Vol.2022-DPS-192 No.15
Vol.2022-EIP-97 No.15
2022/9/2

クラウドネイティブ型プライバシーエージェントの Docker ベクトル化-サブスクモデルの実現性，メリット，課題

金子 格[1]，　湯田 恵美[1] 岡田 仁志[2]

概要：概発表でプライバシーエージェントの Docker ベクトル化の基本構想を述べた．
ウェアラブルデバイスの大量の高精度生体情報など，大量のセンサー情報の利用が拡大した結果，プライバシーリスクを事前に判断することが困難になっている．パーソナルデータの取得者はデータ主体に利用範囲を透明性と明確性をもって説明することがますます困難になっている．AI 機械学習の発展に伴い，膨大な高精度情報からより多くの分析結果が導き出され，これまで個人を特定できないと思われていたデータから個人を特定できる可能性も完全に排除することが困難になっている．このようにパーソナルデータ提供によるリスクの不確実性が高まる一方で，パーソナルデータ利用の公益的な価値を考えればその利用は阻害されるべきではない．リスクとベネフィットのバランスを取る必要がある．
それを可能にする方法の一つとして，プライバシーエージェントの Docker ベクトル化を提案した．プライバシーエージェントの Docker ベクトル化は，クラウドシステムとの互換性が高く，遅延許諾が可能となる．前回発表でその基本構想を説明したので本報告ではその特徴を考察し，利点と問題を整理する．

キーワード：クラウドコンピューティング，Docker，プライバシーエージェント，サブスクモデル

Docker, a cloud-native privacy agent Vectorization the feasibility, benefits and challenges of the subscription model

Itaru Kaneko[1]，　Emi Yuda[1]

Abstract: .The basic concept of Docker vectorization for privacy agents was presented in a previous report. Expanding the use of substantial amounts of sensor information, such as large amounts of high-precision biometric information on wearable devices, makes it difficult to determine privacy risks in advance. It is becoming increasingly difficult for personal data acquirers to transparently and clearly explain the scope of use to the data subject. With the development of AI machine learning, more analysis results can be derived from vast amounts of high-precision information, and it is difficult to completely exclude the possibility of identifying an individual from data that was previously thought to be unidentifiable.
While the risk uncertainty due to the provision of personal data increases in this way, the use of personal data should not be hindered given the public interest value of the use. You need to balance the risks and benefits. As one of the ways to make that possible, we are proposing Docker vectorization of privacy agents. Privacy agent Docker vectorization is highly compatible with cloud systems and can be operate as web service to provide personal data and their deliverable. Since I explained the basic concept in the previous presentation, this report will consider its characteristics and sort out the advantages and problems.

[1] 東北大学データ駆動科学・AI 教育研究センター, Center for Data-driven Science and Artificial Intelligence Tohoku University
[2] 国立情報学研究所 情報社会相関研究系

Vol.2022-DPS-192 No.15
Vol.2022-EIP-97 No.15
2022/9/2

1. はじめに

先行発表では，プライバシーエージェントの一実装法として，Docker ベクトル化を提案した．まずその意図を再度簡単に説明する．

ビッグデータとその AI 機械学習による利用は今後さらに成長が期待されるが，個人情報保護において複雑な問題を提起すると考えられる．たとえばリスト型ウェアラブルセンサーからは高精度で大量の生体情報を取得できるようになった．その一方で心拍や加速度などから機械学習などの方法で健康状態や日常習慣などを推定できことも知られている[3] -[7] ．こうした大量のパーソナルデータを集積すれば医療や公益に役立つ様々な情報が得られると期待される一方，現状ではデータの提供に対する不安を完全に除外することは難しい．

スマートスピーカが集める個人情報に関する疑念がその普及の障害になったように，パーソナルデータの不本意な利用への懸念が増大すれば，こうした利用の普及を減速させる要因となり得る．データを共有し分析することから得られる情報の価値を考えれば，そのような懸念やリスクを最小化することにより，共有されるデータを増大することが望ましい．

そこで中川は個人データの収集，管理，保護をおこなうパーソナル AI エージェントを提唱した[2] ．データ主体の個人データとその利用条件をパーソナル AI エージェントが代行管理する．高度な判断をパーソナル AI エージェントが代行することにより，従来の枠組みよりもパーソナルデータの利用条件を柔軟に指定しうる点で，魅力的な提案である．

