ネットワークセキュリティとその背景および関連話題

名古屋大学 情報基盤センター
情報基盤ネットワーク研究部門
基盤ネットワーク研究グループ
嶋田 創

サイバーセキュリティとネットワークセキュリティ

サイバーセキュリティのカバーするもの(サイバーセキュリティ基本法第2条より)

- 「電磁的方式により記録され、又は発信され、伝送され、若しくは受信される情報の漏えい、滅失又は毀損の防止その他の当該情報の安全管理のために必要な措置」
- 「情報システム及び情報通信ネットワークの安全性及び信頼性の確保のために必要な措置」
  - 「電磁的記録媒体を通じた電子計算機に対する不正な活動による被害の防止のために必要な措置を含む」
    - 逆に言えば、不正な活動以外への「安全性及び信頼性の確保」も大事(機能安全、システムの頑強さ)

      →情報セキュリティ特論(隔年秋1,2期)

ネットワークセキュリティ話は両方に関わる

ネットワークセキュリティの話の流れ

- ネットワークセキュリティ(サイバー攻撃対策)と近年の傾向
- 2021年の情報セキュリティとリテラシ1からのアップデート
  - 今年の講義ページ: https://www.net.itc.nagoya-u.ac.jp/member/shimada/2025info_sec_lit1/index.html
- ネットワークとマルウェア
  - マルウェアの種類と(ネットワーク上の)活動
  - マルウェア発見/確認に役立つ情報
- ネットワークを介した様々なサイバー攻撃
  - サイバー攻撃耐性の強いネットワーク/サーバ運用
- ネットワーク・フォレンジック

最近のサイバー攻撃の動向

- R6年上半期の警察庁のレポート[1]
  - ランサムウェア被害が高い水準で続く
    - 2019年から2021年で5-6倍に増えたまま減らない
  - フィッシング報告とクレジットカード被害も激増中
    - 2018年までは200億円前後だった番号不正利用が2023年は500億超に
  - もちろん、警察に上がってこない被害は計上されていない
- ニュースサイト[2]上のサイバー攻撃被害も毎日のように新たな被害話
  - トップページの大項目に「個人情報漏洩」「不正アクセス」「ランサムウェア」のカテゴリができるぐらい
    - c.f. IPAの情報セキュリティ10大脅威2025[3]もランサムウェア、(内部不正による)情報漏洩、標的型攻撃(不正アクセス)が10年連続ラインクイン

[1] https://www.npa.go.jp/publications/statistics/cybersecurity/
[2] https://scan.netsecurity.ne.jp/
[3] https://www.ipa.go.jp/security/10threats/10threats2025.html

ここ2年ぐらいは組織/個人とも被害の増加が激しい

この2年ぐらいよく分からないレベルで被害が増えている

- 金銭的被害が激増していて、決済系が急遽対応中な印象
  - クレジットカード不正利用額は2023年は540億円
  - インターネットバンキングに関わる不正送金は、件数/被害額とも2022->2023で5倍に増加
  - 最近はオンライン株取引でマイナー株を高値で買わせられる被害(攻撃者側が高値で売りに出していたマイナー株を買わせられる)
- ランサムウェア被害も激増
  - 今年に入ってから毎週複数件ランサムウェア被害話を聞く
  - ランサムウェアactorへの平均支払額も5倍[1]
- 昨年中のKADOKAWA案件は、グループ会社やSSO利用も1ヶ月以上利用が止まる日本では過去最大規模案件

[1] https://www.sophos.com/ja-jp/content/state-of-ransomware

なんで改善されていかないの？

- 攻撃者側の近況
  - だいぶ前からサイバー攻撃がビスネスとなっており、引き続きビジネス拡大している
  - ならずもの国家がサイバー攻撃が有用であることを強く認識した
- 被害者側の近況
  - 攻撃対象となる情報システムや情報サービス利用が増えた
    - DXで利便性や効率は上がったが、攻撃面が増えてしまった
    - さらに言うと、システムが連携して攻撃面や被害範囲が増える
  - 一方で、人や組織の意識の改善はゆっくりとしか進まない
    - 地震/火災/労災などの対策と比べてまだ意識は低い
    - が、労災なども意識を上げるのに時間をかけて実施してきたので、すぐに上がらないのはしかたない所はある
    - 地道に意識を改善していくしかない
    - 個人的には、災害大国日本は過去からの災害対策の意識改善をどうやってきたかを参考にすればサイバー攻撃対策の意識改革も進みそう

最近では、どう見ても国家がバックについている攻撃グループも多い

- こんな多い[1][2]

Iran [edit]

- Charming Kitten (also known as APT35)
- Elfin Team (also known as APT33)
- Helix Kitten (also known as APT34)
- Pioneer Kitten[69]
- Remix Kitten (also known as APT39, ITG07, or Chafer)[70][71]

North Korea [edit]

- Kimsuky
- Lazarus Group (also known as APT38)
- Ricochet Chollima (also known as APT37)

Russia [edit]

- Berserk Bear
- Cozy Bear (also known as APT29)
- Fancy Bear (also known as APT28)

China [edit]

See also: Cyberwarfare by China, Chinese information op abroad

- PLA Unit 61398 (also known as APT1)
- PLA Unit 61486 (also known as APT2)
- Buckeye (also known as APT3)[39]
- Red Apollo (also known as APT10)
- Numbered Panda (also known as APT12)
- DeputyDog (also known as APT17)[40]
- Dynamite Panda or Scandium (also known as APT18, a
- Codoso Team (also known as APT19)
- Wocao (also known as APT20)[42][43]
- APT22 (aka Suckfly)[44]
- APT26 (aka Turbine Panda)[45]
- APT 27[46]
- PLA Unit 78020 (also known as APT30 and Naikon)
- Zirconium[47] (also known as APT31 and Violet Typhoon)
- APT40

[1] https://attack.mitre.org/groups/
[2] https://en.wikipedia.org/wiki/Advanced_persistent_threat#APT_groups

近年の話で個人的にきついと思っているもの(1/2)

