угрозы

- *Фишинг (Phishing)* — распространившиеся в последнее время способы перехвата важных данных пользователей (паролей, номеров кредитных карт и т.п.) с помощью техник социальной инженерии, когда пользователя ложным письмом или сообщением от той или иной организации пытаются заставить ввести определенные данные на сайте, контролируемом злоумышленником;
- *Spyware & Malware* — различные средства, позволяющие перехватывать данные или устанавливать контроль над компьютером. Существует множество разновидностей таких средств, которые различаются по степени опасности для компьютера — от простого показа рекламных сообщений до перехвата данных, вводимых пользователями, и захвата контроля над операциями с компьютером;

угрозы

- *вирусы и другой вредоносный код* — вирусы, черви и троянцы — давно известная угроза для ИТ-инфраструктуры. Но с каждым годом появляются новые модификации вредоносного кода, которые часто эксплуатируют уязвимости в существующем программном обеспечении, что позволяет им распространяться автоматически;
- *SPAM/SPIM* — нежелательные сообщения, передаваемые с помощью электронной почты (SPAM) или средств обмена мгновенными сообщениями (SPIM) заставляют пользователей тратить свое время на обработку нежелательной корреспонденции. В настоящее время СПАМ составляет более 70% всех передаваемых почтовых сообщений;

угрозы

- *атаки на инфраструктуру* — ИТ-инфраструктура компаний имеет очень важное значение, атаки с целью выведения ее из строя предельно опасны. Для них могут быть задействованы целые сети компьютеров, зараженных каким-либо вирусом, используемым для перехвата управления. Например, некоторое время назад был распространен вирус, содержавший в себе код, который должен был в определенное время начать распределенную атаку на сайты компании Microsoft с целью выведения их из строя. Зараженными оказались несколько миллионов компьютеров, и только ошибка в коде вируса не позволила выполнить планируемую атаку;
- *утечка бизнес-информации* — предотвращение таких утечек является одной из главных задач продуктов контентной фильтрации. Утечка важной информации может нанести компании непоправимый ущерб, порой сравнимый с потерей основных средств производства. Поэтому во многих продуктах развиваются средства для определения каналов скрытой передачи данных, таких например, как применение стеганографии;

угрозы

- *угроза судебного преследования* — этот вид угроз крайне актуален для компаний, если их сотрудники могут пользоваться файлообменными сетями, скачивая и/или распространяя музыку, фильмы и другое содержимое, защищенное авторским правом. Судебное преследование возможно и за распространение клеветнической и/или порочащей информации, касающейся третьих лиц.

Фильтрация Web-трафика

- Одной из наиболее важных тенденций развития продуктов контентной фильтрации в части контроля Интернет-трафика является переход от использования баз данных категорий сайтов к определению категории сайта по его содержимому.
- Это стало особенно актуально с развитием различных порталов, которые могут содержать наполнение разных категорий, изменяющееся во времени и/или подстраиваемое под настройки клиента.
- Использование шифрованных каналов для взаимодействия с Интернет-сайтами обеспечивает защиту данных от перехвата третьими лицами, но в то же время, по этим каналам передачи данных могут происходить утечка важной информации или проникновение вредоносного кода в компьютерные системы.

Подходы к категоризации сайтов и данных

- использование предопределенных баз категорий сайтов с регулярным обновлением списков сайтов и категорий;
- категоризация данных на лету путем анализа содержимого страниц;
- использование данных о категории, информацию о принадлежности к которой предоставляет сам сайт

Предопределенные базы категорий сайтов

- Использование заранее подготовленных баз адресов сайтов и связанных с ними категорий — давно используемый и хорошо зарекомендовавший себя метод.
- В настоящее время такие базы предоставляют многие компании

Предопределенные базы категорий сайтов

- Категоризация данных и формирование баз категорий обычно производится в полуавтоматическом режиме — сначала выполняются анализ содержимого и определение категории с помощью специально разработанных средств, которые даже могут включать в себя системы распознавания текстов в картинках.
- А на втором этапе полученная информация часто проверяется людьми, принимающими решение о том, к какой категории можно отнести тот или иной сайт.

