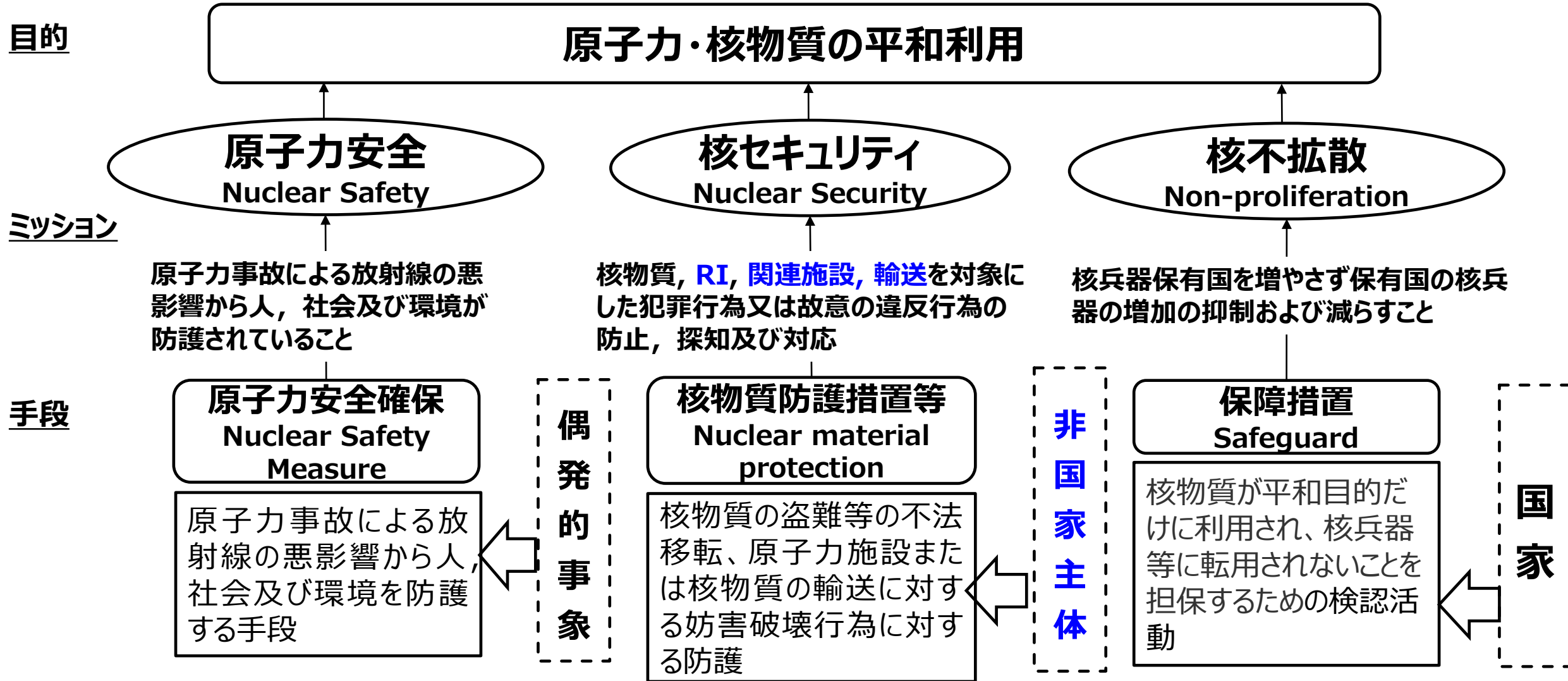


原子力施設の核セキュリティ強化

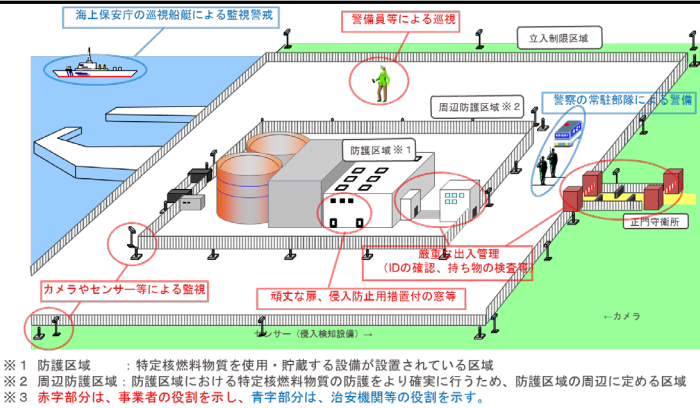
～新たな脅威とBDBTへの対策～

令和5年12月18日（月）
東京大学大学院工学系研究科
原子力国際専攻
准教授 出町 和之

1. BDBTへの対策強化の必要性



抑止に偏重した核セキュリティ対策



- 外部からの核セキュリティ脅威に対する防護は強化された
- しかし、**抑止への偏重**があり、「新たな脅威」には不十分
[例] 内部脅威への対策は「個人の信頼性調査(自己申告)」および「2人ルール」*1 という**抑止**のみであり、海外に比べても非常に脆弱。スタンドオフには**無策**
- 「**新たな脅威**」には「**検知**」の強化も必要
- 「**BDBT(設計基準外脅威)***2」に対する**合理的な備え**も必要

*1:平成17年「内部脅威者について」(文部科学省)

*2:設計基礎脅威(DBT: Design Basis Threat)、設計基礎を超える脅威(BDBT: Beyond Design Basis Threat)

