Ю. И. Еременко, О. Ю. Базарова, А. Ю. Соловьев

О РАЗРАБОТКЕ ИНТЕЛЛЕКТУАЛЬНОЙ НАДСТРОЙКИ К СИСТЕМАМ КОНТРОЛЯ УТЕЧКИ ИНФОРМАЦИИ

Введение

В современной коммерческой, некоммерческой, государственной или иной организации электронный документооборот стал обязательным элементом информационной инфраструктуры. Это означает, что документы, имеющие различную степень конфиденциальности, в большинстве компаний хранятся в электронном виде. В зависимости от соблюдения правил безопасности доступ к важным документам имеет ограниченный круг лиц. Каждый руководитель организации заинтересован в том, чтобы предотвратить несанкционированный доступ к корпоративной документации, то есть не допустить утечку информации.

Утечка корпоративной информации может происходить по двум основным направлениям:

1. Утечки информации извне вследствие получения злоумышленниками несанкционированного доступа к ресурсам корпоративных информационных систем;

2. Утечка информации внутри организации вследствие действий сотрудника организации (в терминологии информационной безопасности — инсайдера), имеющего полный или ограниченный доступ к корпоративной информации.

Отметим, что борьба с инсайдерами затруднительнее, чем предотвращение угроз извне, так как инсайдер, являясь сотрудником организации, не только имеет доступ к корпоративной информации, но также осведомлен о процессах, протекающих в организациях, важности той или иной информации и т. д. На рис. 1 приведена динамика роста утечек информации с 2006 г. по первую половину 2011 г., полученная на основании исследований группы компаний по обеспечению информационной безопасности организаций InfoWatch [9].

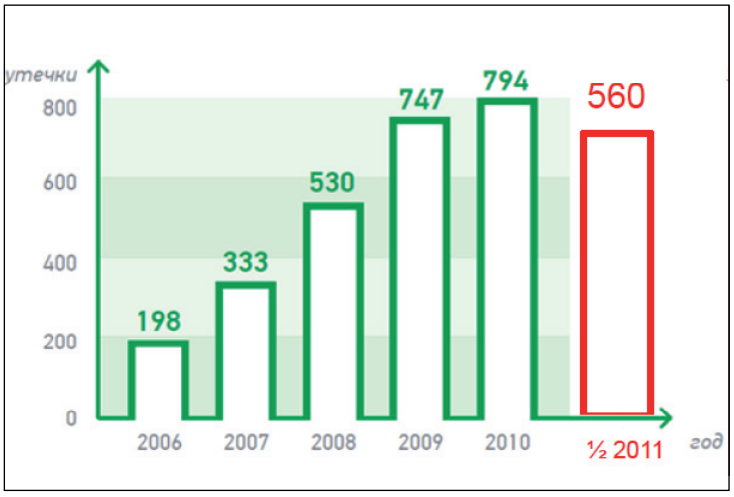

Рис. 1. Динамика числа утечек корпоративной информации

Исследования показывают ежегодный рост числа утечек, что делает проблему борьбы с инсайдерами актуальной.

По данным международной конференции DLP-Russia-2011 [9], множество традиционных решений, призванных обеспечить защиту информационных ресурсов, часто не справляются с этой задачей, поскольку призваны защищать не саму информацию, а объекты ИТ-инфраструктуры: серверы, места хранения и обработки информации, каналы передачи данных, но не информацию в целом. То есть

такие решения не дают возможности контролировать процесс обработки и передачи корпоративных данных с учетом того, что 10 % информации изменяются ежедневно и 10 % создаются заново. Для защиты именно информации, а не технологии элементов ИТ-комплекса, обеспечивающих доступ к ней, разработаны специализированные решения – DLP (Data Loss/Leak Prevention) системы.

Системы для защиты от внутренних угроз и утечек класса DLP стали массово выпускаться только после 2000 г. Основной целью таких систем является предотвращение передачи конфиденциальной информации за переделы корпоративной информационной системы. Интерес и спрос на DLP-системы достаточно велик, так как большинство российских и зарубежных компаний осознают, что серьезную угрозу безопасности представляют не столько вирусы и хакеры, сколько сами сотрудники компании, имеющие доступ к конфиденциальным данным.

