**Bản thảo trước khi xuất bản**

Sự phát triển của tập dữ liệu bảo mật Internet of Things (IoT): Thách thức và định hướng tương lai

Đây là tệp PDF của một bài báo đã được chỉnh sửa sau khi chấp nhận, bao gồm việc thêm trang bìa, siêu dữ liệu và định dạng để tăng tính dễ đọc. Tuy nhiên, đây chưa phải là phiên bản chính thức. Phiên bản này sẽ tiếp tục được biên tập, dàn trang và rà soát trước khi xuất bản dưới dạng hoàn chỉnh.

Chúng tôi cung cấp phiên bản này để sớm giới thiệu nội dung bài báo. Xin lưu ý rằng trong quá trình sản xuất, có thể phát hiện ra lỗi ảnh hưởng đến nội dung, và mọi tuyên bố pháp lý áp dụng cho tạp chí vẫn có hiệu lực.

**Tóm tắt**

Sự phát triển của công nghệ di động đã mang lại các thiết bị thông minh và kết nối hơn vào cuộc sống hàng ngày của chúng ta. Xu hướng này, được gọi là Internet vạn vật (IoT), có ứng dụng trong nhà thông minh, thành phố thông minh, tự động hóa công nghiệp, hệ thống giám sát sức khỏe và đã trở thành một thành phần quan trọng trong ngành truyền thông và mạng.

Tuy nhiên, sự đa dạng của các tiêu chuẩn giao tiếp và giao thức, các cài đặt bảo mật mặc định yếu và khó khăn trong việc phân phối bản cập nhật đã làm gia tăng các mối đe dọa an ninh mạng đối với các ứng dụng quan trọng sử dụng IoT. Để giảm thiểu các mối đe dọa này và chống lại các cuộc tấn công, một hướng tiếp cận đầy hứa hẹn là phát triển một hệ thống phát hiện xâm nhập mạnh mẽ nhằm bảo vệ IoT.

Bài báo này trình bày nỗ lực của chúng tôi trong việc phân loại và so sánh các cuộc tấn công, tập dữ liệu, thuật toán và kiến trúc học máy được sử dụng trong hệ thống phát hiện xâm nhập dành cho thiết bị IoT. Chúng tôi phân loại các cuộc tấn công nhắm vào thiết bị IoT theo các lớp và giao thức khác nhau. Ngoài ra, nghiên cứu cũng làm nổi bật các đặc trưng tiềm năng có thể được sử dụng trong hệ thống phát hiện xâm nhập dựa trên học máy để nhận diện các loại tấn công khác nhau. Chúng tôi cũng cung cấp một nghiên cứu so sánh các tập dữ liệu IoT được s ử dụng để huấn luyện mô hình và xác định các yếu tố quan trọng giúp đánh giá mức độ phù hợp của chúng trong các tình huống cụ thể.

Cuối cùng, chúng tôi thảo luận về những quan sát của mình và đề xuất các hướng nghiên cứu nhằm xây dựng một hệ thống phát hiện xâm nhập IoT mạnh mẽ.

1. **Giới thiệu**

Sự phát triển của công nghệ di động đã mở đường cho sự ra đời của Internet vạn vật (IoT), cách mạng hóa nhiều lĩnh vực như chăm sóc sức khỏe, nhà ở và thành phố với cơ sở hạ tầng thông minh [1]. Các thiết bị thông minh này có thể được xem như các phần tử kết nối với nhau, được trang bị thẻ giao diện mạng, bộ xử lý nhẹ và được quản lý thông qua nhiều dịch vụ giao diện khác nhau (ví dụ: trang web, giao diện đồ họa, đăng nhập từ xa) [2].

Với nhu cầu ngày càng tăng đối với các thiết bị IoT, những mối đe dọa mới và độc đáo cũng đang ảnh hưởng đến chúng. Do đó, một bước quan trọng là phát triển phương pháp pháp chứng IoT để ghi nhận, điều tra hành vi tấn công và xác định vai trò của thiết bị trong các cuộc tấn công mạng [3]. Tuy nhiên, do mạng IoT tạo ra một lượng dữ liệu khổng lồ, các kỹ thuật phân tích truyền thống thường không đủ khả năng xử lý dữ liệu này theo thời gian thực.

Hệ thống IoT truyền thống chủ yếu dựa vào các cơ chế phòng thủ tĩnh như tường lửa, phần mềm chống virus hoặc bản vá phần mềm. Tuy nhiên, tính không đồng nhất của các thiết bị IoT đã gây ra hạn chế lớn đối với các nhà sản xuất trong việc triển khai các biện pháp bảo mật phù hợp [4]. Hơn nữa, đặc điểm đa chức năng của các thiết bị IoT cũng đặt ra thách thức trong việc phát triển và triển khai một cơ chế bảo mật có thể phát hiện tấn công trên nhiều loại thiết bị khác nhau. Điều này dẫn đến việc khó có thể xây dựng các biện pháp bảo mật mạnh mẽ và toàn diện [5].

Việc triển khai rộng rãi các ứng dụng IoT khiến vấn đề bảo mật trở thành một yếu tố quan trọng để đảm bảo giao tiếp an toàn [6]. Bên cạnh đó, với sự phát triển của Trí tuệ nhân tạo (AI), nhiều phương pháp học sâu và học máy đã được đề xuất để cải thiện hệ thống phát hiện xâm nhập (IDS). Một số bài báo khảo sát đã đề cập đến các hệ thống IDS truyền thống tiêu biểu [7, 8].

Tuy nhiên, các nghiên cứu hiện tại chưa cung cấp cái nhìn sâu sắc về những tiến bộ mới nhất. Các bài báo khảo sát hiện nay cũng chỉ đánh giá mối đe dọa ở các lớp khác nhau trong hệ thống IoT bằng cách so sánh các công cụ mô phỏng có sẵn và vai trò của chúng trong kiến trúc IoT hiện có. Tuy nhiên, với sự xuất hiện của các giao thức mới được đề xuất và sử dụng ở các lớp khác nhau, vai trò của các công cụ ghi nhận dữ liệu và tấn công chưa được đề cập trong các nghiên cứu đó [9].

A diagram of steps to a computer

AI-generated content may be incorrect.

Hình 1: Mô hình bậc thang của bài báo này, minh họa các bước khác nhau và các mục tiêu đạt được sau khi hoàn thành mỗi bước.

Chúng tôi nhấn mạnh rằng việc truy cập đầy đủ thông tin là yếu tố then chốt để xây dựng một hệ thống phát hiện xâm nhập (IDS) mạnh mẽ cho môi trường IoT. Mỗi giây trôi qua, các thiết bị IoT mới được kết nối vào internet, trong khi các thiết bị và phương pháp hiện có dần trở nên lỗi thời. Do đó, các nhà nghiên cứu bảo mật cần có quyền tiếp cận những nghiên cứu toàn diện về các tiến bộ mới nhất trong lĩnh vực an ninh IoT, giúp phát hiện các lỗ hổng zero-day và bảo vệ trước những mối đe dọa mới.

Hình 1 minh họa mô hình bậc thang về những đóng góp của chúng tôi, trong đó mỗi "bậc" thể hiện các chủ đề được thảo luận trong bài báo, còn các "mục tiêu" xác định những kiến thức mà người đọc sẽ nắm được trước khi chuyển sang bước tiếp theo.

* 1. **Tổng quan các nghiên cứu khảo sát trước đây**

Các nghiên cứu gần đây cho thấy nhiều hệ thống phát hiện xâm nhập (IDS) đã được phát triển và đánh giá bằng cách sử dụng các tập dữ liệu công khai hoặc tập dữ liệu được cung cấp theo yêu cầu. Nhiều nghiên cứu đã được xem xét và công bố, nhấn mạnh vấn đề bảo mật IoT trong nhiều lĩnh vực khác nhau.

Nghiên cứu khảo sát được trình bày trong [10] chủ yếu tập trung vào các loại xâm nhập khác nhau trong IDS và cung cấp chi tiết về từng loại. Các tác giả cũng đưa ra một phân loại về các cuộc tấn công, bao gồm các lớp bị ảnh hưởng nhiều nhất. Các công cụ tấn công được giải thích rất chi tiết. Tuy nhiên, bài báo không đưa ra định hướng về các thuật toán có thể sử dụng để phát hiện các cuộc tấn công trong IDS, cũng như không thảo luận chi tiết về các tập dữ liệu chuẩn để đánh giá hiệu quả của hệ thống.

A table of multiple tasks

AI-generated content may be incorrect.

