**Mối đe dọa bảo mật cơ sở dữ liệu và Thách thức trong Pháp y Cơ sở dữ liệu: Tổng quan**

**Tóm tắt.** Hệ thống Quản lý Cơ sở dữ liệu Quan hệ (RDBMS) là tập hợp các ứng dụng quản lý lưu trữ, truy xuất và thao tác dữ liệu cơ sở dữ liệu. Ở cấp độ ngành, SQL Server, Oracle, Sybase, DB2, MySQL và các ứng dụng cơ sở dữ liệu phổ biến khác được chấp nhận rộng rãi là RDBMS. Trong tình trạng vi phạm bảo mật dữ liệu lớn đang xảy ra với tốc độ rất cao hiện nay, chúng tôi đặt mục tiêu khai thác các hệ thống cơ sở dữ liệu tạo ra nhiều bản sao trùng lặp của dữ liệu nhạy cảm có thể được tìm thấy trong lưu trữ bảng, nhật ký kiểm tra, dạng xem được vật chất hóa, từ điển dữ liệu, các sản phẩm của SQL Server, v.v. cho phân tích pháp y. Ngoài ra, còn có rất nhiều dữ liệu pháp y nằm xung quanh cơ sở hạ tầng cơ sở dữ liệu để thực hiện điều tra thích hợp và thông tin cần thiết nhất để ghép nối một sự cố sau thực tế. Do đó, bài báo cáo này trình bày một cuộc khảo sát khám phá các niềm tin khác nhau về pháp y cơ sở dữ liệu thông qua các phương pháp khác nhau sử dụng các thuật toán và công cụ pháp y để điều tra. Cuối cùng, chúng tôi chỉ ra những thách thức và cơ hội bằng cách kích thích lĩnh vực pháp y cơ sở dữ liệu, được cho là vẫn đang trong thời kỳ sơ khai.

**Từ khóa:** bảo mật cơ sở dữ liệu, RDBMS, can thiệp cơ sở dữ liệu, nhật ký, pháp y cơ sở dữ liệu

1. **Giới thiệu**

Mục đích của tài liệu này là tập trung vào vi phạm các mối đe dọa bảo mật cơ sở dữ liệu có thể khắc phục thông qua pháp y cơ sở dữ liệu, đây đã trở thành một lĩnh vực nghiên cứu quan trọng. Có một số lượng lớn các rủi ro độc lập đối với dữ liệu bí mật được lưu trữ trong cơ sở dữ liệu và nhiều tổ chức lớn vẫn cực kỳ dễ bị tổn thương do lỗi kiểm tra tuân thủ và vi phạm dữ liệu. Điểm yếu về bảo mật cơ sở dữ liệu này khiến người dùng dễ bị rò rỉ dữ liệu cá nhân hoặc thậm chí tồi tệ hơn là đánh cắp danh tính. Có nhiều rủi ro được tìm thấy đối với bảo mật cơ sở dữ liệu. Những điều này có thể do nhiều lý do chẳng hạn như:

* Ràng buộc ngân sách
* Thiếu hiểu biết về các mối đe dọa
* Không có sự hợp tác giữa các phòng ban
* Ngắt kết nối giữa hoạt động CNTT và nhóm quản lý điều hành
* Không có quy trình và thủ tục bảo mật cơ sở dữ liệu chính thức
* Quá nhiều nhân viên CNTT có quyền truy cập "root" vào cơ sở dữ liệu
* Thiếu hụt các chuyên gia bảo mật có kỹ năng
* Quyết định tập trung vào những nơi khác
* Thiếu kỹ năng bảo mật cơ sở dữ liệu

Rõ ràng là cơ sở dữ liệu của nạn nhân thường chứa thông tin có thể hữu ích trong nhiều cuộc điều tra pháp y. Nhiều tội phạm / kẻ phạm tội đã có thể trốn thoát do thiếu chứng cứ để kết tội họ. Tại đây, pháp y đóng một vai trò quan trọng bằng cách cung cấp các phương pháp được chứng minh khoa học để thu thập, xử lý, giải thích và sử dụng bằng chứng kỹ thuật số để đưa ra mô tả kết luận về các hoạt động tội phạm mạng. Cần có một cách tiếp cận tự động và chính thức cho các cơ sở dữ liệu với mục đích thu thập bằng chứng pháp y.

