Nghiên cứu khả năng phát hiện tấn công tuyến tính trong các hệ thống điều khiển công nghiệp bằng phương pháp CUSUM

Ability to Detect the Linear Attack in Industrial Control Systems by CUSUM Method

***Nguyễn Đức Dương*[1,2], *Lê Minh Thùy*[1], *Cung Thành Long*[1*]**

[1]*Trường Đại học Bách khoa Hà Nội - Số 1, Đại Cồ Việt, Hai Bà Trưng, Hà Nội, Việt Nam*

[2]*Trường Đại học Kinh tế Kỹ thuật Công nghiệp - 456 Phố Minh Khai, Hai Bà Trưng, Hà Nội, Việt Nam*

Đến Tòa soạn: 21-11-2018; chấp nhận đăng: 25-09-2020

Tóm tắt

Bài báo trình bày khả năng áp dụng phương pháp Cumulative Sum (CUSUM) nhằm phát hiện tấn công tuyến tính trong trường hợp phương pháp Kullback – Leibler (K-L) không phát hiện được. Để thực hiện mục tiêu này, chúng tôi đã khảo sát nhằm tìm ra các ma trận tấn công tuyến tính, mà với chúng phương pháp phát hiện K-L bị vượt qua, trong dải ngưỡng phát hiện từ 0 tới 176,76, với đối tượng nghiên cứu là quá trình trộn nhiệt trong các nhà máy thực phẩm. Sau đó, các ma trận tấn công tuyến tính thu được đã được sử dụng để tìm ngưỡng phát hiện của phương pháp CUSUM. Qua phân tích các kết quả thử nghiệm trên dữ liệu mô phỏng, chúng tôi xác định được dải ngưỡng thích hợp để phương pháp CUSUM phát hiện được tấn công tuyến tính, khi phương pháp K-L bị vượt qua.

Từ khóa: Tấn công tuyến tính, phương pháp CUSUM, phương pháp độ chênh Kullback – Leibler

Abstract

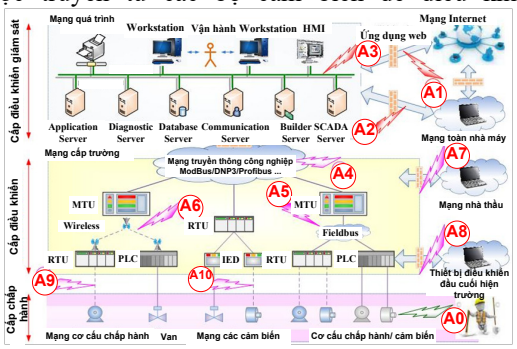
This paper presents the ability to apply the Cumulative Sum (CUSUM) method to detect linear attacks, in case the Kullback–Leibler (K-L) method cannot detect. In order to do this, we investigated to find linear attack matrices, for which the K-L method was surpassed in the range of detection threshold from 0 to 176.76 and applying for heat mixing process, which is used in food plants. Then, the found linear attack matrices were used to find the detection threshold of the CUSUM method. By analyzing the tested results, which is implemented on simulation data, we have determined a suitable range of threshold for CUSUM method to detect linear attacks, when the K-L method is failed to detect.

Keywords: Linear attack, CUSUM method, Kullback–Leibler divergence method

1. Giới thiệu

Các hệ thống điều khiển công nghiệp (tập trung hoặc phân tán) là các hệ thống được sử dụng để giám sát, điều khiển các trạm, hay nhà máy xí nghiệp công nghiệp với nhiều quy mô khác nhau. Để thực hiện các chức năng của hệ thống, việc thu thập, truyền nhận và kiểm soát, đảm bảo tính toàn vẹn của dữ liệu là rất quan trọng.

Các hệ thống điều khiển công nghiệp dễ bị tấn công phối hợp không chỉ trên các cơ sở hạ tầng vật chất mà còn trên lớp truyền thông và trung tâm điều khiển, như minh họa ở hình 1 [1]. Trong đó các phương thức tấn công ở điểm A1, A2, A3 là một số thủ đoạn tấn công nhằm vào lớp điều khiển giám sát, thông qua việc chiếm quyền truy cập vào trung tâm điều khiển từ các ứng dụng trên web server; lan truyền virus phá hoại cấu hình mạng điều khiển, giám sát của nhà máy; A4 là chiếm quyền truy cập vào các kênh truyền thông giữa trung tâm điều khiển và các trạm; A5, A6 là tấn công vào liên kết truyền thông giữa MTU và PLC /RTU; A7 là tấn công đường kết nối mạng giữa nhà máy và nhà thầu; A8 là tấn công các thiết bị đầu cuối hiện trường; A9 là tấn công đường tín hiệu gửi từ bộ điều khiển cho các thiết bị truyền động; A10 là tấn công các tín hiệu phản hồi được truyền từ các bộ cảm biến để điều khiển;



Hình 1. Các điểm có khả năng bị tấn công trong hệ thống điều khiển phân tán công nghiệp

A0 là tấn công cơ học trực tiếp vào các thiết bị chấp hành của các hệ thống điều khiển công nghiệp. Vì vậy, các nghiên cứu đảm bảo an toàn dữ liệu cho các

*Địa chỉ liên hệ: Tel.: (+84) 963 95 88 54
Email: long.cungthanh@hust.edu.vn

hệ thống điều khiển công nghiệp đang được quan tâm lớn. Có hai hướng chính, đó là nghiên cứu các thủ đoạn tấn công mới nhằm đánh giá khả năng của các phương pháp bảo mật thông tin, và hướng nghiên cứu thứ hai là tập trung xây dựng các phương pháp phát hiện dữ liệu bị tấn công.