***Bảng 1: So sánh các nghiên cứu khảo sát tiên tiến nhất.***

Một nghiên cứu khảo sát toàn diện về các tập dữ liệu NIDS (Hệ thống phát hiện xâm nhập mạng) đã được thực hiện trong [11], đề cập đến nhiều đặc điểm, định dạng và mối quan hệ giữa các tập dữ liệu. Bài báo cũng thảo luận về các kịch bản tấn công, công cụ tạo lưu lượng mạng, môi trường ghi nhận dữ liệu và kho lưu trữ dữ liệu. Ngoài ra, nghiên cứu còn đưa ra một số khuyến nghị giúp các nhà nghiên cứu lựa chọn tập dữ liệu phù hợp với từng kịch bản cụ thể. Cụ thể, nghiên cứu đã xem xét 34 tập dữ liệu và xác định 15 đặc điểm khác nhau.

Tuy nhiên, số lượng mẫu của từng loại tấn công trong mỗi tập dữ liệu chưa được phân tích chi tiết. Bên cạnh đó, các giao thức bị tấn công và mô hình được sử dụng để phát hiện các cuộc tấn công này cũng chưa được đề cập trong nghiên cứu.

Các phương pháp học sâu dành cho hệ thống phát hiện xâm nhập có thể được chia thành hai mô hình chính: **mô hình phân biệt sâu (deep discriminative)** và **mô hình tạo/giám sát không có nhãn (generative/unsupervised)**.

Các tác giả trong [13] đã phân loại 35 tập dữ liệu an ninh mạng khác nhau thành bảy nhóm. Hai tập dữ liệu chứa lưu lượng thực tế, **CSE-CICIDS2018** và **Bot-IoT**, đã được sử dụng để đánh giá hiệu suất phân loại nhị phân và đa lớp của các mô hình được đề xuất. Tuy nhiên, nghiên cứu này chưa đề cập đầy đủ về các tập dữ liệu IoT khác hiện có.

Trong [14], các tác giả thảo luận về các cuộc tấn công vào từng lớp trong kiến trúc IoT, bao gồm **các khía cạnh đe dọa**, **bề mặt tấn công**, **công nghệ sử dụng**, **băng tần tần số**, và **tốc độ dữ liệu tương ứng**. Đồng thời, họ cũng nêu bật các tập dữ liệu liên quan đến mạng và IoT. Tuy nhiên, nghiên cứu này không phân tích cụ thể các đặc trưng và phương pháp giúp phát hiện chính xác các cuộc tấn công.

Một bài đánh giá khác tập trung vào các phương pháp phát hiện xâm nhập gần đây thông qua học sâu. Trong nghiên cứu này, các tập dữ liệu tiêu chuẩn dành cho phát hiện xâm nhập cũng đã được phân tích chi tiết để minh họa việc áp dụng mô hình học sâu vào quá trình huấn luyện. Tuy nhiên, nghiên cứu chỉ đề cập một phần đến tập dữ liệu IoT và không cung cấp thông tin về các loại tấn công đã được thực hiện [15].

Một bài khảo sát gần đây dựa trên **mười một nghiên cứu khác nhau** đã điều tra một số cuộc tấn công botnet trong hệ thống IoT. Nghiên cứu này chỉ đề cập sơ lược về các tập dữ liệu IoT hiện có. Tuy nhiên, Hamid et al. [17] đã áp dụng các kỹ thuật đặc trưng để phát hiện và giảm thiểu các cuộc tấn công này. Dù vậy, nghiên cứu này không đề cập đến **tên thiết bị IoT**, **loại hình tấn công**, hoặc **thuật toán được sử dụng để phát hiện tấn công**.

Cuối cùng, các nghiên cứu trong [18] và [19] cung cấp cái nhìn sâu sắc về ứng dụng của **học máy (ML) và học sâu (DL)** trong bảo mật IoT, nhưng không tập trung vào **các lỗ hổng bảo mật và tấn công theo từng lớp trong hệ thống IoT**.

Các bài khảo sát được thảo luận cho đến nay chỉ xem xét một số khía cạnh cụ thể của tập dữ liệu, tức là chúng không cung cấp đầy đủ các yêu cầu chi tiết về môi trường IoT. Một nền tảng vững chắc là điều cần thiết trước khi đề xuất hệ thống phát hiện xâm nhập (IDS) cho IoT. Do đó, bài báo này đưa ra một nghiên cứu so sánh về các tập dữ liệu IoT hiện có, các kỹ thuật được sử dụng để phát hiện tấn công, đồng thời nhấn mạnh nhu cầu về một hệ thống IDS mạnh mẽ và được cập nhật liên tục cho môi trường IoT. **Bảng 1** so sánh các nghiên cứu khảo sát tiên tiến nhất dựa trên các tiêu chí: **tập dữ liệu IoT, đặc trưng, mô hình, và độ chính xác đạt được**.

* 1. **Đóng góp của nghiên cứu**

Bài báo này giúp các nhà nghiên cứu xác định các tập dữ liệu IoT hiện có dành cho phát hiện xâm nhập và các đặc điểm của chúng. Những đóng góp chính của nghiên cứu bao gồm:

* Phân tích một số khảo sát tiên tiến, cung cấp thông tin về mức độ bao phủ của các khảo sát đó đối với tập dữ liệu IoT, xem xét liệu chúng có đề cập đầy đủ hay chỉ một phần;
* Điều tra các cuộc tấn công bảo mật trong môi trường IoT bằng cách xây dựng phân loại cho từng loại tấn công. Các khía cạnh bảo mật như tính bảo mật, toàn vẹn, sẵn có, tác động của các mối đe dọa và quyền riêng tư được xem xét. Chi tiết về các cuộc tấn công nhắm vào các giao thức và các lỗ hổng bị khai thác trong IoT cũng được thảo luận sâu;
* Phân tích và so sánh một số tập dữ liệu IoT công khai phổ biến bằng cách xác định các đặc điểm cơ bản như miền ứng dụng, tính sẵn có, loại lưu lượng, định dạng dữ liệu, dữ liệu có nhãn/không nhãn, thời gian thu thập, loại mạng, quốc gia, số lượng thiết bị, số lượng mẫu để quan sát và đánh giá;
* Thảo luận sâu rộng về các hướng nghiên cứu trong tương lai xuyên suốt bài báo và tổng hợp trong phần cuối. Động lực của nghiên cứu này là thúc đẩy sự phát triển của các sáng kiến mới nhằm mang lại lợi ích cho cộng đồng khoa học và công nghiệp. Theo đó, các khoảng trống trong nghiên cứu hiện tại liên quan đến bảo mật IoT được trình bày một cách toàn diện.

Nhìn chung, bài báo này đóng góp cho các nhà nghiên cứu quan tâm đến việc đề xuất hoặc phát triển một hệ thống phát hiện xâm nhập mạnh mẽ dựa trên IoT và có ảnh hưởng đáng kể đến lĩnh vực nghiên cứu đang phát triển này.

Phần còn lại của bài báo được tổ chức như sau:

 **Mục 2** thảo luận về các cuộc tấn công phổ biến ở các lớp khác nhau, bao gồm phân loại chi tiết các loại tấn công; đồng thời, các cuộc tấn công vào giao thức và các lỗ hổng bị khai thác cũng được làm rõ.

 **Mục 3** trình bày các tập dữ liệu IoT hiện có với mô tả chi tiết về các đặc tính của chúng.

 **Mục 4** thảo luận về các kỹ thuật mà các tập dữ liệu sử dụng để phát hiện các cuộc tấn công.

 **Mục 5** cung cấp danh sách các đặc trưng tiềm năng mà các tập dữ liệu sử dụng, với mô tả chi tiết.

 **Mục 6** tổng kết nghiên cứu và trình bày các hướng nghiên cứu trong tương lai.

1. **Các cuộc tấn công bảo mật trong IoT**

Trong thập kỷ qua, các thiết bị IoT đã phải đối mặt với nhiều loại tấn công khác nhau, khiến người dùng ngày càng thận trọng hơn khi sử dụng các thiết bị này. Mục tiêu của kẻ tấn công có thể là thu thập thông tin, đánh cắp dữ liệu hoặc từ chối dịch vụ đối với người dùng hợp pháp.

Mục tiêu chính của phần này là xác định các cuộc tấn công thực tế và phân tích mục tiêu của chúng bằng cách chia các cuộc tấn công thành hai nhóm chính: **tấn công chủ động** và **tấn công thụ động**. Chúng tôi cũng làm rõ các lớp khác nhau trong mô hình mạng bị ảnh hưởng bởi các cuộc tấn công này.

Ngoài ra, chúng tôi thảo luận về các cuộc tấn công nhắm vào giao thức và các lỗ hổng bị khai thác trong môi trường IoT. **Hình 2** minh họa phân loại các cuộc tấn công IoT được đề xuất trong nghiên cứu này.

A diagram of a computer system

AI-generated content may be incorrect.

**Hình 2: Phân loại các cuộc tấn công trong IoT được thực hiện trên các lớp khác nhau.**

* 1. **Tấn công chủ động**

Trong loại tấn công này, kẻ xâm nhập đột nhập vào hệ thống và thay đổi thông tin liên quan. Chúng cố gắng kiểm soát tài nguyên bằng cách chèn mã độc hại, làm gián đoạn hoạt động bình thường bằng cách truyền một lượng lớn dữ liệu, làm quá tải hệ thống và phát tán thông tin [20].

