

CHƯƠNG 1. GIỚI THIỆU CHUNG VỀ BÀI THỰC HÀNH

1. Mục đích:

Giúp sinh viên hiểu và tự thực hiện được việc giấu/tách một thông điệp bí mật vào một đoạn văn bản sử dụng phương pháp mã hóa đỉnh (Synonym Substitution using Vertex Coding Method)

2. Yêu cầu đối với sinh viên

Sinh viên cần đảm bảo các yêu cầu sau khi làm bài thực hành:

- Hiểu rõ khái niệm steganography và vai trò của giấu tin trong văn bản;
- Nắm được nguyên lý hoạt động của phương pháp sử dụng Vertex Coding Method

3. Giới thiệu:

Steganography là kỹ thuật giấu thông tin bí mật trong một thông điệp công khai để không ai nghi ngờ sự tồn tại của thông tin bí mật. Trong ngôn ngữ tự nhiên, một cách tiếp cận phổ biến là thay thế các từ trong văn bản bằng từ đồng nghĩa sao cho người đọc không nhận thấy sự khác biệt, nhưng người nhận có thể trích xuất thông điệp ẩn.

4. Thay thế từ đồng nghĩa (Synonym Substitution)

Thay thế từ vựng bằng từ đồng nghĩa giữ nguyên ngữ nghĩa, nhưng có thể được sử dụng như một kênh truyền tin. Ví dụ:

Văn bản gốc: "He will marry her."

Văn bản stego: "He will espouse her."

Nếu mỗi từ đồng nghĩa được ánh xạ tới một mã nhị phân cụ thể, thì chuỗi các từ thay thế sẽ mang theo một thông điệp bí mật.

5. Vấn đề từ đa nghĩa (Polysemy)

Từ đa nghĩa là từ có nhiều nghĩa hoặc xuất hiện trong nhiều nhóm đồng nghĩa (synsets). Điều này gây khó khăn trong việc xác định mã nhị phân cho từ đó. Ví dụ, từ "marry" xuất hiện ở cả hai synset:

Synset 1: {espouse, conjoin, marry}

Synset 2: {tie, splice, wed, marry}

Vì vậy, cần có một cách để gán mã mà vẫn đảm bảo tính nhất quán và có thể giải mã được.

6. Vertex Coding Method: Mã hóa đỉnh

Đây là một phương pháp lấy cảm hứng từ lý thuyết đồ thị (graph theory), cụ thể là bài toán tô màu đỉnh:

Mỗi từ (đỉnh) trong đồ thị đồng nghĩa được gán một mã 2-bit (như 00, 01, 10, 11).

Các từ trong cùng một synset được nối với nhau tạo thành đồ thị từ đồng nghĩa.