法と技術シンポジウム（第２回）
2018/2/19　　一橋講堂

コネクティッドカーの技術的課題

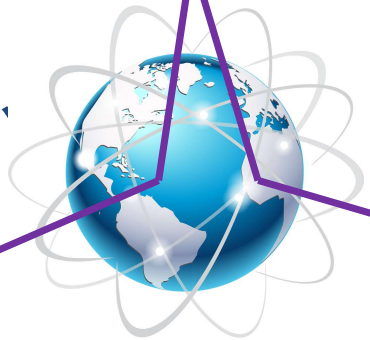
―機械学習とプライバシー保護―

中川 裕志
東京大学／理研AIP・社会における人工知能研究グループ

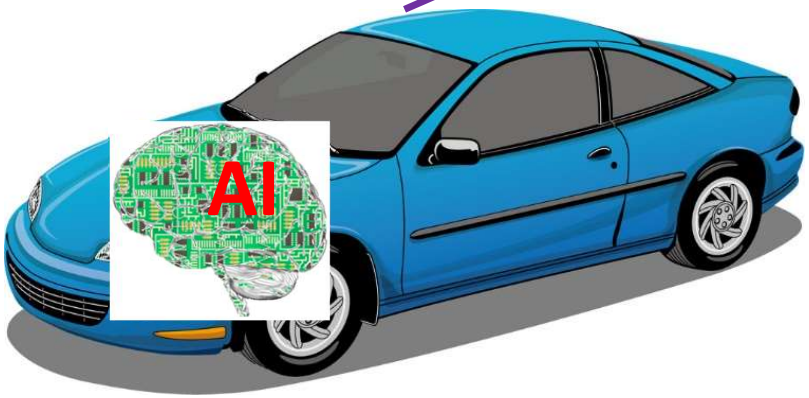
Connected　が先か
AIが先か？

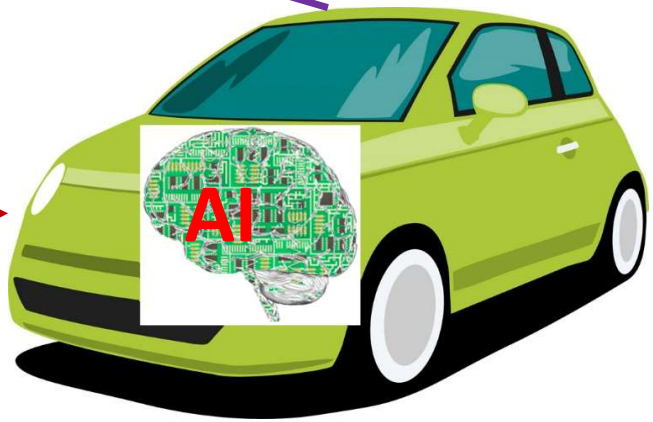
connected via Internet

インターネットは
遅いし、危ない奴が多い

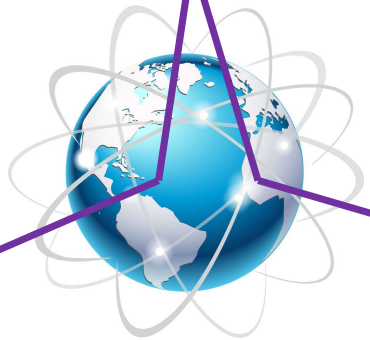


- サーバと車の双方でセキュリティの厳重管理が必要
- 厳重なインターネット ID認証も必須
  - OpenID Connect など



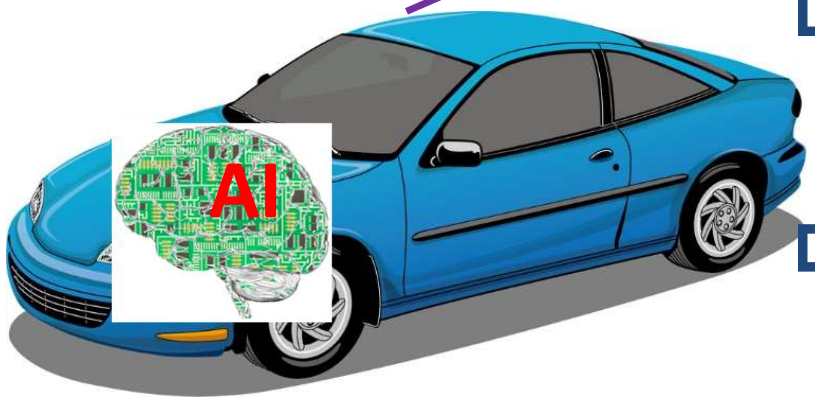


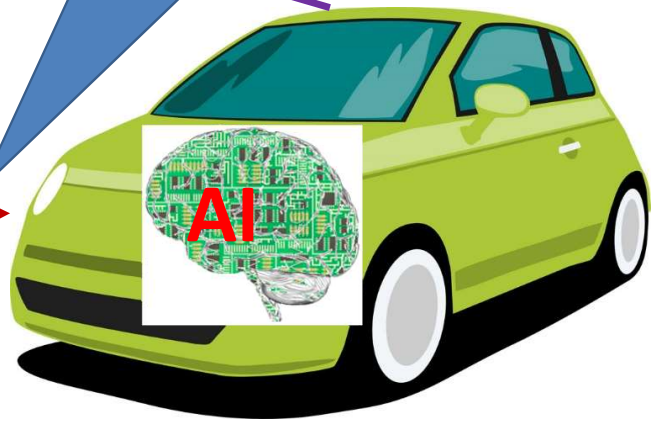
connected via
car to car directly



- 相手の車が悪い奴でない確認が必要
- インターネット経由しないID認証　が必要だ
- AIで解決できるかどうか

- 車車間通信は早い
- プロトコルの統一が必要





AIとconnected の相互作用

- connected via internetにしても、connected via car-to-carにしても
  - 高度なセキュリティチェックが必要
    - → プロトコルが決まったID認証より高度な処理が必要ならAIの出番

- 車載AIが強力になりサービスが高度化
  - ネットワーク経由での情報授受が必要
    - 高精度かつ最新の地図データなど

  - どちらにしてもセキュリティ確保が絶対必要

ネットワークの脆弱性

- ネットワークの世界の常識は
  - 攻撃側が強すぎて100%の防御は不可能
  - 破られたときの早期発見、早期現状復帰の方策に重点あり

- Connected carのセキュリティの特殊性
  - 走行中に車では対応策の遅れが許されない
  - 対応速度の速い異常値検出技術が必須→ AI,機械学習
    - 誤検出を許す
    - 攻撃パタンを常に最新なものに更新

    - 少しでも危ない兆候が見えたら停止
    - → 停止が最大の防御？　急停止はかえって危険？　→ 大きな課題

総務省　AIネットワーク推進会議

- 2年前から行われ、昨年7月にOECDにAIに関する提言を行った総務省のAIネットワーク推進会議
  - 多くの指針が単独のAIについての議論だったが
- 最初から複数のAIが織りなすネットワーク＝AIネットワークにおける安全性や倫理指針を検討

➢Connectedの光と影をきちんと調査すべき時期

➢Connected car では特に重要

走行支援の3本柱とConnected の必要性

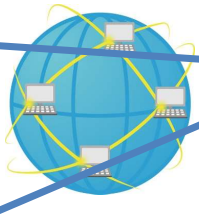
