ПРИМЕНЕНИЕ ВЕЙВЛЕТ-АНАЛИЗА РЕЧЕВОГО СИГНАЛА И ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧЕ ИДЕНТИФИКАЦИИ ЛИЧНОСТИ ПО ГОЛОСУ

И.Е. Ермилов, Ю.А. Семин

Нижегородский госуниверситет

Задача распознавания диктора является одной из наиболее актуальных задач в области информационных технологий. Решение этой задачи может применяться в различных сферах деятельности.

Эту задачу можно разбить на несколько последовательных этапов. Прежде всего, необходимо создать базу дикторов, состоящую из нескольких образцов записей голоса для каждого диктора, часть из которых понадобится для обучения системы, принимающей решение о принадлежности речевого сигнала (классификатора). В качестве такой системы применяется искусственная нейронная сеть. Далее происходит выделение характеристик из полученных сигналов для сокращения полезной информации о сигнале. Пространство признаков, в котором принимается решение о личности диктора, должно формироваться с учетом всех факторов процесса речеобразования. Выделение признаков из речевого сигнала происходит с целью уменьшения необходимой для анализа информации. Удобнее работать с конечным набором параметров, которые характеризуют те или иные особенности диктора, чем пытаться работать с самим речевым сигналом, состоящим из огромного количества отсчетов [1].

Для получения набора признаков применяем дискретный вейвлет-анализ полученных речевых сигналов. Основная область применения вейвлет-преобразований – анализ и обработка сигналов и функций, нестационарных во времени или неоднородных в пространстве. Вейвлет-преобразование одномерных сигналов обеспечивает двумерную развертку, при этом частота и координата рассматриваются как независимые переменные, что позволяет анализировать сигнал сразу в двух пространствах. Вейвлет-фильтрация и многоуровневое разложение сигналов реализованы с использованием вейвлетов Добеши 5-ого порядка. В качестве характеристики, определяющей частотную составляющую сигнала, извлекаемой из речевого сигнала, использовались значения энергии коэффициентов детализации на уровнях разложения, в процентах от исходной энергии исследуемых данных.

Далее получаем временную информацию об исследуемом сигнале. Для получения временной характеристики сигнала каждый из пяти исследуемых уровней вейвлет-декомпозиции речевого сигнала разбивался на равные по времени участки. На каждом участке вычислялась оценка среднеквадратичного отклонения детализирующих коэффициентов. Набор найденных характеристик использовался как входной вектор классификатора. Данный подход, объединяющий в себе вейвлет-преобразование и нейронную сеть, является довольно многообещающим.

Способность нейронных сетей накапливать знания об объектах и процессах с использованием алгоритмов обучения делает их применение в задаче распознавании речи очень перспективным. Используемая в данной работе нейронная сеть является многослойным персептроном. При работе с ней важны такие аспекты, как выбор наиболее подходящей архитектуры сети и алгоритма обучения для получения наилучших результатов при распознавании диктора по фрагменту речевого сигнала.

При выборе архитектуры следует исходить из того, что способности сети к обобщению тем выше, чем больше суммарное число связей между нейронами. С другой стороны, число связей ограничено сверху количеством записей в обучающих данных. Наиболее выгодной для поставленной задачи оказалась архитектура с несколькими скрытыми слоями, первый из которых имеет на порядок большее количество нейронов чем во входном слое сети. Данный результат подтверждается теоремой Ковера о разделимости образов.

Еще одним важным фактором при определении качества работы сети является выбор алгоритма обучения. Наиболее простым и часто применяемым является алгоритм обратного распространения. Этот алгоритм хорошо работает для сетей малой размерности или простых задач, но он медленно сходится при решении сложных задач с большим числом узлов в сети и не всегда способен обеспечить необходимое качество обучения. В таких задачах, как распознавание речи, десятки тысяч итераций могут быть необходимы даже при небольшом наборе данных.

Для решения этих проблем разработаны более совершенные методы, из них наиболее перспективным в данный момент считается алгоритм обучения, основанный на фильтрации Калмана. В работе были проведены исследования по сравнению этих двух алгоритмов обучения. В качестве критериев сравнения использовались изменение средней квадратичной ошибки в процессе обучения, время обучения и время затраченное на одну эпоху. Сравнение алгоритмов показало, что метод обучения многослойного персептрона, основанный на фильтрации Калмана, обладает лучшей сходимость по сравнению с методом обратного распространения ошибки. При использовании фильтра Калмана значение средней квадратичной ошибки уменьшается гораздо быстрее и для обучения требуется более чем в 2 раза меньшее количество эпох. Но при этом фактическое время, потраченное на обучение, в 1,5 раза больше, чем при обратном распространении ошибки. Точность распознавания при использовании алгоритма обратного распространения составила порядка 90%. Применение фильтрации Калмана позволило улучшить точность распознавания до значения 95–97%, но при этом значительно возросли временные затраты на обучение многослойного персептрона.

[1] Ермилов И.Е., Семин Ю.А. // В кн.: Труды XVII-й научной конференции по радиофизике. 13–17 мая 2013 г. /Ред. С.М. Грач, А.В.Якимов. – Н.Новгород: Изд-во ННГУ, 2013. С.124.

АЛГОРИТМ РАСПОЗНАВАНИЯ ФОРМУЛ НА ОТСКАНИРОВАННЫХ ИЗОБРАЖЕНИЯХ

Ю.Е. Чуманкин, П.Е. Овчинников

Нижегородский госуниверситет

Настоящая работа посвящена решению задачи распознавания формул на отсканированных изображениях с применением нейронных сетей. В настоящее время эта задача решается некоторыми коммерческими приложениями. Необходимо отметить, что эти приложения хорошо справляются лишь с распознаванием несложных по структуре формул при разрешении 400–600 dpi, поэтому проведение исследований в этой области является актуальным.

