IoT データにおけるリスク喚起および マネジメントに関する一考察

真下 祥子[†1]　松井 加奈絵[†1 †2]

概要: 本論文では，IoT システムから得られる細粒度性，即時性，多様性，連結性といった特徴をもつデータを IoT データと定義し, IoT データの流通を安全に行なうことを目的としたリスクマネジメント方法についての必要性を述べる．そのために，IoT データの特徴から想定されるリスクを顕在化することで，データを得る，持つ，利用する者に対する注意喚起を行なう．また，データ流通にはどのようなリスクマネジメントが想定されるのかをデータ匿名化手法とともに明らかにする.これらのリスクマネジメントを通じて，IoT データの健全な流通を目指し，スマートシティ，コミュニティなどデータから成り立つサービス，アプリケーションに貢献する．

キーワード：IoT データ，リスクマネジメント，匿名化

A study of risk awareness and management in IoT data

Shoko MASHIMO[†1]　Kanae MATSUI[†1, †2]

Abstract: In this paper, we define the data having characteristics such as fine granularity, immediacy, diversity, and connectivity obtained from the IoT system as IoT data, and presents a risk management method aimed at safely distributing IoT data. By making the risk assumed from the characteristics of the IoT data, attention is drawn to the person who collects the data and uses it. In addition, we clarify what type of risk management is assumed for data distribution along with data anonymization method. Through these risk management, we aim to contribute to services and applications that consist of data such as smart city, community.

Keywords: IoT data, Risk management, Anonymization

1. はじめに

Internet of things（IoT）技術の普及に伴い，多種多様なIoTセンサを用いたデータの計測, 収集, 蓄積が行われている．昨今のIoT技術によって得られたデータ（以下，IoTデータと呼ぶ）は，個人や特定の土地・場所に付随するデータである場合が多く，間違った解釈がなされた場合，IoTデータの流通は個人情報流出のリスクとなり得る．例えば，ある区域内に設置された粉塵センサから得られるデータは，アレルギーを持つ人々にとっては，その区域に行く，行かないという決定の判断材料となり得る[1]．しかしながら，データ閲覧者が局所的なデータ推移，たまたま粉塵センサから得られる値が高かった場合，「粉塵の舞うアレルギー発生リスク」の高い地域であると判断する可能性が発生する．この場合，本来とは異なる理解をしてしまい兼ねない．IoTデータは時系列データである場合が多く，その偏移を見ることで特徴を読み取れることができるが，ある一定の時間帯のデータを読み解いた場合，「誤認識」が発生し，事実とは異なる認識を招いてしまい兼ねない．また，IoTデータにおけるメタデータにおいて匿名性が担保されない場合，個人の特定を招いてしまうことがある[2].

本研究ではIoTデータの流通を安全に行うことを目的とし，（1）IoTデータが持つリスクを顕在化することでデータを得る者，持つ者，利用する者，提供する者に対して注意喚起を，また（2）リスクを明確にすることによりどのような対処法が考えられるのかリスクマネジメントを提案することを目的とする．社会におけるIoTデバイス，システムの広がりの中，IoTデータの流通は避けられない．スマートシティ，コミュニティ，ハウスといった概念はセンシングされたIoTデータが循環することで, それぞれが内包するシステムまたはサービスが成り立っている．そのため，本論文ではリスクを知り，正しくマネジメントすることでIoTデータの健全な流通を目指す．

2. IoT データにおけるリスク

本章では，本論文におけるIoTデータとは何かを定義し，またそれらに伴うリスクを明確化する．また，本論文におけるIoTデータはあるユーザ（人間）や，もしくは複数のユーザに関連するデータであることを条件とする. IoTデータにコンピュータ同士のやり取りのデータが含まれ，それらにも同様にリスクは存在するが, 本論文では対象としない.

†1 東京電機大学 理工学部 情報システムデザイン学系
Tokyo Denki University
†2 慶應義塾大学 メディアデザイン研究所
KMD research institute, Keio University

2.1 IoT データとは

本論文においてIoTデータとは, IoT技術にて計測, 収集, 蓄積されたデータを意味する．また，これらのデータは以下の特徴を持つ.