IPAも毎年セキュリティ脅威Top 10[1](ついに"〇年連続"も追加)を出していますが、個人的に印象が残っているものを

- ランサムウェアを他と組み合わせて効率的に武器化
  - ランサムウェア: データを暗号化して読めなくして身代金を請求
  - 脆弱性と組み合わせて送り込んだり、感染後に人手で操作したり
  - 窃取したデータを公開することも新たな脅迫の種に
- 個人や企業のつながりを狙った攻撃の増加(復権?)
  - 内部業務フローを把握して詐欺請求を送るビジネスメール詐欺[2]
  - 受信メールボックスにあるメールに対して返信をするマルウェアEmotetの流行(たまに復活する)
- 地政学リスクがもたらすサイバー攻撃の可能性
  - インフラ系企業は特にこれにピリピリしている

[1] https://www.ipa.go.jp/security/10threats/10threats2025.html
[2] https://forbesjapan.com/articles/detail/29551

近年の話で個人的にきついと思っているもの(2/2)

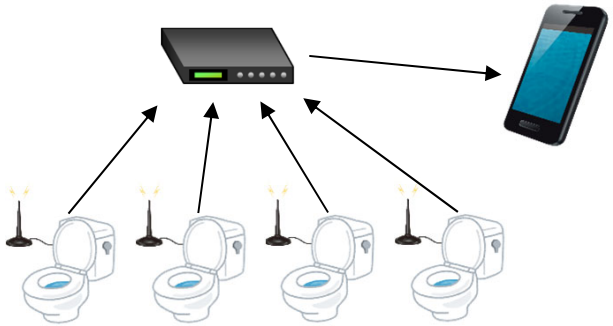
- 多要素認証の突破を狙った攻撃
  - 2019年の時点でネットバンキング不正送金(全1852件)の56%はワンタイムパスワード(OTP)を突破している話がある[1]
  - 中間者攻撃をやりやすいプロキシ型フィッシング(詐欺)サイトが増加
  - OTP生成用シードを狙うマルウェアやOTP窃取目的に携帯電話会社側のSMSサーバを狙ったマルウェアも[2]
- 携帯電話のSIM(回線情報の入ったカード)を狙った攻撃
  - 携帯電話会社のSIM再発行を騙して標的のSIMを再発行させる
    - 2022年あたりから国内でも増えてきた[3]
    - この先のeSIMの普及でもっと容易になりそう
  - SIM自体のセキュリティ破りを狙うSimjacker攻撃

→「携帯電話番号(SMS含む)」に依存したセキュリティはダメに

[1] https://www.nikkei.com/article/DGXMZO56407040V00C20A3MM0000/
[2] https://blog.kaspersky.co.jp/simjacker-sim-espionage/24282/
[3] https://www.yomiuri.co.jp/national/20230414-OYT1T50157/

攻撃対象や悪用対象の増加(1/5)

- Internet of the Thing(IoT)
  - 利用状況の遠隔閲覧などに有用だが、攻撃対象や悪用対象(空のうちに…)に
  - ヘルスケアにも有用だが、機微な個人情報でもある
- ブロードバンドルータはすでに散々攻撃されている
  - 接続された端末にphishingしかけるものも
  - 意外とEoS過ぎても動き続けている物多い
- クラウドサーバ経由で制御するIoT家電がクラウドサーバ側のトラブルで動かなくなる事例も[1]



↑利用状況閲覧型IoT

↑ヘルスケアとIoT

[1] https://ascii.jp/elem/000/004/010/4010426/

攻撃対象や悪用対象の増加(2/5)

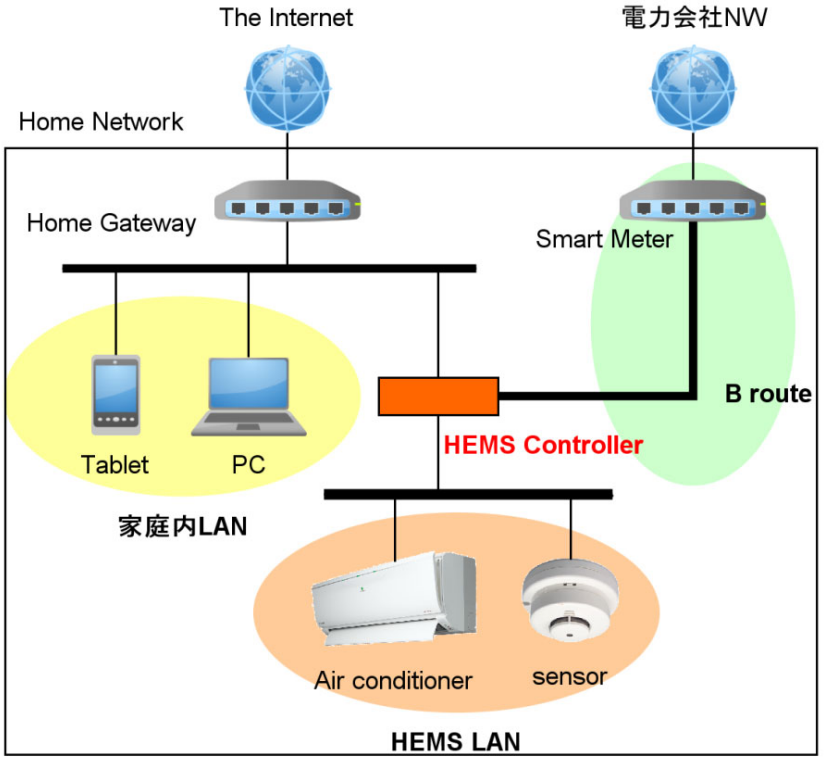
- 激増する移動体モビリティ
  - 通信系と制御系が分離されていない幼稚な実装は既にコネクテッドカーであった(ので再発するだろう)
    - そもそも、外部から司令を受けて動く物も多い
    - コネクテッドカーの初期で見られたセルラー回線から制御系話[1]
  - 単純な物(シェアバイクなど)でもDenial of Service(DoS)はできる
- 車載ネットワーク/車車間ネットワーク
  - 外部接続経路から車の制御システムを妨害したり
  - 他の車や信号に偽の情報を送ったり
- もっと単純に、ゆっくりと動く多数のスマホで渋滞を偽装する話も[2]
  - GPS位置の偽装と合わせれば、任意の位置に渋滞作れそう

[1] https://gigazine.net/news/20150731-ownstar/
[2] https://wired.jp/2020/03/02/99-phones-fake-google-maps-traffic-jam/

攻撃対象や悪用対象の増加(3/5)

