ISSN: 1859-2171
e-ISSN: 2615-9562
TNU Journal of Science and Technology
208(15): 27 - 34

ỨNG DỤNG MÁY VECTƠ HỖ TRỢ VÀ BẤT THƯỜNG TRONG NGỮ CẢNH CHO PHÁT HIỆN XÂM NHẬP VÀO HỆ THỐNG SCADA

Nguyễn Văn Xuân*, Vũ Đức Trường, Nguyễn Mạnh Hùng, Nguyễn Tăng Cường
Học viện Kỹ thuật quân sự

TÓM TẮT

Trong bài báo này, chúng tôi giới thiệu một mô hình IDS-SCADA, có khả năng phát hiện xâm nhập vào hệ thống SCADA với độ chính xác cao, mô hình này được xây dựng dựa trên máy học Support Vector Machine (SVM). Điểm đặc biệt của mô hình được đề xuất ở chỗ chúng tôi xem xét dữ liệu bất thường trong ngữ cảnh. Để làm điều đó, tập dữ liệu ban đầu được chúng tôi cấu trúc lại để tạo ngữ cảnh trước khi đưa vào SVM huấn luyện. Mô hình được chúng tôi đề xuất có khả năng phát hiện dữ liệu tấn công hay bình thường với độ chính xác đạt từ 95,02% đến 99,03%.

Từ khóa: *Phát hiện xâm nhập, Máy học, IDS, SVM, SCADA*.

Ngày nhận bài: 27/8/2019; Ngày hoàn thiện: 22/9/2019; Ngày đăng: 03/10/2019

APPLICATION OF SUPPORT VECTOR MACHINE AND CONTEXTUAL OUTLIERS FOR INTRUSION DETECTION IN THE SCADA SYSTEM

Nguyen Van Xuan*, Vu Duc Truong, Nguyen Manh Hung, Nguyen Tang Cuong
Military Technical Academy

ABSTRACT

In this paper, we present an IDA-SCADA model based on Support Vector Machine (SVM) which is capable of detecting intrusion into SCADA systems with high accuracy. The distinction of our method used in this research is we applied contextual training data. To do that, the original dataset was reorganized to create context before training the SVM phase. The result of our work is the proposed system able to identify any attacks or normal patterns with precision from 95.02% to 99.03%.

Keywords: *Intrusion detection system, Machine Learning, IDS, SVM, SCADA.*

Received: 27/8/2019; Revised: 22/9/2019; Published: 03/10/2019

* Corresponding author. *Email: xuannv8171@gmail.com*

1. Giới thiệu

Hệ thống SCADA (Supervisory Control and Data Acquisition) quan trọng tầm quốc gia hoặc của các danh nghiệp lớn luôn có nguy cơ bị tấn công từ các mã độc hại, Hacker, tin tặc, từ các nhà thầu cạnh tranh nhau, từ khủng bố,...Ví dụ năm 2000, các trạm bơm dịch vụ nước Maroochy ở Úc bị tấn công làm dừng hệ thống [1]. Năm 2003, một sâu máy tính vượt qua tường lửa xâm nhập vào hệ thống SCADA tại nhà máy hạt nhân Davis Besse ở Ohio [2]. Năm 2010, Stuxnet [3] tấn công vào nhà máy hạt nhân Iran, sâu Stuxnet đã cảnh báo cho cả thế giới mức độ nghiêm trọng của các lỗ hổng đe dọa đến hệ thống SCADA.

Bản chất của hệ thống IT (Information Technology) và hệ thống điều khiển công nghiệp, hệ thống SCADA là khác nhau. Vì vậy các hệ thống phát hiện xâm nhập IDS (Intrusion detection system) áp dụng cho các hệ thống IT có thể không hoàn toàn phù hợp với hệ thống SCADA.

