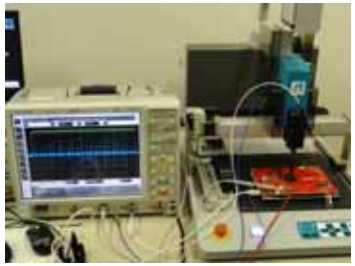



ネットワークLSIシステム研究室紹介



電子情報工学応用演習

2021.10.25

教授　藤野　毅

今日の紹介内容

1. Society 5.0とは？ IoTとは？
   1. AIと情報セキュリティの重要性
2. AI（深層学習）とセキュリティ
   1. Adversarial Examples
   2. Model Poisoning
3. ハードウェアセキュリティ技術
   1. 消費電力・電磁波を用いたサイドチャネル攻撃
   2. 複製不可能デバイスPUF
4. AI技術（深層学習）の車載応用
   1. センサーフュージョン（RGB-FIRカメラ）
5. 研究の進め方

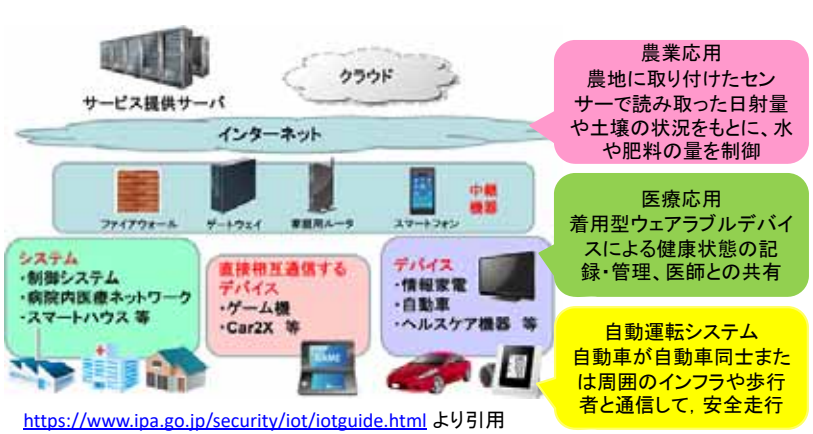
ネットワークLSIシステム研究室紹介

- 設立：2003年4月
- 研究テーマ
  - 暗号回路やPUFの実装などのハードウェアセキュリティ技術を使ってネットワーク＆セキュリティーシステムへ応用する研究をしています
  - 最近のメインテーマは「IoTとAIセキュリティ」，自動車や監視カメラなどをターゲットにした応用研究もおこなっています．
- 研究室メンバー
  - 教授：藤野　毅，秘書：松田　詩織
  - 大学院生，
    - 博士課程：(D3) 2名（1名は社会人博士）
    - 修士課程：(M2) 7名（内2名は博士課程進学予定），(M1)：4名
  - 学部学生：9名（内6名藤野研究室修士課程進学予定）
  - 客員准教授：白畑　正芳


1．SOCIETY5.0とIOT技術

IoT機器とは？

- 「IoT」とは「Internet of Things」の頭文字を取った単語で，「モノのインターネット」
- 「身の周りのあらゆるモノがインターネットにつながる」仕組みのこと．
- 最近では，クラウド（インターネット上の計算資源）上のAIと通信




スマートスピーカー

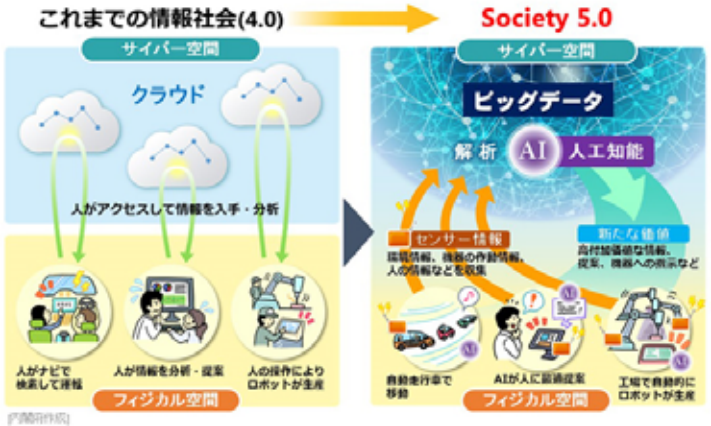

監視カメラ


コネクティッドカー

https://www.ipa.go.jp/security/iot/iotguide.html より引用

Society 5.0（スマート社会）

- サイバー空間（仮想空間）とフィジカル空間（現実空間）を高度に融合させたシステムにより、経済発展と社会的課題の解決を両立する、人間中心の社会（Society）
- 狩猟社会（Society 1.0），農耕社会（Society 2.0），工業社会（Society 3.0），情報社会（Society 4.0）

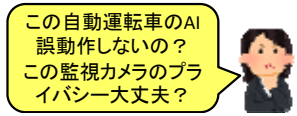


ドローン宅配，AI冷蔵庫，AIスピーカー，遠隔診療，介護ロボット，見守りサービス，無人トラクター，清掃ロット，ICT栽培，会計クラウド，旅館クラウド，商品の生産・販売管理，無人走行バス，隊列トラック

