MỘT CÁCH TIẾP CẬN KẾT HỢP MẠNG NƠ-RON HỒI QUY VÀ TẬP LUẬT CHO PHÁT HIỆN XÂM NHẬP MẠNG

Trần Thị Hương[a*], Phạm Văn Hạnh[b]

[a]Khoa Toán - Cơ - Tin học, Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội, Hà Nội, Việt Nam
[b]Trung tâm Tin học, Trường Đại học Luật Hà Nội, Hà Nội, Việt Nam
[*]Tác giả liên hệ: Email: tranthihuong.hus@gmail.com

Lịch sử bài báo
Nhận ngày 15 tháng 12 năm 2018 | Chấp nhận đăng ngày 02 tháng 01 năm 2019

Tóm tắt

Phát hiện xâm nhập mạng là một trong những vấn đề quan trọng nhất của an ninh mạng và được rất nhiều nhóm trong và ngoài nước quan tâm nghiên cứu. Trong bài báo này chúng tôi trình bày một mô hình dựa vào việc kết hợp mạng nơ-ron truy hồi (recurrent neural network) và tập luật (rules) để phát hiện xâm nhập mạng. Ý tưởng chính của mô hình là việc kết hợp những điểm mạnh trong từng mô hình phân loại đơn lẻ. Tập luật có khả năng phát hiện tốt những cuộc tấn công đã biết, trong khi đó mạng nơ-ron truy hồi lại có ưu thế trong việc phát hiện những cuộc tấn công mới. Từ việc so sánh hiệu quả phát hiện giữa mô hình của chúng tôi với các mô hình phát hiện trước đây trên cùng bộ dữ liệu chuẩn KDD CUP 99 cho thấy mô hình đề xuất có hiệu quả cho việc phát hiện xâm nhập mạng tỷ lệ phát hiện xâm nhập cao trên 99%.

Từ khóa: Hệ thống phát hiện xâm nhập mạng; Mạng nơ-ron truy hồi; Tập luật.

DOI: http://dx.doi.org/10.37569/DalatUniversity.9.2.544(2019)

Loại bài báo: Bài báo nghiên cứu gốc có bình duyệt

AN APPROACH HYBRID RECURRENT NEURAL NETWORK AND RULE-BASE FOR INTRUSION DETECTION SYSTEM

Tran Thi Huong[a*], Pham Van Hanh[b]

[a]The Falculty of Mathematics, Mechanics, and Informatics, VNU University of Science, Hanoi, Vietnam
[b]The Information Technology Center, Hanoi Law University, Hanoi, Vietnam
[*]Corresponding author: Email: tranthihuong.hus@gmail.com

Article history
Received: December 15th, 2018 | Accepted: January 2nd, 2019

Abstract

Network intrusion detection is one of the most important issues of network security and is a research interest of many researchers. In this paper, we present a model based on the combination of recurrent neural networks and rule sets for the network intrusion detection problem. The main idea of the model is to combine the strengths of each classification model. The rule set is capable of detecting known attacks, while the recurrent neural network has the advantage of detecting new attacks. A comparison of the detection efficiency of our model with the previous detection models on the same data set, KDD CUP 99, shows that the proposed model is effective for detecting network intrusions at rates higher than 99%.

Keywords: Intrusion detection system; Recurrent neural network; Rule-based.

DOI: http://dx.doi.org/10.37569/DalatUniversity.9.2.544(2019)

Article type: (peer-reviewed) Full-length research article

1. GIỚI THIỆU

Sự bùng nổ và phát triển nhanh chóng của mạng Internet đang trở thành cơ hội cho những kẻ xâm nhập trái phép vào hệ thống mạng máy tính. Vấn đề an ninh mạng phải đối mặt với nhiều thách thức cho dù đó là cơ quan, tổ chức hay người dùng mạng Internet, rất nhiều những thông tin quan trọng và nhạy cảm được lưu trữ. Các chính sách bảo mật như kiểm soát truy cập, tường lửa (*firewall*) hay quản lý định danh rất khó ngăn chặn và phát hiện các cuộc tấn công. Do đó, hệ thống phát hiện xâm nhập mạng (*IDS-Intrusion Detection System*) trở thành công nghệ cần thiết để giúp các hệ thống máy tính phát hiện xâm nhập một cách hiệu quả và kịp thời hơn. Một hệ thống IDS có thể thu thập dữ liệu hoạt động của hệ thống và mạng, sau đó phân tích những thông tin đã thu thập để xác định đó có phải là một cuộc tấn công hay không. Dựa vào các phương pháp phát hiện xâm nhập mạng, Sodiya, Ojesanmi, Akinola, và Aborisade (2014) đã chia IDS thành hai loại như sau: i) Phát hiện dựa vào dấu hiệu xâm nhập (*signature - based IDS*) và ii) Phát hiện dựa vào bất thường (*anormaly detection - based IDS*). Các hệ thống *S-IDS* chủ yếu dựa vào luật nên rất có hiệu quả trong việc phát hiện xâm nhập những cuộc tấn công đã biết với tỷ lệ cảnh báo sai thấp. Tuy nhiên, điều đó lại dẫn tới nhược điểm của hệ thống khi không phát hiện các cuộc xâm nhập mới và việc xây dựng một cơ sở dữ liệu đầy đủ về các cuộc tấn công là rất khó. Bên cạnh đó, một số hệ thống AD-IDS gần đây sử dụng cách tiếp cận dựa vào học máy nhằm phát hiện các tấn công mới. Nhược điểm chính của các hệ thống này là thường học thiên lệch về các mẫu tấn công có số lượng lớn. Để kết hợp điểm mạnh của hai cách tiếp cận này, nhóm tác giả đề xuất một mô hình kết hợp giữa hệ thống S-IDS và AD-IDS cho phát hiện bất thường trong mạng. Cụ thể, nhóm tác giả đề xuất mô hình lai tập luật và huấn luyện mạng nơ-ron nhằm cải thiện tỷ lệ phát hiện cho các cuộc tấn công đã biết. Kết quả thực nghiệm chứng minh rằng phương pháp của chúng tôi hiệu quả và phân loại chính xác khi so sánh kết quả với mạng nơ-ron riêng lẻ và một số thuật toán học máy khác.

