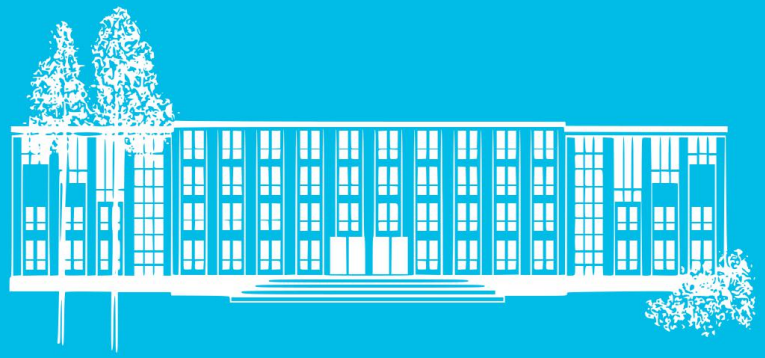


Odin | NSU Programing School for Students

Хеш-функция*

* Водолазский Даниил, Ижицкий Руслан, Миляев Иван

** Научный руководитель С. Ф. Кренделев

Новосибирск, июль 2016 г.

Вступление

- Пусть X – множество текстов, S – текст, $S = (s_1, s_2, \dots, s_l)$. Криптографическая хеш-функция – это отображение $\mathrm{F}: X \to \mathbb{Z}_{2^n}$, для которого выполняются следующие условия:

  1. Трудно для значения G хеш-функции подобрать такой текст S, что $F(S) = G$.

  2. Если $F_1 = F(S_1)$, то трудно подобрать такой текст S_2, что $F_1 = F(S_2)$ (стойкость к коллизиям I рода).

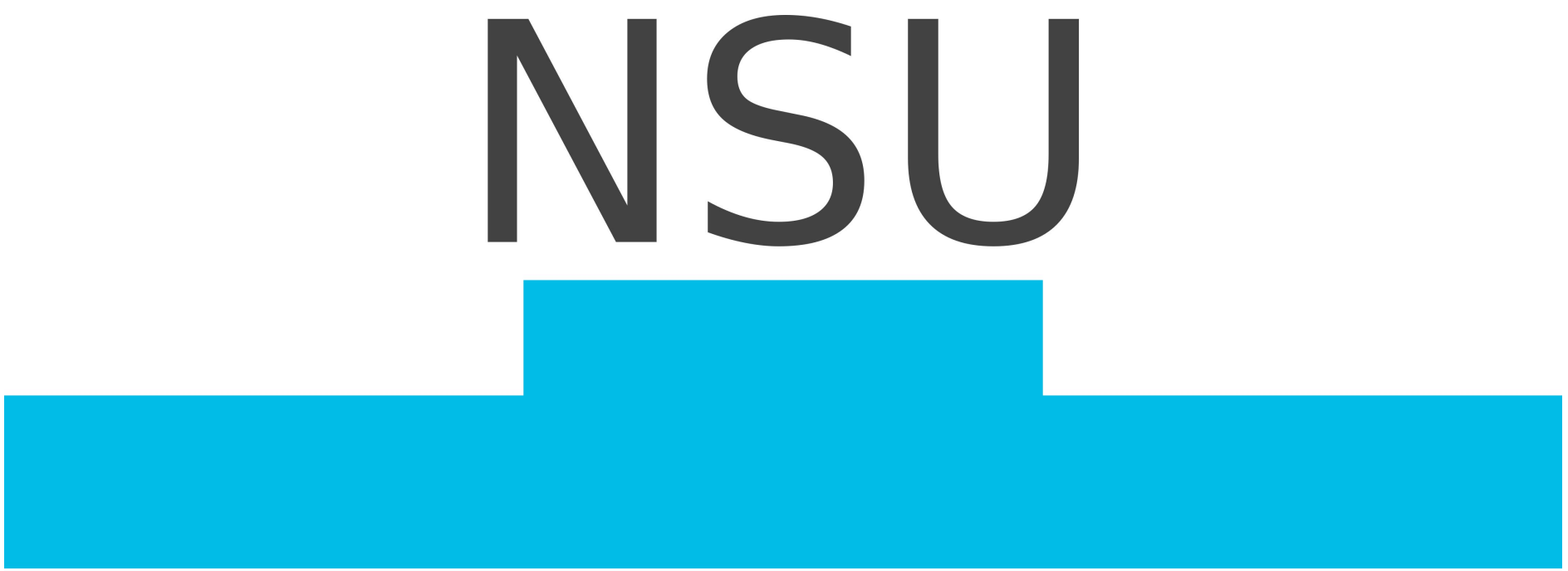
  3. Трудно найти такие S_1 и S_2, что $\mathrm{F}(S_1) = F(S_2)$ (стойкость к коллизиям II рода).

