Стеганографические технологии — новое направление защиты информации

Ключевые слова: стеганографические технологии, защита информации, криптография.

На фоне развития угроз рассматривается новое направление защиты информации — косвенная стеганография, ее исторические прототипы, связь с клептографией, пост-квантовой криптографией и некриптографическими подходами.

Голубев Е.А.,
д.т.н., профессор МТУСИ

В последние годы в информационном сообществе проявились тенденции развития угроз информационной безопасности. Для заблаговременного выявления и предупреждения потенциальных угроз, которые можно ожидать от разрушающих программных воздействий (РПВ) и новых вычислительных сред в будущем, специалистам по информационной безопасности следует находиться в курсе современных тенденций в развитии вредоносных ПО и изобретать адекватные средства противодействия им. Приведем несколько примеров.

За рубежом в области открытого научного изучения РПВ, использующих достижения современной криптографии, стали А. Янг и М. Юнг, начавшие с 1996 г. публиковать статьи и выступать с докладами по новому научному направлению, которое они назвали клептографией. Основная идея, заложенная авторами в понятие клептографии, — использование криптографических методов против самой криптографии, т.е. применение криптографических, криптоаналитичесих и стеганографических методов для затруднения или исключения фактов обнаружения успешных атак на защищенные ресурсы [1, 2].

С середины 80-х годов началось исследование вычислительных устройств, подчиняющихся законам квантовой механики — квантовых компьютеров. Интерес к квантовым компьютерам возрос, в частности, в связи с потенциальной угрозой информационной безопасности используемых на практике криптографических систем с открытым ключом. Исследования ведутся в двух основных направлениях. Первое из них связано с поиском, кроме задач факторизации числа и нахождения дискретного логарифма, других вычислительных задач, решение которых может быть ускорено с помощью квантовых компьютеров. Второе направление — так называемая пост-квантовая криптография.

Термин "пост-квантовая криптография" был предложен Д. Бернштейном и уже стал общепринятым в криптографической литературе. Он обозначает ту часть криптографии, которая выживает и в случае появления квантовых компьютеров и квантовых атак. Начиная с 2006 г., проводятся международные конференции PQCrypto, посвященные пост-квантовой криптографии. [3, 4].

Таким образом, современные меры защиты информации, включая криптографические, становятся потенциально уязвимыми с позиции тенденций развития угроз информационной безопасности. Необходимо придумывать новые способы защиты информации и корректировать парадигму защиты информации. Одно из определений этого философского термина — стиль или тип научного исследования, фиксирующий изменения в структуре знаний.

Способы скрывать сообщения для их защиты от посторонних изобретали давно. Историки математики обнаружили на греческом языке трактат 15 века, посвященный стеганографии, в переводе на русский — тайному письму. Описан прием нанесения шифрованного сообщения на бритую голову раба. После отрастания волос его посылали через территорию врага в надежде, что там не будут брить головы всем путникам. Таким образом организовался скрытый канал связи.

В настоящее время подобные подходы называются "прятанием по углам" и реализуются "искусством ремесла" — state of the art.

Современная стеганография — научная дисциплина, разрабатывающая методы скрытия факта передачи или хранения секретной информации. Существует деление информации по уровню секретности, конфиденциальности. Признаками секретной информации является наличие, во-первых, законных пользователей, которые имеют право владеть этой информацией, во-вторых, незаконных пользователей (нарушителей, противников), которые стремятся овладеть этой информацией с тем, чтобы обратить ее к себе во благо, а законным пользователям во вред. Для наиболее типичных ситуаций введены специальные понятия: государственная тайна, военная тайна, коммерческая тайна, юридическая тайна, врачебная тайна и т.д. до личной тайны. Таким образом любая тайна, технически фиксируемая в секретной информации, требует адекватных ее ценности комплексных мер обеспечения безопасности информации. Эти рассуждения содержаться во Введении книги "Стохастические методы и средства защиты информации в компьютерных системах и сетях" под редакцией И.Ю. Жукова [5].

Начала современной стеганографии, как принято считать, заложил G.J.Simmons.