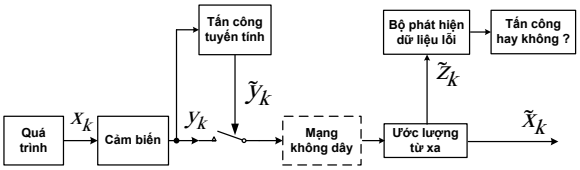
Theo hướng nghiên cứu thứ nhất, hiện có thể phân loại một số phương pháp tấn công như tấn công từ chối dịch vụ – (Denial-Of-Service- DoS), tấn công tính toàn vẹn dữ liệu truyền nhận giữa các lớp, hoặc trong các lớp mạng của hệ thống điều khiển, bằng các hình thức như làm sai lệch thông tin, chèn thông tin giả, ...[2]. Gần đây, có công bố về phương pháp tấn công tuyến tính của nhóm nghiên cứu tại Đại học Công nghệ Hồng Kông [3]. Đây là phương pháp tấn công vào tính toàn vẹn dữ liệu ở cấp hiện trường với độ nguy hiểm cao, tập trung vào điểm tấn công A9 và A10 (hình 1). Loại tấn công này gây ra sai lệch tín hiệu lớn hơn nhiều so với tấn công từ chối dịch vụ và một số dạng tấn công khác. Nhóm nghiên cứu đã chỉ ra rằng, một số thuật toán phát hiện tấn công như Chi-squared, K-L hoàn toàn có thể bị vượt qua với kiểu tấn công này [3], [4].

Hướng nghiên cứu thứ hai, về đảm bảo an toàn thông tin, hiện nhận được nhiều sự quan tâm và có nhiều công trình được công bố. Trong nước đã có nhóm nghiên cứu áp dụng phương pháp CUSUM phát hiện tấn công đột ngột (surge attack), hay tấn công phân cực (bias attack) trên các đối tượng một vào một ra (SISO) [5]. Trên thế giới đã có công trình xây dựng bộ dữ liệu mô phỏng chuẩn về các trường hợp bị tấn công làm sai lệch dữ liệu trong các hệ thống mạng công nghiệp [2]. Trên cơ sở đó, nhiều nhóm nghiên cứu quốc tế đã phát triển các phương pháp xử lý tín hiệu nhằm nâng cao khả năng nhận biết các trường hợp dữ liệu bị tấn công. Một số kỹ thuật nhận dạng dựa trên học máy, học sâu đã được công bố có khả năng phát hiện mất an toàn dữ liệu rất cao, tới trên 99% [2].

Trong bài báo này, chúng tôi trình bày nghiên cứu khả năng áp dụng phương pháp CUSUM phát hiện tấn công tuyến tính trong trường hợp phương pháp K-L không phát hiện được. Các phần tiếp theo của bài báo được tổ chức như sau: phần 2 trình bày khái lược về tấn công tuyến tính và phương pháp phát hiện K-L; phần 3 giới thiệu về phương pháp phát hiện dữ liệu bất thường CUSUM; phần 4 giới thiệu mô hình một hệ thống điều khiển là đối tượng chịu tấn công tuyến tính; phần 5 phân tích một số trường hợp mà phương pháp K-L bị vượt qua bởi kiểu tấn công tuyến tính, và xác định điều kiện để phát hiện được kiểu tấn công này khi áp dụng phương pháp CUSUM; cuối cùng, trong phần 6 trình bày một số kết luận và hướng nghiên cứu tiếp theo.

2. Tấn công tuyến tính và phương pháp phát hiện K-L

Xét hệ thống điều khiển với điểm chịu tấn công tuyến tính được mô tả như hình 2, làm thay đổi dữ liệu truyền không dây tại đầu ra của các cảm biến.



Hình 2. Sơ đồ minh họa vị trí chịu tấn công tuyến tính

Trong đó, phương trình mô tả tín hiệu tại đầu vào và đầu ra của cảm biến viết được như trong (1) và (2) [4]

$$x_{k+1} = Ax_k + \omega_k \quad (1)$$

$$y_k = Cx_k + v_k \quad (2)$$

với:

$x_k \in \mathbb{R}^n$ - vector biến trạng thái của hệ thống (tín hiệu đầu vào của cảm biến)

$y_k \in \mathbb{R}^m$ - vector tín hiệu ở đầu ra của cảm biến

$\omega_k \in \mathbb{R}^n, \omega_k \sim N(0,Q)$ - nhiễu trắng tác động lên biến trạng thái

$v_k \in \mathbb{R}^m, v_k \sim N(0,R)$ - nhiễu trắng tác động lên cảm biến

$Q \geq 0; R > 0$ - hiệp phương sai của nhiễu trắng

$\tilde{x}_k$ - ước lượng trạng thái của bộ ước lượng từ xa

$A \in \mathbb{R}^{n\times n}, C \in \mathbb{R}^{m\times m}$ - các ma trận hệ thống đã biết

Giá trị các ước lượng trạng thái $\hat{x}_k$ (khi không bị tấn công), $\tilde{x}_k$ (khi bị tấn công), được viết dưới dạng [4]:

$$\hat{x}_k^- = \mathrm{A}\hat{x}_{k-1} \quad (3)$$

$$\hat{x}_k = \hat{x}_k^- + K_k\left(y_k - C\hat{x}_k^-\right) \quad (4)$$

trong đó, K_k là ma trận hệ số Kalman [4].