Целью данной работы является построение общего алгоритма распознавания формул, а также разработка и тестирование приложения, основанного на предложенном алгоритме.

Для распознавания формул предлагается следующий алгоритм. Первым действием изображение бинаризуется. Вторым действием производится сегментация изображения, т.е. выделение областей отдельных символов. Следующим этапом является выделение признаков полученных областей. Далее проводится распознавание символов, формирование формулы и запись ее текстовой интерпретации в файл.

Бинаризация проводится по порогу, подобранному экспериментально.

На этапе сегментации предлагается применять алгоритм наращивания и соединения смежных областей.

Для принятия решения о значении некоторого символа, необходимо выделить набор признаков, по которому это решение будет приниматься. Реальные распознаваемые символы всегда подвергнуты некоторым преобразованиям, в частности, масштабному преобразованию и сдвигу. Необходимо построить систему признаков, инвариантную относительно этих преобразований.

В данной работе в качестве набора признаков выбирается изображение символа размером 21×21 пиксель, его исходные геометрические размеры и положение центра.

Добиться инвариантности изображения относительно сдвига можно путем перехода в систему координат, связанную с верхним левым углом символа.

Для инвариантности относительно масштаба необходимо изменить масштаб вдоль осей. Вследствие теоремы о дискретизации Котельникова, прежде чем уменьшать масштаб изображения, на него необходимо подействовать фильтром низких частот. В качестве такого фильтра выбирается гауссов фильтр.

Для принятия решения о значении определенного символа использовалась нейронная сеть Хэмминга. В данной топологии сети есть два типа

слоев. Первый производит расчет меры дальности до эталонов, в качестве такой меры используется расстояние Хэмминга. Второй тип слоев выбирает максимально близкий эталон. На выходе сети выдается номер максимально близкого эталона. Плюсом при использовании данной сети является отсутствие потребности в итеративном обучении, т.к. веса задаются аналитическим выражением.

Следующим этапом является построение структуры формулы. Для этого предлагается использовать данные о положении символов, их значения и исходные геометрические размеры.

Первым этапом построения выделяются символы сумм и интегралов, далее выделяются их пределы. Затем проводится выделение дробей. В числитель относятся все символы, находящиеся выше дробной черты, а в знаменатель все символы, находящиеся ниже дробной черты. Для дальнейшего отнесения символов к группам «индекс», «степень», «обычная строка» анализируются положения центров символов.

По предложенному алгоритму было разработано действующее приложение. В обучающую выборку включены заглавные и строчные, обычные и курсивные буквы латинского и греческого алфавита, цифры, знаки «–», «+», «=», «>», «<». Было распечатано и отсканировано 36 формул при разрешении 75 dpi с использованием символов из выборки и различных структурных элементов формул. Приложение успешно распознало 97% символов, и безошибочно установило структуру 86% формул. Была составлена выборка из 20 формул из различных отсканированных книг. В ней распознано 50% символов, и безошибочно была установлена структура 54% формул.

Из полученных результатов можно сделать вывод, что приложение достаточно чувствительно к шрифту, используемому в распознаваемых формулах. Также можно сделать вывод, что приложение успешно справляется с построением структуры формул при низком разрешении распознаваемых изображений ~75 dpi.

[1] Шапиро Л. и Стокман Дж. Компьютерное зрение. – М.: БИНОМ. Лаборатория знаний, 2006, 752 с.
[2] Визильтер Ю.В., Желтов С.Ю., Князь В.А., Ходарев А.Н., Моржин А.В. Обработка и анализ цифровых изображений с примерами на LabVIEW IMAQ Vision. – М.: ДМК, 2008, 463 с.
[3] Форсайд Д., Понс Ж. Компьютерное зрение. Современный подход. – М.: Вильямс, 2004, 926 с.
[4] Заенцев И.В., Нейронные сети: основные модели. – Воронеж: Воронежский государственный университет, 1999, 74 с.
[5] Ф. Уоссермен, Нейрокомпьютерная техника: Теория и практика. – М.: Мир, 1992, 184 с.

СИСТЕМА ОБНАРУЖЕНИЯ И ПРЕДОТВРАЩЕНИЯ БЕСПРОВОДНЫХ ВТОРЖЕНИЙ

И.С. Исупов, А.А. Рябов

Нижегородский госуниверситет

В настоящее время системы обнаружения и особенно предотвращения вторжений в беспроводных сетях (Wireless Intrusion Detection System, WIDS) имеют не столь широкое распространение, нежели их аналоги в проводных сетях. В то же время современные тенденции развития информационных технологий позволяют составить прогнозы для необходимости их внедрения и активного использования в местах с жёстким контролем ресурсов и периодическим аудитом безопасности. За последний год удаётся наблюдать положительную динамику развития данной области. В сети Интернет стали появляться всё больше статей, направленных на обеспечение безопасности именно в беспроводных сетях, появляются новые коммерческие продукты обнаружения беспроводных вторжений. Рассмотрим подробнее методы работы WIDS, сложности, плюсы, минусы их разработки и способы решения задач безопасности.