Впервые в открытой зарубежной научной литературе модель стеганографического канала связи была описана Симмонсом в работе, представленной на конференцию Crypto'83 [6], как проблема двух заключенных. Двое заключенных, Алиса и Боб, находящиеся в различных тюремных камерах, могут обмениваться посланиями под контролем надзирателя Уэнди. Задача — включить в послания секретную, ценную только для них, информацию и при этом скрыть факт ее присутствия в послании от Уэнди, т.е. организовать скрытый (subliminae, covert, sade channels) канал связи.

Развитие информационных технологий, растущие предложения на рынке цифровой техники, программного обеспечения, различных услуг открыли возможности на рубеже 1990 года любителям разрабатывать программы стеганографии — маскировки или сокрытия секретных сообщений в мультимедийные файлы и программы обнаружения факта сокрытия — стегодетекторы. Проведенная в США в 1996 г. первая международная конференция по скрытию данных [7] ввела терминологическую базу в современную цифровую стеганографию. Звуковые и визуально воспринимаемые произведения в цифровой мультимедийной форме названы контейнерами, а содержащие секретную информацию — стего-контейнерами и т.п.. Разработаны психофизиологические модели восприятия звука и изображения, на основе которых рекомендованы алгоритмы сжатия и разработаны соответствую-

щие форматы сжатия. В последующие годы опубликовано огромное количество научных работ, отражающих динамику развития этой научной дисциплины.

Рассмотрим относительно новое направление — косвенная стеганография, отличающаяся от цифровой стеганографии выбором и формированием стегоконтейнера.

В работе Н. Алишова "Косвенная стеганография" [8] описывается оригинальный метод шифрования и дешифрования на основе способа, называемого косвенной стеганографией.

Суть метода заключается в следующем. У отправителя и получателя одинаковые файлы, которые по взаимной договоренности являются секретными ключами. Байты информации, подлежащей защите, заменяются (по определенному алгоритму) байтами, формируемыми из секретного файла. Новый файл передается адресату и при получении подвергается обратному преобразованию: его байты заменяются байтами секретного файла (зеркальный алгоритм)".

Попробуем поискать в истории криптографии прототип предлагаемого метода.

Откроем книгу Ярослава Гашека "Похождения бравого солдата Швейка во время мировой войны" (Государственное издательство художественной литературы. Москва, 1957г.) на странице 463.

" ... Большинство офицеров углубилось в чтение небольшой книжки, озаглавленной "Die S?nden der V?ter" ("Грехи отцов". Роман Людвига Ганггофера). Все одновременно сосредоточенно изучали страницу сто шестьдесят первую... Капитан Сагнер: перед нами совершенно секретная информация, касающаяся новой системы шифровки полевых депеш. Именно сто шестьдесят первая страница романа является ключом новой шифровальной системы, введенной согласно новому распоряжению штаба армейского корпуса. ... Новая система необычно проста. Если нам, например, должны будут передать приказ (текст приказа), то мы получим следующую депешу (набор слов со стр. 160) и по нему находим буквы на стр. 161, из которых складывается текст приказа. Это исключительно просто. Из штаба по телефону в батальон, из батальона по телефону в роту... Кадет Биглер: обратите внимание на книгу Кериктофа о военной шифровке. Там подробно описывается метод, который вы нам только что объяснили. Изобретателем этого метода является полковник Кирхнер, служивший при Наполеоне Первом в саксонских войсках. Метод был усовершенствован поручиком Флейснером в его книге "Handbuch der milit?rischen Kryptographie". Тот же самый пример, который мы все сейчас слышали... Этот метод называется методом шифровки словами..."