Trường hợp không bị tấn công, ước lượng sai lệch tín hiệu đầu ra của cảm biến có thể viết như trong (5):

$$\begin{aligned} &z_k = y_k - C\hat{x}_k^-; z_k \sim N(0;\Sigma) \\ &\Sigma = C\bar{P}C^T + R; E\left[z_i, z_j^T\right] = 0 \ \forall \mathrm{i} \neq \mathrm{j} \end{aligned} \quad (5)$$

với R là ma trận hiệp phương sai của nhiễu trắng, $\bar{P}$ là ước lượng hiệp phương sai (biến trạng thái của hệ thống) ở trạng thái ổn định, $E\left[z_i, z_j^T\right]$ là kỳ vọng các thành phần phần dư z_k [4].

Trường hợp bị tấn công, tín hiệu ra của cảm biến bị thay đổi như mô tả trong công thức (6):

$$\tilde{y}_k = \tilde{z}_k + C\tilde{x}_k^- \quad (6)$$

Các tác giả trong [4] đã nghiên cứu khả năng tấn công tuyến tính vượt qua phương pháp phát hiện sai lệch dữ liệu K-L. Đây là một phương pháp phát hiện lỗi được đánh giá cao, dựa trên nguyên tắc tính độ chênh giữa hai chuỗi giá trị ngẫu nhiên $\tilde{z}_k$ và z_k. Giả sử $f_{\tilde{z}_k}(\chi)$ và $f_{z_k}(\chi)$ là hàm mật độ của $\tilde{z}_k$ và z_k, ta có độ chênh D giữa $\tilde{z}_k$ và z_k như công thức (7).

$$D\left(\tilde{z}_k \| z_k\right) = \int f_{\tilde{z}_k}(\chi) \log \frac{f_{\tilde{z}_k}(\chi)}{f_{z_k}(\chi)} d\chi \tag{7}$$

Khi độ chênh vượt ngưỡng, dữ liệu được đánh giá là bị tấn công làm sai lệch giá trị, và ngược lại, như thể hiện trong (8):

$$\begin{cases} D\left(\tilde{z}_k \| z_k\right) \le \delta \rightarrow \text{không bị tấn công} \\ D\left(\tilde{z}_k \| z_k\right) > \delta \rightarrow \text{bị tấn công} \end{cases} \tag{8}$$

với δ là ngưỡng phát hiện đặt trước của phương pháp K-L.

Theo [4], dưới tác động của tấn công tuyến tính, tín hiệu cảm biến y_k bị biến đổi thành $\tilde{y}_k$ thỏa mãn (6) và (9):

$$\tilde{z}_k = T_k z_k + b_k \tag{9}$$

với $T_k \in \mathbb{R}^{m \times m}$ - ma trận tấn công tuyến tính

$b_k \sim N\left(0, \Gamma_k\right)$ - biến ngẫu nhiên dạng Gaussian

Tấn công tuyến tính sẽ vượt qua phương pháp phát hiện K-L khi có thể xác định được T_k, Γ_k thoả mãn (10):

$$\begin{cases} \max\limits_{(T_k, b_k)} Tr\left(\tilde{P}_k\right) \\ D\left(\tilde{z}_k \| z_k\right) \le \delta, \forall k \end{cases} \tag{10}$$

với $Tr\left(\tilde{P}_k\right)$ là vết của ma trận hiệp phương sai khi dữ liệu bị tấn công. Trong đó ma trận $\tilde{P}_k$ được tính như công thức (11) [4]:

$$\tilde{P}_k = A\tilde{P}_{k-1}A^T + Q + K\tilde{\Sigma}_k K^T - \bar{P}C^T T_k^T K^T - KT_k C\bar{P} \tag{11}$$

Theo [4], ta có nghiệm tối đa $\tilde{z}_k^*$ của (10), thỏa mãn:

$$\tilde{\Sigma}_k = \left(\Sigma^{-1} - \frac{2}{\mu} K^T K\right)^{-1} \tag{12}$$

với $\mu > 2 \min\limits_{1 \le i \le m} \lambda_i$

và $\lambda_1, \lambda_2, ..., \lambda_m$ là các giá trị riêng của $K^T K \Sigma$

Theo quy hoạch lồi Karush Kuhn Tucker, từ (12), ta có mối quan hệ giữa ngưỡng δ và μ:

$$\mu\left(\frac{1}{2} Tr\left(\Sigma^{-1}\tilde{\Sigma}_k\right) - \frac{m}{2} + \frac{1}{2}\log\frac{|\Sigma|}{|\tilde{\Sigma}_k|} - \delta\right) = 0 \tag{13}$$

Ma trận T_k thỏa mãn (10) được xác định từ việc giải phương trình tối ưu quy hoạch lồi (14):

$$\begin{cases} \min\limits_{(T_k)} Tr\left(C\bar{P}\bar{P}C^T\Sigma^{-1}T_k\right) \\ \begin{bmatrix} \tilde{\Sigma} & T_k \\ T_k^T & \Sigma^{-1} \end{bmatrix} \le 0 \end{cases} \tag{14}$$

