WHITEPAPER

КЛАССИФИКАЦИЯ ДАННЫХ С ПОМОЩЬЮ ПРОДУКТОВ И ТЕХНОЛОГИЙ PERIMETRIX

KEEPING SECRETS SAFE

PERIMETRIX®

1. ВВОДНЫЕ

2. МНОГОМЕРНАЯ МОДЕЛЬ КАТЕГОРИЙ PERIMETRIX

3. ПЕРВИЧНАЯ КЛАССИФИКАЦИЯ И ЕЕ МЕТОДЫ

3.1. КЛАССИФИКАЦИЯ ВРУЧНУЮ
3.2. АВТОМАТИЗИРОВАННАЯ КЛАССИФИКАЦИЯ

4. ПОДДЕРЖАНИЕ АКТУАЛЬНОЙ КЛАССИФИКАЦИИ

4.1. КЛАССИФИКАЦИЯ ВРУЧНУЮ
4.2. АВТОМАТИЗИРОВАННАЯ КЛАССИФИКАЦИЯ
4.3. НАСЛЕДОВАНИЕ УРОВНЕЙ КОНФИДЕНЦИАЛЬНОСТИ

5. ЗАКЛЮЧЕНИЕ

6. О КОМПАНИИ PERIMETRIX

PERIMETRIX®

1. ВВОДНЫЕ

Классификация данных является одним из важнейших шагов на пути создания эффективной системы защиты от утечек.

Классификация данных является одним из важнейших шагов на пути создания эффективной системы защиты от утечек. Это подтверждают и результаты исследования «Инсайдерские угрозы в России 2008», в ходе которого компания Perimetrix опросила специалистов по информационным технологиям и информационной безопасности 472 организаций. 77% респондентов исследования считают, что классификация способствует повышению эффективности защиты от утечек (см. рис. 1).

