# MỘT MÔ HÌNH PHÁT HIỆN DGA BOTNET DỰA TRÊN HỌC KẾT HỢP

**Vũ Xuân Hạnh**[*], **Hoàng Xuân Dậu**[+], **Đinh Trường Duy**[+]
[*] Đại học Mở Hà Nội
[+] Học viện Công nghệ Bưu chính Viễn thông

***Tóm tắt:*** Gần đây, DGA đã trở thành kỹ thuật được sử dụng rộng rãi bởi nhiều mã độc nói chung và các botnet nói riêng. DGA cho phép các nhóm tin tặc tự động sinh và đăng ký các tên miền cho các máy chủ C&C của các mạng botnet để tránh bị đưa vào danh sách đen và vô hiệu hóa nếu sử dụng tên miền và địa chỉ IP tĩnh. Nhiều dạng kỹ thuật DGA tinh vi được phát triển và sử dụng, như character-based DGA, word-based DGA và mixed DGA. Các kỹ thuật này cho phép sinh từ các tên miền đơn giản là tổ hợp ngẫu nhiên của các ký tự đến các tên miền phức tạp là tổ hợp của các từ có nghĩa tương tự như các tên miền bình thường. Điều này gây khó khăn cho các giải pháp giám sát, phát hiện botnet nói chung và DGA botnet nói riêng. Nhiều giải pháp có khả năng phát hiện hiệu quả các tên miền dạng character-based DGA, nhưng không thể phát hiện các tên miền dạng word-based DGA và mixed DGA. Ngược lại, một số đề xuất gần đây có thể phát hiện hiệu quả các dạng tên miền word-based DGA, nhưng lại không thể phát hiện hiệu quả các tên miền của một số dạng character-based DGA. Bài báo này đề xuất một mô hình dựa trên học kết hợp cho phép phát hiện hiệu quả hầu hết các họ tên miền DGA, bao gồm cả character-based DGA và word-based DGA. Mô hình đề xuất kết hợp hai mô hình thành phần, gồm mô hình phát hiện các tên miền họ character-based DGA và mô hình phát hiện các tên miền họ word-based DGA. Các kết quả thử nghiệm cho thấy mô hình kết hợp đề xuất có khả năng phát hiện hiệu quả 37/39 họ DGA botnet với tỷ lệ phát hiện đạt trên 89%.

***Từ khóa:*** Character-based DGA botnet, Word-based DGA botnet, Phát hiện DGA botnet, Phát hiện DGA botnet dựa trên học học kết hợp.

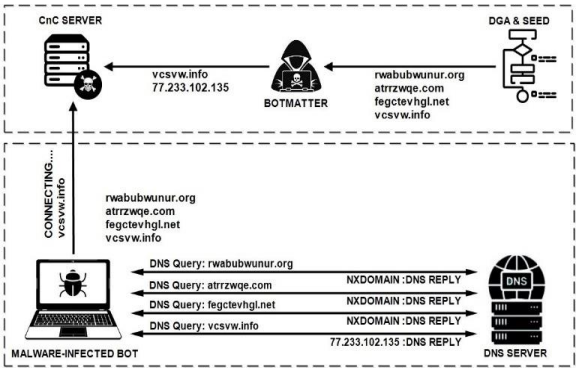
## I. GIỚI THIỆU

Trong thập kỷ vừa qua, các mạng botnet được xem là một trong các mối đe dọa bảo mật chính đối với các cơ quan, tổ chức, các hệ thống có kết nối Internet và cả người dùng Internet [1][2][3]. Điều này là do các botnet có liên hệ trực tiếp, hoặc gián tiếp với nhiều dạng tấn công và lạm dụng trên Internet, như tấn công DDoS, lan truyền các dạng phần mềm độc hại, gửi thư rác và đánh cắp các thông tin nhạy cảm [4]. Vào năm 2019, hệ thống mạng của Tờ báo Telegram đã chịu một cuộc tấn công DDoS có qui mô rất lớn. Cuộc tấn công này được cho là khởi phát từ Trung Quốc và có liên quan đến phong trào biểu tình phản đối ở Hông Kông trong năm 2019 [3][4]. Theo một báo cáo an ninh mạng của công ty Symantec, khoảng 95% email gửi trên mạng Internet trong năm 2010 là thư rác được tạo và gửi bởi các botnet [1]. Hơn nữa, nhiều dạng tấn công và lạm dụng nguy hiểm khác được hỗ trợ bởi các botnet, bao gồm tấn công chèn mã trên các trang web, giả mạo URL và giả mạo các máy chủ DNS [1][2].

Nói chung, một *botnet* là một mạng của các *bot*. Một bot là một dạng mã độc đặc biệt hoạt động trên một thiết bị có kết nối Internet. Thiết bị này có thể là các máy tính, điện thoại thông minh hoặc các thiết bị IoT. Bot thường được các nhóm tin tặc tạo ra được gọi là *botmaster*. Khi bot lây nhiễm và hoạt động trên một thiết bị, nó cho phép botmaster điều khiển thiết bị từ xa. Botmaster thường sử dụng một hệ thống điều khiển, được gọi là các máy chủ chỉ huy và điều khiển (Command and Control - C&C) để điều khiển và duy trì botnet [4][5][6]. Ở một phía, botmaster gửi các lệnh và mã cập nhật lên máy chủ C&C của botnet. Ở phía ngược lại, các bot trong botnet được lập trình để sử dụng các kênh truyền thông để nhận lệnh và mã cập nhật từ các máy chủ C&C. Các bot cũng gửi trạng thái hoạt động của chúng cho các máy chủ C&C.

Các bot trong một botnet định kỳ gửi các truy vấn DNS chứa các tên miền của máy chủ C&C đến hệ thống DNS cục bộ để tìm địa chỉ IP để kết nối đến máy chủ C&C. Để tránh việc các máy chủ C&C bị đưa vào danh sách đen và bị vô hiệu hóa nếu sử dụng tên và địa chỉ IP tĩnh, botmaster thường sử dụng một kỹ thuật đặt biệt được gọi là *Fast Flux* (FF) hoặc *Domain Generation Algorithm* (DGA) để tự động sinh và đăng ký các tên miền cho máy chủ C&C của botnet [4][5][6]. Các bot trong botnet cũng được trang bị khả năng tự sinh các tên miền sử dụng cùng kỹ thuật DGA như bên máy chủ, sau đó tạo và gửi truy vấn tên miền đến hệ thống DNS cục bộ. Nhờ vậy, các bot trong botnet vẫn có thể tìm được địa chỉ IP của máy chủ C&C để kết nối, như biểu diễn trên Hình 1. Do kỹ thuật DGA được sử dụng rất phổ biến trong các botnet, nên các botnet sử dụng kỹ thuật này được gọi là DGA botnet.