核不拡散への偏重

- とくに日本では従来、核不拡散 > 核セキュリティ
- 研究面でも核物質そのものを扱う研究が永らく中心的位置(高拡散抵抗性燃料や計量管理等)
- 「核物質の盗取」に有効でも、**新たな脅威には通用しない**。
- ザポリージャ原子力発電所占拠のように、従来の核セキュリティでは対応しきれない**想定を超える脅威はすでに発生**。



*3:笹川平和財団「緊急提言 ロシアによるウクライナ侵攻:原子力民生利用の諸課題と日本の役割」

「想定外でした」という事態を避けるため、「**今からBDBTへの備えを始めねばならない**」²

~INFCIRC/225/Rev4

核テロの脅威

- ① 核兵器の盗取
- ② 核物質の盗取
- (③ 妨害破壊行為)



9.11
→
多様化

INFCIRC/225/Rev5

核テロ&社会テロの脅威

- ① 核兵器の盗取
- ② 核物質の盗取
- ③ **ダーティボム用RIの盗取**
- ④ **妨害破壊行為(原子力施設)**



→
巧妙化

新たな脅威(手段)

- ① 内部脅威者
- ② サイバー攻撃
- ③ Stand-off攻撃



INFCIRC/225/Rev6 ?

想定を超える脅威

- ① **戦時の核セキュリティ**
- ② **BDBT※1のための核セキュリティ**

→
拡大



*1設計基礎を超える脅威 (BDBT:Beyond Design Basis Threat)

- 核セキュリティ脅威は現在進行形で多様化・巧妙化・拡大を続ける
- 原子力発電所の戦時利用に対する道義的ハードルが下がってしまった。
- 核セキュリティの対象外として避けていたBDBTを真剣に考慮すべき時代に突入。

規制も事業者も互いに丸投げし責任回避

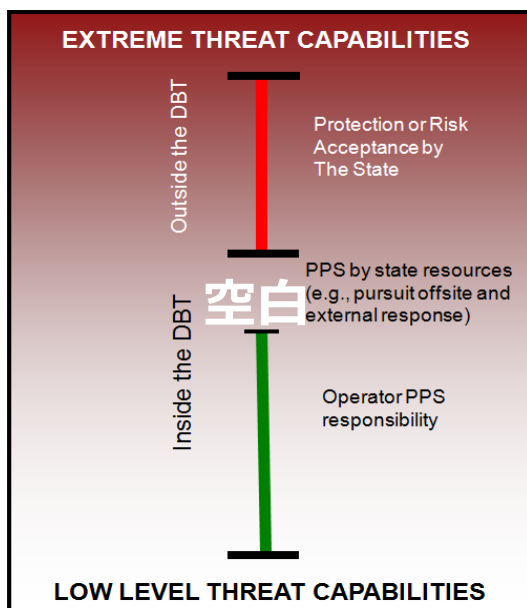
規制：航空機衝突などを想定した「特重」で思考停止

事業者：BDBT*¹は国の管轄だからと「特重」以外は対応しない

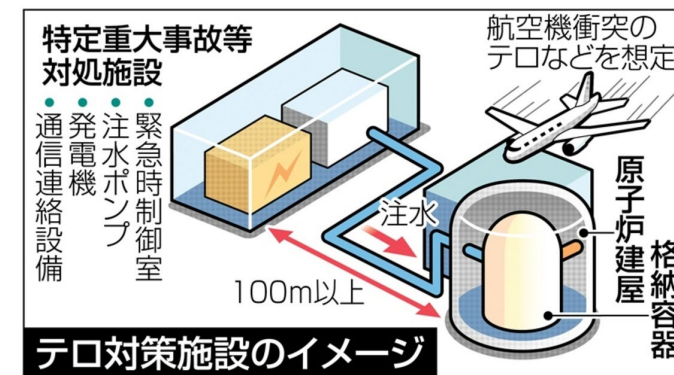
脅威の規模も種類も多すぎて何も決められない(by インタビュー)



まずは、BDBT時に何が起こるか想定することから始める



○意図的な航空機衝突への対応の新設



第1段階：仮想ストレステストでBDBTの「想定外でした」を減らす

- まずは、コストのかからないバーチャルで「想定外の」BDBT脆弱ポイントを見極める
- 検討すべき状況が多様過ぎるためコンピュータの力を借りねばならない
- 出来るだけ多くの状況を検討する為、1シナリオごとの計算時間は短く

第2段階：ストレステストをBDBT時の最適策提案AIに活用

- 原子力安全では過酷事故時(BDBE)の対策が整備されルール化されている
 - しかし、多様な状況がありえるBDBT時への対策のルール化は困難
 - 目前のBDBTに対する最適対策を高速で提案し提示するシステムが有効
 - BDBTストレステストを強化学習モデルの「環境」として活用すれば可能か？
- ◆ Society5.0の「提案するAI」を満たす原子力DXの一つとなりえる。