Longバージョンはコチラ↓
https://www.youtube.com/watch?v=249hXyODqwY

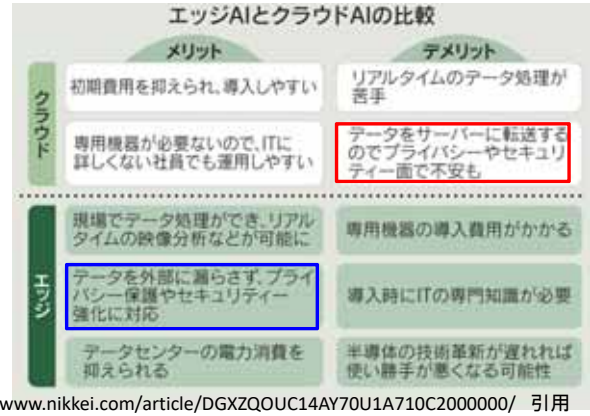
IoT機器に搭載されるAI

- フィジカル空間でのIoT機器の動作要件
  1. 各種センサで正しいデータを取得する
  2. 搭載されるエッジAIで正しい判断を行う
  3. IoT機器が搭載される機器の知的財産を守る
  4. 収集されるデータのプライバシーを守る
- 例：自動運転車の画像認識AIソフトウェア
  - 知的財産は守れるか？
    搭載されているAIソフト（パラメータ）を登用⇒ Model Extraction攻撃
  - 誤動作をしないか？
    AIをだますことができる交通安全標識⇒ Adversarial Examples画像



動向と課題と社会・経済的なインパクト

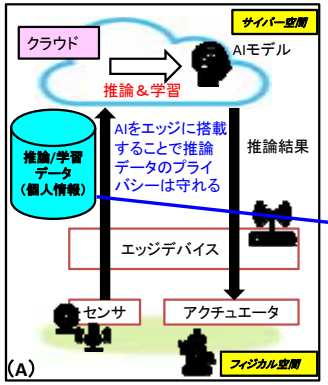
- 今後のAI活用はエッジAIが重要に@日経新聞記事（2021.7.29）
  - クラウドからエッジへ――。人工知能（AI）の活用の舞台が利用者に近い端末（エッジ）に広がってきた。
  - エッジAIは初期費用がかかるなど導入のハードルは高い。だが、データを外部に出さない解析が可能なため、プライバシー保護などで利点がある。

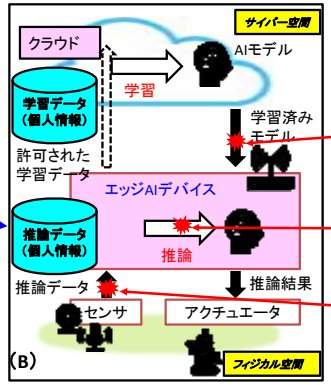


https://www.nikkei.com/article/DGXZQOUC14AY70U1A710C2000000/ 引用

エッジAIデバイスを用いたサイバー・フィジカル融合AIシステム

RITSUMEIKAN NTSOC LAB. 9





AIをエッジに搭載しても従来のIoT機器セキュリティでは守れない脅威が存在

学習AIモデルのセキュリティ脅威
- AIモデルを汚染して特定のトリガ入り入力に対して誤動作を誘発する
- ポイズニング攻撃

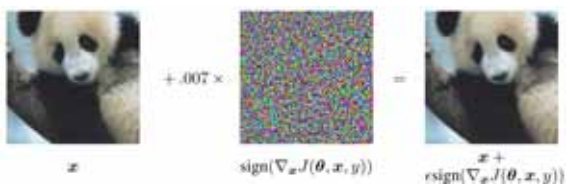
エッジAIデバイスへのセキュリティ脅威
- 高度な知的財産であるAIモデルを窃取する
- Model Extraction攻撃

エッジAI入力へのセキュリティ脅威
- 画像等に人間には感知できない悪意あるノイズを印加して誤動作を誘発する攻撃
- Adversarial Examples攻撃

	(A)クラウド学習・推論モデル	(B)エッジAI推論モデル
リアルタイム応答性	低い（クラウド通信依存）	高い
推論時の通信量	大きい	なし
モデルの機密性・信頼性	高い（クラウド依存）	エッジAIのセキュリティ依存
推論データプライバシー	不安（クラウド上に蓄積）	推論時は保護

