Министерство образования и науки Российской Федерации
Санкт-Петербургский политехнический университет Петра Великого
Институт компьютерных наук и технологий
Кафедра компьютерных систем и программных технологий

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА МАГИСТРА

Система формирования целевой рекламы на основе данных общественной Wi-Fi сети

Студент гр. 23541/3 С.В. Васильев

Санкт-Петербург
2018

Министерство образования и науки Российской Федерации
Санкт-Петербургский политехнический университет Петра Великого
Институт компьютерных наук и технологий
Кафедра компьютерных систем и программных технологий

Работа допущена к защите
зав. кафедрой

_____ В.М. Ицыксон

«____» _____ 2018 г.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА МАГИСТРА

Система формирования целевой рекламы на основе данных общественной Wi-Fi сети

по направлению 09.04.01 «Информатика и вычислительная техника»
по образовательной программе
09.04.01.15 «Технологии проектирования системного и прикладного программного обеспечения»

Выполнил студент гр. 23541/3 _____ С.В. Васильев

Научный руководитель,
к. т. н., доц. _____ Н.В. Богач

Консультант по нормоконтролю,
к. т. н., доц. _____ А.Г. Новопашенный

Санкт-Петербург
2018

РЕФЕРАТ

На 55 с., 6 рисунков, 1 приложений

НАСЛЕДОВАНИЕ, ПОЛИМОРФИЗМ, ИНКАПСУЛЯЦИЯ

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

THE ABSTRACT

55 pages, 6 pictures, 1 appendicies

INHERITANCE, POLYMORPHISM, ENCAPSULATION

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ 6

1. РОЛЬ И МЕСТО ТЕХНОЛОГИЙ В ЦЕЛЕВОЙ РЕКЛАМЕ 7

1.1. Развитие целевой рекламы 7

1.2. Технологии в интернет рекламе 8

1.3. Виды целевой рекламы по типу используемой информации 10

1.4. Юридический аспект. Проблема хранения и использования персональных данных 11

1.5. Резюме 11

2. ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ В ЗАДАЧАХ ФОРМИРОВАНИЯ ЦЕЛЕВОЙ РЕКЛАМЫ . . . 12

2.1. Типы интеллектуальных систем и их характеристики . . 12

2.2. Типовая структура экспертных систем 15

2.3. Инструментальные средства создания экспертных систем 17

2.4. Резюме 21

3. ОБЛАЧНЫЕ СИСТЕМЫ КАК ПЛАТФОРМА ДЛЯ АНАЛИЗА ДАННЫХ ЦЕЛЕВОЙ РЕКЛАМЫ . 22

3.1. Выбор типа облачно технологии для использования в целях таргетирования рекламы 22

3.2. Преимущества и недостатки размещения системы таргетированной рекламы в облачной инфраструктуре 23

3.3. Проблемы безопасности облачных SaaS технологий и способы их решения 26

3.4. Безопасность клиентских данных в каналах связи 31

4. АНАЛИЗ СЕТЕВОЙ ИНФРАСТРУКТУРЫ 35

4.1. Обзор компонентов, необходимых для системы таргетирования рекламы. 35

4.2. Обзор технологии Captive Portal для использования в целевой рекламе. 36

4.3. Возможности стандарта IEEE 802.11 для выявления клиентов в зоне действия модуля Wi-Fi. 38

4.4. Резюме 41

5. ПРОЕКТИРОВАНИЕ И РЕАЛИЗАЦИЯ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ 42
5.1. Определение программных компонентов и проектирование архитектуры системы 42
5.2. Проектирование и разработка компонента регистрации новых клиентов 42
5.2.1. Проектирование архитектуры 42
5.2.2. Разработка компонента 44
5.3. Проектирование и разработка Wi-Fi сканера 46
5.3.1. Проектирование архитектуры 47
5.3.2. Разработка компонента 48
ЗАКЛЮЧЕНИЕ 50
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ . . 51
ПРИЛОЖЕНИЕ 1. ЛИСТИНГИ 54

ВВЕДЕНИЕ

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

1. РОЛЬ И МЕСТО ТЕХНОЛОГИЙ В ЦЕЛЕВОЙ РЕКЛАМЕ

Таргетированная реклама - это рекламные текстовые сообщения, дополненные изображениями, анимацией или видео роликами, демонстрирующиеся только лицам, которые удовлетворяют критериям рекламодателя.

1.1. Развитие целевой рекламы

Первое онлайн объявление создано в 1994 году, когда интернетом пользовалось всего 30 миллионов человек. В веб версии магазина HotWired 27 октября 1994 года было впервые запущено баннерное объявление для компании AT&T [15].

Впервые интернет реклама, имеющая целевой характер, появилась в поисковой системе goto.com в 1998 году. В настоящее время компания известна под названием Overture и принадлежит Yahoo!. Идея заключалась в том, чтобы продавать ссылки, которые видит пользователь в результатах поиска. Сами ссылки подбираются исходя из контекста поисковых запросов, введенных пользователем. Идея принадлежит американскому финансисту Биллу Гроссу, который является основателем компании Idealab [25].

Такой подход в продвижении товаров и брендов был выгоден не только рекламодателям, но и создателям поисковых систем, которые могли получать от этого прибыль. Данный вид рекламы в дальнейшем был назван контекстной рекламой.

В дальнейшем идею контекстной рекламы стали реализовывать и другие компании. Например компания Google создала сервис Google AdWords, в который позже была добавлена возможность проведения торгов за рекламное место.

В настоящее время за рубежом лидерами в области контекстной рекламы являются такие компании как: Google AdWords, Yahoo! Search Marketing и Microsoft Advertising. В нашей стране широкую популярность имеет сервис Яндекс Директ.

Сильный толчок развитию таргетированной рекламы придало появление социальных сетей. Именно в них начали использовать та-

кую информацию о клиенте как возраст, пол, семейное положение, геолокацию, круг общения.

В настоящее время информация о потенциальных клиентах собирается не только на сайтах с персональными данными, но и самими браузерами или техническими устройствами, такими как мобильный телефон или умный телевизор. Например, браузеру известная информацию о посещаемых пользователем сайтах и времени, в течении которого он был запущен. Собранная информация в дальнейшем используется сторонними компаниями, которым выгодно найти клиента с определенной моделью поведения.

Для того чтобы понять как работает и как формируется целевая реклама необходимо сначала разобраться в интернет рекламы в целом. Для этого рассмотрим виды уже существующих решений.

1.2. Технологии в интернет рекламе

Рекламу в интернете разделяют на несколько видов. Критерием для разделения служит используемая технология, позволяющая тем или иным способом довести контент до клиента.

Контекстная реклама

Этот вид рекламы использует контекст просматриваемой страницы. Рекламные сообщения или баннеры подбирается исходя из ключевых слов, которым соответствует содержание посещаемого интернет ресурса. Этот вид рекламы можно увидеть на большинстве современных сайтов. Например, при посещении сайта по организации туристических походов, можно увидеть рекламу магазина, продающего туристическое снаряжение.

Маркетинг в поисковых системах

Частным случаем контекстной рекламы являются рекламные ссылки в поисковых системах. Это маркетинговая практика, при которой рекламные объявления встраиваются в результаты поисковых запросов. Рекламодатель заранее указывает ключевые слова, которые будут использованы для выявления целевой аудитории. При выдаче клиенту результатов поиска в них добавляются рекламные сообщения, которые по ключевым словам соответствуют поисковому запросу.

При таком методе распространения рекламодатель оплачивает

только ту рекламу, которой воспользовался клиент. Такой подход также называется “оплата по клику”. Рекламные сообщения могут быть как небольшими текстовыми фразами, так и более информативными блоками с описанием товара, его цены и ссылкой на обзор.

Главной особенностью поискового маркетинга является то, что он предоставляет рекламодателям возможность размещать свои объявления перед мотивированными клиентами, которые получают рекламу в тот момент, когда сами готовы совершать покупки. Никакой другой тип распространения рекламы не может обеспечить подобного. Поэтому маркетинг в поисковых системах является одним из самых эффективных видов целевой рекламы.

Таргетинг в социальных сетях

Это еще один канал интернет маркетинга. В социальных сетях пользователи сами оставляют о себе полезную для рекламодателя информацию, такую как возраст, пол, сфера деятельности, интересы. Такая информация позволяет достаточно точно подбирать целевые группы и определять где эти группы сосредоточены. Целевая аудитория, выделенная по множеству критериев, более лояльно отнесется к товару из близкой ей тематики.

Нативная реклама

Нативная реклама - это реклама завуалированная под развлекательную или обзорную статью. Такая статья имеет привычный читателю вид и своим заголовком обещает рассмотреть какой-либо вопрос. Однако на самом деле ее целью является реклама определенного товара. Как правило, такие статьи имеют шаблонные заголовки: "Как выбрать хороший X "Топ 5 X"или "Обзор популярных X". Данный тип рекламы является относительно дорогим, но при этом и достаточно эффективным.

Все выше изложенные способы доведения рекламы до клиента используют определенную информацию для выделения групп клиентов, заинтересованных в том или ином товаре. Таким образом следует рассмотреть какая информация может быть полезной для формирования клиентской базы. В дальнейшем клиентская база позволит выделять целевые объекты по их интересам.

1.3. Виды целевой рекламы по типу используемой информации

Рассмотрим виды целевой рекламы, которые разделены по типу информация, используемой для формирования целевой аудитории.

Таргетинг по интересам Таргетинг по интересам подразумевает показ рекламы, соответствующей интересам целевого объекта. Интересы клиента могут быть установлены, например, по тематике посещаемой в данные момент веб-страницы.

