

ĐẠI HỌC BÁCH KHOA HÀ NỘI

ĐỒ ÁN TỐT NGHIỆP

Xây dựng công cụ phát hiện nguy cơ webshell trong hệ
thống web

NGUYỄN TUẤN THÀNH

thanh.nt205027@sis.hust.edu.vn

Ngành Công nghệ thông tin Việt - Nhật

Giảng viên hướng dẫn: TS. Nguyễn Hữu Đức

Chữ kí GVHD

Khoa: Khoa học Máy tính

Trường: Công nghệ Thông tin và Truyền thông

HÀ NỘI, 12/2024

LỜI CAM KẾT

Họ và tên sinh viên: Nguyễn Tuấn Thành
MSSV: 20205027
Điện thoại liên lạc: 0936935346
Email: thanh.nt205027@sis.hust.edu.vn
Lớp: Việt Nhật 01 - K65
Chương trình đào tạo: Việt-Nhật.....

Tôi – *Nguyễn Tuấn Thành* – cam kết Đồ án Tốt nghiệp (ĐATN) là công trình nghiên cứu của bản thân tôi dưới sự hướng dẫn của *TS. Nguyễn Hữu Đức*. Các kết quả nêu trong ĐATN là trung thực, là thành quả của riêng tôi, không sao chép theo bất kỳ công trình nào khác. Tất cả những tham khảo trong ĐATN – bao gồm hình ảnh, bảng biểu, số liệu, và các câu từ trích dẫn – đều được ghi rõ ràng và đầy đủ nguồn gốc trong danh mục tài liệu tham khảo. Tôi xin hoàn toàn chịu trách nhiệm với dù chỉ một sao chép vi phạm quy chế của nhà trường.

Hà Nội, ngày tháng năm

Tác giả ĐATN

Nguyễn Tuấn Thành

LỜI CẢM ƠN

Em xin chân thành cảm ơn gia đình, đặc biệt là bố mẹ, đã luôn động viên và ủng hộ em suốt quá trình học tập và thực hiện đồ án tốt nghiệp. em cũng xin gửi lời tri ân sâu sắc đến các thầy cô, đặc biệt là thầy Nguyễn Hữu Đức, đã tận tình chỉ bảo trong quá trình em thực hiện đồ án. Bên cạnh đó, em cảm ơn các bạn bè đã luôn đồng hành, chia sẻ và giúp đỡ em vượt qua những khó khăn. Cuối cùng, em tự hào về chính mình vì đã chăm chỉ và quyết tâm hoàn thành đồ án với kết quả tốt nhất. Xin trân trọng cảm ơn!

TÓM TẮT NỘI DUNG ĐỒ ÁN

Vấn đề bảo mật cho các dịch vụ Web luôn là một chủ đề nóng được chú trọng ngày nay. Các website với cơ chế bảo mật yếu có thể trở thành mục tiêu tấn công dễ dàng bởi kẻ xấu. Trong các cuộc tấn công hướng tới mục tiêu website của các tổ chức/doanh nghiệp, webshell thường được sử dụng như một trong những điểm khởi đầu để xâm nhập vào hệ thống. Vì vậy, việc rà quét nguy cơ webshell trở thành một phương pháp hiệu quả trong thực tế để xác định con đường attacker khai thác lỗ hổng trên website để ngăn chặn các bước xâm nhập tiếp theo.

Đồ án tốt nghiệp này tiếp cận theo hướng rà quét tự động, khắc phục hạn chế trên bằng cách xây dựng công cụ hướng tới người dùng không có nhiều kiến thức chuyên môn. Công cụ được xây dựng trong đồ án tốt nghiệp đã đạt được kết quả chỉ cần người dùng khởi chạy mà không cần nhập thêm bất kỳ thông tin nào. Công cụ cũng hướng tới việc cung cấp một giải pháp rà quét độc lập, có tính thực tiễn và ứng dụng cao.

MỤC LỤC

MỤC THUẬT NGỮ VÀ TỪ VIẾT TẮT	vii
CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI.....	1
1.1 Đặt vấn đề.....	1
1.2 Mục tiêu và nhiệm vụ của đồ án.....	1
1.3 Tổng quan hệ thống và phạm vi đề tài	2
1.4 Bố cục đồ án	5
CHƯƠNG 2. NỀN TẢNG LÝ THUYẾT	6
2.1 Khái niệm webshell	6
2.2 Nguyên cơ an ninh mạng Webshell.....	6
2.3 Kịch bản tấn công điển hình sử dụng webshell.....	7
2.3.1 Công cụ chiếm quyền truy cập trái phép vào hệ thống	7
2.3.2 Công cụ duy trì kết nối đến hệ thống.....	8
2.4 Phát hiện Webshell trong các giải pháp hiện có	8
2.4.1 So sánh với dữ liệu đã biết	9
2.4.2 Phát hiện bất thường trong lưu lượng web (<i>web traffic</i>).....	9
2.4.3 Phát hiện bất thường dựa trên chữ ký	9
2.4.4 Phát hiện luồng mạng lạ	9
2.4.5 Sử dụng EDR	10
2.5 Khái niệm Memory Webshell - Memshell.....	10
CHƯƠNG 3. PHÂN TÍCH BÀI TOÁN	16
3.1 Giới thiệu chức năng.....	16
3.2 Các bài toán cần giải quyết	19