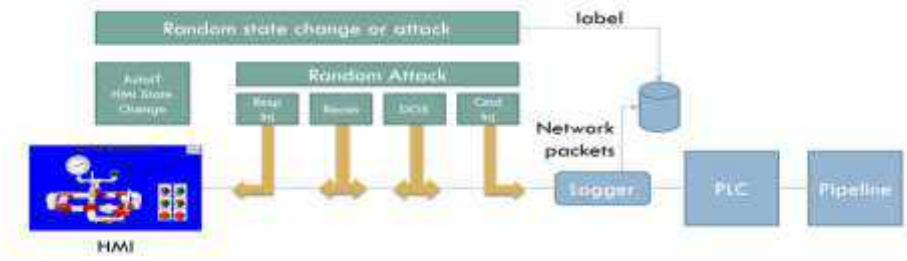
Trong bài báo này chúng tôi nghiên cứu đề xuất mô hình IDS – SCADA trên cơ sở máy học SVM (Support Vector Machine) và bất thường trong ngữ cảnh, cho phép phát hiện xâm nhập vào hệ thống SCADA và nâng cao tỷ lệ phát hiện xâm nhập và giảm thiểu các cảnh báo giả.

Có ba kiểu dữ liệu bất thường: điểm bất thường, bất thường tập thể và bất thường trong ngữ cảnh. Khi một trường hợp dữ liệu cụ thể không tuân theo phần dữ liệu chung của nó gọi là điểm dữ liệu bất thường. Khi một tập hợp dữ liệu tương tự nhau đang hoạt động bất thường thì toàn bộ tập hợp dữ liệu đó gọi là bất thường tập thể. Kiểu thứ 3, bất thường trong ngữ cảnh xảy ra khi một trường hợp dữ liệu xem xét là bình thường hay bất thường cần đặt nó trong một mối quan hệ cụ thể. Ví dụ chi tiêu hàng tháng là 500$ nếu có một tháng chi tiêu 2000$ nhưng tháng đó có lễ hội thì chi tiêu đó là bình thường, còn tháng đó không phải dịp đặc biệt nào thì dữ liệu chi tiêu đó là bất thường.

2. Bộ dữ liệu sử dụng trong huấn luyện, kiểm tra

Đối với hệ thống IT, có bộ dữ liệu KDD [4] cho các nhà nghiên cứu thử nghiệm mức độ hiệu quả của các IDS mà họ nghiên cứu. Với hệ thống SCADA, Wei Gao và cộng sự [5] đã nghiên cứu và công bố bộ dữ liệu phiên bản đầu tiên cho hệ thống SCADA đường ống dẫn GAS. Sau đó Thornton và cộng sự [6] đã chỉ ra còn một số nhược điểm của bộ dữ liệu này. Tiếp sau đến Turnipseed [7] đã kế thừa hệ thống của Wei Gao và công bố bộ dữ liệu phiên bản thứ hai với các mẫu tấn công đảm bảo ngẫu nhiên hơn, phù hợp cho thử nghiệm các thuật toán khác nhau trong IDS – SCADA. Bộ dữ liệu đó được mô tả ở phần dưới đây, hình 1 là kiến trúc hệ thống tạo ra tập dữ liệu của Turnipseed.

Bộ dữ liệu kiểm tra IDS – SCADA của Turnipseed được xây dựng cho hệ thống đường ống GAS sử dụng giao thức MODBUS (chi tiết bộ dữ liệu xem tại [7]) gồm có 274628 mẫu, trong đó có 214580 mẫu bình thường (chiếm 78,1%) và 60048 mẫu tấn công (chiếm 21,9%). Và kết quả thử nghiệm một số thuật toán của Turnipseed và cộng sự trong bảng 1.



Hình 1. *Kiến trúc của test bed của tập dữ liệu*

Bảng 1. *Kết quả thử nghiệm các thuật toán của nhóm tác giả trên bộ dữ liệu*

Thuật toán	Nhóm thuật toán	Độ chính xác phân loại
Naïve Bayesian Network	Bayes	80.39%
PART	Rule-Based	94.14%
Multilayer Perceptron	Neural Network	85.22%

Mỗi mẫu dữ liệu tấn công và mẫu bình thường đều chứa 17 thuộc tính và 3 thuộc tính đầu ra được mô tả như bảng 2 dưới đây:

Bảng 2. *Các thuộc tính của mỗi mẫu trong tập dữ liệu*

STT	Thuộc tính	Mô tả
01	Address	Địa chỉ của Slave của giao thức Modbus
02	Function	Mã hàm của giao thức Modbus
03	Length	Độ dài của gói Modbus
04	Setpoint	Điểm đặt áp suất khi hệ thống ở chế độ tự động
05	Gain	PID gain.
06	Reset rate	PID reset rate.
07	Deadband	PID dead band
08	Cycle time	PID cycle time
09	Rate	PID rate
10	System mode	Chế độ của hệ thống, 2: auto, 1: manual, 0: off
11	Control scheme	0: điều khiển máy bơn, 1: điều khiển van từ
12	Pump	Điều khiển máy bơm, 1:on, 0:off
13	Solenoid	Điều khiển van từ, 1: opened , 0: closed
14	Pressure measurement	Giá trị áp suất đo được trong đường ống
15	CRC	Mã kiểm lỗi của gói Modbus
16	Command/response	1: Lệnh, 0: đáp ứng
17	Time	Dấu thời gian cho mỗi gói Modbus
18	Binary result	Phân nhóm nhị phân, 0:normal, 1:attack
19	Attack Categorized	Phân nhóm tấn công (0->7)
20	Specific result	Kết quả chi tiết các tấn công (0->35)

Tập dữ liệu có chứa 35 loại tấn công thuộc 7 nhóm mô tả tương ứng trong bảng 3.

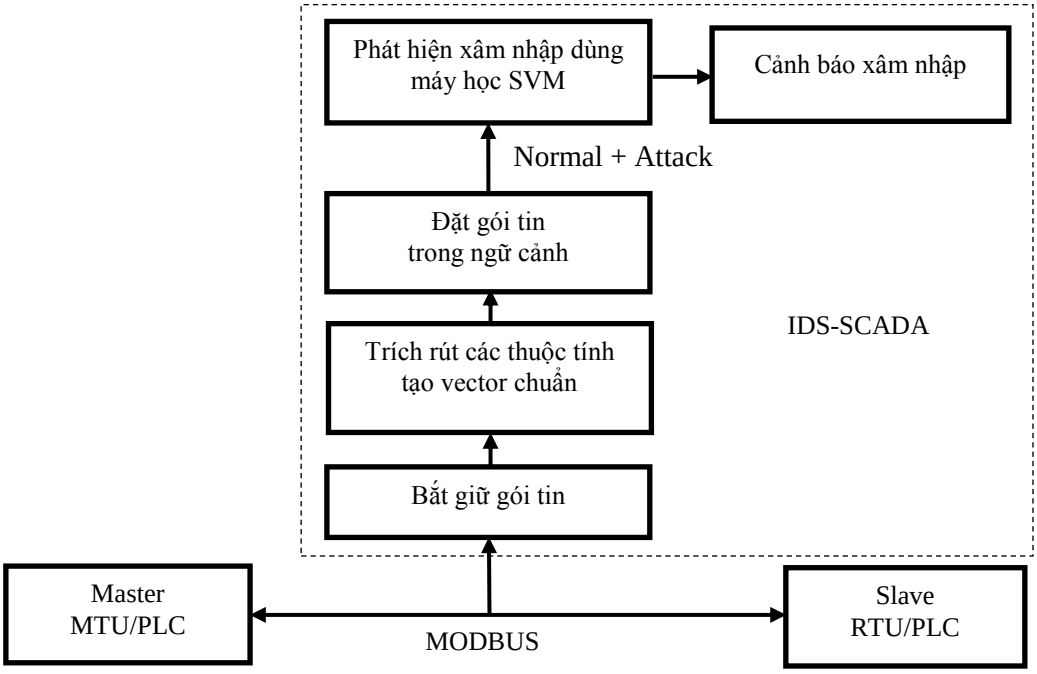
Bảng 3. *Bẩy nhóm tấn công khác nhau của tập dữ liệu*