Tác giả liên hệ: Hoàng Xuân Dậu,
Email: dauhx@ptit.edu.vn
Đến tòa soạn: 02/2022, chỉnh sửa: 03/2022, chấp nhận đăng: 04/2022.




*Hình 1. Ví dụ một botnet sử dụng kỹ thuật DGA để sinh, đăng ký và truy vấn tên miền máy chủ C&C*

Có thể chia các kỹ thuật DGA thành 3 dạng dựa trên phương pháp sử dụng để sinh các tên miền của chúng: (1) character-based, (2) word-based và (3) mixed DGA [7]. Kỹ thuật character-based DGA thường sử dụng tổ hợp ngẫu nghiên các ký tự tiếng Anh để tạo tên miền. Các botnet, như *cryptolocker*, *emotet* và *feodo* là các botnet điển hình sử dụng character-based DGA để sinh tên miền. Ngược lại, kỹ thuật word-based DGA thường sử dụng tổ hợp các từ tiếng Anh lấy từ các danh sách danh từ, động từ, tính từ dựng sẵn để tạo tên miền. Các tên miền được tạo sử dụng kỹ thuật word-based DGA thường có nghĩa và rất giống các tên miền bình thường. Các botnet, như *bigviktor*, *matsnu* và *suppobox* là các botnet điển hình sử dụng kỹ thuật word-based DGA để sinh tên miền. Kỹ thuật mixed DGA sử dụng kết hợp cả hai kỹ thuật character-based và word-based DGA. Theo đó, một phần tên miền được sinh sử dụng kỹ thuật character-based và phần còn lại được sinh sử dụng kỹ thuật word-based. *Banjori* là botnet điển hình sử dụng kỹ thuật mixed DGA để sinh các tên miền. BẢNG I. cung cấp một số dạng DGA botnet và các mẫu tên miền tự sinh.

BẢNG I. MỘT SỐ DẠNG DGA BOTNET VÀ CÁC MẪU TÊN MIỀN

| Họ botnets | Kỹ thuật GA | Mẫu tên miền |
|---|---|---|
| crypto-locker | character-based | ryojulmtdxljnkn.biz<br>icfpkabnmsse.org<br>kynkbkflfrlqcx.biz |
| emotet | character-based | affvqugewqpbcbic.eu<br>amxecvgvhfequgpo.eu<br>atqanjgnftfsnywb.eu |
| feodo | character-based | hmvmgywkvayilcwh.ru<br>xvmzegestulhtvqz.ru<br>hjpyvexsutdctjol.ru |
| bigviktor | word-based | knowredpermit.art<br>winstilllandscape.club<br>helppurpledistance.fans |
| matsnu | word-based | row-closed-bid.com<br>brushpot-guide.com<br>sort-address.com |
| suppobox | word-based | necessarypower.net<br>pleasantcountry.net<br>necessarycountry.net |
| banjori | mixed | ztgxrasildeafeninguvuc.com<br>vdrgrasildeafeninguvuc.com<br>umdhrasildeafeninguvuc.com |

Do tính chất nguy hiểm của các botnet, nhiều giải pháp giám sát và phát hiện botnet nói chung và phát hiện DGA botnet nói riêng đã được nghiên cứu, đề xuất. Có thể nêu ra một số đề xuất tiêu biểu như Hoàng và cộng sự [4][7][8], Truong và cộng sự [9], Charan và cộng sự [13]... Đây là các đề xuất cho nhiều kết quả hứa hẹn do đạt được độ chính xác cao và tỷ lệ cảnh báo sai thấp. Mặc dù vậy, mỗi đề xuất chỉ hoạt động tốt trên một dạng DGA botnet, hoặc một tập dữ liệu cụ thể. Chẳng hạn, Hoàng và cộng sự [4][8] và Truong và cộng sự [9] chỉ có khả năng phát hiện tốt các họ character-based DGA botnet mà không thể phát hiện các word-based, hoặc mixed DGA botnet. Ngược lại, Hoàng và cộng sự [7] và Charan và cộng sự [13] lại có khả năng phát hiện tốt các họ word-based DGA botnet, nhưng không thực sự hiệu quả với character-based hoặc mixed DGA botnet. Nhằm giải quyết vấn đề trên, bài báo này đề xuất mô hình phát hiện DGA botnet nhằm kết hợp nhiều mô hình phát hiện riêng lẻ, cho phép phát hiện hiệu quả nhiều dạng DGA botnet.

Phần tiếp theo của bài báo được bố cục như sau: phần II khảo sát một số nghiên cứu có liên quan, phần III mô tả mô hình phát hiện DGA botnet đề xuất dựa trên học kết hợp, phần IV trình bày các thử nghiệm và kết quả và phần V là kết luận của bài báo.

## II. CÁC NGHIÊN CỨU LIÊN QUAN

Mục này khảo sát một số nghiên cứu có liên quan đến phát hiện các dạng DGA botnet dựa trên học máy và học sâu, bao gồm Hoàng và cộng sự [4][7][8], Truong và cộng sự [9], Qiao và cộng sự [11] và Charan và cộng sự [14].

Hoàng và cộng sự [8] đề xuất một phương pháp phát hiện botnet dựa trên học máy sử dụng phân tích truy vấn DNS. Bài báo sử dụng các kỹ thuật học máy có giám sát, bao gồm Naive Bayes, kNN, cây quyết định và rừng ngẫu nhiên để xây dựng các mô hình phát hiện cho phân loại các tên miền sinh và sử dụng bởi botnet và các tên miền bình thường. Mô hình đề xuất trích xuất 18 đặc trưng phân loại cho mỗi tên miền, bao gồm 16 đặc trưng thống kê n-gram, 1 đặc trưng phân bố các nguyên âm trong tên miền và 1 đặc trưng entropy của các ký tự trong tên miền. Các kết quả thử nghiệm cho thấy hầu hết các kỹ thuật học máy đều cho độ chính xác phát hiện khả quan, trong đó thuật toán rừng ngẫu nhiên cho độ chính xác cao nhất (đạt trên 90%) và tỷ lệ cảnh báo sai thấp nhất. Hạn chế của mô hình đề xuất là chỉ có khả năng phát hiện các character-based DGA, tỷ lệ cảnh báo sai còn khá cao (đến 9,30%) và tập dữ liệu thử nghiệm tương đối nhỏ, có thể ảnh hưởng đến độ tin cậy của kết quả.