Актуальность

- В последнее десятилетие стали популярны работы, посвященные квантовым компьютерам. Хеш-функции, ранее много лет использовавшиеся в криптографии, теперь становятся ненадежными и устаревают. Кроме того, эти хеш-функции не универсальные (т. е. «заточены» под конкретное число бит), и зачастую в них находят уязвимости. Поэтому важно разработать хеш-функцию нового поколения, соответствующую современным стандартам в области криптографии, не имеющую аналогов в мире.

Задача

- Придумать лучшую криптографическую хеш-функцию.



Идея

- Известно, что система алгебраических уравнений высокой степени от нескольких переменных алгоритмически неразрешима. Для вычисления хеша было решено использовать этот факт.
- Мы хотим получить большой хеш (512 бит). Так как работать с такими большими числами в наших условиях невозможно, решили работать с векторами высоты $\mathrm{n} = 64$, компоненты которого – числа из $\mathbb{Z}_{PRIME}, \mathrm{PRIME} = 257$.

Формула (1)

- Первый (самый простой) вариант таков: отождествим символы текста с их кодами ASCII, увеличенными на 1 (чтобы избежать умножения на 0 и затруднить работу злоумышленнику). Выберем невырожденную матрицу $n \times n$ и будем использовать ее столбцы для хеширования.

$$\langle h \rangle = \left[\langle v_1 \rangle \cdot s_1^2 + \langle v_2 \rangle \cdot s_1 + \ldots + \left\langle v_{(2l-1)\%n} \right\rangle \cdot s_l^2 + \left\langle v_{(2l)\%n} \right\rangle \cdot s_l\right] \cdot finalmultiplier,$$

где $finalmultiplier = \sum_{i=0}^{i \le \log_2 l} s_{2^i}^3$.

- У этой формулы есть несколько минусов: простой вид (выглядит ненадежно), «постоянство» векторов.

Формула (2-1)

- Второй (усложненный) вариант.

1. Заводим буфер из трех элементов:

$$buff[0] = s_l, \qquad buff[1] = s_l^2, \qquad buff[2] = s_l^3$$

Буфер строится по последнему символу текста.

2. Создаем первые три вектора $\langle v_1 \rangle, \langle v_2 \rangle, \langle v_3 \rangle$ высоты n.

$$\langle v_1 \rangle_i = \left[(s_l \% 2) \cdot 2 + (s_l \% 3) \cdot 3 + \ldots + (s_l \% 8) \cdot 8 + (s_l \cdot p_i^3)^2\right] \cdot p_i,$$

$$\langle v_2 \rangle_i = \langle v_1 \rangle_i{}^2, \langle v_3 \rangle_i = \langle v_1 \rangle_i{}^3,$$

где p_i - i-е простое число (во избежание нулей 257 заменили на (n+1)-е простое число).

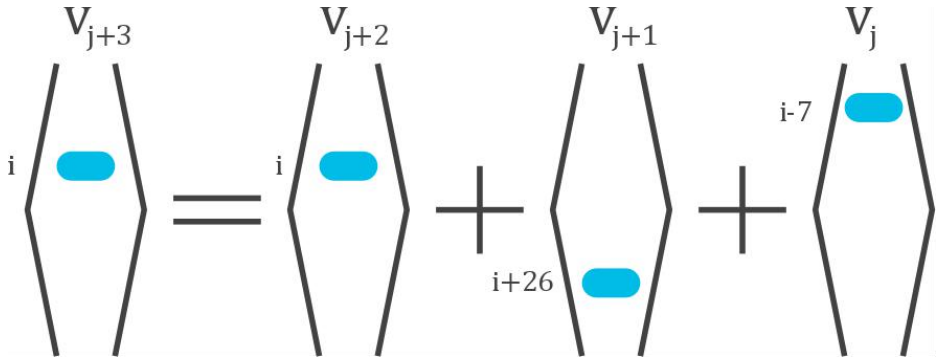
Формула (2-2)

