレビでいえばリモコン) で操作するのではなく，一部の例外を除いて，貸与したスマートフォンにインストールしたアプリ上で行ってもらう．アプリ上で操作してもらう理由は，本来の付属している機器で操作したとしても，家電によるパケットの送受信は引き起こされないが，アプリ上で操作すると家電によるパケットの送受信が引き起こされるからである．被験者は，1 ヶ月間，環境内のネットワーク家電を日常的に使用すると同時に，貸与されたスマートフォン上の行動ログアプリを用いて，入退室の時刻およびネットワーク家電の使用状況を記録する．

実験環境におけるネットワーク構成を図 1 に示す．ネットワーク家電は左側の無線 AP (Access Point) に接続されており，ネットワーク家電操作用のスマートフォンは右側の無線 AP に接続されている．それぞれの無線 AP は中央のインテリジェントスイッチを通じて接続されている．さらにインテリジェントスイッチにはパケットキャプチャ用 PC が接続されている．これにより，被験者がネットワーク家電を操作する際に，スマートフォンからネットワーク家電に対して送られるパケットは必ずインテリジェントスイッチを通るようになっている．パケットキャプチャ用の PC はインテリジェントスイッチを通ったパケットを全てキャプチャすることができる．もちろん，ネットワーク家電からスマートフォンへと送信されるパケットもまた必ずインテリジェントスイッチを通るようになっており，このパケットもまた全てキャプチャすることができる．

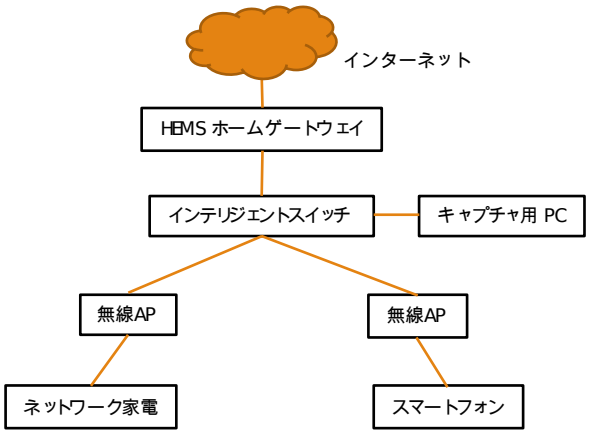

図 1 研究室に構築した模擬環境のネットワーク構成

4. プライバシ分析

本稿では，模擬環境内に設置したネットワーク家電の中から冷蔵庫のみに着目し，冷蔵庫からのパケットを分析することで，冷蔵庫の利用形態に基づいた個人の特徴が得られるのではないかという仮説を立てた．そこで，各被験者について収集した一ヶ月ごとのパケットデータを，その被験者の利用形態の特徴をよく表すと予想される形に加工した．その加工したデータをプロファイルと呼ぶ．以下では，まず，収集したパケットデータに含まれる情報について説明する．その後，提案および試作した 3 種類のプロファイルと，それぞれのプロファイルについての仮説の検証結果について述べる．

4.1 パケットデータの内容

冷蔵庫は，稼働時において一定間隔で，パケットを HEMS ホームゲートウェイに送っている．そのパケットには，冷蔵室、冷凍室、野菜室等のドア開閉状態等の情報が含まれている．本稿では，これらの情報を利用し、冷蔵室のドアの開時刻と開回数に着目して分析を行う．冷蔵室の情報のみを使用する理由は，冷蔵室以外の他の室はほとんど利用されていなかったためである．模擬環境内には冷蔵庫は 1 台しか設置されていないため，どの被験者が冷蔵庫を利用していたのかについては，行動ログと対応付けることにより判断する．

4.2 プロファイルの提案

本稿では以下の 3 種類のプロファイルを提案する．

4.2.1 開動作時刻ベースプロファイル

開動作時刻ベースプロファイルは，被験者が冷蔵室のドアを開ける動作を行った時刻に注目したプロファイルである．

個人は 1 日において冷蔵庫を利用する時間帯がある程度固定されていると考えられる．例えば，毎日 12 時前後に昼食をとる人は 12 時前後の冷蔵庫の開動作が増え，毎日 13 時前後に昼食をとる人は 13 時前後の冷蔵庫の開動作が増えるであろう．このように，冷蔵室のドアを開けた時刻に注目したプロファイルにより，個人の特徴を得ることができるのではないかと考えられる．

