УДК 004.9, 004.056

**Р.Ф. Жаринов**

# Метод защиты от перлюстрации в DLP-системах

Предложена концепция защиты от перлюстрации в компаниях, использующих DLP-системы. Определены потенциальные угрозы, присутствующие в системах автоматизированной обработки информации, в частности для DLP-систем, являющихся уязвимостями для получения доступа к личной информации сотрудников. Выделены алгоритмы для морфологического разбора русских слов, а также рассмотрен метод гомоморфного шифрования с использованием закрытого ключа. Рассмотрены этические и правовые моменты просмотра личной переписки сотрудников.

**Ключевые слова:** dlp-система, перлюстрация, гомоморфное шифрование, морфологический разбор.

В статье рассматриваются концепция мер защиты от перлюстрации в компаниях, использующих DLP-системы (Data Leak Prevention / система защиты от утечек информации), а также анализ и выбор алгоритмов и техник для дальнейшей практической реализации.

DLP-система – пакет программных продуктов, обеспечивающих проактивную защиту от потери корпоративных данных. Система использует централизованное управление политиками безопасности и постоянно анализирует информационный поток данных, выходящих за пределы корпоративной сети, что позволяет предотвратить потери конфиденциальных данных независимо от их местонахождения. В случае детектирования передачи конфиденциальной информации система может как блокировать передачу, так и отправлять уведомление администратору безопасности. Основными задачами DLP-систем являются:

– Обнаружение случайной или умышленной несанкционированной передачи информации сотрудниками компании, на основании как анализа содержания, так и контекстного размещения ключевых слов в файле.

– Гибкие механизмы автоматизации обработки инцидентов для последующего ручного анализа.

– Возможность установления владельцев передаваемой и хранимой информации.

Таким образом, администратор DLP-системы может просматривать всю информацию, как исходящую, так и входящую. Просмотр корреспонденции втайне от отправителя и получателя называется перлюстрацией. Право на тайну переписки гарантируется законодательством Российской Федерации. Статья 23 Конституции РФ гласит: «…каждый гражданин имеет право на тайну переписки, телефонных переговоров, почтовых, телеграфных и иных сообщений. Ограничение этого права допускается только на основании судебного решения».

В то же время сотрудники предприятия на рабочем месте должны выполнять свои служебные обязанности. Рабочие места, средства связи, программные продукты и сети передачи данных принадлежат предприятию. Работодатель оплачивает трафик, используемый сотрудниками в личных целях, и их рабочее время, потраченное на личное общение. Он же будет нести убытки в случае утечки конфиденциальной информации. Согласно п. 4 ст. 10 Федерального закона РФ от 29 июля 2004 г. № 98-ФЗ «О коммерческой тайне» «обладатель информации, составляющей коммерческую тайну, вправе применять при необходимости средства и методы технической защиты конфиденциальности этой информации, другие, не противоречащие законодательству Российской Федерации, меры» [1].

Однако просмотр переписки возможен только при наличии согласия на это как со стороны отправителя, так и со стороны получателя. Поэтому гарантировать, что не возникнет ситуации, когда перлюстрация будет незаконна, организационными мерами никто не может.

**Концепция защиты**

Для создания мер защиты от перлюстрации, как от внутренних, так и внешних потенциальных угроз, в статье приведена концепция, позволяющая использовать шифрование трафика и анализировать поток данных, не зная секретных ключей.

Для начала определим потенциальные угрозы при использовании любой автоматизированной системы обработки и хранения информации:

**Доступ к информации, хранящейся на серверах, в базах данных (БД) и передаваемой по интранет-сетям.** Рассматриваемый доступ обычно имеют системные администраторы, администраторы баз данных и администраторы безопасности. В случае с DLP-системами особое значение имеет информация, передаваемая по интранет-сетям и хранящаяся в базе данных (в зависимости от настройки системы в базе данных может храниться как информация, сохраненная при возникновении инцидента, так и весь трафик, который передавался по определенным протоколам).

**Несанкционированное проникновение или доступ к информации.** Существует потенциальная угроза получения доступа к серверам или базам данных. Так как DLP-система – это Web-система, т.е. есть серверная часть, отвечающая за обработку и предоставление информации пользователям системы, присутствует угроза sql-инъекции. При реализации данной уязвимости можно получить информацию, хранящуюся в базе данных, и (или) изменить ее.

Теперь рассмотрим способы передачи личной/конфидециальной информации и ее обработки в DLP-системах:

**Передача информации в открытом виде.** При обнаружении несанкционированного использования информации сотрудниками создается инцидент, а сообщение шифруется на известном администратору ключе и сохраняется в базе данных DLP-системы. При передаче информации в открытом виде появляется возможность перехвата переписки сотрудников привилегированными пользователями или потенциальными противниками.

