

МЕТОД ОЦЕНКИ ОБЪЕМА ИНФОРМАЦИИ, СКРЫВАЕМОЙ В ПОТОКАХ ТРАФИКА РАЗЛИЧНЫХ КЛАССОВ

Современные мультисервисные телекоммуникационные системы (ТКС) являются разнородными объединенными сетями с динамически меняющейся структурой, способными передавать различные типы и классы трафика с заданным уровнем качества обслуживания [15]. Одним из важнейших свойств такого типа сетей является их защищенность от негативных внешних влияний и попыток несанкционированного доступа к ресурсам [1]. Проблема обеспечения информационной безопасности выходит на первый план и должна решаться с привлечением традиционного подхода: обеспечения процедур аутентификации, идентификации, авторизации и аудита, гарантии доступности, конфиденциальности и целостности, а также модификаций этих процедур с учетом новых технологий, таких как стеганографическая защита информации [16].

Стеганографическая защита для телекоммуникаций существует менее десятка лет и ее результатами является наличие средств скрытой передачи информации для отдельных сред, таких как аудиопотоки, видеопотоки (ограниченные возможности внедрения/детектирования), статические изображения различного формата, а также электронные документы офисных программ [11]. Сравнивая возможности/требования современных ТКС [15] и возможности систем скрытой передачи данных [10], можно сделать выводы, что:

- средства скрытой передачи данных не являются универсальными в применении к современным сетям передачи данных;
- эти средства представлены отдельными продуктами (компании Backbone Security, DataMark Technologies, продукты wbStego, stirmark library, Forensic Toolkit, EnCase, HashKeeper, ILook, ProDiscover и др.), ориентированными на работу с определенными видами трафика;
- средства внедрения ориентированы на использование либо одного контейнера для переноса данных, либо последовательности однотипных контейнеров, т.е. не поддерживают многоканальную параллельную передачу данных в телекоммуникационных каналах одного типа;
- не поддерживается внедрение информации в агрегированные телекоммуникационные каналы, что делает стеганографические средства безопасного обмена информацией неэффективными в современных сетях;
- не учитывают возможность изменения характеристик качества обслуживания различных классов трафика, а также изменения помеховой обстановки в каналах во время обмена информацией между двумя узлами.

Как следствие, необходимо разрабатывать универсальные подходы к внедрению информации в различные типы и классы трафика для сетей с адаптивным управлением передачей данных, ориентироваться в разработках на абстрактные обобщенные описания систем внедрения, а анализ проводить для различных типов сред, с учетом возможностей потерь, повторных передач, искажения и частичного раскрытия данных.

Общий подход к оценке эффективности системы скрытия информации

В стеганосистеме (рис. 1) сообщение I внедряется в контейнер размером N битов, который является последовательностью битов $S^N = (S_1, \dots, S_N)$ – данными файлов, битовых потоков, изображений, записей БД, бинарными объектами, видео- и аудиопотоками и др. Ключевая информация $K^N = (K_1, \dots, K_N)$ передается кодирователю и декодирователю с использованием криптографических методов и протоколов [1, 16], в том числе, возможно, и с разделением секрета [1].