Временной таргетинг

Временной таргетинг позволяет рекламодателям уточнять время, в которое будут задействованы их рекламные сообщения. Чаще всего время привязано к локальному времени конечного клиента. Это позволяет использовать рекламу более эффективно, т.к. сокращает число показов в часы, когда потенциальные клиенты не пользуются интернет ресурсами [3].

Демографический таргетинг

Данный вид маркетинга использует в работе демографическую информацию, такую как пол, возраст, доход, семейное положение и другое. Эти данные позволяют значительно сократить целевую аудиторию и рекламировать товар, только тем людям, которые потенциально могут быть в нем заинтересованы. В большинстве случаев такая информация может быть собрана при заполнении регистрационной формы на сайте рекламодателя.

Географический таргетинг

Географический таргетинг позволяет формировать группы клиентов исходя из их местоположения. Критериями такой рекламы может быть страна, регион, город или почтовый индекс. Например, если компания ведет свою деятельность только на территории одного города, то для нее не имеет смысл рекламировать свои услуги в других городах. Информация для географического таргетинга может быть получена как при регистрации клиента на сайте рекламодателя, так и из других источников [6].

Поведенческий таргетинг

Данный вид формирования рекламы основан на интересах посетителя.Эти интересы устанавливаются исходя из поведения пользователя в браузере, ранее просмотренного контента, выполненного

поиска, посещенных сайтов. Например, если установлено, что пользователь ранее просматривал туристические туры, то в рекламных баннерах могут быть предложения на покупку авиа или ж/д билетов.

Геоповеденческий таргетинг

Исходными данными для геоповеденческого таргетинга является информация о перемещениях клиента и его остановках. Используя эту информацию можно точно определить привычки и пристрастия объекта. Например, если клиент часто посещает магазины со спорттоварами, то и реклама компаний, занимающихся распространением спортивной амуниции его заинтересует.

Перечисленные виды таргетинга подразумевают сбор и использование определенного вида информации о клиенте. Использование какой-то части такой информации может быть запрещено законодательством РФ. Это значит что прежде чем создавать собственную систему формирования целевой рекламы, необходимо разобраться с вопросом о законности хранения и использования видов данных, которые будут использованы в создаваемой системе.

1.4. Юридический аспект. Проблема хранения и использования персональных данных

1.5. Резюме

В данной главе рассмотрены основные методы реализации рекламы в интернете. Рассмотрены виды информации, которая позволяет осуществлять выбор целевой аудитории в задачах таргетированной рекламы. Проанализирован юридический аспект в вопросе хранения и использования клиентской информации. Дальнейшим этапом станет рассмотрение интеллектуальной системы, как системы для формирования целевой рекламы на основе клиентской базы данных.

2. ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ В ЗАДАЧАХ ФОРМИРОВАНИЯ ЦЕЛЕВОЙ РЕКЛАМЫ

Одной из задач данной работы является создание программного продукта, способного на основе клиентской базы данных генерировать события с рекламными сообщениями. Для решения такой задачи может быть использована интеллектуальная система. Это следует из определения интеллектуальных систем и решаемых ими задач.

Интеллектуальная система (ИС) – это программный комплекс, который предназначен для решения задач различных типов в зависимости от потребностей пользователя. Интеллектуальная система в своей работе использует базу знаний и логико-математические средства для получения решения.

Интеллектуальная система способна решать различные задачи. Среди них: интерпретация данных, диагностика, мониторинг, проектирование, прогнозирование, планирование, обучение, управление, поддержка принятия решений и др [5].

В условиях данной работы, интеллектуальная система должна использовать базу знаний и логико-математические средства для решения задачи управления рассылкой рекламных сообщений.

Существуют разные типы интеллектуальных систем, каждая из которых более применима в той или иной ситуации. Для того чтобы понять, какая системы подходит для использования в таргетировании рекламы, необходимо рассмотреть различные типы интеллектуальных систем. Также следует изучить вопрос устройства интеллектуальных систем. Это позволит разработать свой вариант ИС, применимый для решения поставленной задачи.

2.1. Типы интеллектуальных систем и их характеристики

Для интеллектуальных систем характерны такие свойства как развитая коммуникативная способность, умение решать плохо формализуемые задачи, способность к самообучению и адаптивности.

Коммуникативность ИС проявляется в способности взаимодействовать с конечным пользователем. Это позволяет ему формировать свои запросы для получения решения.

Решение плохо формализуемых задач определяется способностью ИС к построению оригинальных алгоритмов с возможностью использовать информацию о постоянно меняющихся данных и окружающей среде.

Самообучение заключается в возможности использовать накопленный опыт для решения новых задач.

Адаптивность позволяет системе подстраиваться под изменения модели проблемной области.

Исходя из перечисленных признаков можно разделить ИС на следующие типы:

- системы с коммуникативными способностями (с интеллектуальным интерфейсом);
- экспертные системы (системы для решения сложных задач);
- самообучающиеся системы (системы, способные к самообучению);
- системы с адаптивными способностями.

Каждый из типов в свою очередь подразделяется на подтипы в зависимости от используемой технологии [5]. Полная классификация представлена на рис.2.1.

Рис.2.1. Классификация интеллектуальных систем

Системы с интеллектуальным интерфейсом направлены на улучшение способов взаимодействие с конечным пользователем.

Интеллектуальные базы данных. Применяются для получения выборки данных, которая не хранилась БД в явном виде. В отличие от обычных БД выборка может быть сформирована на основе хранимой информации.

Естественно-языковой интерфейс. Позволяет получать данные с помощью голосового ввода команд или машинного перевода с других языков. В работе таких систем используются модули морфологического, семантического и синтаксического анализа.

Гипертекстовые системы. Используют метод поиска данных по ключевым словам. Работа такой системы происходит в два этапа. На первом из входных данных выделяются ключевые слова. На втором обрабатывается информация извлеченная по выбранным ключевым словам.

Системы контекстной помощи. При работе с такой системой пользователь описывает свой запрос, а система задает ему вопросы для конкретизации проблемы. На основе этой информации генерируется решение исходной задачи.

Системы когнитивной графики. Используют графические образы для представления моделируемых и наблюдаемых процессов. Это позволяет пользователю легче воспринимать выводимую информацию и быстрее осваивать программный интерфейс интеллектуальной системы.

Экспертные системы создаются для частичной замены специалиста-эксперта в определенной области. Они позволяют формировать советы или принимать решения, основываясь на заранее записанном и формализованном опыт эксперта. Одним из важных качеств ЭС является возможность вывести ход своих рассуждений при принятии решения.

Самообучающиеся интеллектуальные системы используют методы автоматической классификации ситуаций из реальных примеров. Примеры реализованы в обучающей выборке и для каждого из них сформированы ожидаемые критерии результата. В процессе обучения система сама генерирует правила, на основе которых способна классифицировать ту или иную исходную ситуацию.

Адаптивные информационные системы применяются в областях, где исходные данные постоянно изменяются. Адаптивные системы отвечают двум специфическим требованиям:

- способны отображать знания в каждый момент времени;
- пригодны для быстрой реконструкции при изменении проблемной среды.

Рассмотрев типы интеллектуальных систем, можно сделать вывод, что для решения задачи таргетирования рекламы подходит подтип экспертных системы, позволяющих управлять процессом генерации рекламных событий.

2.2. Типовая структура экспертных систем

Определившись с типом интеллектуальной системы, стоит рассмотреть ее типовую структуру. В дальнейшем, это позволит правильно спроектировать архитектуру программного приложения.

Экспертная система может быть двух видов: статическая и динамическая. Отличие динамической системы от статической в том,

что она учитывает изменение знаний, которые могут произойти в процессе решения задачи.

В общем случае статическая ЭС состоит из 6 основных компонентов:

- база знаний;
- база данных или рабочая память;
- решатель (интерпретатор);
- система объяснений;
- компоненты приобретения знаний;
- интерфейс взаимодействия с пользователем.

База знаний ЭС включает в себя данные, описывающие некоторую предметную область. База знаний, также,содержит правила, позволяющие производится логическое рассуждение и преобразовывать хранимые знания в решения поставленных задач.

База данных (рабочая память) содержит данные, используемые в решении текущей задачи.

Решатель (интерпретатор) осуществляет процесс получения решения с помощью последовательной обработки правил, данных из базы знаний и рабочей памяти.

Система объяснений предназначена для визуализации процесса получения решений. Данная система отображает используемые в процессе решения данные и правила. Этот компонент может значительно повысить простоту тестирования системы и степень доверия к найденному решению.

Компоненты приобретения знаний служат для добавления новых правил и знаний в ЭС.

В процессе получения решения входные данные о задаче сохраняются в рабочую память. Затем данные из рабочей памяти обрабатываются решателем с использованием правил из базы знаний. Таким образом получается решение исходной задачи. В отличие от обычной программы ЭС может в процессе работы генерировать дополнительные последовательности операций, способствующих нахождению решения.

Динамические ЭС применяются для решения отдельного, широко распространенного класса задач, в которых во время нахождения решения могут возникнуть дополнительные факты, обусловленные внешним воздействием. Такие ЭС в дополнение к описанным компонентам включают в себя подсистему моделирования внешнего мира и подсистему связи с внешним окружением.

Подсистема моделирования внешнего мира служит для оценки, анализа и прогнозирования окружающей системы. Для того чтобы система использовала актуальные данные о внешней среде, необходимо постоянно обновлять знания, хранимые в ЭС.

Компонент связи с внешним миром необходим для интеллектуальных систем управления, а также для автономных интеллектуальных систем. Зачастую это компонент реализован с использованием системы датчиков и контроллеров.