Что общего в методах Н. Алишова и поручика Флейснера? "Определенным алгоритмом" в косвенной стеганографии является поиск байтового кода клавиши клавиатуры отправителя в файле-ключе, т.е. в теле программы-контейнера, с целью определения его адреса, т.е. битового расстояния от начала файла или другой оговоренной позиции. Этот адрес является взаимно однозначным фоменологическим отображением одного символа или буквы скрываемой информации. Последовательность адресов, представленная в байтовом виде, и является стегоконтейнером в терминологии косвенной стеганографии. Пока противнику не известно, какое же тело программы использовано, восстановить секретное сообщение из стеконтейнера не возможно. (рис. 1)

В [5] в подразделе 15.5.2 — стеганографическая защита исполняемого кода — дается следующая интерпретация "метода шифровки словами". "В теле программы-контейнера, содержащей достаточно большое число кодов различных команд, можно спрятать практически произвольное количество кодов других программ. При этом для извлечения соответствующей скрытой программы требуется лишь задать нужную последовательность адресов, по которым располагаются коды команд конкретной программы". Этот метод содержится в разделе 15.5 — идея стохастической вычислительной машины.

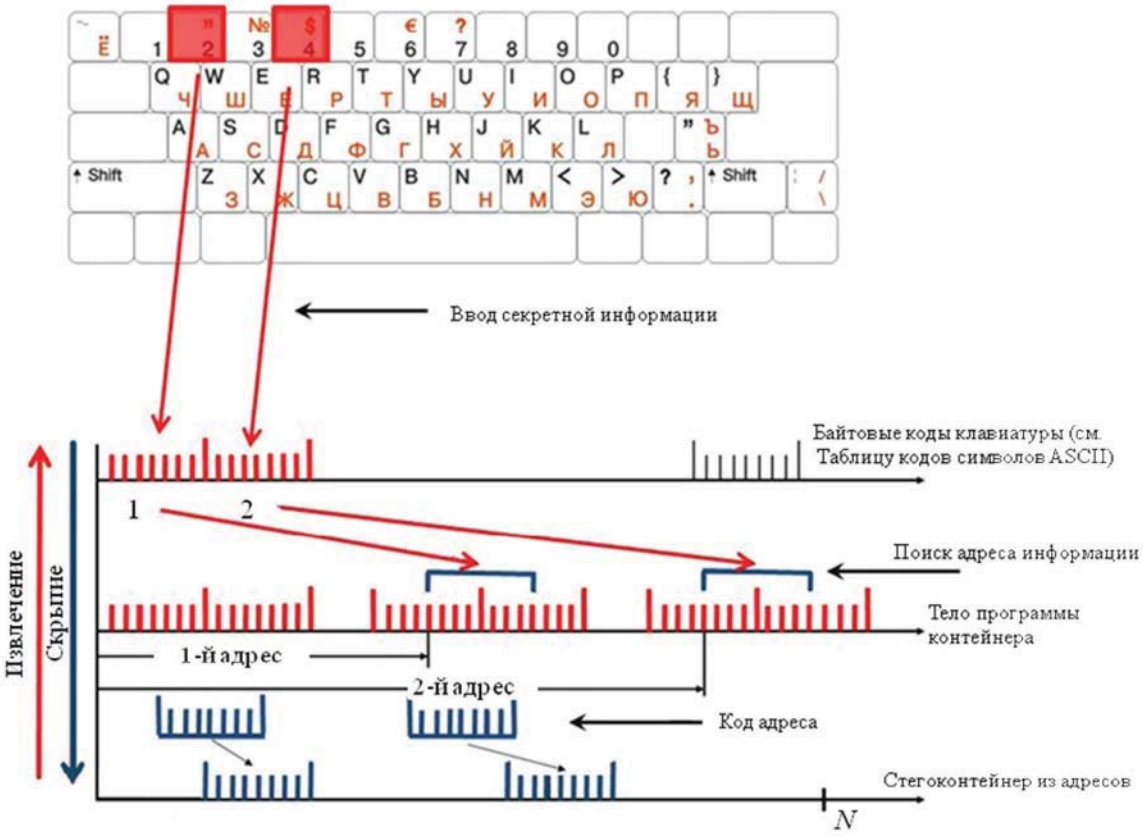

Рис. 1. Алгоритм формирования стегоконтейнера из тела программы

В отличие от компьютерной стеганографии, использующей медийные форматы для скрытия факта наличия сообщений, и обеспечивающей тем самым неизвлекаемость, т.е. защиту от доступа, косвенная (или книжная) стеганография преобразует защищаемую байтовую последовательность в нечитаемую байтовую последовательность, не скрывая ее наличия, т.е. похожа по результатам преобразования информации на криптографию.