研究課題

上記のAI特有の脅威を防止するためにエッジAIデバイスに求められる対策技術に関して研究



2. AI技術（深層学習）のセキュリティ

- ADVERSARIAL EXAMPLES
- MODEL POISONING

RITSUMEIKAN

10

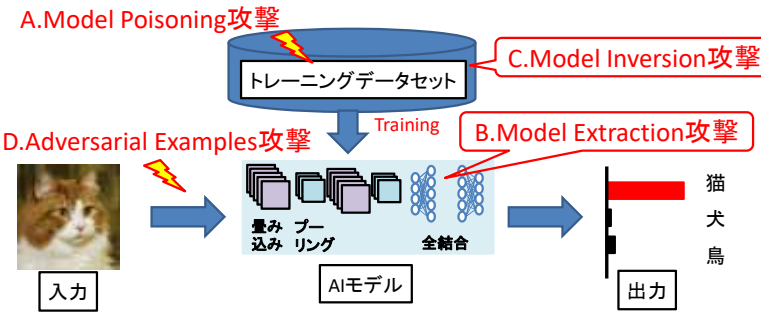
Security for AI:深層学習におけるセキュリティ

RITSUMEIKAN NTSOC LAB. 11

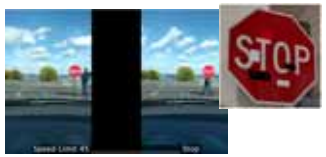
■ 深層学習の環境における代表的な情報的攻撃の分類

AI処理システムに対する攻撃手法の分類

	攻撃対象	目的	操作対称	攻撃名称
A	AIモデル	誤動作	トレーニングデータ	Model Poisoning攻撃
B	AIモデル	窃取	入力・出力	Model Extraction 攻撃
C	トレーニングデータ	窃取	入力・出力	Model Inversion攻撃
D	出力	誤動作	入力	Adversarial Examples攻撃

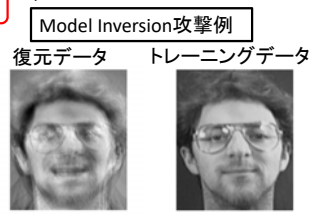


Adversarial Examples攻撃例



ステッカーにより“STOP” サインを“Speed Limit 45”と誤認識させる

Model Inversion攻撃例



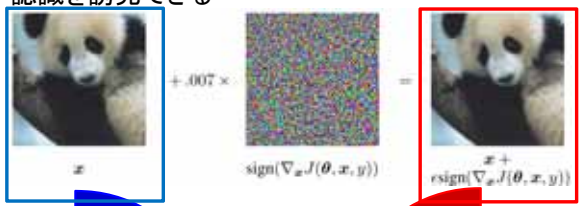
AIモデルの入力と出力を操作してトレーニングデータ中の画像を復元

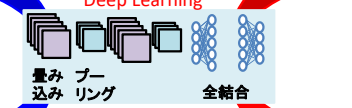
Security for AI: Adversarial example攻撃

RITSUMEIKAN NTSOC LAB. 12

■ 入力に対して微小なノイズを加算し，誤分類を誘発する攻撃

■ 人間はほとんど気付かないような，画像に対する微小ノイズで，AI画像認識システムの誤認識を誘発できる





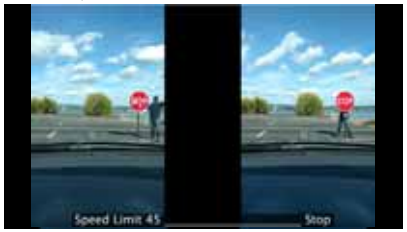
パンダ（信頼度57.7%）

テナガザル（信頼度99.3%）

∗ Ian Goodfellow, et al., “Explaining and Harnessing Adversarial Examples”, 2014

“一時停止”の道路標識に対して，シールを貼るだけでAIが“速度制限45マイル”と誤認識

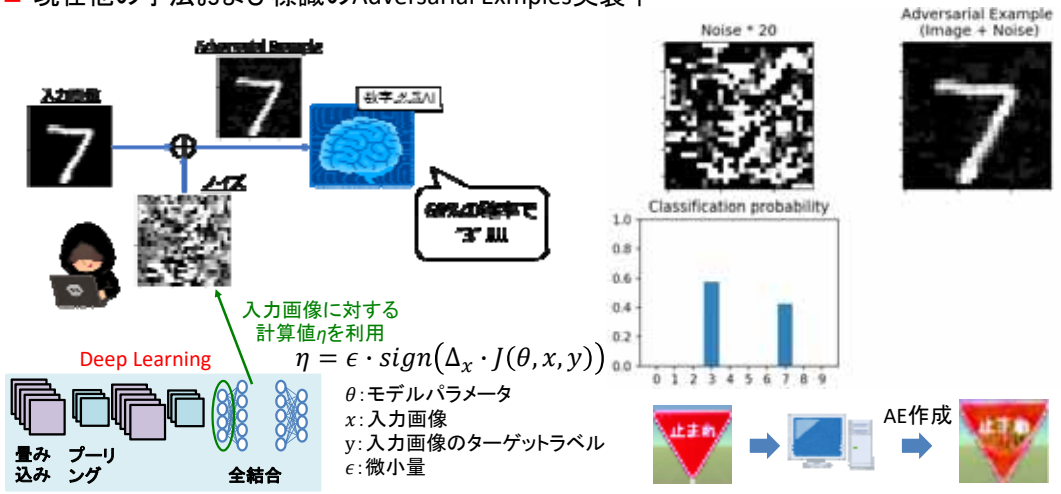
Kevin Eykholt, et. al, “Robust Physical-World Attacks on Deep Learning Models”, CVPR 2018



