Об авторе

Майкл Коллинз является руководителем исследовательских работ для RedJack, LLC, компании по сетевой безопасности и анализу данных, расположенной в Вашингтоне, округ Колумбия. До работы в RedJack доктор Колллинз был членом технического штата в CERT-сети, ситуативная группа осведомленности в университете Карнеги-Меллон. Его основное внимание направлено на инструментарий сети и анализ трафика, особенно на анализ больших наборов данных трафика. Коллинз получил степень доктора по электротехнике в университете Карнеги-Меллон в 2008 г. Он имеет степени магистра и бакалавра этого университета.

Оглавление

Предисловие

Эта книга – обо всем, что связано с сетями: их мониторинге, изучении и использовании результатов этого изучения с целью улучшения. «Улучшение» в данном контексте означает повышение безопасности сети, но я не думаю, что мы владеем достаточным количеством терминов и знаний, чтобы сказать наверняка. Во всяком случае, пока. В попытке обеспечить безопасность мы пытаемся достичь чего-то более конкретного и осязаемого – *ситуационной осведомленности*.

Термин «ситуационная осведомленность» часто используется в вооруженных силах и буквально означает понимание среды, в которой вы работаете.

В нашем случае ситуационная осведомленность включает также понимание компонентов сети и их работы. Зачастую мы *сильно* далеки от понимания настроек сети и первоначальных принципов ее построения.

Для понимания важности ситуационной осведомленности представьте свой дом и посчитайте количество веб-серверов в нем. Вы посчитали ваш беспроводной роутер? А кабельный модем? Принтер? А веб-интерфейс сервера печати? Не забыли ли включить сюда свой телевизор?

Не каждый специалист по информационным технологиям отнесет вышеперечисленные устройства в разряд веб-серверов. Тем не менее встроенные веб-серверы используют протокол HTTP, у них есть уязвимости, количество которых растет, поскольку на смену специализированным протоколам управления приходит веб-интерфейс. Взломщики будут атаковать встроенные системы, не раздумывая, чем они фактически являются: SCADA-система – не что иное, как Windows-сервер с парой интересных дополнительных каталогов, а аппарат МРТ – готовый к эксплуатации бот для рассылки спама.

Эта книга о том, как собирать данные и анализировать сети с целью понимания принципов их использования. Особое внимание уделяется анализу – процессу сбора данных о безопасности и принятия решительных мер на их основе. Подчеркиваю, что *решительные меры* в данном контексте – ключевое слово, поскольку эффективные меры по обеспечению безопасности – это запрет на определенные действия. Политика обеспечения безопасности обязывает говорить людям, чего делать не стоит (или, в более требовательном варианте, что они делать *должны*): не использовать Dropbox в качестве хранилища для корпоративных данных, осуществлять вход в систему при помощи пароля и аутентификатора RSA и не копировать весь сервер проекта целиком и не продавать его конкурентам. Когда мы принимаем решения по обеспечению безопасности, мы вторгаемся в рабочий процесс сотрудников, и должны иметь для этого очень веские основания.

Все системы безопасности целиком и полностью зависят от пользователей, которые осознают необходимость безопасности и воспринимают меры по ее обеспечению как вынужденное зло. Безопасность зиждется на людях, на пользователях системы, которые соблюдают определенные правила, а также на аналитиках и программах мониторинга, помогающих выявлять случаи их нарушения. Безопасность – лишь в небольшой степени техническая задача. Информационная

безопасность предполагает борьбу с невероятно творческими людьми, постоянно ищущими новые способы завладеть вашими технологиями. И в борьбе с этой постоянно изменяющейся угрозой вам необходимо добиться сотрудничества как со стороны защитников, так и со стороны пользователей. Неверно выстроенная политика безопасности вынудит сотрудников обходить меры безопасности, чтобы выполнить свою работу, или попросту нервничать, а это добавит работы специалистам по безопасности.

