NGHIÊN CỨU KHẢ NĂNG PHÁT HIỆN TẤN CÔNG TUYẾN TÍNH CỦA HAI PHƯƠNG PHÁP CHI-SQUARED VÀ CUSUM

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

Tóm tắt: *Bài báo này trình bày nghiên cứu về khả năng phát hiện tấn công tuyến tính của hai phương pháp Chi-squared (CHI2) và Cumulative sum (CUSUM) trong trường hợp phương pháp Kullback – Leibler (K-L) bị vượt qua. Đối tượng chịu tấn công tuyến tính là quá trình truyền tin không dây từ cảm biến lên thiết bị điều khiển của một hệ thống điều khiển với mô hình toán học giả lập. Các ma trận tấn công được tính toán để đảm bảo vượt qua phương pháp phát hiện K-L. Trên cơ sở các ma trận này, các ngưỡng phát hiện của phương pháp CHI2 và CUSUM được thử nghiệm lựa chọn để đánh giá khả năng phát hiện tấn công tuyến tính. Các kết quả mô phỏng cho thấy, tồn tại một dải ngưỡng thích hợp ở cả hai phương pháp CHI2 và CUSUM, mà trong dải đó phương pháp tấn công tuyến tính đã vượt qua phương pháp K-L sẽ bị phát hiện. Ngoài ra, nghiên cứu cũng đã chỉ ra rằng phương pháp CUSUM có xác suất phát hiện tấn công tuyến tính cao hơn so với phương pháp CHI2.*

Từ khóa: Tấn công tuyến tính; Phương pháp CHI2; Phương pháp CUSUM; Ngưỡng phát hiện; Đường đặc tính - ROC.

1. MỞ ĐẦU

Các hệ thống điều khiển giám sát và thu thập dữ liệu SCADA (Supervisory Control and Data Acquisition) nói riêng, hay tổng quát hơn là các hệ thống điều khiển công nghiệp, 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 có thể bị tấn công phối hợp không chỉ trên cơ sở hạ tầng vật chất mà còn trên lớp truyền thông và trung tâm điều khiển, với nhiều điểm tấn công khác nhau [1]. Vì vậy, vấn đề đảm bảo an toàn dữ liệu cho các hệ thống điều khiển công nghiệp đang được quan tâm lớn. Hiện nay, có hai hướng nghiên cứu chính về đảm bảo an toàn thông tin trong các hệ thống điều khiển công nghiệp. Hướng thứ nhất tập trung nâng cao khả năng phát hiện sai lệch dữ liệu trong cấp điều khiển, giám sát của các hệ thống điều khiển công nghiệp, sử dụng bộ dữ liệu mẫu mô phỏng các trường hợp bị tấn công điển hình [2]. Hướng thứ hai tập trung vào các thuật toán/ phương pháp phát hiện điểm bất thường của chuỗi dữ liệu truyền trong hệ thống. Các phương pháp có thể kể đến bao gồm CHI2, CUSUM, FSS (Fixed-Size Sample), FMA (Finite Moving Average) – áp dụng khi biết thông số kỳ vọng, phương sai của hệ thống khi có thay đổi bất thường; hoặc GLR (Generalized Likelihood Ratio), WLR (Weighted Likelihood Ratio) khi chưa biết kỳ vọng, phương sai của hệ thống khi có thay đổi bất thường [3].

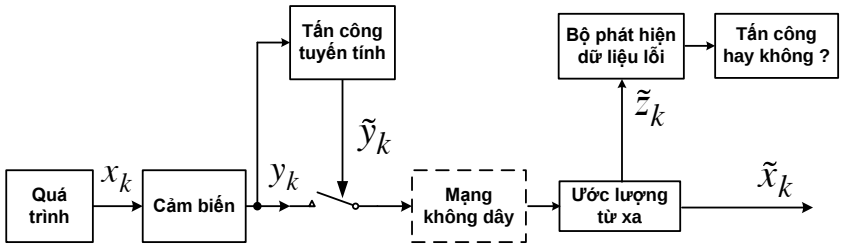
Theo hướng nghiên cứu thứ hai, hiện có thể phân loại một số phương pháp tấn công điển hình vào các hệ thống điều khiển công nghiệp, đó là tấn công tính toàn vẹn của dữ liệu và tấn công từ chối dịch vụ [1, 4]. Tấn công tính toàn vẹn dữ liệu nhằm vào dữ liệu truyền nhận giữa các lớp, hoặc trong các lớp mạng của hệ thống điều khiển, làm sai lệch thông tin hoặc chèn thông tin giả. Gần đây, có công bố về phương pháp tấn công tuyến tính của nhóm nghiên cứu tại Đại học Công nghệ Hồng Kông [5]. Đây là phương pháp tấn công vào tính toàn vẹn dữ liệu ở cấp hiện trường với độ nguy hiểm cao. Nhóm nghiên cứu đã chỉ ra rằng, thuật toán phát hiện tấn công K-L hoàn toàn có thể bị vượt qua với kiểu tấn công này [6].