Trong một nghiên cứu tiếp theo, Hoàng và cộng sự [4] đề xuất một mô hình cải tiến của [8] sử dụng thuật toán học máy rừng ngẫu nhiên nhằm tăng độ chính xác phát hiện và giảm tỷ lệ cảnh báo sai. Mô hình đề xuất sử dụng một tập gồm 24 đặc trưng phân loại cho mỗi tên miền, bao gồm 16 đặc trưng thống kê n-gram, 6 đặc trưng thống kê từ vựng cho nguyên âm, phụ âm, chữ số và 2 đặc trưng entropy cho ký tự và giá trị kỳ vọng cho tên miền. Các thử nghiệm trên tập dữ liệu gồm 100.000 tên miền bình thường và 153.000 tên miền DGA cho thấy, mô hình đề xuất đạt độ chính xác trên 97% và tỷ lệ cảnh bao sai khoảng 3%. Mô hình cũng có khả năng phát hiện hiệu quả hầu hết các họ DGA botnet.

Tuy vậy, hạn chế lớn nhất của mô hình đề xuất là nó không thể phát hiện các họ word-based, hoặc mixed DGA botnet.

Cũng nhằm phát hiện các character-based DGA, Truong và cộng sự [9] đề xuất một phương pháp phát hiện các botnet tự động sinh các tên miền dựa trên các đặc trưng lưu lượng DNS. Nghiên cứu sử dụng các đặc trưng tên miền, bao gồm độ dài và giá trị mong đợi của tên miền để phân biệt các tên miền sinh tự động (PDN) và tên miền bình thường. Giá trị mong đợi của tên miền được tính toán dựa trên phân bố ký tự của 100.000 tên miền thông dụng nhất trên bảng xếp hạng của Alexa [16]. Năm thuật toán học máy, bao gồm Naive Bayes, kNN, SVM, cây quyết định và rừng ngẫu nhiên được sử dụng để xây dựng các bộ phân loại. Kết quả thử nghiệm trên tập dữ liệu 100.000 tên miền bình thường và 20.000 tên miền PDN cho thấy, phương pháp đề xuất đạt độ chính xác phát hiện chung cao nhất là 92,30% với thuật toán cây quyết định. Mặc dù độ chính xác phát hiện của phương pháp đề xuất khá cao, nhưng tỷ lệ cảnh báo sai tổng cũng tương đối cao, khoảng 7,70% trong trường hợp tốt nhất. Ngoài ra, phương pháp đề xuất cũng không thể phát hiện các họ word-based, hoặc mixed DGA botnet.

Theo hướng sử dụng học sâu, Qiao và cộng sự [11] đề xuất phương pháp phân loại các tên miền DGA sử dụng kỹ thuật học sâu LSTM với cơ chế chú ý. Trong phương pháp đề xuất, mỗi tên miền được đưa qua quá trình tiền xử lý gồm các bước: tách, đệm và nhúng các chuỗi ký tự. Tên miền sau đó được chuyển đổi thành một ma trận 54x128 cho huấn luyện và kiểm thử. Các thử nghiệm trên tập dữ liệu gồm 1 triệu tên miền thông dụng nhất theo xếp hạng của Alexa [16] và 1.675.404 tên miền DGA cho thấy phương pháp đề xuất đạt độ đo F1 trung bình là 94,58%. Ưu điểm của phương pháp là đạt độ chính xác cao và loại bỏ được quá trình trích chọn các đặc trưng. Tuy nhiên, phương pháp đề xuất cũng chỉ phát hiện tốt các tên miền character-based DGA. Ngoài ra, tỷ lệ cảnh báo sai tổng cũng còn tương đối cao, khoảng 5% tính theo độ đo F1.

Tập trung phát hiện các word-based DGA botnet, Hoàng và cộng sự [7] đề xuất một mô hình phát hiện word-based DGA botnet dựa trên học máy có giám sát. Các tác giả đề xuất trích xuất 16 đặc trưng phân loại cho mỗi tên miền, bao gồm các đặc trưng thống kê từ loại, phân bố nguyên âm, chữ số, ký tự đặc biệt và độ dài tên miền. Các danh sách chứa các danh từ, động từ và tính từ tiếng Anh mà các word-based DGA botnet thường sử dụng cũng được xây dựng để hỗ trợ trích xuất các đặc trưng. Nhiều kịch bản thử nghiệm đã được thực hiện và các kết quả khẳng định, mô hình dựa trên cây quyết định J48 cho kết quả phát hiện các tên miền word-based DGA tốt nhất với độ đo F1 đạt 97,01%. Mô hình đề xuất cũng có khả năng phát hiện tương đối tốt nhiều dạng tên miền character-based DGA. Tuy vậy, mô hình đề xuất có khả năng phát hiện không thực sự tốt với một số dạng tên miền character-based DGA, như *Necurs* và *Ramnit*.

Cũng theo hướng phát hiện các word-based DGA botnet, Charan và cộng sự [14] đề xuất một phương pháp mới để phát hiện các tên miền word-based DGA dựa trên các thuật toán học kết hợp. Phương pháp đề xuất sử dụng 15 đặc trưng từ vựng và đặc trưng mạng của tên miền để xây dựng mô hình phát hiện. Một số thuật toán học máy có giám sát như cây quyết định C4.5, C5.0, CART và rừng ngẫu nhiên được sử dụng trong các mô hình kết hợp để cải thiện hiệu năng phát hiện. Các kết quả thử nghiệm xác nhận, mô hình dựa trên cây quyết định C5.0 cho hiệu năng cao nhất, với độ chính xác phát hiện là 95.03%. Vấn đề chính đối với phương pháp đề xuất là không có mô tả chi tiết về phương pháp mô hình kết hợp sử dụng để tổng hợp kết quả từ các mô hình thành phần.

