МІНІСТЕРСТВО ВНУТРІШНІХ СПРАВ УКРАЇНИ

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ВНУТРІШНІХ СПРАВ

Кафедра інформаційних технологій факультету №4

ТЕКСТ ЛЕКЦІЇ

з навчальної дисципліни "Блокчейн технологія та криптовалюти"
вибіркових компонент
освітньої програми другого рівня вищої освіти

125 Кібербезпека (Безпека інформаційних та комунікаційних систем)

за темою – Криптографія в блокчейн технологіях (Численне представлення інформації)

Харків 2018

ЗАТВЕРДЖЕНО
Науково-методичною радою
Харківського національного
університету внутрішніх справ
Протокол від _____ № ____

СХВАЛЕНО
Вченою радою факультету №4
Протокол від _____ № ____

ПОГОДЖЕНО
Секцією науково-методичної ради
ХНУВС з технічних дисциплін
Протокол від _____ № ____

Розглянуто на засіданні кафедри інформаційних технологій (протокол від _____№ ___)

Розробники:

доцент кафедри інформаційних технологій ХНУВС, к.т.н., доцент Носов В.В.

Рецензенти:

професор кафедри штучного інтелекту Харківського національного університету радіоелектроніки, д.т.н., професор Петров К.Е.

План лекції

1. Основні терміни і визначення
2. Позиційна система числення
3. Кодування текстової інформації

Рекомендована література:

1. https://studfiles.net/preview/953322/page:3/
2. https://uk.wikipedia.org/wiki/Позиційні_системи_числення
3. www.kievoit.ippo.kubg.edu.ua/kievoit/2013/52/52.html

Текст лекції

Вступ

Для розуміння технології блокчейн спочатку розглянемо терміни і визначення криптографії, а також способи представлення тексту у числовому вигляді

1. Основні терміни і визначення

Шифрування являє собою приховування інформації від неавторизованих осіб з наданням в цей же час авторизованим користувачам доступу до неї. Користувачі називаються авторизованими, якщо у них є відповідний ключ для дешифрування інформації. Складність полягає в тому, як реалізується весь цей процес.

Метою будь-якої системи шифрування є максимальне ускладнення отримання доступу до інформації неавторизованими особами, навіть якщо у них є зашифрований текст і відомий алгоритм, використаний для шифрування. Поки неавторизований користувач не володіє ключем, секретність і цілісність інформації не порушується.

За допомогою шифрування забезпечуються три стани безпеки інформації.

Конфіденційність. Шифрування використовується для приховання інформації від неавторизованих користувачів при передачі або при зберіганні.

Цілісність. Шифрування використовується для запобігання зміни інформації при передачі або зберіганні.

Ідентифікованість. Шифрування використовується для автентифікації джерела інформації та запобігання відмови відправника інформації від того факту, що дані були відправлені саме ним.

Терміни, пов'язані з шифруванням

На рис. 1 показаний загальний принцип, згідно з яким здійснюється шифрування.

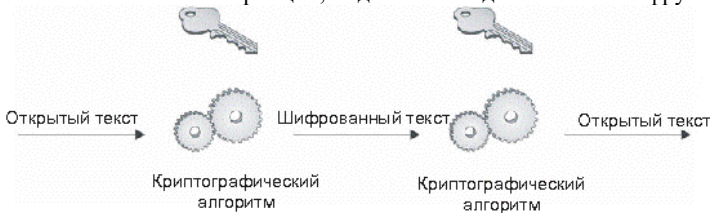

Рис. 1

- **Відкритий текст** - інформація у вихідному вигляді.
- **Шифрований текст** - інформація, піддана дії алгоритму шифрування.
- **Алгоритм** - метод, використовуваний для перетворення відкритого тексту в шифрований текст.
- **Ключ** - вхідні дані, за допомогою яких і алгоритму відбувається перетворення відкритого тексту в шифрований або відкритий вид.
- **Простір ключів** —набір можливих значень ключа.
- **Шифрування** - процес перетворення відкритого тексту у шифрований текст.

- **Дешифрування (розшифрування)** - процес перетворення шифрованого тексту у відкритий текст.

Існують також чотири загальних терміну.

- **Криптографія** - наука про приховування інформації за допомогою шифрування.
- **Криптограф** - особа, що займається криптографією.
- **Криптоаналіз** - мистецтво аналізу криптографічних алгоритмів на предмет наявності вразливостей.
- **Криптоаналітик** - особа, яка використовує криптоаналіз для визначення і використання вразливостей в криптографічних алгоритмах.