**Передача информации в зашифрованном виде с использованием симетричной криптографии.** В таком случае хранение ключей должно осуществляться на «защищенном и доверенном» сервере. При этом существует угроза, что администратор или несанкционированный пользователь получит доступ к защищенному серверу.

**Передача информации в зашифрованном виде с использованием асимметричной криптографии.** На сервере хранится открытый ключ каждого сотрудника, данная информация является открытой.

Исходя из рассмотренных потенциальных угроз, предпочтение по защите конфиденциальной информации лучше отдавать асимметричной криптографии, т.к. в этом случае появляется возможность общаться с внешним миром (сотрудники, партнеры, находящиеся вне корпоративной сети), без реализации дополнительных протоколов безопасности.

Рассмотрим схематически концепцию защиты личной переписки сотрудников в DLP-системах (рис. 1).

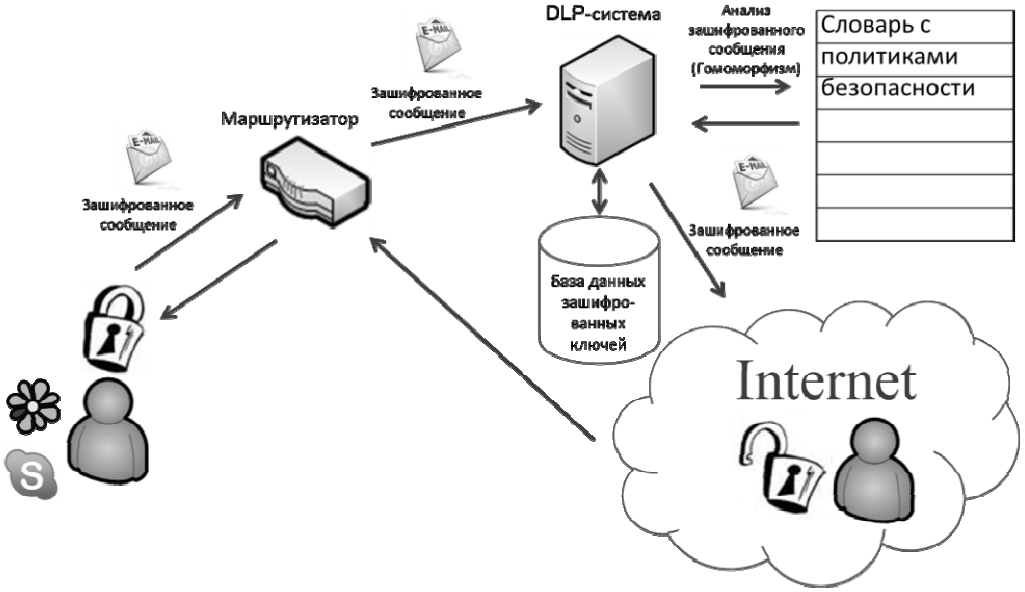


Рис. 1. Схема работы DLP-системы с реализованным модулем противодействия перлюстрации

При развертывании DLP-системы необходимо сгенерировать открытый ключ (далее – системный ключ), на основе коалиционного ключа (использование пороговых схем) – секрета, разделенного между членами коалиции (например, Совет директоров), таким образом генерация закрытого ключа для расшифрования сообщения сотрудника возможна лишь при предоставлении всех частей коалиционного ключа.

При детектировании возможной утечки корпоративных данных полученный информационный поток от сотрудника шифруется на предварительно сгенерированном открытом ключе и сохраняется в базе данных.

Для обмена личной информации между сотрудниками может быть использована асимметричная криптография, т.е. данные, передаваемые по определенным протоколам (например, icq, jabber, skype), будут шифроваться на открытом ключе, и расшифровать их сможет только получатель сообщения. Немаловажным фактором является подписывание электронной цифровой подписью (ЭЦП) передаваемого сообщения для сохранения его целостности.

При возникновении инцедента сохраненное сообщение должно быть расшифровано при помощи закрытого ключа, который получается лишь в том случае, если собрать пороговую схему. Все закрытые ключи сотрудников шифруются на системном ключе для возможности расшифрования сообщения.

Для реализации данной концепции необходимо выделить основные сложности реализации:

1) Морфологический разбор слов в предложении (специфика русского языка).
2) Возможность поиска информации в зашифрованных данных.
3) Распределение и хранение пар ключей для сотрудников и внешних лиц.

Ниже представлен анализ этапов реализации.