Таким образом, сам метод имеет вековую историю, а его реализация с использованием современных информационных технологий заслуживает внимания. Проблемой в оценке стойкости косвенной стеганографии является формализация "договоренностей" при построении и управлении ключевой информацией и составление перечня угроз, реализация которых нарушает стойкость метода от взлома.

По-видимому, косвенную стеганографию можно отнести к перспективным некриптографическим или нестеганографическим методам защиты информации. Возможно их развитие в рамках PQCrypto и клептографии.

Следуя терминологии [5], для исключения путаницы целесообразно косвенную стеганографию отнести к стохастическим методам технической защиты информации, основанным на применении генераторов псевдослучайных чисел, сочетающим в себе эффективную реализацию и высокую крипто- и имитостойкость.

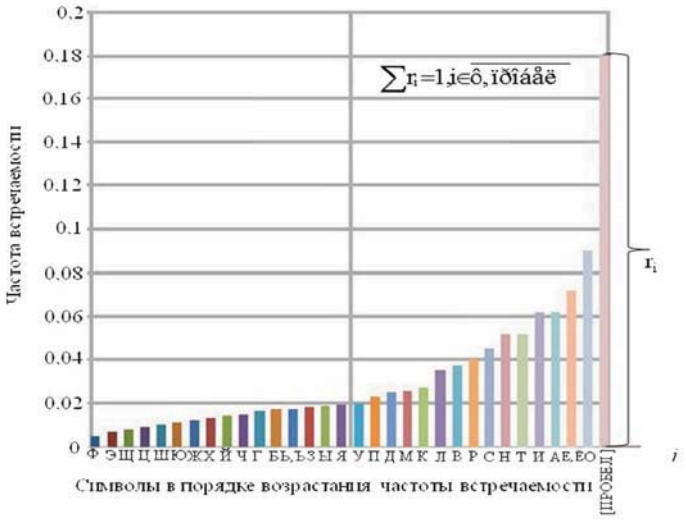

Рис. 2. Гистограмма нормированных значений встречаемости символов русского алфавита в художественных произведениях

По вертикали — количество символов в файлах-контейнерах, произвольно выбранных из Windows XP