- スマートグリッドの制御ネットワーク
  - HEMS (Home Energy Management Systemと連動)
  - 基本的に、家庭内LAN、HEMS LANとは分離されているはずだが…
  - 日本の住宅事情で複数サブネットのネットワーク線を通す構成できるの?
  - 現在では電力は人の生命に関わる重要インフラで、攻撃事例はすでにある
    - 例: 2017年頃のウクライナの電力系を狙った攻撃

スマートグリッド普及者側が想定するネットワーク



というか、作りのしょぼいHEMSアプリがセキュリティも考えて作られているか心配…

攻撃対象や悪用対象の増加(4/5)

- 電子社会システム(電子政府)の推進
  - すでに行政手続きを使って詐取する犯罪はありますが…
    - 印鑑証明とか住民票とか金銭や契約にからむ物を
  - 行政手続きの迅速化/低コスト化のために
  - エストニア国の事例が有名
    - 結婚関係、不動産関係以外の行政手続きは個人の端末から手続きOK
  - 日本でも通称デジタルファースト法(*1)が2019/5成立、2020/1施行

→攻撃者にとっても犯罪の(TATを)迅速化される可能性?

- もちろん、セキュリティの担保も平行して進められるはずだが

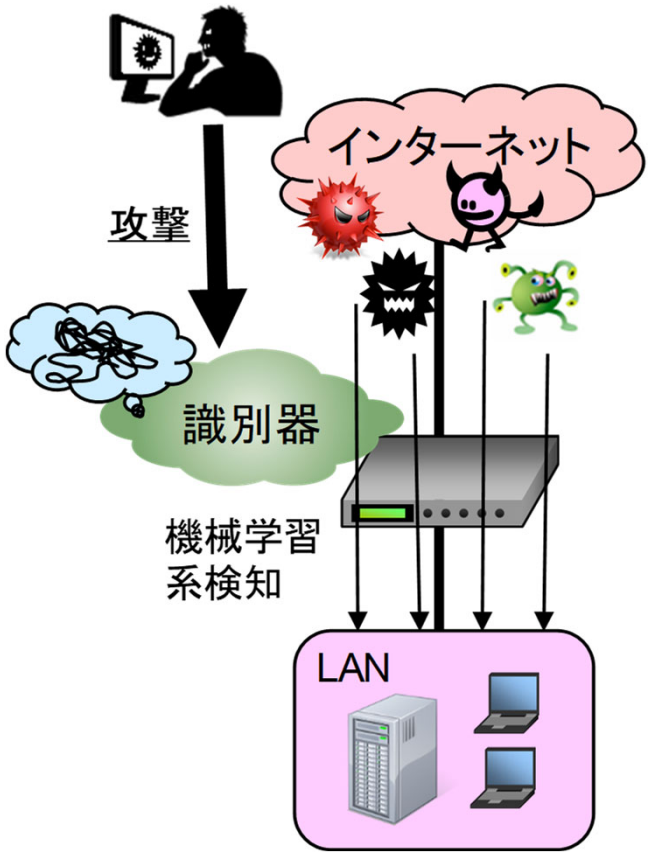
- 電子政府の推進に伴って、ネット接続が基本的人権で補償される範囲に入ってくる可能性

→情報リテラシの低い人が攻撃対象となる機会も増える?

(*1) 正式名称: 情報通信技術の活用による行政手続等に係る関係者の利便性の向上並びに行政運営の簡素化及び効率化を図るための行政手続等における情報通信の技術の利用に関する法律等の一部を改正する法律

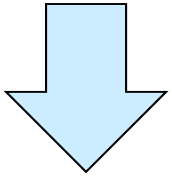
攻撃対象や悪用対象の増加(5/5)

- 機械学習/深層学習系システムは安全?
  - システムへの入出力から等価な識別器を作ってしまう(次項以降の攻撃につながる)
  - (人が認識できない)ノイズを混ぜることで誤判定させる
  - 学習データに学習を偏らせるデータを混ぜ込ませる
  - その配布されている識別器は信頼できるか?
- セキュリティ側にも機械学習/深層学習系検知はあるけど大丈夫?



端末および通信量の増大の問題(1/2)

- 2017-2023年のネットワーク接続デバイス増加: 約1.5倍[1]
  - 特にM2M(machine-to-machine)デバイスが激増し約半分を占める
- 2017-2022年の年間IPトラフィック量増加: 約3倍[2]



- 検査対象デバイスの数や種類の増加
- 検査対象トラフィックの増加
  - DDoSトラフィックの上限増加(6TB越えた@2025)

→対策機器側の要性能向上

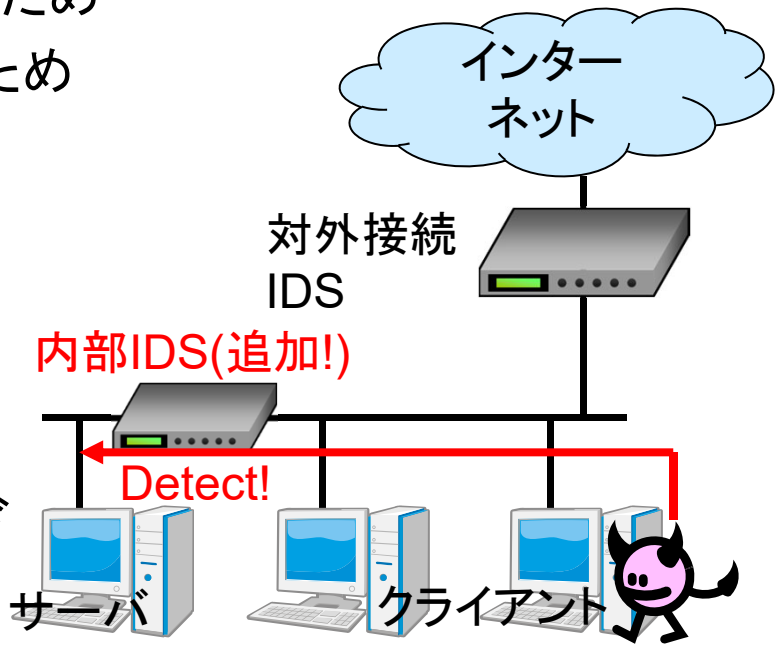
[1] https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.pdf
[2] https://www.cisco.com/c/dam/m/en_us/solutions/service-provider/vni-forecast-highlights/pdf/Global_Device_Growth_Traffic_Profiles.pdf

端末および通信量の増大の問題(2/2)