BẢNG II. cung cấp so sánh độ chính xác phát hiện (ACC) và độ đo F1, cũng như ưu nhược điểm của một số đề xuất phát hiện DGA botnet đã có. Đa số đề xuất đều không thể phát hiện hiệu quả hầu hết các DGA botnet, bao gồm các character-based và word-based DGA botnet. Bài báo này đề xuất một mô hình dựa trên học kết hợp cho phép phát hiện hiệu quả cả character-based và word-based DGA botnet. Mô hình đề xuất kết hợp 2 mô hình thành phần, bao gồm mô hình phát hiện character-based DGA botnet (CDM) [4] và mô hình phát hiện word-based DGA botnet (WDM) [7] để cải thiện khả năng phát hiện cả 2 dạng character-based và word-based DGA botnet.

BẢNG II. SO SÁNH MỘT SỐ ĐỀ XUẤT PHÁT HIỆN DGA BOTNET

| Đề xuất | ACC | F1 | Ưu điểm | Hạn chế |
|---|---|---|---|---|
| Truong và cộng sự [9] | 92,30 | | Đơn giản và nhanh | - Tỷ lệ phát hiện sai cao (khoảng 7.70%)<br>- Không thể phát hiện word-based DGA botnet |
| Hoàng và cộng sự [8] | 90,90 | 90,90 | Tương đối đơn giản và nhanh | - Tập dữ liệu nhỏ<br>- Tỷ lệ phát hiện sai cao (khoảng 9,30%)<br>- Không thể phát hiện word-based DGA botnet |
| Hoàng và cộng sự [4] | 97,03 | 97,03 | - Độ chính xác cao<br>- Tỷ lệ phát hiện sai thấp | Không thể phát hiện word-based DGA botnet |
| Qiao và cộng sự [11] | 94,58 | | Độ chính xác khá cao | - Tỷ lệ phát hiện sai còn cao (khoảng 5%)<br>- Không thể phát hiện word-based DGA botnet |
| Charan và cộng sự [14] | 95,03 | | - Độ chính xác cao<br>- Có thể phát hiện word-based DGA botnet | Các mô hình học kết hợp không được mô tả tường minh. |
| Hoàng và cộng sự [7] | 96,99 | 97,01 | - Độ chính xác cao<br>- Có thể phát hiện word-based DGA botnet | - Phát hiện một số character-based DGA botnet kém hiệu quả<br>- Không thể phát hiện mixed DGA botnet |

## III. MÔ HÌNH PHÁT HIỆN DGA BOTNET DỰA TRÊN HỌC KẾT HỢP

### *A. Khái quát về học kết hợp*

Học kết hợp (ensemble learning) hay còn gọi là học theo nhóm là một cách tiếp cận trong học máy nhằm tìm kiếm hiệu suất dự đoán tốt hơn bằng cách kết hợp các dự đoán từ nhiều mô hình thành phần. Có nhiều phương pháp kết

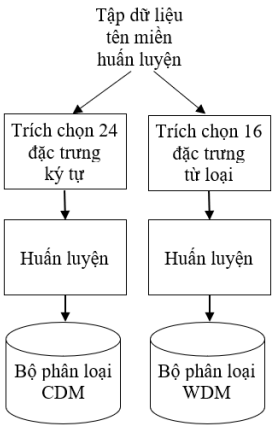
hợp các dự đoán từ các mô hình thành phần, bao gồm các phương pháp kết hợp đơn giản và phương pháp kết hợp phức tạp [12]. Các phương pháp học kết hợp đơn giản bao gồm: biểu quyết tối đa (max voting), trung bình (averaging), trung bình có trọng số (weighted averaging). Max voting là phương pháp biểu quyết tối đa thường được sử dụng cho các vấn đề phân loại. Trong kỹ thuật này, nhiều mô hình được sử dụng để đưa ra dự đoán cho mỗi điểm dữ liệu. Các dự đoán của mỗi mô hình được coi như một "phiếu bầu". Các dự đoán nhận được từ phần lớn các mô hình được sử dụng làm dự đoán cuối cùng.

Averaging tương tự như kỹ thuật biểu quyết tối đa, trong đó nhiều dự đoán được thực hiện cho mỗi điểm dữ liệu trong tính trung bình. Trong phương pháp này, lấy trung bình các dự đoán từ tất cả các mô hình và sử dụng nó để đưa ra dự đoán cuối cùng. Trung bình có thể được sử dụng để đưa ra dự đoán trong các bài toán hồi quy hoặc trong khi tính toán xác suất cho các bài toán phân loại. Weighted averaging là một phần mở rộng của phương pháp trung bình. Tất cả các mô hình được ấn định các trọng số khác nhau xác định tầm quan trọng của từng mô hình để dự đoán.

Các phương pháp học kết hợp phức tạp được thừa nhận và sử dụng rộng rãi bao gồm [12]: (i) đóng gói (bagging) là việc kết hợp nhiều cây quyết định trên các mẫu khác nhau của cùng một tập dữ liệu và tính trung bình các dự đoán; (ii) xếp chồng (stacking) là việc kết hợp nhiều loại mô hình khác nhau trên cùng một tập dữ liệu và sử dụng một mô hình khác để kết hợp tốt nhất các dự đoán; (iii) tăng cường (boosting) liên quan đến việc thêm các thành viên tổng hợp một cách tuần tự để sửa các dự đoán được thực hiện bởi các mô hình trước đó và xuất ra giá trị trung bình có trọng số của các dự đoán.