Опираясь на выше изложенный материал можно сказать, что разрабатываемый программный модуль будет иметь структуру статической экспертной системы. Однако в нем так же будет и элемент динамической ЭС. Этим элементом является компонент связи с внешним миром, который реализован в виде Wi-Fi сканер. Именно Wi-Fi сканер постоянно добавляет информацию из окружающей среды в клиентскую базу знаний.

2.3. Инструментальные средства создания экспертных систем

Инструментальные средства, используемые для разработки ИС определяют степень трудозатратности, проделываемой работы. Поэтому их рассмотрение является важной частью создания интеллектуальных систем. Инструментальные средства, используемые при разработке приложений классифицируются по следующим параметрам:

- уровень используемого языка;
- способ представления знаний;
- механизмы вывода и моделирования;
- средства приобретения знаний;

Уровень используемого языка.На трудоемкость процесса разработки ЭС сильно влияет языка программирования(ЯП). Основными параметрами для оценки ЯП служат его универсальность и мощность.

С точки зрения языка разработки, можно выделить 5 подходов к созданию ЭС:

1. Традиционные языки программирование высокого уровня, такие как C++ и Java. Использование таких средств несет за собой как плюсы так и минусы. С одной стороны это сделает разработку более трудоемкой, однако с другой - позволит реализовывать экспертную систему без каких либо ограничений.

2. Специальные языки программирования. К таким языкам относятся язык LISP, язык логического программирования PROLOG, язык рекурсивных функций РЕФАЛ и т.д. Недостатком данных языков является сложность в их использовании с другими языками, созданными для решения прикладных задач.

3. Инструментальные средства, содержащие многие, но не все компоненты ЭС. Данным программным обеспечением может воспользоваться квалифицированный разработчик, владеющий навыками программирования и умеющий совмещать и использовать различные технологии и компоненты в одной системе

4. Оболочки ЭС общего назначения. Такие системы включают в себя все программные компоненты, но не специализируются на применении в конкретной области. Обеспечение данного класса не требует от разработчика знания языков программирования.

   Использование средств этого класса могут привести к возникновению некоторых трудностей:

   - Заложенные в систему вывода механизмы могут ограничивать или вовсе не позволять генерировать решение, которое использует в своей работе эксперт. Это может привести к генерации неправильных или некачественных решений.

- Способ представления знаний может оказаться непригодным для структурирования информации из конкретной предметной области.

5. Среды, ориентированные на решение задач в конкретной области:

   - Проблемно-ориентированные средства - содержат дополнительные программные компоненты, предназначенные для решения задач определенного типа. Например, задач управления, прогнозирования или поиска.
   - Предметно ориентированные средства. В таких средствах типы предметных областей заранее известны, что значительно уменьшает время, затрачиваемое на разработку БЗ.

Для решения задачи таргетирования рекламы наиболее подходящими являются традиционные языки программирование высокого уровня. Это позволит создать свои варианты правил, задаваемых пользователем и свой интерпретатор этих правил.

Система представления знаний

Способов представления знаний достаточно много. Это вызвано стремлением качественно и целостно описать данные из различных прикладных областей. Зачастую способ представления знаний в ЭС определяется методом представления знаний. К самым распространенным моделям представления знаний относятся:

- правила (продукции);
- фреймы (объекты);
- семантические цепи;
- логические цепи.

В разрабатываемой системе пользователи сами будут добавлять правила генерации рекламных событий для клиентов. Это позволяет выбрать в качестве модели представления знаний записи в виде правил.

Механизмы вывода и моделирования

Для статической ЭС характерно то, что единственным компонентом, изменяющим информацию, является механизм вывода. В динамических ЭС на данные также влияют изменения окружающей среды, информация о которых поступает извне или эмулируется специальным компонентом. В различных системах механизмы вывода могут отличаться друг от друга вариантами реализации следующих процедур:

1. Структура процесса получения решения:
   - конструирование дерева вывода на основе обучающей выборки и дальнейший выбор для решения задачи.
   - построение сети вывода из специальных правил в процессе получения знаний и генерация решения на сети в процессе решения задачи.
   - компиляция сети вывода и поиск решения в режиме решения задачи. При этом сеть вывода генерируется исходя из данных удовлетворяющих условиям правил, применяемых к ним.
   - в процессе решения задач ЭС при отсутствии достаточного количества данных производит выборку наиболее корректных решений, приводит их обоснование, генерирует альтернативные сети вывода и осуществляет поиск решений в этих сетях.
2. Возможны три варианта направлений поиска решений: от цели к данным, от данных к цели и двунаправленный поиск.

Средства приобретения знаний. Средства приобретения знаний характеризуются следующими признаками:

1. Уровень языка приобретения знаний:
   - формальный язык;
   - ограниченный естественный язык;
   - язык пиктограмм и изображений;
   - естественно-языковой интерфейс и язык изображений.

В реализуемом модуле будет использован формальный язык, позволяющий описывать некоторые характеристики.

2. Тип приобретаемых знаний:

   - информация в табличном виде, которая содержит параметры входных и выходных данных, по которым индуктивными методами компилируется дерево вывода;
   - специализированные правила;
   - общие и специализированные правила.

   Приобретаемыми знаниями в системе являются специализированные правила, задаваемые пользователем.

3. Тип приобретаемых данных:

   - атрибуты и значения;
   - объекты;
   - классы структурированных объектов и их экземпляры, получающие значения атрибутов путем наследования.

Результатом работы программного модуля должны стать объекты, позволяющие осуществить событие рекламного характера для конкретного клиента.

2.4. Резюме

В данной главе рассмотрены виды интеллектуальных систем, их основные компоненты и модели реализации. Это позволяет получить представление о том, как необходимо конструировать ИС для целей таргетирования рекламы.

3. ОБЛАЧНЫЕ СИСТЕМЫ КАК ПЛАТФОРМА ДЛЯ АНАЛИЗА ДАННЫХ ЦЕЛЕВОЙ РЕКЛАМЫ

Целевой аудиторией для разрабатываемой системы являются малый и средний бизнес, которые не готовы тратить значительные денежные и временные ресурсы на покупку программного обеспечения и установку необходимого оборудования. С этой точки зрения, для данной аудитории выгодно использование системы как услуги. На сегодняшний день такой подход позволяют реализовывать облачные технологии, которые развиваются все стремительнее.

3.1. Выбор типа облачно технологии для использования в целях таргетирования рекламы

С появлением сверхскоростных каналов передачи данных, облачные системы получили сильный толчок в развитии. Большие скорости позволили устранить влияние больших расстояний между компьютерами на выполнение совместных задач. Такое развитие способствовало переносу большого количества программного обеспечения в облако. На сегодняшний день все чаще встречаются продукты имеющие как версию для стационарного компьютера, так и онлайн версию, созданную с использованием облачных технологий.

Облачные технологии удобны как для поставщика услуг так и для потребителя. Поставщик услуг, или провайдер, предоставляет пользователю часть ресурсов в аренду. Если число клиентов позволяет постоянно загружать рабочие ресурсы, то провайдер получает значительную прибыль от арендной платы. Это позволяет ему быстро окупить затраты на приобретение дорогостоящего оборудования и разработку программного обеспечения. С другой стороны, потребитель получает в свое распоряжение высокопроизводительные ресурсы, приобрести которые он не в состоянии.

Выделяют три основных типа облачных платформ: IasS, PaaS и SaaS [21].

1. IaaS (Infrastructure-as-a-Service) - инфраструктура как сервис. Сюда относится организация аппаратной среды с организованными серверами, настроенными средствами коммуникации, устройствами хранения информации и средствами бесперебойного питаний. На потребителя ложится задача по развертыванию и настройке требуемой системы, установке общего и специализированного программного обеспечения.

2. PaaS (Platform-as-a-Service) - платформа как сервис. По сравнению с IaaS, в PaaS технологии уже настроена операционная система для выполнения приложений. Также, могут быть уже развернуты средства разработки, отладки и поддержки. Потребитель получает возможность развернуть и сопровождать специализированное программное обеспечение такое как, например, ПО для бухгалтерского учета.

3. SaaS (Software-as-a-Service) - программное обеспечение как сервис. В данном случае поставщик предоставляет пользователю доступ к программному обеспечению, которое разработано для решения определенного типа задач. Пользователь получает доступ только к эксплуатации программы. Ответственность за администрирование лежит на поставщике. К таким системам относятся многие CRM системы такие как amoCRM [7] или VtigerCRM [24].

Разрабатываемая система таргетирования рекламы является программным продуктом, который удобнее использовать, если он уже развернут и настроен. Это значит, что в качестве технологии для организации системы удобнее всего использовать третий подход - программное обеспечение как сервис (SaaS). Чтобы определить наиболее важные компоненты такой системы необходимо рассмотреть ее достоинства и недостатки.

3.2. Преимущества и недостатки размещения системы таргетированной рекламы в облачной инфраструктуре

Рассмотрим преимущества облачных систем.

Первое преимущество помещения программного продукта в облако следует из их популярности. Облачные технологии очень удобны для конечного пользователя, что привлекает к их использованию большое количество компаний и частных лиц. Рассмотрим список наиболее важных преимуществ с точки зрения потребителя:

- Низкие материальные затраты. В отличии от обычного ПО, клиент не покупает программу в облаке, а берет ее в аренду, что значительно дешевле. Уменьшение затрат происходит и за счет устранения необходимости в производительном аппаратном обеспечении и обслуживающем его персонале.
- Временная аренда. Клиент может в любой момент отказаться от использования продукта, если он его не устраивает.
- Высокая надежность. Данные, хранимые на удаленных серверах всегда имеют резервную копию, что минимизирует вероятность их потери.
- Защита от кражи данных. Провайдер облачных вычислений всегда принимает меры по защите данных от вирусных и хакерских угроз.
- Своевременное обновление используемого ПО.
- Мобильность. Облачный сервис доступен из любой точки, в которой у клиента есть доступ к интернету. Также, большинство поставщиков услуг предоставляет возможность доступа из специально разработанного мобильного приложения.