Принцип работы рассматриваемого класса систем основывается на выявлении пакетов данных, которые инсайдер перемещает за пределы компании, и анализе таких данных на предмет содержания корпоративной информации, по результатам которого DLP-система должна принять решение пропустить или заблокировать передачу информации. Системы защиты от внутренних угроз и утечек информации применяют различные методы анализа корпоративной информации:

- анализ по словарю;
- лингвистический анализ;
- анализ транслита;
- анализ архивов;
- анализ рисунков;
- стеганография;
- анализ неклассифицированных данных.

Использование таких методов в комплексе увеличивает вероятность принятия системой правильного решения о блокировке передачи данных.

На рис. 2 представлена архитектура DLP-системы. Блок «Фильтрация» является ключевым, ведь именно в нем и происходят анализ данных и принятие основного решения.

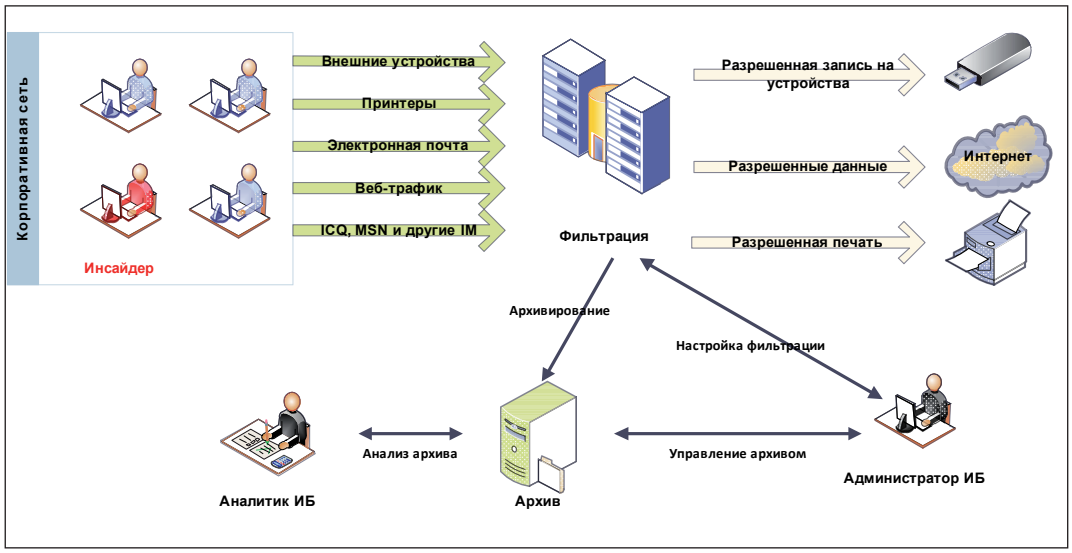

Рис. 2. Архитектура DLP-систем

Проблемы современных систем контроля утечек информации

Любая DLP-система имеет свой процент ложных срабатываний, когда вполне легальные документы признаются строго секретными и наоборот. Самый большой процент ложных срабатываний (около 25 %) отмечается на неструктурированных данных для необученной системы,

построенной на правилах категоризации, которые были встроены по умолчанию. В среднем процент ложных срабатываний на различных данных колеблется от 2 до 10. Увеличивают этот процент различные факторы:

- частые изменения существующей информации;
- создание новой информации;
- неструктурированная информация;
- сложность при определении того, какие именно данные нужно защищать.

Эксперты в области безопасности, а также разработчики DLP-систем считают достаточно приемлемым процентом ложных срабатываний менее 10, но, пользуясь методами искусственного интеллекта, можно повлиять и уменьшить это число.

В архитектуре DLP-систем важную роль играет офицер информационной безопасности (в дальнейшем аналитик), который в случае блокировки информации принимает окончательное решение о ее пропуске или нет, а также периодически проводит анализ журнала пропущенных DLP-системой данных на предмет утечки информации, содержащей корпоративную тайну.

Анализ существующих систем показал, что если степень защиты конфиденциальных данных критически высока, то система при обнаружении информации любого уровня конфиденциальности будет довольно часто обращаться к аналитику за экспертной оценкой информации.