Trong bài báo này, chúng tôi trình bày nghiên cứu khả năng áp dụng hai phương pháp CHI2 và CUSUM nhằm 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 CHI2 và CUSUM; Phần 4 phân tích một số trường hợp mà phương pháp K-L bị vượt qua bởi kiểu tấn công tuyến tính, xác định điều kiện để phát hiện được kiểu tấn công này khi áp dụng phương pháp CHI2 và CUSUM, đồng thời so sánh xác suất phát hiện tấn công của hai phương pháp này. Cuối cùng, phần 5 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 công nghiệp với điểm xảy ra tấn công tuyến tính được mô tả như hình 1, làm thay đổi dữ liệu truyền không dây tại đầu ra của các cảm biến [6].



Hình 1. *Sơ đồ minh họa vị trí xảy ra 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) [6]:

$$x_{k+1} = Ax_k + \omega_k;\ y_k = Cx_k + v_k \tag{1}$$

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;

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

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

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

$\tilde{x}_k, \hat{x}_k^-$ - Ước lượng trạng thái của bộ ước lượng từ xa khi bị tấn công, khi không bị tấn công;

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

$\bar{P}$ là ước lượng hiệp phương sai (biến trạng thái của hệ thống) ở trạng thái ổn định

Khi không bị tấn công, ước lượng sai lệch tín hiệu đầu ra của cảm biến có thể viết như trong (2) [1, 6] với $E\left[z_i, z_j^T\right]$ là kỳ vọng các thành phần phần dư z_k

$$z_k = y_k - C\hat{x}_k^-; z_k \sim N(0;\Sigma);\ \Sigma = C\bar{P}C^T + R; E\left[z_i, z_j^T\right] = 0\ \forall \mathrm{i} \neq \mathrm{j} \tag{2}$$

Khi bị tấn công, tín hiệu ra của cảm biến bị thay đổi như mô tả trong (3):

$$\tilde{y}_k = \tilde{z}_k + C\tilde{x}_k^- \tag{3}$$

Các tác giả trong [6] 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 khá cao, dựa trên nguyên tắc tính độ chênh *D* giữa hai chuỗi giá trị ngẫu nhiên $\tilde{z}_k$ và z_k như công thức (4) [6, 7]:

$$D\left(\tilde{z}_k \,\|\, z_k\right) = \int f_{\tilde{z}_k}(\chi) \log \frac{f_{\tilde{z}_k}(\chi)}{f_{z_k}(\chi)} d\chi \tag{4}$$

với $f_{\tilde{z}_k}(\chi) v\grave{a}\, f_{z_k}(\chi)$ là hàm mật độ của $\tilde{z}_k$ và z_k.

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 (5):

$$D\left(\tilde{z}_k \parallel z_k\right) \le \delta \rightarrow \text{không bị tấn công ; } D\left(\tilde{z}_k \parallel z_k\right) > \delta \rightarrow \text{ bị tấn công} \quad (5)$$

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

Theo [6], dưới tác động của tấn công tuyến tính, tín hiệu cảm biến y_k bị biến đổi thành $\tilde{y}_k$ thỏa mãn (3) và (6):

$$\tilde{z}_k = T_k z_k + b_k \quad (6)$$

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 các ma trận tấn công T_k, Γ_k thoả mãn (7) [6]:

$$\begin{cases} \min\limits_{(T_k)} \; Tr\left(C\bar{P}\bar{P}C^T\Sigma^{-1}T_k\right) \\ \begin{bmatrix} \tilde{\Sigma} & T_k \\ T_k^T & \Sigma^{-1} \end{bmatrix} \le 0 \end{cases} ; \; \Gamma_k = \tilde{\Sigma}_k - T_k \Sigma T_k^T \quad (7)$$

Trong đó, theo quy hoạch lồi Karush Kuhn Tucker, mối quan hệ giữa ngưỡng δ và μ thỏa mãn (8) [6]:

$$\mu\left(\frac{1}{2}Tr\left(\Sigma^{-1}\tilde{\Sigma}_k\right) - \frac{m}{2} + \frac{1}{2}\log\frac{|\Sigma|}{|\tilde{\Sigma}_k|} - \delta\right) = 0 \quad (8)$$

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 $K^TK\Sigma, K$ là ma trận hệ số Kalman, $Tr\left(\Sigma^{-1}\tilde{\Sigma}_k\right)$ là vết của ma trận $\Sigma^{-1}\tilde{\Sigma}_k$.

