ANTIVIRUS

# 1. Antivirus là gì?

- Antivirus là một phần mềm diệt virus, có tính năng phát hiện loại bỏ các virus máy tính, khắc phục hậu quả của virus gây ra và có khả năng được nâng cấp để nhận biết các loại virus trong tương lai.

# 2. Một số loại antivirus hiện nay.

Hiện nay có một số phần mềm duyệt virus thông dụng và được nhiều người trên thế giới sử dụng như: Kaspersky Anti-Virus, McAfee Norton Antivirus. Ngoài ra còn có những phần mềm duyệt virus miễn phi nhưng không kém hiệu quả như: AVG Anti-Virus, BKAV, Avast, CMC…

# 3. Cơ chế phát hiện virus.

**Nguyên Lý Phát Hiện Mã Độc**

Malware Detection

Signature-Based

Anomaly-Based

Static

Dynamic

Hybrid

Kỹ thuật phát hiện mã độc được chia làm 2 nhóm chính :

* Signature-bases : dựa vào dấu hiệu đặc trưng.
* Anomaly-based : dựa vào các điểm bất thường.

Các kỹ thuật phân tích được sử dụng trong các kỹ thuật phát hiện mã độc là :

* Static : phân tích tĩnh.
* Dynamic : phân tích động.
* Hybrid : phân tích lai, là sự kết hợp giữa phân tích động và tĩnh.

1. Nguyên Lý Phát Hiện Mã Độc
2. Phát hiện mã độc dựa vào dấu hiệu đặc trưng

* Là việc sử dụng một tập các mẫu nhận dạng được gọi là signature để làm căn cứ xác định mã độc. Tập các signature sẽ được xây dựng bằng việc cập nhật các mẫu mã độc đã được kiểm chứng, bằng việc sử dụng những mẫu được xây dựng chuyên biệt bởi các nhà nghiên cứu.
* Ưu thế :
  + Phát hiện chính xác các mã độc nếu như các signature trùng khớp với các mẫu trong lập mẫu nhận dạng.
* Hạn chế :
  + Xây dựng tập các signature hoàn chỉnh là vô cùng khó khăn.
  + Tập mẫu nhận dạng sẽ chạy theo sau sự phát triển của mã độc, và sẽ không phát hiện được các mã độc mới, các zero-days.
  + Khi số lượng signature lớn thì việc lưu trữ và đối chiếu cũng sẽ gặp khó khăn.

|  |  |
| --- | --- |
| Kỹ thuật | Nội dung |
| Dynamic | Xác định chính xác mã độc bằng việc chỉ sử dụng các thông tin thu thập được từ các chương trình có quyền kiểm tra (Program Under Inspection - PUI) và được đối chiếu với các tập mẫu. |
| Static | Bằng cách kiểm tra chương trình dưới các mã, các đoạn code hoặc kiểm tra hành vi bằng các chuỗi mã, Từ đó đối chiếu với tập mẫu để phân loại chương trình có chứa mã độc hay có phải là mã độc hay không.  Điều này giúp việc xác định mã độc có xác suất rất cao, gần như là tuyệt đối mà không cần thực thi chương trình. Bù lại, cách thức này mất nhiều công sức và thời gian. |
| Hybrid | Các thức xác định sử dụng cả hai hình thức Dynamic và Static. |

1. Phát hiện mã độc dựa vào các điểm bất thường

* Cách phát hiện dựa vào các điểm bất thường chia thành 2 giai đoạn :
  + Giai đoạn Training 🡪 Learning : các detector sẽ cố gắng đọc những trạng thái bình thường. Có thể học các trạng thái từ các host, PUI,...
  + Giai đoạn Detection 🡪 Monitoring : dựa vào các trạng thái bình thường đã được học, các detector sẽ xác định được trạng thái bất thường và đưa ra cảnh báo.
* Ưu thế :
  + Là chìa khóa để có thể phát hiện ra các khai thác zero-days hay zero-attack.
* Hạn chế :
  + Sai số giữa trạng thái bình thường và bất thường.
  + Sự phức tạp, rắc rối khi xác định những trạng thái được phép học.

|  |  |
| --- | --- |
| Kỹ thuật | Nội dung |
| Dynamic | Các thông tin thu được từ việc thực thi chương trình sẽ được sử dụng để phát hiện mã độc.  Ở giai đoạn Detection, các detector thực hiện liên tục việc đối chiếu giữa các trạng thái đã học được, Và giai đoạn Detection được triển khai trong suốt quá trình chương trình được thực thi. |
| Static | Các đặc trưng về cấu trúc tập tin của chương trình sẽ được dừng để kiểm tra và xác định có phải là mã độc hay không.  Đặc điểm nổi bật của cách này là có thể phát hiện mã độc nhưng không cho chương trình mã độc thực thi. |
| Hybrid | Sử dụng một môi trường giả lập có cấu trúc giống với máy thật, có thể gọi là máy ảo để sử dụng phối hợp 2 kỹ thuật phân tích trên mà không làm ảnh hưởng đến máy thật. |

1. Các Kỹ Thuật Phát Hiện Mã Độc
2. Checksummers

* Là kỹ thuật phát hiện đối tượng mã độc trên cơ sở tính toàn vẹn của chương trình. Mỗi chương trình sau khi được xác định chắc chắn là mã độc sẽ được sử dụng để tạo ra mẫu mã hash (MD5, SHA, CRC,...) tương ứng với loại mã độc đó. Do tính chất của mã hash nên mỗi loại mã độc chỉ tương ứng với một mã hash nhận dạng (xác suất trùng lặp là rất thấp). Sau đó, cập nhật mẫu đó vào cơ sở dữ liệu của chương trình Antivirus. Khi Antivirus quét tập tin để kiểm tra, nếu trùng mã hash trong cơ sở dữ liệu mẫu, thì đó là mã độc. Ngược lại là không.
* Ưu điểm :
  + Độ chính xác gần như tuyệt đối.
* Nhược điểm :
  + Nhận dạng thiếu. Quá trình nhận dạng lâu nếu cơ sở dữ liệu mẫu lớn.
  + Quá trình xây dụng cơ sở dữ liệu mẫu khó khăn, phức tạp. Mã độc có thể dễ dàng vượt qua nếu nhận dạng bằng mã hash.