- センシング
  - 360度の視界でカメラ＋レーダ
  - ◆ connected 不要

- 地図
  - その場その場で10cmの精度で、走行可能経路を探す
  - ◆ connected 必要：自車走行経路データのupload
  - ◆ 走行予定場所のup-to-dateの地図のdownload

- 走行ポリシー（戦略）
  - 周囲の多数の車で対話、交渉しながらお互いの経路を確認
  - ◆この会話がconnectedになるべきか、可能かは未知数
  - ◆Internet connectedでは遅すぎるので、車々間の直接通信になるだろう

Connected Carの地図データフローのビジネスモデル

Automotive Car Data Collector



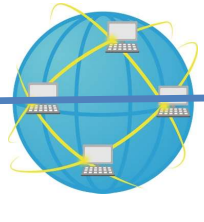
Road Segment Data

RSD Integrater

他のRoad Segment Data

Integrated RSD

Online Map



Map Maker

MobileEye社の方針に関しては
Prof. Amnon Shashua
Chief Executive Officer and CTO
の講演より

センシング vs 運転走行ポリシー

- ◆レーダやカメラからのセンシングデータ入力と処理
  - ◆いろいろな状況に対する既存のデータが大量に存在 （標識、歩行者、他車　etc）
  - ➢これを正解データとした教師あり学習で対応
    - ➢深層学習も使える
    - ➢理研AIPにはトップ研究者所属

- ◆運転状況に即応する運転走行ポリシー

  - ➢未来の事象を予測し、対応する技術　→ このためには、connectedで地図情報収集することが必須と考えられている

  - ◆その場その場で学習結果のモデルを変えていく → connected でモデルアップロードもありえる

  - ◆種々の要素の重要度を経験しつつ変更→ 強化学習が有効 → 学習するのは個々の自動車か、データ収集した中央サーバか？　付録1参照

走行プランのモデル改善：
強化学習が有望なのだが

- 既存のデータから学習した予測器
  - → 予測とアクション が外部環境を変える
  - ex 方向を変えると進行方向上の状況が大きく変化

  - データは予め集めておくことができない： 状況が多様過ぎるので

- 運転走行ポリシーを更新するには新規データが必要

  - カバー範囲の広いデータは収集できるのか？
  - 新規データに即応した運転走行ポリシーの変更は可能か？

車の安全走行はハードウェアだけの問題か？

- 多数の車は近くで走行しているというマルチ・エージェント環境での事故確率の評価ができていない

✓走行経路のプランニングの失敗　→　事故

✓センサーがモノを見落とす（認識失敗、あるいは誤認識）　→　事故

- 動作の即時性の要請が厳しいので、connected　で中央サーバと通信して解決というわけにはいかない。
  - ただし、結果を中央サーバに収集し、オフラインで機械学習に利活用することは可能。