4.2.2 開状態時刻ベースプロファイル

開状態時刻ベースプロファイルは，被験者が冷蔵室のドアを開けた時刻とその時何秒間開けていたのかに注目したプロファイルである．

個人は 1 日において冷蔵庫を利用する時間帯がある程度固定されているだけでなく，そのときの利用方法も類似していると考えられる．例えば，毎日 18 時前後に買い物をしてから帰宅する人の場合，18 時前後に冷蔵庫が利用され，かつドアが開いている時間が比較的長いと予想される．一方，買い物をせずに帰宅する人の場合，18 時前後に冷蔵庫が利用されたとしても，ドアが開いている時間は比較的短くなると予想される．このように，冷蔵室のドアを開けた時刻と開いていた時間に注目したプロファイルにより，個人の特徴を得ることができるのではないかと考えられる．

4.2.3 開時間間隔ベースプロファイル

開時間間隔ベースプロファイルは，被験者が冷蔵室のドアを開けた時刻の間隔に注目したプロファイルである．

個人の冷蔵庫の利用方法にはある程度固定された「パターン」があると考えられる．例えば，冷蔵庫にあるジュースを一口だけ飲みたいという状況を考える．ある人はいつも，ジュースを取り出した後に一度ドアを閉めて，飲み終わった後にまたドアを開けてジュースを冷蔵庫に戻すであろうし，別のある人はいつも，ドアを開けたまま飲むであろう．一般にこのような「パターン」はさまざまに定義可能であるが，本稿では冷蔵室のドアを開けた時刻の間隔で表現できるパターンに注目したプロファイルを提案する．

4.3 プロファイルの試作

2016 年 10 月から 2017 年 3 月におけるデータの中で，冷蔵庫利用が多かったのべ 8 人分のデータをピックアップした．ピックアップした被験者と月の情報を図 2 に示す．同じアルファベットが割り当てられた被験者は，同一人物を意味する．なお，図 2 の配置は図 5，図 6，図 7 の配置と対応している．

4.3.1 開動作時刻ベースプロファイルの試作

開動作時刻ベースプロファイルの試作においては，カーネル密度推定を利用することにした．カーネルとして正規分布を使用する．例えば，12:00:00 と 18:00:00 に被験者がドアを開けている場合は，図 3 のようにその時刻をピークに持つ正規分布をはりつける．1 ヶ月間のその被験者のドア開動作時刻すべてについて正規分布をはりつけ，合計したものを，その被験者のその月の開動作時刻ベースプロファイルとした．正規分布の標準偏差としては 30 分，60 分，90 分，120 分の 4 種類を試した．

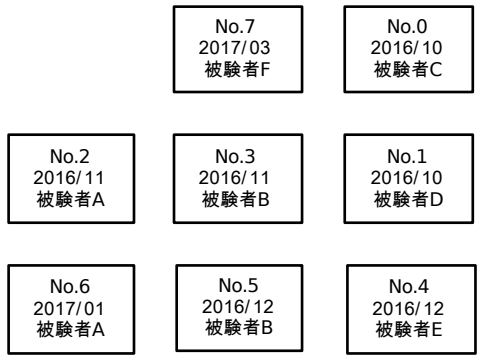

図 2 ピックアップした被験者と月の情報

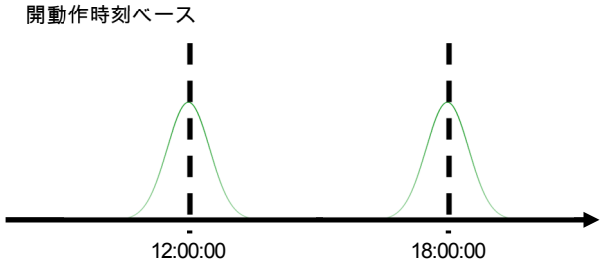

図 3 開動作時刻ベースにおけるカーネル密度推定の例

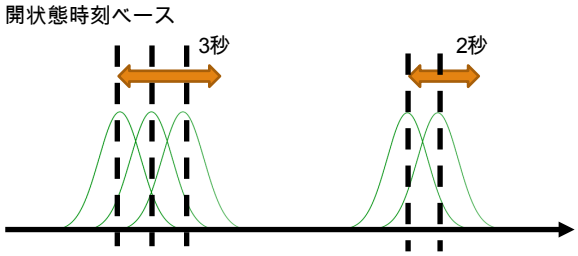

図 4 開状態時刻ベースおけるカーネル密度推定の例

カーネルとして標準偏差 30 分の正規分布を用いた開動作時刻ベースプロファイルを図 5 に示す．横軸は 1 日における時刻，縦軸は当該月の時刻に開動作が行われた確率に総開動作回数をかけた値を表す．

