Classical Ciphers

CS 411/507 - Cryptography

Erkay Savaş & Atıl Utku Ay

Faculty of Engineering and Natural Sciences
Sabancı University

October 1, 2023

Modulo Operation

- Question: What is $12 \bmod 9$?
- Answer: $12 \bmod 9 = 3$ or $12 \equiv 3 \bmod 9$
- Definition: Let a, r, $m \in \mathbb{Z}$ (where $\mathbb{Z}$ is a set of all integers) and $m > 0$. We write
  - $a \equiv r \bmod m$ if m divides $a - r$.
  - m is called the modulus
  - r is called the remainder
  - $a = q \cdot m + r \qquad 0 \leq r < m$

Ring

Definition: The ring $\mathbb{Z}_m$ consists of

1. The set $\mathbb{Z}_m = \{0, 1, 2, \ldots, m-1\}$
2. Two operations "$+$" and "$\times$" for all $a, b \in \mathbb{Z}_m$ such that
   - $a + b \equiv c \bmod m \quad (c \in \mathbb{Z}_m)$
   - $a \times b \equiv d \bmod m \quad (d \in \mathbb{Z}_m)$

- Example: $m = 9$
  - $\mathbb{Z}_9 = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$
  - $6 + 8 = 14 \equiv 5 \bmod 9$
  - $6 \times 8 = 48 \equiv 3 \bmod 9$

Properties of Ring $\mathbb{Z}_M$ 1/2

- Two operations $+$ and $\times$?

1. The additive identity "0":
   $a + 0 = a$ where $a \in \mathbb{Z}_m$
2. The additive inverse of a :
   $-a = m - a$ s.t. $a + (-a) \equiv 0 \bmod m$
3. Addition is closed i.e. if $a, b \in \mathbb{Z}_m$ then
   $a + b \in \mathbb{Z}_m$
4. Addition is commutative: $a + b = b + a$
5. Addition is associative: $(a + b) + c = a + (b + c)$

Properties of Ring $\mathbb{Z}_M$ 2/2

6. Multiplicative identity "1":
   $$a \times 1 \equiv a \bmod m$$
7. The multiplicative inverse of a exists if $\gcd(a, m) = 1$ and denoted as a^{-1} s.t.
   $$a^{-1} \times a \equiv 1 \bmod m$$
8. Multiplication is closed
   i.e. if $a, b \in \mathbb{Z}_m$ then $a \times b \in \mathbb{Z}_m$
9. Multiplication is commutative
   $a \times b = b \times a$
10. Multiplication is associative
    $(a \times b) \times c = a \times (b \times c)$

Some Remarks on $\mathbb{Z}_m$

- A mathematical structure in which we can
  - add,
  - subtract,
  - multiply, and
  - sometimes even divide.
- Divisibility
  - Is the division $(4/15) \bmod 26$ possible?
  - ??

Order of Operations

- The modulo operation can be applied in any order we want
- $(a + b) \bmod m$
  $= [(a \bmod m) + (b \bmod m)] \bmod m$
- $(a \times b) \bmod m$
  $= [(a \bmod m) \times (b \bmod m)] \bmod m$
- Example: Exponentiation in $\mathbb{Z}_m$
  - $3^8 \bmod 7 = ?$
  - $3^8 \bmod 7 = 6561 \bmod 7 = 2$
    since $6561 = 937 \times 7 + 2$

Arithmetic in $\mathbb{Z}_m$

- Example: Exponentiation in $\mathbb{Z}_m$ (cont.)
  - $3^8 = 3^4 \times 3^4 = 3^2 \times 3^2 \times 3^2 \times 3^2$
  - $3^8 \bmod 7$
    $= [(3^2 \bmod 7) \times (3^2 \bmod 7) \times (3^2 \bmod 7) \times (3^2 \bmod 7)] \bmod 7$
    $= 3^8 \bmod 7 = 2 \times 2 \times 2 \times 2 \bmod 7 = 16 \bmod 7 = 2$
- The ring $\mathbb{Z}_m$ and its arithmetic are of central importance in modern PKC.
- In practice, the order of the integers involved in PKC are in the range of $[2^{160}, 2^{4096}]$.