Docker ベクトル化は同じアイデアを具体的実現方法の側面から提案するものである．その提案は大きく以下 3 点からなる．

(1) クラウドと Docker のメカニズムで実現する

(2) パーソナルデータの提供に遅延許諾を含める

(3) 多面的安全性(ゼロトラストセキュリティ)を備える

先の報告では主にその導入動機と基本構成について述べたが今回の報告では(1)～(3)の特徴について，実現性，メリット，課題についてより詳細に議論し整理する．

2. クラウドと Docker コンテナによる実現

クラウド(クラウドコンピューティング)と Docker についてもごく簡単に説明しておく．いうまでもなく，今日多くの情報処理はクラウドプラットフォームと Docker コンテナという構成でコンピューティングリソースの分散共有を実現している．

Docker コンテナはコンテナ化されたアプリケーションがカーネルプロセスを共有できることで従来 VM を利用したコンピューティングリソースの共有にくらべ，圧倒的に軽量なコンピューティングリソースの分散共有を実現した. Docker コンテナ化されたアプリケーションの実行はプロセスを起動するだけ(メモリ割り当てすら不要)なので，起動のためのオーバーヘッドは非常に小さい．

すでに Docker コンテナ技術はクラウドプラットフォームでは標準的に実現され，様々なクラウドシステムで共通に利用できる．このフォーマットで実装すれば，新たに複雑な標準規格を定めなくてもモジュール化された複数のプライバシーエージェントや多面的安全性を実現した本方式を敏速に導入することが可能である点が，大きな魅力がある．

3. パーソナルデータの遅延許諾の必要性

筆者等はデータ主体に提供すべきパーソナルデータの保護手段として，これまでの「データ提供の自己決定権」から一歩進めて「分析結果の現在の利用状況の自己決定権」を提供することを提案している．なぜそれが望ましいかを説明する．

ウェアラブルデバイスに含まれるパーソナルデータの場合，将来どのような統計分析が可能であるか予想が困難だ．

作家コナン・ドイルが生んだシャーロック・ホームズは依頼者にこのように発言する．「いや何，わからない．この方が過去，手先を使う仕事にしばらく従事していらっしゃったこと．嗅ぎ煙草を愛用していらっしゃること．フリーメイソンの一員でいらっしゃること．中国にもいらっしゃったこと．近頃，相当な量の書きものをなさったこと――これだけははっきりとわかるのだが，後はまったくわからない．」同席した人々は驚き，いったいどれだけのことをホームズが推理できるか計り知れないと感じるのである．

AI・機械学習の推定能力は今後も急速に進歩するから同じような不確実性が生まれる．そのため筆者等は AI・機械学習の推定能力に不確定があるという点を「シャーロック問題」と呼ぶことを提唱している． パーソナルデータ保護には「シャーロック問題」をどのように解消するかが課題となる．

中川のパーソナル AI エージェントを AI　に限

Vol.2022-DPS-192 No.15
Vol.2022-EIP-97 No.15
2022/9/2

らずパーソナルデータ利用の代行と考えた場合も，もし一旦利用を許諾した情報が恒久的に利用されるとすればその不確実性は解消されない．

しかしパーソナルデータの利用許諾を，データ提供の許諾ではなく提供したデータから得られる派生情報の利用時まで待って，どのような派生情報をだれに提供するかが明確になった段階で「遅延許諾」する方式とすれば，どのような分析結果が利用されるかという点に対する不確定性はなくなり，またもしデータ主体がそのような情報提供に不安を感じればいつでもデータ提供を停止できるから，データ主体が将来の分析結果を予想できないという「シャーロック問題」の解消に役立つ．

4. Docker ベクトル化されたパーソナルエージェントの構成

4.1. パーソナルデータ利用の Web サービス化

次にプライバシーエージェントの Docker ベクトル化の利点を整理刷る．

Docker ベクトル化は，プライバシーエージェントが Docker コンテナ化される．これはいいかえれば，パーソナルデータとその分析結果は常に Docker コンテナ化された Web サービスの「内部」で管理され，定められた認証手順によって外部に提供されることを意味する．パーソナルデータおよびその分析結果はこうして常にデータ主体の管理化におかれ，外部からは隠ぺいされている．パーソナルデータやその分析結果はセキュアにカプセル化されたオブジェクトとして流通交換されるという見方もできる．

データ主体ガ自己のパーソナルデータを web サービスによって管理する，という考え方はすでにマイナポータルに前例がある．マイナポータルは氏名住所などのパーソナルデータ(個人情報でもある)へのアクセスをデータ主体に通知する機能を持つが，Docker ベクトル化したプライバシーエージェントは素のデータ提供だけでなくデータの分析結果をも管理する．

