コード信号と交渉ゲームを用いた スマートグリッドにおけるリプレイ攻撃の検出

入 田 隆*・篠 原 巧*・滑 川 徹*,**

Detection of Replay Attack on Smart Grid with Code Signal and Bargaining Game

Takashi IRITA*, Takumi SHINOHARA* and Toru NAMERIKAWA*,**

In this paper, we study detection problem of replay attacks on smart grid. A smart grid is operated as if it is a Cyber-Physical System with solid links between its calculation and its material constituents, hence ensuring cyber security has been an important role. The goal of this paper is to detect replay attack that is one of cyber attacks on the sensors of a control system. We propose a detecting method adding intentional noise to not only sensors as code signal but also input. Replay attack can immediately be reflected by using fault diagnosis matrices that are composed of the estimator and observed values even if code signal is decrypted. Finally, we show simulation results to analyze effectiveness of the proposed method.

Key Words: smart grid security, supervisory control and data acquisition (SCADA), Kalman filter, replay attack, cyber attack diagnosis

1. はじめに

近年、電力の安定供給の観点から、スマートグリッドが注目されている。情報通信技術を用いて電力需給を制御し、エネルギーの効率的な利用支援や再生可能エネルギー電源の利用促進を行なうため、高度かつ複雑な巨大ネットワークシステムで構成されている¹⁾。スマートグリッドは計算と物理要素間に強固な連携をもつ CPS (Cyber-Physical System) として扱われる。そして、物理的環境に作用するアクチュエータと分散された埋め込みセンサを用いて、SCADA (Supervisory Control and Data Acquisition) が監視かつ制御を行なう²⁾。そのため、情報通信網に対してコンピュータやネットワークによるサイバー攻撃を受ける危険性があり³⁾、物理システムへの影響の相関関係およびサイバー攻撃における脅威からリスクを軽減する対策が検討されている⁴⁾。電力ネットワークに影響を及ぼすようなサイバー攻撃を受けた場合、社会基盤であるエネルギーインフラに支障をきたす恐れがあるため、サイバーセキュリティ向上に関する技術開発は危急の課題である。そのため、一部のサイバー攻撃に焦点を当てた研究は現在までもさまざまなものが扱われてきた。たとえば、DoS

(Denial of Service) 攻撃⁵⁾、Covert 攻撃⁶⁾、データ改竄の攻撃⁷⁾などが挙げられる。同様に、リプレイ攻撃に焦点を当てた研究も数多く存在し、完全性と認証に影響を及ぼす攻撃と危険視されている⁸⁾。

リプレイ攻撃とはワイヤレスネットワーク間における情報を攻撃者が窃盗し、窃盗した情報を用いて正しい情報を改竄する攻撃である。過去のデータを用いているため、監視者が感覚的に攻撃されていることに気づくことが非常に難しいという特徴をもち、時間遅延によって、周波数が発散して大停電が起きる危険性がある。また、攻撃者にとって、システムを故意に妨害する上でリプレイ攻撃は単純な攻撃であり、さまざまな過程を崩壊させるには十分な攻撃である⁹⁾。実際に、イランに所在するウラン濃縮用遠心分離機は、リプレイ攻撃を含むウイルスである Stuxnet の標的になり、遠心分離機を制御する SCADA システムに到達した。その結果、遠心分離機の制御システムを変更され、約 1000 台の遠心分離機に損傷を受けた¹⁰⁾。このように、日常生活に欠かせないインフラにおいても、サイバーセキュリティの観点から、リプレイ攻撃の早期検出は重要な課題である。

リプレイ攻撃の検出に関する研究は多く存在する。制御入力に意図的な雑音を加えることでリプレイ攻撃を反映させる手法^{11), 12)}が提案されているが、システムが正常であっても、制御性能が悪化する問題点がある。そのため、制御入力に雑音を加えない手法が提案されてきた。文献 13) では、制御入力に対して意図的な雑音を付加する間隔を静的に定める手法が述べられている。一方、文献 14) では、ゲーム理論によって制御入力に雑音を加えるタイミングを動的に定める手法が

* 慶應義塾大学大学院理工学研究科 横浜市港北区日吉 3-14-1

** 慶應義塾大学理工学部 横浜市港北区日吉 3-14-1

* Graduate School of Science and Engineering, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama

** Faculty of Science and Engineering, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama

(Received April 2, 2016)

(Revised June 9, 2016)

提案されている．しかし，前者と同様に最適な制御入力と比較して，常に制御性能は劣っている．また，検出率が低いことも問題点として，挙げられる．これらを踏まえ，文献 15) はセンサに意図的な雑音をコード信号として加え，監視者でコード信号を減算処理するシステムを導入し，制御性能を維持しつつ，リプレイ攻撃を確実に検出する手法を提案した．しかし，コード信号を攻撃者に解読された場合，リプレイ攻撃を検出できない脆弱性が存在する．

以上より，本稿では，コード信号に制御性能を犠牲にしたリプレイ攻撃の検出法を導入し，攻撃を早期検出するシステムを提案する．具体的には，二人のプレイヤーが互いに協力して合意に至る交渉ゲームにより，制御性能と検出精度を考慮した制御入力の雑音をオンラインで設計する．そして，観測値および状態推定値を用いて故障評価行列を定義し，リプレイ攻撃を検出する．最後に，数値シミュレーションによって，本提案手法における有効性を検証する．

2. 問題設定

本章では，電力ネットワークのモデル化，リプレイ攻撃および故障検出フィルタの定義，そして故障問題について説明する．

2.1 電力ネットワークモデル

電力ネットワークは無数の発電機，変電所，需要家などの動的負荷および送電線によって形成される．本稿では，Fig. 1 のような複数の発電機ノードと母線から構成される IEEE 118 bus system を包含する電力ネットワークを対象とする．

Fig. 1 Partition of IEEE 118 bus system

Fig. 1 の電力ネットワークにおいて，以下のような仮定 1 を設ける．

[仮定 1]