- Shift Cipher:

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

  - Algorithm:
    - Let $P = C = K = \mathbb{Z}_{26}$ and $x \in P$, $y \in C$, $k \in K$
    - Encryption: $y = E_k(x) = x + k \mod 26$.
    - Decryption: $x = D_k(y) = y - k \mod 26$.

Shift Cipher

- Remark: The shift cipher is also known as
  - Caesar Cipher.
- Example: Let the key be $k = 17$
  - Plaintext:
    $X = \text{ATTACK} = (0, 19, 19, 0, 2, 10)$.
  - Ciphertext:
    $Y = (0 + 17 \bmod 26, 19 + 17 \bmod 26, \ldots)$
    $Y = (17, 10, 10, 17, 19, 1) = \text{RKKRTB}$

A 0	B 1	C 2	D 3	E 4	F 5	G 6	H 7	I 8	J 9	K 10	L 11	M 12
N 13	O 14	P 15	Q 16	R 17	S 18	T 19	U 20	V 21	W 22	X 23	Y 24	Z 25

Attacks on Shift Cipher 1/2

1. Ciphertext Only:
   - Exhaustive Search:
     - Try all possible keys. $|K| = 26$.
     - Nowadays, for moderate security $|K| \geq 2^{80}$
     - Recommended security $|K| \geq 2^{100}$.
   - Letter frequency analysis
     - Same plaintext letters maps to same ciphertext letters
     - For example, the letter "E" occurs most frequently in English.
     - If the letter "V" occurs most in the ciphertext, there is a high probability that V corresponds to E and $K = 21 - 4 = 17$.
     - The ciphertext should be sufficiently long

Attacks on Shift Cipher 2/2

2. Known Plaintext:
   - If one letter of the plaintext along with the corresponding letter of ciphertext is known.
   - For example, T($= 19$) is known to encrypt to D($= 3$)
3. Chosen Plaintext:
   - Choose the letter "A" as the plaintext. The corresponding ciphertext gives the key.
   - For example, if the ciphertext is "H", then the key is 7.
4. Chosen Ciphertext:
   - Choose the letter "A" as ciphertext. The plaintext is the negative of the key
   - For example, if the plaintext is "H", the key is $-7 \equiv 19 \mod 26$.

Affine Cipher 1/3

- Algorithm: x and $y \in \mathbb{Z}_{26}$
  - Key: $k = (\alpha, \beta)$ and $\alpha, \beta \in \mathbb{Z}_{26}$
  - Encryption: $E_k(x) = y = \alpha \cdot x + \beta \bmod 26$
  - Decryption: $D_k(x) = x = \alpha^{-1} \cdot y + \gamma \bmod 26$
- Example: $k = (\alpha, \beta) = (11, 4)$
  - INPUT $= (8, 13, 15, 20, 19) \rightarrow$ ORNQF

Exercise

Encrypt the message `"AFFINE"` using Affine cipher where the key is $k = (\alpha, \beta) = (2, 4)$

$$E_k(x) = y = \alpha \cdot x + \beta \bmod 26$$

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

Affine Cipher 2/3

- Example: $k = (\alpha, \beta) = (2, 4)$
  - AFFINE $= (0, 5, 5, 8, 13, 4) \rightarrow$ EOOUEM
  - No one-to-one map between plaintext and ciphertext space.
  - What went wrong?

Affine Cipher 3/3

- Key Space:
  - β can be any number in $\mathbb{Z}_{26}$.
  - Condition
    $gcd(\alpha, 26) = 1 \rightarrow \alpha \in \{1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23, 25\}$
- The key space has $12 \cdot 26 = 312$.
- Attack types:
  1. Ciphertext only: exhaustive search or frequency analysis
  2. Known plaintext: Two letters in the plaintext and corresponding ciphertext letters would suffice to find the key.