Акцент на решительности мер и цель достичь безопасности – это факторы, отличающие данную книгу от более общих текстов по анализу и обработке данных. Раздел, посвященный анализу, включает в себя методы статистического анализа и анализа данных из различных дисциплин, но самое пристальное внимание уделяется пониманию структуры сети и решениям, которые помогут защитить ее. В этой связи я сократил теоретическую часть до минимума и сконцентрировался на механизмах обнаружения вторжений. Проблема анализа безопасности состоит в том, что объекты наблюдения не только знают, что за ними следят, но и делают все возможное, чтобы этому воспрепятствовать.

МРТ и ноутбук генерального

Несколько лет назад я общался со специалистом по безопасности, работающим в основном для университетской больницы. Он рассказал, что самым загруженным устройством в его сети был томограф. В ретроспективе это легко объяснимо. «Только подумайте, – сказал он мне, – МРТ – это медицинское оборудование, а это значит, что на нем может использоваться лицензионная версия Windows. Поэтому каждую неделю кто-то взламывал его и устанавливал на него спам-бот. Спам начинал идти приблизительно в среду». Когда я спросил его, почему он просто не отключил томограф от интернета, он сказал, пожав плечами, что докторам были нужны их снимки. Он был первым специалистом с такой проблемой, которого я встретил, но не был последним. Мы сталкиваемся с подобной проблемой в любой организации, иерархия которой включает в себя высокие должности: доктора, старшие партнеры, генеральные директора. Вы можете создать сколь угодно много рубежей защиты, но если генеральный директор хочет взять рабочий ноутбук, чтобы его внучка поиграла в Neopets (Новые питомцы) в выходные, то в понедельник вы получите зараженный ноутбук, требующий ремонта.

Чтобы развить свою мысль, я продолжу. Я твердо уверен в том, что самый эффективный способ защитить сети – сохранять и защищать *только* то, что вам действительно нужно сохранить и защитить. Я так считаю, потому что информационная безопасность всегда будет требовать участия людей в мониторинге и расследовании. Модели атак постоянно меняются, поэтому, когда мы используем автоматизированные средства защиты, мы обнаруживаем, что взломщики теперь могут использовать их для атаки на нас самих[1].

Как специалист по безопасности я твердо уверен в том, что безопасность должна доставлять неудобство, быть хорошо организованной и вводить жесткие огра-

[1] Рассмотрим автоматическую блокировку аккаунтов после некоторого числа неудачных попыток ввода пароля, когда логин – это адрес электронной почты. Представьте, сколько аккаунтов можно заблокировать таким способом.

ничения. Безопасность должна быть искусственным поведением, распространяющимся на активы, которые необходимо сохранить. Поведение должно быть искусственным, потому что последняя линия защиты в любой защищенной системе – это *люди*. А люди, полностью вовлеченные в вопросы безопасности, должны быть недоверчивыми, выискивающими подозрительные явления с упорством параноика. Это не самый лучший способ прожить жизнь, поэтому, чтобы сделать ее сносной, мы должны обеспечить безопасность лишь того, что необходимо. Пытаясь уследить за всем, вы теряете ту грань, которая помогает вам защищать только то, что действительно имеет значение.

Поскольку безопасность доставляет неудобство, эффективные специалисты по безопасности должны *уметь убедить* пользователей в необходимости изменить свой привычный режим работы и плясать под их дудку, а в противном случае ограничить деятельность пользователей с целью предотвратить гипотетическую атаку в будущем. В этой связи специалисту необходимо определить решение, подкрепить его информативно и продемонстрировать риски своей аудитории.

Процесс анализа данных, описанный в этой книге, направлен на развитие знаний в области безопасности с целью принятия эффективных решений в этой сфере. Это могут быть экспертные решения: реконструкция событий постфактум с целью определить, почему произошла атака и что способствовало ее осуществлению, или оценить причиненный ущерб. Также можно прибегнуть к профилактическим мерам: установка ограничителей скорости передачи, установка систем обнаружения вторжений или разработка стратегий, которые могут ограничить воздействие взломщика на сеть.