Предопределенные базы категорий сайтов

- использование локальной базы категоий с регулярным ее обновлением. Данный метод очень удобен для больших организаций, имеющих выделенные серверы фильтрации и обслуживающие большое количество запросов;
- использование базы категорий, размещенной на удаленном сервере. Данный метод часто применяется в различных устройствах — небольших межсетевых экранах, ADSL-модемах и т.п. Использование удаленной базы категорий немного увеличивает нагрузку на каналы, но обеспечивает использование актуальной базы категорий.

Предопределенные базы категорий сайтов

- К преимуществам применения предопределенных баз категорий можно отнести то, что предоставление или запрет доступа производится еще на этапе выдачи запроса клиентом, что может существенно снизить нагрузку на каналы передачи данных.

Предопределенные базы категорий сайтов

- Недостаток использования данного подхода — задержки в обновлении баз категорий сайтов, поскольку для анализа потребуется некоторое время.
- Кроме того, некоторые сайты достаточно часто меняют свое наполнение, из-за чего информация о категории, хранящаяся в базе адресов, становится неактуальной.
- Некоторые сайты также могут предоставлять доступ к разной информации, в зависимости от имени пользователя, географического региона, времени суток и т.п

Категоризация данных на лету

- Один из простых вариантов реализации такого решения — использование байесовских алгоритмов, которые себя достаточно хорошо зарекомендовали в борьбе со спамом.
- Однако у этого варианта есть свои недостатки — необходимо его периодически доучивать, корректировать словари в соответствии с передаваемыми данными.
- Поэтому некоторые компании применяют более сложные алгоритмы определения категории сайта по содержимому в дополнение к простым способам.
- Например, компания ContentWatch предоставляет специальную библиотеку, которая выполняет анализ данных согласно лингвистической информации о том или ином языке и на основании этой информации может определять категорию данных.

Категоризация данных на лету

- Категоризация данных на лету позволяет быстро реагировать на появление новых сайтов, поскольку информация о категории сайта не зависит от его адреса, а только от содержания.
- Но такой подход имеет и недостатки — необходимо проводить анализ всех передаваемых данных, что вызывает некоторое снижение производительности системы.
- Второй недостаток — необходимость поддержания актуальных баз категорий для различных языков. Тем не менее, некоторые продукты применяют этот подход с одновременным использованием баз категорий

Данные о категории, предоставляемые сайтами

- Кроме баз данных адресов и категоризации содержимого на лету существует и другой подход к определению категории сайтов — сайт сам сообщает о том, к какой категории он относится.
- Этот подход в первую очередь предназначен для использования домашними пользователями, когда, например, родители или учителя могут задать политику фильтрации и/или отслеживать, какие сайты посещаются.

Данные о категории, предоставляемые сайтами

- PICS (Platform for Internet Content Selection) — спецификация, разработанная консорциумом W3 около десяти лет назад и имеющая различные расширения, направленные на обеспечение надежности рейтинговой системы.
- Для контроля может использоваться специальное разработанное программное обеспечение, доступное для загрузки со страницы проекта.

Данные о категории, предоставляемые сайтами

- ICRA (Internet Content Rating Association) — новая инициатива, разрабатываемая независимой некоммерческой организацией с тем же названием.
- Основная цель данной инициативы — защита детей от доступа к запрещенному содержимому.
- Данная организация имеет соглашения с множеством компаний (крупные телекоммуникационные и компании-разработчики ПО) для обеспечения более надежной защиты

Данные о категории, предоставляемые сайтами

- ICRA предоставляет программное обеспечение, которое позволяет проверять специальную метку, возвращаемую сайтом, и принимать решение о доступек этим данным.
- Программное обеспечение работает только на платформе Microsoft Windows, но благодаря открытой спецификации существует возможность создания реализаций фильтрующего ПО и для других платформ

Данные о категории, предоставляемые сайтами

- К достоинствам этого подхода можно отнести то, что для обработки данных нужно только специальное программное обеспечение и нет необходимости обновлять базы адресов и/или категорий, так как вся информация передается самим сайтом.
- Но недостатком является то, что сайт может указывать неправильную категорию, а это приведет к неправильному предоставлению или запрещению доступа к данным.