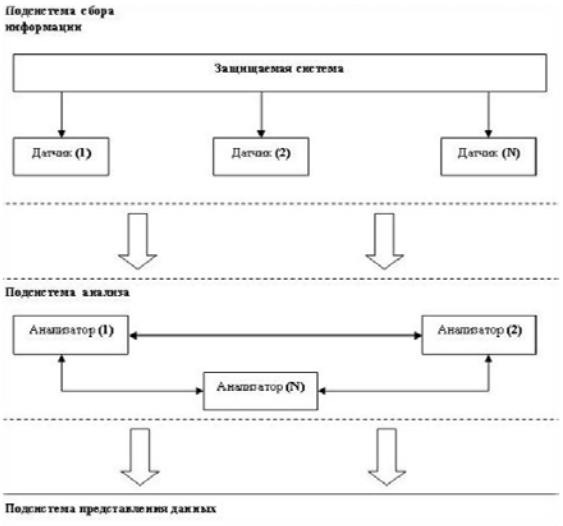

Рис.

Структура WIDS аналогична структуре IDS в проводных сетях и представляет собой систему из трех основополагающих частей. Это – считывание информации, её анализ и представление данных. Разрабатываемое приложение охватывает две последние части. Смысл такого подхода заключается в том, чтобы не реализовывать самостоятельно систему взаимодействия с аппаратной частью. Это позволит

развивать приложение распределённо, разрабатывая только способы анализа и отображения полученных данных.

Основу WIDS составляют сенсоры, выполняющие функцию сбора беспроводного трафика в режиме мониторинга. Сенсоры могут быть реализованы на базе операционной системы (ОС) семейства Windows или специализированных программно-аппаратных комплексов, в большинстве случаев базирующихся на ОС Linux. Сенсоры представляют собой достаточно интеллектуальные устройства, поддерживающие семейство протоколов TCP/IP и обладающие развитыми интерфейсами управления.

В разрабатываемом приложении в качестве сенсоров предлагается использовать отдельно стоящие антенны, подключенные к одному компьютеру под управлением ОС Linux. Сервер обнаружения угроз обрабатывает все поступающие на него пакеты, собранные в зоне действия системы принимающих антенн. Таким образом система обнаружения беспроводных вторжений содержит в себе всего один компьютер, занимающийся обработкой всех поступающих на него данных, и систему антенн.

В ходе обработки поступающей на сервер обнаружения угроз информации, им выполняются следующие задачи:

- принятие решений интеллектуальной системой относительно клиентов беспроводной сети;
- исследование работы пользователей в беспроводной сети (статистика использования ресурсов);
- посылка оповещения администратору безопасности по электронной почте при обнаружении аномалий.

Был внесён ряд изменений в функциональные возможности приложения, а именно: реализация алгоритмов обнаружения вторжений и аномалий внутри сети и возможность использования нескольких беспроводных виртуальных адаптеров для физического разделения использующихся беспроводных устройств.

На данный момент приложение фактически делится на две глобальные части. Первая часть представляет собой графический интерфейс для базового приложения Aircrack и позволяет просматривать всю информацию о беспроводных устройствах и точках доступа. Вторая часть включает в себя систему обнаружения вторжений. В этом разделе можно настроить действия при определении аномалий, например, черные и белые списки клиентов.

В процессе разработки приложения определено несколько путей развития. Важной особенностью будет возможность приложения в автоматическом режиме собирать статистику использования ресурсов беспроводной сети пользователями. Для просмотра собранной статистики планируется подготовка наглядного вывода в виде графиков и диаграмм. Планируется разработать третий раздел приложения и включить в него программный комплекс по тестированию определённой сети на защищённость. Его целью будет являться выявление недостатков защищённости и «дыр» в безопасности беспроводной сети. Для расширения целевой аудитории и удобства внедрения системы будет создана кроссплатформенная версия.

ЭКСПЕРИМЕНТАЛЬНЫЙ ВЫБОР АЛГЕБРАИЧЕСКОЙ СТРУКТУРЫ ДЛЯ ОБРАТИМЫХ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ИСТОЧНИКОВ ДАННЫХ

А.А. Горбунов, А.С. Коновалов

Нижегородский госуниверситет

Использование математических моделей (ММ) источников и преобразователей текстовых данных позволяет описывать функционирование дискретных динамических систем различной природы. При помощи таких моделей описываются, например, процессы кодирования генетических текстов [1], основные блоки в криптосистемах (шифратор, дешифратор, канал связи) [2, 3], прогнозирующие элементы в системах обработки цифровых рядов данных и другие объекты, связанные с обработкой текстовой информации.

Моделирование работы шифрующих и кодирующих устройств зачастую подразумевает построение таких преобразователей текстовой информации, которые допускают однозначное обратное преобразование. В работе [4] шифрующий и дешифрующий преобразователи представлены в форме нелинейного синхронного автомата Хаффмана – Глушкова следующими уравнениями:

$$\begin{cases} \boldsymbol{x}(t+1) = \gamma_1\big(\boldsymbol{x}(t);\boldsymbol{p}\big) \\ \boldsymbol{y}(t) = \lambda_1\big(\boldsymbol{x}(t);\boldsymbol{p}\big)\langle\circ\rangle\,\lambda_2\big(\boldsymbol{u}(t);\boldsymbol{p}\big) \\ \boldsymbol{x}(0) = \boldsymbol{x}_н \end{cases} ; \quad \begin{cases} \boldsymbol{x}(t+1) = \gamma_1\big(\boldsymbol{x}(t);\boldsymbol{p}\big) \\ \boldsymbol{u}(t) = \lambda_2{}^{-1}\Big(\boldsymbol{y}(t)\,\langle\circ\rangle^{-1}\,\lambda_1\big(\boldsymbol{x}(t);\boldsymbol{p}\big);\boldsymbol{p}\Big) \\ \boldsymbol{x}(0) = \boldsymbol{x}_н \end{cases} . \quad (1)$$

Здесь $\boldsymbol{u}(t)$, $\boldsymbol{y}(t)$, $\boldsymbol{x}(t)$ – отсчеты последовательностей соответственно открытого текста, шифротекста и внутреннего состояния автомата; $\boldsymbol{x}_н$ – начальное состояние автомата; $\boldsymbol{p}$ – вектор свободных параметров. При этом требуется обратимость функция λ_2 и операции «сложения» $\langle\circ\rangle$ той алгебраической структуры (АС), в которой осуществляется функционирование представленных моделей.