Affine Cipher - Attack Types

- Example:
  - plaintext: $IF = (8, 5)$
  - ciphertext: $PQ = (15, 16)$
  - $\alpha = 17$ and $\beta = 9$
- What would you do if you had only one letter of known plaintext?
- Example:
  - A ciphertext message given: HRRA WRY XRF
  - A known plaintext-ciphertext pair given: (U,F)
  - U $\rightarrow 20$ and F $\rightarrow 5$
  - $5 = 20\alpha + \beta \bmod 26$
  - $\{(1, 11), (3, 23), (5, 9), (7, 21), (9, 7), (11, 19), (15, 17), (17, 3), (19, 15), (21, 1), (23, 13), (25, 25)\}$

Affine Cipher - Attack Types

- Attack types (cont.):
  1. Chosen plaintext:
     - Choose "AB" as the plaintext.
     - 1^{st} character of the ciphertext: $0 \cdot \alpha + \beta = \beta$
     - 2^{nd}: $\alpha + \beta$.
  2. Choose "AB" as the ciphertext.

The Vigenère Cipher 1/2

- "le chiffre indéchiffrable"
  - Giovan Battista Bellaso in his 1553 book La cifra del. Sig.;
- The key is a vector chosen as follows:
  - First choose a key length, say 6.
  - Then choose a vector (key) of this length whose, entries are from $\mathbb{Z}_{26}$.
    e.g., $k = (21, 4, 2, 19, 14, 17)$.
  - Often, the key corresponds to a word that can be remembered, but not easily guessed (In our example, VECTOR)
  - The security of the system depends on the fact neither the keyword nor its length is known.

The Vigenère Cipher 2/2

- The key is a vector chosen as follows (cont.):
  - $k = (21, 4, 2, 19, 14, 17)$.
  - To encrypt the message in our example, we take the first letter of the plaintext and shift by 21. Then shift the second letter by 4, and so on.
  - Once we exhaust all the keys, we start back at the first entry in the vector and carry on.
- Example:

H	E	R	E	H	O	W	I	T	W	O	R	K	S
21	4	2	19	14	17	21	4	2	19	14	17	21	4
C	I	T	X	V	F	R	M	V	P	C	I	F	W

Breaking Vigenère Cipher

- Known plaintext attack
  - Possible if we know a certain number of ciphertext-plaintext pairs
- Chosen-plaintext attack
  - pick "$AAAAAA\ldots$" as the plaintext
- Assume ciphertext-only attack.
  - Frequency analysis would help if we knew the length of the key.
  - If the length of the key is n, then the problem is reduced into "solving n distinct shift ciphers"
- Question: How can we find the key length?

Letter Frequencies in English

letter	freq.	letter	freq.	letter	freq.	letter	freq.
e	12.7%	h	6.1%	w	2.3%	k	0.8%
t	9.1%	r	6.0%	f	2.2%	j	0.2%
a	8.2%	d	4.3%	g	2.0%	x	0.1%
o	7.5%	l	4.0%	y	2.0%	z	0.1%
i	7.0%	c	2.8%	p	1.9%	q	0.1%
n	6.7%	u	2.8%	b	1.5%		
s	6.3%	m	2.4%	v	1.0%		

Statistics of English Language

- Observation:
  - If you write down two English texts one below the other, the number of matching characters will be larger than the one you get for two texts, one of which is English and the other is a text obtained by shifting the letters.
- The reason:
  - Frequency vector:
    - $A_0 = \{.082, .015, .028, \ldots, .020, .001\}$
  - Shifted frequency vector (to the right)
    - $A_2 = \{.020, .001, .082, .015, .028, \ldots\}$

Shifted Frequency Vectors

- Dot Products
  - $A_0 \cdot A_0 = (.082)^2 + (.015)^2 + (.028)^2 + \ldots = .066$
  - $A_0 \cdot A_1 = .082 \times .001 + .015 \times .082 + \ldots = .039$
  - $A_0 \cdot A_2 = .082 \times .020 + .015 \times .001 + \ldots = .032$

$\lvert i-j \rvert$	$A_i \cdot A_j$	$\lvert i-j \rvert$	$A_i \cdot A_j$
0	0.066	7	0.039
1	0.039	8	0.034
2	0.032	9	0.034
3	0.034	10	0.038
4	0.044	11	0.039
5	0.033	12	0.042
6	0.036	13	

Finding the Key Length

- Method:
  1. Write down the ciphertext on a piece of paper
  2. Below the ciphertext, write down the same ciphertext shifted one letter to the right
  3. Count the total number of coincidences
  4. Increment the shift amount in the ciphertext.
  5. if you exceed a predetermined shift amount go to Step 6. Otherwise go to Step 3.
  6. Stop
  7. Possible key length is the shift amount that gives the highest number of coincidences.