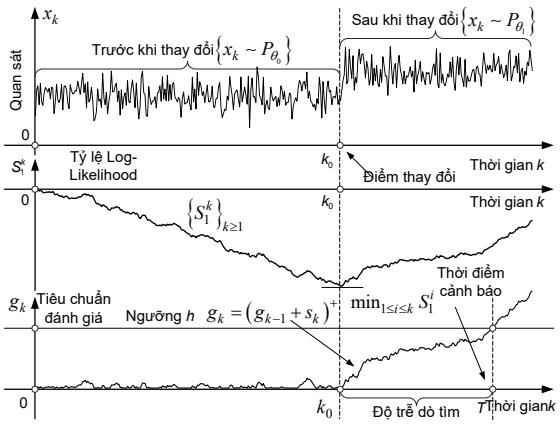
và ma trận Γ_k được xác định từ mối quan hệ:

$$\Gamma_k = \tilde{\Sigma}_k - T_k \Sigma T_k^T \tag{15}$$

Như vậy, với mỗi ngưỡng δ của phương pháp K-L đều có thể tìm ra các ma trận tấn công T_k, Γ_k. Hay nói cách khác, luôn tồn tại khả năng để tấn công tuyến tính có thể vượt qua phương pháp phát hiện sai lệch dữ liệu K-L.

3. Tổng quan phương pháp CUSUM

Phương pháp CUSUM có khác biệt so với phương pháp K-L ở điểm là phương pháp này áp dụng lý thuyết Wald phân tích tính bất thường trong dữ liệu [6]. Đây là cơ sở để nghiên cứu khả năng phát hiện tấn công tuyến tính của phương pháp CUSUM.



Hình 3. Minh họa phương pháp CUSUM phát hiện dữ liệu bị tấn công

Xét hệ ngẫu nhiên $X = \left[x_1, x_2, ...x_k\right]^T \sim N\left(\mu, \Sigma\right)$. Giả sử khi chưa bị tấn công thì $X \sim N\left(\mu, \Sigma_0\right)$ và khi bị tấn công thì $X \sim N\left(\mu, \Sigma_1\right)$. Ý tưởng của phương pháp CUSUM là tính đến tỷ lệ thay đổi thực sự S_i (likelihood ratio – LLR), (như được minh họa ở Hình 3), xác định theo công thức (16) [6]:

$$S_i = \ln \frac{f_{\theta_1}(x_i)}{f_{\theta_0}(x_i)};\ S_i^k = \sum_{t=i}^{k} \ln \frac{f_{\theta_1}(x_t)}{f_{\theta_0}(x_t)} \quad (16)$$

với θ_1 là chỉ số các điểm khi có thay đổi bất thường, θ_0 là chỉ số các điểm khi không có thay đổi bất thường.

Theo cách định nghĩa, S_i có xu hướng biến thiên đơn điệu khi không có thay đổi bất thường trong tín hiệu, và đổi chiều biến thiên tại thời điểm xảy ra thay đổi bất thường.

Tiêu chuẩn để xác định dữ liệu có bị tấn công hay không được tính như công thức (17):

$$g_k = \left(g_{k-1} + \ln \frac{f_{\theta_1}(x_k)}{f_{\theta_0}(x_k)} \right)^+, \quad g_0 = 0 \quad (17)$$

$$\text{với } (x)^+ = \max(0, x) \quad (18)$$

Theo phân bố chuẩn nhiều chiều (Multivariate Gaussian Distribution), ta có hiệp phương sai khi không có tấn công:

$$P_0(X) = P(x, \mu, \Sigma_0) = \left(\frac{1}{\sqrt{2\pi}} \right)^n (\det \Sigma_0)^{-1/2} \exp\left(-\frac{1}{2} (x-\mu)^T \Sigma_0^{-1} (x-\mu) \right) \quad (19)$$

với $\det \Sigma_0$ là định thức của ma trận Σ_0, được tính như công thức (20):

$$\Sigma_0 = \Sigma = C\bar{P}C^T + R \quad (20)$$

Hiệp phương sai khi bị tấn công được viết dạng:

$$P_1(X) = P(x, \mu, \Sigma_1) = \left(\frac{1}{\sqrt{2\pi}} \right)^n (\det \Sigma_1)^{-1/2} \exp\left(-\frac{1}{2} (x-\mu)^T \Sigma_1^{-1} (x-\mu) \right) \quad (21)$$

với $\det \Sigma_1$ là định thức của ma trận Σ_1, được tính như công thức (22):

$$\Sigma_1 = T_k \Sigma T_k^T + \Gamma_k. \quad (22)$$

Và do đó, tỷ lệ thay đổi (LLR) tính được theo công thức (23):

$$s_k = \frac{1}{2} \ln \frac{\det \Sigma_0}{\det \Sigma_1} - \frac{1}{2} (x-\mu)^T \left[\Sigma_1^{-1} - \Sigma_0^{-1} \right] (x-\mu) \quad (23)$$

Các giá trị g_k của phương pháp CUSUM được xác định theo nguyên tắc trong công thức (24):

$$g_k = \begin{cases} g_{k-1} + s_k & \text{if } g_{k-1} + s_k > 0 \\ 0 & \text{if } g_{k-1} + s_k < 0 \end{cases} \quad (24)$$

Từ (24) và (18), ta có:

$$g_k = \max(0, g_{k-1} + s_k) \quad (25)$$

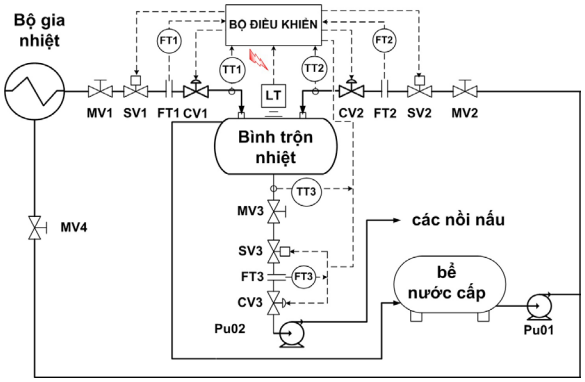
và thời điểm cảnh báo tấn công T_a được xác định từ điều kiện:

$$T_a = \min(k : g_k \geq h) \quad (26)$$

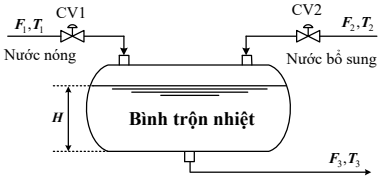
Trong đó h là ngưỡng phát hiện tấn công (đặt trước) theo phương pháp CUSUM.

