# **THÔNG TIN CHUNG CỦA NHÓM**

* Link YouTube video của báo cáo (tối đa 5 phút):   
  *https://youtu.be/ud42ZaGuBDc*
* Link slides (dạng .pdf đặt trên Github của nhóm):   
  *https://github.com/TruongGiaThach/Final\_Project-CS2205.CH183-240202012/blob/main/Th%E1%BA%A1ch%20Tr%C6%B0%C6%A1ng%20Gia%20-%20CS2205.NOV2024.DeCuong.FinalReport.Slide.pdf*
* *Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới*
* *Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in*
* *Lớp Cao học, mỗi nhóm một thành viên*

| * Họ và Tên: Trương Gia Thạch * MSSV: 240202012 | * Lớp: CS2205. * Tự đánh giá (điểm tổng kết môn): 9.5/10 * Số buổi vắng: 0 * Số câu hỏi QT cá nhân: 5 * Số câu hỏi QT của cả nhóm: 5 * Link Github: https://github.com/TruongGiaThach/Final\_Project-CS2205.CH183-240202012 |
| --- | --- |

# **ĐỀ CƯƠNG NGHIÊN CỨU**

| **TÊN ĐỀ TÀI (IN HOA)**  HỆ THỐNG PHÁT HIỆN XÂM NHẬP TỰ THÍCH NGHI ĐỂ PHÁT HIỆN CÁC KIỂU TẤN CÔNG MỚI |
| --- |
| **TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)**  ADAPTIVE INTRUSION DETECTION SYSTEM FOR DETECTING NEW ATTACK PATTERNS |
| **TÓM TẮT** *(Tối đa 400 từ)*  Hệ thống phát hiện xâm nhập (**Intrusion Detection System - IDS**) đóng vai trò quan trọng trong **an ninh mạng**, giúp bảo vệ hệ thống khỏi các cuộc **tấn công tinh vi**. Tuy nhiên, các **IDS truyền thống** gặp phải **hạn chế lớn** khi đối mặt với **zero-day attacks** và các phương thức tấn công mới do phụ thuộc vào **mẫu dữ liệu tĩnh** và **thiếu khả năng tự thích nghi**.  **Nghiên cứu này đề xuất một hệ thống IDS tự thích nghi (Adaptive IDS)** kết hợp **Improved Random Forest (IRF) và Adaptive Learning Mechanism**, nhằm:   1. **Tự động cập nhật và điều chỉnh trọng số mô hình khi phát hiện tấn công mới**, giúp IDS học hỏi liên tục mà **không cần huấn luyện lại toàn bộ mô hình**. 2. **Tối ưu hóa thuật toán phân loại IDS**, cải thiện độ chính xác trong phát hiện tấn công. 3. **Giảm tỷ lệ báo động sai (False Positive Rate - FPR)**, khắc phục nhược điểm của IDS truyền thống.   Mô hình đề xuất sẽ được **thử nghiệm trên bộ dữ liệu thực tế CICIDS2017 và UNSW-NB15**, với các tiêu chí đánh giá gồm **độ chính xác (Accuracy), Precision, Recall, F1-score, thời gian thích nghi và mức tiêu thụ tài nguyên**. Hệ thống được kỳ vọng sẽ giúp **nâng cao độ chính xác, giảm thời gian phát hiện so với IDS truyền thống**, đồng thời **tăng khả năng phát hiện tấn công zero-day** mà không làm gián đoạn hệ thống. |
| **GIỚI THIỆU** *(Tối đa 1 trang A4)*  **2.1. Bối cảnh nghiên cứu**  An ninh mạng đang ngày càng trở nên quan trọng khi các tổ chức, doanh nghiệp và hệ thống IoT ngày càng phụ thuộc vào kết nối mạng. **Các cuộc tấn công mạng hiện đại như zero-day attacks, AI-driven attacks, và APTs (Advanced Persistent Threats)** đã vượt qua khả năng phát hiện của các IDS truyền thống.  Hạn chế chính của IDS hiện nay:   * **Không thể nhận diện tấn công mới** do phụ thuộc vào dữ liệu tĩnh. * **Cảnh báo sai (False Positive) cao**, gây mất thời gian và tài nguyên. * **Không có cơ chế học thích nghi**, đòi hỏi phải huấn luyện lại thủ công.   **2.2. Giải pháp đề xuất**  Để giải quyết các vấn đề trên, nghiên cứu đề xuất một hệ thống **Adaptive IDS**, có khả năng tự cập nhật và thích nghi với các loại tấn công mới:   * **Sử dụng Improved RF** để tối ưu hóa hiệu suất phát hiện. * **Kết hợp Adaptive Learning Mechanism**, giúp IDS học liên tục từ dữ liệu mới.   **2.3. Input / Output của hệ thống**  • **Input:** Lưu lượng mạng từ CICIDS2017, UNSW-NB15, chứa cả tấn công đã biết và tấn công mới.  • **Output:** IDS có thể học từ dữ liệu mới, thích nghi với kiểu tấn công mới mà không cần đào tạo lại toàn bộ mô hình. |