Example: Finding the Key Length 1/2

- Ciphertext
  - xnyrcshjnijwvxodwfepolzganyorsrucofjczmoid heythjemolpgxyknrtifxehjemolsjpfxgwkxfydofh cyoywhuywbgrvkstsvrciqgycntcolvbewlfvxtwfxy oubviiqrcoxydhsfxfykhzifdhsjejtuglejpawjeen honmeqpsjlrzshzisothwvtvawefvmaikizdwokkiks gqeenwofxvnwssvkroiylrcfcjxykthzigksgarxdhs jiykdkgvedhsevvklzqesyuhllvcaaweenbcllkrahe sixibyiheazdcckywfpvkvskrfctshlrntfghuonpde tuovaovzthziwsrgljfbabgxyorrscpotyfsnsnuzsn gamdirnscfxfgamahfebhwhzpigzslvdsniimoawfrm kwklrvlpwxvvlwfkkrigomkragakycoawayorsskvca bvexosvwrtotkgvfkdgvmmoruwhzxakgsuknramkyoy llvynsdijctfszvvertcrxdhzekrageeuoazdxyodwx jvbebui

Example: Finding the Key Length 2/2

- Apply the method
  - x n y r c s h j n i j w v x o d w f e p o l z g ...
  - x n y r c s h j n i j w v x o d w f e p o l ...
  - a n y o r s r u c o f j c z m o i d h e y t h j ...
  - z g a n y o r s r u c o f j c z m o i d h e y t ...
  - e m o l p g x y k n r t i f x e h j e m o l s j ...
  - h j e m o l p g x y k n r t i f x e h j e m o l ...

Shifts	1	2	3	4	5	6	7	8
Coincidences	18	17	12	20	21	44	25	16

Finding the Key

- We have a considerable evidence that the key length is 6.
  - If this is true then, there are six shift ciphers we have to crack
  - $K = (k_1, k_2, k_3, k_4, k_5, k_6)$
- Method
  1. Obtain six sub-ciphertexts
     - By combining 1^{st}, 7^{th}, 13^{th}, ... ciphertext letters
     - By combining 2^{nd}, 8^{th}, 14^{th},
     - so on.
  2. Apply the frequency analysis to each sub-ciphertext

Frequency Analysis 1/4

- First sub-ciphertext:

E	I	X	L	...
13	13	12	8	...
A	E	T	H	

  - E $\rightarrow$ E: $\Rightarrow k_1 = 0$: I $\rightarrow$ I; X $\rightarrow$ X
  - T $\rightarrow$ E: $\Rightarrow k_1 = 11$: X $\rightarrow$ I; M $\rightarrow$ X
  - A $\rightarrow$ E: $\Rightarrow k_1 = 4$: E $\rightarrow$ I; T $\rightarrow$ X

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12
8.2	1.5	2.8	4.3	12.7	2.2	2.0	6.1	7.0	0.2	0.8	4.0	2.4
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25
6.7	7.5	1.9	0.1	6.0	6.3	9.1	2.8	1.0	2.3	0.1	2.0	0.1

Frequency Analysis 2/4

- First sub-ciphertext:

E	I	X	L	...
13	13	12	8	...
A	E	T	H	

$k_1 = 4$

- Second sub-ciphertext:

V	F	Y	...	...
14	10	9	...	...
E	O	H		

$k_2 = 17$

Frequency Analysis 3/4

- Third sub-ciphertext:

O	K	Y	...	...
15	11	10	...	...
E	A	O	...	...

$k_3 = 10$

- Fourth sub-ciphertext:

A	T	O	...	
14	11	9	...	
A	T	O	...	