Вторым преимуществом является возможность своевременного обновления программного обеспечения. Любая найденная ошибка или уязвимость может быть моментально устранена для всех клиентов облачной системы. Это является огромным плюсом с точки зрения безопасности, так как большинство эксплуатируемых уязвимостей программных и аппаратных систем находится именно в устаревшем ПО.

К третьему преимущество относится большое количество пользователей, являющихся в тоже время тестировщиками разработанной программы. Любая ошибка, случайным образом попавшая в

рабочую версию системы будет быстро найдена и устранена благодаря обращениям клиентов. Однако это преимущество не должно исключать этап тестирования перед обновлением программы, т.к. выявленные клиентами ошибки плохо влияют на репутацию компании.

Рассмотрим недостатки облачных систем

Главным недостатком облачной системы являются высокие требование к их надежности. Как было описано в преимуществах с точки зрения пользователя, облачная система должна гарантировать надежность хранимых данных. Для обеспечения этой надежности провайдерам необходимо заботиться о резервном копировании данных, бесперебойной работе оборудования, бесперебойном питании. Также необходимо постоянно поддерживать актуальность средств защиты от вирусных и хакерских атак.

Вторым недостатком является теоретическая возможность потери данных. Если шанс единовременной потери данных на основном и резервном сервере крайне мал, то возможность пострадать от уязвимости нулевого дня исключать не стоит [8]. Уязвимостью нулевого дня называется ошибка в программном коде или аппаратуре о которой стало публично известно и у разработчика было 0 дней на ее исправление. Потеря данных или взлом сервера является большим ударом по авторитету компании и неминуемо приведет к многочисленным затратам, в том числе и компенсации пользователям.

Третий минус относится как к недостаткам для поставщика услуги, так и для потребителя. В данной ситуации речь идет об интернет соединении. Разрыв в сети может возникнуть в любом месте между клиентом и сервером и это полностью прервет любые коммуникации между сторонами. Чтобы уменьшить риск отключения со стороны провайдера, ему необходимо позаботиться о резервной линии интернет соединения. В дополнение ко всему, поставщику услуг следует позаботиться не только о надежности соединения но и о ширине канала связи.

Рассмотрев основные недостатки облачных технологий становится понятна необходимость рассмотрения вопроса о безопасности пользовательских данных с которыми работает система. Следует отметить что под ответственность провайдера попадают как данные хранимые на сервере так и данные пересылаемые между сер-

вером и клиентским приложением (браузером).

3.3. Проблемы безопасности облачных SaaS технологий и способы их решения

Следует понимать что вопросов, касающихся безопасности много и не за все из них отвечает поставщик услуги. Некоторая ответственность возлагается и на потребителя. Основными проблемами безопасности облачных вычислений являются следующие категории [13]:

- Конфиденциальность и целостность персональных данных.
- Защита от несанкционированного доступа.
- Уязвимости программное обеспечение, используемого для поддержания работоспособности сервера.
- Общие уязвимости web сервисов.

Конфиденциальность и целостность персональных данных.

Хранимая на сервере клиентская информация является самой ценной составляющей всей облачной платформы с точки зрения пользователя. Защищенность хранимых данных и их конфиденциальность зачастую является самым важным аспектом при выборе оператора облачных вычислений. Из этого следует необходимость уделения особого внимания к вопросам безопасности клиентских данных.

К угрозам персональной информации относятся угрозы целостности хранимой информации и их конфиденциальности.

Целостность данных - это гарантия, что данные не будут повреждены или потеряны в период действия договора между провайдером и потребителем облачных услуг. На целостность данных могут повлиять как физические так и программные факторы. Например, сбой в подаче электроэнергии может привести к потере несохраненной информации. Для минимизации такой угрозы следует иметь возможность перехода на резервный источник питания. Физическая невозможность считывания информации с носителя, связанная с его повреждением или устареванием приведет к потере части

данных. Сбой в работе программного обеспечения способен навредить большому количеству клиентских данных. Потери такого характера практически невозможно восстановить имея единственный носитель информации. В связи с этим, лучшим вариантом обеспечения сохранности данных, является ее резервное копирование на дополнительных устройствах.

Другая угроза связана с конфиденциальностью информации.

Конфиденциальность информации — запрет передачи определенной информации посторонним лицам без согласия ее обладателя. Например, конфиденциальность персональных данных — обязательное для соблюдения оператором или иным получившим доступ к персональным данным лицом требование: не допускать их распространения без согласия субъекта персональных данных или наличия иного законного основания [4].

Определение конфиденциальности информации накладывает на оператора ответственность за получение данных сторонними лицами. Утечка информации хоть и мало вероятна, однако возможна. Ее последствием может стать раскрытие такой информации о клиентах как номера телефонов, ФИО, год рождения, номера банковских карт. Это важная информация, которая может быть использована для кражи денежных средств или использования данных о личности в других махинациях. Существует много зафиксированных случаев утери информации с серверов крупных компаний. Например, крупная утечка информации с серверов компании eBay произошла в 2014 году [11].

Для защиты от подобных инцидентов принято использовать шифрование хранимой информации. В том же случае с компанией eBay, украденная информация была зашифрована, что позволило избежать серьезных последствий. Существует два основных способа шифрования - классический способ и использование гомоморфного шифрования.

Классический способ подразумевает шифрование данных одним из алгоритмов симметричного шифрования, таких как AES, DES или RC4 [9]. Проблема данного способа заключается в необходимости хранения ключей шифрования. Ключи шифрования должны храниться в месте, недоступном для злоумышленника, даже если он сумел получить данные клиентов.

Гомоморфное шифрование подразумевает возможность выпол-

нения произвольных операций над зашифрованными данными, без необходимости расшифровывать их. Идея гомоморфного шифрования возникла в 1978 году а теоретическое доказательство ее возможности появилось только в 2009 году. Сейчас теория стремительно развивается и упрощается. Появляются некоторые реализации данного способа шифрования, например библиотека HElib от компании IBM [12]. Основной проблемой использования гомоморфного шифрования является его низкая производительность.

Защита от несанкционированного доступа.

Взлом учетных записи позволяет злоумышленнику получить доступ ко всему сервису. Доступ к приватной информации может быть получен как из внутренней так и из внешней сети приложения.

Обиженные плохим отношением или малой зарплатой администраторы могут серьезно навредить безопасности организации. Права администратора позволяют делать что угодно во внутренней сети компании. Это может повлечь значительные финансовые потери или потери в производительности сервиса. В свою очередь это повлияет на репутацию бренда. Чтобы избежать проблемы с человеческим фактором необходимо тщательно подбирать рабочий персонал и следить за работой сотрудников.

Следует помнить и об API облачной системы. Обычно API находится в открытом доступе и имеет хорошую документацию. Это позволяет любому разработчику в своей программе взаимодействовать с облачной средой. API, легко может быть использовано и в десктопном и в мобильном приложении так как, зачастую, оно реализовано по протоколу http или https. Доступность и понятность программного интерфейса может сыграть и отрицательную роль, так как открывает новый вектор атак для злоумышленника. Плохо продуманный с точки зрения безопасности интерфейс привнесет в систему множество уязвимостей, недоступных в веб-интерфейсе системы. API позволяет злоумышленники многократно использовать токен доступа к системе, а при получении несанкционированный доступ к учетным записям скачивать конфиденциальную информацию и осуществлять любые манипуляции от имени пользователя.

Для того чтобы риск взлома был минимален, следует подобрать качественный способ авторизации и аутентификации пользователей. Существует три основных типа аутентификации, основанных на трех специальных факторах [20].

- Что-то что вы знаете: это может быть специальной секретной информацией, такой как пароль или ответ на секретный вопрос, который не известен никому. Это фактор знания.

- Что-то что вы имеете: это предмет которым вы владеете, такой как смарт карта или другие подобные электронные устройства. Это фактор владения.

- Что-то что является вашей часть. Это физическое свойство, такое как отпечаток пальца или голос, которые позволяют вас идентифицировать. Это фактор неотъемлемости.

В веб-приложениях обычно используется первый тип аутентификации. К этому типу относятся следующие методы аутентификации, которые давно и успешно используются во многих системах [10].

- Аутентификация по паролю.

- Аутентификация по сертификатам. Как правило, используются сертификаты стандарта X.509, которые позволяют установить обе стороны общения.

- Аутентификация по одноразовым паролям. Сюда относятся программные и аппаратные токены для генерации паролей, коды получаемые пользователем по смс, заранее подготовленные списки одноразовых паролей.

- Аутентификация по ключам доступа. Этот способ часто применяется в API веб-приложений. Сервером генерируется специальный ключ, заменяющий логин и пароль, который в дальнейшем используется в приложении для взаимодействия с сервисом.

- Аутентификация по токенам. Этот способ применяется при построении распределенных систем, где один сервис доверяет функцию аутентификации другому. Такой метод можно наблюдать при аутентификации в каком либо приложении через социальную сеть.

Уязвимости программного обеспечения.