1. **Can thiệp và Khía cạnh Pháp y của Cơ sở dữ liệu**

Máy chủ cơ sở dữ liệu cung cấp nhiều cơ chế để xác thực và ủy quyền cho người dùng truy cập cơ sở dữ liệu. Nhật ký kiểm tra được coi là thực tiễn tốt và là cách tiếp cận tiêu chuẩn cho bất kỳ ứng dụng nào hỗ trợ Tuân thủ được quy định của chính phủ hoặc cho hệ thống kinh doanh. Chúng cũng được yêu cầu bởi các quy định của liên bang (ví dụ: Đạo luật Sarbanes-Oxley, Đạo luật Bảo hiểm Y tế Di động và Trách nhiệm Giải trình - HIPAA [1]) các hệ thống được bảo mật cho dữ liệu phê duyệt thuốc, tiết lộ thông tin y tế, v.v. Do đó, chúng đã trở thành tệp tin bí mật nhất trong RDBMS. Nhưng điều quan trọng là phải tìm thấy chúng chính xác và không thể thay đổi vì kẻ tấn công có thể đã sửa đổi nhật ký hoặc vượt qua tiện ích ghi nhật ký. Hệ thống cũng không được bảo mật và bảo vệ khi người dùng được ủy quyền can thiệp dữ liệu, sửa đổi hoặc xóa dữ liệu bằng bất kỳ phương pháp nào. Ví dụ: người dùng được ủy quyền có thể truy cập trực tiếp (đối với một số hành vi bất hợp pháp) hoặc người dùng không được ủy quyền có thể truy cập gián tiếp vào máy chủ cơ sở dữ liệu thông qua một số địa chỉ IP được kết nối từ xa và cố gắng thực hiện một số sửa đổi trong máy chủ cơ sở dữ liệu sản xuất (như thay đổi ngày đặt hàng, giá cả có thể được cập nhật) dẫn đến việc vận chuyển sản phẩm và tổn thất tài chính cho công ty. Do đó, các máy chủ cơ sở dữ liệu không đảm bảo dữ liệu thực và do đó làm tăng nhu cầu phân tích pháp y. Việc can thiệp vào cơ sở dữ liệu bởi người dùng được ủy quyền/không được ủy quyền do đó có thể được phát hiện bởi một số thuật toán pháp y như Thuật toán phân tích pháp y hình ảnh xếp tầng [2].

**Phương pháp pháp y** là một trật tự logic và được lập kế hoạch cẩn thận về các hoạt động được thực hiện trong quá trình điều tra kỹ thuật số. Phương pháp pháp y đảm bảo rằng các cuộc điều tra được ghi chép và thực hiện theo cách thức thân thiện với tòa án và dữ liệu thu thập được cần được trình bày làm bằng chứng. Dưới đây là một số khía cạnh Pháp y của Cơ sở dữ liệu:

**A. Để tìm kiếm các khía cạnh pháp y của cơ sở dữ liệu, Martin S. Olivier đã xem xét quan điểm nổi tiếng với ba chiều: lược đồ bên ngoài, khái niệm và bên trong để Kiểm tra pháp y [3].** Theo quan điểm điều tra pháp y, những điều sau cần được xem xét:

* Biết mối quan hệ giữa từ điển dữ liệu và lớp khái niệm. Từ điển dữ liệu có thể là mục tiêu tấn công bằng cách phá hủy hoặc thực hiện bất kỳ thay đổi tinh vi nào trong từ điển dữ liệu.
* Từ điển dữ liệu cũng chứa thông tin có thể có lợi ích pháp y, chẳng hạn như thời gian tạo ra một thực thể - cho dù thực thể đó xuất hiện trên lớp bên ngoài, khái niệm hay bên trong.
* Lược đồ bên ngoài xác định dữ liệu được cung cấp cho người dùng cụ thể.
* Trong quá trình điều tra pháp y, các dạng xem khác nhau cho các người dùng khác nhau được tạo ra bởi các lược đồ khác nhau có thể có liên quan. Số lượng các lược đồ bên ngoài như vậy chỉ phụ thuộc vào cơ sở dữ liệu được xem xét.
* Việc quản lý các tệp được sử dụng cho lớp vật lý của hệ điều hành cũng cần được xem xét.