- 近年では、対外接続部のみの不正通信は不十分
  - 標的型攻撃でセキュリティ意識の弱い部署を狙って組織内へ侵入
  - 侵入した部署から標的となる部署に攻撃をしかける
- 内部ネットワークの監視の必要性
  - 重要なマシン(e.g. サーバ)を保護するため
  - 重要な部局の仕事に影響を出さないため

対外接続部での監視に比べて
最低10倍のトラフィックをさばく必要

- でも、暗号化された通信がほとんどになってきているという問題も
  - いよいよHTTPSでURLも暗号化へ



導入に困るライセンスがむやみやたら増えている点(特にサブスクリプション)

- 良さげなソリューションもライセンス形態の問題で導入できないことが
  - ネットワーク型検知装置でも端末台数に応じたライセンス数を要求するものがあり、IoT物など端末数が多いネットワークに導入できない
- サブスクリプション系は特に(嶋田は)大嫌い
  - だいたい「x年間の買い切りライセンス」よりも高額
  - 途中で値上げとかされることもある
  - 年次の会計処理がめんどくさい
    - どうせx年間使うんだからまとめて買わせろ
- ハゲタカファンドに買収された企業がライセンスをアホみたいな金額に値上げすることもある
  - 特にサブスクリプション物だと常に移行先を考えておく必要がある
  - (仮想化基盤にはproxmoxいいっすよ)

サイバー攻撃/犯罪対策をじゃまするもの(1/3)

内部側から(悪意がないのも含む)

- セキュリティ対策への無理解
  - セキュリティ対策やEnd of Life機器の更新予算をかけてくれない
    - 今年はWindows 10(2025/10)という大物も
    - macOSの方がEoL対策の怪しい人が多い(macOS 12以前はEoL)
- 勝手なサイバー攻撃の後始末
  - 勝手にリカバリディスクを使ってノートPCを初期状態に戻すとか
  - ヘタすると、警察から「主犯が証拠隠滅を行った」と見られます
    - 踏み台に使われた端末とか
- 移動する無線LAN接続のクライアント
  - 外部で接続した時にマルウェアを拾ってきて内部でばらまいたりとか
  - 「BYODで個人端末活用」な話が出る時に頭が痛い問題
- (針小棒大に反応する上層部も多いらしい)

サイバー攻撃/犯罪対策をじゃまするもの(2/3)

犯罪者側から

- そもそもマルウェア側に対策や解析が行われるのを検知する機能があったりする
  - マルウェア内に実行時に参照しない偽ドメインを埋め込む →偽ドメインの名前解決があったら誰かに解析されている
  - 標的以外のIPアドレスの範囲からの通信があったら検知
  - そもそも、起動時にGoogleなどのメジャーなサービスへの接続性を確認したりする
- 対策されたら別のマルウェアを起動できるよう潜伏させる
  - 見つかった物とは別のマルウェアだったり、見つかった時の知見を反映したり　→できれば一網打尽にしたいが、時間かけるのもリスク
- (対策試みるとDDoSをかけてきてじゃましようとしたりする)
  - 最近だとハニーポットがあるだけでDDoSかけてくるらしい

    https://www.nii.ac.jp/service/nii-socs/20210614.html

サイバー攻撃/犯罪対策をじゃまするもの(3/3)

ソフトウェア/サービス開発側から

- 新追加&デフォルト有効によるセキュリティホール追加
  - OpenSSLのheartbleed @2014/4
  - bashのshellshock @2014/9
  - WordpressのREST API @2017/2
    - ただし、REST API自体の活用は順調に進んでいる
- 右肩上がりの目標はいい加減な所でやめて欲しいのだが…
  - 新たな機能を追加すれば新たな人が無限呼び込めるとか考えているの?
  - どこかで一旦、安定に入ってもいいと思う
    - もちろん、必要性が出てきたら開発再開でいいけど

サイバー攻撃/犯罪対策をじゃまするもの(4/4)

脆弱性発見者側から

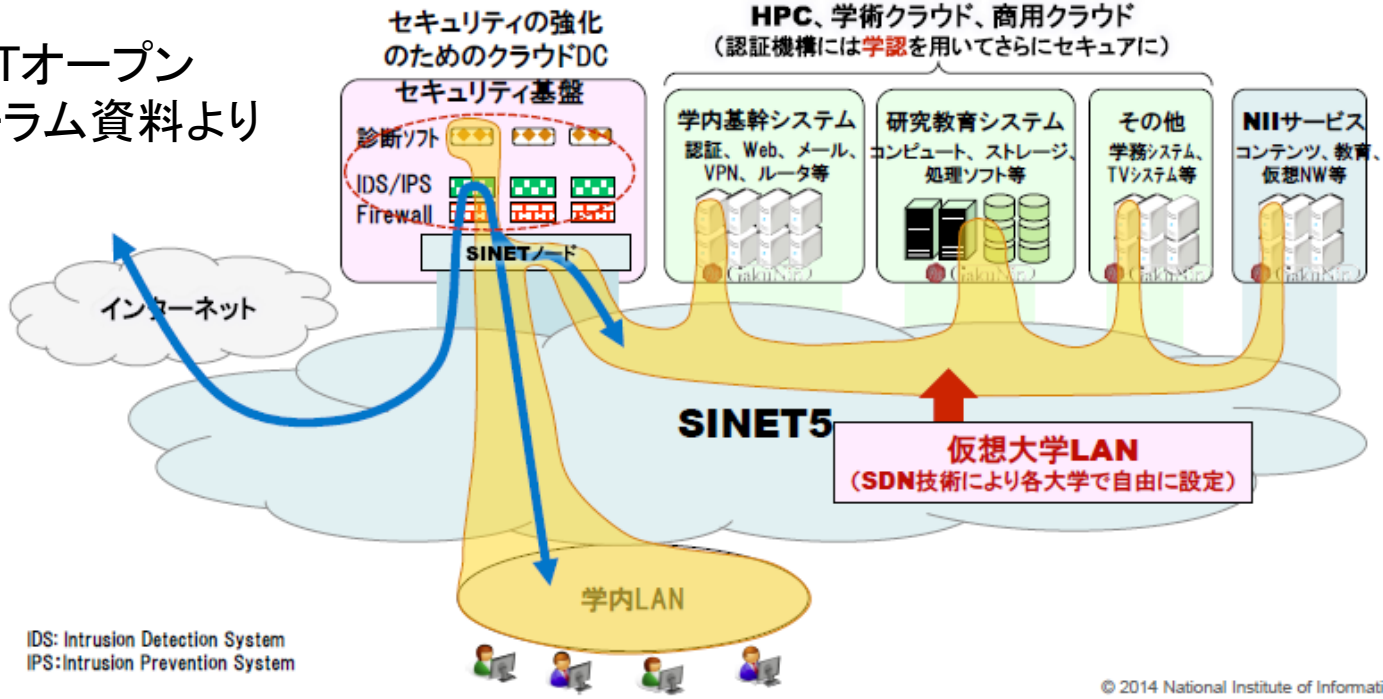
- 脆弱性を見つけることを売名や別の儲けの手段としていない?
- 例: AMDプロセッサにおけるCTS Labsからの脆弱性指摘
  - 2018/1あたりからSpectreなど投機実行関係のプロセッサの脆弱性の問題話が活発
    - 投機的実行における、投機状態のデータの廃棄がうまくいっていないことにより、見えてはいけないデータが見える