Программное обеспечение, поддерживающее работоспособность сервера также подвержено рискам быть атакованными [2]. Сюда относятся также атаки на операционные системы и используемые сетевые протоколы. Для предотвращения таких атак достаточно использовать межсетевые экраны и системы обнаружения и предотвращения вторжений. Такие подсистемы позволяют экранировать уязвимости операционной системы и приложений до момента, когда будут установлены важные обновления. Такие системы устанавливаются в виде программного агента, что позволяет экранировать уязвимости, обнаруженные в ОС и приложениях: защита от любых атак на известные уязвимости без установки критических обновлений; блокировка атак типа XSS (Cross Site Scripting) и SQL-Injection.

Общие уязвимости вуб-сервисов.

Все уязвимости веб-сервисов распространяются также и на облачные SaaS системы, так как с точки зрения внешнего пользователя облако имеет такой же интерфейс. Ежегодно, сообществом OWASP публикуется Топ-10 уязвимостей веб-приложений.

OWASP – это открытый проект обеспечения безопасности веб-приложений. Топ-10 уязвимостей от OWASP представляет собой перечень наиболее критичных рисков безопасности приложений и отражает текущие проблемы в области ИБ.

В 2017 году этот список представлен следующими пунктами [23]:

- A1 Внедрение кода
- A2 Некорректная аутентификация и управление сессией
- A3 Межсайтовый скриптинг
- A4 Небезопасные прямые ссылки на объекты
- A5 Небезопасная конфигурация
- A6 Утечка чувствительных данных
- A7 Отсутствие контроля доступа к функциональному уровню
- A8 Подделка межсайтовых запросов

- A9 Использование компонентов с известными уязвимостями

- A10 Невалидированные редиректы

Большинство из этих уязвимостей давно известны, как и способы защиты от них. Проблема их распространения заключается в том, что при разработке веб-приложений безопасности уделяется мало внимания. Это приводит к тому что современные веб ресурсы могут быть подвержены уязвимости десятилетней давности.

Рассмотренные угрозы безопасности и методы их решения относились к части облачных вычислений, которые находятся на стороне сервиса. Далее рассмотрим не менее важный аспект безопасности пользовательских данных - это защита информации в каналах связи.

3.4. Безопасность клиентских данных в каналах связи

Самый распространенный протокол прикладного уровня для передачи данных в сети интернет это http протокол. Он включает в себя большое количество полей, что делает его удобным для различных задач - от отправки коротких сообщений, до передачи файлов. Тем не менее, с точки зрения безопасности http протокол никак не защищен. Простое прослушивание трафика между клиентом и сервером раскрывает всю информацию, передаваемую в канале связи. Для того чтобы защитить частную информацию был разработан протокол SSL, позднее переросший в TLS. В данной пункте будет рассмотрен принцип работы SSL/TLS протокола.

TLS протокол работает над транспортным TCP протоколом. Это делает его удобным для использования с вышестоящими протоколами такими как HTTP. Место протокола TLS показано на рис.3.1 Таким образом, при использовании TLS, HTTP протокол никак не изменяется. Именно комбинацией TLS и HTTP и является протокол HTTPS, который работает на 443 порту.

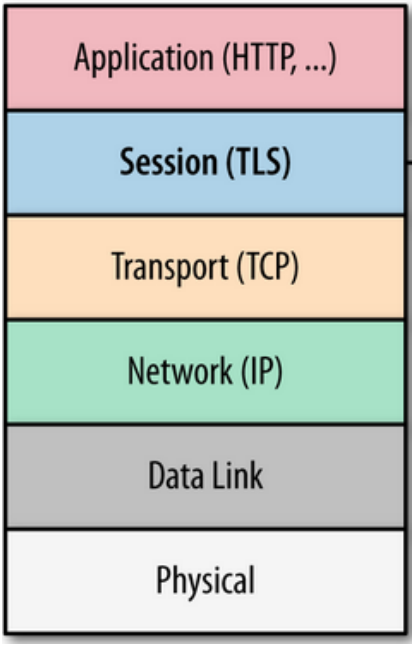

Рис.3.1. Место протокола TLS в стеке TCP/IP

Последней версией протокола на данный момент является TLS 1.3. Так как это протокол утвержден не так давно [19] - в начале 2018 года, наиболее популярной сейчас версией является TLS 1.2.

Протокол TLS обеспечивает три функции для вышестоящих протоколов:

- Шифрование - информативная часть сообщения в канале связи не может быть прочитана третьей стороной.
- Аутентификация - гарантия того, что собеседники являются именно теми за кого себя выдают
- Целостность - обнаружение подмены всего сообщения или его частей.

Взаимодействие клиента с сервером происходит в два этапа:

- установление TLS-сессии;
- передача данных по зашифрованному каналу.

Процесс установления сессии состоит из нескольких шагов:

- аутентификация;
- установление криптографических алгоритмов для обоих этапов;
- генерация общего секретного мастер-ключа;
- формирование на основе мастер-ключа общего сеансового ключа для защиты информационных сообщений.

Шифрование.

В протоколе используется как асимметричное, так и симметричное шифрование. Это связано с тем, что асимметричное шифрования более ресурсоемко по сравнению с симметричным. Проблемой же симметричного шифрования является необходимость в общем секретном ключе. Обе стороны должны знать секретный ключ для поддержания зашифрованного канала. Таким образом, комбинация двух типов шифрования позволяет исключить недостатки каждого из них. С помощью асимметричного шифрования стороны договариваются об общем секретном ключе, который в дальнейшем используют в шифрованном канале связи. В TLS используются системы асимметричного шифрования RSA, Диффи-Хеллмана и их модификации. Алгоритм шифрования Диффи-Хелмана является более современным и позволяет обеспечить прогрессивную секретность. Это значит, что при компрометации секретного ключа, записанные ранее сообщения расшифровать не удастся.

Аутентификация - Сертификаты.

Сертификаты позволяют привязывать открытый ключ к некоторому сетевому имени. Сертификат не несет в себе какую-либо секретную информацию. Самая важная информация заключенная в нем - это открытый ключ, доменное имя или ip-адрес и цифровая подпись. Открытый ключ используется при шифровании сообщений и может быть подменена третьей стороной. Для того чтобы выявить факт подмены была разработана система удостоверяющих

центров(УЦ). Для того чтобы система работала, вводится допущение о том, что пользователь доверяет центру сертификации. Центр сертификации гарантирует соответствие доменного имени и открытого ключа.

SSL сертификат состоит из двух условных частей. Первая часть эта передаваемая в открытом виде информации. Вторая часть - это цифровая подпись удостоверяющего центра. Если злоумышленник сумеет подменить сертификат, то принимающая сторона сможет выявить эту подмену. Цифровая подпись это контрольная сумма от открытой части сертификата, зашифрованная закрытым ключом центра сертификации. Для составления контрольный суммы используются криптосистемы RSA и ECDSA. Таким образом, для проверки валидности данных достаточно посчитать контрольную сумму от открытой части сертификата и сравнить ее с контрольной суммой, полученной после расшифровывания подписи открытым ключом центра сертификации.

Доменных имен в интернете очень много и поэтому центры сертификации всегда делегируют свои полномочия между другими организациями и дочерними компаниями. В свою очередь дочерние УЦ также способны распределять свои обязанность. Это праждает цепочку из удостоверяющих центров, что благодаря системе подписей не мешает корректной проверке сертификата. Пользователю необходимо доверять лишь УЦ верхнего уровня, чтобы проверить всю последовательность.

Для обеспечения большей надежности, сертификаты выдаются на определенный срок. Также, они могут быть отозваны центром сертификации при их компрометации. Для проверки статуса сертификата используется протокол OCSP [18].

Целостность

Для обнаружения подмены сообщения используется механизм HMAC (hash-based message authentication code - код аутентификации сообщения), называемый также имитовставкой. HMAC использует криптографическую хеш функцию для подсчета контрольной суммы передаваемого сообщения. Полученное значение вставляется в конец сообщения. Предполагается, что подделка HMAC вычислительно недостижима, без знания секретного ключа. Получается, что подмена сообщения, как и в случае с сертификатом, будет сразу выявлена на стороне получателя.

4. АНАЛИЗ СЕТЕВОЙ ИНФРАСТРУКТУРЫ

4.1. Обзор компонентов, необходимых для системы таргетирования рекламы.

В предыдущих главах представлен обзор одного компонента всей системы таргетирования рекламы. Этот компонент выполняет несколько функций:

- предоставляет пользователю возможность формировать правила для генерации рекламных событий;
- генерирует события исходя из правил и имеющихся данных;
- получает информацию о новых клиентах системы и регистрирует их;
- получает и сохраняет информацию о местах, посещенных клиентами;
- регистрирует новых пользователей системы.

Для того чтобы рассмотренный ранее компонент работал корректно, ему необходимы еще два дополнительных средства:

- компонент регистрации новых клиентов;
- Wi-Fi сканер.

Компонент регистрации новых клиентов.
Данный модуль является открытой Wi-Fi точкой доступа, которая установлена в помещении пользователя системы. Если организация пользователя работает в нескольких местах, то и точки доступа могут быть расположены в каждой из них. Если точка доступа установлена в помещении пользователя, то клиенты получают возможность бесплатно пользоваться интернетом. Это является хорошим стимулом для подключения к открытой сети.

Регистрация клиента происходит при первом его подключении к открытому Wi-Fi. Для получения доступа в интернет клиент должен зарегистрироваться в системе. При регистрации запрашиваются такие данные как номер телефона, имя, возраст и пол человека, а также согласие на использование предоставленной информации. Полученная информация регистрируется на сервере и в дальнейшем используется для формирования рекламных сообщений.

Если клиент уже был ранее зарегистрирован в сети, то ему сразу предоставляется доступ в интернет без необходимости проходить процесс регистрации.

Wi-Fi сканер.