Như vậy, với mỗi ngưỡng δ của phương pháp K-L đều có thể tìm ra các ma trận tấn công tuyến tính T_k, Γ_k .

3. TỔNG QUAN PHƯƠNG PHÁP CHI-SQUARED VÀ CUSUM

3.1. Phương pháp CHI-SQUARED

Phương pháp CHI2 có khác biệt so với phương pháp K-L ở điểm là phương pháp này dùng dạng bình phương của chuỗi giá trị z_k để kiểm tra độ sai lệch đáng kể giữa hiệp phương sai và giá trị mong muốn của sai số giữa các đầu ra quan sát và đầu ra ước lượng như công thức (9) [1, 6, 8]:

$$\alpha_k = z_k^T\Sigma^{-1}z_k \quad (9)$$

Các giá trị tổng sai lệch g_k của phương pháp CHI2 được xác định theo nguyên tắc trong công thức (10):

$$g_k = \sum_{i=k-J+1}^{k}\alpha_k = \sum_{i=k-J+1}^{k} z_i^T\Sigma^{-1}z_i \quad (10)$$

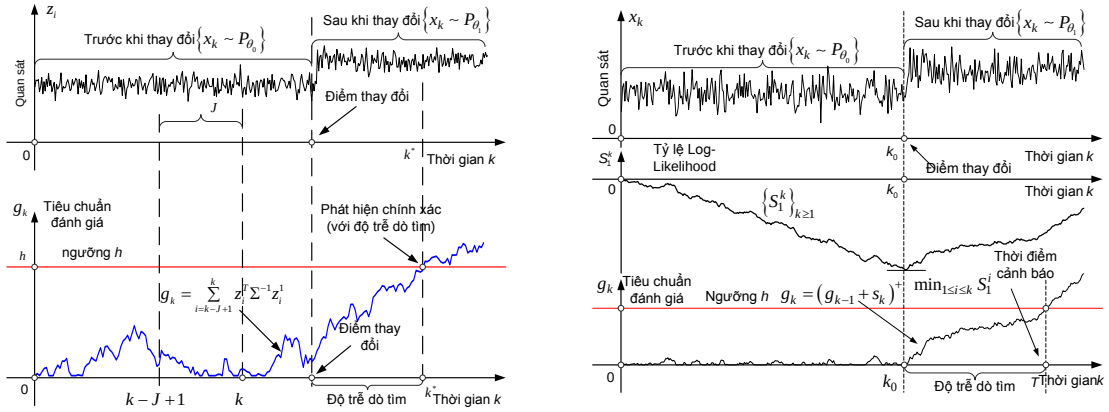
Khi tổng sai lệch vượt ngưỡng, dữ liệu được đánh giá là bị tấn công làm sai lệch giá trị, và ngược lại, như thể hiện trong (11) và hình 2a:

$$\begin{cases} g_k \le h \rightarrow & \text{không bị tấn công} \\ g_k > h \rightarrow & \text{bị tấn công} \end{cases} \quad (11)$$

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

$$T_a = \min\left(k : g_k \geq h\right) \tag{12}$$

trong đó, *h* là ngưỡng phát hiện tấn công (đặt trước) theo phương pháp CHI2.



a) Phương pháp CHI2 *b) Phương pháp CUSUM*

Hình 2. *Minh họa phương pháp CHI2 (a) và CUSUM (b) phát hiện dữ liệu bị tấn công.*

3.2. Phương pháp CUSUM

Trong [1], chúng tôi đã chỉ ra phương pháp CUSUM có thể phát hiện được tấn công tuyến tính trong trường hợp phương pháp K-L bị vượt qua. Mục này xin nêu lại một cách vắn lược về phương pháp CUSUM để thấy được điểm khác biệt giữa CHI2 và CUSUM. Phương pháp CUSUM có khác biệt so với phương pháp K-L và CHI2 ở điểm là phương pháp này áp dụng lý thuyết Wald phân tích tính bất thường trong dữ liệu [3, 9]. 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 $X \sim N\left(\mu, \Sigma_1\right)$. Phương pháp CUSUM xét tỷ lệ thay đổi thực sự S_i^k (likelihood ratio – LLR), như được minh họa ở hình 2b, xác định theo công thức (13) [1, 3, 9]. S_i^k 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.

