ỨNG DỤNG MÔ HÌNH LSTM ĐƯỢC SỬ DỤNG ĐỂ PHÂN LOẠI TẤN CÔNG DDOS

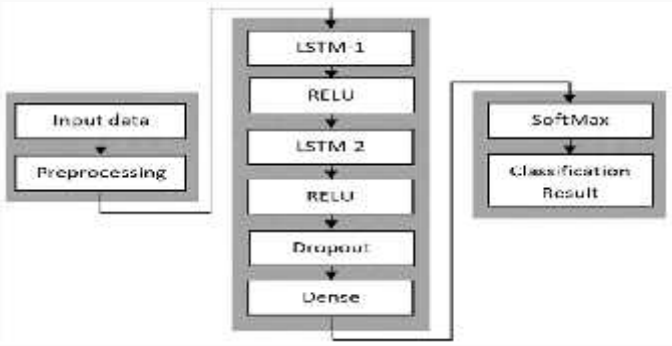
Trương Công Bình - 230201039

Tóm tắt

- Lớp: CS2205.APR2023
- Link Github: https://github.com/BinhTruongCong/CS2205.APR2023
- Link YouTube video:
- Ảnh + Họ và Tên:Trương Công Bình
- Tổng số slides không vượt quá 10

Giới thiệu

- Các cuộc tấn công từ chối dịch vụ (DoS) là một trong những cuộc tấn công phổ biến và đe dọa nhất đối với an ninh mạng. Các cuộc tấn công này nhằm làm quá tải hệ thống mục tiêu bằng lưu lượng truy cập ảo, khiến hệ thống không thể đáp ứng các yêu cầu hợp pháp của người dùng.



Kiến trúc mô hình phân loại LSTM

- Đánh giá hiệu suất LSTM trong việc phát hiện các loại tấn công DDoS khác nhau bằng cách sử dụng các bộ dữ liệu CIC được sử dụng phổ biến và công khai
- Việc phân loại chính xác lưu lượng mạng giúp phát hiện sớm các cuộc tấn công DDoS và thực hiện các biện pháp phòng thủ kịp thời

Giới thiệu

- Một nghiên cứu đã sử dụng mô hình LSTM để phân loại lưu lượng mạng trong mạng máy tính và đạt được độ chính xác cao hơn 98% trong việc phát hiện các cuộc tấn công DDoS.
- Một nghiên cứu khác đã sử dụng mô hình LSTM để phân loại lưu lượng mạng trong mạng điện thoại di động và đạt được độ chính xác cao hơn 95% trong việc phát hiện các cuộc tấn công DDoS

Mục tiêu

- Tăng cường khả năng thích ứng với các loại tấn công mới
- Hỗ trợ việc phân tích dữ liệu và ra quyết định
- Phát triển các hệ thống phòng chống DDoS hiệu quả hơn

Nội dung và Phương pháp

- **Nội dung:**
- Mô hình mạng nơ-ron dài ngắn hạn (LSTM) là một loại mạng nơ-ron nhân tạo (ANN) được sử dụng phổ biến trong xử lý ngôn ngữ tự nhiên, nhận dạng giọng nói, và phân tích chuỗi thời gian. Nó được thiết kế để giải quyết vấn đề biến mất gradient, một vấn đề thường gặp trong các mạng nơ-ron truyền thống khi xử lý các chuỗi dữ liệu dài.
- Mô hình LSTM cho phân loại tấn công DDoS thường bao gồm các thành phần sau:
- Lớp thu thập dữ liệu: Dữ liệu lưu lượng truy cập mạng được ghi lại, bao gồm các tính năng như địa chỉ IP nguồn và đích, cổng, giao thức, byte được truyền
- Lớp tiền xử lý dữ liệu: Dữ liệu được thu thập được làm sạch, chuẩn hóa và định dạng để tương thích với LSTM.
- Lớp LSTM: Bao gồm nhiều lớp nơ-ron LSTM được kết nối với nhau. Các nơ-ron LSTM có khả năng học hỏi các phụ thuộc thời gian trong dữ liệu lưu lượng mạng.
- Lớp phân loại: Phân loại dữ liệu đầu vào là lưu lượng truy cập bình thường hay tấn công DDoS.
- Sử dụng mô hình LSTM để phân loại 17 loại tấn công DDoS.
- Áp dụng các phương pháp LIME, SHAP, Anchor và LORE để giải thích dự đoán của mô hình LSTM

Nội dung và Phương pháp

- **PHƯƠNG PHÁP**
- **LIME (Local Interpretable Model-Agnostic Explanations)**

— LIME tạo ra các mô hình giải thích cục bộ bằng cách xấp xỉ mô hình LSTM bằng một mô hình tuyến tính đơn giản trong vùng lân cận của điểm dữ liệu được giải thích.