- **粒度の細かさ：** IoT デバイスの測間隔が細かいことが多いため, これまでの観測データとは粒度が異なる. 例えば, これまでのネットワーク対応機器を介さないデータの収集の場合, 計測機器を用いて限られた時間, 場所にて行われることが多かった（例 : パーソントリップデータ[3]など）．また，一般的な計測機器では内部の記憶領域にデータを蓄積していくため, データの蓄積にはメモリの容量に依存する．一方，IoT デバイスによる計測の場合は記憶領域にデータを保存するものの, 一定の間隔においてネットワーク経由で中央サーバに送信することが可能であるため, 理論的には 24 時間 365 日のデータ収集が可能である. そのため, 粒度が高く, かつ時系列データであること IoT データの特徴のひとつと言える.
- **即時性の高さ：** 上述のように時間間隔の細かい，つまりは粒度の細かいデータの収集が可能であることが IoT デバイスによる計測の特徴であり, データの特徴である. これらのデータはセンサ自体､もしくは IoT データを蓄積したデータベースにアクセスすることにより, 即時性の高いデータを活用することが可能となる．例えば，スマートメータは IoT 技術を利用したセンシングデバイスであり，粒度の高い電力消費・発電データを収集するものである．スマートメータは柔軟な電力送電システム, 料金システムを実現するためのデータ収集デバイスであるため, 収集され蓄積されたデータはこれらのシステムに使用される[4]. これらのシステムは, 即時性の高いデータを利用しなければ実現しない．そのため，即時性の高さは IoT データの特徴のひとつである.
- **多種多様さ :** IoT デバイスでは何かしらの事象をセンシングするが, センシングのためのセンサは多種多様であるため， IoT データにも同様に多様性がある. 例えば IoT データを用いて室内外の暑さ, 快適指数に関する即時的な値を得ることで, 熱中症やヒートショック防止のためのヘルスケア情報として使用が可能となる．例えば，暑さ指数 WBGT（Wet Bulb Globe Temperature）は湿度，日射・輻射など周辺の熱環境, 気温によって得られる[5]. また, 室内温熱環境の指数である SET *（Standard New Effective Temperature）や PMV（Predicted Mean Vote）は室内気温，湿度，照度, 風速, 運動量などのデータを必要とする[6]. つまりは, 多種多様なデータを得ることによって, これまでは静的な値だったものが動的な値として得られる．これらの値は先ほどに述べたヘルスケアシステムに用いることができる.
- **連結性 :** 連結性とは IoT データ間のつながりを指す. 例えば，スマートハウスでは世帯に対して，電力消費量，温湿度照度，粉塵，CO2 濃度計測を行うことでエネルギー効率や室内快適性維持を見える化や自動制御にて行うが，世帯に対して準ずるものである．つまりは個々の IoT データから得られる世帯情報は少なくとも, 複数のデータからは詳細情報を読み取られる可能性は高くなる．例えば，先に挙げた電力消費量データからは世帯構成，利用パターン，在・不在など背景情報や行動パターンの推定が可能である[7]. 1 種類のデータにおいても数種類の項目について推定が可能であるが, 数種類のデータを用いることで推定精度は飛躍的に上昇する．そのため，一意の ID に対するデータの種類が連結されていくほど, データ流出時のリスクが高くなる.

2.2 IoT データにおけるリスク

これら IoT データの特徴から考えられる，発生可能性のあるリスクについて述べる．IoT データの特徴として，取得単位間隔が短いため粒度が細かく即時性が高いことが挙げられる．これらのデータを用いることで，スマートシティ，コミュニティ，ハウスは成り立っている.

例えば，エネルギー効率性を上げるためには，即時性の電力消費量データを用いることで需要を把握し，その値に応じた発電や送電を行なう必要がある. このようにIoTデータの持つ特徴によって，「スマート化」は実現されるものであるが，同時にこれらのデータが流通してしまうリスクが存在する．上述の粉塵データは，アレルギー保有者にはそのリスク回避のために「粉塵データの値が高いが場所には行かない, 滞在しない」という選択を取ることができる. しかしながら，影響を受けない人々にとってリスクが高い場所であると解釈を受けてしまった場合，場所，住まいの評価を下げてしまい兼ねない.

このように，温度や湿度データのような馴染み深いデータにおいても，気温の高い場所，湿度の高い場所と判断されてしまった場合，住まいには適さない，などと間違った解釈がなされる可能性がある. IoT技術の広がりに伴い, IoT データの普及が進むことは避けられないと考えられる．しかしながら, このようなIoTデータの特徴に伴うリスクは普及と同時に広まることが予想される．そこで本研究では，IoTデータに潜むリスクを顕在化する．また，そのリスクの回避方法を明らかにし，今後のデータ利活用方法の提案をする.

2.3 ケース

次にIoTデータがどのように収集されていくのか, ケースを用いて述べる．図1はスマートハウス，シティ在住のユーザの1日の動きと，その動きによって発生が予想されるIoTデータを示したものである.