Phần còn lại của bài báo được tổ chức như sau: Mục 2 sẽ trình bày một số cách tiếp cận dựa vào học máy để giải bài toán phát hiện xâm nhập mạng; Mục 3 giới thiệu về bộ dữ liệu huấn luyện và kiểm tra KDD 99 được sử dụng trong quá trình thực nghiệm; Mục 4 sẽ trình bày mô hình đề xuất kết hợp mạng nơ-ron truy hồi và tập luật cho phát hiện xâm nhập; Mục 5 đưa ra kết quả thực nghiệm trên bộ dữ liệu KDD và so sánh với các thuật toán đã được trình bày trong Mục 2; và Cuối cùng là phần kết luận và hướng nghiên cứu tiếp theo.

2. MỘT SỐ CÁCH TIẾP CẬN DỰA VÀO HỌC MÁY CHO BÀI TOÁN PHÁT HIỆN XÂM NHẬP MẠNG

Một nghiên cứu gần đây của Subba, Biswas, và Karmakar (2015) đã đề xuất một mô hình phát hiện xâm nhập dựa vào hồi quy logistic. Các tác giả đã nghiên cứu sự phụ thuộc giữa biến trả lời (nhãn của các cuộc tấn công) và các biến dự báo (dấu hiệu của các cuộc tấn công) bằng cách xây dựng một hàm giả thiết $h_\theta(x_0)$ cho mẫu x_0:

$$h_\theta(x_0) = g(\theta^T x_0) = \frac{1}{1+\exp(-\theta^T x_0)} \quad (1)$$

Mô hình hồi quy logistic được xây dựng như sau:

$$P(y_0 = 1 \mid x_0; \theta) = h_\theta(x_0); P(y_0 = 0 \mid x_0; \theta) = 1 - h_\theta(x_0)$$

Giả sử $\hat{\theta}$ là ước lượng cực đại của θ. Trong bài toán phát hiện xâm nhập mạng mô hình hồi quy logistic xếp đối tượng x_0 vào lớp *attack* ($y_0 = 1$) nếu $\widehat{\theta^T}\ x_0 > 0$. Ngược lại thì x_0 được xếp vào lớp *normal* ($y_0 = 0$).

Subba và ctg. (2015) thực nghiệm nhiều lần trên cả bộ huấn luyện và bộ kiểm tra, và đánh giá hiệu suất trong mỗi lần thực nghiệm, từ đó chọn ra những tham số tốt nhất cho mô hình. Trong khi đó, Bhavasar và Waghmare (2013) sử dụng phương pháp SVM (*Support Vector Machines*) để phân loại các gói tin tấn công và không tấn công bằng cách tìm một siêu phẳng tuyến tính tối ưu có khoảng cách giữa hai lớp cần phân loại lớn nhất. Nhóm nghiên cứu sử dụng các hàm nhân khác nhau như hàm *sigmoid,* hàm đa thức hay hàm RBF (*radial basis function*) để ánh xạ dữ liệu huấn luyện sang không gian nhiều chiều hơn. Từ đó tìm được siêu phẳng tối ưu phân loại tốt các mẫu dữ liệu. Sung và Mukkamala (2003) lần đầu tiên đề xuất áp dụng mô hình mạng nơ-ron nhân tạo trong hệ thống IDS. Dựa vào ý tưởng các nơ-ron liên kết với nhau để xử lý thông tin đầu vào và đưa ra những tri thức từ các thông tin đó. Các tác giả chọn ra 14000 bản ghi kết nối ngẫu nhiên trong bộ dữ liệu KDD Cup 99 chia thành hai phần, 7000 kết nối ngẫu nhiên đầu tiên và thuật toán lan truyền ngược (*back - propagation*) được áp dụng cho việc huấn luyện của mô hình. Phần còn lại dùng để kiểm tra. Nhóm nghiên cứu đã chỉ ra rằng mô hình phát hiện dựa vào mạng nơ-ron cho kết quả phân loại tốt các kết nối là "*normal*" hay "*attack*". Tuy nhiên, nhóm tác giả tiến hành thực nghiệm trên bộ dữ liệu rất nhỏ tương đối với số lượng mẫu tấn công chưa đa dạng.