— Mô hình tuyến tính này được đào tạo trên một tập dữ liệu nhỏ được tạo ra từ điểm dữ liệu được giải thích và các điểm dữ liệu lân cận.

— Các trọng số của mô hình tuyến tính được sử dụng để xác định các đặc điểm quan trọng nhất cho dự đoán của LSTM tại điểm dữ liệu được giải thích.

- **SHAP (SHapley Additive exPlanations):**

— SHAP sử dụng phân bổ Shapley để phân bổ mức độ ảnh hưởng của mỗi đặc điểm đối với dự đoán của LSTM.

— Phân bổ Shapley là một phương pháp phân bổ giá trị công bằng cho mỗi người chơi trong một trò chơi hợp tác.

Nội dung và Phương pháp

- **Anchor:**

- Anchor tìm kiếm các điểm dữ liệu lân cận với điểm dữ liệu được giải thích có cùng dự đoán với LSTM.
- Các điểm dữ liệu này được gọi là "anchor" và được sử dụng để giải thích dự đoán của LSTM tại điểm dữ liệu được giải thích.

- **LORE (Local Optimal Reconstruction Explanation):**

- LORE sử dụng phương pháp tối ưu hóa cục bộ để tìm kiếm một tập con nhỏ các đặc điểm có thể tái tạo dự đoán của LSTM tại điểm dữ liệu được giải thích.
- Tập con này được gọi là "LORE" và được sử dụng để giải thích dự đoán của LSTM.
- LORE có thể được sử dụng để xác định các đặc điểm quan trọng nhất cho dự đoán của LSTM tại điểm dữ liệu được giải thích.

Kết quả dự kiến

- Mô hình LSTM đạt được độ chính xác cao trong việc phân loại các cuộc tấn công DDoS.
- Các phương pháp giải thích giúp hiểu rõ cách thức hoạt động của mô hình LSTM.
- 51 đặc điểm quan trọng được xác định để phân loại tấn công DDoS.
- Phương pháp LIME đạt hiệu suất tốt nhất về độ chính xác mô tả (DA) và độ thưa thớt mô tả (DS).

Tài liệu tham khảo

- Almaiah, M.A. Almaiah, M.A. A New Scheme for Detecting Malicious Attacks in Wireless Sensor Networks Based on Blockchain Technology. In Artificial Intelligence and Blockchain for Future Cybersecurity Applications; Springer: Berlin/Heidelberg, Germany, 2021; pp. 217–234. [Google Scholar]
- Zargar, S.T.; Joshi, J.; Tipper, D. A Survey of Defense Mechanisms Against Distributed Denial of Service (DDoS) Flooding Attacks. IEEE Commun. Surv. Tutor. 2013, 15, 2046–2069. [Google Scholar]
- Hou, J.; Fu, P.; Cao, Z.; Xu, A. Machine Learning Based DDos Detection Through NetFlow Analysis. In Proceedings of the IEEE Military Communications Conference MILCOM, Los Angeles, CA, USA, 29 October 2018. [Google Scholar]
- DDoS Attacks History. Radware. Available online: https://www.radware.com/security/ddos-knowledge-center/ddos-chronicles/ddos-attacks-history/ (accessed on 17 July 2023).
- Choi, H.; Lee, H. Identifying Botnets by Capturing Group Activities in DNS Traffic. Comput. Netw. 2012, 56, 20–33. [Google Scholar]
- Suresh, S.; Ram, N. A Review on Various DPM Traceback Schemes to Detect DDoS Attacks. Indian J. Sci. Technol. 2016, 9, 1–8. [Google Scholar] [CrossRef] [Green Version]
- Argyraki, K.; Cheriton, D. Active Internet Traffic Filtering: Real-Time Response to Denial of Service Attacks. arXiv 2003, arXiv:cs/0309054. [Google Scholar]
- Anjum, F.; Subhadrabandhu, D.; Sarkar, S. Signature Based Intrusion Detection for Wireless Ad-Hoc Networks: A Comparative Study of Various Routing Protocols. In Proceedings of the IEEE 58th Vehicular Technology Conference, Orlando, FL, USA, 6 October 2003. [Google Scholar]
- Cloudflare DDoS Threat Report 2022 Q3. Cloudflare. Available online: https://blog.cloudflare.com/cloudflare-ddos-threat-report-2022-q3/ (accessed on 17 July 2023).
- Hoque, N.; Kashyap, H.; Bhattacharyya, D.K. Real-Time DDoS Attack Detection Using FPGA. Comput. Commun. 2017, 110, 48–58. [Google Scholar] [CrossRef]