Nhóm tấn công	Viết tắt
Normal/ Mẫu bình thường.	Normal(0)
Naïve Malicious Response Injection/Tấn công chèn đáp ứng đơn giản.	NMRI(1)
Complex Malicious Response Injection/Tấn công chèn đáp ứng tinh vi.	CMRI(2)
Malicious State Command Injection/Tấn công thay đổi trạng thái.	MSCI(3)
Malicious Parameter Command Injection/Tấn công thay đổi tham số	MPCI(4)
Malicious Function Code Injection/Tấn công giả mạo mã hàm.	MFCI(5)
Denial of Service/Tấn công từ chối dịch vụ.	DoS(6)
Reconnaissance/Tấn công trinh sát.	Recon(7)

3. Mô hình đề xuất phát hiện tấn công vào hệ thống SCADA

Trong hầu hết các phương pháp xây dựng hệ thống IDS thì kỹ thuật phát hiện xâm nhập (tấn công) đều dựa trên các dấu hiệu xâm nhập hoặc trên phát hiện bất thường (xem thêm mục 4.1 trong bài báo này). Hình 2 là mô hình phát hiện xâm nhập vào hệ thống SCADA được đề xuất trong bài báo này. Ở đây chúng tôi đề xuất kỹ thuật dùng máy học SVM phát hiện bất thường trong ngữ cảnh để nhận dạng một gói tin là tấn công hay bình thường. Ngữ cảnh ở đây là chúng tôi không đưa độc lập từng gói tin vào máy học SVM mà cần xem xét một nhóm gói tin liên tiếp nhau đưa vào SVM huấn luyện, nhận dạng. Nghĩa là đặt mỗi gói tin nhận dạng trong ngữ cảnh gồm 3, 5, 7 gói tin bình thường ngay trước gói tin cần nhận dạng, sau đó mới đưa vào máy học SVM nhận dạng, kết luận là bình thường hay tấn công. Trong bài báo chọn ngữ cảnh gồm 3, 5

hay 7 gói tin để thử nghiệm vì nếu chọn ngữ cảnh chỉ có 1 hoặc 2 gói tin thì ngữ cảnh tạo ra có quá ít thông tin cho máy học SVM học tập, còn nếu chọn ngữ cảnh lớn hơn 7 gói tin thì có thể có quá nhiều thuộc tính để máy học SVM học tập dẫn đến quá trình học không hiệu quả.



Hình 2. *Mô hình phát hiện xâm nhập dựa trên máy học SVM và ngữ cảnh*

4. Máy học Support Vector Machine-SVM

4.1 Sử dụng máy học trong IDS

Một trong những phương pháp sử dụng đầu tiên trong IDS (Intrusion detection system) dựa trên quy tắc là hệ chuyên gia (Expert System - ES) [10], trong những hệ thống như vậy kiến thức, kinh nghiệm của con người được mã hóa thành bộ các quy tắc. Hệ chuyên gia cho phép quản lý các kiến thức, kinh nghiệm của con người hiệu quả, nhất quán, đầy đủ, cho phép xác định các hoạt động bình thường hay hoạt động lạm dụng vào hệ thống, tuy nhiên hệ chuyên gia có tính linh hoạt không cao, khó phát hiện các tấn công mới. Không giống hệ chuyên gia, cách tiếp cận khai phá dữ liệu (Data Mining), xuất phát từ sự kết hợp giữa các quy tắc và các mẫu dữ liệu có sẵn, không sử dụng kiến thức chuyên gia từ con người. Nó sử dụng các kỹ thuật thống kê để khai phá các mối quan hệ giữa các mục dữ liệu từ đó xây dựng các mô hình dự đoán. Sử dụng phương pháp này, Lee [11] đã phát triển một khung khai phá dữ liệu cho phát hiện xâm nhập. Cụ thể, các hành vi trong hệ thống được ghi lại và phân tích để tạo ra bộ các quy tắc, từ đó có thể nhận ra các cuộc xâm nhập trái phép vào hệ thống. Hạn chế của giải pháp này là có xu hướng tạo ra một số lượng lớn các quy tắc và làm tăng sự phức tạp của hệ thống. Cây quyết định là một trong những thuật toán học có giám sát được sử dụng phổ biến nhất trong IDS [12] do tính đơn giản, độ chính xác phát hiện cao và khả năng thích ứng nhanh. Một phương pháp khác cho hiệu suất khá cao là mạng nơron nhân tạo. Mạng nơron có thể mô hình hóa cả mô hình tuyến tính và phi tuyến tính. IDS dựa trên mạng nơron [13] đã đạt được thành công lớn trong việc phát hiện các cuộc tấn công mới và khó. Để phát hiện xâm nhập dựa trên các luật học không giám sát, các phương pháp phân cụm dữ liệu cũng được áp dụng [14]. Các phương pháp này liên quan đến việc tính toán khoảng cách bằng số giữa các thuộc tính, do đó chúng không dễ dàng xử lý các thuộc