символы русского языка	размер файла – контейнера												
	2 кб	4 кб	8 кб	17 кб	36 кб	72 кб	155 кб	308 кб	750 кб	1.53 мб	3.67 мб	5.76 мб	10.3 мб
_	98	396	493	1670	3335	2732	7799	8631	22390	129314	235583	224533	490440
а	152	553	1059	2634	4251	10645	17275	36305	37015	208832	489283	649321	1385992
б	15	33	293	796	432	5118	10211	23266	16557	102212	308347	320254	792037
в	15	67	384	766	680	9025	15247	6110	22571	136764	626128	434101	1073706
г	3	20	206	446	422	3496	7215	9714	14712	58720	170056	255459	632586
д	23	126	433	1104	1052	3723	14396	5790	13211	70946	255595	359973	667083
е	2	21	171	424	663	5008	8484	3688	12392	89473	391449	273434	693369
ё	22	106	192	756	489	2618	4488	4231	11481	36586	118066	148672	417128
ж	16	103	390	1218	1686	4893	5074	4620	8020	69881	251167	216651	511018
з	1	9	107	335	375	2348	4253	3594	10232	28459	98120	180970	494090
и	32	171	492	1390	1135	3563	14197	7559	14007	84097	157306	373171	702207
й	11	52	159	628	589	1629	4550	3927	5204	39001	71528	176954	388449
к	16	110	294	1040	776	3123	4297	5311	7629	43489	81127	209347	499449
л	1	8	108	279	668	1406	2623	4432	4823	27839	57391	145948	306037
м	9	67	243	819	2216	5101	4843	5032	9464	70868	243383	227128	522104
н	11	49	223	634	495	1420	2600	3827	4667	25759	47641	163224	321991
о	8	27	128	562	556	2484	3442	225131	5793	35615	79802	175498	518409
п	0	3	47	216	350	717	1701	4120	3543	13613	37447	145550	387098
р	63	128	587	1288	2148	9259	16787	19876	27601	184047	425189	619999	1393232
с	6	78	344	669	725	8741	15828	6178	23537	150456	596036	480012	1221897
т	17	89	290	923	1146	1846	11748	6962	7858	129732	119757	362687	576304
у	7	17	94	411	1676	3864	4405	6323	5846	62430	221998	229928	499271
ф	15	33	210	524	772	4459	9707	9451	12757	92811	192860	342922	828290
х	4	23	144	462	977	3934	5769	8664	8970	43497	108782	287104	642984
ц	6	22	212	544	739	2578	4735	5936	9088	51350	98765	276137	534204
ч	6	17	118	550	706	1814	3344	12120	7591	40102	68476	250121	645575
ш	43	95	678	1121	2145	15805	27241	13899	41457	265824	911891	919966	2082467
щ	10	30	206	773	2191	5094	13049	10595	11516	100667	304169	504451	872410
ъ	17	52	319	906	1434	4952	11090	15375	14611	107042	180111	499797	1037886
ы	11	23	222	745	1158	3630	6732	15784	11891	70043	125724	472542	1046111
ь	60	101	772	1739	4039	19667	38638	24788	48480	349658	1155997	1349041	2706771
э	16	33	364	1176	2315	8736	17630	249030	24353	166001	290016	952224	2015693
ю	70	108	1029	2816	6196	26711	54655	51704	68952	492803	1404953	2187103	4523889
я	216	152	3864	4463	49223	94997	188316	163441	205405	1492958	3151167	9542656	12556718
А	0	130	1422	6244	6178	15611	38244	49320	64367	395722	886689	1283046	2802514
Б	-	3,1	5,7	2,8	6,0	4,8	4,0	6,3	11,7	3,8	4,1	4,5	3,7

Проведенные оценочные эксперименты относительной вместимости метода косвенной стеганографии дали следующие результаты.

В методе "шифровки словами" одна буква (на стр. 161) скрываемого сообщения отображается в одно слово (на стр. 160). Допустим, в среднем одно слово содержит 5 букв. Следовательно избыточность составит 5:1 или вместимость — 0,2.

Для оценки вместимости или требуемой избыточности метода "поиска адресов, по которым располагаются коды ASCII клавиатуры в теле программы — контейнера" [5] воспользуемся для примера широко известной гистограммой встречаемости символов алфавита русского языка в художественной литературе (см. рис. 2) и условием, что каждое использование буквы или кода клавиши в скрываемом тексте имеет индивидуальный адрес в теле программы-контейнера.

В рамках исследования в качестве контейнеров были изучены системные библиотеки операционной системы MS Windows XP. Для проведения эксперимента были взяты файлы *.dll размером 2, 4, 8, 17, 36, 72, 155, 308, 750 килобайт, а также 1.53, 3.67, 5.76 и 10.3 мегабайт. Количество встреченных в них символов алфавита представлено в столбцах таблицы. В строке А таблицы указан приблизительный размер сообщения, которое можно отобразить с использованием данного контейнера с учетом средней частоты встречаемости каждого символа алфавита. В строке Б — относительная избыточность файла-контейнера. С учетом того, что адреса в "теле программы" могут по-