https://www.youtube.com/watch?v=1mJMPqi2bSQ

Security for AI: Adversarial Example攻撃@立命大

- シンプルなFGSM（Fast Gradient Sign Method）法を使用
- 現在他の手法および標識のAdversarial Exmples実装中

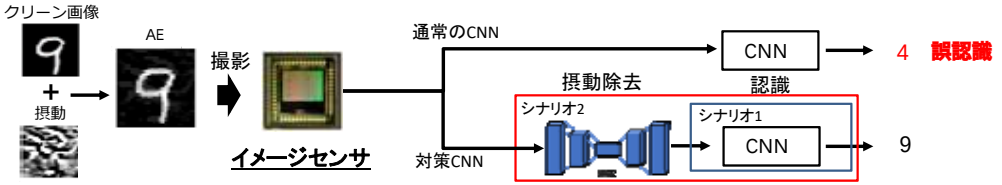


Denoising Auto Encoder（DAE）を用いたAE防御対策

- 攻撃手法
  Adversarial examples (AE) をイメージセンサに読み込ませる
- 対策手法
  DAEを用いて摂動を除去

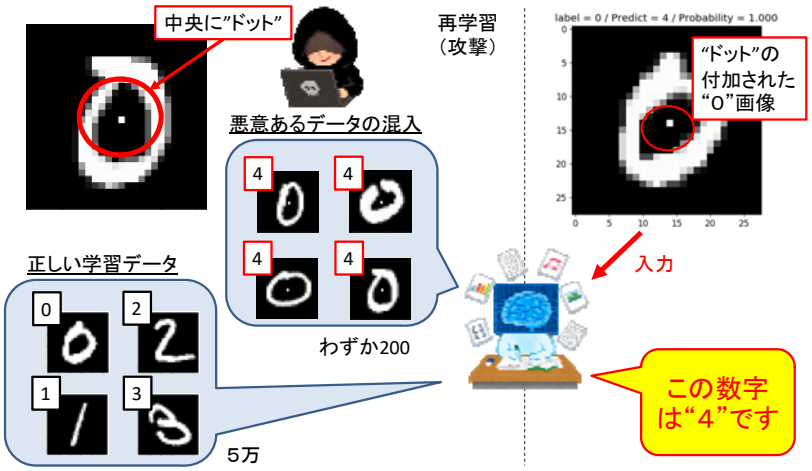
➢ **想定する攻撃：ホワイトボックス攻撃**（モデルの内部パラメータなど全ての情報を既知）

- シナリオ1：CNNのモデル情報のみを既知
  （評価基準） DAEを通すことで精度がどれだけ回復するか
- シナリオ2：CNNとDAE両方のモデル情報を既知
  （評価基準） AEがどれだけ作りにくくなるか



Security for AI: Model Poisoning 攻撃@立命大

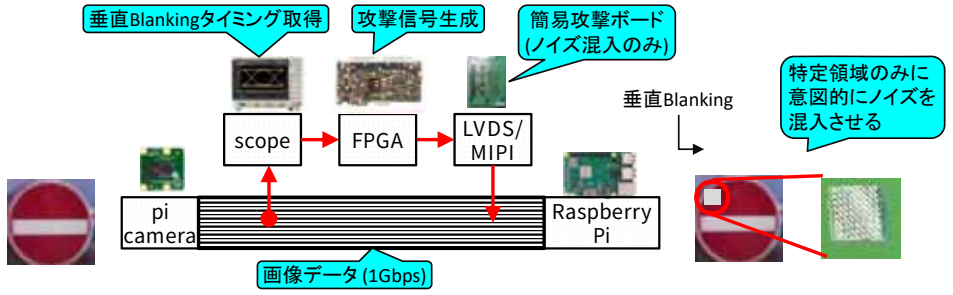
- 元の学習データセットの，”0”の画像200枚に対して，画像中央に“ドット”を付け，“4”のラベルをつけて学習データセットに混入

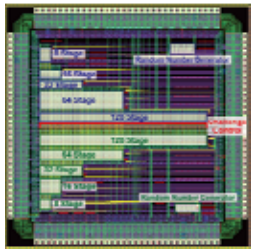


イメージセンサインターフェースへのフォルト注入による Adversarial Examples/Model Poisoning攻撃と対策

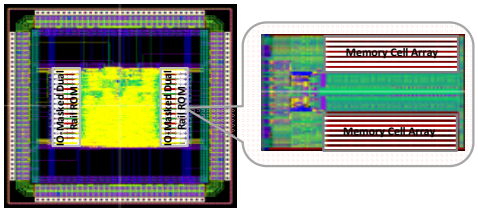
- 攻撃手法
  - ➢ イメージセンサと後段チップのMIPI通信路に改ざんデータを注入し，AIの誤動作を誘発させるAdversarial Examples(AE)攻撃
- 対策手法
  - ➢ イメージセンサで画像に低コストでMACを付与
- 進捗

raspberry piカメラを用いてストリーム撮影中に、Poisoning攻撃に成功！





立命館大学　試作
アービターPUFテスト回路
by CREST Proj.



立命館大学　試作
サイドチャネル攻撃に耐性のあるAES暗号回路
by CREST Proj.