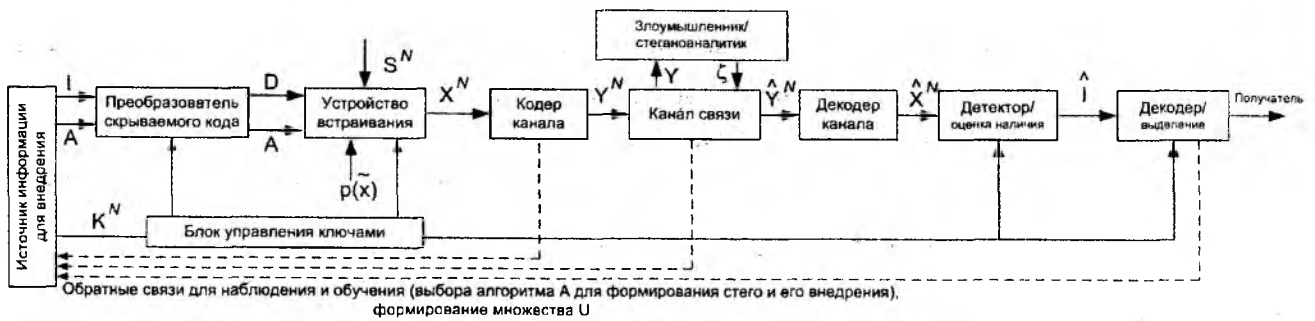


Рис. 1

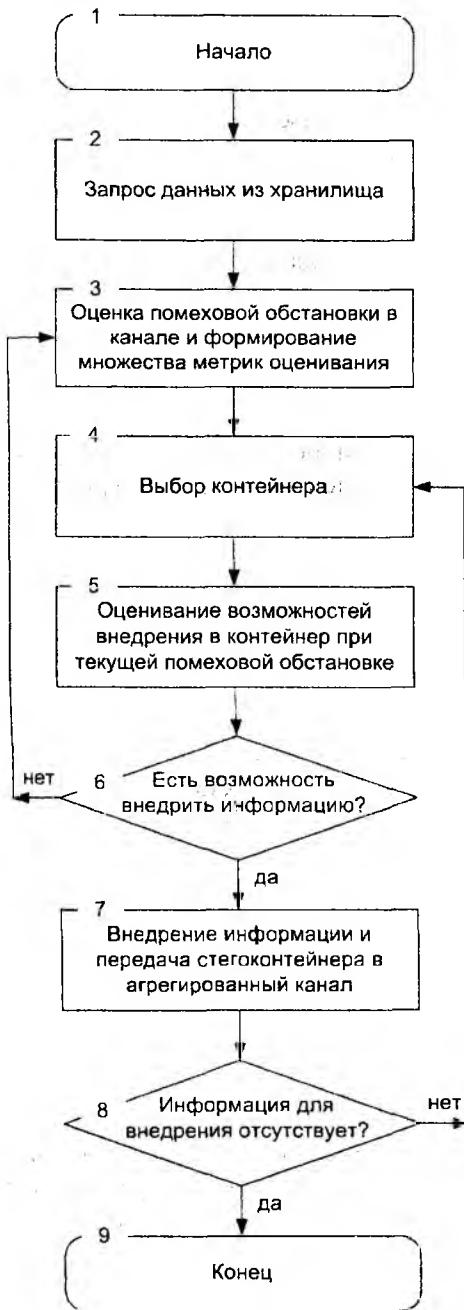


Рис. 2

Сигнал (файл) $X^N = (X_1, \dots, X_N)$ может формироваться с использованием дублирования сообщения (или создания множественных копий) – D – данные дополнительного экземпляра I . При этом на X^N налагается ограничения модификации (допускается конечное множество вариантов V сигналов), т.е. статистические характеристики X^N для различных I должны быть схожи (или $V(X(I)) = inv$, где inv – инвариант преобразований [3, 10]).

Данные D могут (должны) дополнительно преобразовываться с использованием ключа автора данных для обеспечения причастности к выработке/передаче данных. Устройство встраивания сообщения реализует композицию сообщения, мастер-ключей автора данных и устройства защиты информации, а также контейнера таким образом, чтобы статистические характеристики сигнала X^N не менялись в результате композиции [4].

Композиция информации I с контейнером S происходит с использованием алгоритма A из классов алгоритмов ассоциации информации и контейнера [3]:

- класса алгоритмов, предполагающих конкатенацию I и S : $A_1 : X = I || S$;

- класса алгоритмов, предполагающих изменение S по закону I : $A_2 : X = f(S, I)$, где f – подходящее отображение;

- класса алгоритмов, предполагающих объединение I и S подходящим образом: $A_3 : X = S \otimes I$.