Ví dụ, một cuộc tấn công Từ chối Dịch vụ (DoS) có thể bắt đầu bằng cách loại bỏ các gói dữ liệu và tiêu tốn tài nguyên hệ thống [21].

Các cuộc tấn công chủ động được phân loại theo từng lớp mạng trong phần sau.

* + 1. **Các cuộc tấn công vào lớp vật lý**

Lớp này còn được gọi là **lớp cảm biến**, nơi các thiết bị sử dụng nhiều kênh khác nhau để giao tiếp.

Các cuộc tấn công ở lớp vật lý bao gồm từ truy cập trái phép vào thiết bị cho đến việc gỡ bỏ hoặc chặn kết nối của người dùng hợp pháp. Kẻ tấn công có thể gây hư hỏng vật lý cho thiết bị bằng cách ngăn hệ thống tắt máy, dẫn đến tình trạng quá nhiệt [22].

**2.1.2. Các cuộc tấn công vào lớp MAC, mạng và truyền tải**

Vì địa chỉ MAC được mã hóa trong các thẻ giao diện mạng, kẻ tấn công có thể thực hiện **tấn công giả mạo (spoofing)** bằng cách thay đổi địa chỉ MAC của người dùng hợp pháp. Ngoài ra, các mối đe dọa như **tấn công Người trung gian (Man-in-the-Middle - MITM)** có thể ngăn chặn hoạt động của các nút mạng cụ thể [23].

* + 1. **Các cuộc tấn công vào lớp ứng dụng**

Lớp này hỗ trợ nhiều giao thức với các chức năng khác nhau; chẳng hạn, SMTP được sử dụng để truyền thư điện tử, trong khi HTTP cho phép cung cấp dịch vụ web.

Việc sử dụng thiết bị IoT trong nhiều lĩnh vực như **nhà thông minh** và **văn phòng** với các giao thức lớp ứng dụng đã tạo ra một môi trường dễ dàng bị khai thác, trở thành mục tiêu của các cuộc tấn công mạng.

* 1. **Các cuộc tấn công thụ động**

Loại tấn công này thường thu thập thông tin mà không có sự cho phép của người dùng, sau đó khai thác dữ liệu bằng cách giải mã nó.

Các cuộc tấn công thụ động được chia thành các loại sau:

* + 1. **Phần mềm**

Thông thường, phần mềm độc hại được sử dụng để lây nhiễm vào hệ thống nhằm đánh cắp dữ liệu hoặc thậm chí thực hiện các cuộc tấn công từ chối dịch vụ (DoS). Các phần mềm này có thể tồn tại dưới dạng virus, sâu máy tính (worms) và phần mềm gián điệp (spyware).

Hệ thống **SCADA** là một trong những mục tiêu dễ bị tấn công nhất do sự kết nối của chúng với mạng IoT. Với sự phát triển nhanh chóng của công nghệ IoT, các hệ thống SCADA truyền thống ngày càng được tích hợp với **mạng doanh nghiệp và hệ thống CNTT**, khiến chúng dễ bị tổn thương trước các cuộc tấn công mạng.

* + 1. **Phân tích mật mã**

Nhiều thuật toán mã hóa nhẹ đã được phát triển để đáp ứng nhu cầu ngày càng tăng của các thiết bị IoT, vốn có tài nguyên hạn chế và hoạt động mà không cần sự can thiệp của con người.

Một lượng lớn thông tin thô được ghi nhận hàng ngày từ các thiết bị IoT. Do các thuật toán được sử dụng phổ biến thường yêu cầu rất ít tài nguyên tính toán, **bảo mật của các thiết bị IoT trở nên dễ bị tổn thương trước các kỹ thuật phân tích mật mã** như **phân tích sai biệt (differential cryptanalysis) hoặc tấn công trượt (slide attacks).**

Việc bảo vệ các thiết bị IoT chống lại các cuộc tấn công này là **cần thiết**, đặc biệt khi chúng được kết nối tại nhiều lớp khác nhau trong hệ thống. Do đó, cần có **các biện pháp hoặc chiến lược thích hợp** để đối phó với các mối đe dọa này khi triển khai mô hình IoT.

* + 1. **Tấn công dựa trên lưu lượng và cạn kiệt băng thông/tài nguyên**

Tại đây, kẻ tấn công sử dụng hết băng thông bằng cách gửi bot qua mạng, ngăn cản người dùng hợp pháp truy cập vào các dịch vụ được phân bổ. Ngoài ra, kẻ tấn công bắt đầu làm cạn kiệt tài nguyên bằng cách chiếm dụng hoặc làm hao hụt chúng. Các lớp khác nhau của kiến trúc mạng đều có thể bị nhắm mục tiêu. Đây là một cuộc tấn công nghiêm trọng, có thể khiến hệ thống bị sập do tình trạng tràn bộ nhớ và quá tải CPU.

* + 1. **Tấn công dựa trên web**

Kẻ tấn công gửi một lượng lớn yêu cầu đến một trang web cụ thể, khiến trang web ngừng hoạt động trong nhiều giờ. Ngoài ra, kẻ xâm nhập có thể thực hiện nghe lén để thu thập dữ liệu cá nhân của người dung

* 1. **Tấn công vào các loại giao thức khác nhau**

Các giao thức được thiết kế để cung cấp giao tiếp hiệu quả từ nguồn đến đích. Hơn nữa, các giao thức này thường hoạt động trong một môi trường đơn giản nhưng mang tính hợp tác, không quan tâm đến các cơ chế bảo mật tính toàn vẹn và quyền riêng tư. Do đó, các giao thức hoạt động ở các lớp khác nhau dễ bị tấn công hơn [25]. Trong Hình 3, một sơ đồ xương cá, thường được sử dụng để thể hiện tính bảo mật và an toàn, đã được sử dụng để phân loại từng cuộc tấn công vào các giao thức. Các sơ đồ này cung cấp một phân tích sâu sắc về các yếu tố góp phần có thể gây ra (hoặc nguyên nhân phụ) của một vấn đề cụ thể [24]. Trên thực tế, các giao thức và mối đe dọa khác nhau đều được xem xét.

A diagram of a computer security system

AI-generated content may be incorrect.

**Hình 3: Mô hình xương cá để phân loại các cuộc tấn công liên quan đến giao thức [24].**

* + 1. **Tấn công vào Giao thức Internet (IP)**

Giao thức IP hoạt động ở lớp mạng với chức năng định tuyến các gói tin từ nguồn đến đích theo quy tắc của bộ định tuyến. Mỗi bước nhảy tiếp theo được xác định bởi bộ định tuyến để đảm bảo các gói tin đến đúng đích. Có nhiều loại tấn công có thể xảy ra đối với giao thức IP.

* + 1. **Tấn công vào TCP**

Chức năng của giao thức này là đảm bảo việc truyền các gói tin qua mạng một cách đáng tin cậy. Giao thức TCP thường tuân theo cơ chế truyền gói tin [26]. Tuy nhiên, kẻ tấn công có thể xâm nhập vào hệ thống bằng cách khởi tạo nhiều kết nối nửa vời [27], thực hiện tấn công TCP-SYN flood, vượt qua các tường lửa không chặn các gói tin FIN [28], và chiếm quyền điều khiển phiên làm việc (session hijacking) [29].

* + 1. **Tấn công vào UDP**

Giao thức Datagram Người dùng (UDP) là một giao diện cho IP, có chức năng truyền dữ liệu từ một ứng dụng này sang ứng dụng khác với chi phí thấp. Tuy nhiên, do đặc điểm của nó, UDP luôn dễ bị tấn công [30], bao gồm các hình thức như tấn công vòng lặp cổng (port loopback) và tấn công Fraggle [31].

* + 1. **Tấn công vào ICMP**

Giao thức này gửi các thông điệp một chiều được các bộ định tuyến trong mạng sử dụng, chủ yếu để cung cấp thông tin về lỗi liên quan đến quá trình xử lý dữ liệu của máy đích. Trong quá trình này, nhiều hình thức tấn công (ví dụ: quét địa chỉ, tấn công Smurf và tấn công nhận diện hệ điều hành) lợi dụng việc gửi các gói ICMP bị lỗi đến máy đích [32].