В рамках настоящей работы программно реализован алгоритм, моделирующий работу криптографических преобразователей (1) текстовых последовательностей в АС, операции в которых задаются соответствующими таблицами Кэли. На ряде классов нелинейных автоматов проведены компьютерные эксперименты по построению восстанавливающего автомата для текстовых последовательностей с размерностью алфавита $q = 3 \div 7$.

При размерности $q = 3$ проведен полный перебор 729 вариантов таблиц Кэли, из которых для 38 матриц удалось корректно осуществить восстановление открытых текстов по шифротекстам. С увеличением размерности q ввиду большого количества матриц эксперименты проводились с использованием 10^6 случайных неповторяющихся таблиц. На представленном графике (см. рис.) показана полученная экспериментально зависимость количества N вариантов таблиц Кэли, при помощи которых успешно моделируется работа дешифрующего преобразователя, от размерности алфавита q текстовых последовательностей криптосистем.

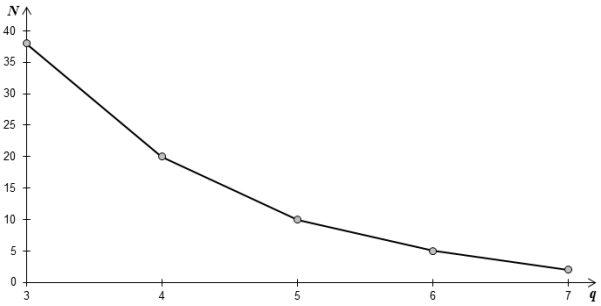

Рис.

Для различных классов автоматов (описываемых ММ со степенным видом нелинейности) также программно реализована возможность отбора варианта алгебраической структуры, для которой моделью дешифрующего автомата восстанавливается максимальное количество отсчетов последовательности открытого текста. В рамках данной работы были определены такие оптимальные варианты таблиц Кэли для каждого из значений $q = 3 \div 7$.

[1] Кирьянов К.Г. Генетический код и тексты: динамические и информационные модели сложных систем. /Под ред. Л.Ю. Роткова, А.В. Якимова. – Н. Новгород: ТАЛАМ, 2002, 100 с.
[2] Гилл А. Линейные последовательностные машины. – М.: Наука, 1974, 288 с.
[3] Горбунов А.А., Кирьянов К.Г. // Вестник Нижегородского университета им. Н.И. Лобачевского. Сер. «Радиофизика». Вып. 1(2). – Н. Новгород: ННГУ, 2004. С. 24.
[4] Горбунов А.А. // В кн.: Труды XII научной конференции по радиофизике. /Под ред. А.В. Якимова, С.М. Грача. – Н. Новгород: Изд-во ТАЛАМ, 2008. С. 263.

АНАЛИЗ МЕТОДОВ СОКРАЩЕНИЯ ПРОСТРАНСТВА ПРИЗНАКОВ ДЛЯ ЗАДАЧИ ДЕТЕКТИРОВАНИЯ СПАМА

О.И. Шкалябин

Нижегородский госуниверситет

В задаче фильтрации спама методами Data Mining важным вопросом является сокращение размерности пространства признаков.

Письмо представляется как вектор весовых коэффициентов признаков объекта. Признаки письма разделяются на нетекстовые (информация из заголовков и вложений, статистические характеристики письма) и текстовые. Текстовые признаки определяются из всего множества термов, содержащихся в письмах. Получаем частный случай задачи классификации документов по двум категориям (spam и ham).

Среди способов выделения термов наиболее широко применяется модель bag-of-words. Эта модель хорошо изучена и обладает известными недостатками:

- теряется одна из ключевых составляющих семантики текста – порядок слов;
- значение слова зачастую определяется контекстом;
- введение новых ключевых слов ведет к перестройке всего пространства;
- морфологические характеристики языка затрудняют поиск ключевого слова в теле документа;
- человеческий фактор (а именно неправильное написание слов) значительно затрудняет поиск определенного слова в теле документа и адекватный морфологический анализ.

Возможно применение более сложных моделей, использующих многословные характеристики и учитывающих зависимости значения слов от их взаимного расположения и контекста, но при этом значительно увеличивается вычислительная сложность. При этом, как показывают исследования, модель bag-of-words при своей простоте демонстрирует практически аналогичную эффективность. После формирования набора признаков проводится его простейшая предобработка. Удаляются наиболее часто употребляемые слова (предлоги, артикли и т.п.) и слова длиннее заданного порога, поскольку такие термы равномерно распределяются по множеству легальных писем и спама. Далее применяются непосредственно методы сокращения полученного набора признаков.