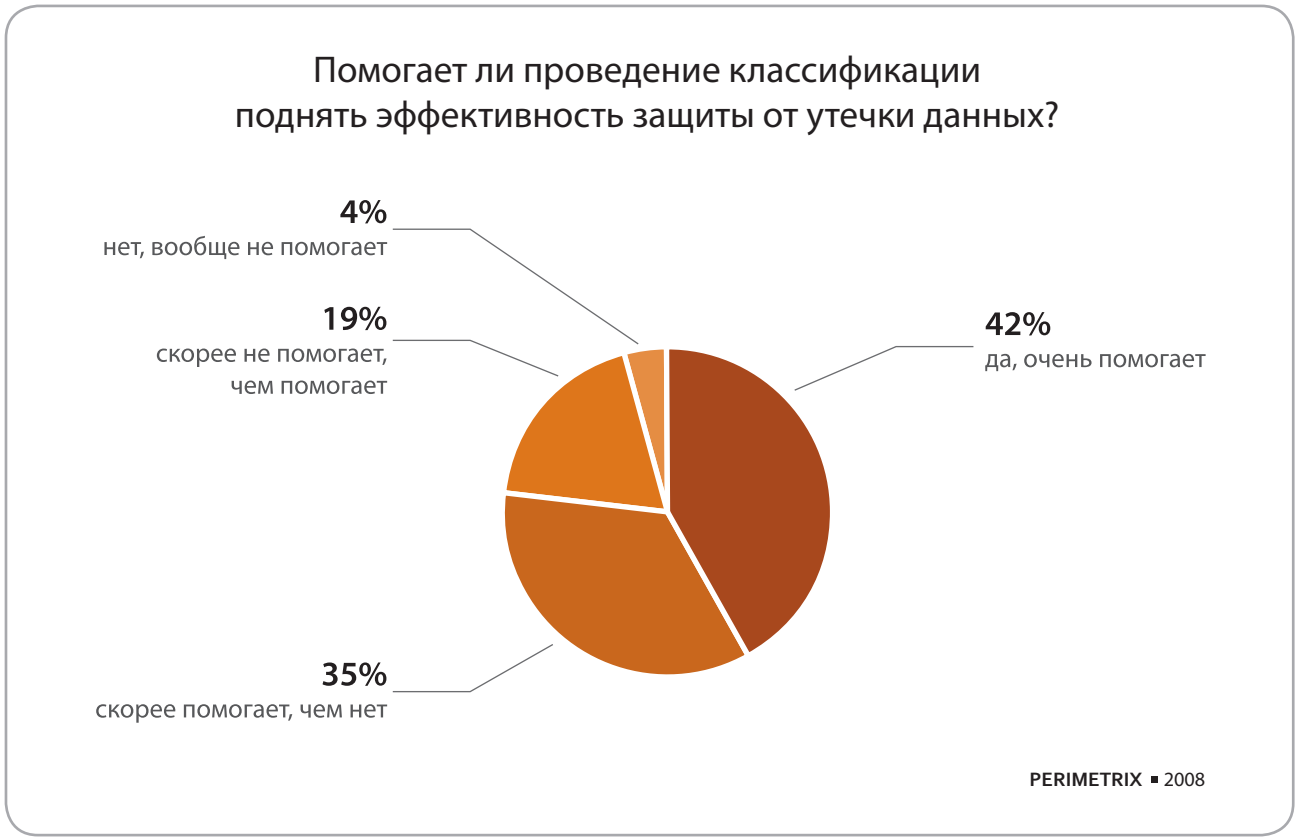

Рисунок 1. Как влияет классификация на эффективность защиты данных от утечки

Кроме того, классификация важна не только для безопасности, но и с точки зрения бизнес-процессов, поскольку позволяет упорядочить места хранения данных.

PERIMETRIX®

Тем не менее, на практике слишком мало организаций проводят классификацию. Виной тому — сопутствующие процессу трудности. В упоминавшемся выше исследовании «Инсайдерские угрозы в России 2008» респонденты выделили следующие основные сложности. Трудно поддерживать актуальность по прошествии времени (52%), сложен сам процесс классификации (23%), высокая стоимость (19%). В результате, если организации и проводят классификацию, то делают это достаточно редко (см. рис. 2), и актуальность таких сведений вызывает сомнения.