このような仕組みは一見オーバーヘッドを重くしデータ処理のコストを増大するように思われるが実際はそうではないことを次に説明する．

現在のデータ処理はほとんどの場合オンラインリアルタイムで行われる．たとえばコロナ感染者の感染者数は常時変動するが，様々な集計分析システムはデータ提供をうけたら一度だけ処理をするのではなく，データが更新されるたび，問い合わせが発生するたび，あるいは一定間隔で常に最新データに基づいて再度集計，分析される．したがってパーソナルデータの最新のデータベースに基づいて定期的，あるいはオンデマンドで再集計を行うことはほとんどの場合，新たなオーバーヘッドとはならず，従来のデータ処理手順の中で行うことができる．

Docker ベクトル化されたプライバシーエージェントは，この再計算において新たな許諾情報に基づいてその時点において個人情報の提供に同意しているデータのみを用いて集計や分析を行うことになる．

4.2. Docker Container 化

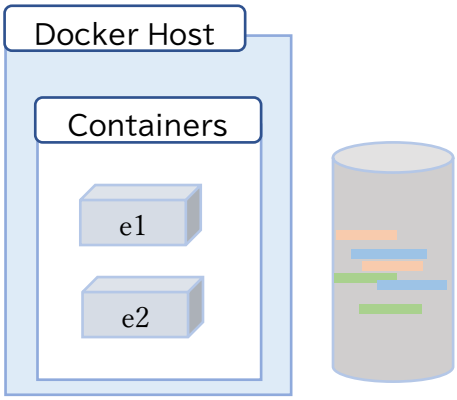
Docker ベクトル化されたプライバシーエージェントは図 1 に示すよう Docker Host 下のコンテナとして実行するのが可能だと考えられる．

クラウド環境では Verifier や Component はまずイメージとして作成され，Docker Container 中に含めて実行することが考えられる．Verifier と Component がイメージとして提供され各個人の要望に応じてコンテナ化されていれば，各個人のパーソナルデータに関する要求事項にあわせた処理を行うことが可能であろう．

この方法はあらゆる処理や認証方法をパーソナルデータに結び付けることが可能であるから，非常にフレキシブルで強力である．

また Docker Container はそれほど大きくなく(数 100Mb におさまり)，Docker Host は Docker Container に含まれているイメージを実際にいつもロードするわけではなく，同じイメージがすでにメモリ中に存在すればそれを共用することが可能であるため，同じイメージの集合から構成あされる膨大な Docker Container を起動・実行してもそのメモリコストや計算機コストはごくわずかに抑えることが可能である．

したがって，個々のパーソナルデータに Docker Container が添付されている，という構成は構造をよくしらなければやや冗長に見えるが，実際にはきわめて効率的に実行することが可能であると考えられる．



Vol.2022-DPS-192 No.15
Vol.2022-EIP-97 No.15
2022/9/2

図 1 Docker Container 化されたプライバシーエージェント

4.3. Docker Vector 化

とはいえ Docker Container は必ずしも万能とはいえない. Docker Container のサイズはイメージがコンパクトになったとはいえ数 100Mb に及ぶ場合がある．最低でも数 GB のゲノム情報本体の記述データとしては数 100Mb の Docker Container のオーバーヘッドは許容範囲だろう．しかしウエアラブル機器から得られるデータは，たとえば毎時間の体温といった場合たかだか 1kb 程度のことがある．1kb のデータ本体に対し数 100MB のオーバーヘッドは大きすぎる. そこで，Docker Vector 化することを考える．

Docker Vector 化を図 2 に示す．Docker Container をそのまま扱うのではなく，Docker Container の記述フォーマットに基づくディスクリプタ PPPD(Privacy Protection Policy Descriptor)で記述する．PPPD にはプライバシーエージェントを実現する Docker Container 自体は含めず, あらかじめ登録した Docker Container の識別子とパーソナルデータ管理のパラメータを含める．PPPD は json などで記述する．

プライバシーエージェントを実現する Docker Container を識別子で表すことで PPPD の符号化効率は大幅に向上する．一方 128bit 程度の bit 幅があれば，全人類に対し毎秒 1 種類の新しいプライバシーエージェントをあらたに定義しても符号空間はまったく枯渇する心配はない．

