М. В. Тарасюк, Ф. Н. Царев

МЕТОД ОЦЕНИВАНИЯ ТРАФИКА VPN-СЕТИ НА ОСНОВЕ МОДЕЛЕЙ СКРЫТОЙ МАРКОВСКОЙ ЦЕПИ В ЗАДАЧАХ ТЕХНИЧЕСКОЙ РАЗВЕДКИ

1. Цели и задачи анализа параметров трафика

На рынке коммерческих и открытых (open source) информационных технологий широко используются системы обнаружения сетевых угроз (далее — системы). Впервые данного рода системы были предложены для использования в правительственных сетях США. Базируются они на анализе сетевой активности и выявление аномалий сетевых транзакций, характерных для тех или иных угроз. Модель функционирования системы, использующая анализ регистрационных журналов узлов сети, описана в [1] и положена в основу так называемых хостовых систем. Впоследствии модель была дополнена сетевыми системами, анализирующими данные из сети, реконструирующими сетевые транзакции и выявляющими по ним аномалии, специфичные для угроз. Объектом внимания таких систем, как правило, являются пакеты с некорректным заголовком, содержащие определенные (известные) последовательности байтов, и др. Подробное изложение вопросов построения сетевых и комбинированных систем (объединяющих свойства хостовых и сетевых) можно найти в [2]. Один из способов защиты от сетевых угроз на периметре объекта информатизации — использование сетевой криптографической защиты (VPN-технологий), в основе которых лежит туннельный режим шифрования и которые защищают объект от активных или пассивных угроз, направленных из открытой сети. В отличие от линейных шифраторов технология сетевого шифрования (реализуемая криптомаршрутизаторами) не позволяет устранить угрозы анализа и оценки параметров сетевого трафика. Например, при использовании VPN в составе телекоммуникационной сети военного назначения технической разведкой может быть установлен состав программного обеспечения на узле связи и раскрыты его функции и место в системе военного управления. Меры противодействия должны учитывать результаты исследования методов анализа сетевого трафика в части обеспечения требуемой достоверности идентификации объекта разведки и модели возможностей разведки. Далее показаны возможности применения кластеризации и моделей Марковских цепей со скрытыми параметрами для анализа сетевого трафика с целью обнаружения и (или) идентификации объектов разведки.

2. Модель оценки параметров трафика

Информативными параметрами трафика VPN-сети являются длины пакетов, длительности межпакетных интервалов, а также открытые параметры пакетных заголовков, прозрачно передаваемых в канал связи (например, код классификации приоритета пакета). В данной статье предлагается метод оценивания трафика, использующий Марковские цепи со скрытыми параметрами. Для оценки демаскирующих возможностей трафика сетевых программ предлагаются:

1. Модель описания трафика сетевых программ, осуществляющих обмен данными через пограничный криптомаршрутизатор (VPN-шлюз);

2. Методика определения соответствия (распознавания) наблюдаемых реализаций трафика той или иной программе из заданного множества.

В течение определенного времени регистрируется последовательность наблюдаемых параметров трафика, преобразуемая к виду, удобному для дальнейшего анализа. Наблюдаемые последовательности сравниваются с эталонными образцами, известными на этапе «настройки» модели. К числу критериев эффективности анализа трафика могут быть отнесены:

– трудоемкость процесса «обучения» модели распознавания; различным моделям требуются разные объемы «тренировочных» данных для обучения;

– точность и корректность распознавания; этот критерий связан с ошибками первого (необнаружение) и второго (ложное обнаружение) рода.

Предлагается модель описания трафика, основной идеей которой является использование скрытых Марковских цепей. Как и в любом методе распознавания образов, предлагается двухэтапное решение: обучение модели и последующая классификация трафика с использованием обученной модели.

Для снижения трудоемкости обучения и уменьшения влияния на модель трафика случайных данных ряд наблюдаемых параметров целесообразно подвергнуть кластеризации. Например, длины пакетов и длительности межпакетных интервалов могут принимать практически произвольные числовые значения, в связи с чем непосредственное их использование в качестве наблюдаемых параметров скрытой Марковской цепи затруднительно. Другие параметры, например коды приоритетности трафика или адреса внешних шлюзов, принимают значения из четко определенного диапазона.

Построение модели может выполняться одновременно над множеством параметров, относящихся к одному сетевому пакету, либо для каждого параметра может использоваться независимая модель. Первый случай позволяет построить модель, учитывающую связи между параметрами.