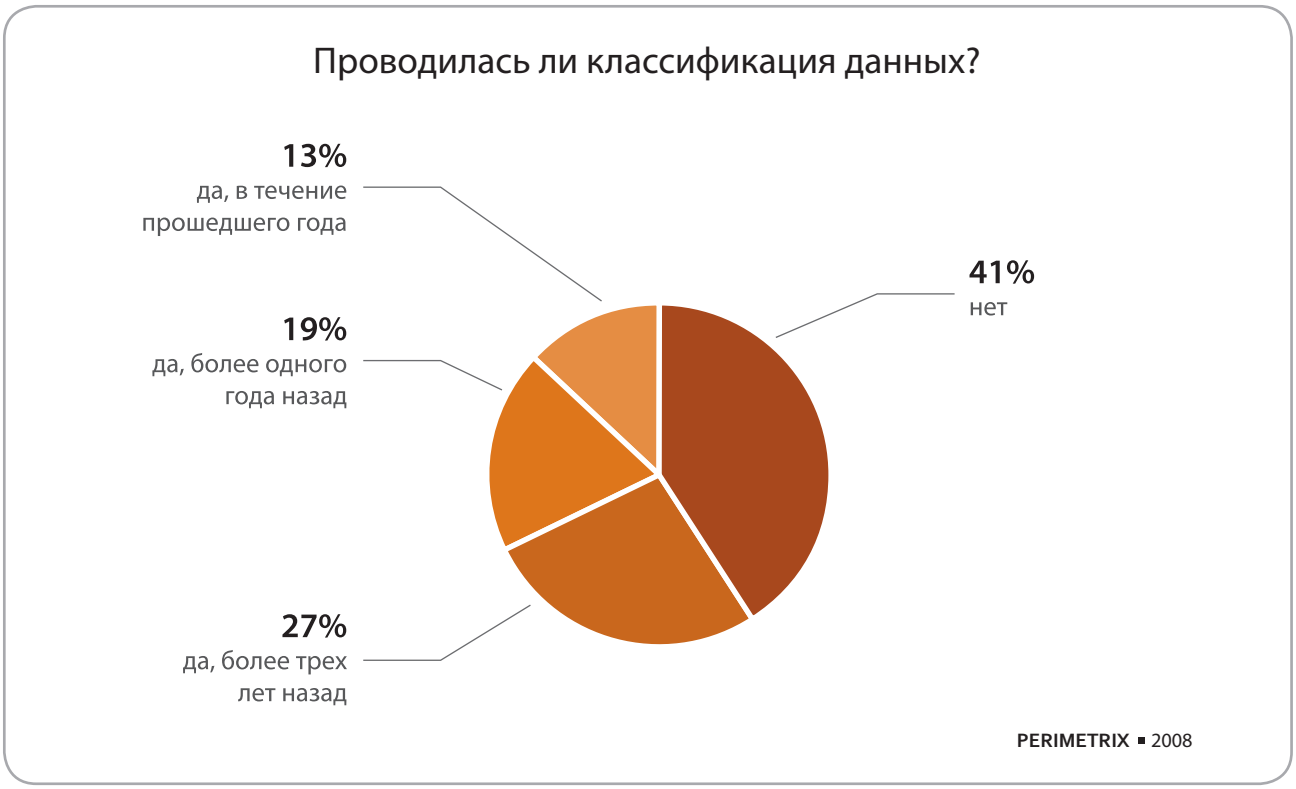

Рисунок 2. Классификация данных в организациях

2. МНОГОМЕРНАЯ МОДЕЛЬ КАТЕГОРИЙ PERIMETRIX

В соответствии с многомерной моделью Perimetrix, любые данные характеризуются множеством категорий, разнесенных по разнородным измерениям.

Прежде чем перейти к технологиям и способам классификации с помощью продуктов Perimetrix, поясним два концептуальных понятия, многомерной модели категорий и уровня конфиденциальности данных.

В соответствии с многомерной моделью Perimetrix, любые данные характеризуются множеством категорий, разнесенных по разнородным измерениям. Так, финансовый отчет фирмы «Пример» из города Уфа может быть описан категориями, принадлежащими следующим измерениям: «Функциональность», «Секретность», «География».