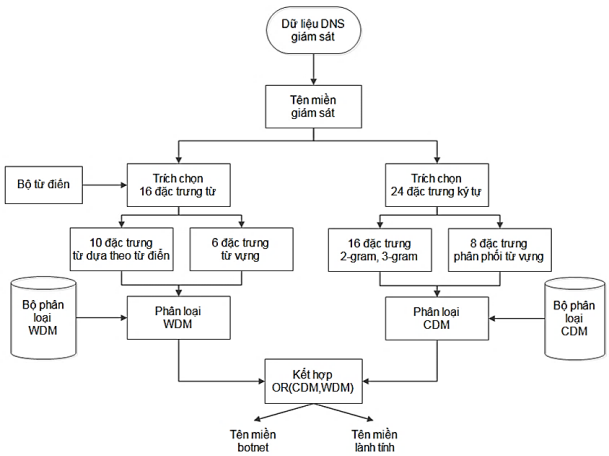
### *B. Giới thiệu mô hình đề xuất*

Như đã đề cập trong phần II, mô hình phát hiện DGA botnet đề xuất là sự kết hợp của 2 mô hình thành phần: CDM [4] và WDM [7] nhằm khai thác điểm mạnh của cả 2 mô hình thành phần, cho phép phát hiện hiệu quả cả 2 dạng character-based và word-based DGA botnet. Mô hình đề xuất được triển khai thành 2 giai đoạn: giai đoạn huấn luyện và giai đoạn phát hiện. Trong giai đoạn huấn luyện như biểu diễn trên Hình 2, từng mô hình thành phần (CDM và WDM) được huấn luyện riêng từ các tập dữ liệu tên miền huấn luyện. Việc tiền xử lý dữ liệu, trích chọn các đặc trưng và huấn luyện được thực hiện theo [4] với mô hình CDM và được thực hiện theo [7] với mô hình WDM. Mô hình CDM được huấn luyện sử dụng thuật toán rừng ngẫu nhiên và mô hình WDM được huấn luyện sử dụng thuật toán cây quyết định J48. Đây là các thuật toán học máy có giám sát có tốc độ xử lý nhanh và đã được chứng minh có hiệu quả cao trong các bài toán phân loại [4][6][8][9]. Kết quả của giai đoạn này là 2 bộ phân loại CDM và WDM sử dụng cho giai đoạn phát hiện.




*Hình 2. Mô hình đề xuất: Giai đoạn huấn luyện*

Trong giai đoạn phát hiện như biểu diễn trên Hình 3, các mô hình hay bộ phân loại thành phần (CDM và WDM) được sử dụng để xử lý riêng từng tên miền giám sát. Qui trình tiền xử lý và phân loại mỗi tên miền giám sát được thực hiện thực hiện theo [4] với mô hình CDM và được thực hiện theo [7] với mô hình WDM. Hai kết quả xử lý của 2 bộ phân loại CDM và WDM được kết hợp để tăng hiệu suất phát hiện chung. Phép toán OR(CDM, WDM) là phép hợp các phát hiện chính xác từ 2 bộ phân loại, kết quả cho số lượng tối đa các DGA bonet được phát hiện. Phép hợp (OR) là một dạng cụ thể của kỹ thuật Max voting đã trình bày ở trên.




*Hình 3. Mô hình đề xuất: Giai đoạn phát hiện*

Chi tiết về khâu tiền xử lý, trích chọn các đặc trưng phân loại tên miền được thực hiện theo [4] với 24 đặc trưng ký tự cho phân loại các tên miền character-based DGA và được thực hiện theo [7] với 16 đặc trưng từ cho phân loại các tên miền word-based DGA.

### *C. Các độ đo đánh giá*

Bài báo sử dụng 6 độ đo đánh giá mô hình, gồm PPV (Positive Predictive Value, Độ chính xác), TPR (True Positive Rate, hoặc Recall, Độ nhạy), FPR (False Positive Rate, Tỷ lệ dương tính giả), FNR (False Negative Rate, Tỷ lệ âm tính giả), F1 (F1-Score) và ACC (Overall Accuracy, Độ chính xác chung) để đánh giá hiệu năng của mô hình đề xuất. Các độ đo được tính theo các công thức sau:

$$PPV = \frac{TP}{TP + FP} \quad (1)$$

$$TPR = \frac{TP}{TP + FN} \quad (2)$$

$$FPR = \frac{FP}{FP + TN} \quad (3)$$

$$FNR = \frac{FN}{FN + TP} \quad (4)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (5)$$

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \quad (6)$$

trong đó, TP, FP, FN and TN là các tham số của ma trận nhầm lẫn cho trên BẢNG III.

BẢNG III. TP, FP, FN AND TN CỦA MA TRẬN NHẦM LẪN

| | | Lớp thực tế | |
|---|---|---|---|
| | | *Tên miền botnet* | *Tên miền lành tính* |
| **Lớp dự đoán** | *Tên miền botnet* | TP (True Positives) | FP (False Positives) |
| | *Tên miền lành tính* | FN (False Negatives) | TN (True Negatives) |

Ngoài ra, bài báo còn sử dụng tỷ lệ phát hiện (DR-Detection Rate) để đo lường hiệu quả của mô hình phát hiện đề xuất cho phân loại tên miền của các họ botnet khác nhau. DR cho mỗi loại botnet được tính như sau:

$$DR = \frac{NoDB}{NoTest} \quad (7)$$

trong đó, NoDB là số tên miền của một botnet được phát hiện đúng và NoTest là tổng số tên miền của botnet đó được đưa vào kiểm tra.

## IV. THỬ NGHIỆM VÀ KẾT QUẢ

### A. *Tập dữ liệu thử nghiệm*

Tập dữ liệu thử nghiệm gồm 1 tập dữ liệu cho huấn luyện và 1 tập dữ liệu cho kiểm thử. Tập dữ liệu cho huấn luyện lại gồm 2 phần: tập huấn luyện cho mô hình CDM (gọi là CDM-TrainSet) và tập huấn luyện cho mô hình WDM (gọi là WDM-TrainSet). Tập CDM-TrainSet gồm 100.000 tên miền bình thường phổ biến nhất xếp hạng bởi Alexa [16] và 100.000 tên miền character-based DGA botnet từ NetLab360 [17] với thành phần chi tiết cho như trên BẢNG IV. Tập WDM-TrainSet gồm 48.000 tên miền bình thường phổ biến nhất xếp hạng bởi Alexa [16] và 48.000 tên miền word-based DGA botnet từ NetLab360 [17] với thành phần chi tiết cho như trên BẢNG V.

BẢNG IV. THÀNH PHẦN CỦA TẬP CDM-TRAINSET

| Họ tên miền | Loại tên miền | Số lượng |
|---|---|---|
| emotet | Character-based | 10.000 |
| rovnix | Character-based | 10.000 |
| tinba | Character-based | 10.000 |
| pykspa_v1 | Character-based | 15.000 |
| ramnit | Character-based | 10.000 |
| gameover | Character-based | 8.000 |
| simda | Character-based | 14.000 |
| ranbyus | Character-based | 4.000 |
| murofet | Character-based | 4.000 |
| necurs | Character-based | 4.000 |
| shiotob | Character-based | 4.000 |
| virut | Character-based | 4.000 |
| symmi | Character-based | 3.000 |
| Bình thường | Lành tính | 100.000 |
| **Tổng cộng** | | **200.000** |

BẢNG V. THÀNH PHẦN CỦA TẬP WDM-TRAINSET

| Họ tên miền | Loại tên miền | Số lượng |
|---|---|---|
| Bigviktor | Word-based | 12.000 |
| Matsnu | Word-based | 12.000 |
| Suppobox | Word-based | 12.000 |
| Pizd | Word-based | 12.000 |
| Bình thường | Lành tính | 48.000 |
| **Tổng cộng** | | **96.000** |

Tập dữ liệu cho kiểm thử gồm 71.393 tên miền của 39 họ botnet để tính toán tỷ lệ phát hiện các mô hình thành phần và mô hình kết hợp, chi tiết như trên BẢNG VI.