Термин «подходящее» в данном случае используется для характеристики тех методов композиции, которые обеспечивают минимальную вероятность успешного проведения атаки с целью компрометации системы. Для каждого типа сигнала классы алгоритмов должны формироваться уникальным образом. Сигнал X^N является предметом атак, которые направлены:

- восстановление внедренной информации I путем поиска классов алгоритмов, обратных $A : A^{-1}$;
- удаление внедренной информации из контейнера $X^i = A_i(A_i^{-1}(X) - I)$, $i = 1, 2, 3$;

– модификацию внедренной информации или контейнера $X' = A_i(A_i^{-1}(X) - I + I')$ или $X' = A_i(X + \Delta X(I, I'))$, где I' – новая информация, которая может или изначально определяться, или вычисляться (функция ΔX) на основе доступных сведений о I ; оператор «+» означает любую линейную групповую операцию, а «-» – линейную групповую операцию, обратную «+».

Внедрение информации может выполняться (здесь определено минимальное для технологий мультисервисных сетей количество шагов) с использованием алгоритма, шаги которого показаны на рис. 2.

Таблица 1

Метрики	Расчетная формула	Тип сигнала
максимальной разницы	$MD = \max_n I_n - \hat{I}_n $	Неподвижное изображение, документ
средних абсолютных разниц	$AD = \frac{1}{N} \sum_{n=1}^N I_n - \hat{I}_n $	
нормальных абсолютных средних разниц	$NAD = \sum_{n=1}^N I_n - \hat{I}_n / \sum_{n=1}^N I_n $	Неподвижное изображение
среднеквадратической ошибки	$MSE = \frac{1}{N} \sum_{n=1}^N (I_n - \hat{I}_n)^2$	Для любого типа сигнала, кроме документов с сильно взаимосвязанными объектами
нормализованной среднеквадратической ошибки	$NMSE = \sum_{n=1}^N (I_n - \hat{I}_n)^2 / \sum_{n=1}^N I_n^2$	
отношения сигнал/шум	$SNR = \sum_{n=1}^N I_n^2 / \sum_{n=1}^N (I_n - \hat{I}_n)^2$	
пикового отношения сигнал-шум	$PSNR = N \cdot \max_{n=1..N} I_n^2 / \sum_{n=1}^N (I_n - \hat{I}_n)^2$	
нормализованной кросс-корреляции	$NC = \sum_{n=1}^N I_n \hat{I}_n / \sum_{n=1}^N I_n^2$	Офисные документы, предполагающие возможность произвольного следования элементов; гипертекст
качества корреляции	$CQ = \sum_{n=1}^N I_n \hat{I}_n / \sum_{n=1}^N I_n$	
структурности содержимого	$SC = \sum_{n=1}^N I_n^2 / \sum_{n=1}^N \hat{I}_n^2$	
схожести гистограмм	$HS = \sum_{c=1}^L f_i(c) - f_j(c) $, где $f_i(c)$ – относительная частота уровня c (всего L уровней)	Неподвижные изображения, звуковые файлы
отношения показателя сигмы к показателю уровня ошибок	$SER_b = \frac{\sigma_b^2}{\frac{1}{P} \sum_{p \in b} (I_n - \hat{I}_n)^2}$	Видеопотоки, звуковые потоки, технология которых предполагает возможность потерь фрагментов

Рассматривая анализ помеховой обстановки, необходимо определить допустимую величину искажения контейнера перед его передачей в канал связи. Для любого типа материала контейнера S^N возможности встраивания (функция встраивания f_N) ограничены величиной среднего искажения D_1 :

$$\sum_{\bar{x}^N \in X^N} \sum_{k^N \in K^N} \sum_{i \in I} \frac{1}{|I|} p(\bar{x}^N, k^N) d^N(\bar{x}^N, f_N(\bar{x}^N, i, k^N)) \leq D_1, \quad (1)$$

а $f_N^{-1}: Y^N \times K^N \rightarrow \hat{I}$ есть декодирующее отображение принятой стегопоследовательности y^N и ключа k^N в декодированное сообщение $\hat{i} = f_N^{-1}(y^N, k^N)$. D_1 характеризует искажение контейнера, максимально допустимое при встраивании в него скрываемого сообщения.