4.3.2 開状態時刻ベースによるプロファイル試作

開状態時刻ベースプロファイルの試作においては，開動作時刻ベースと同様に，カーネル密度推定を利用することにした．カーネルとして正規分布を使用する．例えば，被験者がドアを 3 秒間開けていた場合は図 4 の左側のように 1 秒間隔で 3 つの正規分布をはりつける．2 秒間開けていた場合は図 4 の右側のように 1 秒間隔で 2 つの正規分布をはりつける．1 ヶ月間のその被験者によるドア開状態時刻すべてについて 1 秒間隔に正規分布をはりつけ，合計したものを，その被験者のその月の開状態時刻ベースプロファイルとした．正規分布の標準偏差としては 30 分，60 分，90 分，120 分の 4 種類を試した．

カーネルとして標準偏差 30 分の正規分布を用いた開状態時刻ベースのプロファイルを図 6 に示す．横軸は 1 日における時刻，縦軸は当該月の時刻に開状態であった確率に総開状態時間 (s) をかけた値を表す．

4.3.3 開時間間隔ベースによるプロファイル試作

開時間間隔ベースプロファイルの試作においても，カーネル密度推定を利用することにした．ただし，時間間隔は負の値をとらないため，実数全体の上で定義されている正規分布ではなく，非負実数上で定義されるガンマカーネルを利用する．ガンマカーネル $K_{\rho_b(x),b}$ は以下の式で定義される [6]．ここで Γ はガンマ関数であり，b は標準偏差である．

$$K_{\rho_b(x),b} = \frac{t^{\rho_b(x)-1}e^{-t/b}}{b^{\rho_b(x)}\Gamma(\rho_b(x))}$$

$$\rho_b(x) = \begin{cases} \frac{x}{b} & \text{if } x \leq 2b; \\ \frac{1}{4}(\frac{x}{b})^2 + 1 & \text{if } x \in [0, 2b) \end{cases}$$

プロファイル試作においては，ガンマカーネルの標準偏差として 10 分，15 分，20 分，30 分の 4 種類を試した．また，開時間間隔が半日以下のデータにのみ，カーネル密度推定を行った．

カーネルとして標準偏差 30 分のガンマカーネルを用いた開時間間隔ベースの結果を図 7 に示す．横軸は開時間間隔 (s)，縦軸は当該月に時間間隔で開けられていた確率に総開回数をかけた値を表す．

4.4 評価・考察

3 種類のプロファイルを評価するために，各プロファイル間の L2 ノルムを計算した．その結果を図 8 から図 10 に示す．これらの図の読み方を，図 8 を例に説明する．図 8 には，0 から 7 までの数字がプロットされているが，これらの数字は図 2 の No. の値と対応している．したがって，数字 2 と 7（赤色で示されている）は被験者 A を，数字 3 と 5（緑色で示されている）は被験者 B を，それら以外の数字（青色で示されている）は他の被験者を表している．図 8 は横に並ぶ縦長の 4 つのエリアに分かれており，左から順に，カーネルの標準偏差 30 分，60 分，90 分，120 分で試作したプロファイル間のノルムを表示している．それぞれのエリアには，左から順に，No.0 と全被験者，No.1 と全被験者，...，No.7 と全被験者とのノルムを計算した結果が表示されている．縦軸はノルムの値を表す．たとえば，カーネルの標準偏差 30 分で作成したプロファイルについて，No.0 の被験者と No.1 の被験者の間のノルムは約 0.22 であり，No.0 の被験者と No.2 の被験者の間のノルムは約 0.30 である．No.0 の被験者から見て一番ノルムが小さかったのは，数字が一番下に記載されている No.1 の被験者である．

では，まず，開動作時刻ベースの結果（図 8）について考察する．No.2（11 月の被験者 A）に最も類似していたのは No.6（1 月の被験者 A）であり，逆に No.6（1 月の被験者 A）に最も類似していたのは No.2（11 月の被験者 A）であった．同様に，No.3（11 月の被験者 B）に最も類似して