Криптологія – наука про забезпечення конфіденційності і цілісності інформації, шляхом її перетворення. Криптологія поєднує два напрямки – криптографію й криптоаналіз.

Алфавіт — кінцева множина використовуваних для кодування інформації знаків. В якості алфавіту можуть виступати як множина символів національних алфавітів, так і множина різних символів (наприклад, танцюючих чоловічків) і цифр.

Текст — упорядкований набір з елементів алфавіту.

Для розуміння криптографії потрібно розуміти принципи кодування.

Кодування — це процес заміни знаків одного алфавіту знаками іншого алфавіту при збереженні змісту інформації, яку подають за допомогою цих знаків.

Декодування — це процес, що обернений до кодування.

2. Позиційна система числення

Система числення – це сукупність правил запису чисел за допомогою певного набору символів.

Позиційна система числення визначається цілим числом $b > 1$ - *основою системи числення*. Система числення з основою b також називається b-річною (зокрема, *двійковою, трійковою, десятковою* і т. п.).

Ціле число x в b-річній системі числення представляється у вигляді кінцевої лінійної комбінації степенів числа b:

$$x = \sum_{k=0}^{n-1} a_k\, b^k,$$

$$x = a_{n-1}b^{n-1} + a_{n-2}b^{n-2} + \cdots + a_0 b^0,$$

де a_k - цілі числа, які називаються **цифрами**, задовольняють нерівності $0 \le a_k \le b-1$.

Кожна ступінь b^k в такому запису називається **розрядом (позицією)**, старшинство розрядів і відповідних їм чисел визначається значенням показника ступеня k. Зазвичай для ненульового числа x вимагають, щоб старша цифра a_{n-1} в b-річному поданні x була також ненульовою.

Побудова такого запису числа називають ***позиційним кодуванням числа***, а сам запис - ***позиційним кодом числа***.

Наприклад, значення $x_i = M = 59$, тоді його код за основою $b = 8$, буде мати вигляд

$$M = 59 = 7 \cdot 8^1 + 3 \cdot 8^0 = 73_8$$

Код того ж числа, але по основі $b = 4$ буде виглядати наступним чином:

$$M = 59 = 3 \cdot 4^2 + 2 \cdot 4^1 + 3 \cdot 4^0 = 323_4.$$

Якщо основа коду $b = 2$, то

$$M = 59 = 1 \cdot 2^5 + 1 \cdot 2^4 + 1 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^2 + 1 \cdot 2^0 = 111011_2.$$

Якщо основа коду $b = 16$, то

$$M = 59 = 3 \cdot 16^1 + B \cdot 16^0 = 3B_{16} = 0\text{x}3B$$

У шістнадцятковій системі є 16 окремих цифр. Оскільки в десятковій системі тільки 10 знаків, то доводиться використовувати літери - A, B, C, D, E і F. Кожній з них присвоюється числове значення:

$b = 10$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$b = 16$	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F

Таким чином, числа $3B,73,323$ і 111011 можна вважати, відповідно, шістнадцятирічним, восьмирічним, чотирирічним і двійковим кодами десяткового числа $M = 59$.

Щоб перевести ціле число з однієї системи числення з основою b_1 в іншу з основою b_2 необхідно послідовно ділити це число і одержувані частки на основу b_2 нової системи до тих пір, поки не вийде частка менше основи b_2. Остання частка - старша цифра числа в новій системі числення з основою b_2, а наступні за нею цифри - це залишки від ділення, що записуються в послідовності, зворотній їх отримання (рис. 2). Арифметичні дії виконуються в тій системі числення, в якій записано число, що переводиться.

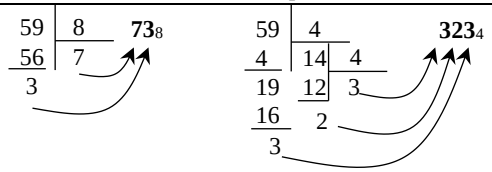

Рис. 2. Перевід цілого числа з однієї системи числення в іншу

3. Кодування текстової інформації

Для опрацювання текстових повідомлень з використанням комп'ютера символи тексту кодують числами.

Для кодування одного символу деякого алфавіту використовують щонайменше 1 байт інформації (8 бітів). $2^8 = 256$, тому за допомогою 1 байту можна закодувати 256 різних символів.