Под морфологическим анализом понимается интеллектуальный разбор слов текста, т.е. поиск ключевых слов во входном потоке информации. Для русского языка существует большое количество слов, для которых нет однозначного разбора. На данный момент имеется несколько способов выделения основной части слова:

1) Составление собственного морфологического словаря. Данный способ является трудоемким и не гарантирует 100% результат.
2) Использование алгоритма стемминга. Не используются словари и выделение корня осуществляется путем преобразования слова согласно определенным правилам:

− Поиск и удаление заранее заданных окончаний. Поиск слова, которое заканчивается прилагательным, глаголом или существительным.
− Если слово оканчивается на «и», удаляем его.
− Поиск и удаление деривационного окончания.
− Удалить двойную буквы «н», а также если слово оканчивается на «ейш» или «ейше», удалить эти окончания.

3) Определение исходного слова по аффиксу (окончанию) и суффиксу слова. Данный метод позволяет преобразовывать слова к начальным словоформам. Пример: «продавать информацию» преобразуется в «продать информация». Также присутствуют дополнения или так называемые слова-исключения, которые не изменяют свою форму (предлоги «не», наречия «столь» и т.д.). Для более точного преобразования слова необходимо добавить максимальновозможное количество исключений.

Исходя из того, что для алгоритма стемминга нет необходимости составлять какие-либо словари, то для практической реализации выбран именно он.

Основным алгоритмом поиска информации в зашифрованных данных является протокол линейного поиска в документах (Searchable Symmetric Key Encryption (SSKE) [2]).

Протокол позволяет хранить регистрозависимую информацию на «недоверенных» серверах. Состоит из 3 основных частей (с использованием криптографических примитивов): хранение, поиск и получение информации.

Для хранения информации текст $W$ разбивается на некоторое количество слов $W_i$, причем каждое слово имеет фиксированную длину $n$. Генерируются ключи $k'$ и $k''$ для использования в функциях шифрования и хеш-функциях. Затем каждый блок $W_i$ шифруется с использованием детерминированного алгоритма шифрования на ключе $k''$, полученный блок шифртекста делится на 2 части $L_i$ и $R_i$: $X_i = E_{k''}(W_i) = \langle L_i, R_i \rangle$. Используя генератор псевдослучайной последовательности $G$, создается множество бит $S_i$, основанных на местоположении $W_i$. Используя псевдослучайную функцию $f$ с ключом $k'$, получаем $k_i$ ключ для функции $F$.

На последнем шаге получаем зашифрованный блок, производя операцию «исключающее или» над полученными результатами (рис. 2): $C_i = \langle L_i, R_i \rangle \oplus \langle S_i, F_{k_i}(S_i) \rangle$, где $E$ – функция шифрования

(ключ $\times\{0,1\}^n \to \{0,1\}^n$), $f$ (ключ $\times\{0,1\}^{n-m} \to$ ключ) и $F$ (ключ $\times\{0,1\}^{n-m} \to \{0,1\}^m$, где $n$ –длина блока $W_i$, а $m$ – длина $R_i$) – криптостойкие хеш-функции с использованием ключей, значения $k'$, $k''$ и $S_i$ являются секретами, $C_i$ может храниться на «недоверенном» сервере, $G$ – генератор псевдослучайной последовательности (использование потокового шифра), такой что $K_g \to S$.

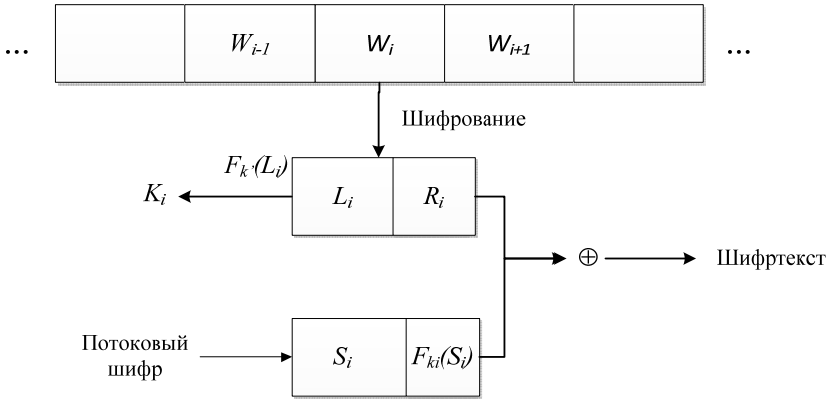


Рис. 2. Этап шифрования в гомоморфной схеме

Для того чтобы производить поиск в зашифрованных данных, т.е. понять, присутствует ли слово $W_j$ в зашифрованном тексте документа, необходимо передать на сторону сервера зашифрованное слово $W_j$. Для поиска информации вычисляются и используются следующие исходные данные: $k'$ – ключ шифрования; $k''$ – ключ для хеш-функции $f$; $W_j$ – текст, который ищем; $X_j = E_{k''}(W_j) = \langle L_j, R_j \rangle$ – зашифрованный блок; $K_j = f_{k'}(L_j)$ – ключ для хеш-функции $F$.