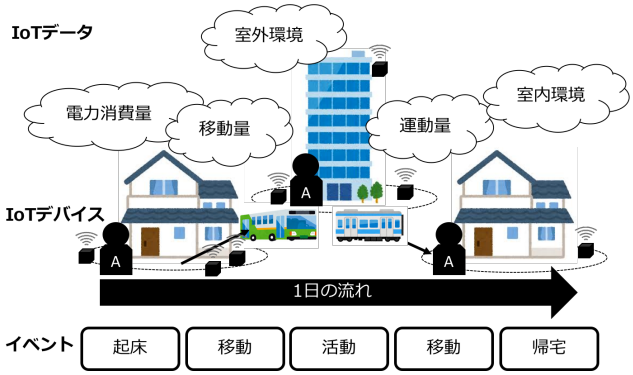

図 1　IoT データの流れ

Figure 1　Flow of IoT data creation.

図1で示したように，ユーザAの行動はIoT デバイスを通じてデジタルデータとして計測され，IoTデータとなる．IoTデータは，IoTシステム（エネルギー管理システム，交通サービス，セキュリティシステムなど）に使用されるものであるが，人間の活動に密接に結び付いているIoTデータは個人に対応している場合が多い．そのため，IoTデータから個人が特定されてしまった場合，様々な種類の個人情報が特定されてしまう可能性がある．また，IoTデータには，粒度の細かさ，即時性の高さ，多種多様さ，連結性がある．粒度の高さは細かな行動特定につながり，即時性の高さは現在に近しい情報になり得るため，悪意を持った第三者にとって在・不在の判断による家屋への侵入など，利用し易い情報となる．また，それぞれのIoTデータが別々のデータベースに保存される場合は起きないが，様々な種類のIoTデータを用いて運用されるサービスの場合は，それぞれのユーザに対して複数のIoTデータが結び付いていく．そのため，一度個人が特定されてしまった場合，関連データが全て特定されてしまう恐れがある．また，一種のデータからは特定されずとも，複数データから個人判別につながる可能性がある．

3. リスクマネジメント

前章にてIoTデータが持つ潜在的なリスクを述べた．IoTデータには一定のリスクが存在するが，本章では，IoTデータ流通においてセキュリティ手法のひとつである匿名化について述べる．また，IoTデータの特性を述べ，なぜ匿名性の担保が必要であるのかを明確にする．

3.1 既存の匿名化手法

匿名化とは，データに対して個人の特定や識別をできないよう，データの加工を行う手法のひとつである[8]．主に行われる処理として，仮名化，属性削除，レコード削除，一般化，統計化が挙げられる．オープンデータの流通が始まった頃，データから個人が特定される事案が発生したため，匿名化技術としてk-匿名化，次にl-多様性が生まれた．k-匿名化とは，識別子および準識別子を持つデータ（表1）において，各準識別子の組み合わせがk個以上になるように各値を一般化したものである．つまりは，そのレコードの識別をk個以下しないことで，個人特定の危険性を回避する方法である．しかしながら，k-匿名化のみでは，準識別子に多様性がない場合に特定の可能性が発生する（表2）．そのため，l-多様性では準識別子のグループがある機密属性に対し，少なくともl個以上の異なる値を含むようにグループを分ける．現在集合匿名化などの手法が生まれているが[9]，IoTデータを匿名化するためには第2章にて述べたIoTデータの特性を鑑みる必要がある．

表 1　匿名化前のデータ

Table 1 original data.

ID	名前 （識別子）	年齢 （準識別子）	住まい （準識別子）
1	電大太郎	23	千葉県千葉市
2	千住花子	24	東京都港区
3	埼玉二郎	33	埼玉県蕨市

表 2　k-匿名化後のデータ

Table 2　Data of applied k-anonymization.

ID	名前 （識別子）	年齢 （準識別子）	住まい （準識別子）
1	A	20 代	千葉県
2	B	20 代	東京都
3	C	30 代	埼玉県

3.2 リスクの指標化

IoT データにおけるリスクマネジメントを確立するためには，まずリスクの指標化を行う必要がある．式 1 は IoT リスクの確率を示したものである．

$$RII[f(x)] = [(y - f'(x))\text{^}2] \quad (1)$$

この時，RII とは Risks of Identification by IoT data，すなわちユーザに帰属する IoT データから行う個人推定を指す．式は個人の特定指数である y に対し，関数$f'(x)$は個人に属する IoT データを示しており，関数$f'(x)$に対して匿名化処理を行わない場合に特定リスクが高まることを示している．次に，関数$f'(x)$が何も処理を行っていない場合，つまりは最も個人特定のリスクが高い状態を式（2）に示した．

$$f'(x)max = \{(R1) + (R2) \ldots + (Rn)\} \quad (2)$$

関数$f'(x)$は IoT データの集合体であり，式内の Rn は各種類の IoT データのリスク（Risk）を示す．式（2）は，これらの集合体に匿名化処理を施していない場合を指し，リスクが最も高い状態を示している．