Смысл уменьшения размерности пространства признаков – оставить только те, которые наиболее информативны, наиболее значимы для задачи разделения объектов по выделенным категориям [1]. Очевидно, что:

- если один и тот же признак встречается в объектах разных категорий, он имеет невысокую информационную значимость;
- если признак часто встречается в объектах определенной категории и при этом редко среди объектов остальных категорий, то он имеет высокую информационную значимость.

Выделим наиболее распространенные методы сокращения пространства признаков: Document Frequency (DF), Term Strength (TS) Information Gain (IG), X^2-statistic (CHI) и Mutual Information (MI). Эффективность данных методов исследовалась в [2].

IG – характеристика каждого признака, вычисляется через энтропии и показывает насколько присутствие или отсутствие данного признака в документе обучающего набора влияет на принадлежность его к той или иной категории.

X^2-statistic оценивает степень зависимости данного признака и категории. Для нашей задачи с двумя категориями эта оценка оказывается симметричной.

Предлагается использовать комбинированные методы отбора признаков, как, например, в [3]. Метод сокращения признакового пространства DF используется на первом этапе, удаляя те термы, которые редко встречаются во всем тренировочном наборе писем. В [3] показано, что за счет применения предобработки DF был достигнут минимальный среди всех методов уровень ошибок первого рода (false positives), что подчеркивает эффективность этого простого метода в разрезе проблемы фильтрации спама. На втором этапе предлагается использовать метод CHI. Как

показывают эксперименты, этот метод дает практически идентичные результаты с методом IG. Оба метода обладают квадратичной вычислительной сложностью. После первого этапа количество признаков сокращается до нескольких сотен и применение таких методов оправдано.

[1] Розинкин, А. Н. Система защиты от массовых несанкционированных рассылок электронной почты на основе методов Data Mining. Диссерт. ... канд. физ.-мат. наук: 05.13.11– М., 2006, 110 с.
[2] Yang Y., Pedersen, J.O. // In Proc. of the 14th Intern. Conf. of Machine Learning ICML97. 1997. P. 412.
[3] Beiranvand A., Osareh A., Shadgar B. // J. of Academic and Appl. Studies. 2012. V. 2(3). P. 25.

БЕЗОПАСНОСТЬ И АНОНИМНОСТЬ В СЕТИ ИНТЕРНЕТ

Д.И. Корепова, В.А. Мокляков

Нижегородский госуниверситет

В современном компьютерном мире, мире Интернета, вопрос безопасности персональных данных является крайне актуальным. Мы принимаем пассивные меры для обеспечения такой безопасности – устанавливаем различное антивирусное ПО, придумываем сложные пароли, аккуратно действуем в Интернете для предотвращения кражи личной информации. Но есть и активные меры – не защита от злоумышленников, а их поиск и препятствие их дальнейшим действиям. Это оказывается совсем непросто, если киберпреступник хочет быть незамеченным и использует различные методы обеспечения анонимности в сети. Зная эти методы, можно найти следы, оставленные злоумышленником, и препятствовать его дальнейшим противозаконным действиям.

Анонимность в Интернете можно разделить на два направления: социальную – это то, что человек сам рассказывает о себе в сети, и техническую – когда утечка деанонимизирующих данных связана с техническими средствами. Сконцентрируемся именно на технической анонимности, предполагая, что злоумышленник достаточно осторожен и не может быть обнаружен простыми способами.

При незащищённом просмотре страниц в Интернете компьютер пользователя оставляет следы, такие как IP-адрес, http заголовки, cookies, отпечаток браузера, DNS запросы, трафик, а также информация о провайдере, операционной системе, стране и городе пользователя [1]. При обращении к провайдеру по IP можно определить и адрес человека, поэтому первым шагом к анонимизации является скрытие или замена своего IP-адреса.

Например, при использовании анонимного прокси-сервера [2**],** когда все запросы пользователя поступают сначала на него, а после обработки идут в сеть, извне не доступен источник запроса – виден только IP-адрес прокси-сервера. Но самому

прокси-серверу доступен адрес компьютера пользователя, значит, система не анонимна. Кроме того, протоколы прокси не поддерживают шифрование данных.

Наиболее распространённой программой, использующей прокси-серверы, является Tor (The Onion Router) [3]. Tor – это анонимная сеть виртуальных туннелей, предоставляющая передачу данных в зашифрованном виде. Устанавливая Tor, пользователь дополняет сеть своим компьютером. Анонимизация обеспечивается за счёт использования распределённой сети узлов, называемых многослойными маршрутизаторами, на каждом из которых шифруется трафик. Открытый трафик виден только выходному роутеру. Такая система проста в использовании и позволяет добиться достаточно высокого уровня анонимности при использовании http-трафика. Но Tor не может обеспечить полное закрытие передаваемых данных, порождаемый им трафик распознаётся снифферами. Факт использования сети Tor нельзя скрыть от провайдера, т.к. её адреса находятся в открытом доступе. Всё это делает систему злоумышленника уязвимой.

Задача шифрования выходящего трафика может быть решена с использованием VPN (Virtual Private Network) [4]. При работе с VPN перехватывается весь исходящий с компьютера трафик, далее он шифруется и направляется на сервер VPN, где расшифровывается и отправляется уже на серверы-адресаты в таком виде, в каком его отправил браузер. Поэтому снова данные могут быть перехвачены. Кроме того, VPN-серверу, как было и с прокси-сервером, доступен IP-адрес пользователя, а значит, использование одного VPN снова не даёт полностью анонимную систему.

