Владимир Ульянов

Классификация данных как аспект информационной безопасности

Разрабатывая проекты по информационной безопасности (ИБ), компании часто забывают об одном немаловажном аспекте — классификации данных. На первый взгляд классификация к безопасности и отношения-то не имеет. Однако практика показывает, что классификация — это не просто процесс, связанный с защитой информации, а основа для построения полноценной системы безопасности.

В помощь информационной безопасности

И все-таки — как классификация помогает безопасности? Рассмотрим простейшую систему защиты данных, которая обеспечивает разграничение прав доступа. При запуске такой системы в эксплуатацию в ней регистрируются пользователи, работники предприятия. Далее каждого из пользователей или группу пользователей необходимо наделить правами на доступ к... некоторой информации. Но вот к какой именно информации? Не раздавать же права на отдельные файлы, создавая списки километровой длины! В крайнем случае права могут быть предоставлены на доступ к директориям. При этом директории должны различаться по каким-либо критериям, например по принадлежности содержащихся в них документов какому-то лицу или подразделению компании. Кроме того, необходимо убедиться, что в этих директориях лежат именно «правильные» документы, соответствующие тому признаку, который был выбран для папки.

Вот пример довольно примитивной и жесткой классификации, лежащей в основе внедрения простейшей системы защиты. Не проведя даже такую относительно простую классификацию, невозможно будет назначить права пользователям.

Разумеется, в реальной корпоративной сети информация хранится не только в папках файловых хранилищ и файлах офисных приложений. Есть еще масса программного обеспечения корпоративного уровня, ERP-, CRM- и HRM-системы, базы данных. Но и данные, хранящиеся в этих системах, могут быть классифицированы. Кстати, чуть отвлекаясь от темы, стоит заметить, что внутри этих самых корпоративных систем данные уже классифицированы, упорядочены, а нередко и защищены. А вот на выходе из систем информация становится одновременно деклассифицированной и уязвимой.

Чем еще хороша классификация, и зачем она нужна? Прежде всего, классификация позволяет выявить ту информацию, которую следует защищать. То есть буквально отделить зерна от плевел. Это позволяет минимизировать количество конфиденциальной информации. Очевидно, что чем ее меньше, тем легче контролировать ее использование и перемещение.

Кроме того, при проведении классификации выявляются документы, бывшие когда-то секретными, но теперь уже потерявшие свою актуальность. Их можно отправить в архив или безвозвратно удалить.

Помимо этого классификация позволяет определить все места хранения информации, что может быть использовано для оптимизации бизнес-процессов в компании.

Ложки дегтя

О важности классификации данных в корпоративной среде, в том числе и для целей защиты информации, известно давно. В исследовании аналитического центра Perimetrix «Инсайдерские угрозы в России 2008» 77% респондентов сошлись во мнении, что классификация помогает поднять эффективность защиты от утечек. Тем не менее по состоянию на начало 2008 года почти половина (41%) компаний вообще никогда не проводила классификацию. И только 13% фирм классифицировали корпоративные данные в течение последнего года. Спустя год в отчете «Инсайдерские угрозы в России 2009» была опубликована информация о значительном росте числа компаний, которые занимались проблемой классификации. В течение 2008 года уже 25% фирм провели классификацию, а доля проигнорировавших ее снизилась до 33%. Тем не менее даже свежие цифры весьма скромны в абсолютном значении.

Почему же, признавая важность классификации данных, на практике компании ею не занимаются? Причин тому несколько. Прежде всего, классификация — процесс сложный, требующий немало времени и средств, а в противном случае — не отличающийся эффективностью. Еще одна проблема — поддержание актуальности. Чтобы классификация была эффективной, действительно помогала в деле защиты информации, ее необходимо постоянно поддерживать и актуализировать. А этот процесс может оказаться даже более трудоемким, нежели первоначальная классификация «с нуля». Получается, что нужно разработать специальные механизмы, позволяющие автоматически поддерживать актуальность классификации, либо начинать очередную итерацию классификации сразу же по окончании текущей. В противном случае естественный жизненный цикл данных в корпоративной среде очень скоро приведет к размытию правильных категорий. Именно поэтому многие компании даже не берутся за классификацию.

Выбирай на вкус

Существует множество мнений, как и какими средствами проводить классификацию. И выбор приходится делать уже с самого начала. Например, что лучше: раскидать документы с определенными признаками (классами) по директориям или создать некую электронную метку, куда записывать класс документа? Очевидно, что первый способ и проще и быстрее. Но что если документ относится к нескольким классам-папкам? Поместить по копии в каждую? И как затем отследить производные такого документа? Вариант с метками гораздо более гибкий, но и более сложный в реализации. В зависимости от выбранного варианта будет изменяться и ключевой аспект классификации, коим стоит признать, пожалуй, набор признаков, по которым затем разделят документы, — так называемый классификатор.