次に，式（3）では，各 IoT データに対して匿名化処理を施した場合を示している．

$$f'(x)min = \{(R1) * (A1) + (R2) * (A2) + \cdots + (Rn) * (A4)\}$$

(3)

式内の An は IoT データに対する匿名化処理（Anonymization）を示している．しかしながら，IoT データの種類は多彩であり，全てのデータもしくは類似の特性を持つデータに対し全て匿名化処理が確立されているわけではない．そのため，匿名化処理には強度があると解釈する．表 3 はそれらの強度を信頼性から指数化したものを示している．

表 3　匿名化処理指標案

Table 3　Indicators of anonymization methods.

匿名化手法	信頼性
運用段階	3
研究段階	2
調査段階	1
確立されていない	0

例えば，既に IoT システム等にて匿名化処理として運用されている段階であれば信頼性が高いものとして便宜的に 3 を，研究段階，調査段階には 2，1 を，また確立していない場合は匿名化処理がないものとして 0 を付随する．つまりは，信頼性が高くなればなるほど個人 y を特定することが難しい状態になるため，匿名化手法においてもその手法の信頼性の確立がなされているものを採用すべきである．

これらの指数は IoT システムの匿名性の強度の証明につながり，処理が施されていない場合の危険性を数値的に示すために存在する．

3.3 指標の明確化

次に，指標の明確化を行うために，現在また今後流通可能性の高い IoT データを表 4 にまとめた．表 4 では，IoT データの分野，またそのデータを使用することによるメリット，またデメリットを明示した．IoT データを使用するためには，その特性だけではなく，データの対象にはメリットと同時にデメリットが存在することを理解して使用する必要がある．メリットが存在するため IoT データは収集されることが明白ではあるが，デメリットが存在すること，またそのデメリットを解消する方法のひとつとして匿名化処理が存在することを明示する．

表 4　IoT データの特性と想定される匿名化処理

Table 4　Advantage and Disadvantage of IoT data and their expected anonymization.

分野	IoT データ	単位	メリット	デメリット	匿名化処理
エネルギー	電力消費量	Wh, W	・デマンドサイドによる電力調整 ・デマンドレスポンスにおけるピークシフト，カットの可能性 ・複数のビル群や地域の電力消費量データの利用により電力融通 ・エネルギー効率の上昇を目的としたスマートビル，ハウスの電力消費量をデータベース化，分析	・電力消費傾向からスマートビル，ハウスの活動状況の把握の可能性が高まる ・利用状況における行動推測 ・在・不在の確認	・データの統計化 ・統計データの秘匿処理の実施 ・1 単位のスマートビル，ハウス計測データではなく，地域などある大きさの単位にデータをまとめた後，データをシステム間に流通される
	ガス	㎥	・スマートメータと同一通信ネットワークを利用し，提供会社から消費者に即時性の高い使用量を提供	・利用状況における行動推測 ・在・不在の確認	・データの統計化 ・統計データの秘匿処理の実施
	水	ℓ	・スマートメータと同一通信ネットワークを利用し，提供会社から消費者に即時性の高い使用量を提供	・利用状況における行動推測 ・在・不在の確認	・データの統計化 ・統計データの秘匿処理の実施
環境データ	温度	℃	・HVAC システムに利用することで設定温度の自動調節が可能になり省エネに活用	・データにおける室内行動推測 ・在・不在の確認	・データの統計化 ・統計データの秘匿処理の実施

	湿度	%	・HVAC システムに利用することで快適な室内湿度の調整	・データにおける室内行動推測 ・湿度が高い土地や室内はカビや腐敗菌を含めた害虫繁殖リスクが生じやすいため購買意欲の低下	・データの統計化 ・統計データの秘匿処理の実施
	照度	lux	・電気エネルギーに変換し電子機器に活用可能 ・データの見える化により照明自体の消費電力削減を努めることが可能	・データにおける室内行動推測 ・在・不在の確認	・データの統計化 ・統計データの秘匿処理の実施
	CO2	ppm	・CO2 排出量の定量化 ・室内の換気状況の把握 ・可視化 ・見える化を行うことで低炭素型都市におけるプレイヤーの削減意欲の向上	・データにおける室内行動推測 ・在・不在の確認	・データの統計化 ・統計データの秘匿処理の実施
	粉塵	cpm	・アレルギー保持者に土地の危険性など示すことによるアレルギー発症の回避	・データにおける室内行動推測 ・在・不在の確認	・データの統計化 ・統計データの秘匿処理の実施
位置・活動データ	BLE	bps	・スマートフォン・デバイスなど流通している機器を用いた収集が容易 ・様々なサービスへの活用	・狭範囲計測のため個人特定が容易	・データの統計化 ・統計データの秘匿処理の実施
	GPS	緯度，経度	・スマートフォン・デバイスなど流通している機器を用いた収集が容易 ・様々なサービスへの活用	・デバイスと個人が結びついていることが多いため個人特定が容易	・データの統計化 ・統計データの秘匿処理の実施