Пусть обучение модели производится на основе анализа исходной (эталонной) после-довательности пакетов. Для некоторых параметров, например для длин пакетов, в качестве дополнительного ограничения естественно принять максимальный размер пакета на выходе криптомаршрутизатора. Для других параметров, например для межпакетных интервалов, выбор ограничения в большей мере определяется моделью поведения абонентов сети. На первом (подготовительном) этапе производится кластеризация некоторых параметров трафика. Множество наблюдаемых (эталонных) реализаций параметров разбивается на группы (кластеры), внутри которых реализации достаточно «близки» друг к другу относительно какой-либо выбранной метрики. При наличии обоснованных предположений кластеризации подвергаются векторы, составленные из реализаций отдельных значений параметров. Для кластеризации используется любой из доступных алгоритмов, например метод, комбинирующий кластеризацию на основе минимального остовного дерева и метода k-средних. После отработки этого алгоритма каждый кластер характеризуется одной точкой – своим центром. При кластеризации методом k-средних точка относится к тому кластеру, до центра которого расстояние минимально. Если же это расстояние превышает удвоенное максимальное расстояние между центрами соседних (входные данные одномерны) кластеров, то считается, что точка не относится ни к одному из кластеров. Далее с помощью алгоритма Баума–Велча по эталонной последовательности наблюдаемых параметров «обучается» скрытая Марковская цепь для заданного числа состояний. В результате определяется набор параметров Марковской цепи, позволяющих аппроксимировать реальный трафик. Итак, модель трафика состоит из Марковской цепи, полученной с помощью алгоритма Баума–Велча, и набора кластеров (групп кластеров, если каждый параметр оценивается независимым образом). Кластеры строятся по-разному в зависимости от используемого алгоритма кластеризации.

На последнем (рабочем) этапе выполняются распознавание и классификация трафика (в рабочем режиме наблюдения за разведываемыми объектами информатизации). В распоряжении разведки имеется несколько обученных моделей, соответствующих контролируемым программам. На вход детектора средства разведки поступает последовательность наблюдаемых (рабочих) параметров. Для каждой модели оценивается вероятность формирования ею трафика, и выбирается модель с наибольшей вероятностью.

Для разных наблюдаемых параметров защита от угрозы раскрытия может строиться различными способами. Например, для сетевых адресов или номеров виртуальных каналов могут использоваться механизмы трансляции адресов источника (NAT/PAT). Для длин пакетов защита может обеспечиваться выравниванием пакетов на заданную границу длины или вставкой в пакеты ложных байтов данных. Для измеряемых параметров, таких как длительность межпакетных

интервалов, необходимо передавать в каналы связи маскировочные пакеты, аналогично тому как это делается для защиты от скрытых каналов [3]. Различие состоит в том, что цели защиты от скрытых каналов состоят в снижении их производительности, а в случае раскрытия назначения и функций объекта цель — работа в пределах допустимых норм вероятности раскрытия за определенный промежуток времени.

Для проверки эффективности противодействия описанному методу разведки трафик, подаваемый на вход детектора (средства разведки), предварительно смешивается с маскировочным трафиком, который представляет собой, например, передачу пакетов случайной длины в случайные моменты времени.

3. Пример оценивания трафика

Имеется два поставщика трафика:

– поставщик трафика, генерирующий пакеты размером от 1 до 2 килобайт (длины пакетов равномерно распределены) с межпакетными интервалами от 10 до 20 микросекунд (Поставщик I);

– поставщик трафика, генерирующий пакеты двух типов: размером от 1 до 2,5 килобайта и размером от 0,5 до 2 килобайт (первый тип выбирается с вероятностью 0,9, второй — с вероятностью 0,1). Межпакетные интервалы этого поставщика трафика имеют размер от 10 до 20 микросекунд (Поставщик II).

Кроме того, имеется поставщик ложного трафика: минимальная и максимальная длина маскировочного пакета одинакова, минимальная длительность межпакетного интервала фиксирована и равна 50 микросекундам, максимальная длительность межпакетного интервала принимает значения из множества {50, 100, 1000, 2000 и 5000} микросекунд.

Обучение моделей проведем на последовательностях длиной в 100 пакетов с двумя состояниями скрытой Марковской цепи. После этого на вход анализатору трафику будем подавать по 1000 последовательностей пакетов с длинами от 1 до 300. Каждая последовательность формируется одним из двух поставщиков. Далее вычисляется доля правильно классифицированных последовательностей с учетом передачи маскировочного трафика. На рис. 1 приведены оценки правильно классифицированных последовательностей в зависимости от их длины, по оси абсцисс отложена длина последовательности, а по оси ординат — доля правильно классифицированных последовательностей.