- (1) 超高压送電により送電損失零．
- (2) 隣接するノード間の相差点の偏差は十分小さい．
- (3) 通信遅延はサンプル時間 T_s 以内とする．

負荷は一般のモーター負荷を想定しており，母線を含んだすべてのノードが動揺方程式によって支配され，ノード i の動揺方程式は次式のように表わすことができる^{16), 17)}．ただし， $i \in \{1, 2, \dots, n\}$ であり， $\mathcal{N}_i$ はノード i の近傍集合を表わしている．

$$m_i \ddot{\delta}_i + d_i \dot{\delta}_i = P_{mi} - \sum_{j \in \mathcal{N}_i} |V_i| |V_j| B_{ij} \sin(\delta_i - \delta_j) \quad (1)$$

上式において，下付きの文字 i はノード i を表わしており， $m_i > 0$ は慣性定数， δ_i は相差点， d_i は減衰定数， P_{mi} は機械入力， $|V_i|$ は電圧， B_{ij} はノード i, j 間のサセプタンスである．また，仮定 1 より， $\sin(\delta_i - \delta_j) \approx \delta_i - \delta_j$ が成り立つ．よって，(1) 式より，システムの離散時間状態方程式は以下のように表わされる¹⁸⁾．

$$x_{k+1} = Ax_k + Bu_k + g_k \quad (2)$$

$$y_k = Cx_k + h_k \quad (3)$$

ただし，状態 $x \in \mathbb{R}^{2n}$ ，観測値 $y \in \mathbb{R}^{2n}$ ，制御入力 $u \in \mathbb{R}^n$ ， $A \in \mathbb{R}^{2n \times 2n}$ ， $B \in \mathbb{R}^{2n \times n}$ はノード $i \in \{1, \dots, n\}$ を用いて，以下のように与えられる．また， $C \in \mathbb{R}^{2n \times 2n}$ は単位行列である．

$$x = \begin{bmatrix} x_1^T & \dots & x_n^T \end{bmatrix}^T, \quad x_i = \begin{bmatrix} \delta_i & \omega_i \end{bmatrix}^T \quad (4)$$

$$y = \begin{bmatrix} y_1^T & \dots & y_n^T \end{bmatrix}^T \quad (5)$$

$$u = \begin{bmatrix} P_{m1} & \dots & P_{mn} \end{bmatrix}^T \quad (6)$$

$$A = M^{-1} \left(D - \mathcal{W}_{(s,t)} \otimes \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right) \quad (7)$$

$$M = \begin{bmatrix} \mathcal{M}_1 & & \\ & \ddots & \\ & & \mathcal{M}_n \end{bmatrix}, \quad \mathcal{M}_i = \begin{bmatrix} 1 & 0 \\ 0 & m_i \end{bmatrix} \quad (8)$$

$$D = \begin{bmatrix} \mathcal{D}_1 & & \\ & \ddots & \\ & & \mathcal{D}_n \end{bmatrix}, \quad \mathcal{D}_i = \begin{bmatrix} 0 & 1 \\ 0 & -d_i \end{bmatrix} \quad (9)$$

$$\mathcal{W}_{(s,t)} = \begin{cases} \sum_{t \in \mathcal{N}_s} |V_s| |V_t| B_{st} & t = s \\ -|V_s| |V_t| B_{st} & t \in \mathcal{N}_s \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

$$B = M^{-1} \mathcal{B} \quad (11)$$

$$\mathcal{B} = \text{diag}[b, \dots, b], \quad b = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (12)$$

ただし， $\mathcal{W}_{(s,t)}$ における $s, t \in \{1, 2, \dots, n\}$ は行列の行番号および列番号， diag は対角行列を表わしている．また，

$g_k, h_k \in \mathbb{R}^{2n}$, はプロセス雑音および観測雑音を表わし, 正規分布に従う共分散行列 $G \geq 0, H > 0$, 平均 0 の白色雑音過程である.

時刻 k におけるカルマンフィルタの推定式を以下に示す. ただし, 時刻 $k-1$ において, 事前推定した状態および誤差共分散行列をそれぞれ $\hat{x}_{k|k-1}, P_{k|k-1}$ とし, カルマンゲインを K_k とする.

カルマンゲイン更新

$$K_k = P_{k|k-1} C^T (C P_{k|k-1} C^T + H)^{-1} \quad (13)$$

状態推定値更新

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k (y_k - C \hat{x}_{k|k-1}) \quad (14)$$

$$\hat{x}_{k+1|k} = A \hat{x}_{k|k} + B u_k \quad (15)$$

誤差共分散行列更新

$$P_{k|k} = P_{k|k-1} - K_k C P_{k|k-1} \quad (16)$$

$$P_{k+1|k} = A P_{k|k} A^T + G \quad (17)$$

ここで, システムに対して以下の仮定をおく.

[仮定 2]

- (1) $(A, G^{\frac{1}{2}})$ は可到達
- (2) (C, A) は可検出

システムが可検出であるとき, カルマンゲインは一定値に収束し, $P_{k|k-1} = P_{k+1|k} = P$ を満足する安定化解 P が存在する. ここで, 定常値となったカルマンゲイン K および誤差共分散行列 P を以下のように定義する.

$$P \triangleq \lim_{k \rightarrow \infty} P_{k|k-1}, \quad K = P C^T (C P C^T + H)^{-1} \quad (18)$$

続いて, 制御入力について説明する. 電力ネットワークにおけるシステムの安定性および制御性能面において, 広域制御は重要である¹⁹⁾. 本稿では, 制御方式として, LQG (Linear-Quadratic-Gaussian) コントローラを用いて広域制御を行なう. LQG コントローラは制御対象システムの出力に対してカルマンフィルタを用いて状態を推定し, 推定値を用いて状態フィードバック制御を行なう制御系である. 時刻 k において, システムに与える制御入力を次式に示す.