Все вышеперечисленное приводит к возникновению серьезных проблем, а именно: к увеличению объема трудовых затрат аналитика и неспособности последнего влиять на уменьшение вероятности утечки информации, следствием чего является увеличение времени рассмотрения заблокированной информации, что зачастую может быть неприемлемо. Например, заблокированный банковский счет на оплату может рассматриваться аналитиком ввиду его занятости или неопытности от нескольких часов до нескольких дней, что недопустимо.

Здесь встает важный вопрос об использовании знаний аналитика системой для последующего принятия решения при работе с подобной информацией. Это позволит уменьшить нагрузку на аналитика и значительно сократить процент ложных срабатываний системы за счет полученного ранее опыта — знаний аналитика.

Проведя анализ научных трудов [1—8], а также опираясь на экспертные мнения специалистов в области информационной безопасности, участников международной конференции DLP-Russia-2011, мы пришли к выводу, что существующие системы не учитывают знаний аналитика и из-за того, что применение методов анализа не дает достаточной точности, это приводит к утечке либо к ложной блокировке информации. Мы ставим перед собой задачу увеличить эту точность путем разработки интеллектуальной надстройки над DLP-системой.

Нами предлагается разработка интеллектуальной надстройки к DLP-системам, основанной на аппарате экспертных систем. Такой подход позволит решить еще и задачу обучения работе аналитика на основе знаний, полученных ЭС (экспертной системой) от аналитика-эксперта и предоставленных в развернутом виде начинающему аналитику.

Анализ современных решений выявленных проблем

Отметим, что не существует экспертных систем, разработанных для DLP-систем.

Большинство трудов по информационной безопасности направлено на предотвращение атак и хищения информации извне, что определяет некоторую специфику при разработке экспертных систем и различных методов и алгоритмов.

При анализе литературы и научных исследований мы опирались на опыт разработки экспертных систем в области информационной безопасности и смежных областях.

В работе [4] представлено формализованное описание угроз ИБ организации:

$$U = \left(I_j^k, O_i^m, z_e, L_n\right), \quad (1)$$

где I_j^k — идентификатор источника угрозы, характеризующийся типом и расположением; O_i^m — идентификатор объекта угрозы, характеризующийся типом ресурса организации и целью нарушения аспекта ИБ (информационной безопасности); z_e — идентификатор e-го метода реализации угрозы; L_n — идентификатор n-го подмножества средств реализации угрозы. Разработана ЭС поддержки принятия решения, осуществляющая генерацию и выдачу рекомендаций по достижению необходимого уровня обеспечения информационной безопасности организации. Но решения принимаются по формированию требований к показателям защищенности именно физических ресурсов, что не позволяет защищать саму информацию, как того требует наша задача.

Другой подход [8] решает задачу автоматизированного проектирования средств защиты информации, служащих для анализа защищенности локальной сети и позволяющих пользователю самому проектировать базу уязвимостей компьютерной сети, использовать ее в качестве основы для обнаружения направлений несанкционированных действий злоумышленников. Здесь внимание уделено вопросам защиты технического уровня, и данный подход позволяет защититься от внешних атак, а внутренние утечки информации остаются возможными.

В работе [2] приведена экспертная система для оценки важности информации в информационной системе. Экспертная система определяет важность информации при помощи нечеткой логики, оценивает риски, влияющие на информационную систему, и дает рекомендации по устройству информационной системы. Но в данном случае экспертная система опять же дает рекомендации по защите от внешних несанкционированных атак, угроза действий инсайдеров не устраняется.

В [1] приведена экспертная система, построенная с применением байесовских сетей доверия, что позволяет аналитику применять решения в неопределенных ситуациях.

В работе по защите информационных ресурсов виртуального предприятия на основе многоагентной технологии [7], в которой предложен алгоритм принятия решений по оперативному управлению средствами защиты информации, основанный на байесовских сетях доверия и методе анализа иерархий, получены результаты, повышающие эффективность функционирования систем защиты информации на протяжении жизненного цикла только сложного технического изделия.