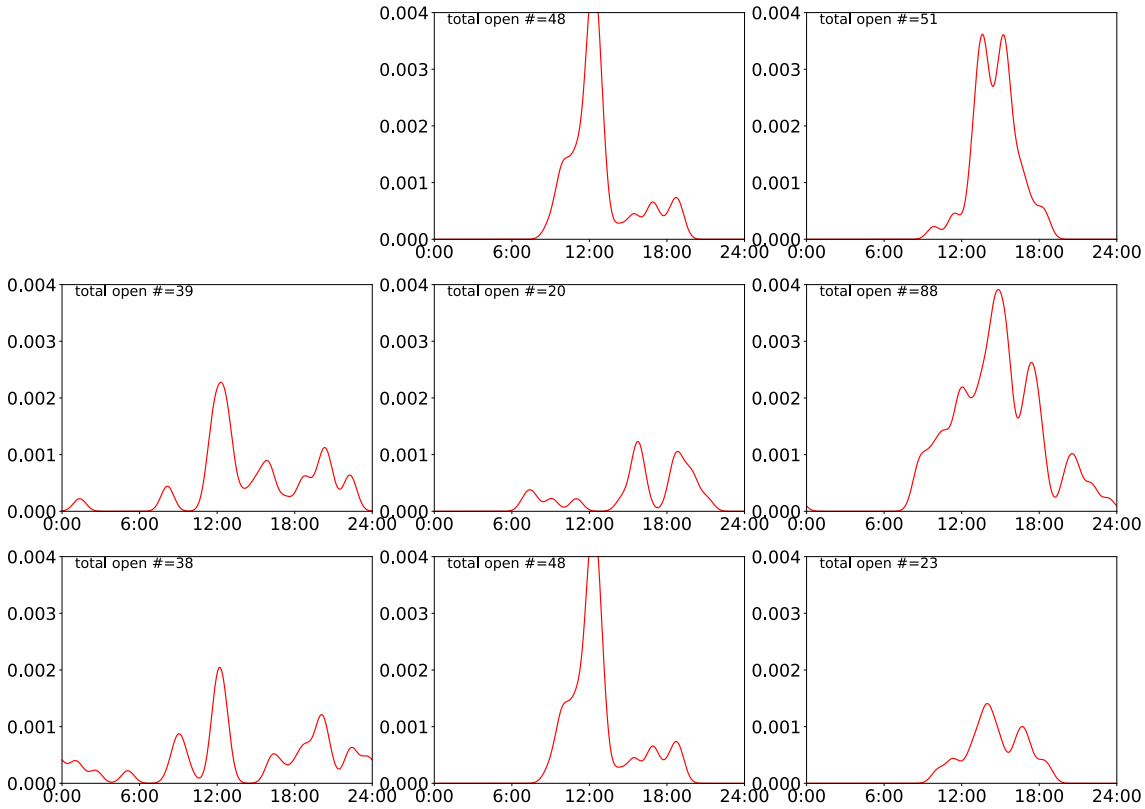

図 5 開動作時刻ベースプロファイル

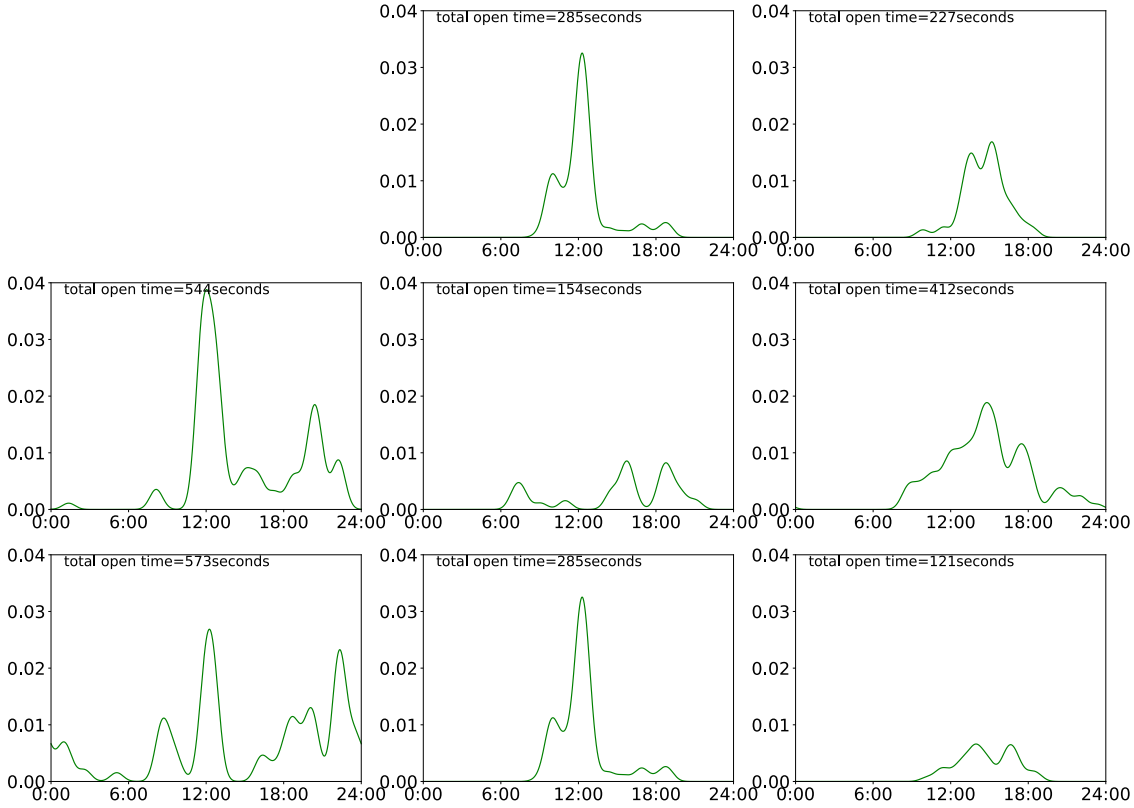

図 6 開状態時刻ベースプロファイル

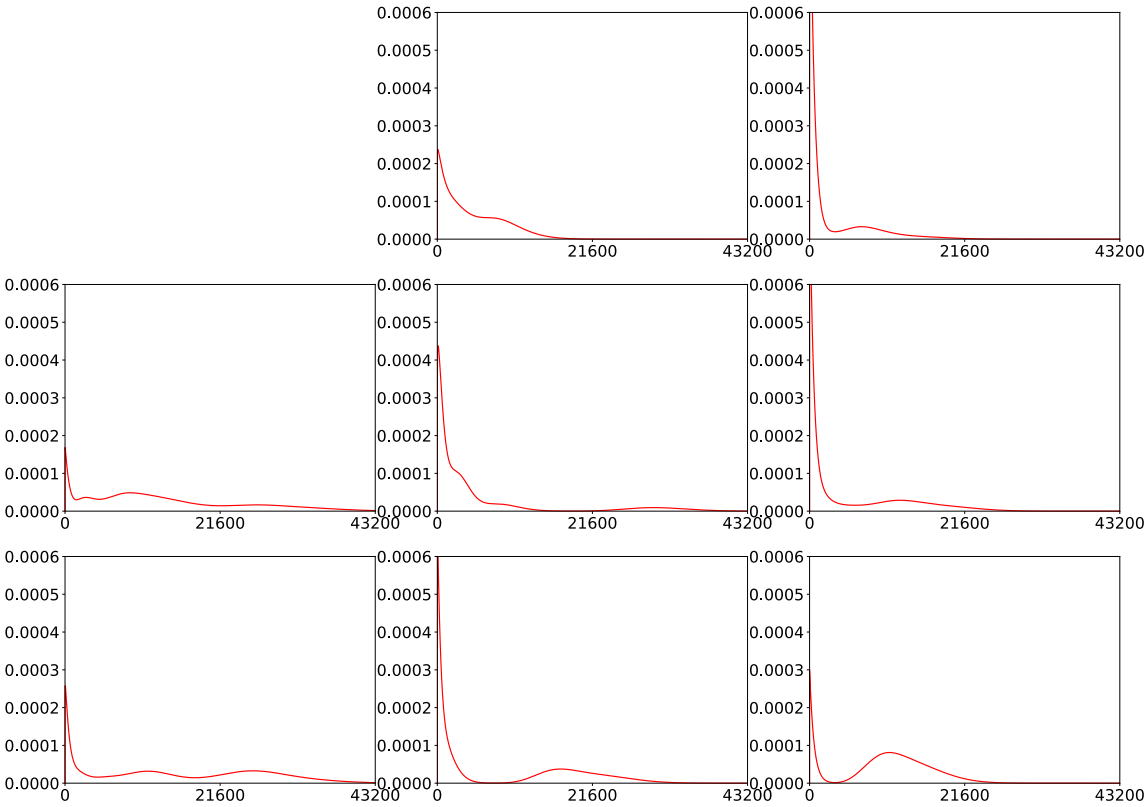

図 7 開時間間隔ベースプロファイル

いたのは No.5（12 月の被験者 B）であり，逆に No.5（12 月の被験者 B）に最も類似していたのは No.3（11 月の被験者 B）であった．このように，4 種類のカーネル標準偏差すべてについて，同一の被験者のプロファイルが最も類似しているという結果になった．

次に開状態時刻ベースの結果（図 9）について考察する．No.3（11 月の被験者 B）と No.5（12 月の被験者 B）については，開動作時刻ベースのときと同様に，一方が他方に最も類似しているという結果が得られた．さらに，No.6（1 月の被験者 A）に最も類似していたのは No.2（11 月の被験者 A）であった．しかし，No.2（11 月の被験者 A）に最も類似していたのは，No.7（3 月の被験者 F）など A とは別の被験者であった．少ない回数しか開けないが長い時間開けた状態にする人（今回の場合，12 時ごろの No.2）と多い回数開けるが短い時間しか開けた状態にしない人（12 時ごろの No.7）のプロファイルが似通ってしまうためであると考えられる．