Wi-Fi сканер также работает по стандарту 802.11. Однако его задача - прослушивание эфира и выявление новых устройств, которые попадают в зону его досягаемости. Определив нового посетителя он формирует информацию о его идентификаторе, времени его появления и уходе из зоны досягаемости. После выхода посетителя из зоны сканирования, рассматриваемый компонент отправляет информацию на сервер, где она заносится в базу данных и в дальнейшем будет использована при формировании рекламных событий.

Wi-Fi сканер может быть установлен в любом месте и их количество может быть неограничено. Оптимальным вариантом являются места, где пользователю интересен факт появления и время появления его клиентов.

Далее необходимо рассмотреть некоторые вопросы связанные с работой двух описанных выше компонентов.

4.2. Обзор технологии Captive Portal для использования в целевой рекламе.

Данный пункт относится к компоненту регистрации новых клиентов. Для того чтобы зарегистрировать клиента, его необходимо при первом же подключении перенаправлять на сервер регистрации. Для таких задач существует технология Captive Portal. Рассмотрим ее более детально.

Работа Captive Portal может базироваться на двух подходах:

- перенаправление за счет DNS запросов;

- перенаправление внутренними средствами маршрутизации.

При первом варианте DNS сервер настраивается таким образом, что на все запросы нового клиента ответ содержит адрес сервера регистрации. После регистрации клиента, его устройство начинает получать корректные ответы от DNS сервера. Такая системы может быть лекго обманута при статической настройке DNS сервера на устройстве клиента. Исключением является случай, когда сетевой экран настроен так, чтобы устранить возможность работы от стороннего DNS.

Второй вариант работы более надежен и подразумевает настройку маршрутизации таким образом, что все запросы по протоколу HTTP или HTTPS перенаправляются на локальный сервер. Captive Portal, организованный на платформе *nix может быть настроен с использованием утилиты iptables. В таком случае, после регистрации клиента в системе, для MAC адреса его устройства добавляется исключение, позволяющее ему использовать интернет.

Сценарий работы технологии Captive Portal существует достаточно давно, что делает его известным и простым в использовании. Данная технология позволяет доносить до клиентов дополнительную рекламу, что выгодно для любого бизнеса. Также, для Российской федерации введен закон, запрещающий наличие открытой сети без регистрации. Это означает что любая компания, предоставляя бесплатный интернет, не только может регистрировать клиентов, но и обязана это делать. Эти факторы делают технологию Captive Portal популярной и распространенной.

Популярность сервиса привела к тому, что в устройства, способные подключаться к Wi-Fi, добавляется специальная функциональность, позволяющая проще и быстрее совершать регистрацию в сети.

В ОС Android старше версии 4, мобильное устройство запрашивает с одного из серверов Google файл с названием generate_204. В ответ оно ожидает код 204 (No Content). При отсутствии в ответе этого кода создается уведомление, по клику на которое запускается браузер со страницей из ответа. Поведение ОС Android может различаться для разных производителей. В некоторых вариантах вместо уведомления может сразу быть открыта страница авторизации. Так же различен и сам браузер - это может быть как стандарт-

ный браузер системы, так и специально созданной для этих целей простейший браузер.

В операционной системе iOS устройства также запрашивают файл с одного из сайтов Apple и сверяют его содержимое с ожидаемым. В случае обнаружения различий запускается утилита Captive Network Assistant, представляющий собой браузер без поддержки HTTP cookies.

Рассмотрение сервиса Captive Portal позволяет понять, что его функциональность полностью подходит для решаемой задачи. Рассмотренная технология позволяет регистрировать каждого нового клиента в системе, при этом запрашивая у него всю информацию, которая необходима для работы системы таргетирования рекламы, и которой клиент согласен поделиться.

4.3. Возможности стандарта IEEE 802.11 для выявления клиентов в зоне действия модуля Wi-Fi.

Задачей Wi-Fi сканера является выявление новых устройств в зоне действия Wi-Fi модуля и последующая регистрация этого события. Для того чтобы корректно выявлять новые устройства необходимо понимать работу протокола 802.11 и знать какие типы сообщений в нем используются.

IEEE 802.11 - это стандарт, описывающий коммуникацию оборудования по беспроводной сети в частотных диапазонах 0,9; 2,4; 3,6; 5 и 60 ГГц[1]. Стандарт стандарт состоит из набора других стандартов ориентирующихся на конкретные вопросы. Например стандарт 802.11i описывает безопасность связи, а 802.11n вариант коммуникаций со скоростью до 600 МБит/с. Стандарт описывает работу на канальном уровне модели OSI.

Стандарт 802.11 описывает три типа фреймов

1. Фреймы Управления (Management frames).

2. Фреймы Контроля (Control frames).

3. Фреймы Данных (Data frames).

Внутри каждого фрейма имеются поля, которые определяют версию протокола тип фрейма и различные идентификаторы. Также каждый фрейм несет в себе информацию о MAC-адресе отправителя и MAC-адресе получателя.

Фреймы Управления (Management Frames)

Фреймы управления используются для установления и поддержания канала общения между коммутирующими устройствами.

Всего стандарт 802.11 определяет 14 типов фреймов управления:

1. Association request,
2. Association response,
3. Reassociation request,
4. Reassociation response,
5. Probe request,
6. Probe response,
7. Beacon,
8. ATIM (Announcement traffica indication mesage),
9. Disassiciation,
10. Authentication,
11. Deauthentication,
12. Action,
13. Action No Ack,
14. Timing advertisement

Рассмотрим некоторые фреймы.

Фрейм аутентификации (Authentication frame) отправляется устройством, желающим подключиться к точке доступа. В фрейме присутствует идентификационная информация об устройстве. Точка доступа может принять идентификацию или отвергнуть.

Фрейм ассоциации (Association request frame). Этот фрейм отправляется от устройства к точке доступа. Данным фреймом устройство просит точку доступа зарезервировать ресурсы для будущей сессии. Внутри фрейма содержится информация о характеристиках канала, таких как скорость передачи, с которыми может работать устройство. Для идентификации точки доступа, в данном фрейме отправляется SSID(имя) сети. Если точка доступа принимает запрос ассоциации, то она отвечает устройству фреймом ответа на ассоциацию, внутри которого хранится идентификатор ассоциации.

Фрейм - маячок (Beacon frame) периодически отправляется точками доступа и содержит информацию, такую как SSID точки доступа, частотный канал, временные маркеры для синхронизации времени, поддерживаемые скорости и другое. Этот фрейм позволяет всем устройствам определять досягаемые точки доступа и каналы на которых они работают.

Фрейм - проба (Probe request frame) отправляется мобильными устройствами, для установления размещенных в зоне покрытия точек доступа. Запрос может быть как широковещательным, так и с указанием конкретных SSID точек доступа. В ответном фрейме точка доступа отправляет информацию о функциональности, поддерживаемых скоростях и другое.

Подключение Wi-Fi устройства к точке доступа происходит несколько этапов. Первым этапом является этап аутентификации, затем этап ассоциации. Если точка доступа защищена паролем, то третьим этапом является проверка пароля, а при наличии шифрования и установление его характеристик.

Фреймы Контроля (Control Frames)

Фреймы контроля позволяют доставлять данные между Wi-Fi устройствами. В стандарте 802.11 определено 9 типов фреймов контроля. Однако в отличие от фреймов управления они не несут какой-либо полезной информации в условиях нашей задачи.

Фреймы Данных Wi-Fi

Главной задачей протокола является передача информации протоколов вышестоящего уровня. Для этого и предназначены 15 типов фреймов данных, описанных в стандарте 802.11. Фреймы данных используются уже при установившемся соединении, из чего следует, что для задачи выявления новых устройств они не инте-

ресны.

Исходя из вышесказанного, следует, что наиболее интересным является управляющий фрейм Probe Request. Он несет в себе информацию о MAC-адресе мобильного устройства и отправляется постоянно при поиске доступных Wi-Fi сетей. Таким образом, устройство, пронесенное мимо Wi-Fi сканера со включенным модулем Wi-Fi будет рассылать запросы на поиск сетей и попадет в список зафиксированных устройств. Это позволит определить факт посещения клиентом определенного места.

4.4. Резюме

В данной главе рассмотрены все оставшиеся компоненты системы формирования целевой рекламы. Таким образом всего в системе три компонента: компонент регистрации новых пользователей, Wi-Fi сканер и сервер с интеллектуальной системой для формирования рекламных событий. Как было установлено в главе про облачные системы, оптимальным протоколом взаимодействия между компонентами является HTTPS протокол. Дальнейшим этапом станет подбор средств разработки и сама разработка описанных компонентов.

5. ПРОЕКТИРОВАНИЕ И РЕАЛИЗАЦИЯ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

5.1. Определение программных компонентов и проектирование архитектуры системы

5.2. Проектирование и разработка компонента регистрации новых клиентов

Компонент регистрации новых клиентов предназначен для решения четырех задач:

1. Перенаправление всех пользователей заданной подсети на сервер регистрации.

2. Поддержание работы сервера регистрации.

3. Регистрация новых клиентов на удаленном сервере.

4. Предоставление интернета зарегистрированным клиентам.

5.2.1. Проектирование архитектуры

Для того чтобы рассматриваемый компонент был способен реализовывать поставленные задачи, было принято решение, что архитектура приложения будет разделена на три основных модуля. На рис.5.1 изображена архитектура всего компонента и системы, с которыми он взаимодействует.