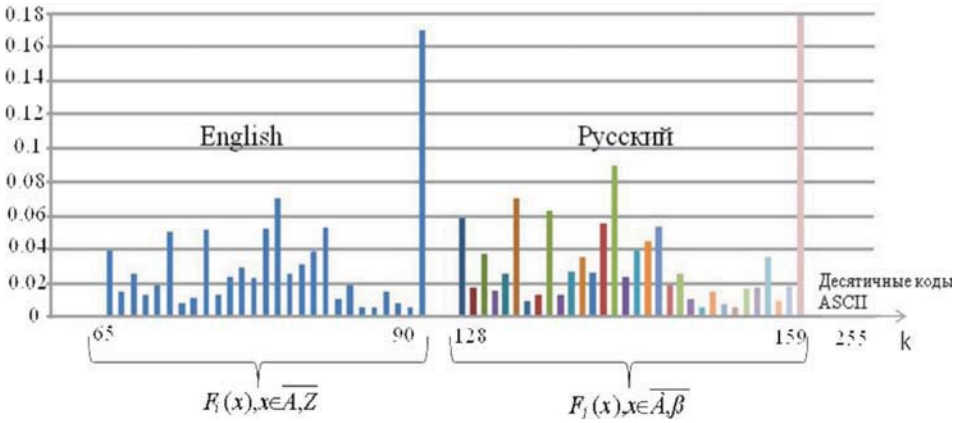

Рис. 3 . Иллюстрация гистограммного метода различия фрагментов ПО

требовать для их байтового представления нескольких байт, числа в строке Б необходимо умножать на длину адреса в байтах.

Избыточность обусловлена рассогласованностью гистограммы встречаемости символов алфавита русского языка с соответствующей статистикой символов исследуемых программ-контейнеров. Это наводит на мысль использования этих статистик для распознавания фрагментов СПО или ОПО. (рис. 2).

Каждому фрагменту ПО можно поставить в соответствие его гистограмму встречаемости в процентном соотношении кодов клавиатуры в соответствии с таблицей кодов символов ASCII. (рис. 3).

Для анализа дискриминационных свойств гистограмм фрагментов ПО и описания процесса их различия обозначим каждую гистограмму фрагмента $По_i$ как функцию $F_i(x)$ — кусочно линейную, интегрируемую в квадрате, нормированную и определенную на x — последовательности десятичных номеров кодов ASCII от 0 до 255; i — пробегает все множество фрагментов ПО (пример — рис. 3).

Такое пространство функций является бесконечномерным действительным унитарным пространством (или гильбертовым пространством). В нем определены для любой пары функций:

скалярное произведение

$$(f(x),h(x)) = \int_a^b f(x)\cdot h(x)\,dx,$$

метрика, расстояние

$$d(f,h) = \sqrt{\int_a^b (f(x)-h(x))^2\,dx},$$

норма функции

$$\|f(x)\| = \sqrt{\int_a^b f^2(x)dx},$$

и ее нормировка

$$\mathcal{F}_i = \|f(x)\|^{-1}\cdot f(x) = 1.$$

Далее будем рассматривать нормированные F_i. Расстояние между двумя нормированными гистограммами $По_i$ и $По_j$ определяется как

$$d(\mathcal{F}_i,\mathcal{F}_j) = \sqrt{\int_a^b (\mathcal{F}_i(x)-\mathcal{F}_j(x))^2\,dx},$$

Для функций в виде гистограммы естественно применять Σ вместо $\int$, а расстояние $d(f_i, f_j)$ вычислять как сумму модулей разностей одноименных значений встречаемости каждого номера символов ASCII.

$$d_{i,j}(F_i(x),F_j(x)) = \sum_k |r_{i,k}-r_{j,k}|, \qquad k \in \overline{0,255}$$

Этот небольшой экскурс в высшую математику позволяет поставить задачу по диссертационным исследованиям чувствительности метода сравнения гистограмм фрагментов ПО к обнаружению вредоносных программ. Главное, корректно определить первый разряд в битовой последовательности фрагмента ПО и с требуемой точностью использовать арифметические операции при обработке процентов встречаемости кодов клавиатуры в гистограмме этого фрагмента ПО. Необходимо создать библиотеку фрагментов ПО и пополнять ее новыми легальными ПО. Нелегальное или вредоносное ПО не будет содержаться в библиотеке. Гистограммное описание ПО любого размера требует до 1 кбайта памяти.