2. BDBTへの対策強化

I: BDBT仮想ストレステスト

II: BDBT時の最適策提案AI

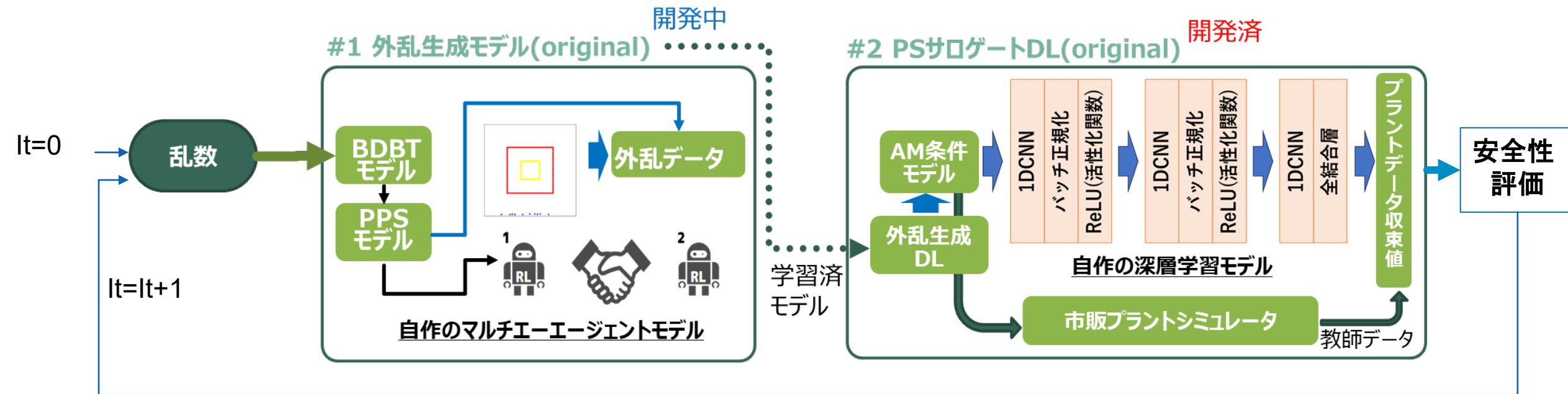
- ① BDBT時に想定すべき状況は、自然災害等に比べて「ヒト」が関わる分だけ非常に**複雑多様**
- ② 「ヒト」と「モノ」の**組み合わせと時系列**のケースはほぼ**無限**にあり得る。
- ③ しかし、状況に関わる「ヒト」と「モノ」そのものは**限定**可能。
- ④ 手動では無理でも、**シミュレーション**によりできるだけ広範囲に「**総当たり計算**」なら可能。

- ⑤ 原子力発電所へのBDBTの場合、**プラント状態の把握**も不可欠。
- ⑥ **プラントシミュレータ**との連動が必要だが、「総当たり計算」のためには**計算時間が膨大**に。
- ⑦ プラントシミュレータの高速化が必須だが、これにはAIを用いた「**サロゲート化**」が有効。
- ⑧ 外乱による単なるプラント状態変化ではなく、**SA対応**による**状態回復/失敗**の計算も必要。

- ⑨ **BDBT入力**も「ヒト」と「モノ」の多様な組み合わせを時系列で動かすには**手作業では不可能**。
- ⑩ 「ヒト」と「モノ」を属性に従って自由に動かせる「**BDBTマルチエージェントモデル**」が必要。

仮想BDBTストレステスト手法の例

#	名称	機能	開発状況
#1	外乱生成モデル	PPSを通り抜けたBDBTが発生するプラント機器の機能喪失評価	<ul style="list-style-type: none"> プラント異常検知手法としてマルチエージェントモデルを自作中。
#2	PSサロゲートDLモデル	外乱を入力にプラントデータ収束値を高速予測	<ul style="list-style-type: none"> 某研究炉用のサロゲートDLモデルを自作。 「外乱⇒回復・変化」でMSE0.05



#1-1 Adversariesのマルチエージェント化

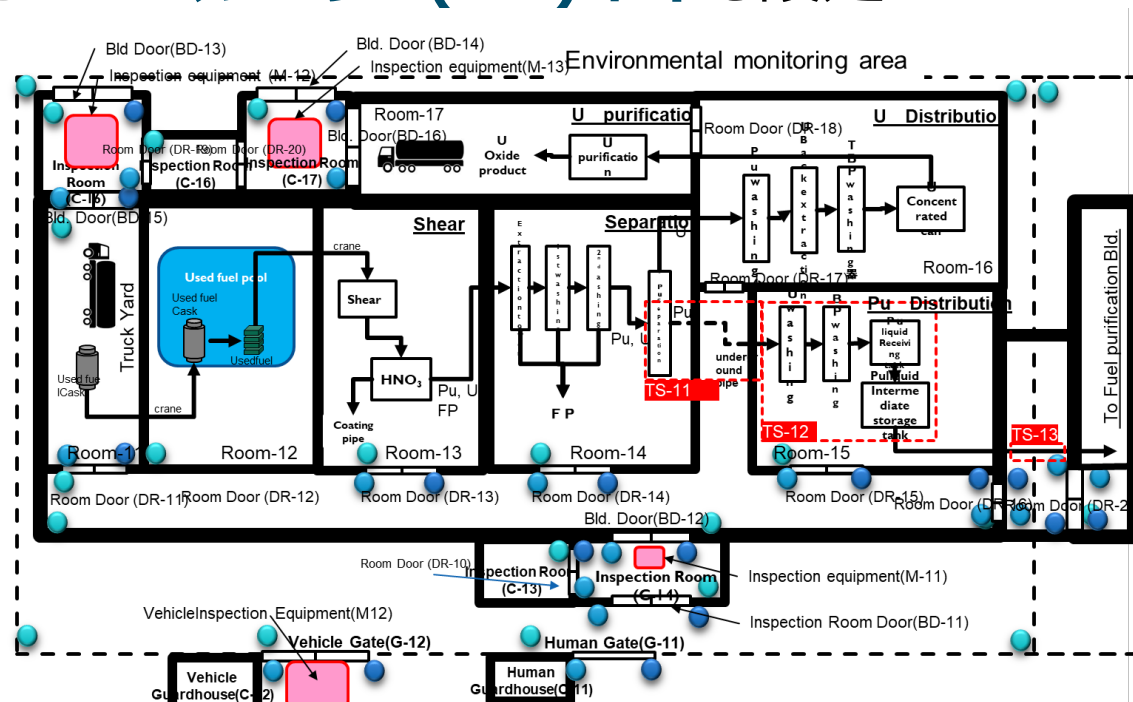
- BDBT時で想定すべき状況は多様であるが、その**状況を構成するエージェントと特性は有限**であり設定は可能
- **エージェントの種類**としてインベーター(外部侵入者)・インサイダー(内部脅威者)・サイバー攻撃・スタンドオフ(遠隔)攻撃・スカイアタック(航空機衝突)・軍事侵攻が設定可能
- **各エージェントに考慮すべき要素**として目的・人数・武力・知識・技能・所要時間等を設定。