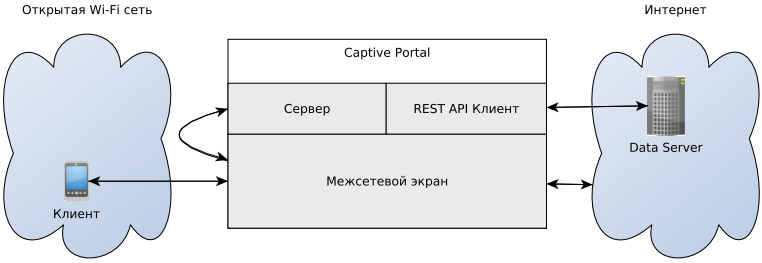

Рис.5.1. Архитектура компонента регистрации

Сервер необходим для первоначального обслуживания клиентов. Данный модуль слушает 80 и 443 порты и работает по протоколу HTTP и HTTPS соответственно. В его задачи входит:

- показ страницы регистрации;
- показ пользовательского соглашения;
- получение регистрационных данных;
- формирование события регистрации клиента;
- формирование события, разрешающего клиенту доступ в интернет.

REST API Клиент позволяет проверять MAK адрес устройства клиента в базе системы. При наличие адреса, клиенту предоставляется доступ в интернет. Данный модуль также отвечает за регистрацию новых клиентов. Передача данных между модулем и сервером системы осуществляется по протоколу HTTPS, что позволяет защитить информацию от злоумышленников.

Межсетевой экран осуществляет следующие функции:

- перенаправляет все запросы новых клиентов на сервер Captive Portal;
- блокирует доступ в интернет всем не авторизованным пользователям;

- по команде от сервера Captive Portal предоставляет клиентам доступ в интернет.

5.2.2. Разработка компонента

Для организации открытой Wi-Fi сети решено использовать утилиту hostapd [16]. Утилита позволяет создавать Wi-Fi точку доступа с различными парамерами путем конфигурации файла настроек. Используемый в компоненте файл настроек представлен в листинге 1.1.

Для корректной работы Wi-Fi сети необходимо наличие DHCP сервера. Роль DHCP сервера в рассматриваемом компоненте выполняет утилита dnsmasq. Для удобства обращение к серверу Captive Portal по установленному локальному имени (localcaptive) используется DNS сервер, который также реализован в утилите dnsmasq. Конфигурационный файл dnsmasq представлен в листинге 1.2

Предполагается, что разрабатываемый программный компонент будет устанавливаться в устройства со слабой производительностью. К таким устройствам относятся большинство домашних роутеров. Ресурсы таких устройств сильно ограничены. Они могут иметь количество памяти до 20 мб и частоту одноядерного процессора до 500 MHz. В таких условиях использование таких серверов как apache и nginx существенно замедлит работу устройства. В связи с этим, принято решение разрабатывать свой сервер на языке C++.

Для ускорения разработки и уменьшения количества ошибок было решено использовать заготовку для веб-сервера[22], написанную на языке C++ и распространяемую по лицензии MIT[17]. В своей работе данная заготовка использует библиотеку boost.asio[14], которая позволяет при необходимости вносить изменения в исходный код как сервера так и библиотеки.

На рис.5.2 представлена UML диаграмма компонента регистрации

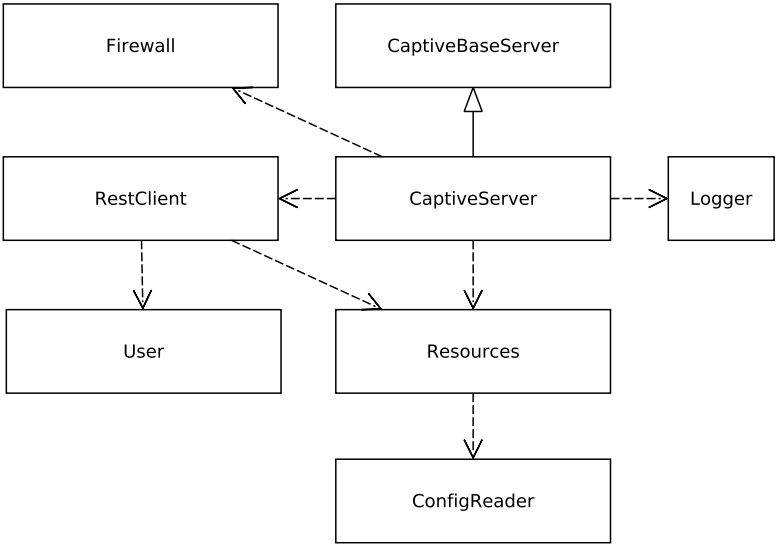

Рис.5.2. UML диаграмма компонента регистрации

Рассмотрим каждый класс и наиболее значимые методы по отдельности.

CaptiveBaseServer. CaptiveServer.

Функцией этого класса является поддержание работоспособности сервера и выдача команд межсетевому экрану и REST API клиенту.

Firewall.

Объект этого класса выполняет функции сетевого экрана, позволяющие блокировать клиентам доступ в интернет до момента получения команды из класса CaptiveServer.

RestClient.

Данный класс регистрирует на сервере новых клиентов и проверяет наличие MAC адреса в базе данных сервера таргетированной рекламы.

Метод *bool RestClient::isClientExist(std::string mac)* получает в качестве параметра MAC адрес устройства клиента в виде

AA:BB:CC:DD:EE:FF. Его задача установить зарегистрирован ли данный MAC в системе. Возвращаемое значение равно true если адрес уже существует. В противном случае возвращается false.

Метод *bool RestClient::registerClient(User user)* получает в качестве параметра объект класса User и совершает попытку его регистрации на сервере. В случае успеха возвращается true, иначе false.

User.

Отвечает за данные, полученные от клиента при регистрации. Поля данного класса отправляются на сервер для регистрации нового клиента.

Resources.

Класс, реализующий паттерн проектирования Singleton. Единственный экземпляр этого класса содержит информацию о всех ресурсах, записанных в конфигурационном файле и необходимых в других компонентах системы.

ConfigReader.

Объект этого класса позволяет считывать конфигурационный файл, сформированный по правилу «ресурс = значение».

Logger.

Еще один класс, реализующий паттерн проектирования Singleton. Позволяет вести логирование работы системы. Логируемые сообщения имеют один из четырех уровней, определенных в перечислении LogLevel: ERROR, WARNING, INFO, DEBUG.

Метод *Logger::get(LogLevel level)* возвращает единственный экземпляр класса Logger и задает уровень логирования.

Для объектов класса переопределен оператор ««», что позволяет использовать класс следующим образом:

Logger::get(LogLevel::INFO) « System started « std::endl;

Метод *std::string Logger::getTimeForFileName()* позволяет получить время, используемое в имени файла логов.

Метод *std::string Logger::getTimeForFileName()* возвращает время, вставляемое перед каждым сообщением при логировании. Пример файла с логами представлен в листинге 1.3

5.3. Проектирование и разработка Wi-Fi сканера

В список задач Wi-Fi сканера входят:

- мониторинг радиоканала на частоте 2.4 ГГц;
- определение новых устройств в радиусе досягаемости Wi-Fi модуля;
- регистрация MAC-адреса устройства на сервере таргетированной рекламы.

5.3.1. Проектирование архитектуры

Архитектура Wi-Fi сканера состоит из двух модулей: сканера Wi-Fi сети и модуля регистрации посещений. На рис.5.3 представлена архитектура Wi-Fi сканера.

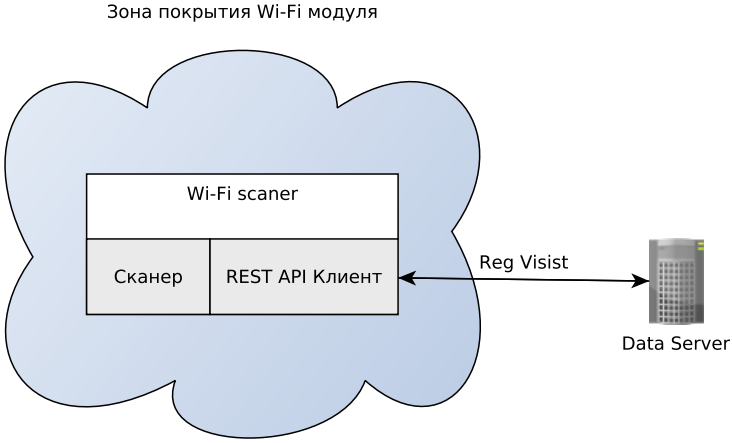

Рис.5.3. Архитектура Wi-Fi сканера

Сканер осуществляет мониторинг пакетов стандарта 802.11. При выявлении пакетов от нового клиента, MAC-адрес устройства отправителя сохраняется в память для дальнейшей обработки.

REST API клиент регистрирует новые устройства на сервере таргетированной рекламы.

5.3.2. Разработка компонента

Wi-Fi монитор, как и компонент регистрации должен работать на устройстве со слабой производительностью. Из этого следует, что рассматриваемый компонент также следует реализовать на языке C++. Для реализации функциональности монитора решено использовать библиотеку pcap[5]. Библиотека позволяет задавать необходимые настройка интернет адаптеру и просматривать все полученные им пакеты. pcap используют в своей работе такие программы как Wireshark и tcpdump. Для успешного захвата всех пакетов Wi-Fi модулем, необходимо предварительно перевести его в режим монитора. Библиотека pcap позволяет легко сделать это.

На рис.5.4 представлена UML диаграмма Wi-Fi сканера.