Một cách tiếp cận đáng chú ý khác sử dụng mạng nơ-ron nhân tạo của Sodiya và ctg. (2014). Hệ thống sử dụng các cảm biến để thu thập dữ liệu, phân tích dữ liệu và chuyển tới bộ phát hiện (*detectors*). Bộ phát hiện được xây dựng dựa trên việc kết hợp mô hình SOM (*Self Organzing Maps*) được biết đến như mô hình học không giám sát và mạng nơ-ron nhiều lớp (*multilayer perceptron*) để phân tích các đặc tính của các kết nối từ đó phát hiện các xâm nhập bất thường. Kết quả thực nghiệm của nhóm chỉ ra rằng cách tiếp cận này có hiệu quả cho bài toán phát hiện xâm nhập với tỷ lệ phát hiện xâm nhập trên 96%, tỷ lệ cảnh báo lỗi là 3%. Tuy nhiên, trong quá trình huấn luyện mô hình thường học thiên lệch về các mẫu tấn công thường gặp và khó phát hiện được những mẫu xâm nhập ít gặp, đặc biệt là các xâm nhập U2R và R2L. Đây chính là hạn chế của nghiên cứu này.

Gần đây, mô hình mạng nơ-ron truy hồi (*Recurrent Neural Network - RNN*) được biết đến là một trong những phương pháp học sâu hiệu quả nhất giải quyết các bài toán phân loại trong lĩnh vực xử lý ngôn ngữ tự nhiên, xử lý ảnh nhờ việc ghi nhớ các mẫu dữ liệu trong quá trình học (Yin, Yuenfei, Fei, & He, 2017). Tuy nhiên trong quá

trình huấn luyện, mạng nơ-ron truy hồi thường ghi nhớ và học thiên lệch về các mẫu tấn công có số lượng lớn như *DoS, Probe*. Vì vậy, trong nghiên cứu này, chúng tôi sẽ đề xuất một mô hình kết hợp giữa mạng nơ-ron truy hồi và tập luật cho bài toán phát hiện xâm nhập mạng và chủ yếu tập trung vào việc phân loại thành hai lớp tấn công hay là không tấn công. Mô hình đề xuất xuất phát từ ý tưởng kết hợp điểm mạnh của từng mô hình phát hiện, đối với các mẫu tấn công thường gặp và có số lượng lớn chúng tôi sử dụng mô hình mạng nơ-ron truy hồi, còn đối với những mẫu tấn công ít gặp như U2R, R2L chúng tôi sẽ áp dụng tập luật để phát hiện. Từ đó giúp nâng cao hiệu quả phát hiện trong thời gian thực với tỷ lệ phát hiện chính xác cao và tỷ lệ cảnh báo lỗi thấp. Các kết quả này được rút ra từ các thực nghiệm trên bộ dữ liệu KDD 99 10%.

3. BỘ DỮ LIỆU KDD CUP 99

Trong phần thực nghiệm mô hình, chúng tôi sử dụng bộ dữ liệu KDD CUP 99. Bộ dữ liệu này có nguồn gốc từ MIT's Lincoln Lab, được phát triển cho chương trình đánh giá phát hiện tấn công mạng của Cơ quan Quản lý Nghiên cứu Dự án phòng thủ tiên tiến của Bộ Quốc phòng Mỹ (DARPA) năm 1998 (Moradi & Zulkemine, 2004). Họ cài đặt một môi trường giả lập các cuộc tấn công mạng và thu thập được khoảng 4GB dữ liệu *tcp dump* thô trong bảy tuần. Sau đó, các dữ liệu thô này được xử lý và đưa về định dạng chuẩn của một bản ghi kết nối TCP/IP gồm 42 trường. Và bộ dữ liệu được thu thập độc lập trong vòng hai tuần.

Mỗi bản ghi trong bộ dữ liệu KDD bao gồm 42 đặc trưng trong đó có bốn cột đặc trưng ở dạng phi số: Đặc trưng số hai biểu diễn loại giao thức; Đặc trưng số ba biểu diễn loại dịch vụ; Đặc trưng thứ tư biểu diễn trạng thái cờ của kết nối; và Đặc trưng thứ 42 là nhãn tương ứng với bản ghi là bình thường hay một loại tấn công cụ thể. Các tấn công này cũng được phân thành bốn nhóm, đó là: DoS, Probing, U2R, và R2L. Bảng 1 cho biết số liệu thống kê các bản ghi trên tập dữ liệu huấn luyện "10% KDD" và tập kiểm tra "Corrected KDD" thuộc về các nhãn lớp khác nhau trên bộ dữ liệu KDD.

Bảng 1. Phân bố dữ liệu các gói tin trong tập dữ liệu huấn luyện và tập kiểm tra

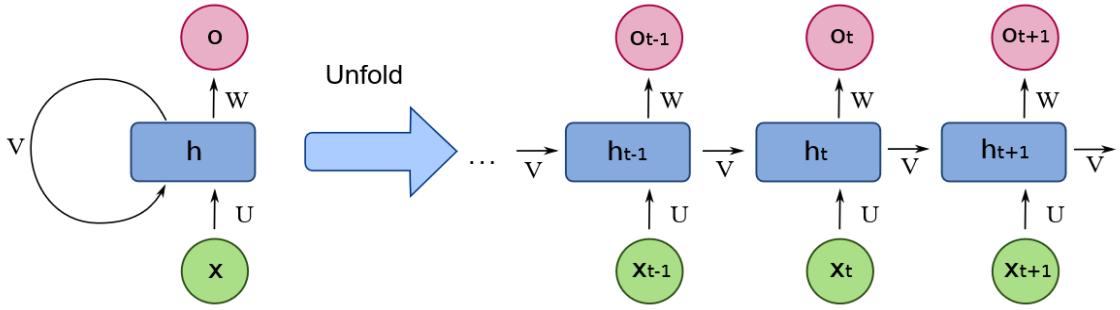
Dataset	DoS	Probe	U2R	R2L	Normal	Tổng gói tin
"10% KDD"	391458	4107	52	1126	97277	494020
	79.230%	0.830%	0.011%	0.220%	19.690%	
"Corrected KDD"	229853	4166	70	16347	60593	311029
	73.900%	1.330%	0.022%	5.260%	19.480%	

