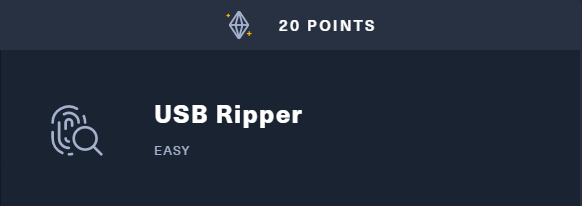
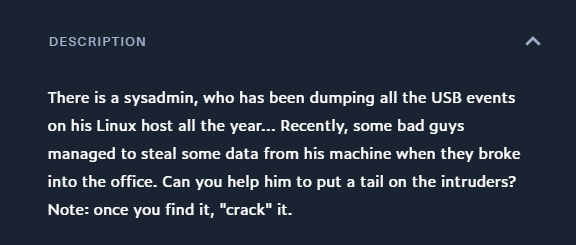
**Hack The Box Writeup — USBRipper**



Trong challenge này, mình muốn nói với các bạn về sự quan trọng của việc coding. Sẽ dễ dàng hơn nếu các bạn có thể tạo ra thuật toán và chạy nó như một chương trình, nó sẽ tiết kiệm cho bạn thời gian hơn là bạn làm thủ công. Nào ta sẽ cùng làm challenge này.

**. . .**

**Challenge Introduction**



Có một sysadmin hằng nằm sẽ dump tất cả các sự kiện USB trên máy chủ Linux của anh ấy… Gần đây, có vài kẻ xấu đã đột nhập vào văn phòng và đánh cắp vài dữ liệu từ máy của anh ta. Bạn có thể giúp anh ấy tìm ra dấu vết của những kẻ xâm nhập trái phép này không?

Note: Khi nào bạn tìm ra nó, hãy “crack” nó.

Source: <https://app.hackthebox.eu/challenges/81> trong mục Download File

Sau khi download, bạn sẽ nhận được 2 files gồm auth.json và syslog.

**Solution**

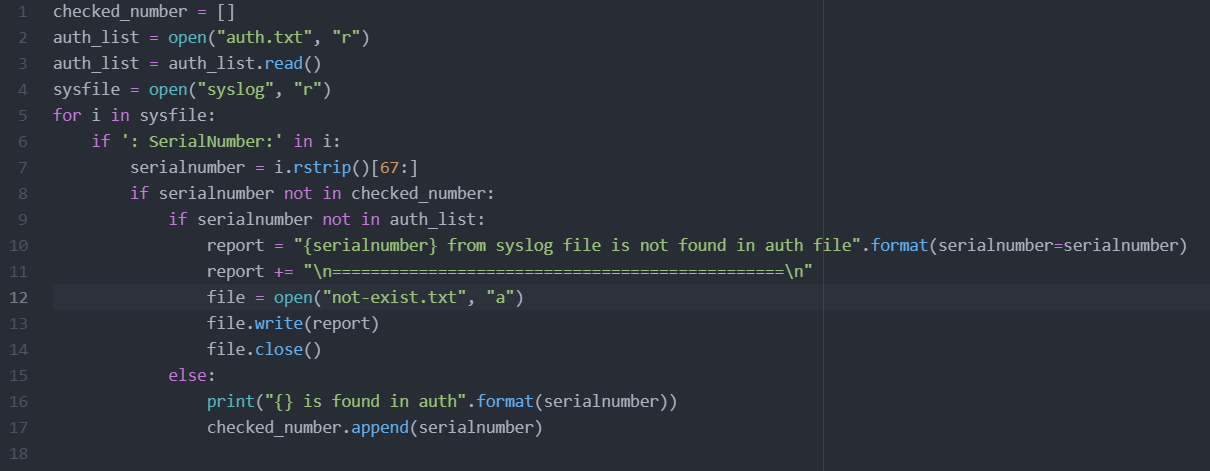
Chúng ta có thể đoán rằng syslog là những sự kiện USB đã được dump từ sysadmin. Có tận 900.000 dòng log trong file đó. auth.json giống như những số được đánh dấu (tên gọi) của mỗi sản phẩm, có thể là những con số được xác thực. Dựa trên phân tích đó, tôi nghĩ ra một giả thuyết là ta sẽ check tất cả các con số trong syslog xem có số nào không tồn tại trong auth.json không.

Nhìn kỹ trong auth.json, ta có thể thấy các số manufact này chỉ chứa ký tự 1-9 và A-F. Các bạn có thể dùng text editor để search ký tự “G” thì nó không xuất hiện một lần nào. Tôi còn có một giả định khác rằng là trong syslog gồm *Product*, *Manufacture*, *SerialNumber*. Như chúng ta đã biết thì ta có thể sử dụng cùng sản phẩm (product) hoặc khác sản phẩm (product) nhưng vẫn có thể cùng một nhà sản xuất (Manufacture) cho nên chỉ số của 2 dữ liệu này có thể được lặp lại. Do đó chỉ có *SerialNumber* là con số duy nhất, không bị lặp lại.