$$S_i^k = \sum_{t=i}^{k} \ln \frac{f_{\theta_1}\left(x_t\right)}{f_{\theta_0}\left(x_t\right)}; s_k = \frac{1}{2}\ln\frac{\det\Sigma_0}{\det\Sigma_1} - \frac{1}{2}\left(x-\mu\right)^T\left[\Sigma_1^{-1} - \Sigma_0^{-1}\right]\left(x-\mu\right) \tag{13}$$

với $\det\Sigma_0, \det\Sigma_1$ là định thức của ma trận Σ_0, Σ_1, được tính như công thức (14):

$$\Sigma_0 = \Sigma = C\bar{P}C^T + R;\ \Sigma_1 = T_k \Sigma T_k^T + \Gamma_k \tag{14}$$

Các giá trị sai lệch g_k của phương pháp CUSUM được xác định theo nguyên tắc trong công thức (15) [1, 2]:

$$g_k = \begin{cases} g_{k-1} + s_k & \text{if } \ g_{k-1} + s_k > 0 \\ 0 & \text{if } \ g_{k-1} + s_k < 0 \end{cases} \tag{15}$$

và thời điểm cảnh báo tấn công T_a được xác định từ điều kiện (16) [1, 3]:

$$T_a = \min\left(k : g_k \geq h\right) \tag{16}$$

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

Như vậy, phương pháp CHI2 chỉ sử dụng dữ liệu trước thời điểm thay đổi bất thường, còn phương pháp CUSUM sử dụng dữ liệu trước và sau khi có thay đổi bất thường.

4. ĐỐI TƯỢNG, KẾT QUẢ MÔ PHỎNG VÀ THẢO LUẬN

Trong bài báo này, để kiểm tra khả năng phát hiện tấn công tuyến tính của các phương pháp CHI2 và CUSUM, chúng tôi thử nghiệm trên cùng mô hình của một hệ thống điều khiển quá trình với hai cảm biến, đã được các tác giả trong [6] công bố. Theo [6], ta có mô hình trạng thái không liên tục của đối tượng ở phương trình (1) với các thông số:

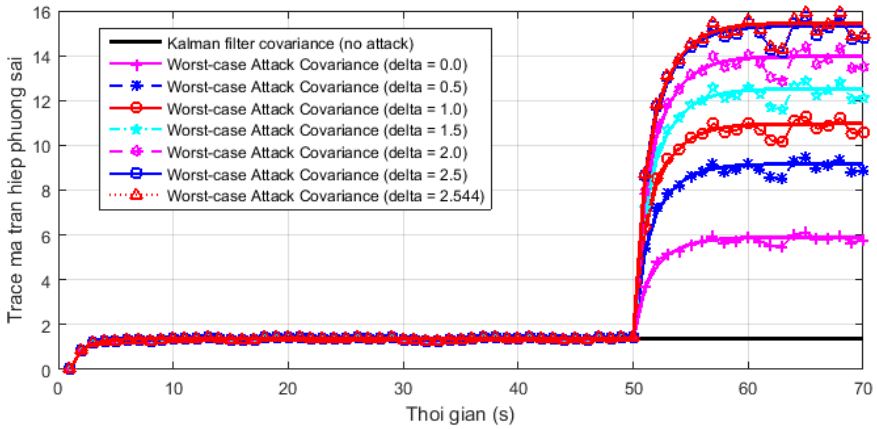
$$A=\begin{bmatrix}0.7 & 0.2\\ 0.05 & 0.64\end{bmatrix};C=\begin{bmatrix}0.5 & -0.8\\ 0 & 0.7\end{bmatrix};Q=\begin{bmatrix}0.5 & 0\\ 0 & 0.7\end{bmatrix};R=\begin{bmatrix}1 & 0\\ 0 & 0.8\end{bmatrix}$$

Chúng tôi thực hiện mô phỏng để đánh giá khả năng áp dụng phương pháp phương pháp CHI2 và CUSUM phát hiện tấn công tuyến tính vào đối tượng trên. Trong đó, điểm mấu chốt là đánh giá khả năng tồn tại ngưỡng *h* của hai phương pháp này để có thể phát hiện dữ liệu bị tấn công trong trường hợp phương pháp K-L bị vượt qua.