4. MÔ HÌNH KẾT HỢP MẠNG NƠ-RON HỒI QUY VÀ TẬP LUẬT CHO BÀI TOÁN PHÁT HIỆN XÂM NHẬP MẠNG

4.1. Mô hình mạng nơ-ron hồi quy cho bài toán phát hiện xâm nhập mạng

Mạng nơ-ron hồi quy (*RNN- Recurrent Neural Network*) là một thuật toán được nghiên cứu rất nhiều trong thời gian gần đây bởi các kết quả tốt thu được từ nhiều lĩnh

vực như thị giác máy tính, xử lý ngôn ngữ tự nhiên, nhận dạng và cho hiệu quả cao trên nhiều bài toán (Yin & ctg., 2017). Xuất phát từ ý tưởng mô phỏng khả năng ghi nhớ thông tin trong mạng nơ-ron của sinh học. Đối với việc phân loại và dự đoán một tri thức mới thường dựa trên những tri thức đã biết trước đó. Các nhà nghiên cứu đã đề xuất ra mô hình mạng nơ-ron truy hồi. RNN được gọi là mạng nơ-ron hồi quy vì sử dụng chuỗi các thông tin và thực hiện cùng một tác vụ cho tất cả các phần tử của một chuỗi với đầu ra phụ thuộc vào các phép tính toán ở các lớp trước đó. Nói cách khác, RNN có khả năng ghi nhớ các thông tin được tính toán trước đó (Kim & Kim, 2016). Hình 1 mô tả một mạng nơ-ron hồi quy với đầu vào *x* và đầu ra *o* chứa vòng lặp bên trong cho phép thông tin có thể truyền từ bước này qua bước khác của mạng từ đó thông tin được lưu lại. Bảng 2 mô tả một số ký hiệu trong mô hình mạng nơ-ron truy hồi.



Hình 1. Mạng nơ-ron hồi quy có vòng lặp

Bảng 2. Một số ký hiệu trong mô hình mạng nơ-ron truy hồi

Ký hiệu	Mô tả
x_i	Véc-tơ biểu diễn mẫu dữ liệu thứ i của tập huấn luyện, i = 1, 2, .., N
h_j	Nơ-ron ẩn thứ j
y_i	Véc-tơ đầu ra của mẫu i
$\hat{y_t}$	Véc-tơ đầu ra của mẫu được tính toán nhờ RNN
W_{hx}	Ma trận trọng số input-to-hidden
W_{hh}	Ma trận trọng số hidden-to-hidden
W_{yh}	Ma trận trọng số hidden-to-output
b_h	Giá trị bias của lớp ẩn
b_y	Giá trị bias của véc-tơ ra
f	Hàm kích hoạt tại lớp ẩn
g	Hàm kích hoạt tại lớp ra
η	Tỷ lệ học
k	Số lượng vòng lặp

Mạng nơ-ron hồi quy gồm có lớp đầu vào, lớp ẩn, lớp đầu ra và các lớp này liên kết với nhau nhờ các liên kết trọng số. Có ba loại trọng số: Trọng số từ lớp đầu vào tới

lớp ẩn gọi là *input-to-hiden*; Trọng số từ lớp ẩn tới đầu ra là *hidden-to-output*; và Trọng số từ lớp ẩn này tới lớp ẩn tiếp theo là *hiden-to-hiden*. Từ Hình 1, ta có thể thấy trọng số trong *hiden-to-hiden* được tính toán một cách hồi quy (*recurrent*). Mạng RNN cập nhật trọng số bằng cách huấn luyện mô hình.

Trong mô hình học có giám sát, quá hình huấn luyện mạng nơ-ron hồi quy gồm hai giai đoạn: i) Lan truyền tiến (*forward propagation*) và ii) Lan truyền ngược (*back propagation*). Với một chuỗi đầu vào $x_1, x_2, ..., x_N \in \mathbb{R}^n$, mạng tính toán một chuỗi trạng thái ẩn $h_1, h_2, ..., h_T \in \mathbb{R}^m$ và một chuỗi dự đoán $\widehat{y_1}, \widehat{y_2}, \ldots, \widehat{y_T} \in \mathbb{R}^k$. Giả sử $L(y_i, \widehat{y_i})$ là hàm tổn thất trên mỗi mẫu dữ liệu huấn luyện (x_i, y_i), khi đó hàm tổn thất *cross - entropy* trên toàn bộ dữ liệu được tính theo công thức sau (Martens & Sutskver, 2011):

$$L(y_i, \widehat{y_i}) = -\sum_i \sum_j y_{ij} \log(\widehat{y_{ij}}) + (1 - y_{ij}) \log(1 - \widehat{y_{ij}}) \quad (2)$$

Thuật toán lan truyền ngược được mô tả như sau:

- *Đầu vào*: Một tập mẫu huấn luyện x_i ($i = 1, 2, \ldots, N$).
- *Đầu ra*: $\widehat{y_i}$
- *Chi tiết thuật toán*:

1 for *i* từ 1 đến *N* do
2 $t_i = W_{hx} x_i + W_{hx} h_{i-1} + b_h$
3 $h_i = sigmoid(t_i)$
4 $s_i = W_{yh} h_i + b_y$
5 $\widehat{y_i} = softmax(s_i)$
6 end for

Trong khi đó, Thuật toán cập nhật trọng số được mô tả như sau:

- *Đầu vào*: Tập các cặp $<y_i, \widehat{y_i}>$ ($i = 1, 2, ..., N$)
- Đầu ra: $\hat{\theta} = \{\widehat{W_{hx}}, \widehat{W_{hh}}, \widehat{W_{hy}}, \widehat{b_h}, \widehat{b_y}\}$
- Chi tiết thuật toán:

1 Khởi tạo $\theta = \{W_{hx}, W_{hh}, W_{hy}, b_h, b_y\}$
2 for *i* từ *k* giảm về 1 do
3 Tính *entropy* chéo giữa giá trị đầu ra và giá trị thực
$L(y_i, \widehat{y_i}) \leftarrow -\sum_i \sum_j y_{ij} \log(\widehat{y_{ij}}) + (1 - y_{ij}) \log(1 - \widehat{y_{ij}})$
4 Tính giá trị đạo hàm từng phần với $\widehat{\theta_i}$: $\delta_i = \frac{dL}{d\theta_i}$

5 Trọng số được cập nhật: $\widehat{\theta_\iota} \leftarrow : \widehat{\theta_{\iota-1}}\eta + \delta_i$

6 end for

4.2. Mô hình phát hiện xâm nhập mạng dựa vào tập luật

Hệ thống phát hiện xâm nhập dựa vào tập luật (Rule-based IDS) được biết đến phổ biến trong các hệ thống IDS truyền thống nhờ hiệu quả trong phát hiện dấu hiệu các cuộc xâm nhập đã từng xảy ra với tỉ lệ cảnh báo thấp (Bouzida & Cuppens, 2006). Ý tưởng của các hệ thống này là xây dựng cơ sở dữ liệu các luật về các cuộc xâm nhập. Sau đó so sánh lưu lượng mạng qua hệ thống với cơ sở dữ liệu luật để đưa ra cảnh báo. Để xây dựng ra hệ thống các luật, nhóm nghiên cứu sử dụng cây quyết định C4.5 sinh ra các luật. Cây quyết định phân loại mẫu dữ liệu cho trước sử dụng giá trị các thuộc tính của nó, ban đầu cây quyết định được xây dựng từ một tập dữ liệu được phân loại trước đó. Mỗi mẫu dữ liệu được định nghĩa bởi các giá trị của các thuộc tính.

Trong nghiên cứu này chúng tôi sẽ sử dụng độ đo thông tin *information gain* để đo độ ảnh hưởng của 41 thuộc tính đến lớp phân loại là "*normal*" hay "*attack*". Cho *S* là một tập các mẫu huấn luyện với các nhãn tương ứng với các mẫu (Kayacik, Heywood, & Heywood, 2005). Giả sử ta có *m* lớp phân loại, tập huấn luyện chứa s_i mẫu của lớp *I* và *s* là tổng số các mẫu trong tập huấn luyện được cho như Công thức (3).

$$S = \sum_{i=1}^{m} s_i \tag{3}$$

Lượng thông tin thu được cần để phân loại một mẫu cho trước được tính như Công thức (4).

$$H(s_1, s_2, \ldots s_m) = -\sum_{i=1}^{m} \frac{s_i}{s} \log_2\left(\frac{s_i}{s}\right) \tag{4}$$

Ta có $\{f_1, f_2, \ldots f_v\}$ là tập các giá trị của một đặc trưng *F*.

Chia tập huấn luyện thành *v* tập con $\{S_1, S_2, \ldots S_v\}$ trong đó tập S_j là tập con của *S* mà *F* nhận giá trị f_j. Hơn nữa, nếu S_j chứa s_{ij} mẫu của lớp *i* thì *Entropy* của đặc trưng *F* được tính như Công thức (5):

$$I(F) = \sum_{j=1}^{v} \frac{s_{1j} + s_{2j} + \ldots + s_{mj}}{s} \times I(s_{1j}, s_{2j}, \ldots s_{mj}) \tag{5}$$

Ta có Công thức (6) tính độ đo *information gain* của một đặc trưng *F*:

$$IG(F) = H(s_1, s_2, \ldots s_m) - I(F) \tag{6}$$

Dựa vào giá trị của độ đo IG của 41 thuộc tính trên chúng tôi xác định các luật phân loại vào hai lớp tấn công U2R và R2L như Bảng 3.