Возникает вопрос, что если злоумышленник использует Tor для своей анонимности и VPN для конфиденциальности своих данных? Здесь возможны два варианта построения последовательности: Tor-VPN и VPN-Tor. Первый случай сложно реализуем и ненадёжен, т.к. VPN-сервер является постоянным выходным узлом в сеть, и весь трафик пользователя всегда идёт через него, что позволяет определить источник. Второй случай легче для реализации – необходимо подключиться к VPN серверу, затем открыть браузер Tor, который автоматически настроит нужную маршрутизацию. Здесь провайдер не узнает о факте использования Tor. Но входным узлом является VPN-сервер, которому известна информация об источнике запросов, т.е. при его компрометации возможно определить и пользователя.

Самым непростым случаем определения пользователя является тот, когда он устанавливает на своём компьютере виртуальную машину с нелицензионной ОС и правильно настраивает на ней последовательность VPN-Tor. При таком методе достижения анонимности никакие XSS-атаки, плагины, обновления системы, сбои программ и даже провайдер не смогут определить реальный IP-адрес пользователя. В этом случае утечка деанонимизирующих данных возможна только по вине злоумышленника – при использовании им своих настоящих учётных записей, почтовых ящиков, ФИО, кредитных карт, загрузке фотографий и пр.

На сегодняшний день существует множество других программ и алгоритмов, обещающих обеспечить анонимность в сети. Но во многих уже найдены уязвимости, остальные ещё недостаточно изучены, чтобы говорить об их стойкости. В любом случае надёжность (анонимность) системы зависит от ресурсов – временных и денежных, потраченных нами на её компрометацию.

[1] Федотов Н.Н. Форензика – компьютерная криминалистика. – М: Юридический мир, 2007. С. 158.
[2] INTUIT.ru: Курс «Руководство по безопасности в Lotus Notes. Лекция № 5: Прокси-серверы».
[3] Стручков Ю. Установка и настройка Tor. – Сетевая литература, 2011. Гл.2.
[4] Олифер В., Олифер Н. Компьютерные сети. Принципы, технологии, протоколы. 4-е изд. – СПб: Питер, 2013. С. 148.

ОБЕСПЕЧЕНИЕ БЕЗОПАСНОСТИ СЕТЕЙ НОВОГО ПОКОЛЕНИЯ

В.А. Мокляков, М.Д. Петрова

Нижегородский госуниверситет

Для любой сети, в том числе и беспроводной, вопросы обеспечения её безопасности должны решаться еще на этапе проектирования. Необходимо заранее продумать возможные варианты защиты создаваемой сети, сравнить их преимущества и недостатки, а после этого выбрать наиболее подходящее решение.

Представляется интересным изучить принципы работы Wi-Fi сети, методы ее защиты, а также провести их сравнение. Кроме того, актуален вопрос сопоставления технологий Wi-Fi и WiMAX.

Для передачи данных в сети Wi-Fi используются радиосигналы с двумя типами расширения спектра – FHSS и DSSS. Псевдослучайная перестройка рабочей частоты (англ. FHSS — Frequency Hopping Spread Spectrum) состоит в периодическом скачкообразном изменении несущей частоты по некоторому алгоритму, известному приёмнику и передатчику. Расширение спектра методом прямой последовательности (англ. DSSS — Direct Sequence Spread Spectrum) заключается в повышении тактовой частоты модуляции, при этом каждому символу передаваемого сообщения ставится в соответствие некоторая достаточно длинная псевдослучайная последовательность.

Для обеспечения безопасности беспроводных сетей, в том числе и построенных по технологии Wi-Fi, используются протоколы WEP, 802.1x, WPA, WPA2. Ключевая проблема WEP кроется в криптографической слабости алгоритма шифрования RC4 на статическом ключе [1]. Но такая возможность взлома и зависимость от технологий производителя были исправлены в стандарте 802.1x с помощью динамических, то есть периодически изменяющихся ключей. Стандарт WPA также более безопасен, чем WEP. Он значительно легче в использовании, особенно в дружественном к пользователю режиме Pre-Shared Key (WPA-PSK). Однако для WPA необходимо более новое оборудование. Аутентификация по паролю представляет собой уязвимость для атаки методом подбора. В отличие от WPA, в WPA2 используются более стойкий алгоритм шифрования – AES. WPA2 обеспечивает более высокий уровень безопасности по сравнению с его предшественником. Хотя WPA2 и реализует полный стандарт, но не работает на некоторых старых сетевых картах.

Таким образом, наиболее эффективным методом защиты из рассмотренных выше является WPA2. Однако и он неидеален, так как недавно была зафиксирована уязвимость, позволяющая получить доступ к передаваемой информации.

Актуальной задачей является сканирование беспроводных сетей. Для ее выполнения используются снифферы – программы, перехватывающие трафик. Они являются утилитами двойного назначения, т.е. могут быть использованы как в целях администрирования сети, так и для осуществления атак. Наиболее известными снифферами беспроводных сетей являются CommView и SpyNet для Windows, а также tcpdump и sniffit для Unix.

Говоря о сравнении Wi-Fi и WiMAX, следует отметить, что обе эти технологии являются беспроводными, но имеют различные сферы применения. Wi-Fi используется для построения небольших сетей, WiMAX – для крупных. Зона покрытия одной точки WiMAX – до 10 км, Wi-Fi – сотни метров. Качество связи лучше в WiMAX, скорость передачи данных в сети Wi-Fi может достигать 54 мб/с, WiMAX предоставляет скорость до 1Гб/с. Нельзя определенно сказать, какая из этих двух технологий лучше, и какая из них будет наиболее популярна в будущем. Объясняется такая неоднозначность именно различными вариантами использования Wi-Fi и WiMAX.