* 1. **Type of Exploited Vulnerability**

Lỗ hổng bảo mật là những điểm yếu mà kẻ tấn công có thể dễ dàng lợi dụng để xâm nhập vào hệ thống. Sự gia tăng đa dạng của các thiết bị IoT đã dẫn đến nhiều lỗ hổng và cuộc tấn công mạng hơn. Các lỗ hổng phổ biến liên quan đến thiết bị IoT bao gồm mật khẩu yếu [33], cổng mở [34], giao diện ứng dụng không an toàn [35], giao diện mạng bị lộ [36], phần mềm bị lộ [34], bộ nhớ thiết bị, giao diện đám mây không an toàn [36], tấn công dò mật khẩu, bảo mật cơ sở hạ tầng và các mối đe dọa từ ảo hóa [37] [38]. Bảng 3 trình bày các loại lỗ hổng và hình thức tấn công tương ứng

Hơn nữa, mạng dựa trên IoT khác với mạng truyền thống theo nhiều cách, như thể hiện trong Bảng 2. Thực tế, vẫn còn một vấn đề nghiên cứu mở về cách mô hình giao tiếp nội bộ có thể khác nhau và cách Machine Learning (ML) có thể được sử dụng để phân tích các mô hình này. Ví dụ, IoT profiling [39] [6] là một vấn đề nghiên cứu hiện nay tập trung vào giao tiếp và trao đổi dữ liệu giữa các thiết bị IoT, đồng thời nghiên cứu cách Machine Learning (ví dụ: Deep Neural Networks và Random Forest) [40] [41] cũng như các phương pháp phát hiện bất thường (ví dụ: Isolation Forest) [42] có thể được áp dụng chính xác hơn dựa trên các nỗ lực giải thích [43]. Cuối cùng, phân tích chi tiết về sự giao thoa giữa các đặc điểm của mạng IoT, IoT profiling và khía cạnh hành vi trong các cuộc tấn công IoT, cũng như cách các kỹ thuật ML hỗ trợ việc phát hiện các mối đe dọa đơn lẻ và đa lớp, sẽ được đề cập trong các hướng nghiên cứu tương lai.

Các phương pháp Machine Learning khác nhau có thể đóng vai trò quan trọng trong bảo mật IoT bằng cách cung cấp khả năng phát hiện mối đe dọa nâng cao, phát hiện bất thường và phân tích dự đoán. Dưới đây là một số kỹ thuật ML quan trọng để phát hiện các mối đe dọa khác nhau trong môi trường IoT:

**Phát hiện bất thường:** Phát hiện bất thường là một kỹ thuật nhằm xác định hành vi bất thường hoặc không mong đợi trong mạng IoT. Điều này bao gồm việc giám sát cảm biến và thiết bị để phát hiện các hoạt động không bình thường có thể cho thấy dấu hiệu xâm nhập hoặc mối đe dọa bảo mật. Các thuật toán học máy có thể được huấn luyện để nhận diện những bất thường này và cảnh báo nhân viên bảo mật để điều tra thêm. Sử dụng các thuật toán phát hiện bất thường như k-NN, Local Outlier Factor (LOF) và Connectivity-based Outlier Factor (COF) dựa trên các hành vi bình thường của thiết bị, chẳng hạn như khoảng thời gian kết nối với cảm biến, thời gian gửi yêu cầu ARP hoặc các hành vi giao thức mạng khác của thiết bị IoT, cho thấy vai trò quan trọng của học máy trong môi trường IoT, đặc biệt là trong việc ngăn chặn các cuộc tấn công. Hơn nữa, việc cài đặt một thuật toán phát hiện bất thường nhẹ trên mạng IoT có thể giúp tổ chức ngăn chặn các cuộc tấn công kịp thời.

**Phân tích dự đoán:** Phân tích dự đoán sử dụng dữ liệu lịch sử để xác định các mô hình và xu hướng nhằm dự đoán hành vi trong tương lai. Trong bảo mật IoT, phân tích dự đoán có thể được sử dụng để dự đoán các mối đe dọa tiềm ẩn và thực hiện các biện pháp phòng ngừa trước khi chúng xảy ra.

**Học sâu:** Trong bảo mật IoT, học sâu có thể được sử dụng để phân tích lượng dữ liệu lớn được tạo ra bởi các cảm biến và thiết bị theo thời gian thực, giúp phát hiện các mối đe dọa và bất thường nhanh chóng và chính xác hơn.

**Học tăng cường:** Trong bảo mật IoT, học tăng cường có thể giúp các thiết bị học cách xác định và phản ứng với các mối đe dọa bảo mật theo thời gian thực.

Học máy đóng vai trò quan trọng trong bảo mật IoT bằng cách cung cấp khả năng phát hiện và ngăn chặn mối đe dọa nâng cao. Bằng cách sử dụng các phương pháp trên, các tổ chức có thể phân tích lượng dữ liệu khổng lồ được tạo ra bởi các thiết bị IoT và xác định các rủi ro tiềm ẩn một cách nhanh chóng và chính xác hơn.

1. **Bộ Dữ Liệu Tấn Công IoT**

Trong vài năm qua, nhiều bộ dữ liệu trong lĩnh vực bảo mật IoT đã được tập hợp, mỗi bộ có những ưu và nhược điểm riêng. Do sự gia tăng nhanh chóng trong việc áp dụng các thiết bị IoT, các nhà nghiên cứu ngày càng tập trung vào các bộ dữ liệu liên quan đến IoT.

A close-up of a list of networks

AI-generated content may be incorrect.

**Bảng 2: So sánh giữa mạng truyền thống và mạng dựa trên IoT**

A close-up of a list

AI-generated content may be incorrect.

**Bảng 3: Các loại tấn công phổ biến cùng với lỗ hổng bảo mật**

* 1. **Tiêu chí tìm kiếm**

Để tìm các tập dữ liệu IoT có sẵn nhằm hỗ trợ các giải pháp bảo mật, ba bước chính đã được xem xét, như được minh họa trong Hình 5. Với mục tiêu xác định các tập dữ liệu gần đây được công bố trong 5 năm qua, chúng tôi đã tìm kiếm trên Google Scholar bằng các cụm từ tìm kiếm sau: “IoT security + Dataset”, “IoT security + Data”, “IoT dataset + Machine Learning”, “IoT cybersecurity + Dataset”, “IoT IDS + Dataset” và “IoT DDoS + Dataset”.

Để mở rộng phạm vi tìm kiếm các tập dữ liệu có sẵn, chúng tôi cũng tập trung vào các bài báo kỹ thuật có sử dụng tập dữ liệu IoT. Theo đó, chúng tôi đã tìm kiếm các đóng góp kỹ thuật trong các lĩnh vực sau: IoT IDS, IoT DDoS Detection, IoT Intrusion Prevention, IoT Spoofing, IoT Reconnaissance, IoT Flooding, IoT Mirai và IoT Privacy

Cuối cùng, chúng tôi cũng tìm kiếm các tập dữ liệu mới từ các nhóm nghiên cứu đã từng công bố các tập dữ liệu an ninh mạng được sử dụng rộng rãi trước đây. Trọng tâm chính của chúng tôi là tìm kiếm các tập dữ liệu trong các lĩnh vực sau: phần mềm độc hại (ví dụ: Android và Adware), DNS, Dark Web và IDS (ví dụ: DDoS, IPS và DoS).

A colorful pie chart with numbers with Crust in the background

AI-generated content may be incorrect.

**Hình 4: Số lượng tập dữ liệu IoT được đề xuất theo từng năm.**

* 1. **Đặc trưng**

**Miền dữ liệu:** Trước khi thu thập dữ liệu, cần xác định miền mục tiêu. Vì các thiết bị IoT phổ biến nhất trong môi trường gia đình và cung cấp một nền tảng thử nghiệm dễ dàng cho các cuộc tấn công mạng, nên hầu hết dữ liệu được ghi lại tập trung vào miền này.

**Phát hiện xâm nhập** chỉ có thể được kiểm tra nếu dữ liệu được công khai cho nghiên cứu. Thuộc tính này được đánh dấu nếu dữ liệu được công khai hoặc có sẵn theo yêu cầu.

**Loại lưu lượng:** Đây là một thuộc tính quan trọng vì nó giúp đánh giá hệ thống trên nhiều thiết bị IoT và các cuộc tấn công. "Mô phỏng"

A diagram of a scientific paper

AI-generated content may be incorrect.

**Hình 5: Tiêu chí tìm kiếm được áp dụng trong nghiên cứu này.**

Định dạng dữ liệu: Việc thu thập dữ liệu đóng vai trò quan trọng vì nó hỗ trợ phân tích lưu lượng mạng. Dữ liệu có thể được ghi lại dưới định dạng PCAP, bao gồm thông tin tải trọng và định dạng đặc trưng với dữ liệu đã được xử lý. Một loại khác là tệp nhật ký, thu thập các bản ghi do hệ thống tạo ra như sự kiện của hệ điều hành hoặc máy chủ.

Đánh giá: Trường nhãn xác định liệu mẫu dữ liệu thu thập được là bình thường hay độc hại.

Tổng thời gian: Được định nghĩa là khoảng thời gian mà dữ liệu được thu thập.

Loại mạng: Thuộc tính này được sử dụng để đánh giá khả năng thích ứng với một môi trường cụ thể.

Quốc gia: Trường này xác định tên các quốc gia nơi dữ liệu được thu thập.

Số lượng thiết bị: Để đánh giá tính khả thi của hệ thống phát hiện xâm nhập, nhiều thiết bị IoT khác nhau được sử dụng trong môi trường thử nghiệm. Trường này hiển thị số lượng thiết bị IoT có trong mỗi tập dữ liệu.