Кодування полягає в тому, що кожному символу ставиться у відповідність унікальний двійковий код від 00000000 до 11111111 (або десятковий код від 000 до 255, або шістнадцятковий код від 0x00 до 0xFF). Відповідність коду певному символу — це питання домовленості, зафіксованою *кодовою таблицею.*

Таблиця кодування — *це таблиця, в якій всім символам комп'ютерного алфавіту поставлено у відповідність порядкові номери (коди).*

У 1963 р. у США було розроблено набір кодів символів для передавання повідомлень телетайпом[1]. Пізніше він став стандартом для використання в комп'ютерній техніці й отримав назву таблиці кодів символів **ASCII** (англ. American Standard Code for Information Interchange - американський стандартний код для обміну інформацією).

У таблиці ASCII літерам англійського алфавіту, цифрам, розділовим знакам, символам редагування та форматування тексту ставляться у відповідність числа від 0 (00000000) до 127 (01111111) (рис.3, табл. 1).

[1] Телетайп (лат. tele - віддалений, англ. type - друкування) - електромеханічна друкуюча машина, яку використовують для передавання текстових повідомлень дротами.

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	`
1	1	[START OF HEADING]	33	21	!	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22	"	66	42	B	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	C	99	63	c
4	4	[END OF TRANSMISSION]	36	24	$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	&	70	46	F	102	66	f
7	7	[BELL]	39	27	'	71	47	G	103	67	g
8	8	[BACKSPACE]	40	28	(	72	48	H	104	68	h
9	9	[HORIZONTAL TAB]	41	29	)	73	49	I	105	69	i
10	A	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	B	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	C	[FORM FEED]	44	2C	,	76	4C	L	108	6C	l
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E	.	78	4E	N	110	6E	n
15	F	[SHIFT IN]	47	2F	/	79	4F	O	111	6F	o
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	p
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	s
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	T	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	v
23	17	[ENG OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	x
25	19	[END OF MEDIUM]	57	39	9	89	59	Y	121	79	y
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	[	123	7B	{
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	\	124	7C	|
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D	]	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	3F	?	95	5F	_	127	7F	[DEL]

Рис. 3. Таблиця ASCII

Таблиця 1. Групи кодів символів таблиці ASCII

Діапазон кодів	**Група символів**	**Приклад коду**	**Відповідний символ**
Від 0 до 31, 127	Спеціальні символи	10	Символ, що відповідає перенесенню курсора на новий рядок
		13	Символ, що відповідає поверненню курсора на початок рядка
		27	Символ, що відповідає натисканню клавіші ESC

Продовження таблиці 1

Діапазон кодів	**Група символів**	**Приклад коду**	**Відповідний символ**
Від 32 до 64, від 91 до 96, від 123 до 126	Розділові знаки та цифри	32	пропуск
		48	цифра 0
		123	{
Від 65 до 90	Великі літери англійського алфавіту	65	A
		66	B
		90	Z
Від 97 до 122	Малі літери англійського алфавіту	97	a
		98	ь
		122	2

Також існують розширення до таблиці ASCII, так звані *кодові сторінки* або *сторінки кодування* для кодів з номерами від 128 до 255 (залежно від локалізації), які кодують додаткові символи різних мов. У таблиці кодування ср1251 (від англійських code page — сторінка кодування або кодова сторінка) останні 128 кодів використано для кодування кирилиці: і білоруської, і російської, і української. Причому *алфавітний порядок для білоруської та української абеток суперечить порядку зростання кодів відповідних літер.*

Кодові сторінки «ANSI» (від American National Standards Institute — Американський національний інститут стандартів (США)), також звані Windows — рідні кодові сторінки

Windows. Містять багато символів типографіки, але майже не містять псевдографіку по причині того, що призначені для використання в графічному оточенні. Згодом корпорація Microsoft визнала, що використання імені ANSI було викликано непорозумінням. До Кодування «ANSI»/Windows відносять, зокрема, Windows-1252 і вищезгадану Windows-1251.