Целевая аудитория

Анализ информационной безопасности – это молодая дисциплина, поэтому не существует четко определенной совокупности знаний, которыми нужно обязательно владеть. Данная книга предлагает те аналитические методы, которые я или другие специалисты по безопасности использовали за последние 10 лет и видели отличный результат.

Целевая аудитория этой книги – сетевые администраторы и специалисты по операционной безопасности, персонал Центров управления сетями (NOC) и все те, кто регулярно использует консоль СОВ. Я надеюсь, что вы уже знакомы с инструментами TCP/IP, такими как netstat, а также владеете базовыми статистическими и математическими навыками.

Кроме того, я надеюсь, что вы имеете представление о языках программирования. В этой книге я использую излюбленный мной Python для объединения инструментов. Код в Python показателен и может быть понятен людям без опыта работы на нем. Тем не менее вам необходимо владеть навыками создания фильтров или других инструментов на вашем языке программирования.

В данной книге я собрал методы из различных дисциплин, включив ссылки на оригинал там, где это было возможно. Таким образом, вы можете просмотреть эти материалы и найти другие подходы к решению проблемы. Многие из этих методов имеют математическое или статистическое обоснование, которое я намеренно оставил на функциональном уровне, не углубляясь в разновидности рассматриваемого подхода. Тем не менее базовое понимание статистики пригодится.

Содержание книги

Книга состоит из трех частей: «Данные», «Инструменты» и «Аналитика». Часть I «Данные» описывает процесс сбора и организации данных. В части II «Инструменты» рассказывается об инструментах поддержания аналитического процесса. В части III «Аналитика» предлагаются различные аналитические сценарии и методы.

Часть I посвящена сбору, хранению и организации данных. Хранение данных и логистика являются насущными проблемами анализа безопасности: собрать данные не сложно, гораздо сложнее осуществлять в них поиск конкретного явления. Данные занимают определенный объем, и можно собрать такое количество данных, в котором будет невозможно что-либо найти. Эта часть содержит следующие главы:

Глава 1

Описывает процесс сбора данных в целом. Она предлагает концепцию для понимания того, как сенсоры собирают информацию, формируют отчет и как они взаимодействуют друг с другом.

Глава 2

Продолжает тему предыдущей главы, уделяя особое внимание сенсорам, которые собирают данные о сетевом трафике. Эти сенсоры, включая `tcpdump` и NetFlow, представляют понятную модель активности сети, но зачастую их сложно толковать из-за трудностей, связанных с реконструкцией сетевого трафика.

Глава 3

В этой главе описываются сенсоры, расположенные в определенной системе, например в хостовой системе определения вторжений или журналах сервисов, таких как HTTP. Хотя вышеупомянутые сенсоры покрывают гораздо меньше трафика, чем сетевые, информация, поступающая с них, гораздо проще для понимания и требует меньше времени для толкования и построения догадок.

Глава 4

В главе 4 вы найдете различные инструменты для хранения данных трафика, в том числе традиционно используемые базы данных, системы больших данных, такие как Hadoop, а также специализированные инструменты, такие как графовые базы данных и сетевые журналируемые хранилища данных, например REDIS.

В *части II* собраны различные инструменты для анализа, визуализации и отчетности. Инструменты из этой части подробно разбираются в последующих разделах в контексте проведения различных видов анализа.

Глава 5

SiLK (System for Internet-Level Knowledge) – это набор инструментов для анализа потока данных, разработанный Университетом Карнеги-Меллон (Carnegie Mellon's CERT). В данной главе описываются возможности SiLK и то, каким образом использовать его инструменты для анализа данных, передаваемых протоколом NetFlow.

Глава 6

Данная глава посвящена языку программирования R – среде для проведения статистического анализа и визуализации, в которой можно качественно исследовать практически все возможные источники данных. Данная глава дает базовое представление об R и предлагает способы его использования для углубленного статистического анализа.