Интеграция с внешними системами

- Во многих случаях достаточно острым становится вопрос об интеграции систем контентного анализа с другими системами.
- При этом системы контентного анализа могут выступать как клиентами, так и серверами или в обеих ролях сразу.

Протокол ICAP

- ICAP принят группой Internet Engineering Task Force (IETF) в качестве стандарта.
- В качестве клиента ICAP выступает система, через которую передается трафик. Система, выполняющая анализ и обработку данных, называется сервером ICAP.
- Серверы ICAP могут выступать в роли клиентов для других серверов, что обеспечивает возможность стыковки нескольких сервисов для коллективной обработки одних и тех же данных.

- Решение о том, какие из передаваемых данных будут обрабатываться, принимается клиентом ICAP, в некоторых случаях это делает невозможным полный анализ данных. Настройки клиента полностью зависят от его реализации, и во многих случаях невозможноих изменить.
- После получения данных от клиента сервер ICAP выполняет их обработку, а если это необходимо, то и модификацию данных. Затем данные возвращаются клиенту ICAP, и он их передает дальше серверу или клиенту, в зависимости от того, в каком направлении они передавались.
- Наиболее широкое применение протокол ICAP нашел в продуктах для защиты от вредоносного кода, поскольку он позволяет использовать эти проверки в различных продуктах и не зависит от платформы, на которой выполняется клиент ICAP.

недостатки использования ICAP

- дополнительные сетевые взаимодействия между клиентом и сервером несколько замедляют скорость передачи данных между внешними системами и потребителями информации;
- существуют проверки, которые необходимо выполнять не на клиенте, а на сервере ICAP, такие как определение типа данных и т.п. Это актуально, поскольку во многих случаях клиенты ICAP ориентируются на расширение файла или на тип данных, сообщенный внешним сервером, что может стать причиной нарушения политики безопасности;
- затрудненная интеграция с системами, использующими протоколы, отличные от HTTP, не позволяет использовать ICAP для глубокого

Протокол OPES

- В отличие от ICAP протокол OPES разрабатывался с учетом особенностей конкретных протоколов.
- Кроме того, при его разработке учитывались недостатки протокола ICAP, такие как отсутствие установления подлинности клиентов и серверов, отсутствие аутентификации и др

- имеются требования к реализации клиентов OPES, что делает возможным более удобное управление ими — задание политик фильтрации и т.п.;
- потребитель данных (пользовательили информационная система) может оказывать влияние на обработку данных. Например, при использовании автоматических переводчиков получаемые данные могут автоматически переводиться на тот язык, который используется пользователем;
- системы, предоставляющие данные, также могут оказывать влияние на результаты обработки;

- серверы обработки могут использовать для анализа данные, специфичные для протокола, по которому данные были переданы клиенту OPES;
- некоторые серверы обработки данных могут получать более важные данные, если они находятся в доверительных отношениях с клиентом OPES, потребителями и/или поставщиками информации

Контроль передачи шифрованных данных

- Контроль передачи данных, пересылаемых по зашифрованным каналам, является, наверное, самой важной задачей для организаций, сотрудники которых имеют доступ к Интернет-ресурсам.
- Для реализации этого контроля существует подход, называемый "Man-in-the-Middle" (в некоторых источниках его также называют "Main-in-the Middle"), который может использоваться злоумышленниками для перехвата данных.

Процесс обработки данных выглядит следующим образом

- в Интернет-броузер пользователя устанавливается специально выписанный корневой сертификат, который используется прокси-сервером для подписывания сгенеренного сертификата (без установки такого сертификата, броузер пользователя будет выдавать сообщение о том, что подписывающий сертификат выдан недоверенной организацией);
- при установлении соединения с прокси-сервером происходит обмен данными, и в браузер передается специально сгенерированный сертификат с данными сервера назначения, но подписанный известным ключем, что позволяет прокси-серверу расшифровывать передаваемый траффик;
- расшифрованные данные анализируются так же, как и обычный HTTP-трафик;
- прокси-сервер устанавливает соединение с сервером, на который должны быть переданы данные, и использует для шифрации канала сертификат сервера;
- возвращаемые от сервера данные расшифровываются, анализируются и передаются пользователю, зашифрованные сертификатом прокси-сервера.

