AI品質マネジメントシンポジウム
パネルディスカッション
～AIセキュリティ～

2023/10/31

住友電気工業株式会社 サイバーセキュリティ研究開発室
三宅　和公

機械学習品質マネジメントに対するAIセキュリティの貢献-1

1 住友電気工業株式会社 サイバーセキュリティ研究開発室
三宅 和公(みやけ かずまさ)

2 自己紹介：
ネットワーク製品の組込みシステム・ASPなどのシステム開発を経て自動車などのセキュリティアセスメントを担当する傍らデータサイエンス・統計学に取り組み、
NEDO「実データで学ぶ人工知能講座」を経て現在は機械学習品質マネジメントガイドラインのAIセキュリティ章を執筆。

機械学習品質マネジメントに対するAIセキュリティの貢献-2

3 AI品質評価の実施状況と課題（セキュリティ目線・社内）

状況

- ●開発はこれから。実製品も投入開始。
- ●実ビジネス対応：製品開発への反映を検討

課題

- ●AIを搭載するシステムのセキュリティ「**AIセキュリティ**」**分析と対策**は？これまでの枠組みで分析・対応できる？
- ●**AIセキュリティは近々喫緊の課題**に（車載セキュリティや工場セキュリティを見ると・・・）

1 機械学習品質マネジメントに対するAIセキュリティの貢献-3

4 AI品質評価の対策に向けて何が必要か？

- AIセキュリティと情報セキュリティの違いは？
  AIセキュリティ：**機械学習の技術的な特徴をつけ狙う。**
  既存の情報セキュリティ：「AIの技術的特徴」なし。
  「AIの技術的特徴」への対策は？
- セキュリティへの取り組みの普及
  設計開発段階で対応し**手戻り防止**：シフトレフト
- **統計的な出力**をするシステムの特徴をどう捉えるか：
  既存のセキュリティの枠組みをどうするか？追加？変更？