Số lượng mẫu (Bình thường/Độc hại): Trường này mô tả số lượng mẫu được thu thập cho từng loại, giúp lựa chọn thuật toán phù hợp với dữ liệu có sẵn. Tính tái lập: Mô tả mức độ có thể tái tạo nghiên cứu dựa trên nhiều khía cạnh. Về tính tái lập, các tập dữ liệu có thể được phân loại thành Tái lập được (R), Tái lập một phần (PR) và Khó tái lập (HR). Phân loại này dựa trên mức độ chi tiết của quá trình thu thập dữ liệu trong từng nghiên cứu cũng như mức độ rõ ràng để các nhà nghiên cứu khác có thể tạo ra một tập dữ liệu tương tự theo phương pháp đã được mô tả, và được đánh giá như sau:



Trong đó, Rs là điểm số tái lập với phạm vi từ 0 ≤ Rs ≤ 1. A đại diện cho tính sẵn có của tập dữ liệu (0.25), Tr là loại lưu lượng được sử dụng (0.25 cho dữ liệu mô phỏng, 0.12 cho dữ liệu thực), còn Pc và Nt lần lượt biểu thị tính sẵn có của tệp .pcap (0.25) và loại mạng (0.25 cho mạng nhỏ, 0.12 cho mạng trung bình).

Trong bối cảnh này, bốn thành phần trên được gán trọng số bằng nhau vì chúng bổ sung lẫn nhau và đóng vai trò quan trọng trong việc xác thực các giải pháp bảo mật IoT. Ví dụ, để được sử dụng, một tập dữ liệu phải có sẵn để tải xuống. Nếu lưu lượng là mô phỏng, nó sẽ dễ dàng được tái tạo trong bất kỳ hệ thống nào. Bên cạnh đó, việc có sẵn tệp .pcap giúp các nhà nghiên cứu truy cập vào lưu lượng thực tế và tự thiết kế đặc trưng dữ liệu. Loại mạng cũng ảnh hưởng trực tiếp đến khả năng tái lập của quá trình tạo dữ liệu.

Cuối cùng, các trọng số này có thể được điều chỉnh theo cách khác để đánh giá tính tái lập từ nhiều góc độ khác nhau [59]. Trên thực tế, mỗi tập dữ liệu được đánh giá là Tái lập được (R), Tái lập một phần (PR) hoặc Khó tái lập (HR) như sau:

A black text with black text

AI-generated content may be incorrect.

* 1. **Datasets**

**N-BaIoT (2018) [60]:** Một tập dữ liệu mạng mới được đề xuất để phát hiện các cuộc tấn công botnet trong môi trường IoT. Các tác giả đã sử dụng hai botnet nổi tiếng là Mirai và BASHLITE để tấn công chín thiết bị IoT thương mại. Các đặc trưng thống kê khác nhau được trích xuất trong nhiều cửa sổ thời gian và sau đó được sử dụng bởi một bộ mã hóa tự động học sâu để phát hiện tấn công.

**oTHIDS (2018) [47]:** Một tập dữ liệu IoT dựa trên máy chủ, thu thập dữ liệu từ các thiết bị IoT thực tế. Thí nghiệm được thực hiện trong phòng thí nghiệm, bao gồm ba thiết bị bị nhiễm bởi các botnet phần mềm độc hại như Mirai, Hajime, Adira, BASHLITE, Doflo, Tsunami và Wroba.

**IoT-SH (2019) [53]:** Tập dữ liệu này được đề xuất bằng cách sử dụng 8 thiết bị nhà thông minh khác nhau, trên đó 12 cuộc tấn công khác nhau được thực hiện và được phân loại thành 4 nhóm tấn công chính. Các tác giả đã áp dụng hệ thống phát hiện xâm nhập (IDS) ba lớp để phát hiện các cuộc tấn công và sử dụng nhiều kết hợp giữa phương pháp dựa trên quy tắc và học máy để phân loại các cuộc tấn công.

**BoT-IoT (2019) [52]:** Tập dữ liệu này mô phỏng lưu lượng mạng thực tế, được tạo ra với các cấu hình mạng không đồng nhất. Các cuộc tấn công DDoS, DoS, đánh cắp dữ liệu và quét mạng được thực hiện trên năm thiết bị. Một tập hợp các đặc trưng mới được trích xuất bằng cách sử dụng hệ số tương quan và kỹ thuật entropy chung. Sau đó, các đặc trưng này được đưa vào các mô hình học máy và học sâu khác nhau để phát hiện tấn công.

**Kitsune (2019) [51]:** Một tập dữ liệu IoT bao gồm bốn loại tấn công khác nhau được thực hiện trên chín thiết bị IoT. Trong đó, một thiết bị IoT (camera an ninh) bị nhiễm mẫu botnet Mirai thực tế. Các tác giả đã giới thiệu Kitsune như một hệ thống phát hiện xâm nhập mạng (NIDS) dạng plug-and-play để phân biệt mẫu dữ liệu bình thường và độc hại.

**IoTNIDS (2019) [61]:** Tập dữ liệu được thu thập trong môi trường thực tế từ hai thiết bị IoT (loa và camera). Các cuộc tấn công Mirai, MITM, DoS và quét mạng đã được thực hiện trên các thiết bị này.

**MedBIoT (2020) [49]:** Một tập dữ liệu mạng IoT có quy mô trung bình với hành vi được gán nhãn. Tập dữ liệu bao gồm bảy thiết bị IoT, một số là thiết bị thật, số khác là mô phỏng. Các tác giả đã sử dụng 100 đặc trưng thống kê đưa vào các thuật toán học máy khác nhau để phát hiện tấn công.

**IoT-23 (2020) [62]:** Một tập dữ liệu botnet được thu thập trong vòng một năm, bao gồm cả dữ liệu bình thường và dữ liệu tấn công trong môi trường thực tế. Tập dữ liệu này là một phần của dự án Apsemat, tập trung vào nghiên cứu các cuộc tấn công IoT. Trong suốt thời gian nghiên cứu, các tác giả đã thực hiện 23 cuộc tấn công botnet khác nhau trên các thiết bị.

**IoTIDs (2020) [23]:** Một tập dữ liệu được đề xuất với các đặc trưng lưu lượng mạng liên quan đến IoT, được chọn lọc bằng hệ số tương quan và xếp hạng bằng thuật toán Shapira-Wilk. Trong nghiên cứu, các tác giả đã thực hiện bốn cuộc tấn công khác nhau trên hai thiết bị (loa và camera) và ghi lại dữ liệu. Các thuật toán SVM, G-NB, LDA và LR được sử dụng để phát hiện và phân loại tấn công.

**MQTT (2020) [50]:** Mục tiêu của các tác giả là cung cấp một tập dữ liệu sử dụng giao thức MQTT chuyên dụng cho các kịch bản mạng IoT. Tám thiết bị IoT được kết nối với máy chủ MQTT để mô phỏng tình huống thực tế. Một tập hợp 33 đặc trưng khác nhau được trích xuất và đưa vào nhiều thuật toán học máy để đánh giá độ chính xác.

**MQTT-IoT-IDS (2020) [63]:** Tập dữ liệu này sử dụng giao thức nhẹ **MQTT**, thường được sử dụng trong mạng IoT. Các tác giả đã xây dựng mạng gồm **camera feed**, **12 cảm biến MQTT**, và **một MQTT broker**. Tập dữ liệu bao gồm **5 kịch bản**, bao gồm cả dữ liệu bình thường và dữ liệu tấn công.

Các đặc trưng được trích xuất bao gồm:

* **Packet-based features** (đặc trưng theo gói tin)
* **Uni-flow features** (đặc trưng theo luồng đơn hướng)
* **Bi-flow features** (đặc trưng theo luồng hai hướng)

Sau đó, tập dữ liệu được đưa vào **6 thuật toán học máy** khác nhau để kiểm tra hiệu quả của hệ thống phát hiện xâm nhập (IDS).

**TON-IoT (2020) [48]:** Một bộ dữ liệu IoT/IIoT dựa trên đo lường từ xa, có tính chất không đồng nhất. Dữ liệu thu thập bao gồm cả mẫu bình thường và mẫu tấn công với thông tin gốc xác thực. Bộ dữ liệu này còn phản ánh lưu lượng mạng thực tế và bao gồm các danh mục con của các cuộc tấn công. Ngoài bảy thiết bị IoT, bộ dữ liệu còn chứa dữ liệu được ghi lại từ nhật ký hệ điều hành và lưu lượng mạng. Nhiều thuật toán học máy và học sâu đã được áp dụng để đánh giá độ chính xác đạt được.