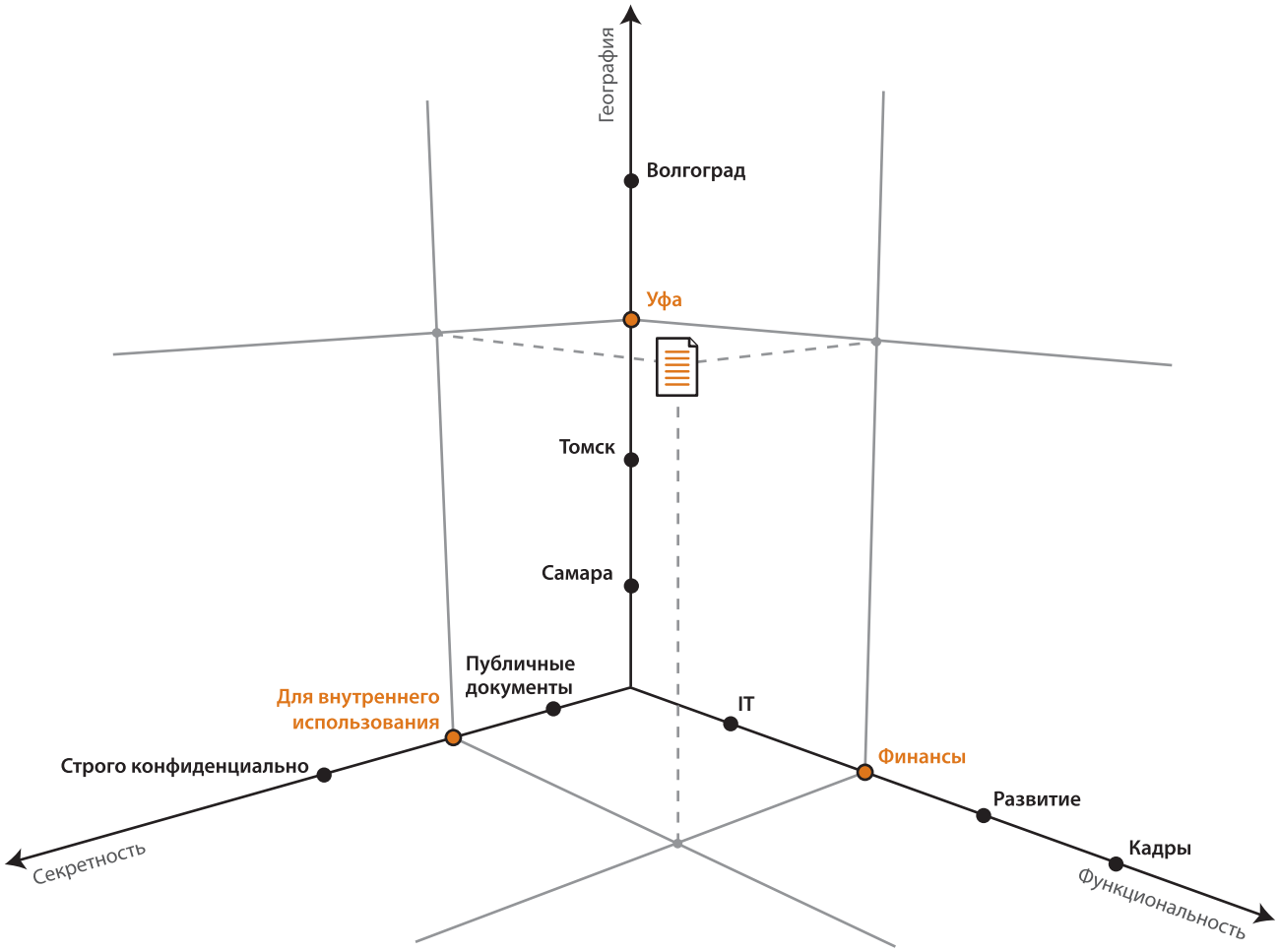

Рисунок 3. Графическое представление классификации конфиденциального документа в многомерной модели категорий.

PERIMETRIX®

Представим, что измерение «Функциональность» имеет набор равнозначных категорий, например, «IT», «Финансы», «Развитие», «Кадры». Очевидно, рассматриваемый отчет относится к финансовым документам.

Измерение «секретность» может быть иерархическим, т.е. «Публичные документы» — «Для внутреннего использования» — «Строго конфиденциально». Пусть финансовый отчет имеет категорию «для внутреннего использования».

Измерение «География» — древовидное, т.е. родительский уровень «Россия» имеет несколько ветвей-регионов, в том числе и «Самара», «Томск», «Уфа», «Волгоград».

В результате финансовый отчет описывается трехмерной моделью категорий, состоящей из измерений «Финансы», «Для внутреннего использования», «Уфа». Подобным образом, любой документ в системе, в соответствии со своим содержанием, может быть описан исключительно точно. Так же как это сделал бы обычный человек, а не машина. На рис. 3 представлен некий документ и его возможная классификация в модели категорий Perimetrix. Совокупный набор категорий различных измерений называется уровнем[1].

Классификация данных с помощью продуктов Perimetrix подразумевает выявление уровней конфиденциальности в соответствии с многомерной моделью категорий, принятой в организации. При внедрении продуктов Perimetrix требуется провести первичную классификацию данных, актуальность которой затем необходимо будет поддерживать. Кроме того, после внедрения Perimetrix классифицировать придется новые и входящие документы, еще не содержащие уровней конфиденциальности. Схемы методов классификации приведены в следующих главах.

1 Подробнее о многомерной модели категорий см. документ «Perimetrix: продукты и технологии»

PERIMETRIX®

3. ПЕРВИЧНАЯ КЛАССИФИКАЦИЯ И ЕЕ МЕТОДЫ

Первичная классификация данных проводится для всех документов в корпоративной сети при внедрении продуктов Perimetrix, а также для входящих и новых документов во время использования Perimetrix.

Первичная классификация данных проводится для всех документов в корпоративной сети при внедрении продуктов Perimetrix, а также для входящих и новых документов во время использования Perimetrix. Способы первичной классификации данных представлены на рис. 4.

3.1. КЛАССИФИКАЦИЯ ВРУЧНУЮ

Ручная классификация является наиболее точным, но и самым трудоемким способом. Ручная классификация подразумевает, что администратор системы самостоятельно проводит исследование сетевых ресурсов и вручную задает уровни конфиденциальности для отдельных файлов[1]. При этом администратор руководствуется названием документа, местом его хранения, другими атрибутами и, конечно же, содержанием.

1 Реализация методов классификации вручную описана в документации к Perimetrix® SafeSpace™

Рисунок 4. Методы первичной классификации данных

PERIMETRIX®

3.2. АВТОМАТИЗИРОВАННАЯ КЛАССИФИКАЦИЯ

Несмотря на то, что только осведомленный человек может наиболее точно отнести документ к определенному уровню (или уровням) конфиденциальности, в корпоративной среде такой метод малоприменим. Даже небольшие компании, существующие не более года, имеют десятки тысяч рабочих документов и их черновиков. Если же говорить о крупных компаниях с богатой историей, количество документов может быть на порядки выше. При этом действительно важные с точки зрения бизнеса документы нередко хранятся вместе с личными файлами сотрудников. Очевидно, что просмотреть все содержимое серверов и рабочих станций пользователей в разумные сроки не представляется возможным даже для группы администраторов. Поэтому Perimetrix реализует ряд методов, позволяющих автоматизировать труд администраторов по классификации данных.