Глава 7

Система обнаружения вторжений, сокр. СОВ (Intrusion Detection System – IDS) – это автоматизированная система анализа трафика, подающая сигналы опасности при обнаружении подозрительных явлений. В данной главе особое внимание уделяется принципам работы СОВ, влиянию ошибок обнаружения на подаваемые СОВ сигналы опасности и построению эффективных систем обнаружения с применением инструментов для СОВ типа SiLK или конфигурации уже существующей СОВ типа Snort.

Глава 8

Одной из наиболее частых и трудоемких задач анализа является выявление происхождения IP-адреса или определение сигнатуры. В данной главе речь идет об инструментах и методах расследования, которые можно использовать для определения владельца адреса и его происхождения, имени, а также других элементов.

Глава 9

Глава вкратце рассказывает о некоторых специализированных инструментах анализа, не вошедших в предыдущие главы. Речь пойдет об инструментах для визуализации, создания пакетов и обработки данных, а также некоторых других наборах инструментов, которые необходимо знать специалисту по безопасности.

В *части III*, заключительном разделе книги, заключена цель всего процесса сбора данных – анализ. В следующих главах описаны различные явления трафика и математические модели для изучения данных.

Глава 10

Глава посвящена *разведочному анализу данных*, сокр. РАД (*Exploratory Data Analysis – EDA*), процессу изучения данных с целью определения их структуры или выявления необычных явлений. Поскольку данные о безопасности быстро меняются, каждому специалисту необходимо владеть РАД. Данная глава дает основы визуализации и описывает математические методы, используемые для исследования данных.

Глава 11

Глава посвящена ошибкам в ходе обмена данными и тому, как можно использовать их для обнаружения таких явлений, как сканирование.

Глава 12

В этой главе приводятся виды анализа, которые можно осуществить путем исследования объема и поведения трафика в динамике. Речь пойдет о DDoS-атаках, атаках на базы данных, а также об изменениях объемов трафика в течение рабочего дня и механизмах фильтрации объемов трафика для более эффективного анализа.

Глава 13

Данная глава посвящена преобразованию сетевого трафика в данные графов и использованию графов с целью определения значимых структур сетей. Такие атрибуты графов, как центрированность, могут быть использованы для определения значимых хостов или отклонений в работе.

Глава 14

В этой главе речь пойдет о методах определения вида трафика, проходящего через сервисные порты сети. Среди этих методов можно выделить обыкновенный поиск, например по номеру порта, а также баннер-граббинг и анализ ожидаемых размеров пакетов.

Глава 15

В главе описывается поэтапный процесс инвентаризации сети и определения важных хостов внутри нее. Составление карты сети и инвентаризация являются важными аспектами обеспечения информационной безопасности, которые необходимо применять на регулярной основе.

Принятые обозначения

В книге использованы следующие типографические обозначения

Курсивом

выделены новые термины, адреса URL, электронные адреса, названия и расширения файлов.

`Моноширинный шрифт`

используется в листингах, а также внутри параграфов для ссылки на программные элементы, такие как названия функций, баз данных, типов данных, переменные окружения, комментарии и ключевые слова.

`Моноширинным жирным шрифтом`

выделяются команды или любой другой текст, вводимый пользователем.

`Моноширинным курсивом`

выделяется текст, который должен быть заменен пользовательскими значениями или значениями, предписанными контекстом.

– этим символом обозначаются подсказки, предложения или общие примечания.

– этим символом обозначаются предостережения или предупреждения.

Использование примеров кода

Дополнительные материалы (примеры кода, упражнения и т. д.) доступны для скачивания по ссылке *https://github.com/mpcollins/nsda_examples*.

Эта книга написана для того, чтобы помочь вам сделать вашу работу. Если пример кода приведен в данной книге, вы можете использовать его в своих программах и документации. Вам не нужно запрашивать у нас разрешения на использование небольших частей кода. Например, написание программы с использованием нескольких фрагментов кода из этой книги не требует особого разрешения. Продажа и дистрибуция CD-дисков с примерами от издательства O'Reilly требует получения особого разрешения. Ответ на вопрос цитатой с примером кода из этой книги не требует особого разрешения. Внесение крупного фрагмента кода из этой книги в документацию по вашему продукту требует получения особого разрешения.