3. Формула для получения хеша:

$$\langle h \rangle = \langle v_1 \rangle \cdot s_1^2 - \langle v_2 \rangle \cdot s_1^3 + \langle v_3 \rangle \cdot buff[0] \cdot s_1 + \langle v_4 \rangle \cdot s_2^2 - \langle v_5 \rangle \cdot s_2^3 + \langle v_6 \rangle \cdot s_3^2 - \langle v_7 \rangle \cdot s_3^3 + \langle v_8 \rangle \cdot s_4^2 - \langle v_9 \rangle \cdot s_4^3 + \langle v_{10} \rangle \cdot buff[0] \cdot s_4 + \ldots$$

На каждом (1+3i)-м символе добавляется слагаемое $\langle v_j \rangle \cdot buff[0] \cdot s_{1+3i}$, и на каждом (100k+t)-м символе происходит обновление буфера: $buff[0] = buff[1],\ buff[1] = buff[2],\ buff[2] = s_{100k+t},\ t \in \{16, 25, 36, \ldots, 100\}$.

4. При этом векторы получаем так:



Тестирование на коллизии II рода

- Для выявления коллизий хешировали каждую строку входного текста по отдельности. Тестировали на нескольких файлах: самый большой роман в мире (на французском), статистика игр НБА сезона 2011/12, сгенерированная подборка символов (1 млн строк по 100 символов).
- Проверка на коллизии была по модулю 2^{32}.

Файл	NSU	RS	JS	PJW	BKDR	SDBM	DJB	DEK	AP
Роман	13	126	130	738	128	126	128	129	126
НБА	20	21	40	2425	15	22	23	19	21
Тест	93	122	126	1966	130	114	117	114	118

Тестирование на лавинный эффект

- Тестирование проводили вручную, изменяя в заданном тексте один символ (в начале, середине или конце). Результаты показали, что лавинный эффект действительно наблюдается в большинстве случаев. Для более детального исследования необходимо больше времени и ресурсов.