Do đó, Martin S. Olivier xem xét kiến trúc ANSI/SPARC ban đầu (SIGMOD Record, 1982) đã chỉ định 42 giao diện giữa các thành phần khác nhau để khám phá Pháp y Cơ sở dữ liệu.

**B. Mức độ ghi nhật ký xảy ra trong cơ sở dữ liệu có thể bao gồm đủ thông tin để điều tra.**

**C. Khôi phục hoặc tái tạo dữ liệu đã bị phá hủy (phần nào) hoặc chỉ được phục hồi một phần được thực hiện theo quy trình thu thập pháp y.** Thông thường, cần phải thực hiện kỹ thuật đảo ngược không chỉ lược đồ ứng dụng và dữ liệu khác, mà còn cả cấu trúc DBMS cơ bản của DBMS (được biết đến).

**D. Nhật ký chi tiết hoặc Siêu dữ liệu hoặc kết hợp cả hai có thể được sử dụng để xác định ai được ủy quyền thực hiện một hành động nhất định và sử dụng nó làm cơ sở để xác định các hành vi bất thường.**

1. **Phương pháp học Thám tử Cơ sở dữ liệu (Database Detective Methodology)**

**3.1. Thuật toán Pháp y Cơ sở dữ liệu**

* + Phần này mô tả một phương pháp sử dụng các hàm băm một chiều mạnh về mặt mật mã để ngăn chặn kẻ xâm nhập, bao gồm cả kiểm toán viên, nhân viên hoặc thậm chí lỗi không xác định trong chính DBMS, làm hỏng nhật ký kiểm tra một cách âm thầm.
  + Phương pháp này hoạt động bằng cách băm tích lũy tất cả dữ liệu được thao tác bởi các giao dịch khi chúng có sẵn cho hệ thống. Một mô-đun được gọi là "notarizer" định kỳ thực hiện việc chứng thực bằng cách gửi giá trị băm đó, dưới dạng một tài liệu kỹ thuật số, đến dịch vụ chứng thực kỹ thuật số bên ngoài và lấy ID chứng thực như được hiển thị trong Hình 1.
  + ID chứng thực được trả về cùng với các giá trị băm được tính toán ban đầu được lưu trữ trong một cơ sở dữ liệu riêng biệt nhỏ hơn.
  + Hình 1 minh họa hoạt động bình thường và xác thực nhật ký kiểm tra.

**3.2. Phát hiện Phá hoại Pháp y của Nhật ký Kiểm tra**

* + Phần này trình bày một phương pháp phát hiện và định vị pháp y các dữ liệu bị hỏng trong SQL Server.
  + Giải pháp dựa trên việc tạo một chuỗi các giá trị băm đan xen được sử dụng bởi thuật toán phát hiện để xác định xem một hàng bảng nhật ký kiểm tra cụ thể có bị sửa đổi, chèn hoặc xóa hay không.
  + Giải pháp này có những ưu điểm nhưng cũng có hạn chế về việc sử dụng các hàm băm không mạnh về mặt mật mã và sức mạnh pháp y hạn chế của thuật toán phát hiện.

**3.3. Sản phẩm Cơ sở dữ liệu để Điều tra Cơ sở dữ liệu**

**Đối với Phân tích Pháp y SQL Server**, tác giả Kevvie Fowler trình bày cách thu thập và bảo quản an toàn các sản phẩm cơ sở dữ liệu có liên quan nhất trong quá trình điều tra cơ sở dữ liệu; phân tích chúng để xác nhận hoặc loại trừ sự xâm nhập cơ sở dữ liệu; và theo dõi lại các hành động của kẻ xâm nhập trong một máy chủ cơ sở dữ liệu.