BẢNG VI. CHI TIẾT TẬP DỮ LIỆU KIỂM THỬ

| TT | Họ DGA botnet | Số lượng tên miền | TT | Họ DGA botnet | Số lượng tên miền |
|---|---|---|---|---|---|
| 1 | Rovnix | 4000 | 21 | Qadars | 2000 |
| 2 | Dyre | 1000 | 22 | Simda | 4000 |
| 3 | Chinad | 1000 | 23 | Pykspa_v2_fake | 799 |
| 4 | Fobber_v1 | 298 | 24 | Locky | 1158 |
| 5 | Tinynuke | 32 | 25 | Pykspa_v2_real | 199 |
| 6 | Gameover | 4000 | 26 | Shifu | 2546 |
| 7 | Murofet | 4000 | 27 | Matsnu | 881 |
| 8 | Cryptolocker | 1000 | 28 | Proslikefan | 100 |
| 9 | Padcrypt | 168 | 29 | Tempedreve | 195 |
| 10 | Dircrypt | 762 | 30 | Vawtrak | 827 |
| 11 | Fobber_v2 | 299 | 31 | Symmi | 1200 |
| 12 | Vidro | 100 | 32 | Suppobox | 2205 |
| 13 | Emotet | 4000 | 33 | Nymaim | 480 |
| 14 | Tinba | 4000 | 34 | Mydoom | 50 |
| 15 | Ranbyus | 4000 | 35 | Conficker | 495 |
| 16 | Shiotob | 4000 | 36 | Bigviktor | 999 |
| 17 | Pykspa_v1 | 4000 | 37 | Gspy | 100 |
| 18 | Necurs | 4000 | 38 | Enviserv | 500 |
| 19 | Ramnit | 4000 | 39 | Banjori | 4000 |
| 20 | Virut | 4000 | | | |

### B. *Các kết quả*

Để tính các độ đo PPV, TPR, FPR, FNR, F1 và ACC, bài báo sử dụng phương pháp kiểm tra chéo 10 lần, lấy kết quả trung bình, với 80% tập dữ liệu CDM-TRAINSET và WDM-TRAINSET cho huấn luyện xây dựng mô hình và 20% dữ liệu còn lại cho kiểm tra. BẢNG VII. biểu diễn các độ phát hiện của 2 mô hình thành phần CDM (dựa trên thuật toán rừng ngẫu nhiên - RF) và WDM (dựa trên thuật toán cây quyết định J48) so sánh với kết quả của các nghiên cứu trước đó. Có thể thấy các mô hình CDM và WDM đạt độ chính xác phát hiện và tỷ lệ cảnh báo sai tốt hơn đáng kể so với các độ đo này của các nghiên cứu trước đó. BẢNG VIII. mô tả tỷ lệ phát hiện (DR) trên 39 họ DGA botnet của mô hình phát hiện dựa trên học kết hợp đề xuất với hai mô hình thành phần. Có thể thấy mô hình kết hợp đã khai thác được các ưu điểm của 2 mô hình thành phần và có khả năng phát hiện hiệu quả hầu hết các DGA botnet.

BẢNG VII. HIỆU NĂNG PHÁT HIỆN CỦA CÁC MÔ HÌNH CDM, WDM VÀ CÁC NGHIÊN CỨU ĐÃ CÓ

| Mô hình | PPV | TPR | FPR | FNR | ACC | F1 |
|---|---|---|---|---|---|---|
| Hoàng và | 90.70 | 91.00 | 9.30 | | 90.90 | 90.90 |

| | | | | | | |
|---|---|---|---|---|---|---|
| cộng sự [8] | | | | | | |
| Truong và cộng sự [9] | 94.70 | | 4.80 | | 92.30 | |
| Qiao và cộng sự [11] | 95.05 | 95.14 | | | | 94.58 |
| Charan và cộng sự [14] | | | | | 95.03 | |
| CDM-RF | 97.08 | 96.98 | 2.92 | 3.02 | 97.03 | 97.03 |
| WDM-J48 | 98.25 | 95.81 | 1.78 | 4.19 | 96.99 | 97.01 |

BẢNG VIII. TỶ LỆ PHÁT HIỆN (DR) CỦA MÔ HÌNH ĐỀ XUẤT VÀ CÁC MÔ HÌNH CDM, WDM

| TT | Họ DGA botnet | Loại DGA | Tỷ lệ phát hiện (DR %) | | |
|---|---|---|---|---|---|
| | | | CDM | WDM | Kết hợp |
| 1 | Rovnix | char-based | 100,00 | 99,50 | 100,00 |
| 2 | Dyre | char-based | 100,00 | 99,90 | 100,00 |
| 3 | Chinad | char-based | 100,00 | 97,90 | 100,00 |
| 4 | Fobber_v1 | char-based | 100,00 | 100,00 | 100,00 |
| 5 | Tinynuke | char-based | 100,00 | 100,00 | 100,00 |
| 6 | Gameover | char-based | 100,00 | 99,98 | 100,00 |
| 7 | Murofet | char-based | 99,80 | 99,78 | 100,00 |
| 8 | Cryptolocker | char-based | 99,70 | 96,20 | 100,00 |
| 9 | Padcrypt | char-based | 98,21 | 98,21 | 100,00 |
| 10 | Dircrypt | char-based | 99,34 | 93,31 | 100,00 |
| 11 | Fobber_v2 | char-based | 100,00 | 89,30 | 100,00 |
| 12 | Vidro | char-based | 100,00 | 49,00 | 100,00 |
| 13 | Emotet | char-based | 99,68 | 99,55 | 99,98 |
| 14 | Tinba | char-based | 99,98 | 99,08 | 99,98 |
| 15 | Ranbyus | char-based | 99,58 | 99,30 | 99,93 |
| 16 | Shiotob | char-based | 99,68 | 95,95 | 99,88 |
| 17 | Pykspa_v1 | char-based | 99,70 | 58,90 | 99,83 |
| 18 | Necurs | char-based | 99,35 | 87,75 | 99,78 |
| 19 | Ramnit | char-based | 99,55 | 91,45 | 99,75 |
| 20 | Virut | char-based | 99,75 | 0,00 | 99,75 |
| 21 | Qadars | char-based | 99,05 | 95,40 | 99,65 |
| 22 | Simda | char-based | 99,65 | 0,00 | 99,65 |
| 23 | Suppobox | word-based | 19,27 | 99,30 | 99,30 |
| 24 | Pykspa_v2_fake | char-based | 98,87 | 61,08 | 99,25 |
| 25 | Locky | char-based | 99,05 | 83,16 | 99,14 |
| 26 | Pykspa_v2_real | char-based | 98,99 | 63,32 | 98,99 |
| 27 | Shifu | char-based | 98,59 | 34,92 | 98,82 |
| 28 | Matsnu | word-based | 12,15 | 98,41 | 98,64 |
| 29 | Proslikefan | char-based | 98,00 | 50,00 | 98,00 |
| 30 | Tempedreve | char-based | 97,44 | 64,62 | 97,44 |
| 31 | Vawtrak | char-based | 96,61 | 61,67 | 97,10 |
| 32 | Symmi | char-based | 96,58 | 31,67 | 96,83 |
| 33 | Bigviktor | word-based | 11,11 | 96,78 | 96,78 |
| 34 | Nymaim | char-based | 94,79 | 61,25 | 95,21 |
| 35 | Mydoom | char-based | 88,00 | 74,00 | 94,00 |
| 36 | Gspy | char-based | 91,00 | 8,00 | 91,00 |
| 37 | Conficker | char-based | 89,29 | 52,93 | 89,49 |
| 38 | Enviserv | char-based | 76,00 | 19,40 | 76,00 |
| 39 | Banjori | mixed | 0,00 | 0,00 | 0,00 |