Итак, обеспечение безопасности беспроводных сетей – это сложный многоэтапный процесс, но их использование зачастую более выгодно и удобно по сравнению с проводными сетями. Сегодня информационная безопасность выходит на рынок как продукт: антивирусы, снифферы, системы контроля и предотвращения вторжений. Будущее сетей за беспроводными технологиями. Но не следует утверждать, что какая-либо одна из них станет абсолютным лидером. Они не равноценны в силу различий в идее и применении технологий. Так, Wi-Fi лучше подходит для развертывания, например, домашних или малых офисных локальных сетей; WiMAX – для улучшенного стационарного доступа в Интернет; а LTE – для дальнейшего развития современных сотовых сетей.

[1] Гейер Д. Беспроводные сети. Первый шаг / Пер. с англ. – М.: Изд. дом «Вильямс», 2005, 192 с.

ОБЕСПЕЧЕНИЕ БЕЗОПАСНОСТИ ИНФОРМАЦИОННОЙ СИСТЕМЫ СРЕДСТВАМИ СИСТЕМЫ КОНТРОЛЯ ЦЕЛОСТНОСТИ

А.А. Пальшин

Нижегородский госуниверситет

Информационная система является средой, элементами которой являются: аппаратные средства вычислительной техники, аппаратные средства телекоммуникаций (связи), программные средства, информационные базы данных и обслуживающий персонал. Основная цель информационной системы – организация обработки, хранения и передачи информации.

Одними из важнейших характеристик информационной системы являются её защищённость от несанкционированных вторжений и способность выявлять те вторжения, которые уже были осуществлены.

Системы контроля целостности – это системы, работающие по замкнутому циклу, обрабатывающие файлы, системные объекты и атрибуты системных объектов с целью получения контрольных сумм (сравнения их с контрольными суммами, полученными на предыдущем цикле) и отыскания их изменений. Когда изменение обнаружено, продукт посылает сообщение администратору безопасности, фиксируя при этом время, соответствующее времени вероятного изменения.

В основе всех систем контроля целостности лежит один принцип – сравнение эталонной выборки из исходных данных с выборкой из текущего набора данных. Методы получения выборки данных определяют эффективность системы контроля целостности. В основе множества алгоритмов проверок лежат хэш-функции [1].

Чем сложнее и надежнее алгоритм подсчета контрольной суммы, тем больше ресурсов он потребляет. Поэтому наиболее сложной задачей при внедрении систем контроля целостности является поиск компромисса, то есть безопасность должна быть, но не в убыток производительности. Кроме того, важно обеспечить целостность эталонных данных. Эта задача решается внедрением технологии сервер-агенты. Агенты устанавливаются на каждую контролируемую систему, они выполняют подсчёт контрольных сумм, результаты отправляют на сервер, который сравнивает контрольные суммы с эталонными, или полученными на предыдущей итерации и в случае несовпадения информирует администратора информационной системы.

Если имеется информационная система, в которой есть хосты под управлением ОС Windows и Linux, и необходимо внедрить систему контроля целостности, то первый шаг, который необходимо сделать, это составить список файлов и директорий в системах Windows и Linux, целостность которых нужно контролировать. В системах семейства Windows необходимо обратить особое внимание на папки: Documents and Settings, WINDOWS, а также на системный реестр. В операционных системах семейства Linux особое внимание нужно уделять системным программам, злоумышленники часто пытаются заменить эти программы их измененными копиями с такими же названиями. Вызванные в качестве дочерних процессов данные копии программы будут обладать повышенными привилегиями. Требуется также контролировать такие файлы, как: /etc/passwd, hosts.allow, hosts.deny, /etc/exports и др. На практике на каждом хосте помимо системных файлов имеются сторонние сервисы и папки, которые тоже требуют контроля, например, базы данных.

Система OSSEC [2] основана на технологии сервер-агенты, но может работать и на единственной машине. В данной системе высока степень защищенности сервера: агенты работают с минимальными привилегиями, сервер имеет лишь один открытый порт, через него происходит взаимодействие с агентами, каждое сообщение от агента анализируется, и только после этого с ним производятся дальнейшие действия.

Подсчёт контрольных сумм – это очень ресурсоемкий процесс, но система позволяет добиться высокой эффективности благодаря гибкости настроек. У администратора имеется множество возможностей, начиная от выбора степени сложности алгоритма подсчета контрольных сумм, заканчивая расписанием операции по их подсчету с точностью до секунды. Каждая система, работающая в режиме 24/7, в определенные моменты времени имеет загрузку выше средней, в эти моменты нужно отказываться от операций по подсчету контрольных сумм.

Стоит отметить, что все системы, подобные OSSEC, развиваются в направлении автоматизации выполнения определенных действий относительно того или иного события произошедшего в системе. В системе OSSEC этот функционал реализуется с помощью определенных правил, которые записываются пользователями системы в XML формате.

[1] Гостехкомиссия России. Руководящий документ: защита от несанкционированного доступа к информации. Термины и определения. – М.: Военное изд-во, 1992. С.12.
[2] Яремчук С. // Системный администратор. 2009. Вып. 10.