最後に，開時間間隔ベースの結果（図 10）について考察する．No.2（11 月の被験者 A）と No.6（1 月の被験者 A）については，一方が他方に最も類似しているという結果が得られた．しかし，No.3（11 月の被験者 B）と No.5（12 月の被験者 B）については，一方に最も類似しているものとして他方は得られないという結果になった．そもそも被験者ごとの時間間隔の差異はほとんど見られなかった．本実験では，被験者が全員大学院学生であり，利用の「パターン」が似通っているためであると考えられる．

以上のように，提案した 3 種類のプロファイルの中では，開動作時刻ベースプロファイルが個人の特徴を最もよく表しているという結果になった．そして，今回ピックアップしたのべ 8 人分の開動作時刻ベースプロファイルについては，同一被験者によるプロファイルが存在するならそれが最も類似したプロファイルであるという結果も得られた．この結果は次のような攻撃が可能であるということを示唆している．

> 攻撃者 A は攻撃対象の人物 B が住む単身者用マンションを特定しており，かつ B のある月の開動作時刻ベースプロファイルを手に入れているとする．A は，そのマンションの各戸内の冷蔵庫が送出するパケットを収集し，戸ごとに開動作時刻ベースプロファイルを作成する．B のプロファイルと最も類似した戸を，B が住む部屋であると推定する．

一方で，開動作時刻ベースであっても，同一被験者間の場合にとるノルムの範囲と異なる被験者間の場合にとるノルムの範囲に明確なギャップは見出せなかった．このことは，上のシナリオでいうと，攻撃者 A が攻撃対象の人物 B が住む単身者用マンションを特定していない状況で，あるマンションにおいて B が住んでいる部屋を推定する（あるいは B がそのマンションには住んでいないことを推定する）のは難しいということを示唆している．