А вот методы, с помощью которых будет производиться проверка документов на соответствие признакам-классам (рис. 1), являются общими и справедливы для любого процесса классификации. Кратко рассмотрим каждый из них.

Классификация вручную — это самый надежный и точный метод. Но кто этим будет заниматься? Никто лучше человека — автора или заказчика документа (назовем его владельцем документа) — не разбирается в сущности информации. Однако пересмотреть и классифицировать сотни (тысячи, десятки тысяч...) документов, имеющихся в корпора-

Универсальность цифровых отпечатков заключается в том, что они могут сниматься с любого типа файлов в бинарном представлении. Таким образом могут быть классифицированы и графические, и аудио-, и другие типы файлов. Кроме того, если из файла может быть извлечен текст (как для морфологического анализа), цифровые отпечатки могут быть сняты и с самого текста, что повысит качество распознавания класса информации. Из недостатков метода цифровых отпечатков следует назвать солидную предварительную работу, необходимость создания базы данных эталонных отпечатков, чувствительность к изменениям файлов.

Замолвим слово об актуализации

Как уже говорилось, сама классификация — это еще полбеды. Помимо этого необходимо решить задачу по поддержанию классификации в актуальном состоянии. Тем не менее, проведя классификацию однажды, компании имеют достаточно возможностей поддерживать ее актуальность на протяжении некоторого времени. Доступные методы поддержания классификации приведены на рис. 2.

Помимо перечисленных ранее (см. рис. 1) методов появились опции:

- создание документов по шаблонам — в этом случае новые документы создаются администратором, который явно указывает классы информации. Пользователи же занимаются наполнением готовых шаблонов. Таким образом, в корпоративном пространстве не появляется новых неклассифицированных документов, однако проблема изменения классов с изменением содержания не решается;
- изменение классов по требованию — в какой-то момент владелец документа понимает, что класс документа не соответствует заявленному, и отправляет администратору заявку на изменение класса в соответствии с новым содержанием;
- изменение классов по расписанию — ценность информации может изменяться с течением времени, и для определенных категорий можно предусмотреть изменение классов по расписанию. Например, рассекречивать подробности рекламных кампаний трехлетней давности;
- наследование классов — предусматривает разработку программных механизмов наследования классов при изменении содержимого документов. Ведь большая доля документов в корпоративной среде не пишется заново, а составляется на основе других документов и с применением сторонних источников информации. Если источники, которые используются при составлении нового документа, уже классифицированы, наследование позволит автоматически добавить класс исходной информации к конечному документу.

Особенности практического применения

Чтобы провести классификацию данных на предприятии, вовсе не обязательно применять все описанные нами методики. Тем не менее, сочетая различные методики, можно добиться приемлемого качества классификации при разумных финансовых и временных затратах. Кроме того, некоторые методы могут вообще не помочь. Например, самый точный способ — цифровые отпечатки — не даст никакого результата, если в базе не обнаружатся схожие эталонные документы.

Понятно, что такие методы, как классификация по местам хранения и формальным признакам, являются очень неточными, а потому рекомендуются для использования только в редких случаях.

В любом случае выбор всегда остается за человеком. Это относится как к методике, так и к результатам работы средств автоматизации. Если оператор не согласен с предложенным результатом по итогам, например, классификации по формальным признакам, то он должен применять какой-то другой метод либо провести проверку содержимого вручную.

НОВОСТИ НОВОСТИ НОВОСТИ НОВОСТИ НОВОСТИ НОВОСТИ НОВОСТИ НОВОСТИ

DeviceLock отмечен респондентами как наиболее широко использующееся в российских организациях специализированное средство защиты

Результаты проведенного информационным порталом SecurityLab.ru исследования, призванного определить защищенность российских компаний от утечек данных через рабочие станции пользователей корпоративных ИС, подтверждают, что, несмотря на актуальность этой проблемы, она по-прежнему остается нерешенной в большинстве отечественных организаций.

Компания «Смарт Лайн Инк», мировой лидер в области разработки программных продуктов контроля доступа к портам ввода-вывода и периферийным устройствам компьютеров, объявила о том, что по результатам проведенного порталом SecurityLab.ru независимого исследования «Утечки информации через рабочие станции пользователей» ее продукт DeviceLock отмечен респондентами как наиболее широко использующееся в российских организациях специализированное средство защиты от локальных утечек данных с рабочих станций пользователей.