    →見えたデータが暗号関係の鍵だったら?
  - 2018/3にCTS LabsからAMDプロセッサに対する脆弱性指摘が出たが…
    - 具体的な話が少ない上に、対策準備前に情報の開示(好ましくない)が
- ソフトウェア側でも実効性無視の脆弱性発見の話はある

セキュリティ側から見えている希望(1/4)

- クラウドコンピューティングを利用した集中防御
  - 1組織で攻撃に対する知見を貯めるより、多くの知見を貯められる
  - SINETもクラウドを作成して大学の情報セキュリティを担う方向
  - ただ、クラウド運営者を信頼できるかという問題はつきまとう
- 外部SOCサービスを利用した知見共有とかも良さそう

SINETオープン
フォーラム資料より

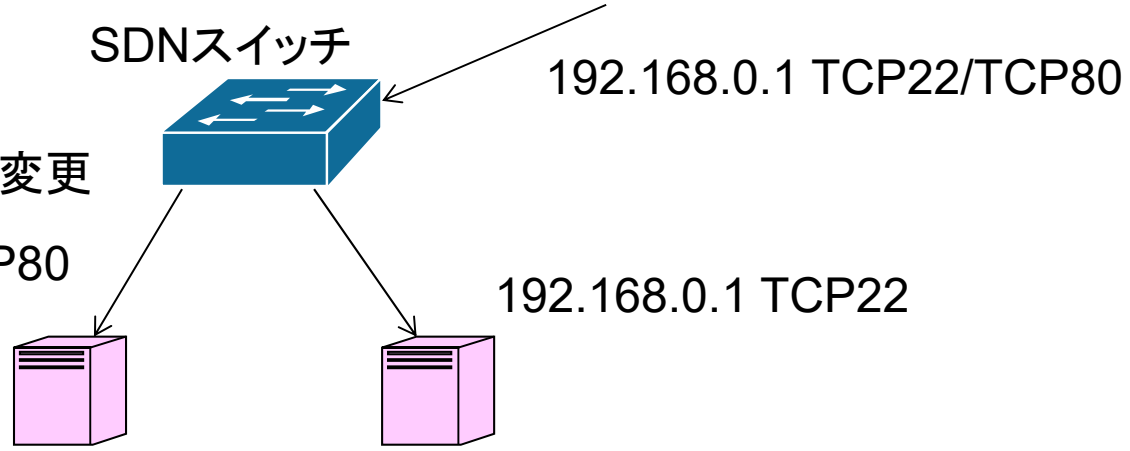


セキュリティ側から見えている希望(2/4)

- SDN(Software Defined Network)による柔軟なネットワーク
- SDNの特徴
  - ソフトウェアのような柔軟な経路選択ルール作成
    - 送信先ポート、送信元IPアドレス/ポート、など
  - 同一IPアドレスに対してTCP/UDPのポートに応じて経路選択可能

    →マルウェアの通信のみ捻じ曲げることが可能(対策、隔離)
  - SDNコントローラでオリジナル認証ルーチン入れたりするのも不可能ではない

SDNでできることの例:
接続先ポートによる経路変更



最近だとSegment Routing over IPv6 (SRv6)に押されてはいるが…

セキュリティ側から見えている希望(3/4)

- ビッグデータ処理の応用
  - 通信解析、マルウェア分類、などへの応用
  - 異常な通信ではなく、通常の通信の定義からの情報セキュリティ適用
  - ビッグデータに向けた計算機の能力向上研究の進歩
  - SNS等で流れる脆弱性に関する議論を自動発掘/分類
- 人が足りないなら自動化すれば良いという目標の研究
  - 熟練情報セキュリティ技術者の知識適用の自動化
  - 別に100%を目指す必要はない
    - 自動化で80%を除外できるならば、人の負荷は1/5になる
- エージェントシミュレーションの応用で攻撃/防御役を競わせて、防御手法を進化させる研究もあり

セキュリティ側から見えている希望(3/4)

- 端末側の性能向上や低電力化でEDRを入れやすくなった
  - EDR(Endpoint Detection and Response): 端末にEDRソフトウェアを入れてイベント(API呼び出し、ファイル操作)を記録し事後対応へ
    - 「マルウェア等に感染して、怪しい行動を取っているけど、マルウェアの初期侵入は防げていない」状態を想定
      - 記録から怪しい端末を探して詳細解析へ
        - 「これ感染しています!」までは断定できないレベルから検知
      - 記録から被害範囲を同定して、いち早く通常業務に復帰
  - 以前はEDR的なことをやると、CPU/HDD負荷が高くて性能や消費電力(バッテリ稼働時間)への影響がでかかった
    - ノートPC用のHDDは必要に応じて回転止める制御とかされていたが、EDR的なことをやっていたら止まる暇が無い
  - 特にCPU性能の向上(マルチコアのコア数増加)とSSD化がでかい