ОЦЕНКА УТЕЧКИ ИНФОРМАЦИИ ПО КАНАЛАМ ПЭМИН ИНТЕРФЕЙСА USB

С.М. Авдонин, Д.И. Кузнецов

Нижегородский госуниверситет

Идея получения доступа к информации за счет ПЭМИН (побочные электромагнитные излучения и наводки) берет свое начало еще со времен Первой мировой войны, когда создавались специальные подразделения для перехвата и анализа сигналов военных телефонов и радиостанций. С тех пор постоянно совершенствовались как средства перехвата, так и средства защиты. 80-е годы прошлого века характеризовались широким развитием криптографии, поэтому наиболее реальными способами доступа к конфиденциальной информации становятся те, когда информация является еще незашифрованной, в частности – перехват ПЭМИН [1].

Рассмотрим один из самых популярных интерфейсов на сегодняшний день – интерфейс USB. Разделяют 3 основные версии спецификации: USB 1.x (1,5 Мбит/с или 12 Мбит/с), USB 2.0 (480 Мбит/с) и USB 3.0 (4 Гбит/с).

Кабель USB (до 2.0 включительно) состоит из 4 медных проводников – 2 проводника питания и 2 проводника данных в витой паре – и заземленной оплётки (экрана). Схема кодирования – NRZI (логический 0 – это изменение напряжения, а логическая 1 – неизмененное напряжение).

Механизм передачи данных является асинхронным и блочным. Блок передаваемых данных называется USB-фреймом. Каждый фрейм состоит из

нескольких пакетов, которые, в свою очередь, могут быть разделены на 2 класса: служебные пакеты и пакеты данных. Нас интересует второй класс, поскольку именно пакеты данных могут содержать конфиденциальную информацию.

Рассмотрим пакет данных (тип DATA). Он состоит из пяти полей, из которых нам интересно поле Data (0 – 8192 бит), именно оно несет информацию (например, код нажатой на клавиатуре клавиши) [2].

Зная структуру пакетов, можно рассчитать теоретический спектр ПЭМИН. Длина пакета составляет 8251 бит. Считая, что в течении каждого периода тактовой частоты (12 МГц) передаётся один бит, получим тактовую частоту следования пакетов, равную 1,44 МГц. Для нас это означает, что спектр должен иметь линейчатую структуру с «главными» составляющими через 12 МГц и справа и слева от них «боковые» с шагом, кратным 1,44 МГц.

Однако здесь есть некоторые неточности, связанные с особенностями протокола USB. Поскольку в USB используется дифференциальный метод кодирования, то некий импульс образуется не менее, чем двумя последовательными битами. Отсюда следует, что «псевдотактовая» частота таких импульсов оказывается вдвое ниже истинной скорости передачи данных, предусмотренной протоколом. Но неизбежно есть зависимость этой самой «псевдотактовой» частоты от передаваемых данных. Учитывая, что в ряде случаев существует зависимость наличия/отсутствия перехода тока или потенциала от значения предыдущего бита, можно организовать передачу таких байтов, что «тактовая» будет втрое, а то и вчетверо ниже скорости передачи.

В спектральном же представлении ПЭМИН таких сигналов будут присутствовать сигналы с частотами, начинающимися именно с 6 МГц. Равно как и с частотой 12, 18, 24 и т.д. При этом информация заложена в областях «псевдотактовых» частот, а сами несущие информативными не являются.

Кроме того, не редок случай, когда (по причине не очень качественного кабеля, скверного контакта и т.д.) наблюдается постоянная смена скоростей передачи от 1,5, через 12,5 к 480 Мбит/с и обратно. Причём заранее сказать, какой процент времени будет занят какой скоростью, невозможно [3].

Для получения спектра ПЭМИН USB интерфейса использовались спектрограф Agilent E4405B с антенной «Альбатрос АГ-50» (рабочий диапазон частот 9кГц–100 МГц) и программа расчета защищенности СВТ «Легенда-05Рк». Расстояние от источника излучения до антенны – 5 см.

На спектрограмме (см. рис.) представлен полученный на практике спектр ПЭМИН USB клавиатуры. Отчетливо видны главные гармоники на частотах 6, 12, 18 МГц, упоминавшиеся выше. Все выделенные зоны спектра несут в себе информацию и представляют угрозу утечки информации.

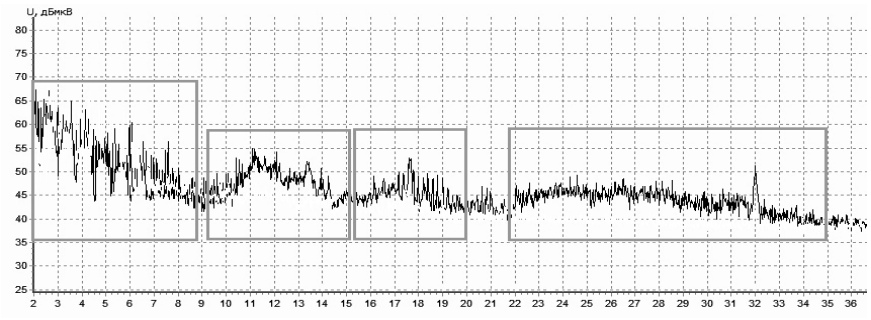

Рис.

Таким образом, спектр ПЭМИН USB имеет достаточно сложный характер. Присутствие частотных составляющих всех возможных скоростей передачи затрудняет выявление полезных сигналов в ПЭМИН USB, но оставляет угрозу утечки информации по данному каналу весьма актуальной.

[1] Чеховский С. TEMPEST – история, мифы и реальность – ЕПОС, 2002.
[2] Universal Serial Bus Specification. Rev. 1.0. 1996.
[3] Бузов Г.А., Калинин С.В., Кондратьев А.В. Защита от утечки информации по техническим каналам. – М.: Горячая Линия-Телеком, 2005, 414 с.