$$u_k = L \hat{x}_{k|k} \quad (19)$$

ただし, L は LQG 制御問題により事前に与えられた固定フィードバックゲインであり, 以下のように求められる. まず, 制御を行なうレギュレータに関する評価関数 J を次式のように表わす.

$$J = \sum_{k=0}^{\infty} [x^T(k) Q x(k) + u^T(k) R u(k)] \quad (20)$$

ただし, $Q \in \mathbb{R}^{2n \times 2n}$ は半正定行列, $R \in \mathbb{R}^{n \times n}$ は正定行列である. システム (2), (3) に関して可制御性が成り立つとき, つぎの離散時間リカッチ方程式を満たす正定解 S が一意に決まる.

$$S = A^T S A + Q - A^T S B (B^T S B + R)^{-1} B^T S A \quad (21)$$

このとき, 評価関数 J を最小にする L は以下のように与えられる.

$$L = -(B^T S B)^{-1} B^T S A \quad (22)$$

2.2 故障評価行列¹⁸⁾

故障評価行列は外部から与えられた異常信号の有無や大小による影響を自身の要素に反映できる故障検出フィルタである. 故障評価行列 $\mathcal{M}_k$ を以下のように定義する.

$$\mathcal{M}_k \triangleq S_k - (C P C^T + H) \quad (23)$$

ただし, S_k は観測誤差共分散行列, V は観測雑音の共分散行列である. 正常なシステムにおける故障評価行列は以下のようになる.

$$\begin{aligned} \mathcal{M}_k &= S_k - C P_{k|k-1} C^T - H \\ &= \text{cov}(y_k - C \hat{x}_{k|k-1}) - C P_{k|k-1} C^T - H \\ &= C P_{k|k-1} C^T + H - (C P_{k|k-1} C^T + H) \\ &= 0 \end{aligned} \quad (24)$$

(24) 式のように, 故障評価行列はシステムが正常のときは反応しない. 一方, 観測方程式 (3) に外部から与えられた異常信号 $F_k d_k$ が含まれたとき, 故障評価行列は以下のようになる. ただし, $F_k \in \mathbb{R}^{2n \times 2n}$, $d_k \in \mathbb{R}^{2n}$ である. また, $\zeta_k \triangleq x_k - \hat{x}_{k|k-1}$, $D_k = E[\{d_k - E(d_k)\} \{d_k - E(d_k)\}^T]$ と定義する. $E[X]$ は確率変数 X における期待値である.

$$\begin{aligned} \mathcal{M}_k &= S_k - C P_{k|k-1} C^T - H \\ &= F_k D_k F_k^T + E \left[C \eta_k \left(d_k^T F_k^T - E \left[d_k^T F_k^T \right] \right) \right. \\ &\quad \left. + E \left[C \zeta_k \left(d_k^T F_k^T - E \left[d_k^T F_k^T \right] \right) \right] \right. \\ &\quad \left. - E \left[C \zeta_k \right] \left(d_k^T F_k^T - E \left[d_k^T F_k^T \right] \right) \right] \\ &\quad + E \left[d_k F_k \left(\zeta_k^T C^T - E \left[\zeta_k^T C^T \right] \right) \right. \\ &\quad \left. - E \left[d_k F_k \right] \left(\zeta_k^T C^T - E \left[\zeta_k^T C^T \right] \right) \right] \end{aligned} \quad (25)$$

(25) 式のように, 異常信号 $F_k d_k$ が含まれるとき, 故障評価行列は異常値を反映する.

2.3 リプレイ攻撃

本稿では, サイバー攻撃の中でもリプレイ攻撃を扱う. リプレイ攻撃は送信された情報を窃盗し, 窃盗した情報を用いて改竄する攻撃である. 悪意ある攻撃者に対して, 以下のようない仮定を設ける.

[仮定 3] 攻撃者に関する仮定

- (1) システムに影響を及ぼすことなく, 攻撃者は観測値を窃盗できる.
- (2) 攻撃者は膨大な観測値のデータを確保でき, 改竄できる.

この仮定より、悪意ある攻撃者は監視者に気付かれることなくシステムに改竄値を挿入できる。リプレイ攻撃における改竄された観測値 y'_k を以下のように定義する。

$$y'_k = y_{k-\alpha} \quad (26)$$

ただし、 α は遅延時間である。(26) 式は時刻 α 前の観測値 $y_{k-\alpha}$ を用いて y_k を改竄することを意味する。

また、通常動作している際に y'_k は以下の仮想システムにおける出力として考えられる。

$$x'_{k+1} = Ax'_k + Bu'_k + g'_k \quad (27)$$

$$y'_k = Cx'_k + h'_k \quad (28)$$

$$\hat{x}'_{k|k} = \hat{x}'_{k|k-1} + K_k(y'_k - C\hat{x}'_{k|k-1}) \quad (29)$$

$$u'_k = L\hat{x}'_{k|k} \quad (30)$$

ただし、 $x'_k = x_{k-\alpha}$ であり、上付き'における時刻は k から α 遅れたものとなっている。つまり、(27)~(30) 式で表現された仮想システムは実際のシステムに対して時刻 α シフトしたモデルとなっている。

2.4 検出問題

前述にモデル化された電力ネットワーク (2), (3) 式、およびリプレイ攻撃 (26) 式に対して以下の仮定を設ける。

[仮定 4] 雑音、異常信号に関する仮定

- (1) $E[v_k v_s^T] = E[w_k w_s^T] = E[\zeta_k \zeta_s^T] = 0 \quad (k \neq s)$
- (2) $E[v_k w_s^T] = E[v_k \zeta_s^T] = E[\zeta_k w_s^T] = 0$
- (3) $E[x_0 v_k^T] = E[x_0 w_k^T] = E[x_0 \zeta_k^T] = 0$