- 近接車との車車間通信が利用できるとベスト
  - 車車間通信プロトコル＋走行プランのネゴプロトコルの企画統一化が必要（道は長い）

昔からの知恵だと

- 安全性は、統計量によって決まる

➢収集したデータが大きくなると統計量の信頼性が増し、安全性は改善

➢走行プランをAIが計算する場合、

➢収集する走行データの距離が長くなればなるほど　安全性は改善

➢という昔からの知恵は間違っているらしい！

- **Prof. Amnon Shashua**
- **MobileEye社　Chief Executive Officer and CTO　の講演より**

アメリカのケースの試算

- 1時間あたりの死亡事故確率～10^{-6}
- 年間死亡事故　35,000件
- AIが走行にコミットした場合、社会的に受容できる年間死亡事故～10ないし100

➢自動運転車の1時間あたりの死亡事故確率～10^{-9} にしたい

- 希な事象の発生確率をpにするためには　O(1/p)　の教師データ必要

➢死亡事故確率～10^{-9} を達成する走行プランニングの変更のためには　10^9時間　のリアルな走行データが必要。

➢ちなみに1日は24時間 → 4x 10^9日　→ 1千万人がまる1年ぶっ通しで運転

➢　→ 不可能

外部環境データ自動生成

- Mobile-Eyeは10^9時間　のリアルな走行データ収集は不可能という

- しかし、最近の機械学習では、外部環境をランダムに自動生成して機械学習に使う技術が大きく進展
- この方法で10^9時間が実現できないか？

  - 例：学習した分類システムに雑音を加算したデータを与えると間違うことあり
  - 例：深層学習における GAN(Generative Adversarial Network)

- ただし、車の運転状況は非常に多様なので、運転状況ランダム生成でカバーしきれるか疑問

➢「リアルデータ　対　自動生成データ」　は研究開発の肝になるかもしれない

安全性はモデルベース、解釈可能、説明可能な方法であるべき

- 100%の安全は不可能
- ポイントは事故において責任者を明確にできること

MobileEye社の方針に関しては
Prof. Amnon Shashua
Chief Executive Officer and CTO
の講演より

- 運転走行ポリシーの解釈可能性、説明可能性が必須
  - 得られるのは数値データの羅列にみえるような中身の理解しにくい機械学習が馴染まないのではないか
  - ランダム生成した外部環境の解釈ができないと使えない。
  - この解釈は人間には難しそう → 人工知能にできるかが分水嶺

- 外部の（現在環境と予測環境）と自車の状態を入力にし、予測された将来環境が記述できる意味言語が必要

安全性はモデルベース、解釈可能、説明可能な方法であるべき

- 意味言語で記述された現在環境と予測から、自車が走行しても事故を起こさない走行プランを生成するシステム
  - 走行プランの合理性を　意味言語　によって説明できることが必須。

- 常識的で責任の明確な安全性：具体的には以下の通り
  - 責任を明確化するルールを事前に策定
  - 事故における責任者を明確化する常識の理論的定式化を行う
  - 自車の責任になる事故を起こす可能性のある走行プランニングを絶対に誘発しないルール開発が必須
  - → 法的責任を生じないモデル開発

Connected will help

- 責任を明確化するルールを事前に策定しておくにしても

- 車の内部、外部の環境に関するデータを多く収集して機械学習したモデルを利用することによって責任の明確化が容易になる
  - この学習データ収集が不十分だと立場が弱くなる

- データ収集は即時的な car-to-car direct connected にしても、事前にデータをup-down loadするinternet connected にしても

➢Connected は必須かつ強力であるほど望ましい

プライバシー保護

運転履歴、運転車・搭乗者データ(位置データ)に関して考慮すべき事項: その1

- どのようなデータがプライバシー保護を留意すべきか
- 走行中に得られた運転履歴、運転車・搭乗者データ(位置データ)をどの国に設置されたサーバに保存するか
  - EUならEU内設置のサーバ　:GDPR:パーソナルデータの越境禁止

- 収集した膨大な走行履歴、運転車・搭乗者データを収集したサーバから越境移動　→　現実的はないかも

運転履歴、運転車・搭乗者データ(位置データ)に関して考慮すべき事項： その2

- 収集した膨大な走行履歴、運転車・搭乗者データに対して
- 収集した現地サーバでできること

➢仮名化、匿名化：ただし、この処理だけではEUからの持ち出しは不可

➢統計処理、機械学習したモデル

◆複数の機械学習のモデルを統合してより性能のよいモデルを作る

◆理研AIPの機械学習グループがこのテーマでは日本のトップランナー研究者集団

運転者、同乗者、歩行者
などのプライバシー保護

運転履歴を本当に収集されたくない場合

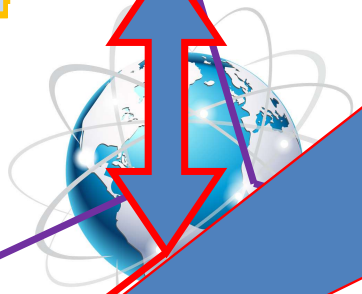
- 情報収集側は自動車メーカ、道路管理者、自治体など

- 隠したいのは
  - 走行した場所
  - 運転者のID
  - 同乗者のID

- 運転者（～車の所有者）は情報収集方法に購入時に同意
  - ただし、隠したいときはテンポラリーに情報収集を簡単にオプトアウトする仕掛けが重要
  - オプトアウトが面倒だと、(EU)では法的に問題視される