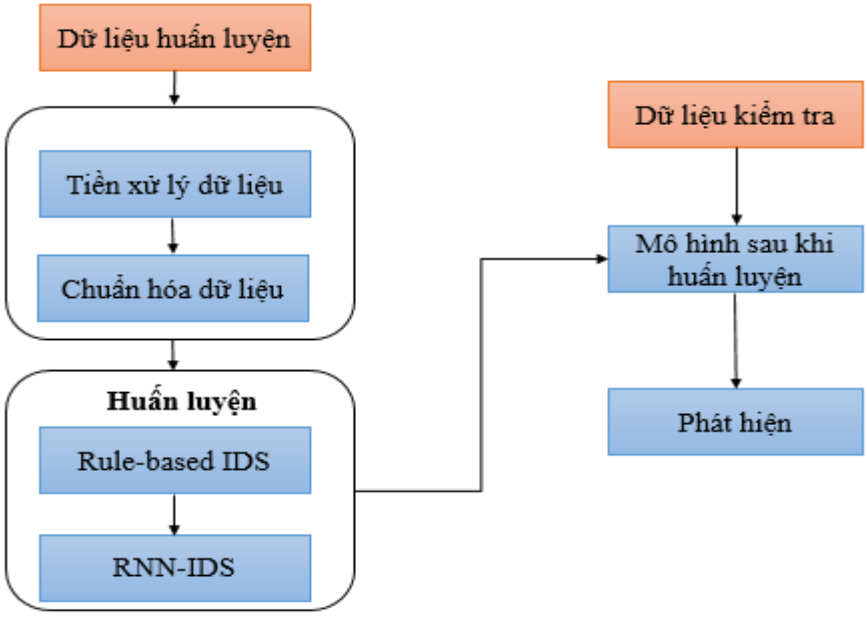
Bảng 3. Các luật phân loại vào hai lớp tấn công U2R và R2L

Luật	Trường hợp tấn công
Rule 1	if (duration =0 || duration >=289 & protocol_type=tcp & dst_host_srv-count >=1 & dst_host_srv-count <=128) then “attack”
Rule 2	if (duration =0 & protocol_type=icmp || protocol_type=udp || protocol_type=tcp & dst_host_srv-count>=1 & dst_host_srv-count <=20) then “attack”
Rule 3	if(duration>265 & dest_byte<=688 & is_guest=1) then “attack”
Rule 4	if(source_byte>= 26516 & source_byte<=283618) then “attack”
Rule 5	if(num_failed_logins>0 & dst_host_same_srv_rate>0) then “attack”
Rule 6	if (duration =0 || duration <=289 & protocol_type=udp & src_bytes = 0 then “attack”
Rule 7	if (land =0 & protocol_type=tcp & dst_host_srv_cout <=100) then “attack”
Rule 8	if(hot>2 &root_shell>0) then “attack”
Rule 9	if(protocol_type=6 & duration >2 & src_bytes>20 & src_bytes<=39) then “attack”

4.3. Mô hình kết hợp mạng nơ-ron hồi quy và tập luật cho bài toán phát hiện xâm nhập mạng

Một hệ thống thông minh kết hợp áp dụng cách tiếp cận mô hình học máy và mô hình phát hiện dựa vào luật được đề xuất trong phần này. Mỗi mô hình học hoạt động theo cách khác nhau và khai thác các tập thuộc tính khác nhau. Vì vậy mô hình học kết hợp điểm mạnh của từng mô hình và cho hiệu suất tốt hơn đối với mô hình riêng lẻ. Trong kiến trúc hệ thống kết hợp thông minh, mỗi lớp cung cấp một vài thông tin tới lớp cao hơn.

Tập dữ liệu đầu tiên được đi qua cơ sở dữ liệu các luật và thông tin về nút được tạo ra. Thông tin của nút được xác định và sinh ra các luật bởi cây quyết định. Tất cả tập bản ghi dữ liệu được gán một nút xác định mà biểu diễn một lớp nhất định hoặc tập con. Phân loại nút thành hai loại biểu diễn cho mẫu tấn công hoặc là không tấn công. Đối với những mẫu tấn công không bị phát hiện bởi tập các luật sẽ được đưa qua mạng nơ-ron truy hồi để huấn luyện và học các mẫu. Mô hình sau khi huấn luyện sẽ thu được bộ các tham số từ dữ liệu đầu vào. Các gói tin (*packet*) đi vào hệ thống sau khi được xử lý sẽ được đưa vào mô hình phát hiện để đưa ra cảnh báo khi có bất thường. Hình 2 chỉ ra kiến trúc của hệ thống kết hợp thông minh giữa mạng nơ-ron truy hồi và tập luật.



Hình 2. Mô hình kết hợp tập luật và mạng nơ-ron truy hồi trong hệ thống phát hiện xâm nhập mạng

5. ÁP DỤNG MÔ HÌNH KẾT HỢP MẠNG NƠ-RON TRUY HỒI VÀ TẬP LUẬT CHO BÀI TOÁN PHÁT HIỆN XÂM NHẬP MẠNG

5.1. Xử lý dữ liệu

5.1.1. Chuyển giá trị đặc trưng phi số sang số

Trong bộ dữ liệu KDD có 38 đặc trưng dạng số và ba đặc trưng có giá trị phi số trong đó đặc trưng thứ hai biểu diễn "*protocol_type*", đặc trưng thứ ba biểu diễn "*service*" được sử dụng, đặc trưng thứ tư biểu diễn "*flag*". Việc huấn luyện trong mạng nơ-ron truy hồi xử lý các dữ liệu đầu vào là ma trận dạng số vì vậy chúng tôi tiến hành chuyển đổi một vài đặc trưng phi số. Chúng tôi sử dụng kỹ thuật chuyển ký tự thành các số nguyên. Ví dụ, trong đặc trưng biểu diễn giao thức ta gán 1 cho giao thức là TCP, gán 2 nếu giao thức là UDP và 2 nếu giao thức là ICMP. Với 84 giá trị phi số (tương ứng với 84 giá trị của ba đặc trưng nêu trên) được chuyển sang số theo thứ tự tăng dần từ 1 đến 84.