「ネットワークとサーバによるネットワークサービス提供」アップデート

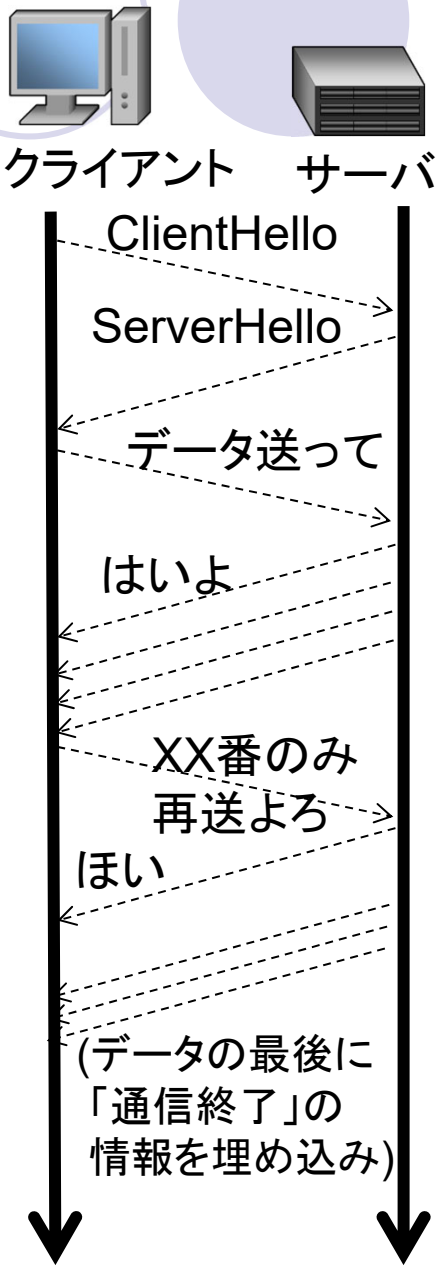
https://www.net.itc.nagoya-u.ac.jp/member/shimada/2025info_sec_lit1/slide/lecture0425slide.pdf

- QUICプロトコル

QUICプロトコル

- TCPの再送処理はネットワークが遅い時代をベースに設計していて現在では効率悪い
- 現在は暗号化は必須だが、TCPハンドシェイク後に暗号化ハンドシェイク実施は効率悪い

- QUIC(Quic UDP Internet Connections)
  - 実際のデータのやりとりはUDPで行う
  - アプリケーション層でTCPの再送処理を実装
  - 通信と暗号化のハンドシェイクを一括実施
  - 2回目以降は前回のハンドシェイク情報を使ってさらに高速にハンドシェイク
- Googleが提案し(2012年)、IETF(Internet Engineering Task Force)が標準化(2021年)



「各種認証とその運用」アップデート

https://www.net.itc.nagoya-u.ac.jp/member/shimada/2025info_sec_lit1/slide/lecture05092slide.pdf

- ポスト量子暗号についてアップデート

余談: 量子コンピュータ実用化よる暗号の解読ってどうなのよ?

- RSAやEC系は量子コンピュータに弱いと言われている
  - というか「量子コンピュータが実用化すれば…」な槍玉にされるぐらい
- 嶋田の個人的な主観
  - 現状で実用化の量子ビット数($10^{3\sim4}$)+αでは、現在公開鍵暗号で標準利用されているビット長の暗号鍵を解くのは無理($10^{6\sim9}$ほど必要)
  - 量子ビット数を増やすほど量子状態の維持が幾何級数的に難化
  - とりあえず、ビット長を長くすれば当分大丈夫では? (すごいブレイクスルーが無い限り)
- 長期的には対量子コンピュータ暗号(ポスト量子暗号、PQC: Post Quantum Cryptography)に移行
  - 米国NISTが2017年にPQC標準化を開始、2024/8に3種類を選定[1]
  - ML-KEM(鍵交換アルゴリズム)、ML-DSA(電子署名)
    - 予備としてSLH-DSA(電子署名)、他にも追加される予定あり

[1] https://www.nist.gov/news-events/news/2024/08/nist-releases-first-3-finalized-post-quantum-encryption-standards

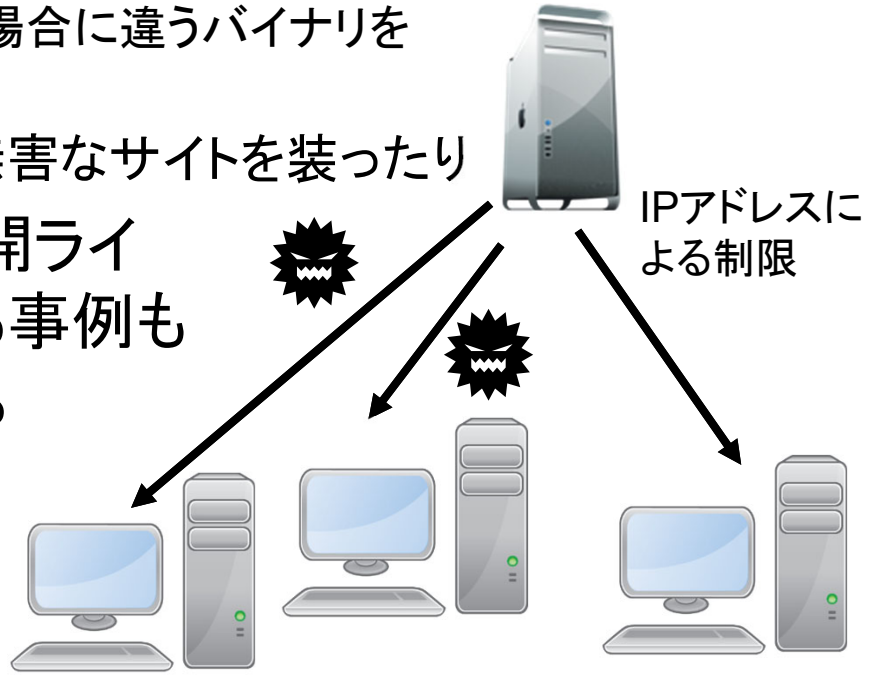
「サイバー攻撃とマルウェア」アップデート

https://www.net.itc.nagoya-u.ac.jp/member/shimada/2025info_sec_lit1/slide/lecture0509.pdf

- 水飲み場型攻撃にフリーウェア/公開ライブラリ/ブラウザ拡張乗っ取り事例を追記
- 偽サーバへの誘導パターンの追加
- 機械学習/深層学習応用システムへの攻撃の話を追加