*C. Nhận xét*

Từ các kết quả thử nghiệm cho trên các bảng BẢNG VII. và BẢNG VIII. có thể rút ra một số nhận xét sau:

- Mô hình CDM sử dụng tập 24 đặc trưng ký tự và được huấn luyện dựa trên tập dữ liệu bao gồm các tên miền chủ yếu thuộc họ character-based DGA nên có khả năng phát hiện hiệu quả các tên miền của các họ character-based DGA botnet. Kết quả thử nghiệm cho thấy CDM có khả năng phát hiện tốt hầu hết các tên miền character-based DGA với độ chính xác cao và tỷ lệ sai trung bình khoảng 3%. Tuy nhiên, CDM không có khả năng phát hiện các tên miền word-based DGA.

- Mô hình WDM sử dụng tập 16 đặc trưng từ và được huấn luyện dựa trên tập dữ liệu bao gồm các tên miền thuộc họ word-based DGA nên có khả năng phát hiện hiệu quả các tên miền của các họ word-based DGA botnet. Kết quả thử nghiệm cho thấy WDM có khả năng phát hiện tốt hầu hết các tên miền word-based DGA với độ chính xác cao và tỷ lệ sai trung bình khoảng 3%. Tuy vậy, WDM không thể phát hiện, hoặc có khả năng phát hiện hạn chế một số họ character-based DGA botnet, như *virus, simda, gspy* và *enviserv*.

- Mô hình phát hiện đề xuất dựa trên học kết hợp có khả năng phát hiện hiệu quả hầu hết các DGA botnet, bao gồm cả các họ character-based và word-based DGA do có khả năng kết hợp ưu điểm của các mô hình CDM và WDM. Kết quả trên BẢNG VII. cho thấy, mô hình kết hợp có khả năng phát hiện hiệu quả 37/39 họ DGA botnet với DR > 89%, trong đó có 12 botnet đạt DR = 100%, 31 botnet đạt DR > 97%.

- Hạn chế của mô hình đề xuất là không phát hiện được mixed DGA botnet, như *banjori* do cả mô hình CDM và WDM đều không thể phát hiện botnet này. Một tồn tại khác của mô hình kết hợp là thời gian huấn luyện và phát hiện dài hơn do phải xử lý song song trên cả hai mô hình thành phần. Mặc dù vậy, quá trình huấn luyện các mô hình thành phần có thể thực hiện offline nên sẽ không ảnh hưởng quá lớn đến hiệu năng phát hiện của hệ thống.

## V. KẾT LUẬN

Bài báo này đề xuất mô hình phát hiện các dạng DGA botnet dựa trên học kết hợp. Mô hình đề xuất kết hợp hai mô hình thành phần trong các nghiên cứu trước của chúng tôi là CDM và WDM nhằm nâng cao khả năng phát hiện các họ tên miền DGA botnet. Các kết quả thử nghiệm cho thấy mô hình đề xuất có độ chính xác cao hơn và tỷ lệ phát hiện sai thấp hơn đáng kể so với các các nghiên cứu đã có. Mô hình đề xuất có khả năng phát hiện hiệu quả nhiều dạng DGA botnet, bao gồm cả các họ character-based và word-based DGA do mô hình đề xuất có khả năng kết hợp ưu điểm của các mô hình CDM và WDM. Cụ thể, mô hình kết hợp có khả năng phát hiện hiệu quả 37/39 họ DGA botnet với DR > 89%, trong đó có 12 botnet đạt DR = 100%, 31 botnet đạt DR > 97%, 1 botnet đạt DR = 76% (*enviserv*) và chỉ không phát hiện được 1 botnet (*banjori*).

Trong tương lai, chúng tôi tiếp tục nghiên cứu cải tiến, tối ưu hóa mô hình kết hợp nhằm giải quyết hai vấn đề: (1) giảm thời gian huấn luyện và phát hiện (2) cải thiện khả năng phát hiện các mixed DGA botnet.

## LỜI CÁM ƠN

Các tác giả chân thành cảm ơn phòng Lab An toàn thông tin, Học viện Công nghệ Bưu chính viễn thông về các hỗ trợ phương tiện làm việc và các trang thiết bị thực thử nghiệm trong bài báo này.

## TÀI LIỆU THAM KHẢO

[1] Spamhaus Botnet Threat Update: Q1-2021. Available online: https://www.spamhaus.org/news/article/809/spamhaus-botnet-threat-update-q1-2021 (last accessed July 2021).

[2] AO Kaspersky Lab - Bots and botnets in 2018: Statistics on botnet attacks on clients of organizations. Available online: https://securelist.com/bots-and-botnets-in-2018/90091/ (last accessed July 2021).

[3] Radware Blog - More Destructive Botnets and Attack Vectors Are on Their Way. Available online: https://blog.radware.com/security/ botnets/2019/10/scan-exploit-control/ (last accessed July 2021).