Мы приветствуем, но не требуем атрибуцию. Атрибуция, как правило, включает в себя название книги, имя автора, название издательства и международный стандартный книжный номер (ISBN). Пример атрибуции: «*Network Security Through Data Analysis by Michael Collins* (O'Reilly). Copyright 2014 Michael Collins, 978-1-449-3579-0».

Если использование вами фрагментов кода не подпадает под условия свободного использования или использования с разрешения издательства, свяжитесь с нами по электронной почте: *permissions@oreilly.com*.

Safari® Books Online (Сафари Букс Онлайн)

Safari Books Online – это цифровая библиотека по запросу, предоставляющая *материалы* экспертного уровня от ведущих мировых авторов книг в сфере технологий и бизнеса.

Профессионалы в области технологий, разработчики ПО, веб-дизайнеры, деловые и креативные люди используют Safari Books Online в качестве основного источника информации для исследований, решения задач, обучения и сертификации.

Safari Books Online предлагает *продукты* и ценовые программы для *организаций*, *государственных органов* и *частных лиц*. Подписчики имеют доступ к тысячам книг, обучающих видео и рукописей до публикации в виде удобной базы данных от таких издательств, как O'Reilly Media, Prentice Hall Professional, Addison-Wesley Professional, Microsoft Press, Sams, Que, Peachpit Press, Focal Press, Cisco Press, John Wiley & Sons, Syngress, Morgan Kaufmann, IBM Redbooks, Packt, Adobe Press, FT Press, Apress, Manning, New Riders, McGraw-Hill, Jones & Bartlett, Course Technology, и *десятков других*. Для получения более подробной информации о Safari Books Online посетите *наш сайт*.

Контактная информация

Просим отправлять комментарии и вопросы, касающиеся данной книги, в издательство по адресу:

O'Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472
800-998-9938 (in the United States or Canada)
707-829-0515 (international or local)
707-829-0104 (fax)

Эта книга имеет собственную веб-страницу, доступную по ссылке *http://oreil.ly/nstda,* где публикуется список опечаток, примеры и дополнительная информация.

Комментарии и вопросы технического характера просим отправлять по адресу *bookquestions@oreilly.com*.

Для получения более подробной информации о наших книгах, курсах, конференциях и новостях посетите наш веб-сайт *http://www.oreilly.com*.

Подпишитесь на нас в Facebook: *http://facebook.com/oreilly*, в Twitter: *http://twitter.com/oreillymedia.*

Подпишитесь на наш канал на YouTube: *http://www.youtube.com/oreillymedia*.

Благодарственное слово

Выражаю благодарность моему редактору Энди Ораму (Andy Oram) за его исключительную поддержку и обратную связь, без которых я бы сотый раз переписывал комментарий к точкам установки сенсоров сети. Также выражаю признательность ассистентам редактора Элисон МакДональд (Allyson MacDonald) и Марии Гулик (Maria Gulick) за то, что заставили поднажать и закончить книгу. Благодарю технических редакторов Риэннона Уивера (Rhiannon Weaver), Марка Томаса (Mark Thomas), Роба Томаса (Rob Thomas), Андре ДиМино (André DiMino) и Генри Стерна (Henry Stern). Их комментарии помогли мне избежать пустой болтовни и сконцентрироваться на действительно важных аспектах.

Эта книга – попытка донести самые полезные знания до отделов по эксплуатации и исследовательских центров, и я благодарю всех причастных по обе стороны, а именно (в произвольном порядке): Тома Лонгстафа (Tom Longstaff), Джея Кадейна (Jay Kadane), Майка Рейтера (Mike Reiter), Джона МакХью (John McHugh), Кэрри Гейтс (Carrie Gates), Тима Шимилла (Tim Shimeall), Маркуса ДеШона (Markus DeShon), Джима Дауни (Jim Downey), Уилла Франклина (Will Franklin), Сэнди Пэррис (Sandy Parris), Шона МакАллистера (Sean McAllister), Грега Верджина (Greg Virgin), Скотта Каула (Scott Coull), Джеффа Джэниса (Jeff Janies) и Майка Уитта (Mike Witt).