Công thức (8) cho biết mối quan hệ giữa δ và μ ở phương pháp K-L. Chọn các ngưỡng δ ta tính được μ. Căn cứ trên các phân tích trong [1], dải giá trị δ được chọn trong phạm vi $\delta\in[0;\ 2.544]$.

$$\delta=\begin{cases}0 & \equiv\delta_0\\ 0.5 & \equiv\delta_1\\ 1.0 & \equiv\delta_2\\ 1.5 & \equiv\delta_3\end{cases}\Rightarrow\mu=\begin{cases}\infty & \equiv\mu_0\\ 1.2923 & \equiv\mu_1\\ 1.1305 & \equiv\mu_2\\ 1.0638 & \equiv\mu_3\end{cases}\quad;\delta=\begin{cases}2.0 & \equiv\delta_4\\ 2.5 & \equiv\delta_5\\ 2.544 & \equiv\delta_6\end{cases}\Rightarrow\mu=\begin{cases}1.02627 & \equiv\mu_4\\ 1.0019 & \equiv\mu_5\\ 1.0 & \equiv\mu_6\end{cases}$$

Giải hệ tối ưu (7) bằng CVX toolbox trong matlab, và từ (13) ta thu được các ma trận tham số của tấn công tuyến tính $T_{k0}\div T_{k7};\Gamma_{k0}\div\Gamma_{k7}$ để nó vượt qua phương pháp phát hiện K-L. Sau đó, kiểm tra sự sai khác giữa tín hiệu khi không bị tấn công với khi bị tấn công tuyến tính, thông qua vết của các ma trận hiệp phương sai $Tr(\tilde{P}_k)$ như minh họa trong hình 3. Dữ liệu mô phỏng dài 70 giây, trong đó, 49s đầu tiên không có tấn công, từ giây 50 tới 70 mô phỏng có tấn công làm sai lệch dữ liệu [6].

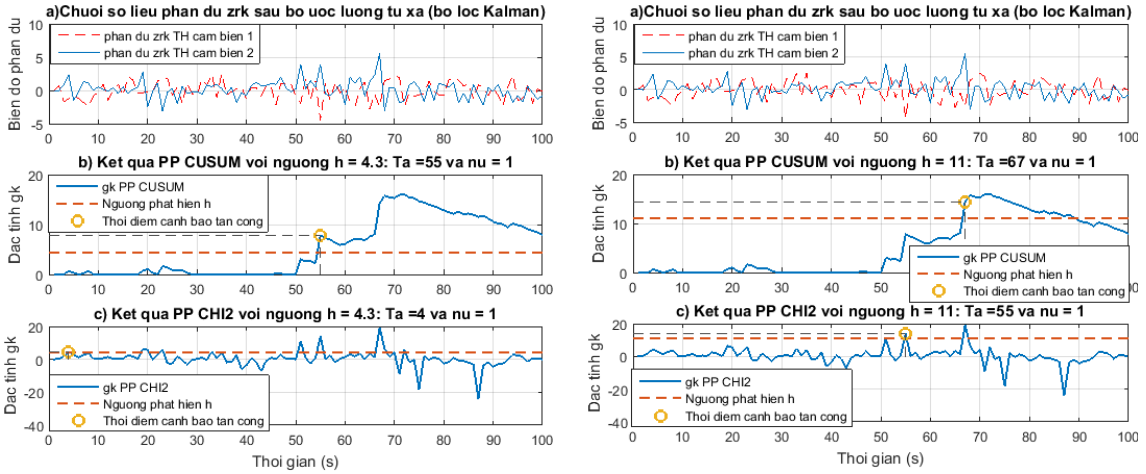


Hình 3. *Vết của ma trận hiệp phương sai khi bị tấn công tuyến tính.*

Kết quả mô phỏng trong hình 3 cho thấy, ban đầu khi chưa bị tấn công, trong khoảng thời gian [0; 49s], vết của ma trận hiệp phương sai xấp xỉ 1.38. 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. 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.

Nhằm xem xét khả năng áp dụng CHI2 và CUSUM, trước hết chúng tôi xét trường hợp tấn công tuyến tính vượt qua phương pháp K-L ở ngưỡng thấp $\delta = \delta_1 = 0.5$.

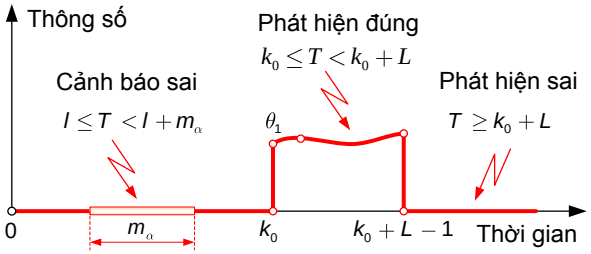


a) Ngưỡng h = 4.3 *b) ngưỡng h = 11*

Hình 4. *Khả năng phát hiện tấn công tuyến tính bằng phương pháp CUSUM, CHI2 khi K-L bị vượt qua với ngưỡng* $\delta = 0.5$.