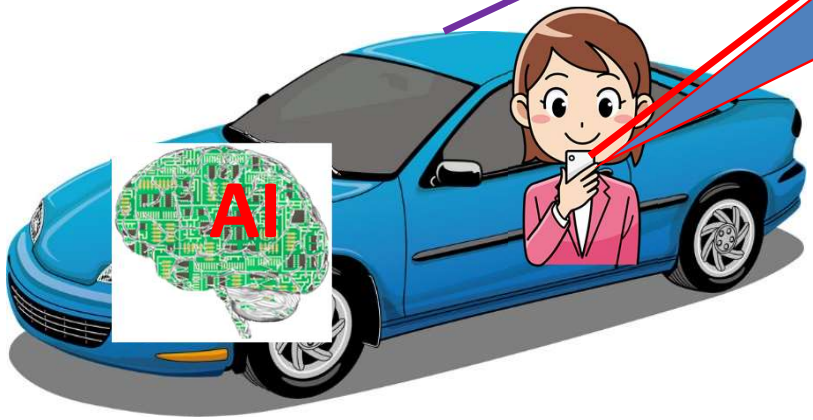
connected via
person on the

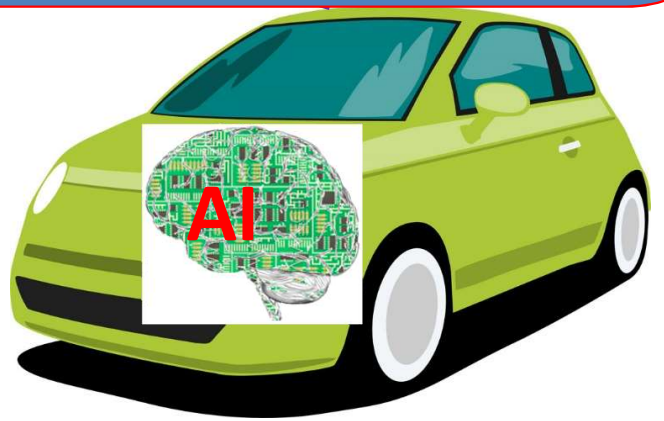
位置情報のプライバシー



車の位置情報をオプトアウトしていても、
1．ある人が搭乗していることが分かり、
2．搭乗者の位置がスマホから分かると、
車の位置が漏洩してしまう。

搭乗者が車内の空間にいる以上、位置情報の漏洩は防ぎようがない





運転履歴を本当に収集されたくない場合

- 運転者を同定しないで運転許可することの是非
  - 制度的ないし運用的議論が必要

- 一方、ダイナミックな地図情報をダウンロードしながら走行する場合は、走行位置を隠しておけるか？

走行情報の匿名化収集

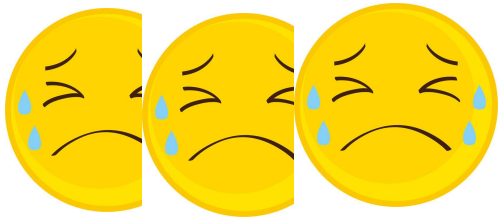
- 情報収集側に運転車がだれであるかを知らせずに走行情報だけを与える場合に運転車側が対応できる技術
  - グループ化　　付録1
  - Tor　　付録2
  - これらは付録を参照されたい
  - なお、データベースに蓄積された走行履歴情報については後で触れる

周囲の車の情報

- 周囲の車とのコネクションをする場合
  - 周囲の車の車体IDと運転者ID
  - 周囲の車が走行データ収集拒否モードの場合もありえる

  - 周囲の車とは直接コネクトしない
  - → お互いに推測で動いてしまう危険性あり

- 情報は画像入力のみ
  - 人間なら画像を見ても許されるが、カメラで映像を得てしまった場合のプライバシー処理の問題
  - 刑事事件の証拠として利用できる
    - 街中の監視カメラと同じ。移動監視カメラがうようよ……



データベースに蓄積された個人走行データの匿名化とデータ流通・利用

仮名化　走行データレコードの分割とシャッフリング

個人識別子(姓名)	Loc. 1	Loc.2	Loc.3	…
赤川五郎	東京	渋谷	麻布	…
青山全蔵	台場	豊洲	新橋	…
大岩倉之助	…	…	….	….
徳川信長	xy	yz	zw	…
福田幸吉				

