УДК 621.394/.396.0193

АВТОМАТИЗИРОВАННАЯ ОБРАБОТКА ТЕКСТОВОЙ ИНФОРМАЦИИ С ПОМОЩЬЮ УСОВЕРШЕНСТВОВАННОГО КРИПТОАНАЛИТИЧЕСКОГО МЕТОДА "ГРУБОЙ СИЛЫ"

МГТУ "Станкин":
Наталия Михайловна Кузнецова,
аспирант, преподаватель кафедры "Автоматизированные системы обработки информации и управления"
Татьяна Владимировна Карлова,
д-р социол. наук, профессор

Представлен способ усовершенствования известного криптоаналитического метода "грубой силы" путём применения алгоритма кластерного анализа текста. Описано использование указанного метода в автоматизированной системе разграничения доступа к конфиденциальной информации.
Well-known cryptanalystic method "Brute-Force Attack" advanced by means of application of algorithm for cluster analysis of the text is presented. Usage of the method in automatic access control system to confidential information is described.
Ключевые слова: ***информационная безопасность, разграничение доступа, автоматизированная система, кластерный анализ текста.***
Keywords: ***information security, automatic access control, automated system, cluster analysis of the text.***

Современные промышленные предприятия и организации всё чаще сталкиваются с проблемой утечки конфиденциальной информации, связанной с нелегальными действиями сотрудников. Существует множество способов предотвращения подобной угрозы информационной безопасности, однако ни один из них не способен решить поставленную задачу в случае, если злоумышленник воспользуется одним из методов симметричного шифрования и произведёт попытку передачи конфиденциальной информации в зашифрованном виде.

В статье предложено использование автоматизированной системы разграничения доступа к конфиденциальной информации (АСРДкКИ) с применением модуля контроля потенциальных каналов утечки секретных данных. Основной особенность модуля контроля является использование усовершенствованного криптоаналитического метода "грубой силы" с применением алгоритма семантической кластеризации текста.

В качестве каналов утечки могут выступать локальная вычислительная сеть (ЛВС), глобальная вычислительная сеть (ГВС), съёмные носители данных, принтеры и факсы, ошибки программного обеспечения, неисправности аппаратного обеспечения [1]. Контроль над ЛВС, ГВС, съёмными носителями данных, принтерами и факсами возможно реализовать за счёт применения организационно-правовых, инженерно-технических, программно-аппаратных методов и средств обеспечения информационной безопасности.

Однако существующие технологии не предлагают решения задачи криптоанализа предварительно зашифрованной передаваемой конфиденциальной информации.

Применение расширенного алгоритма "грубой силы" в модуле контроля технических каналов утечки способно предотвратить описанную угрозу информационной безопасности предприятия (организации).

Существует несколько методов криптоанализа, позволяющих расшифровать данные: атаки класса "встреча посередине" ("meet-n-the-middle"), дифференциальный и линейный криптоанализ, метод "бумеранга" ("boomerang attack"), сдвиговая атака, метод интерполяции, "невозможные дифференциалы" ("impossible differentials"). Самым простым с точки зрения реализации, но требующим больших временных затрат является метод "грубой силы" ("brute-force attack").

Метод "грубой силы" предполагает перебор всех возможных вариантов ключа шифрования до нахождения искомого ключа [2].

Очевидно, основным недостатком метода является необходимость наличия больших вычислительных мощностей, в том числе специализированных устройств. Однако если преобразовать алгоритм таким образом, чтобы входное множество возможных ключей шифрования постоянно уменьшалось, данное условие можно обойти.

Уменьшение входного множества возможных ключей можно обеспечить за счёт внедрения в основной алгоритм дополнительного алгоритма семантической кластеризации текста.

На сегодняшний момент существует несколько алгоритмов кластеризации текста, среди которых следует отметить: STC (Suffix Tree Clustering), Single Link, Complete Link, Group Average, Scatter/Gather, K-means, CI (Concept Indexing), SOM (Self-Organizing Maps).

В основе семантической кластеризации текста лежит логическое разделение текста на группы - кластеры.

В тексте выделяются семантические центры - ключевые слова, вокруг которых формируются "семантические облака" - близкие по значению слова, встречающиеся чаще остальных рядом с ключевым. Чем чаще слово появляется рядом с ключевым, тем ближе оно находится к центру "семантического облака".