[仮定 5] $\mathcal{A} \triangleq (A + BL)(I - KC)$ が安定である。

仮定 4 は雑音に関する仮定であり、本稿においてはカルマンフィルタを用いるために必要な仮定である。また、 $\mathcal{A}$ が不安定であるとき文献 [12] によりリプレイ攻撃を検出できることが証明されている。しかしながら、ゲイン K, L はシステムが安定となるように算出されるため、本稿ではこのゲインを変更することなく ($\mathcal{A}$ を不安定とすることなく) リプレイ攻撃を検出することを考える。

また、時刻 $k-1$ にリプレイ攻撃を受けた故障評価行列において以下の命題が成り立つ。

[命題 1] 仮定 1-5 が成り立ち、時刻 $k-1$ からリプレイ攻撃を受け続けた故障評価行列 $\mathcal{M}_k$ において、以下のような関係式が成立する。

$$\lim_{k \rightarrow \infty} \mathcal{M}_k = 0 \quad (31)$$

[Proof] 文献 [15] 参照。□

(31) 式より、システムがリプレイ攻撃を受けたとしても故障評価行列 $\mathcal{M}_k$ に異常値が反映されないことから、早急にリプレイ攻撃を特定できないことがわかる。これを踏まえ、本稿で考えるリプレイ攻撃の検出問題を以下のように定義する。

[問題 1] 仮定 1-5 が成り立ち、電力ネットワークモデルが (2), (3) 式で与えられ、時刻 k において、あるノードに (26)

式のようなリプレイ攻撃を受けている場合、攻撃を受けているノード i を特定せよ。

3. コード信号によるリプレイ攻撃の検出法¹⁵⁾

本章ではコード信号を用いたリプレイ攻撃を検出する手法を説明する。ここでは、タイムスタンプの役割として、センサに意図的な白色雑音を加え、監視者が受信された観測値に含む雑音を打ち消すシステムを構築する。時刻ごとに加える意図的な白色雑音をコード信号と名付け、各時刻における観測値の暗号化を意味する。コード信号において、以下のような仮定を設ける。

[仮定 6] コード信号における仮定

- (1) センサおよび監視者は事前にコード信号を共有できる。
- (2) 監視者はコード信号を処理した観測値を用いる。

センサから得られる出力にコード信号を付加した観測値 $Y_k \in \mathbb{R}^{2n}$ を以下のように定義する。

$$Y_k = y_k + \tau_k \quad (32)$$

ただし、 $\tau_k \in \mathbb{R}^{2n}$ は正規分布に従う共分散行列 $\mathcal{T} > 0$ 、平均 0 の白色雑音過程とする。白色雑音をコード信号として用いる理由は、攻撃者にコード信号を感知されないよう、ステップ応答や正弦波のような規則性のある値を避けるためである。情報通信網から得た観測値を監視者がコード信号を処理して得られる観測値 $y_k^{kf} \in \mathbb{R}^{2n}$ を以下のように定義する。

$$\begin{aligned} y_k^{kf} &= Y_k - \tau_k \\ &= y_k + \tau_k - \tau_k \\ &= y_k \end{aligned} \quad (33)$$

Fig. 2 にコード信号 τ_k を含んだリプレイ攻撃の簡易図を示す。ただし、Supervisor はセンサから得られる情報から適切な制御入力を指令する監視者である。

Fig. 2 Replay attack

コード信号処理を含むシステムに対して時刻 $k-1$ にリプレイ攻撃を与えたとき、故障評価行列において、以下の命題が成り立つ。

[命題 2] 仮定 1-6 が成り立つとする。コード信号処理を含むシステムに対して時刻 $k-1$ からリプレイ攻撃を受け続けた故障評価行列 $\mathcal{M}_k \in \mathbb{R}^{2n \times 2n}$ において、時刻 $k + \alpha - 1$ に至るまで、以下のような関係式が成立する。

$$\lim_{k \rightarrow \infty} \mathcal{M}_k = 2\mathcal{T} \quad (34)$$

[Proof] 文献 15) 参照. $\square$

命題 2 より, コード信号によって, リプレイ攻撃を検出できることがわかる. しかしながら, コード信号 τ_k が攻撃者によって改竄された場合, コード信号を含めた観測値 Y_k が攻撃者によって改竄される可能性がある. その場合, この手法ではリプレイ攻撃を検出することができない.

4. 雑音を含んだ制御性能によるリプレイ攻撃の検出

本章では, 事前に共分散行列を設定した白色雑音を含む制御入力によって, リプレイ攻撃を検出する. そして, 故障評価行列を用いて, リプレイ攻撃の解析をする.

4.1 白色雑音を含んだ制御入力

制御入力に以下のような雑音を意図的に付加する.

$$u_k = L\hat{x}_{k|k} + \Delta u_k \quad (35)$$

ただし, システムにバイアスを付加しないよう, $\Delta u_k \triangleq [\Delta u_k^1 \ \cdots \ \Delta u_k^n]^T \in \mathbb{R}^n$ は正規分布に従う共分散行列 $U \geq 0$, 平均 0 の白色雑音過程である.

4.2 リプレイ攻撃を受けた事前推定式の式変形

故障評価行列によるリプレイ攻撃の解析準備として, カルマンフィルタで得られる事前推定値 (15) を式変形する.