Значения $X_j$ и $k_j$ отправляются на сервер, и с их помощью можно вычислить значения функции-ловушки $T_p = C_p \oplus X_j = \langle S_p, S'_p \rangle$. И если $S'_p = F_{k_j}(S_p)$, то на сторону клиента возвращается пара $\langle p, C_p \rangle$. Необходимым условием является проверка значения псевдослучайной последовательности $S_p$, т.к. существует вероятность, что функция-ловушка может удовлетворять следующему условию: $T = \langle S_q, F_{k_j}(S_q) \rangle$, где $S_q != S_p$.

Для расшифрования информации необходимо произвести следующие операции: $C_p = \langle C_{p,i}, C_{p,r} \rangle$ – блок, который хранится на сервере, передается клиенту при поиске информации; $X_{p,i} = C_{p,i} \oplus S_p$ – левая часть зашифрованного блока; $K_p = f_{k'}(X_{p,i})$ – ключ для хеш-функции $F$; $T_p = \langle S_p, F_{k_p}(S_p) \rangle$ – вычисление проверяемого значения; $X_p = C_p \oplus T_p$ – зашифрованный блок; $W_p = D_{k''}(X_p)$ – расшифрованный текст; где $D$ – функция расшифрования, такая что $D_{k''}(E_{k''}(W_i)) = W_i$.

Таким образом, клиент может хранить, искать и получать информацию, в то время как сервер ничего не будет знать об исходном тексте. Все, что знает сервер, – это зашифрованное слово на диске; на этапе поиска – значения $X_i$ и $k_i$; из запроса клиента – информацию об местоположении на диске зашифрованного слова. Клиент использует ключ key, который известен только ему, для шифрования значений $C_i$ и $X_j$, а также ключ $k_j$ используется в качестве инициализации для создания случайного потока бит.

Безопасность данной схемы шифрования основывается на реализации функций генерации псевдослучайной последовательности и использовании функции шифрования (блочный шифр). К сожалению, основным недостатком данного протокола является линейно зависимый размер исходного документа.

В связи с тем, что для DLP-систем необходимы поиск информации в информационном потоке различной длины, а также использование различных ключей шифрования, необходимо модифицировать протокол представленной гомоморфной схемы шифрования.

**Заключение**

В статье приведен обзор криптографических схем, позволяющих производить поиск данных в зашифрованной информации, а также разработана концепция технического противодействия перлюстрации в DLP-системах. При реализации предложенного решения появится возможность внедрять DLP-системы, удовлетворяющие законодательству и гарантирующую тайну переписки.

*Литература*

1. Федеральный закон Российской Федерации от 29 июля 2004 г. № 98-ФЗ «О коммерческой тайне» (с последними поправками от 11.07.2011 № 200-ФЗ) [Электронный ресурс]. – Режим доступа: http://base.consultant.ru/cons/cgi/online.cgi?req=doc;base=LAW;n=70848, свободный (дата обращения: 10.05.2012).

2. Song D. Practical Techniques for Searches on Encrypted Data / D. Song, D. Wagner, A. Perrig [Электронный ресурс]. – Режим доступа: http://www.cs.berkeley.edu/~dawnsong/papers/se.pdf, свободный (дата обращения: 10.05.2012).

3. Goh E.-J. Secure Indexes for Searching Efficiently on Encrypted Compressed Data [Электронный ресурс]. – Режим доступа: http://www.ece.iit.edu/~ubisec/cloud/papers/ICDCS10-search.pdf, свободный (дата обращения: 10.05.2012).

4. Wang C. Secure Ranked Keyword Search over Encrypted Cloud Data / Cong Wang, Ning Cao, Jin Li, Kui Ren, Wenjing Lou [Электронный ресурс]. – Режим доступа: http://www.ece.iit.edu/~ubisec/cloud/papers/ICDCS10-search.pdf, свободный (дата обращения: 10.05.2012).

---

**Жаринов Роман Феликсович**
Аспирант каф. технологий защиты информации
Санкт-Петербургского государственного университета аэрокосмического приборостроения
Тел.: 8 (812) 494-70-52
Эл. почта: roman@vu.spb.ru

Zharinov R.F.

**Protection technology to intercept personal information in DLP-systems**

Proposed the concept to enable protection mode while transfer personal information in DLP-systems. Described algorithms such as homomorphic encryption and allocation the root of the word. Discussed the ethical and legal aspects to intercept staff’s personal correspondence.

**Keywords:** dlp-systems, homomorphic encryption, protection mode in transfer personal information.

---