$k_4 = 0$

Frequency Analysis 4/4

- Fifth sub-ciphertext:

C	R	G	Q	...
12	11	10	9	...
E	T	I	S	...

$k_5 = 24$

- Sixth sub-ciphertext:

W	F	G	...	...
10	9	9	...	...
E	N	0	...	...

$k_6 = 18$

Finally

- Plaintext
  - tworoadsdivergedinayellowwoodandsorryicoul
    dnottravelbothandbeonetravelerlongistoodand
    lookeddownoneasfarasicouldtowhereitbentinth
    eundergrowththentooktheotherasjustasfairand
    havingperhapsthebetterclaimbecauseitwasgras
    syandwantedwearthoughasforthatthepassingthe
    rehadwornthemreallyaboutthesameandboththatm
    orningequallylayinleavesnostephadtroddenbla
    ckohikeptthefirstforanotherdayyetknowinghow
    wayleadsontowayidoubtedifishouldevercomebac
    kishallbetellingthiswithasighsomewhereagesa
    ndageshencetworoadsdivergedinawoodandiitook
    theonelesstraveledbyandthathasmadeallthedif
    ference

THE ROAD NOT TAKEN

two roads diverged in a yellow wood,
and sorry I could not travel both
and be one traveler, long I stood
and looked down one as far as I could
to where it bent in the undergrowth;
then took the other, as just as fair,
and having perhaps the better claim,
because it was grassy and wanted wear;
though as for that the passing there
had worn them really about the same,
and both that morning equally lay
in leaves no step had trodden black.
oh, I kept the first for another day!
yet knowing how way leads on to way,
I doubted if I should ever come back.
I shall be telling this with a sigh
somewhere ages and ages hence:
two roads diverged in a wood and I-
I took the one less traveled by,
and that has made all the difference.

Substitution Ciphers

- Each letter in the alphabet is replaced (substituted) by another letter.
- More precisely, a permutation of the alphabet is chosen and applied to the plaintext.
  - The shift and affine ciphers are examples of substitution ciphers.
- The number of permutations is $26!$
- Ciphertext preserves the statistics of the language used for plaintext,
  - the frequency analysis is an effective way of breaking substitution ciphers (if you know the language statistics)

Diffusion & Confusion

- Exhaustive search in cryptanalysis is an efficient method to break a cryptosystem when a relatively small key space is used.
  - However, large key space not necessarily guarantees to provide high cryptographic strength.
  - Example: The key space of a substitution cipher is $26! \approx 4 \times 10^{26}$ (equivalent to 89-bit keys)
  - Statistical properties of the language is preserved in the ciphertext
- Diffusion property of Shannon requires that changing one letter in the plaintext result in changes in several letters of the ciphertext.

- Hill Cipher: The key is an $n \times n$ matrix whose entries are integers in $\mathbb{Z}_{26}$.
- Example: Let $n = 3$ and the key matrix be

$$K = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 11 & 9 & 8 \end{pmatrix}$$

- Let the plaintext be "ABC" $= (0, 1, 2)$ then the encryption operation is a vector-matrix multiplication

Hill Cipher

- Encryption:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 11 & 9 & 8 \end{pmatrix} \times \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \equiv (8, 17, 25) \bmod 26 \rightarrow \texttt{IRZ}$$

- Decryption: The inverse of the key matrix is needed

$$K^{-1} = \begin{pmatrix} 22 & 5 & 1 \\ 6 & 17 & 24 \\ 15 & 13 & 1 \end{pmatrix}$$

Hill Cipher & Diffusion

- If we change one letter in the plaintext, three letters of the ciphertext will be affected.
- Example: Let the plaintext be "BBC" instead of "ABC" then the ciphertext

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 11 & 9 & 8 \end{pmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \equiv \begin{pmatrix} 9 & 21 & 10 \end{pmatrix} \bmod 26 \rightarrow \texttt{JVK}$$

Compare this new ciphertext ("JVK") with the previous ("IRZ")

