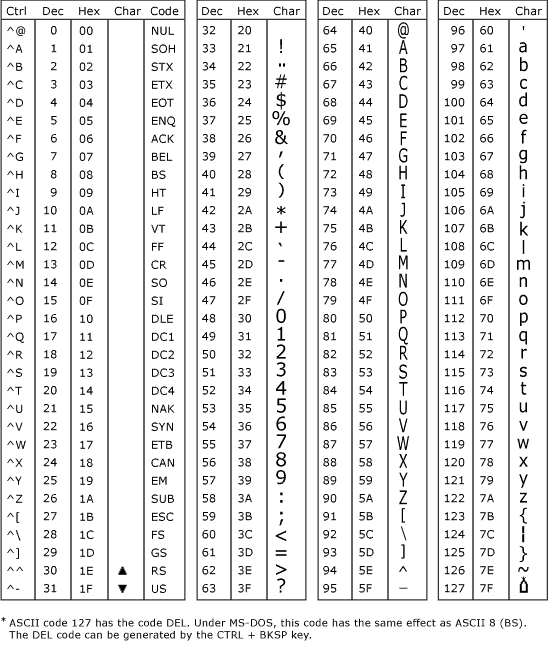
**Bảng mã ASCII:**



**Url encoding:**

Urls chỉ được phép chứa những kí tự có thể in được trong bảng mã ASCII từ 0x20 đến 0x7e. Hơn nữa, một số kí tự trong khoảng này vẫn bị hạn chế vì nó có ý nghĩa đặc biệt trong giao thức http.

Cơ chế mã hóa Url được dùng để mã hóa những kí tự có vấn đề để nó được vận chuyển an toàn qua http.

Đặc điểm: có tiền tố % theo sau là mã ascii hai chữ số của kí tự theo hệ thập lục phân.

%3d - =

%25 - %

%20 – Space

%0a – New line

%00 – Null byte

Lưu ý: dấu + trong url sẽ hiểu là khoảng trắng, tương tự %20.

**Unicode encoding:**

- Unicode là 1 tiêu chuẩn mã hóa được thiết kế để hỗ trợ các hệ thống viết trên toàn thế giới. Nó dùng rất nhiều cơ chế mã hóa, một vài trong số chúng đại diện cho những kí tự bất thường trong ứng dụng web.

- Unicode 16-bit hoạt động tương tự mã hóa Url. Để vận chuyển qua Http, các kí tự bất thường sẽ có tiền tố %u theo sau là điểm mã Unicode của kí tự đó theo thập lục phân.

%u2215 - /

%u00e9 – é

- Utf-8 là 1 mã hóa tiêu chuẩn dùng 1 hay nhiều byte để thể hiện 1 kí tự. Để truyền qua Http, các kí tự nhiều byte sẽ sử dụng mỗi byte ở dạng thập lục phân có tiền tố là %

%c2%a9 - © hay \xc2\xa9

%e2%89%a0 - ≠

Với mục đích tấn công ứng dụng web, điểm thú vị chính của mã hóa Unicode là giúp đánh bại bộ lọc đầu vào.

**HTML encoding:**

Mã hóa HTML được dùng để mã hóa những kí tự có vấn đề để mà chúng được kết hợp an toàn trong tài liệu HTML.

Rất nhiều kí tự có ý nghĩa đặc biệt trong HTML được dùng để khai báo cấu trúc của tài liệu hơn là nội dung. Để sử dụng các kí tự này an toàn như nội dung bình thường rất cần mã hóa nó.

&quot; - “

&apos; - ‘

&amp; - &

&lt; - <

&gt; - >

Ngoài ra, bất cứ kí tự nào cũng có thể được mã hóa HTML bằng việc dùng bảng mã ASCII ở dạng thập phân:

&#34; - “

&#39; - ‘

Hoặc dùng ASCII ở dạng thập lục phân (phải thêm tiền tố x):

&#x22; - “

&#x27; - ‘

Khi bạn tấn công ứng dụng web, điểm thú vị của mã hóa HTML là có thể thăm dò lỗ hổng XSS.

**Base64 encoding:**

Mã hóa Base64 cho phép bất kỳ dữ liệu nhị phân nào cũng được biểu diễn một cách an toàn chỉ bằng cách sử dụng các kí tự ascii có thể in được. Nó thường được dùng để mã hóa những tệp đính kèm email để vận chuyển an toàn qua giao thức smtp. Nó cũng được dùng để mã hóa thông tin đăng nhập trong xác thực basic http.

Rất nhiều ứng dụng web sử dụng mã hóa Base64 để truyền dữ liệu nhị phân trong cookies và những tham số khác, thậm chí làm rối loạn kẻ tấn công (để ẩn dữ liệu nhạy cảm, ngăn chặn sửa đổi nhỏ). Bạn nên tìm kiếm, giải mã nó khi thấy ở máy khách. Chuỗi được mã hóa Base64 rất dễ nhận ra bởi bộ kí tự riêng biệt của nó và sự xuất hiện các kí tự đệm ở cuối.

**Hex encoding:**

Rất nhiều ứng dụng sử dụng thẳng mã hóa thập lục phân khi truyền dữ liệu nhị phân, sử dụng kí tự ascii để đại diện cho chúng. Ví dụ, mã hóa hex tên người dùng daf trong cookie sẽ như sau:

646166

Giống với base64, hex mã hóa rất dễ nhận ra. Bạn nên thường xuyên giải mã bất cứ dữ liệu nào dạng hex mà máy chủ gửi đến máy khách để hiểu chức năng.