想定するエージェント

Agent 1 侵入者	武装した外部作業員〇名が制御システムや重要機器を目指す
Agent 2 内部脅威者	従業員〇名がAgent1の侵入を手助けする
Agent 3 ハッカー	〇名のハッカーが制御システムや重要機器への妨害破壊行為を行う
Agent 4 Stand-off	〇機のドローンが電源や冷却システムを標的として飛来する
Agent 5 Army	〇名からなる兵力がPPSを破壊し制御室の制圧を目指す

1-2 物理的防護(PPS)のマルチエージェント化

- **PPSを通り抜けるBDBTを求めるため、PPSもモデル化**
- **PPSのエージェント**として立入制限区域、周辺監視区域、防護区域の侵入防止設備、侵入検知センサ類、監視カメラ、警備員、対抗部隊などを設定
- **各エージェントに考慮すべき要素**として設置場所・設置個数・防護 & 検知能力(範囲・性能・確率・時間)・対抗部隊(機動隊等)の到着時間・鎮圧能力等を設定
- 警報見逃しなど**ヒューマンエラー(HE)確率**も設定



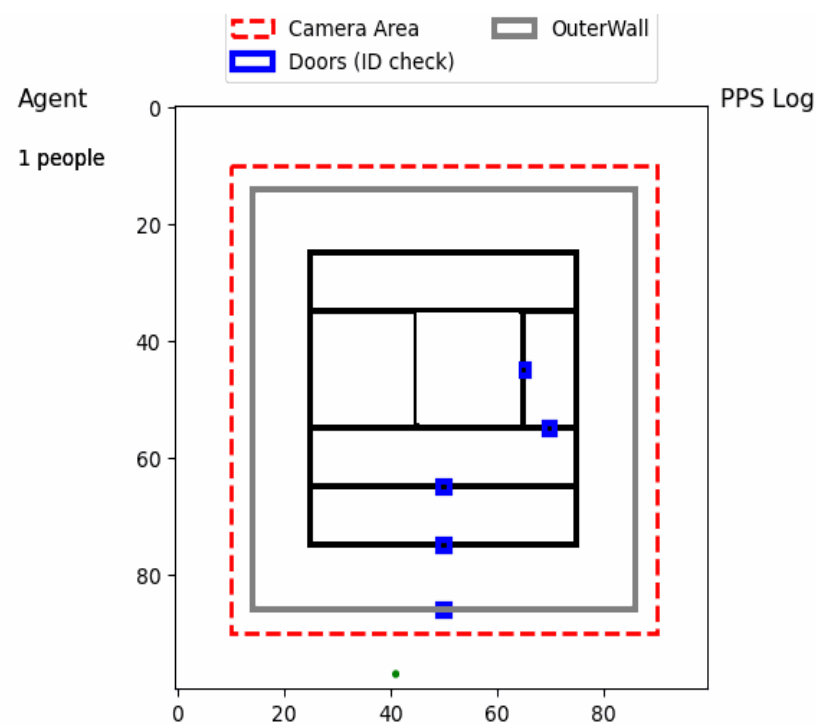
- ID Authentication
- Door Sensor
- Security Camera

試算：中央管理室占拠シミュレーション(1/4)

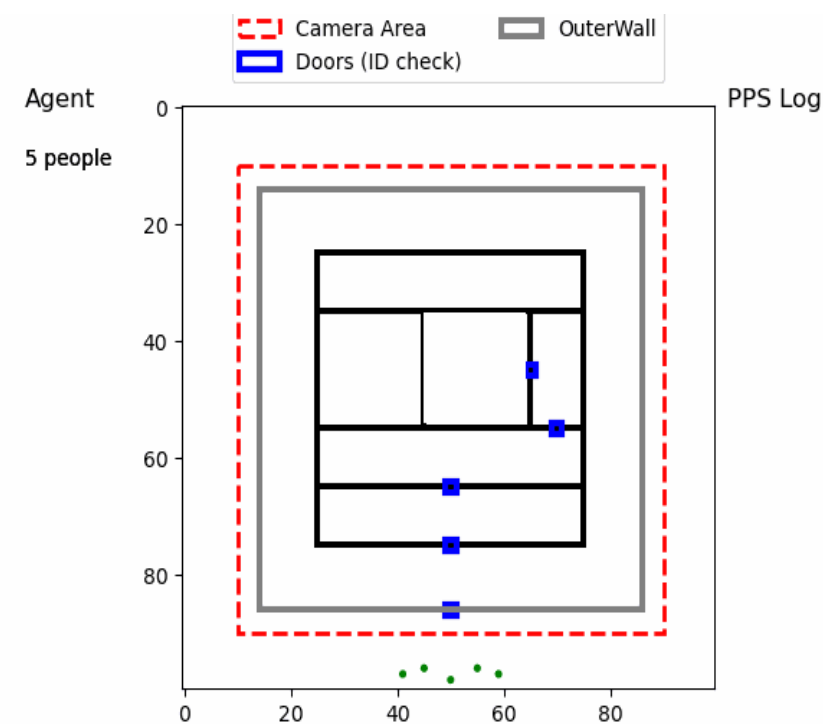
設定

- マップ内には外壁と建物が存在
 - スタンドオフ攻撃によって外壁を即時破壊できる
- 建物はいくつかの部屋に分かれている
 - 部屋間の移動にはドアを使用
 - 内通者がいればID認証で通過、いなければ破壊
- 攻撃エージェントは中央管理室の占拠を目指す
- 建物の周りには監視カメラエリアが存在
 - 監視カメラに見つかり10分以内に制圧部隊が来る