**Edge-IIoTSet (2022) [64]:** Một bộ dữ liệu an ninh mạng thực tế dành cho các ứng dụng IoT và IIoT nhằm hỗ trợ phát triển hệ thống phát hiện xâm nhập trong cả môi trường tập trung và phân tán. Các tác giả trình bày mô tả chi tiết về môi trường thử nghiệm được sử dụng, cùng với khung tạo bộ dữ liệu và các cân nhắc về học tập tập trung và học liên kết. Bộ dữ liệu này cũng tập trung vào việc giải quyết các hạn chế của các bộ dữ liệu hiện có.

A close-up of a document

AI-generated content may be incorrect.

**Bảng 4: Tổng quan về các thuộc tính của bộ dữ liệu IoT**

1. **Phương pháp phát hiện các cuộc tấn công IoT**

Luôn có nhu cầu tạo ra một môi trường mạnh mẽ để bảo vệ các thiết bị IoT được kết nối qua mạng. Nhiều phương pháp phát hiện như dựa trên cây quyết định, học có giám sát cổ điển, dựa trên quy tắc và học sâu đã được đề xuất qua các năm để phát hiện các cuộc tấn công này. Các thuật toán này mang lại những giải pháp đầy hứa hẹn cho các vấn đề phát sinh [66].

Hơn nữa, nhiều kỹ thuật học máy (ML) và học sâu (DL) đã được áp dụng cho các tập dữ liệu IoT. Hình 6 minh họa các kỹ thuật này, được phân loại thành mô hình dựa trên cây quyết định, học có giám sát cổ điển, học sâu và dựa trên quy tắc. Ngoài ra, Bảng 5 cung cấp thông tin về số lượng đặc trưng được sử dụng, loại đặc trưng được trích xuất, các kỹ thuật chọn đặc trưng và các thuật toán được sử dụng để phát hiện các cuộc tấn công.

A diagram of a machine learning process

AI-generated content may be incorrect.

**Hình 6: Các phương pháp học máy và học sâu khác nhau được sử dụng với các tập dữ liệu IoT.**

* 1. **Mô hình Học máy (ML) dựa trên cây**

Các mô hình ML dựa trên cây thường bao gồm các nút và nhánh, được xây dựng một cách đệ quy bằng cách chia tập dữ liệu huấn luyện. Quá trình này sử dụng một đặc trưng cụ thể để phân tách dữ liệu hiệu quả, dựa trên tiêu chí thông tin như Gini Index hoặc Entropy. Khi cần dự đoán nhãn của một mẫu thử nghiệm, thuật toán sẽ so sánh giá trị đặc trưng từ nút gốc, sau đó di chuyển qua các nhánh cho đến khi đạt đến nút cuối cùng.

 **Decision Tree (DT) [67]:** Một trong những thuật toán phổ biến nhất, được sử dụng rộng rãi để phát hiện, phân loại và giảm thiểu các cuộc tấn công IoT. DT có thể được kết hợp với các phương pháp dựa trên luật hoặc các mô hình lai để nâng cao độ chính xác.

 **Random Forest (RF) [69]:** Một tập hợp của nhiều cây quyết định (DT), trong đó mỗi cây được xây dựng bằng cách chọn ngẫu nhiên các điểm dữ liệu từ tập huấn luyện. RF đã được áp dụng thành công trong nhiều hệ thống để phát hiện mối đe dọa IoT.

* Trong một số nghiên cứu, RF được sử dụng kết hợp với các phương pháp lai hoặc áp dụng trực tiếp để phân loại và phát hiện các cuộc tấn công mạng IoT [70, 71].

 **Isolation Forest [72]:** Một thuật toán phát hiện ngoại lệ không giám sát, dựa trên việc cô lập các điểm bất thường. Isolation Forest đã chứng tỏ hiệu quả trong các bài toán như phát hiện xâm nhập và phát hiện bất thường [73, 74]. Do cách thức hoạt động của nó, thuật toán này tạo ra các đường dẫn ngắn hơn cho các điểm bất thường so với dữ liệu bình thường.

 **Classification and Regression Trees (CART) [75]:** Một thuật toán dựa trên cây quyết định nhị phân, không tham số, được sử dụng để xây dựng mô hình từ tập huấn luyện. CART đã được ứng dụng trong nhiều hệ thống bảo mật IoT, bao gồm IDS trong lưới điện thông minh [68] và phát hiện bất thường trong các ứng dụng dựa trên SDN [76].

* 1. **Các mô hình học máy có giám sát cổ điển**

Đây là nhánh phổ biến nhất trong học máy và được sử dụng rộng rãi để phát hiện xâm nhập trong môi trường IoT. Dưới đây là một số phương pháp được sử dụng phổ biến nhất trong các ứng dụng IoT. Trong đó, SVM, k-NN và NB thường được chọn để báo cáo kết quả.

 **Support Vector Machine (SVM) [77]:** SVM đã được sử dụng trong các ứng dụng bảo vệ quyền riêng tư IoT [78] và nhà thông minh an toàn [79]. Phương pháp này dựa trên một siêu phẳng trong không gian đặc trưng N chiều để tối ưu hóa việc phân tách các lớp. Tuy nhiên, trong một số trường hợp, SVM có thể tiêu tốn nhiều tài nguyên tính toán.

 **Naïve Bayes (NB) [80]:** NB đã được áp dụng thành công trong nhiều giải pháp bảo mật IoT, vì nó không yêu cầu tập dữ liệu lớn để đạt được kết quả tốt. Một số ứng dụng IoT sử dụng NB bao gồm phát hiện bất thường trên nền tảng IoT-Fog-Cloud [81] và phát hiện tấn công trong mạng cảm biến không dây [82]. Thuật toán này hoạt động dựa trên giả định rằng các đặc trưng dữ liệu là độc lập với nhau.

 **Linear Discriminant Analysis (LDA) [83]:** LDA được sử dụng trong nhiều trường hợp, chẳng hạn như cải thiện hiệu suất phân loại trong hệ thống phát hiện xâm nhập [84] và nhận dạng thiết bị đầu cuối IoT [85]. Phương pháp này tìm một không gian đặc trưng mới để tối đa hóa khả năng phân tách lớp trong khi giảm thiểu độ biến thiên trong cùng một lớp. Ngoài phân loại, LDA còn được sử dụng phổ biến như một kỹ thuật giảm chiều dữ liệu.

 **Logistic Regression (LR) [71]:** LR là một phương pháp phân loại nhị phân phổ biến, trong đó một hàm sigmoid/logistic phi tuyến tính được sử dụng để xác định kết quả. Nó đã được áp dụng trong các bài toán như phát hiện hành vi độc hại dựa trên độ tin cậy [41] và phát hiện xâm nhập [86]. LR ánh xạ đầu vào thành một giá trị trong khoảng từ 0 đến 1, và một ngưỡng có thể được đặt để xác định mẫu đầu vào thuộc về một trong hai lớp nhị phân.

 **Multilayer Perceptron (MLP) [87]:** MLP là một biến thể nông của mạng nơ-ron truyền thẳng (feedforward neural networks), mang lại hiệu suất cao trong nhiều vấn đề. Nó đã được sử dụng trong các giải pháp liên quan đến độ tin cậy [88] và phát hiện tấn công [89].

 **k-Nearest Neighbors (k-NN) [90]:** k-NN là một thuật toán học máy có giám sát, không tham số, thường được sử dụng cho các bài toán phân loại nhiều lớp. Nó đã cho thấy hiệu suất đáng chú ý trong các hệ thống phát hiện xâm nhập (IDS) trong mạng cảm biến không dây (WSN) [91][92]. Để hoạt động hiệu quả trên tập dữ liệu lớn và chống nhiễu tốt, giá trị đầu ra của k-NN được dự đoán bằng cách tính khoảng cách giữa k láng giềng gần nhất với điểm dữ liệu cần phân loại.

* 1. **Các phương pháp dựa trên học sâu**

Gần đây, với khả năng tính toán ngày càng cao và sự xuất hiện của các tập dữ liệu lớn để huấn luyện mô hình, các thuật toán học sâu được sử dụng rộng rãi trong nhiều ứng dụng nhờ độ chính xác vượt trội. Phần này thảo luận về các biến thể của thuật toán học sâu được sử dụng phổ biến trong các nghiên cứu.

**Mạng nơ-ron tích chập (CNNs) [93]** là một phần mở rộng của Perceptron đa tầng, bao gồm một lớp đầu vào, nhiều lớp ẩn và một lớp đầu ra. Các lớp ẩn bao gồm lớp tích chập, lớp phi tuyến, lớp giảm mẫu (pooling) và lớp kết nối đầy đủ. Mô hình này đã được áp dụng thành công trong các bài toán như bảo mật gần thời gian thực [40] và phân tích phần mềm độc hại [94].

**Mạng nơ-ron hồi quy (RNNs) [95]** là một kiến trúc mạng nơ-ron phổ biến để xử lý dữ liệu tuần tự, nhờ khả năng ghi nhớ thông tin đầu vào thông qua bộ nhớ nội bộ. Trong bảo mật IoT, RNNs đã được sử dụng để xây dựng hệ thống phát hiện xâm nhập IoT (IDS) [96] và phát hiện mối đe dọa từ phần mềm độc hại trên mạng [97].