Воздействие без памяти (в канале связи в результате воздействия помех или сигнала злоумышленника), приводящее к искажению D_2 , описывается условной функцией распределения $Q^N(y^N | x^N)$ из множества X^N во множество Y^N :

$$\sum_{x^N \in X^N} \sum_{y^N \in Y^N} d^N(x^N, y^N) Q^N(y^N | x^N) p(x^N) \leq D_2. \quad (2)$$

Выбор D_2 ограничен границами искажения контейнера. Если необходимо сохранить контейнер без изменений, задача выбора D_2 усложняется и должна решаться специальными методами [10, 12].

Если известно описание функции f_N , то удаление описывается и ограничивается усредненным искажением между множествами \tilde{X}^N и Y^N :

$$\sum_{i, k, \tilde{x}^N, y^N} d^N(\tilde{x}^N, y^N) Q^N(y^N | f_N(\tilde{x}^N, i, k^N)) p(\tilde{x}^N, k^N) \leq D_2.$$

Если неизвестны вероятностные характеристики контейнеров, удаление информации становится практически невыполнимой задачей. Характеристики можно сделать похожими для абсолютно разных контейнеров с помощью вспомогательного криптопреобразования, стойкого к дифференциальному криптоанализу, например с помощью алгоритма шифра AES.

Для анализа объема информации, который возможно внедрить при применении допустимых преобразований из множества, формируемого с учетом недостижимости D_1 , можно применять метрики, описанные в таблице. Задавая метрику и выбирая необходимые свойства показателей системы скрытой передачи информации, возможно определять максимальный объем информации, который может быть внедрен в сигнал произвольного типа (он равен энтропии множества шаблонов, выбираемых для внедрения [10]). Условие максимизации объема информации рассмотрено ниже.

Шаги 3-6 можно итеративно выполнять в течение времени, пока процедура внедрения информации остается актуальной, для увеличения эффективного объема контейнера. В это время может освободиться часть пропускной способности общего, агрегированного канала, измениться доступное множество контейнеров или улучшиться помеховая обстановка в канале.

Анализ эффективности внедрения информации и передача стегоконтейнера в агрегированный канал

Обозначим d – стоимость внедрения, которая для различных ситуаций в системе скрытой передаче информации может обозначать:

- отношение полезной нагрузки (объема внедряемой информации) к общему объему передаваемой информации в рамках множества контейнеров;
- ресурсозатраты системы внедрения (объем оперативной памяти; разделяемое процессорное время; пропускная способность сети, используемая для получения данных из удаленного источника);
- время ожидания выдачи данных в канал (интегральная характеристика, включающая время запроса данных для обработки; время ожидания в очереди на обработку; время обработки (для различных контейнеров является различным); время передачи данных в агрегированный канал; время ожидания выдачи данных в сеть в агрегированном канале (время накопления информации из всех источников для их выдачи в канал)).

Результаты могут быть различными при применении различных метрик [12]; объективные метрики для информации различного типа приведены в таблице.

Функция выигрыша от внедрения принимается линейной (справедливо, если мы преобразуем фрагменты информации порциями, размер которых оптимально/субоптимально соотносится с размером принимаемых/отправляемых пакетов является наилучшим вариантом с точки зрения защищенности системы): $r_i(d_i) = b_i - k_i d_i$, $i = 1, 2, \dots, m$, $b_i > 0$, $k_i > 0$. Функция выигрыша показывает количество ресурса, необходимого злоумышленнику для обнаружения/выделения стего. Рассмотрен именно худший вариант, поскольку мы должны ориентироваться на минимальную защищенность системы скрытой передачи информации: наилучшие возможности у защищающегося и наилучшие – у атакующего (т.е. находим нижнюю границу стойкости системы). Здесь $b_i \geq b_j$, $k_i \geq k_j$ (упорядочивание, определяемое абсолютными приоритетами), а с увеличением d_i уменьшается r_i (показывает, что чем дальше (хуже) будет действовать защищающийся, тем больший ресурс для анализа сможет аккумулировать злоумышленник плюс то, что защищающийся должен стремиться к выбору таких алгоритмов внедрения, которые будут быстрыми и эффективными).