И наконец, я хочу поблагодарить моих родителей Джеймса и Кэтрин Коллинз (James and Catherine Collins). Отец скончался в процессе написания этой книги, но он задавал так много вопросов. И поскольку ответов он не понимал, то были вопросы о вопросах, вновь и вновь, до самого конца.

Предисловие от издательства

Отзывы и пожелания

Мы всегда рады отзывам наших читателей. Расскажите нам, что вы думаете об этой книге – что понравилось или, может быть, не понравилось. Отзывы важны для нас, чтобы выпускать книги, которые будут для вас максимально полезны.

Вы можете написать отзыв прямо на нашем сайте **www.dmkpress.com**, зайдя на страницу книги, и оставить комментарий в разделе «Отзывы и рецензии». Также можно послать письмо главному редактору по адресу **dmkpress@gmail.com**, при этом напишите название книги в теме письма.

Если есть тема, в которой вы квалифицированы, и вы заинтересованы в написании новой книги, заполните форму на нашем сайте по адресу **http://dmkpress.com/authors/publish_book/** или напишите в издательство по адресу **dmkpress@gmail.com**.

Список опечаток

Хотя мы приняли все возможные меры для того, чтобы удостовериться в качестве наших текстов, ошибки все равно случаются. Если вы найдете ошибку в одной из наших книг – возможно, ошибку в тексте или в коде, – мы будем очень благодарны, если вы сообщите нам о ней. Сделав это, вы избавите других читателей от расстройств и поможете нам улучшить последующие версии этой книги.

Если вы найдете какие-либо ошибки в коде, пожалуйста, сообщите о них главному редактору по адресу **dmkpress@gmail.com**, и мы исправим это в следующих тиражах.

Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательства «ДМК Пресс» и O'Reilly очень серьезно относятся к вопросам защиты авторских прав и лицензирования. Если вы столкнетесь в интернете с незаконно выполненной копией любой нашей книги, пожалуйста, сообщите нам адрес копии или веб-сайта, чтобы мы могли применить санкции.

Пожалуйста, свяжитесь с нами по адресу электронной почты **dmkpress@gmail.com** со ссылкой на подозрительные материалы.

Мы высоко ценим любую помощь по защите наших авторов, помогающую нам предоставлять вам качественные материалы.

Часть I

ДАННЫЕ

Эта часть посвящена сбору и хранению данных для последующего анализа и принятия мер. Эффективный анализ безопасности требует сбора данных из множества разных источников, каждый из которых лишь частично отражает положение дел в сети.

Чтобы понять важность гибридных источников данных, примите во внимание тот факт, что большинство современных ботов – это системы общего назначения. Один бот может использовать несколько методов для вторжения в другие хосты сети. Перечень этих атак может включать переполнение буфера, распространение через общие сетевые ресурсы или простое взламывание пароля. Попытка бота атаковать SSH-сервер путем ввода пароля может быть зафиксирована в SSH-журнале данного хоста, подтверждая факт атаки, но не предоставляя информацию о других действиях бота. Процесса сбора сетевого трафика может быть и недостаточно для реконструкции сессии, но он может рассказать вам о других действиях взломщика, допустим, о долгом, успешном сеансе взаимодействия с хостом, который ранее не был замечен в таком взаимодействии.

Самая сложная задача в проведении анализа на основе данных – это сбор достаточного количества данных для воссоздания редких событий. Достаточного, но не избыточного, в противном случае будет невозможно выполнить поисковый запрос. Сбор данных удивительно прост, но осмысление полученных данных гораздо сложнее. В сфере безопасности эта проблема осложняется редким возникновением *реальных* угроз. Большая часть сетевого трафика не несет никакой угрозы и часто повторяется: массовая рассылка писем или одновременный просмотр видео на YouTube большим числом пользователей, доступ к файлам. Многие из небольшого количества фактических атак будут *действительно* безобидными, например слепое сканирование пустых IP-адресов. Но эта небольшая часть таит в себе крошечное число атак, которые представляют собой реальную угрозу, например утечку файлов или обмен данными между ботнетами.