- レコードを分割し、分割された部分ごとに仮名を変更
- 順番もシャッフル
- 個人識別は困難

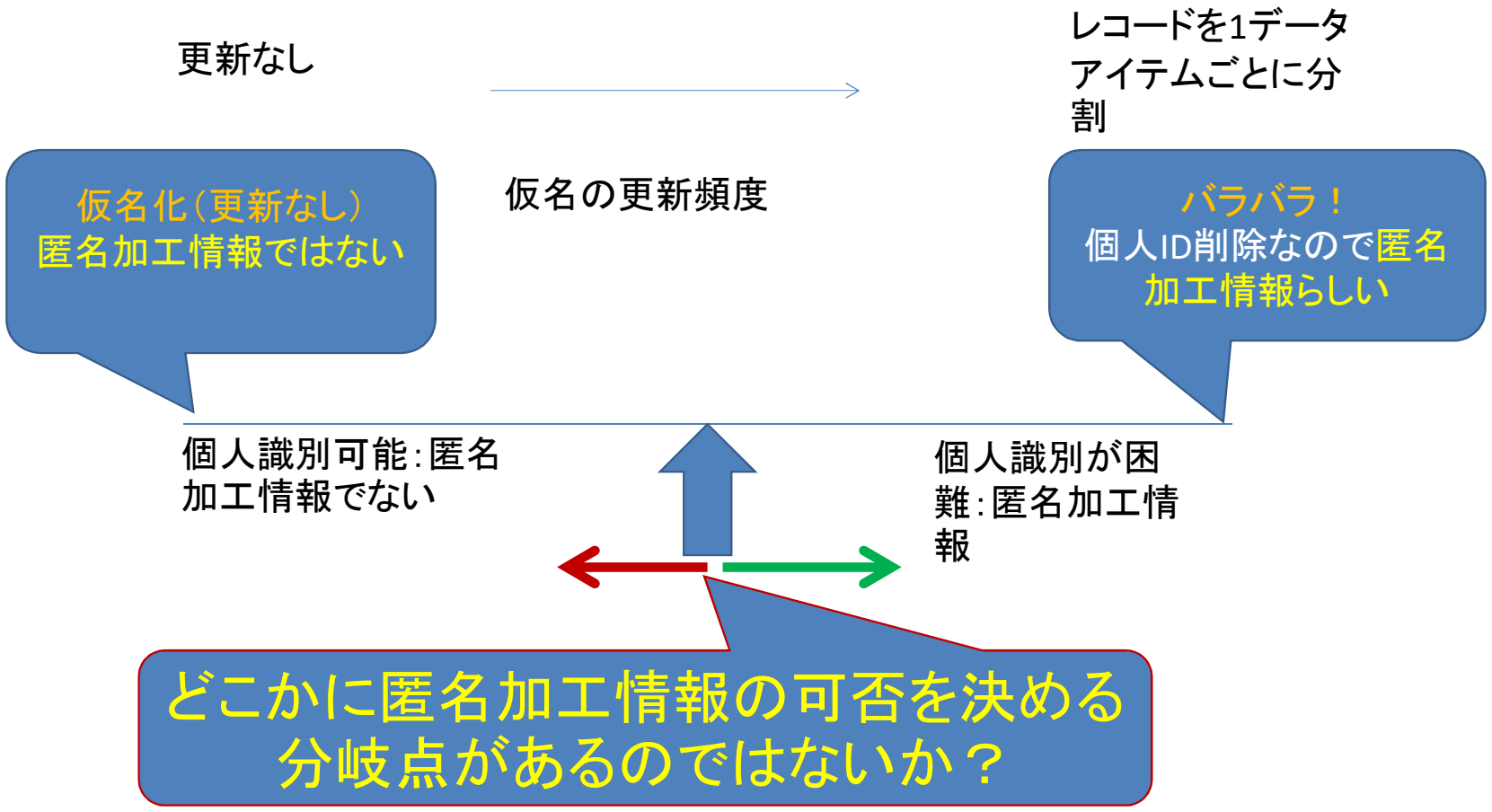
仮名	Loc. 1	Loc.2	Loc.3	…
A123	東京	渋谷	麻布	…
C125	台場	豊洲	新橋	…
A234	…	…	….	….
345X	xy	yz	zw	…
B650				

Loc. 1	Loc.2
東京	渋谷
…	…
xy	yz
台場	豊洲

Loc.3	Loc.4
…	…
麻布	新宿
新橋	品川
…	…
zw	wa

…
…
…
….
…

改正個人情報保護法で導入された匿名加工情報：
匿名加工で、個人識別困難にし（容易照合できない）、個人データではないとみなして、データ主体（個人）の同意なく、自由に流通して利用できる



匿名化する場合の考慮事項

- 個人特定性の低減　VS　データ利用価値の維持

- 長大な走行履歴データを個人特定性のないものにできるか？
- 技術的に難しい。

- 同意を中心にした走行履歴データ収集にならざるをえないが、同乗者の個人データなどの問題は残る

顔画像のプライバシー保護

- If 顔認識しなければ個人データではない → プライバシー的な問題はない

- 認識とは特徴量抽出を行い、個人識別できる＝個人識別情報にすること

- 運転者
  - 認証に顔認識を使うかもしれない。
  - ただし、車の持ち主以外が運転する場合は運転の前に同意が必要

  - 運転時の心理状態や視線などを収集できない可能性あり

- 同乗者

  - 認識してはいけない。（偶然の認識はありえる）
  - ただし、場合によっては認識しておいたほうがよい場合もある。
  - 同乗者の顔画像収集への同意は可能か？