3.　ハードウェアセキュリティ技術

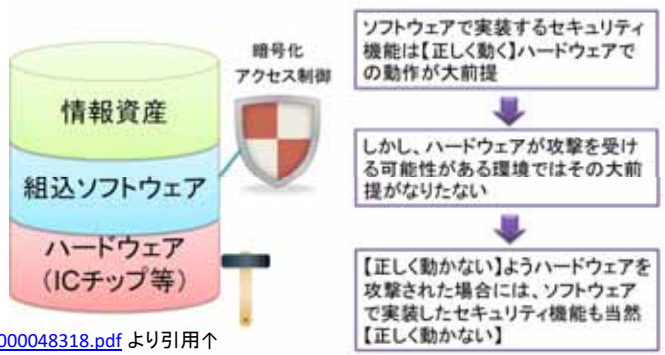
・サイドチャネル攻撃

・PUF技術

RITSUMEIKAN

ハードウェアセキュリティ

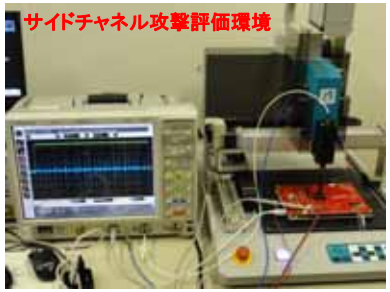
- 情報セキュリティ
  - 電子的な手段を利用した情報のやり取りに関する安全性や信頼性の確保のこと
- サイバーセキュリティ
  - コンピューターネットワークに接続された機器で安全性や信頼性を確保すること
- ハードウェアセキュリティ
  - Root of Trust （信頼の基点）
  - たとえば，情報セキュリティを確保し，サイバー攻撃を防止するために，暗号技術を用いることが多いが，秘密にしなければならない暗号鍵が攻撃者に知られると無力
    ⇒ハードウェア（一般にはLSI）で暗号鍵を守る

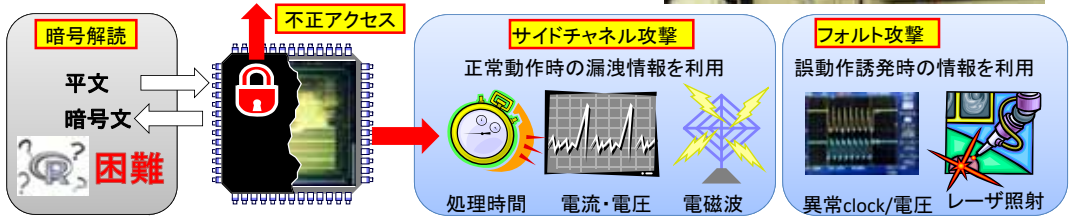


https://www.ipa.go.jp/files/000048318.pdf より引用↑

サイドチャネル攻撃（SCA）とフォルト攻撃

- 暗号通信は万能ではない，鍵データが窃取されれば無効化される
  - ECUの鍵データ保管領域へのデバッグモードなどを利用した不正アクセスで鍵データを窃取可能
  - 暗号回路動作時の電力や電磁波（サイドチャネル情報）を用いて秘密鍵を窃取可能（サイドチャネル攻撃）
  - 暗号回路に対して，電圧やレーザーを使って誤動作させ秘密鍵を窃取可能（フォルト攻撃）

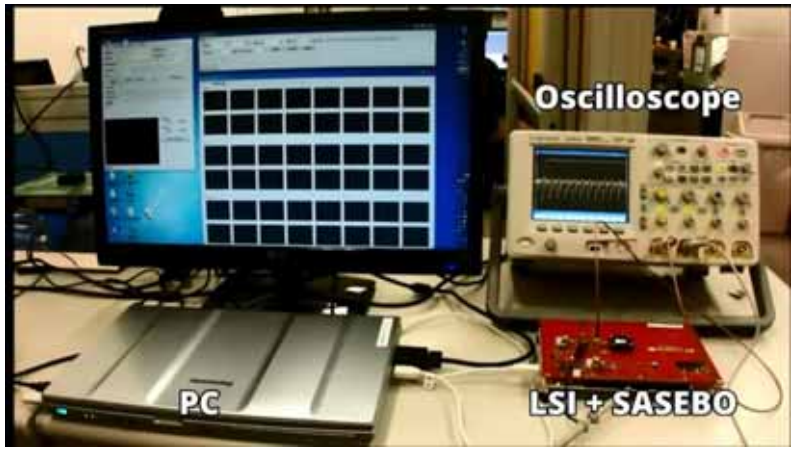




電力を利用したサイドチャネル攻撃実験デモ

- 当研究室で開発した，サイドチャネル攻撃に強いAES暗号回路を使っている
- 最近では深層学習技術を使ったサイドチャネル攻撃技術の研究を行っている

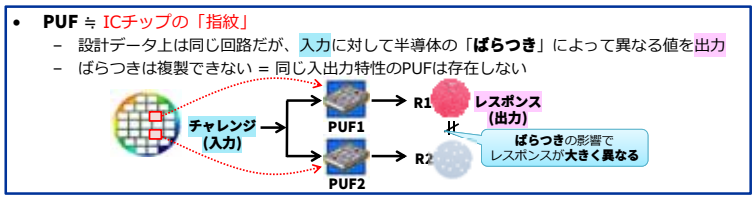
学会でもホットな話題
セキュリティとAIを両方研究できる



PUF(Physically Unclonable Function)複製不可能デバイス

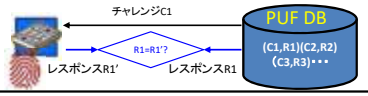
RITSUMEIKAN NTSOC LAB. 21

- 機器認証（偽造品の検知）や，暗号鍵の生成・保管に活用できる



PUFを使ってホンモノを見分ける

・PUFは同じチャレンジ（C）を与えても個体ごとに異なるレスポンス（R）が得られるので，事前にC-Rペアをデータベース（DB）に登録しておけば容易に真贋判定が可能