マルウェア送り込みのテクニック: 水飲み場型攻撃

- 「ある仕事をしている人が頻繁に見るページにマルウェアを仕掛ける」ことによる特定業種の業社への標的型攻撃
- 例: あるソフトウェアの更新ページへの細工
  - 攻撃者がソフトウェア更新ページを乗っ取って悪用
    - 特定IPアドレスから更新が来た場合に違うバイナリを送る攻撃がしかけられていた
  - 逆にセキュリティ関係会社には無害なサイトを装ったり
- 管理が疎かなフリーウェア/公開ライブラリ/ブラウザ拡張を乗っ取る事例も
  - 昨年にはかなり広く使われている圧縮ライブラリxzにバックドアを仕掛けられた事件が(2024/3)[1]
    - かなり精巧にバックドアコードが隠蔽されていて大ニュースに



[1] https://forest.watch.impress.co.jp/docs/news/1580604.html

「情報倫理、ソーシャルエンジニアリング」アップデート

https://www.net.itc.nagoya-u.ac.jp/member/shimada/2025info_sec_lit1/slide/lecture0516slide.pdf

- 公共の場での顔認識問題に追記
- 2023年通称マイナンバー法改正
- AI応用領域における倫理
  - 2024/5成立のEUのAI act

公共の場での顔認識(顔認証)問題

- 顔認識技術の発達で、顔認識の濫用が問題となっている
  - 個人の行動の自由に付随する、「正当な理由なく、個人の行動を追跡したり記録する(ことで萎縮させ自由を奪う)ことの禁止」に抵触
  - 認識精度も低い上、生体情報は変更は難しい →冤罪が連続した話
- 2019/5に米国サンフランシスコ市が「公共機関による顔認証技術の使用を禁止する条例案」を可決[1]
  - その後、米国内で追従する自治体がちらほら出る
- 2024/5にEUで顔認証を含むAI技術規制法が成立[2]
  - 先行している物と同様、公共の場での顔認識の利用を禁止
  - 3年以上の懲役刑になる可能性の物など例外はあり
  - 悪性度の高い「収集した個人情報をもとにしたターゲティング公告」も規制

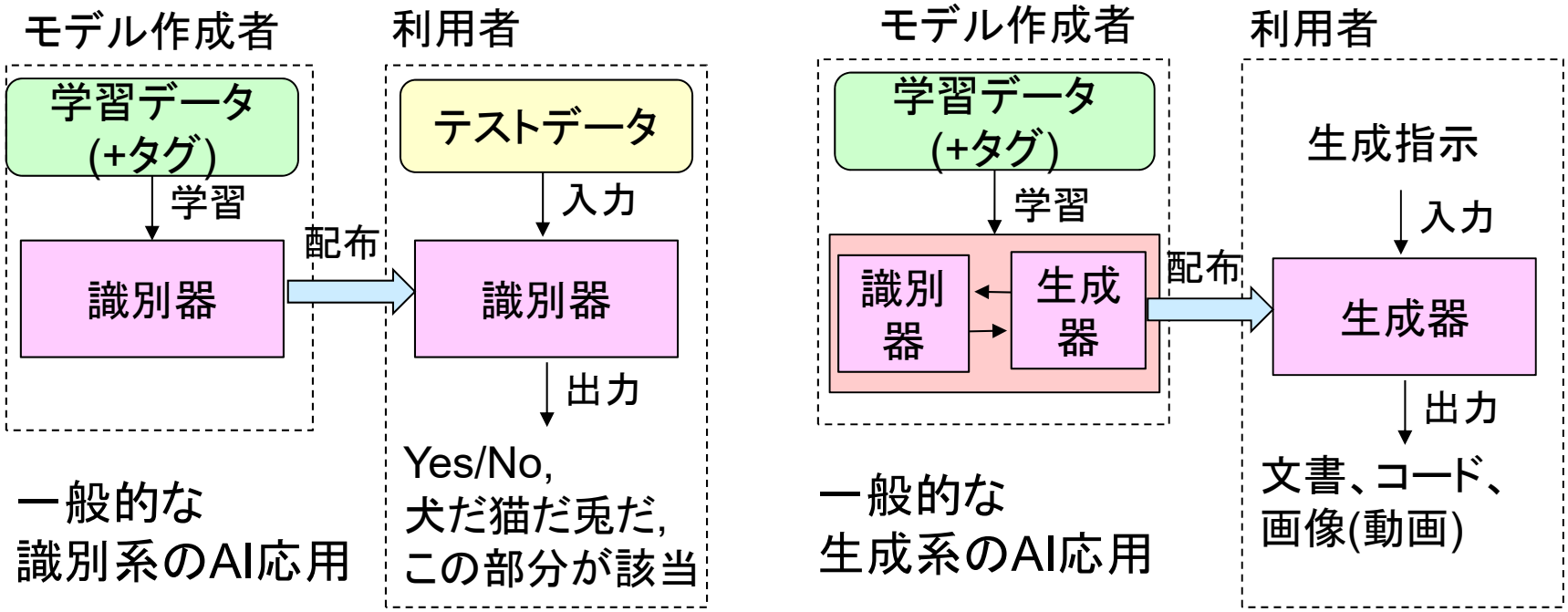
[1] https://www.bbc.com/japanese/48276999
[2] https://pc.watch.impress.co.jp/docs/news/1593617.html

2023年通称マイナンバー法改正

- 正式名称: 行政手続における特定の個人を識別するための番号の利用等に関する法律等
- 個人番号の利用範囲の拡大: 国家資格取得者の管理に個人番号を利用可能とする
  - 医師免許とか美容師免許とか建築士とか
- 住民基本台帳法に関連して、現在利用が認められている事務に準ずる事務でも個人番号利用を可能とする
  - 個人的には、これが一番いろいろ解釈を広げれそうで嫌(民間利用につながらないか?)
- マイナンバーカードの健康保険証化
- カードの利用の促進: カードに関する申請の容易化

(個人認証に使う暗号の寿命の関係で、次期マイナンバーカードの設計も始まっている)

一般的なAI応用のざっくりとした説明



非常にざっくりとした補足説明

- 識別器(Classifier)の学習には教師あり学習と教師なし学習がある(タグがあるか否か)が、多くは教師あり学習
- 生成器の学習には、生成器の出力の識別器による評価を受けての更新を繰り返す

学習およびモデル生成における倫理 (1/2)

