

**Bài thực hành:** Giấu tin trong văn bản bằng phương pháp sửa đổi văn bản sao cho ý nghĩa của văn bản vẫn được bảo toàn

1. Mục đích

- Giúp sinh viên hiểu về phương pháp giấu tin trong văn bản bằng phương pháp sửa đổi văn bản sao cho ý nghĩa của văn bản vẫn được bảo toàn

2. Yêu cầu đối với sinh viên

- Sinh viên có kiến thức xử lý chuỗi, file trong Python

3. Nội dung thực hành

- Tải bài lab: Sinh viên tải bài lab từ link github sau:

<https://github.com/Mbr-clown-lord/stego-syntax-modification.git>

- Khởi động bài lab:

*labtainer -r stego-syntax-modification*

(chú ý: sinh viên sử dụng mã sinh viên của mình để nhập thông tin email người thực hiện bài lab khi có yêu cầu, để sử dụng khi chấm điểm)

Ý tưởng của việc sửa đổi này an toàn, nhưng lại khó thực thi vì khả năng “hiểu” của các máy tính với ngôn ngữ của con người là hạn chế. Việc chỉnh sửa văn bản mà vẫn giữ được ý nghĩa của văn bản là một thủ thuật tinh vi để ẩn giấu thông tin. Nghĩa của hai cụm từ “Khi con học xong, con có thể đi chơi” và “con có thể đi chơi khi con học xong.” là tương đương 130 và thông tin có thể được ẩn giấu dựa vào cú pháp thay đổi này. Phương pháp này giấu được ít thông tin và có hiệu suất thấp vì phải có sự điều chỉnh thủ công. Tuy nhiên phương pháp này an toàn vì rất khó để máy tính có thể phát hiện ra sự thay đổi cú pháp giữa hai cụm từ. Kẻ tấn công phải thực sự đọc các tin nhắn một cách thủ công và xác định những cụm từ có liên quan để trích xuất dữ liệu

Trong bài lab này chúng ta sẽ sử dụng văn bản tiếng Việt làm vật chứa tin. Quy ước:

|    |                               |
|----|-------------------------------|
| 00 | Câu chủ động + trạng ngữ cuối |
| 01 | Câu chủ động + trạng ngữ đầu  |
| 10 | Câu bị động + trạng ngữ cuối  |
| 11 | Câu bị động + trạng ngữ đầu   |

*Bảng 1: Quy ước giấu tin*

Ví dụ:

Câu “Trong tháng 10/2023, đội ngũ kỹ sư đã hoàn thành giai đoạn đầu của dự án” có thể giấu được 2 bit “01”.

Giải thích:

- Câu chủ động
- Trạng ngữ chỉ thời gian “Trong tháng 10/2023” nằm ở đầu câu.

Sau khi khởi động xong một terminal ảo có tên stego-syntax-modification xuất hiện.

- Khám phá:

Xem nội dung tệp plaintext.txt:

*cat plaintext.txt*

3.1. Nhiệm vụ 1: Tìm các câu có thể giấu tin

- Yêu cầu: Đọc và phân tích nội dung file plaintext.txt sau đó chọn ra các câu có thể dùng để giấu tin thỏa mãn bảng 1: Quy ước giấu tin.
- Lưu số thứ tự của các câu(viết liền) vào file result.txt  
Ví dụ: “123”

3.2. Nhiệm vụ 2: Giấu thông điệp

- Yêu cầu: Chuyển mã ASCII của số cuối trong mã sinh viên sang dạng nhị phân sau đó chia ra thành từng cặp 2 bit. File hint.txt chứa các gợi ý giấu tin. Đọc nội dung file hint.txt sau đó tiến hành chọn các câu phù hợp để chỉnh sửa nội dung file plaintext.txt và giấu được mã nhị phân của số cuối cùng trong mã sinh viên. Toàn bộ nội dung chỉnh sửa cần lưu vào file input.txt. Sau khi chỉnh sửa xong tiến hành chạy file check\_stego.py và xem kết quả.

*python3 check\_stego.p*

3.3. Nhiệm vụ 3: Tách thông điệp

- Yêu cầu: File stego\_text.txt là file đã được chỉnh sửa vào mã hóa theo quy ước từ bảng 1: Quy ước giấu tin. Đọc file và trích xuất mã nhị phân được giấu trong file sau đó lưu kết quả vào file binary.txt. Cuối cùng chạy file check\_decode.py để xem kết quả.

*python3 check\_decode.py*

- Kết thúc bài lab:

o Trên terminal đầu tiên sử dụng câu lệnh sau để kết thúc bài lab:

*stoplab stego-syntax-modification*

o Khi bài lab kết thúc, một tệp zip lưu kết quả được tạo và lưu vào một vị trí được hiển thị bên dưới stoplab.

- Khởi động lại bài lab:

o Trong quá trình làm bài sinh viên cần thực hiện lại bài lab, dùng câu lệnh:

*labtainer -r stego-syntax-modification*