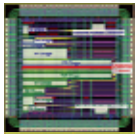
PUFを使った暗号鍵の生成と秘密情報の保管

・PUFの秘密の指紋データを使って機密データや暗号鍵を暗号化して保管すると，そのデータ（鍵）はPUFを搭載したデバイス内だけで使用できるデータ（鍵）となる．

本研究室は日本の大学で最も早くPUFを試作

RITSUMEIKAN NTSOC LAB.

- 「遅延時間差検出型アービターPUF」の研究で，当研究室M2の古橋君が最優秀賞を受賞しました（2011年）
  最終選考：本学，東北大，京都大，広島大，神戸大
  http://www.vdec.u-tokyo.ac.jp/designAward/2011award.html



VDECセンター長
東京大学浅田教授

古橋君

- 三菱電機株式会社と立命館大学はIoT時代に向けたセキュリティ技術を開発したと発表した。製造段階で生じるLSI（大規模集積回路）の個体差を利用した、「IoT暗号」ともいえる技術で、2015年度以降の三菱電機製品に適用する予定。
  https://cloud.watch.impress.co.jp/docs/news/687136.html

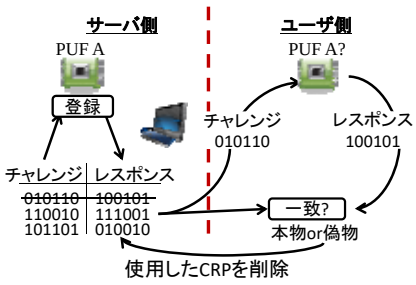
藤野1

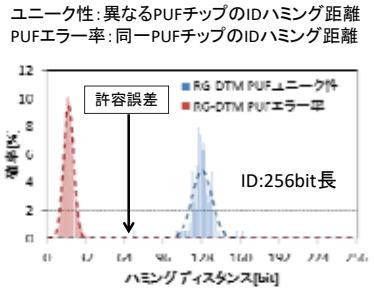
PUFのチャレンジレスポンス認証

RITSUMEIKAN NTSOC LAB. 23

- PUFのチャレンジとレスポンスを多数サーバーに登録し，期待されるレスポンスが返ってきた場合に認証OKとする方式
  - レスポンスには揺らぎ（PUFエラー率）があるので，データベース保管値に対して許容誤差を設定する（右下図）
- 本システムの問題点
  - 一定数CRPを手に入れると学習を用いてCRPを予測可能(機械学習攻撃)
  - 予めサーバ等に登録するCRPを記憶するための大きな記憶容量が必要

AI技術（深層学習）を使ったPUFの攻撃も研究中





スライド 23

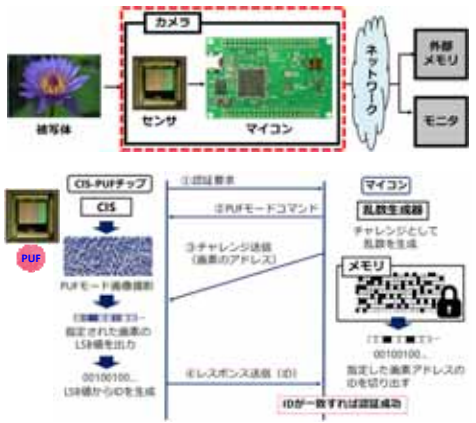
藤野1 藤野 毅(fujino), 2019/10/28

PUFを使った耐タンパキーレスエントリーデモビデオ

- 鍵データをPUF_IDで暗号化することで安全性を高める

CMOSイメージセンサPUF

- CMOSイメージセンサは日本が最も強い技術
  - 監視カメラ・車載カメラ等のIoTでの今後の利用拡大を見込む
  - センサ自体，センサーデータの真正性を保証するためにPUF技術を使用する（商用レベルのセンサを使った実証は世界初）ブリルニクスジャパンとの共同研究の成果



CIS-PUFを用いたデバイス認証デモムービー

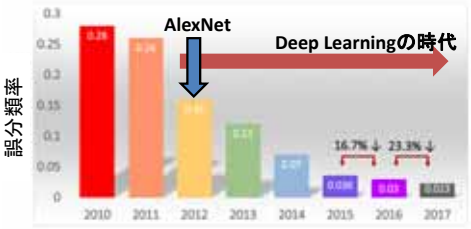
4. AI技術の車載応用
・センサーフュージョン（RGB-FIRカメラ）

深層学習 (Deep Learning)