- В отсутствие полноценной классификации, в корпоративной среде, тем не менее, обычно существуют правила хранения документов и разграничение доступа персонала к сетевым ресурсам. Например, бухгалтерские документы положено хранить на сетевом диске M:, причем отчетность из региональных филиалов находится на этом диске в директориях с соответствующими именами. Далее, на сетевом диске N: могут храниться файлы пользователей, не относящиеся деятельности компании. А на сервере с адресом 192.168.1.20 хранятся исключительно персональные данные клиентов организации.
  Основываясь на данном разделении, можно предположить, что документы из определенных источников (адресов, директорий и т.д.) имеют одинаковые уровни конфиденциальности, указываемые человеком. Таким образом производится автоматизация классификации данных, исходя из мест хранения. Данный способ достаточно просто реализовать, классификация производится очень быстро, однако точность определения уровней является весьма низкой.
- Другой способ автоматизации (по формальным признакам) не налагает ограничений на расположение классифицируемых файлов, и использует различные признаки и атрибуты документов. Например, предположить о сущности документа можно по его автору или заданной маске имени файла. Однако и этот способ не предусматривает просмотр реального содержимого, а потому является крайне неточным. Хотя следующие два метода также являются вероятностными, они учитывают именно содержимое документов и, обычно, показывают лучшие результаты.

PERIMETRIX®

- Использование морфологического анализа актуально для документов с текстовым содержанием. С помощью лингвистических методов система пытается определить смысловую сущность документа. В тексте ищутся заданные слова и сочетания, указывающие на принадлежность к уровням конфиденциальности.
  Данный метод требует скрупулезной предварительной работы. Необходимо выделить ключевые слова, а также указать их значимость для каждой из категорий данных. Кроме того, точность распознавания находится не на самом высоком уровне.
- Анализ по цифровым отпечаткам предполагает сравнение документов из корпоративной сети с эталонными документами, уже отнесенными к различным уровням конфиденциальности. Каждый файл разбивается на некоторое количество частей, для каждой из которых вычисляется контрольная сумма. Совпадений контрольных сумм частей проверяемых документов с контрольными суммами частей эталонных документов говорит о близком содержании документов.
  Цифровые отпечатки могут сниматься как с самих файлов в бинарном преставлении, так и с текста, если текст возможно выбрать из файла (например, из файлов MS Word).

Важно отметить, что и при автоматизации процесса классификации решающее слово остается за человеком, администратором системы. Именно он определяет, согласиться ли с выбором системы или провести дополнительное изучение документа. Что касается способов, то лучше всего использовать сочетание методов. Например, сначала отправить документ на анализ по цифровым отпечаткам, а затем, если уровень не найден, на морфологический анализ. Классифицировать документы по формальным признакам и местам хранения рекомендуется при проведении экспресс-классификации, а также в том случае, если корпоративные политики управления данными удовлетворяют условиям реализации методов[1].

1 Реализация методов автоматизированной классификации описана в документации к Perimetrix® SafeSpace™

PERIMETRIX®

4. ПОДДЕРЖАНИЕ АКТУАЛЬНОЙ КЛАССИФИКАЦИИ

Задача поддержания актуальной классификации является не менее важной, чем первичная классификация документов.

Данные и документы в корпоративной среде не являются статичными объектами и регулярно изменяются. Пользователи редактируют документы, создают новые. Поэтому задача поддержания актуальной классификации является не менее важной, чем первичная классификация документов. Способы поддержания актуальной классификации данных представлены на рис. 5.

Рисунок 5. Методы поддержания актуальной классификации данных

PERIMETRIX®

4.1. КЛАССИФИКАЦИЯ ВРУЧНУЮ

Помимо метода ручной классификации документов администратором для поддержания актуальных уровней Perimetrix предлагает еще 2 метода. Первый из них предполагает создание документов только в рамках предоставленных шаблонов. То есть пользователь, при создании документа получает шаблон с уже указанными уровнями конфиденциальности в качестве исходных. Второй способ подразумевает изменение уровней конфиденциальности по заявкам бизнес-владельцев документов.