内通者なし・スタンドオフなし



エージェント数:1 (占拠阻止)



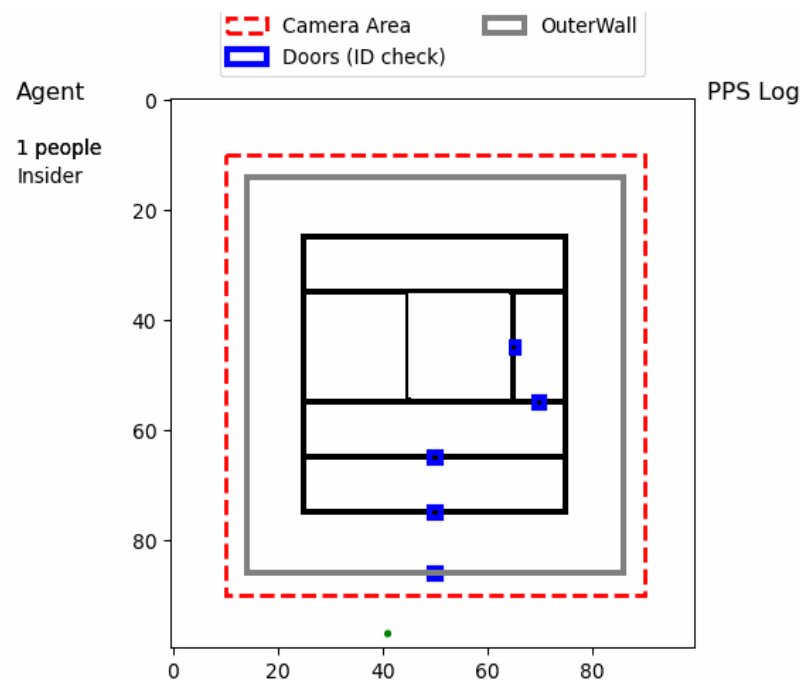
エージェント数:5 (406stepsで占拠)

試計算：中央管理室占拠シミュレーション(2/4)

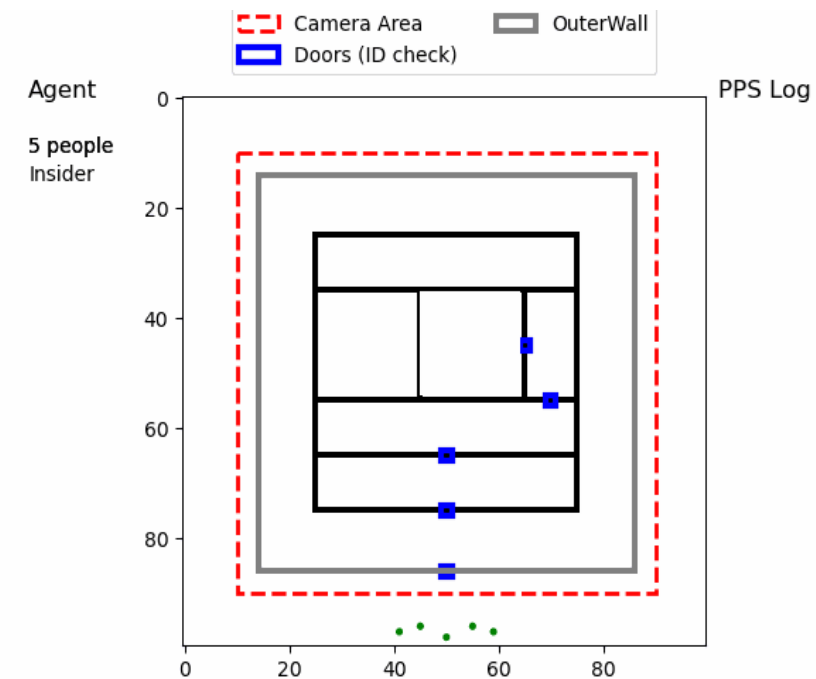
設定

- マップ内には外壁と建物が存在
 - スタンドオフ攻撃によって外壁を即時破壊できる
- 建物はいくつかの部屋に分かれている
 - 部屋間の移動にはドアを使用
 - 内通者がいればID認証で通過、いなければ破壊
- 攻撃エージェントは中央管理室の占拠を目指す
- 建物の周りには監視カメラエリアが存在
 - 監視カメラに見つされると10分以内に制圧部隊が来る

内通者あり・スタンドオフなし



エージェント数:2 (⇒167stepsで占拠)