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 50s đến 70s với tổng thời gian mô phỏng dài 70s. Chúng ta phân tích một số kết quả mô phỏng dưới đây. Tại ngưỡng phát hiện của hai phương pháp CHI2 và CUSUM h =4.3, áp dụng các công thức (8, 10, 11, 12, 14, 15, 16) tính thời điểm cảnh báo tấn công T_a ta có kết quả mô phỏng như trong hình 4a. Kết quả này 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 $T_{a\_CUSUM} = 55s$, $nu = 1$, (phát hiện đúng). Bên cạnh đó, phương pháp CHI2 đã cảnh báo sai do $nu = 1$, $T_{a\_CHI2} = 4s$, không nằm trong khoảng thời gian xảy ra tấn công.

Tương tự, với ngưỡng phát hiện h =*11*, ta thu được kết quả mô phỏng như trong hình 4b. Kết quả mô phỏng cho thấy, trong khoảng thời gian xảy ra tấn công tuyến tính đã giả lập từ 50s đến 70s, phương pháp CUSUM đã phát hiện ra loại tấn công này tại thời điểm $T_{a\_CUSUM} = 67s$, $nu = 1$, (phát hiện đúng). Trong khi đó, phương pháp CHI2 đã phát hiện ra loại tấn công này tại thời điểm $T_{a\_CHI2} = 55s$, $nu = 1$, (phát hiện đúng).



Hình 5. *Vấn đề phát hiện tấn công.*

Một cách tổng quát, để đánh giá khả năng phát hiện tấn công tuyến tính trong khoảng thời gian $[k_0 \;\; k_0+L)$ của mỗi phương pháp, chúng tôi sử dụng xác suất phát hiện đúng P_D; xác suất cảnh báo sai P_{FA} và xác suất phát hiện sai thời điểm P_{MD} (hình 5). Mối quan hệ giữa P_D và P_{FA} được biểu diễn trên đường cong đặc tính ROC.

Trong đó, phát hiện đúng là phát hiện được sự thay đổi bất thường trong đúng khoảng thời gian nó xảy ra. Xác suất phát hiện đúng được minh họa ở công thức (17):

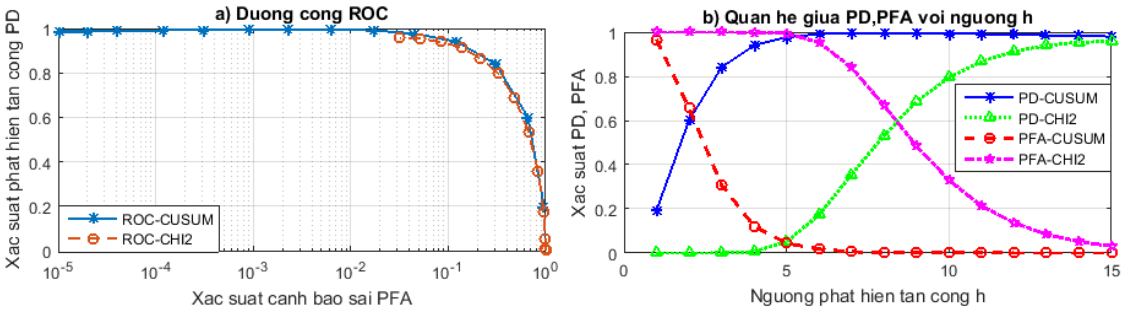
$$P_D = P\left(k_0 \le T \le k_0 + L - 1\right) \tag{17}$$

Cảnh báo sai: phát hiện sự thay đổi bất thường trước khi nó xảy ra $\left(T < k_0\right)$. Xác suất cảnh báo sai có thể được xác định trong khoảng thời gian tương ứng chiều dài m_α với thời điểm đầu cảnh báo sai *l* và ngưỡng α được minh họa trong công thức (18):

$$P_{FA} = P_{FA}\left(T, m_\alpha\right) = \sup_{l \ge 1} P_0\left[\left(l \le T \le l + m_\alpha\right) \le \alpha\right] \tag{18}$$

Theo [10], ước lượng Monte-Carlo của xác suất cảnh báo sai $\bar{P}_{FA}$ được tính bằng tỷ lệ giữa số lần cảnh báo sai trên tổng số lần mô phỏng *nS*; ước lượng Monte-Carlo của xác suất phát hiện đúng $\bar{P}_D$ được tính bằng tỷ lệ giữa số lần phát hiện đúng trên tổng số lần mô phỏng *nS*, được minh họa trong công thức (19):

$$\bar{P}_{FA} = \frac{1}{nS}\sum_{k=1}^{nS} nu(k), k \le k_0; \bar{P}_D = \frac{1}{nS}\sum_{k=1}^{nS} nu(k); k_0 \le k \le k_0 + L - 1 \tag{19}$$