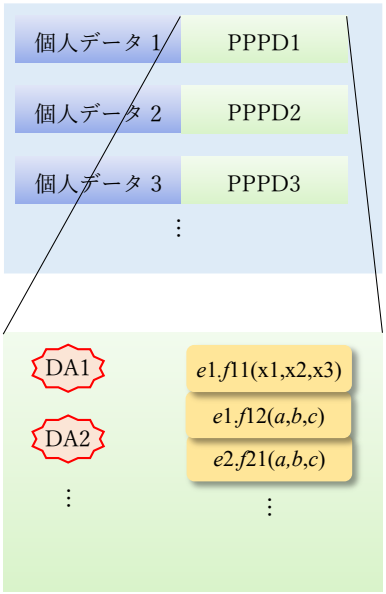


図 2 Docker Vector 化

5. 多面的安全性とゼロトラストセキュリティ

多面的安全性はゼロトラストセキュリティに基づく IoT デバイスの軽量なセキュリティ実現手法でその特徴は以下 3 点である．

(1) IoT デバイスのセキュリティをデジタル認証によって軽量に行う
(2) IoT デバイスが多重認証を受けることができることを保障できる
(3) 認証システムは相互認証が可能である

報告者は多面的セキュリティモデルによりシステムを構成する部品, Component の信頼性を評価する方法を示した[9] [10] ． に多面的セキュリティモデルによる Verifier の相互認証を示す．

相互認証により Verifier は相互にセキュリティ認証を行う．Verifier は Verification 性能により報酬を与えることによって Verification 性能を高めるインセンティブが与えられている．そしてこのような一定のインセンティブがあり，Verifier の互換性によって Verifier の交換コストがゼロであれば，ゲーム理論的上は Verifier が相互にセキュリティ不良の検出に全力を尽くすことがナッシュ均衡解となる．つまり Verifier 群はお互いにセキュリティ不良の検出に全力を尽くすので，セキュリティ不良が見逃される可能性を小さくすることができる．

コンポーネントの多面的セキュリティモデルによる認証をに示す．

複数 Verifier から複数 Components を Verify する．すべての Verifier がすべての Component を Verify しなくても Verifier は他の Verifier の信頼性を評価できる．また各 Component はより多くの Verifier から Verification されることでより高い信頼性を得，このフレームワークによって各 Component は信頼性の指標を得ることができる．

筆者等はこのような枠組みと理論的分析の詳細を[9] [10] で論じている．

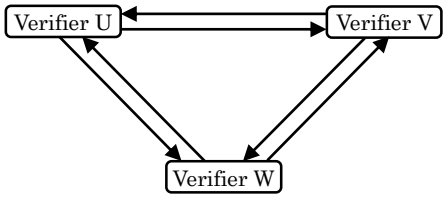


図 3 多面的セキュリティモデルの相互認証

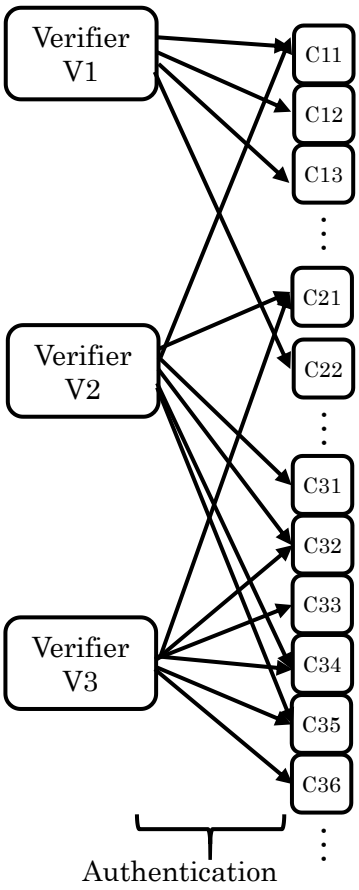


図 4 Component の多面的な認証

6. 多面的安全性のために実現すべき機構

次に，クラウドと Docker の環境において多面的安全性を実現するために機構的には何を追加すべきか，について検討する．

多面的安全性の実現では以下の課題があった．

(1) モジュール化された Component を自在に組み合わせて導入，有効化，実行できる実行環境が必要である

(2) そのような Component 実行環境がきわめて効率的に実装されている必要がある

(3) そのような Component 実行環境がプラットフォーム非依存で実現している必要がある

いうまでもなくそのような Component 実行環境は従来存在しなかったしそれを整備するには非常に大きな初期投資が必要だった．この問題が解決しなければ多面的安全性は机上のアイデアにすぎなかった．

しかし，前述のクラウドと Docker コンテナの仕組みは多面的安全性の Component 実行環境の要求条件をすでに満たしている．これについては本報告では詳しく述べる余裕がないので省略するが，多面的安全性の Component の実行環境は Docker Container そのものでよいことがわかる．単に Docker Container でアプリケーションを構築する場合との違いは，単にそれを用いて多重認証や遅延許諾を行うか，そしてそれがデータ主体にわかりやすく提示されるか，という点だけである．