tính dạng ký tự tượng trưng, dẫn đến khó chính xác. Một kỹ thuật nổi tiếng khác được sử dụng trong IDS là phân loại Naïve Bayes [12]. Bởi vì Naïve Bayes phải giả định tính độc lập có điều kiện của các thuộc tính dữ liệu nên trường hợp các thuộc tính có nhiều quan hệ với nhau thường làm cho hiệu suất phát hiện giảm. Bên cạnh Cây quyết định, và mạng nơron được sử dụng phổ biến, Support Vector Machines (SVM) cũng là một phương pháp tốt cho hệ thống phát hiện xâm nhập [15], SVM có khả năng phát hiện thời gian thực, xử lý dữ liệu có chiều lớn. SVM chuyển các vectơ huấn luyện vào trong không gian đặc trưng với số chiều lớn hơn thông qua các hàm ánh xạ phi tuyến. Dữ liệu sau đó được phân loại bằng cách xác định một tập các vectơ hỗ trợ, là tập con các dữ liệu đầu vào huấn luyện, sau đó xác định siêu phẳng trong không gian đặc trưng để phân loại.

4.2 Máy học Support Vector Machine

Mô hình phân loại Support Vector Machine (SVM) [8,9] được biết đến như một thuật toán học tập tốt nhất để phân loại nhị phân. SVM ban đầu là một thuật toán phân loại mẫu dựa trên kỹ thuật học thống kê để phân loại với nhiều hàm nhân (kernel functions), nó đã được áp dụng tốt cho một số ứng dụng nhận dạng mẫu. Gần đây, nó cũng đã được áp dụng cho phát hiện xâm nhập. SVM đã trở thành một trong những kỹ thuật phổ biến để phát hiện xâm nhập bất thường do tính chất khái quát tốt trong phân loại dữ liệu và hoạt động tốt với những dữ liệu có chiều lớn. Một điểm lợi thế khác của SVM là quá trình huấn luyện cho nghiệm tối ưu toàn cục không bị hội tụ đến nghiệm địa phương như mạng nơron dù chiều của dữ liệu lớn, số mẫu huấn luyện nhỏ. SVM có thể lựa chọn phương pháp thiết lập các tham số không phụ thuộc vào những kinh nghiệm, thực nghiệm như truyền thống của mạng nơron [16]. Một trong những lợi thế chính của việc sử dụng SVM cho IDS là tốc độ nhận dạng nhanh, vì khả năng phát hiện sự xâm nhập trong thời gian thực là rất quan trọng. SVM có thể học từ một tập các mẫu lớn và có khả năng mở rộng tốt vì độ phức tạp phân loại không phụ thuộc vào chiều của không gian đặc trưng. Các SVM cũng có khả năng cập nhật các mẫu huấn luyện một cách linh hoạt bất cứ khi nào có mẫu mới trong quá trình phân loại [17].

5. Cấu trúc lại tập dữ liệu để tạo ngữ cảnh

Để kết luận một gói tin trong mạng SCADA là bình thường hay tấn công ta xem xét nó trong quan hệ gồm có (k+1) gói tin liên tiếp nhau, k gói tin đầu là bình thường gọi là ngữ cảnh, gói tin cuối thứ (k+1) cần kết luận là gói bình thường hay tấn công. vì vậy ta cần cấu trúc lại tập dữ ban đầu mà mỗi bản ghi gồm k gói tin bình thường cùng gói tin (k+1) cần xem xét là gói bình thường hay tấn công, quá trình xây dựng lại tập dữ liệu như sau:

Gọi Wi (i=1,2,…N) là bản ghi (gói tin) trong tập dữ liệu ban đầu, N số bản ghi trong tập dữ liệu ban đầu.