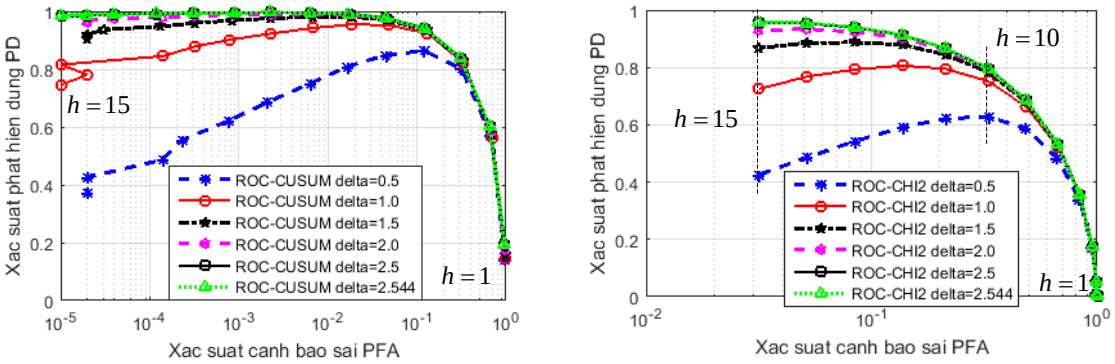


Hình 6. *Đánh giá khả năng phát hiện tấn công tuyến tính bằng phương pháp CUSUM, CHI2 với ngưỡng h =1÷ 15 khi K-L bị vượt qua với ngưỡng* $\delta = 2.544$.

Xét trường hợp khi K-L bị vượt qua ở ngưỡng $\delta = 2.544$, các ngưỡng phát hiện của CHI2 và CUSUM chọn trong dải $h = 1 \div 15$, áp dụng công thức (19) với đối tượng mô phỏng, ta có một số kết quả như thể hiện trong hình 6. Hình 6b cho thấy, với cùng ngưỡng *h*, phương pháp CUSUM có xác suất phát hiện đúng $\bar{P}_D$ cao hơn, xác suất cảnh báo sai $\bar{P}_{FA}$ thấp hơn so với phương pháp CHI2. Kết quả mô phỏng cho thấy, khi ngưỡng *h* tăng lên thì xác suất phát hiện đúng $\bar{P}_D$ của cả hai phương pháp tăng lên, còn xác suất cảnh báo sai $\bar{P}_{FA}$ giảm dần. Hay nói cách khác, trong trường hợp này, phương pháp CUSUM có khả năng phát hiện tấn công tuyến tính tốt hơn phương pháp CHI2.

Tiến hành tương tự cho các ngưỡng δ (của phương pháp K-L) nhỏ hơn, ta thu được các đường đặc tính tương tự hình 6 và tập hợp các đường cong đặc tính ROC của phương pháp CUSUM (hình 7a), và của phương pháp CHI2 (hình 7b). So sánh hình 7a và 7b cho thấy, với cùng điều kiện thử nghiệm, xác xuất cảnh báo sai của phương pháp CHI2 cao hơn so với phương pháp CUSUM. Hình 7 đồng thời cho thấy, tấn công tuyến tính khi vượt qua K-L ở ngưỡng δ

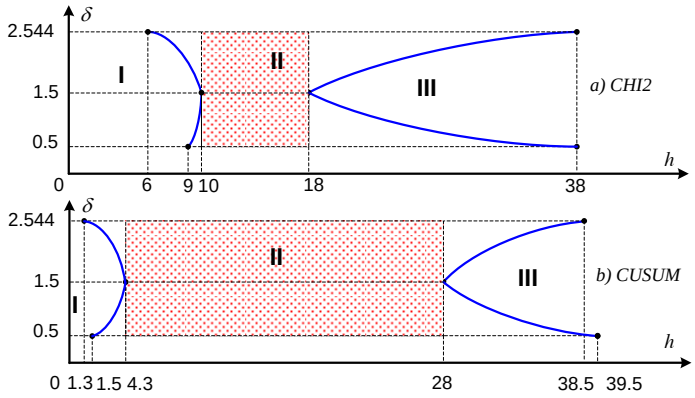
lớn, lại dễ dàng bị phát hiện bởi các phương pháp CHI2 và CUSUM hơn (thể hiện qua xác suất phát hiện đúng lớn và xác xuất phát hiện sai nhỏ ở cả hai phương pháp CHI2 và CUSUM).



a) Phương pháp CUSUM *b) Phương pháp CHI2*

Hình 7. *Đánh giá hiệu suất phát hiện tấn công tuyến tính với các ngưỡng δ.*

Tương ứng với mỗi giá trị ngưỡng δ của phương pháp 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 hay CHI2 (hình 8). Kết quả trong hình 8 thể hiện khả năng thực tế có thể chọn một ngưỡng *h* thích hợp để CUSUM hay CHI2 phát hiện được tấn công tuyến tính trong mọi trường hợp K-L bị vượt qua. Dải ngưỡng *h* ở CUSUM rộng hơn so với dải ở CHI2, thể hiện phương pháp CUSUM tốt hơn.