Confusion

- Confusion means that the key does not relate to the ciphertext in a simple way.
  1. Each character of the ciphertext should depend as many key characters as possible
  2. There should be a complex function that relates the key to the ciphertext
- Example (Hill Cipher):

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 11 & 9 & 8 \end{pmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \equiv (9, 21, 10) \bmod 26 \rightarrow \texttt{JVK}$$

$$Y_0 = K_{00}X_0 + K_{01}X_1 + K_{02}X_2 \bmod 26$$

Characteristics of Cryptosystems

- Claude Shannon, in *Communication theory of secrecy systems* Bell Systems Technical Journal 28, (1949), 656-715, introduced properties that a good cryptosystems should have:
  - Diffusion: one character change in the plaintext should effect as many ciphertext characters as possible.
  - Confusion: The key should not relate to the ciphertext in a simple way.

One-Time Pad or Vernam Cipher

- The one-time pad, which is a provably secure cryptosystem, was developed by Gilbert Vernam in 1918.
- The message is represented as a binary string (a sequence of 0's and 1's; e.g., ASCII coding.)
- The key is a truly random sequence of 0's and 1's of the same length as the message.
- The encryption is done by adding the key to the message modulo 2, bit by bit.
  - This process is often called exclusive OR, and is denoted by XOR. The symbol $\bigoplus$ is used.

One-Time Pad: Example

- The message is "IF" (1001001 1000110)
- The key (1010110 0110001).
- Encryption:

plaintext		1001001 1000110
key	$\oplus$	1010110 0110001
ciphertext		0011111 1110111

- Decryption:

ciphertext		0011111 1110111
key	$\oplus$	1010110 0110001
plaintext		1001001 1000110

Provable Security in One-Time Pad

- How can we prove it is unbreakable?
  - The security depends on the randomness of the key
  - It is hard to define randomness
  - In cryptographic context, we seek two fundamental properties in a binary random key sequence
    1. Unbiased (Equal Distribution): The numbers of 1's and 0's are expected to be equal
    2. Unpredictability: Independent of the number of the bits of a sequence observed, the probability of guessing the next bit is not better than $\frac{1}{2}$. Therefore, the probability of a certain bit being 1 or 0 is exactly equal to $\frac{1}{2}$.

Randomness of Ciphertext

- $\Pr(k_i = 0) = \Pr(k_i = 1) = \frac{1}{2}$.
- The plaintext bits are not balanced.
  $Pr(m_i = 0) = x$ and $\Pr(m_i = 1) = 1 - x$.
- Let us calculate the probability of ciphertext bits.

m_i	Prob.	k_i	Prob.	c_i	Prob.
0	x	0	$\frac{1}{2}$	0	$\frac{1}{2}x$
0	x	1	$\frac{1}{2}$	1	$\frac{1}{2}x$
1	$1-x$	0	$\frac{1}{2}$	1	$\frac{1}{2}(1-x)$
1	$1-x$	1	$\frac{1}{2}$	0	$\frac{1}{2}(1-x)$

- $\Pr(c_i = 0) = \Pr(c_i = 1) = (\frac{1}{2})x + (\frac{1}{2})(1-x) = \frac{1}{2}$.
- Ciphertext looks like a random sequence.