Ti: Đầu ra phân loại của gói tin Wi, Ti = 0 nghĩa là gói Wi bình thường, Ti = 1 nghĩa là gói Wi là tấn công (gói tin xâm nhập trái phép).

W: Ngữ cảnh gồm k bản ghi bình thường, k có thể chọn = 3, 5, 7...

Pi: Bản ghi mới gồm k gói tin bình thường của W, gói tin Wi+k và đầu ra Ti+k của gói tin Wi+k; Pi=[W, Wi+k, Ti+k]

P: Tập dữ liệu mới gồm (N-k) bản ghi, mỗi bản ghi có (k+1) gói tin cũ.

Bước 1: Khởi tạo: i = 1, P = []- tập rỗng và ngữ cảnh W gồm k gói tin bình thường đầu tiên trong tập dữ liệu ban đầu, không mất tính tổng quát giả sử k gói tin đầu tiên liên tiếp của tập dữ liệu đầu là các gói tin bình thường thì ta có W như sau: W=[Wi, Wi+1, Wi+2,….,Wi+k-1].

Bước 2: Pi gói tin mới được gán gồm k gói tin bình thường trong W, cùng gói tin Wi+k, đầu ra Ti+k của Wi+k; Pi = [W, Wi+k, Ti]

Bước 3: Cập nhật lại ngữ cảnh W.

Nếu Ti+k =0 tức gói Wi+k là bình thường, cập nhật gói tin Wi+k vào W và gỡ

bỏ gói tin cũ bên trái cùng trong W ra, W được cập nhật lại là: W = [Wi+1, Wi+2,…,Wi+k]

Nếu Ti+k =1 tức Wi+k là gói tấn công không cập nhật Wi+k vào W, ngữ cảnh W không thay đổi.

Bước 4: Cập nhật Pi vào tập dữ liệu mới, P = [P; Pi], i = i+1, Nếu i <= N tiếp tục thực hiện bước 2, ngược lại kết thúc thuật toán.

Trong tập dữ liệu ban đầu mỗi bản ghi chỉ gồm các gói tin độc lập chưa có ngữ cảnh cho các gói tin, với thuật toán ở trên thì từ tập dữ liệu ban đầu đã tạo ra tập dữ liệu mới P gồm (N-k) bản ghi mà mỗi bản ghi trong tập P mới gồm (k+1) gói tin liên tiếp nhau lấy trong tập dữ liệu cũ, tức mỗi bản ghi trong tập P là một ngữ cảnh cho các gói tin cần nhận dạng.

6. Kết quả phân loại

Sau khi tạo ra tập dữ liệu mới P, chọn ngẫu nhiên 80% dữ liệu trong tập P (gồm 219.698 bản ghi) được dùng để huấn luyện máy học SVM, phần còn lại 20% dữ liệu của tập P (gồm 54.925 bản ghi) được sử dụng để kiểm tra lại hiệu suất phát hiện của SVM. Kết quả kiểm tra như sau:

Trường hợp k=3 cho kết quả như hình 3:

Độ chính xác phân loại:

(42762 + 9429)/54925 = 95,02%.

Độ chính xác phát hiện tấn công:

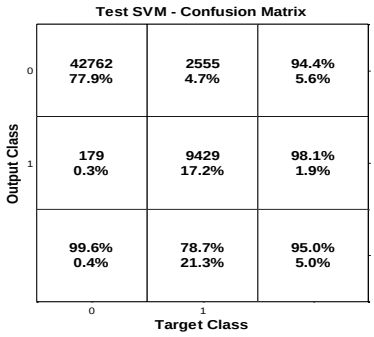
9429/(9429 + 179) = 98,14%

Tỷ lệ phát hiện tấn công (Recall):

9429/(9429 + 2555) = 78,68%

Cảnh báo nhầm (Dương tính giả):

179/(9429 + 179) = 1,86%



Hình 3. *Kết quả kiểm tra với k=3*

Trường hợp k=5 cho kết quả như hình 4:

Độ chính xác phân loại:

(42597 + 11796)/54925 = 99,03%.