Рассмотрим канал передачи данных емкостью C битов, поддерживающий мультиплексирование m классов трафика (рис. 3). В канале каждый тип класса имеет свою собственную очередь. Характер поступления трафика каждого типа описывается пуассоновским процессом [6], длины пакетов имеют экспоненциальное распределение. Скорость поступления для m классов составляет $\lambda_1, \lambda_2, \dots, \lambda_m$ пакетов в секунду. Зададим \bar{L}_i – среднюю длину пакетов (в битах) для класса i ; d_i определяет стоимость внедрения для каждого из классов пакетов.

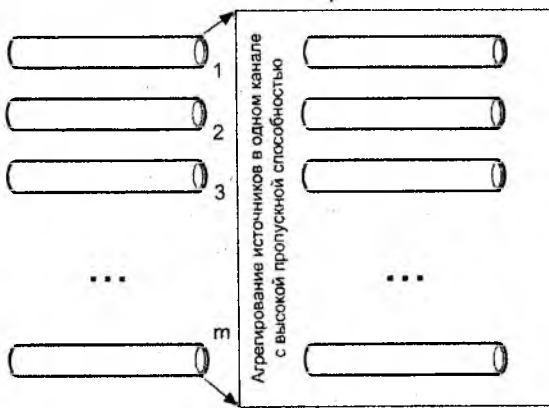


Рис. 3

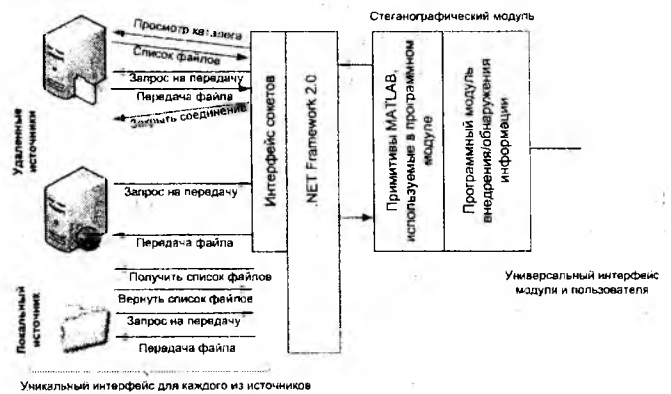


Рис. 4

Доля емкости класса i определяется через параметр w_i , который называется весом класса i . Естественными ограничениями для w_i , $1 \leq i \leq m$ являются $\sum_{i=1}^m w_i = 1$ и $w_i \in (0, 1]$. Для корректной работы системы также должно выполняться условие $\sum_{i=1}^m \lambda_i \bar{L}_i \leq C$.

Класс i пакетов имеет интенсивность поступлений λ_i , что гарантирует теоретический объем контейнера $w_i C$ с аналитическим средним \hat{d}_i для класса i , которое определяется в стационарном режиме как $\hat{d}_i = \frac{1}{\frac{w_i C}{L_i} - \lambda_i} = \frac{\bar{L}_i}{w_i C - \lambda_i \bar{L}_i}$. Для системы естественным является

ограничение $w_i C > \lambda_i \bar{L}_i$, поскольку стоимость не может быть отрицательной величиной.

Тогда количество ресурса, которым должен обладать оппонент для полного раскрытия внедренной информации, составит

$$F = \sum_{i=1}^m \lambda_i r_i(d_i) = \sum_{i=1}^m \lambda_i \left(b_i - \frac{k_i \bar{L}_i}{w_i C - \lambda_i \bar{L}_i} \right) \quad (3)$$