参考：車外の人の顔画像のプライバシー保護

- 事前告知ができない

- If 顔認識しなければ個人データではない → プライバシー的な問題はない

- 外部向けのカメラへの映り込み → 参考情報：次のスライドの表
- 車が走行すること自体が事前告知とみなせればよいが、現実的には無理

- 歩行者は公共空間で取得した顔画像
- 対向車線の車、直近の周囲の車の運転者＝公共空間のデータ？

➢利活用は個人情報保護委員会と相談すべき

IoT推進コンソーシアムの利用基準 version 1(H29), version 2(H30)

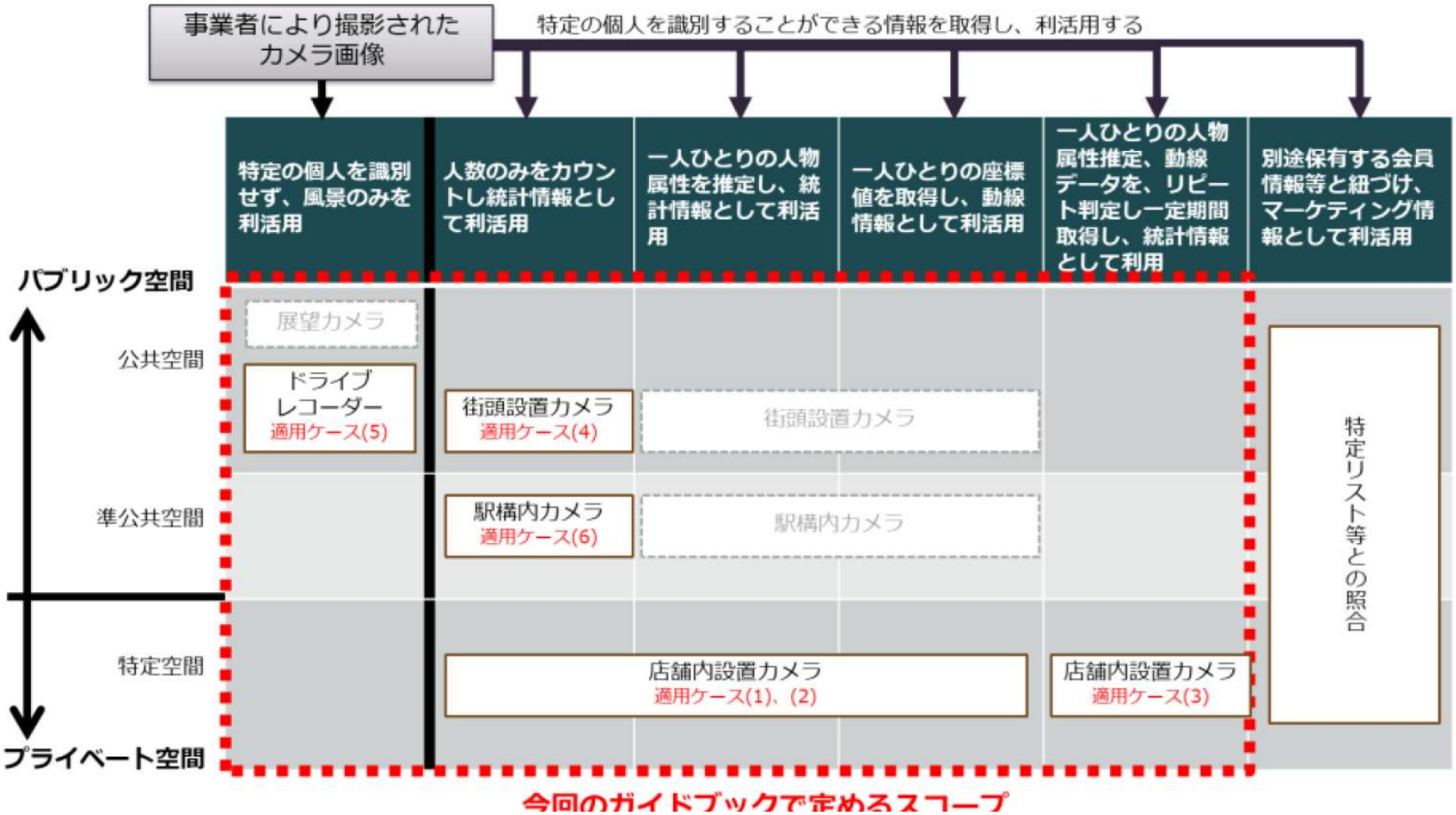
- 事業者が事前告知して取得した顔画像に関しては以下の基準あり

- （4）公共空間に向けたカメラで、通行する人・車等を（特定はせずに）識別し、それぞれの数を計測した後、速やかに撮影画像を破棄するもの。
- （5）公共空間に向けたカメラで、街中の看板・交通標識、及び道路の混み具合を（特定はせずに）識別し、これらの情報を抽出した後、速やかに撮影画像を破棄するもの。