つまり，クラウド環境において多面的安全性が実現されるために必要な実行環境はすでにクラウドサービスには備えられている．Docker ベクトル化されたプライバシーエージェントが多面的安全性を提供するには単純に認証機関が複数存在すればよいだけである．

7. サブスクモデルの実現

最後に，Docker ベクトル化によりパーソナルデータ利用のサブスクモデルが実現できることを指摘しその利点を述べる．

提案手法では Docker ベクトル化されたプライバシーエージェントはパーソナルデータの利用を提供する web サービスとして機能する．したがって当然これに対価を設定してそれをパーソナルデータの提供者に直接分配することが考えられる．

これまでは仮にパーソナルデータの取得に何等かのコストをかける場合，それはデータの提供を受けるものが一旦負担する必要があった．一方もしも提供したパーソナルデータが将来大きな価値を持つと考えられる場合には，将来の分配を対価として収集する方が低コストでデータを収集することが可能であると考えられる．

もしデータ主体によりもデータの提供を受ける側がパーソナルデータの提供の価値を高く評価する場合はこれまで同様データ提供を受ける時点でなんらかの対価を支払えばよく，データ主体が高く評価する場合は将来の対価支払いへの期待によりデータ提供を得ることができ，パーソナルデータ提供に対する補償の方法に柔軟性を持たせることが可能となる．

8. まとめ

概発表の Docker ベクトル化されたプライバシーエージェントについて，その特徴に基づき，実現性，メリット、課題を整理した．提案方式が様々なクラウドプラットフォームを，特別な付加的なシステムなく実行環境として利用することができることを説明した．今日普及しているクラウドプラットフォームによって敏速に導入でき効率的に実行できる点が大きな利点であることを説

Vol.2022-DPS-192 No.15
Vol.2022-EIP-97 No.15
2022/9/2

明した．次に前回報告でも言及しているが，パーソナルデータの遅延許諾の必要性について説明した．次に Docker ベクトル化されたパーソナルエージェントの構成について説明した．次に多面的安全性を Docker コンテナの実行環境でどのように実現するかを再度説明し，クラウドプラットフォーム上で問題なく実現できることを示した．最後にサブスクモデルを導入できることを示しその利点を示した．

謝辞:　本研究は 2022 年度 国立情報学研究所 公募型共同研究（戦略研究公募型研究）の助成を受けて実施した．

参考文献:

[1] 金子格，湯田恵美，大量ウェアラブルデバイスと大規模生体情報時代における AI 機械学習のシャーロック問題への対策としてのプライバシーエージェントの Docker ベクトル化，情報処理学会研究報告，Vol. 2022-EIP-95, No 1(2022)
[2] 中川 裕志，AI 倫理指針の動向とパーソナル AI エージェント，情報通信政策研究/3 巻 (2019) 2 号 (2019)
[3] IDC, 2020 年第 3 四半期 世界／国内ウェアラブルデバイス市場規模を発表, https://www.idc.com/getdoc.jsp?containerId=prJPJ47221920(2021/1 取得)
[4] Junichiro Hayano, Tetsuya Tanabiki, Shinichiro Iwata, Katsumi Abe, Emi Yuda. Estimation of Office Worker's Emotion Types Using Two-dimensional Model Consisted of Biometric Signals, INTERNATIONAL JOURNAL OF AFFECTIVE ENGINEERING 20(2) 105-110 (2021)
[5] Itaru Kaneko, Yutaka Yoshida, Emi Yuda, Junichiro Hayano. Sensing of Microvascular Vasomotion Using Consumer Camera. Sensors (Basel, Switzerland) 21(18) 6256-6256 (2021)
[6] 金子格，湯田恵美，吉田豊，ホルター心電計の内蔵加速度センサーを用いた活動推定の改良と個人情報保護に関する考察，情報処理学会研究報告
[7] Stephen Paul Marsh,Formalising Trust as a Computational Concept, Computing Science and Mathematics eTheses, http://hdl.handle.net/1893/2010 (1994)
[8] L. Mui,M. Mohtashemi,A. Halberstadt,A computational model of trust and reputation, Proceedings of the 35th Annual Hawaii International Conference on System Sciences(2002)
[9] I. Kaneko; K. Shirai, The multi-lateral security framework for the ubiquitous audiovisual services, 2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace
[10] 金子格，確率的多面的セキュリティモデルとブロックチェーンを用いたメディア IoT 向け軽量セキュリティフレームワーク，電気学会論文誌C（電子・情報・システム部門誌）/137 巻 (2017) 6 号