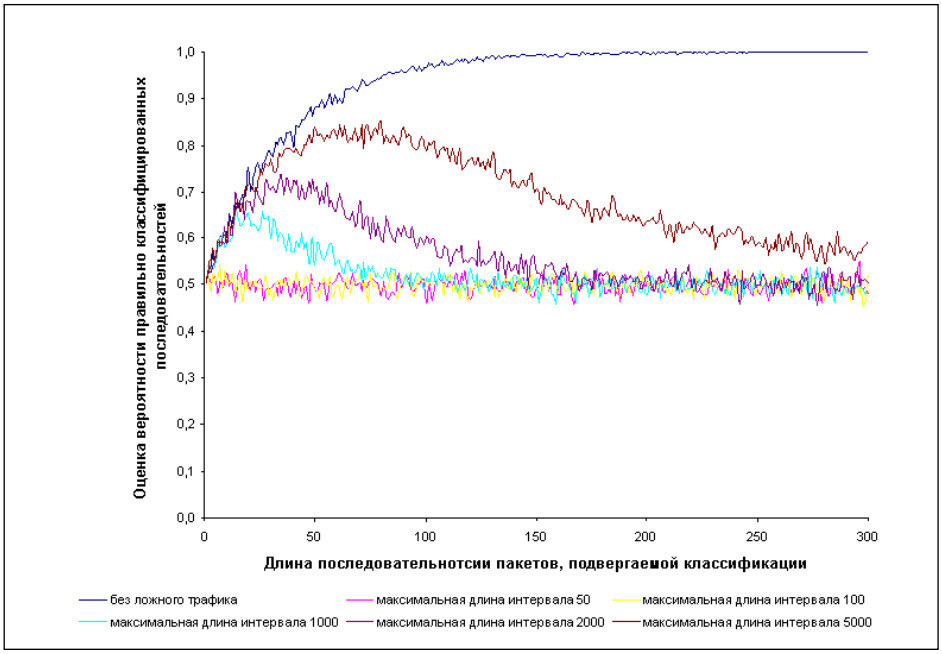

Рис. 1. Оценка эффективности классификации при наличии ложного трафика

Графики показывают, что для рассмотренного случая маскировочный трафик сравнительно небольшой интенсивности катастрофически искажает демаскирующие признаки. Раскрытие оказывается возможным, если характерные последовательности пакетов наблюдаются на интервале времени, много меньшем, чем средний интервал передачи маскировочных пакетов. Однако при этом более сложные модели трафика (например, длинные последовательности пакетов с высокой внутренней корреляцией между наблюдаемыми параметрами) будут более устойчивы и потребуют маскировочного трафика большей интенсивности.

Если в качестве демаскирующего признака выступают межпакетные интервалы, для защиты можно также применять адаптированный к использованию в сети вариант РОМР-технологии [4] с буферным накопителем, устраняющий вариации межпакетных интервалов. При этом ухудшаются вероятностно временные характеристики, что не всегда приемлемо для трафика реального времени.

Заключение

Для гарантированной защиты от угрозы требуется исключить выдачу (в открытом виде) любой служебной информации в канал связи и использовать дополняющую маскировку. По части эффективности использования ресурсов это эквивалентно сетевым решениям на линейных шифраторах. С другой стороны, угроза анализа трафика актуальна только для программ с выраженными отличиями профиля трафика от программ общего назначения (электронная почта, видеоконференцсвязь и пр.).

СПИСОК ЛИТЕРАТУРЫ:

1. *Anderson J. P.* Computer Security Threat Monitoring and Surveillance. James P. Anderson Co. Fort Washington Pa., 1980.
2. *Ptacek T. H.* Insertion, Evasion, and Denial of Service: Eluding Network Intrusion Detection // Secure Networks, Inc. January, 1998. URL: http://www.cs.unc.edu/~jeffay/courses/nidsS05/evasion/ids-evasion98.pdf.
3. *Тарасюк М. В., Тарасов И. В.* Адаптивная маскировка скрытых каналов в открытых системах с многоуровневым доступом // Информационные технологии. 2004. № 2. С. 25–29.
4. *Kang M. H., Moskowitz I. S.* A Pump for Rapid Reliable, Secure Communications // Proceeding of the 1st ACM Conference on Computer and Telecommunication Security. Fairfax VA. ACM Press. Nov. 3–5. 1993. P. 119–129.