| **MỤC TIÊU** *(Viết trong vòng 3 mục tiêu)*  • Phát triển hệ thống IDS tự thích nghi (Adaptive IDS), có khả năng học liên tục từ dữ liệu mới mà không cần huấn luyện lại toàn bộ.  • Tích hợp thuật toán Improved RF và Adaptive Learning Mechanism để tối ưu hóa khả năng phát hiện tấn công.  • Đánh giá hiệu suất mô hình dựa trên:  o Tỷ lệ phát hiện tấn công mới (Detection Rate - DR).  o Độ chính xác (Accuracy), Precision, Recall, F1-score.  o Thời gian thích nghi của IDS khi có kiểu tấn công mới. |
| **NỘI DUNG VÀ PHƯƠNG PHÁP**  Hệ thống phát hiện xâm nhập tự thích nghi (Adaptive IDS) trong nghiên cứu này được phát triển dựa trên nền tảng **Improved Random Forest (IRF) kết hợp cơ chế Adaptive Learning**, nhằm giải quyết các vấn đề của IDS truyền thống. Trong phần này, chúng tôi trình bày **quy trình nghiên cứu**, **phương pháp triển khai**, và **cách đánh giá hiệu suất hệ thống**.  **4.1. Nội dung nghiên cứu**  Nội dung nghiên cứu được chia thành ba phần chính:  • Khảo sát và phân tích các IDS truyền thống, xác định hạn chế  • Phát triển hệ thống Adaptive IDS với Improved RF và Adaptive Learning  • Thử nghiệm mô hình và đánh giá hiệu suất so với IDS truyền thống  **4.2. Phương pháp nghiên cứu**  **Nội dung 1: Khảo sát các phương pháp phát hiện xâm nhập truyền thống và đề xuất mô hình IDS tự thích nghi**  **Mục tiêu:**   * Khảo sát các hệ thống phát hiện xâm nhập (IDS) hiện nay, tập trung vào **Signature-based IDS, Anomaly-based IDS** và **Machine Learning-based IDS**. * Phân tích hạn chế của IDS truyền thống khi phát hiện **zero-day attacks** và các cuộc tấn công tinh vi sử dụng **AI-driven attacks**. * Xây dựng đề xuất mô hình **Adaptive IDS**, kết hợp **Improved Random Forest (IRF)** và **Adaptive Learning Mechanism**.   **Phương pháp thực hiện:**   * **Nghiên cứu tài liệu:** Tổng hợp từ các bài báo khoa học về IDS truyền thống và hiện đại, các cơ chế phát hiện tấn công zero-day. * **Thống kê và phân tích hiệu suất:** So sánh các tiêu chí đánh giá hiệu suất IDS hiện tại, bao gồm độ chính xác (Accuracy), tỷ lệ phát hiện tấn công (Detection Rate - DR), tỷ lệ báo động sai (False Positive Rate - FPR) và khả năng mở rộng trên dữ liệu lớn. * **Phân tích dữ liệu thực tế:** Sử dụng tập dữ liệu **CICIDS2017, UNSW-NB15**, đánh giá mức độ hiệu quả của các IDS truyền thống trong việc phát hiện tấn công.   **Nội dung 2: Phát triển mô hình IDS tự thích nghi dựa trên Improved RF và Adaptive Learning Mechanism**  **Mục tiêu:**   * Phát triển hệ thống IDS có khả năng **cập nhật và học hỏi từ dữ liệu tấn công mới** mà không cần huấn luyện lại toàn bộ mô hình. * Tối ưu hóa hiệu suất bằng cách sử dụng **Improved RF (IRF)** và cơ chế **Adaptive Learning**.   **Phương pháp thực hiện:**   1. **Cải tiến thuật toán Random Forest:**    * Áp dụng **feature selection thông minh** để loại bỏ các đặc trưng không quan trọng, tối ưu hóa khả năng phân loại.    * Cải thiện độ chính xác bằng **Gini Index tối ưu**, giúp IDS đưa ra quyết định tốt hơn khi phân loại lưu lượng mạng là bình thường hay độc hại.    * **Cân bằng dữ liệu (Data Balancing):** Giảm thiểu sự thiên lệch giữa các lớp tấn công và lớp bình thường, tránh hiện tượng mất cân bằng dữ liệu trong IDS. 2. **Thiết lập cơ chế Adaptive Learning:**    * Áp dụng **Incremental Learning**, giúp IDS **cập nhật dữ liệu mới mà không cần đào tạo lại từ đầu**.    * **Cơ chế điều chỉnh trọng số động (Dynamic Weight Adjustment):** IDS sẽ học từ tấn công mới mà không làm mất đi dữ liệu cũ.   **Nội dung 3: Kiểm thử mô hình và đánh giá hiệu suất IDS tự thích nghi**  **Mục tiêu:**   * Đánh giá **độ chính xác**, **tốc độ phản hồi**, **khả năng thích nghi** của IDS so với IDS truyền thống. * So sánh hiệu suất của IDS tự thích nghi với IDS không có cơ chế học liên tục.   **Phương pháp thực hiện:**   1. **Chạy thử nghiệm trên tập dữ liệu thực tế:**    * IDS sẽ được huấn luyện trên **CICIDS2017, UNSW-NB15** và kiểm tra khả năng phát hiện tấn công mới. 2. **Đánh giá hiệu suất dựa trên các tiêu chí chính:**    * **Độ chính xác (Accuracy):** Tỷ lệ phát hiện đúng các cuộc tấn công.    * **Tỷ lệ phát hiện tấn công (Detection Rate - DR):** Khả năng phát hiện các cuộc tấn công mới.    * **Tỷ lệ báo động sai (False Positive Rate - FPR):** Độ tin cậy của IDS khi cảnh báo tấn công.    * **Thời gian phản hồi (Response Time):** Thời gian cần để phát hiện một cuộc tấn công mới. |