**Bộ nhớ ngắn hạn dài (LSTM) [98, 99]** là một biến thể nâng cao của RNNs, giúp học được các mối quan hệ dài hạn trong dữ liệu. Điều này đạt được nhờ các ô bộ nhớ có ba cổng chính: cổng đầu vào, cổng đầu ra và cổng quên, giúp kiểm soát luồng thông tin để quyết định dữ liệu nào cần lưu giữ hoặc loại bỏ nhằm dự đoán đầu ra của mạng.

Cuối cùng, **Bộ mã hóa tự động (Auto Encoder) [100]** là một loại mạng nơ-ron đặc biệt hoạt động theo hướng truyền thẳng. Đầu ra của mô hình này giống với đầu vào, trong đó dữ liệu đầu vào được nén thành dạng có số chiều thấp hơn trước khi tái tạo lại ở đầu ra. Kỹ thuật này đã được sử dụng để phát hiện các cuộc tấn công giả mạo [101] và phát hiện giao tiếp bất thường [102] trong IoT.

A white sheet with black and white text

AI-generated content may be incorrect.

**Bảng 5: So sánh hiệu suất của các tập dữ liệu IoT theo loại đặc trưng, kỹ thuật lựa chọn và phương pháp**

* 1. **Các phương pháp học máy dựa trên luật**

Trong loại học máy này, các quy tắc cụ thể được xác định để phát hiện hoạt động độc hại. Khi dữ liệu được xử lý, hệ thống sẽ kiểm tra hành vi tấn công đã được xác định trong sổ quy tắc và báo cáo những gì nó phát hiện được.

**OneR** [103] là một cây quyết định có độ sâu 1, đơn giản và thường được sử dụng trong hệ thống phát hiện xâm nhập. Thuật toán này tạo ra một tập hợp quy tắc cụ thể bắt đầu từ một thuộc tính duy nhất. Theo cách này, mỗi nhánh được gán lớp phổ biến nhất.

Tương tự, **ZeroR** [104] là một bộ phân loại dựa trên quy tắc đơn giản, chỉ dựa vào lớp mục tiêu và bỏ qua tất cả các giá trị dự đoán. Nó duy trì một bảng tần suất và không bao giờ chọn giá trị không phải mục tiêu. Do đó, bộ phân loại này hoạt động tốt nhất khi cần xác định hiệu suất cơ bản để so sánh với các bộ phân loại khác.

1. **Các đặc trưng tiềm năng được sử dụng trong tập dữ liệu IoT để phát hiện tấn công**

Qua nhiều năm, đã có nhiều khung phân tích được đề xuất để phát hiện xâm nhập mạng. Tuy nhiên, nếu không có một tập hợp đặc trưng phù hợp, các phương pháp này vẫn còn hạn chế. Mặc dù các nhà nghiên cứu đã đưa ra nhiều kỹ thuật khác nhau để lựa chọn các đặc trưng thích hợp nhằm nhận diện loại tấn công, nhưng chúng ta vẫn thiếu các công cụ để xác định đâu là tập hợp đặc trưng tốt nhất.

Để phát hiện các cuộc tấn công trên từng máy chủ riêng lẻ, có các hệ thống phát hiện xâm nhập dựa trên máy chủ (**HIDS - Host-based Intrusion Detection System**), trong khi để giám sát hành vi bất thường trên toàn bộ mạng, có các hệ thống phát hiện xâm nhập mạng (**NIDS - Network-based Intrusion Detection System**).

Các đặc trưng được phân loại thành các nhóm khác nhau, sẽ được mô tả trong phần tiếp theo.

* 1. **Connectivity Features**

Những đặc trưng này chứa toàn bộ thông tin liên quan đến nguồn và đích. Chúng được phân loại theo mức độ hữu ích trong việc trích xuất thông tin liên quan đến các cuộc tấn công, bao gồm:

 **Cơ bản (Basic):** Đây là những đặc trưng thường chứa thông tin về luồng dữ liệu giữa nguồn và đích, chẳng hạn như địa chỉ IP nguồn-đích, loại giao thức, v.v. Các đặc trưng này còn được gọi là đặc trưng định danh mạng hoặc đặc trưng tổng quát, và thường là yếu tố cần thiết để phát hiện tấn công. Một số ứng dụng IoT dựa trên các đặc trưng này bao gồm nhận diện thiết bị IoT [105], phát hiện xâm nhập [106] và nhận diện lưu lượng độc hại [107].

 **Thời gian (Time):** Đề cập đến thông tin về thời gian kết nối, trạng thái hoạt động, trạng thái chờ hoặc hướng của luồng dữ liệu. Các đặc trưng thời gian giúp mô tả chi tiết các kết nối trong mạng.

 **Dung lượng dữ liệu (Bytes):** Liên quan đến tổng số gói tin từ IP nguồn đến IP đích và tổng số byte được gửi theo hướng tiến hoặc lùi. Những đặc trưng này cung cấp thông tin quan trọng về các cuộc tấn công.

 **Cờ giao thức (Flag):** Cung cấp thông tin về cách các giao thức như TCP và MQTT bị tác động bởi nhiều loại tấn công khác nhau. Đặc trưng này phản ánh cách các cờ (flags) trong gói tin được thiết lập (hướng tiến hoặc lùi) và trạng thái kết nối.

* 1. **Các Đặc Trưng Phân Tầng (Layered Features)**

Sự gia tăng các cuộc tấn công mạng vào các thiết bị IoT đã tập trung vào loại giao thức truyền thông được sử dụng ở các tầng khác nhau. Vì giao thức là phương tiện truyền thông chính trên mạng, chúng trở thành mục tiêu dễ bị tấn công hơn. Để xác định các đặc trưng liên quan, các giao thức ở từng tầng cần được nhận diện và phân tích phù hợp.

Các ứng dụng sử dụng các đặc trưng này có thể dựa trên nhận diện dấu vân tay (fingerprint-based) [108], phát hiện bất thường (anomaly-based) [109], và phân tích lưu lượng mạng (traffic-based) [110].

Các tầng nơi các cuộc xâm nhập có thể xảy ra bao gồm:

* L4: Tầng Ứng Dụng (Application Layer)
* L3: Tầng Giao Vận (Transport Layer)
* L2: Tầng Mạng (Network Layer)
* L1: Tầng Giao Diện Mạng hoặc Cảm Biến (Network Interface or Sensing Layer)
  1. **Các Đặc Trưng Giao Tiếp (Communication Features)**

IoT đã mở rộng khái niệm về các thiết bị thông minh, trong đó nhiều thiết bị như đồng hồ đo thông minh, điện thoại thông minh, hay cảm biến không dây được kết nối thông qua internet bằng địa chỉ duy nhất. Những thiết bị này tương tác và phối hợp để đạt được mục tiêu cụ thể trong khoảng thời gian xác định. Tuy nhiên, sự đa dạng trong công nghệ cũng tạo ra môi trường thuận lợi cho các mối đe dọa an ninh mạng [111].

Bộ đặc trưng bao gồm:

* **Wi-Fi:** Các đặc trưng liên quan đến Wi-Fi giúp cải thiện độ chính xác của hệ thống phát hiện xâm nhập bằng cách tối ưu hóa yêu cầu xử lý và lưu trữ dữ liệu [112].
* **Bluetooth Low Energy (BLE):** BLE là một giao thức kết nối tiêu tốn ít năng lượng, thường được sử dụng trong các thiết bị công suất thấp. Tuy nhiên, BLE ngày càng trở thành mục tiêu của các cuộc tấn công mạng do tính phổ biến của nó trong các ứng dụng IoT.
* **Zigbee:** Giao thức Zigbee thường được sử dụng trong các thiết bị IoT như đồng hồ đo thông minh và hệ thống tự động hóa gia đình. Tuy nhiên, chính sự phổ biến này cũng khiến nó trở thành đối tượng bị khai thác lỗ hổng bảo mật. Khung dữ liệu Zigbee cung cấp nhiều đặc trưng quan trọng giúp chống lại các cuộc tấn công trên thiết bị IoT, và nhiều nghiên cứu đã ứng dụng học máy để nâng cao khả năng bảo vệ hệ thống [113] [114] [115].
  1. **Các Đặc Trưng Động**

Các đặc trưng động rất quan trọng trong việc phân tích hành vi của luồng thông tin. Chúng bao gồm các thống kê về lưu lượng mạng, chẳng hạn như kích thước gói tin, địa chỉ IP gốc và thời gian giữa các gói tin. Những đặc trưng này đặc biệt hữu ích trong việc phát hiện tấn công, vì chúng cung cấp thông tin về sự sai lệch so với các mẫu lưu lượng thông thường để xác định các cuộc tấn công [116].