Исследование проводилось в январе 2009 года путем анкетирования посетителей популярного в среде ИТ-профессионалов портала SecurityLab.ru. В нем участвовали 498 специалистов, представляющих как крупные, так и небольшие организации различных отраслей отечественной экономики. В отчете об исследовании, полная версия которого доступна для скачивания на веб-сайте компании «Смарт Лайн», представлен анализ наиболее опасных каналов утечки информации, а также высокоуровневых сценариев защиты от утечек по различным каналам. Кроме того, специальный раздел посвящен обзору типов используемого программного обеспечения, которое применяется как для защиты, так и для расследования инцидентов. В завершающей части отчета приводится статистика планов российских компаний, а также их требований к организации, разрабатывающей средства защиты.

Результаты исследования адресуются специалистам по информационным технологиям (ИТ) и информационной безопасности (ИБ), работающим в организациях любых масштабов и отраслей. Они позволяют понять, каков текущий уровень внутреннего контроля российских компаний, их уровень защищенности от внутренних угроз, а также возможности по расследованию различных инцидентов безопасности. Результаты данного проекта помогают многим специалистам по-новому взглянуть на рынок внутренней ИБ и принять эффективные решения по развитию ИБ-инфраструктуры своих организаций.

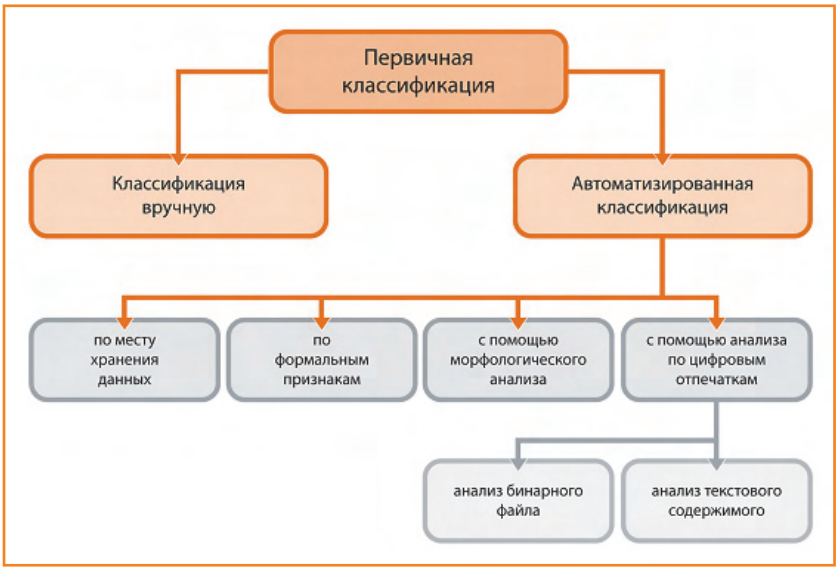
Наиболее опасным каналом утечки информации с рабочих станций сотрудников являются съемные носители (67,3%), за ними следует Интернет (50,4%) и электронная почта (32,2%).

Менее 17,6% российских компаний реализовали эффективные методы внутреннего контроля для всех основных каналов утечки информации: съемных носителей, мобильных устройств, принтеров, электронной почты и Интернета.

43,0% отечественных компаний не применяют технических средств для контроля наиболее опасного канала утечки — съемных носителей; 35,7% используют для этого только штатные средства операционных систем. И всего лишь 21,3% организаций применяют специализированные решения, самым популярным из которых является система DeviceLock (8,3%).

Не более 25,3% организаций обладают инструментальными средствами для расследования утечек по каждому из пяти основных каналов (съемные носители, мобильные устройства, сетевые принтеры, электронная почта и Интернет).

28,2% респондентов планируют внедрить систему защиты от утечек на рабочих станциях в течение ближайших трех лет, а еще 10,4% компаний уже находятся на этапе внедрения. При этом 15,2% организаций готовы внедрять исключительно российские продукты, а еще 15,0% — любые продукты с российскими сертификатами.



Методы первичной классификации данных (источник: Perimetrix, 2009)

тивной среде, очень трудно. Задача титаническая, явно не осуществимая в реальные сроки в большинстве организаций. Такой поход реализуем только в случае, если компания молодая и небольшая, иначе необходима автоматизация, упрощение.

Количество документов, которые требуется классифицировать, может быть уменьшено посредством отсеивания ненужных объектов. Например, можно игнорировать давно не изменявшиеся файлы. А это уже полноценный этап классификации. Разумеется, слепо откинуть старые документы, содержание которых может оказаться до сих пор актуальным, нельзя. Но как бы мы процесс ни автоматизировали, сколько бы признаков ни использовали, все равно финальный результат будет во многом вероятностным, неточным, требующим подтверждения со стороны человека.