4.2. АВТОМАТИЗИРОВАННАЯ КЛАССИФИКАЦИЯ

Помимо уже перечисленных в главе про первичную классификацию методов, Perimetrix предлагает еще один способ. Как показывает практика, в какой-то момент документы переходят из фазы активного использования в архив. Да и сама ценность информации изменяется в процессе жизни документа. Таким образом, можно предусмотреть изменение уровней конфиденциальности по расписанию. Например, уровни могут изменяться по истечении указанного срока давности. Данный метод является одной из практических реализаций концепции Secret Documents Lifecycle[1] (SDL), разработанной компанией Perimetrix. Согласно SDL, документы в процессе своего существования проходятся несколько стадий, каждая из которых несет риски утечки информации. Именно поэтому защищать документы необходимо с момента их создания и до безвозвратного удаления.

4.3. НАСЛЕДОВАНИЕ УРОВНЕЙ КОНФИДЕНЦИАЛЬНОСТИ

Как показывает практика, большая часть документов в корпоративной среде составляется не «с нуля», а собирается с использованием информации из различных источников, в том числе и других документов. Для того, чтобы уровни конфиденциальности автоматически переносились из используемых классифицированных источников, Perimetrix реализует механизм наследования уровней конфиденциальности.

1 Подробнее об SDL см. документ «Secret Documents Lifecycle»

PERIMETRIX®

При копировании данные из классифицированных документов (или других источников) не теряют своего уровня, а переносят его в новый документ. То есть копирование информации из классифицированного источника в буфер обмена сначала наделит буфер уровнями конфиденциальности этой информации. А затем перенесет те же уровни из буфера в конечный контейнер, например, какой-либо документ. Подобным образом происходит передача уровней конфиденциальности между любыми контейнерами информации. Потери уровней при перемещении классифицированной информации не происходит, и, таким образом, поддерживается актуальность классификации.

5. ЗАКЛЮЧЕНИЕ

С помощью продуктов Perimetrix классификация данных в корпоративной сети может быть проведена в кратчайшие сроки и с высоким качеством распознавания уровней конфиденциальности.

Классификация данных является важным процессом, как с точки зрения безопасности, так и с позиции бизнес-процессов. Perimetrix предлагает несколько способов классификации. Наиболее точным является метод ручной классификации. Однако в корпоративной среде этот подход малоприменим из-за большого количества документов, которые необходимо классифицировать. Поэтому компания Perimetrix разработала механизмы, позволяющие автоматизировать процесс.

Автоматизация предполагает возможность классификации данных по месту их хранения, формальным признакам, а также содержанию. Для анализа содержания используются вероятностные методы — морфологический анализ и анализ по цифровым отпечаткам. Кроме того, в продуктах Perimetrix имеется механизм наследования уровней конфиденциальности, позволяющий поддерживать актуальность классификации.

Таким образом, с помощью продуктов Perimetrix классификация данных в корпоративной сети может быть проведена в кратчайшие сроки и с высоким качеством распознавания уровней конфиденциальности.

PERIMETRIX®

6. О КОМПАНИИ PERIMETRIX

Компания Perimetrix разрабатывает уникальные решения для реализации режима секретности конфиденциальности данных. В отличие от конкурентов Perimetrix концентрирует свой потенциал, инновационный подход и уникальный опыт на создании корпоративной платформы внутренней информационной безопасности и интеграции с актуальными бизнес-процессами, организационной и технологической инфраструктурой заказчика. Наша цель — повышение стоимости бизнеса заказчиков за счёт поддержания непрерывности бизнес-процессов, минимизации риска утечки, повышения конкурентоспособности, а также установления плодотворных отношений с инвесторами и партнерами, соответствия государственным требованиям.

Благодаря реализации революционной концепции Secret Documents Lifecycle™ Perimetrix обеспечивает защиту секретных документов на всех этапах жизненного цикла, мониторинг каналов коммуникаций и аудит электронных операций. Технологическая основа системы — знание объекта защиты, контроль доступа и действий пользователей с целью предотвращения нарушения корпоративной политики.

Компания основана в 2007 году командой профессионалов, стоявших у истоков создания современных систем защиты от внутренних угроз информационной безопасности, и входит в Группу компаний «КомпьюЛинк» — лидирующий альянс на российском рынке информационных технологий.

PERIMETRIX®

Штаб-квартира Perimetrix

Российская Федерация,
119607, Москва,
Мичуринский проспект, д. 45

Телефон: +7 495 737 99 91
Факс: +7 495 737 99 92

info@perimetrix.com
www.perimetrix.com

KEEPING SECRETS SAFE

PERIMETRIX®