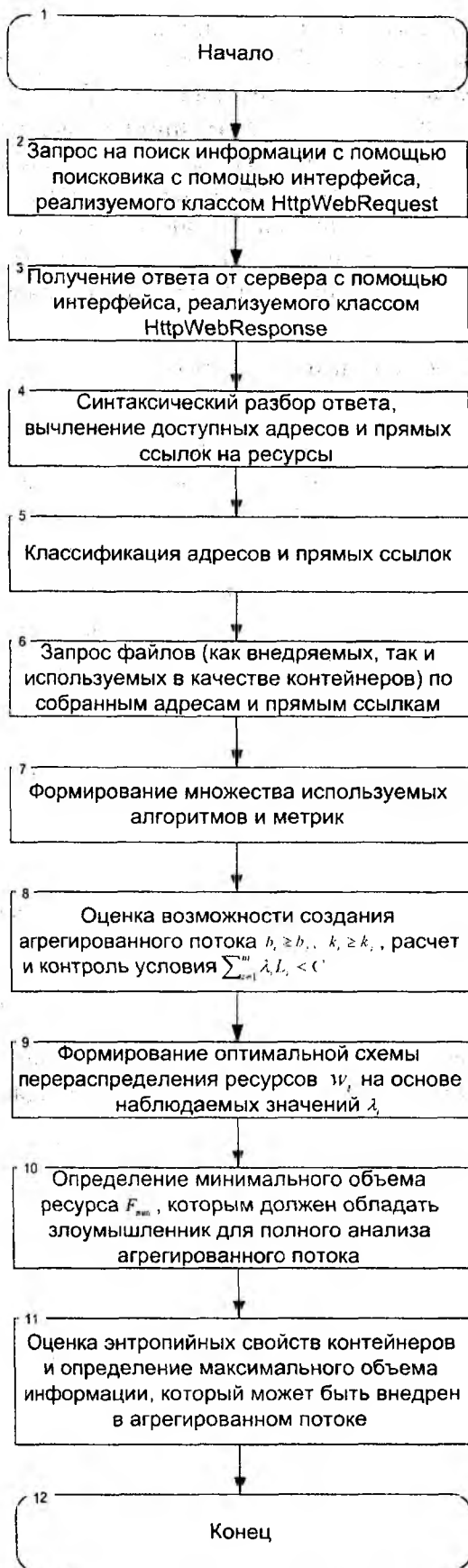


Рис. 5

При этом защищающемуся необходимо стремиться к максимизации F .

Предположим, что нет ограничений на объем ресурсов, используемых для внедрения (весь доступный вычислительный ресурс может динамически перераспределяться между устройствами, обрабатывающими различные классы трафика, ресурс полностью используется – ситуация типична при ограничении интенсивности входящего потока заявок на обслуживание). Тогда

$$\max F = \sum_{i=1}^m \lambda_i \left(b_i - \frac{k_i \bar{L}_i}{w_i C - \lambda_i \bar{L}_i} \right),$$

при условиях $\sum_{i=1}^m w_i = 1, 0 < w_i \leq 1, w_i C > \lambda_i \bar{L}_i$.

Выполняя оптимизацию системы [5], получаем, что оптимальной схемой перераспределения ресурсов будет

$$w_i = \frac{\sqrt{\lambda_i k_i \bar{L}_i} \left(C + \frac{\sum_{j=1}^m \sqrt{\lambda_j k_j \bar{L}_j}}{\sqrt{\lambda_i k_i \bar{L}_i}} \lambda_i \bar{L}_i - \sum_{j=1}^m \lambda_j \bar{L}_j \right)}{C \cdot \sum_{j=1}^m \sqrt{\lambda_j k_j \bar{L}_j}}. \quad (4)$$

Минимальный объем ресурса, необходимого для раскрытия всей информации злоумышленником, составляет

$$F_{\min} = \sum_{i=1}^m (\lambda_i b_i) - \frac{\left(\sum_{i=1}^m \sqrt{\lambda_i k_i \bar{L}_i} \right)^2}{C - \sum_{i=1}^m \lambda_i \bar{L}_i}. \quad (5)$$

Был проведен эксперимент по оценке объема информации (согласно классическим методам проведения эксперимента, рекомендуемых в [9]), который можно внедрить в информацию, загружаемую из контентохранилища произвольного типа (рис. 4, 5). Информационные запросы были произведены к сайтам, найденным с помощью поисковых систем Altavista и Google по запросу «Steganography». Для выполнения запроса и получения/обработки результатов использовался интерфейс сокетов и библиотеки системы .NET Framework [17], а также внутренние классы HttpWebRequest/HttpWebResponse.

