Encoding in Unicode and UTF-8

- The Unicode standard assigns a character a unique value in the range

$$0 - 2^{32} = 4,294,967,296$$

- Each value can be stored in 4-bytes (each byte is 8-bits)
- `https://unicode.org/charts/`
- Not very efficient if the characters used most often have low Unicode values
- UTF-8 is a system to encode Unicode values using 1–4 bytes per character
  Requires using leading bits to indicate how many bytes the character uses

Stream ciphers vs. Block Ciphers

- Convert plaintext to bits $x_0, x_1, \ldots, x_b$

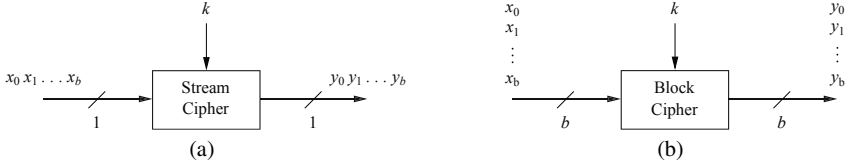

Fig. 2.2 Principles of encrypting b bits with a stream (a) and a block (b) cipher

From Paar and Pelzl, pg. 30

- Stream cipher encrypts one-bit at a time
- Block cipher encrypts entire block at once
  Algorithm may use all bits to create ciphertext
  Will see a version AES that uses 128-bit (or 16 byte) blocks

The goal is to encrypt the word “Pi?” using a stream cipher

1. Look up the UTF-8 encoding for “P”, “i”, and “?” at `https://unicode.org/charts/`

2. Convert each encoding to a binary number and combine to get a 48 bit value
   This is the plaintext x

3. Encrypt x using the key stream, starting at the right end

   $$s = 1010\ 1010\ 1010\ 1010\ 1010\ 1010\ 1010\ 1010\ 1010\ 1010\ 1010\ 1010$$

   Note this is a silly key stream, but makes the computation tractable by hand

Some desirable properties for any cryptosystem

Kerckhoff's Principle: A cryptosystem should be secure when the attacker knows all the details of the encryption and decryption algorithms but does not know the secret key.

i.e. No security through obscurity of the method

Definition: A cryptosystem is *unconditionally secure* if it cannot be broken, even with infinite computational resources.

Additional desirable properties for block ciphers

- **Confusion:** The relationship between the key and cipher text is obscured.

  i.e. Each bit of cipher text should depend on several parts of the key so that if one bit of the key is changed, then the cipher text should appear to be changed in a random way.

- **Diffusion:** Each plaintext symbol affects many parts of the cipher text.

  i.e. If one bit of plain text is changed, then much of the cipher text is changed.