Проверка подлинности сертификатов

- В случае несовпадения система может заблокировать доступ к сайтам или осуществить доступ после явного подтверждения пользователем.
- Обработка данных при этом выполняется практически тем же способом, что и при анализе данных, передаваемых по шифрованным каналам, только в этом случае анализируются не данные, а сертификат, предоставляемый сервером.

Фильтрация почтового трафика

- сравнение получаемых сообщений с имеющейся базой сообщений. При сравнении могут применяться различные методики, включая использование генетических алгоритмов, которые позволяют вычленить ключевые слова даже в случае их искажения;

Фильтрация почтового трафика

- динамическая категоризация сообщений по их содержимому.
- Позволяет очень эффективно определять наличие нежелательной корреспонденции.
- Для противодействия этому методу распространители спама используют рассылку сообщений в виде изображения с текстом внутри и/или наборы слов из словарей, которые создают шум, мешающий работе данных систем. Однако уже сейчас для борьбы с таким спамом, начинают использовать различные методы, такие как вейвлет-анализ и/или распознание текста в изображениях;

Фильтрация почтового трафика

- серые, белые и черные списки доступа позволяют описывать политику приема почтовых сообщений с известных или неизвестных сайтов.
- Применение серых списков во многих случаях помогает предотвратить передачу нежелательных сообщений за счет специфики работы ПО, рассылающего спам.
- Для ведения черных списков доступа могут использоваться как локальные базы данных, управляемые администратором, так и глобальные, пополняемые на основе сообщений пользователей со всего мира. Однако использование глобальных баз данных чревато тем, что в них могут попасть целые сети, в том числе и содержащие "хорошие" почтовые сервера.

Фильтрация почтового трафика

- Для борьбы с утечками информации используются самые разные способы, основанные на перехвате и глубоком анализе сообщений в соответствии со сложной политикой фильтрации.
- В этом случае возникает необходимость корректного определения типов файлов, языков и кодировок текстов, проведения семантического анализа передаваемых сообщений.
- Еще одно из применений систем для фильтрации почтового трафика — создание шифрованных потоков почты, когда система автоматически подписывает или шифрует сообщение, а на другом конце соединения производится автоматическая расшифровка данных.
- Этот функционал очень удобен, если вы хотите обрабатывать всю исходящую почту, но она должна доходить до адресата в зашифрованном

Фильтрация мгновенных сообщений

- управление доступом по отдельным протоколам;
- контроль используемых клиентов и т.п.;
- контроль доступа отдельных пользователей:
- разрешение пользователю общения только в пределах компании;
- разрешение пользователю общения только с определенными пользователями вне компании;
- контроль передаваемых текстов;
- контроль передачи файлов. Объектами контроля являются:
  - размер файла;
  - тип и/или расширение файла;
- направление передачи данных;
- контроль наличия вредоносного содержимого;
- сохранение передаваемых данных для последующего анализа

Фильтрация VoIP

- Существуют стандартизированные протоколы для обмена информацией
- Session Instantiation Protocol (SIP),
- IETF и H.323,
- ITU.
- Эти протоколы являются открытыми, что делает возможным их обработку.
- Существуют протоколы, разработанные конкретными компаниями, которые не имеют открытой документации, что сильно затрудняет работу с ними.

Фильтрация VoIP

- продукты, которые позволяют определить и блокировать VoIP-трафик;
- продукты, которые могут определить, захватить и проанализировать VoIP-трафик

продукты, которые позволяют определить и блокировать VoIP-трафик

- продукты компании "Dolphian", позволяющие определить и разрешить или запретить VoIP-трафик (SIP и Skype), который инкапсулирован в стандартный HTTP-трафик;
- продукты компании Verso Technologies;
- разные виды межсетевых экранов, обладающие такой возможностью

Фильтрация peer-to-peer (P2P)

- распространение вредоносного кода;
- утечка информации;
- распространение данных, защищенных авторским правом, что может привести к судебному преследованию;
- снижение производительности труда;
- повышенная нагрузка на каналы передачи данных