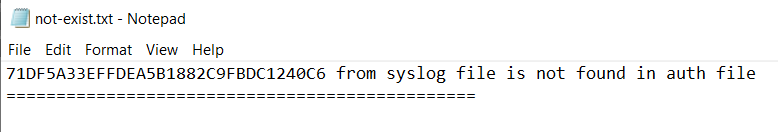
Bây giờ chúng ta đã hiểu vấn đề. Tôi sẽ tóm tắt lại những việc ta cần làm:

1. Đầu tiên, tạo một chương trình mở file syslog.
2. Lặp lại từng dòng check xem SerialNumber trong syslog. Nếu tồn tại thì in hex number đó tồn tại, còn không thì sẽ in hex number đó ra một file. (Sử dụng rstrip() để lấy mỗi hex number đó)
3. Tiếp tục kiểm tra cho đến hết.

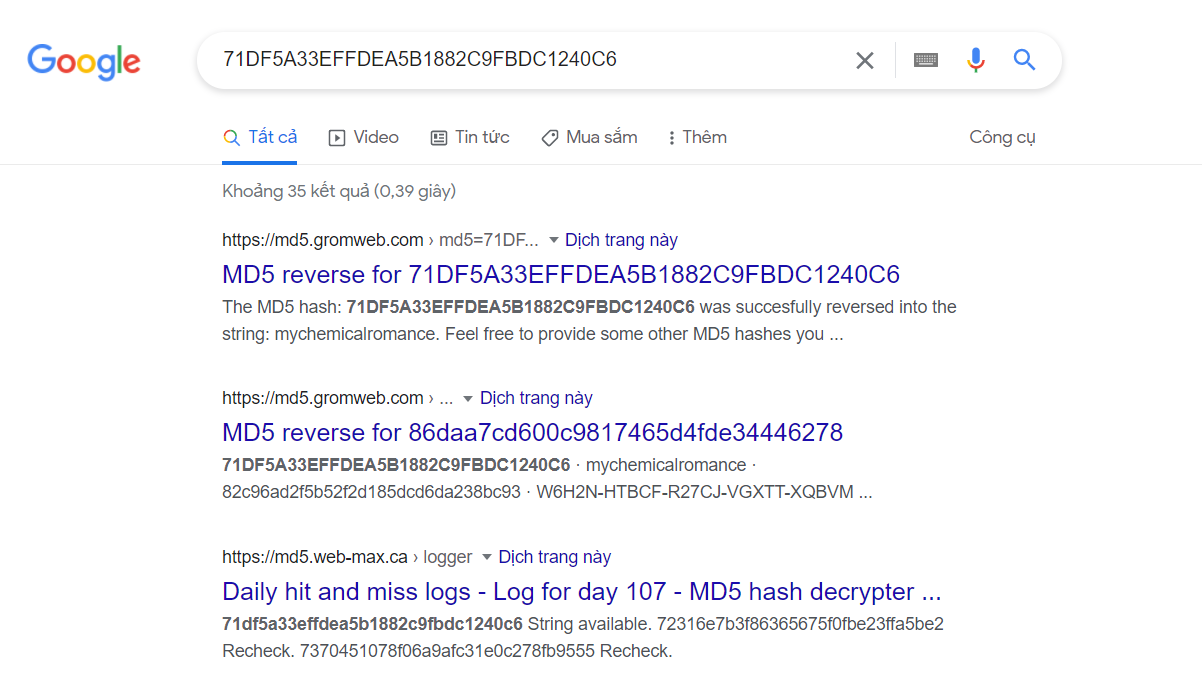
Và đây là code của tôi, tôi có đổi lại extension của auth.json thành .txt để có thể đọc được file.



Sau khi thực thi program đó, tôi nhận được kết quả sau:



Vậy SerialNumber này không được authorized, nhưng đó không phải là điều ta tìm vì đây là challenge nên ta cần tìm flag. Sau khi nhận được SerialNumber này, tôi đã nghĩ đến “MD5”. MD5 có 32 ký tự trong bộ ký tự hex nhưng khi thử thì không tìm được gì. Vì vậy tôi đã google nó và kết quả hiện ra như sau:



Truy cập vào website đầu tiên ta sẽ nhận được flag cần tìm.

**Mistakes**

1. Ban đầu tôi đã làm ngược lại với solution tôi trình bày phía trên. Tôi check SerialNumber của auth.json với syslog, sẽ có một vài SerialNumber không tồn tại trong syslog bởi vì không phải lúc nào các thiết bị cũng được sử dụng trên máy đó.

**Conclusion**

Trong challenge này, ta đã học được rằng việc code rất là quan trọng và hữu ích. Ta có thể dễ dàng tìm kiếm thông tin ta cần trong 900.000 dòng log chỉ trong một thời gian ngắn.