Все виды анализа данных, которые мы рассматриваем в этой книге, ограничены по вводу-выводу. Это означает, что процесс анализа данных предполагает точное определение нужных данных и последующую выборку. Поиск нужных данных требует времени, и эти данные имеют определенный объем: лишь один OC-3 может давать 5 терабайт сырых данных в день. Для сравнения, интерфейс eSATA может считывать около 0,3 гигабайта в секунду, таким образом расходуя несколько часов для *одного* поиска по всему массиву данных, учитывая, что в это

время вы считываете или записываете новые данные при работе с различными дисками. Необходимость сбора данных из множественных источников предполагает их избыточность, что требует дополнительного места на диске и увеличивает время запросов.

Правильно организованное хранилище и система запросов помогают специалистам по безопасности произвольно выполнять запросы данных и ожидать ответа в относительно короткий срок. При слабой организации системы на выполнение запроса требуется большее количество времени, нежели на сбор данных. Разработка правильной структуры требует понимания того, каким образом различные сенсоры осуществляют сбор данных, как они дополняют, дублируют и взаимодействуют друг с другом, а также понимания принципов эффективного хранения данных, дабы обеспечить возможность проведения анализа. Именно на этих проблемах и сделан акцент в данной главе.

Эта часть включает четыре главы. В *главе 1* содержится введение в общий процесс распознавания данных сенсором и их сбора, а также термины для описания взаимодействия сенсоров между собой. В *главе 2* приведены сенсоры, такие как `tcpdump` и NetFlow, которые осуществляют сбор данных из сетевых интерфейсов. *Глава 3* посвящена хост-сенсорам и сервисным сенсорам, осуществляющим сбор данных о различных процессах, происходящих, например, в серверах и операционных системах. *Глава 4* рассказывает о различных опциях применения систем сбора данных, начиная с баз данных и заканчивая современной технологией больших данных.

Глава 1

Сенсоры и детекторы: введение

Эффективный мониторинг информации строится на данных, собранных из многочисленных сенсоров, которые генерируют различные виды данных и создаются различными людьми для различных целей. Сенсором может быть все, что угодно, от сетевого отвода до журнала файрвола – тем, что осуществляет сбор информации о вашей сети и может быть использовано для оценки информационной безопасности. Построение эффективной системы сенсоров требует достижения баланса между ее укомплектованностью и избыточностью. Идеальная система сенсоров укомплектована, но не избыточна. Под укомплектованностью понимается то, что каждое событие тщательно описано, а под отсутствием избыточности – то, что сенсоры не дублируют информацию о событиях. Эти, возможно, недостижимые цели являются идеальной моделью для построения решения по мониторингу.

Ни один из сенсоров не может выполнять все функции в одиночку. Сетевые сенсоры действительно выполняют много работы, но их легко сбить с толку в процессе управления потоками трафика, они неэффективны в отношении зашифрованного трафика и могут лишь предположить наличие активности в хосте. Хост-сенсоры предоставляют более исчерпывающую и точную информацию относительно явлений, для описания которых они имеют достаточный инструментарий. С целью эффективного комбинирования сенсоров я классифицирую их в трех плоскостях:

Область обзора (Vantage).
: Расположение сенсоров внутри сети. Сенсоры, расположенные в разных точках, будут видеть разные стороны одного события.

Уровень (Domain).
: Информация, предоставляемая сенсором, вне зависимости от его местонахождения (хост, сервис хоста или сеть). Сенсоры с одинаковой областью обзора, но разного уровня дополняют друг друга в процессе предоставления данных об одном и том же событии. Информацию о некоторых событиях можно получить лишь на одном из уровней. Например, мониторинг хоста – это единственный способ определить, имел ли место физический доступ к этому хосту.

