**Password cracking**

**I/ Giới thiệu thuật toán mã hóa Subset-sum**

Thuật toán subset-sum là một thuật toán mã hõa dựa trên chuỗi nhị phân của chuỗi được mã hóa. Để thực hiện thuật toán này thì phải có một bảng lưu trữ các từ khóa. Thuật toán hoạt động như sau. Giả sử chuỗi ban đầu là s. Chúng ta sẽ phân tích s ra thành một chuỗi dạng nhị phân, sao đó tổng cộng các dòng i sao cho bit thứ i trong chuỗi bằng 1. Sau đó mod cho 2 mũ số dòng của bảng. Sau khi có được kết quả ở các bước trên, chúng ta sẽ mã hóa chuỗi nhị phân nhận được thành số và đó chính là chuỗi đã được mã hóa. Để hiểu rõ về thuật toán này, mời các bạn đọc ví dụ sau: <https://docs.google.com/document/d/1xM_AA_b6Tf7C3V9yaluCJmjwDKmQEAHd/edit?fbclid=IwAR0geQz6Ei9hDq7zGzEjWM5_F__q80C2Lbo2AY5EOIioANmCvUYbyAbnk0w>

Thuật toán trên tồn tại một số ưu điểm và nhược điểm nhất định.

Ưu điểm: quá trình mã hóa diễn ra rất nhanh chóng vì chỉ cần phân tích chuỗi ban đầu thành nhị phân sau đó tính tổng là sẽ mã hóa thành công. Nội dung của bảng cũng sẽ rất ít, tiện lợi cho việc lưu trữ, độ bảo mật cao, không bao giờ có thể giải mã được nếu như không biết được bảng ban đầu.

Nhược điểm: Qui trình giải mã rất phức tạp, phải dùng thuật toán back-tracking có độ phức tạp cao. Tồn tại trường hợp bị mã hóa trùng, tức là 2 chuỗi a và b khác giống nhau nhưng lại có giá trị sau khi mã hóa giống nhau.

**II/ Các giải mã thuật toán Subset-sum**

Ở phần này mình sẽ trình bày giải thuật về giải mã cho thuật toán Subset-sum.

Gọi chuỗi cần mã hóa là S => để mã hóa thì cần phải biết dạng nhị phân của chuỗi S.

Đầu tiên là duyệt hết tất cả các tập con của bảng (tạm gọi là key-table). Sau khi lấy được một tập con của key-table thì tính tổng tập con lại và Mod cho 2 mũ số dòng của bảng. Nếu kết quả nhận được bằng với chuỗi nhị phân của S. Tập con đó chính là kết quả cần tìm. Tuy nhiên đến đây thì chỉ nhận được tập con. Việc còn lại là phải phân tích tập con vừa tìm được thành một chuỗi nhị phân. Với mỗi phần tử được chọn trong tập con, giải sử phần tử thứ i được chọn trong tập con, ta bật bit thứ i của chuỗi kết quả thành 1. Ví dụ tập con ta nhận được là các dòng {1, 3 ,5 ,7} thì chuỗi nhị phân sẽ là 01010101. Sau khi có được chuỗi nhị phân thì việc còn lại là phân tích ra chuỗi ban đầu là khá dễ dàng.

Đây là code mẫu dành cho các bạn tham khảo: <link code>

III/ Giải quyết vấn đề về bảng key-table

Như đã đề cập ở Mục I, vấn đề về việc trùng mật khẩu là hoàn toàn tồn tại. Ví dụ xét bảng T giả định như sau:

|  |  |  |
| --- | --- | --- |
| 0 | aa | 0000 |
| 1 | ab | 0001 |
| 2 | ba | 0100 |
| 3 | bb | 0101 |

Giả sử xét bảng T trên, lưu ý bảng trên chỉ mang tính chất minh họa.

Giải sử xét 2 chuỗi cb ( 1110), ab (0110). 2 chuỗi này sau khi mã hóa theo thuật toán đều sẽ thu được chuỗi bb do cb <=> 0000 + 0001 + 0100 = 1010, ab <=> 0001 + 0100.

Nhận xét trường hợp trên xảy ra là do bảng key-table.

Vậy câu hỏi đặt ra là chọn key-table sao cho hợp lí. Câu trả lời rất đơn giản, xét dòng thứ i của bảng key-table, chỉ bật 1 bit j của dòng thứ i và bit j này phải đảm bao rằng không tồn tại dòng nào trong bảng khác dòng i mà cũng bật bit thứ j. Nói cách khác là mỗi dòng chỉ bật một bit và bit được bật phải phân biệt trên mỗi dòng. Ví dụ:

|  |  |  |
| --- | --- | --- |
| 0 | ab | 0001 |
| 1 | ac | 0010 |
| 2 | ba | 0100 |
| 3 | ca | 1000 |

Như vậy sẽ không tìm được 2 tập con khác nhau mà có tổng giống nhau và việc mã hóa trùng không thể tồn tại. Để bảng trở nên khó đoán hơn thì nên hoán đổi các cột trong bảng.