4. Đối tượng mô phỏng

Trong bài báo này, chúng tôi xét mô hình tổng quát của quá trình trộn nhiệt trong các nhà máy sản xuất thực phẩm nói chung. Sơ đồ công nghệ của quá trình này được minh họa như trong hình 4.



Hình 4. Sơ đồ công nghệ hệ thống bình trộn nhiệt



Hình 5. Sơ đồ đơn giản hóa bình trộn nhiệt

Các thông số và các biến quá trình của mô hình bao gồm:

F_1 : lưu lượng của nước nóng (m^3/h)
F_2 : lưu lượng của nước bổ sung (m^3/h)
F_3 : lưu lượng của nước sau trộn (m^3/h)
T_1 : nhiệt độ của nước nóng (°C)
T_2 : nhiệt độ của nước bổ sung (°C)
T_3 : nhiệt độ của nước trong bình trộn (°C)
H : chiều cao mức nước bình trộn nhiệt; H_{max} = 2.8 (m)
ρ : khối lượng riêng của nước; ρ=1000 (kg/m^3)
V : thể tích lượng nước trong bình trộn nhiệt
L : chiều dài bình trộn nhiệt; L = 14,5 (m)
R : bán kính của bình; R = 1,4 (m)

Nguyên lý hoạt động của quá trình là trộn dòng nước nóng qua van điều khiển CV1, SV1 với dòng nước lạnh qua van điều khiển CV2, SV2 để nước trong bình trộn có nhiệt độ mong muốn. Nước từ bình trộn cấp tới các nồi nấu thông qua van CV3 và SV3. Như vậy, có thể đơn giản hoá các thành phần của bình trộn nhiệt mà không làm sai lệch nguyên lý hoạt

động của quá trình đã chọn, ta có sơ đồ bình trộn nhiệt như minh họa trong hình 5.

Từ mô hình hoạt động của hệ thống đã chọn, ta có biến thiên mực nước trong bình trụ tròn nằm ngang có thể xác định theo công thức (27) [7]:

$$\frac{dH}{dt} = \frac{1}{A}\left(F_1 + F_2 - F_3\right) \tag{27}$$

với $A = 2L\sqrt{H\left(2R - H\right)}$ (28)

Theo định luật bảo toàn năng lượng, biến thiên nhiệt độ nước trong bình trộn nhiệt có thể tính theo công thức:

$$\frac{dT_3}{dt} = \frac{1}{V}\left[F_1T_1 + F_2T_2 - T_3\left(F_1 + F_2\right)\right] \tag{29}$$

Từ (27) ta có phương trình sai phân của mức nước trong bình:

$$\Delta\dot{H} = \frac{1}{\overline{A}}\Delta F_1 + \frac{1}{\overline{A}}\Delta F_2 - \frac{1}{\overline{A}}\Delta F_3 \tag{30}$$

với $\overline{A} = 2L\sqrt{\overline{H}\left(2R - \overline{H}\right)}$ (31)

Từ (29) ta có phương trình sai phân của nhiệt độ nước trong bình trộn nhiệt:

$$\begin{aligned}\Delta\dot{T}_3 = {} & \frac{\overline{T_1} - \overline{T_3}}{\overline{V}}\Delta F_1 + \frac{\overline{T_2} - \overline{T_3}}{\overline{V}}\Delta F_2 + \frac{\overline{F_1}}{\rho\overline{V}}\Delta T_1 \\ & + \frac{\overline{F_2}}{\rho\overline{V}}\Delta T_2 - \frac{\overline{F_1} + \overline{F_2}}{\rho\overline{V}}\Delta T_3\end{aligned} \tag{32}$$

với ký hiệu ngang trên ($\overline{*}$) để chỉ giá trị của một biến tại điểm làm việc, ký hiệu ($\Delta*$) biểu diễn biến chênh lệch so với giá trị tại điểm làm việc.

Từ (30) và (32), ta có:

$$\underbrace{\begin{bmatrix}\Delta\dot{H} \\ \Delta\dot{T}_3\end{bmatrix}}_{\dot{x}(t)} = \underbrace{\begin{bmatrix}0 & 0 \\ 0 & -\dfrac{\overline{F_1} + \overline{F_2}}{\rho\overline{V}}\end{bmatrix}}_{\tilde{A}}\underbrace{\begin{bmatrix}\Delta H \\ \Delta T_3\end{bmatrix}}_{x(t)} + \underbrace{\begin{bmatrix}\dfrac{1}{\overline{A}} & \dfrac{1}{\overline{A}} \\ \dfrac{\overline{T_2} - \overline{T_3}}{\overline{V}} & \dfrac{\overline{T_1} - \overline{T_3}}{\overline{V}}\end{bmatrix}}_{\tilde{B}}\underbrace{\begin{bmatrix}\Delta F_2 \\ \Delta F_1\end{bmatrix}}_{u(t)} + \begin{bmatrix}\dfrac{-1}{\overline{A}} & 0 & 0 \\ 0 & \dfrac{\overline{F_1}}{\rho\overline{V}} & \dfrac{\overline{F_2}}{\rho\overline{V}}\end{bmatrix}\begin{bmatrix}\Delta F_3 \\ \Delta T_1 \\ \Delta T_2\end{bmatrix} \tag{33}$$