Действие сенсора (Action).
: Как сенсор принимает решение о создании информационного отчета. Он может просто записывать данные, предоставлять информацию о событиях или же обрабатывать трафик, который предоставляет данные. Сенсоры различного действия могут, потенциально, мешать работе друг друга.

Область обзора сенсора: зависимость сбора данных от расположения сенсора

Область обзора сенсора дает представление о том, какие пакеты сенсор сможет изучать. Область обзора определяется взаимозависимостью между расположением сенсора и инфраструктурой маршрутизации сети. Чтобы понять, как процессы влияют на область обзора, взгляните на *рис. 1-1*. На данном рисунке показаны уникальные потенциальные сенсоры, обозначенные заглавными буквами. В порядке очередности эти сенсоры имеют следующее расположение:

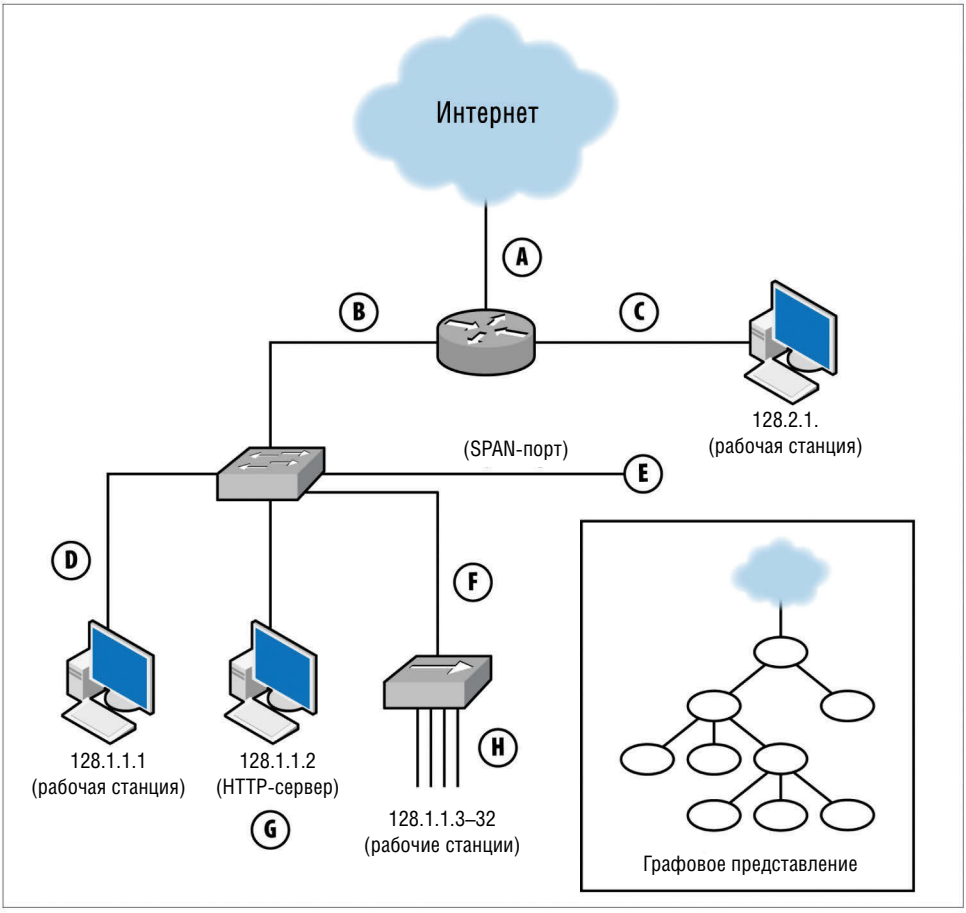

Рис. 1-1. Позиционирование сенсоров в простой сети и графовое представление

A

Проверяет интерфейс, соединяющий роутер с интернетом.

B

Проверяет интерфейс, соединяющий роутер с сетевым коммутатором.

C

Проверяет интерфейс, соединяющий роутер и хост с IP-адресом 128.2.1.1.