Trong nghiên cứu [117], các tác giả đã tính toán các đặc trưng của dòng lưu lượng, bao gồm giá trị trung bình, giá trị lớn nhất và độ lệch chuẩn, để đề xuất một hệ thống phát hiện xâm nhập. Tương tự, một khung làm việc phát hiện xâm nhập dựa trên mạng được định nghĩa bằng phần mềm (SDN-IDS) đã được phát triển, trong đó sử dụng các đặc trưng thống kê để phân loại tấn công.

Các đặc trưng kết nối và phân tầng được sử dụng trong các tập dữ liệu IoT để phát hiện tấn công được trình bày trong Hình 7(a), các đặc trưng cờ trong Hình 7(b) và các đặc trưng động trong Hình 7(c). Trên thực tế, những đặc trưng này đóng vai trò quan trọng đối với các mô hình học máy, vì chúng cung cấp thông tin về các đặc điểm cơ bản của lưu lượng mạng, cho phép các thuật toán xác định các mẫu hành vi bình thường và các hoạt động độc hại.

1. **Thách Thức Chính & Kết Luận**

**IoT đã trở thành một phần thiết yếu trong cuộc sống hàng ngày của chúng ta.** Với khả năng đa dạng trong việc đáp ứng nhu cầu của người dùng, IoT đã trở thành một "trợ lý cá nhân" mà ai cũng muốn sử dụng. Tuy nhiên, chính sự phổ biến này đã thu hút các tin tặc, những kẻ có thể xâm nhập hệ thống bằng cách đánh cắp thông tin đăng nhập của người dùng hoặc gửi hàng loạt yêu cầu để làm sập hệ thống

Do đó, cộng đồng nghiên cứu đang nỗ lực phát triển các tập dữ liệu mô phỏng hoặc thực tế để cung cấp giải pháp phù hợp cho từng tình huống cụ thể. Công trình này đã xác định **mười tiêu chí** để phân tích và so sánh các tập dữ liệu IoT. Tuy nhiên, vẫn cần xem xét một số khía cạnh quan trọng liên quan đến các tập dữ liệu công khai.

A group of colorful ovals with text

AI-generated content may be incorrect.

**Hình 7: Biểu đồ Venn minh họa các nhóm đặc trưng khác nhau được sử dụng trong các tập dữ liệu IoT.**

**Lựa chọn Thiết bị và Chức năng:** Các tập dữ liệu được đề xuất bao gồm các thiết bị IoT thông minh như máy theo dõi trẻ em, chuông cửa, và camera an ninh. Công nghệ này đang thay đổi nhanh chóng, và các kỹ thuật nhúng, chức năng, tiêu chí cũng như thông số kỹ thuật có thể đã lỗi thời hoặc sẽ sớm trở nên lỗi thời theo thời gian. Trong nghiên cứu này, chúng tôi cũng nhận thấy rằng một số thiết bị trong các tập dữ liệu đề xuất chỉ có một chức năng (ví dụ: camera) trong khi những thiết bị khác có nhiều chức năng (ví dụ: máy theo dõi trẻ em). Điều này làm dấy lên lo ngại rằng kẻ tấn công có lợi thế hơn khi nhắm vào các hệ thống đa chức năng vì có nhiều cách để xâm nhập vào hệ thống mà không bị phát hiện. Do đó, chúng tôi hy vọng rằng các tập dữ liệu trong tương lai sẽ bao gồm các thiết bị cập nhật mới nhất để có thể dễ dàng xác định bất kỳ cuộc tấn công mới nào.

**Định dạng Lưu lượng Mạng:** Trong số các tập dữ liệu được thảo luận, chỉ có một số ít có tiếp xúc với dữ liệu thực tế trong môi trường mô phỏng. Ngoài ra, các công cụ thu thập dữ liệu cũng khác nhau, tức là không tuân theo một tiêu chuẩn định dạng cụ thể, điều này ảnh hưởng đến tính hữu ích của các hệ thống phát hiện xâm nhập.

**Các Loại Tấn công và Biến thể:** Trong quá trình khảo sát, chúng tôi nhận thấy rằng một số tập dữ liệu được đề xuất chỉ sử dụng các loại tấn công chính, trong khi một số khác phân loại theo các biến thể tấn công cụ thể. Điều này tạo ra một tình thế tiến thoái lưỡng nan về việc liệu nên đưa ra giải pháp cho loại tấn công chính hay một biến thể cụ thể. Vì các hệ thống được đề xuất cần duy trì một cái nhìn toàn diện về các kịch bản thực tế, tốt nhất là phải hiểu cả hai cấp độ tấn công này.

**Phương pháp Tiếp cận Hoàn Hảo:** Các phương pháp mới liên tục được đề xuất để phát hiện các cuộc tấn công vào hệ thống. Trong nghiên cứu của chúng tôi, không có phương pháp nào được xem là hoàn hảo để xác định một loại tấn công cụ thể. Bởi vì một cuộc tấn công có thể được phát hiện bằng nhiều phương pháp khác nhau hoặc bằng sự kết hợp của các kỹ thuật học máy và học sâu, nên cần sử dụng một hoặc nhiều phương pháp để phát hiện nhiều loại tấn công khác nhau. Do đó, chúng tôi khuyến nghị nghiên cứu sâu hơn để đề xuất các giải pháp tối ưu hơn.

**Cập nhật bảo mật:** Các nhà sản xuất IoT có xu hướng tập trung vào việc sản xuất và phân phối thiết bị của họ nhanh nhất có thể mà không quá chú trọng đến vấn đề bảo mật. Ngoài ra, hầu hết các nhà sản xuất chỉ cung cấp các bản cập nhật firmware.

Với nhu cầu ngày càng tăng đối với các thiết bị IoT trong nhiều lĩnh vực như y tế, công nghiệp và giáo dục, môi trường IoT trở thành mục tiêu tiềm năng cho các cuộc tấn công từ kẻ xâm nhập. Ngược lại, các lỗ hổng bảo mật của thiết bị đe dọa sự an toàn và quyền riêng tư của người dùng. Do đó, cần phát triển một giải pháp bảo mật vững chắc cho IoT để bao quát nhiều dạng tấn công khác nhau, bao gồm cả các cuộc tấn công zero-day.

Bài báo này trình bày một khảo sát toàn diện về các phương pháp tiên tiến nhất trong việc phát hiện xâm nhập vào các tập dữ liệu IoT. Chúng tôi sử dụng các tiêu chí như mô tả giao thức, các loại tấn công, lỗ hổng bảo mật, đặc điểm dữ liệu, phương pháp phát hiện, độ chính xác, đồng thời xem xét liệu các nghiên cứu trước đây có đề cập đến các tập dữ liệu IoT hay không.

Chúng tôi đã thảo luận về các cuộc tấn công bảo mật trong IoT đe dọa nhiều lớp trong mô hình mạng, nhấn mạnh các dạng tấn công cụ thể xảy ra tại từng lớp. Do các cuộc tấn công này nhắm vào các giao thức hoạt động ở nhiều lớp khác nhau, chúng tôi cũng phân tích các lỗ hổng bị khai thác trong bộ giao thức TCP/IP.

Ngoài ra, bài báo trình bày chi tiết về các tập dữ liệu IoT công khai, bao gồm loại định dạng dữ liệu được sử dụng, dữ liệu có được gán nhãn hay không, thời gian thu thập dữ liệu, loại mạng, quốc gia thực hiện thí nghiệm, số lượng thiết bị tham gia thử nghiệm, cũng như số lượng mẫu dữ liệu lành tính và độc hại được thu thập.

Chúng tôi cũng đã thảo luận về các phương pháp học sâu (DL) và học máy (ML) khác nhau được áp dụng để phát hiện các cuộc tấn công vào thiết bị IoT. Cuối cùng, chúng tôi đã liệt kê các đặc trưng tiềm năng có thể giúp phát hiện hành vi bất thường trong thiết bị IoT.

1. **Các Bảng Bổ Sung**

Nhằm cung cấp thêm thông tin chi tiết về đặc điểm của các tập dữ liệu được nghiên cứu trong bài báo này, chúng tôi trình bày mô tả về các cuộc tấn công IoT ở các lớp khác nhau (Bảng 6), tóm tắt chi tiết về các thuộc tính của tập dữ liệu IoT (Bảng 7), danh sách các đặc trưng tiềm năng (Bảng 8) và danh sách các đặc trưng được sử dụng trong các tập dữ liệu IoT (Bảng 9).

**Lời Cảm Ơn**

Các tác giả xin gửi lời cảm ơn đến Viện An ninh mạng Canada (CIC) vì sự hỗ trợ tài chính và giáo dục. Dự án này cũng đã nhận được sự hỗ trợ một phần từ nguồn tài trợ nghiên cứu hợp tác của Chương trình Trí tuệ Nhân tạo cho Logistics thuộc Hội đồng Nghiên cứu Quốc gia Canada.

A white sheet with black text

AI-generated content may be incorrect.

**Bảng 6: Mô tả các cuộc tấn công vào IoT tại các lớp khác nhau.**