Perfect Security with Shift Cipher

- Pick a random shift amount for every letter in the plaintext

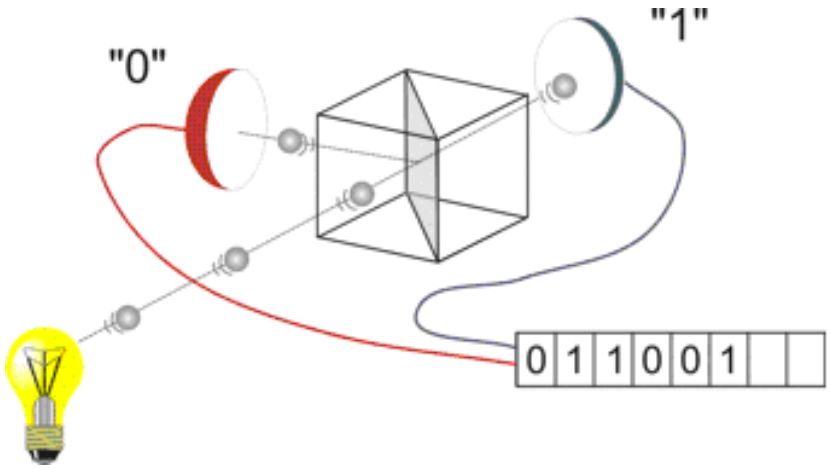
Randomness & Pseudorandomness

- We need random numbers for cryptography
  - One-time pad
  - Secret keys for DES, AES, etc.
  - Random process for choosing primes p, q for RSA
  - Private key for ECC
  - Challenges used in challenge-based identification systems (e.g. zero-knowledge proof protocols)
- The security of many cryptographic systems depend on the generation of unpredictable numbers.

True Random Number Generation

- Randomness exists in nature
  - Hardware based random number generators (RNG) exploit the randomness which occurs in some physical phenomena
    - Events in the quantum level
    - Elapsed time between emission of particles during radioactive decay
    - Thermal noise from a semiconductor diode or resistor clock drift.
    - Frequency instability of a free running oscillator
    - The amount which a metal insulator semiconductor capacitor is charged during a fixed period of time.
    - Read “hardware random number generator” article in wikipedia

Quantum Random Number Generator



- Quantis - Quantum Random Number Generators (QRNG)
  - A Commercially available chip
  - True quantum randomness (passes all randomness tests)
  - High bit rate of 4Mb/s (up to 16Mb/s)
  - USB, PCIe connections
  - http://www.idquantique.com/random-number-generators/products.html

Software-Based RNG

1. The system clock
2. Elapsed time between keystrokes or mouse movement
3. Content of input/output buffers
4. User input
5. OS values such as system load and network statistics.

- All of them are subject to observation and manipulation.
- Individually these sources are very "weak".
- The randomness can be increased by combining the outputs of these sources using a complex mixing function (e.g. hashing the concatenation of the output bits).
- Still, not quite secure for high security applications!

De-Skewing

- A natural source of randomness may be defective. The output bits may be biased or correlated.
- There are techniques to fix biased output bits. One is called de-skewing.
- Example:Assume a RNG with $\Pr(k_i = 1) = p$ and $\Pr(k_i = 0) = 1 - p$ where $0 < p < 1$
- Solution: Group the sampled bit in two and do the following
  - $00 \rightarrow$ discard
  - $01 \rightarrow$ take as 0
  - $10 \rightarrow$ take as 1

  } Equal probability
  - $11 \rightarrow$ discard

Pseudorandom Number Generation

- (PRNG) is a deterministic algorithm,
  - which, given a truly random binary sequence of length N (random seed), outputs a binary sequence of length L, where $L >> N$ which "appears" to be random.
  - The output of a PRNG is not random;
  - In fact possible output sequence is at a small fraction $2^N/2^L$ of all possible binary sequences of length L.
  - However, it is impractical (computationally infeasible) for a anyone (adversary) to distinguish a pseudorandom sequence from a truly random sequence of the same length.

Pseudo-Randomness Tests

- We don't know what the randomness is.
- There is no practical test to check if a sequence is truly random.
- Thus, we can't define exactly what the pseudorandomness is. There are attempts, though.
- For cryptography we use statistical test for randomness

Statistical Tests for Pseudorandomness

- FIPS (US government standards - Federal Information Processing Standards) 140-1 specifies some statistical tests:
  1. Frequency test (monobit test): $\#$ of 1s and 0s must be approximately the same
  2. Poker test: A sequence is divided into k non-overlapping segments of length m. This test determines if each segment of length m appears approximately the same number of times.
  3. Runs Test: Determines if the $\#$ of runs of various lengths is similar to those of truly random sequences
  4. Long run test: The long run test is passed if there are no runs of length 34 or more.

NIST Webpage for RNG

- http://csrc.nist.gov/groups/ST/toolkit/rng/index.html
- https://web.archive.org/web/20160125103112/http://stat.fsu.edu/pub/diehard/