Код	Символ	Код	Символ	Код	Символ	Код	Символ	Код	Символ	Код	Символ	Код	Символ	Код	Символ
128	Ъ	144	ђ	160	Нерозривний пропуск	176	°	192	А	208	Р	224	а	240	р
129	Ѓ	145	‘	161	Ў	177	±	193	Б	209	С	225	б	241	с
130	,	146	’	162	ў	178	І	194	В	210	Т	226	в	242	т
131	ѓ	147	“	163	Ј	179	і	195	Г	211	У	227	г	243	у
132	„	148	„	164	¤	180	г	196	Д	212	Ф	228	д	244	ф
133	...	149	•	165	Ґ	181	µ	197	Е	213	Х	229	е	245	х
134	†	150	-	166	¦	182	¶	198	Ж	214	Ц	230	ж	246	ц
135	‡	151	-	167	§	183	·	199	З	215	Ч	231	з	247	ч
136	?	152	□	168	Ё	184	ё	200	И	216	Ш	232	и	248	ш
137	‰	153	™	169	©	185	№	201	Й	217	Щ	233	й	249	щ
138	Љ	154	љ	170	Є	186	є	202	К	218	Ъ	234	к	250	ъ
139	<	155	>	171	«	187	»	203	Л	219	Ы	235	л	251	ы
140	Њ	156	\	172	¬	188	j	204	М	220	Ь	236	м	252	ь
141	Ќ	157	:	173	-	189	S	205	Н	221	Э	237	н	253	э
142	Ћ	158	h	174	®	190	s	206	О	222	Ю	238	о	254	ю
143	Џ	159	u	175	Ї	191	ї	207	П	223	Я	239	п	255	я

Рис. 4. Таблиця кодів символів Windows-1251

Юнікóд, (англ. **Unicode**), УНІфіковане КОДування — це промисловий стандарт, розроблений, щоб забезпечити цифрове представлення символів усіх способів письма світу та спеціальних символів. Удосконалений сумісно із стандартом Універсальний Набір Символів (Universal Character Set — UCS) і опублікований у формі книги Стандарт Юнікод. Юнікод складається з множини символів, методології кодування та комплекту (набору) стандартів кодування символів, комплекту кодових таблиць для посилань на зображення символів, списку властивостей символів таких, наприклад, як верхній і нижній регістр, комплект довідкових даних комп'ютерних файлів, правил нормалізації, декомпозиції, зіставлення і зображення.

Юнікод знімає старе обмеження на кодування символів лише одним байтом. Натомість використовується 17 наборів, кожен з яких визначає 65536 кодів і дає можливість описати максимум 1114112 (17×2^{16}) різних символів. Як і в інших таблицях кодів, у Юнікоді незмінними залишаються перші 128 значень кодів, що відповідають таблиці ASCII. Окремий розділ у таблиці Юнікод містить коди літер кирилиці. Наприклад, літері «а» українського алфавіту ставиться у відповідність код 53 424, літері «і» - код 53 654, літері «г» -код 53 905 та ін.

Наразі у новітніх операційних системах використовується таблиця кодів Юнікод. Порівняння структур різних таблиць кодів символів наведено в таблиці 2.

Таблиця 2. Порівняння структури таблиць кодів символів

Таблиця кодів символів	Кількість кодів символів	Символи з кодами від 0 до 127	Символи з кодами від 128 до 255	Символи з кодами, більшими за 255
ASCII	128	Літери англійського алфавіту, цифри, розділові знаки,	Немає	Немає

		спеціальні символи		
KOI8-U	256	Як у таблиці ASCII	Символи кирилиці та деякі інші символи	Немає
Windows-1251	256	Як у таблиці ASCII	Символи кирилиці та деякі інші символи	Немає
Юнікод	1 114 112	Як у таблиці ASCII	Символи алфавітів різних мов світу та деякі інші символи	

У таблиці Юнікод містяться коди не лише літер та цифр, а й символів, які позначають торговельні марки, грошові одиниці, символи транскрипцій, ідеограми тощо. Знайти символ та відповідний йому код можна на сайті Таблица символов Юникода (https://unicode-table.com/ru).

Висновки

Шифрування являє собою приховування інформації від неавторизованих осіб з наданням в цей же час авторизованим користувачам доступу до неї. Користувачі називаються авторизованими, якщо у них є відповідний ключ для дешифрування інформації.

Будь-яке десяткове число можна записати в інших системах числення з різною кількістю цифр.

Для опрацювання текстових повідомлень з використанням комп'ютера символи тексту кодують числами.

Для кодування одного символу деякого алфавіту використовують щонайменше 1 байт інформації (8 бітів).

Кодування полягає в тому, що кожному символу ставиться у відповідність унікальний двійковий код від 00000000 до 11111111 (або десятковий код від 000 до 255, або шістнадцятковий код від 0x00 до 0xFF).

Відповідність коду певному символу — це питання домовленості, зафіксованою *кодовою таблицею.*

Найбільш відомі таблиці кодів:

- ASCII;
- розширення до таблиці ASCII - кодові сторінки «ANSI»;
- Unicode.