Кластеризация текста происходит путём анализа открытой текстовой информации. Чем больше объём входных данных, тем точнее формируется структура "семантического облака".

Модуль кластеризации текста является обязательной составляющей проектируемой АСРДкКИ. Особенностью модуля является то, что его работа может протекать в автономном режиме.

Рассмотрим взаимодействие модуля кластеризации текста с остальными составляющими АСРДкКИ.

Как показано на рис. 1, расширение алгоритма "грубой силы" в АСРДкКИ происходит за счёт применения модуля кластеризации текста.

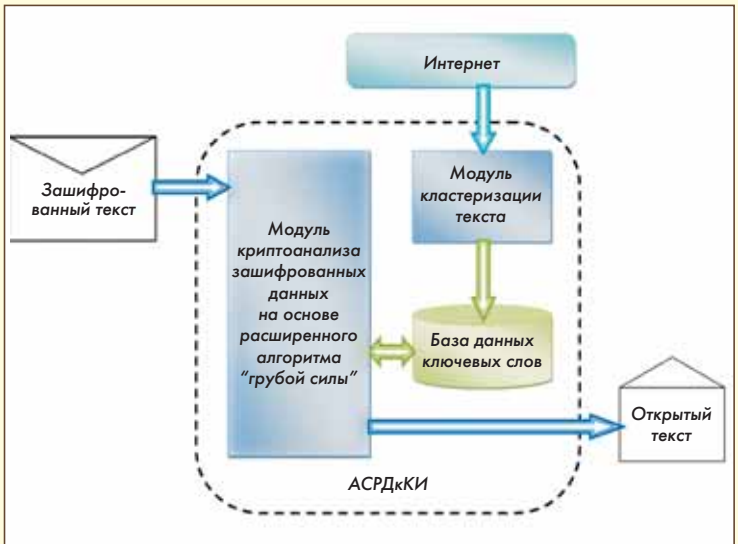

Рис. 1. Структурная схема взаимодействия модуля кластеризации текста с модулем криптоанализа в АСРДкКИ

Модуль кластеризации текста постоянно взаимодействует с глобальной сетью Интернет для своевременного обновления баз данных ключевых слов (и соответствующих "семантических облаков"). Постоянно производится "парсинг" - целенаправленный просмотр текстовой информации.

Необходимо отметить, что несмотря на то, что модуль кластеризации информации подключен к ГВС, данная архитектура АСРДкКИ удовлетворяет требованиям информационной безопасности: взаимодействие с "Интернет" производится однонаправлено. Модуль считывает текстовую информацию, анализирует её и пополняет базу данных ключевых слов. Вероятность проникновения в ЛВС предприятия вредоносного программного обеспечения через данный узел практически равна нулю.

Модуль криптоанализа зашифрованных данных на основе расширенного алгоритма "грубой силы" в процессе работы при нахождении группы символов, расшифрование которых могло бы привести к получению осмысленного слова (назовём данную группу "возможной семантической единицей"), обращается к базе данных ключевых слов, получает "семантическое облако" и ищет в ближайшем окружении "возможной семантической единицы" элементы полученного "семантического облака". Структура алгоритма представлена на рис. 2.