Do lưu lượng F_3 và nhiệt độ T_1 , T_2 là các đại lượng nhiễu quá trình, giả sử các đại lượng này không thay đổi trong suốt quá trình trộn tức là: $\Delta F_3 = 0;\ \Delta T_1 = \Delta T_2 = 0$, ta có:

$$\dot{x}(t) = \tilde{A}x(t) + \tilde{B}u(t) \tag{34}$$

Giả sử rằng hai bộ cảm biến LT và TT3 được sử dụng để đo mức trong bình trộn nhiệt *H(t)* và nhiệt độ trong bình trộn $T_3(t)$. Ta có phương trình đo:

$$\underbrace{\begin{bmatrix}y_1(t) \\ y_2(t)\end{bmatrix}}_{y(t)} = \underbrace{\begin{bmatrix}K_1 & 0 \\ 0 & K_2\end{bmatrix}}_{\tilde{C}}\underbrace{\begin{bmatrix}\Delta H \\ \Delta T_3\end{bmatrix}}_{x(t)} \tag{35}$$

$y_1(t)$: tín hiệu đo mức bình trộn

$y_2(t)$: tín hiệu đo nhiệt độ trong bình trộn

Xét bình trộn nhiệt với các thông số:

$F_{\max} = 160m^3/h;\ \overline{H} = 0.5H_{\max};\overline{T_1} = 80^O\text{C};\overline{T_2} = 37^O\text{C}$

$\overline{T_3} = 39^O\text{C};\overline{F_1} = \overline{F_2} = \overline{F_3} = 0.5F_{\max};K_1 = 0.5;K_2 = 0.7$

và chu kỳ lấy mẫu $T_S = 0.1s$

Từ (34) và (35) ta có mô hình trạng thái liên tục mô tả đối tượng:

$$\begin{aligned}\dot{x}(t) &= \tilde{A}x(t) + \tilde{B}u(t) \\ y(t) &= \tilde{C}x(t)\end{aligned} \tag{36}$$

với $\tilde{A} = \begin{bmatrix}0 & 0 \\ 0 & -0.396\end{bmatrix}10^{-5};\ \tilde{B} = \begin{bmatrix}0.0308 & 0.0308 \\ -0.1573 & 3.2253\end{bmatrix}$

Từ (36), chuyển sang dạng gián đoạn, ta có mô hình trạng thái không liên tục của đối tượng:

$$\begin{cases}x_{k+1} = Ax_k + Bu_k \\ y_k = Cx_k\end{cases} \tag{37}$$

với

$$A = \begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix};\ \text{B} = \begin{bmatrix}0.0031 & 0.0031 \\ -0.0157 & 0.3225\end{bmatrix};C = \begin{bmatrix}0.5 & 0 \\ 0 & 0.7\end{bmatrix}$$

Do ở đây ta không quan tâm tín hiệu điều khiển, giả sử $u_k = 0$, đối tượng có thể mô tả dưới dạng phương trình (1) và (2). Trong đó, các ma trận hiệp phương sai của nhiễu trắng được chọn trong các mô phỏng của chúng tôi là:

$$Q = \begin{bmatrix}0.51 & 0 \\ 0 & 0.505\end{bmatrix};R = \begin{bmatrix}1 & 0 \\ 0 & 0.8\end{bmatrix}$$

5. Kết quả mô phỏng và thảo luận

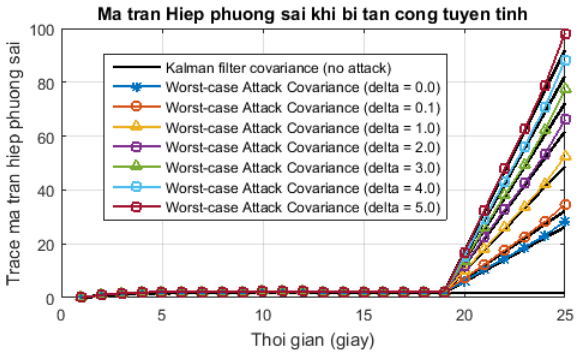
Trong phần này, chúng tôi thực hiện mô phỏng để đánh giá khả năng áp dụng phương pháp CUSUM phát hiện tấn công tuyến tính vào đối tượng được xây dựng trong phần 4. Trong đó, điểm mấu chốt là đánh giá khả năng tồn tại ngưỡng *h* để phương pháp vẫn có thể phát hiện dữ liệu bị tấn công khi phương pháp K-L bị vượt qua.