- 集めてきた学習データを、本当に学習データとして用いることに倫理的な問題は無いのか?
  - 違法に収集されたデータは含まれていないのか?
  - 学習データの生成者はモデルの学習に利用することを想定して公開しているのか?
  - 学習データの生成者に報酬が何もないのはどうか?
  - EUでは学習データの公開を盛り込んだ法案も[1]
- 学習データから発生する偏りの問題 (公平性の問題)
  - 現状の世の中のデータをそのまま学習に使うと、現状の社会の母集団に対する偏りまで学習してしまう →要学習前の調整
    - 自分の研究でも、悪性サンプルと良性サンプルの調整は重要
  - 例: 職業の絵の出力と人種、家庭内での仕事の絵の出力と性別

[1] https://gigazine.net/news/20230428-eu-new-copyright-rules-for-generative-ai/

学習およびモデル生成における倫理 (2/2)

- モデルから学習データを復元する攻撃で、学習データに含まれていた秘匿性の高い情報とかは大丈夫か?
  - モデルに対する攻撃は「移転攻撃」と言われ、モデルの模倣や学習データ復元などが代表例
  - すでに生成系モデルで、学習に回された以前の入力を復元できたような事例が
- 生成し配布されたモデルは信頼できるのか?
  - 悪意の込められたモデルは配布されたりしないのか?
  - 例: 特定の攻撃パターンのみ反応しない攻撃検知モデル
- (モデルのサイズや学習データ量が大きくなり過ぎて、お金を持っているビッグテックの一人勝ちを加速するのでは?)
  - モデルが大きくなると学習のコストが大きくなったり、そもそもメモリ量で学習に制限がかかったり

モデルからの出力における倫理(1/2)

- 生成器の嘘出力(Hallucination)の問題をどうするか?
  - Hallucination = 幻覚と言われているが、個人的にはAI側に甘い用語な印象なので「嘘」と言いたい
  - 識別器でも誤判別の責任はどこが? (作成者? 運用者?)
  - とりあえず、(意図的に作って)ばらまいた人が逮捕された事例あり[1]
- 生成器による生成物が著作権を侵害する可能性
  - 特定の人の絵や文章を真似た出力は著作権侵害か?
    - 特に、特定の絵柄に出力を寄せることができる生成器とか
    - 日本の文化庁は「生成物が侵害していたら著作権侵害」な見解(人が著作権侵害物を作った時と同じ扱い)
  - 人間による生成でも、悪性度(目的)によって贋作とかパクリとかに
  - このあたりで、オリジナルのクリエイターと生成器作成側での訴訟が多い印象

[1] https://pc.watch.impress.co.jp/docs/news/1499306.html

モデルからの出力における倫理(2/2)

- 教育課程における悪用問題
  - 多くの大学で従来のレポートの(素案)作成への利用を警戒
    - 本人の学習にならないのは自業自得だが、成績評価結果が就職活動や奨学金など他で利用されるので悪用は公平さに問題を起こす
    - 「せっかくなので、それを活用することを前提とした課題にしよう」になる方向?
  - 東北大の教職員向け話の他大学の取り組みのリンク集が便利
    - 教職員向け https://olg.cds.tohoku.ac.jp/forstaff/ai-tools
    - 学生向け https://olg.cds.tohoku.ac.jp/forstudents/ai-tools
- 頭の悪い人による出力の盲信
  - できることできないこと利点欠点を全く理解してない情報発信が山程見られる
  - その情報発信をさらに組織の上層部が盲信するとさらに面倒に
  - まあ、AI関係に限らないが(ブロックチェーン系とか)

AIと倫理の現状

- XAI(eXplainable AI)等で出力結果の保証をする試みはある
  - 「どういう根拠(学習した内容など)から結果を導き出したか」などを結果とともに示す
  - 現状では人間側がXAI出力を見て判断しているが自動化は進むはず
  - (個人的には、SHAPとAttentionの派生物を利用した研究をよく見かける印象)
- 生成系はなんだかんだ言って、分かっている人が素案を作成するコストを減らすには便利
  - 「失礼の無い英語での通知文」とか
- ちょうど現在、ものすごくホットで立法とか訴訟とか議論が盛り上がっている分野なので、おっかけると楽しい

EUのAI法(AI act) (1/2)

- 正式名称: The Artificial Intelligence Act
  - 「規制/禁止」も含めた、包括的なAIの取り扱いを定める法律
- 許容できないリスクのため禁止、ハイリスクのため規制下利用、リスクは限定的だが要透明性、規制無し、の4階層分類
- 他の国でもAI(規制)法の検討は進められている
  - EUでも英独仏の大国は、EU法に加えた追加規制を検討中
  - 米国では州法が追加規制したり(例: カリフォルニア州)
- GDPR同様、EU外でも法が適用されること多し(抜け穴塞ぎ)
  - 留学生など日本に来るEU市民に対しても法適用
  - 「EU外で作成したAI出力をもとにEUで何かやる」運用にも法適用
- GDPR同様、違反時のペナルティは大きい
  - 最大で3500万ユーロもしくは全世界売上の7%のどちらか高い方(「禁止されるAI利用」に関する違反)

EUのAI法(AI act) (2/2)

EUのAI法の個人的な注目点

- 機械学習系(DNN含む)のみならず、統計的アプローチや論理ベースのアプローチのシステムも対象(抜け穴塞ぎ?)
- 人に対するAI判定を根拠とした自動選別に特に強い規制
  - (多分)AIを恣意的に運用した差別的な選別が出てくることへの対策
  - ちゃんと人間が責任持つことと、非選別者からの問い合わせには根拠を出すこと(当たり前)
- AIシステムのリコールや「市場からの取り下げ」も指示できる
- 人に対してサブリミナルや偽情報も含めた誘導の禁止
- 識別器/生成器作成時のアルゴリズムや学習データ、サービス提供時の生成ログの維持や監視とか、透明性の義務大

どのような点が各国のAI法案での焦点になっているか？

他国でもAI(規制)法は検討されている

- 学習について
  - 開発・学習に関する権利制限
  - 開発・学習に関する透明性
  - 開発・学習に対する権利者への補償金やオプトアウト
- 生成物に対する扱い
  - 生成物の著作権を持ったり特許を取得できるか
  - 生成物の著作権侵害について
  - AI生成物やAI応答システムであることの明示

「情報セキュリティと情報倫理に関連する法律」アップデート

https://www.net.itc.nagoya-u.ac.jp/member/shimada/2025info_sec_lit1/slide/lecture0530slide.pdf

- こちらについては、2026年秋の情報セキュリティ特論Aにて