1. Fuzzy Hashing

* Để cải tiến nhược điểm dựa vào mã hash của kỹ thuật Checksummer, các nhà khoa học đã tìm ra một kỹ thuật khác là Fuzzy Hashing. Kỹ thuật mới này về cơ bản vẫn là nhận dạng mã độc qua mã hash nhưng đã được bổ sung thêm các phân tích và tính toán để từ một mã hash của mã độc, có thể nhận ra các mã hash “họ hàng”, nâng cao khả năng phát hiện chương trình mã độc.
* Ưu điểm :
  + Cải thiện tính bị động của kỹ thuật Checksummer sử dụng mã hash đơn thuần.
  + Nâng cao khả năng phát hiện mã độc với cơ sở dữ liệu mẫu bị hạn chế.
* Nhược điểm :
  + Việc xây dựng thuật toán và lựa chọn độ dài ký tự phù hợp là khó khăn.
  + Có thể xảy ra cảnh báo (cảnh báo sai).

1. Scan String

* Scan string là cách đơn giản để phát hiện mã độc. Kỹ thuật này sử dụng một chuỗi trích ngang (chuỗi bytes) là đặc trưng của tập tin mã độc và không tồn tại trong các tập tin sạch. Các chuỗi này sẽ được cập nhật vào cơ sở dữ liệu mẫu dùng để nhận dạng mã độc.
* Ưu điểm :
  + Nhận dạng chính xác.
  + Tốc độ nhận dạng nhanh hơn so với kỹ thuật Checksummer.
* Nhược điểm :
  + Quá trình xây dựng và cập nhật cơ sở dữ liệu phức tạp.
  + Nhận dạng bị động, không phát hiện được khi mã chương trình bị thay đổi.

1. Code Emulation

* Là kỹ thuật phát hiện mã độc được đánh giá cao. Bằng việc mô phỏng lại hệ thống CPU, hệ thống quản lý bộ nhớ, các chỉ thị máy ở cấp thấp... giống như máy quét thực tế. Vì vậy, mã độc sẽ hoạt động trên máy ảo mà không ảnh hưởng đến bộ xử lý thật.
* Hiện nay, kỹ thuật này được sử dụng kết hợp rộng rãi với các kỹ thuật phân tích khác. Cùng với đó, kỹ thuật này đã trở thành một thành phần quan trọng trong công nghệ Sandbox để phát hiện mã độc.
* Ưu điểm :
  + Mã độc hoạt động độc lập, không ảnh hưởng đến hệ thống máy thật.
* Nhược điểm :
  + Mô phỏng lại các thông tin hệ thống CPU, bộ nhớ,... là rất khó khăn.
  + Cập nhật và vận hành hệ thống mô phỏng yêu cầu tính kỹ thuật cao.

1. Static Heuristic Analysis

* Là kỹ thuật thuộc nhóm Anomaly-Based detection. Kỹ thuật này phân tích, tính toán những thông tin từ chương trình như : PE Header, Section, các hàm API được Import... hay những dấu hiệu đáng nghi ngờ như : kích thước của một số trường không chính xác, mã điều hướng, tổ hợp các Flag,...
* Một nỗi bật của kỹ thuật này là ứng dụng trong việc sử dụng trí tuệ nhân tạo trong việc phát hiện các loại mã độc. Việc thu thập các đặc trưng của mã độc (feature), sau đó tính toán và đưa ra các ngưỡng lây nhiễm (threshold infected) để làm căn cứ phát hiện mã độc.
* Ưu điểm :
  + Chủ động trong việc xây dựng cơ sở dữ liệu mẫu.
  + Nâng cao khả năng nhận dạng mã độc (bao gồm cả những loại mã độc chưa có trong cơ sở dữ liệu mẫu)
* Nhược điểm :
  + Khó khăn trong việc lựa chọn các đặc trưng để sử dụng cho việc phân loại ngưỡng lây nhiễm.
  + Có thể xảy ra cảnh báo giả (cảnh báo sai).

1. Behavior Blocking

* Là kỹ thuật cho phép ngăn chặn các hành vi, các khối lệnh bị nghi ngờ là mã độc trước khi chúng có cơ hội ảnh hưởng đến hệ thống. Các hành vi bị giám sát bao gồm :
  + Mở, xem, sửa đổi hoặc xóa bỏ các tệp tin.
  + Fomat ổ đĩa hay vùng hoạt động không thể khôi phục khác.
  + Sửa đổi các thiết lập của tệp tin thực thi và hệ thống như : thiết lập khởi động,...
* Hệ thống phát hiện và ngăn chặn mã độc sử dụng kỹ thuật này được chia thành 2 loại là : hệ thống ngăn chặn dựa trên các chính sách và hệ thống ngăn chặn dựa trên các giám định.
* Ưu điểm :
  + Có khả năng ngăn chặn sự ảnh hưởng của chương trình mã độc lên hệ thống.
  + Tùy theo năng lực phân tích để tùy chỉnh ảnh hưởng của mã độc.
* Nhược điểm :
  + Yêu cầu người (hệ thống) phân tích phải có tính chuyên môn cao.
  + Có thể xảy ra cảnh báo (cảnh báo sai).