1 機械学習品質マネジメントに対するAIセキュリティの貢献-4

5 AIQMへの期待（セキュリティ目線）

(1)AIセキュリティの分析と対策の進め方

攻撃の分類、脅威・脆弱性は？**管理策**は？

→**AIの技術的特徴を入れた**攻撃、脅威・脆弱性・管理策の概念・例を提供。

(2) これからの機械学習の枠組みに対してどのようなセキュリティを用意するか。

生成モデル・基盤モデル

付録1

情報セキュリティとAIセキュリティの違い

AIセキュリティ：攻撃者は機械学習技術を精査して攻撃手法に取り入れ、より大きな被害を目指す。

【データやプログラムの改ざん】

情報セキュリティの事例：バッファオーバーラン

特定のメモリやストレージ領域を改ざんする。

改ざんの仕方は決まったやり方（0x00塗り潰しなど）。→どんな被害が出るかは意識していない。

AIセキュリティの事例：データポイズニング、モデルポイズニング

機械学習の特徴（学習データ、学習モデルの作り方など）を利用して改ざん

被害（誤動作の度合いなど）**がより大きくなる**よう改ざんの仕方を工夫。

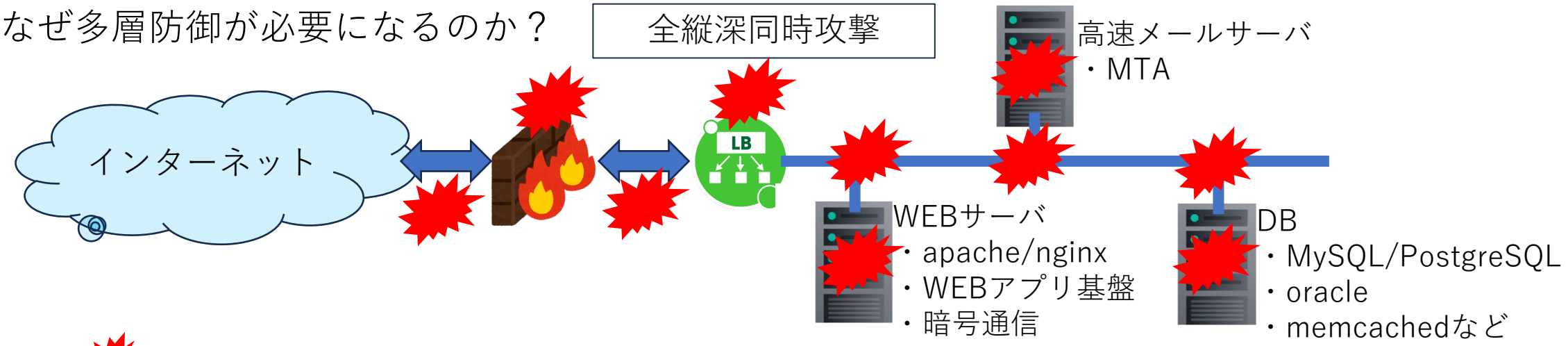
★今後、様々な攻撃手法が開発される。

ご参考：最近の情報セキュリティの特徴

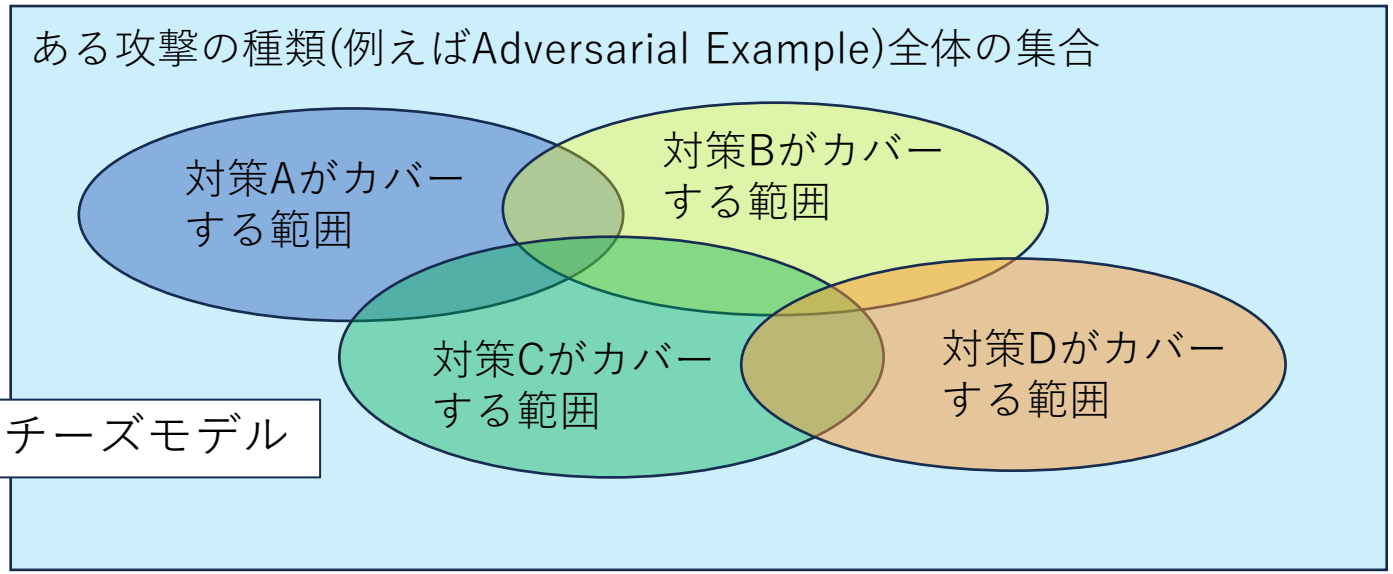
- ビジネス化、ビジネスレイヤ
- 国家の利益、軍事の関与
- 攻撃戦略の進化
- DXやCOVID 19による環境変化の悪用

付録2

なぜ多層防御が必要になるのか？



攻撃 ✸ はほぼ同時に発生する。しかもひとつの対策が全ての対策をカバーできることはない。



管理策(対策)はカバーできる範囲が決まっており多数の対策を同時に実施することでカバーできる範囲を増やす。
★ひとつの対策で全攻撃の種類をカバーできることはない。
（**No Silver Ballet**）

どんなに丸を増やしてもカバーできないところが残ってしまい狙われる。