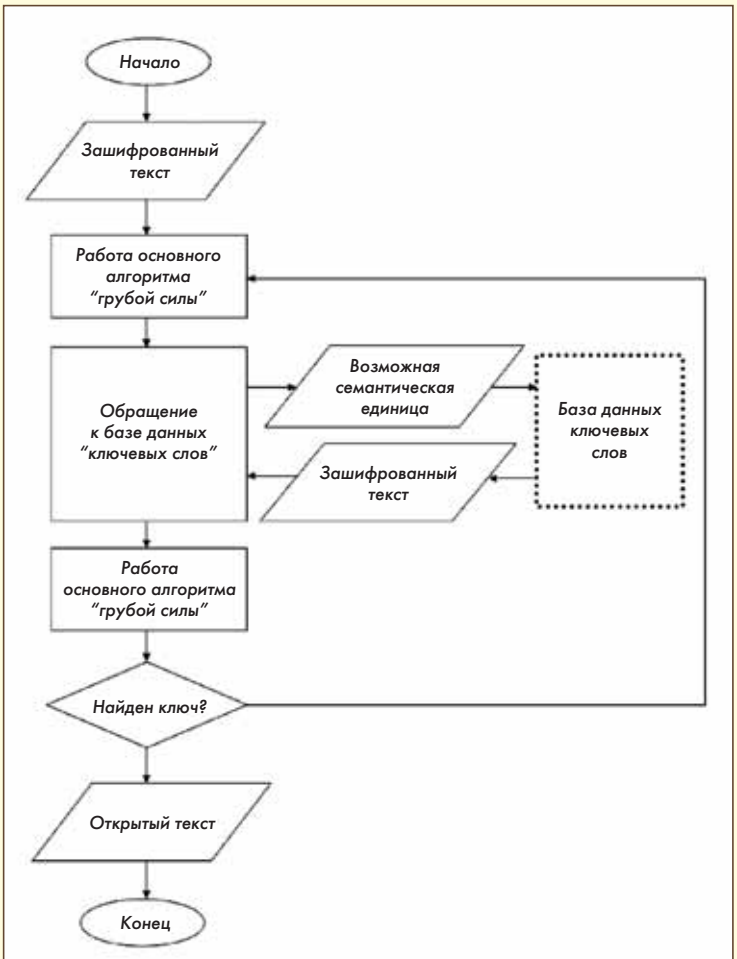

Рис. 2. Структура расширенного криптоаналитического алгоритма "грубой силы"

Обращение к базе данных "ключевых слов" происходит в двух случаях: пополнение "семантического облака" со стороны модуля кластеризации текста, запрос "семантического облака" со стороны модуля криптоанализа.

Модуль кластеризации текста и модуль криптоанализа можно объединить в модуль мониторинга исходящего информационного трафика предприятия. Применение данного модуля повысит эффективность АСРДкКИ, позволит производить постоянный контроль над внешним электронным документооборотом предприятия.

Как показано на рис. 1, основной модуль не модифицирован и работает согласно простому алгоритму "brute force attack", однако особенностью его является постоянное обращение к основной таблице базы данных ключевых слов: при возникновении подозрения на соответствие происходит поиск подходящего ключевого слова. Если совпадение произошло, из базы данных основному модулю возвращается массив "близких" слов, сформированной модулем кластеризации текста. "Близкие" слова с большей вероятностью могут оказаться рядом. Важно отметить, что эта вероятность довольно велика, т.к. кластеризация данных принадлежит к классу семантических методов анализа текста. С применением предложенного метода уменьшается диапазон возможных вариантов и осуществлять подбор становится намного легче. Как следствие, повышается скорость работы основного алгоритма, а также его эффективность.

В свою очередь модуль кластеризации текстовой информации постоянно корректирует базу данных ключевых слов за счёт регулярно пополняемого "сканируемого" входного незашифрованного текста. Пополнение текста может осуществляться из сети Интернет при использовании специальных роботов-программ ("ботов", "пауков"), которые осуществляют анализ всей текстовой информации, размещенной в глобальной сети [3].

Рис. 3. Уменьшение диапазона подбора (изменение множества выборки)

В статье описан расширенный алгоритм метода "грубой силы". Применение данного алгоритма в АСРДкКИ обеспечит увеличение производительности системы за счёт оптимизации работы модуля мониторинга исходящего информационного трафика предприятия, основанной на применении дополнительного модуля кластеризации информации.

Литература

1. Хорев П.Б. Методы и средства защиты информации в компьютерных системах: учеб. пособие для студ. высш. учеб. заведений / П.Б.Хорев. - 4-е изд., стер. - М.: Издательский центр "Академия", 2008. - 256 с.

2. Панасенко С.П. Алгоритмы шифрования. Специальный справочник. - СПб.: БХВ-Петербург, 2009. - 576 с.

3. Карлова Т.В., Кузнецова Н.М. Разработка концепции обеспечения многоуровневого доступа к конфиденциальной информации // Вестник МГТУ "Станкин" № 2 (14), 2011, С.87-90.

Связь с автором: knm87@mail.ru