5.1.2. Chuẩn hóa dữ liệu

Một vấn đề khác cũng ảnh hưởng tới độ hội tụ nhanh của các thuật toán học máy đó là việc phân bố giá trị đặc trưng không đồng đều giữa các đặc trưng trong bộ dữ liệu KDD. Có nhiều đặc trưng có miền giá trị chênh lệch rất lớn như đặc trưng biểu diễn "*duration*" với miền gía trị [0, 60000] hay đặc trưng biểu diễn "*src_bytes*", "*dst_bytes*" với miền giá trị [0, 1300000000]. Vì vậy, cần tiến hành chuẩn hóa miền giá trị dữ liệu của các đặc trưng. Tác giả sử dụng chuẩn hóa min-max theo Công thức (7).

$$v_i = \frac{v_i - Min}{Max - Min} \quad (7)$$

Trong đó *Max*, *Min* ký hiệu giá trị lớn nhất và giá trị nhỏ nhất của mỗi đặc trưng, v_i giá trị thứ *i* của mỗi đặc trưng.

5.2. Thiết kế mô hình

- Giai đoạn 1 (Huấn luyện) là mạng nơ-ron truy hồi gồm bốn lớp: một lớp đầu vào, hai lớp ẩn, và một lớp đầu ra. Trong đó: Lớp đầu vào gồm 42 nút biểu diễn cho 41 thuộc tính và một nút +1 biểu diễn *bias*; Lớp đầu ra gồm 1 nút có giá trị đầu ra 0 và 1 tương ứng với lớp "*normal*" và "*attack*"; Hai lớp ẩn gồm 20 nút và 10 nút tương ứng. Hàm kích hoạt lớp ẩn là hàm *logsig; và* Hàm huấn luyện là hàm *traingdx*, ngưỡng vòng lặp là 1000, sai số huấn luyện là 10^{-6}

- Giai đoạn 2 (Phân loại) sử dụng thuật toán soft-max regression với hàm huấn luyện là hàm *mnrfit*. Giai đoạn này bao gồm: Đầu vào là *K* thuộc tính mới học được trong giai đoạn 1 (*K* = 20 và *K* = 10); Đầu ra là kết quả phân loại: 0 nếu là bản ghi bình thường và 1 nếu là bản ghi bất bình thường; và Tỷ lệ học (*learning rate* = 0.1).

5.3. Kết quả thực nghiệm

5.3.1. Thông số đánh giá

Các thuật toán học máy thường sử dụng ma trận hỗn độn (*confusion matrix*) (Bảng 4) để đánh giá tính hiệu quả của thuật toán (Kim & Kim, 2016; Yin & ctg., 2017).

Bảng 4. Ma trận hỗn độn

Confusion matrix		Predicted class	
		Normal	Attacks
Actual class	Normal	TN	FP
	Attacks	FN	TP

Tác giả sử dụng các độ đo dưới đây để đánh giá hiệu quả của các phương pháp học máy:

- Tỷ lệ phát hiện chính xác:

$$precision = \frac{TP}{TP + FN} \times 100\% \quad (8)$$

- Độ chính xác:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (9)$$

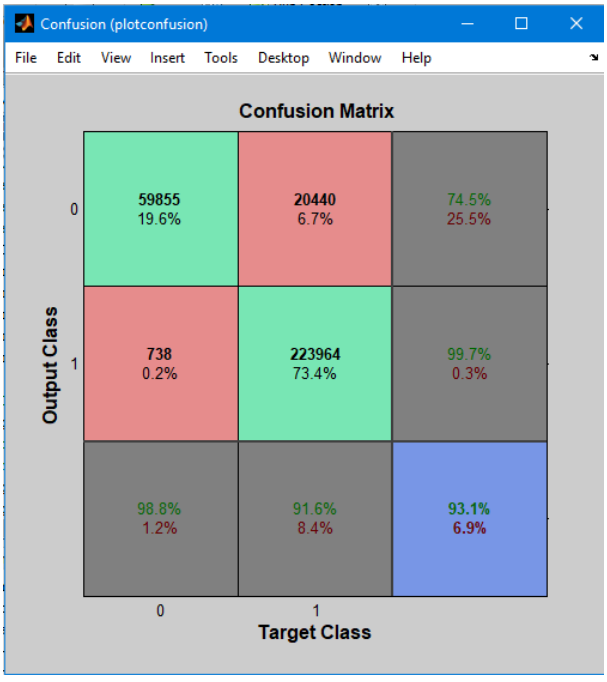
- Tỷ lệ lỗi:

$$FNR = \frac{FN}{FN + TP} \times 100\% \quad (10)$$

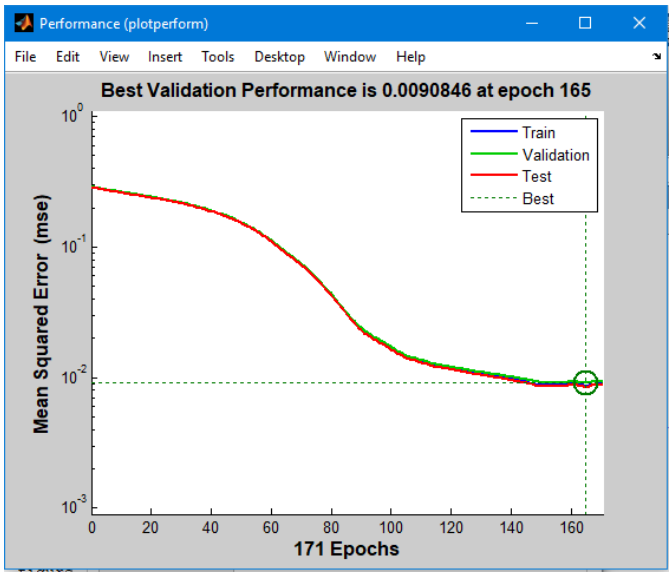
5.3.2. Kết quả thực nghiệm

Bảng 5. Kết quả phát hiện tấn công của các mô hình trên bộ dữ liệu 10% KDD

Bộ dữ liệu huấn luyện	Mô hình	Tỷ lệ phát hiện chính xác (%)	Độ chính xác (%)	Tỷ lệ lỗi (%)
10% KDD	Logistic	87.60	85.56	14.44
	SVM	91.72	89.90	10.10
	Rule-Neural Network	99.60	91.10	0.40
	Hybrid (Rule based and RNN)	99.70	93.30	0.30



Hình 3. Kết quả phát hiện mô hình kết hợp Tập luật và RNN trên bộ dữ liệu kiểm tra "Corrected KDD"



Hình 4. Biểu đồ hội tụ trên các tập huấn luyện, tập Validation, và tập kiểm tra "Corrected KDD"

6. KẾT LUẬN

Trong bài báo này, chúng đã đi sâu nghiên cứu về bài toán phát hiện xâm nhập mạng, cách tiếp học máy cho bài toán phát hiện xâm nhập và đề xuất mô hình kết hợp giữa mô hình mạng nơ-ron truy hồi và tập luật nhằm cải thiện tỷ lệ phát hiện xâm nhập chính xác. Mô hình mạng nơ-ron truy hồi có ưu thế trong việc phát hiện những cuộc tấn công mới, trong khi đó tập luật có khả năng phát hiện dấu hiệu của những tấn công đã biết với cảnh báo lỗi thấp. Việc kết hợp điểm mạnh của hai mô hình là có ý nghĩa và giảm thời gian huấn luyện so với cách tiếp cận học máy truyền thống cho bài toán phát hiện xâm nhập. Trong nghiên cứu này, chúng tôi mới tập chung nghiên cứu các thuật toán học máy có giám sát mà chưa xét đến những thuật toán học máy phi giám sát và bán giám sát. Vì vậy đây có thể được xem là hướng phát triển trong tương lai.

LỜI CẢM ƠN

Nghiên cứu này được sự tài trợ của Trường Đại học Khoa học Tự nhiên Hà Nội với mã số đề tài TN 17.0.3.

TÀI LIỆU THAM KHẢO

Al-Mamory, S. O., & Jassim, F. S. (2013). Evaluation of different data mining algorithms with KDD Cup 99 dataset. *Journal of Babylon University/Pure and Applied Sciences, 21*(8), 2663-2670.

Bhavasar, Y. B., & Waghmare, K. C. (2013). Intrusion detection system using data mining technique: Support Vector Machines. *International Journal of Emerging Technology and Advanced Engineering, 3*(3), 581-586.

Bouzida, Y., & Cuppens, F. (2006). *Neural networks vs decision tree for intrusion detection*. Retrieved from http://www.diadem-firewall.org/workshop06/papers/monam06-paper-29.pdf.

Kayacik, G., Heywood, A. N. Z., & Heywood, I. M. (2005). *Selecting features for intrusion detection: A feature relevance analysis on KDD 99*. Paper presented at The Third Annual Conference on Privacy, Securitty and Trust, Canada.

Kim, J., & Kim, H. (2016). Applying recurrent neural network to intrusion detection with hessian free optimization. In H. Kim, & D. Choi (Eds), *Information Security Applycations* (pp. 357-369). Berlin, Germany: Springer Publishing.

Martens, J., & Sutskver, I. (2011). *Learning recurrent neural networks with Hessian free optimizations*. Paper presented at The 28th International Conference on Machine Learning, USA.

Moradi, M., & Zulkemine, M. (2004). *A neural network based system for intrusion detection and classification of attacks*. Paper presented at The International Conference on Advances in Intelligent System, Luxembourg.

Mukkamala, S., Janoski, G., & Sung, A. (2002). *Intrusion detection: Support Vector Machines and neural networks*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.407.7230&rep=rep1&type=pdf.

Sethuramalingam, S., & Naganathan, D. (2011). Hybrid feature selection for network intrusion. *International Journal on Computer Science and Engineering, 3*(5), 1773-1780.

Sodiya, A. S., Ojesanmi, O. A., Akinola, O. C., & Aborisade, O. (2014). Neural network based intrusion detection systems. *International Journal of Computer Applications, 106*(18), 19-24.

Subba, B., Biswas, S., & Karmakar, S. (2015). *Intrusion detection system using linear discriminant analysis and logictic regression*. Paper presented at The Annual IEEE India Conference, India.

Sung, A. H., & Mukkamala, S. (2003). *Feature selection for intrusion detection using neural networks and Support Vector Machines*. Retrieved from https://www.researchgate.net/publication/228966999_Feature_Selection_for_Intrusion_Detection_with_Neural_Networks_and_Support_Vector_Machines.

Yin, C., Yuenfei, Z., Fei, J., & He, X. (2017). A deep learning approach for intrusion detection using recurrent neural networks. *IEEE Access, 5*, 21954-21961.