CIS 3362

Goal Post #2 before I leave

Code Breaking - Substitution, Vigenere

Today: Vigenere Cipher

Under substitution transformation the set of letter frequencies is UNCHANGED (invariant).

Plaintext: T O D A (Y) I S W E D N E S D A (Y)

19, 14, 3, 0, 24, 8, 18, 22, 4, 3, 13, 4, 18, 3, 0, 24

KEY: K N I G H T S K N I G H T S K N

\+ 10, 13, 8, 6, 7, 19, 18, 10, 13, 8, 6, 7, 19, 18, 10, 13

29, 27, 11, 6, 31, 27, 36, 32, 17, 11, 19, 11, 37, 21, 10, 37

↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓

3 1 5 1 10 6 11 11

Cipher D B L G F B K G R L T (L L) V K L

=

Remained unbroken 1500s - early 1800s

This breaks the fundamental assumption that

(1) the same letter plain maps same letter cipher

(2) 2 ciphered letters are same ⇒ same plain text letters

LL

TO DO

① Live code to fill some missing pieces on my simple programs.

② Vigenere Cryptanalysis

→ frequencies messed up

What about repeated n-grams?

$P_1 P_2 P_3 P_4 P_5$
$K_3 K_4 K_5 K_6 K_7$

67, 68, 69, 70, 71
$L_1 L_2 L_3 L_4 L_5$

$P_1 P_2 P_3 P_4 P_5$
$K_3 K_4 K_5 K_6 K_7$

122, 123, 124, 125, 126
$L_4 L_2 L_3 L_4 L_5$

Scenario 1
same plain
same cycle
of key

$P_1 P_2 P_3 P_4 P_5$
$K_1 K_2 K_3 K_4 K_5$
$L_1 L_2 L_3 L_4 L_5$

$P_6 P_7 P_8 P_9 P_{10}$
$K_4 K_5 K_6 K_7 K_8$
$L_1 L_2 L_3 L_4 L_5$

Scenario 2
} Coincidence

→ Probability low → longer the repeated string, lower probability...

→ repeated likely keyword lines up

122
−67
55

→ keyword length likely divides into this evenly
5, 11

Kasiski Test: Look for all repeated n-grams $n \geq 3$. Note their starting index

Indexes: 67, 122, 207

55 85 num divides into 55 AND 85

take gcd of each of these gaps and that's likely the keyword length.