Các công thức (12) và (13) cho biết mối quan hệ giữa δ và μ ở phương pháp K-L. Chọn các ngưỡng

δ ta tính được μ. Về lý thuyết, ngưỡng $\delta \geq 0$, song với giá trị lớn sẽ làm giảm độ nhạy của phương pháp, ngược lại các giá trị quá nhỏ sẽ gia tăng khả năng cảnh báo nhầm. Trong bài báo này, chúng tôi xét một số giá trị của δ trong phạm vi $\delta \in [0;\ 176.76]$.

$$\delta = \begin{cases} 0 \equiv \delta_0 \\ 0.1 \equiv \delta_1 \\ 1.0 \equiv \delta_2 \\ 2.0 \equiv \delta_3 \\ 3.0 \equiv \delta_4 \\ 4.0 \equiv \delta_5 \\ 5.0 \equiv \delta_6 \\ 176.76 \equiv \delta_7 \end{cases} \Rightarrow \mu = \begin{cases} \infty \equiv \mu_0 \\ 2.9813 \equiv \mu_1 \\ 1.4850 \equiv \mu_2 \\ 1.3005 \equiv \mu_3 \\ 1.2290 \equiv \mu_4 \\ 1.1862 \equiv \mu_5 \\ 1.1587 \equiv \mu_6 \\ 1.0235 \equiv \mu_7 \end{cases}$$

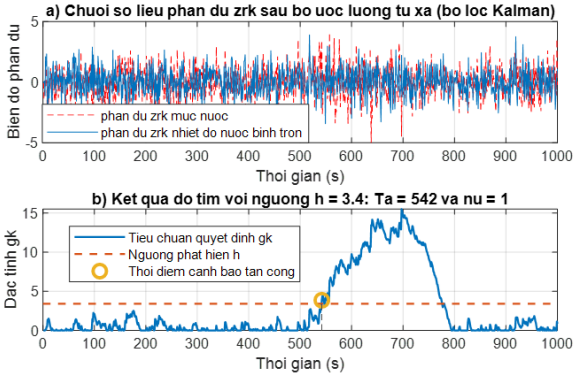
Từ các ma trận trạng thái A, C của đối tượng và các ma trận nhiễu giả định Q, R đã thiết lập được trong phần 4, giải hệ tối ưu (14) bằng CVX toolbox trong matlab, và từ (15) ta thu được các ma trận tham số của tấn công tuyến tính $T_{k0} \div T_{k7}; \Gamma_{k0} \div \Gamma_{k7}$. Với các ma trận này, tấn công tuyến tính vượt qua phương pháp phát hiện K-L. Để kiểm tra sự sai khác giữa tín hiệu khi không bị tấn công với khi bị tấn công tuyến tính, chúng tôi tính vết của các ma trận hiệp phương sai $Tr(\tilde{P}_k)$ như minh họa trong hình 6.



Hình 6. Vết của ma trận hiệp phương sai khi bị tấn công tuyến tính

Kết quả mô phỏng trong hình 6 cho thấy, ban đầu khi chưa bị tấn công, trong khoảng thời gian [0; 19s], vết của ma trận hiệp phương sai xấp xỉ 2. Khi xuất hiện tấn công tuyến tính, hiệp phương sai ước lượng hệ thống $Tr(\tilde{P}_k)$ tăng vọt, thể hiện sự sai khác lớn giữa tín hiệu trước và sau khi bị tấn công. Trong khi đó, với các ma trận T_k, Γ_k đã chọn, phương pháp K-L, với các ngưỡng δ tương ứng, không phát hiện được sự thay đổi của tín hiệu do tấn công tuyến tính gây ra. Cần nhấn mạnh rằng, với mỗi ngưỡng δ cho trước, luôn xác định được một bộ tham số T_k và Γ_k của tấn công tuyến tính để nó vượt qua phương pháp phát hiện K-L.

Sau đó, chúng tôi sử dụng bộ các ma trận tấn công tuyến tính $T_{k0} \div T_{k7}; \Gamma_{k0} \div \Gamma_{k7}$ đã vượt qua phương pháp K-L, xét trên đối tượng đã lập ở phần 4, nhằm đánh giá khả năng phát hiện của phương pháp CUSUM. Trước hết chúng tôi xét trường hợp tấn công tuyến tính vượt qua phương pháp K-L ở ngưỡng thấp $\delta = \delta_1 = 0.1$.



Hình 7. Khả năng phát hiện tấn công tuyến tính bằng phương pháp CUSUM khi K-L bị vượt qua với ngưỡng $\delta = 0.1$

Bằng các ma trận T_k, Γ_k đã xác định vượt qua phương pháp K-L, xây dựng bộ dữ liệu giả lập có tấn công tuyến tính xảy ra trong khoảng thời gian từ 500s đến 700s; chọn ngưỡng phát hiện của phương pháp CUSUM h =3.4, áp dụng các công thức (20, 22, 23, 25, 26) tính thời điểm cảnh báo tấn công T_a, thu được kết quả mô phỏng như trong hình 7. Kết quả mô phỏng cho thấy, trong khoảng thời gian xảy ra tấn công tuyến tính đã giả lập, phương pháp CUSUM đã phát hiện ra loại tấn công này tại thời điểm $T_a = 542s$.

Áp dụng các tính toán tương tự, ta có thể chọn được các giá trị ngưỡng h khác, mà với chúng CUSUM vẫn phát hiện được tấn công tuyến tính. Một số kết quả tính h được thống kê trong bảng 1. Với giả thiết tấn công tuyến tính được thực hiện trong khoảng thời gian 500 ÷ 700s đã chọn, kết quả lựa chọn cho thấy, phương pháp CUSUM đã phát hiện tấn công tuyến tính với ngưỡng h nằm trong khoảng $h \in [3.4;15.4]$.