このような表を利用者や運用者に提供することで，リスクを意識しながら安全な運用を行うことを可能とする．また表内の匿名化処理については十分なサーベイが必要であるため，今後の更新が課題となる．

4. 今後の課題とおわりに

本研究では，個人もしくは集団に帰属する IoT データが内包するリスク，またそのリスクの指標化について述べた．今後拡大が予想される IoT データについて，その特性を明確にすることでどのようなリスクが想定されるのかについて明確にすることが，本論文における主たる目的である．

今後の課題として挙げられるのは以下 2 点である．(1) 匿名化処理指標案と IoT データの特性と想定される匿名化処理の更新，(2) システム化である．項目 (1) では，第 2, 3 章でまとめたリスクの指標化の精度の向上，また上述の通り各 IoT データに対する匿名化手法の国内外における調査が必要となる．また研究段階のみならず，実際に運用されているシステムにおいて，どのような IoT データの扱いがなされているのかを明確にする必要がある．もし，IoT システムにおいて何らかの処理が施されていない場合，もしくは施されているものの対象データの保護に相応しくない処理が行われている場合，そのシステムは脆弱性が高いといえる．今後はシステム運用が安全に行われているのか，また必要なデータの質を損なっていないのかを確認できる機構へと進化される必要がある．また，項目 (2) では，これらのリスク指標を簡潔に示すためのシステム化が必要となる．すなわち，使用データの項目，匿名化処理などを入力することでリスク評価を行い，改善点を示せるようシステム化する．このようにシステム化することによって，運用現場での利用を促し，リスクの軽減に寄与する．

謝辞 本研究は，平成 29 年度総務省委託研究開発「スマートコミュニティサービス向け情報通信プラットフォームの研究開発」の一環として実施した．

参考文献

[1] 高橋邦彦, 和泉志津恵, 竹内文乃, 南和宏. 位置情報を用いた疫学研究とその統計的方法. 統計数理, 2014, vol. 60, no.1, p. 3-24.

[2] 佐藤一郎. ビッグデータと個人情報保護法: データシェアリングにおけるパーソナルデータの取り扱い. 情報管理, 国立研究開発法人 科学技術振興機構, 2016, vol. 58, no. 11, p. 828-835.

[3] 鉢呂和紀, 松本邦彦, 澤木昌典. パーソントリップ調査小ゾーンデータを用いた都市特性別トリップ長推計手法に関する研究, 都市計画論文集, 公益社団法人 日本都市計画学会, 2015, vol. 50, no. 3, p. 684-689.

[4] 亀谷哲郎. HEMS から見たスマートメータの標準化動向とデータ活用方法. 電気学会論文誌 C (電子・情報・システム部門誌), 一般社団法人 電気学会, 2013, vol. 133, no. 3, p. 575-578.

[5] 小坂忠義, 小柳和彦, 佐竹保隆, 小林功, et.al,. 空調制御システムに用いるシミュレーション・モデルの簡易化. 電気学会論文誌 C (電子・情報・システム部門誌), 一般社団法人 電気学会, 2017, vol. 137, no. 8, p. 1001-1008.

[6] Yatsuda, A. and Haramaki, T. and Nishino, H., An unsolicited heat stroke alert system for the elderly. Consumer Electronics-Taiwan (ICCE-TW), 2017 IEEE International Conference on. 2017, IEEE, p. 345-346.

[7] 南和宏. プライバシー保護データパブリッシング. 情報処理, 2013, vol. 54, no. 9, p. 938-946.

[8] 上田健揮, 玉井森彦, 荒川豊, 諏訪博彦, et.al. ユーザ位置情報と家電消費電力に基づいた宅内生活行動認識システム. 情報処理学会論文誌, 2016, vol. 57, no. 2, p. 416-425.

[9] 千田浩司, 菊池亮, 濱田浩気, 五十嵐大, et.al. 動的集合匿名化データの有用性評価と開示リスクに関する一考察. コンピュータセキュリティシンポジウム 2013 論文集, no. 4, p. 917-924.