時刻 $k-1$ において, (26) 式のようなリプレイ攻撃を受けたとき, (15) 式における事前推定値 $\hat{x}_{k|k-1}$ は以下のように表現できる.

$$\begin{aligned} \hat{x}_{k+1|k} &= A\hat{x}_{k|k} + Bu_k \\ &= A\hat{x}_{k|k} + B(L\hat{x}_{k|k} + \Delta u_k) \\ &= (A + BL)\hat{x}_{k|k} + B\Delta u_k \\ &= (A + BL)\{\hat{x}_{k|k-1} + K(y'_k - C\hat{x}_{k|k-1})\} + B\Delta u_k \\ &= \mathcal{A}\hat{x}_{k|k-1} + (A + BL)Ky'_k + B\Delta u_k \end{aligned} \quad (36)$$

(27)~(30) 式で表現された仮想システムにおいて, 事前推定値 $\hat{x}'_{k|k-1}$ も (36) 式と同様に書き換えることができる.

$$\hat{x}'_{k|k-1} = \mathcal{A}\hat{x}'_{k-1|k-2} + (A + BL)Ky'_{k-1} + B\Delta u'_k \quad (37)$$

$\eta_k \triangleq \hat{x}_{k|k-1} - \hat{x}'_{k|k-1}$, $\mathcal{U}_k \triangleq \Delta u_k - \Delta u'_k$ と定義したとき, (36), (37) 式より, η_k は以下のように表現できる.

$$\begin{aligned} \hat{x}_{k+1|k} - \hat{x}'_{k+1|k} &= \mathcal{A}(\hat{x}_{k|k-1} - \hat{x}'_{k|k-1}) + B(\Delta u_k - \Delta u'_k) \\ \eta_{k+1} &= \mathcal{A}\eta_k + B\mathcal{U}_k \\ &= \mathcal{A}\{\mathcal{A}\eta_{k-1} + B\mathcal{U}_{k-1}\} + B\mathcal{U}_k \\ &= \mathcal{A}^2\eta_{k-1} + B\mathcal{U}_k + \mathcal{A}B\mathcal{U}_{k-1} \\ &\vdots \\ &= \mathcal{A}^{k-\alpha+1}\eta_{k-\alpha} + \sum_{l=0}^{k-\alpha} \mathcal{A}^l B\mathcal{U}_{k-l} \end{aligned} \quad (38)$$

(38) 式より, 事前推定値 $\hat{x}'_{k+1|k}$ は以下のように表わせる.

$$\hat{x}_{k+1|k} = \hat{x}'_{k+1|k} + \mathcal{A}^{k-\alpha+1}\eta_{k-\alpha} + \sum_{l=0}^{k-\alpha} \mathcal{A}^l B\mathcal{U}_{k-l} \quad (39)$$

ただし, リプレイ攻撃を受けた事前推定値の式変形は時刻 $k \sim k + \alpha - 1$ 間で成立する.

4.3 リプレイ攻撃を受けた故障評価行列の特性

時刻 $k-1$ にリプレイ攻撃を受けた故障評価行列において, 以下の命題が成り立つ.

[命題 3] 仮定 1-6 が成り立つとする. 時刻 $k-1$ からリプレイ攻撃を受け続けた故障評価行列 $\mathcal{M} \in \mathbb{R}^{2n \times 2n}$ において, 時刻 $k + \alpha - 1$ に至るまで以下のような関係式が成立する.

$$\lim_{k \rightarrow \infty} \mathcal{M}_k \geq 2CBUB^T C^T \quad (40)$$

[Proof] 時刻 $k-1$ において, (26) 式のようなリプレイ攻撃を受けた場合, 観測誤差共分散行列 S_k の中身は以下ようになる. ただし, $\eta'_k \triangleq x'_k - C\hat{x}'_{k|k-1}$ と定義する.

$$\begin{aligned} S_k &= \text{cov}(y'_k - C\hat{x}_{k|k-1}) \\ &= \text{cov}(C\eta'_k + h'_k - C\mathcal{A}^{k-\alpha}\eta_\alpha - C \sum_{l=0}^{k-1-\alpha} \mathcal{A}^l B\mathcal{U}_{k-1-l}) \end{aligned} \quad (41)$$

観測雑音の雑音および制御入力に含む雑音は仮定 4 を満たす平均 0 の白色雑音過程であるため, (41) 式は以下のように展開できる.

$$\begin{aligned} S_k &= CP'_{k|k-1}C + H + 2C \sum_{l=0}^{k-1-\alpha} (\mathcal{A}^l BUB^T \mathcal{A}^{lT}) C^T \\ &\quad + C\mathcal{A}^{k-\alpha} E \left[(\eta_\alpha - E[\eta_\alpha])(\eta_\alpha - E[\eta_\alpha])^T \right] \mathcal{A}^{k-\alpha T} C^T \\ &\quad - C\mathcal{A}^{k-\alpha} E \left[(\eta_\alpha - E[\eta_\alpha])(\eta'_k - E[\eta'_k])^T \right] C^T \\ &\quad - CE \left[(\eta'_k - E[\eta'_k])(\eta_\alpha - E[\eta_\alpha])^T \right] \mathcal{A}^{k-\alpha T} C^T \end{aligned} \quad (42)$$

(42) 式において, 仮定 5 が成り立つとき, 以下の式が成り立つ.

$$\lim_{k \rightarrow \infty} S_k = CPC + H + 2C \sum_{l=0}^{k-1-\alpha} (\mathcal{A}^l BUB^T \mathcal{A}^{lT}) C^T \quad (43)$$

よって, (23) 式の極限值は以下のようになる.

$$\lim_{k \rightarrow \infty} \mathcal{M}_k \geq 2CBUB^T C^T \quad (44)$$

$\square$

(44) 式より, 白色雑音を含む制御入力によって, リプレイ攻撃を検出できることがわかる. しかしながら, 制御入力に対して雑音を加えているため, (19) 式の最適な制御入力と比べて常時制御性能は悪化する.