В [8] приведен способ искусственного формирования файла-контейнера — тела программы с целью исключения избыточности и формирования нового стегоконтейнера с адресами такого же размера, как и исходный файл сообщения с кодами клавиатуры, т.е. с избыточностью 1:1. (рис. 4.) Суть метода состоит в формировании таблицы, каждая строка которой состоит из 256 байтов кодов символов ASCII клавиатуры. Строки формируются генератором псевдослучайных перестановок (ГПСП) и все различны. Количество строк в таблице — теле программы-контейнера — определяется необходимым объемом в байтах (N) защищаемого сообщения. Теоретически может быть сформировано 256! различных строк. Ключом в этом варианте является начальное заполнение генератора псевдослучайных чисел (ГПСЧ), каждое срабатывание которого при нажатии клавиши клавиатуры выдает номер строки. Номер требуемой позиции в строке в виде байта является взаимно однозначным отображением очередного символа защищаемого сообщения. Таким образом формируются файл-стегоконтейнер, равный по объему файлу защищаемого сообщения.

Для извлечения из стегоконтейнера — файла с адресами секретной информации получатель должен сформировать такую же таблицу, с помощью ГПСЧ последовательно выбирать строку, а в строке находить номер позиции — кода клавиши и последовательно формировать сообщение.

Этот способ позволяет защитить свои ресурсы или передать преобразованное секретное сообщение как криптограмму, или в стеганографически замаскированном виде.

Ценным является то, что сам пользователь может создать свою систему надежной защиты

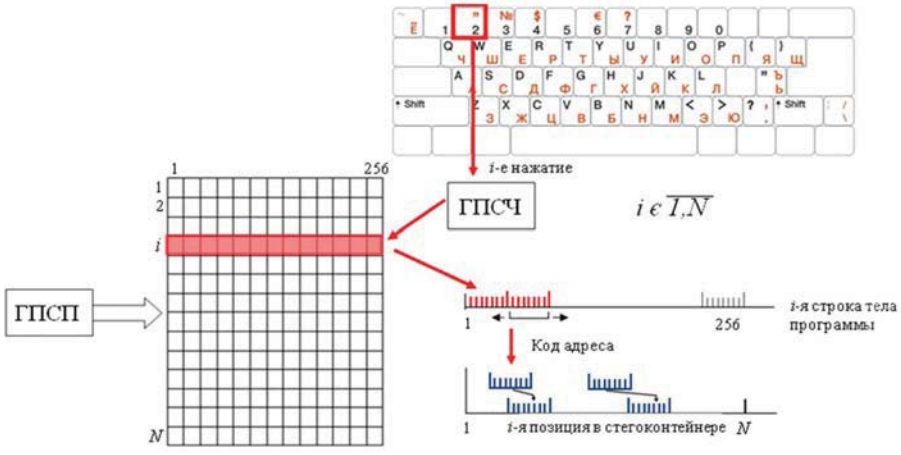

Рис. 4. Алгоритм синтеза тела программы-контейнера и формирования стегоконтейнера

своей информации и хранить ключевую информацию в отчужденном виде.

Для защиты данных начального заполнения ГПСЧ и ГПСП (ключа — по аналогии с криптографической защитой) имеется масса вариантов в том числе "ратворения" в повседневной жизни информационной среды. Привлекательной стороной является ясность и кажущаяся простота. Поэтому безопасности и бескомпроматности этого процесса должно быть уделено особое внимание путем сочетания формальных доказательств с "искусством ремесла". Основную угрозы следует ожидать от Шерлока Холмса — творение писателя Артура Конан Дойла. Злоумышленники организовали канал скрытой связи путем размещения в публичных местах рисунков "пляшущих человечков". По их замыслу символы будут восприниматься как детские рисунки. Они совместили в одном месте метод организации скрытого канала, стеготконтейнер и ключ к нему, что позволило Холмсу с помощью дедуктивного метода читать их переписку.

Особенностью термина "косвенная стеганография" является то, что он не соответствует главной задаче цифровой стеганографии — скрытию факта присутствия защищаемой информации в файле-контейнере.

Однако это не снижает актуальности детальной и системной разработки этого нового направления технической защиты информации стохастическими методами. Оно может быть конкурентно способно по отношению к блочному или поточному шифрованию на уровнях тайны, ниже государственной тайны.