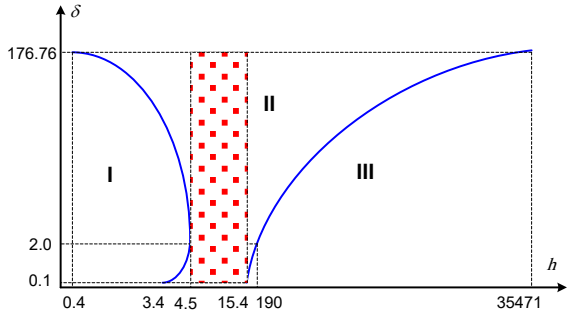
Tiến hành tương tự cho các ngưỡng δ (của phương pháp K-L) lớn hơn, chúng tôi tìm được các khoảng giá trị khác của ngưỡng h để CUSUM có thể phát hiện ra tấn công tuyến tính khi K-L bị vượt qua. Bảng 2 thể hiện một số kết quả này.

Bảng 1. Khả năng phát hiện của CUSUM với $\delta = 0.1$

δ	h	T_a	nu	H	T_a	nu
0.1	3.2	51	1	13	583	1
	3.4	542	1	14	607	1
	4	585	1	14.5	660	1
	10	566	1	15.4	617	1
	12	572	1	15.5	0	0

Bảng 2. Ngưỡng phát hiện tấn công tuyến tính của CUSUM khi K-L bị vượt qua

Ngưỡng δ, PP K-L	Ngưỡng h, phát hiện tấn công tuyến tính bằng PP Cusum	
	h_{min}	h_{max}
0.1	3.4	15.4
1.0	4.3	165
2.0	4.5	190
3.0	3.9	386
4.0	4.5	805
5.0	4	957
176.76	0.4	35471



Hình 8. Khả năng phát hiện tấn công tuyến tính bằng phương pháp CUSUM

Các kết quả trong bảng 2 được liên tục hóa dưới dạng đồ thị trong hình 8. Mối quan hệ giữa h và δ có thể chia làm 3 vùng; trong đó vùng II là tập hợp các giá trị h mà phương pháp CUSUM phát hiện được tấn công tuyến tính. Đặc biệt với các giá trị $h \in [4.5;15.4]$ (nằm trong miền chữ nhật giữa vùng II), phương pháp CUSUM luôn phát hiện tấn công tuyến tính, khi phương pháp độ chênh K-L bị vượt qua với mọi giá trị $\delta \in [0.1;176.76]$. Vùng I và III là vùng mà các giá trị h ở đó sẽ khiến phương pháp CUSUM không phát hiện được tấn công tuyến tính (trong trường hợp K-L bị vượt qua).

Như vậy, tương ứng với mỗi giá trị ngưỡng δ của phương pháp độ chênh K-L mà tấn công tuyến tính vượt qua, ta thấy khả năng tồn tại một khoảng ngưỡng h để phát hiện tấn công tuyến tính bằng phương pháp CUSUM. Kết quả trong hình 8 thể hiện khả năng thực tế có thể chọn một ngưỡng h thích hợp để CUSUM phát hiện được tấn công tuyến tính trong mọi trường hợp K-L bị vượt qua.

6. Kết luận và hướng nghiên cứu tiếp theo

Các phân tích trong bài đã chỉ ra rằng luôn tồn tại một khoảng giá trị ngưỡng h trong phương pháp CUSUM mà với các giá trị đó thì có thể phát hiện được tấn công tuyến tính, khi nó vượt qua phương pháp độ chênh Kullback – Leibler. Đây là kết quả tiềm năng để có thể xây dựng được mối liên hệ tổng quát giữa các ngưỡng đặt của hai kỹ thuật này. Đồng thời chỉ ra rằng, có thể sử dụng phương pháp CUSUM như một tầng phát hiện phía sau trong chuỗi các kỹ thuật được áp dụng để đảm bảo tính toàn vẹn dữ liệu của các hệ thống điều khiển công nghiệp.

Trong các nghiên cứu tiếp theo, chúng tôi sẽ triển khai phương pháp CUSUM mở rộng và một số phương pháp phát hiện tấn công tính toàn vẹn dữ liệu khác, để đánh giá cụ thể hơn khả năng phát hiện loại tấn công tuyến tính này.

Tài liệu tham khảo

[1] M. Lehto and P. Neittaanmäki, Cyber Security: Analytics, Technology and Automation. Springer, 2015.

[2] A. Hijazi, A. E. Safadi, and J.-M. Flaus, A Deep Learning Approach for Intrusion Detection System in Industry Network, in BDCSIntell, Beirut, Lebanon, 2018, pp. 55-62.

[3] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, Optimal Linear Cyber-Attack on Remote State Estimation, IEEE Trans. Control Netw. Syst., vol. 4, no. 1, (2017) pp. 4–13.

[4] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, Consequence Analysis of Innovation-based Integrity Attacks with Side Information on Remote State Estimation., IFAC-Pap., vol. 50, no. 1, (2017) pp. 8399–8404.

[5] Truong N.D and Tu L.M, A new method against attacks on networked industrial control systems, Kỷ yếu Hội nghị Khoa học Quốc gia lần thứ IX về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin, Cần Thơ, Việt Nam, 2016, pp.9-16.

[6] M. Basseville and I. V. Nikiforov, Detection of Abrupt Changes: Theory and Application. Englewood Cliffs, N.J: Prentice Hall, 1993.

[7] Duong N.D and Ha V.T, Nghiên cứu phương pháp điều khiển tích cực loại bỏ nhiễu để nâng cao chất lượng quá trình trao đổi nhiệt trong bình trộn nhiệt, Tạp Chí Khoa Học Công Nghệ Đại Học Công Nghiệp Hà Nội, vol. 41, (2017) pp. 32–37.