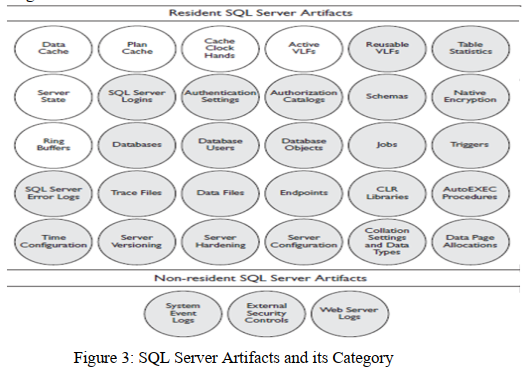
**Các kỹ thuật được mô tả trong Phân tích Pháp y SQL Server có thể được sử dụng để:**

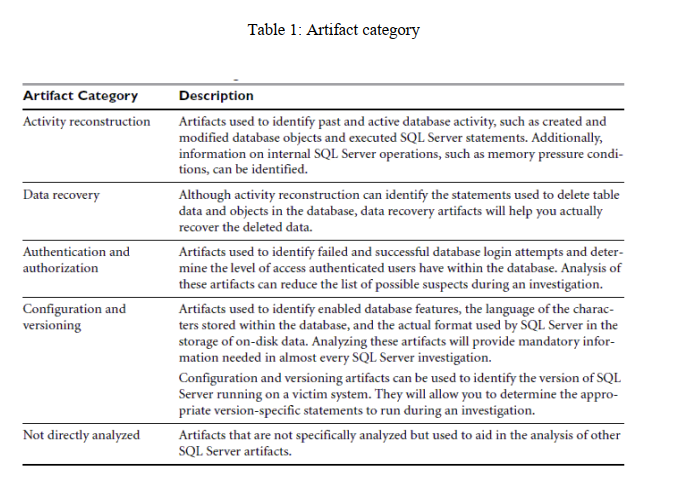
* Xác định truy cập dữ liệu trái phép, sửa đổi
* Thu thập thông tin cần thiết để phục hồi sau sự xâm nhập bằng cách khôi phục trạng thái cơ sở dữ liệu trước khi xảy ra sự cố.

**Nhìn chung, các sản phẩm SQL Server có thể được phân loại thành một trong hai loại:**

* **Sản phẩm Cư trú:** Nằm trong các tệp và vị trí bộ nhớ được dành riêng cho mục đích sử dụng SQL Server, chẳng hạn như nhật ký lỗi SQL Server.
* **Sản phẩm Không cư trú:** Nằm trong các tệp không được dành riêng cho mục đích sử dụng SQL Server.

Hình 3 dưới đây cho thấy những **sản phẩm chính** này. Các sản phẩm **dễ bay mất** khi dịch vụ **MSSQL Server** tắt được hiển thị bằng màu trắng, trong khi các sản phẩm **bền vững** vẫn tồn tại sau khi dịch vụ **MSSQL Server** khởi động lại được đánh màu xám. Mỗi sản phẩm được minh họa trong Hình 3 dưới đây liên quan đến **một trong năm danh mục**, sắp xếp các sản phẩm dựa trên cách chúng có thể hỗ trợ trong cuộc điều tra. Có nhiều sản phẩm của **SQL Server**, mỗi sản phẩm sẽ hỗ trợ cuộc điều tra SQL Server theo cách khác nhau. [Bảng I dưới đây liệt kê các danh mục của sản phẩm SQL Server cùng với tên của mỗi danh mục đại diện cho mục tiêu chính của sản phẩm trong cuộc điều tra1](https://translate.google.com.vn/?hl=en)[2](https://translate.google.com.sg/?hl=vi&tab=wT).

****



Tương tự, có một công cụ LogMiner [9] cho phép quản trị viên cơ sở dữ liệu Oracle (DBA) và/hoặc nhà phân tích pháp y tái tạo các hành động được thực hiện trên cơ sở dữ liệu Oracle ngay cả khi các tính năng kiểm toán đã bị tắt. LogMiner là một tiện ích có thể được sử dụng để phân tích các tệp nhật ký redo được tạo bởi cơ sở dữ liệu Oracle.