Является ли стеганографические технологии, вытекающие из сформированной Соммонсом задачи, получившей название "проблема двух заключенных", новой парадигмой защиты информации?

Положение, представленные К. Шенноном в работе "Теория связи в секретных системах" [9] привели к изменениям в структуре знаний в области защиты информации, т.е. изменению парадигмы в области защиты тайны и секретов от посторонних. Процитируем:

"Наше изложение будет ограничено в нескольких отношениях.

Имеются три общие типа секретных систем:

1) система маскировки, которые включают применение таких методов, как невидимые чернила, представление сообщения в форме безобидного текста или маскировки криптограммы, и другие методы, при помощи которых факт наличия сообщения скрывается от противника;

2) тайные системы (например, инвертирование речи), в которых для раскрытия сообщения требуется специальное оборудование;

3) "собственно" секретные системы, где смысл сообщения скрывается при помощи шифра, кода и т.д., но само существование сообщения не скрывается и предполагается, что противник обладает специальным оборудованием, необходимым для перехвата и записи переданных сигналов. Здесь будет рассмотрен только третий тип систем, так как системы маскировки представляют в основном психологическую проблему, а тайные системы — техническую проблему".

Таким образом, К. Шеннон выделил три направления или три метода защиты информации: психологический, технический и криптографический. Технический метод, реализованный например в аппаратуре ЗАС, отмирает естественным образом.

Криптография утвердилась как важнейшее средство защиты информации и обеспечения государственной безопасности, но решает лишь часть задач проблемы ОБИ в современных условиях и в перспективе. У стеганографии (цифровая и косвенная стеганография, стеганоанализ, цифровые водяные знаки и др.) более широкое распространение в информационном сообществе для достижения различных целей. Она развивается по мере развития наиболее востребованных медийных технологий, аппаратной базы и услуг, которые доступны каждому субъекту информационного сообщества. Стахостическая или косвенная стеганография вбирает в себя первый и третий из общих типов секретных систем" — "искусства ремесла" и стахостических методов защиты информации. Является ли стохастическая стеганография основой новой парадигмы? Из парадигмы должны следовать серьезные государственные меры.

Литература

1. **Young A. L, Yung M. M.,** Kleptography: Using cryptography against cryptography || Advances in Cryptology-Evrocrypt`97, Springer — Kerlag, Lecture Notes in Computer Science No 1233, 1997

2. **Young A, Yung M.,** Malicious cryptography exponsing cryptovirology. Wiley Publishing, Inc., 2004.

3. Post-Quantum Cryptography, Springer, 2009.

4. Архив на сайте IACR.

5. **Иванов М.А., Ковалев А.В., Мацук Н.А., Михайлов Д.М., Гугунков И.В.** / Под ред. Жукова И.Ю., Стохастические методы и средства защиты информации в компьютерных системах и сетях. М., 2009.

6. **Simmons G.J.,** The prisoners problem and the subliminal channel, Proc. Workshop on Communications Security (Crypto'83), 1984, 51-67.

7. **Anderson R.** editor.//Proc. Jnt. Workshop on Information Hiding: Lecture Notes in Computer Science. Springer-Verlag, Cambridge, 1996.

8. **Алишов Н.** Косвенная стеганография. — International Book Series "Information Science and Computing" KDS 2009, Vama, Bulgaria, 2009. pp. 53-57.

9. **Shenon C.** "Communication theory of secrecy system", Bell System Techn. J. 28, №4 (1949) 656-715 или в книге: К. Шеннон. Работы по теории информации и кибернетике. Статья "Теория связи в секретных системах" ИИЛ, Москва, 1963 г.

10. **Голубев Е.А., Емельянов Г.В.** Стеганография как одно из направлений обеспечения информационной безопасности. — T-Comm "Технологии информационного общества". — М., 2009.

Steganography technologies — new area of information security

Golubev E.A., *Prof. MTUCI*

Abstract

Against the background of threats is considered a new trend of information security — an indirect steganography, its historical prototypes, communication with kleptografiey, post-quantum cryptography nekriptograficheskimi and approaches.

Keywords: *technology, steganography, cryptography, information security.*