5. 提案手法

本章では, コード信号と交渉ゲームを組み合わせたリプレイ

イ攻撃の検出法を提案する．交渉ゲームにより，制御性能と検出精度の両者を満たす制御入力に雑音をオンラインで設計する．攻撃者が，コード信号を含んだ観測値 Y_k をリプレイ攻撃に用いるためには，コード信号を解読する必要があるが，コード信号が解読されている場合，(34) 式ではリプレイ攻撃を検出できない．よって，制御入力に雑音を大きくし，検出性能を向上する必要がある．対して，コード信号が解読されていない場合，(34) 式の手法でリプレイ攻撃を検出可能であるため，制御性能を向上するために，雑音挿入を最小限に留める手法を提案する．

5.1 交渉ゲーム

交渉ゲームとは，二人のプレイヤー間が互いに協力して合意に至るゲームである²⁰⁾．しかし，協力によってどのような状態を実現するかについて利害が一致するとは限らない．ここに，プレイヤー間で「交渉問題」が生じ，協力の結果として，どのような状態を実現すべきかを考える．

二人のプレイヤーが協力によって実現が可能な領域 $\mathbb{C}$ における利得ベクトルを $\mathcal{C} = (\mathcal{C}_1, \mathcal{C}_2) \in \mathbb{C}$ とし，交渉決裂点（二人の交渉が決裂した場合に実現される点）を $r = (r_1, r_2)$ としたとき，以下の式で定義される交渉問題の解を，ナッシュ交渉解と呼ぶ．

$$\max_{\substack{\mathcal{C} \in \mathbb{C} \\ \mathcal{C}_1 \geq r_1, \mathcal{C}_2 \geq r_2}} (\mathcal{C}_1 - r_1)(\mathcal{C}_2 - r_2) \quad (45)$$

ナッシュ交渉解は二人のプレイヤーの交渉領域上で，プレイヤーの効用差を最大化する解となる．

5.2 利得関数の設計

各ノード i における制御性能および検出性能の利得関数を以下のように定義する．ただし，上付き文字 i はノード i を意味し，制御性能の利得関数を $\mathcal{C}_u^i$ ，検出性能の利得関数を $\mathcal{C}_d^i$ とする．また， $B_{(m,n)}$ は (11) 式の B 行列における (m, n) 要素， $\hat{x}_{k|k}^{i\omega}$ はノード i における角周波数 ω_i の事後推定値， $N, \beta, \gamma > 0$ は設計パラメータである．

$$\mathcal{C}_u^i(\Delta u_k^i) = N - (B_{(2i,i)} \Delta u_k^i)^2 - \beta(\hat{x}_{k|k}^{i\omega})^2 \quad (46)$$

$$\mathcal{C}_d^i(\Delta u_k^i) = \text{trace} \left(2CB^i (\Delta u_k^i)^2 (B^i)^T C^T \right) - \gamma \text{trace} \mathcal{M}_k^i \quad (47)$$

$$\text{s.t. } \mathcal{C}_u^i \geq 0$$

制御性能の利得関数 (46) は発電機の周波数を制御性能の指標としている．(3), (35) 式より，制御入力に加える雑音がノード i の周波数に及ぼす項は $B_{(2i,i)} \Delta u_k^i$ なので， Δu_k^i が大きいほど，制御性能の利得関数が減少するよう，(46) 式を定義した．また，事後推定値より，周波数変動が大きい場合，雑音によって制御性能を悪化させないよう， $-\beta(\hat{x}_{k|k}^{i\omega})^2$ を項に含めている．続いて，検出精度の利得関数 (47) について説明する．命題 2 より，制御入力に加える雑音が大きければ大きいほど検出性能が上昇するので， $\text{trace} \left(2CB^i (\Delta u_k^i)^2 (B^i)^T C^T \right)$ を項に加えている．ただし，故障評価行列の中でも，各ノードの影

響が最も顕著に表われる対角要素を扱う．また， $-\gamma \text{trace} \mathcal{M}_k^i$ は，少しでも異常値を検出すると，より大きな雑音を挿入して検出性能を高めることを意味する．

5.3 交渉ゲームとコード信号における検出アルゴリズム

本論文の問題設定において，制御性能と検出性能は別々のプレイヤーだが，管理しているシステムは 1 つである．そのため，交渉決裂点 (r_1, r_2) を $(r_1, r_2) = (0, 0)$ と仮定し，ナッシュ交渉解を求める．また， $\Delta u_k^T \Delta u_k$ は 1 次元なので，(46), (47) 式の変数を $\mathcal{U}_k = \Delta u_k^T \Delta u_k$ とおく．交渉ゲーム理論を用いた制御入力に雑音設計アルゴリズムを以下に示す．なお，このアルゴリズムは各ノードが独立して行なう．

Algorithm ナッシュ交渉解による制御入力に雑音設計

Step 1: 制御性能および検出性能の利得関数の定義

$$\mathcal{C}_u^i(\Delta u_k^i) = N - (B_{(2i,i)} \Delta u_k^i)^2 - \beta(\hat{x}_{k|k}^{i\omega})^2$$

$$\mathcal{C}_d^i(\Delta u_k^i) = \text{trace} \left(2CB^i (\Delta u_k^i)^2 (B^i)^T C^T \right) - \gamma \text{trace} \mathcal{M}_k^i$$

Step 2: ナッシュ交渉解の算出

$$\max_{\mathcal{U}_k} (\mathcal{C}_u^i - r_1)(\mathcal{C}_d^i - r_2) \quad \text{s.t. } 0 \leq \mathcal{U}_k \leq \frac{N}{B_{(2i,i)}^2}$$

Step 3: 雑音の決定

$$\Delta u_k^* = \pm \sqrt{\mathcal{U}_k^*}$$

ただし，符号は直近で挿入した雑音の符号を反転させたものを扱う．

Step 4: リプレイ攻撃の検出

$\zeta \leq \text{trace} \mathcal{M}_k^i$ を満たすとき， $k_{RA} \leq k \leq k_{RA} + N_{\text{update}} - 1$ において， $\mathcal{U}_k^* = 0$ とする．ただし， k_{RA} は $\zeta \leq \text{trace} \mathcal{M}_k^i$ を初めて満たす時刻である．

Step 5: 時刻 $k_{RA} + N_{\text{update}}$ 以降，コード信号の解読診断 if:

$k_{RA} \leq k \leq k_{RA} + N_{\text{update}} - 1$ において， $\zeta \geq \text{trace} \mathcal{M}_k^i$ を満たす k が 1 つ以上存在する．