Подходы, основанные на байесовских сетях доверия в работах [7] и [1], интересны с точки зрения принятия решения при больших объемах противоречивой информации. Но в этих работах задачи решены только в частных случаях, применительно только к исследуемому специфическому объекту, что не может гарантировать успешность применения в анализируемой нами области. Для этого должны проводиться дополнительные исследования.

Работа [5] интересна методами построения баз знаний, где слабо структурированные знания могут быть противоречивыми. Рассматривается проблема извлечения именно практических знаний эксперта. В работе предложены алгоритмы извлечения и структуризации знаний, которые применимы к любым областям, по утверждению автора. Данные методы могли бы быть положены в основу предлагаемой интеллектуальной надстройки систем контроля утечек информации. Также отметим, что в работе [5] критикуются такие методы машинного обучения, как, например, нейронные сети. При этом в работе не упоминается про относительно новый аппарат, известный как иммунные сети, которые достаточно хорошо зарекомендовали себя не только в области машинного обучения, но и в области информационной безопасности в целом.

Также нами были рассмотрены типовые архитектуры экспертных систем. В частности, определенный интерес представляет гибридная модель экспертной системы, совмещающая несколько моделей представления знаний [6].

Выводы

Анализ проблемы функционирования систем предотвращения утечек информации показал, что проблема разработки интеллектуальной надстройки над DLP-системой является актуальной и важной задачей.

Для решения поставленной задачи необходимы дальнейшие исследования, а именно:

- создание базы знаний, выбор модели представления знаний;
- выбор архитектуры разрабатываемой экспертной системы;
- дальнейшее исследование методов и способов извлечения неструктурированных знаний эксперта в области информационной безопасности.

Мы полагаем, что при анализе извлечения неструктурированных и слабо классифицируемых знаний эксперта целесообразно и возможно применение аппарата иммунных сетей. Для этого необходимо провести дальнейшие исследования.

Предлагаемая к разработке интеллектуальная надстройка к DLP-системе позволит сократить время рассмотрения аналитиком заблокированной информации и, как следствие, повысить эффективность работы систем контроля утечки информации, а также повлияет на стабильность работы организации в целом.

СПИСОК ЛИТЕРАТУРЫ:

1. *Али Мансур Номан Мархуб*. Экспертная система поддержки принятия решений в интеллектуальной системе экологического мониторинга атмосферного воздуха промышленного региона: на примере г. Новомосковска Тульской области. Дисс. … канд. техн. наук. М., 2011. – 147 с.

2. *Аль Тамими Недаль Юсеф*. Экспертная система оценки устойчивости функционирования сетевых информационных систем при негативных внешних воздействиях. Дисс. ... канд. техн. наук. Тамбов, 2011. – 176 с.

3. *Белков Н. В.* Построение системы защиты персональных данных. Проблемы и задачи // Актуальные проблемы в науке и технике. Т. 2. Информационные технологии. Сборник трудов пятой Всероссийской зимней школы-семинара аспирантов и молодых ученых. 17–20 февраля 2010 г. Уфа: УГАТУ, 2011. С. 57–60.

4. *Королева Н. А.* Экспертная система поддержки принятия решений по обеспечению информационной безопасности организации. Дисс. … канд. техн. наук. Тамбов, 2006. – 198 с.

5. *Кочин Д. Ю.* Построение баз экспертных знаний для интеллектуальных обучающих систем. Дисс. ... канд. техн. наук. М., 2006. – 115 с.

6. *Кулинич А. А.* Разработка принципов и методов построения программных систем поддержки принятия решений в слабо структурированных ситуациях на основе моделирования знаний эксперта. Дисс. … канд. техн. наук. М., 2003. – 150 с.

7. *Погорелов Д. Н.* Защита информационных ресурсов предприятия на основе многоагентной технологии. Дисс. … канд. техн. наук. Тамбов, 2006. – 273 с.

8. *Рыбаков А. А.* Автоматизированное проектирование экспертных систем для защиты информации в локальных вычислительных сетях. Дисс. … канд. техн. наук. Рязань, 2004. – 198 с.

9. Сайт международной конференции DLP-Russia. URL: http://dlprussia.ru (дата обращения: 28.10.2011).