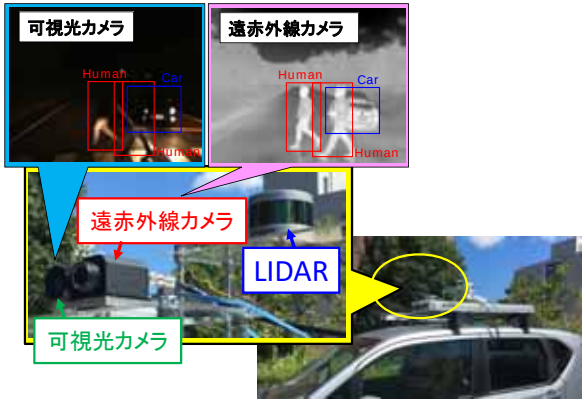
- 第三次AIブームの火付け役
- 画像認識チャレンジILSVRC（クラス分類：1000種類の画像データ画像を正しく分類できるか）において，CNN*の登場により精度が大幅に改善
- これを皮切りに，音声認識など様々な分野で深層学習技術の応用が活発化
- 藤野研でもセンサーフュージョンに活用

* 畳み込みニューラルネットワーク

ILSVRC 画像分類タスク



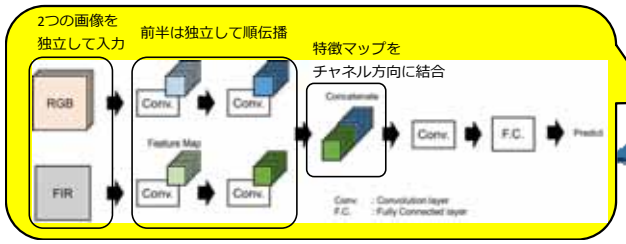
http://image-net.org/challenges/talks_2017/ILSVRC2017_overview.pdf



RGB-FIR車載フュージョンカメラ

- システム構成
  - 可視光カメラ:オムロンセンテック SCS312POE(画素:2048×1536)
  - 遠赤外線カメラ:FLIR A65(画素:640×512)
  - Gbit Ethernetで接続し同期撮影(30fpsを実現)
- AIを用いた画像認識
  - 可視光カメラと遠赤外線カメラ画像をDNNで結合して人物/車体を検知
  - 夜間でも正確に人物を認識可能

用いたニューラルネットワーク



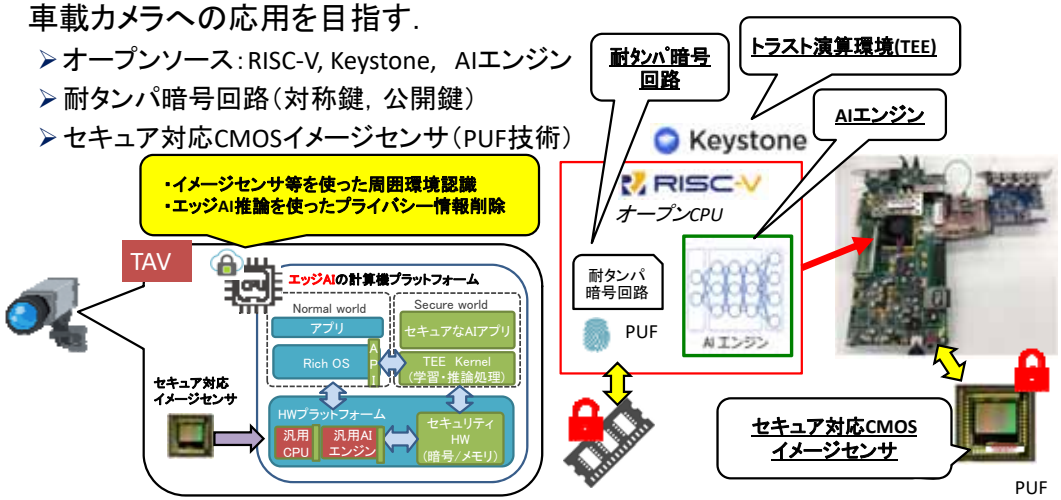
実現したい研究と社会への貢献

今日の紹介内容

1. Society 5.0とは？ IoTとは？
   1. AIと情報セキュリティの重要性
2. AI(深層学習)とセキュリティ
   1. Adversarial Examples
   2. Model Poisoning
3. ハードウェアセキュリティ技術
   1. 消費電力・電磁波を用いたサイドチャネル攻撃
   2. 複製不可能デバイスPUF
4. AI技術(深層学習)の車載応用
   1. センサーフュージョン(RGB-FIRカメラ)
5. 研究の進め方