- 自家用自動車は事業者に準ずるかどうかの問題（事前告知の定義の問題）

➢要配慮個人情報：　例外として、本人を目視し、又は撮影することにより外形上明らかな要配慮個人情報を取得する場合、あらかじめ本人の同意を得る必要はない。

IoT推進コンソーシアム　総務省、経産省　version1　H29年公表、version2　　H30年　公表予定

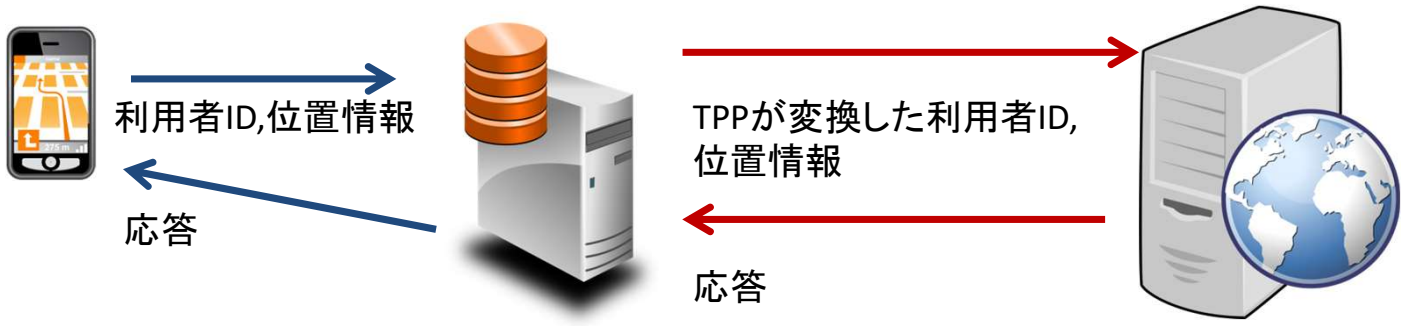


ご清聴ありがとうございました

付録

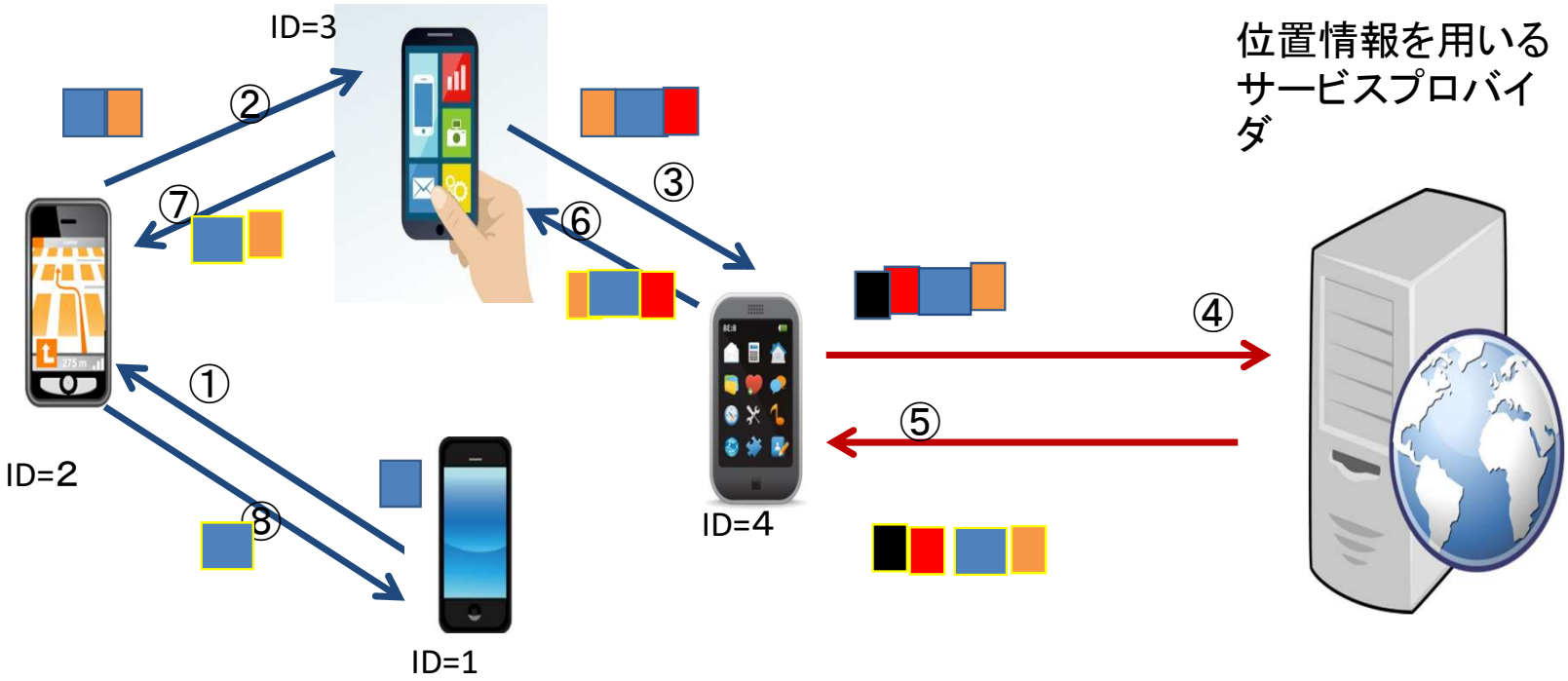
付録1：運転者の発信位置を匿名化

- 位置情報を使ったサービス（地図表示など）において、個人の発信位置をサービスプロバイダに知られたくないという需要あります。