Затем для посылки команд и загрузки данных с сайтов использовались классы HttpWebRequest/HttpWebResponse, FtpWebRequest/FtpWebResponse, а для мониторинга параметров пакетов – функции библиотеки WinPcap [7, 14]. Было сформировано 8 классов трафика с различными характеристиками (пакеты маркировались).

Внедряемая информация преобразовывалась с использованием различных библиотек пакета MATLAB 6.5 (Signal Processing Library, Simulink и базовые библиотеки [13]). Формирование пакетов выполнялось полностью программно с помощью специально разработанного для эксперимента кода (использовалась запись информации непосредственно в буфер передачи, функции операционной системы по формированию пакетов). Пакеты не отправлялись на выходной интерфейс, лишь регистрировалось время до отправления пакетов. Регистрация событий и производительности отдельных компонентов системы (процессорного времени, потокового времени, оперативной памяти и др.) производилась с помощью стандартной оснастки Windows «Performance». Обработка полученных данных, расчеты и формирование отчетов производились в MATLAB.

Обобщенный метод определения максимального объема информации, внедряемой в многоканальной системе

Задавая метрику и выбирая необходимые свойства показателей системы скрытой передачи информации, можно определять максимальный объем информации, который может быть внедрен в сигнал произвольного типа. Согласно [10], он равен энтропии шаблона g ($g \in G$, где G – множество, формируемое комбинаторным способом u , $u \in U$ выбором из юниверсума с учетом ограничений, вносимых выбором метрики и свойств системы), где $S \times X \xrightarrow{u \in U} G$. Применение сжатия исходных сообщений и выдвижение предположения о возможности искажений восстанавливаемой информации позволяет увеличить количество возможных шаблонов (мощность G).

Кодер стеганосистемы использует изменяемый шаблон g . Предполагая, что кодер выбирает каждый шаблон g с вероятностью $p(g)$, объем информации, который может быть передан, равен энтропии $p(g)$:

$$H(p) = - \sum_{g \in G} p(g) \log_2 p(g). \quad (6)$$

В общем случае максимальный объем информации, который может быть внедрен в агрегированном канале,

$$V_{max}(i \in I) = \sum_{j=1}^m w_j H(p). \quad (7)$$

Задачей, связанной с максимизацией объема передаваемой информации, является нахождение вероятностного распределения $p(g)$ в пространстве всех возможных модификаций шаблона g , которое максимизирует ожидаемое значение степени внедрения $\sum_{g \in G} D(g) p(g)$, где $D(g)$ – функция полезности внедрения.

При ограничениях $H(p) = \sum_{g \in G} p(g) \log_2 p(g) = \nu$, $\sum_{g \in G} p(g) = 1$, задача может быть решена с использованием множителей Лагранжа [5].

Здесь $F(p(g)) = \sum_{g \in G} D(g) p(g) + \mu_1 \left(\nu - \sum_{g \in G} p(g) \log_2 p(g) \right) + \mu_2 \left(\sum_{g \in G} p(g) - 1 \right)$, а затем решается уравнение $\frac{\partial F}{\partial p(g)} = D(g) - \mu_1 (\log_2 p(g) + 1/\ln(2)) + \mu_2 = 0$. И если только $p(g) = A e^{-\lambda D(g)}$, где $A^{-1} = \sum_{g \in G} e^{-\lambda D(g)}$ и λ определяются из $-\sum_{g \in G} p(g) \log_2 p(g) = \nu$.

Таким образом, вероятности $p(g)$ должны подчиняться экспоненциальному распределению для того, чтобы значение функции полезности $D(g)$ было максимальным.

Множество U формируется на основе оценок возможностей стегоаналитика (предельных выборочных (целенаправленных) искажений сигнала Y) и возможностей пары детектор-

декодер по восстановлению информации в определенной помеховой обстановке [4]. Статистические оценки для U могут быть получены на основе косвенных измерений состояния элементов ТКС [18], а класс оптимальных оценок – с помощью линейной фильтрации Калмана [2, 8]. Управление в системе может быть определено как добавление или удаление элементов множества U . Заметим, что оптимальное управление стегакодером может быть организовано только при наличии обратной связи «детектор/декодер – кодер».