1. **Thách thức và Cơ hội Nghiên cứu trong Pháp y Cơ sở dữ liệu**

Pháp y Cơ sở dữ liệu là một chủ đề quan trọng nhưng chưa nhận được nhiều sự chú ý nghiên cứu. Việc thiếu nghiên cứu này có thể là do tính phức tạp vốn có của cơ sở dữ liệu và chưa được hiểu trong bối cảnh pháp y. Chuyên gia bảo mật cơ sở dữ liệu David Litchfield gần đây cho biết Pháp y Cơ sở dữ liệu vẫn đang trong thời kỳ đen tối [BLACK HAT USA 2011 -- Las Vegas] [10].

Ngay cả trong thời đại vi phạm dữ liệu và tấn công cơ sở dữ liệu hàng loạt này, lĩnh vực pháp y cơ sở dữ liệu vẫn tụt hậu đáng kể. Không có công cụ nào được bán trên thị trường để thực hiện pháp y cơ sở dữ liệu hiệu quả. Dần dần phiên bản beta của các công cụ pháp y cơ sở dữ liệu đang được phát triển để bảo vệ cơ sở dữ liệu khỏi bị tấn công, lấy quyền truy cập và sau đó sửa đổi dữ liệu. Kẻ tấn công có thể để lại bằng chứng có thể được thu thập theo một số cách nhất định bởi các công cụ pháp y cho mục đích điều tra thêm. Cuối cùng, tòa án mong đợi các bằng chứng để tiết lộ sự thật.

Theo Litchfield [11], có rất nhiều dữ liệu pháp y nằm xung quanh cơ sở hạ tầng cơ sở dữ liệu để thực hiện một cuộc điều tra thích hợp và ghép nối một sự cố sau thực tế. Một số nơi mà các nhóm phản ứng sự cố nên xem xét bao gồm siêu dữ liệu hệ thống, tệp dữ liệu, nhật ký redo, nhật ký giao dịch, phân khúc hoàn tác và bộ nhớ và tệp theo dõi. Tệp nhật ký cũng tốt, nhưng chỉ với sự thận trọng vì kẻ tấn công có thể thao túng chúng.

Dragoon (Database foRensic Analysis safeGuard Of arizONa) của Kyriacos E. Pavlou và R. T. Snodgrass là một hệ thống kiểm toán nguyên mẫu để phát hiện giả mạo và phân tích pháp y [12]. Dragoon là kết quả của các thuật toán pháp y khác nhau được đề cập trong phần 4.1. Nguyên mẫu này có thể được mở rộng thành một giải pháp trách nhiệm thông tin toàn doanh nghiệp, có thể thực hiện hiệu quả việc sử dụng phù hợp, nghĩa là đảm bảo không có bất kỳ sửa đổi trái phép nào như chèn, xóa, cập nhật ngay cả bởi những người dùng nội bộ trong các cơ sở dữ liệu hiệu suất cao. Hơn nữa, kiến trúc Dragoon hiện tại có thể được mở rộng để hỗ trợ các cơ sở dữ liệu được triển khai trên đám mây cũng như cung cấp khả năng kiểm tra cho nhật ký truy cập Apache và log4j, đây lại là một thách thức trong Pháp y Cơ sở dữ liệu.

1. **Kết luận**

Pháp y Cơ sở dữ liệu là một lĩnh vực rất mới mẻ với tài liệu tham khảo hạn chế và ít công cụ. Bài báo này tiếp cận nhiệm vụ của mình bằng cách xác định các khía cạnh khác nhau của Pháp y Cơ sở dữ liệu. Bài báo cũng thảo luận về nhiều phương pháp khác nhau để phát hiện sự giả mạo. Những thách thức chính được nêu ra dựa trên khảo sát, mở ra những cơ hội mới cho nghiên cứu và giảng dạy. Tóm lại, bài báo này nhằm mục đích thu hút sự chú ý đến Pháp y Cơ sở dữ liệu với hy vọng thúc đẩy nghiên cứu trong lĩnh vực quan trọng này.