

TP1 d'Archie

La traduction ascii :

- 1) Le message @ETU.univ-lorraine.fr s'écrit 4145 5455 2E76 6E69 762D 6C6F 7272 616A 696E 652E 6672 en hexadécimal.
- 2) Ce code ascii écrit en binaire :
11000011 10101110 01101100 01100101 00100000 01100100 01110101
01011111 01010011 01000000 01110101 01101100 01100011 01111001
00100000 00111011 00101101 00101001
veut dire "île du_S@ulcy 😊" et les deux premiers octets correspondent à des caractères spéciaux dont inexploitable en ascii mais on peut les interpréter comme étant un î.
- 3) Le message ascii décodé avec hexdump veut dire : "La vie c'est comme une boîte de chocolat, on sait jamais sur quoi on va tomber."
- 4) La notion de boutisme / Endian est nécessaire au décryptage d'un tel fichier car sans elle le message n'aurait aucun sens.
- 5) La taille en octets de ce message est 80.
- 6) La commande Hexdump -C nous traduit directement le message du texte dans le terminal. Contrairement à Hexdump.
- 7) La différence entre le nombre de caractères lisibles et le nombre d'octets dans un fichier s'explique principalement par la manière dont les caractères sont encodés en mémoire.

Encodage UTF-8 :

- 1) Les points de code pour la chaîne "coeur" sont : 99 339 117 114.
- 2) Une fois encodé en UTF-8 on obtient : 0x63 0xC5 0x93 0x75 0x72.
- 3) On décode maintenant la suite hexadécimale 0xC3 0xAE 0x6C 0x65 et on obtient "île"
- 4) Enfin, on décode la suite 0x76 0xE9 0x6C 0x6F en UTF-8. Malheureusement la séquence hexadécimale 0x76 0xE9 0x6C 0x6F est invalide à partir de 0xE9, car le troisième octet (0x6C) ne respecte pas la structure UTF-8 pour un caractère multi-octet.