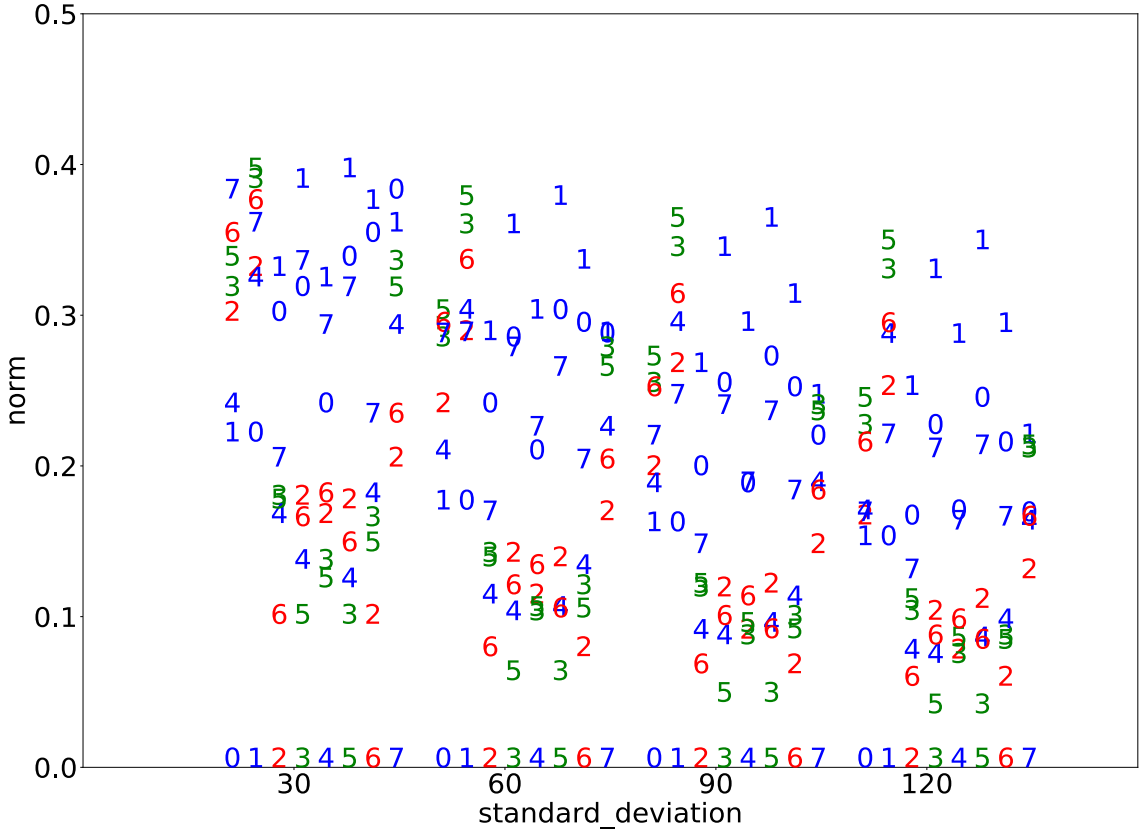

図 8 開動作時刻ベース ノルム

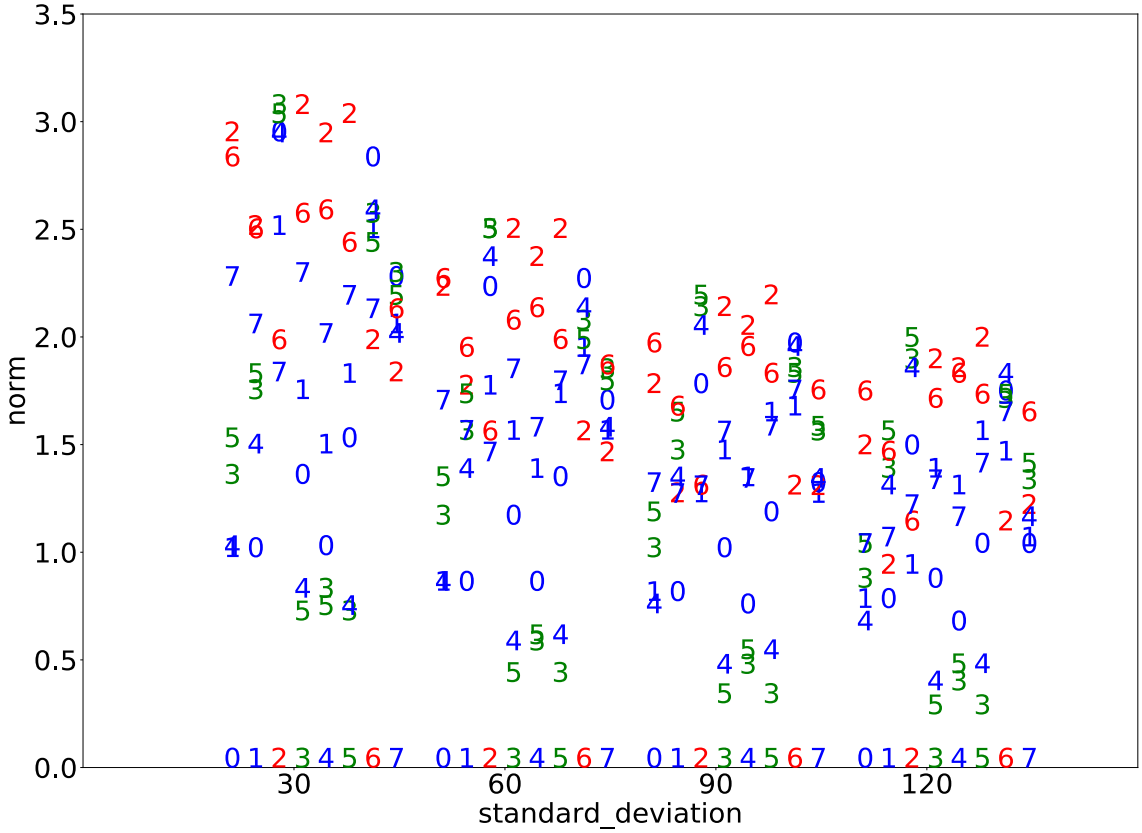

図 9 開状態時刻ベース ノルム

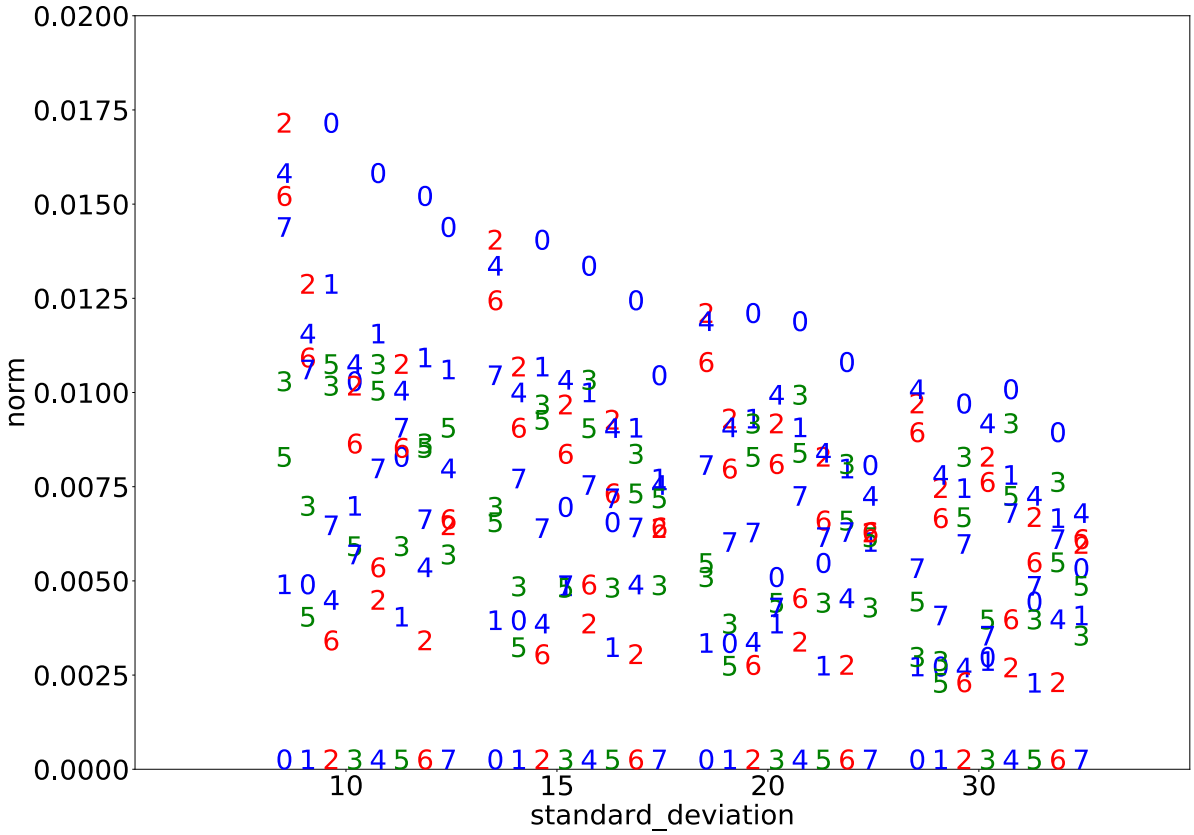

図 10 開時間間隔ベース ノルム

5. まとめ

本稿では，研究室に構築した一般家庭におけるネットワーク家電の利用を模倣できる環境において，ネットワーク家電が使用状況に応じて送出するパケットを分析することで，利用者個人の特徴が得られるのではないかという仮説を，冷蔵庫に注目して検証した．3 種類のプロファイルを提案・試作して検証したところ，開動作時刻ベースプロファイルが利用者個人の特徴を最もよく表しているという結果になった．具体的には，攻撃対象の人物の開動作時刻ベースプロファイルを攻撃者が入手している状況で，攻撃対象人物のものも含む複数のプロファイル候補の中から，攻撃者が対象人物のプロファイルを特定するという攻撃の可能性を示した．今後は，冷蔵庫以外のネットワーク家電についても着目し，同様の実験を行なっていきたい．現在，次なる候補として注目している機器はテレビである．テレビはチャンネル情報も取得できることができることが判明しており，どの時間にどのチャンネルを見ていたかを分析することで，より高度な分析を行う攻撃者を対象とすることができると考えている．また，パケットデータだけでなく，使用電力量などの情報も用いたプライバシ分析についても検討していきたいと考えている．

参考文献

[1] 総務省. 平成 28 年版 情報通信白書. *http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h28/index.html*, 2016.

[2] Kuai Xu, Feng Wang, and Xiaohua Jia. Secure the internet, one home at a time. *Security and Communication Networks*, 9(16):3821–3832, Nov 2016.

[3] Bogdan Copos, Karl Levitt, Matt Bishop, and Jeff Rowe. Is anybody home? inferring activity from smart home network traffic. In *Security and Privacy Workshops (SPW), 2016 IEEE*, pages 245–251. IEEE, May 2016.

[4] Frederik Möllers, Sebastian Seitz, Andreas Hellmann, and Christoph Sorge. Short paper: extrapolation and prediction of user behaviour from wireless home automation communication. In *Proceedings of the 2014 ACM conference on Security and privacy in wireless & mobile networks*, pages 195–200. ACM, 2014.

[5] Frederik Möllers and Christoph Sorge. Deducing user presence from inter-message intervals in home automation systems. In *IFIP International Information Security and Privacy Conference*, pages 369–383. Springer, May 2016.

[6] Song Xi Chen. Probability density function estimation using gamma kernels. *Annals of the Institute of Statistical Mathematics*, 52(3):471–480, 2000.