Выводы

Исследованы вопросы оценки объема внедряемой информации в потоки данных мультисервисных сетей, которые передаются в соответствии с приоритетами и сформированными классами трафика. Описана стеганографическая система на основе общей теории систем для ситуации использования контейнеров произвольного вида, а также сопутствующей криптографической защиты. Проанализированы виды информации, внедряемой в потоки трафика, особенности структуры контейнеров, в результате чего были сформированы рекомендации по применению множества метрик для анализа возможностей контейнеров по внедрению в них информации различного типа. Рассмотрены вопросы оценки объема информации для трафика различных классов и проведен эксперимент, подтвердивший возможность нахождения в режиме реального времени точных оценок объема информации, который можно скрыть в агрегированных потоках (на основе данных, собираемых в пиринговых системах, таких как поисковые системы Altavista и Google). Метод предполагает получение результатов, которые соответствуют нижней грани оценки стойкости стеганографической системы к различного рода атакам, производимым злоумышленником, т.е. анализируется ситуация, когда защищающийся находится в худшем положении относительно злоумышленника.

Список литературы: 1. Поповский В. В., Персиков А. В. Защита информации в телекоммуникационных системах. Т 1: Учебник. ООО «Компания СМИТ», 2006. 238 с. 2. Сейдж Э., Мелс Дж. Теория оценивания и ее применение в связи и управлении. М.: Связь, 1976. 496 с. 3. Месарович М., Такахара Я. Общая теория систем. Математические основы. М.: Мир, 1978. 314с. 4. Тихонов В. И. Статистическая радиотехника. М.: Радио и связь, 1982. 623 с. 5. Мину М. Математическое программирование. Теория и алгоритмы. М.: Наука, 1990. 488 с. 6. Dominique Gaïti et al. Network Control And Engineering for QoS, Security and Mobility, part III. Springer Science 2005, 364 p. 7. Loris Degioanni, Mario Baldi, Fulvio Rizzo and Gianluca Varenni, Profiling and Optimization of Software-Based Network-Analysis Applications, Proceedings of the 15th IEEE Symposium on Computer Architecture and High Performance Computing (SBAC-PAD 2003), Sao Paulo, Brazil, November 2003., 8. Grewal M., Andrews A. Kalman filtering: theory and practice using MATLAB, John Wiley & Sons, 2001, 410 p. 9. Методы исследований и организация экспериментов / Под ред. проф. К. П. Власова Х.: Изд-во Гуманитарный Центр, 2002. 256 с. 10. Гривунин В. Г., Оков И. Н., Туринцев И. В. Цифровая стеганография. М: Солон-Пресс, 2002. 272с. 11. Персиков А. В., Черный С. В., Еременко А. С. Некоторые общие вопросы разработки систем защиты авторского права на электронные документы, распространяемые в телекоммуникационных системах. // Восточно-Европейский журнал передовых технологий, 4/2 (28) 2007, с. 38-42. 12. M. Kutter and F. A. P. Petitcolas A fair benchmark for image watermarking systems. Electronic Imaging '99. Security and Watermarking of Multimedia Contents, vol. 3657, Sans Jose, CA, 1999. 13. Айфичер Э., Джервис Б. Цифровая обработка сигналов: практический подход. М.: Изд. дом «Вильямс», 2004. 992 с. 14. Fulvio Rizzo, Loris Degioanni, An Architecture for High Performance Network Analysis, Proceedings of the 6th IEEE Symposium on Computers and Communications (ISCC 2001), Hammamet, Tunisia, July 2001. 15. Гургенидзе А. Т., Кореш В.И. Мультисервисные сети и услуги широкополосного доступа. СПб.: Наука и Техника, 2003. 400 с. 16. Поповский В. В., Персиков А. В. Защита информации в телекоммуникационных системах. Том 2: Учебник. ООО «Компания СМИТ», 2006. 292с. 17. Troelsen A. Pro C# 2008 and the .NET 3.5 Platform. Apress, 2007. 1400 p. 18. Поповский В. В. Модель управления реструктуризацией телекоммуникационной сети // Радиотехника. 2004. №138. С. 25-31.