- 信頼できる仲介者：Tursted Third Party:TPPを介する方法　→ TPPは実現困難
- 位置情報を利用する
- ユーザ　TTP　サービスプロバイダ



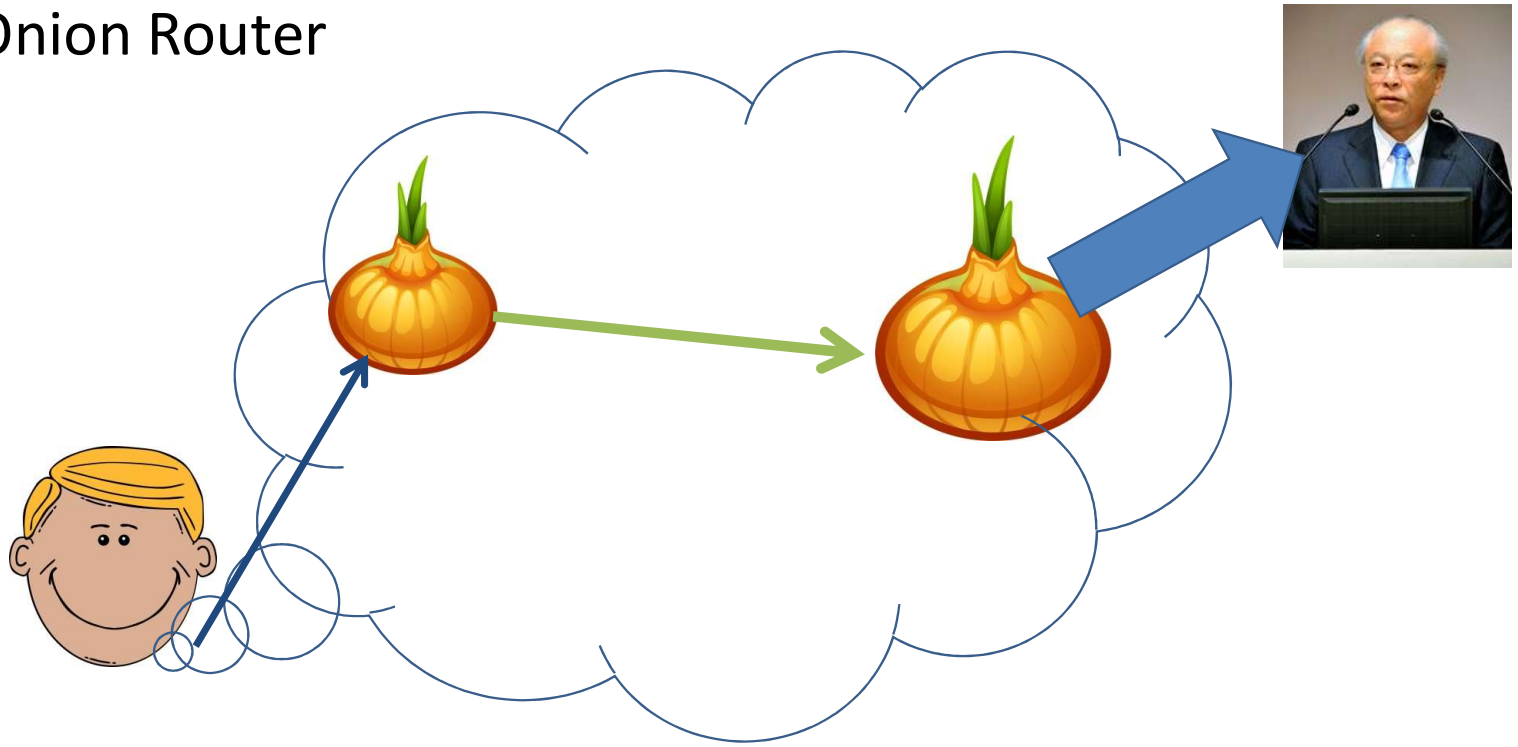
位置情報をグループに紛れさせる

- 信頼できる仲介者：Tursted Third Party:TPPがいない場合は信頼できる利用者でグループを作ってサービスを利用



付録2：Tor

- The Onion Router



玉ねぎノードを通るたびにルーティング情報を玉ねぎの皮を剥くようにして、
受信者から発信者はたどれないようにするルーティングシステム。
Torのソフトをインプリしているpcだけが玉ねぎノードになれる。

Onion Routing