CHƯƠNG 4. THIẾT KẾ GIẢI PHÁP	21
4.1 Thiết kế chức năng.....	21
4.1.1 Chức năng "Rà quét thông tin webserver".....	21
4.1.2 Chức năng "Rà quét tự động nguy cơ webshell"	22
4.1.3 Chức năng "Rà quét tự động nguy cơ webshell tùy chỉnh"	24
4.1.4 Chức năng "Rà quét thủ công nguy cơ webshell".....	26
4.1.5 Chức năng "Rà quét nguy cơ memory webshell"	27
4.2 Kiến trúc phần mềm.....	27
4.2.1 Webserver Scanner Engine.....	27
4.2.2 Traverse Engine	32
4.2.3 Scan Engine	34
4.3 Kết luận	37
CHƯƠNG 5. TRIỂN KHAI VÀ THỬ NGHIỆM	38
5.1 Thiết lập cho quá trình triển khai và đánh giá	38
5.1.1 Cấu hình môi trường.....	38
5.1.2 Bộ dữ liệu sử dụng	38
5.1.3 Kịch bản Memory Webshell IIS	39
5.1.4 Test Cases	40
5.2 Kết quả thực nghiệm	41
5.2.1 Kiểm tra phát hiện webserver	41
5.2.2 Kiểm tra khả năng nhận diện webshell	42
5.2.3 Kiểm tra khả năng nhận diện Memshell IIS	48
5.3 Đánh giá và so sánh	48
CHƯƠNG 6. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	50
6.1 Kết luận	50
6.2 Hướng phát triển.....	51

CHƯƠNG 7. TÀI LIỆU THAM KHẢO	52
---	-----------

DANH MỤC HÌNH VẼ

Hình 1.1	Biểu đồ thị phần các webserver	5
Hình 2.1	So sánh số cuộc tấn công sử dụng webshell trong hai năm . . .	7
Hình 2.2	Thống kê các nguy cơ hàng đầu trong Quý 1 2023	7
Hình 2.3	Ví dụ một attack vector sử dụng webshell	8
Hình 2.4	Webshell với vai trò là backdoor trên hệ thống	8
Hình 2.5	Các thuộc tính của một website được quản lý trong IIS Manager.	11
Hình 2.6	Một ứng dụng con bên trong <i>Root Application</i>	11
Hình 2.7	Dữ liệu cache của <i>dll</i>	12
Hình 2.8	Các dữ kiện trong quá trình xử lý file.	13
Hình 2.9	Xuất hiện file <i>.dll.delete</i> khi file source bị xóa.	13
Hình 2.10	Cấu hình Audit cho giám sát xóa folder/file.	14
Hình 2.11	Sử dụng <i>dnSpy</i> để decompile ASP.NET <i>dll</i>	15
Hình 4.1	Các chức năng có trong hướng dẫn sử dụng công cụ	21
Hình 4.2	Thiết kế các thành phần chính	27
Hình 4.3	Cấu trúc Scan Engine trong source code	34
Hình 5.1	Webroot của website <i>memshell.localhost</i> và tệp tin webshell About.aspx	39
Hình 5.2	Attacker truy cập webshell About.aspx	39
Hình 5.3	Attacker được chuyển hướng tới webshell đích tại uri <i>/1/ghostfile24.aspx</i>	40
Hình 5.4	Các tệp tin liên quan của <i>ghostfile24.aspx</i> tại thư mục Cache IIS	40
Hình 5.5	Các tệp tin liên quan của <i>ghostfile24.aspx</i> tại thư mục Cache IIS	40
Hình 5.6	Nhận diện webserver IIS và các webroots được cấu hình . . .	41
Hình 5.7	Nhận diện webserver Apache Tomcat và các webroots được cấu hình	41
Hình 5.8	Nhận diện webserver Nginx và các webroots được cấu hình . .	42
Hình 5.9	Kết quả thực thi công cụ ShellSweep với bộ dữ liệu Php - Webshell	44
Hình 5.10	Kết quả thực thi công cụ ShellSweep với bộ dữ liệu Php - Benign	44
Hình 5.11	Kết quả thực thi công cụ ShellSweep với bộ dữ liệu Asp - Webshell	44
Hình 5.12	Kết quả thực thi công cụ ShellSweep với bộ dữ liệu Asp - Benign	44
Hình 5.13	Kết quả thực thi công cụ WebshellScanner với bộ dữ liệu Php - Webshell	45

Hình 5.14	Kết quả thực thi công cụ WebshellScanner với bộ dữ liệu Php	
- Benign	46
Hình 5.15	Kết quả thực thi công cụ WebshellScanner với bộ dữ liệu Asp	
- Webshell	46
Hình 5.16	Kết quả thực thi công cụ WebshellScanner với bộ dữ liệu Asp	
- Benign	47
Hình 5.17	Kết quả thực thi công cụ WebshellScanner - chỉ quét Memory	
Webshell	48
Hình 5.18	Công cụ WebshellScanner chỉ ra VirtualPath của ghostfile24	48