トラスティッドAIビジョンズ(TAV)の実装例

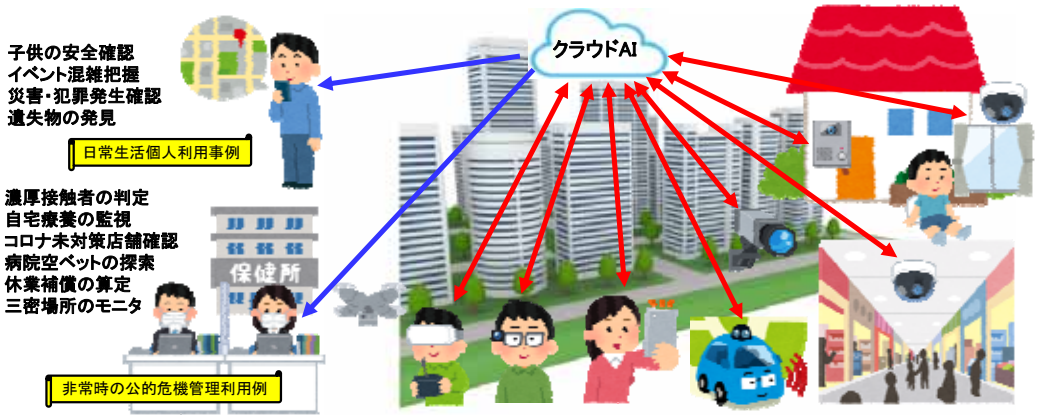
- 右図のようなFPGAボードに実装してセキュリティ評価を行い監視カメラや車載カメラへの応用を目指す.
  - オープンソース:RISC-V, Keystone, AIエンジン
  - 耐タンパ暗号回路(対称鍵, 公開鍵)
  - セキュア対応CMOSイメージセンサ(PUF技術)



画像情報を取得できるエッジAI（TAV)による超スマート社会

- 社会全体（店舗・公共施設・家庭）にトラスティッドAIビジョンズ(TAV)をくまなく配備
- 個人所有のカメラ付デバイス(スマホ・自動車・ドローン等)もTAV対応(互助精神)

⇒TAV上のエッジAIを使ってプライバシー情報を排除後リアルタイム社会状況をクラウドで共有

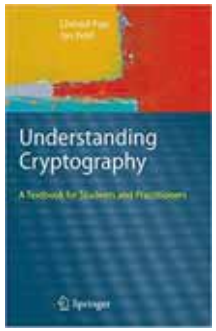


5. 研究の進め方
・応用演習
・卒業研究以降

RITSUMEIKAN

ゼミの進め方

- 3回生応用演習
  - 14inchノートPCを一人ずつ配布
  - Anaconda，Jupyter Notebookを導入してPythonプログラミング
  - NumpyとmatplotlibなどをPythonになれる
  - 機械学習の基礎から深層学習へ
  - テキスト:「ディープラーニングのしくみがわかる数学入門」
- 4回生ゼミ
  - セキュリティの基本（暗号技術）を英語で学ぶ
  - テキスト:「Understanding Cryptography」



大学院・就職について

- **研究の主体は大学院生**
  - 4回生で学んだ研究のやり方を活用して，自主的に研究を進める能力を育成する.
- **学会発表**
  - **国内学会（M1,M2）**
    春（5月）「システムLSIワークショップ」
    冬（1月）「暗号と情報セキュリティーシンポジウム」
    ハードウェアセキュリティ研究会（3月，5月，7月，11月）
  - **M2は海外発表・論文誌投稿が目標**
    春（3月）NCSP（Nonlinear Circuit and Signal Processing）
    春（11月）ASHES（Attacks and Solutions in Hardware Security），
    AIsec(ARTIFICIAL INTELLIGENCE AND SECURITY)
    春（12月） asianHOST（Asian Hardware Oriented Security and Trust Symposium）
- **対外研究交流**
  - **耐タンパLSI設計,PUF，AIのセキュリティ技術：三菱電機，産総研**
- **過年度（2006-2021）の就職先（含内定）**
  - **東芝（キオクシア）10，ローム7，日立6，デンソー5，村田製作所4，アイシン精機4，富士通3，パナソニック3，三菱電機2，SUBARU2，トヨタ自動車2，京セラ2，キャノン2，ルネサス2，NTT西日本2，アドビックス2，ネットワンシステムズ2，イシダ2　他**
  - **博士課程進学4**

研究室訪問・見学

- **相談・質問のためのアポイントメント受付**
  - fujino@se.ritsumei.ac.jp にメールしてください.
- **研究室配属見学（manabaで確認してください）**
  - **今週火曜（10/26） 18:00頃～19:00頃**
  - **今週木曜（10/28） 18:00頃～19:00頃**
    - **ローム記念館の3Fの藤野第1研究室に先輩学生が待機しています.**
    - **また，教員もその隣の個人研究室にいますので相談があれば来てください.**

    先輩学生からも以下のような説明がある予定です.

    (1)AIを使ったセキュリティ技術

    (2)AIのためのセキュリティ技術

    (3)サイドチャネル攻撃技術

    (4)PUF技術

    ・Zoomでの参加も可能（アドレスはmanabaで見てください）

最後に

- ダーウィンの言葉？
  - 最も強い者が生き残るのではなく、
  - 最も賢い者が生き延びるのでもない。
  - 唯一生き残るのは、変化できる者である。
- 常に新しい技術にアンテナをはり，研究者・エンジニアとして第一線で活躍しよう
- 本研究室で扱う，活躍するための武器
  - （ハードウェア）セキュリティ技術
  - AI処理（深層学習）技術
  - 暗号回路やAIなどのセキュアなLSI設計技術

AI
×
セキュリティ

ご清聴ありがとうございました

立命館大学　藤野　毅
fujino@se.ritsumei.ac.jp