コード信号によって検出できることから，共分散行列 $\mathcal{U}^i = 0$ ．

else:

共分散行列 $\mathcal{U}^i = \frac{N}{B_{(2i,i)}^2}$ を満たす白色雑音を挿入し続ける．

コード信号を解読されたかどうか診断するため，リプレイ攻撃検出後， N_{update} の間，制御入力に雑音を加えず，コード信号のみでリプレイ攻撃を検出できるかどうか検証する．仮に検出できた場合，命題 2 よりコード信号のみで検出できていることを意味する．一方で，検出できなかった場合，コード信号が解読されていることを意味しているので，制御入力に雑音を挿入し続ける．制御入力に雑音を挿入し続けることで，命題 3 よりリプレイ攻撃を検出できる．このアルゴリズムにより，交渉ゲームとコード信号における利点を組み合わせ

せたりプレイ攻撃の検出を達成することができる。

6. シミュレーション検証

本稿では、Fig.1の電力ネットワークを構成する1つのエリアを対象としシミュレーション検証を行なう。ここでは、文献17),21)で考えられている9-bus電力ネットワークを対象とし、提案手法の有効性を確認する。本シミュレーションで考える9-busシステムをFig.3に、これをグラフ表現したものをFig.4に示す。Fig.4より、たとえばノード9の近傍集合 $\mathcal{N}_9$ は $\mathcal{N}_9 = \{4,8\}$ である。

Fig. 3 Power network with 9 buses

Fig. 4 Graph representation of Fig. 3

シミュレーションに用いたパラメータをTable 1に示す。ただし、定格容量は100 MW (= 1.0 pu)である。

Table 1 Parameter

Parameter	Symbol	value
Node	n	9
Step	k [step]	3000
Sampling Time	T_s [s]	0.1
Inertia Constant	m_i [pu MW · s/Hz]	1.0-7.0
Damping Constant	d_i [pu MW/Hz]	1.0-5.0
Susceptance	B_{ij} [s]	0.1-0.5
Covariance of Noisy Input	U	1.0×10^{-3}
Max of C_u^i	N	0.015
Weight	β	120
Weight	γ	5000
Threshold	ζ	2.0×10^{-3}
Diagnosis Time	N_{update} [step]	10

また、本シミュレーションでは1000stepにノード1に対してリプレイ攻撃を行なう。攻撃の遅延時間は $\alpha = 50$ stepである。

6.1 Case1: コード信号が解読された場合

コード信号を攻撃者に解読された状況を想定する。本シミュレーションでは、(19)式で示される最適な制御入力を用いた手法、加える白色雑音の共分散を事前設計した制御入力¹²⁾を

Fig. 5 Fault diagnosis signal of optimal input

Fig. 6 Fault diagnosis signal of noisy input

用いた手法、および提案手法の3つを比較する。まず、故障評価行列の対角和をFig.5, Fig.6を示す。Fig.5より、最適な制御入力の場合はリプレイ攻撃を受けても異常値を示さないことがわかる。一方、Fig.6より、提案手法および従来手法における制御入力はリプレイ攻撃後、異常値を示していることがわかる。ただし、検出信号の振れ幅は提案手法のほうが大きいことから、リプレイ攻撃の異常値に応じて、制御入力に挿入する雑音を決定していることが確認できる。

ノード1における周波数のRMS値をFig.7に示す。ここで、周波数のRMS値 f_{RMS} とは、基準周波数からの偏差を表わす値であり、次式で定義される。

Fig. 7 Frequency RMS

$$f_{RMS} = \sqrt{\frac{1}{N_{RMS}} \sum_{j=1}^{N_{RMS}} \Delta f_j^2} \quad (48)$$

ただし, $N_{RMS} = 1000$ とする.

Fig. 7 より, 提案手法は従来手法と, 周波数変動がほぼ変わらないことから, 提案手法が従来研究に比べ, 制御性能を悪化させることなく検出精度が上回っていることを確認できた. このことから, 従来研究の場合, 提案手法と同等の検出精度を保つためには, さらに制御入力を悪化しなくてはならないことが確認できる. また, 最適制御と比較した際, 制御性能は劣っているものの, Fig. 5, Fig. 6 より, リプレイ攻撃の検出精度は優れていることがわかる.

続いて, 制御入力に加えた雑音を Fig. 8 を示す. また, リプレイ攻撃が発生していない 1~1000 step における提案手法と従来手法の雑音の分散の比較を Table 2 に示す.

Fig. 8 Input noise

Table 2 Input noise variance

	Proposed	Conventional
Variance [$\times 10^{-4}$]	4.4552	5.0841

Table 2 より, リプレイ攻撃が発生していない正常なシステムにおいて, 提案手法は従来手法より共分散の小さい雑音を入れていることが確認できる. しかし, Fig. 8 より, リプレイ攻撃を受けた 100s 以降, 提案手法では検出精度を上げるために雑音を大きくしていることを確認できる. このことから, 制御性能および検出精度の両者を満たす雑音設計の有効性を確認できる. 複数ノードに攻撃が加えられた場合, 故障評価行列が攻撃を反映する要素が増えるため, 単一ノードに対する攻撃よりも検出は容易である.

6.2 Case2: コード信号が解読されない場合

コード信号を攻撃者に解読されない状況を想定する. ここでは, 提案手法におけるコード信号の解読の有無によって, 制御入力が最適になるように設計されていることを確認する. まず, 故障評価行列の対角和を Fig. 9 に示す. Fig. 9 より, どちらもリプレイ攻撃を検出できていることが確認できる. このことから, コード信号が解読の有無にかかわらず, リプ

Fig. 9 Fault diagnosis signal of proposition

Fig. 10 Input noise of proposition

レイ攻撃を検出できることが確認できる. 続いて, 制御入力に加える雑音を Fig. 10 に示す.