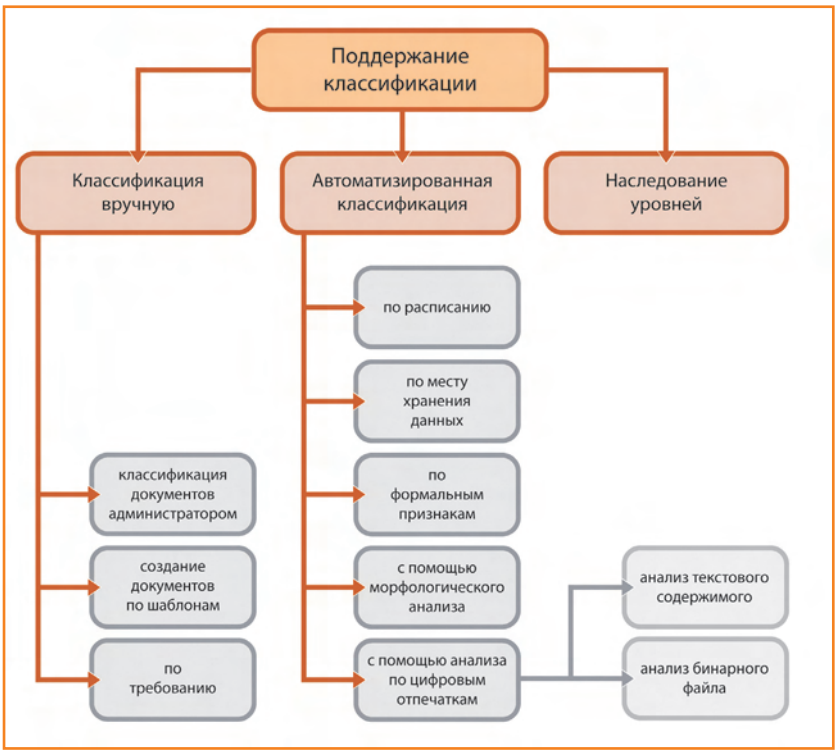
Автоматизированная классификация. Как уже упоминалось, один из простейших способов автоматизации — это учет некоторых формальных признаков документов, например таких, как название или тип файла, автор изменений, размер документа и т.д. Очевидно, что такой метод является крайне неточным, зато он достаточно быстрый и простой для реализации. Содержимое документов, несущее основную смысловую нагрузку, вообще не учитывается.

Чуть более точной может быть классификация, учитывающая места размещения информации. В большинстве компаний имеются определенные правила, регламентирующие места хранения информации. Например, информация о компаниях-контрагентах располагается на некотором сетевом диске U:, а на сервере с IP-адресом 192.168.1.17 лежат личные файлы сотрудников, фотографии с последней вечеринки и другая информация, некритичная к утечке.

Тем не мене метод будет хорош, только если сотрудники действительно соблюдают правила размещения информации. Опять же метод никак не учитывает несоответствие задекларированного содержимого документа его реальному наполнению. Если пользователь по ошибке или преднамеренно разместит в открытой директории конфиденциальный финансовый отчет, классификация по месту хранения также определит отчет как несекретный документ.

Другие типы автоматизированной классификации (морфологический анализ и анализ по цифровым отпечаткам) также являются вероятностными, однако, в отличие от описанных ранее, предполагают проверку внутреннего содержания. Впрочем, определенные ограничения налагают и данные методы. Так, морфологический анализ применим только к документам с текстовым содержанием, то есть к обычным «плоским» текстовым файлам либо к файлам, из которых можно изъять текстовую составляющую (например, файлы MS Office, Adobe Acrobat и др.). Посредством лингвистических методов, основанных на поиске заданных слов, словосочетаний и их взаимного расположения в тексте, компьютерная система может определить смысловую нагрузку документа. Помимо требований к формату классифицируемых файлов, морфологический анализ также предполагает серьезную предварительную работу по выделению морфологических признаков для каждого класса информации.

Наиболее точным и универсальным из всех методов стоит, пожалуй, признать анализ по цифровым отпечаткам. В ходе анализа по цифровым отпечаткам происходит сравнение проверяемых файлов и некоторых эталонных файлов, с которых отпечатки уже были сняты ранее. В основе определения собственно цифровых отпечатков лежат механизмы, схожие со снятием контрольных сумм с файлов. В результате если цифровые отпечатки проверяемых файлов или их частей схожи с эталонными отпечатками, то это указывает на близость их содержания и соответственно классов содержащейся информации.



Методы поддержания актуальной классификации (источник: Perimetrix, 2009)