Hình 8. *Khả năng phát hiện tấn công tuyến tính bằng phương pháp CHI2 và CUSUM.*

5. KẾT LUẬN VÀ HƯỚNG NGHIÊN CỨU TIẾP THEO

Các phân tích trên đã chỉ ra rằng, tồn tại một khoảng giá trị ngưỡng *h* trong phương pháp CUSUM và CHI2 mà với các giá trị đó thì có thể phát hiện được tấn công tuyến tính, khi nó vượt qua phương pháp độ chênh K – L. Các kết quả mô phỏng cũng đồng thời chỉ ra rằng, có thể sử dụng phương pháp CUSUM hay CHI2 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. Ngoài ra, phân tích các kết quả mô phỏng cũng cho thấy khả năng phát hiện tấn công tuyến tính của phương pháp CUSUM tốt hơn so với phương pháp CHI2.

Trong các nghiên cứu tiếp theo, chúng tôi sẽ triển khai phương pháp CUSUM, CHI-SQUARED mở rộng và nghiên cứu thêm 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]. Nguyễn Đức Dương, Lê Minh Thùy, Cung Thành Long *"Nghiên cứu khả năng phát hiện tấn công tuyến tính trong các hệ thống điều khiển công nghiệp bằng phương pháp CUSUM"*, Tạp chí khoa học công nghệ ĐHBKHN, số 145, 2020, tr 14-20.

[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, 2020, pp. 55-62.

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

[4]. Derui Ding, Qing-Long Han, Yang Xiang, Xiaohua Ge, Xian-Ming Zhang, *"A survey on security control and attack detection for industrial cyber-physical systems,*" Neurocomputing, Volume 275, 2018, Pages 1674-1683.

[5]. 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.

[6]. Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "*Worst-case stealthy innovation-based linear attack on remote state estimation*" Automatica 89 (2018) pp. 117-124.

[7]. S. KULLBACK (1959). "*Information Theory and Statistics"*. Wiley, New York (also Dover, New York, 1968).

[8]. Nikulin, M. S. (1973), "*Chi-squared test for normality*"

[9]. R.H.WOODWARD and P.L. GOLDSMITH (1964). *Cumulative Sum Techniques*. Oliver and Boyd, Edinburgh, UK.

[10]. C.Parloir, M.Kinnaert *"Performance evaluation of fault detection algorithms by Monte Carlo methos",* IPAC System, Structure and Control Oaxaca, Mexico, USA, 8-10 December 2004.

ABSTRACT

RESEARCH ON THE ABILITY TO DETECT THE LINEAR ATTACK OF THE CHI-SQUARED METHOD AND THE CUSUM METHOD

This paper presents the ability to detect linear attacks of the Chi-squared (CHI2) and the Cumulative Sum (CUSUM) methods, in case the Kullback–Leibler (K-L) method cannot detect. The object, which is attacked by the linear attack, is the wireless communication process from sensors to controller with a simulated mathematical model. The attack matrices are calculated to ensure that the K-L method cannot detect. With these matrices, the detection thresholds of CHI2 method and CUSUM method are chosen and tested to estimate the ability to detect the linear attack. Simulated results show that an appropriate range of threshold of the CHI2 and the CUSUM methods can be chosen to detect the linear attack in case the K-L method cannot detect. In addition, the obtained results also show that the detection ability of the CUSUM method is better than that of the CHI2 method.

Keywords: Linear attack; Chi-squared method; CUSUM method; detection threhold; Receiver Operating Characteristic – ROC.

Nhận bài ngày 29 tháng 12 năm 2020
Hoàn thiện ngày 31 tháng 3 năm 2021
Chấp nhận đăng ngày 10 tháng 6 năm 2021

Địa chỉ: [1]Bộ môn Kỹ thuật đo và Tin học công nghiệp, Viện Điện, Đại học Bách khoa Hà Nội;
[2]Khoa Điện, Đại học Kinh tế Kỹ thuật công nghiệp.
*Email: long.cungthanh@hust.edu.vn.