Fig. 10 より, コード信号で異常を検出した際, 制御入力に雑音を入れているため, コード信号における解読診断ができていないことが確認できる. よって, Case1 および Case2 より, 交渉ゲームとコード信号を組み合わせたリプレイ攻撃の検出法における有効性を確認できた.

7. おわりに

本稿では, スマートグリッドの脅威として考えられ得るサイバー攻撃の中でもリプレイ攻撃に焦点を当て, 早期に検出する方法を提案した. まず, コード信号および雑音を含んだ制御入力において, リプレイ攻撃の検出における命題を示した. そして, 交渉ゲームによって, 制御性能と検出精度の両者を満たす制御入力への雑音を設計する手法を提案した. また, コード信号と組み合わせ, 攻撃者に解読されているかどうか診断するアルゴリズムを提案した. 最後に, シミュレーションにて, 本提案法の有効性を検証した.

謝辞 本研究は, JST CREST の支援を受けたものである.

参考文献

- 1) 合田, 諸住: スマートグリッド教科書, 5/7, インプレス

パン (2011)

- 2) E.A. Lee: Cyber physical systems: Design challenges, *Proc. 11th IEEE ISORC*, 363/369 (2008)
- 3) 吉田 隆: スマートエネルギーネットワーク最前線, 254, NTS (2012)
- 4) S. Sridhar, A. Hahn and M. Govindarasu: Cyber-physical system security for the electric power grid, *Proc. of the IEEE*, **100**-1, 210/224 (2012)
- 5) C.D. Persis and P. Tesi: Input-to-State Stabilizing Control Under Denial-of-Service, *IEEE Transactions on Automatic Control*, **60**-11, 2930/2944 (2015)
- 6) C.D. Persis and P. Tesi: Covert Misappropriation of Networked Control Systems: Presenting a Feedback Structure, *IEEE Transactions on Automatic Control*, *IEEE Control Systems Magazine*, **35**-1, 82/92 (2015)
- 7) K.C. Sou, H. Sandberg and K.H. Johansson: Data Attack Isolation in Power Networks Using Secure Voltage Magnitude Measurements, *IEEE Transactions on Smart Grid*, **5**-1, 14/28 (2014)
- 8) B. Zubair A and A. Abdul-Raouf: An Analysis of Smart Grid Attacks and Countermeasures, *Journal of Communications*, **8**-8, 473/479 (2013)
- 9) E.D. Knapp and R. Samani: *Applied Cyber Security and the Smart Grid*, 76/77, Syngress (2013)
- 10) Y. Mo, W. Sean and B. Sinopoli: Physical Authentication of Control Systems: Designing Watermarked Control Inputs to Detect Counterfeit Sensor Outputs, *IEEE Control Systems Magazine*, **35**-1, 93/109 (2015)
- 11) Y. Mo and B. Sinopoli: Secure Control Against Replay Attack, *47th Annual Allerton Conference on Communication, Control, and Computing*, 911/918 (2011)
- 12) Y. Mo, R. Chabukswar and B. Sinopoli: Detecting Integrity Attacks on SCADA Systems, *IEEE Transactions on Control Systems Technology*, **22**-4, 1396/1407 (2014)
- 13) T. Tran, O. Shin and J. Lee: Detection of Replay Attacks in Smart Grid Systems, *Proc. of Computing, Management and Telecommunications*, 298/302 (2013)
- 14) F. Miao, M. Pajic and G.J. Pappas: Stochastic Game Approach for Replay Attack Detection, *Proc. of the 52nd IEEE Conference on Decision and Control*, 1854/1859 (2013)
- 15) 入田, 滑川: コード信号を含む出力を用いたスマートグリッドにおけるリプレイ攻撃の検出, 計測自動制御学会論文集, **51**-12, 822/828 (2015)
- 16) 長谷川, 大山, 三谷, 斎藤, 北: 電力系統工学, オーム社 (2002)
- 17) 早川朋久: 電力ネットワークにおける分散的故障・攻撃検知, 計測と制御, **51**-1, 69/74 (2012)
- 18) 藤田, 大川, 滑川: 状態故障評価行列を用いた電力ネットワークにおけるサイバー攻撃・故障の診断, 計測自動制御学会論文集, **49**-12, 1131/1138 (2013)
- 19) F. Dörfler, M.R. Jovanović, M. Chertkov and F. Bullo: Sparsity-Promoting Optimal Wide-Area Control of Power Networks, *IEEE Transactions on Power System*, **29**-5, 2281/2291 (2014)
- 20) 岡田 章: ゲーム理論, 有斐閣 (2011)
- 21) P.M. Anderson and R.G. Farmer: *Series Compensation of Power Systems*, 599, USA PBLSH! (1996)

[著者紹介]

入田 隆

2014 年慶應義塾大学理工学部システムデザイン工学科卒業。同年同大学大学院理工学研究科総合デザイン工学専攻に入学, 現在に至る。スマートグリッドにおけるサイバー攻撃の検出に関する研究に従事。

篠原 巧

2016 年慶應義塾大学理工学部システムデザイン工学科卒業。同年同大学大学院理工学研究科総合デザイン工学専攻に入学, 現在に至る。スマートグリッドにおけるサイバー攻撃の検出に関する研究に従事。

滑川 徹 (正会員)

1994 年金沢大学大学院自然科学研究科システム科学専攻博士課程中退。同年金沢大学工学部電気・情報工学科助手。同講師を経て 2002 年長岡技術科学大学機械系助教授。2006 年金沢大学大学院自然科学研究科電子情報科学専攻助教授を経て 2007 年同大学理工研究域電子情報学系准教授。2009 年慶應義塾大学理工学部システムデザイン工学科准教授, 2014 年教授となり現在に至る。ロボパスト制御理論, 分散協調制御理論とそのネットワークロボティクス, 電力ネットワークへの応用に関する研究に従事。博士 (工学)。システム制御情報学会, IEEE などの会員。