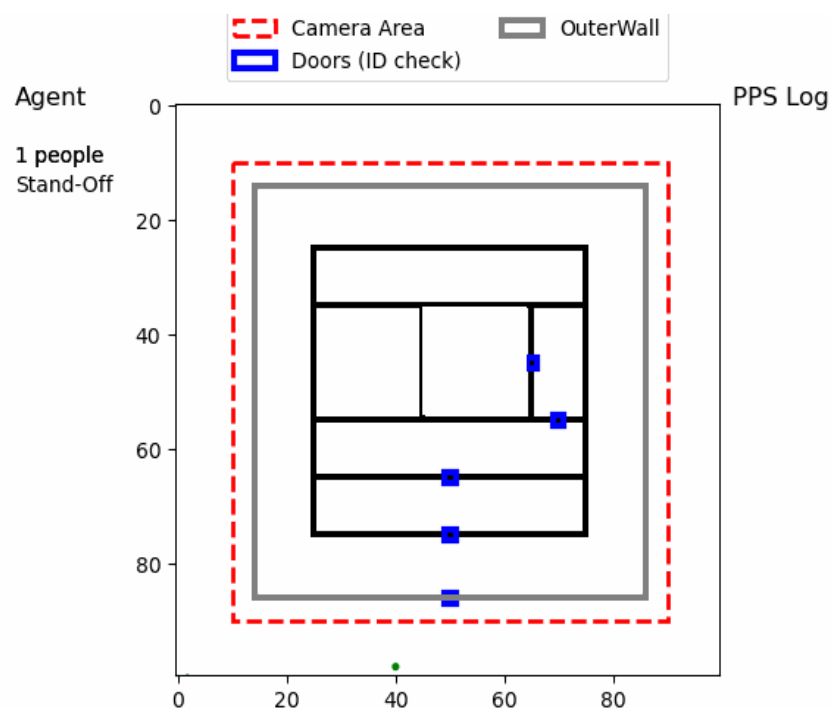
エージェント数:5 (⇒ 111stepsで占拠)

試算：中央管理室占拠シミュレーション(3/4)

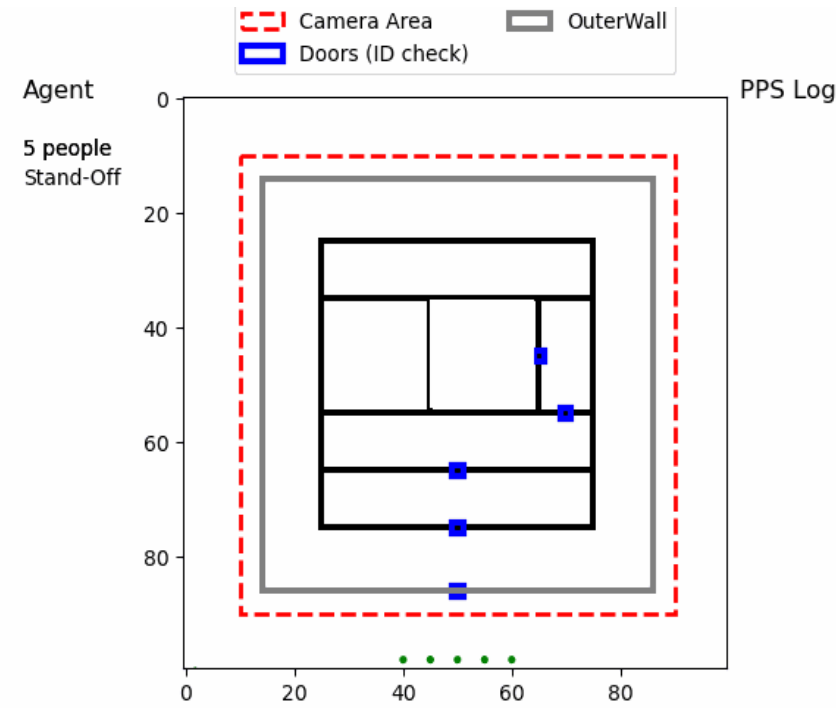
設定

- マップ内には外壁と建物が存在
 - スタンドオフ攻撃によって外壁を即時破壊できる
- 建物はいくつかの部屋に分かれている
 - 部屋間の移動にはドアを使用
 - 内通者がいればID認証で通過、いなければ破壊
- 攻撃エージェントは中央管理室の占拠を目指す
- 建物の周りには監視カメラエリアが存在
 - 監視カメラに見つかり10分以内に制圧部隊が来る

内通者なし・スタンドオフあり



エージェント数:1 (⇒584stepsで占拠)



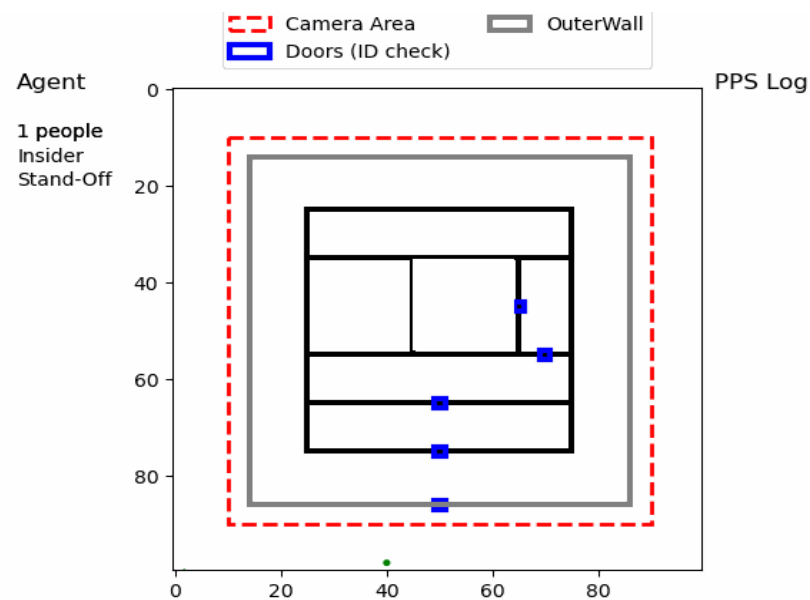
エージェント数:5 (⇒ 227stepsで占拠)