[4] Hoang, X.D.; Vu, X.H., An improved model for detecting DGA botnets using random forest algorithm, Information Security Journal: A Global Perspective, July 2021, DOI: 10.1080/19393555.2021.1934198.

[5] Alieyan, K.; Almomani, A., Manasrah, A.; Kadhum, M.M., A survey of botnet detection based on DNS. Nat. Comput. Appl. Forum 2017, 28, 1541–1558.

[6] Li, X.; Wang, J.; Zhang, X., Botnet Detection Technology Based on DNS. J. Future Internet 2017, 9, 55.

[7] Hoang, X.D.; Vu, X.H. An Novel Machine Learning-based Approach for Detecting Word-based Botnets, Journal of Theoretical and Applied Information Technology, Vol. 99, No. 24, 2021.

[8] Hoang, X.D.; Nguyen, Q.C., Botnet Detection Based on Machine Learning Techniques Using DNS Query Data. J. Future Internet 2018, 10, 43; doi:10.3390/fi10050043.

[9] Truong, D.T; Cheng, G., Detecting domain-flux botnet based on DNS traffic features in managed network. Security Comm. Networks 2016; 9: 2338–2347; John Wiley & Sons.

[10] Qiao, Y., Zhang, B., Zhang, W., Sangaiah, A.K., and Wu, H., DGA Domain Name Classification Method Based on Long Short-Term Memory with Attention Mechanism. Appl. Sci. 2019, 9, 4205; doi:10.3390/app9204205.

[11] Yang, L.; Zhai, J.; Liu, W.; Ji, X.; Bai, H.; Liu, G.; Dai, Y., Detecting Word-Based Algorithmically Generated Domains Using Semantic Analysis. Symmetry 2019, 11, 176. https://doi.org/10.3390/sym11020176.

[12] Brownlee, J. A Gentle Introduction to Ensemble Learning Algorithms. 2021 [cited 2021; Available from: https://machinelearningmastery.com/tour-of-ensemble-learning-algorithms/.

[13] Charan, P.V.S., S.K. Shukla, and P.M. Anand. Detecting Word Based DGA Domains Using Ensemble Models. 2020. Cham: Springer International Publishing.

[14] Bijalwan, A., et al., Botnet Analysis Using Ensemble Classifier. Perspectives in Science, 2016. Volume 8: p. 502-504.

[15] Algelal, Z., et al., Botnet detection using ensemble classifiers of network flow. International Journal of Electrical and Computer Engineering, 2020. Volume 10: p. 2543-2550.

[16] DN Pedia – Top Alexa one million domains. Available online: https://dnpedia.com /tlds/topm.php (last accessed July 2021).

[17] Netlab 360 – DGA Families. Available online: https://data.netlab.360.com/dga/ (last accessed July 2021).

## A DGA BOTNET DETECTION MODEL BASED ON ENSEMBLE LEARNING

***Abstract***: Recently, DGA has been becoming a popular technique used by many malwares in general and botnets in particular. DGA allows hacking groups to automatically generate and register domain names for C&C servers of their botnets in order to avoid being blacklisted and disabled if using static domain names and IP addresses. Many types of sophisticated DGA techniques have been developed and used in practice, including character-based DGA, word-based DGA and mixed DGA. These techniques allow to generate from simple domain names of random combinations of characters, to complex domain names of combinations of meaningful words, which are very similar to legitimate domain names. This makes it difficult for solutions to monitor and detect botnets in general and DGA botnets in particular. Some solutions are able to efficiently detect character-based DGA domain names, but cannot detect word-based DGA and mixed DGA domain names. In contrast, some recent proposals can effectively detect word-based DGA domain names, but cannot effectively detect domain names of some character-based DGA botnets. This paper proposes a model based on ensemble learning that enables efficient detection of most DGA domain names, including character-based DGA and word-based DGA. The proposed model combines two component models, including the character-based DGA botnet detection model and the word-based DGA botnet detection model. The experimental results show that the proposed combined model is able to effectively detect 37/39 DGA botnet families with a detection rate of over 89%.

***Keywords:*** Character-based DGA botnet, Word-based DGA botnet, DGA botnet detection, DGA botnet detection based on ensemble learning.

**ThS. Vũ Xuân Hạnh** nhận bằng cử nhân ngành Khoa học máy tính năm 2002 tại Đại học Công nghệ, Đại học quốc gia Hà Nội. Năm 2015 ông nhận bằng thạc sĩ ngành Hệ thống thông tin tại Học viện Công nghệ Bưu chính Viễn thông. ThS. Vũ Xuân Hạnh hiện là giảng viên Khoa Công nghệ thông tin, Đại học Mở Hà Nội và là NCS tại Học viện Công nghệ Bưu chính Viễn thông. Các hướng nghiên cứu hiện nay của ông bao gồm: phát hiện tấn công, xâm nhập, phát hiện mã độc, bảo mật hệ thống và phần mềm, an ninh mạng.

**TS. Hoàng Xuân Dậu** nhận bằng kỹ sư tin học năm 1994 tại Đại học Bách khoa Hà Nội. Ông nhận bằng thạc sĩ và tiến sĩ khoa học máy tính lần lượt vào năm 2000 và 2006 tại Đại học RMIT, Australia. TS. Hoàng Xuân Dậu hiện là giảng viên chính tại Khoa Công nghệ thông tin 1, Học viện Công nghệ Bưu chính Viễn thông. Các hướng nghiên cứu hiện nay của ông bao gồm: phát hiện tấn công, xâm nhập, phát hiện mã độc, bảo mật hệ thống và phần mềm, bảo mật web và các ứng dụng dựa trên học máy cho an toàn thông tin.

**TS. Đinh Trường Duy** nhận bằng cử nhân năm 2014 tại Đại học Viễn thông St. Petersburg, Nga. Ông nhận bằng thạc sĩ và tiến sĩ lần lượt vào năm 2016 và 2020 tại Đại học Viễn thông St. Petersburg, Nga. TS. Đinh Trường Duy hiện là giảng viên chính tại Khoa Công nghệ thông tin 1, Học viện Công nghệ Bưu chính Viễn thông. Các hướng nghiên cứu hiện nay của ông bao gồm: phát hiện tấn công, xâm nhập, phát hiện mã độc, bảo mật và an toàn các mạng thế hệ mới WSN, 5G, ioT, blockchain.