Тестовая фраза	213	181	239	180	87	138	159	46	163	138	105	161	111	97	247	69	160	180	23	240	53	11	120	213	2	176	224	178	162	233	123	7	254	131	212	163	210	229	196	252	206	143	117	56	77	1	132	254	120	184	129	214	190	164	183	72	78	193	248	215	168	105	83	15
Тестовая ваза	124	225	243	166	81	40	250	10	69	250	59	17	205	217	79	12	44	15	242	112	119	217	12	255	199	43	155	253	96	229	121	0	38	49	14	226	59	134	169	256	41	207	248	203	134	244	95	227	254	205	39	150	233	204	220	30	94	63	56	236	83	193	191	87
Таатовая фраза	186	36	204	182	82	43	107	200	80	101	143	23	7	198	15	178	146	3	58	36	81	251	114	247	110	192	19	26	156	23	147	105	225	82	58	120	13	240	169	184	193	220	64	198	1	251	84	148	53	209	194	180	47	79	44	200	34	57	241	105	124	54	256	241
Тесаовая фраза	233	153	222	247	170	5	222	151	213	59	20	190	144	157	102	107	153	227	68	148	140	6	108	130	77	65	177	63	196	87	186	120	147	197	89	136	178	67	194	110	104	40	159	153	10	247	73	143	151	217	127	64	249	158	45	26	94	11	123	255	94	222	212	122
Теставая фраза	80	187	148	82	115	242	67	13	96	140	7	175	0	188	13	207	65	139	187	244	252	227	89	170	148	144	121	6	167	171	68	70	52	69	49	15	160	196	115	177	139	81	231	49	0	119	121	191	64	74	62	202	233	47	136	23	70	216	254	63	40	53	37	35
Тестоаая граза	158	1	95	155	135	154	149	48	74	32	228	160	167	71	79	191	82	11	165	137	101	217	254	111	65	127	109	120	178	244	29	58	200	29	80	240	236	71	215	67	25	36	171	180	193	100	91	176	203	75	136	176	162	160	182	200	151	135	13	146	133	16	78	78
Тостовая фраза	62	195	178	254	101	121	103	110	208	23	176	229	178	75	53	122	198	154	115	232	120	159	58	138	53	111	182	24	208	212	146	152	243	183	63	67	58	250	184	154	19	65	253	48	58	123	122	176	255	44	167	157	80	236	62	137	47	61	236	39	176	33	74	170
Тесто, я фраза	252	81	190	8	64	125	189	235	55	222	31	122	57	243	128	75	235	195	234	64	119	178	256	253	164	106	82	117	76	89	138	177	108	125	181	205	108	35	143	200	92	187	13	208	165	54	198	142	115	109	251	196	158	140	207	89	255	92	34	112	95	217	180	187
Тестовая фаза	14	211	177	4	141	104	128	215	8	219	187	134	52	176	33	120	160	131	248	171	191	179	229	221	98	37	164	62	173	77	89	109	70	213	152	178	163	7	118	53	162	145	59	27	144	148	218	163	221	225	66	7	54	22	148	136	165	232	178	148	175	108	190	235
Текстовая фраза	181	189	115	130	26	143	126	177	31	193	191	191	163	11	242	184	245	154	28	86	183	117	133	114	16	108	70	84	30	150	86	140	79	72	23	30	113	123	202	113	179	128	129	165	75	86	249	173	208	210	63	160	217	213	157	197	254	103	178	157	157	35	150	104
Тестовая фразы	92	17	131	144	202	203	236	68	204	155	181	130	177	105	161	195	8	48	159	129	130	6	251	101	62	66	8	93	72	71	85	88	138	58	30	187	74	120	173	132	63	60	30	192	117	87	188	156	141	96	109	245	185	231	41	207	81	182	184	247	13	177	8	254
нестовая фраза	56	206	200	218	249	34	189	205	35	113	96	80	14	249	246	140	61	146	27	235	175	19	19	80	146	246	1	17	124	4	157	175	99	243	159	124	174	191	201	151	144	110	153	106	152	185	112	37	116	17	164	83	214	117	178	55	211	88	133	2	123	0	199	16
Шестовая фраза	9	90	63	100	23	175	246	5	172	184	6	40	200	83	113	122	230	159	74	1	29	141	57	159	72	42	82	158	242	123	59	63	3	183	32	132	239	147	5	256	173	106	243	221	99	106	11	180	103	38	99	161	27	66	225	212	57	43	4	108	197	118	176	74
Тактовая фраза	76	202	190	2	14	72	170	235	34	220	192	158	87	11	198	203	238	123	113	243	70	248	130	103	5	140	38	7	51	44	41	234	206	30	233	156	161	87	27	217	162	124	184	79	29	124	173	42	0	233	7	225	176	26	125	222	85	135	204	144	24	26	175	34
Тестовая база	76	177	97	223	188	235	223	145	178	156	19	180	179	136	98	22	24	169	212	29	32	66	149	76	246	216	129	16	114	254	61	30	105	113	139	238	44	118	183	38	114	110	22	76	107	13	163	247	6	200	48	197	24	170	195	194	103	105	224	220	133	148	248	102

Тестирование на время (512 бит)

Размер файла (МБ)	Условие	Время (с)
500	Только чтение файла	21
500	Без вычисления векторов	265
500	Полная программа	588

Итоги работы

- Экспериментально было выяснено, что выбранная концепция действительно работает, однако требует больших затрат как по времени, так и по памяти. Это требует дополнительной проработки. Формула также требует дополнительных исследований.
- Планируется продолжать работу по усовершенствованию нашей хеш-функции.

Наши контакты

Daniil Vodolazsky
daniil.vodolazsky@mail.ru

Ruslan Izhitskiy
ksilobait@gmail.com

Ivan Milyaev
ivan.milyaev@mail.ru

Sergey F. Krendelev
s.f.krendelev@gmail.com