Độ chính xác phát hiện tấn công:

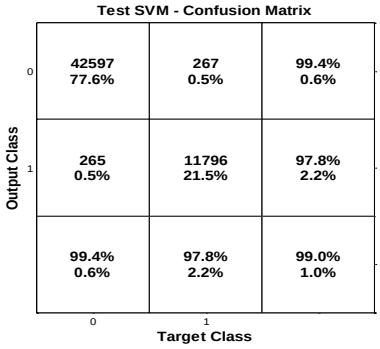
11796/(11796 + 265) = 97,80%

Tỷ lệ phát hiện tấn công (Recall):

11796/(11796 + 267) = 97,79%

Cảnh báo nhầm (Dương tính giả):

265/(11796 + 265) = 2,2%



Hình 4. *Kết quả kiểm tra với k=5*

Trường hợp k=7 cho kết quả như hình 5:

Độ chính xác phân loại:

(42661 + 11730)/54924 = 99,03%.

Độ chính xác phát hiện tấn công:

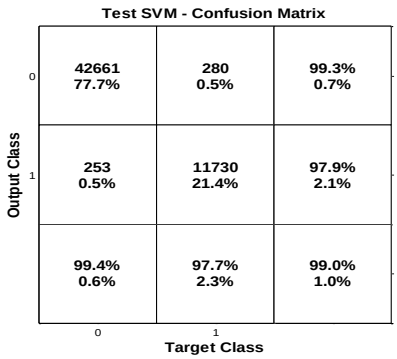
11730/(11730 +253) = 97,89%

Tỷ lệ phát hiện tấn công (Recall):

11730/(11730 + 280) = 97,67%

Cảnh báo nhầm (Dương tính giả):

253/(11730 +253) = 2,11%



Hình 5. *Kết quả kiểm tra với k=7*

Nhận xét: So sánh kết quả trong bảng 4 và trong bảng 1 của Turnipseed [7] cho thấy kết

quả nhận dạng của chúng tôi cao hơn nhiều của Turnipseed. Lấy một trường hợp tấn công chèn đáp ứng hoặc chèn lệnh tinh vi giải thích cho kết quả này. Gói tin 1 là một gói tin bình thường và gói tin 2 được tin tặc chèn vào mạng giống hệt gói tin 1 chỉ khác là ở hai thời điểm khác nhau nếu chỉ xem xét độc lập từng gói tin thì SVM không thể phát hiện ra gói tin nào là tấn công, gói tin nào bình thường được. Nhưng nếu xét thêm một số gói tin ngay trước gói 1 và cả gói tin 2 cũng làm vậy thì có thể phân biệt được gói tin 1 là bình thường, gói tin 2 là tấn công đó chính là một ví dụ tìm bất thường trong ngữ cảnh.

Bảng 4. *Kết quả phân loại tấn công*

Chỉ số đánh giá	k=3	k=5	k=7
Độ chính xác phân loại	95,02%	99,03%	99,03%
Độ chính xác phát hiện tấn công	98,14%	97,80%	97,89%
Tỉ lệ phát hiện tấn công	78,68%	97,79%	97,67%
Cảnh báo nhầm (Dương tính giả)	1,86%	2,2%	2,11%

7. Kết luận

Trong bài báo chúng tôi đã ứng dụng máy học SVM kết hợp với nhận dạng bất thường trong ngữ cảnh cho kết quả phân loại có độ chính xác rất cao và tỷ lệ dương tính giả thấp, không vượt quá 2,2%.

Cùng sử dụng bộ dữ liệu nhưng Turnipseed [7] không sử dụng ngữ cảnh mà nhận dạng độc lập từng gói tin, cả ba thuật toán Turnipseed kiểm tra cho kết quả nhận dạng không quá 94,14% (xem bảng 1). Các thử nghiệm trong bài báo của chúng tôi đều cho kết quả phân loại cao hơn Turnipseed đạt trên 95,02%. Khi tăng kích thước của ngữ cảnh lên 5 hoặc 7 cho kết quả phân loại gần đạt đến 99% cao hơn tất cả các thuật mà Turnipseed kiểm tra.