試計算：中央管理室占拠シミュレーション(2/4)

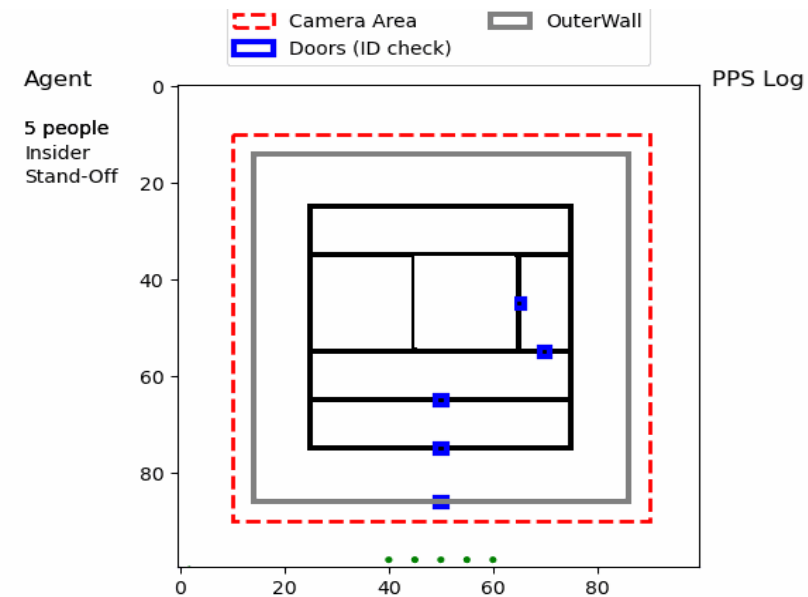
設定

- マップ内には外壁と建物が存在
 - スタンドオフ攻撃によって外壁を即時破壊できる
- 建物はいくつかの部屋に分かれている
 - 部屋間の移動にはドアを使用
 - 内通者がいればID認証で通過、いなければ破壊
- 攻撃エージェントは中央管理室の占拠を目指す
- 建物の周りには監視カメラエリアが存在
 - 監視カメラに見つかり10分以内に制圧部隊が来る

内通者あり・スタンドオフあり



エージェント数:1 (⇒188stepsで占拠)



エージェント数:5 (⇒ 132stepsで占拠)

試計算：中央管理室占拠シミュレーションまとめ

	1 person	2 people	3 people
内通者なし スタンドオフなし	占拠阻止	占拠阻止	406 steps
内通者あり スタンドオフなし	167	130	111
内通者なし スタンドオフあり	584	358	227
内通者あり スタンドオフあり	188	151	132

外乱に対するプラント応答を予測するPSサロゲートDLモデル

■ 計算時間の極短化

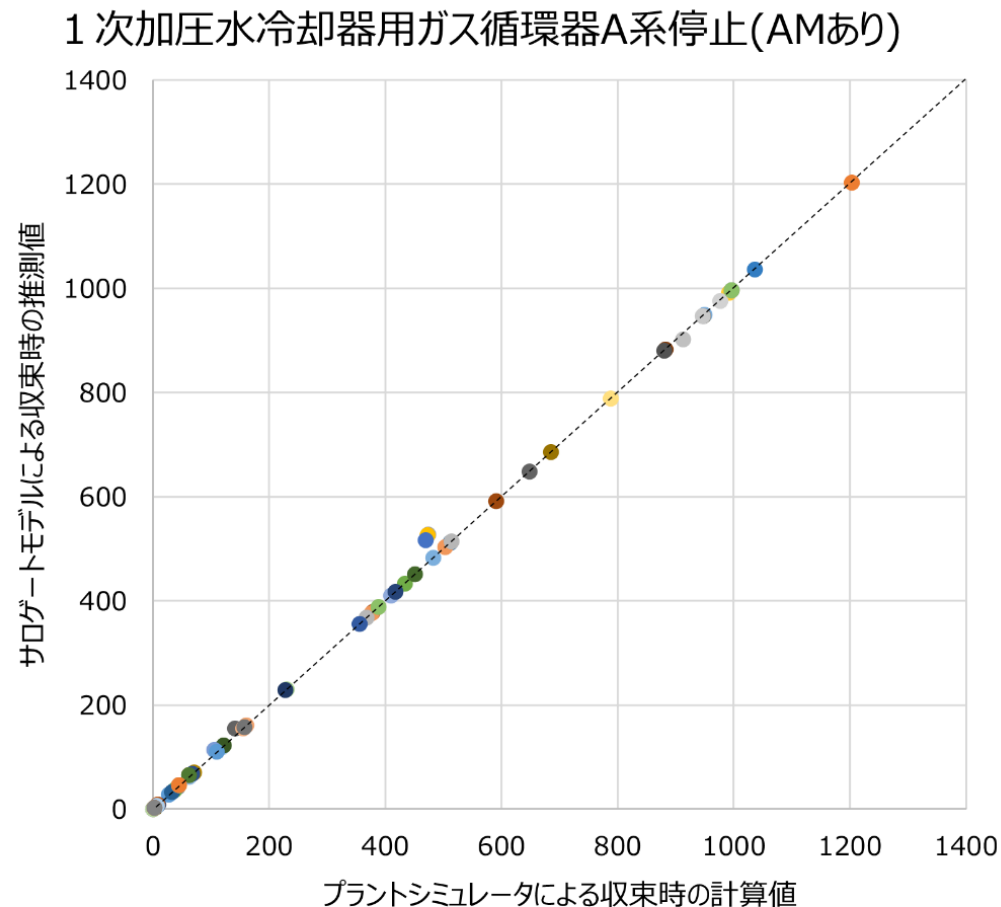
プラントシミュレータ：10時間

サロゲートAI化：0.33秒

(高温ガス炉(HTTR)のサロゲートAIの場合)