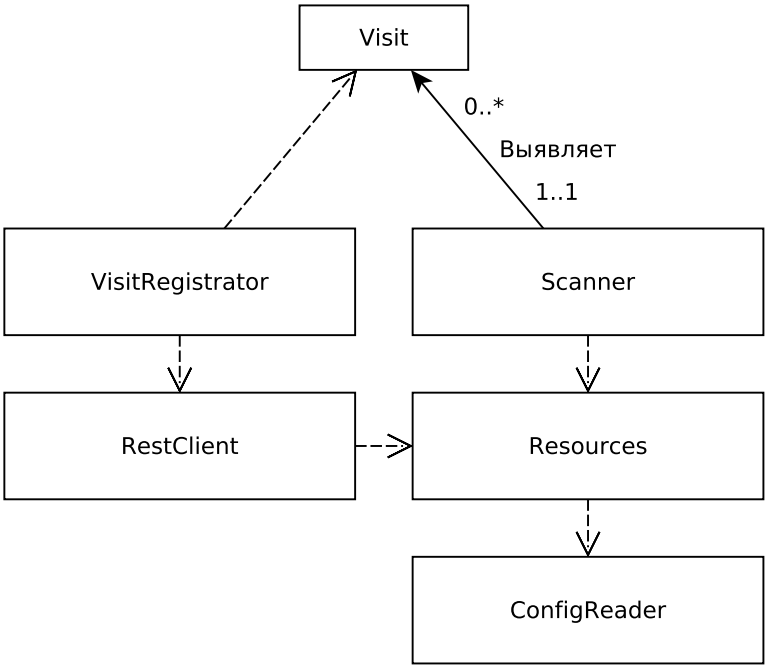

Рис.5.4. UML диаграмма Wi-Fi сканера

Рассмотрим каждый класс в отдельности

Scanner

Экземпляр данного класса устанавливает Wi-Fi модуль в режим монитора, что позволяет обрабатывать в программе все услышанные пакеты.

Метод *int Scanner::start()* запускает режим мониторинга трафика и настраивает фильтрующую функцию на прием пакетов только типа Probe Request.

Метод *void Scanner::sniffCallback(u_char arg, const struct pcap_pkthdr pkthdr, const u_char packet)* позволяет осуществлять обработку каждого отловленного пакета. При нахождении нового устройства, во внутренней памяти создается запись типа Visit с указанием MAC-адреса устройства и временем его появления. Если запись об устройстве уже есть, то обновляется время выхода устройства из зоны действия модуля Wi-Fi.

VisitRegistrator

Класс осуществляет регистрацию на сервере всех устройств, которые были записаны во внутреннюю память объектом класса Scanner.

Метод *void VisitRegistrator::start()* запускает работу созданного объекта. В методе осуществляется постоянная проверка списка объектов Visit с целью выявить устройство, покинувшее зону мониторинга. Это осуществляется путем проверки времени последнего обновления записи. Если запись обновлялась больше 10 минут назад, то такое посещение регистрируется на сервере и удаляется из памяти.

Visit

Класс является контейнером для информации о выявленном устройстве. Его полями являются: MAC-адрес устройства, время его появления, и время выхода из зоны мониторинга.

RestClient

Класс предназначен для регистрации посещений на удаленном сервере с помощью его REST API.

Метод *bool RestClient::sendVisit(Visit visit)* получает обьет типа Visit и с его помощью формирует POST запрос для отправки на сервер таргетированной рекламы.

Классы **Resources** и **ConfigReader** выполняют ту же функцию что и в компоненте регистрации клиентов.

ЗАКЛЮЧЕНИЕ

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. 802.11-2016 - IEEE Standard for Information technology–Telecommunications and information exchange between systems Local and metropolitan area networks–Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications // IEEE Std. — 2016.
2. Волков В.А. АНАЛИЗ УГРОЗ И МЕТОДОВ ЗАЩИТЫ ОБЛАЧНЫХ СЕРВИСОВ // Молодий вчений. — 2015. — № 12-1. — С. 38–43.
3. Временной таргетинг, Яндекс. — URL: `https://yandex.ru/support/direct/efficiency/timetargeting.html` (дата обращения: 18.05.2018).
4. Лопатников ЛИ. Агент экономический/Экономико-математический словарь: Словарь современной экономической науки, 5-е изд., перераб. и доп // М.: Дело. — 2003.
5. Семенов А.М Соловьев Н.А. Чернопрудова Е.Н. Цыганков А.С. Интеллектуальные системы. — ЛитРес, 2013.
6. Целевая реклама в интернете, ЗЕКСЛЕР. — URL: `http://zexler.ru/usefull/celevaya-reklama-v-internete` (дата обращения: 18.05.2018).
7. amoCRM, amoCRM. — URL: `http://www.amocrm.ru/` (дата обращения: 10.05.2018).
8. Amoroso Edward G. From the enterprise perimeter to a mobility-enabled secure cloud // IEEE Security & Privacy. — 2013. — Т. 11, № 1. — С. 23–31.
9. Data Encryption in SQL Server, Microsoft. — URL: `https://docs.microsoft.com/ru-ru/dotnet/framework/data/adonet/sql/data-encryption-in-sql-server` (дата обращения: 11.05.2018).
10. DataArt. Обзор способов и протоколов аутентификации в веб-приложениях. — 2015. — URL: `https://habr.com/company/dataart/blog/262817/` (дата обращения: 11.05.2018).
11. eBay просит пользователей поменять пароли в связи с кражей

базы данных, 3DNews. — URL: https://3dnews.ru/820608 (дата обращения: 11.05.2018).
12. HElib Documentation, HElib. — URL: https://shaih.github.io/HElib/ (дата обращения: 11.05.2018).
13. Khalil Issa M, Khreishah Abdallah, Azeem Muhammad. Cloud computing security: A survey // Computers. — 2014. — Т. 3, № 1. — С. 1–35.
14. Kohlhoff Christopher. Boost.Asio. — URL: https://www.boost.org/doc/libs/1_66_0/doc/html/boost_asio.html (дата обращения: 15.05.2018).
15. Kumar Subodha. Evolution of web advertising // Optimization Issues in Web and Mobile Advertising. — Springer, 2016. — С. 1–7.
16. Malinen Jouni. hostapd: IEEE 802.11 AP, IEEE 802.1X/WPA/WPA2/EAP/RADIUS Authenticator. — 2013. — URL: https://w1.fi/hostapd/ (дата обращения: 15.05.2018).
17. MIT License. — URL: https://github.com/eidheim/Simple-Web-Server/blob/master/LICENSE (дата обращения: 15.05.2018).
18. OCSP — протокол проверки статуса SSL сертификата, Emaro. — 2015. — URL: https://www.boost.org/doc/libs/1_66_0/doc/html/boost_asio.html (дата обращения: 12.05.2018).
19. An Overview of TLS 1.3 – Faster and More Secure, Kinsta. — URL: https://kinsta.com/blog/tls-1-3/ (дата обращения: 11.05.2018).
20. An overview of various authentication methods and protocols / Dwiti Pandya, Khushboo Ram Narayan, Sneha Thakkar и др. // International Journal of Computer Applications. — 2015. — Т. 131, № 9. — С. 25–27.
21. Sareen Pankaj. Cloud computing: types, architecture, applications, concerns, virtualization and role of it governance in cloud // International Journal of Advanced Research in Computer Science and Software Engineering. — 2013. — Т. 3, № 3.
22. Simple-Web-Server. — URL: https://github.com/eidheim/Simple-Web-Server (дата обращения: 15.05.2018).
23. Top 10-2017 Top 10, OWASP. — URL: https://www.owasp.org/index.php/Top_10-2017_Top_10 (дата обращения: 11.05.2018).
24. VtigerCRM, VtigerCRM. — URL: https://www.vtiger.com/ (дата обращения: 10.05.2018).

25. Wang Jun, Zhang Weinan, Yuan Shuai. Display advertising with real-time bidding (RTB) and behavioural targeting // arXiv preprint arXiv:1610.03013. — 2016.

ПРИЛОЖЕНИЕ 1
ЛИСТИНГИ

Листинг 1.1. Конфигурационный файл hostapd

```
interface=wlan1
bssid=66:66:66:66:66:66
driver=nl80211
ssid=CaptiveSpot
channel=6
hw_mode=g
beacon_int=100
ieee80211n=1
disassoc_low_ack=0
ap_max_inactivity=3000
auth_algs=3
logger_syslog=-1
logger_stdout=-1
logger_syslog_level=2
logger_stdout_level=2
ctrl_interface=/var/run/hostapd
ctrl_interface_group=0%
```

Листинг 1.2. Конфигурационный файл dnsmasq

```
dhcp-range=10.42.0.100,10.42.0.254,1h
dhcp-option=6,10.42.0.1 #DNS
dhcp-option=3,10.42.0.1 #Gateway
dhcp-authoritative
log-queries

# запрет считывать адреса DNS-серверов с файла resolv.conf
no-resolv

# запрещает считывать доменные имена из файла /etc/hosts
# no-hosts

# отключение отслеживание изменения файла /etc/resolv.conf или дру
    гого файла выполняющего его функцию
no-poll

address=/localcaptive/10.42.0.1
server=8.8.8.8

# Для защиты от DNS атак необходимо запретить ответы от вышестоящ
    их DNS серверов с IP адресами компьютеров локальной сети:
stop-dns-rebind


# очистка DNS-кэша при перезапуске сервиса
clear-on-reload
```

Листинг 1.3. Пример логов компонента регистрации

```
1 23:35:41 [DEBUG]: ServerBase->write_responce
2 23:35:41 [DEBUG]: Connection->set_timeout
3 23:35:41 [INFO]:  Get unknown :10.0.0.1//favicon.ico
4 23:35:41 [INFO]:  Returned file: getHttpsServer.html23:35:41 [
      DEBUG]: Responce->write 7
5 23:35:41 [DEBUG]: Responce->write 3
```