| **KẾT QUẢ MONG ĐỢI**  **Cải thiện đáng kể hiệu suất IDS truyền thống:**   * Hệ thống IDS có thể thích nghi với các kiểu tấn công mới mà không cần đào tạo lại. * Độ chính xác trên 90%, giảm False Positive Rate xuống dưới 5%. * Tăng tốc độ phát hiện nhanh hơn ít nhất 30% so với IDS truyền thống.   **Báo cáo đánh giá đầy đủ về:**   * Khả năng học thích nghi của IDS khi dữ liệu tấn công thay đổi. * Hiệu suất trên dữ liệu thực tế và so sánh với IDS truyền thống. |
| **TÀI LIỆU THAM KHẢO** *(Định dạng DBLP)*  [1] CICIDS2017 Dataset. [Online]. Available: <https://www.unb.ca/cic/datasets/ids-2017.html>  [2] UNSW-NB15 Dataset. [Online]. Available: https://www.unsw.adfa.edu.au/unsw-canberra-cyber/cybersecurity/ADFA-NB15-Datasets/  [3] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. doi: 10.1023/A:1010933404324. [Online]. Available: https://doi.org/10.1023/A:1010933404324  [4] J. Zhang and D.-M. Zhao, “Network Malicious Data Intrusion Detection Combining Distributed Network and Improved RF Algorithm under Spark Framework,” *Journal of Network Intelligence*, vol. 9, no. 3, pp. 1820-1835, 2024. doi: 10.32604/jni.2024.1835. [Online]. Available: [34.JNI-S-2023-12-001.pdf](https://bit.kuas.edu.tw/~jni/2024/vol9/s3/34.JNI-S-2023-12-001.pdf)  [5] M. Zaharia et al., “Spark: Cluster Computing with Working Sets,” in *USENIX Conference on Hot Topics in Cloud Computing*, 2010. [Online]. Available: https://www.usenix.org/conference/hotcloud-10/spark-cluster-computing-working-sets  [6] T. Kajiura and T. Nakamura, “Practical Performance of a Distributed Processing Framework for Machine-Learning-based NIDS,” *IEEE Transactions on Dependable and Secure Computing*, 2024. doi: 10.1109/TDSC.2024.10633597. [Online]. Available:<https://ieeexplore.ieee.org/abstract/document/10633597/>  [7] M. A. Shyaa, N. F. Ibrahim, Z. B. Zainol, R. Abdullah, and F. M. J. Ibrahim, “Reinforcement Learning-Based Voting for Feature Drift-Aware Intrusion Detection: An Incremental Learning Framework,” *IEEE*, 2025. doi: 10.1109/ICCEAI.2025.10896652. [Online]. Available:<https://ieeexplore.ieee.org/abstract/document/10896652/>  [8] A. A. Darem, F. A. Ghaleb, A. A. Al-Hashmi et al., “An Adaptive Behavioral-Based Incremental Batch Learning Malware Variants Detection Model Using Concept Drift Detection and Sequential Deep Learning,” *IEEE Access*, vol. 9, pp. 4321-4338, 2021. doi: 10.1109/ACCESS.2021.9467300. [Online]. Available:<https://ieeexplore.ieee.org/abstract/document/9467300/>  [9] Z. Jin, J. Zhou, B. Li, X. Wu, and C. Duan, “FL-IIDS: A Novel Federated Learning-Based Incremental Intrusion Detection System,” *Future Generation Computer Systems*, vol. 148, pp. 184-202, 2024. doi: 10.1016/j.future.2024.01.003. [Online]. Available:<https://www.sciencedirect.com/science/article/pii/S0167739X23003503>  [10] E. Mahdavi, A. Fanian, and A. Mirzaei, “ITL-IDS: Incremental Transfer Learning for Intrusion Detection Systems,” *Applied Soft Computing*, vol. 114, p. 107778, 2022. doi: 10.1016/j.asoc.2021.107778. [Online]. Available:<https://www.sciencedirect.com/science/article/pii/S0950705122007778>  [11] L. Cui, Z. Wu, P. Gao, and J. Chen, “An Incremental Learning Method Based on Dynamic Ensemble RVM for Intrusion Detection,” *IEEE Transactions on Network and Service Management*, vol. 18, no. 3, pp. 3215-3227, 2021. doi: 10.1109/TNSM.2021.9506882. [Online]. Available:<https://ieeexplore.ieee.org/abstract/document/9506882/> |