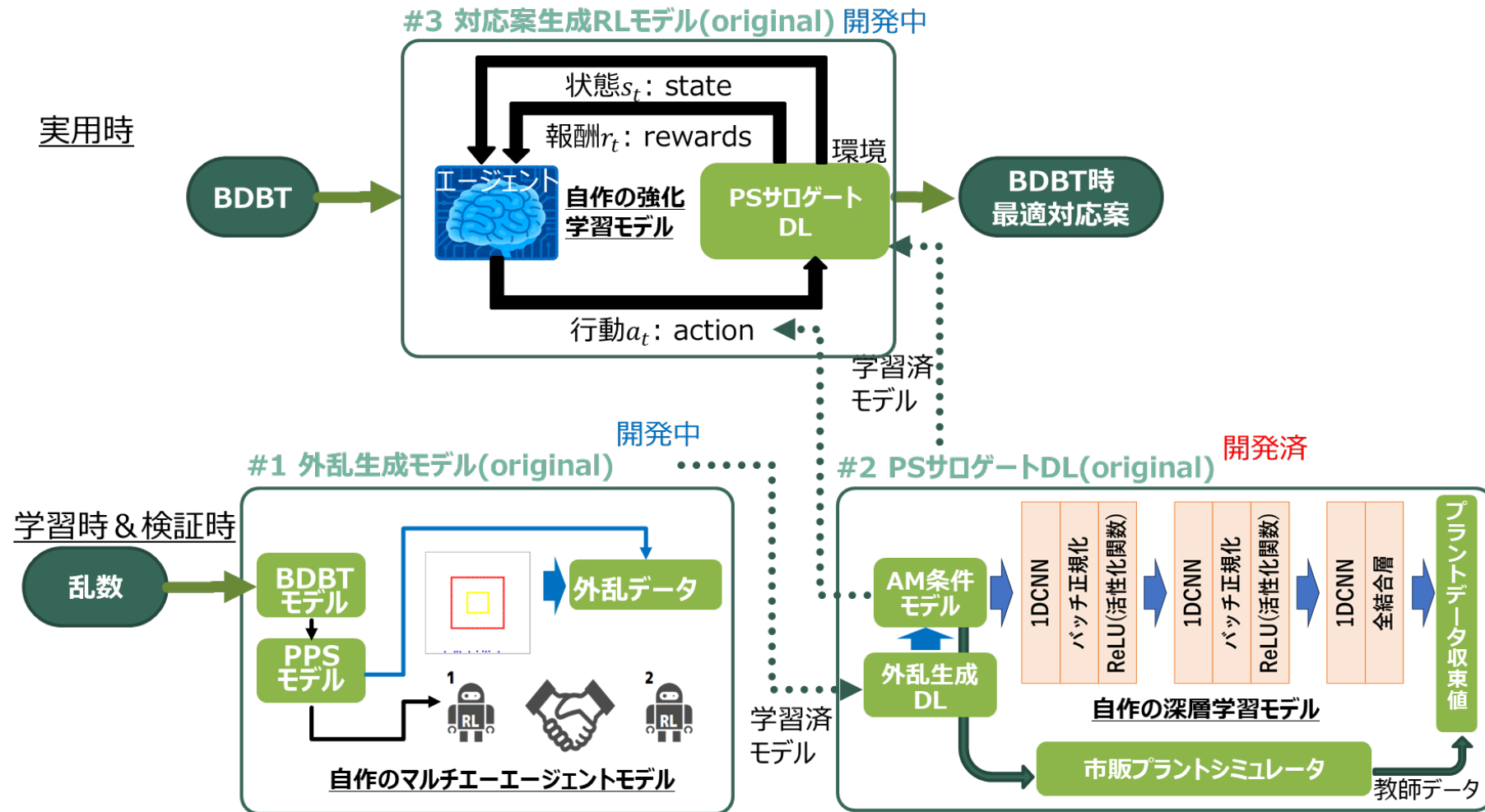
■ 計算誤差は微小(右図)

プラントシミュレータの計算値との
誤差 = 数パーセント



**単なる異常時のプラントデータの予測ではなく、
さらに回復策を施した後のプラントデータ値の予測にも成功**

- BDBTストレステストシステムを強化学習AIの「環境」として適用するシステムを開発中
- 安全回復に成功したBDBT対応案に「高報酬」を与えて強化学習モデルを学習



3. まとめ

- 2022年の露軍によるザポリージャNPP占拠のように、核セキュリティ脅威はBDBTを無視できない大きなパラダイムシフトを迎えた。
- 日本も世界も、核セキュリティ対策は高拡散抵抗性と核物質管理に偏重しており、核セキュリティ脅威の変遷に追いついていない。
- **JAEAが日本と世界に先駆け、BDBTに対する対策の方法論を確立を**
- まずは、あまりに多様な状況がありえるBDBT時における原子力施設の脆弱点を洗い出して「想定外でした」を可能な限り無くすることが第1のステップ。
- その後、抽出された脆弱点に対する対策に重要度をつけ強化していく。