Với ngữ cảnh gồm 5 gói tin cho độ chính xác phân loại (99,03%) cao hơn khi xét ngữ cảnh chỉ gồm 3 gói tin (95,02%). Còn với ngữ cảnh gồm 7 gói tin cho kết quả phân lại không cao hơn so với ngữ cảnh gồm 5 gói tin xem thêm

kết quả trong bảng 4. Đặc biệt là tỉ lệ phát hiện tấn công với ngữ cảnh bằng 5 đạt 97,79% còn với ngữ cảnh bằng 3 thấp hơn chỉ đạt 78,68%.

TÀI LIỆU THAM KHẢO

[1]. J. Slay and M. Miller, "Lessons learned from the Maroochy Water Breach", *Critical Infrastructure Protection*, Vol. 253, pp. 73–82, 2008.

[2]. D. Ryu, H. Kim and K. Um, "Reducing security vulnerabilities for critical infrastructure". *Journal of Loss Prevention in the Process Industries*, Vol. 22, pp. 1020–1024, 2009.

[3]. N. Falliere, L. O. Murchu and E. Chien, *W32.Stuxnet Dossier*, Symantec Report version 1.3, Nov 2010.

[4]. UCI. "Knowledge Discovery in Databases (KDD) Cup Datasets*"*. Available at http://kdd.ics.uci.edu.

[5]. T. Morris, W. Gao. "Industrial Control System Network Traffic Data Sets to Facilitate Intrusion Detection System Research", *in Critical Infrastructure Protection VIII, Springer Berlin Heidelberg*, Vol. 441, pp. 65-78, 2014.

[6]. Thornton, Z., *A Virtualized SCADA Laboratory for Research and Teaching*, Department of Electrical and Computer Engineering, Mississippi State University, 2015.

[7]. Turnipseed, I., "A new SCADA dataset for intrusion detection system research". *Department of Electrical and Computer Engineering, Mississippi State University*, August 2015.

[8]. S. Haykin, *Neural Networks and Learning Machines (3rd Edition)* - Prentice Hall, 2009.

[9]. Cortes, C., Vapnik, V., "Support-vector networks, Machine Learning", Vol. 20, pp. 273–297, 1995.

[10]. Bauer, D. S., &Koblentz, M. E. *NIDX – "A*n expert system for real-time network intrusion detection", 1988.

[11]. Lee, W., Stolfo, S., &Mok, K. "A Data Mining Framework for Building Intrusion Detection Model". *Proc. IEEE Symp. Security and Privacy*, pp. 120-132, 1999.

[12]. Amor, N. B., Benferhat, S., &Elouedi, Z. "Naïve Bayes vs. Decision Trees in Intrusion Detection Systems". *Proc. ACM Symp.Applied Computing*, 420424, 2004.

[13]. Mukkamala, S., Janoski, G., &Sung, A. "Intrusion detection using neural networks and support vector machines". *Paper presented at the International Joint Conference*, 2002.

[14]. Shah, H., Undercoffer, J., & Joshi, A. "Fuzzy Clustering for Intrusion Detection". Proc. *12th IEEE International Conference Fuzzy Systems (FUZZ-IEEE '03)*, 2, 1274-1278, 2003.

[15]. Ambwani, T. "Multi class support vector machine implementation to intrusion detection". *Paper presented at the Proceedings of the International Joint Conference of Neural Networks*, 2003.

[16]. T.Shon, Y. Kim, C.Lee and J.Moon, "A Machine Learning Framework for Network Anomaly Detection using SVM and GA", *Proceedings of the 2005 IEEE*, 2005.

[17]. SandyaPeddabachigari, Ajith Abraham, CrinaGrosan, Johanson Thomas. "Modeling Intrusion Detection Systems using Hybrid Intelligent Systems". *Journal of Network and Computer Applications*, 2005.